DIFFUSION TENSOR FIELD RESTORATION AND SEGMENTATION

By

ZHIZHOU WANG

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY UNIVERSITY OF FLORIDA 2004
I dedicate this work to my family.
ACKNOWLEDGMENTS

I would like to first thank my advisor, Dr. Baba C. Vemuri, for everything he has done during my doctoral study. This dissertation would not be possible without him. Dr. Vemuri introduced me to the field of medical imaging and always helps and encourages me. His insight and experience have guided me through my research and he gave numerous suggestions to this dissertation. I have been very lucky to work with him. I would also like to thank my excellent committee, Dr. Yunmei Chen, Dr. Anand Rangarajan, Dr. Murali Rao, Dr. Sartaj K. Sahni and Dr. Baba C. Vemuri, for providing valuable advice. In addition, special thanks go to Dr. Arunava Banerjee for attending my defense.

My doctoral research is a happy cooperation with many people. Dr. Vemuri has been involved with the whole progress, Dr. Chen has contributed a lot in the DTI restoration part, Dr. Thomas H. Mareci and Evren Özarslan have been my resource of DT-MRI physics and have kindly provided the data for DTI segmentation, Dr. Rachid Deriche has provided excellent comments on my work in DTI segmentation and Tim McGraw has provided beautiful fiber visualizations to make my work more appealing.

I would like to thank the editors in the Editorial office for their careful and thorough corrections. I would also like to thank the staff and the faculty members who have been very patient and helpful during my study here. In particular, Dr. David Gu and Dr. Arunava Banerjee have attended several of my seminars and showed me quite a few interesting points. I have also benefited a lot from the course taught by Dr. David Metzler and Dr. Anand Rangarajan.
Friends in my lab, my department and the University of Florida are much appreciated for sharing my joys and sadnesses. Special thanks go to Xiaobin Wu, Fangwen Chen, Fuwang Chen and Daying Zhang who have helped me to pull through the busiest time in my life. I am also fortunate to enjoy the friendship of Jundong Liu, Ju Wang, Haibin Lu, Fei Wang, Jie Zhou, Tim McGraw, Eric Spellman, Bing Jian, Neeti Vohra, Andy Shiue Le-Jeng, Lingyun Gu, Hongyu Guo and their families.

Though it is a distant memory, I have always cherished the mentorship from some of my advisors in the long past: Anan Sha, Yining Hu, Shengtao Qin, Yuan Zhu and Yuqing Gu. They guided me through different stages of study and life, I would not have been able to achieve what I have today without them.

Last, but not least, I would like to thank my parents, my sister and my wife Jing for their understanding and love during the past few years. Their support and encouragement are my source of strength. I am also thankful for my son who has inspired me a lot with his little discoveries.

This research was supported in part by the grants NIH R01-NS42075. I also have received travel grants from IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2003 and the Department of Computer and Information Science and Engineering, College of Engineering, University of Florida.

Chapter 2 is copyrighted by IEEE and was published in Proceedings of the 2004 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, co-authored with Baba C. Vemuri, Evren Özarslan, Yunmei Chen and Thomas H. Mareci.

Chapter 3 is copyrighted by IEEE and will appear in IEEE Transaction on Medical Imaging, co-authored with Baba C. Vemuri, Yunmei Chen and Thomas H. Mareci.

A preliminary version of chapter 4 is copyrighted by IEEE and will appear in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, co-authored with Baba C. Vemuri.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>x</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>xii</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1  INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 ABC of Magnetic Resonance Imaging</td>
<td>1</td>
</tr>
<tr>
<td>1.1.1 Nature of a Proton Spin with a Magnetic Field</td>
<td>1</td>
</tr>
<tr>
<td>1.1.2 Net Magnetization and Its Detection</td>
<td>2</td>
</tr>
<tr>
<td>1.1.3 Magnetic Resonance Imaging</td>
<td>3</td>
</tr>
<tr>
<td>1.2 Background of DT-MRI</td>
<td>5</td>
</tr>
<tr>
<td>1.2.1 Diffusion Process</td>
<td>5</td>
</tr>
<tr>
<td>1.2.2 Diffusion Weighted Images</td>
<td>5</td>
</tr>
<tr>
<td>1.2.3 Application of DT-MRI</td>
<td>7</td>
</tr>
<tr>
<td>1.3 Literature Review</td>
<td>9</td>
</tr>
<tr>
<td>1.3.1 Optimal $b$ Values</td>
<td>9</td>
</tr>
<tr>
<td>1.3.2 DTI Restoration</td>
<td>10</td>
</tr>
<tr>
<td>1.3.3 DTI Segmentation</td>
<td>12</td>
</tr>
<tr>
<td>1.4 Contributions</td>
<td>14</td>
</tr>
<tr>
<td>2  STATISTICAL ANALYSIS OF A NONLINEAR ESTIMATOR FOR ADC</td>
<td>17</td>
</tr>
<tr>
<td>AND ITS APPLICATION TO OPTIMIZING $b$ VALUES</td>
<td></td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>17</td>
</tr>
<tr>
<td>2.2 Theory</td>
<td>18</td>
</tr>
<tr>
<td>2.2.1 A Nonlinear Estimator for ADC</td>
<td>18</td>
</tr>
<tr>
<td>2.2.2 Variance of the Nonlinear Estimate for ADC</td>
<td>20</td>
</tr>
<tr>
<td>2.2.3 Weighted COV of the Nonlinear Estimator for ADC</td>
<td>21</td>
</tr>
<tr>
<td>2.2.4 Optimal Diffusion Weighting Factors</td>
<td>22</td>
</tr>
<tr>
<td>2.3 Results</td>
<td>22</td>
</tr>
<tr>
<td>2.3.1 Simulation Results of Different Estimators</td>
<td>22</td>
</tr>
<tr>
<td>2.3.2 Variance of Different Estimates</td>
<td>23</td>
</tr>
<tr>
<td>2.3.3 Weighted COV and Optimal Diffusion Weighting Factors</td>
<td>25</td>
</tr>
</tbody>
</table>
# A Constrained Variational Principle for Direct Estimation and Smoothing of the DTI from Complex DWIs

## 3.1 Introduction

## 3.2 Constrained Variational Principle Formulation

- **3.2.1 The Complex Nonlinear Data Constraint**
- **3.2.2 $L^p$ Smoothness Norm**
- **3.2.3 The Positive Definite Constraint**
- **3.2.4 Existence of a Solution**

## 3.3 Numerical Methods

- **3.3.1 Discretized Constrained Variational Principle**
- **3.3.2 Augmented Lagrangian Method**
- **3.3.3 Limited Memory Quasi-Newton Method**
- **3.3.4 Line Search**
- **3.3.5 Implementation Issues**

## 3.4 Experimental Results

- **3.4.1 Complex Valued Synthetic Data**
- **3.4.2 Complex DWI of a Normal Rat Brain**

# DTI Segmentation Using an Information Theoretic Tensor Dissimilarity Measure

## 4.1 Introduction

## 4.2 A New Tensor Distance and Its Properties

- **4.2.1 Definitions**
- **4.2.2 Affine Invariant Property**
- **4.2.3 Tensor Field Mean Value**

## 4.3 Active Contour Models for Image Segmentation

- **4.3.1 Snake Energy Functional**
- **4.3.2 Curve Evolution**
- **4.3.3 Level Set Formulation**

## 4.4 Piecewise Constant DTI Segmentation Model

- **4.4.1 Curve Evolution Equation**
- **4.4.2 Level Set Formulation**
- **4.4.3 Implementation and Numerical Methods**

## 4.5 Piecewise Smooth DTI Segmentation Model

- **4.5.1 Discontinuity Preserving Smoothing**
- **4.5.2 Curve Evolution Equation and Level Set Formulation**
- **4.5.3 Implementation and Numerical Methods**

## 4.6 Experimental Results

- **4.6.1 Synthetic Tensor Field Segmentation**
- **4.6.2 Single Slice DTI Segmentation**
- **4.6.3 DTI Segmentation in 3D**
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>MRI properties of different tissues.</td>
<td>4</td>
</tr>
<tr>
<td>2.1</td>
<td>Optimum $\beta$-value for typical ADC range of human brains with multiple measurements.</td>
<td>26</td>
</tr>
<tr>
<td>2.2</td>
<td>Optimum diffusion weighting factors with multiple measurements for normal rat brains.</td>
<td>26</td>
</tr>
<tr>
<td>3.1</td>
<td>Comparison of the accuracy of the estimated DEVs using different methods for different noise levels.</td>
<td>48</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Proton spin and its interaction.</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>Equilibrium magnetization of proton spins.</td>
<td>3</td>
</tr>
<tr>
<td>1.3</td>
<td>Detection of net magnetization.</td>
<td>4</td>
</tr>
<tr>
<td>1.4</td>
<td>Ellipsoid visualization of diffusion tensors.</td>
<td>5</td>
</tr>
<tr>
<td>1.5</td>
<td>A slice of DWI taken from a normal rat brain.</td>
<td>7</td>
</tr>
<tr>
<td>1.6</td>
<td>A slice of DWI taken from a normal rat brain with various diffusion weighting factors.</td>
<td>8</td>
</tr>
<tr>
<td>1.7</td>
<td>Fiber tract maps computed from diffusion tensor imaging</td>
<td>9</td>
</tr>
<tr>
<td>1.8</td>
<td>A general framework of DT-MRI analysis</td>
<td>10</td>
</tr>
<tr>
<td>2.1</td>
<td>Distribution of ADC values in the white matter and the gray matter of a normal rat brain.</td>
<td>22</td>
</tr>
<tr>
<td>2.2</td>
<td>Simulation results of different ADC estimates.</td>
<td>24</td>
</tr>
<tr>
<td>2.3</td>
<td>Variance of different ADC estimates when $\sigma = 0.05$.</td>
<td>25</td>
</tr>
<tr>
<td>2.4</td>
<td>Weighted COV of the nonlinear ADC estimate versus the diffusion weight factors.</td>
<td>27</td>
</tr>
<tr>
<td>3.1</td>
<td>A slice of several volumetric complex DWIs</td>
<td>50</td>
</tr>
<tr>
<td>3.2</td>
<td>A slice of the original (ground truth) and the estimated diffusion tensor fields</td>
<td>51</td>
</tr>
<tr>
<td>3.3</td>
<td>A slice of the computed DEV field for the noisy synthetic data</td>
<td>52</td>
</tr>
<tr>
<td>3.4</td>
<td>A slice of the normal rat brain DTI estimated using multivariate linear regression without smoothing, viewed channel by channel.</td>
<td>53</td>
</tr>
<tr>
<td>3.5</td>
<td>A slice of the normal rat brain DTI estimated using multivariate nonlinear regression without smoothing, viewed channel by channel.</td>
<td>54</td>
</tr>
<tr>
<td>3.6</td>
<td>A slice of the normal rat brain DTI estimated using our proposed method, viewed channel by channel.</td>
<td>55</td>
</tr>
</tbody>
</table>
3.7 A slice of the computed DEV field from a normal rat brain. . . . . . . 56
3.8 A slice of the normal rat brain DTIs in the region of the cerebellum viewed channel by channel. . . . . . . . . . . . . . . . . . . . . . 57
4.1 Transformation of tensor fields. . . . . . . . . . . . . . . . . . . . 62
4.2 Evolution of a curve and its corresponding higher dimension function. 71
4.3 Segmentation of a synthetic tensor field where two regions differs only in the orientations, . . . . . . . . . . . . . . . . . . . . . . . . . . . . 82
4.4 Segmentation of a synthetic tensor field where two regions differs only in scale, . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 83
4.5 Segmentation of a synthetic tensor field with two roughly homogeneous regions that only differ in the orientations. . . . . . . . . . . . . . 84
4.6 Segmentation of a synthetic tensor field with two roughly homogeneous regions that only differ in scale. . . . . . . . . . . . . . . . . 85
4.7 A slice of the DTI of a normal rat spinal cord. . . . . . . . . . . . . 86
4.8 Segmentation of the slice of DTI shown in Fig. 4.7 . . . . . . . . . . 87
4.9 A slice of the DTI of a normal rat brain. . . . . . . . . . . . . . . . 88
4.10 Segmentation of the corpus callosum from a slice of DTI shown in Fig. 4.9. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 89
4.11 Segmentation of the corpus callosum from a slice of the DTI in Fig. 4.9 using piecewise smooth model. . . . . . . . . . . . . . . . . 90
4.12 3D Segmentation of the DTI of a normal rat spinal cord. . . . . . . 90
4.13 3D Segmentation of the corpus callosum from the DTI of a normal rat brain. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 91
DIFFUSION TENSOR FIELD RESTORATION AND SEGMENTATION

By Zhizhou Wang

August 2004

Chair: Baba C. Vemuri
Major Department: Computer and Information Science and Engineering

Diffusion Tensor Magnetic Resonance Imaging (DT-MRI) is a relatively new MRI modality that can be used to study tissue microstructure, e.g., white matter connectivity in brain in vivo. It has attracted vast attention during the past few years. In this dissertation, novel solutions to two important problems in DT-MRI analysis: diffusion tensor field (aka diffusion tensor image, DTI) restoration and segmentation are presented.

For DTI restoration, we first develop a model for estimating the accuracy of a nonlinear estimator used in estimating the apparent diffusivity coefficient (ADC). We also use this model to design optimal diffusion weighting factors by accounting for the fact that the ground truth of the ADC is a distribution instead of a single value. The proposed method may be extended to study the statistical properties of DTI estimators and to design corresponding optimal acquisition parameters.

We then present a novel constrained variational principle for simultaneous smoothing and estimation of the DTI from complex valued diffusion weighted images (DWI). The constrained variational principle involves the minimization of a regularization term of $L^p$ smoothness norms, subject to a nonlinear inequality constraint on the data. The data term we employ is the original Stejskal-Tanner equation instead of
the linearized version usually employed in literature. The complex valued nonlinear form leads to a more accurate (when compared to the linearized version) estimate of the DTI. The inequality constraint requires that the nonlinear least squares data term be bounded from above by a known tolerance factor. Finally, in order to accommodate the positive definite constraint on the diffusion tensor, it is expressed in terms of Cholesky factors and estimated. The constrained variational principle is solved using the augmented Lagrangian technique in conjunction with the limited memory quasi-Newton method.

For DTI segmentation, we present a novel definition of tensor “distance” grounded in concepts from information theory and incorporate it in the segmentation of DTI. Diffusion tensor is a symmetric positive definite (SPD) tensor at each point of a DTI and can be interpreted as the covariance matrix of a local Gaussian distribution. Thus, a natural measure of dissimilarity between diffusion tensors would be the Kullback-Leibler (KL) divergence or its relative. In particular, we define a tensor “distance” as the square root of the J-divergence (symmetrized KL) between two Gaussian distributions corresponding to the tensors being compared. Our definition leads to novel closed form expressions for the “distance” itself and the mean value of a DTI. We then incorporate this new tensor “distance” in a region based active contour model for DTI segmentation and the closed expressions we derived greatly help the computation.
CHAPTER 1
INTRODUCTION

In their seminal work \cite{1}, Basser et al., introduced diffusion tensor magnetic resonance imaging (DT-MRI) as a new MRI modality from which anisotropic water diffusion can be inferred quantitatively. Since then, DT-MRI has became a powerful method to study the tissue microstructure, e.g., white matter connectivity in the brain in vivo. DT-MRI analysis consists of a variety of interesting problems: diffusion weighted image (DWI) acquisition parameter optimization, diffusion tensor field (aka diffusion tensor image, DTI) restoration, DTI segmentation, DTI registration, DTI atlas construction, fiber tracking and visualization. Since this research area is still very young, most questions raised in the above problems have not been solved yet. In this chapter, the basics of MRI and DT-MRI are reviewed, followed by a literature review of the state of art research in DT-MRI analysis.

1.1 ABC of Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is a powerful noninvasive imaging modality used mainly to “see” the soft tissues inside the human body in medical sciences. It has its beginnings in the early 1970s and is based on the physics of nuclear magnetic resonance that was thoroughly studied in the first half of the 20th century. Here only the most fundamental concepts and ideas in MRI are briefly presented; readers are referred to Haacke et al. \cite{2} for an excellent and complete coverage of this topic.

1.1.1 Nature of a Proton Spin with a Magnetic Field

$^1H$ or proton MR is the most common technique for imaging biological systems due to its high concentration in the human body and high sensitivity. A classical view of a proton is a tiny spinning positive charge. The high angular rotation of
Figure 1.1: Proton spin and its interaction. (a) Proton spin. (b) Interaction of a proton spin with an external magnetic field.

a proton causes a circular current loop $I$, which in turn brings a magnetic moment $\mathbf{\mu}$ (see Fig. 1.1(a)). As shown in Fig. 1.1(b), a spinning proton immersed in an external magnetic field $B_0$ will have a precessional motion around the field direction if $\mathbf{\mu}$ is not aligned with $B_0$. The frequency of this precession is called the Larmor frequency and is given by

$$\omega_0 = \gamma B_0$$

where the constant $\gamma$ is called the gyromagnetic ratio.

1.1.2 Net Magnetization and Its Detection

Consider an object that contains numerous protons and place it in a large static magnetic field. Eventually the precessing protons will be aligned with the external magnetic field; however, the alignments for a group of protons might be parallel or antiparallel due to thermal agitation with the absolute temperature $T$. Still, the excess of parallel alignments will bring a net equilibrium magnetization $\mathbf{M}_0$ that is proportional to the spin density $\rho_0$ everywhere (Fig. 1.2).

Even in a large body, the non-vanishing net magnetization is not significant enough to be detected when it is aligned with the applied magnetic field. However,
Figure 1.2: Equilibrium magnetization of proton spins. The black colored ones are parallel while the the gray colored ones are antiparallel to the external magnetic field.

as shown in Fig. 1.3, the magnetization vector can be tipped away by a nearby radio-frequency (rf) magnetic pulse and then precess around the static magnetic field and its frequency as given by equation (1.1). It is now possible to detect this precessing magnetization using a closely placed coil in which an electronic signal is induced and measured. There are three tissue properties that have important impact on the strength of the signal. One of them is the spin density, the other two are the longitudinal relaxation time $T_1$ and the transversal relaxation time $T_2$. Both $T_1$ and $T_2$ are tissue-specific time constants for protons and measure the rate of change in net magnetization. While $T_1$ decides the time taken for spinning protons to realign with the external magnetic field, $T_2$ measures the decay of magnetization perpendicular to the main magnetic field.

1.1.3 Magnetic Resonance Imaging

Imaging aims to relate signal with the spatial localization of various factors. This is achieved by using a spatially variant magnetic field that leads to a spatially changing distribution of resonance frequencies $\omega(x) = \gamma(B_0 + G \cdot x)$, where $G$ is the gradient of the magnetic field. Now the measured signal contains a spectrum of
Figure 1.3: Detection of net magnetization. (a) Net magnetization is tipped away by a rf pulse. (b) Processed net magnetization is detected by a nearby coil.

Table 1.1: MRI properties of different tissues

<table>
<thead>
<tr>
<th></th>
<th>white matter</th>
<th>grey matter</th>
<th>water</th>
<th>bone</th>
<th>air</th>
<th>tumor</th>
<th>infarct</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>bright</td>
<td>grey</td>
<td>dark</td>
<td>dark</td>
<td>dark</td>
<td>dark</td>
<td>dark</td>
</tr>
<tr>
<td>T2</td>
<td>bright</td>
<td>interm.</td>
<td>bright</td>
<td>dark</td>
<td>dark</td>
<td>bright</td>
<td>bright</td>
</tr>
</tbody>
</table>

received signals:

\[
S(t) = \int M(x)e^{i\gamma G \cdot x}dx
\] (1.2)

where \( M(x) \) is the magnetization that is a combination of the spin density, longitudinal relaxation, transversal relaxation and other factors like diffusion. \( M(x) \) can be computed by inverting the received signals using Fourier transformation:

\[
M(x) = \int S(k)e^{-i k \cdot x}dk
\] (1.3)

where \( k = \frac{\gamma G t}{2\pi} \) is called the reciprocal space vector to make the Fourier transformation obvious. In practice, a particular slice in a 3D volume is selected using a magnetic field with changing magnitude along a certain direction. Different kinds of contrast images can be derived from the effective spin density. Table 1.1 shows the imaging properties of various normal and abnormal tissues in the brain [3], with the abbreviation interm. referring to an intermediate brightness value between dark and bright.
1.2 Background of DT-MRI

1.2.1 Diffusion Process

Diffusion is a process of intermingling molecules as a result of random thermal agitation, and in our context, refers specifically to the random translational motion of water molecules in the part of the anatomy being imaged with MR. In three dimensional space, diffusivity can be described by a $3 \times 3$ matrix $D$ called diffusion tensor that is intimately related to the geometry and organization of the microscopic environment.

The diffusion tensor $D$ can be easily visualized as ellipsoids shown in Fig. 1.4, where the axes of each ellipsoid correspond to the eigenvector directions of the diffusion tensor and are scaled by the eigenvalues. The color of the ellipsoid ranging from blue to red shows the lowest to highest degree of anisotropy. For example, diffusion tensors in free water region are shown as large round blue ellipsoids that are just blue spheres.

1.2.2 Diffusion Weighted Images

Diffusion weighted image $S_l$ and the diffusion tensor $D$ are related through the Stejskal-Tanner equation as given by [1]

$$S_l = S_0 e^{-b \cdot D} = S_0 e^{-\sum_{i=1}^{3} \sum_{j=1}^{3} b_{ij} D_{ij}}$$  \hspace{1cm} (1.4)
where $b_l$ is the diffusion weighting matrix of the $l$-th magnetic gradient, ":" denotes the generalized inner product for matrices. Equation (1.4) is a compact form suitable for mathematical analysis, some authors prefer to use the LeBihan’s $b$-factor and the diffusion sensitizing gradient direction $g = [g_x, g_y, g_z]^T$ to reflect the imaging physics as the follows:

$$S_l = S_0 e^{-b g^T D g}$$

It is not hard to verify that

$$b_{l,11} = b g_x^2, \quad b_{l,12} = b g_x g_y, \quad b_{l,13} = b g_x g_z$$
$$b_{l,21} = b g_y g_x, \quad b_{l,22} = b g_y^2, \quad b_{l,23} = b g_y g_z$$
$$b_{l,31} = b g_z g_x, \quad b_{l,32} = b g_z g_y, \quad b_{l,33} = b g_z^2$$

The popular phase-encoding method for acquiring DT-MRI images yields complex measurements; thus, $S_l$ and $S_0$ will be complex variables and equation (1.4) still holds in such cases. In the following text, we assume $S_l = R_l + iI_l$ and $S_0 = R_0 + iI_0$ are complex variables, where $R_l = real(S_l)$, $I_l = imaginary(S_l)$, $R_0 = real(S_0)$ and $I_0 = imaginary(S_0)$. Taking the magnitude on both sides in equation (1.4) yields

$$\|S_l\| = \|S_0\| e^{-b_l : D} = \|S_0\| e^{-\sum_{i=1}^{3} \sum_{j=1}^{3} b_{l,ij} D_{ij}} \quad (1.5)$$

Taking log on both sides of equation (1.5) leads to the following linearized Stejskal-Tanner equation:

$$\log(\|S_l\|) = \log(\|S_0\|) - b_l : D = \log(\|S_0\|) - \sum_{i=1}^{3} \sum_{j=1}^{3} b_{l,ij} D_{ij} \quad (1.6)$$

Figure 1.5 shows a slice of the magnitude, real part and imaginary part of a DWI. Note that in all of published literature to date, what have been considered is only the magnitude of the complex measurements and in particular the linearized equation (1.6).
1.2.3 Application of DT-MRI

Given several non-collinear diffusion weighted intensity measurements, $\mathbf{D}$ can be estimated via multivariate regression models from equation (1.4) or its variants such as the linearized version in equation (1.6). To improve the accuracy, different $b$ values are also employed to acquire the raw DWIs. Figure 1.6 shows the complex valued DWIs in different directions and with different $b$ values, the changes in DWIs are clearly evident with different $b$ values. Though the difference in DWIs is not as obvious with changing directions, it is still visible inside the corpus callosum where the diffusion tensor is anisotropic since the value of DWI at a pixel changes with directions only when the diffusion tensor at that location is anisotropic. A general principle is that water diffuses preferably along ordered tissues e.g., the brain white matter; thus, diffusion anisotropy can then be computed to show microstructural and physiological features of tissues [4]. Especially in highly organized nerve tissue, like white matter, diffusion tensor provides a complete characterization of the restricted motion of water through the tissue that can be used to infer fiber tracts. Another possible application of the diffusion tensor field is segmentation of anisotropic regions.
Figure 1.6: A slice of DWI taken from a normal rat brain with various diffusion weighting factors. Left to right: Fixed direction, varying $b$ values. Top to bottom: fixed $b$ value, changing directions.

DT-MRI has recently became a powerful method to study the tissue microstructure e.g., white matter connectivity in the brain in vivo. To this end, Fig. 1.7 is an example depicting the fiber tract map which is traced using the diffusion tensor image inside the corpus callosum of a human brain and is visualized as colored streamtubes [5]. A general framework of DT-MRI analysis is shown in Fig. 1.8. The first step of DT-MRI analysis is the optimal acquisition of the diffusion weighted images and is of particular interest to imaging physics as well as analysts. The second step is the restoration of DTI and may involve both estimation and smoothing. Given the restored DTI, a large amount of published work in the literature has been focused on fiber tracking and visualization. However, with the rich information contained in DTI, it is also beneficial to segment some structures which might not be easy to achieve from the standard MRI images. This segmentation can also assist in focusing on a smaller region of interest (ROI) for fiber tracking, a useful task in many clinical
Figure 1.7: Fiber tract maps computed from diffusion tensor imaging visualized as colored streamtubes.

applications. Finally, it is crucial to validate the tracked fiber pathways with higher resolution images like fluro images of histology [6].

1.3 Literature Review

1.3.1 Optimal b Values

In DT-MRI, the most important control parameters in data acquisition are the diffusion weighting factors (aka. b-values) as shown in equation (1.4). By choosing these b-values carefully, one can obtain high quality diffusion weighted images without additional cost. Da Xing et al. [7] proposed a scheme to find optimal b-values for measurement of the apparent diffusion coefficient (ADC) based on the use of two diffusion weighted images acquired with $b_1$ and $b_2$. Armitage and Bastin [8] extends Xing’s result to DT-MRI. However, in [7] and [8], the reported methods use the linear equation (1.6). Though not aiming at optimizing b values, Pierpaoli and Basser [9], Basser and Pajevic [10], Anderson [11] and others have studied the effects of noise on
Figure 1.8: A general framework of DT-MRI analysis: from raw data to fiber tract map computation, visualization and segmentation.

DTI estimation. Earlier works \cite{9, 10} are based on numerical simulations while the latest work \cite{11} aims to provide a theoretical analysis. Anderson \cite{11} uses a power series of equation (1.5) to the first order to get a noise model for the linearized equation and then evaluates the perturbation of the estimated eigenvalues and eigenvectors. Just recently, Brihuega-Moreno et al. \cite{12} aimed to optimize diffusion measurements by minimizing the Cramer-Rao lower bound (CRLB), which is a function of the diffusion weighting factors. Their approach is independent of the estimation method; however, this poses a practical difficulty in choosing an estimator which can achieve the Cramer-Rao lower bound.

1.3.2 DTI Restoration

Currently there are two popular approaches, one involves smoothing the magnitude of the raw data $|S_t|$ while preserving relevant detail and then estimating the
diffusion tensor $D$ from the smoothed raw data $[13, 14]$. The raw data in this context consist of several diffusion weighted images acquired for varying magnetic field strengths and directions. Note that at least seven values at each 3D grid point in the data domain are required to estimate the six unknowns in the 3x3 symmetric tensor $D$ and one scale parameter $\|S_0\|$. The raw data smoothing or de-noising can be formulated using variational principles which in turn require solution to PDEs or at times directly using PDEs which are not necessarily arrived at from variational principles $[15, 16, 17, 18, 19, 20]$.

Another approach to restore the DTI is to smooth the dominant eigenvector (also refereed as principal diffusion direction) after the diffusion tensor has been estimated from the raw noisy measurements $\|S_l\|$. In Poupon et al. $[21]$, an energy function based on a Markovian model was used to regularize the noisy dominant eigenvector field computed directly from the noisy estimates of $D$ obtained from the measurements $\|S_l\|$ using the linearized Stejskal-Tanner equation (1.6). Coulon et al. $[22]$ proposed an iterative restoration scheme for principal diffusion direction based on direction map restoration work reported in Tang et al. $[23]$. Other sophisticated vector field restoration methods $[24, 25, 26]$ can potentially be applied to the problem of restoring the dominant eigenvector fields computed from the noisy estimates of $D$. Weickert $[27]$ used a nonlinear diffusion technique to restore a given tensor field. Their scheme was a generalization of the nonlinear anisotropic regularization of vector-valued image to the matrix-valued image case. The regularization term there involved the trace of a penalized measure applied to the sum of the component-wise structure tensors of the given matrix. Recently, Chefd’Hotel et al. $[28, 29]$ presented an elegant geometric solution to the problem of smoothing a noisy $D$ that was computed from $\|S_l\|$ using the log-linearized model (1.6) described above. They assume that the given (computed) tensor field $D$ from $\|S_l\|$ is positive definite and develop a clever approach based on differential geometry of manifolds to achieve constrained smoothing where
the smoothed tensor field is constrained to be positive semi-definite. Interesting results of mapped fibers are shown for human brain MRI.

The idea of simultaneous estimation and smoothing of the diffusion tensors from DWI was pioneered by our earlier work [30] and improved upon in [31]. This improvement involved methods to overcome the problem of manual choice of regularization control parameters. In both these works [30, 31], the estimated smooth tensor field was guaranteed to be positive semi-definite. Moreover, these works were a report of the first use of the nonlinear Stejskal-Tanner equation in the estimation of the diffusion tensors. Recently, in Tschumperlé and Deriche [32], a robust version of the linearized Stejskal-Tanner equation is used as the data term along with a robust regularization term in a unified variational principle to estimate a smooth $D$ from the noisy signal measurements. Note that the data term uses a linearized version of the Stejskal-Tanner equation as in earlier works [22, 28, 21, 14].

1.3.3 DTI Segmentation

One key factor in tensor field analysis is a proper choice of tensor distance that measures the similarity or dissimilarity between tensors and is particularly important in DTI segmentation and registration.

In general, any kind of matrix norm can be used to induce a tensor distance. One such example is the tensor Euclidean distance obtained by using the *Frobenius norm*. Due to its simplicity, tensor Euclidean distance has been used extensively in tensor field restoration [28, 27]. Alexander et al. [33] considered a number of tensor similarity measures for matching diffusion tensor images and concluded empirically that the Euclidean difference measure yields the best results. Though not many sophisticated tensor distances have been proposed in tensor field analysis, there are quite a few in the machine learning literature. Miller and Chefd’hotel [34] proposed an interesting measure on transformation groups to design an invariant kernel for non-parametric density estimation. What is most closely related to our present work
was proposed by Tsuda et al. [35]. They introduced information geometry in the space of positive definite matrices to derive a Kullback-Leibler divergence for two matrices and then used it in an “em” algorithm (not the well known EM - expectation maximization) to approximate an incomplete kernel.

In the context of DT-MRI segmentation, recently, Zhukov et al. [36] proposed a level set segmentation method which is in fact a segmentation of a scalar anisotropic measure of the diffusion tensor. The fact that a scalar field computed from the diffusion tensor field is used implies that they have ignored the direction information contained in the tensor field. Thus, this method will fail if two homogeneous regions of tensor field have the same anisotropy property but are oriented in a totally different direction!

To the best of our knowledge, before the acceptance of our work in [37], the only published work which aims to segment matrix-valued images was reported in Feddern et al. [38]. In their work, Feddern et al. extended the concept of a structure tensor to a tensor field and then used it for extending the geodesic active contours to tensor fields. The stopping criterion in this case is chosen as a decreasing function of the trace of the sum of the structure tensors formed from individual elements of the given tensor field. This amounts to taking the Frobenius norm of the tensor field formed by the gradient magnitude of the individual channels of the given tensor field. This scheme is a gradient based active contour(snake) whose performance is lacking in absence of significant gradient in the measurements. Moreover, norm used there is not invariant to affine transformations of the input coordinates on which the original tensor field is defined. Affine invariance is a very desirable property for the segmentation scheme when dealing with data sets obtained from different hardware or from subjects exhibiting anatomical changes due to development of pathology etc. in medical imaging applications.
In [37], we applied a region based model for tensor field segmentation by incorporating a tensor distance based on matrix Frobenius norm. Simultaneously, Rousson et al. [39] extended the classical surface evolution segmentation model by incorporating region statistics defined on the tensor fields for DT-MRI segmentation. In both works [37, 39], the diffusion tensor is treated as a simple matrix and every component is treated equally. However, the diffusion tensor is actually the covariance matrix of a local diffusion process and different components have different importance/weight. In [40], we were the first to use this fact in tensor field segmentation. In particular, we proposed a novel tensor “distance” based on information theory and incorporated it in region based models for tensor field segmentation. Very recently, the concepts presented therein have been extended by Lenglet et al. [41], to the case of regions with piecewise constant non-unit-variances. They also use the Fisher information metric on the manifold of local Gaussians to achieve segmentation of the diffusion tensor field.

1.4 Contributions

• Statistical Analysis of a Nonlinear Estimator for ADC and Its Application to Optimizing $b$ Values

We develop a model for estimating the accuracy of a nonlinear estimator used in computing the apparent diffusivity coefficient (ADC) which provides useful information about the structure of tissue being imaged with diffusion weighted MR. Further, we study the statistical properties of the nonlinear estimator and use them to design optimal diffusion weighting factors. Specifically, we show that a weighted linear estimator can well approximate the nonlinear estimator and thus can be used to analyze the statistical properties of the nonlinear estimator. Furthermore, to account for the fact that the ground truth of the ADC is a distribution instead of a single value, a weighted coefficient of variance (COV) is proposed as a criteria to be minimized for the determination of the
optimal diffusion weighting factors. Our model has the potential to be extended to analyze the statistical properties of a nonlinear estimator for diffusion tensor and thus obtain the optimal $b$ values for DTI acquisition.

- A Constrained Variational Principle for DTI Restoration. We present a novel constrained variational principle for simultaneous smoothing and estimation of the diffusion tensor field from complex valued diffusion weighted images (DWI). The constrained variational principle involves the minimization of a regularization term using $L^p$ smoothness norm, subject to a nonlinear inequality constraint on the data. The data term we employ is the original Stejskal-Tanner equation instead of the linearized version usually employed in literature. We empirically show that the complex valued nonlinear form leads to a more accurate (when compared to the linearized version) estimate of the DTI. The inequality constraint requires that the nonlinear least squares data term be bounded from above by a known tolerance factor which can be computed from the data. Finally, in order to accommodate the positive definite constraint on the diffusion tensor, it is expressed in terms of Cholesky factors and estimated. The constrained variational principle is solved efficiently by using the augmented Lagrangian technique in conjunction with the limited memory quasi-Newton method.

- DTI Segmentation Using an Information Theoretic Tensor Dissimilarity Measure. We aim to segment the DTI using all the information contained in the tensors defining the field, not only the scalar anisotropic properties, but also the orientation information contained therein. The key step will be a novel definition of the matrix distance which can be used to measure heterogeneity of the diffusion tensor field quantitatively. We present a novel definition of tensor “distance” grounded in concepts from information theory and incorporate it in
the segmentation of tensor-valued images. In DTI, the symmetric positive definite (SPD) diffusion tensor at each point can be interpreted as the covariance matrix of a local Gaussian distribution. Thus, a natural measure of dissimilarity between diffusion tensors would be the Kullback-Leibler (KL) divergence or its relative. In particular, we define a tensor “distance” as the square root of the J-divergence (symmetrized KL) between two Gaussian distributions corresponding to the tensors being compared. Our definition leads to novel closed form expressions for the “distance” itself and the mean value of a DTI. Unlike the traditional Frobenius norm-based tensor distance, our “distance” is affine invariant, a desirable property in many applications. We then incorporate this new tensor “distance” in a region-based active contour model for DTI segmentation, and the closed expressions we derived greatly helps the computation.

*To our knowledge, this is the first use of a region-based active contour model for diffusion tensor field segmentation.*
CHAPTER 2
STATISTICAL ANALYSIS OF A NONLINEAR ESTIMATOR FOR ADC AND ITS APPLICATION TO OPTIMIZING b VALUES

2.1 Introduction

Prior to the invention of DT-MRI, apparent diffusion coefficient (ADC) has been used intensively as a contrast mechanism in clinical MRI. The principle underlying both ADC and DT-MRI is the sensitivity of the magnetic resonance (MR) signal to the random motion of water molecules. In the case of ADC, the diffusion weighted image (DWI) is also given by the Stejskal-Tanner equation:

\[ S = S_0 e^{-bD} \]  

(2.1)

where \( D \) is the ADC, and \( S_0 \) is the non-DW image.

DWIs are often corrupted by noise, the noise model for complex valued data is [42]

\[ S = S_0 e^{-bD} + n_r + in_i \]  

(2.2)

where \( n_r \) and \( n_i \) are uncorrelated Gaussian noise with zero mean and standard deviation \( \sigma \). Given a set of DWIs acquired with different \( b \)-values, ADC can be estimated via linear or nonlinear regression.

Linear estimators of the ADC and the diffusion tensor have been well studied and also used in designing optimal \( b \) values. However, analysis regarding the nonlinear estimation has not been considered before. In this chapter, statistical analysis of a nonlinear estimator for ADC and its application to optimizing \( b \) values are carried out. The method used here also sheds some lights on how to analysis the nonlinear estimator for diffusion tensor and design a corresponding optimal acquisition method.
2.2 Theory

2.2.1 A Nonlinear Estimator for ADC

Let \( S_0 = R_0 + iI_0 \) be the echo intensity without magnetic gradient, \( D \) be the ADC, \( S_k = R_k + iI_k \) be the measurement for the \( k \)th magnetic gradient, \( k = 1, \ldots, K \). Since \( S_k = S_0 e^{-b_kD} + n_r + in_i \) as in equation (2.2), the sum of squared errors (SSE) that measures the goodness of estimation is

\[
E(S_0, D) = \sum_k \left[ (R_k - R_0 e^{-b_kD})^2 + (I_k - I_0 e^{-b_kD})^2 \right] \tag{2.3}
\]

Then we have a nonlinear estimate \( (S_0^*, D^*) \) that is the solution of a nonlinear optimization problem, i.e.

\[
(S_0^*, D^*) = \min_{S_0, D} E(S_0, D) \tag{2.4}
\]

Though there is no analytical form for the nonlinear estimate \( (S_0^*, D^*) \), it can be efficiently solved numerically. Levenberg-Marquardt (LM) method [43] is ideal for nonlinear least squares optimization problems like equation (2.4). Let

\[
F(S_0, D) = \begin{bmatrix}
R_1 - R_0 e^{-b_1D} \\
\vdots \\
R_K - R_0 e^{-b_KD} \\
I_1 - I_0 e^{-b_1D} \\
\vdots \\
I_K - I_0 e^{-b_KD}
\end{bmatrix}
\]

Then

\[
E(S_0, D) = F(S_0, D)^T F(S_0, D)
\]
The Jacobian matrix $\mathbf{J}(S_0, D)$ of $\mathbf{F}(S_0, D)$ is

$$
\mathbf{J}(S_0, D) = 
\begin{bmatrix}
-\epsilon^{-b_1D} & 0 & b_1R_0e^{-b_1D} \\
\vdots & \vdots & \vdots \\
-\epsilon^{-b_KD} & 0 & b_KR_0e^{-b_KD} \\
0 & -\epsilon^{-b_1D} & b_1I_0e^{-b_1D} \\
\vdots & \vdots & \vdots \\
0 & -\epsilon^{-b_KD} & b_KI_0e^{-b_KD}
\end{bmatrix}
$$

Then the gradient $\mathbf{G}(S_0, D)$ of $E(S_0, D)$ is given by

$$
\mathbf{G}(S_0, D) = 2\mathbf{J}(S_0, D)^T\mathbf{F}(S_0, D)
$$

If we denote the Hessian matrix of the $k_{th}$ component of $\mathbf{F}(S_0, D)$ as $\mathbf{H}_k(S_0, D)$, then the Hessian matrix $\mathbf{H}(S_0, D)$ of $E(S_0, D)$ is

$$
\mathbf{H}(S_0, D) = 2\mathbf{J}(S_0, D)^T\mathbf{J}(S_0, D) + 2\sum_{k=1}^{K} F_k(S_0, D)\mathbf{H}_k(S_0, D)
$$

The essence of the LM method is to compute a search direction at iteration $n$ according to

$$
(J_n^TJ_n + \lambda_n \mathbf{I})\mathbf{d}_n = -J_n \mathbf{F}_n \quad (2.5)
$$

where $\lambda_k \geq 0$ is a control parameter. When $\lambda_k$ is zero, the search direction is computed as the Gauss-Newton method. As $\lambda_k$ goes to infinity, the search direction approaches the steepest descent direction. With proper choices of $\lambda_k$, the LM method can perform like the gradient descent method when the current solution is far from the final solution and it can approximate the Gauss-Newton method when the current solution is close to the final solution.
2.2.2 Variance of the Nonlinear Estimate for ADC

The nonlinear estimate given in (2.4) can only be computed numerically, thus it is impossible to analytically compute its statistical properties. We propose to use a weighted linear estimator to approximate the nonlinear estimator. Since the weighted linear estimator has a close form solution, we can analytically approximate the statistical behavior of the nonlinear estimator.

The sum of weighted squared errors of the linearized form of the model in (2.1) is

\[ E_w(||S_0||, D) = \sum_k w_k^2 (\log(||S_k||) - \log(||S_0||) + b_k D)^2 \]  

(2.6)

where \( w_k = ||S_0 e^{b_k D_g}||, \ k = 1, ..., K \) are the weights, \( D_g \) is the ground truth. A corresponding weighted linear estimate is given by

\[ (||S_0||^*, D^*) = \min_{||S_0||, D} E_w(||S_0||, D) \]  

(2.7)

The solution \((||S_0||^*, D^*)\) can be easily computed as

\[
\begin{bmatrix}
\ln(||S_0||^*) \\
D^*
\end{bmatrix} = \alpha A \begin{bmatrix}
\sum_k (w_k^2 b_k \ln(||S_k||)) \\
\sum_k w_k^2 \ln(||S_k||)
\end{bmatrix}
\]  

(2.8)

where

\[
\alpha = \frac{1}{-(\sum_k |w_k|^2 b_k)^2 + (\sum_k |w_k|^2)(\sum_k w_k^2 b_k^2)}
\]

\[
A = \begin{bmatrix}
-\sum_k w_k^2 b_k & \sum_k w_k^2 b_k^2 \\
-\sum_k w_k^2 & \sum_k w_k^2 b_k
\end{bmatrix}
\]
It is pointed out in [42] that \( \|S_k\| \sim \|S_0\|e^{-b_k D} + n \) when \( SNR > 3 \), then by using Taylor expansion on \( ln(\|S_k\|) \), we have

\[
ln(\|S_k\|) \sim ln(\|S_0\|e^{-b_k D} + n) \sim ln(\|S_0\|e^{-b_k D}) + \frac{n}{\|S_0\|e^{-b_k D}}
\]  

(2.9)

thus \( \sigma_{ln(\|S_k\|)} \sim \frac{\sigma}{\|S_0\|e^{-b_k D}} \) and we have

\[
var(D^*) = \frac{\alpha^2 \sigma^2}{\|S_0\|^2} \sum_k \left\{ w_k^2 e^{b_k D} \left[ \sum_j (b_j - b_k) w_j^2 \right] \right\}^2
\]  

(2.10)

It is easy to prove that equation (2.10) is the same as the CRLB given in [12] for two \( b \)-values and we found empirically that this is true even for multiple \( b \)-values more than two. However, whether equation (2.10) and the CRLB given in [12] are actually the same remains to be explored. Note that this weighted linear estimator is NOT a realistic estimator since the weighting are related to the ground truth, however, it serves well as a tool to analyze the nonlinear estimator which is validated by simulation experiments later.

### 2.2.3 Weighted COV of the Nonlinear Estimator for ADC

Equation (2.10) shows the variance of the weighted linear estimator when the ground truth is a single ADC value. However, the ground truth \( D_g \) is usually a distribution instead of a single value as shown in Fig. 2.1. Thus we propose to use the following weighted COV of \( D^* \).

\[
COV_w(D^*) = \int COV(D^*|D_g)p(D_g)dD_g
\]  

(2.11)

where \( COV(D^*|D_g) = \sqrt{Var(D^*|D_g)}/D_g \), \( p(D_g) \) is the probability density function of the ground truth.
2.2.4 Optimal Diffusion Weighting Factors

The weighted COV of $D^*$ in equation (2.11) depends on the ground truth distribution, total number of measurements and the choice of diffusion weighting factors. The ground truth distribution cannot be controlled. Though increasing total number of measurements will decrease the weighted COV, this will also increase the acquisition time which is not desirable. When the total number of measurements is fixed, what one can do is to choose the diffusion weighting factors carefully to get a minimum weighted COV.

2.3 Results

2.3.1 Simulation Results of Different Estimators

The simulation procedure that we developed consists of the following steps:

1. Choose a ground truth $D_g$ and $S_0$.
2. Choose a set of diffusion weighting factors $b_1, ..., b_N$.
3. Generate $n$ complex measurements using (2.2) with the selected set of $b$-values.
4. Estimate $D^*$ using the nonlinear estimator in (2.4), the weighted linear estimator in (2.7) and the linear estimator in Xing et al. [7] respectively.

The last two steps are executed for several trials. The following settings were used in our simulation here:

\[
D_g = 0.001 \text{mm}^2/\text{s}, \quad S_0 = 10 + 10i, \quad b_1 = 100 \text{s/mm}^2, \quad b_2 = 500 \text{s/mm}^2 \quad \text{and} \quad b_3 = 1000 \text{s/mm}^2, \quad \sigma = 0.5 \quad \text{or} \quad \sigma = 1.0.
\]

The results of the simulation are shown in Fig. 2.2. It is evident that the nonlinear estimator yields the best results, the weighted linear least square estimator has similar outcomes like the nonlinear estimator while the linear least square model does not perform as good as the other two. This indicates that the nonlinear estimator should be used for better accuracy.

2.3.2 Variance of Different Estimates

The variance of the nonlinear estimate can only be estimated by simulation. This simulation will be similar as described earlier. However, the total number of trial is set to be 2000 for a stable and accurate estimation of the variance. The variance of the linear (LSQR) estimate is given in Bito et al. [44] and the weighted linear estimate (weighted LSQR) can be computed from equation (2.10). The settings used here are:

\[
D_g = 0.001 \text{mm}^2/\text{s}, \quad S_0 = 10 + 10i, \quad b_1 = 0, \quad b_3 = 2000 \text{s/mm}^2, \quad b_2 \text{ will vary from 0 to 3000 s/mm}^2 \quad \text{and} \quad \sigma = 0.05.
\]

The results of the computation are shown in Fig. 2.3. It is clear that the variance of the weighted linear estimate is a good approximation of the nonlinear estimate. Also we see that in most cases, the nonlinear estimate yields a smaller variance than the linear estimate. However, they happen to have the same variance when $b_2 = b_1$ or $b_2 = b_3$. 
Figure 2.2: Simulation results of different ADC estimates. Top: Simulation results when $\sigma = 0.5$. Bottom: Simulation results when $\sigma = 1.0$. 
2.3.3 Weighted COV and Optimal Diffusion Weighting Factors

It is easy to compute the weighted COV in equation (2.11) by numerical integration given the ground truth ADC and the diffusion weighting factors and it is also possible to find optimal diffusion weighting factors numerically. Currently we use the medium-scale sequential quadratic programming (SQP) algorithm in Matlab as the numerical optimization procedure to achieve this.

For our first experiment, we use the typical ADC range found in the human brain as given in Brihuega-Moreno et al. [12] and assume the distribution to be uniform. In this case, we can use the dimensionless $\beta$ value which is defined as $\min(D_g) \times b$. The optimal $\beta$-values with multiple measurements up to five are summarized in table 2.1. Notice that our result is quite different from [12] and these differences are caused by the different precision in the numerical integration techniques used in the step to compute the weighted COV. Our integration has higher accuracy than the one
Table 2.1: Optimum $\beta$-value for typical ADC range of human brains with multiple measurements. $m$ is the total number of measurements, the number of measurements at a particular $\beta$-value are shown in the bracket in the event that it is more than one and the unit of the $D$ values is $mm^2/s$.

\[
\begin{array}{|c|c|c|c|}
\hline
D \text{ range is} & \beta_1 & \beta_2 & \beta_3 \\
[0.3 \times 10^{-3}, 3 \times 10^{-3}] & & & \\
\hline
m & \beta_1 & \beta_2 & \beta_3 \\
2 & 0 & 0.210 & - \\
3 & 0 & 0.223(2) & - \\
4 & 0 & 0.190(2) & 0.380 \\
5 & 0 & 0.201(3) & 0.519 \\
\hline
\end{array}
\quad
\begin{array}{|c|c|c|}
\hline
D \text{ range is} & \beta_1 & \beta_2 & \beta_3 \\
[0.1 \times 10^{-3}, 3 \times 10^{-3}] & & & \\
\hline
m & \beta_1 & \beta_2 & \beta_3 \\
2 & 0 & 0.080 & - \\
3 & 0 & 0.055 & 0.170 \\
4 & 0 & 0.064(2) & 0.276 \\
5 & 0 & 0.070(3) & 0.328 \\
\hline
\end{array}
\]

Table 2.2: Optimum diffusion weighting factors with multiple measurements for normal rat brains. $m$ is the total number of measurements, the number of measurement at a particular $b$-value is shown in the bracket in the event that it is more than one and the unit of the $b$-values is $s/mm^2$.

\[
\begin{array}{|c|c|c|c|}
\hline
Grey matter & White matter \\
\hline
m & b_1 & b_2 & m & b_1 & b_2 \\
2 & 0 & 2370 & 2 & 0 & 2830 \\
3 & 0 & 2530(2) & 3 & 0 & 3010(2) \\
4 & 0 & 2650(3) & 4 & 0 & 3150(3) \\
5 & 0(2) & 2460(4) & 5 & 0 & 3250(4) \\
\hline
\end{array}
\]

used in [12] and hence yields higher accuracy results. For our second experiment, we use real distributions of the ADC values estimated from a normal rat brain. Histograms of the ADC values from different parts of a normal rat brain are used as these distributions. These are not uniform as evident from the Fig. 2.1. Results of weighted COV versus $b_2$ and $b_3$ using $(0, b_2, b_3)$ as diffusion weighting factors are shown in Fig. 2.4. Furthermore, the optimal $b$-values with multiple measurements of up to five are summarized in table 2.2. Notice that we get two different $b$ values as the optimal diffusion weighting factors for these real distributions.
Figure 2.4: Weighted COV of the nonlinear ADC estimate versus the diffusion weight factors. $COV_w(D^*)$ of the gray matter (First row) and the white matter (Second row) versus $b_2$ and $b_3$ using $(0, b_2, b_3)$ as diffusion weighting factors.
CHAPTER 3
A CONSTRAINED VARIATIONAL PRINCIPLE FOR DIRECT ESTIMATION
AND SMOOTHING OF THE DTI FROM COMPLEX DWIS

3.1 Introduction

We are the first to propose the idea of simultaneous estimation and smoothing of the diffusion tensors from DWI [30] and gave a significant improvement later on in [31]. We further extend our model [31] to restoration of DTI from complex valued diffusion weighted images. Specifically, we propose a novel formulation of the DTI estimation and smoothing as a constrained optimization problem. The specific approach we use is called the augmented Lagrangian technique which allows one to deal with inequality constraints. The novelty of our formulation lies in the ability to directly, in a single step process, estimate a smooth $D$ from the noisy complex measurements $S_l$ with the preservation of its positiveness. The formulation does not require any adhoc methods of setting tuning parameters to achieve the solution. These are the key features distinguishing our solution method from methods reported in literature to date.

In contrast to our solution (to be described subsequently in detail), most of the earlier approaches used a two step method involving (i) computation of a $D$ from $\|S_l\|$ using a linear least-squares approach and then (ii) computing a smoothed $D$ via either smoothing of the eigenvalues and eigenvectors of $D$ or using the matrix flows approach in Chefd’hotel et al. [28]. The problem with the two step approach to computing $D$ is that the estimated $D$ in the first step using the log-linearized model need not be positive definite or even semi-definite. Moreover, it is hard to trust the fidelity of the eigenvalues and vectors computed from such matrices even if they are to be smoothed subsequently prior to mapping out the nerve fiber tracts.
Briefly, our model seeks to minimize a cost function involving, the sum of an $L^p$ norm based gradient of the Cholesky factor $L$ which ensure the positiveness of $D$ by the Cholesky factorization $LL^T$ – and an $L^p$ norm based gradient of $S_0$, subject to a nonlinear data constraint based on the complex (not linearized) Stejskal-Tanner equation (1.4). The model is posed as a constrained variational principle which can be minimized by either discretizing the variational principle itself or the associated Euler-Lagrange equation. We choose the former and use the augmented Lagrangian method together with the limited memory quasi-Newton method to achieve the solution.

This chapter is organized as follows: in section 3.2, the detailed variational formulation is described along with the nonlinear data constraints, the positive definite constraint and the augmented Lagrangian solution. Section 3.3 contains the detailed description of the discretization as well as the algorithmic description of the augmented Lagrangian framework. In section 3.4, we present experiments on application of our model to synthetic as well as real data. Synthetic data experiments are conducted to present comparison of tensor field restoration results with a recently presented work of Coulon et al. [22]. Moreover, results of comparison between the use of the linearized Stejskal-Tanner model and the nonlinear form of the same are presented as well.

3.2 Constrained Variational Principle Formulation

Our solution to the recovery of a piecewise smooth diffusion tensor field from the complex measurements $S_i$ is posed as a constrained variational principle. We seek to minimize a measure of lack of smoothness in $S_0$ and the diffusion tensor $D$ being estimated using an $L^p$ norm of the gradient in $S_0$ and an $L^p$ norm of the gradient in the Cholesky factor $L$. This measure is then constrained by a nonlinear data fidelity term related to the complex Stejskal-Tanner equation (1.4). The nonlinear data term is constrained by an inequality which requires that it be bounded from above by a possibly known tolerance factor. The positiveness constraint on the diffusion tensor
being estimated is achieved via the use of the Cholesky factorization theorem from computational linear algebra. The constrained variational principle is discretized and posed using the augmented Lagrangian technique [43]. Each subproblem in the augmented Lagrangian framework is then solved using the limited memory Quasi-Newton scheme [43]. The novelty of our formulation lies in the unified framework for recovering and smoothing of the diffusion tensor field from the raw data $S_l$. In addition, to our knowledge, this is the first formulation which allows for simultaneous estimation and smoothing of $D$ as well as one in which the regularization parameter is not set in an ad hoc way.

Let $S_0(x) = R_0(x) + iI_0(x)$ be the complex DWI when no diffusion-encoding magnetic gradient is present, $D(x)$ be the unknown symmetric positive definite diffusion tensor, $LL^T(x)$ be the Cholesky factorization of the diffusion tensor with $L$ being a lower triangular matrix, $S_l(x) = R_l(x) + iI_l(x), l = 1, \ldots, N$ are the complex DWIs measured after application of a diffusion-encoded magnetic gradient of known strength and direction and $N$ is the total number of measured DWIs. The constrained variational principle is

$$\min_{S_0, L} \mathcal{E}(S_0, L) = \int_{\Omega} \left[ |\nabla R_0|^{p_1} + |\nabla I_0|^{p_1} + |\nabla L|^{p_2} \right] dx$$

s.t. $\mathcal{C}(S_0, L) = \alpha \sigma^2 - \int_{\Omega} \sum_{l=1}^{N} (FR_l^2 + FI_l^2) dx \geq 0 \quad (3.1)$

where $\Omega$ is the image domain, the first and the second terms in the variational principle are $L^p$ smoothness norm on the real and imaginary part of $S_0$, the third term is an $L^p$ smoothness norm on $L$, where $p_1 > 6/5$ for $R_0$ and $I_0$ and $p_2 \geq 1$ for $L$. $|\nabla L|^{p_2} = \sum_d |\nabla L_d|^{p_2}$, where $d \in D = \{xx, yy, zz, xy, yz, xz\}$ are indices to the six nonzero components of $L$. The lower bounds on the value of $p_1$ and $p_2$ are chosen so as to make the proof of existence of a solution for this minimization (see theorem 3.2.4 and its proof) mathematically tractable. $\alpha$ is a constant scale factor and $\sigma^2$ is
the noise standard deviation in the measurements $S_t$. $FR_t$ and $FI_t$ are defined as

$$FR_t = R_t - R_0 e^{-b_t LL^T}, \quad FI_t = I_t - I_0 e^{-b_t LL^T}$$

3.2.1 The Complex Nonlinear Data Constraint

The Stejskal-Tanner equation (1.4) shows the relation between the complex diffusion weighted image $S_t$ and the diffusion tensor $D$. However, multivariate linear regression based on (1.6) has been used to estimate the diffusion tensor $D$ [1]. It was pointed out in [1] that these results agree with nonlinear regression based on the magnitude Stejskal-Tanner equation (1.5). However, if the signal to noise ratio (SNR) is low and the number of $S_t$ is not very large (unlike in [1] where $N = 315$ or $N = 294$), the result from multivariate linear regression will differ from the nonlinear regression significantly. A robust estimator belonging to the M-estimator family was used by Poupon et al. [21], however, its performance is not discussed in detail. In Westin et al. [45]), an analytical solution is derived from (1.6) by using a dual tensor basis, however, it should be noted that this can only be used for computing the diffusion tensor $D$ when there is no noise in the measurements $S_t$ or the SNR is extremely high.

Our aim is to provide an accurate estimation of the diffusion tensor $D$ for practical use, where the SNR may not be high and the total number of DWIs $N$ is restricted to a moderate number. The nonlinear data fidelity term based on the complex Stejskal-Tanner equation (1.4) is fully justified for use in such situations. This nonlinear data term is part of an inequality constraint that imposes an upper bound on the closeness of the measurements $S_t$ to the mathematical model $S_0 e^{-b_t LL^T}$. The bound $\alpha \sigma^2$ may be estimated automatically [46, 47].
3.2.2 $L^p$ Smoothness Norm

There are numerous image smoothing techniques, however, not many of them are suitable for keeping sharp discontinuities. Partial Differential Equation based (PDE) methods are the few that can capture edges and in particular total variation (TV) norm based restoration [18] introduced by Rudin et al. is an excellent PDE-based edge preserving denoising method. TV restoration aims to minimizing the total variation functional

$$TV(u) = \int_\Omega |\nabla u|$$  \hspace{1cm} (3.2)

subject to the noise constraint:

$$\|Au - z\|^2 \leq \sigma^2$$  \hspace{1cm} (3.3)

where $z$ is a noisy scalar image defined on domain $\Omega$, $u$ is the smoothed image, $A$ is bounded linear operator and it is an identity for image denoising, and $\sigma$ quantifies the amount of Gaussian noise added to the observed image. TV based restoration can be extended to $L^p$ smoothness norm based method [48] by replacing the total variational functional with

$$L^p(u) = \int_\Omega |\nabla u|^p$$  \hspace{1cm} (3.4)

where $p \geq 1$. When $p = 1$, equation (3.4) becomes total variation norm, when $p = 2$, equation (3.4) becomes $H^1$ norm. In Blomgren et al. [48], it was shown that $L^p$ smoothness norm based restoration doesn’t admit discontinuous solutions as the TV-norm does when $p > 1$. However, when $p$ is chosen close to 1, its behavior is close to the TV-norm for restoring edges. We adopt $L^p$ smoothness norm in our constrained model. In particular, we need $p_1 > 6/5$ for regularizing $S_0$ and $p_2 \geq 1$ for $L$ to ensure existence of the solution described in 3.2.4. Note that what is of importance here is the estimation the diffusion tensor $D$ and therefore, the edge-preserving property.
in the estimation process is more relevant for the case of $D$ than for $S_0$. In our experiment, we choose $p_1 = 1.205$ for $S_0$ and $p_2 = 1.00$ for $L$.

3.2.3 The Positive Definite Constraint

In general, a matrix $A \in \mathbb{R}^{n \times n}$ is said to be positive definite if $x^T A x > 0$, for all $x \neq 0$ in $\mathbb{R}^n$. The diffusion tensor $D$ happens to be a symmetric positive definite matrix but due to the noise in the data $S_l$, it is hard to recover a $D$ that retains this property unless one includes it explicitly as a constraint. One way to impose this constraint is using the Cholesky factorization theorem, which states that: *If $A$ is a symmetric positive definite matrix then, there exists a unique factorization $A = LL^T$ where, $L$ is a lower triangular matrix with positive diagonal elements.* After doing the Cholesky factorization, we have transferred the positive definiteness constraint on the matrix $D$ to an inequality constraint on the diagonal elements of $L$. This is however still hard to satisfy theoretically because, the set on which the minimization takes place is an open set. However, in practise, with finite precision arithmetic, testing for a positive definiteness constraint is equivalent to testing for positive semi-definiteness. This is because for any symmetric positive definite matrix $D$, its machine representation $\tilde{D} = D + E$ with $\| E \| \leq \epsilon \| D \|$, where $\epsilon$ is a small multiple of the machine precision. When $D$ is a small symmetric positive definite matrix, $\tilde{D}$ can become a semi-definite matrix, it follows that in finite precision arithmetic, testing for definiteness is equivalent to testing for semi-definiteness. Thus, we repose the positive definiteness constraint on the diffusion tensor matrix as, $x^T \tilde{D} x \geq 0$ which is satisfied when $\tilde{D} = LL^T$.

3.2.4 Existence of a Solution

Justification for using the augmented Lagrangian method for constrained problems is given in Nocedal and Wright [43], thus we only need to prove there is a solution
for the following subproblem:

\[
\min_{(S_0, L) \in \mathcal{A}} \mathcal{L}(S_0, L; \lambda, \mu) = \begin{cases} 
\mathcal{E}(S_0, L) - \lambda \mathcal{C}(S_0, L) + \frac{c^2(S_0, L)}{2\mu} & \text{if } \mathcal{C}(S_0, L) \leq \mu \lambda \\
\mathcal{E}(S_0, L) - \frac{\mu}{2} \lambda^2 & \text{otherwise}
\end{cases}
\]  

(3.5)

where \( \lambda \geq 0 \) is an estimate of the Lagrange multiplier, \( \mu > 0 \) is a penalty parameter and \( \mathcal{A} = \{(S_0, L) \mid L \in L^p(\Omega) \text{ where } p \geq 1 \text{ and } R_0, I_0 \in W^{1,p}(\Omega) \text{ where } p > 6/5\} \). Here \( \Omega \subset \mathbb{R}^d \), \( L^p(\Omega) \) denotes the space of functions with bounded \( L^p \) norms, \( L^2(\Omega) \) is the space of square integrable functions on \( \Omega \) and \( W^{1,p}(\Omega) \) denotes the Sobolev space of order \( p \) on \( \Omega \) [49]. Consider the augmented Lagrangian formulation (3.5) which serves as a subproblem of (3.1), the existence theorem will be stated and proved after the following two lemmas. If not pointed out, the definitions and theorems employed in the proof can be found in Evans [49].

**Lemma 1** Let \( \mathcal{A}_1 = \{(S_0, L) \mid (S_0, L) \in \mathcal{A} \text{ and } \mathcal{C}(S_0, L) \leq \mu \lambda, \text{ and suppose } R_0, I_0 \in L^2(\Omega), \text{ then the following minimization problem (3.6) has a solution } (S_0^*, L^*) \in \mathcal{A}_1 \text{ if } \mathcal{A}_1 \neq \emptyset:} \]

\[
\min_{(S_0, L) \in \mathcal{A}_1} \mathcal{L}(S_0, L; \lambda, \mu) = \mathcal{E}(S_0, L) - \lambda \mathcal{C}(S_0, L) + \frac{1}{2\mu} \mathcal{C}^2(S_0, L)
\]

(3.6)

**Proof 1** We will verify the following three statements one by one and then prove this lemma:

- The first term \( \mathcal{E}(S_0, L) \) in (3.6) is lower semi-continuous with respect to \( L \) in \( L^p(\Omega), p \geq 1 \) and weakly lower semi-continuous with respect to \( S_0 \) in \( W^{1,p}(\Omega), p \geq q_1 \).

- The second term \( \mathcal{C}(S_0, L) \) in (3.6) is continuous with respect to \( S_0 \) in \( W^{1,p}(\Omega) \) when \( p > 6/5 \), and is continuous with respect to \( L \) in \( L^p(\Omega) \) when \( p \geq 1 \).

- The third term \( \mathcal{C}^2(S_0, L) \) in (3.6) has the same continuity property as the second term.
As $\mathcal{L}$ in (3.6) is lower bounded, there exists a minimizing sequence $(S_0^n, L^n)$ for it, where $L^n_d \in L^p(\Omega), d \in \mathcal{D}, p \geq 1, R^n_0$ and $I^n_0 \in W^{1,p}(\Omega), p > 6/5$ and $C(S_0^n, L^n) \leq \mu \lambda$. Then

\begin{align*}
\{L^n_d\}_{n=1}^{\infty}, d \in \mathcal{D} & \text{ is bounded in } L^p(\Omega), p \geq 1. \\
\{R^n_0\}_{n=1}^{\infty} \text{ and } \{I^n_0\}_{n=1}^{\infty} & \text{ are bounded in } W^{1,p}(\Omega).
\end{align*}

Therefore there is a subsequence $\{(S_0^{n_k}, L^{n_k})\}, L^n_d \in L^p(\Omega), p \geq 1, d \in \mathcal{D}$ and $R^n_0, I^n_0 \in W^{1,p}(\Omega), p > 6/5$ such that when $n_k \to \infty$,

- $\{L^{n_k}_d\} \to L^*_d, d \in \mathcal{D}$ in $L^1(\Omega)$ and a.e. on $\Omega$. (From the compactness property of $L^p(\Omega), p \geq 1$).
- $\{R^{n_k}_0\} \to R_0^*$ and $\{I^{n_k}_0\} \to I_0^*$ in $W^{1,p}(\Omega)$. (From the weak compactness of $W^{1,p}$)
- $\{R^{n_k}_0\} \to R_0^*$ and $\{I^{n_k}_0\} \to I_0^*$ in $L^2(\Omega)$ and a.e. on $\Omega$. (From Rellich-Kondrachov Compactness Theorem when $p > 6/5$ and $n = 3$, this is why we need $p > 6/5$!)

1) Now we are ready to demonstrate the lower semi-continuity of the first term $\mathcal{E}(S_0, L)$ in (3.6).

By the lower semi-continuity of $L^p$ norm in $L^p(\Omega), p \geq 1$, we have

\begin{align*}
\int_{\Omega} |\nabla L^*|^p d\mathbf{x} \leq & \int_{\Omega} \sum_{d \in \mathcal{D}} |\nabla L_d|^p d\mathbf{x} \\
\leq & \liminf_{n_k \to \infty} \int_{\Omega} \sum_{d \in \mathcal{D}} |\nabla L^{n_k}_d|^p d\mathbf{x} \leq \liminf_{n_k \to \infty} \int_{\Omega} |\nabla L^{n_k}|^p d\mathbf{x}
\end{align*}

By the lower semi-continuity of $L^p$ norm in $W^{1,p}(\Omega)$, we have

\begin{align*}
\int_{\Omega} |\nabla R_0^*|^p d\mathbf{x} \leq & \liminf_{n_k \to \infty} \int_{\Omega} |\nabla R^{n_k}_0|^p d\mathbf{x} \\
\int_{\Omega} |\nabla I_0^*|^p d\mathbf{x} \leq & \liminf_{n_k \to \infty} \int_{\Omega} |\nabla I^{n_k}_0|^p d\mathbf{x}
\end{align*}
Thus
\[
\mathcal{E}(S_0^*, L^*) = \int_{\Omega} \left[ |\nabla R_0^*(x)|^{p_1} + |\nabla I_0^*(x)|^{p_1} + |\nabla L^*(x)|^{p_2} \right] dx
\leq \liminf_{n_k \to \infty} \int_{\Omega} \left[ |\nabla R_0^{n_k}(x)|^{p_1} + |\nabla I_0^{n_k}(x)|^{p_1} + |\nabla L^*(x)|^{p_2} \right] dx \quad (3.7)
\]

2) Next we claim
\[
\mathcal{C}(S_0^*, L^*) = \lim_{n_k \to \infty} \mathcal{C}(S_0^{n_k}, L^{n_k}) \quad (3.8)
\]

Since \( \mathcal{C}(S_0^*, L^*) = \alpha \sigma^2 - \int_{\Omega} \sum_{l=1}^{N} \| S_l - S_0^* e^{-b_l L^* L^{n_k}} \|^2 dx \), we only need to show
\[
\int_{\Omega} \sum_{l=1}^{N} \| S_l - S_0^* e^{-b_l L^* L^{n_k}} \|^2 dx = \lim_{n_k \to \infty} \int_{\Omega} \sum_{l=1}^{N} \| S_l - S_0^{n_k} e^{-b_l L^{n_k} (L^{n_k})^T} \|^2 dx \quad (3.9)
\]

This will be proved in several stages

(a) Let \( FR_t^* = R_t - S_0^* e^{-b_l L^* L^T} \) and \( FR_{t,n_k} = R_t - R_0^{n_k} e^{-b_l L^{n_k} (L^{n_k})^T} \), then
\[
\int_{\Omega} \left[ FR_{t,n_k} - FR_t^* \right]^2 dx
= \int_{\Omega} \left[ R_0^{n_k} e^{-b_l L^{n_k} (L^{n_k})^T} - R_0^* e^{-b_l L^* L^T} \right]^2 dx
\leq 2 \int_{\Omega} \left[ R_0^{n_k} (e^{-b_l L^{n_k} (L^{n_k})^T} - e^{-b_l L^* L^T}) \right]^2 dx + 2 \int_{\Omega} \left[ (R_0^{n_k} - R_0^*) e^{-b_l L^* L^T} \right]^2 dx
\leq 2A + 2B
\]

Since \( \| R_0^{n_k} \|_{W^{1,p}(\Omega)} \) is uniformly bounded in \( n_k \), by the Sobolev embedding theorem, for \( R^3 \), we have
\[
\| R_0^{n_k} \|_{L^2(\Omega)} \leq C \| R_0^{n_k} \|_{W^{1,p}(\Omega)} \leq C
\]

Noting the fact that \( L^{n_k} \) has a strong convergence towards \( L^* \) in \( L^1(\Omega) \), we have from the Dominant Convergence Theorem [49]
\[
\lim_{n_k \to \infty} \| e^{-b_l L^{n_k} (L^{n_k})^T} - e^{-b_l L^* L^T} \|_{L^2(\Omega)} = 0
\]
Thus we have

\[
A \triangleq \int_{\Omega} \left[ R_{0}^{n} (e^{-b_{i} : L^{T} (L^{n} k)^{T} - e^{-b_{i} : L^{*} L^{T}}}) \right] dx \\
\leq \| R_{0}^{n} \|_{L^{2}(\Omega)}^{2} \| e^{-b_{i} : L^{T} (L^{n} k)^{T} - e^{-b_{i} : L^{*} L^{T}}} \|_{L^{2}(\Omega)}^{2} \xrightarrow{n \rightarrow \infty} 0
\]

From the strong convergence of \( R_{0}^{n} \) to \( R_{0}^{*} \) in \( L^{2}(\Omega) \) and \( e^{-b_{i} : L^{*} L^{T}} \leq 1 \), we have

\[
|B| \leq \int_{\Omega} |R_{0}^{n} - R_{0}^{*}|^{2} dx \xrightarrow{n \rightarrow \infty} 0
\]

Now as \( A \xrightarrow{n \rightarrow \infty} 0 \) and \( B \xrightarrow{n \rightarrow \infty} 0 \), we have

\[
\lim_{n \rightarrow \infty} \| FR_{l}^{*} - FR_{l, n} \|_{L^{2}(\Omega)} \xrightarrow{n \rightarrow \infty} \int_{\Omega} [FR_{l}^{*} - FR_{l, n}] dx = 0 \quad (3.10)
\]

(b) We now prove

\[
\int_{\Omega} FR_{l}^{*2} dx = \lim_{n \rightarrow \infty} \int_{\Omega} FR_{l, n}^{2} dx \quad (3.11)
\]

First of all, if \( f \in L^{2}(\Omega) \) and \( g \in L^{2}(\Omega) \), then we have \( f - g \in L^{2}(\Omega) \) and \( f + g \in L^{2}(\Omega) \). Further we have

\[
\left| \| f \|_{L^{2}(\Omega)}^{2} - \| g \|_{L^{2}(\Omega)}^{2} \right| = \left| \int_{\Omega} f^{2} dx - \int_{\Omega} g^{2} dx \right| = \int_{\Omega} (f^{2} - g^{2}) dx \leq \int_{\Omega} |f^{2} - g^{2}| dx \\
= \int_{\Omega} |f - g||f + g| dx \leq \| f - g \|_{L^{2}(\Omega)} \| f + g \|_{L^{2}(\Omega)} \quad (\text{Hölder's inequality}) (3.12)
\]

It is easy to verify that \( FR_{l}^{*} \in L^{2}(\Omega) \), \( FR_{l, n} \in L^{2}(\Omega) \) and are uniformly bounded in \( L^{2}(\Omega) \). Thus applying the above equation, we have

\[
\left| \int_{\Omega} FR_{l}^{2} dx - \int_{\Omega} FR_{l, n}^{2} dx \right| = \| FR_{l}^{*} \|_{L^{2}(\Omega)} - \| FR_{l, n} \|_{L^{2}(\Omega)} \leq C\| FR_{l}^{*} - FR_{l, n} \|_{L^{2}(\Omega)} \rightarrow 0, \quad \text{as} \ n \rightarrow \infty
\]
(c) Similarly as previous step, we can prove

\[
\int_{\Omega} \left[ I_t - I_0^* e^{-b_t L^* T} \right]^2 \, dx = \lim_{n_k \to \infty} \int_{\Omega} \left[ I_t - I_0^{n_k} e^{-b_t L^{n_k} (L^{n_k})^2} \right]^2 \, dx
\]  

(3.13)

Combining with (b), it is easy to verify equation (3.9).

3) Now we will show that

\[
C(S_0^*, L^*)^2 = \lim_{n_k \to \infty} C(S_0^{n_k}, L^{n_k})^2
\]  

(3.14)

The above can be easily verified since \( C(S_0^*, L^*), C(S_0^{n_k}, L^{n_k}) \) are bounded and

\[
C(S_0^*, L^*) = \lim_{n_k \to \infty} C(S_0^{n_k}, L^{n_k})
\]  

(3.15)

Finally, we have from 1), 2) and 3)

\[
\mathcal{L}(S_0^*, L^*; \lambda, \mu) \leq \liminf_{n_k \to \infty} \mathcal{L}(S_0^{n_k}, L^{n_k}; \lambda, \mu) = \inf \mathcal{L}(S_0, L; \lambda, \mu)
\]  

(3.16)

Therefore, \((S_0^*, L^*)\) is a minimizer of \(\mathcal{L}(S_0, L; \lambda, \mu)\) as defined in (3.6).

**Lemma 2** Let \( A_2 = \{ (S_0, L) \mid (S_0, L) \in A \text{ and } C(S_0, L) > \mu \lambda \} \), and suppose \( R_t, I_t \in L^2(\Omega) \), then the following minimization problem (3.17) has a solution \((S_0^*, L^*) \in A_2\) if \( A_2 \neq \emptyset \):

\[
\min_{(S_0, L) \in A_2} \mathcal{L}(S_0, L; \lambda, \mu) = \mathcal{E}(S_0, L) - \frac{\mu}{2} \lambda^2
\]  

(3.17)

The proof is similar as in the lemma 1.

**Theorem 1** Suppose \( R_t, I_t \in L^2(\Omega) \), then the augmented Lagrangian formulation (3.5) has a solution \((S_0^*, L^*) \in A\).

**Proof 2** It is easy to see \( A \neq \emptyset \), as a matter of fact, constant functions will be members of \( A \). Thus, there will be three cases:

1. \( A_1 \neq \emptyset \) and \( A_2 = \emptyset \)

2. \( A_2 \neq \emptyset \) and \( A_1 = \emptyset \)
3. \( \mathcal{A}_1 \neq \emptyset \) and \( \mathcal{A}_2 \neq \emptyset \).

Here we provide a proof for case 3, case 1 and 2 are trivial to prove. Let \((S^1_0, L^1)\) be the solution for (3.6) and \((S^2_0, L^2)\) be the solution for (3.17), it is not hard to see that the solution of (3.5) is

\[
(S^*_0, L^*) = \begin{cases} 
(S^1_0, L^1) & \text{if } \mathcal{L}(S^1_0, L^1; \lambda, \mu) \leq \mathcal{L}(S^2_0, L^2; \lambda, \mu) \\
(S^2_0, L^2) & \text{otherwise}
\end{cases}
\]

Finding a solution of the constrained variation principle (3.1) involves solving a sequence of (3.5) with fixed \( \lambda \) and \( \mu \) at each stage. It is much more difficult than when dealing with the problems of recovering and smoothing separately. However, there are benefits of posing the problem in this constrained unified framework, namely, one does not accumulate the errors from a two stage process. Moreover, this framework incorporates the nonlinear data term which is more appropriate for low SNR values prevalent when the magnitude of the diffusion-encoded magnetic gradient is high. Also, the noise model is correct for the nonlinear complex data model unlike the log-linearized case. Lastly, in the constrained formulation, it is now possible to pose mathematical questions of existence and uniqueness of the solution – which was not possible in earlier formulations reported in literature.

### 3.3 Numerical Methods

The minimization problem given by (3.1) can only be solved numerically. Here, we discretize the constrained variational principle (3.1), transform it into a sequence of unconstrained problems by using the augmented Lagrangian method and then employ the limited Quasi-Newton technique [43] to solve them. Note that this framework allows us to solve the minimization without resorting to adhoc methods of choosing the ”tuning” parameters. Also limited memory Quasi-Newton is the method of choice here due to the advantages it affords in the context of memory/storage savings.
Proper line search technique is employed once the search direction is found by using limited memory Quasi-Newton method.

### 3.3.1 Discretized Constrained Variational Principle

We use the standard finite difference method to discretize the problem. Let

\[
FR_{l,ijk} = R_{l,ijk} - R_{0,ijk} e^{-b_i L_{ijk} L_{ijk}^T} \\
FI_{l,ijk} = I_{l,ijk} - I_{0,ijk} e^{-b_i L_{ijk} L_{ijk}^T} \\
F_{l,ijk} = FR_{l,ijk} + FI_{l,ijk}
\]

\[
|\nabla R_0|_{ijk} = \left[\sqrt{(\Delta_x^+ R_0)^2 + (\Delta_y^+ R_0)^2 + (\Delta_z^+ R_0)^2 + \epsilon}\right]_{ijk} \\
|\nabla I_0|_{ijk} = \left[\sqrt{(\Delta_x^+ I_0)^2 + (\Delta_y^+ I_0)^2 + (\Delta_z^+ I_0)^2 + \epsilon}\right]_{ijk} \\
|\nabla L_d|_{ijk} = \left[\sqrt{(\Delta_x^+ L_d)^2 + (\Delta_y^+ L_d)^2 + (\Delta_z^+ L_d)^2 + \epsilon}\right]_{ijk} \\
|\nabla L|_{ijk}^p = \sum_{d \in D} |\nabla L_d|_{ijk}^p
\]

where \(\Delta_x^+, \Delta_y^+\) and \(\Delta_z^+\) are forward difference operators, \(\epsilon\) is a small positive number used to avoid singularities of the \(L^p\) norm when \(p < 2\). Now the discretized constrained variational principle can be written as

\[
\min_{S_0, L} \mathcal{E}(S_0, L) = \sum_{i,j,k} (|\nabla R_0|_{ijk}^{p_1} + |\nabla I_0|_{ijk}^{p_1} + |\nabla L|_{ijk}^{p_2}) \\
s.t. C(S_0, L) = \alpha \sigma^2 - \sum_{i,j,k} \sum_{l=1}^N \|F_{l,ijk}\|^2 \geq 0 \tag{3.18}
\]

### 3.3.2 Augmented Lagrangian Method

The above problem is now posed using the augmented Lagrangian method, where a sequence of related unconstrained subproblems is solved, and the limit of these solutions is the solution to (3.18). Following the description in Sijbers [43], the \(k\)-th
subproblem of (3.18) is given by
\[
\min \mathcal{L}(S_0, L, s; \lambda_k, \mu_k) = \mathcal{E}(S_0, L) - \lambda_k [\mathcal{C}(S_0, L) - s] + \frac{1}{2\mu_k} [\mathcal{C}(S_0, L) - s]^2 \quad (3.19)
\]
where \( s \geq 0 \) is a slack variable, \( \mu_k, \lambda_k \) are the barrier parameter and the Lagrange multiplier estimate for the \( k \)-th subproblem respectively.

One can explicitly compute the slack variable \( s \) at the minimum as
\[
s = max \left[ \mathcal{C}(S_0, L) - \mu_k \lambda_k, 0 \right]
\]
and substitute it in (3.19) to get an equivalent subproblem in \((S_0, L)\) given by
\[
\min \mathcal{L}(S_0, L; \lambda_k, \mu_k)
\]
\[
= \begin{cases} 
\mathcal{E}(S_0, L) - \lambda_k \mathcal{C}(S_0, L) + \frac{\mathcal{C}^2(S_0, L)}{2\mu_k} & \text{if } \mathcal{C}(S_0, L) \leq \mu_k \lambda_k \\
\mathcal{E}(S_0, L) - \frac{\mu_k}{2} \lambda_k^2 & \text{otherwise}
\end{cases}
\]

The following algorithm summarizes the procedure to find the solution for (3.18):

**Algorithm 1** Augmented Lagrangian Algorithm

1. Initialize \( S_0(0), L(0) \) using the nonlinear regression, choose an initial \( \mu_0 \) and \( \lambda_0 \).
2. for \( k = 1, 2, ... \)
3. Find an approximate minimizer \( S_0(k), L(k) \) of \( \mathcal{L}(\cdot, \cdot; \lambda_k, \mu_k) \) as in (3.20) starting with \( S_0(k - 1), L(k - 1) \);
4. If final convergence test is satisfied
5. STOP with an approximate solution \( S_0(k), L(k) \);
6. Update the Lagrange multiplier using \( \lambda_{k+1} = max(\lambda_k - \mathcal{C}(S_0, L)/\mu_k, 0) \);
7. Choose a new penalty parameter \( \mu_{k+1} = \mu_k/2 \);
8. Set the new starting point for the next iteration to \( S_0(k), L(k) \);
9. end(for)

3.3.3 Limited Memory Quasi-Newton Method

Due to the large number of unknown variables in the minimization, we solve the subproblem using limited memory Quasi-Newton technique. Hessian matrix at each iteration of the optimization by using only the first derivative information. In the
Limited-Memory Broyden-Fletcher-Goldfarb-Shano (BFGS) method, search direction is computed without storing the approximated Hessian matrix which can be a very large matrix in general ($O(N^6)$ size for $O(N^3)$ unknowns).

Let $\mathbf{x} = (S_0, \mathbf{L})$ be the vector of variables, and $f(\mathbf{x}) = \mathcal{L}(S_0, \mathbf{L}; \lambda, \mu)$ denote the augmented Lagrangian function (3.20) to be minimized. For simplicity, we write $f(\mathbf{x}) = \mathcal{L}(S_0, \mathbf{L})$ by omitting the fixed parameter $\lambda$ and $\mu$ in the following description.

At $k$th iteration, let $s_k = x_{k+1} - x_k$ be the update of the variable vector $x$, $y_k = \nabla f_{k+1} - \nabla f_k$ the update of the gradient and $H_k^{-1}$ the approximation of the inverse of the Hessian. The inverse of the approximate Hessian can be approximated using the BFGS updating formula:

$$H_{k+1}^{-1} = V_k H_k^{-1} V_k^T + \frac{S_k S_k^T}{y_k^T s_k}$$  \label{eqn:bfgs}

where $V_k = I - \frac{y_k s_k^T}{y_k^T s_k}$.

Then we can use the following L-BFGS two-loop recursion iterative procedure, which computes the search direction $H_k^{-1} \nabla f_k$ efficiently by using last $m$ pairs of $(s_k, y_k)$ [43].

\begin{algorithm}
\caption{Search Direction Update Algorithm}
\begin{algorithmic}[1]
\State $q \leftarrow \nabla f_k$;
\For{$i = k - 1, k - 2, ..., k - m$}
\State $\alpha_i \leftarrow \rho_i s_i^T q$;
\State $q \leftarrow q - \alpha_i y_i$;
\EndFor
\State $r \leftarrow (H_k^0)^{-1} q$;
\For{$i = k - m, k - m - 1, ..., k - 1$}
\State $\beta \leftarrow \rho_i y_i^T r$;
\State $r \leftarrow r + s_i (\alpha_i - \beta)$
\EndFor
\State \textbf{stop} with result $H_k^{-1} \nabla f_k = r$.
\end{algorithmic}
\end{algorithm}

where $\rho_k = \frac{1}{y_k^T s_k}$ and $(H_k^0)^{-1}$ is the initial approximation of the inverse of the Hessian, we set $(H_k^0)^{-1} = \gamma_k I$ where $\gamma_k = \frac{s_k^T y_{k-1} y_{k-1}}{y_{k-1}^T y_{k-1}}$. 
The gradient of our energy function is

\[ \nabla f(x) = \left( \frac{\partial L(S_0, L)}{\partial R_0}, \frac{\partial L(S_0, L)}{\partial I_0}, \frac{\partial L(S_0, L)}{\partial L_{xx}}, \frac{\partial L(S_0, L)}{\partial L_{yy}}, \frac{\partial L(S_0, L)}{\partial L_{zz}}, \frac{\partial L(S_0, L)}{\partial L_{xy}}, \frac{\partial L(S_0, L)}{\partial L_{xz}}, \frac{\partial L(S_0, L)}{\partial L_{yz}} \right) \]

(3.22)

where

\[
\frac{\partial L(S_0, L)}{\partial R_{0,ijk}} = \begin{cases} 
\sum_{i'j'k'} \frac{\partial |\nabla R_0|_p^{i'j'k'}}{\partial R_{0,ijk}} & \text{if } C(S_0, L) - \mu_k \lambda_k > 0 \\
\sum_{i'j'k'} \frac{\partial |\nabla R_0|_p^{i'j'k'}}{\partial R_{0,ijk}} + 2(\lambda - \frac{C(S_0, L)}{\mu}) \sum_{l=1}^{N} FR_{l,ijk} \frac{\partial FR_{l,ijk}}{\partial R_{0,ijk}} & \text{otherwise}
\end{cases}
\]

\[
\frac{\partial L(S_0, L)}{\partial I_{0,ijk}} = \begin{cases} 
\sum_{i'j'k'} \frac{\partial |\nabla I_0|_p^{i'j'k'}}{\partial I_{0,ijk}} & \text{if } C(S_0, L) - \mu_k \lambda_k > 0 \\
\sum_{i'j'k'} \frac{\partial |\nabla I_0|_p^{i'j'k'}}{\partial I_{0,ijk}} + 2(\lambda - \frac{C(S_0, L)}{\mu}) \sum_{l=1}^{N} FI_{l,ijk} \frac{\partial FI_{l,ijk}}{\partial I_{0,ijk}} & \text{otherwise}
\end{cases}
\]

\[
\frac{\partial L(S_0, L)}{\partial L_{d,ijk}} = \begin{cases} 
\sum_{i'j'k'} \frac{\partial |\nabla L_d|_p^{i'j'k'}}{\partial L_{d,ijk}} & \text{if } C(S_0, L) - \mu_k \lambda_k > 0 \\
\sum_{i'j'k'} \frac{\partial |\nabla L_d|_p^{i'j'k'}}{\partial L_{d,ijk}} + 2(\lambda - \frac{C(S_0, L)}{\mu}) \sum_{l=1}^{N} (FR_{l,ijk} \frac{\partial FR_{l,ijk}}{\partial L_{d,ijk}} + FI_{l,ijk} \frac{\partial FI_{l,ijk}}{\partial L_{d,ijk}}) & \text{otherwise}
\end{cases}
\]

\[
d = xx, yy, zz, xy, yz, xz
\]

(3.23)

Here \( \sum_{i'j'k'} \) is computed over a neighborhood of the voxel \((i, j, k)\) where the forward differences involves the variables \( R_{0,ijk}, I_{0,ijk} \) or \( L_{d,ijk} \). Each term in equation (3.23) can be computed analytically, for example,

\[
\frac{\partial FR_{l,ijk}}{\partial L_{xx,ijk}} = R_{0,ijk} e^{-b_1 L_{ijk}^T L_{ijk}^T} \times \frac{\partial b_1}{\partial L_{xx,ijk}} L_{ijk}^T L_{ijk}^T
\]

\[
\frac{\partial FI_{l,ijk}}{\partial L_{xx,ijk}} = I_{0,ijk} e^{-b_1 L_{ijk}^T L_{ijk}^T} \times \frac{\partial b_1}{\partial L_{xx,ijk}} L_{ijk}^T L_{ijk}^T
\]

\[
\frac{\partial b_1}{\partial L_{xx,ijk}} = (2b_{l,xx} L_{xx,ijk} + 2b_{l,xy} L_{xy,ijk} + 2b_{l,xz} L_{xz,ijk})
\]
3.3.4 Line Search

The limited Quasi-newton method computes a search direction in which the minimum or maximum is estimated to lie and yield a one dimension optimization problem as

$$
\min_{\alpha} f(x + \alpha \cdot d)
$$

(3.24)

where \(d\) is computed by algorithm 2 described in the previous subsection. Equation (3.24) needs to be solved by a search procedure where the solution is find by the following update iteration:

$$
x_{k+1} = x_k + \alpha_k \cdot d
$$

There are many line search techniques can can be roughly categorized as line search without derivatives like Fibonacci method and line search with derivatives like polynomial method. As we have an analytical form of derivative, we use a cubic interpolation [43] that is generally the most efficient for continuous functions like our energy function.

3.3.5 Implementation Issues

The constraint in (3.18) is directly related to the standard deviation \(\sigma\) of the noise which can be computed as in Seber [47]. Since there are \(N\) complex measurements and 8 unknown parameters at each voxel (note: \(S_0\) is complex-valued, so it is treated as 2 unknowns), we have

$$
\sigma = \frac{\sum_{l=1}^{N} \|S_l(x) - S_0(x)e^{-b_{l} : L(x)L(x)^{T}}\|^2}{2N - 8}
$$

Initialization is very crucial for nonlinear optimization. In our case, we use the following nonlinear regression with positive definiteness constraint as the initial guess:

$$
\min \mathcal{E}(S_0(x), L(x)) = \sum_{l=1}^{N} \|S_l(x) - S_0(x)e^{-b_{l} : L(x)L(x)^{T}}\|^2
$$

(3.25)
The above minimization is a simple nonlinear least square problem and can be efficiently solved by the Levenberg-Marquardt method [43] using the results of the corresponding linearized least square problem as initial guess.

There are a few practical issues in implementing the augmented Lagrangian method and the Quasi-Newton method, these are settled by using the suggestions in Nocedal and Wright [43] or empirically. For example, in the augmented Lagrangian method (see algorithm 1), we start with a penalty control parameter $\mu = 5.0$, decrease it by a factor of 2 in each step until it is less than 0.01. We also choose $\lambda_0 = 1.0$. Note that the augmented Lagrangian method is quite robust with respect to the choice of $\mu_0$ and $\lambda_0$ since $\mu_0$ will eventually decrease to 0 and $\lambda_0$ approaches the Lagrange multiplier. The final convergence test has two criteria: The subproblem converges and $\mu < 0.01$. As the subproblem is just a standard unconstrained minimization problem, the criteria to check whether it converges or not is achieved using any of the standard criteria in iterative optimization schemes [43] and for the line search, we employ cubic interpolation and Wolfe convergence criterion, see Nocedal and Wright [43] for more details. For the limited memory Quasi-Newton method, we use the last 5 update pairs to update the search direction.

3.4 Experimental Results

In this section, we present three sets of experiments on the application of our direct tensor smoothing and estimation model. One is on complex valued synthetic data sets and the other two are on a complex valued DWI data acquired from a normal rat brain. For the synthetic data example, we compare the results obtained from our estimation procedure with competing methods published in literature.

3.4.1 Complex Valued Synthetic Data

We synthesized an anisotropic tensor field on a 3D lattice of size $32 \times 32 \times 8$. The volume consists of two homogeneous regions with the following values for $S_0$ and
Region 1: $S_0 = 10.0 e^{i\theta}$

$$D = 0.001 \times [0.970 \ 1.751 \ 0.842 \ 0.0 \ 0.0 \ 0.0],$$

Region 2: $S_0 = 8.0 e^{i\theta}$

$$D = 0.001 \times [1.556 \ 1.165 \ 0.842 \ 0.338 \ 0.0 \ 0.0]$$

where the tensor $D$ is depicted as

$$D = [d_{xx}, d_{yy}, d_{zz}, d_{xy}, d_{xz}, d_{yz}]$$

the dominant eigenvector (DEV) of the first region is along the $y$ axis, while that of the second region is in the $xy$ plane and inclined at $60^\circ$ to the $y$ axis. $\theta$ is chosen to be $45^\circ$ to yield an even distribution of the real and the imaginary part.

The complex diffusion weighted images $S_l$ are generated using the Stejskal-Tanner equation at each voxel $x$ given by

$$S_l(x) = S_0(x)e^{-b_l \cdot D(x)} + n(x), \quad (3.26)$$

where $n(x) \sim N(0, \sigma_N) + iN(0, \sigma_N)$, $N(0, \sigma_N)$ is a zero mean Gaussian noise with standard deviation $\sigma_N$. As the signal measured before Fourier transform in MRI is complex, it is reasonable to assume the noise is an additive complex Gaussian noise. The noise in the DWIs remains to be complex valued after the Fourier Transform. Thus our simulated data reflects the physics of MRI imaging. Note that the noise in the magnitude of the complex DWIs have a Rician distribution and approximates a Gaussian distribution when the SNR is high [42]. We choose the 7 commonly used configurations $x, y, z, xy, xz, yz, xyz$ for the directions of the diffusion-encoded magnetic gradient as in Basser et al. [1] and use 3 different field strengths in each direction (100, 500 and 1000s/mm$^2$). Thus we have a total of 21 different $b$ matrix. A slice of the generated data set is shown in Fig. 3.1, note that the DWIs are different
when either the directions or the magnitudes of diffusion-encoded magnetic gradient are different.

For better illustration of the superior performance of our model, we compare performance with the following methods in our experiments: (i) **Linear** - linear regression on (1.6) as used in [1], (ii) **Nonlinear** - nonlinear regression applied to (1.4), (iii) **Linear + EVS** (Eigenvector smoothing) - linear regression followed by the DEV smoothing method described in Coulon et al. [22], (iv) **Nonlinear + EVS** - nonlinear regression plus the smoothing as in (iii), and (v) **Ours**-Our method. Note that the EVS method [22] is a direction field restoration scheme that preserves discontinuities and is based on Chan and Shen’s work [50].

Figure 3.2 shows an ellipsoid visualization of the restored diffusion tensor field for the synthetic data set with $\sigma_N = 0.5$. It is evident that our method restored the noisy tensor field quite well in comparison to the nonlinear regression method which did not perform well and the linear regression technique which performed worst.

For further comparison, Fig. 3.3 shows the DEV computed from the original and the restored diffusion tensor field using all five methods as mentioned before. This figure clearly shows that our model yielded the best estimation of the original DEV field. The method in Coulon et al. [22], however, did not work well at voxels where the estimated DEVs are almost orthogonal to those in their neighborhoods. In such cases, Coulon et al.’s method will treat them as discontinuities and does not smooth them. Though it is possible to treat these locations as outliers, it is difficult to set a reasonable criteria to achieve the same. It is interesting to notice that the Nonlinear+EVS method which serves as an improvement of the existing Linear+EVS method can diminish this problem. Additional quantitative measures, described below, confirm the visual comparison results.

To quantitatively assess the proposed model, we compare the accuracy of the DEV computed using the previously mentioned methods. Let $\theta$ be the angle (in
Table 3.1: Comparison of the accuracy of the estimated DEVs using different methods for different noise levels.

\[ \sigma_n = 0.5 \]

<table>
<thead>
<tr>
<th>( \mu_\theta )</th>
<th>Linear</th>
<th>Nonlinear</th>
<th>Linear+EVS</th>
<th>Nonlinear+EVS</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.57</td>
<td>7.44</td>
<td>1.63</td>
<td>1.19</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>6.93</td>
<td>5.08</td>
<td>1.57</td>
<td>0.84</td>
<td>0.96</td>
<td></td>
</tr>
</tbody>
</table>

\[ \sigma_n = 1.0 \]

<table>
<thead>
<tr>
<th>( \mu_\theta )</th>
<th>Linear</th>
<th>Nonlinear</th>
<th>Linear+EVS</th>
<th>Nonlinear+EVS</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>22.28</td>
<td>16.94</td>
<td>6.67</td>
<td>3.78</td>
<td>1.99</td>
<td></td>
</tr>
<tr>
<td>17.46</td>
<td>13.86</td>
<td>13.06</td>
<td>8.58</td>
<td>2.74</td>
<td></td>
</tr>
</tbody>
</table>

\[ \sigma_n = 1.5 \]

<table>
<thead>
<tr>
<th>( \mu_\theta )</th>
<th>Linear</th>
<th>Nonlinear</th>
<th>Linear+EVS</th>
<th>Nonlinear+EVS</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>33.14</td>
<td>26.40</td>
<td>14.55</td>
<td>9.09</td>
<td>4.08</td>
<td></td>
</tr>
<tr>
<td>22.60</td>
<td>20.19</td>
<td>22.39</td>
<td>17.46</td>
<td>4.70</td>
<td></td>
</tr>
</tbody>
</table>

degrees) between the estimated DEV and the original DEV, Table 3.1 shows the mean \( \mu_\theta \) and standard deviation \( \sigma_\theta \) of \( \theta \) using different methods for the synthetic data with different levels of additive Gaussian noises. A better method is one that yields smaller values. From this table, we can see that our model yields lower error values than all other methods under various noise levels. It is also clear from this table that the methods using the original nonlinear complex Stejskal-Tanner equation (1.4) are more accurate than those using the linearized one (1.6). The advantage of our method and the nonlinear approaches are more apparent when the noise level is higher, which supports our discussion in section 3.2.1.

3.4.2 Complex DWI of a Normal Rat Brain

The normal rat brain data is imaged using a 17.6T (750MHz) Bruker Avance Imaging Spectrometer system with the following settings: \( TR = 3058ms, TE = 28.8ms, \Delta = 17.8ms, \delta = 2.4ms, diffusion time = 17.0ms, bandwidth = 40kHz \). The field of view is \( 15 \times 15 \times 21mm^3 \) with a resolution of \( 117 \times 117 \times 270\mu m^3 \). The same set of 7 diffusion-encoded magnetic directions as the synthetic data are used
with two different magnitudes (100, 500s/mm²). With a number of averages equal to 8 for each signal measurement \( S_i \), the raw data is a set of 14 complex DWI volume data, each with a size of 128 × 128 × 78.

We extract a 128 × 128 × 10 volume in the region of the corpus callosum for our first experiment. Figure 3.4.2 depicts the restored images of the six independent components of the estimated diffusion tensor. As a comparison, Figs. 3.4 and 3.5 show the same images computed using linear regression applied to (1.6) and the nonlinear regression applied to (1.4) from the raw data respectively. For display purposes, we use the same brightness and contrast enhancement for displaying the corresponding images in all the three figures. We also present the computed DEV of the estimated diffusion tensor in Fig. 3.7. We didn’t compare with the EVS methods because the sorting problem of the eigenvectors is very severe in free water or other isotropic regions, thus it is necessary to exclude those regions to make effective usage of EVS methods. This involves a segmentation issue which is non-trivial.

We then extracted a 10×127×78 volume in the region of the cerebellum and show the sagittal view of the results in Fig. 3.8. The brightness and contrast enhancement of the figures are the same as in the previous experiment. In Figs. 3.4.2 and 3.8, the edge preserving smoothing is evident especially in the off diagonal terms of the diffusion tensor \( D \) which are essential in evaluating the structural anisotropy. We also notice that there are some differences in the region of free water between Figs. 3.4 and 3.5 visible in the off-diagonal terms while there is no difference visible inside the corpus callosum between these two figures. However, Fig. 3.7 gives more insight into this via the depiction of the computed DEVs. Note that smoothing effect on DEV is evident in the shadowed box in Fig. 3.7. Our model for estimating a smooth diffusion tensor field from the noisy DWIs may be used to achieve better accuracy in fiber tractography. Our future efforts will focus on applying the model described in this paper to achieve better fiber tract maps.
Figure 3.1: A slice of several volumetric complex DWIs generated with $\sigma_N = 0.5$. Left to right: real and imaginary pairs of complex DWIs with varying magnitude of diffusion encoded magnetic gradient. Top to bottom: complex DWIs for varying directions of diffusion encoded magnetic gradient.
Figure 3.2: A slice of the original (ground truth) and the estimated diffusion tensor fields for the noisy synthetic data with $\sigma_N = 0.5$. 
Figure 3.3: A slice of the computed DEV field for the noisy synthetic data with $\sigma_N = 1.5$. Top left image is the DEV field computed from the original tensor field, and the other images arranged from left to right, top to bottom are the DEV field computed from estimated tensor field using the following methods: linear, nonlinear, linear+EVS, nonlinear+EVS and our model.
Figure 3.4: A slice of the normal rat brain DTI estimated using multivariate linear regression without smoothing, viewed channel by channel. First row, left to right: $D_{xx}$, $D_{xy}$ and $D_{xz}$. Second row, left to right: $S_0$, $D_{yy}$ and $D_{yz}$. Last row, left to right: $FA$, $<D>$ and $D_{zz}$. 
Figure 3.5: A slice of the normal rat brain DTI estimated using multivariate nonlinear regression without smoothing, viewed channel by channel. Arrangement of the figures are the same as in Fig. 3.4.
Figure 3.6: A slice of the normal rat brain DTI estimated using our proposed method, viewed channel by channel. Arrangement of the figures are the same as in Fig. 3.4.
Figure 3.7: A slice of the computed DEV field from a normal rat brain. Top to bottom: Linear, nonlinear regression and our model. Left column: Region of interest for depicting the DEV indicated by the white box superposed on the $D_{xx}$ image. Right column: The computed DEV field inside the white rectangle on the left and the smoothing effect of our model is clearly visible in the shaded box.
Figure 3.8: A slice of the normal rat brain DTIs in the region of the cerebellum viewed channel by channel. The DTIs are estimated using (a) multivariate non-linear regression without smoothing and (b) our proposed method (bottom row). Both (a) and (b) are sagittal views and are arranged as in Fig. 3.4.
CHAPTER 4
DTI SEGMENTATION USING AN INFORMATION THEORETIC TENSOR DISSIMILARITY MEASURE

4.1 Introduction

We tackle the DTI segmentation problem using region-based active contour model by incorporating an information theoretic tensor dissimilarity measure. Geometric active contour models have long been used in scalar and vector image segmentation [51, 52, 53]. Our work is based on the active contour models derived from the Mumford-Shah functional [54], and can be viewed as a significant extension of the work on active contours without edges, by Chan and Vese [55] and the work on curve evolution implementation of the Mumford-Shah functional by Tsai et al. [53] to diffusion tensor images. Our key contributions are, (i) the definition of a new discriminant of tensors based on information theory, (ii) a theorem and its proof that allows for the computation of the mean value of an SPD tensor field required in the active contour model in closed form and facilitates the efficient segmentation of the DTI, (iii) extension of the popular region-based active contour model to handle matrix-valued images, e.g., DTI.

Rest of this chapter is organized as follows: in section 4.2, the new definition of tensor distance is introduced and its properties are discussed in detail. In section 4.3, active contour model for image segmentation is reviewed. Then in section 4.4, the piecewise constant region-based active contour models for DTI segmentation is discussed. We describe the associated Euler-Lagrange equation, the curve evolution equation, the level set formulation and the implementation details. Similarly in section 4.5, we discuss the piecewise smooth region-based active contour models for DTI
segmentation. Finally in section 4.6, we present experiments on application of our model to synthetic tensor fields as well as real DTIs.

4.2 A New Tensor Distance and Its Properties

4.2.1 Definitions

In the context of DT-MRI, diffusion of water molecules may be characterized by a 2-tensor $D$ which is symmetric positive definite. This $D$ is related to the displacement $r$ of water molecules at each lattice point in the volumetric data at time $t$ via

$$p(r|t, D) = \frac{1}{\sqrt{(2\pi)^n|2tD|}} e^{-\frac{r^T D^{-1} r}{4t}}$$  \hspace{1cm} (4.1)

Note that the mean of $r$ is $E(r) = 0$ and the covariance matrix of $r$ is $D(r) = 2tD$ [47]. Thus, it is natural to use the distance measure between Gaussian distributions to induce a distance between these tensors. The most frequently used information theoretic “distance” measure is the Kullback-Leibler divergence defined as

$$KL(p\|q) = \int p(x) \log \frac{p(x)}{q(x)} dx$$  \hspace{1cm} (4.2)

for two given densities $p(x)$ and $q(x)$.

KL divergence is not symmetric and a popular way to symmetrize it is

$$J(p, q) = \frac{1}{2}(KL(p\|q) + KL(q\|p))$$  \hspace{1cm} (4.3)

which is called the J-divergence. We propose a novel definition of tensor “distance” as the square root of the J-divergence, i.e.

$$d(T_1, T_2) = \sqrt{J(p(r|t, T_1), p(r|t, T_2))}$$  \hspace{1cm} (4.4)

It is known that twice the KL divergence and thus twice the J-divergence is the squared distance of two infinitesimally nearby points on a Riemannian manifold of parameterized distributions [56]. Thus taking the square root in equation (4.4) is
justified. Furthermore, equation (4.4) turns out to have a very simple form given by

\[ d(T_1, T_2) = \frac{1}{2} \sqrt{\text{tr}(T_1^{-1}T_2 + T_2^{-1}T_1)} - 2n \] (4.5)

where \( \text{tr}(\cdot) \) is the matrix trace operator, \( n \) is the size of the square matrix \( T_1 \) and \( T_2 \).

**Derivation 1**: For Gaussian distributions \( p(r|t, T_1) \) and \( p(r|t, T_2) \) given as in equation (4.1), we have

\[
KL(p(r|t, T_1) \| p(r|t, T_2)) \\
= \int p(r|t, T_1) \log \frac{p(r|t, T_1)}{p(r|t, T_2)} dr \\
= \int p(r|t, T_1) \log \left[ \sqrt{\frac{|T_2|}{|T_1|}} e^{-\frac{r^T T_1^{-1} r}{4} + \frac{r^T T_2^{-1} r}{4t}} \right] dr \\
= \int p(r|t, T_1) \left[ \frac{1}{2} \log \left| \frac{T_2}{T_1} \right| - \frac{r^T T_1^{-1} r}{4t} + \frac{r^T T_2^{-1} r}{4t} \right] dr \\
= \frac{1}{2} \log \left| \frac{T_2}{T_1} \right| \int p(r|t, T_1) dr - \int p(r|t, T_1) \frac{r^T T_1^{-1} r}{4t} dr + \int p(r|t, T_1) \frac{r^T T_2^{-1} r}{4t} dr \\
= \frac{1}{2} \log \left| \frac{T_2}{T_1} \right| E(r T_2^{-1} r) - E(r T_2^{-1} r) \\
= \frac{1}{2} \log \left| \frac{T_2}{T_1} \right| - \text{tr} \left[ \frac{T_1^{-1}}{4t} (2tT_1) \right] + \text{tr} \left[ \frac{T_2^{-1}}{4t} (2tT_1) \right] \\
= \frac{1}{2} \log \left| \frac{T_2}{T_1} \right| - \text{tr} \left( \frac{I_n}{2} \right) + \text{tr} \left( \frac{T_2^{-1} T_1}{2} \right) \\
= \frac{1}{2} \left[ \log \left| \frac{T_2}{T_1} \right| - n + \text{tr}(T_2^{-1} T_1) \right] \quad (4.6)
\]

Switching the role of \( p(r|t, T_1) \) and \( p(r|t, T_2) \) in the above steps leads to

\[
KL(p(r|t, T_2) \| p(r|t, T_1)) = \frac{1}{2} \left[ \log \left| \frac{T_1}{T_2} \right| - n + \text{tr}(T_1^{-1} T_2) \right] \quad (4.7)
\]
Combined with the definition of J-divergence, we have

\[
J(p(r|t, T_1), p(r|t, T_2)) = 
\frac{1}{2} \left[ KL(p(r|t, T_1)||p(r|t, T_2)) + KL(p(r|t, T_2)||p(r|t, T_1)) \right]
\]

\[
= \frac{1}{4} \left[ \log \frac{|T_2|}{|T_1|} - n + tr(T_2^{-1}T_1) \right] + \frac{1}{4} \left[ \log \frac{|T_1|}{|T_2|} - n + tr(T_1^{-1}T_2) \right]
\]

\[
= \frac{1}{4} \left[ tr(T_2^{-1}T_2 + T_2^{-1}T_1) - 2n \right]
\]

Thus we have equation (4.8).

Now we give an example of the above “distance”. When \( T_1 \) and \( T_2 \) can both be diagonalized using the same orthogonal matrix \( O \), which means \( T_1 = ODO^T \) and \( T_2 = OEO^T \), where \( D = \text{diag}\{d_1, ..., d_n\} \) and \( E = \text{diag}\{e_1, ..., e_n\} \). Then we have

\[
tr(T_1^{-1}T_2) = \sum_{i=1}^{n} \frac{e_i}{d_i}, \quad tr(T_2^{-1}T_1) = \sum_{i=1}^{n} \frac{d_i}{e_i}
\]

Thus

\[
d(T_1, T_2) = \frac{1}{2} \sqrt{\sum_{i=1}^{n} \frac{e_i}{d_i} + \sum_{i=1}^{n} \frac{d_i}{e_i} - 2n} = \frac{1}{2} \sqrt{\sum_{i=1}^{n} \left( \frac{d_i}{e_i} + \frac{e_i}{d_i} - 2 \right)}
\]

\[
= \frac{1}{2} \sqrt{\sum_{i=1}^{n} \left( \frac{d_i^2 + e_i^2 - 2d_i e_i}{e_i d_i} \right) = \frac{1}{2} \sqrt{\sum_{i=1}^{n} \left( \frac{(d_i - e_i)^2}{e_i d_i} \right)}
\]

(4.9)

4.2.2 Affine Invariant Property

When a coordinate system undergoes an affine transformation, the DTI will also be transformed. If the coordinate system undergoes an affine transform \( y = Ax + b \), then the displacement of the water molecules will be transformed as \( \hat{r} = Ar \). Since \( r \) has a Gaussian distribution with covariance matrix \( 2ttT \), the transformed displacement \( \hat{r} \) has a covariance matrix of \( 2ttATA^T \). Thus, the transformed DTI will be

\[
\hat{T}(y) = AT(x)A^T, \quad y = Ax + b
\]

(4.10)
Figure 4.1 shows an example of a $32 \times 32$ 2D diffusion tensor field where the affine transformation is

$$A = \begin{bmatrix} 0.8 & 0.0 \\ 0.0 & 0.8 \end{bmatrix}, \quad b = \begin{bmatrix} 2.0 \\ 2.0 \end{bmatrix}$$

Our definition of tensor “distance” is invariant to such transformations, i.e.

$$d(T_1, T_2) = d(AT_1 A^T, AT_2 A^T) \quad (4.11)$$

Though the transformation of the diffusion tensor is actually a congruent transformation, however, we still refer the above invariance as “affine” invariance because in the context of segmentation the invariance property has been linked with the transformation of the coordinate system.
Derivation 2: First of all, we have

\[ \text{tr} \left[ (AT_1A^T)^{-1}(AT_2A^T) \right] \]
\[ = \text{tr}(A^{-T}T_1^{-1}A^{-1}AT_2A^T) \]
\[ = \text{tr}(A^{-T}T_1^{-1}T_2A^T) \]
\[ = \text{tr}(T_1^{-1}T_2A^T A^{-T}) \]
\[ = \text{tr}(T_1^{-1}T_2) \]  \hspace{1cm} (4.12)

Similarly we get \( \text{tr} \left[ (AT_2A^T)^{-1}(AT_1A^T) \right] = \text{tr}(T_2^{-1}T_1) \), thus we have

\[ d(AT_1A^T, AT_2A^T) \]
\[ = \frac{1}{2} \sqrt{\text{tr} \left[ (AT_1A^T)^{-1}(AT_2A^T) + (AT_2A^T)^{-1}(AT_1A^T) \right] - 2n} \]
\[ = \frac{1}{2} \sqrt{\text{tr}(T_1^{-1}T_2 + T_2^{-1}T_1) - 2n} \]
\[ = d(T_1, T_2) \]  \hspace{1cm} (4.13)

Thus our “distance” is invariant under affine coordinate transformation.  \[ \blacksquare \]

It is easy to show that Frobenius norm of the tensor difference commonly used in published work [28, 27, 33] does not have this property.

4.2.3 Tensor Field Mean Value

We now develop a theorem that allows us to compute the mean value of a tensor field. This is essential for the region-based active contour model used in the segmentation algorithm.

Theorem 2 The mean value of a tensor field defined as

\[ \bar{M}(T, R) = \min_{M \in \text{SPD}(n)} \int_R d^2 [M, T(x)] \, dx \]  \hspace{1cm} (4.14)
is given by

\[ M = \sqrt{B^{-1}} \left[ \sqrt{BA} \sqrt{B} \right] \sqrt{B^{-1}} \]  

(4.15)

where \( A = \int_R T(x)dx \), \( B = \int_R T^{-1}(x)dx \) and \( SPD(n) \) denotes the set of symmetric positive definite matrices of size \( n \).

**Proof 3** Let \( E(M) = \int_R d^2 [M, T(x)] dx \), we have

\[
E(M) = \int_R d^2 [M, T(x)] dx \\
= \int_R \left[ \frac{1}{4} tr(M^{-1}T(x) + T^{-1}(x)M) - \frac{n}{2} R \right] dx \\
= \frac{1}{4} tr \left[ M^{-1} \left( \int_R T(x)dx \right) + \left( \int_R T^{-1}(x)dx \right) M \right] - \frac{n}{2} |R| \\
= \frac{1}{4} tr(M^{-1}A + BM) - \frac{n}{2} |R|
\]

For a small perturbation in the neighborhood \( N(M) \subset SPD(n) \) represented by \( M + tV \), where \( t \) is a small enough positive number, \( V \) is symmetric matrix, we have

\[(M + tV)^{-1} = M^{-1}(I - tVM^{-1} + o(t))\]

Thus

\[
E(M + tV) \\
= \frac{1}{4} tr \left[ (M + tV)^{-1} A + B(M + tV) \right] - \frac{n}{2} |R| \\
= \frac{1}{4} tr \left[ M^{-1}(I - tVM^{-1} + o(t))A + B(M + tV) \right] - \frac{n}{2} |R| \\
= \frac{1}{4} tr(M^{-1}A + BM) - \frac{n}{2} |R| \\
+ \frac{1}{4} tr(BV - M^{-1}VM^{-1}A)t + o(t) \\
= E(M) + \frac{1}{4} tr(BV - M^{-1}AM^{-1}V)t + o(t)
\]

Thus at the critical point, we need

\[
tr(BV - M^{-1}AM^{-1}V) = 0, \quad \forall \text{symmetric } V
\]  

(4.16)
This is actually equivalent to \( \tilde{M}B\tilde{M} = A \) and can be solved in closed form [57] yielding the desired result in equation (4.15). Since \( A \) and \( B \) are both SPD matrices, \( M \) is also an SPD matrix, thus it is indeed a solution to the minimization equation (4.14).

In some special cases, the tensor field mean value can be computed easily. If the tensor field is constant, \( T(x) \equiv T_c \), then we have \( A = T_c|R| \) and \( B = T_c^{-1}|R| \). Substituting into equation (4.15) we have \( M = T_c \). So the mean value of a constant tensor field is the constant itself which makes perfect sense. Now if all the tensors in the tensor field is diagonal, \( T(x) \equiv D(x) = \text{diag}\{d_1(x), ..., d_n(x)\} \), we have

\[
A = \int_R D(x)dx = \int_R \text{diag}\{d_1(x), ..., d_n(x)\}dx
= \text{diag}\{\int_R d_1(x)dx, ..., \int_R d_n(x)dx\}
\]

Similarly

\[
B = \text{diag}\{\int_R \frac{1}{d_1(x)}dx, ..., \int_R \frac{1}{d_n(x)}dx\}
\]

As both \( A \) and \( B \) are diagonal matrices, their polynomial forms are all diagonal and the multiplication between these terms are commutable. Then we have

\[
M = \sqrt{B^{-1}}\left[\sqrt{\sqrt{B}A\sqrt{B}}\right]\sqrt{B^{-1}}
= B^{-\frac{1}{2}}B^{\frac{1}{2}}A^{\frac{1}{2}}B^{\frac{1}{2}}B^{-\frac{1}{2}}
= B^{-\frac{1}{2}}A^{\frac{1}{2}} \quad (= A^{\frac{1}{2}}B^{-\frac{1}{2}})
= \text{diag}\left\{\sqrt{\frac{\int_R d_1(x)dx}{\int_R d_1(x)dx}}, ..., \sqrt{\frac{\int_R d_n(x)dx}{\int_R d_n(x)dx}}\right\}
\]

In general, \( AB \neq BA \) won’t commute and equation (4.15) can not be simplified further and needs to resort to the matrix diagonalization.
4.3 Active Contour Models for Image Segmentation

Active contour models (aka snakes) have been successfully used in numerous applications including computer vision, medical imaging, computer graphics etc. Generally speaking, active contour models deal with hyper-surfaces in $\mathbb{R}^n$ and they are often referred as curve evolution models in 2D and surface propagation models in 3D. The basic idea of active contour models is simple, it is just the movement of a hyper-surface in $\mathbb{R}^n$ (that represents the boundary of regions) in the direction of the normal with an application driven velocity. Level set methods serve as a powerful mathematical tool to analyze and implement the hyper-surface evolution equations in active contour models. They involve expressing the geometric hyper-surface evolution in an Eulerian form allowing the use of efficient and robust numerical methods. In addition, topology change of the hyper-surface is transparent. The time dependent level set formulation was proposed independently by Dervieux and Thomasset [58] to study multifluid incompressible flows, and Osher and Sethian [59] to study front propagation. In the context of computer vision, the history of active contour models began with the pioneering work by Kass, Witkin and Terzopoulos [60]. A few years later, Malladi et al. [61], and Caselles et al. [62] independently proposed the seminal idea of modeling image segmentation with geometric snakes in a level set formulation. There are also many other significant contributions in image segmentation using active contour model and level set methods. A brief review can be found in Vemuri et al. [63]. For a complete discussion on the general curve evolution and level set methods, the readers are referred to two outstanding books, one by Sethian [64] and the other by Osher and Paragios [65]. In this section, only basic concepts and methods of 2D active contour models are presented for the sake of self containedness.

4.3.1 Snake Energy Functional

Most active contour models aim to minimize a snake energy functional. The minimization of a functional is also called a variational principle and offers several
advantages over other methods in image segmentation. The most significant one is that the incorporation of prior shape information into this framework is straightforward. The vast majority of active contour models can be categorized as two types: edge-based snakes and region-based snakes.

The classical edge-based snake was first proposed by Kass et al. [60] and involved minimizing the following energy functional:

$$E(C) = \int_0^1 \left[ E_{\text{int}}(C(p)) + E_{\text{image}}(C(p)) + E_{\text{con}}(C(p)) \right] dp$$

(4.17)

where $C$ is a parameterized curve that separates the different regions, $E_{\text{int}} = \alpha |C_p|^2 + \beta |C_{pp}|^2$ is the internal energy that measures the length and the stiffness of the curve, $E_{\text{image}}$ represents the image force that attracts the curve to locations with large image gradient (as edges), $E_{\text{con}}$ is given by the external constraint forces that usually refine the curve $C$ in the segmentation and is not used in non-interactive segmentation tasks, $\alpha$ and $\beta$ in the internal energy are control parameters that balance the geometric behavior of the snake.

Though the classical snake solves the problem of linking edges to form a segmentation, it has several disadvantages. The most severe one is that it can not converge when far away from the true solution. The balloons model introduced by Cohen [66] incorporates an additional constant inflating or shrinking external force as well as stable numerical techniques to achieve a larger range of convergence. This constant force was shown to be part of an area energy minimization term by K. Siddiqi et al. [67]. The resulting energy functional is

$$E(C) = \int_0^1 \left[ E_{\text{int}}(C(p)) + E_{\text{image}}(C(p)) \right] dp + \gamma \int_{R_i} d\mathbf{x}$$

(4.18)

where $R_i$ is the region inside the boundary $C$.

The geometric active contour models were introduced by Malladi et al. [61, 51] and Caselles et al. [62] in the curve evolution context by adding an image stopping
term, and can also be derived as part of a weighted length energy functional [52]:

\[ E(C) = \int_0^1 g(|\nabla I|)|C'(p)|dp \]  

(4.19)

where \( g(\cdot) \) is an inverse function of the image gradient \( |\nabla I| \).

The above three types of active contour models are all edge based since the stopping criterion in all of them is a function of the image gradient. A more robust type of active contour models is the regions-based model and is based on the Mumford-Shah functional [54]:

\[ E(f, C) = \gamma \int_{\Omega} (f - g)^2 d\mathbf{x} + \alpha \int_{\Omega\setminus C} |\nabla f|^2 d\mathbf{x} + \beta |C| \]  

(4.20)

where \( \Omega \) is the image domain, \( g \) is the original noisy image data, \( f \) is a piecewise smooth approximation of \( g \) with discontinuities only along \( C \), \( |C| \) is the arclength of the curve \( C \), \( \alpha, \beta \) and \( \gamma \) are control parameters. The first term models the data fidelity and \( \gamma \) is inversely proportional to the variance of the noise process, the second term measures the smoothness of the approximation \( f \) and can be viewed as a prior model of \( f \) given the discontinuity at \( C \), the third term aims to remove tiny disconnected regions and keep a smooth object boundary. With all these three terms, the Mumford-Shah functional provides an elegant framework for simultaneous image segmentation and smoothing. However, it is a difficult problem to solve the original Mumford-Shah functional and numerous methods were proposed to tackle its various approximations. For example, Chan and Vese [55] and Tsai et al. [53] presented a two-stage implementation of (4.20) where the first stage involves, for a fixed \( C \), constructing a constant or smooth approximation to the image function inside and outside of \( C \) and the second-stage evolves \( C \) for a fixed \( f \).
4.3.2 Curve Evolution

The exact form of curve evolution is

\[
\frac{\partial C(x, t)}{\partial t} = \alpha(x)T(x) + \beta(x)N(x) \tag{4.21}
\]

where \( \alpha \) and \( \beta \) are the speeds of the curve along the tangential and normal directions respectively, \( T \) is the tangent and \( N \) is the unit outer normal to the curve. However without loss of generality, equation (4.21) can be rewritten as a deformation along the normal only in the form of the following PDE \[68]\:

\[
\frac{\partial C(x, t)}{\partial t} = F(x)N(x) \tag{4.22}
\]

where \( F(x) = \beta(x) \) is the speed of the curve at location \( x \) along the normal and might depend on various factors like local curvature, global properties of the curve etc. In the following text, we will not explicitly specify the domain of a function when it is clear from the context. For example, we will use \( F \) instead of \( F(x) \), \( N \) instead of \( N(x) \) and etc. Particular form of \( F \) is the focus of many current researchers and are usually derived from application related variational principles. With proper design of the speed function, curve evolution can be used in the context of image segmentation where \( F \) is related to the image data and the curve smoothing term.

Though curve evolution can be used directly for various applications, a more natural and elegant way is to derive it from the first principles as gradient flows that minimize the snake energy functional. For example, the gradient flow that minimize (4.19) yields \[52]\:

\[
F = -(gk + \nabla g \cdot N)
\]

For computing the curve evolution equation from region based active contour energy functionals, we refer the readers to Aubert et al. \[69]\ for the shape gradient method.
4.3.3 Level Set Formulation

Though traditional techniques for solving (4.22) including the marker/string methods and the volume-of-fluid techniques [64] are useful under certain circumstance, they have quite a few draw backs. For example, the marker/string methods provide numerical scheme based on parameterized curves and have instability and topological limitations. The volume-of-fluid techniques are not accurate. In contrast, level-set methods [59] offer several advantages at the same time as follows:

1. Topological changes as the evolving curves merge and split are transparent in an implicit formulation.
2. Efficient and accurate numerical methods can yield stable solutions when shocks form.
3. Extension from 2D curve evolution to 3D surface propagation is easy.
4. Can use fast solution techniques such as narrow banding and fast marching for speedup.

The key idea of the level set methods is very simple. Given a higher dimensional function $\phi$ whose zero level set is the curve $C$ (see Fig. 4.2), it is not hard to derive a corresponding update equation for $\phi$ as

$$\frac{\partial \phi(x, t)}{\partial t} = -F|\nabla \phi|$$  \hspace{1cm} (4.23)

The higher dimensional function $\phi$ is usually chosen to be the signed distance function of the curve $C$. In such a case, $|\nabla \phi| = 1$ and will ensure numerical stability.

4.4 Piecewise Constant DTI Segmentation Model

Our model for DTI segmentation in $\mathbb{R}^2$ is posed as minimization of the following variational principle based on the Mumford-Shah functional [54]:

$$E(T, C) = \int_{\Omega} d^2(T(x), T_0(x))dx + \alpha \int_{\Omega \setminus C} |\nabla dT(x)|^2dx + \beta |C|$$ \hspace{1cm} (4.24)
where the curve $C$ is the boundary of the desired unknown segmentation, $\Omega \subset \mathbb{R}^2$ is the image domain, $T_0$ is the original noisy tensor field, $T$ is a piecewise smooth approximation of $T_0$ with discontinuities only along $C$, $|C|$ is the arclength of the curve $C$, $\alpha$ and $\beta$ are control parameters, $dist(.,.)$ is a measure of the distance between two tensors, $\nabla_d T(x) \in \mathbb{R}^2$ is defined as

$$\nabla_d T(x) \cdot v = \lim_{dt \to 0} \frac{dist(T(x), T(x + dtv))}{dt}$$

where $v \in \mathbb{R}^2$ is a unit vector. $\nabla_d T$ can be called the gradient of $T$ under the tensor distance $dist(.,.)$ and its magnitude is a measure of the smoothness of the tensor field $T$ using $dist(.,.)$ as a tensor distance. The extension of the Mumford-Shah functional to 3D is straightforward by replacing the curve $C$ with a surface $S$ and the implementation in 3D is similar as in 2D.

The above variational principle (4.24) will capture piecewise smooth regions while maintaining a smooth boundary, the balance between the smoothness of the DTI in each region and the boundaries is controlled by $\alpha$ and $\beta$. When $\alpha$ is extremely large, equation (4.24) is reduced to a simplified form which aims to capture piecewise
constant regions of two types:

\[
E(C, T_1, T_2) = \int_R \text{dist}^2(T(x), T_1)dx + \int_{R^c} \text{dist}^2(T(x), T_2)dx + \beta |C| \tag{4.25}
\]

where \( R \) is the region enclosed by \( C \) and \( R^c \) is the region outside \( C \), \( T_1 \) and \( T_2 \) are the mean values of the diffusion tensor field in region \( R \) and \( R^c \) respectively.

The above model can be viewed as a modification of the active contour model without edges for scalar valued images by Chan and Vese [55]. It can segment tensor fields with two constant regions (each region type however can have disconnected parts) in a very efficient way. We incorporate our new tensor “distance” in this active contour model to achieve DTI segmentation.

4.4.1 Curve Evolution Equation

The Euler Lagrange equation for the variational principle (4.25) is

\[
\left[ \beta k - d^2(T, T_1) + d^2(T, T_2) \right] N = 0
\]

\[
T_1 = M(T, R), \quad T_2 = M(T, R^c)
\]

where \( k \) is the curvature of the curve \( C \), \( N \) is the outward normal to the curve. When \( T_1 \) and \( T_2 \) are fixed, we have the following curve evolution form for the above equation:

\[
\frac{\partial C}{\partial t} = - \left[ \beta k - d^2(T, T_1(t)) + d^2(T, T_2(t)) \right] N
\]

\[\text{where } T_1 = M(T, R), \quad T_2 = M(T, R^c) \tag{4.26}\]

4.4.2 Level Set Formulation

The curve evolution equation (4.26) can be easily implemented in a level set framework. As shown in section 4.3, we have

\[
\frac{\partial \phi}{\partial t} = \left[ \beta \nabla \cdot \frac{\nabla \phi}{|\nabla \phi|} - d^2(T, T_1) + d^2(T, T_2) \right] |\nabla \phi| \tag{4.27}
\]
where \( \phi \) is the signed distance function of \( C \).

### 4.4.3 Implementation and Numerical Methods

We follow the two stage implementation as described in Chan and Vese [55] and our modified procedure is as follows:

**Algorithm 3** Two Stage Piecewise Constant Segmentation of DTIs

1: Set initial curve \( C_0 \) and compute its signed distance function \( \phi_0 \).
2: Compute \( T_1 \) and \( T_2 \) according to equation (4.15).
3: Update signed distance function \( \phi \) according to equation (4.27).
4: Reinitialize \( \phi \) using the updated zero level set.
5: Stop if the solution is achieved, else go to step 2.

The key to the computation of \( T_1 \) and \( T_2 \) is the computation the square root of an SPD matrix. We use the matrix diagonalization to achieve this, a real symmetric matrix \( A \) can be diagonalized as

\[
A = OD^T \quad O^T
\]

where \( O \) is an orthogonal matrix and \( D = diag\{d_1, d_2, ..., d_n\} \) is a diagonal matrix. Then the polynomial form of \( A \) is

\[
A^\alpha = OD^\alpha O^T
\]

where \( D^\alpha = diag\{d_1^\alpha, d_2^\alpha, ..., d_n^\alpha\} \). Note that in equation (4.15), \( \sqrt{BA\sqrt{B}} \neq B^{\frac{1}{2}}A^\frac{1}{2}B^{\frac{1}{2}} \), it has to be computed as follows:

**Algorithm 4** Computing \( \sqrt{BA\sqrt{B}} \)

1: Diagonalize \( B \) as \( O_B D_B O_B^T \).
2: Compute \( P = \sqrt{B} \) as \( O_B \sqrt{D_B} O_B^T \).
3: Compute \( Q \) as \( PAP \).
4: Diagonalize \( Q \) as \( O_Q D_Q O_Q^T \).
5: Compute \( \sqrt{Q} \) as \( O_Q \sqrt{D_Q} O_Q^T \).

Equation (4.27) can be easily discretized using an explicit Euler scheme. We can assume the spatial grid size to be 1, then the finite differences of the partial
derivatives are
\[
\Delta^i \phi_{i,j} = \frac{1}{2}(\phi_{i+1,j} - \phi_{i-1,j}), \quad \Delta^j \phi_{i,j} = \frac{1}{2}(\phi_{i,j+1} - \phi_{i,j-1})
\]
\[
\Delta^i \phi_{i,j} = \phi_{i,j} - \phi_{i-1,j}, \quad \Delta^j \phi_{i,j} = \phi_{i,j} - \phi_{i,j-1}
\]
\[
\Delta^i \phi_{i,j} = \phi_{i+1,j} - \phi_{i,j}, \quad \Delta^j \phi_{i,j} = \phi_{i,j+1} - \phi_{i,j}
\]
\[
\Delta^i \phi_{i,j} = \phi_{i+1,j} - 2\phi_{i,j} + \phi_{i,j-1}
\]
\[
\Delta^j \phi_{i,j} = \frac{1}{4}(\phi_{i+1,j+1} - \phi_{i+1,j-1} - \phi_{i-1,j+1} + \phi_{i-1,j-1})
\]
\[
\Delta^j \phi_{i,j} = \phi_{i,j+1} - 2\phi_{i,j} + \phi_{i,j-1}
\]

In this case, we have the following update equation:
\[
\frac{\phi_{i,j}^{n+1} - \phi_{i,j}^n}{\Delta t} = \left[ \beta k_{i,j}^n - d^2(T_{i,j}, T_1^n) + d^2(T_{i,j}, T_2^n) \right] 
\]
\[
\times \sqrt{(\Delta^i \phi_{i,j})^2 + (\Delta^j \phi_{i,j})^2}
\]
(4.28)

where the curvature \( k_{i,j}^n \) of \( \phi^n \) can be computed as
\[
k_{i,j}^n = \frac{\Delta^j \phi_{i,j}^n(\Delta^i \phi_{i,j})^2 - 2\Delta^i \phi_{i,j}^n \Delta^j \phi_{i,j}^n \Delta^j \phi_{i,j}^n + \Delta^j \phi_{i,j}^n (\Delta^j \phi_{i,j})^2}{[(\Delta^i \phi_{i,j})^2 + (\Delta^j \phi_{i,j})^2]^{3/2}}
\]

Update according to equation (4.28) on the whole domain \( \Omega \) has a complexity of \( O(|\Omega|) \) and is going to be slow when the the domain is huge. As we are most interested in the evolving zero level set, updating only a narrow band around the zero level set will be sufficient and this can be achieved using the narrow band method described in [70, 71]. In order to maintain \( \phi \) as a signed distance function of \( C \), it is necessary to reinitialize \( \phi \) and can also be done only within a narrow band. There are also several other efficient numerical schemes that one may employ for example the multigrid scheme as was done in Tsai et al. [53]. At this time, our explicit Euler scheme with the narrow band method yielded reasonably fast solutions (3-5 secs. for the 2D synthetic data examples and several minutes for the 3D real DTI examples on a 1Ghz Pentium-3 CPU).
4.5 Piecewise Smooth DTI Segmentation Model

In certain cases, the piecewise constant assumption will be violated and the piecewise smooth model (4.24) has to be employed in such cases. Following the curve evolution implementation of the Mumford-Shah functional by Tsai et al. [53], we have the following two-stage scheme. In the smoothing stage, the curve is fixed and a smoothing inside the curve and outside the curve is done by preserving the discontinuity across the curve. In the curve evolution stage, the inside and outside of the smoothed tensor field are fixed while the curve is allowed to move.

4.5.1 Discontinuity Preserving Smoothing

If the curve is fixed, we have

\[ E_C(T) = \int_{\Omega} d^2(T(x), T_0(x)) dx + \alpha \int_{\Omega/C} |\nabla_d T(x)|^2 dx \]  

(4.29)

For implementation, the above VP is discretized as follows:

\[ E_C(T) = \sum_x d^2(T(x), T_0(x)) + \alpha \sum_{(x,y) \in N_C} d^2(T(x), T(y)) \]  

(4.30)

where \( N_C \) defines a collection of neighboring pixels. If a pair \((x,y)\) cuts across the boundary, it is excluded from \( N_C \).

As we have an energy functional of a tensor field on a discrete grid, we can compute its gradient with respect to this discrete tensor field. A straightforward way to do this is to treat all the independent components of the tensors as the components of a vector and compute the gradient of this energy function with respect to this vector. However, the form of the gradient will not be compact. Instead, we use the derivative of a matrix function \( f(A) \) with respect to its matrix variable as follows:

\[ \frac{\partial f(A)}{\partial A} = \begin{bmatrix} \frac{\partial f(A)}{\partial A(i,j)} \end{bmatrix} \]  

(4.31)
It is not hard to see that
\[
\frac{\partial f(A)}{\partial A(i,j)} = \lim_{dt \to 0} \frac{f(A + dtE_{ij}) - f(A)}{dt}
\]  
(4.32)
where \(E_{ij}\) is a matrix with 1 at location \((i,j)\) and 0 elsewhere.

As the directional derivative with respect to a perturbation \(V\) on \(T(x)\) is given by
\[
E_C(T(x) + V) - E_C(T(x))
\]
\[
= \left[ d^2(T(x) + V, T_0(x)) + \alpha \sum_{y \in N_C(x)} d^2(T(x) + V, T(y)) \right]
- \left[ d^2(T(x), T_0(x)) + \alpha \sum_{y \in N_C(x)} d^2(T(x), T(y)) \right]
\]
\[
= \frac{1}{4} tr \left[ (T^{-1}_0(x) - T^{-1}(x)T_0(x)T^{-1}(x))V \right]
\]
\[
+ \alpha \sum_{y \in N_C(x)} tr \left[ (T^{-1}(y) - T^{-1}(x)T(y)T^{-1}(x))V \right]
\]
\[
= \frac{1}{4} tr \left[ (B - T^{-1}(x)AT^{-1}(x))V \right]
\]  
(4.33)
where
\[
A = \alpha \sum_{y \in N_C(x)} T^{-1}(y) + T^{-1}_0(x)
\]
\[
B = \alpha \sum_{y \in N_C(x)} T(y) + T_0(x)
\]

We have the gradient from equation (4.32) and equation (4.33):
\[
\frac{\partial E_C}{\partial T(x)} = \frac{1}{4} [B - T^{-1}(x)AT^{-1}(x)]
\]  
(4.34)
So the minimizer of the VP (4.30) satisfies
\[
B = T^{-1}(x)AT^{-1}(x)
\]  
(4.35)
4.5.2 Curve Evolution Equation and Level Set Formulation

Once the boundary discontinuity preserving smoothed tensor field is fixed, we have

\[ E_T(C) = \int_R d^2(T_R(x), T_0(x))dx + \int_{R^c} d^2(T_{R^c}(x), T_0(x))dx \]
\[ + \alpha \int_R |\nabla dT_R(x)|^2 dx + \alpha \int_{R^c} |\nabla dT_{R^c}(x)|^2 dx + \beta |C| \quad (4.36) \]

Thus the curve evolution equation will be

\[ \frac{\partial C}{\partial t} = \{ [d^2(T_R, T_0) - d^2(T_{R^c}, T_0)] + \alpha (|\nabla dT_{R^c}|^2 - |\nabla dT_R|^2) - \beta k \} N \quad (4.37) \]

Again for implementation, we have

\[ \frac{\partial C}{\partial t} = [d^2(T_R, T_0) - d^2(T_{R^c}, T_0)] N \]
\[ + \alpha \left[ \sum_{\mathbf{y} \in N_{R^c}(x)} d^2(T_{R^c}, T_{R^c}(\mathbf{y})) - \sum_{\mathbf{y} \in N_R(x)} d^2(T_R, T_R(\mathbf{y})) \right] N \]
\[ - \beta k N \quad (4.38) \]

The level set formulation for (4.38) is then given by

\[ \frac{\partial \phi}{\partial t} = \left[ \beta \nabla \cdot \frac{\nabla \phi}{|\nabla \phi|} + F \right] |\nabla \phi| \quad (4.39) \]

where \( \phi \) is the signed distance function of \( C \) and the data dependent speed \( F \) is given by

\[ F = - [d^2(T_R, T_0) - d^2(T_{R^c}, T_0)] \]
\[ - \alpha \left[ \sum_{\mathbf{y} \in N_{R^c}(x)} d^2(T_{R^c}, T_{R^c}(\mathbf{y})) - \sum_{\mathbf{y} \in N_R(x)} d^2(T_R, T_R(\mathbf{y})) \right] \]

4.5.3 Implementation and Numerical Methods

Similarly as in algorithm 3 and also as in Tsai et al. [53], we have the following two-stage implementation. The major difference here lies in the computation of \( T_R \).
Algorithm 5 Two Stage Piecewise Smooth Segmentation for DTI

1: Set initial curve $C_0$ and compute its signed distance function as $\phi_0$.
2: Compute $T_R$ and $T_{R^e}$.
3: Update level set $\phi$ according to equation (4.39).
4: Reinitialize $\phi$ using the updated zero level set.
5: Stop if the solution is achieved. Otherwise repeat from step 2.

and $T_{R^e}$. As the gradient can be computed, it is easy to design efficient numerical algorithm to achieve the discontinuity preserving smoothing. Currently, we use gradient descent with adaptive step size due to its simplicity.
4.6 Experimental Results

In this section, we present several sets of experiments on the application of our DTI segmentation model. The first one is on 2D synthetic data sets, the second one is on single slices of real DTIs estimated from DWIs, the last one is on 3D real DTIs. In the experiments below, if not explicitly stated, the segmentation model used is the piecewise constant model.

4.6.1 Synthetic Tensor Field Segmentation

The purpose of these experiments is to demonstrate the need to use the full information contained in the tensors for segmentation purposes. To this end, we synthesize two tensor fields, both are $2 \times 2$ symmetric positive definite matrix valued images on a $128 \times 128$ lattice and have two homogeneous regions. The two regions in the first tensor field only differ in the orientations while the two regions in the second tensor field only differ in the scales. These two tensor fields are visualized as ellipses as shown in Fig. 4.3(a) and Fig. 4.4(a) respectively. Each ellipse’s axes correspond to the eigenvector directions of the diffusion tensor and are scaled by the eigenvalues. With an arbitrary initialization of the geometric active contour, our proposed model can yield high quality segmentation results as show in Figs. 4.3 and 4.4. The evolving boundaries of the segmentation are shown as curves in red. Note that the first tensor field can not be segmented by using only the scalar anisotropic properties of tensors as in [36] and the second tensor field can not be segmented by using only the dominant eigenvectors of the tensors. These two examples show that one must use the full information contained in tensors in achieving quality segmentation. As our model is a region based segmentation method, it is more resistant to noise than the gradient-based snakes. Figures 4.5 and 4.6 depict the segmentation process for synthetic noisy tensor fields and the results are very satisfactory.
4.6.2 Single Slice DTI Segmentation

Figure 4.7 shows a slice of the DTI estimated from the DWIs of a normal rat spinal cord. Each of the six independent components of the individual symmetric positive definite diffusion tensors in the DTI is shown as a scalar image in the top row. The arrangements of the components from left to right are - $D_{xx}$, $D_{yy}$, $D_{zz}$, $D_{xy}$, $D_{yz}$ and $D_{xz}$. The off diagonal terms are greatly enhanced by brightness and contrast factors for better visualization. An ellipsoid visualization of same slice as the top row is shown Fig. 4.7 (g). Each ellipsoid’s axes correspond to the eigenvector directions of the 3x3 diffusion tensor and are scaled by the eigenvalues and its color from blue to red shows the lowest to highest degree of anisotropy. For example, diffusion tensors in free water region are shown as large round blue ellipsoids. Figure 4.8 demonstrates the separation of the gray matter and white matter inside the normal rat spinal cord with the evolving curve in red superimposed on the ellipsoid visualization of the DTI. Similarly, Fig. 4.9 shows a slice of DTI of a normal rat brain in the region of corpus callosum and Fig. 4.10 depicts the segmentation procedure of the corpus callosum. In the final step, the essential part of the corpus callosum is captured by the proposed piecewise constant segmentation model. To further get the bending horns of the corpus callosum, we use the segmentation results of the piecewise constant model as initialization and apply the piecewise smooth model. The result is shown in Fig. 4.11. Now we have a much better refinement. In all the above experiments, we exclude the free water regions which are not of interest in the biological context.

4.6.3 DTI Segmentation in 3D

Now we demonstrate some segmentation results for 3D DTI. Figure 4.12 depicts the segmentation procedure for a normal rat spinal cord DTI of size $108 \times 108 \times 10$. Figure 4.12 (a)-(d) clearly depict the surface evolution in 3D and Fig. 4.12 (e)-(h) depict the intersection of the propagating surface in (a)-(d) with a slice of the $D_{xx}$ component of the DTI. The separation of the gray matter and white matter inside
the normal rat spinal cord is achieved with ease. Similarly, Fig. 4.13 (a)-(h) show the segmentation procedure for a normal rat brain DTI of size $114 \times 108 \times 12$. In addition, intersections of the final 3D segmentation with different slices of the $D_{xx}$ component of the DTI are shown in Fig. 4.13 (i)-(l). It is evident that a majority of the corpus callosum inside this volume is captured. Again, we exclude the free water regions which are not of interest in the biological context.
Figure 4.3: Segmentation of a synthetic tensor field where two regions differ only in the orientations, (b),(c) and (d) are the initial, intermediate and final steps of the curve evolution process in the segmentation.
Figure 4.4: Segmentation of a synthetic tensor field where two regions differ only in scale, (b), (c) and (d) are the initial, intermediate and final steps of the curve evolution process in the segmentation.
Figure 4.5: Segmentation of a synthetic tensor field with two roughly homogeneous regions that only differ in the orientations. (b)-(d) are the initial, intermediate and final steps of the curve evolution process for segmentation of (a).
Figure 4.6: Segmentation of a synthetic tensor field with two roughly homogeneous regions that only differ scale. (b)-(d) are the initial, intermediate and final steps of the curve evolution process for segmentation of (a).
Figure 4.7: A slice of the DTI of a normal rat spinal cord. (a)-(f): viewed channel by channel as gray scale images. (g): viewed using ellipsoids.
Figure 4.8: Segmentation of the slice of DTI shown in Fig. 4.7. (a)-(d): initial, intermediate and final steps in separating the gray and white matter inside the rat spinal cord.
Figure 4.9: A slice of the DTI of a normal rat brain. (a)-(f): viewed channel by channel as gray scale images. (g): viewed using ellipsoids.
Figure 4.10: Segmentation of the corpus callosum from a slice of DTI shown in Fig. 4.9. (a)-(d): initial, intermediate and final steps in segmenting the corpus callosum.
Figure 4.11: Segmentation of corpus callosum from a the slice of the DTI in Fig. 4.9 using piecewise smooth model. (a)-(b): initial and final steps in separating the corpus callosum.

Figure 4.12: 3D Segmentation of the DTI of a normal rat spinal cord. (a)-(d): initial, intermediate and final steps in separating the gray and white matter inside the rat spinal cord. (e)-(h): a 2D slice of the corresponding evolving segmentation in (a)-(d) superimposed on the $D_{xx}$ component.
Figure 4.13: 3D Segmentation of the corpus callosum from the DTI of a normal rat brain. (a)-(d): initial, intermediate and final steps in separating the corpus callosum. (e)-(h): a 2D slice of the corresponding evolving 3D segmentation in (a)-(d) superimposed on the $D_{xx}$ component. (i)-(l): different 2D slices of the final segmentation superimposed on the $D_{xx}$ component.
CHAPTER 5
CONCLUSIONS

5.1 Optimal b Values

We first proposed a weighted linear estimator to analyze the statistical properties of a nonlinear estimator of ADC. This allows us to derive a closed form expression to approximate the variance of the nonlinear estimator. Our method was validated later by simulation experiments. Next, we showed that the ground truth of ADC is often a distribution instead of a simple interval of values or just a single value and proposed to use weighted COV as a criteria for determining the optimal diffusion weighting factors. Application of this approach to human brains using uniform distribution yield results compatible with previous work, though ours are more precise as we use a more accurate numerical integration step. In addition, we applied our approach to the case of a normal rat brain with the real distributions obtained from segmented regions of gray and white matter and computed the optimal diffusion weighting factors that are not provided in any of the existing literature.

5.2 DTI Restoration

We presented a novel constrained variational principle formulation for simultaneous smoothing and estimation of the positive definite diffusion tensor field from complex diffusion weighted images (DWI). To our knowledge, this is the first attempt at simultaneous smoothing and estimation of the positive definite diffusion tensor field from the complex DWI data. We used the Cholesky decomposition to incorporate the positive definiteness constraint on the diffusion tensor to be estimated. The constrained variational principle formulation is transformed into a sequence of
unconstrained problems using the augmented Lagrangian technique and solved numerically. Proof of the existence of a solution for the minimization problem posed in the augmented Lagrangian framework is presented.

Results of comparison between our method and a representative [22] from the competing schemes are shown for synthetic data under a variety of situations involving the use of linearized and nonlinear data acquisition models depicting the influence of the choice of the data acquisition model on the estimation. It was concluded that using the complex nonlinear data model yields better accuracy in comparison to the log-linearized model. Also, superior performance of our method in estimating the tensor field over the results of comparison between our method and a representative [22] from the competing schemes are shown for synthetic data under a variety of situations involving the use of linearized and nonlinear data acquisition models depicting the influence of the choice of the data acquisition model on the estimation. It was concluded that using the complex nonlinear data model yields better accuracy in comparison to the log-linearized model. Also, superior performance of our method in estimating the tensor field over the chosen competing method was demonstrated for the synthetic data experiment. The estimated diffusion tensors are quite smooth without loss of essential features when inspected visually via the use of ellipsoid visualization as well as principal diffusion direction (PDD) field visualization. The superior performance is borne out via a quantitative statistical comparison of the angle between estimated PDD and ground truth PDD. Additional quantitative validation may be performed by comparing the estimated fiber tracts from the smooth tensor field obtained here to those obtained from histology as was done in our earlier work [6] and will be the focus of our future efforts.

Though the presented work focuses on Diffusion Tensor imaging, the constrained variational principle and the applied numerical methods can be easily tailored to processing the high angular resolution diffusion imaging (HARDI) data and other image
datasets. Our ongoing efforts are in fact focussed on application of the constrained variational principle formulation presented here to HARDI data sets. Three significant advantages of our approach will carry over to any application of our method, (i) a unified variational framework for direct smoothing and estimation, (ii) no ad-hoc methods of choosing the smoothing control parameters and (iii) efficient numerical methods applied in solving the nonlinear large scale problem. Finally, in the context of tensor field estimation, the Cholesky decomposition method employed here to maintain the positive definiteness of the diffusion tensor provides an easy and effective way to satisfy such a constraint if required in other tensor field processing applications.

5.3 DTI Segmentation

We presented a novel tensor field segmentation method by incorporating a new discriminant for tensors into the region based active contour models [55, 53]. The particular discriminant we employed is based on information theory which offers several advantages: It naturally follows from the physical phenomenon of diffusion, is affine invariant and is computationally tractable. The computational tractability follows from a theorem that we proved which allows for the computation of the mean of the tensor field in closed form. By using a discriminant on tensors, as opposed to either the eigenvalues or the eigenvectors of these tensors, we make full use of all the information contained in tensors. This proposed model is then implemented in a level set framework to take advantage of the easy ability of this framework to change topologies when desired. Our approach was applied to synthetic and real diffusion tensor field segmentation. The experimental results are very promising, essential part of the regions are well captured and topological changes are handled naturally.

The tensor “distance” we used is not a true distance as it violates the triangle inequality. We could use a true tensor distance (e.g., Rao’s distance [72]) that is defined as the geodesic distance between Gaussian distributions in an information
geometry framework. However, the computation difficulties associated with such a
tensor distance in the context of segmentation remain to be investigated. The active
model can be extended by incorporating shape statistics to improve the robustness
in situations when the object of interest is partially occluded. In addition, we can
apply similar ideas to other tensor fields like the structure tensor field for texture
image segmentation. These will be our future direction of research.
REFERENCES


BIOGRAPHICAL SKETCH

Zhizhou Wang was born in Wu Wei, AnHui, P. R. China. He received his Bachelor of Science degree from the Special Class for Gifted Young at the University of Science and Technology of China, P. R. China, in 1996. He earned his Master of Science degree from the Institute of Software, Chinese Academy of Science, in 1999. He received his Doctor of Philosophy degree in computer engineering from University of Florida, Gainesville, in August 2004. His research interests include medical imaging, computer vision, scientific visualization and computer graphics.