Estimation and Processing of Ensemble Average Propagator and Its Features in Diffusion MRI

Jian Cheng

PhD THESIS

prepared at
INRIA Sophia-Antipolis Méditerranée

and presented at the
University of Nice-Sophia Antipolis
Graduate School of Information and Communication Sciences

A dissertation submitted in partial fulfillment
of the requirements for the degree of

DOCTOR OF SCIENCE
Specialized in Control, Signal and Image Processing

Estimation and Processing of Ensemble Average Propagator and Its Features in Diffusion MRI

Jian CHENG

Advisers	Dr. Rachid Deriche	INRIA Sophia Antipolis, France
	Dr. Tianzi Jiang	Institute of Automation, CAS, China
Reviewers	Dr. Maxime Descoteaux	Sherbrooke University, Canada
	Dr. Hongtu Zhu	UNC-Chapel Hill, USA
	Dr. Gaolang Gong	Beijing Normal University, China
Examiners	Dr. Théodore Papadopoulo	INRIA Sophia Antipolis, France
	Dr. Yong Fan	Institute of Automation, CAS, China
Invited Members	Dr. Shan Yu	LIAMA, INRIA, France
	Dr. Rong Xue	Institute of Biophysics, CAS, China

UNIVERSITÉ NICE-SOPHIA ANTIPOLIS - UFR Sciences

École Doctorale STIC
(Sciences et Technologies de l'Information et de la Communication)

THÈSE

pour obtenir le titre de
DOCTEUR EN SCIENCES
de l'UNIVERSITÉ de Nice-Sophia Antipolis
Discipline: Automatique, Traitement du Signal et des Images

présentée et soutenue par

Jian CHENG

Estimation et Traitement du Propagateur et de ses Caractéristiques en IRM de diffusion

Thèse dirigée par Dr. Rachid DERICHE et Dr. Tianzi JIANG

Date prévue de soutenance, 30 mai 2012

Composition du jury:

Rapporteurs	Dr. Maxime Descoteaux	Sherbrooke University, Canada
	Dr. Hongtu Zhu	UNC-Chapel Hill, USA
	Dr. Gaolang Gong	Beijing Normal University, China
Examinateurs	Dr. Théodore Papadopoulo	INRIA Sophia Antipolis, France
	Dr. Yong Fan	Institute of Automation, CAS, China
Membres invités	Dr. Shan Yu	LIAMA, INRIA, France
	Dr. Rong Xue	Institute of Biophysics, CAS, China

To my wife Ge-Li HU,
father Xi-Ming CHENG, mother Shu-Jiao LIU

Contents

List of Figures

List of Tables

Symbols

Numbers in italic indicate the section number, numbers in bold indicate page numbers where the main definition occurs.

Notation	Description	Index
$\mathbb{R}^d$	d dimentional Euclidean space	*1.0*, *3.1*, *4.3*, **19**
$\mathbb{S}^{d-1}$	$d-1$ dimensional sphere in $\mathbb{R}^d$	*1.0*, *2.5*, *3.1*, *4.3*, **19**
$\mathbb{L}^2(\chi, w(\boldsymbol{x}))$	space of square integrable functions in domain χ with the inner product based on weight $w(\boldsymbol{x})$	**14**
	$\mathbb{L}^2(\chi, w(\boldsymbol{x}))$ when $\chi = [a, b]$	*5.7*, **14**
	$\mathbb{L}^2(\chi, w(\boldsymbol{x}))$ when $\chi = \mathbb{S}^2$ and $w(x) = 1$	*2.5*, **22**
	$\mathbb{L}^2(\chi, w(\boldsymbol{x}))$ when $\chi = \mathbb{S}^1$ and $w(x) = 1$	**24**
	$\mathbb{L}^2(\chi, w(\boldsymbol{x}))$ when $\chi = (-\infty, \infty)$ and $w(x) = e^{-x^2}$	*2.7*, **28**
	$\mathbb{L}^2(\chi, w(\boldsymbol{x}))$ when $\chi = [0, \infty)$ and $w(x) = x^\alpha e^{-x}$	*2.7*, **29**
	$\mathbb{L}^2(\chi, w(\boldsymbol{x}))$ when $\chi = [0, \infty)$ and $w(x) = x^2$	*5.1*, **30**
	$\mathbb{L}^2(\chi, w(\boldsymbol{x}))$ when $\chi = \mathbb{R}^3$ and $w(x) = 1$	*5.1*, *6.1*, **30**
	$\mathbb{L}^2(\chi, w(\boldsymbol{x}))$ when $w(x) = 1$	**178**
$P_n^m(x)$	associated Legendre polynomial	**21**
	Legendre polynomial	*2.5*, **16**
$H_n(x)$	Hermite polynomial	*2.5*, **16**
$L_n^\alpha(x)$	Associated Laguerre polynomials	*2.2*, *2.5*, *2.6*, **16**
$\mathscr{F}\{\cdot\}$	$\hat{f}(\xi) = \mathscr{F}\{f(x)\}(\xi)$ is the Fourier transform of $f(x)$	*2.3*, **65**
$N(\boldsymbol{x}\|\boldsymbol{\mu}, \Sigma)$	Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance Σ	*2.3*, *3.3*, *4.2*, *4.7*, *5.1*, *6.1*, *6.6*, **63**, *A.0*
	$N(\boldsymbol{x}\|\boldsymbol{\mu}, \Sigma)$ with mean $\boldsymbol{\mu} = 0$	*3.3*, *4.4*, *5.8*, *6.1*, *6.2*, **76**
$\mathcal{P}_N(\mathbb{R}^d)$	space of polynomials in $\mathbb{R}^d$ with orders no more than N	*5.3*, **19**
$\mathcal{HP}_n(\mathbb{R}^d)$	space of homogeneous polynomials with order n in $\mathbb{R}^d$	**19**
$\binom{a}{k}$	binomial coefficient	*2.4*, *5.1*, **29**
$\mathcal{P}_N(\mathbb{S}^{d-1})$	restriction of $\mathcal{P}_N(\mathbb{R}^d)$ in $\mathbb{S}^{d-1}$	**19**
$\mathcal{HP}_n(\mathbb{S}^{d-1})$	restriction of $\mathcal{HP}_n(\mathbb{R}^d)$ in $\mathbb{S}^{d-1}$	**19**
$\bigoplus$	direct sum of Hilbert spaces	**19**
Δ	Laplace (aka Laplace-Beltrami) operator	*2.5*, *4.3*, **19**
$\mathcal{H}_n(\mathbb{R}^d)$	space of homogeneous harmonic polynomials with order n in $\mathbb{R}^d$	**20**

Notation	Description	Index
$\mathcal{H}_n(\mathbb{S}^{d-1})$	restriction of $\mathcal{HP}_n(\mathbb{R}^d)$ in $\mathbb{S}^{d-1}$	**20**
Δ_b	Laplace-Beltrami operator restricted in $\mathbb{S}^2$	*4.3*, **21**
y_l^m	complex Spherical Harmonic	**22**
Y_l^m	real Spherical Harmonic	*2.7*, *4.3*, *6.3*, **22**
$W_{ll'\alpha}^{mm'\beta}$	integration of $y_l^m y_{l'}^{m'} y_\alpha^\beta$ in $\mathbb{S}^2$	*6.2*, **25**
$Q_{ll'\alpha}^{mm'\beta}$	integration of $Y_l^m Y_{l'}^{m'} Y_\alpha^\beta$ in $\mathbb{S}^2$	*6.2*, **25**
$n!!$	$\prod_{j=0}^{[\frac{n-1}{2}]}(n-2j)$ for integer $n>0$ and 1 for $n=0,-1$	*2.6*, **25**
$J_\alpha(x)$	Bessel function	*2.6*, *4.3*, *6.2*, **26**
$j_n(x)$	Spherical Bessel function	*6.2*, **26**
$\Gamma(x)$	Gamma function	*2.5*, **31**
${}_nF_m$	generalized hypergeometric function	**31**
	ordinary hypergeometric function	**30**
${}_1F_1(a;c;x)$	confluent hypergeometric function	**31**
$\gamma(s,x)$	lower incomplete Gamma function	**31**
Sym_K^+	the space of $K\times K$ symmetric positive definite matrix	**55**
	the space of 3×3 symmetric positive definite matrix	*3.0*, *3.1*, *4.4*, *5.5*, **70**
$B(p,R)$	open ball in Euclidean space, i.e. $B(p,R)=\{x\in\mathbb{R}^d:\ \|p-x\|<R\}$	*3.1*, **37**
Sym_K	the space of $K\times K$ symmetric matrix	**56**
	the space of 3×3 symmetric matrix	**38**
$\mathbf{C}^k$	k times differentiable	**39**
$B_\mathcal{X}(p,R)$	open ball in Riemannian manifold $\mathcal{X}$, i.e. $B_\mathcal{X}(p,R)=\{x\in\mathcal{X}:\ d(p,x)<R\}$	*3.2*, **46**
$E_\chi(f(x))$	expectation of $f(x)$ over $x\in\chi$	*5.8*, **48**
$\mathbf{q}$	$\mathbf{q}$ vector in $\mathbf{q}$-space, $\mathbf{q}=q\mathbf{u}$, $\mathbf{u}$ is unit vector, $q=\|\mathbf{q}\|$	*6.1*, **68**
$\mathbf{R}$	$\mathbf{R}$ vector in $\mathbf{R}$-space, $\mathbf{R}=R\mathbf{r}$, $\mathbf{r}$ is unit vector, $R=\|\mathbf{R}\|$	*5.1*, *6.1*, **68**
$\Phi(\mathbf{r})$	ODF, defined in the sphere of $\mathbf{R}$-space	
	ODF by Wedeen with solid angle, $\Phi_t(\mathbf{r})\stackrel{\text{def}}{=}\int_0^\infty P(\mathbf{R})R^2\mathrm{d}R$	*5.1*, *6.2*, **80**
	ODF by Tuch, $\Phi_t(\mathbf{r})\stackrel{\text{def}}{=}\frac{1}{Z}\int_0^\infty P(\mathbf{R})\mathrm{d}R$	*6.2*, **83**
	k-order radial mode of EAP, $\Phi_k(\mathbf{r})=\frac{1}{Z}\int_0^\infty P(\mathbf{R})R^k\mathrm{d}R$	*5.8*, **90**
$\Pi_\mathbf{r}$	the plane orthogonal to $\mathbf{r}$, i.e. $\Pi_\mathbf{r}=\{q\mathbf{u}:\ \mathbf{u}^T\mathbf{r}=0\}$	*4.3*, *5.1*, **82**
PF_K	Parametric Family with K parameters	**178**
PS_K	Parameter Space with K parameters	**178**

Acronyms

Numbers in italic indicate the section number, numbers in bold indicate page numbers where the main definition occurs.

Acronym	Full Name	Index
MSE	Mean Squared Error	*5.1*, *5.6*, **14**
	Normalized Mean Squared Error	**233**
1D	one dimension	*2.3*–*2.5*, *3.3*, *4.3*, *5.2*, **63**, *A.0*
3D	three dimension	*2.5*, *4.3*, *5.2*, **63**
SH	Spherical Harmonic	*4.3*, *6.1*, **22**
2D	two dimension	*2.5*, **65**
SHO	Simple Harmonic Oscillator	*6.1*, **29**
	Simple Harmonic Oscillator in 1D	*2.5*, *5.2*, **95**
	Simple Harmonic Oscillator in 3D	*2.5*, *5.1*, *5.2*, **96**
	Simple Harmonic Oscillator Reconstruction and Estimation	*5.2*, **95**
	the product of SHO-1D basis in 3D space	*4.3*, **144**
QBI	Q-Ball Imaging	*2.7*, *5.2*, **83**
HARDI	High Angular Resolution Diffusion Imaging	**74**
	single shell HARDI	*2.7*, **74**
	multiple shell HARDI	*2.7*, **74**
ODF	Orientation Distribution Function	*2.7*, *3.0*, *3.4*, **80**
EAP	Ensemble Average Propagator	*2.7*, *3.0*, *3.4*, **68**
PGA	Principal Geodesic Analysis	*3.2*, *6.0*, *6.1*, **36**
PDF	Probability Density Function	*3.2*, *3.3*, *4.6*, *6.1*, **63**
PCA	Principal Component Analysis	*6.1*, **50**
PF	Parametric Family	*6.1*, **53**
PS	Parametric Space	*6.1*, **53**
dMRI	diffusion Magnetic Resonance Imaging	**62**
MSD	Mean Squared Displacement	*4.3*, *5.1*, *5.8*, **63**

Acronym	Full Name	Index
MRI	Magnetic Resonance Imaging	**63**
PGSE	Pulsed Gradient Spin-Echo	**66**
DWI	Diffusion Weighted Imaging	*6.1*, **67**
ADC	Apparent Diffusion Coefficient	**67**, *A.0*
DTI	Diffusion Tensor Imaging	*6.4*, **67**
FA	Fractional Anisotropy	*5.1*, *6.1*, **72**
MD	Mean Diffusivity	**72**
GDTI	Generalized Diffusion Tensor Imaging	*5.3*, **75**
HOT	High Order Tensor	*5.1*, **77**
DSI	Diffusion Spectrum Imaging	**80**
RTO	Return-To-Origin probability	*4.3*, *5.1*, *5.8*, *6.2*, *6.3*, **82**
FRT	Funk-Radon Transform	*5.1*, **83**
SNR	Signal-to-Noise Ratio	*4.3*, *5.1*, **84**, *A.0*
FA	Generalized Fractional Anisotropy	*5.1*, **86**
DOT	Diffusion Orientation Transform	*5.1*, *5.2*, **91**
DPI	Diffusion Propagator Imaging	*5.1*, *5.2*, **94**
SPF	Spherical Polar Fourier	*5.1*, *6.1*, *6.3*, **97**
	Spherical Polar Fourier Imaging	*5.0*, *6.1*, **97**
	Fourier dual Spherical Polar Fourier	*5.1*, *6.1*, **116**
GA	Geodesic Anisotropy	*6.0*, *6.1*, *6.4*, **100**
SRPE	Square Root Parameterized Estimation	*4.6*, *6.2*, *6.4*, **177**
SPNP	Spherical Polar Non-Polynomial	*5.2*, **124**
GHOT	Generalized High Order Tensor	*6.2*, **124**
AFT-SC	Analytical Fourier Transform in Spherical Coordinate	**132**
SPP	Spherical Polar Polynomial	**145**
MDA	Mean Difference of Angle	*5.4*, **233**
DA	Difference of Angle	**233**

Abstract

Diffusion MRI (dMRI) is the unique technique to infer the microstructure of the white matter in vivo and noninvasively, by modeling the diffusion of water molecules. Ensemble Average Propagator (EAP) and Orientation Distribution Function (ODF) are two important Probability Density Functions (PDFs) which reflect the water diffusion. Estimation and processing of EAP and ODF is the central problem in dMRI, and is also the first step for tractography. Diffusion Tensor Imaging (DTI) is the most widely used estimation method which assumes EAP as a Gaussian distribution parameterized by a tensor. Riemannian framework for tensors has been proposed successfully in tensor estimation and processing. However, since the Gaussian EAP assumption is oversimplified, DTI can not reflect complex microstructure like fiber crossing. High Angular Resolution Diffusion Imaging (HARDI) is a category of methods proposed to avoid the limitations of DTI. Most HARDI methods like Q-Ball Imaging (QBI) need some assumptions and only can handle the data from single shell (single b value), which are called as single shell HARDI (sHARDI) methods. However, with the development of scanners and acquisition methods, multiple shell data becomes more and more practical and popular. This thesis focuses on the estimation and processing methods in multiple shell HARDI (mHARDI) which can handle the diffusion data from arbitrary sampling scheme.

There are many original contributions in this thesis.

- First, we develop the analytical Spherical Polar Fourier Imaging (SPFI), which represents the signal using SPF basis and obtains EAP and its various features including ODFs and some scalar indices like Generalized Fractional Anisotropy (GFA) from analytical linear transforms. In the implementation of SPFI, we present two ways for scale estimation and propose to consider the prior $E(0) = 1$ in estimation process.

- Second, a novel Analytical Fourier Transform in Spherical Coordinate (AFT-SC) framework is proposed to incorporate many sHARDI and mHARDI methods, explore their relation and devise new analytical EAP/ODF estimation methods.

- Third, we present some important criteria to compare different HARDI methods and illustrate their advantages and limitations.

- Fourth, we propose a novel diffeomorphism invariant Riemannian framework for ODF and EAP processing, which is a natural generalization of previous Riemannian framework for tensors, and can be used for general PDF computing by representing the square root of the PDF called wavefunction with orthonormal basis. In this Riemannian framework, the exponential map, logarithmic map and geodesic have closed forms, the weighted Riemannian mean and median

uniquely exist and can be estimated from an efficient gradient descent. Log-Euclidean framework and Affine-Euclidean framework are developed for fast data processing.

- Fifth, we theoretically and experimentally compare the Euclidean metric and Riemannian metric for tensors, ODFs and EAPs.
- Finally, we propose the Geodesic Anisotropy (GA) to measure the anisotropy of EAPs, Square Root Parameterized Estimation (SRPE) for nonnegative definite ODF/EAP estimation, weighted Riemannian mean/median for ODF/EAP interpolation, smoothing, atlas estimation. The concept of *reasonable mean value interpolation* is presented for interpolation of general PDF data.

Resumé

L'IRM de diffusion est à ce jour la seule technique à même d'observer in vivo et de façon non-invasive les structures fines de la matière blanche, en modélisant la diffusion des molécules d'eau. Le propagateur moyen (EAP pour Ensemble average Propagator en anglais) et la fonction de distribution d'orientation (ODF pour Orientation Distribution Function en anglais) sont les deux fonctions de probabilités d'intérêt pour caractériser la diffusion des molécules d'eau. Le problème central en IRM de diffusion est la reconstruction et le traitement de ces fonctions (EAP et ODF); c'est aussi le point de départ pour la tractographie des fibres de la matière blanche. Le formalisme du tenseur de diffusion (DTI pour Diffusion Tensor Imaging en anglais) est le modèle le plus couramment utilisé, et se base sur une hypothèse de diffusion gaussienne. Il existe un cadre riemannien qui permet d'estimer et de traiter correctement les images de tenseur de diffusion. Cependant, l'hypothèse d'une diffusion gaussienne est une simplification, qui ne permet pas de décrire les cas où la structure microscopique sous-jacente est complexe, tels que les croisements de faisceaux de fibres. L'imagerie à haute résolution angulaire (HARDI pour High Angular Resolution Diffusion Imaging en anglais) est un ensemble de méthodes qui permettent de contourner les limites du modèle tensoriel. La plupart des méthodes HARDI à ce jour, telles que l'imagerie sphérique de l'espace de Fourier (QBI pour Q-Ball Imaging en anglais) se basent sur des hypothèses réductrices, et prennent en compte des acquisitions qui ne se font que sur une seule sphère dans l'espace de Fourier (sHARDI pour single-shell HARDI en anglais), c'est-à-dire une seule valeur du coefficient de pondération b. Cependant, avec le développement des scanners IRM et des techniques d'acquisition, il devient plus facile d'acquérir des données sur plusieurs sphères concentriques. Cette thèse porte sur les méthodes d'estimation et de traitement de données sur plusieurs sphères (mHARDI pour multiple-shell HARDI en anglais), et de façon générale sur les méthodes de reconstruction indépendantes du schéma d'échantillonnage.

Cette thèse présente plusieurs contributions originales.

- En premier lieu, nous développons l'imagerie par transformée de Fourier en coordonnées sphériques (SPFI pour Spherical Polar Fourier Imaging en anglais), qui se base sur une représentation du signal dans une base de fonctions à parties radiale et angulaire séparables (SPF basis pour Spherical Polar Fourier en anglais). Nous obtenons, de façon analytique et par transformations linéaires, l'EAP ainsi que ses caractéristiques importantes : l'ODF, et des indices scalaires tels que l'anisotropie fractionnelle généralisée (GFA pour Generalized Fractional Anisotropy en anglais). En ce qui concerne l'implémentation de SPFI, nous présentons deux méthodes pour déterminer le facteur d'échelle, et nous prenons en compte le fait que $E(0) = 1$ dans l'estimation.

- En second lieu, nous présentons un nouveau cadre pour une transformée de Fourier analytique en coordonnées sphériques (AFT-SC pour Analytical Fourier Transform in Spherical Coordinate en anglais), ce qui permet de considérer aussi bien les méthodes mHARDI que sHARDI, d'explorer les relations entre ces méthodes, et de développer de nouvelles techniques d'estimation de l'EAP et de l'ODF.

- Nous présentons en troisième lieu d'importants critères de comparaison des différentes méthodes HARDI, ce qui permet de mettre en lumière leurs avantages et leurs limites.

- Dans une quatrième partie, nous proposons un nouveau cadre riemannien invariant par difféomorphisme pour le traitement de l'EAP et de l'ODF. Ce cadre est une généralisation de la méthode riemannienne précédemment appliquée au tenseur de diffusion. Il peut être utilisé pour l'estimation d'une fonction de probabilité représentée par sa racine carrée, appelée fonction d'onde, dans une base de fonctions orthonormale. Dans ce cadre riemannien, les applications exponentielle et logarithmique, ainsi que les géodésiques ont une forme analytique. La moyenne riemannienne pondérée ainsi que la médiane existent et sont uniques, et peuvent être calculées de façon efficace par descente de gradient. Nous développons également un cadre log-euclidien et un cadre affine-euclidien pour un traitement rapide des données.

- En cinquième partie, nous comparons, théoriquement et sur un plan expérimental, les métriques euclidiennes et riemanniennes pour les tenseurs, l'ODF et l'EAP.

- Finalement, nous proposons l'anisotropie géodésique (GA pour Geodesic Anisotropy en anglais) pour mesurer l'anisotropie de l'EAP; une paramétrisation par la racine carrée (SRPE pour Square-Root Parameterized Estimation en anglais) pour l'estimation d'un EAP et d'une ODF positifs; la médiane et la moyenne riemanniennes pondérées pour l'interpolation, le lissage et la construction d'atlas basés sur l'ODF et de l'EAP. Nous introduisons la notion de valeur moyenne raisonnable pour l'interpolation de fonction de probabilités en général.

Acknowledgments

This thesis was performed jointly in CCM group in CASIA and in Odyssée/Athena group in INRIA-Sophia Antipolis Méditerranée, under the supervision of Prof. Tianzi Jiang and Prof. Rachid Deriche. This thesis was partially granted by the French Government Award Program, the Natural Science Foundation of China (30730035,81000634), the French ANR "Neurological and Psychiatric diseases" NucleiPark and the France-Parkinson Association.

First of all, I would like to thank my advisors. I want to thank Prof. Tianzi Jiang for supervising me for seven years in CASIA. In my seven years' MS & PhD study, I have worked on several topics. Every time when I changed my research topic, Prof. Jiang always supported me and trusted me, although my research interests are different from his at most of time. I really cherished the freedom and the unconditional support he granted for my study, without which the thesis can not be finished with the current contents. I also want to express my gratitude to Prof. Rachid Deriche for accepting me and supervising me in INRIA, for his trust and invaluable support. Prof. Deriche introduced me the area of advanced dMRI, i.e. HARDI, and showed me the "KISS" principle for excellent papers. Although currently some of my papers seem to violate this principle, I will consider it as the golden rule for all my future works. I also want to thank Prof. Oliver Faugeras for accepting me in Odyssée/Athena group.

Next, I am grateful to all the reviewers and the jury members, i.e. Prof. Maxime Descoteaux, Gaolang Gong, Hongtu Zhu, Théodore Papadopoulo, Yong Fan, Shan Yu and Rong Xue, for their valuable review comments and for their participation in my defense.

Many people contributed in different ways to the contents of this thesis. I want to express my deep thanks to Feng Shi in CCM group for his long term help not only in academic works but also in daily life, to Auro in Athena group for our collaboration in programming and discussion, especially when I first came to France. I want to thank Max and Christophe for their works and codes they left, which makes me start in a very high level in dMRI domain. I would like to thank Manu for his help in the French abstract, in my living in France, and for the collaboration on sampling scheme. I would specially thank Dr. Haz-Edine Assemlal for introducing me his works on SPF basis.

Many other people in CCM group and Odyssée/Athena group help me in variant ways in my works and in my daily life. I would like to thank my collogues in CCM: Gaolang Gong, Fuchun Lin, Ni Shu, Yonghui Li, Qifeng Wang, Xiuchao Sui, Jiajia Li, Lijuan Xu, Kun Wang, Jiefeng Jiang, Bing Liu, Ming Song, Nianming Zuo, Yong Liu, Xinpeng Jiang, Lei Lin, Yuanchao Zhang, Hongming Li, Di Jiang, Sangma Xie, Jiaojian Wang, Chongying Chu, Yuhui Du, Yu Zhang, Haixia Long, Qiaojun Li, etc. I also would like to thank my collogues in Odyssée/Athena group: Olivier, Theo, Maureen, Pierre, Demian, Manu Olivi, Anne-Charlotte, Jaime, etc.

This joint thesis can not be finished without the administrative support by Ms. Bizhen Hong, Lei Lee, Shaomei Wang, Marie-Cecile Lafont, Claire Senica, Juliette Jin, Ghislaine Rodriguez. Thank you for your help.

I would like also to thank my Chinese friends in France: Ting Peng, Yiyi Wei, Yizhen Zhang, Junbo Huang, Ke Huang, Nanjiang Chen, Yinan Liu, Zhigang Liu, Zhengqin Luo, Ken Wong, etc.

I want to thank my roommates and classmates in HIT and CASIA for their long term support and help in last years: Aiwen Jiang, Yanming Zhang, Wei Li, Xiaoyu Zhang, Haoshui Zhang, Deliang Yu, Linjing Li, Mingwei Li, Zheng Wen, etc.

Last, I deeply thank my grandparents, parents, uncles, aunts, and my wife, for their invaluable love, endless patience and encouragement during these years.

Thank you all!

Jian Cheng
2012. 04

CHAPTER 1

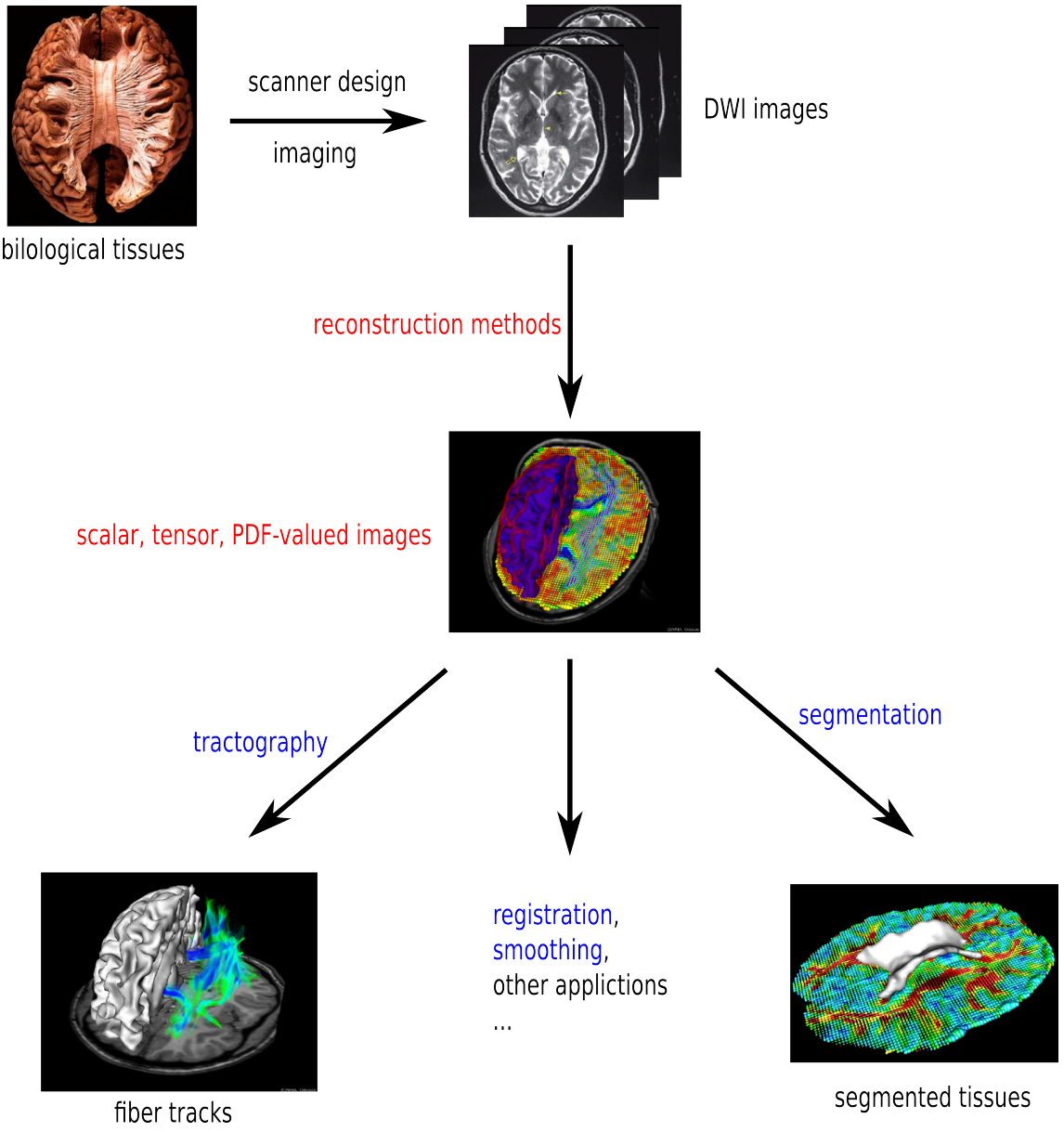
INTRODUCTION

CONTEXT

Diffusion Magnetic Resonance Imaging (dMRI) is a relatively recent MRI method, introduced in the middle of the 80's by [Bihan et al., 1986; Merboldt et al., 1985; Taylor and Bushell, 1985]. The diffusion of water molecules in tissues is hindered by many obstacles, such as macromolecules, fibers, membranes, etc. Thus the diffusion of water molecules can be used as a probe to reveal microscopic details about tissue microstructure. dMRI is a powerful and the unique technique to study the white matter in vivo and noninvasively by modeling the water diffusion. There are many research directions in dMRI, from the imaging technique for Diffusion Weighted Images (DWIs), to reconstruction of scalar/tensor/function valued images, to segmentation, registration, tractography and to clinical applications, etc. Please see Fig. 1.1. In this thesis we are interested in the reconstruction and processing of the Ensemble Average Propagator (EAP) and its various features like Orientation Distribution Functions (ODFs).

Diffusion Tensor Imaging (DTI) is the most widely used reconstruction method based on free diffusion assumption [Basser et al., 1994]. The EAP in DTI is assumed to be Gaussian distribution parameterized by diffusion tensor $\mathbf{D}$. Thus The EAP field can be represented by a tensor field and visualized by ellipsoids. The Mean Diffusivity (MD) and Fractional Anisotropy (FA) are two useful scalar measurements in clinical studies [Pierpaoli and Basser, 1996]. The Riemannian framework for Gaussian distribution has been successfully used in tensor estimation [Lenglet et al., 2006b; Fillard et al., 2007], regularization [Pennec et al., 2006], segmentation [Lenglet et al., 2006a], Principal Geodesic Analysis (PGA) [Fletcher et al., 2004], statistical test [Schwartzman, 2006], etc. In Odyssée/Athena group, Dr. Lenglet has made a great contribution on the Riemannian framework for tensors in DTI [Lenglet, 2006; Lenglet et al., 2006a,b]. In CCM group, we has published many works on neuroscience applications based on the scalar indices or brain network analysis in DTI model [Gong et al., 2005; Lin et al., 2006; Yu et al., 2008; Shu et al., 2009; Li et al., 2009; Wang et al., 2011; Li et al., 2012]. Although the Gaussian assumption for free diffusion is adequate for voxels with isotropic diffusion or coherent single direction diffusion, it fails for voxels with more complicated microstructure, because the water diffusion in tissues is generally the hindered diffusion. Tensor model in DTI cannot resolve complex fiber

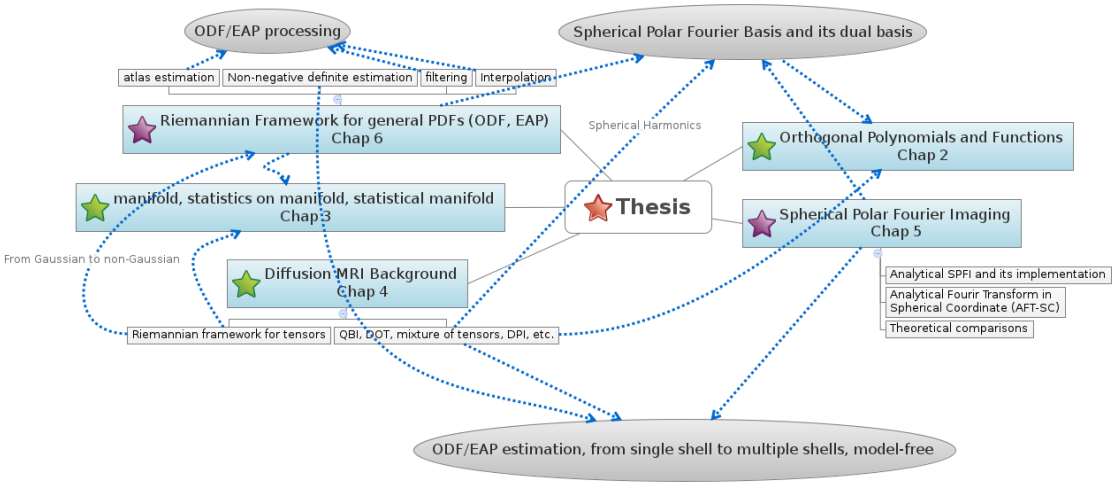
Figure 1.1: Research contents in dMRI. In this thesis, we focus on the reconstruction and processing of the EAP and its features like ODFs. Chapter 5 focuses on estimation of the EAP and its features. Chapter 6 focuses on the Riemannian framework for processing of ODFs and EAPs.



configurations like crossing, merging, kissing, etc. While it was reported that between one third to two thirds of imaging voxels in the human brain contain have more than one fiber orientation [Behrens et al., 2007].

High Angular Resolution Diffusion Imaging (HARDI) is a category of reconstruction methods proposed to avoid the Gaussian EAP assumption and resolve the complex fiber configurations. In this thesis, we consider HARDI as all methods beyond DTI, although HARDI was proposed initially as a mixture of tensor model [Tuch et al., 2002]. HARDI methods can be separated into two classes, i.e. single shell HARDI (sHARDI) and multiple shell HARDI (mHARDI). sHARDI methods like the most famous one Q-Ball Imaging (QBI) [Tuch, 2004; Anderson, 2005; Hess et al., 2006;

Figure 1.2: Sketch of the chapters and their relations in the thesis. The green stars are the chapters in background part. The purple stars are the chapters in contribution part.



Descoteaux et al., 2007] only can handle the data with single b value (single shell), because the data from different shell obtain different results. In Odyssée/Athena group, Dr. Descoteaux did a great job in analytical QBI and other Spherical Harmonic (SH) related sHARDI methods [Descoteaux, 2008; Descoteaux et al., 2007, 2008a]. mHARDI methods normally can work for both single shell data and multiple shell data by considering a basis defined in $\mathbb{R}^3$, not SH basis in $\mathbb{S}^2$. Compared to sHARDI methods, mHARDI methods need less assumptions and are model-free if complete basis is used. In this thesis we focus on mHARDI methods for the estimation and processing of EAP and its features.

Organization and Contributions of this Thesis

This thesis is organized mainly in two parts, i.e. the background part and the contribution part. In the background part, we give a quick review for the background knowledge on dMRI, orthonormal polynomials and special functions, manifold, statistics on manifold and statistical manifold, which are used in the contribution part. In the contribution part, we list two main contributions. One is the analytical Spherical Polar Fourier Imaging (SPFI) and the Analytical Fourier Transform in Spherical Coordinate (AFT-SC) framework in Chapter 5, which can be seen as a generalization of Dr. Descoteaux's work on analytical QBI from $\mathbb{S}^2$ to $\mathbb{R}^3$. The other one is the diffeomorphism invariant Riemannian framework for processing of ODFs and EAPs in Chapter 6, which can be seen as a generalization of Dr. Lenglet's work on DTI from Gaussian distribution family to general non-Gaussian distribution family. We now give an overview for each chapter in turn.

Part I: Background

The Background part describes some useful mathematical backgrounds required to understand the technical contents in this thesis, and the basic concepts and knowl-

edge of dMRI.

Chapter 2: This chapter covers the mathematical materials on orthogonal polynomials and special functions, which are needed to understand Chapter 4 and 5. We first introduce the basic concepts of the orthonormal functions and polynomials in section 2.1 and 2.2. Then in section 2.3 Sturm-Liouville theory is introduced and diffusion equation in one dimension space is solved in Example 2.3. Next in section 2.4 we review some basic concepts and results on polynomials, homogeneous polynomials and homogeneous harmonic polynomials in $\mathbb{R}^d$ and $\mathbb{S}^{d-1}$. The most important theoretical result is the Harmonic Decomposition Theorem 2.4 and its various corollaries, which directly results in Theorem 4.1 on the equivalence of High Order Tensor (HOT) basis, polynomial basis and SH basis in $\mathbb{S}^2$. Then some orthonormal basis and their properties are introduced in section 2.5, including associate Legendre polynomial, Spherical Hamonics (SHs), Bessel function and spherical Bessel function, Hermite polynomial, associate Laguerre polynomial and hypergeometric function. Three important theorems about SHs, i.e. Funk-Hecke theorem 2.7, addition theorem 2.6 and plane wave expansion 2.8, are listed in this section. Their relation can be found in Appendix 5.7. This section also introduced Simple Harmonic Oscillator basis in 1D space in Eq. (2.60), called SHO-1D basis, and Simple Harmonic Oscillator basis in 3D space in Eq. (2.69), called as SHO-3D basis. We introduce the Gamma function and some other useful integral formulae in section 2.6.

Chapter 3: This chapter covers the mathematical concepts and results on manifold, statistics on manifold and statistical manifold. Section 3.1 reviews the three layers of structural concepts of Riemannian manifold, i.e. the topological structure, differentiable structure and the Riemannian structure. The geometry of high dimension sphere $\mathbb{S}^{d-1}$ is well studied in Example 3.2 on topology, Example 3.5 on tangent space, Example 3.7 on geodesic, Example 3.8 on exponential and logarithmic maps. Riemannian framework on sphere including geodesic, exponential map and logarithmic map can be used to devise algorithms to process the data on sphere. Section 3.2 presents the statistical analysis on Riemannian manifold, including the definition, existence and uniqueness of weighted Riemannian mean and median, PGA. Section 3.3 presents the basic concepts on information geometry which considers the probability family as a manifold called statistical manifold. The Fisher information metric is the natural Riemannian metric in the statistical manifold. We list two examples on the Riemannian framework for different probability family. Example 3.10 is for the Gaussian distribution family which is used in DTI. Example 3.11 is for the multinomial distribution family which is used in Riemannian framework for ODFs [Goh et al., 2011] by considering the ODF represented by its histogram. Our formulation in Chapter 6 is more general and efficient by considering the orthonormal basis representation in continuous case.

Chapter 4: This chapter presents the fundamental concepts and knowledge of Magnetic Resonance Imaging (MRI) and diffusion MRI (dMRI), reviews the reconstruction methods including DTI, sHARDI and mHARDI, and reviews the previous studies on metrics and processing frameworks for Probability Density Function (PDF) valued

data. First, we describe the basic principles of MRI and dMRI in section 4.1, including the dualities between **k**-space and **x**-space, between **q**-space and **R**-space. Next, we review different reconstruction methods in dMRI, including DTI in section 4.2, sHARDI methods and mHARDI methods in section 4.3. We list the advantages and limitations for every method, especially the assumptions used in each method. See Table 4.1 and Appendix 4.7 for the assumptions in QBI and exact QBI. We also propose two correct ways for ODF normalization in Appendix 4.6. Finally we review the metrics and processing frameworks proposed for tensors (Gaussian EAPs) and ODFs in section 4.4. To our knowledge, there is no work on the metric and processing framework for EAPs so far, which will be addressed in Chapter 6.

Part II: Contributions

This part contains two main contributions in this thesis.

Chapter 5: This chapter proposes a novel analytical Spherical Polar Fourier Imaging (SPFI) reconstruction for EAP and its various features. First, Dr. Assemlal's work on SPF basis, least square estimation with regularization, and numerical inner product to estimate EAP and its features are presented in subsection 5.1.1. Second, the analytical forms for EAP, two kinds of ODFs, three scalar indices including Return-To-Origin probability (RTO), Mean Squared Displacement (MSD), generalized FA (GFA), are proposed in subsection 5.2.2. These analytical forms are linear transforms. Note the GFA we proposed is for EAP in $\mathbb{R}^3$ which is the generalization of FA for tensors and GFA for ODFs. Next, the implementation of SPFI is shown in Table 5.1, which contains three independent steps. In the scale estimation step, we propose two ways. One is based on the typical ADC, and the other one is based on fitting Generalized HOT (GHOT) model. In the least square estimation part, we propose to consider the prior $E(0) = 1$ in estimation. The third step is the analytical linear transforms demonstrated in subsection 5.2.2. Section 5.2.2 proposes an novel Analytical Fourier Transform in Spherical Coordinate (AFT-SC) framework to compare and analyze different sHARDI and mHARDI methods in a unified framework. Many HARDI methods can be explained in AFT-SC framework, such as QBI, Diffusion Orientation Transform (DOT), Diffusion Propagator Imaging (DPI), Simple Harmonic Oscillator Reconstruction and Estimation (SHORE) and SPFI. Moreover many new methods can be proposed in AFT-SC for analytical ODF and EAP estimation. Please see Table 5.2 for these analytical reconstruction methods and their possible variants. For QBI in AFT-SC framework, we propose the EAP for QBI, and demonstrate the EAP and ODF by Wedeen for QBI are actually impractical since they have large modeling error. For SHORE, we propose the analytical forms for EAP, ODFs and variant scalar indices. For DPI and DOT, we analyze their limitations and propose several variants to avoid the limitations. AFT-SC framework makes the analytical ODF and EAP estimation become an easy job. Then two questions arise naturally: How to evaluate all these analytical methods? Which analytical method is better? In section 5.3, we propose some criteria for theoretical comparisons. See Table 5.3 for an overview. These sHARDI and mHARDI methods are compared in experiments by synthetic data, phantom data and real data in section 5.4. The experimental results validated the methods and our theoretical comparisons.

Chapter 6: This chapter proposes a general state-of-the-art Riemannian framework as a mathematical tool to process PDF data like ODF/EAP valued images, by representing the square root of the PDF, called wavefunction based on quantum mechanics, as a linear combination of some orthonormal basis functions. The proposed Riemannian framework is the natural extension of previous Riemannian framework for tensors. In theoretical part in section 6.1, we deduce the Riemannian metric, i.e. the Fisher information metric, for PDFs based on orthonormal basis representation and show the properties of the statistical manifold which is a convex subset of a high dimension sphere. In this framework, the exponential map, logarithmic map and geodesic have closed forms, and weighted Riemannian mean (Fréchet mean) and weighted Riemannian median uniquely exist. Moreover, we present two efficient frameworks, i.e. Affine-Euclidean framework and Log-Euclidean framework, for fast processing of data in subsection 6.1.6, and generalize the Geodesic Anisotropy (GA) form tensors to ODFs and EAPs in subsection 6.1.5, which is the Riemannian distance from the ODF/EAP to the nearest isotropic ODF/EAP and is closely related to the Rényi entropy. It should be noted that our theoretical results can be used for any probability density function (PDF) besides ODF/EAP under any orthonormal basis representation. Furthermore we analyze theoretically the similarities and differences between Riemannian frameworks for EAPs, ODFs and for tensors. We analyzed theoretically and experimentally the similarities and differences between Riemannian framework for tensors and for ODFs and EAPs, between Riemannian metric and Euclidean metric. The proposed Riemannian metric is diffeomorphism invariant which is the natural extension of the previous affine-invariant metric for tensors. Based on this Riemannian framework of PDFs, we demonstrate some potential applications in section 6.2 via the Riemannian operations for ODF and EAP computation, such as anisotropy description via GA, nonnegative definite ODF/EAP estimation, interpolation, filtering, Principle Geodesic Analysis (PGA) and atlas estimation. In the interpolation part, we propose the concept of *reasonable mean value interpolation* on general PDF data. The *swelling effect* on tensor interpolation is a specific phenomenon of *unreasonable mean value interpolation*. In section 6.3, the proposed Riemannian framework and its applications are validated in synthetic, phantom and real data. The experiments demonstrate that the Riemannian framework is very useful for ODF and EAP computing.

Part III: Conclusions

Chapter 7: This is the conclusion chapter which summarizes our contributions in Chapter 5 and 6. It also summarizes the journey of my study for this thesis and presents the potential perspectives.

Part IV: Appendices

Appendix A: This appendix describes the synthetic data generation and evaluation of reconstruction methods in this thesis. The synthetic data is generated by mixture of tensor model or mixture of Söderman cylinder model. The evaluation can be performed by comparing the estimated ODFs/EAPs with ground truth ODFs/EAPs based on mean squared error or fiber directions which are considered to be the maxima of

ODF/EAP profiles. The maxima of ODF/EAP profiles are detected by a hybrid way which combines the discrete mesh search and gradient ascent.

Appendix B: This appendix describes two real data set. One is the public phantom data from LNAO used in fiber cup MICCAI 2009. The other one is the monkey data from Xuanwu Hospital, Capital Medical University.

Appendix C: This appendix lists the publications from the author for this thesis.

Part I

Background

CHAPTER 2

ORTHOGONAL POLYNOMIALS AND FUNCTIONS

"You don't have to believe in God, but you should believe in The Book."

– Paul Erdös

Contents

OVERVIEW

Special functions are some particular mathematical functions which have more or less established names and notations due to their importance in mathematical analysis, functional analysis, physics, or other applications. An orthogonal polynomial family is a set of polynomials which are orthogonal to each other under a certain inner product. Classic orthogonal polynomials are some widely used special functions and can be represented by hypergeometric functions based on Askey scheme [Askey and Wilson, 1985]. Special functions and their properties are also widely used in HARDI. For example, spherical harmonics and its properties have been deeply explored and applied in QBI [Descoteaux et al., 2007; Aganj et al., 2010b], DOT [Özarslan et al., 2006], etc.

This chapter gives a quick overview of some classic special functions and their properties that are used in the thesis. Please refer some textbooks on Mathematical Physics, functional analysis and partial differential equation and Fourier analysis for more details on the contents of this chapter [Arfken et al., 2005; Conway, 1990; Haberman, 1987; Byron and Fuller, 1992; Axler et al., 2001; Andrews et al., 1999].

Organization of this chapter:
First, orthogonal functions in piecewise continuous function space are introduced in Section 2.1. Second, orthogonal polynomials are introduced based on Stone-Weierstrass theorem and Gram-Schmidt orthogonalization in Section 2.2. Then Sturm-Liouville theory on the connection between complete orthogonal function system and second order ordinary differential equation is described in Section 2.3, which provides another way to construct an orthonormal basis in addition to Gram-Schmidt orthogonalization. Next, we list some basic theoretical results on decomposition of polynomial space in Section 2.4. Then some important orthonormal bases and their properties used in this thesis, i.e. Spherical Harmonics, Laguerre Polynomial and Hermite Polynomial, Bessel function etc., are introduced from partial differential equation and orthogonalization in Section 2.5. At last, some other special functions and integral formulae used in the following chapters are listed in Section 2.6.

2.1 ORTHOGONAL FUNCTIONS

The function space of all piecewise continuous complex valued ($\mathbb{C}^1$) functions defined in a given real interval $[a, b] \subset \mathbb{R}^1$ forms a **vector space**. There are many possibilities to define such an **inner product** $\langle \cdot, \cdot \rangle$ in the function space such that it is a **Hilbert space**. Then the **norm** and **metric** introduced by the inner product are given as $\|f(x)\| = \sqrt{\langle f(x), f(x) \rangle}$ and $d(f(x), g(x)) = \|f(x) - g(x)\|$, where $f(x)$ and $g(x)$ are any two functions in the function space. A normally used inner product is defined as $\langle f(x), g(x) \rangle = \int f(x)\overline{g(x)}\mathrm{d}x$, where $\overline{g(x)}$ means the conjugate of $g(x)$. This inner product introduces the well-known ℓ_2 norm and metric in the function space. A more general and widely used inner product is in Eq. (2.1), where $w(x)$ is a positive definite weight function.

$$\langle f(x), g(x) \rangle = \int_a^b f(x)\overline{g(x)}w(x)\mathrm{d}x \tag{2.1}$$

After an inner product is defined, many useful tools can be introduced based on it, such as the norm (or called length) of one function, distance and angle between two functions, etc.

Definition 2.1. *Let $f(x)$ and $g(x)$ be two piecewise continuous $\mathbb{C}^1$ (1D complex valued) functions. For one function, the **length** or **norm** of $f(x)$ is defined as $\|f(x)\| = \sqrt{\langle f(x), f(x)\rangle}$. $f(x)$ is said to be **normalized** if it has unit norm. For two functions, $f(x)$ and $g(x)$ are said to be **orthogonal** if $\langle f(x), g(x)\rangle = 0$. The **metric** or **distance** is introduced as $d(f(x), g(x)) = \|f(x) - g(x)\|$. The **angle** between $f(x)$ and $g(x)$ is defined as* $\arccos\left(\frac{\langle f(x),g(x)\rangle}{\|f(x)\|\|g(x)\|}\right)$.

When $f(x)$ is orthogonal to $g(x)$, the angle between these two functions is $90°$. When $f(x)$ is not orthogonal to $g(x)$, $f(x)$ can be separated into two parts, where one is $f(x) - \frac{\langle f(x),g(x)\rangle}{\langle g(x),g(x)\rangle} g(x)$ that is orthogonal to $g(x)$, and the other one is $\frac{\langle f(x),g(x)\rangle}{\langle g(x),g(x)\rangle} g(x)$ that has angle of $0°$ with $g(x)$. We can also normalize these two parts such that they have unit norm. This process is called as **Gram–Schmidt orthogonalization**. Then these two parts are just like two axes in Euclidean space $\mathbb{R}^2$ and we can represent $f(x)$ with two scalar coefficients along these two axes.

The above concepts are only for one or two functions. For a set of functions $\{f_n(x)\}|_{n\in J}$ (maybe infinite) where J is the index set, the functions are called to be **linearly independent** if the equation $\sum_{n\in J} c_n f_n(x) = 0$ has only the trivial solution $c_n = 0$, $\forall n \in J$. Otherwise, they are **linearly dependent**. A function set is an **orthonormal system** if the functions all have unit norm and any two different functions are orthogonal, i.e. $\langle f_n(x), f_m(x)\rangle = \delta_{nm}$. It is easy to check that the functions in an orthonormal system are linearly independent. A linearly independent function set can be always linearly transformed into an **orthonormal system** by the Gram–Schmidt process performed in one dimension after another as we have described above in two dimension.

The **linear span** of a linearly independent function set $\{f_n(x)\}|_{n\in J}$ defined as $\text{Span}\{\{f_n(x)\}|_{n\in J}\} = \{\sum_{n\in J} c_n f_n(x) \mid c_n \in \mathbb{C}\}$ forms a vector space, which is homeomorphism to the Euclidean space $\mathbb{C}^{|J|}$ when J is a finite set[1]. Based on the definition of the linear span, any function in this spanned space can be linearly represented by these functions. Thus we say that the function set is **complete** in the space or the function set is a **basis** of the space. A ***weaker*** and more ***formal*** definition of **completeness** is given as follows.

Definition 2.2. *A function set $\{f_n(x)\}_{n\in J}$ is called to be **complete** in a function space $\mathbb{FS}$ [Arfken et al., 2005; Courant and Hilbert, 1989], if $\forall g(x) \in \mathbb{FS}$, $\exists\ c_1, c_2, ..., c_N$, such that*

$$\lim_{N\to\infty} \|g(x) - \sum_{n=1}^{N} c_n f_n(x)\|^2 = \lim_{N\to\infty} \int_a^b \left|g(x) - \sum_{n=1}^{N} c_n f_n(x)\right|^2 w(x)\mathrm{d}x = 0 \tag{2.2}$$

*Complete orthonormal system in $\mathbb{FS}$ is called as **orthonormal basis** of $\mathbb{FS}$.*

Example 2.1. *A simple example of orthonormal basis is the well-known Fourier basis $\{e^{-i2\pi nx}\}_{n=0,\pm1,\cdots}$ for the functions defined in $[0, 1]$.*

[1] $|J|$ means the number of members in J

The coefficients for the general orthonormal basis are called as **generalized Fourier coefficients** and the partial summation $\sum_{n=1}^{N} c_n f_n(x)$ is called as **generalized Fourier series**, which are analogical to the Fourier coefficients and Fourier series.

As we have shown about, any orthonormal system is an orthonormal basis for its linear span. However, for a given orthonormal system and given function space, Eq. (2.2) is not easy to be verified. Eq. (2.2) based on Mean Squared Error (MSE) means the generalized Fourier series $\sum_{n=1}^{N} c_n f_n(x)$ **converges in the mean** to $g(x)$, where $w(x)$ is like a probability density function. **Convergence in the mean** is different from **pointwise convergence**. For example, when Fourier basis is used to represent discontinuous function like square wave function, Gibbs phenomenon occurs in the discontinuous points. That is understandable because we use continuous basis functions to represent discontinuous function. In this case, Fourier series of square wave function does not converge at the discontinuous points, while Eq. (2.2) is still satisfied. However, in practice when MSE is used to evaluate the difference between signals (functions), we can simply ignore the discontinuous points because they have no contribution for MSE. In this thesis, we are interested in DWI signals which can be seen as continuous function based on biological priors. So we assume the functions studied in this thesis are continuous, not piecewise continuous. Moreover, like Fourier basis, general orthonormal basis satisfies the Bessel's inequality, i.e. $\langle g(x), g(x)\rangle^2 \geq \sum_{n\in J} c_n^2$, because $\|g(x) - \sum_{n\in J} c_n f_n(x)\|^2 \geq 0$. The equality holds only if Eq. (2.2) is satisfied. So in order to guarantee the coefficients $\{c_n\}$ exist, we need to assume the norm of function $g(x) \in \mathbb{FS}$ is bounded. In other words, $\mathbb{FS}$ is square integrable function space, denoted by $\mathbb{L}^2([a,b], w(x))$, where $w(x)$ is the weight function, $[a,b]$ is the real interval. Let $\mathbb{L}^2(\chi, w(\boldsymbol{x}))$ denote the space of square integrable functions in domain χ and with weight function $w(\boldsymbol{x})$. ***In the following parts of the thesis, if not explicitly stated, we always assume the functions are continuous and in*** $\mathbb{L}^2(\chi, w(\boldsymbol{x}))$ ***with different*** $w(\boldsymbol{x})$ ***and in different*** χ**.** Actually square integrable condition is easy to be satisfied. For example, if $g(x)$ is a bounded function defined in a bounded space, its norm is always bounded.

The generalized Fourier coefficients for orthonormal basis $\{f_n(x)\}$ can be easily obtained as $c_n = \langle g(x), f_n(x)\rangle$, which is the unique solution of the least square problem $\{c_n\} = \arg\min \int_a^b \left| g(x) - \sum_{j=1}^{\infty} c_j f_j(x) \right|^2 w(x)\mathrm{d}x$. Then the generalized Fourier series is given in Eq. (2.3), where $\sim$ means the approximation in the mean, and the approximation becomes equality in every point if $g(x)$ is continuous and square integrable.

$$g(x) \sim \sum_{n=1}^{\infty} \langle g(x), f_n(x)\rangle f_n(x) \tag{2.3}$$

In this case c_n is only dependent on the basis function $f_n(x)$ and $g(x)$, independent with $\{f_m(x)\}_{m\neq n}$.

2.2 ORTHOGONAL POLYNOMIALS

Compared to Trigonometric function in Fourier basis, polynomial is the another kind of basic elementary function. Historically it has been proved useful in func-

tion representation and approximation. A fundamental theorem on this topic is the famous **Stone-Weierstrass theorem**.

Theorem 2.1 (Stone-Weierstrass theorem). *Let $g(x)$ be a continuous complex valued function defined in $[a, b]$. for any given ϵ, there exists a polynomial $p(x)$ defined in $[a, b]$, such that $|g(x) - p(x)| < \epsilon$, $\forall x \in [a, b]$.*

In other words, the function defined on a closed and bounded interval can be **uniformly approximated** by polynomials to any degree of accuracy. Thus no matter how complex the function $g(x)$ is, we can always study its approximated polynomial instead of study the function itself. Please note that the **uniform convergence** in this theorem is stronger than convergence in the mean in last subsection. Although the proof of Stone-Weierstrass theorem is a ***constructive proof*** based on **Bernstein polynomial**, the proof only shows an asymptotic solution that there exists a Bernstein polynomial with a large enough order which can guarantee the absolute difference is less than the given tolerance ϵ. See [Byron and Fuller, 1992, chap. 5] for the details of the proof. However, in practice we are more interested in the following two questions.

- For a given order N, how to estimate polynomial $p_N(x)$ with the minimal MSE $\int_a^b \|g(x) - p_N(x)\|^2 w(x)\mathrm{d}x$?
- For a given MSE $= \epsilon$, how to estimate $p_N(x)$ with the minimal order N ?

These two questions are like two faces of one coin. In order to answer them, we need to better understand the polynomial space.

Let $p_n(x)$ denote the polynomial of order n and let $\mathcal{P}_N$ denote the polynomial function space contains all polynomial with the order no more than N, i.e. $\mathcal{P}_N = \{p_n(x) : n \leq N\}$. An obvious basis of $\mathcal{P}_N$ is the set of $N+1$ monomials $\{x^n\}_{n=0}^N$, which is not orthonormal system because $\langle x^n, x^m \rangle \neq \delta_{nm}$. For a given N, we can represent the polynomial as $p_N(x) = \sum_{n=0}^N c_n x^n$ and minimize the cost function $M(\boldsymbol{c}) = \int_a^b |g(x) - \sum_{n=0}^N c_n x^n|^2 w(x)\mathrm{d}x$, where $\boldsymbol{c} = (c_1, ..., c_N)^T$. This is a least square problem. After letting $\frac{\partial M(\boldsymbol{c})}{\partial c_n}$ zero, we have $\sum_{n=0}^N c_n \langle x^n, x^m \rangle = \langle g(x), x^m \rangle$. Then the unique solution is $\boldsymbol{c} = X^{-1}\boldsymbol{g}$, where $\boldsymbol{g} = (\langle g(x), x^0 \rangle, ..., \langle g(x), x^N \rangle)^T$, and $X_{nm} = [\langle x^n, x^m \rangle]$ is called as **Hilbert matrix** when $w(x) = 1$. Hilbert matrix is known to have large condition number and is a classical example for demonstrating round-off error difficulties. Another issue is that we need to calculate the inverse of X, which means c_n and c_m are coupled together.

Historically, in order to solve these two issues, we can perform Gram–Schmidt orthogonalization to $\{x^n\}_{n=0}^N$ and obtain the orthonormal system $\{u_n(x)\}_{n=0}^N$. The orthogonalization is done for each dimension as follows.

$$u_0(x) \leftarrow \frac{x^0}{\|x^0\|} \tag{2.4}$$

$$\forall n > 0, \quad u_{n+1}(x) \leftarrow x^n - \sum_{m=0}^{n} \langle x^n, u_m(x) \rangle, \quad u_{n+1}(x) \leftarrow \frac{u_{n+1}(x)}{\|u_{n+1}(x)\|} \tag{2.5}$$

Then we can assume $p_N(x) = \sum_{n=0}^N c_n u_n(x)$, and obtain the solution $\sum_{n=0}^N c_m \langle u_n(x), u_m(x) \rangle = \langle g(x), u_m(x) \rangle = c_m$. In this case we do not need to calculate

Table 2.1: Some classical orthogonal polynomials generated by Gram–Schmidt orthogonalization of $\{x^n\}$ and second order ordinary differential equations.

Polynomial	Legendre $P_n(x)$	Hermite $H_n(x)$	Laguerre $L_n(x)$	Associated Laguerre $L_n^\alpha(x)$
interval	$[-1,1]$	$(-\infty,\infty)$	$[0,\infty)$	$[0,\infty)$
$w(x)$	1	e^{-x^2}	e^{-x}	$x^\alpha e^{-x}$
squared norm	$\langle P_n(x),P_n(x)\rangle = \frac{2}{2n+1}$	$\langle H_n(x),H_n(x)\rangle = 2^n\sqrt{\pi}n!$	$\langle L_n(x),L_n(x)\rangle = 1$	$\langle L_n^\alpha(x),L_n^\alpha(x)\rangle = \frac{(n+\alpha)!}{n!}$
$p_0(x)$	$1-x^2$	1	x	x
$p_1(x)$	$-2x$	$-2x$	$1-x$	$\alpha+1-x$
$R(x)=e^{\int^x \frac{p_1(t)}{p_0(t)}\mathrm{d}t}$	$1-x^2$	e^{-x^2}	xe^{-x}	$x^{\alpha+1}e^{-x}$
$\lambda(n)$	$n(n+1)$	$2n$	n	n

the inverse of the matrix since it is an identity matrix. c_n is only dependent on the basis function $u_n(x)$ and $g(x)$, as we have shown in the last section. Please note for $\{x^n\}|_{n=0}^N$ that is just independent, the optimal c_n is dependent on the whole function set, not just $u_n(x)$. Moreover, for two different order N_1 and N_2, the optimal first $\min\{N_1,N_2\}$ coefficients are the same for orthonormal basis, but different for non-orthonormal basis. Both orthonormal and non-orthogonal basis functions are used in dMRI field. We will compare them in the following chapters .

Example 2.2 (Orthogonalization and Legendre Polynomials)**.** *For* $\mathbb{L}^2([-1,1],1)$ *where the function domain is* $[-1,1]$ *and weight function* $w(x)=1$*, the obtained orthonormal basis from* $\{x^n\}_{n=0}^N$ *is proportional to the famous* ***Legendre polynomial*** $\{P_n(x)\}$ *which is not normalized. The Legendre polynomial is actually the unique form of orthonormal system for closed interval, because if* $[a,b]$ *is not* $[-1,1]$ *we can always perform a linear transform* $y=\frac{2x-b-a}{b-a}$ *such that the interval becomes* $[-1,1]$.

If we choose another $w(x)$ or consider a or b is infinity, Gram-Schmidt orthogonalization will result in another polynomial family as an orthonormal basis. Table 2.1 lists some classical orthogonal polynomials used in this thesis, where $L_n^\alpha(x)$ becomes $L_n(x)$ when $\alpha=0$. The normalized version of these polynomials are orthonormal basis from Gram–Schmidt Orthogonalization.

2.3 STURM-LIOUVILLE THEORY

It is easy to test whether a given function set under the given weight function is orthonormal system or not. However, it is hard to verify whether the given function set is complete based on the definition 2.2. Since independent function set can be always orthogonalized to orthonormal system, historically Gram–Schmidt orthogonalization becomes an important way to study orthogonal functions and Stone-Weierstrass theorem is a way to prove completeness. Another way is the **Sturm-Liouville theory** for second order ordinary differential equations.

In this section, we are interested in the **Sturm-Liouville equation** in Eq. (2.6), where $p(x)>0$, $w(x)>0$, and $p(x)$, $q(x)$, $w(x)$ are continuous function defined in $[a,b]$ and determined by their physical meaning in different applications. The boundary conditions are given as $g(a)$, $g(b)$, $g'(a)$, $g'(b)$ or their linear combination. The value of λ is not specified in the equation. Finding the values of λ such that there exists a non-trivial solution satisfying the boundary conditions is part of the problem called

the **Sturm-Liouville problem**.

$$\frac{d}{dx}[p(x)\frac{dg(x)}{dx}] + q(x)g(x) + \lambda w(x)g(x) = 0, \quad x \in (a,b) \tag{2.6}$$

Many famous equations in physics are in this form or can be transformed into this form through **separation of variables**.

Example 2.3 (**Diffusion Equation** in 1D)**.** *Diffusion equation (or called heat equation) in one dimension is*

$$\frac{\partial g}{\partial t} = D\frac{\partial^2 g}{\partial x^2}, \quad x \in (-\infty, \infty), t \in [0, \infty) \tag{2.7}$$

*where D is the **diffusion coefficient** used to describe the rate of the diffusion. After considering $g(x,t) = X(x)T(t)$ in the technique of separation of variables, we have two equations*

$$X'' + kX = 0, \quad T' + kDT = 0$$

*where k is the so called **separation constant**. Because we are interested in stale solutions after large enough time, k should be positive, which means $X'' + kX = 0$ is a Sturm-Liouville equation. Then the solutions of above two equations are*

$$X(x) = A(k)\sin(\sqrt{k}x) + B(k)\cos(\sqrt{k}x), \quad T(t) = C(k)e^{-kDt}$$

where $A(k), B(k), C(k)$ are constants dependent on k and need to be set by considering boundary conditions. Then the final solution is

$$g(x,t) = (A(k)\sin(\sqrt{k}x) + B(k)\cos(\sqrt{k}x))e^{-kDt} \tag{2.8}$$

Besides, any arbitrary summation of such $g(x,t)$ with different k is still a solution of diffusion equation. In the following, we will obtain the solution of the diffusion equation under different boundary conditions.

- *Consider the following boundary condition*

  $$g(a,t) = g(b,t) = 0, \quad \forall t \geq 0$$

  which is the boundary condition for the diffusion in a given stick with length $b-a$, then we have $k = \frac{n^2\pi^2}{(b-a)^2} \geq 0$, $n = 0, 1, 2\ldots$ for non-trivial solutions. Then the final solution is

  $$g(x,t) = \sum_{n=1}^{\infty} A_n \sin\left(\frac{n\pi}{b-a}(x-a)\right)e^{-\frac{Dn^2\pi^2}{(b-a)^2}t}$$

- *Consider another boundary condition as*

  $$g(x,0) = f(x), \quad x \in (-\infty, \infty)$$

  *which is the boundary condition of **random walk** of a particle when $f(x)$ is the initialization probability in spatial space. In Eq. (2.8), we let $k = 4\pi^2\xi^2$ and replace trigonometric functions with exponential functions. Then we have*

  $$g(x,t) = A(\xi)e^{i2\pi\xi x}e^{-4\pi^2\xi^2 Dt}$$

The integration over ξ is still a solution given as

$$g(x,t) = \int_{-\infty}^{\infty} A(\xi) e^{i2\pi\xi x} e^{-4\pi^2\xi^2 Dt} \mathrm{d}\xi$$

which is the inverse Fourier transform of $A(\xi)e^{-4\pi^2\xi^2 Dt}$. Then $A(\xi)$ is the Fourier transform of $f(x)$, because

$$f(x) = g(x,0) = \int_{-\infty}^{\infty} A(\xi) e^{i2\pi\xi x} \mathrm{d}\xi = \mathscr{F}^{-1}\{A(\xi)\}(x)$$

Thus considering

$$\mathscr{F}^{-1}\{e^{-4\pi^2\xi^2 Dt}\}(x) = \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{x^2}{4Dt}}$$

we have the final solution as the convolution of $f(x)$ and $\mathscr{F}^{-1}\{e^{-4\pi^2\xi^2 Dt}\}(x)$, i.e.

$$g(x,t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{(y-x)^2}{4Dt}} f(y)\mathrm{d}y \tag{2.9}$$

When $f(x) = \delta(x)$, we have the so called Green's function as

$$g(x,t) = \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{x^2}{4Dt}} = N(x|0,2Dt) \tag{2.10}$$

*which is the probability at spatial and temporal position (x,t) in **Brownian motion**. Its 2-order moment is*

$$\langle x^2 \rangle \overset{\text{def}}{=} \int_{-\infty}^{\infty} x^2 N(x|0,2Dt)\mathrm{d}x = 2Dt \tag{2.11}$$

The classic polynomials described in the last section are proved to be the solutions of some ordinary differential equations which are in the general form.

$$p_0(x)g''(x) + p_1(x)g'(x) + p_2(x)g(x) + \lambda g(x) = 0 \tag{2.12}$$

When $p_1(x) = p_0'(x)$, this form can be written directly as equation Eq. (2.6). If $p_1(x) \neq p_0'(x)$, let $R(x) = e^{\int^x \frac{p_1(t)}{p_0(t)}\mathrm{d}t}$ and a multiplier $\frac{R(x)}{p_0(x)}$ can be performed to the general form, then we have $[R(x)g'(x)]' + \frac{R(x)p_2(x)}{p_0(x)}g(x) + \lambda\frac{R(x)}{p_0(x)}g(x) = 0$, which is exactly the form in Eq. (2.6) with $p(x) = R(x)$, $q(x) = \frac{R(x)p_2(x)}{p_0(x)}$ and $w(x) = \frac{R(x)}{p_0(x)}$. Table 2.1 lists $p_0(x)$, $p_1(x)$, $R(x)$, $w(x)$ for the classic polynomials used in this thesis.

Formula Eq. (2.6) can be written as $\mathcal{L}g(x) = \lambda g(x)$, where $\mathcal{L} = -\frac{1}{w(x)}\Big(\frac{d}{dx}[p(x)\frac{d}{dx}] + q(x)\Big)$ is a **self-adjoint operator**, because $\int_a^b (\mathcal{L}u(x))\overline{u(x)}\mathrm{d}x = \int_a^b u(x)\overline{(\mathcal{L}u(x))}\mathrm{d}x$, $\forall u(x)$. Thus Eq. (2.6) is actually to find the eigen-decomposition of the self-adjoint operator $\mathcal{L}$.

Sturm-Liouville problem is called **regular** if the boundary conditions are given as

$$\begin{aligned} \alpha_1 g(a) + \alpha_2 g'(a) = 0 \quad (\alpha_1^2 + \alpha_2^2 > 0) \\ \beta_1 g(b) + \beta_2 g'(b) = 0 \quad (\beta_1^2 + \beta_2^2 > 0) \end{aligned} \tag{2.13}$$

Theorem 2.2 (Sturm-Liouville Theorem). *Regular Sturm-Liouville problem has the following theoretical results.*

- *The eigenvalues of a self-adjoint operator are real and can be ordered such that* $\lambda_1 < \lambda_2 < \cdots < \lambda_n < \cdots$.

- *The eigenfunctions of a self-adjoint operator is an orthogonal basis in function space* $\mathbb{L}^2([a,b], w(x))$.

The second result is another way to prove the completeness of the function set. Please see [Arfken et al., 2005, chap. 10] for the proof of this theorem.

2.4 Polynomials in $\mathbb{R}^d$ and $\mathbb{S}^{d-1}$

The polynomials and functions in previous sections are defined in one dimension (1D). This section overviews some important concepts and results of functions and polynomials in d dimensional Euclidean space $\mathbb{R}^d$ and their restrictions in sphere $\mathbb{S}^{d-1}$.

If $\{f_1(x)\}$ and $\{f_2(y)\}$ are orthonormal bases in $\mathbb{L}^2([a_1,b_1], w_1(x))$ and $\mathbb{L}^2([a_2,b_2], w_2(y))$, obviously $\{f_1(x)f_2(y)\}$ is a orthonormal basis in $\mathbb{L}^2([a_1,b_1]\times[a_2,b_2], w_1(x)w_2(y))$ based on the inner product defined in the product space. So most concepts in $\mathbb{R}^1$ can be trivially generalized into $\mathbb{R}^d$. Stone-Weierstrass theorem can be applied into $\mathbb{R}^d$, i.e. the continuous function defined on a closed and bounded interval of $I \subset \mathbb{R}^d$ can be approximated uniformly by a d dimensional polynomial defined in I.

Let $\mathcal{P}_N(\mathbb{R}^d)$ denote the space of the polynomials with the order no more than N in $\mathbb{R}^d$. The monomial basis in $\mathcal{P}_N(\mathbb{R}^d)$ is $\{\prod_{j=1}^d x_j^{n_j}\}|_{\sum_{j=1}^d n_j<N}$. A polynomial $p_n(\boldsymbol{x})$ is called to be **homogeneous** with order n if $p_n(t\boldsymbol{x}) = t^n p_n(\boldsymbol{x})$. Based on the definition, homogeneous polynomial $p_n(\boldsymbol{x})$ is even when n is even, and is odd when n is odd. $\forall p_m(\boldsymbol{x}) \in \mathcal{P}_N(\mathbb{R}^d)$, $p_m(\boldsymbol{x})$ can be separated uniquely such that $p_m(\boldsymbol{x}) = \sum_{n=0}^m q_n(\boldsymbol{x})$, where $q_n(\boldsymbol{x}) \in \mathcal{HP}_n(\mathbb{R}^d)$, $\mathcal{HP}_n(\mathbb{R}^d) = \text{Span}\{\prod_{j=1}^d x_j^{n_j}\}|_{\sum_{j=1}^d n_j=n}$ is the space of homogeneous polynomials with order n. The dimension of $\mathcal{HP}_n(\mathbb{R}^d)$, i.e. the number of monomials in $\mathcal{HP}_n(\mathbb{R}^d)$, is $\binom{n+d-1}{d-1}$, because the index of the monomial in $\mathcal{HP}_n(\mathbb{R}^d)$ is determined uniquely by choosing different $d-1$ numbers from $(0, 1, ..., n+d-2)$.

Denote respectively the restrictions of $\mathcal{P}_N(\mathbb{R}^d)$ and $\mathcal{HP}_n(\mathbb{R}^d)$ in $\mathbb{S}^{d-1} \subset \mathbb{R}^d$ by $\mathcal{P}_N(\mathbb{S}^{d-1})$ and $\mathcal{HP}_n(\mathbb{S}^{d-1})$. Considering the natural inner product in $\mathbb{S}^{d-1}$ defined as $\langle p(\boldsymbol{x}), q(\boldsymbol{x})\rangle_{\mathbb{S}^{d-1}} = \int_{\mathbb{S}^{d-1}} p(\boldsymbol{x})\overline{q(\boldsymbol{x})}\mathrm{d}\boldsymbol{x}$, we have some interesting results.

Proposition 2.1. *$\mathcal{HP}_n(\mathbb{S}^{d-1}) \subset \mathcal{HP}_{n+2}(\mathbb{S}^{d-1})$, $\forall n \geq 0$. $\mathcal{P}_N(\mathbb{S}^{d-1}) = \mathcal{HP}_N(\mathbb{S}^{d-1}) \bigoplus \mathcal{HP}_{N-1}(\mathbb{S}^{d-1})$, $\forall N \geq 1$, where $\bigoplus$ means the **direct sum** of Hilbert spaces.*

Proof. For any given polynomial $p_n(\boldsymbol{x}) \in \mathcal{HP}_n(\mathbb{R}^d)$, $\boldsymbol{x} \in \mathbb{R}^d$, obviously $\|\boldsymbol{x}\|^2 p_n(\boldsymbol{x}) \in \mathcal{HP}_{n+2}(\mathbb{R}^d)$, which means $\mathcal{HP}_n(\mathbb{S}^{d-1}) \subset \mathcal{HP}_{n+2}(\mathbb{S}^{d-1})$, $\forall n \geq 0$. Thus $\forall p_n(\boldsymbol{x}) \in \mathcal{P}_N(\mathbb{S}^{d-1})$, $p_n(\boldsymbol{x}) = q_n(\boldsymbol{x}) + q_{n-1}(\boldsymbol{x})$, where $q_n(\boldsymbol{x}) \in \mathcal{HP}_n(\mathbb{S}^{d-1}) \subset \mathcal{HP}_N(\mathbb{S}^{d-1})$ and $q_{n-1}(\boldsymbol{x}) \in \mathcal{HP}_{n-1}(\mathbb{S}^{d-1}) \subset \mathcal{HP}_{N-1}(\mathbb{S}^{d-1})$. $\forall p_N(\boldsymbol{x}) \in \mathcal{HP}_N(\mathbb{R}^d)$ and $p_{N-1}(\boldsymbol{x}) \in \mathcal{HP}_{N-1}(\mathbb{R}^d)$, we have $\langle p_N(\boldsymbol{x}), p_{N-1}(\boldsymbol{x})\rangle_{\mathbb{S}^{d-1}} = 0$, because one is odd and the other one is even. So $\mathcal{HP}_N(\mathbb{R}^d)$ and $\mathcal{HP}_{N-1}(\mathbb{R}^d)$ are orthogonal. Then $\mathcal{P}_N(\mathbb{S}^{d-1}) = \mathcal{HP}_N(\mathbb{S}^{d-1}) \bigoplus \mathcal{HP}_{N-1}(\mathbb{S}^{d-1})$, $\forall N \geq 1$. □

Definition 2.3. *A function $g(\boldsymbol{x})$ is **harmonic** if $\Delta g(\boldsymbol{x}) = 0$, where $\boldsymbol{x} \in \mathbb{R}^d$ and $\Delta \stackrel{\text{def}}{=} \sum_{j=1}^{d} \frac{\partial^2}{\partial x_j^2}$ is the **Laplace operator** (aka **Laplace-Beltrami operator**). A polynomial $p_n(\boldsymbol{x})$ is called as a **harmonic polynomial** if $\Delta p_n(\boldsymbol{x}) = 0$,*

Denote the space of **homogeneous harmonic polynomials** with order n by $\mathcal{H}_n(\mathbb{R}^d)$, and its restriction in $\mathbb{S}^{d-1}$ as $\mathcal{H}_n(\mathbb{S}^{d-1})$. Harmonic polynomial plays an important role in polynomial theory. First, $\forall p_n(\boldsymbol{x}) \in \mathcal{HP}_n(\mathbb{R}^d)$, if $\Delta p_n(\boldsymbol{x}) \neq 0$, then $\Delta p_n(\boldsymbol{x}) \in \mathcal{HP}_{n-2}(\mathbb{R}^d)$. $\forall p_{n-2}(\boldsymbol{x}) \in \mathcal{HP}_{n-2}(\mathbb{R}^d)$, $\exists\, q_n(\boldsymbol{x}) \in \mathcal{HP}_n(\mathbb{R}^d)$ such that $\Delta q_n(\boldsymbol{x}) = p_{n-2}(\boldsymbol{x})$, which means Δ is a **surjection** from $\mathcal{HP}_n(\mathbb{R}^d)$ to $\mathcal{HP}_{n-2}(\mathbb{R}^d)$, and $\mathcal{H}_n(\mathbb{R}^d)$ is the **kernel**. So the dimension of $\mathcal{H}_n(\mathbb{R}^d)$ is

$$\binom{n+d-1}{d-1} - \binom{n+d-3}{d-1} \tag{2.14}$$

Second, Harmonic polynomials can be used to separate the homogeneous polynomial space based on the following theorems.

Theorem 2.3 (Orthogonality of homogeneous harmonic polynomials). *With the definition of the natural inner product in $\mathbb{S}^{d-1}$, $\mathcal{H}_n(\mathbb{S}^{d-1}) \perp \mathcal{H}_m(\mathbb{S}^{d-1})$, if $n \neq m$, where $\perp$ means two spaces are orthogonal, i.e. $\forall p_n(\boldsymbol{x}) \in \mathcal{H}_n(\mathbb{S}^{d-1})$ and $p_m(\boldsymbol{x}) \in \mathcal{H}_m(\mathbb{S}^{d-1})$, we have $\langle p_n(\boldsymbol{x}), p_m(\boldsymbol{x}) \rangle = 0$.*

Theorem 2.4 (Harmonic Decomposition Theorem). *$\forall p_n(\boldsymbol{x}) \in \mathcal{HP}_n(\mathbb{R}^d)$, $x \in \mathbb{R}^d$, it can be uniquely separated in the form $p_n(\boldsymbol{x}) = \sum_{j=0}^{[\frac{n}{2}]} \|\boldsymbol{x}\|^{2j} h_{n-2j}(\boldsymbol{x})$, where $h_{n-2j}(\boldsymbol{x}) \in \mathcal{H}_{n-2j}(\mathbb{R}^d)$, and $[\frac{n}{2}]$ is the maximal integer no more than $\frac{n}{2}$.*

Theorem 2.3 shows the orthogonality of homogeneous harmonic polynomial spaces with different orders, whose proof based on Green's second identity can be found in [Axler et al., 2001, chap. 5]. Theorem 2.4 shows that every multivariate polynomial over a field can be decomposed as a finite sum of products of a radical polynomial and a harmonic polynomial. Please refer [Axler et al., 2001, chap. 5] for the rigorous proof of it. A straightforward result based on these two theorems is

Corollary 2.1. *$\mathcal{HP}_n(\mathbb{S}^{d-1}) = \bigoplus_{j=0}^{[\frac{n}{2}]} \mathcal{H}_{n-2j}(\mathbb{S}^{d-1})$, where $[\frac{n}{2}]$ is the maximal integer no more than $\frac{n}{2}$.*

Based on Proposition 2.1 and Corollary 2.1, we have Corollary 2.2. This corollary shows that for every $p_n(x) \in \mathcal{P}_N(\mathbb{R}^d)$, although $p_n(x)$ may be not harmonic there always exists a harmonic polynomial $q_n(x) \in \bigoplus_{j=0}^{N} \mathcal{H}_j(\mathbb{R}^d)$, such that $p_n(x) = q_n(x)$, $\forall x \in \mathbb{S}^{d-1}$.

Corollary 2.2. $\mathcal{P}_N(\mathbb{S}^{d-1}) = \bigoplus_{j=0}^{N} \mathcal{H}_j(\mathbb{S}^{d-1})$.

Considering Stone-Weierstrass theorem in Theorem 2.1, we have

Corollary 2.3. $\mathbb{L}^2(\mathbb{S}^{d-1}, 1) = \lim_{N\to\infty} \mathcal{P}_N(\mathbb{S}^{d-1}) = \bigoplus_{j=0}^{\infty} \mathcal{H}_j(\mathbb{S}^{d-1})$.

2.5 Some Orthonormal Bases and Their Properties

In the previous sections, we have given an overview of some abstract concepts and results of orthogonal function, Gram–Schmidt orthogonalization, Stone-Weierstrass theorem, and Sturm-Liouville theory. Every specific polynomial (or function) has a lot of useful properties which can be used in different applications. In this section, we just list some properties of the polynomials and other special functions used in this thesis.

2.5.1 Associated Legendre Polynomial

We have demonstrated Legendre polynomial $\{P_n(x)\}$ from Gram-Schmidt orthogonalization in Section 2.2. Associated Legendre polynomial $\{P_n^m(x)\}$ is a generalized version of $\{P_n(x)\}$. It is defined as

$$P_n^m(x) = (-1)^m(1-x^2)^{\frac{m}{2}}\frac{d^m}{dx^m}(P_n(x)) \tag{2.15}$$

It satisfies the general Legendre equation in Eq. (2.16).

$$(1-x^2)g'' - 2xg' + (l(l+1) - \frac{m^2}{1-x^2})g = 0 \tag{2.16}$$

$P_n^m(x)$ becomes the Legendre polynomial $P_n(x)$ when $m = 0$. It satisfies the following orthogonality.

$$\int_{-1}^{1} P_n^m(x)P_{n'}^m(x)\mathrm{d}x = \frac{2(n+m)!}{(2n+1)(n-m)!}\delta_{nn'} \tag{2.17}$$

Please note that $P_n^m(x)$ is defined in the same domain $[-1,1]$ and has the same weight function $w(x) = 1$ as $P_n(x)$. See Table 2.1. However Legendre polynomial $\{P_n(x)\}$ is the unique form from Gram-Schmidt orthogonalization. Associated Legendre polynomials $\{P_n^m(x)\}$ actually form a larger space than polynomial space. For example, $P_1^1(x) = -(1-x^2)^{1/2}$ is not a polynomial, although normally we call $P_n^m(x)$ as associated Legendre polynomial. See the discussions in [Arfken et al., 2005, chap. 12] for more details.

2.5.2 Spherical Harmonics

Laplace's Equation in $\mathbb{R}^3$

In 3D space, the Laplace equation in spherical coordinate (R,θ,ϕ) is in Eq. (2.18), where $R \geq 0$, $\theta \in [0,\pi]$ and $\phi \in [0,2\pi)$, and Δ_b is the Laplace-Beltrami operator restricted in $\mathbb{S}^2$.

$$\Delta g = \frac{1}{R^2}\frac{\partial}{\partial R}\Big(R^2\frac{\partial g}{\partial R}\Big) + \frac{1}{R^2}\Delta_b g = 0 \tag{2.18}$$

$$\Delta_b g = \frac{1}{\sin^2\theta}\frac{\partial^2 g}{\partial\phi^2} + \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\Big(\sin\theta\frac{\partial g}{\partial\theta}\Big) \tag{2.19}$$

By assuming $g(R,\theta,\phi) = F(R)\Theta(\theta)\Phi(\phi)$, we have the following three equations based on separation of variables, where m and λ are separation constants which need to be determined for non-trivial solutions.

$$\Phi'' + m^2\Phi = 0 \tag{2.20}$$

$$\Theta'' + \frac{\cos\theta}{\sin\theta}\Theta' + (\lambda - \frac{m^2}{\sin^2\theta})\Theta = 0 \tag{2.21}$$

$$R^2F'' + 2RF' - \lambda F = 0 \tag{2.22}$$

From Eq. (2.20), $\Phi(\phi) = C_1e^{im\phi}+C_2e^{-im\phi}$, where C_1 and C_2 are constants. Because $\Phi(\phi) = \Phi(\phi+2\pi)$, m should be integer and $\Phi(\phi) = Ce^{im\phi}$. By letting $x = \cos\theta$, equation Eq. (2.21) becomes

$$(1-x^2)\frac{d^2\Theta}{dx^2} - 2x\frac{d\Theta}{dx} + (\lambda - \frac{m^2}{1-x^2})\Theta = 0 \tag{2.23}$$

This equation has bounded solution $P_l^m(x)$ if $\lambda = l(l+1)$, $l \geq |m|$, l and m are integers. So equation Eq. (2.22) becomes

$$R^2F'' + 2RF' - l(l+1)F = 0 \tag{2.24}$$

which has solutions $F(R) = C_1R^l + C_2R^{-(l+1)}$. Then without considering the constant, the final independent solutions of the 3D Laplace equation are

$$R^le^{im\phi}P_l^m(\cos(\theta)) \quad \text{and} \quad R^{-(l+1)}e^{im\phi}P_l^m(\cos(\theta)), \quad l \geq 0, l \geq |m| \tag{2.25}$$

The second kind of solution is the irregular solution which is singular at the origin point. Then the final general solution for Laplace's equation is

$$g(R,\theta,\phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}(a_l^mR^l + b_l^mR^{-(l+1)})y_l^m(\theta,\phi) \tag{2.26}$$

Definition, Orthogonality, Completeness

Considering the natural inner product in $\mathbb{S}^2$ defined as $\langle f(\theta,\phi), g(\theta,\phi)\rangle_{\mathbb{S}^2} = \int_0^{2\pi}\int_0^{\pi} f(\theta,\phi)\overline{g(\theta,\phi)}\sin(\theta)d\theta d\phi$, the spherical part in Eq. (2.25), i.e. $e^{im\phi}P_l^m(\cos(\theta))$ is orthogonal but not normalized.

Definition 2.4. *The Spherical Harmonics (SHs) are the normalized angular portion of the solution to Laplace's equation in spherical coordinates. Spherical Harmonics of order l and degree m, denoted by $y_l^m(\theta,\phi)$, are defined as*

$$y_l^m(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}e^{im\phi}P_l^m(\cos\theta) \tag{2.27}$$

***Real Spherical Harmonics**, called also spherical harmonics for short in this thesis, are defined as*

$$Y_l^m(\theta,\phi) = \begin{cases} \sqrt{2}Re(y_l^{|m|}(\theta,\phi)) & if \quad -l \leq m < 0 \\ y_l^m(\theta,\phi) & if \quad m = 0 \\ \sqrt{2}Im(y_l^m(\theta,\phi)) & if \quad l \geq m > 0 \end{cases} \tag{2.28}$$

where $Re(\cdot)$ and $Im(\cdot)$ mean the real and imaginary parts.

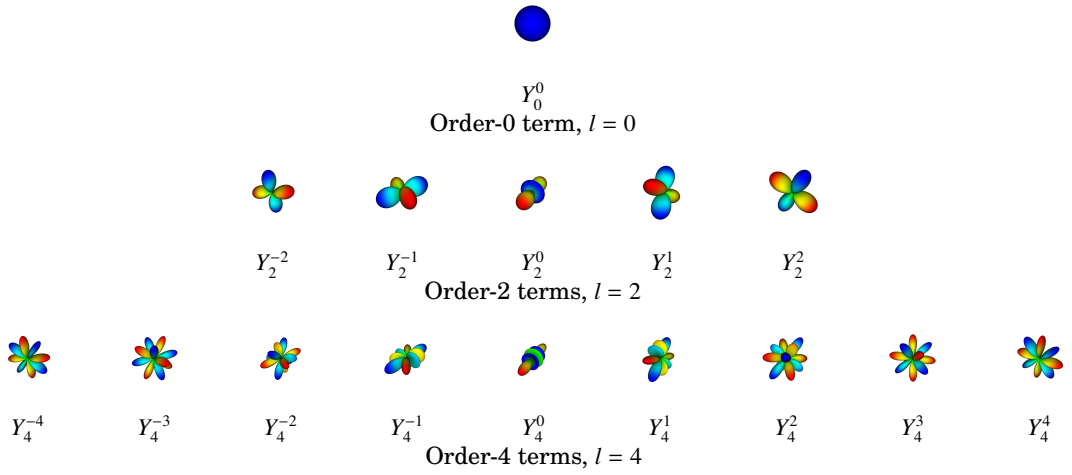


Figure 2.1: Symmetric real spherical harmonics up to order $L = 4$. The blue color in SHs with $l = 2, 4$ means negative values.

Thus $\{y_l^m(\theta,\phi)\}$ is an orthonormal system in $\mathbb{L}^2(\mathbb{S}^2, 1)$ by definition. It is also complete because $e^{im\phi}$ is complete in $[0, 2\pi)$ and $P_l^m(x)$ is complete in $[-1, 1]$. So $\forall g(\theta,\phi) \in \mathbb{L}^2(\mathbb{S}^2, 1)$, it can be represented as the so called **Laplace series** in Eq. (2.29)

$$g(\theta,\phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c_l^m y_l^m(\theta,\phi), \quad \text{where } c_l^m = \langle g(\theta,\phi), y_l^m(\theta,\phi)\rangle_{\mathbb{S}^2} \tag{2.29}$$

$\{Y_l^m(\theta,\phi)\}$ is an orthonormal basis in real valued functions in $\mathbb{L}^2(\mathbb{S}^2, 1)$.

We have shown that Spherical Harmonics $\{y_l^m\}$ can be obtained from Laplace's equation. SHs can be also obtained from Gram-Schmidt orthogonalization, because $\{\frac{1}{\sqrt{2\pi}} e^{-in\phi}\}$ and $\{P_n(x)\}$ can be obtained respectively from the orthogonalization of monomial basis in $\mathbb{S}^1$ and $[-1, 1]$. Please refer [Byron and Fuller, 1992, chap. 5] for the derivation of SHs from Gram-Schmidt orthogonalization.

Based on the definition in Eq. (2.27) and $P_l^{-m}(x) = (-1)^m \frac{(l-m)!}{(l+m)!} P_l^m(x)$, we have the property

$$\overline{y_l^m(\theta,\phi)} = (-1)^m y_l^{-m}(\theta,\phi) \tag{2.30}$$

$$y_l^m(\pi-\theta,\pi+\phi) = \begin{cases} y_l^m(\theta,\phi) & \text{if } l \text{ is even} \\ -y_l^m(\theta,\phi) & \text{if } l \text{ is odd} \end{cases} \tag{2.31}$$

Since $(\pi-\theta,\pi+\phi)$ is the antipodal point of (θ,ϕ), formula Eq. (2.31) means $y_l^m(\theta,\phi)$, as well as $Y_l^m(\theta,\phi)$, is even function if l is even, and is odd function if l is odd. So $\{Y_l^m(\theta,\phi)\}$ with even l, named as **symmetric real Spherical Harmonics**, are orthonormal basis of the space of all the antipodal symmetric spherical functions in $\mathbb{L}^2(\mathbb{S}^2, 1)$. Fig. 2.1 shows the symmetric real spherical harmonics up to order 4. Symmetric real SH basis plays an important role in HARDI methods as shown in Chapter 4, and it is used intensively throughout this thesis.

Based on the above analysis in Laplace's equation, we have Eq. (2.32) which is also applicable for $Y_l^m(\theta,\phi)$.

$$\Delta_b y_l^m(\theta,\phi) = \frac{1}{\sin^2\theta} \frac{\partial^2 y_l^m(\theta,\phi)}{\partial\phi^2} + \frac{1}{\sin\theta} \frac{\partial}{\partial\theta}\Big(\sin\theta \frac{\partial y_l^m(\theta,\phi)}{\partial\theta}\Big) = -l(l+1) y_l^m(\theta,\phi) \tag{2.32}$$

Based on the definition of SHs, the two kinds of solutions in Laplace's equation become $R^l y_l^m(\theta,\phi)$ which is called as **regular solid Harmonic**, and $R^{-(l+1)} y_l^m(\theta,\phi)$ which is called as **irregular solid harmonic**.

Theorem 2.5. *The regular solid harmonic* $\{R^l y_l^m(\theta,\phi)\}_{m=-l}^l$ *is a basis in* $\mathcal{H}_l(\mathbb{R}^3)$. *Spherical harmonic* $\{y_l^m(\theta,\phi)\}_{m=-l}^l$ *is an orthonormal basis in* $\mathcal{H}_l(\mathbb{S}^2)$.

Proof. $\forall p_l(R,\theta,\phi) \in \mathcal{HP}_l(\mathbb{R}^3)$, we have $p_l(R,\theta,\phi) = R^l q_l(\theta,\phi)$, where $q_l(\theta,\phi) = p_l(1,\theta,\phi) \in \mathcal{HP}_l(\mathbb{S}^2)$. Then

$$\begin{aligned}\Delta p_l(R,\theta,\phi) &= \frac{1}{R^2}\frac{\partial}{\partial R}\Big(R^2 \frac{\partial R^l q_l(\theta,\phi)}{\partial R}\Big) + \frac{1}{R^2}\Delta_b\Big(R^l q_l(\theta,\phi)\Big) \\ &= R^{l-2}\Big(l(l+1)q_l(\theta,\phi) + \Delta_b q_l(\theta,\phi)\Big)\end{aligned}$$

Thus, $\Delta p_l(R,\theta,\phi) = 0$ is equivalent to $\Delta_b q_l(\theta,\phi) = -l(l+1)q_l(\theta,\phi)$, which has the solution $\sum_{m=-l}^l C_m y_l^m(\theta,\phi)$, where $\{C_m\}$ is complex constants. So $p_l(R,\theta,\phi) = \sum_{m=-l}^l C_m R^l y_l^m(\theta,\phi)$, which means $\{R^l y_l^m(\theta,\phi)\}_{m=-l}^l$ is a basis in $\mathcal{H}_l(\mathbb{R}^3)$. Restricting $\{R^l y_l^m(\theta,\phi)\}_{m=-l}^l$ in $\mathbb{S}^2$, $\{y_l^m(\theta,\phi)\}_{m=-l}^l$ is a basis in $\mathcal{H}_l(\mathbb{S}^2)$. It is an orthonormal basis because $\langle y_l^m, y_{l'}^{m'}\rangle = \delta_{ll'}\delta_{mm'}$. □

The above theorem shows that

- The dimension of $\mathcal{H}_l(\mathbb{R}^3)$ is $\sum_{m=-l}^l 1 = 2l+1$, which agrees with formula Eq. (2.14).
- $\mathcal{H}_l(\mathbb{S}^2) \perp \mathcal{H}_{l'}(\mathbb{S}^2)$ if $l \neq l'$, which agrees with Theorem 2.3.
- For any polynomial $p(R,\theta,\phi) \in P_N(\mathbb{R}^3)$, we have $p(1,\theta,\phi) = \sum_{l=0}^N \sum_{m=-l}^l c_l^m y_l^m(\theta,\phi)$. No matter whether $p(R,\theta,\phi)$ is harmonic or not, $q(R,\theta,\phi) = \sum_{l=0}^N \sum_{m=-l}^l c_l^m R^l y_l^m(\theta,\phi)$ is always harmonic, which agrees with Corollary 2.2.

If we restrict the coefficients of monomials in polynomials to be real, $\{R^l Y_l^m\}_{m=-l}^l$ is a basis in $\mathcal{H}_l(\mathbb{R}^3)$ and $\{Y_l^m\}_{m=-l}^l$ is an orthonormal basis in $\mathcal{H}_l(\mathbb{S}^2)$.

Addition Theorem

$\{R^n \frac{1}{\sqrt{2\pi}} e^{-in\phi}\}_{n=0,\pm1,\cdots}$ are independent solutions in polar coordinate (R,ϕ) of Laplace's equation in two dimension (2D) space. While $\{R^l y_l^m(\theta,\phi)\}_{|m|\le l, l=0,1,\cdots}$ are independent solutions in spherical coordinate (R,θ,ϕ) of Laplace's equation. In this sense, Spherical Harmonic basis $\{y_l^m(\theta,\phi)\}$ can be seen as a generalization of Fourier basis $\{\frac{1}{\sqrt{2\pi}} e^{-in\phi}\}$ from $\mathbb{L}^2(\mathbb{S}^1, 1)$ to $\mathbb{L}^2(\mathbb{S}^2, 1)$.

Since $\mathrm{Re}(e^{-in\phi}e^{in\phi'}) = \mathrm{Re}(e^{-in\phi})\mathrm{Re}(e^{in\phi'}) - \mathrm{Im}(e^{-in\phi})\mathrm{Im}(e^{in\phi'})$, we have $\cos(n(\phi'-\phi)) = \cos(n\phi)\cos(n\phi') + \sin(n\phi)\sin(n\phi')$, which is the **addition theorem** in $\mathbb{S}^1$.

For SHs in $\mathbb{S}^2$, we have the following addition theorem.

Theorem 2.6 (Addition Theorem). *Let* $\boldsymbol{u}$ *and* $\boldsymbol{v}$ *are two unit vectors in* $\mathbb{S}^2$, *then*

$$P_l(\boldsymbol{u}^T\boldsymbol{v}) = \frac{4\pi}{2l+1}\sum_{m=-l}^{l} y_l^m(\boldsymbol{u})\overline{y_l^m(\boldsymbol{v})} \tag{2.33}$$

Please refer the detailed proof in [Arfken et al., 2005, chap. 12].

Product of Spherical Harmonics

Since $\{y_l^m(\theta,\phi)\}$ is an orthonormal basis in $\mathbb{L}^2(\mathbb{S}^2,1)$, $y_l^m y_{l'}^{m'}$ can be represented as [2] $\sum_{l=\alpha}^{\infty}\sum_{\beta=-\alpha}^{l} W_{ll'\alpha}^{mm'\beta}\overline{y_\alpha^\beta}$, where the coefficients $W_{ll'\alpha}^{mm'\beta}$ is

$$W_{ll'\alpha}^{mm'\beta} = \int_0^{2\pi}\int_0^{\pi} y_l^m y_{l'}^{m'} y_\alpha^\beta \sin\theta \mathrm{d}\theta\mathrm{d}\phi = \sqrt{\frac{(2l+1)(2l'+1)(2a+1)}{4\pi}}\begin{pmatrix} l & l' & \alpha \\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix} l & l' & \alpha \\ m & m' & \beta\end{pmatrix} \tag{2.34}$$

$\begin{pmatrix} l & l' & \alpha \\ m & m' & \beta\end{pmatrix}$ is the **Wigner 3j-symbol**, which is nonzero only if $m+m'+\beta=0$ and l, l', α satisfy triangular inequalities. Please refer [Arfken et al., 2005, chap. 12] for the proof.

Based on the definition of real spherical harmonics in Eq. (2.28), the integration

$$Q_{ll'\alpha}^{mm'\beta} = \int_0^{2\pi}\int_0^{\pi} Y_l^m Y_{l'}^{m'} Y_\alpha^\beta \sin\theta \mathrm{d}\theta\mathrm{d}\phi \tag{2.35}$$

is determined by $W_{ll'\alpha}^{mm'\beta}$. For example, if $m<0$, $m'<0$, $\beta<0$, and $\beta=m+m'$, then $Q_{ll'\alpha}^{mm'\beta} = \frac{(-1)^\beta}{\sqrt{2}}\int_{\mathbb{S}^2} y_l^m y_{l'}^{m'} y_\alpha^{-\beta}\sin\theta\mathrm{d}\theta\mathrm{d}\phi = \frac{(-1)^\beta}{\sqrt{2}} W_{ll'\alpha}^{mm'(-\beta)}$. $Q_{ll'\alpha}^{mm'\beta}$ in other cases is omitted.

Funk-Hecke Theorem

The Funk-Hecke theorem was proposed by Funk and Hecke. It is applicable in high dimensional sphere $\mathbb{S}^{d-1}$, whose proof can be found in [Andrews et al., 1999, chap. 9]. In this thesis, we only use the 3D case of this theorem.

Theorem 2.7. *Let $g(t)$ be a continuous function in $[-1,1]$ and y_l^m is the l order m degree spherical harmonic, then for given unit vector $\boldsymbol{u}$, we have*

$$\int_{\mathbb{S}^2} g(\boldsymbol{u}^T\boldsymbol{v}) y_l^m(\boldsymbol{v})\mathrm{d}\boldsymbol{v} = \lambda_l y_l^m(\boldsymbol{u}) \tag{2.36}$$

$$\lambda_l = 2\pi\int_{-1}^{1} P_l(t)g(t)\mathrm{d}t \tag{2.37}$$

where $P_l(t)$ is the Legendre polynomial of order l.

A corollary based on this theorem is

Corollary 2.4. *Let $\delta(t)$ be the Dirac delta function and y_l^m is the l order m degree spherical harmonic, then for given unit vector $\boldsymbol{u}$, we have*

$$\int_{\mathbb{S}^2} \delta(\boldsymbol{v}^T\boldsymbol{u}) y_l^m(\boldsymbol{v})\mathrm{d}\boldsymbol{v} = 2\pi P_l(0) y_l^m(\boldsymbol{u}) \tag{2.38}$$

where $P_l(0)$ is the Legendre polynomial of order l evaluated at 0.

$$P_l(0) = \begin{cases} 0 & \text{if } l \text{ is odd} \\ (-1)^{l/2}\dfrac{1\cdot 3\cdot 5\cdots(l-1)}{2\cdot 4\cdot 6\cdots l} = (-1)^{l/2}\dfrac{(l-1)!!}{l!!} & \text{if } l \text{ is even}\end{cases} \tag{2.39}$$

where $n!!$ is defined as $\prod_{j=0}^{[\frac{n-1}{2}]}(n-2j)$ for $n>0$ and 1 for $n=0,-1$.

[2] If $\sum_{l=\alpha}^{\infty}\sum_{\beta=-\alpha}^{l} W_{ll'\alpha}^{mm'\beta} y_\alpha^\beta$ is used, $\sum_{l=\alpha}^{\infty}\sum_{\beta=-\alpha}^{l} W_{ll'\alpha}^{mm'\beta} y_\alpha^\beta = \sum_{l=\alpha}^{\infty}\sum_{\beta=-\alpha}^{l} W_{ll'\alpha}^{mm'\beta}(-1)^\beta\overline{y_\alpha^\beta}$.

Since $\delta(t)$ is discontinuous at 0, the proof of this corollary is to define a sequence of functions $\{\delta_n(t)\}$ such that $\forall n$, $\delta_n(t)$ is continuous, and $\lim_{n\to\infty}\delta_n(t) = \delta(t)$. This sequence can be chosen as $\{\frac{n}{\sqrt{\pi}}e^{-n^2t^2}\}$ [Descoteaux et al., 2007]. If l is odd, $P_l(0) = 0$ because $P_l(x)$ is odd function for odd l. For even l, considering the recursive relation

$$(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x) \tag{2.40}$$

we have $\frac{P_{l+2}(0)}{P_l(0)} = -\frac{l+1}{l+2}$ and $P_l(0) = (-1)^{l/2}\frac{(l-1)!!}{l!!}$. Please note that based on the definition of y_l^m in Eq. (2.27) there is an equivalent solution for $P_l(0)$ that

$$P_l(0) = \sqrt{\frac{4\pi}{2l+1}}y_l^0(\frac{\pi}{2}, 0) \tag{2.41}$$

2.5.3 Bessel Function and Spherical Bessel Function

Definition, Orthogonality

Bessel function of the first kind, denoted by $J_\alpha(x)$ and called as Bessel function for short, is the solution of the Bessel's differential equation:

$$x^2g'' + xg' + (x^2 - \alpha^2)g = 0 \tag{2.42}$$

Normally α is used as integer or half of integer, although it can be chosen as real or complex number. When $\alpha = n$ is integer, the Bessel function $J_n(x)$ can be represented as the following integrals as

$$J_n(x) = \frac{1}{\pi}\int_0^\pi \cos(nt - x\sin t)\,\mathrm{d}t, \quad J_n(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-i(nt - x\sin t)}\mathrm{d}t \tag{2.43}$$

The **spherical Bessel function**, denoted by $j_n(x)$, is defined as

$$j_n(x) \overset{\text{def}}{=} \sqrt{\frac{\pi}{2x}}J_{n+1/2}(x) \tag{2.44}$$

There is an irregular spherical Bessel function

$$y_n(x) \overset{\text{def}}{=} (-1)^{n+1}\sqrt{\frac{\pi}{2x}}J_{-n-1/2}(x) \tag{2.45}$$

which is singular at $x = 0$. If not explicitly stated, spherical Bessel function in this thesis always means $j_n(x)$.

The Bessel equation in Eq. (2.42) can be transformed into the standard form in Eq. (2.6). Then by analyzing the regular conditions, we have the orthogonality of $\{J_\alpha\}$ as follows.

$$\int_0^\infty xJ_\alpha(bx)J_\alpha(cx)\mathrm{d}x = \frac{1}{b}\delta(b-c), \quad \text{where } \alpha > -\frac{1}{2} \tag{2.46}$$

Based on the definition of $j_n(x)$, we have the orthogonality for $\{j_n(x)\}$

$$\int_0^\infty x^2 j_n(bx)j_n(cx)\mathrm{d}x = \frac{\pi}{2b^2}\delta(b-c), \quad \text{where } n > -1 \tag{2.47}$$

Plane Wave Expansion Theorem

The well-known **Helmholtz equation** in $\mathbb{R}^d$ is in Eq. (2.48), which becomes the Laplace's equation when $k = 0$.

$$\Delta g(\boldsymbol{x}) + k^2 g(\boldsymbol{x}) = 0 \tag{2.48}$$

On the one hand, a heuristic solution of this equation is

$$g(\boldsymbol{x}) = e^{i\boldsymbol{x}^T\boldsymbol{k}} \tag{2.49}$$

where $\boldsymbol{k}$ can be any vector that satisfies $\|\boldsymbol{k}\|^2 = k^2$. This heuristic solution is called as **plane wave** which satisfies Helmholtz equation for any dimension d.

On the other hand, in three dimension (3D) space, similarly with Laplace's equation in Eq. (2.18) we can use separation of variables technique to solve Helmholtz equation. Let $g(R, \theta, \phi) = F(R)\Theta(\theta)\Phi(\phi)$, after some tedious derivations, we have the following general solution.

$$g(R, \theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (a_l^m j_l(kR) + b_l^m y_l(kR)) y_l^m(\theta, \phi) \tag{2.50}$$

So the heuristic solution can be represented in the form of Eq. (2.50). Considering the second term $y_l(kR)$ is singular at $R = 0$, we have

$$e^{i\boldsymbol{x}^T\boldsymbol{k}} = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_l^m j_l(\|\boldsymbol{k}\|R) y_l^m(\theta, \phi) \tag{2.51}$$

a_l^m is dependent only on $\boldsymbol{k}$ and can be solved analytically, which results in the Plane Wave Expansion theorem.

Theorem 2.8 (Plane Wave Expansion Theorem). *Let $\boldsymbol{x} = R_{\boldsymbol{x}}\boldsymbol{u}$ and $\boldsymbol{k} = R_{\boldsymbol{k}}\boldsymbol{v}$ be two 3D vectors in $\mathbb{R}^3$, and $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{S}^2$, we have*

$$e^{\pm i\boldsymbol{x}^T\boldsymbol{k}} = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} 4\pi(\pm i)^l j_l(R_{\boldsymbol{x}}R_{\boldsymbol{k}}) y_l^m(\boldsymbol{u}) \overline{y_l^m(\boldsymbol{v})} \tag{2.52}$$

The proof of this theorem is based on the **Green's function** representation of the general solution of Helmholtz equation. Please refer [Arfken et al., 2005, chap. 9] for the details. Note that the real symmetric part of $e^{\pm i\boldsymbol{x}^T\boldsymbol{k}}$ is $\cos(\boldsymbol{x}^T\boldsymbol{k})$ which can be written as

$$\cos(\boldsymbol{x}^T\boldsymbol{k}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} 4\pi(-1)^{l/2} j_l(R_{\boldsymbol{x}}R_{\boldsymbol{k}}) Y_l^m(\boldsymbol{u}) Y_l^m(\boldsymbol{v}) \tag{2.53}$$

where $\sum_{l=0}^{\infty}$ means the summation of only even order l. It can be proved easily by the Theorem 2.8, the definition of $\{Y_l^m\}$ in Eq. (2.28), and the property $y_l^{-m}(\mathbf{u}) = (-1)^m \overline{y_l^m(\mathbf{u})}$.

2.5.4 Hermite Polynomial

Physicists' Hermite polynomials

There are two kinds of standard notations for Hermite polynomials. One is the **probabilists' Hermite polynomial** which defines the polynomial as $He_n(x) = (-1)^n e^{\frac{x^2}{2}} \frac{d^n}{dx^n} e^{-\frac{x^2}{2}}$. The other one is the **physicists' Hermite polynomials** which defines the polynomial as

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2} \tag{2.54}$$

They are closely related, because $H_n(x) = 2^{n/2} He_n(\sqrt{2}x)$. In this thesis, unless otherwise specifically stated, we always use the physicists' Hermite polynomials in Eq. (2.54), which satisfies the equation $H_n''(x) - 2xH_n'(x) + 2nH_n(x) = 0$ and orthogonality in Eq. (2.55). See Table 2.1.

$$\int_{-\infty}^{\infty} H_n(x) H_m(x) e^{-x^2} \mathrm{d}x = 2^n \sqrt{\pi} n! \delta_{nm} \tag{2.55}$$

Based on Sturm-Liouville theorem 2.2, $\{H_n(x)\}$ is an orthogonal basis in $\mathbb{L}^2((-\infty, \infty), e^{-x^2})$.

Hermite Polynomial in Fourier Transform

Hermite polynomial can also be defined by the generating function in Eq. (2.56).

$$e^{-t^2+2tx} = \sum_{n=0}^{\infty} H_n(x) \frac{t^n}{n!} \tag{2.56}$$

Considering the **Fourier transform** in $\mathbb{R}^1$ defined in Eq. (4.11) we have

$$\mathscr{F}\{e^{-t^2+2tx-\frac{x^2}{2}}\}(\xi) = \sqrt{2\pi} e^{t^2-4\pi i t\xi-\frac{(2\pi\xi)^2}{2}}$$

So

$$\mathscr{F}\{\sum_{n=0}^{\infty} e^{-\frac{x^2}{2}} H_n(x) \frac{t^n}{n!}\} = \sqrt{2\pi} \sum_{n=0}^{\infty} e^{-\frac{(2\pi\xi)^2}{2}} H_n(2\pi\xi) \frac{(-it)^n}{n!}$$

Thus $e^{-\frac{x^2}{2}} H_n(x)$ is an eigenfunction of Fourier transform, i.e.

$$\mathscr{F}\{e^{-\frac{x^2}{2}} H_n(x)\}(\xi) = \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} H_n(x) e^{-2\pi i x\xi} \mathrm{d}x = \sqrt{2\pi}(-i)^n e^{-\frac{(2\pi\xi)^2}{2}} H_n(2\pi\xi) \tag{2.57}$$

Hermite Polynomial in Schrödinger Equation

In quantum mechanics, the square root of the probability of finding a particle at a certain time and position is called as **wavefunction**. Wavefunction is a complex valued function and it satisfies the well-known **Schrödinger equation**.

For single particle with a spherically symmetric **potential energy** $V(\|\boldsymbol{x}\|)$, the time-independent Schrödinger equation is given as

$$\Big(-\frac{\hbar^2}{2m}\Delta + V(\|\boldsymbol{x}\|)\Big)\psi(\boldsymbol{x}) = E\psi(\boldsymbol{x}) \tag{2.58}$$

where $\hbar$ is the reduced Planck constant, m is the mass, and E is the energy.

Quantum Simple Harmonic Oscillator (SHO) problem is one of the few quantum-mechanical systems for which an exact, analytical solution is known. The Schrödinger equation for SHO problem in 1D called as SHO-1D, is

$$\Big(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}m\omega^2x^2\Big)\psi(x) = E\psi(x) \tag{2.59}$$

where $V(x) = \frac{1}{2}m\omega^2x^2$ and ω is the angular frequency. The solution of this equation is

$$\psi_n(x) = 2^{-\frac{n}{2}}(n!)^{-\frac{1}{2}}(\pi\zeta)^{-\frac{1}{4}}e^{-\frac{x^2}{2\zeta}}H_n(\frac{x}{\sqrt{\zeta}}) \tag{2.60}$$

where $\zeta = \frac{\hbar}{m\omega}$. The corresponding energy is $E = \hbar\omega(n + \frac{1}{2})$. $\{\psi_n(x)\}$ is an orthonormal basis in $\mathbb{L}^2((-\infty,\infty),1)$ for given ζ, called as SHO-1D basis.

2.5.5 Associated Laguerre Polynomial

Orthogonality and Completeness

Associated Laguerre polynomial $L_n^\alpha(x)$ satisfies the following differential equation.

$$xg''(x) + (\alpha + 1 - x)g'(x) + ng(x) = 0 \tag{2.61}$$

For given α, $\{L_n^\alpha(x)\}_{n=1}^\infty$ is an orthogonal basis in $\mathbb{L}^2([0,\infty), x^\alpha e^{-x})$. It satisfies the following orthogonality:

$$\int_0^\infty x^\alpha e^{-x}L_n^\alpha(x)L_m^\alpha(x)\mathrm{d}x = \frac{\Gamma(n+\alpha+1)}{n!}\delta_{nm} \tag{2.62}$$

where $\Gamma(x)$ is the Gamma function, which can be seen in section 2.6.1. $L_n^\alpha(x)$ has the explicit form as

$$L_n^\alpha(x) = \sum_{j=0}^{n}(-1)^j\binom{n+\alpha}{n-j}\frac{x^j}{j!} \tag{2.63}$$

$L_n^\alpha(x)$ becomes the Laguerre polynomial $L_n(x)$ when $\alpha = 0$. Please note $\binom{n+\alpha}{n-j}$ has real number $n+\alpha$ in it. In this thesis, the binomial symbol is defined as binomial series

$$\binom{a}{k} = \frac{a(a-1)(a-2)\cdots(a-k+1)}{k(k-1)(k-2)\cdots 1} = \frac{\prod_{j=0}^{k-1}(a-j)}{k!} \tag{2.64}$$

where k is an integer number and a is arbitrary.

Relation with Hermite Polynomial

Compared to $\{L_n^\alpha(x)\}$ that is a basis in $\mathbb{L}^2([0,\infty), x^\alpha e^{-x})$, $\{H_n(x)\}$ is a basis in $\mathbb{L}^2((-\infty,\infty), e^{-x^2})$. When $\alpha = \pm\frac{1}{2}$, we have the following relation:

$$H_{2n}(x) = (-1)^n 2^{2n} n! L_n^{-\frac{1}{2}}(x^2) \tag{2.65}$$

$$H_{2n+1}(x) = (-1)^n 2^{2n+1} n! x L_n^{\frac{1}{2}}(x^2) \tag{2.66}$$

Relation with Monomials

Since $\{L_n^\alpha(x)\}$ is a basis, monomials can be represented by linear combination of $\{L_n^\alpha(x)\}$.

$$x^n = n! \sum_{j=0}^{n} (-1)^j \binom{n+\alpha}{n-j} L_j^\alpha(x) \tag{2.67}$$

Laguerre Polynomial in Schrödinger Equation

Historically in quantum mechanics, associated Laguerre polynomial $L_n^\alpha(x)$ occurs normally as the radial part of the solutions of Schrödinger equation in 3D space. The spherical harmonic y_l^m occurs in spherical part of the solution. A classic example is to determine the atomic orbitals of hydrogen-like ions in 3D space.

In Eq. (2.58) the Schrödinger equation for SHO problem in $\mathbb{R}^3$ called as SHO-3D, is given as

$$\Big(-\frac{\hbar^2}{2m}\Delta + \frac{1}{2}m\omega^2\|\boldsymbol{x}\|^2\Big)\psi(\boldsymbol{x}) = E\psi(\boldsymbol{x}) \tag{2.68}$$

The solution [Arfken et al., 2005, pp. 847] based on separation of variables in (R,θ,ϕ) is

$$\psi_{nlm}(R,\theta,\phi) = \left[\frac{2}{\zeta^{3/2}}\frac{(n-l/2)!}{\Gamma(n+l/2+3/2)}\right]^{1/2} (\frac{x^2}{\zeta})^{l/2} \exp\left(-\frac{x^2}{2\zeta}\right) L_{n-l/2}^{l+1/2}(\frac{x^2}{\zeta}) y_l^m(\theta,\phi) \tag{2.69}$$

where $\zeta = \frac{\hbar}{m\omega}$. The energy eigenvalue is $E = \hbar\omega(2n+\frac{3}{2})$. Please note that the notation n used here is shown as $\frac{n-1}{2}$ in [Arfken et al., 2005, pp. 847]. The normalization factor which makes the unit norm of ψ_{nlm} is determined based on Eq. (2.62).

Actually it is easy to verify that for given l the radial part are a set of orthonormal basis in $\mathbb{L}^2([0,\infty), q^2)$. Since the spherical part $\{y_l^m\}$ also forms an orthonormal basis in $\mathbb{L}^2(\mathbb{S}^2, 1)$, the function set $\{\psi_{nlm}(R,\theta,\phi)\}$ is an orthonormal basis in $\mathbb{L}^2(\mathbb{R}^3, 1)$ for given ζ, called as SHO-3D basis.

2.5.6 Confluent Hypergeometric Function

Kummer's Function

In mathematics, the **ordinary hypergeometric function** ${}_2F_1(a,b;c;x)$ is a special function represented by the hypergeometric series. It can represent many other special functions as specific or limiting cases.

In this thesis, we are more interested in **confluent Hypergeometric Function** ${}_1F_1(a;c;x)$, which satisfies the Kummer's equation.

$$xg'' + (c-x)g' - ag = 0 \tag{2.70}$$

This equation has two independent solutions. We are only interested in the one called as Kummer's function and given by the following form:

$${}_1F_1(a;c;x) \overset{\text{def}}{=} \sum_{n=0}^{\infty} \frac{a^{(n)}x^n}{b^{(n)}n!} \tag{2.71}$$

where $a^{(n)} \overset{\text{def}}{=} \prod_{j=0}^{n-1}(a+j) = a(a+1)\cdots(a+n-1)$. ${}_1F_1(a;c;x)$ is a limit of ${}_2F_1(a,b;c;x)$, i.e. $\lim_{b\to\infty} {}_2F_1(a,b;c;x/b) = {}_1F_1(a;c;x)$.

${}_1F_1(a;c;x)$ can be used to many functions. For example,

$$e^x = {}_1F_1(a;a;x) \tag{2.72}$$

$$L_n^{\alpha} = \binom{n+\alpha}{n} {}_1F_1(-n;\alpha+1;x) \tag{2.73}$$

Generalized Hypergeometric Function

The **generalized Hypergeometric function** is defined as

$${}_nF_m(a_1,\cdots,a_n;b_1,\cdots,b_m;x) \overset{\text{def}}{=} \sum_{j=0}^{\infty} \frac{(a_1)^{(j)}\cdots(a_n)^{(j)}}{(b_1)^{(j)}\cdots(b_m)^{(j)}} \frac{x^j}{j!} \tag{2.74}$$

2.6 Other Special Functions and Useful Equations

2.6.1 Gamma Function

The **gamma function** is defined as

$$\Gamma(x) \overset{\text{def}}{=} \int_0^{\infty} e^{-t}t^{x-1}\mathrm{d}t \tag{2.75}$$

It satisfies the following properties:

$$\Gamma(x+1) = x\Gamma(x), \quad \Gamma(\frac{1}{2}) = \sqrt{\pi} \tag{2.76}$$

So if n is a positive integer number, we have

$$\Gamma(n+1) = n\Gamma(n) = n! \tag{2.77}$$

$$n!! \overset{\text{def}}{=} \prod_{j=0}^{[\frac{n-1}{2}]}(n-2j) = \begin{cases} 2^{\frac{n}{2}}(\frac{n}{2})! = 2^{\frac{n}{2}}\Gamma(\frac{n}{2}+1) & \text{if } n \text{ is even} \\ \pi^{-\frac{1}{2}}2^{\frac{n+1}{2}}\Gamma(\frac{n}{2}+1) & \text{if } n \text{ is odd} \end{cases} \tag{2.78}$$

The **lower incomplete Gamma function** is defined as:

$$\gamma(s,x) \overset{\text{def}}{=} \int_0^x t^{s-1}e^{-t}\mathrm{d}t \tag{2.79}$$

2.6.2 Some Useful Integral Formulae

Laguerre polynomial $L_n^{\alpha}(x)$ and other useful special functions plays an important role throughout this thesis. Thus we list some useful integral formulae in this section. Please refer [Gradshteyn and Ryzhik, 2007] for these integral formulae. The page number after each equation shows where to find it in [Gradshteyn and Ryzhik, 2007].

$$\int_0^{\infty} x^m \exp(-\beta x^n)\mathrm{d}x = \frac{\Gamma(\frac{m+1}{n})}{n\beta^{\frac{m+1}{n}}} \quad (\beta > 0, m \geq 0, n > 0)^{3} \quad \text{[pp. 337]} \tag{2.80}$$

$$\int_0^{\infty} x^{\nu+1} e^{-\beta x^2} L_n^{\nu}(\alpha x^2) J_{\nu}(xy)\mathrm{d}x = 2^{-\nu-1}\beta^{-\nu-n-1}(\beta-\alpha)^n y^{\nu} e^{-\frac{y^2}{4\beta}} L_n^{\nu}\left(\frac{\alpha y^2}{4\beta(\alpha-\beta)}\right) \quad \text{[pp. 812]} \tag{2.81}$$

$$\int_0^{\infty} x^{\mu} \exp(-\alpha x^2) J_{\nu}(\beta x)\mathrm{d}x = \frac{\beta^{\nu}\Gamma(0.5\nu + 0.5\mu + 0.5)}{2^{\nu+1}\alpha^{0.5(\mu+\nu+1)}\Gamma(\nu+1)} {}_1F_1\left(\frac{\mu+\nu+1}{2}; \nu+1; -\frac{\beta^2}{4\alpha}\right) \tag{2.82}$$
$$(\alpha > -1, s > 0) \quad \text{[pp. 706]}$$

$$\int_0^{\infty} \exp(-st) t^{\alpha} L_n^{\alpha}(t)\mathrm{d}t = \frac{\Gamma(\alpha+n+1)(s-1)^n}{n! s^{\alpha+n+1}} \quad (\alpha > -1, s > 0) \quad \text{[pp. 809]} \tag{2.83}$$

$$\int_0^{\infty} \exp(-bx) L_n^{a}(t)\mathrm{d}x = \sum_{m=0}^{n} \binom{a+m+1}{m} \frac{(b-1)^{n-m}}{b^{n-m+1}} \quad (b > 0) \quad \text{[pp. 809]} \tag{2.84}$$

$$\int_0^{\infty} t^{b-1} {}_1F_1(a; c; -t)\mathrm{d}t = \frac{\Gamma(b)\Gamma(c)\Gamma(a-b)}{\Gamma(a)\Gamma(c-b)} \quad (0 < b < a) \quad \text{[pp. 821]} \tag{2.85}$$

$$\int_0^{\infty} x^{\mu} J_{\nu}(ax)\mathrm{d}x = 2^{\mu} a^{-\mu-1} \frac{\Gamma(\frac{1}{2} + \frac{\nu}{2} + \frac{\mu}{2})}{\Gamma(\frac{1}{2} + \frac{\nu}{2} - \frac{\mu}{2})} \quad (-\nu - 1 < \mu < \frac{1}{2}, \ a > 0) \quad \text{[pp. 676]} \tag{2.86}$$

2.7 SUMMARY

This chapter made a quick tutorial for some special functions and polynomials which play important role in many applications including diffusion MRI. Please note that these special functions and polynomials all have exhaustive useful properties, and we just showed only those properties used in this thesis. Please refer some textbooks for more properties.

We now summarize the important and useful mathematical properties of specific polynomials and functions which will be used in the following chapters.

[3] [Gradshteyn and Ryzhik, 2007] refers $m > 0$, however, we have tested the equation also holds when $m = 0$.

1. **Sturm-Liouville theorem** 2.2 and **Stone-Weierstrass theorem** 2.1 are very useful to prove the **completeness** and **orthogonality** of given function set.

2. **Harmonic Decomposition theorem** 2.4 is useful to study different function spaces in $\mathbb{R}^3$ parameterized by different models.

3. **Real spherical harmonic basis** $\{Y_l^m\}$ with even order l, i.e. **symmetric real SH basis**, is a useful orthonormal basis to represent antipodal symmetric spherical functions in dMRI. There are already exhaustive papers on Q-ball Imaging related methods which use SH basis to represent diffusion signal or other functions in dMRI.

4. **Plane wave expansion theorem** 2.8 is the most crucial theorem in this thesis, which will be used exhaustively in the following chapters to **solve Fourier transform analytically**.

5. **Funk-Hecke theorem** 2.7, **addition theorem** 2.6, the relation between Laplace-Beltrami operator and SHs in Eq. (2.32) are useful to derive analytic forms related with spherical harmonics.

6. **Hermite polynomial** $\{H_n(x)\}$ and **Laguerre polynomial** $\{L_n^\alpha\}$ are orthogonal bases respectively in $\mathbb{L}^2((-\infty,\infty), e^{-x^2})$ and $\mathbb{L}^2([0,\infty), x^\alpha e^{-x})$. They will be used to represent Gaussian-like functions in the following chapters.

7. Some important **integral formulae** listed in this chapter are useful to derive analytic forms of integrations.

Spherical Harmonic basis has becomes a natural and common tool in single shell HARDI (sHARDI) method [Frank, 2002; Hess et al., 2006; Anderson, 2005; Özarslan et al., 2006; Descoteaux et al., 2007; Tournier et al., 2007; Canales-Rodrıguez et al., 2009; Aganj et al., 2010b; Tristán-Vega et al., 2009; Tristán-Vega et al., 2010]. Many papers converged to use SH basis to represent diffusion signals in $\mathbb{S}^2$ and estimate the ODFs. In this thesis, one of the results is to generalize the sHARDI methods to multiple shell HARDI (mHARDI) methods, by considering not only SH basis in $\mathbb{S}^2$ but also an additional basis in radial part. We now on the way to describe how to devise an appropriate basis in 3D $\mathbf{q}$-space and analytically estimate the ODFs and EAPs from diffusion signals.

CHAPTER 3

MANIFOLD, STATISTICS ON MANIFOLD AND STATISTICAL MANIFOLD

"Geometry, which is the only science that it hath pleased God hitherto to bestow on mankind."

– Thomas Hobbes

Contents

OVERVIEW

Riemannian manifold was developed in the nineteenth century. It is very useful to analyze data in non-linear space. When analyzing and processing real data with given coordinates, sometimes addition and subtraction between two data points or multiplication with a negative value have no physical meaning. For example, when processing matrices in Sym_3^+, multiplying a 3×3 positive definite matrices with a negative value will obtain a symmetric matrix that is no longer in Sym_3^+. In this case, manifold based methods which consider the intrinsic properties of data space can be appropriate candidates of mathematical tools for data processing.

This chapter overviews some basic concepts about Riemannian manifold, statistics on manifold and statistical manifold which are the background knowledge of the Chapter 6 on the Riemannian framework for ODFs and EAPs. Please note that the subjects described in this chapter contain a lot of materials, while we only list some self-contained contents which will be used in the following Chapter 6. Please refer some textbooks [Boothby, 1986; Do Carmo, 1992; Spivak, 1999; Amari, 1985; Amari and Nagaoka, 2000] for more details of the contents.

Organization of this chapter:
First, some basic concepts on three layers of Riemannian manifold, i.e. topological structure, differentiable structure and Riemannian metric, are introduced in Section 3.1. We also give some examples on the geometry of high dimensional sphere to demonstrate the concepts. Second, in a computation point of view, Section 3.2 shows the definition, estimation, existence and uniqueness of weighted Riemannian mean and median, and demonstrate Principal Geodesic Analysis (PGA) method to find the principal components in manifold. At last, statistical manifold and Fisher information metric are introduced in Section 3.3, where we also provide two examples of statistical manifolds, which are the multivariate Gaussian distribution and multinomial distribution families.

3.1 RIEMANNIAN MANIFOLD

Riemannian manifold has three layers of structural concepts. The first one, which is defined for continuity and convergence, is the topological structure for topological manifold. The second one is the differentiable structure for differentiable manifold, which generalizes the differentiability from Euclidean space. The third layer for Riemannian manifold defines the Riemannian metric in infinitesimal neighborhood.

3.1.1 Topological Manifold

Definition 3.1. *A **topological space** $(\mathcal{X}, \mathcal{T})$ is a set $\mathcal{X}$ together with a topology $\mathcal{T}$ (a collection of subsets of $\mathcal{X}$) which satisfies:*

(i) The empty set $\varnothing$ and $\mathcal{X}$ are in $\mathcal{T}$.

(ii) Arbitrary union of elements of $\mathcal{T}$ is in $\mathcal{T}$.

(iii) Finite intersection of elements of $\mathcal{T}$ is in $\mathcal{T}$.

A topological space $(\mathcal{X}, \mathcal{T})$ is a set of points with a collection of well defined subsets. $(\mathcal{X}, \mathcal{T})$ is usually denoted by $\mathcal{X}$ for short. Elements in topology $\mathcal{T}$ are called as **open sets** of $\mathcal{X}$. Topology $\mathcal{T}$ can be generated from a topological basis.

Definition 3.2. *A **topological basis** on a set $\mathcal{X}$ is a collection $\mathcal{B}$ of subsets of $\mathcal{X}$ satisfying:*

(i) $\forall x \in \mathcal{X}$, $\exists B \in \mathcal{B}$, such that $x \in B$.

(ii) Let $B_1, B_2 \in \mathcal{B}$, $\forall x \in B_1 \cap B_2$, $\exists B_3 \in \mathcal{B}$, such that $x \in B_3$.

Example 3.1. *In Euclidean space $\mathbb{R}^n$, the standard basis is the open ball $B(x, r) = \{x' \in \mathbb{R}^n : \|x' - x\| < r\}$, which generates the open sets as the unions of open balls in the topology called **standard topology**.*

The open balls in $\mathbb{R}^n$ defined based on the distance $\|x' - x\|$ can be generalized to general topological space.

Definition 3.3 (Metric Space). *A **metric space** is a set $\mathcal{X}$ with a binary function $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ called **distance function** which satisfies the following items $\forall x, y, z \in \mathcal{X}$.*

(i) $d(x, y) \geq 0$, $d(x, y) = 0$ if and only if $x = y$.

(ii) $d(x, y) = d(y, x)$.

(iii) $d(x, z) \leq d(x, y) + d(y, z)$

If there exists such a distance function d in $\mathcal{X}$, the topology **induced** from d is the collection of all open balls defined as $B(x, r) = \{x' \in \mathcal{X} : d(x, x') < r\}$. For every two points $x, y \in \mathcal{X}$ with such induced topology, there exist at least two disjoint open sets U and V such that $x \in U$, $y \in V$. For example, $U = B(x, r)$, $V = B(y, r)$ and $r = \frac{1}{2}d(x, y)$. Note that this separation property, called as **Hausdorff**, is not satisfied in general topological space.

Definition 3.4. *A topological space $\mathcal{X}$ is called **Hausdorff** if for any two different point $x, y \in \mathcal{X}$ there exist two open sets U and V such that $U \bigcap V = \varnothing$ and $x \in U$, $y \in V$. $\mathcal{X}$ is called **connected** if there is no such pair of two open sets which satisfy $U \bigcap V = \varnothing$ and $\mathcal{X} = U \bigcup V$.*

The above concepts are for single set $\mathcal{X}$. It is useful to consider the relation between two topological spaces, which brings the concept of function.

Definition 3.5. *A **function** (aka **map**) $f : \mathcal{X} \to \mathcal{Y}$ is a relation between two topological spaces $\mathcal{X}$ and $\mathcal{Y}$ such that for any element $x \in \mathcal{X}$, there exists only one element $f(x) \in \mathcal{Y}$. It is a **bijection** if $\forall y \in \mathcal{Y}$, $f^{-1}(y) \in \mathcal{X}$ exists and is unique. It is **continuous** if for each open set $U \subset \mathcal{Y}$, the inverse image $f^{-1}(U) \subset X$ is also open.*

It can be seen that the above definition of continuity is a generalization of the well-known $\epsilon-\delta$ definition for the function between two Euclidean spaces. The function can be used to define **topological equivalence** between two topological spaces. When two topological spaces are equivalent, all their elements and open sets are related by a continuous bijection, called homeomorphism.

Definition 3.6. *A function $f : \mathcal{X} \to \mathcal{Y}$ is a **homeomorphism** between two topological spaces $\mathcal{X}$ and $\mathcal{Y}$, if f is a bijection and both f and f^{-1} are continuous. If there is a homeomorphism between $\mathcal{X}$ and $\mathcal{Y}$, we call $\mathcal{X}$ is **homeomorphic** to $\mathcal{Y}$.*

Since Euclidean space is well studied, it would be useful to generalize concepts from Euclidean space to topological space by locally relating the topological space to an Euclidean space via a homeomorphism.

Definition 3.7 (Topological Manifold). *A d dimensional **topological manifold** $\mathcal{X}$ is a **Hausdorff space** with a countable basis such that every point x has a neighborhood $U \in \mathcal{X}$ homeomorphic to an open set $\varphi_U(U) \in \mathbb{R}^d$, where the local homeomorphism $\varphi_U : U \in \mathcal{X} \to \varphi_U(U) \in \mathbb{R}^d$ is called a **coordinate chart** on U. If $\varphi_U(U)$ is an open set of half plane $\mathbb{R}^d_+ = \{\boldsymbol{x} = (x_1, \cdots, x_n)^T \in \mathbb{R}^d : x_n \geq 0\}$, then $\mathcal{X}$ is a **topological manifold with boundary**.*

Based on the definition, topological manifold $\mathcal{X}$ is covered by $\bigcup_{x\in\mathcal{X}} U$. The **cover** together with the charts, i.e. $\bigcup_{x\in\mathcal{X}}(U, \varphi_U)$, is called an **atlas** of $\mathcal{X}$.

Example 3.2 (Topology of Sphere)**.** *The sphere $\mathbb{S}^2$ is a subset of $\mathbb{R}^3$. The topology of $\mathbb{S}^2$ is the **subspace topology** of $\mathbb{R}^3$, i.e. a set $U \bigcap \mathbb{S}^2$ is called to be open in $\mathbb{S}^2$ if $U \subset \mathbb{R}^3$ is open in $\mathbb{R}^3$. A small enough local neighborhood of $\mathbb{S}^2$ is a surface patch can be represented as two parameters, which means the dimension of $\mathbb{S}^2$ is 2 and the neighborhood is homeomorphic to a neighborhood of $\mathbb{R}^2$. For $i = 1, 2, 3$, consider the hemispheres $U_i^+ = \{(x_1, x_2, x_3) \in \mathbb{S}^2 : x_i > 0\}$, $U_i^- = \{(x_1, x_2, x_3) \in \mathbb{S}^2 : x_i < 0\}$. These 6 hemispheres are open sets of $\mathbb{S}^2$. Then the 6 homeomorphisms $\varphi_i^\pm : U_i^\pm \subset \mathbb{S} \to \varphi(U_i^\pm) \subset \mathbb{R}^2$ can be simply chosen as $\varphi_i^\pm(x_1, x_2, x_3) \mapsto (x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_3)$. These 6 hemispheres and corresponding homeomorphisms form an atlas in $\mathbb{S}^2$. Note that the neighborhood can be even much bigger. For example the neighborhood can be $U = \mathbb{S}^2 - \{(0, 0, 1)\} = \{(x_1, x_2, x_3) \in \mathbb{S}^2 : (x_1, x_2, x_3) \neq (0, 0, 1)\}$, and the homeomorphism can be chosen as the **stereographic projection**. However we can not find a global coordinate chart from $\mathbb{S}^2$ to $\mathbb{R}^2$ because they are is not homeomorphic. The analysis here can be used in $d-1$ dimensional sphere $\mathbb{S}^{d-1}$.*

Example 3.3 (Topology of the Space of Positive Definite Matrices)**.** *Sym_3^+ is a topological manifold which is locally homeomorphic to the space of 3×3 symmetric matrix Sym_3. The space of 3×3 symmetric matrix with non-negative eigenvalues is a topological manifold with boundary. The boundary contains all matrices with non-negative eigenvalues and at least one zero eigenvalues. The coordinate chart for these two spaces can be chosen as the identity map to Sym_3.*

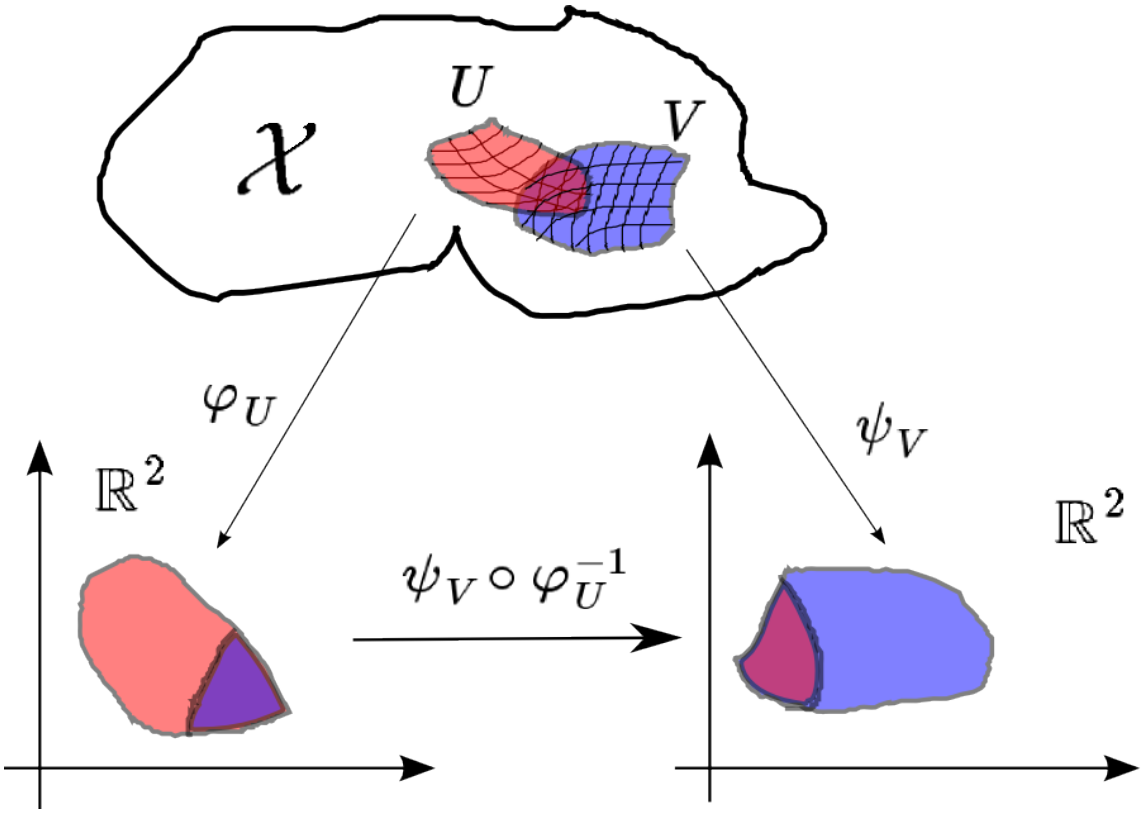


Figure 3.1: Sketch map for local coordinate charts and transition map.

3.1.2 Differentiable Manifold

Differentiable Structure and Definition

For two charts φ_U and ψ_V with overlapping domains U and V, the **transition function**

$$\psi_V \circ \varphi_U^{-1} : \ \varphi_U(U \cap V) \subset \mathbb{R}^d \to \psi_V(U \cap V) \subset \mathbb{R}^d \tag{3.1}$$

is a homeomorphism from a subset of $\mathbb{R}^d$ to another subset of $\mathbb{R}^d$. $\psi_V \circ \varphi_U^{-1}$ is continuous based on the definition. Please see Fig. 3.1 for the sketch map of local coordinate charts and transition map.

In order to perform calculus in $\mathcal{X}$, we need this transition function is not only continuous but also differentiable. The transition function $\psi_V \circ \varphi_U^{-1}$ is called **C^k differentiable**, or $\mathbf{C}^k$ for short, if its partial derivatives are all k times differentiable. An atlas is $\mathbf{C}^k$ if all possible transition functions in it are $\mathbf{C}^k$. A topological manifold can have many atlases. For two atlases $\mathcal{A}_1$ and $\mathcal{A}_2$ of a manifold, their combination is still an atlas. However, if $\mathcal{A}_1$ and $\mathcal{A}_2$ are $\mathbf{C}^k$ differentiable, their combination may be not $\mathbf{C}^k$ differentiable. If their combination is still $\mathbf{C}^k$ differentiable, these atlases are called to be **compatible**. The **maximal $\mathbf{C}^k$ atlas**, also called **differentiable structure**, is the atlas which contains all compatible $\mathbf{C}^k$ atlases. The maximal atlas is proved to be unique.

Definition 3.8 (Differentiable Manifold). *A $\mathbf{C}^k$ **differentiable manifold** is a topological manifold equipped with the maximal $\mathbf{C}^k$ atlas.*

Example 3.4 (Sphere as a Differentiable Manifold)**.** *Recall the atlas of $\mathbb{S}^2$ which contains 6 hemispheres shown in Example 3.2. It is easy to verify that these atlases are compatible and there exists a maximal atlas which contains all of them.*

Normally when we talk about differentiable manifold, we mean the $\mathbf{C}^k$ differentiable manifold with $k \geq 1$. In the following text, unless specifically stated otherwise, the term "differentiable manifold" always means a $\mathbf{C}^k$ differentiable manifold with $k \geq 1$.

The transition map has been used to define the differentiability of coordinate maps. It can also be used to define the differentiability of a general function between two manifolds.

Definition 3.9 (Diffeomorphism)**.** *For differentiable manifold $\mathcal{X}$ and $\mathcal{Y}$ with dimension d_1 and d_2, a function $f : \mathcal{X} \to \mathcal{Y}$ is **differentiable** if for every two charts $\varphi_U : U \subset \mathcal{X} \to \varphi_U(U) \in \mathbb{R}^{d_1}$ and $\psi_V : V \subset \mathcal{Y} \to \varphi_V(V) \in \mathbb{R}^{d_2}$, the function $\psi_V \circ f \circ \varphi_U^{-1} : \varphi_U(U) \in \mathbb{R}^{d_1} \to \varphi_V(V) \in \mathbb{R}^{d_2}$ is differentiable. $f : \mathcal{X} \to \mathcal{Y}$ is a **diffeomorphism** if f is a differentiable homeomorphism.*

Note that when $f : \mathcal{X} \to \mathcal{Y}$ is a diffeomorphism, based on **inverse function theorem** these two manifolds have the same dimension, i.e. $d_1 = d_2$. Homeomorphism is the topological equivalence between topological spaces $\mathcal{X}$ and $\mathcal{Y}$. While diffeomorphism is the equivalence between differentiable manifolds.

When $\mathcal{Y} = \mathbb{R}^1$ is a Euclidean space, $f : \mathcal{X} \to \mathbb{R}^1$ is a scalar function defined in $\mathcal{X}$, which is essential to introduce tangent vector and tangent space.

Tangent Vector and Tangent Space

In Euclidean space $\mathbb{R}^d$, let $\gamma : [-\epsilon, \epsilon] \to \mathbb{R}^d$ be a smooth curve passing across $p \in \mathbb{R}^d$ with $\gamma(0) = p \in \mathbb{R}^d$. In the coordinate $\boldsymbol{x} = (x_1, x_2, \cdots, x_d)^T \in \mathbb{R}^d$, $\gamma(t) = \boldsymbol{x}(t) = (x_1(t), x_2(t), \cdots, x_d(t))^T$, $t \in [-\epsilon, \epsilon]$. Then $\boldsymbol{v} = \gamma'(0) = (x_1'(0), x_2'(0), \cdots, x_d'(0))^T \in \mathbb{R}^d$. Let $f : \mathbb{R}^d \to \mathbb{R}^1$ be a differentiable scalar function defined in a neighborhood of p. Then the **directional derivative** of f along $\boldsymbol{v}$ is $\frac{d(f\circ\gamma)}{dt}|_{t=0} = \sum_{i=1}^d \frac{\partial f}{\partial x_i}|_{\boldsymbol{x}=p} \frac{dx_i}{dt}|_{t=0} = \sum_{i=1}^d (v_i \frac{\partial}{\partial x_i}) f$. So the tangent vector $\boldsymbol{v}$ can be seen as an operator on the differentiable functions. This concept can be generalized into differentiable manifold.

Definition 3.10 (**Tangent Vector and Tangent Space**)**.** *Let $\gamma : [-\epsilon, \epsilon] \to \mathcal{X}$ with $\gamma(0) = p \in \mathcal{X}$ be a differentiable curve cross p in a differentiable manifold $\mathcal{X}$. Let f be any differentiable function defined in a neighborhood of p. Then the **tangent vector** to the curve $\gamma(t)$ at $t = 0$ is given by the operator $\gamma'(0)$ that maps function f to its directional derivative, i.e.*

$$\gamma'(0) f = \left. \frac{df \circ \gamma}{dt} \right|_{t=0} \tag{3.2}$$

*The space of all tangent vectors in a given point $p \in \mathcal{X}$ is called **tangent space** and indicated by $T_p\mathcal{X}$.*

Let (U, φ) be local coordinate system around p, then $f(p) = f(\varphi(p)) = f(x_1, x_2, \cdots, x_d)$, $\boldsymbol{x} = \varphi(p) = (x_1, x_2, \cdots, x_d)^T \in \mathbb{R}^d$. Then $\varphi \circ \gamma : [-\epsilon, \epsilon] \in \mathbb{R}^1 \to \varphi(\gamma([-\epsilon, \epsilon]) \bigcap U) \in \mathbb{R}^d$ is a curve in $\mathbb{R}^d$, which is the image of the overlap between

$\gamma([-\epsilon,\epsilon])$ and U. Thus denote $\varphi(\gamma(t)) = \boldsymbol{x}(t) = (x_1(t), x_2(t), \cdots, x_d(t))^T \in \mathbb{R}^d$, $t \in [-\epsilon,\epsilon]$, then the directional derivative of f can be represented as

$$\begin{aligned}\gamma'(0)f &= \frac{d}{dt}(f\circ\gamma)\Big|_{t=0} = \frac{d}{dt}f(x_1(t), x_2(t), \cdots, x_d(t))\Big|_{t=0} \\ &= \sum_{i=1}^{d}\frac{dx_i(t)}{dt}\Big|_{t=0}\frac{\partial f}{\partial x_i}\Big|_{\boldsymbol{x}(t)=\varphi(p)} = (\sum_{i=1}^{d}x_i'(0)\frac{\partial}{\partial x_i}\Big|_{\boldsymbol{x}(t)=\varphi(p)})f\end{aligned}$$

So the vector $\gamma'(0)$ can be represented in coordinate chart φ by

$$\gamma'(0) = \sum_{i=1}^{d}x_i'(0)\frac{\partial}{\partial x_i}\Big|_p \tag{3.3}$$

Eq. (3.3) shows the tangent vector at p depends only on the derivative $\gamma'(0)$ in a given coordinate chart. It can be proved that the tangent space $T_p\mathcal{X}$ with the usual operations of functions forms a vector space with the same dimension d as $\mathcal{X}$, and a given coordinate φ induces a basis $\{\frac{\partial}{\partial x_i}\}_{i=1}^d$ in $T_p\mathcal{X}$.

$$T_p\mathcal{X} = \left\{\sum_{i=1}^{d}c_i\frac{\partial}{\partial x_i}\Big|_p : \quad (c_1, c_2, \cdots, c_d)^T \in \mathbb{R}^d\right\} \tag{3.4}$$

Example 3.5 (Tangent Space of Sphere)**.** *Let $p \in \mathbb{S}^{d-1}$ and $\gamma : [-\epsilon,\epsilon] \to \mathbb{S}^{d-1}$ with $\gamma(0) = p$. Let $f(\gamma(t)) = \|\gamma(t)\|^2 = 1$ be a function restricted in $\gamma(t)$ which is a constant since $\gamma(t) \in \mathbb{S}^{d-1}$, $\forall t \in [-\epsilon,\epsilon]$. Choose a coordinate such that $p = (p_1, \cdots, p_d)^T$ and $\gamma(t) = (x_1(t), \cdots, x_d(t))^T$. Then $\frac{\partial f}{\partial x_i}|_p = 2\gamma(0) = 2p$, and $0 = \frac{\partial f(\gamma(t))}{\partial t} = \sum_{i=1}^d x_i'(0)\frac{\partial f}{\partial x_i}|_p = 2\sum_{i=1}^d x_i'(0)p_i$. So the tangent space $T_p\mathbb{S}^{d-1}$ is*

$$T_p\mathbb{S}^{d-1} = \left\{\boldsymbol{v} = (v_1, v_2, \cdots, v_d)^T \in \mathbb{R}^d : \ \boldsymbol{v}^T p = 0\right\} \tag{3.5}$$

The inverse of the coordinate chart, i.e. φ^{-1}, is a diffeomorphism which actually maps the local basis system $\{\frac{\partial}{\partial x_i}\}_{i=1}^d$ from $T_{\boldsymbol{x}(0)}\mathbb{R}^d$ to $T_p\mathcal{X}$. This can be generalized into differentiable function between two manifolds.

Definition 3.11. *Let $\varphi : \mathcal{X} \to \mathcal{Y}$ be a differentiable function between two differentiable manifolds $\mathcal{X}$ and $\mathcal{Y}$. For a given point $p \in \mathcal{X}$, the **pushforward map of** φ (aka derivative of φ) is a linear map from the $T_p\mathcal{X}$ to $T_{\varphi(p)}\mathcal{Y}$, i.e. $\varphi_*|_p : T_p\mathcal{X} \to T_{\varphi(p)}\mathcal{Y}$. The **pullback map of** φ is a linear map from $T_{\varphi(p)}\mathcal{Y}$ to $T_p\mathcal{X}$, i.e. $\varphi^*|_p : T_{\varphi(p)}\mathcal{Y} \to T_p\mathcal{X}$.*

In other words, $\forall X \in T_p\mathcal{X}$ and differentiable function $f : \mathcal{Y} \to \mathbb{R}^1$, we have $f\circ\varphi$ is a differentiable function defined in $\mathcal{X}$. Then $(\varphi_*X)_p \in T_{\varphi(p)}\mathcal{Y}$ is defined as $(\varphi_*X)(f) = X(f\circ\varphi)$. $\forall Y \in T_{\varphi(p)}\mathcal{Y}$ and differentiable function $g : \mathcal{X} \to \mathbb{R}^1$, we have $g\circ\varphi^{-1}$ is a differentiable function defined in $\mathcal{Y}$. Then $(\varphi^*Y)_{\varphi(p)} \in T_p\mathcal{X}$ is defined as $(\varphi^*Y)(g) = Y(g\circ\varphi^{-1})$.

3.1.3 Riemannian Manifold

Riemannian Metric

A distance function used to distinguish two different points in a manifold is essential for manifold statistics and devising algorithms.

Let $\gamma : [a, b] \to \mathbb{R}^d$ be a differentiable curve in Euclidean space $\mathbb{R}^d$. Then $\gamma'(t_0)$ for any given $t_0 \in [a, b]$ is the velocity of the curve at $t = t_0$, and the length of the curve is given as

$$L(\gamma) = \int_a^b \|\gamma'(t)\| \mathrm{d}t$$

The line segment connecting two points $p, q \in \mathbb{R}^d$ is proved to have the minimal length among all such differentiable curves $\gamma(t)$ with $\gamma(a) = p$ and $\gamma(b) = q$. The length of the line segment is normally used as the distance between p and q.

For $\gamma : [a, b] \to \mathcal{X}$ as a differentiable curve in a general differentiable manifold $\mathcal{X}$, we have shown in last subsection that $\gamma'(t_0)$ is a tangent vector in the tangent space $T_{\gamma(t_0)}\mathcal{X}$. Thus we need to define the length of the curve $\gamma(t)$ in a general differentiable manifold, which needs to equip a **norm** induced by an **inner product** in $T_{\gamma(t)}$.

Definition 3.12 (Riemannian Metric and Riemannian Manifold). *Let $\mathcal{X}$ be a differentiable manifold. A **Riemannian metric** on $\mathcal{X}$ is defined as an inner product (a symmetric, bilinear, positive definite form) in the tangent space $T_p\mathcal{X}$, i.e.*

$$g_p = \langle \cdot, \cdot \rangle_p : T_p\mathcal{X} \times T_p\mathcal{X} \to \mathbb{R}^1, \quad p \in \mathcal{X} \tag{3.6}$$

*such that with respect to the coordinate $\varphi : U \in \mathcal{X} \to \mathbb{R}^d$, $\varphi(p) = (x_1, x_2, \cdots, x_d) \in \varphi(U)$, the function $g_{ij}(x_1, x_2, \cdots, x_d) \overset{\text{def}}{=} \langle \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \rangle_p$ is differentiable in U. **Riemannian manifold** $(\mathcal{X}, g)$, or written as $(\mathcal{X}, \langle \cdot, \cdot \rangle)$, is a differentiable manifold $\mathcal{X}$ together with a Riemannian metric g or $\langle \cdot, \cdot \rangle$.*

Recall that for a given coordinate $\varphi : U \in \mathcal{X} \to \mathbb{R}^d$, $\varphi(p) = (x_1, x_2, \cdots, x_d) \in \varphi(U)$, $\{\frac{\partial}{\partial x_i}\}_{i=1}^d$ forms a basis in $T_p\mathcal{X}$, which defines an $d \times d$ positive definite matrix $G = [g_{ij}]$ with the elements g_{ij}. The matrix G is the local representation of Riemannian metric under coordinate φ. Thus for two tangent vectors $X = \sum_{i=1}^d X_i \frac{\partial}{\partial x_i}$ and $Y = \sum_{i=1}^d Y_i \frac{\partial}{\partial x_i}$, their inner product is

$$\langle X, Y \rangle_p = \langle \sum_{i=1}^d X_i \frac{\partial}{\partial x_i}, \sum_{i=1}^d Y_i \frac{\partial}{\partial x_i} \rangle_p = \boldsymbol{x}^T G \boldsymbol{y} \tag{3.7}$$

where $\boldsymbol{x} = (X_1, X_2, \cdots, X_d)^T$ and $\boldsymbol{y} = (Y_1, Y_2, \cdots, Y_d)^T$ are the local representations of X and Y under φ. Please note that for given X, Y and metric g_p, different coordinates will obtain different representations of the basis $\{\frac{\partial}{\partial x_i}\}_{i=1}^d$, $\boldsymbol{x}$, $\boldsymbol{y}$ and G. But the inner product $\langle X, Y \rangle_p = \boldsymbol{x}^T G \boldsymbol{y}$ remains the same.

With the Riemannian metric, the norm of tangent vector $X = \sum_{i=1}^d X_i \frac{\partial}{\partial x_i}$ is

$$\|X\|_p \overset{\text{def}}{=} \sqrt{\langle X, X \rangle_p} = \sqrt{\boldsymbol{x}^T G \boldsymbol{x}} \tag{3.8}$$

For the last two layers, we have shown homeomorphism and diffeomorphism as the equivalence respectively between two topological manifolds and differentiable manifolds. With the additional Riemannian metric in the third layer, the equivalence between two Riemannian manifolds is called as isometry which means in the sense of preserving lengths of curves in addition to diffeomorphism.

Definition 3.13. *Let $(\mathcal{X}, \langle\cdot,\cdot\rangle_{\mathcal{X}})$ and $(\mathcal{Y}, \langle\cdot,\cdot\rangle_{\mathcal{Y}})$ be two Riemannian manifolds. Let $f : \mathcal{X} \to \mathcal{Y}$ be a diffeomorphism. Then f is called as an **isometry** if $\langle \boldsymbol{u}, \boldsymbol{v}\rangle_{\mathcal{X}} = \langle f_* \boldsymbol{u}, f_* \boldsymbol{v}\rangle_{\mathcal{Y}}$, $\forall \boldsymbol{u}, \boldsymbol{v} \in T_p\mathcal{X}$, $p \in \mathcal{X}$. If the metric in $\langle\cdot,\cdot\rangle_{\mathcal{Y}}$ is defined such that diffeomorphism f becomes an isometry, we call the metric $\langle\cdot,\cdot\rangle_{\mathcal{Y}}$ is **induced** by $(\mathcal{X}, \langle\cdot,\cdot\rangle_{\mathcal{X}})$ and $f : \mathcal{X} \to \mathcal{Y}$.*

Please note that the term "metric" in "Riemannian metric" is different from "metric" in "metric space" in Definition 3.3. The "metric" here is a metric tensor which describes an inner product in the tangent space of one point. While the "metric" in Definition 3.3 is a distance function defined for any two points, which is the geodesic distance shown as follows.

Geodesic

With the Riemannian metric, the length of a differentiable curve $\gamma : [a, b] \to \mathcal{X}$ in Riemannian manifold $\mathcal{X}$ is given as

$$L(\gamma) = \int_a^b \|\gamma'(t)\|_{\gamma(t)} \mathrm{d}t = \int_a^b \sqrt{\langle \gamma'(t), \gamma'(t)\rangle_{\gamma(t)}} \mathrm{d}t \tag{3.9}$$

Note the length is independent with the chosen coordinate system, since the inner product $\|\gamma'(t)\|_{\gamma(t)} = \sqrt{\langle \gamma'(t), \gamma'(t)\rangle_{\gamma(t)}}$ is independent of chosen coordinate.

Definition 3.14 (Geodesic). *The **geodesic distance** between two points p, q in a Riemannian manifold $\mathcal{X}$ is the infimum of the length among all differentiable curves $\gamma : [a, b] \to \mathcal{X}$ with $\gamma(a) = p$, $\gamma(b) = q$, i.e.*

$$d(p, q) = \inf\{L(\gamma) : \gamma(t) \text{ is continuous and piecewise differentiable, } \gamma(a) = p, \gamma(b) = q\} \tag{3.10}$$

*A **geodesic** between p and q is a piecewise differentiable curve with the minimal length $d(p, q)$.*

Please note three things. First, we relax the constraint such that the curves may not be differentiable in some discrete points, because the discrete points have no contribution in the integral. Second, the geodesic may not exist, which means the infimum of length can not be reached by a curve. Third, if a geodesic exists, there may be more than one geodesic between two points.

It can be proved that the distance defined in Eq. (3.10) satisfies the conditions for distance function listed in Definition 3.3. So with the Definition 3.14 of geodesic distance and Definition 3.3 of the metric space, the Riemannian manifold becomes a metric space if any two points have a geodesic.

The geodesic can be obtained by minimizing the length functional in Eq. (3.9), which is a **Euler-Lagrange equation** and can be solved by standard **calculus of variations** technique [Spivak, 1999, chap. 9]. Then we obtain the **geodesic equation** as

$$\frac{d^2 x_k}{dt^2} + \sum_{i,j=1}^{d} \Gamma_{ij}^k(\gamma(t)) \frac{dx_i}{dt} \frac{dx_j}{dt} = 0 \tag{3.11}$$

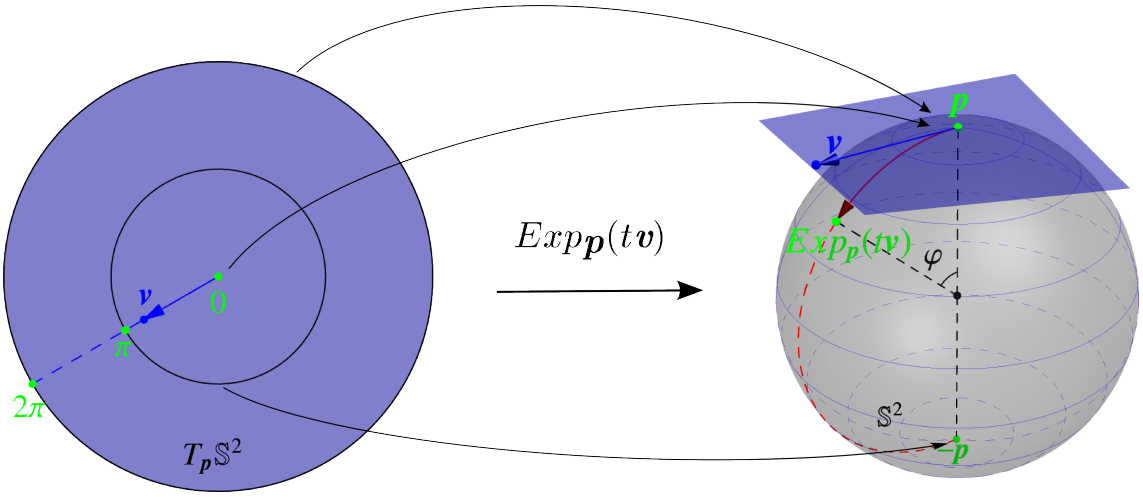


Figure 3.2: Tangent space $T_{\boldsymbol{p}}\mathbb{S}^2$, exponential map $Exp_{\boldsymbol{p}}(\boldsymbol{v})$, geodesic $Exp_{\boldsymbol{p}}(t\boldsymbol{v})$ and cut locus $\boldsymbol{p}$ for $\boldsymbol{p} \in \mathbb{S}^2$.

where $x_i(t)$ is the i-th component of $\gamma(t)$ under a given coordinate, i.e. $\gamma(t) = (x_1(t), x_2(t), \cdots, x_d(t))^T$. Γ_{ij}^k is the **Christoffel symbol** defined as

$$\Gamma_{ij}^k = \frac{1}{2}\sum_{m=1}^{d} g^{km}\Big(\frac{\partial g_{jm}}{\partial x_i} + \frac{\partial g_{im}}{\partial x_j} - \frac{\partial g_{ij}}{\partial x_m}\Big) \tag{3.12}$$

where $\{g^{jm}\}$ is the components of G^{-1} that is the inverse of the metric tensor matrix G.

Example 3.6 (Geodesic in Euclidean Space)**.** *In $\mathbb{R}^d$, $g_{ij} = \delta_{ij}$ and G is identity matrix. So $\Gamma_{ij}^k = 0$. Then $\frac{d^2x_k}{dt^2} = 0$. So $\gamma(t)$ is the straight line segment connecting two points.*

Example 3.7 (Geodesic in Sphere [Spivak, 1999, chap. 9])**.** *As we have shown that the tangent space $T_p\mathbb{S}^{d-1}$ for a given point p is an $d-1$ dimensional Euclidean space. See Fig 3.2. So the Riemannian metric for T_p is naturally chosen as the Euclidean metric $g_{ij} = \delta_{ij}$. For given two points $p, q \in \mathbb{S}^{d-1}$, they are two vectors in $\mathbb{R}^d$ starting from the original point. Let E^2 be the plane spanned by these two vectors, and let $C = E^2 \bigcap \mathbb{S}^{d-1}$ be the great circle (equator) through these two points. Then consider an isometry $f : \mathbb{S}^{d-1} \to \mathbb{S}^{d-1}$ defined as the reflection through the plane E^2. Then if γ is the geodesic between p and q, $I(\gamma)$ is also the geodesic between them, which means $\gamma = I(\gamma)$ [1]. In other words, $\gamma \subset C = E^2 \bigcap \mathbb{S}^{d-1}$. So the geodesic between two points in $\mathbb{S}^{d-1}$ is a part of the great circle which is smaller than semi-circle. For two points which are not antipodal, there exists only one great circle $C = E^2 \bigcap \mathbb{S}^{d-1}$ cross these two points. Then the geodesic is the smaller arc of the great circle connecting these two points, which is unique. For two antipodal points, there exist infinite great circles $C = E^2 \bigcap \mathbb{S}^{d-1}$. Then there are infinite geodesics which are semi-circle connecting these two points.*

Proposition 3.1. *If $\gamma(t) : [a, b] \to \mathcal{X}$ is a geodesic between two points $p = \gamma(a)$ and $q = \gamma(b)$ in Riemannian manifold $(\mathcal{X}, \langle\cdot,\cdot\rangle_{\mathcal{X}})$, then $\|\gamma'(t)\|_{\gamma(t)} = \|\gamma'(a)\|_{\gamma(a)} = \|\gamma'(b)\|_{\gamma(b)}$ is a constant for $t \in [a, b]$. So $\int_a^b \|\gamma'(t)\|_{\gamma(t)}\mathrm{d}t = (b-a)\|\gamma'(a)\|_{\gamma(a)}$.*

[1] Here we assume p and q are close enough in a neighborhood such that there exists a unique geodesic between them. See the Proposition 3.2 for the existence of such neighborhood in a general Riemannian manifold.

Note that the proof of the above proposition is based on **affine connection** in Riemannian manifold. Geodesic can be defined as a 1-dimensional **auto-parallel** curve based on affine connection, which is a more formal definition used in many textbooks and can also result in the geodesic equation in Eq. (3.11). However, in order to make the text more understandable, we are trying to avoid the technical concepts which are not used in the following chapters, e.g. affine connection. Thus please refer some textbooks like [Do Carmo, 1992, chap. 3] for affine connection and the proof of the above proposition.

Exponential Map and Logarithmic Map

Based on the existence and uniqueness of the solution of ordinary partial equations, the solution of Eq. (3.11) uniquely exists in a sufficiently small neighborhood when the initial point $\gamma(a) = p$ and initial velocity $\gamma'(a) = \boldsymbol{v}$ are given. This essentially defines a map from the tangent vector $\gamma'(a) = \boldsymbol{v} \in T_p\mathcal{X}$ to $\gamma(b) = q$. This map is formally called as the exponential map shown as follows.

Based on the geodesic equation in Eq. (3.11), let $t' = \lambda t$, $\lambda > 0$, then we will obtain the same geodesic with scaled parameterization and domain.

Corollary 3.1 (Homogeneity of Geodesics). *If the geodesic $\gamma(t)$ determined by Eq. (3.11) and conditions $\gamma(a) = p$ and $\gamma'(a) = \boldsymbol{v}$ uniquely exists in $[-\epsilon, \epsilon]$, we denote the geodesic by $\gamma(t; p, \boldsymbol{v})$, $t \in [-\epsilon, \epsilon]$. Then the geodesic $\gamma(t; p, \lambda\boldsymbol{v})$, $\lambda > 0$ uniquely exists in $[-\frac{\epsilon}{\lambda}, \frac{\epsilon}{\lambda}]$ and $\gamma(t; p, \lambda\boldsymbol{v}) = \gamma(\lambda t; p, \boldsymbol{v})$ with $t \in [-\frac{\epsilon}{\lambda}, \frac{\epsilon}{\lambda}]$.*

This corollary shows that it is possible for a geodesic to increase its interval of definition by decreasing the velocity $\boldsymbol{v}$, or vice-versa, which means the definition domain can be always scaled to $[-1, 1]$ by scaling the velocity $\boldsymbol{v}$.

Definition 3.15 (Exponential Map). *Let $\mathcal{X}$ be a Riemannian manifold, $p \in \mathcal{X}$ and $\boldsymbol{v} \in T_p\mathcal{X}$. Then there exists in a local neighborhood $[-\epsilon, \epsilon]$ a curve $\gamma(t; p, \boldsymbol{v})$ such that $\gamma(0) = p$, $\gamma'(0) = \boldsymbol{v}$. Assume $\gamma(t)$ exits for $t \in [0, 1]$, then $\gamma(t; p, \boldsymbol{v})$ becomes the geodesic between $\gamma(0)$ and $\gamma(1)$, and the exponential map is defined as*

$$Exp_p(\boldsymbol{v}) = \gamma(1; p, \boldsymbol{v}) \tag{3.13}$$

Based on the homogeneity of geodesics in Corollary 3.1 and the definition of exponential map, we have

$$\gamma(t; p, \boldsymbol{v}) = \gamma(1; p, t\boldsymbol{v}) = Exp_p(t\boldsymbol{v}) \tag{3.14}$$

which shows that $Exp_p(t\boldsymbol{v})$ as a geodesic locally exists in a sufficiently small interval $[-\epsilon, \epsilon]$, $\forall \boldsymbol{v} \in T_p\mathcal{X}$.

Definition 3.16 (Completeness). *$\mathcal{X}$ is called **geodesically complete**, if $\forall p \in \mathcal{X}$, the exponential map $Exp_p(\boldsymbol{v})$ is defined $\forall \boldsymbol{v} \in T_p\mathcal{X}$. In other words, for any given $p \in \mathcal{X}$, $\boldsymbol{v} \in T_p\mathcal{X}$, the curve $\gamma(t) = Exp_p(t\boldsymbol{v})$ exists $\forall t \in \mathbb{R}^1$.*

If the manifold is not geodesically complete, each point can have a cut locus which can be defined in the manifold or in tangent space.

Definition 3.17 (Cut Locus). *For $p \in \mathcal{X}$, the **cut locus of** p **in the tangent space** is defined as the set of all vectors $\boldsymbol{v} \in T_p\mathcal{X}$ such that $\gamma(t; p; \boldsymbol{v}) = Exp_p(t\boldsymbol{v})$ is a geodesic for $t \in [0, 1]$ and is not the curve with minimal length for $t \in [0, 1+\epsilon)$, $\forall \epsilon > 0$. The **cut locus of** p **in the manifold** is defined as the image of the cut locus in $T_p\mathcal{X}$, which is the set of points in the manifold where the geodesics starting at p stop being minimizing.*

Proposition 3.2. *Exp_p is a local diffeomorphism from open ball $B(0, R) \subset T_p\mathcal{X}$ to $Exp_p(B(0, R)) \subset \mathcal{X}$. The intersection of the cut locus of p in $T_p\mathcal{X}$ and $B(0, R)$ is empty.*

Proof. Exp_p is a function from a subset of $T_p\mathcal{X}$ to a subset of $\mathcal{X}$. Thus its pushforward map is $Exp_{p*}(\boldsymbol{v}) = \frac{dExp_p(t\boldsymbol{v})}{dt}|_{t=0} = \frac{d\gamma(t;p,\boldsymbol{v},)}{dt}|_{t=0} = \boldsymbol{v}$, which means Exp_{p*} is an identity map in $T_p\mathcal{X}$. So based on inverse function theorem, there exists an open ball $B(0, R) \subset T_p\mathcal{X}$ such that $Exp_p : B(0, R) \subset T_p\mathcal{X} \to Exp_p(B(0, R)) \subset \mathcal{X}$ is a diffeomorphism. $\forall \boldsymbol{v} \in B(0, R)$, $Exp_p(t\boldsymbol{v})$ with $t \in [0, 1]$ is a geodesic, which means it has not reached the cut locus of p based on the Definition 3.17. □

The above proposition shows that the maximal radius for such open ball $B(0, R)$ which makes Exp_p as a diffeomorphism is dependent on the cut locus. The open ball $B(0, R)$ is in $T_p\mathcal{X}$. Denote its image $Exp_p(B(0, R))$, a **geodesically open ball** in $\mathcal{X}$, by $B_{\mathcal{X}}(p, R) = \{x \in \mathcal{X} : \ d(p, x) < R\}$. Since exponential map is a local diffeomorphism, its inverse is well defined locally and is called as the logarithmic map.

Definition 3.18 (Logarithmic Map). *For two points $p, q \in \mathcal{X}$, if there exits $\boldsymbol{v} \in T_p\mathcal{X}$ such that $Exp_p(\boldsymbol{v}) = q$, the logarithmic map is defined as*

$$Log_p(q) = \boldsymbol{v} \tag{3.15}$$

Based on the definitions of exponential map and logarithmic map, we have

$$Log_p(Exp_p(\boldsymbol{v})) = \boldsymbol{v}, \quad Exp_p(Log_p(q)) = q \tag{3.16}$$

The geodesic $\gamma(t)$ connecting p and q such that $\gamma(0) = p$ and $\gamma(1) = q$, if it exists, is

$$\gamma(t) = Exp_p(tLog_p(q)), \quad t \in [0, 1] \tag{3.17}$$

Example 3.8 (Exponential Map and Logarithmic Map in Sphere)**.** *We have shown in Example 3.7 the Riemannian metric in $\mathbb{S}^{d-1}$ is δ_{ij}, and the geodesic γ between two points $p, q \in \mathbb{S}^{d-1}$ is in the great circle $C \subset E^2$ where E^2 is the 2 dimensional plane spanned by p, q and original point. Let $\boldsymbol{v} = Log_p(q) \in T_p\mathbb{S}^{d-1}$, then p, q and $\boldsymbol{v}$ are all in E^2, because $\gamma \subset E^2$. Since $\boldsymbol{v} \perp p$, $\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}$ and p form an orthonormal basis in E^2. So $\gamma(t)$ can be represented as*

$$\gamma(t) = a(t)p + b(t)\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}, \quad t \in [0, 1] \tag{3.18}$$

Considering $1 = \|\gamma(t)\|^2 = a^2(t) + b^2(t)$, $\gamma(0) = p$, we can set $a(t) = \cos(h(t))$ and $b(t) = \sin(h(t))$, then

$$\gamma(t) = \cos(h(t))p + \sin(h(t))\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}, \quad \textit{where } h(0) = 0$$

Then considering $\|\gamma'(t)\| = \|\gamma'(0)\| = \|\boldsymbol{v}\|$ based on Proposition 3.1, we have

$$\|\boldsymbol{v}\| = \|\gamma'(t)\| = \left\| -\sin(h(t))h'(t)p + \cos(h(t))h'(t)\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} \right\| = |h'(t)|$$

Considering $h(0) = 0$, we have $h(t) = \|\boldsymbol{v}\|t$. Then the geodesic is

$$\gamma(t) = \cos(\|\boldsymbol{v}\|t)p + \sin(\|\boldsymbol{v}\|t)\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}, \quad t \in [0, 1] \tag{3.19}$$

The exponential map is

$$Exp_p(\boldsymbol{v}) = \gamma(1) = \cos(\|\boldsymbol{v}\|)p + \sin(\|\boldsymbol{v}\|)\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} \tag{3.20}$$

Please note that the geodesic distance $\int_0^1 \gamma(t)\mathrm{d}t = \|\boldsymbol{v}\|$ is the distance of the arc, also the angle φ between the vector p and $\gamma(1) = q$ because the radius of $\mathbb{S}^{d-1}$ is 1. The logarithmic map is the inverse of exponential map, so it is

$$Log_p(q) = \frac{q - p\cos\varphi}{\|q - p\cos\varphi\|}\varphi, \quad \textit{where } \varphi = \arccos(p^T q) \tag{3.21}$$

Based on the Definition 3.17, the cut locus at p in $\mathbb{S}^2$ is its antipodal point $-p$, and the cut locus at p in $T_p\mathbb{S}^{d-1}$ is the circle with radius of π. Exp_p maps all vectors in $T_p\mathbb{S}^{d-1}$ with norm $2\pi n$ to p, where n is any non-negative integer. Note that $\mathbb{S}^{d-1}$ is geodesically complete based on Definition 3.16. The maximal radius of the local open ball $B(p, R)$ in Proposition 3.2, which makes Exp_p be a diffeomorphism, is $R = \pi$. See Fig. 3.2 for the specific case of $\mathbb{S}^2$.

3.2 Statistics on Riemannian Manifold

Compared to the materials in the last section which can be found in exhaustive textbooks, statistics on Riemannian manifold is a recent emerging area. There are indeed some monographs which focus on theoretical statistics in some special manifolds [Chikuse, 2003; Mardia and Jupp, 2000]. However in the view of medical image analysis applications in this thesis, we are interested in statistical computing on manifold valued data. Most materials for this specific topic are distributed in some papers or theses in last several decades [Kendall, 1990; Karcher, 1977; Pennec, 2006; Fletcher, 2004]. Thus we only list here some theoretical results used in Chapter 6, which have been also used in computer vision [Li et al., 2008; Subbarao and Meer, 2009], statistical shape analysis [Dryden and Mardia, 1998], medical image analysis [Fletcher et al., 2009; Pennec, 2006; Cheng et al., 2009b,a, 2011a], etc.

For statistical computing on manifold valued data, there are two ways, i.e. extrinsic way and intrinsic way. In extrinsic way, the manifold $\mathcal{X}$ is embedded into a higher dimensional Euclidean space $\mathbb{R}^d$. This embedding always exists based on **Whitney embedding theorem**. Thus statistical computing can be performed in Euclidean space, then the final results in Euclidean space are projected back to the manifold $\mathcal{X}$. Extrinsic way is dependent on the chosen coordinate and maps. While intrinsic way is independent on chosen coordinate, which is shown in this section.

With the tools described in last section, we can perform integral in Riemannian manifold $\mathcal{X}$. The length of a geodesic is the integral along the geodesic which is a 1-dimensional auto-parallel submanifold. It is possible to perform integral in high dimension, where the measure, or called the infinitesimal volume element, is defined as $\mathrm{d}\mathcal{V} = \sqrt{|\det(G)|}\mathrm{d}\boldsymbol{x}$ [2].

[2] It needs more technical materials which I am trying to avoid in this thesis.

This measure can be used to define distributions in manifold $\mathcal{X}$ such that they have the PDF $p_{\boldsymbol{X}}(\boldsymbol{x}) \geq 0$, $\int_{\mathcal{X}} p_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}\mathcal{V} = 1$. Then it is possible to calculate some statistical quantities, to compare measurements, to perform hypothesis test, etc [Kendall, 1990; Karcher, 1977; Pennec, 2006; Fletcher, 2004].

With a distribution $p_{\boldsymbol{X}}(\boldsymbol{x})$ defined in a Riemannian manifold $\mathcal{X}$, the expectation of function $f(\boldsymbol{x})$, denoted by $E_{\mathcal{X}}[f(\boldsymbol{x})]$, is

$$E_{\mathcal{X}}[f(\boldsymbol{x})] = \int_{\mathcal{X}} f(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}\mathcal{V} = \int_{\mathcal{X}} f(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x}) \sqrt{|\det(G)|} \mathrm{d}\boldsymbol{x} \tag{3.22}$$

In discrete case with n samples $\{\boldsymbol{x}^{(i)} \in \mathcal{X}\}_{i=1}^{n}$, $p_{\boldsymbol{X}}(\boldsymbol{x}) \sqrt{|\det(G)|} \mathrm{d}\boldsymbol{x}$ is given by weight vector $\boldsymbol{w} = (w_1, w_2, \cdots, w_n)^T$ and w_i is the probability for $\boldsymbol{x}^{(i)}$, $w_i \geq 0$, $\sum_{i=1}^{n} w_i = 1$. Then for the function f defined in the samples given by a vector $\boldsymbol{f} = (f_1, f_2, \cdots, f_n)^T$, its expectation is

$$E_{\{\boldsymbol{x}^{(i)} \in \mathcal{X}\}_{i=1}^{n}}[\boldsymbol{f}] = \sum_{i=1}^{n} w_i f_i = \boldsymbol{w}^T \boldsymbol{f} \tag{3.23}$$

In practice, normally we do not know the probability $p_{\boldsymbol{X}}(\boldsymbol{x})$ in the manifold $\mathcal{X}$, What we obtain is a limited number of random samples $\{\boldsymbol{x}^{(i)} \in \mathcal{X}\}$ and the measured values of function $f(\boldsymbol{x})$ at these samples. How to estimate the probability density function $p_{\boldsymbol{X}}(\boldsymbol{x})$ in Riemannian manifold from random samples is another topic [Pelletier, 2005], which is much difficult in practice especially when only limited number of samples obtained in a high dimensional manifold. In this thesis, we only consider the expectation given in the discrete case in Eq. (3.23), where the weight $\boldsymbol{w}$ is normally pre-fixed.

3.2.1 Riemannian Mean and Variance

For $\boldsymbol{x} \in \mathcal{X}$ and n samples $\{\boldsymbol{x}^{(i)} \in \mathcal{X}\}$, consider the variance function

$$\sigma^2(\boldsymbol{x}) \stackrel{\text{def}}{=} \sum_{i=1}^{n} w_i d(\boldsymbol{x}, \boldsymbol{x}^{(i)})^2 \tag{3.24}$$

where $d(\boldsymbol{x}, \boldsymbol{x}^{(i)})$ is the geodesic distance from $\boldsymbol{x}$ to $\boldsymbol{x}^{(i)}$, and $\boldsymbol{w} = (w_1, w_2, \cdots, w_n)^T$ is the weight vector.

Definition 3.19 (Weighted Riemannian Mean [Fréchet, 1948; Buss and Fillmore, 2001; Karcher, 1977; Kendall, 1990]**).** *The weighted Riemannian mean (or called Fréchet mean, or Karcher mean) is defined as the minimizer of the variance function $\sigma^2(\boldsymbol{x})$, denoted by $\boldsymbol{\mu}_{\boldsymbol{w}}$.*

$$\boldsymbol{\mu}_{\boldsymbol{w}} \stackrel{\text{def}}{=} \arg\min_{\boldsymbol{x} \in \mathcal{X}} \sigma^2(\boldsymbol{x}) = \arg\min_{\boldsymbol{x} \in \mathcal{X}} \sum_{i=1}^{n} w_i d(\boldsymbol{x}, \boldsymbol{x}^{(i)})^2 \tag{3.25}$$

When $w_i = \frac{1}{n}$, $\boldsymbol{\mu}_{\boldsymbol{w}}$ is called as the Riemannian mean, denoted by $\boldsymbol{\mu}$, and $\sigma^2 \stackrel{\text{def}}{=} \sigma^2(\boldsymbol{\mu})$ is called the variance.

In Euclidean space, the minimization in Eq. (3.25) has global solution $\boldsymbol{\mu}_{\boldsymbol{w}} = \sum_{i=1}^{n} w_i \boldsymbol{x}^{(i)}$ which is the weighted Euclidean mean. However in a general Riemannian manifold, weighted Riemannian mean may not exist (unreachable), or may have more than one solutions which obtain the global minimum of variance function.

In [Karcher, 1977], Karcher considered also the local minima when the variance function is not convex and has multiple local minima. The local minima of the variance function are called as **Riemannian centers of mass**. In this thesis, we just simply call the local minima as **weighted Riemannian means**. [Karcher, 1977; Kendall, 1990] gave some sufficient conditions for the existence and uniqueness of the global minima of variance function based on the concept of regular geodesic ball.

Definition 3.20 (Regular Geodesic Ball [Karcher, 1977; Kendall, 1990]**).** *The open ball $B_{\mathcal{X}}(p,R)$ in Riemannian manifold $\mathcal{X}$ is called a regular geodesic ball if*

(i) $HR < \frac{1}{2}\pi$, where H^2 is the supremum of sectional curvatures in $B_{\mathcal{X}}(p,R)$, or zero if the supremum is negative.

(ii) the cut locus of p does not meet $B_{\mathcal{X}}(p,R)$.

The local open ball which makes Exp_p as a diffeomorphism has been discussed in Proposition 3.2. The regular geodesic ball is such a local open ball which also satisfies $HR < \frac{1}{2}\pi$.

Note that [Kendall, 1990] also proved an important property of regular geodesic ball.

Proposition 3.3. *Let $B_{\mathcal{X}}(p,R)$ be a regular geodesic ball in Riemannian manifold $\mathcal{X}$ as the Definition 3.20, and $x, y \in B_{\mathcal{X}(p,R)}$. Then x, y are connected by one and only one geodesic within $B_{\mathcal{X}}(p,R)$.*

This proposition shows that for any point x in a regular geodesic ball $B_{\mathcal{X}}(p,R)$, the cut locus of x does not meet $B_{\mathcal{X}}(p,R)$. So (ii) in Definition 3.20 can also be replaced by "(ii) the cut locus of q does not meet $B_{\mathcal{X}}(p,R)$ for at least one $q \in B_{\mathcal{X}}(p,R)$".

Theorem 3.1 (Existence and Uniqueness of Riemannian Mean)**.** *Based on the analysis of regular geodesic ball, Kendall [Kendall, 1990] and Karcher [Karcher, 1977] proved the following sufficient conditions for the existence and uniqueness of Riemannian mean in continuous case.*

- *[Kendall, 1990] shows that if the support of $p_{\boldsymbol{X}}(\boldsymbol{x})$ is contained in a regular geodesic ball $B_{\mathcal{X}}(p,R)$, then there exists one and only one Riemannian center of mass $x \in B_{\mathcal{X}}(p,R)$.*
- *[Karcher, 1977] shows that if the support of $p_{\boldsymbol{X}}(\boldsymbol{x})$ is contained in regular geodesic ball $B_{\mathcal{X}}(p,R)$ and $B_{\mathcal{X}}(p,2R)$ is also a regular geodesic ball, then the variance function $\sigma^2(\boldsymbol{x})$ is a convex function with respect to x and has a global minima in $B_{\mathcal{X}}(p,R)$.*

Note that these two results can be also used in discrete case by considering $p_{\boldsymbol{X}}(\boldsymbol{x})$ as discrete probability distribution defined in points $\{\boldsymbol{x}^{(i)} \in \mathcal{X}\}_{i=1}^{n}$. So if all samples are in a regular geodesic ball $B_{\mathcal{X}}(\boldsymbol{x},R)$, the Riemannian mean defined in Eq. (3.25) uniquely exists.

These two results are applicable to general Riemannian manifold. For S^{d-1}, the curvature is the constant 1, which means for any $p \in \mathbb{S}^{d-1}$, $B_{\mathbb{S}^{d-1}}(p, R)$ is a regular geodesic ball if $R < \frac{\pi}{2}$. Based on Proposition 3.3, if all samples $\{\boldsymbol{x}^{(i)}\}_{i=1}^n$ are all in a hemispheres, the Riemannian mean $\boldsymbol{\mu}_{\boldsymbol{w}}$ uniquely exists. Note that Theorem 1 in [Buss and Fillmore, 2001] also proved this result for $\mathbb{S}^{d-1}$ based on the analysis specific to $\mathbb{S}^{d-1}$. [Buss and Fillmore, 2001] also proved some other interesting results.

Theorem 3.2 ([Buss and Fillmore, 2001, theorem 7]). *Suppose $\{\boldsymbol{x}^{(i)} \in \mathbb{S}^{d-1}\}_{i=1}^n$ are n points in a hemisphere of $\mathbb{S}^{d-1}$, and that it is not the case that $n = 2$ with $x^{(1)}$ and $x^{(2)}$ antipodal. Then the **convex hull** C of $\{\boldsymbol{x}^{(i)}\}_{i=1}^n$ [3] exists and is equal to the set of weighted Riemannian mean $\boldsymbol{\mu}_{\boldsymbol{w}}$ generated by all possible $\boldsymbol{w}$, $\sum_{i=1}^n w_i = 1$, $w_i \geq 0$.*

To find the local minima of the variance function $\sigma^2(\boldsymbol{x})$, we can perform gradient descent method [Pennec et al., 2006]. The gradient of $\sigma^2(\boldsymbol{x})$ is

$$\nabla\sigma^2(\boldsymbol{x}) = -2\sum_{i=1}^n w_i Log_{\boldsymbol{x}}(\boldsymbol{x}^{(i)}) \tag{3.26}$$

Then the gradient descent iteration is given as

$$\boldsymbol{\mu}_{\boldsymbol{w}}^{k+1} = Exp_{\boldsymbol{\mu}_{\boldsymbol{w}}^k}(-\frac{1}{2}\tau_k\nabla\sigma^2(\boldsymbol{x})) = Exp_{\boldsymbol{\mu}_{\boldsymbol{w}}^k}\Big(\tau_k\sum_{i=1}^n w_i Log_{\boldsymbol{\mu}_{\boldsymbol{w}}^k}(\boldsymbol{x}^{(i)})\Big) \tag{3.27}$$

where $\boldsymbol{\mu}_{\boldsymbol{w}}^k$ is the estimated $\boldsymbol{\mu}_{\boldsymbol{w}}$ in k-th iteration and τ_k is the step size.

Note that like the standard gradient descent in Euclidean space normally we need to choose an appropriate initialization $\boldsymbol{\mu}_{\boldsymbol{w}}^0$ and choose τ_k as a sufficiently small value for convergence. While in $\mathbb{S}^{d-1}$, [Buss and Fillmore, 2001] has proved that a constant $\tau_k = 1$ is sufficient for convergence of this gradient descent method. [Buss and Fillmore, 2001] also proposed to use the normalized Euclidean mean $\frac{\sum_{i=1}^n w_i x_i}{\|\sum_{i=1}^n w_i x_i\|}$ as the initialization for fast convergence.

3.2.2 Principal Geodesic Analysis

In Euclidean space, after mean is obtained, Principal Component Analysis (PCA) can be used to transform data into principal subspace spanned by some principal components. PCA minimizes the projection of data to the subspace. PGA was proposed to generalize the PCA from Euclidean space to Riemannian manifold [Fletcher et al., 2004]. After obtaining the Riemannian mean $\boldsymbol{\mu}$, PGA projects all points $\{\boldsymbol{x}^{(i)} \in \mathcal{X}\}_{i=1}^n$ onto the tangent space $T_{\boldsymbol{\mu}}\mathcal{X}$, and calculate the covariance matrix as

$$\Sigma = \frac{1}{n-1}\sum_{i=1}^n Log_{\boldsymbol{\mu}}(\boldsymbol{x}^{(i)})Log_{\boldsymbol{\mu}}(\boldsymbol{x}^{(i)})^T \tag{3.28}$$

Then like PCA a set of eigenvectors with large eigenvalues can be obtained by eigendecomposition of this covariance matrix. For the eigenvector $\boldsymbol{v}$ with eigenvalue λ, the principal component is $Exp_{\boldsymbol{\mu}}(\alpha\boldsymbol{v})$, where $\alpha \in \mathbb{R}^1$. α is normally chosen in $[-3\sqrt{\lambda}, 3\sqrt{\lambda}]$, which can be used to demonstrate the variance in that component [Fletcher et al.,

[3] Set C is the convex hull of a set D if C is the unique smallest convex set containing D.

2004; Fletcher, 2004]. So essentially PCA is performed in the tangent space $T_{\boldsymbol{\mu}}\mathcal{X}$ and then the principal subspace in $T_{\boldsymbol{\mu}}\mathcal{X}$ is mapped to manifold $\mathcal{X}$ by exponential map $Exp_{\boldsymbol{\mu}}$. PGA has been successfully used in shape analysis [Fletcher et al., 2004], tensor processing [Fletcher and Joshi, 2007].

3.2.3 Riemannian Median

Definition 3.21 (Weighted Riemannian Median). *Weighted Riemannian median is defined as the minimizer of the weighted sum of distance [Fletcher et al., 2009].*

$$\boldsymbol{m_w} \overset{\text{def}}{=} \arg\min_{\boldsymbol{x}\in\mathcal{X}} \sigma^1(\boldsymbol{x}) = \arg\min_{\boldsymbol{x}\in\mathcal{X}} \sum_{i=1}^{n} w_i d(\boldsymbol{x}, \boldsymbol{x}^{(i)}) \tag{3.29}$$

where $\boldsymbol{w}$ is the weight vector with $\sum_{i=1}^{n} w_i = 1$, $w_i \geq 0$. When $w_i = \frac{1}{n}$, it is called the Riemannian median, denoted by $\boldsymbol{m}$.

When $\mathcal{X}$ is a Euclidean space, the minimization of $\sigma^1(\boldsymbol{x})$ has global minimum which is the median in Euclidean space. When $\mathcal{X}$ is a general Riemannian manifold, the minimization may have no solution or may have many solutions as local minima. [Fletcher et al., 2009] proposed a sufficient condition for the existence and uniqueness of such minimization, based on the results on the convex analysis in [Karcher, 1977].

Theorem 3.3 ([Fletcher et al., 2009])**.** *In a Riemannian manifold $\mathcal{X}$, the weighted Riemannian median $\boldsymbol{m_w}$ uniquely exists if $diam(U) \cdot H < \frac{\pi}{2}$, where the diameter of $U = \{\boldsymbol{x}^{(i)}\}_{i=1}^{n}$, denoted by $diam(U)$, is the maximal distance between any two points in $\{\boldsymbol{x}^{(i)} \in \mathcal{X}\}_{i=1}^{n}$, $H = \sqrt{\Delta}$ if the sectional curvatures of $\mathcal{X}$ are positive and bounded by $\Delta > 0$, and $H = 0$ if the sectional curvatures are non-positive.*

Note that based on the triangle inequality and Proposition 3.3, if all points $U = \{\boldsymbol{x}^{(i)}\}_{i=1}^{n}$ are contained in a regular geodesic ball $B_{\mathcal{X}}(p, R)$, the diameter $\text{diam}(U) < 2R$. So if $B_{\mathcal{X}(p,2R)}$ is still a regular geodesic ball, then $\text{diam}(U) \cdot H < 2R \cdot H < \frac{\pi}{2}$ when $H > 0$ for positive sectional curvatures. Thus we have another sufficient condition which derives Theorem 3.3.

Theorem 3.4. *In a Riemannian manifold $\mathcal{X}$, the weighted Riemannian median $\boldsymbol{m_w}$ uniquely exists if all points $U = \{\boldsymbol{x}^{(i)}\}_{i=1}^{n}$ are contained in a regular geodesic ball $B_{\mathcal{X}}(p, R)$ such that $B_{\mathcal{X}}(p, 2R)$ is still a regular geodesic ball.*

In order to find the weighted Riemannian median, we can still use the gradient descent. The gradient of $\sigma^1(\boldsymbol{x})$ is

$$\nabla\sigma^1(\boldsymbol{x}) = -\sum_{i=1}^{n} \frac{w_i Log_{\boldsymbol{x}}(\boldsymbol{x}^{(i)})}{d(\boldsymbol{x}, \boldsymbol{x}^{(i)})} \tag{3.30}$$

Because the gradient has no definition when $\boldsymbol{x} = \boldsymbol{x}^{(i)}$, we need to require $\boldsymbol{x} \neq \boldsymbol{x}^{(i)}$. Then the gradient descent, which is **iteratively re-weighted least square method**, is

given as

$$\boldsymbol{v}_{\boldsymbol{m}_{\boldsymbol{w}}}^{k} = \sum_{i=1}^{n} \frac{w_i / d(\boldsymbol{m}_{\boldsymbol{w}}^{k}, \boldsymbol{x}^{(i)})}{\sum_{j=1}^{n} w_j / d(\boldsymbol{m}_{\boldsymbol{w}}^{k}, \boldsymbol{x}^{(j)})} Log_{\boldsymbol{m}_{\boldsymbol{w}}}(\boldsymbol{x}^{(i)}) \tag{3.31}$$

$$\boldsymbol{m}_{\boldsymbol{w}}^{k+1} = Exp_{\boldsymbol{m}_{\boldsymbol{w}}}^{k}\left(\tau_k \boldsymbol{v}_{\boldsymbol{m}_{\boldsymbol{w}}}^{k}\right) \tag{3.32}$$

where $\boldsymbol{m}_{\boldsymbol{w}}^{k}$ is the estimated $\boldsymbol{m}_{\boldsymbol{w}}$ in k-th iteration, τ_k is the step size. Like the gradient ascent method for $\boldsymbol{\mu}_{\boldsymbol{w}}$, normally we need a good initialization $\boldsymbol{m}_{\boldsymbol{w}}^{0}$ and a good τ_k in each step. [Fletcher et al., 2009] proved a result for τ_k for convergence.

Theorem 3.5 ([Fletcher et al., 2009]). *If the sectional curvatures of $\mathcal{X}$ are nonnegative and bounded by $\Delta > 0$ and $diam(U) \cdot \sqrt{\Delta} < \frac{\pi}{2}$, then the gradient descent converges for $\tau_k \in [0, 2]$.*

Riemannian median has been proved to be more robust than Riemannian mean and is more appropriate for atlas estimation in medical image analysis, computer vision, etc [Fletcher et al., 2009].

3.2.4 Riemannian L^p Mean

Recently [Afsari, 2011] proposed a general result on existence and uniqueness of Riemannian L^p mean (center of mass). Consider the function $\sigma^p(\boldsymbol{x})$ as

$$\sigma^p(\boldsymbol{x}) = \begin{cases} \int_{\mathcal{X}} d(\boldsymbol{x}, \boldsymbol{y})^p p_{\boldsymbol{X}}(\boldsymbol{y}) \mathrm{d}\boldsymbol{V} & \text{if} \quad 1 \le p < \infty \\ \arg\max_{y \in \overline{\mathrm{supp}}(p_{\boldsymbol{X}})} d(\boldsymbol{x}, \boldsymbol{y}) & \text{if} \quad p = \infty \end{cases} \tag{3.33}$$

where $\overline{\mathrm{supp}}(p_{\mathcal{X}})$ is the closure of the support of the probability $p_{\boldsymbol{X}}$. The local minima of $\sigma^p(\boldsymbol{x})$ are called Riemannian L^p mean, which is Riemannian mean when $p = 2$ and Riemannian median when $p = 1$.

[Afsari, 2011] proved a upper bound of R such that the Riemannian L^p mean uniquely exists in regular geodesic ball $B_{\mathcal{X}}(p, R)$. First we introduce the concept of injectivity radius. The injectivity radius at $p \in \mathcal{X}$ is the largest radius for which the exponential map at p is a diffeomorphism. The injectivity radius of Riemannian manifold $\mathcal{X}$ is the infimum of the injectivity radii at all points.

Theorem 3.6 ([Afsari, 2011]). *In a complete Riemannian manifold $\mathcal{X}$ with sectional curvatures upper bound of Δ and the injective radius of injX, define*

$$R_{\Delta,p} \overset{\text{def}}{=} \begin{cases} \frac{1}{2} \min\{injX, \frac{\pi}{2H}\} & if \quad 1 \le p < 2 \\ \frac{1}{2} \min\{injX, \frac{\pi}{H}\} & if \quad 2 \le p \le \infty \end{cases} \tag{3.34}$$

where H is $\sqrt{\Delta}$ if $\Delta \ge 0$, 0 if $\Delta < 0$. Then if $\overline{supp}(p_{\mathcal{X}}) \subset B_{\mathcal{X}}(p, R) \subset \mathcal{X}$, the Riemannian L^p mean uniquely exists in except some degenerate cases and lies in $B_{\mathcal{X}}(p, R)$, if $R < R_{\Delta,p}$.

The above theorem on Riemannian L^p mean in [Afsari, 2011] agrees with Theorem 3.1 [Karcher, 1977; Kendall, 1990] when $p = 2$ and agrees with Theorem 3.4 which can derive Theorem 3.3 [Fletcher et al., 2009] when $p = 1$.

3.3 Statistical Manifold

In last two sections, we have reviewed the basic concepts of Riemannian manifold and some important theoretical results on how to perform statistical computing on manifold valued data. This section shows the basic theory on how to endow a sophisticated Riemannian manifold structure to a parametric family of distributions, so that we can perform statistical computing on the data which are distributions from the parametric family.

It starts in the original **Mahalanobis distance** in 1936 [Mahalanobis, 1936] to endow a geometric structure in a parametric distribution family and define the distance between different parameterized distributions. The pioneer work in [Rao, 1945] generalizes the Mahalanobis distance and opens a new field called **Information Geometry**, which has many applications in quantum mechanics [Wootters, 1981; Braunstein and Caves, 1994; Brody and Hughston, 1998], computer vision [Srivastava et al., 2007; Maybank, 2004], machine learning [Lebanon, 2006; Lafferty and Lebanon, 2006], etc. In this section, we will review some basic useful concepts in Information Geometry theory. Please refer [Amari, 1985; Amari and Nagaoka, 2000] for more formal details.

3.3.1 Basic Concepts in Information Geometry

Parametric Family (PF)

Let $\Pr(\chi)$ be the space of all probabilities defined in a field χ, i.e.

$$\Pr(\chi) = \left\{ p(\boldsymbol{x}) \ : \ \int_{\chi} p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = 1, \ p(\boldsymbol{x}) \geq 0 \right\} \tag{3.35}$$

Parametric Family (PF), aka **statistical model**, is a family of parametric distributions on the field χ in Eq. (3.36), where $\boldsymbol{c}$ is a K dimensional parameter vector in Parametric Space (PS), denoted by $PS \subset \mathbb{R}^K$ [Rao, 1945; Amari and Nagaoka, 2000]. Thus $PF \subset \Pr(\chi)$.

$$PF = \left\{ p(\boldsymbol{x}|\boldsymbol{c}) \ : \ \int_{\chi} p(\boldsymbol{x}|\boldsymbol{c}) \mathrm{d}\boldsymbol{x} = 1, \ p(\boldsymbol{x}|\boldsymbol{c}) \geq 0, \ \boldsymbol{c} \in PS \subset \mathbb{R}^K \right\} \tag{3.36}$$

Statistical manifold

Let's consider the following mapping as

$$\varphi : PS \subset \mathbb{R}^K \to \varphi(PS) = PF \subset \Pr(\chi), \quad \varphi : \boldsymbol{c} \mapsto p(\boldsymbol{x}|\boldsymbol{c}) \tag{3.37}$$

Obviously φ is not a **surjective** to $\Pr(\chi)$, because there exist some probabilities in $\Pr(\chi)$ which can not be represented by parameterized model in Eq. (3.36). Normally we require φ is a **injective** such that the model in Eq. (3.36) is **identifiable**. Otherwise there exist $\boldsymbol{c}^{(1)}$ and $\boldsymbol{c}^{(2)}$ such that $p(x|\boldsymbol{c}^{(1)}) = p(x|\boldsymbol{c}^{(2)}), \forall x \in \chi$.

So we assume φ is a **bijective** from $PS \subset \mathbb{R}^K$ to $PF \subset \Pr(\chi)$. We also assume $p(\boldsymbol{x}|\boldsymbol{c})$ is differentiable with respect to $\boldsymbol{c}$, which means φ is differentiable. Then φ becomes a diffeomorphism. Thus we can identify parametric distribution $p(\boldsymbol{x}|\boldsymbol{c})$ in the family PF as points $\boldsymbol{c} \in PS$. Next we assume PS is a differentiable manifold embedded in $\mathbb{R}^K$. Then PF with distributions as elements is a differentiable manifold, called **statistical manifold**, which is diffeomorphic to PS. The parameter vector $\boldsymbol{c}$ is called the **coordinate**.

Please note that all assumptions we made here are very weak, which can be satisfied by most parametric family used in practice.

Example 3.9. *Consider the Gaussian distribution in 1D, i.e.* $N(x|\mu,\sigma)$. *Thus* $\boldsymbol{c} = (\mu,\sigma)^T$, $PS = \{(\mu,\sigma)^T \in \mathbb{R}^2 : \mu \in (-\infty,\infty), \sigma \in (0,\infty)\}$ *which is a differentiable manifold in* $\mathbb{R}^2$. $N(x|\boldsymbol{c})$ *is differentiable with respect to* $\boldsymbol{c}$.

Fisher information metric as the Riemannian metric

Based on Section 3.1.2, the tangent space in PS is generated by local basis $\{\frac{\partial}{\partial c_i}\}_{i=1,\dots,K}$. [Rao, 1945] considered $\{\frac{\partial \log p(\boldsymbol{x}|\boldsymbol{c})}{\partial c_i}\}_{i=1,\dots,K}$ as the **local basis** on the **tangent space** at $p(\boldsymbol{x}|\boldsymbol{c})$ which is isomorphic to the tangent space of PS. The vector $\boldsymbol{u} = \sum_{i=1}^K u_i \frac{\partial}{\partial c_i}$ corresponds to the vector $\boldsymbol{u}(x) = \sum_{i=1} u_i \frac{\partial \log p(\boldsymbol{x}|\boldsymbol{c})}{\partial c_i}$ with the same coordinate [Amari, 1985, chap. 2]. Thus we identify these two tangent vector and denote them by $\boldsymbol{u}$, and we also identify the two tangent spaces and denote them by $T_{\boldsymbol{c}}$. The expectation of the random variable $\boldsymbol{u} = \sum_{i=1} u_i \frac{\partial \log p(\boldsymbol{x}|\boldsymbol{c})}{\partial c_i}$ is 0 because

$$E_\chi\left(\frac{\partial \log p(\boldsymbol{x}|\boldsymbol{c})}{\partial c_i}\right) = \int_\chi \frac{\partial p(\boldsymbol{x}|\boldsymbol{c})}{\partial c_i} \mathrm{d}\boldsymbol{x} = \frac{\partial}{\partial c_i} \int_\chi p(\boldsymbol{x}|\boldsymbol{c}) \mathrm{d}\boldsymbol{x} = 0 \tag{3.38}$$

where we assume the integration over $\boldsymbol{x}$ and differentiation with respect to $\boldsymbol{c}$ can be freely rearranged.

As we have shown in Section 3.1.3, Riemannian manifold is a differentiable manifold equipped with a Riemannian metric g in its tangent space. For the statistical manifold PF, there exist many possible inner products to make it a Riemannian manifold. The **Fisher information matrix** with its (i,j)-th entry in Eq. (3.39) naturally defines a positive definite inner product, i.e. a Riemannian metric called the **Fisher information metric**.

$$g_{ij}(\boldsymbol{c}) \overset{\text{def}}{=} E_\chi\left(\frac{\partial \log p(\boldsymbol{x}|\boldsymbol{c})}{\partial c_i} \frac{\partial \log p(\boldsymbol{x}|\boldsymbol{c})}{\partial c_j}\right) = \int_\chi \frac{\partial \log p(\boldsymbol{x}|\boldsymbol{c})}{\partial c_i} \frac{\partial \log p(\boldsymbol{x}|\boldsymbol{c})}{\partial c_j} p(\boldsymbol{x}|\boldsymbol{c}) \mathrm{d}\boldsymbol{x} \tag{3.39}$$

$$= 4 \int_\chi \frac{\partial \sqrt{p(\boldsymbol{x}|\boldsymbol{c})}}{\partial c_i} \frac{\partial \sqrt{p(\boldsymbol{x}|\boldsymbol{c})}}{\partial c_j} \mathrm{d}\boldsymbol{x} \tag{3.40}$$

It is easy to see this metric has another equivalent form in Eq. (3.41) when the order of differentiation and integration can be changed.

$$g_{ij}(\boldsymbol{c}) = -E_\chi\left(\frac{\partial^2}{\partial c_i \partial c_j} \log p(\boldsymbol{x}|\boldsymbol{c})\right) = -\int_\chi \frac{\partial^2 \log p(\boldsymbol{x}|\boldsymbol{c})}{\partial c_i \partial c_j} p(\boldsymbol{x}|\boldsymbol{c}) \mathrm{d}\boldsymbol{x} \tag{3.41}$$

The above three formulae Eq. (3.39) Eq. (3.40) Eq. (3.41) are normally used to calculate g_{ij} from $p(\boldsymbol{x}|\boldsymbol{c})$.

For given two **tangent vectors** represented by the basis as $\boldsymbol{v} = \sum_{i=1}^{K} v_i \frac{\partial \log p(\boldsymbol{x}|\boldsymbol{c})}{\partial c_i}$ and $\boldsymbol{u} = \sum_{j=1}^{K} u_j \frac{\partial \log p(\boldsymbol{x}|\boldsymbol{c})}{\partial c_j}$, the inner product is given as

$$\langle \boldsymbol{v}, \boldsymbol{u} \rangle_{\boldsymbol{c}} = E_\chi(\boldsymbol{u}, \boldsymbol{v}) = \sum_{ij} v_i u_j g_{ij}(\boldsymbol{c}) \tag{3.42}$$

Note $E_\chi(\boldsymbol{u}, \boldsymbol{v}) = Cov(\boldsymbol{u}, \boldsymbol{v})$ is the covariance of $\boldsymbol{v}$ and $\boldsymbol{u}$ because $E_\chi(\boldsymbol{u}) = E_\chi(\boldsymbol{v}) = 0$.

Geodesic, exponential map and logarithmic map

Since we have constructed the statistical manifold and equip it with the Fisher information metric, the concepts and results in Section 3.1.3 can be used in statistical manifold.

The length of the tangent vector $\boldsymbol{v} \in T_{\boldsymbol{c}}$ is $\|\boldsymbol{v}\| = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}$. Then the length of the curve $\gamma : [a, b] \to PS$ is $L(\gamma) = \int_a^b \|\gamma'(t)\| \mathrm{d}t$, where $\gamma'(t)$ is a tangent vector in $T_{\gamma(t)}$. For given two points (distributions) in the statistical manifold, the curve connecting them with the minimal length is the **geodesic**. A given point (distribution) $\boldsymbol{c}$ with a given tangent vector $\boldsymbol{v}$ determines the unique geodesic $\gamma(t)$ with $\gamma(0) = \boldsymbol{c}$ and $\gamma'(0) = \boldsymbol{v}$. The **exponential map** and **logarithmic map** are diffeomorphisms defined in a local star-shaped subset $U \in T_{\boldsymbol{c}}$. The boundary of U, if it exists, is the cut locus in tangent space.

The geodesic, exponential map and logarithmic map can be used as a framework for statistical computing on distributions from a parametric family.

3.3.2 Examples on Gaussian Distribution and Multinomial Distribution Families

We have shown in Example 3.9 that the parametric family of Gaussian distributions in 1D can be seen as a statistical manifold. Here we show another example for the statistical manifold of Gaussian distributions in K dimension. Please refer [Atkinson and Mitchell, 1981; Skovgaard, 1984] for the proofs and more details on the geometry for Gaussian distribution family.

Example 3.10 (Gaussian Distribution Family in K Dimension)**.** *Multi-variant Gaussian distribution family in K dimension is given as*

$$PF = \left\{ N(\boldsymbol{x}|\mu, \Sigma) = \frac{1}{(2\pi)^{K/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \mu)^T \Sigma^{-1} (\boldsymbol{x} - \mu) \right) \ : \mu \in \mathbb{R}^K, \ \Sigma \in Sym_K^+ \right\} \tag{3.43}$$

We consider PF for fixed Σ and fixed μ respectively.

Considering $PF = \left\{ N(\boldsymbol{x}|\mu, \Sigma) : \mu \in \mathbb{R}^K, \ \Sigma = \Sigma_0 \right\}$, then $g_{ij} = \Sigma_0^{-1}(i, j)$ based on Eq. (3.41) which is a constant. So let $\boldsymbol{y} = \Sigma_0^{-1/2} \boldsymbol{x}$, $\boldsymbol{y}$ is in Euclidean space with $g_{ij} = \delta_{ij}$. Thus the geodesic, exponential map, logarithmic map of PF are induced from Euclidean space by the transform $\boldsymbol{x} = \Sigma_0^{1/2} \boldsymbol{y}$. For two Gaussian PDFs with the different mean vectors μ_1 and μ_2 but the same covariance matrix Σ_0, the geodesic distance is the well-known Mahalanobis distance [Atkinson and Mitchell, 1981]

$$d(N(\boldsymbol{x}|\mu_1, \Sigma), N(\boldsymbol{x}|\mu_2, \Sigma)) = \sqrt{(\mu_1 - \mu_2)^T \Sigma^{-1} (\mu_1 - \mu_2)} \tag{3.44}$$

Thus the geodesic distance in information geometry is a generalization of Mahalanobis distance that is only for Gaussian distributions with the same covariance.

Now for fixed μ_0, we consider

$$PF = \left\{ N(\boldsymbol{x}|\mu_0, \Sigma) : \quad \Sigma \in Sym_K^+ \right\}$$

Note that $N(\boldsymbol{x}|\mu, \Sigma)$ is defined $\forall \boldsymbol{x} \in \mathbb{R}^K$, the space PF is invariant under shift $\boldsymbol{y} = \boldsymbol{x} + \mu_0$. So the PF can be written as

$$PF = \left\{ N(\boldsymbol{x}|\Sigma) = \frac{1}{(2\pi)^{K/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x}\right) : \quad \Sigma \in Sym_K^+ \right\} \tag{3.45}$$

where $\mu_0 = \boldsymbol{0}$. Sym_K^+ is open convex cone embedded in $K \times K$ symmetric matrix space Sym_K which is isomorphic to Euclidean space $\mathbb{R}^{K^2}$. Matrix exponential function is a diffeomorphism between Sym_K and Sym_K^+. The tangent space of a given $\Sigma \in Sym_K^+$, i.e. $T_\Sigma(Sym_K^+)$, is Sym_K. For given two symmetric matrices $A, B \in Sym_K$, the inner product in Sym_K is naturally defined as $\text{Trace}(AB)$. *Based on directional derivative analysis of matrix function, the inner product in $T_\Sigma(Sym_K^+)$ is [Atkinson and Mitchell, 1981; Skovgaard, 1984]*

$$\langle A, B\rangle_\Sigma = \frac{1}{2}\text{Trace}(\Sigma^{-1}A\Sigma^{-1}B), \quad A, B \in T_\Sigma(Sym_K^+) \tag{3.46}$$

The geodesic $\gamma(t)$ with $\gamma(0) = \Sigma$ and $\gamma'(0) = A \in T_\Sigma(Sym_K^+)$ is given as

$$\gamma(t) = \Sigma^{\frac{1}{2}}\log(\Sigma^{-\frac{1}{2}}At\Sigma^{-\frac{1}{2}})\Sigma^{\frac{1}{2}}, \quad t \in [0, 1] \tag{3.47}$$

The exponential map is

$$Exp_\Sigma(A) = \gamma(1) = \Sigma^{\frac{1}{2}}\exp(\Sigma^{-\frac{1}{2}}A\Sigma^{-\frac{1}{2}})\Sigma^{\frac{1}{2}}, \quad A \in Sym_3 \tag{3.48}$$

The logarithmic is

$$Log_\Sigma(\Lambda) = \Sigma^{\frac{1}{2}}\log(\Sigma^{-\frac{1}{2}}\Lambda\Sigma^{-\frac{1}{2}})\Sigma^{\frac{1}{2}}, \quad \Lambda \in Sym_3^+ \tag{3.49}$$

For two Gaussian PDFs with the same mean vector μ but different covariance matrices Σ_1 and Σ_2, the geodesic distance is [Atkinson and Mitchell, 1981; Skovgaard, 1984]

$$d(N(\boldsymbol{x}|\mu, \Sigma_1), N(\boldsymbol{x}|\mu, \Sigma_2)) = \sqrt{\frac{1}{2}\sum_{i=1}^{K}\log^2(\lambda_i)} \tag{3.50}$$

where λ_i is the eigenvalues of $\Sigma_1^{-1}\Sigma_2$. For an affine transform $A : \chi \to \chi$, $A\boldsymbol{x} = \boldsymbol{y}$, we have

$$\begin{aligned} d(N(\boldsymbol{y}|\mu, \Sigma_1), N(\boldsymbol{y}|\mu, \Sigma_1)) &= d(N(\boldsymbol{x}|A\mu, A\Sigma_1A^T), N(\boldsymbol{x}|A\mu, A\Sigma_2A^T)) && (3.51)\\ &= d(N(\boldsymbol{x}|\mu, \Sigma_1), N(\boldsymbol{x}|\mu, \Sigma_1)) && (3.52) \end{aligned}$$

which means the distance is affine-invariant. The affine-invariant distance has been successfully used in tensor processing [Pennec et al., 2006; Moakher, 2005; Batchelor et al., 2005; Lenglet et al., 2006b; Fletcher and Joshi, 2007].

Example 3.11 (Multinomial Distribution Family)**.** *Let's consider the multinomial distribution family given in Eq.* (3.53)*.*

$$PF = \left\{ p(x|\boldsymbol{p}) = \sum_{i=1}^{K} p_i \delta(x=i) : \ \boldsymbol{p} \in \mathbb{R}^K, \ p_i \geq 0, \ \sum_{i=1}^{K} p_i = 1, \right\} \tag{3.53}$$

The PS is the probability simplex defined by

$$PS = \left\{ \boldsymbol{p} = (p_1, p_2, \cdots, p_K)^T \in \mathbb{R}^K : \sum_{i=1}^{K} p_i = 1, \ p_i \geq 0 \right\} \tag{3.54}$$

If $K = 3$, PS is the equilateral triangle with vertices $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$. The tangent space of every point is the plane $T_{\boldsymbol{c}} = \{\boldsymbol{v} \in \mathbb{R}^K : \sum_{i=1}^{K} v_i = 0\}$. Please see the left side of Fig. 3.3. Although it is possible to calculate the metric and geodesic based on the formulation in Eq. (3.53)*, it is convenient to consider the new coordinates under square root parametrization, where $c_i = \sqrt{p_i}$. Then the PF is Eq.* (3.55)

$$PF = \left\{ p(x|\boldsymbol{c}) = \sum_{i=1}^{K} c_i^2 \delta(x=i) : \ \boldsymbol{c} \in \mathbb{R}^K, \ c_i \geq 0, \ \sum_{i=1}^{K} c_i^2 = 1, \right\} \tag{3.55}$$

The new PS is the positive orthant of the unit sphere $\mathbb{S}^{K-1}$, i.e.

$$PS = \left\{ \boldsymbol{c} = (c_1, c_2, \cdots, c_K)^T \in \mathbb{S}^2 : c_i \geq 0 \right\} \tag{3.56}$$

The new tangent space at $\boldsymbol{c}$ is $T_{\boldsymbol{c}} = \{\boldsymbol{v} \in \mathbb{R}^K : \sum_{i=1}^{K} v_i c_i = 0\}$ based on Example 3.5. Based on Eq. (3.40)*, the Fisher information metric is*

$$g_{ij}(\boldsymbol{c}) = 4 \sum_{i=1}^{K} \frac{\partial \sqrt{p(x|\boldsymbol{c})}}{\partial c_i} \frac{\partial \sqrt{p(x|\boldsymbol{c})}}{\partial c_j} = 4\delta_{ij} \tag{3.57}$$

It can be proved that the geodesic distance between $p(x|\boldsymbol{c})$ and $p(x|\boldsymbol{c}')$ is twice of the geodesic on $\mathbb{S}^{K-1}$, i.e. $d(p(x|\boldsymbol{c}), p(x|\boldsymbol{c}')) = 2\arccos(\boldsymbol{c}^T\boldsymbol{c}')$. Please see the right side of Fig. 3.3. The square root parametrization for multinomial distribution was explored early in [Wootters, 1981] and in example 2.4 of [Amari, 1985], and was applied in different domains in [Srivastava et al., 2007; Lebanon, 2006; Goh et al., 2011]. Please note the difference between our description and example 2.4 of [Amari, 1985]. [Amari, 1985] considered the new coordinates as $c_i = 2\sqrt{p_i}$, which results in PS as a sphere with radius 2, i.e. $2\mathbb{S}^2$, and $g_{ij}(\boldsymbol{c}) = \delta_{ij}$. However the geodesic does not change under different parametrizations.

Normally a constant multiplication on metric can be ignored in the statistical computing on manifold valued data. Thus we can assume $g_{ij} = \delta_{ij}$ after the transform $c_i = \sqrt{p_i}$. Then based on Example 3.5, 3.7, 3.8, we show two parameter spaces under $\boldsymbol{p}$ and $\boldsymbol{c}$ representations in Fig. 3.3. In the right side of Fig. 3.3, two points $\boldsymbol{c}$, $\boldsymbol{c}'$, the geodesic, and the tangent vector $\boldsymbol{v}_{\boldsymbol{c}} = Log_{\boldsymbol{c}}(\boldsymbol{c}')$ are shown in tangent space $T_{\boldsymbol{c}}$. The left side of the figure shows the two corresponding points $\boldsymbol{p}$, $\boldsymbol{p}'$, the geodesic and tangent vector $\boldsymbol{p}' - \boldsymbol{p}$ for the representation with $\boldsymbol{p}$.

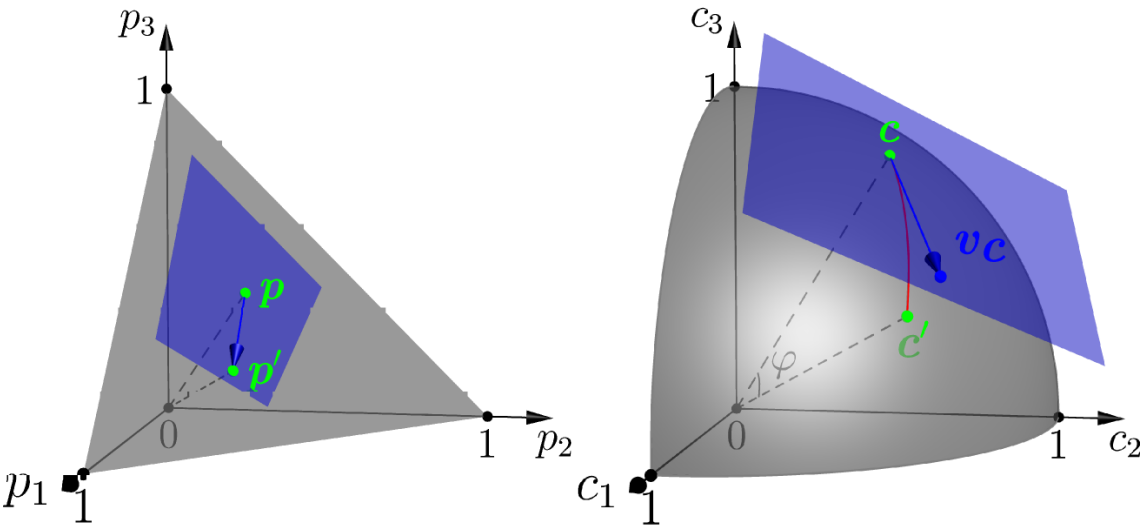


Figure 3.3: Statistical manifold for multinomial distribution in 3D. From left to right: parameter space PS represented by $\{p_i\}$ in Eq. (3.54); PS represented by $\{c_i\}$ in Eq. (3.56). The tangent vector, exponential map and geodesic distance are also shown respectively in the tangent space T_p and T_c.

3.4 SUMMARY

In this chapter, we made an overview of basic concepts and results of Riemannian manifold, statistical computing on manifold and statistical manifold, which are background knowledge for Chapter 6. We also gave some examples for better understanding the materials.

We now summarize the important and useful materials listed in this chapter which will be used in the Chapter 6.

1. **The basic notions of Riemannian geometry** described in Section 3.1 will be implicitly used in Chapter 6.

2. **Weighted Riemannian mean and median** described in Section 3.2 will be used in statistical computing on Gaussian distributions and non-Gaussian distributions in Chapter 6. The **existence and uniqueness** of weighted Riemannian mean and median are essential for the analysis in Chapter 6.

3. **Example 3.10 on Gaussian distributions** lists theoretical results on previous **Riemannian framework for tensors (Gaussian distributions)** [Atkinson and Mitchell, 1981; Skovgaard, 1984; Moakher, 2005; Pennec et al., 2006; Moakher, 2005; Batchelor et al., 2005; Lenglet et al., 2006b; Fletcher and Joshi, 2007]. Chapter 6 will generalize the Riemannian framework from Gaussian distributions to general distributions which are not necessary to be Gaussian.

4. **The geometry of high dimensional sphere** $\mathbb{S}^{d-1}$ is well studied in Example 3.2 on topology, Example 3.5 on tangent space, Example 3.7 on geodesic, Example 3.8 on exponential and logarithmic maps. Riemannian framework on sphere including geodesic, exponential map and logarithmic map can be used to

devise algorithms to process the data on sphere. For example, in [Cheng et al., 2009b] we proposed to use mean shift on sphere to analyze functional MRI data.

5. Example 3.11 showed **the statistical manifold of multinomial distribution**, which is essential for our Riemannian framework on ODFs and EAPs in Chapter 6. [Goh et al., 2011] directly used the results of multinomial distribution family by considering the ODF represented by its histogram. Our formulation in [Cheng et al., 2009a, 2011a] is quite different. We consider the PDFs (both ODFs and EAPs) whose square roots are represented by linear combination of orthonormal basis functions. Please see Chapter 6 for more details and comparisons.

CHAPTER 4

DIFFUSION MRI

"If I have seen further it is by standing on ye sholders of Giants."

– Isaac Newton

Contents

OVERVIEW

During the last several decades, the principles of diffusion Magnetic Resonance Imaging (dMRI) have been largely developed, which makes dMRI become the unique noninvasive technique to study white matter and fibers in human brain. This chapter covers the basic principles of dMRI on diffusion phenomenon, MRI measurements, the concepts of Ensemble Average Propagator (EAP) and Orientation Distribution Function (ODF), and some state-of-the-art reconstruction methods which estimate EAPs/ODFs and infer some other meaningful information from measured dMRI signals. Since EAP and ODF are essentially Probability Density Functions (PDFs) which describe the water diffusion, we also review some works on metric selection for processing distribution valued data. This introductory chapter is inspired from [Johansen-Berg and Behrens, 2009; Huettel et al., 2009; Tuch, 2002; Le Bihan et al., 2003; Lenglet, 2006; Hagmann et al., 2006; Descoteaux, 2008; Ghosh, 2011; Assemlal et al., 2011]. It contains the background knowledge and motivations of the Chapter 5 and 6,

Organization of this chapter:
We first introduce some basic concepts of diffusion process in subsection 4.1.1, Magnetic Resonance Imaging (MRI) and the duality between **k**-space and **x**-space in subsection 4.1.2. Then the PGSE sequence, Stejskal-Tanner equation, the duality between **q**-space and **R**-space and EAP are introduced in subsection 4.1.3. Section 4.2 gives a review on Diffusion Tensor Imaging (DTI). Section 4.3 lists the start-of-the-art reconstruction methods for ODFs and EAPs in High Angular Resolution Diffusion Imaging. Section 4.4 describes the possible metric and processing framework for tensors, ODFs and EAPs.

4.1 BASIC OF DIFFUSION MRI

Diffusion Magnetic Resonance Imaging (dMRI) is a widely used in-vivo imaging technique to explore the information of neural micro-structure by probing the diffusion of water molecules. So far it is still the unique non-invasive method to reveal the micro-geometry of nervous tissues noninvasively and to explore the neural connectome in living human subjects. The diffusion of water molecules is constrained by the surrounding structures including nerves, cells and surrounding tissue. For example, qualitatively water molecules diffuse fast along fibers and slowly across fibers. Thus measuring the diffusion process quantitatively is crucial to understanding the neural micro-structure and fiber directions.

4.1.1 Diffusion Process, Free Diffusion and Gaussian Propagator

A drop of colored fluorescent dye placed in a glass of water will spread out in the glass. The color becomes less intense and finally the glass is filled with a solution of uniform color. This phenomenon, called as the Brownian motion, can be described by

Fick's first law [Fick, 1855], which states that the net particle flux $\mathbf{J}$ is proportional to the spatial gradient of the particle concentration C, i.e.

$$\mathbf{J} = -D\frac{\partial}{\partial x}C$$

where D is the diffusion coefficient. Then consider the continuity equation, the net influx equals the increase in the concentrate, i.e. $-\frac{\partial}{\partial x}\mathbf{J} = \frac{\partial}{\partial t}C$, then we have

$$\frac{\partial}{\partial t}C = D\frac{\partial^2}{\partial x^2}C \tag{4.1}$$

This is the diffusion equation in 1D space. The solution of this equation with the boundary condition $C(x,0) = \delta(x - x_0)$ is

$$C(x,t) = \frac{1}{\sqrt{4\pi t D}}e^{-\frac{(x-x_0)^2}{4Dt}} = N(x|x_0, 2tD) \tag{4.2}$$

where $N(x|x_0, 2tD)$ is the Gaussian distribution with mean x_0 and variance $2Dt$. The derivation of this solution can be found in Example 2.3 in details. The free diffusion in 3D space can be described by diffusion tensor $\mathbf{D}$ which is a 3×3 symmetric positive definite matrix.

$$\mathbf{J} = -\mathbf{D}\nabla C, \quad -\nabla\cdot\mathbf{J} = \frac{\partial}{\partial t}C$$

Then the solution with the boundary condition $C(\boldsymbol{x},0) = \delta(\boldsymbol{x} - \boldsymbol{x}_0)$ is

$$C(\boldsymbol{x},t) = \frac{1}{\sqrt{(4\pi t)^3|\mathbf{D}|}}e^{-\frac{(\boldsymbol{x}-\boldsymbol{x}_0)^T\mathbf{D}^{-1}(\boldsymbol{x}-\boldsymbol{x}_0)}{2t}} = N(\boldsymbol{x}_0|0, 2t\mathbf{D}) \tag{4.3}$$

Note that Dr. Einstein also obtained the solution by considering the diffusion displacement Probability Density Function (PDF) or called diffusion propagator $P(\mathbf{R},t)$ of the particle [Einstein, 1956].

$$P(\mathbf{R},t) = \frac{1}{\sqrt{(4\pi t)^3|\mathbf{D}|}}e^{-\frac{\mathbf{R}^T\mathbf{D}^{-1}\mathbf{R}}{2t}} = N(\mathbf{R}|0, 2t\mathbf{D}) \tag{4.4}$$

Then the Mean Squared Displacement (MSD) is

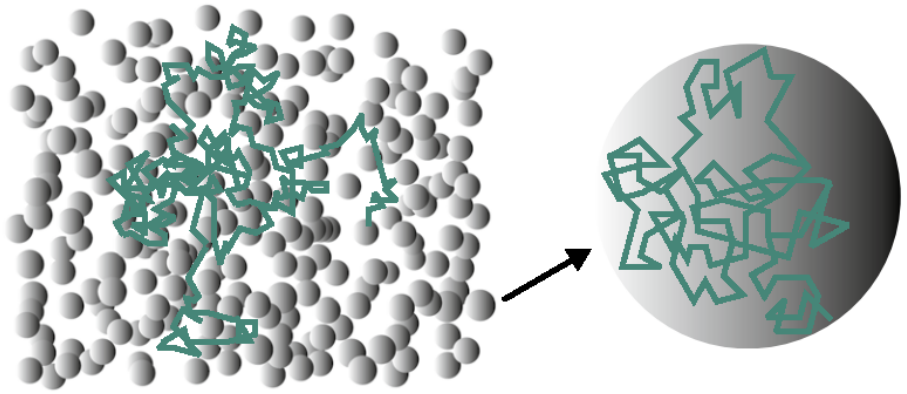
$$\langle\mathbf{R}\mathbf{R}^T\rangle = 2t\mathbf{D} \tag{4.5}$$

Note that the above equations are for free diffusion. However the water diffusion in biological tissues may be hindered by biological cells and other environment surrounding tissues. See Fig. 4.1. So far we still do not know the diffusion propagator in complex environment because it depends on the complex boundary conditions which are hard to be quantified.

4.1.2 Magnetic Resonance Imaging (MRI), k-space and x-space

The principles of Magnetic Resonance Imaging (MRI) are based on **spin** which is the rotation of a particle around some axis. Spin is a fundamental quantum characteristic of elementary particles like protons, electrons. Some nuclei have the property to align with a magnetic field $\mathbf{B}_0$ if their mass number, i.e. the summation number of protons

Figure 4.1: The water diffusion in biological tissues may be hindered by biological cells and other environment surrounding tissues. The figure is from [Johansen-Berg and Behrens, 2009].



and neutrons, is odd. Essentially their spin aligned along $\mathbf{B}_0$. Without the external stimulus by magnetic field $\mathbf{B}_0$, the macroscopic magnetization $\mathbf{M} = 0$. In MRI, the particles considered are hydrogen nucleus because human body is largely composed of water molecules. Each water molecule has two hydrogen nuclei or protons. When a person is inside the magnetic field $\mathbf{B}_0$ of the scanner, the average magnetic moment of many protons becomes aligned with the direction of the field $\mathbf{B}_0$, which is assumed as the z-axis by convection Meanwhile, the magnetization vector of spins precesses around $\mathbf{B}_0$ with an angular frequency known as the **Larmor frequency**, i.e.

$$\omega_0 = \gamma \mathbf{B}_0$$

where γ is the **gyromagnetic ratio** dependent in the particle. Then the net magnetization satisfies

$$\frac{\mathrm{d}\mathbf{M}}{\mathrm{d}t} = \gamma \mathbf{M} \times \mathbf{B}_0 \tag{4.6}$$

When a Radio-Frequency (RF) is applied to the spins with the resonance frequency, the energy of RF is absorbed by the spins with low energy configuration and changes them into high energy configuration. Then the spins change their alignments. After RF is turned off, the spins begin to recover the alignment with $\mathbf{B}_0$, and finally return to the thermal equilibrium with low energy configuration. This is called as the **relaxation phase**. Normally 90° and 180° RFs are used, which change the direction of spin with 90° or 180°. Assume the magnetization $\mathbf{M} = \mathbf{M}_x + \mathbf{M}_y + \mathbf{M}_z$, then $\mathbf{M}(t)$ in the relaxation phase satisfies the famous **Bloch equation** as follows[1].

$$\frac{\mathrm{d}\mathbf{M}}{\mathrm{d}t} = \gamma \mathbf{M} \times \mathbf{B} - \frac{1}{T_1}(\mathbf{M}_z - \mathbf{M}_0) - \frac{1}{T_2}(\mathbf{M}_y + \mathbf{M}_x) \tag{4.7}$$

where $\mathbf{B}$ is the magnetic field, $\mathbf{M}_0$ is the original $\mathbf{M}$ when relaxation phase starts, T_1 and T_2 are relaxation time for M_z in z-axis and $\mathbf{M}_x$, $\mathbf{M}_y$ in x-y plane. The equation characterizes the relaxation process.

Add a gradient field $\mathbf{G}(t) = G_x(t)\boldsymbol{e}_x + G_y(t)\boldsymbol{e}_y + G_z(t)\boldsymbol{e}_z$ into the static field $\mathbf{B}_0 = B_0\boldsymbol{e}_z$, then the total magnetic field is

$$\mathbf{B}(t) = B_0\boldsymbol{e}_z + G_x(t)\boldsymbol{e}_x + G_y(t)\boldsymbol{e}_y + G_z(t)\boldsymbol{e}_z$$

[1] http://en.wikipedia.org/wiki/Bloch_equations

Denote $\mathbf{M}_{xy}(x,y,z,t) = \mathbf{M}_x + i\mathbf{M}_y$, based on the Bloch equation, we have

$$\mathbf{M}_{xy}(x,y,z,t) = \mathbf{M}_{xy}(x,y,z,0)e^{-t/T_2}e^{-i\gamma B_0 t}e^{-i\gamma\int_0^t(G_x(\tau)\boldsymbol{e}_x+G_y(\tau)\boldsymbol{e}_y+G_z(\tau)\boldsymbol{e}_z)\mathrm{d}\tau}$$

The magnetic resonance signal $s(t)$ is the spatial summation of $\mathbf{M}_{xy}(x,y,z,t)$ [Huettel et al., 2009], i.e.

$$s(t) = \int_x\int_y\int_z \mathbf{M}_{xy}(x,y,z,0)e^{-t/T_2}e^{-i\omega_0 t}e^{-i\gamma\int_0^t(G_x(\tau)\boldsymbol{e}_x+G_y(\tau)\boldsymbol{e}_y+G_z(\tau)\boldsymbol{e}_z)\mathrm{d}\tau}\mathrm{d}x\mathrm{d}y\mathrm{d}z$$

In practice, we do not need to consider $e^{-i\omega_0 t}$ because modern MRI scanners demodulate the detected signal with the resonance frequency ω_0. The term e^{-t/T_2} is independent of the spatial position $\boldsymbol{x}$. By ignoring these two terms, we have

$$s(t) = \int_x\int_y\int_z \mathbf{M}_{xy}(x,y,z,0)e^{-i\gamma\int_0^t(G_x(\tau)\boldsymbol{e}_x+G_y(\tau)\boldsymbol{e}_y+G_z(\tau)\boldsymbol{e}_z)\mathrm{d}\tau}\mathrm{d}x\mathrm{d}y\mathrm{d}z$$

For slice z_0 with thickness $\triangle z$, we consider

$$\mathbf{M}(x,y,z_0) = \int_{z_0-\frac{\triangle z}{2}}^{z_0+\frac{\triangle z}{2}} \mathbf{M}_{xy}(x,y,z,0)\mathrm{d}z$$

The magnetization $\mathbf{M}(x,y,z_0)$ describes the tissue property at position (x,y,z_0) in **x**-space. Then we have

$$s(t) = \int_x\int_y \mathbf{M}_{z_0}(x,y)e^{-i\gamma\int_0^t G_x(\tau)\boldsymbol{e}_x+G_y(\tau)\boldsymbol{e}_y\mathrm{d}\tau}\mathrm{d}x\mathrm{d}y \tag{4.8}$$

Define **k** vector as

$$\mathbf{k} = (k_x,k_y)^T,\quad k_x = \frac{\gamma}{2\pi}\int_0^t G_x(\tau)\mathrm{d}\tau,\quad k_y = \frac{\gamma}{2\pi}\int_0^t G_y(\tau)\mathrm{d}\tau \tag{4.9}$$

Then the magnetic resonance signal $s(k_x,k_y,z)$ in **k**-space is related with net magnetic $\mathbf{M}(x,y,z)$ with a two dimension (2D) Fourier transform, i.e.

$$s(k_x,k_y,z) = \int_x\int_y \mathbf{M}(x,y,z)e^{-2\pi i(k_x(t)x+k_y(t)y)}\mathrm{d}x\mathrm{d}y = \mathscr{F}_{2D}\{\mathbf{M}(x,y,z)\}(k_x,k_y,z) \tag{4.10}$$

where $\mathscr{F}_{2D}\{\cdot\}$ denotes the Fourier transform in 2D space.

There are several common conventions for defining the **Fourier transform**[2]. Different convention obtains different formulae in form, although theoretically they are equivalent. In dMRI domain, different papers use different conventions, which sometimes makes readers confused. Throughout this thesis, we use the convention of Fourier transform defined as

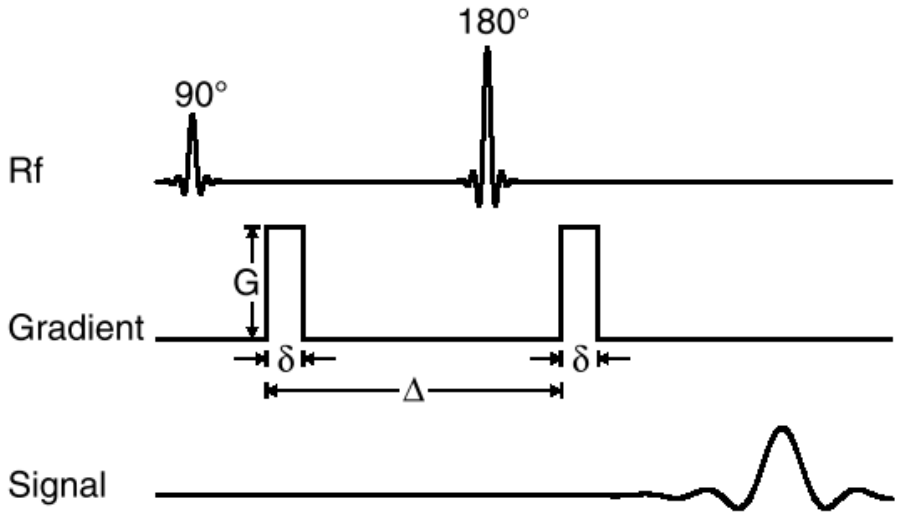
$$\hat{f}(\xi) = \mathscr{F}\{f(x)\}(\xi) \overset{\text{def}}{=} \int_{-\infty}^{\infty} f(x)e^{-2\pi i x\xi}\mathrm{d}x \tag{4.11}$$

In this convention, the inverse Fourier transform is

$$f(x) = \mathscr{F}^{-1}\{\hat{f}(\xi)\}(x) = \int_{-\infty}^{\infty}\hat{f}(\xi)e^{2\pi i x\xi}\mathrm{d}\xi \tag{4.12}$$

[2] http://en.wikipedia.org/wiki/Fourier_transform

Figure 4.2: Pulsed Gradient Spin-Echo (PGSE) sequence introduced by Stejskal and Tanner [Stejskal and Tanner, 1965]. δ is the duration of the diffusion gradient pulses and Δ is the time between two diffusion gradient pulses.



4.1.3 Diffusion Weighted Imaging (DWI), q-space and R-space

Diffusion Gradient Sequence

The classical diffusion gradient sequence used in dMRI is the Pulsed Gradient Spin-Echo (PGSE) sequence proposed by Stejskal and Tanner [Stejskal and Tanner, 1965]. See Fig. 4.2 for the sketch map of this sequence. PGSE sequence has two gradient pulses $\mathbf{G}(t)$ with duration time δ. The 90° RF pulse, translates the spins into the transverse plane, i.e. x-y plane, considering the $\mathbf{B}_0$ is along z-axis. Then the spins precess around $\mathbf{B}_0$ with the resonance frequency ω_0. Due to local magnetic field inhomogeneities, some spins slow down and some spins speed up.[3] After time Δ between two pulses, the 180° RF pulse refocuses the phase of spins so that slower spins lead ahead and the fast ones trail behind. The spin echo process occurs when the spins recover their net magnetization.

Note that the PGSE sequence uses rectangular gradient lobes. There are also other kinds of gradient lobes commonly used in dMRI [Bernstein et al., 2004, chap. 9].

Free Diffusion, Stejskal-Tanner Equation

If there is spin displacement as a result of Brownian motion, we can add the diffusion term in Bloch equation in 4.7.

$$\frac{\mathrm{d}\mathbf{M}}{\mathrm{d}t} = \gamma\mathbf{M}\times\mathbf{B} - \frac{1}{T_1}(\mathbf{M}_z - \mathbf{M}_0) - \frac{1}{T_2}(\mathbf{M}_y + \mathbf{M}_x) + \nabla\cdot\mathbf{D}\nabla(\mathbf{M} - \mathbf{M}_0) \tag{4.13}$$

where $\mathbf{D}$ is the diffusion tensor used in Diffusion Tensor Imaging in subsection 4.2. To eliminate the dependence of spin density, we need at least two measurements

[3] http://en.wikipedia.org/wiki/Spin_echo

of Diffusion Weighted Imaging (DWI) signals, i.e. $S(b)$ with the diffusion weighting factor b in Eq. (4.14) introduced by Dr. Lebihan in [LeBihan et al., 1986], and $S(0)$ with $b = 0$ which is the baseline signal without any gradient.

$$b = \gamma^2\delta^2(\Delta - \frac{\delta}{3})\|\mathbf{G}\|^2 \tag{4.14}$$

In the b value in Eq. (4.14), γ is the proton gyromagnetic ratio, $\mathbf{G} = \|\mathbf{G}\|\mathbf{u}$ is the diffusion sensitizing gradient pulse.

$$\tau = \Delta - \frac{1}{3}\delta \tag{4.15}$$

is normally used to describe the effective diffusion time [Bihan et al., 1986; Basser et al., 1994]. The b value is dependent on the sequence, and it is different in different kinds of lobes in diffusion sequence [Bernstein et al., 2004, chap. 9]. The signal intensity at each voxel in DWI is dependent on both surrounding structures and given weighted magnetic gradient [Bihan et al., 1986]. See Fig. 4.3 for the DWI images $S(b)$ with different b values and different gradient directions $\mathbf{u}$. It can be seen that the DWI images are very noise, especially for large b values.

With the PGSE sequence described above, the diffusion weighted **signal attenuation** $E(b) = \frac{S(b)}{S(0)}$ is given by Stejskal-Tanner equation [Stejskal and Tanner, 1965]

$$E(b) = \frac{S(b)}{S(0)} = \exp(-bD) \tag{4.16}$$

where D is known as the Apparent Diffusion Coefficient (ADC) which reflects the property of surrounding tissues. Note that in general case ADC D is also dependent on $\mathbf{G}$ in a complex way, however free diffusion assumes D is only dependent on the direction of $\mathbf{G}$, i.e. $\mathbf{u} = \frac{\mathbf{G}}{\|\mathbf{G}\|}$.

The early works in dMRI reported that the ADC D is dependent on gradient direction $\mathbf{u}$ and used two or three DWI images in different directions to detect the properties of tissues [Moseley et al., 1990; Douek et al., 1991]. Then Dr. Basser introduced diffusion tensor [Basser et al., 1994] to represent ADC as

$$D(\mathbf{u}) = \mathbf{u}^T\mathbf{D}\mathbf{u} \tag{4.17}$$

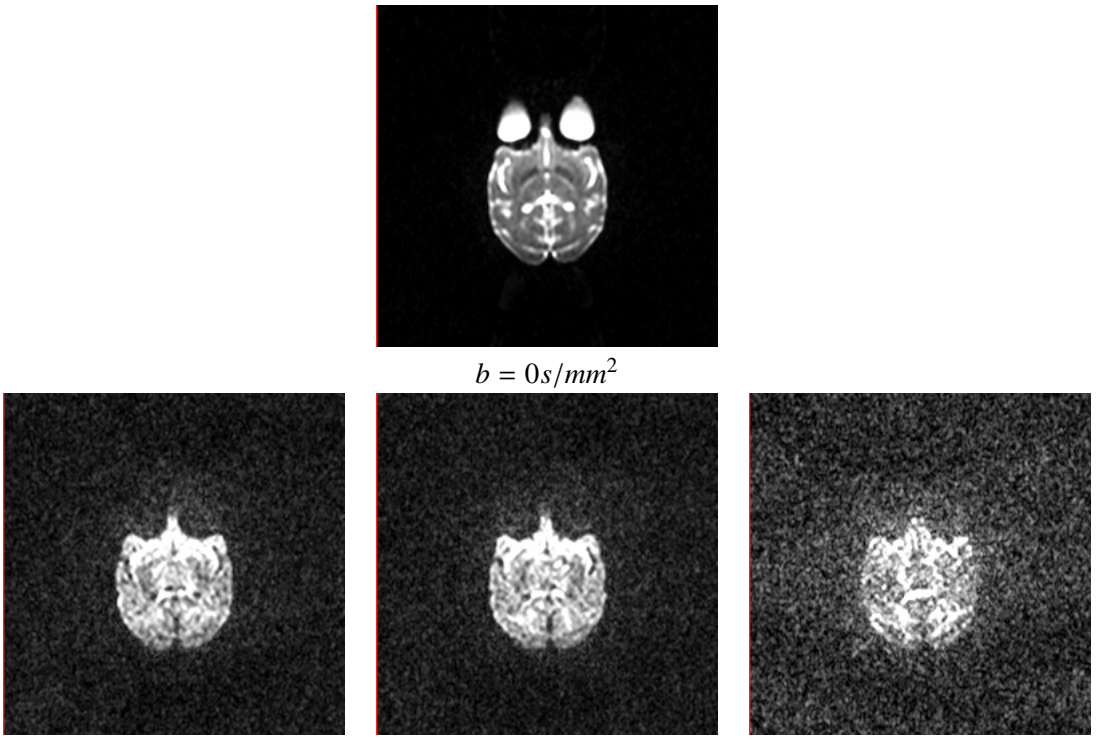
$\mathbf{D}$ is called as the **diffusion tensor**, which is a 3×3 symmetric positive definite matrix independent of $\mathbf{u}$. This method is called as Diffusion Tensor Imaging (DTI), which is the most common method nowadays in dMRI field. See Section 4.2 for more materials in DTI.

Narrow Pulse Condition, q-space, Ensemble Average Propagator

Based on free diffusion assumption, Stejskal-Tanner equation reveals the underlying ADC value (or diffusion tensor) from the measured diffusion signal attenuation. However, the diffusion of water molecules is hindered by surrounding tissues, especially in white matter. See Fig. 4.1 for the hindered diffusion. In a general diffusion process, the Gaussian propagator assumption is not satisfied.

For each voxel in $\mathbf{x}$-space, let $\rho(\mathbf{R}_0)$ denote the spin density at initial time $t = 0$ and $P(\mathbf{R}_\Delta|\mathbf{R}_0)$ denote the probability that the spin moves from $\mathbf{R}_0$ at $t = 0$ to $\mathbf{R}_\Delta$ at

Figure 4.3: DWI images for different b-values and gradients. The data is from one of the subjects in the real monkey dataset described in Appendix B.2.



$b = 0s/mm^2$

$b = 1500s/mm^2$
$\mathbf{u} = (-0.204, 0.515, 0.833)^T$

$b = 1500s/mm^2$
$\mathbf{u} = (0.198, 0.515, 0.834)^T$

$b = 3000s/mm^2$
$\mathbf{u} = (-0.204, 0.515, 0.833)^T$

$t = \Delta$. Then the diffusion signal attenuation $E(\mathbf{G}, \Delta, \delta) = \frac{S(\mathbf{G},\Delta,\delta)}{S(0)}$ can be represented as [Stejskal and Tanner, 1965; Callaghan, 1991]

$$E(\mathbf{G}, \Delta, \delta) = \int_{\mathbb{R}^3} \rho(\mathbf{R}_0) \int_{\mathbb{R}^3} P(\mathbf{R}_\Delta|\mathbf{R}_0) \exp\left(i\gamma(\mathbf{R}_\Delta - \mathbf{R}_0)^T \left(\int_0^\delta \mathbf{G}(t)\mathrm{d}t\right)\right) \mathrm{d}\mathbf{R}_\Delta \mathrm{d}\mathbf{R}_0 \tag{4.18}$$

where γ is gyromagnetic ratio, $S(0)$ is the baseline DWI signal without diffusion gradient, $S(\mathbf{G}, \Delta, \delta)$ is the DWI signal with imaging parameters $(\mathbf{G}, \Delta, \delta)$. Under narrow pulse condition, i.e. the duration time δ is much smaller than the separation time between two pluses Δ, $\mathbf{G}(t)$ is a constant $\mathbf{G}$ during δ. Then we introduce $\mathbf{q}$ vector as

$$\mathbf{q} = q\mathbf{u} = (2\pi)^{-1}\gamma \int_0^\delta \mathbf{G}(t)\mathrm{d}t = (2\pi)^{-1}\gamma\delta\mathbf{G} \tag{4.19}$$

which can be seen as a vector in $\mathbf{q}$-space. We also define the Ensemble Average Propagator (EAP) as

$$P(\mathbf{R}) = \int_{\mathbb{R}^3} \rho(\mathbf{R}_0) P(\mathbf{R}_0 + \mathbf{R}|\mathbf{R}_0)\mathrm{d}\mathbf{R}_0, \quad \mathbf{R} = \mathbf{R}_\Delta - \mathbf{R}_0 \tag{4.20}$$

where $\mathbf{R} = R\mathbf{r}$ is the displacement vector in $\mathbf{R}$-space. $\mathbf{u}$ and $\mathbf{r}$ are unit vectors. Then the signal attenuation can be written as the inverse Fourier transform of EAP $P(\mathbf{R})$, i.e.

$$E(\mathbf{q}) = \int_{\mathbb{R}^3} P(\mathbf{R}) \exp\left(2\pi i \mathbf{q}^T \mathbf{R}\right) \mathrm{d}\mathbf{R} = \mathscr{F}_{3D}^{-1}\{P(\mathbf{R})\}(\mathbf{q}) \tag{4.21}$$

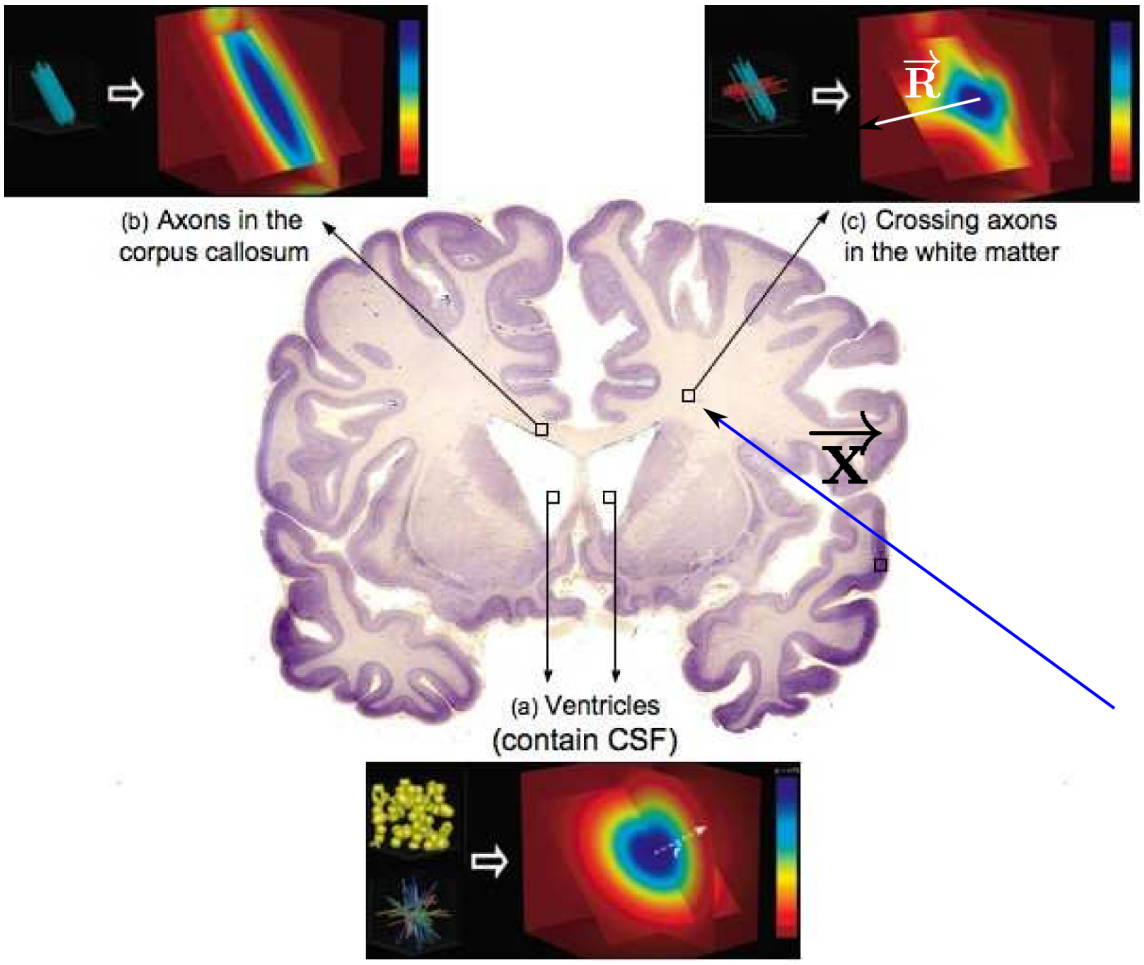


Figure 4.4: 3D **x**-space and 3D **R**-space. EAPs in different regions in brain reflect different micro-structures with isotropic diffusion, single fiber and crossing fibers. The image is taken from [Descoteaux, 2008] with the original figures adapted from [Hagmann et al., 2006] and the brain museum (www.brainmuseum.org/Specimens).

So based on the narrow pulse assumption, the diffusion process at each voxel is fully described by the so called EAP $P(\mathbf{R})$ which is the average displacement probability in 3D **R**-space. Then the diffusion weighting factor can be represented by **q**, i.e.

$$b = \gamma^2\delta^2(\Delta - \frac{\delta}{3})\|\mathbf{G}\|^2 = 4\pi^2\tau q^2 \tag{4.22}$$

where $\tau = \Delta - \frac{1}{3}\delta$ is the effective diffusion time [Bihan et al., 1986; Basser et al., 1994]. In the following part of this thesis, $E(\mathbf{q})$ (or written as $E(b)$) is also called as diffusion signal if no confusion in context. EAPs in different regions in brain reflect the different micro-structures and reveal fiber directions. Please see Fig. 4.4, which also demonstrates the diffusion data is in six dimensional space. The acquisition of dMRI data is performed in a 6D space, i.e. 3D **k**-space and 3D **q**-space. Then the DWI data $S(\mathbf{q})$ is the Fourier transform of **k**-space signal, and the EAP is another Fourier transform of $E(\mathbf{q})$. Thus there are two dualities in dMRI, i.e. **k**-space and **x**-space, **q**-space and **R**-space.

Note that $P(\mathbf{R})$ is related with $E(\mathbf{q})$ by a Fourier transform only when narrow pulse

assumption is satisfied. However in practice, this assumption is not always satisfied. Several works [Mair et al., 2002; Bar-Shir et al., 2008] showed that even when this assumption does not hold, the Fourier relation can be still used to obtain good results. In this thesis we always assume this assumption holds and use this Fourier relation to obtain analytical formulae in Chapter 5 and 6.

Note that $P(\mathbf{R})$ is assumed to be antipodally symmetric (or called radially symmetric), i.e. $P(\mathbf{R}) = P(-\mathbf{R})$, because asymmetrical diffusion propagator violates the principle of microscopic detailed balance. Because of $P(\mathbf{R}) = P(-\mathbf{R})$, we have $E(\mathbf{q}) = E(-\mathbf{q})$. Since $E(\mathbf{q}) = E(-\mathbf{q})$, Eq. (4.21) can be written as

$$E(\mathbf{q}) = E(-\mathbf{q}) = \int_{\mathbb{R}^3} P(\mathbf{R}) \exp(-2\pi i \mathbf{q}^T \mathbf{R}) d\mathbf{R} = \mathscr{F}\{P(\mathbf{R})\}(\mathbf{q}) = \int_{\mathbb{R}^3} P(\mathbf{R}) \cos(2\pi \mathbf{q}^T \mathbf{R}) d\mathbf{R} \quad (4.23)$$

Then the EAP is the Fourier transform of signal, i.e.

$$P(\mathbf{R}) = \mathscr{F}\{E(\mathbf{q})\}(\mathbf{R}) = \int_{\mathbb{R}^3} E(\mathbf{q}) \exp(-2\pi i \mathbf{q}^T \mathbf{R}) d\mathbf{q} = \int_{\mathbb{R}^3} E(\mathbf{q}) \cos(2\pi \mathbf{q}^T \mathbf{R}) d\mathbf{q} \quad (4.24)$$

Both Eq. (4.21), Eq. (4.23) and Eq. (4.24) are commonly used in papers. Because of the antipodal symmetry of $E(\mathbf{q})$, the DWI samples in $\mathbf{q}$-space can be performed only in a half space, e.g. $\{\mathbf{q} \in \mathbb{R}^3 : \ (0,0,1)\mathbf{q} = q_z \geq 0\}$.

Historically, people first measured ADC values from two DWI images based on Eq. (4.16), then tensor image from at least 7 DWI images in DTI [Basser et al., 1994] in Section 4.2, and finally 3D images of the full PDF-valued images from more DWI images in High Angular Resolution Imaging (HARDI) in Section 4.3. All these kinds of images (scalar-valued, tensor-valued, or PDF-valued) estimated from DWIs let us understand better and better the microstructure of biological tissues.

In dMRI field, there are many research subareas as shown in Fig. 1.1. In this thesis we are interested in reconstruction methods which estimate scalar, tensor, PDF-valued images from DWI images, and the metric used to process pre-estimated PDF-valued data. So in the following sections of this chapter, we will review some reconstruction models and metrics for PDF-valued data. Different reconstruction model/method normally uses different sampling scheme in $\mathbf{q}$-space. Please see Fig. 4.5 on several kinds of sampling schemes used in reconstruction models/methods.

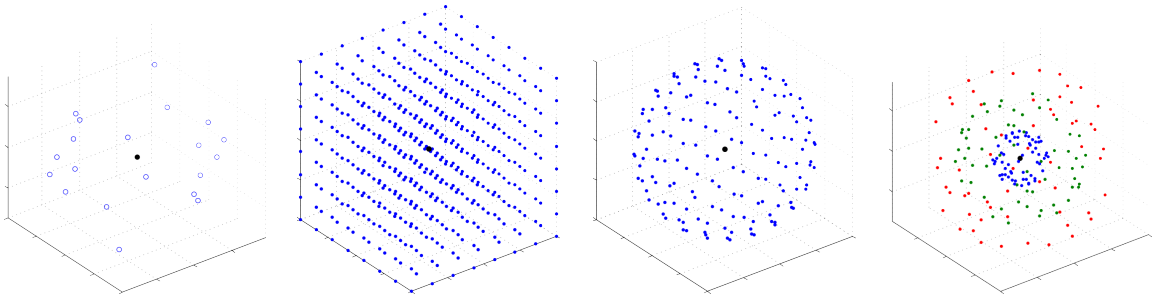
4.2 DIFFUSION TENSOR IMAGING (DTI)

Dr. Basser proposed to model the ADC as a quadratic form parameterized by the diffusion tensor $\mathbf{D}$ in Eq. (4.17) [Basser et al., 1994]. Then the Stejskal-Tanner equation becomes

$$E(b) = \frac{S(b)}{S(0)} = \exp(-b\mathbf{u}^T \mathbf{D}\mathbf{u}) \quad (4.25)$$

The diffusion tensor $\mathbf{D} \in Sym_3^+$ is independent of b value and gradient direction $\mathbf{u}$, where Sym_3^+ is the space of 3×3 symmetric positive definite matrix. $\mathbf{D}$ can be eigen-decomposed into three positive eigenvalues and corresponding eigenvectors, which is useful to define some scalar indices containing biological meaning. See subsection 4.2.2.

$$\mathbf{D} = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{pmatrix} = \lambda_1 \boldsymbol{v}_1 \boldsymbol{v}_1^T + \lambda_2 \boldsymbol{v}_2 \boldsymbol{v}_2^T + \lambda_3 \boldsymbol{v}_3 \boldsymbol{v}_3^T \quad (4.26)$$



(a) sampling in DTI (b) sampling in DSI (c) single shell sampling (d) sparse sampling

Figure 4.5: Several kinds of sampling in $\mathbf{q}$-space. The black dot in $\mathbf{q} = (0,0,0)^T$ is the baseline image without diffusion gradient. Note that although we showed sampling in $\mathbb{R}^3$, normally only samples in a half space is used, e.g. $(0,0,1)\mathbf{q} = q_z \geq 0$. (a) sampling used in DTI, normally less than 20 DWI images are used; (b) dense Cartesian sampling used in DSI. Note in practice the Cartesian samples inside a given Ball are used; (c) single shell sampling used in sHARDI methods, e.g. QBI, DOT etc; (d) sparse sampling used in mHARDI methods, e.g. DPI, SHORE, SPFI. Note although normally multiple shell sampling is used, any sampling scheme can be used in mHARDI methods.

The free diffusion in coherent fibers can be represented by Gaussian propagator parameterized by diffusion tensors. See Fig. 4.6 for the sketch map of tensor representation and free diffusion along fibers.

In general diffusion process, the EAP $P(\mathbf{R})$ is not a Gaussian distribution, especially in the area with crossing fibers. In free diffusion, $P(\mathbf{R})$ is Gaussian distribution parameterized by the diffusion tensor $\mathbf{D}$, which is the Fourier transform of $E(\mathbf{q})$ in Eq. (4.25), i.e.

$$P(\mathbf{R}) = \mathscr{F}\{\exp(-4\pi^2\tau\mathbf{q}^T\mathbf{D}\mathbf{q})\} = N(\mathbf{R}|2\tau\mathbf{D}) = \frac{1}{\sqrt{(4\pi\tau)^3|\mathbf{D}|}}\exp\left(\frac{-\mathbf{R}^T\mathbf{D}^{-1}\mathbf{R}}{4\tau}\right) \tag{4.27}$$

The covariance matrix of the EAP is $\Sigma = 2\tau\mathbf{D}$. Please see Appendix 5.8 for the derivation.

4.2.1 Tensor Estimation

For given measured diffusion signal samples $\{E(b_i)\}_{i=1}^{N_s}$, rewrite Eq. (4.25) as

$$-\ln E(b_i) = b_i\mathbf{u}_i^T\mathbf{D}\mathbf{u}_i = \mathbf{B}_i^T\phi(\mathbf{D})$$

where

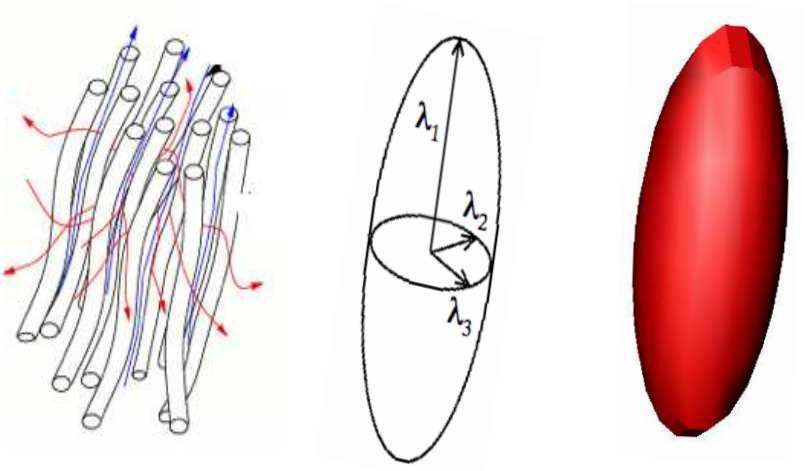
$$\phi(\mathbf{D}) = (D_{xx}, D_{xy}, D_{xz}, D_{yy}, D_{yz}, D_{zz})^T$$
$$\mathbf{B}_i = b_i(u_i^x u_i^x, 2u_i^x u_i^y, 2u_i^x u_i^z, u_i^y u_i^y, 2u_i^y u_i^z, u_i^z u_i^z)^T, \quad \mathbf{u}_i = (u_i^x, u_i^y, u_i^z)^T$$

Then the solution of least square estimation is

$$\phi(\mathbf{D}) = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{Y} \tag{4.28}$$

where

$$\mathbf{B} = \begin{pmatrix}\mathbf{B}_1^T\\ \vdots \\ \mathbf{B}_{N_s}^T\end{pmatrix} = \begin{pmatrix} b_1u_1^xu_1^x & 2b_1u_1^xu_1^y & 2b_1u_1^xu_1^z & b_1u_1^yu_1^y & 2b_1u_1^yu_1^z & b_1u_1^zu_1^z \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{N_s}u_{N_s}^xu_{N_s}^x & 2b_{N_s}u_{N_s}^xu_{N_s}^y & 2b_{N_s}u_{N_s}^xu_{N_s}^z & b_{N_s}u_{N_s}^yu_{N_s}^y & 2b_{N_s}u_{N_s}^yu_{N_s}^z & b_{N_s}u_{N_s}^zu_{N_s}^z \end{pmatrix}_{N_s\times 6}, \quad \mathbf{Y} = \begin{pmatrix}-\ln E(b_1)\\ \vdots \\ -\ln E(b_{N_s})\end{pmatrix}_{N_s\times 1}$$



H_2O Brownian motion along the fibers [Poupon, 1999]

Eigen decomposition of the DT

Ellipsoidal visualization of the DT

Figure 4.6: Diffusion tensor representation from [Descoteaux, 2008].

Diffusion tensor $\mathbf{D}$ can be estimated from measured samples of diffusion signal $\{E(b_i)\}$ through the above simple least square method or weighted least square method [Basser et al., 1994], or more complex methods which consider positive definite constraint or Rician noise [Tschumperlé and Deriche, 2003; Chefd'hotel et al., 2004; Koay et al., 2006; Fillard et al., 2007].

Note that although many works estimate tensor $\mathbf{D}$ from single shell data, i.e. the data with single b value, based on Eq. (4.25) the tensor estimation is actually independent of the sampling in $\mathbf{q}$-space. Once more than six DWI images and one baseline image without diffusion are measured, the tensor $\mathbf{D}$ can be estimated from various methods. However, the different sampling has different estimation quality. If single shell data is used, the optimal b value was reported to range in $(700, 1500) s/mm^2$ [Jones et al., 1999; Alexander and Barker, 2005], and normally about twenty DWI images are used in DTI in clinical study. Please see Fig. 4.5(a) for the sketch map of the sampling scheme normally used in DTI.

4.2.2 Scalar Indices of Tensor

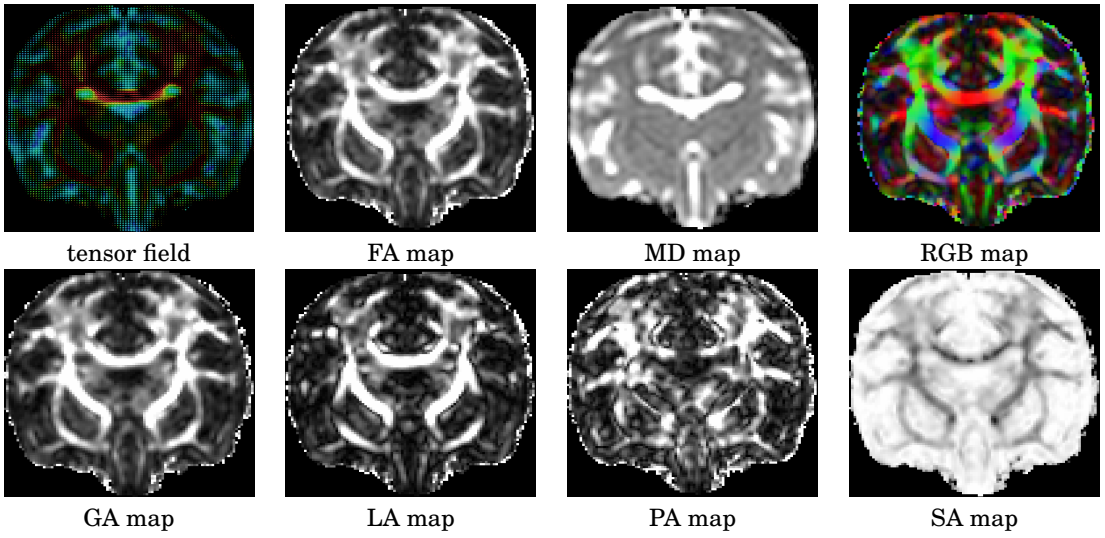
Some useful scalar indices can be obtained from tensor $\mathbf{D}$. The most important two indices are Fractional Anisotropy (FA) and Mean Diffusivity (MD) [Pierpaoli and Basser, 1996] defined as

$$\mathrm{FA} = \frac{\sqrt{3}\|\mathbf{D} - \frac{1}{3}\mathrm{Trace}(\mathbf{D})\mathbf{I}\|}{\sqrt{2}\|\mathbf{D}\|} = \sqrt{\frac{3}{2}}\sqrt{\frac{(\lambda_1-\bar{\lambda})^2+(\lambda_2-\bar{\lambda})^2+(\lambda_3-\bar{\lambda})^2}{\lambda_1^2+\lambda_2^2+\lambda_3^2}} \tag{4.29}$$

$$\mathrm{MD} = \frac{1}{3}\mathrm{Trace}(\mathbf{D}) = \frac{\lambda_1+\lambda_2+\lambda_3}{3} \tag{4.30}$$

where we assume the eigen-decomposition is $\mathbf{D} = \lambda_1\boldsymbol{v}_1\boldsymbol{v}_1^T + \lambda_2\boldsymbol{v}_2\boldsymbol{v}_2^T + \lambda_3\boldsymbol{v}_3\boldsymbol{v}_3^T$, $\lambda_1 \geq \lambda_2 \geq \lambda_3$, and $\bar{\lambda} = \frac{\lambda_1+\lambda_2+\lambda_3}{3}$. MD and FA have been used in many clinical applications [Mori, 2007;

Figure 4.7: Tensor field and the scalar maps estimated from the monkey data with $b = 1500 s/mm^2$ described in Appendix B.



Johansen-Berg and Behrens, 2009]. For example, MD is known to be useful in stroke study. [Westin et al., 2002] proposed the linear, planar and spherical anisotropies which describe the elongated, oblate and spherical shape configurations of tensors.

$$\text{LA} = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3}, \quad \text{PA} = \frac{2(\lambda_2 - \lambda_3)}{\lambda_1 + \lambda_2 + \lambda_3}, \quad \text{SA} = \frac{3\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3} \tag{4.31}$$

These indices can be obtained from the eigenvalues of $\mathbf{D}$. The eigenvectors of $\mathbf{D}$ are also very useful. The first eigenvector corresponding to the largest eigenvalue can be used as the estimated fiber direction. It can also be encoded in Red-Blue-Green (RGB) map to describe the fiber directions. The tensor $\mathbf{D}$ itself can be visualized by a ellipsoid, then the tensor field becomes ellipsoid field. Please refer Fig. 4.7 for the tensor field and various scalar maps estimated from the monkey data with $b = 1500 s/mm^2$ described in Appendix B, where the Geodesic Anisotropy (GA) is introduced in the Riemannian framework for tensors in subsection 4.4.1.

Summary of Advantages and Limitations:

- ✔ DTI only needs more than 6 diffusion weighted images and one baseline image. Normally around 20 DWIs are used in DTI model. See Fig. 4.5
- ✔ DTI provides some useful scalar indices like FA and MD which have many applications in clinical studies.
- ✔ The Gaussian propagator is adequate to model the areas with isotropic diffusion or single dominant direction of diffusion.
- ✘ Gaussian assumption is not appropriate in whole brain. DTI model can not represent complex fiber configuration like fiber crossing.

4.3 HIGH ANGULAR RESOLUTION DIFFUSION IMAGING (HARDI)

The term High Angular Resolution Diffusion Imaging (HARDI) was first proposed by Tuch [Tuch et al., 1999, 2002], where a finer angular resolution sampling scheme than conventional DTI sampling scheme was considered. The original HARDI term in [Tuch et al., 1999, 2002] means single shell sampling (only one b value). See Fig. 4.5(c). However the mixture of tensor model in [Tuch et al., 1999, 2002] actually can be also used in Cartesian sampling in Fig. 4.5(b) and sparse sampling (multiple b values) in Fig. 4.5(d). With the development of MRI scanner, the acquisition time is reduced, which makes multiple shell data more practical and maybe available in clinical study in the near future. Some research works proposed to estimate Orientation Distribution Functions or EAPs in multiple shell sampling [Liu et al., 2004; Assemlal et al., 2009a; Özarslan et al., 2009; Descoteaux et al., 2010]. Thus in this thesis, the term **HARDI methods** include all modeling methods beyond DTI. The HARDI methods which only can be used in single shell data are called as sHARDI methods. The HARDI methods which can be used in multiple shell data are called as mHARDI methods.

4.3.1 Generalization of Diffusion Tensor Imaging

Since DTI has been successfully used in clinical study since 1990s, although it is based on Gaussian propagator, a straightforward idea is to generalize the tensor model in DTI to non-Gaussian case. There are several ways to perform this generalization, including mixture of tensor model [Tuch et al., 2002; Hosey et al., 2005; Assaf et al., 2004], Generalized DTI [Liu et al., 2004], High Order Tensor (HOT) [Özarslan and Mareci, 2003].

Mixture of Tensor Model

Mixture of tensor model is a natural generalization of tensor model, where the signal is assumed to be a mixture of signals generated from tensors $\{\mathbf{D}_i\}_{i=1}^{K}$.

$$E(b) = \sum_{i}^{K} w_i \exp(-b\mathbf{u}^T \mathbf{D}_i \mathbf{u}) \tag{4.32}$$

Based on some biological priors, the number of tensors is normally less than 3, typically $K = 2$. Unlike tensor model this model can not be solved by least square. People normally use gradient descent method (typically the Levenberg-Marquardt minimization) [Tuch et al., 2002] to find a local minimum of the cost function in

$$\min_{\{w_i, \mathbf{D}_i\}} \sum_{j=1}^{N_s} \left(E_j - \sum_{i}^{K} w_i \exp(-b_j \mathbf{u}_j^T \mathbf{D}_i \mathbf{u}_j) \right)^2 \tag{4.33}$$

which is unstable and the result is sensitive to the initial point. Note that for single shell data, the isotropic part of $\mathbf{D}_i$, i.e. $\frac{\text{Trace}(\mathbf{D}_i)}{3}$, and w_i are undistinguishable [Kreher

et al., 2005], because the above cost function for single shell data can be written as

$$\min_{\{w_i,\mathbf{D}_i\}} \sum_{j=1}^{N_s} \left(E_j - \sum_{i}^{K} w_i \exp\left(-b\frac{\text{Trace}(\mathbf{D}_i)}{3}\right) \exp\left(-b\mathbf{u}_j^T \left(\mathbf{D}_i - \frac{\text{Trace}(\mathbf{D}_i)}{3}\mathbf{I}\right)\mathbf{u}_j\right)\right)^2$$

Thus if we choose the new weights and tensors as $\{w_i \exp\left(-b\frac{\text{Trace}(\mathbf{D}_i)}{3}\right)\}$ and $\{\mathbf{D}_i - \frac{\text{Trace}(\mathbf{D}_i)}{3}\mathbf{I}\}$, then the cost function has the same minimal value as the cost function with the weights and tensors $\{w_i\}$, $\{\mathbf{D}_i\}$. Thus normally some constraints on tensors are considered in this model. For example, the two minimal eigenvalues λ_2 and λ_3 in $\mathbf{D}_i$ can be chosen as the same value. Tensors $\{\mathbf{D}_i\}$ can be chosen as one isotropic tensor and other anisotropic tensor with $\lambda_2 = \lambda_3 = 0$, which is the **ball and stick model** [Hosey et al., 2005]. CHARMED model [Assaf et al., 2004] proposed to consider $\{\mathbf{D}_i\}$ as a hindered diffusion part which is close to Gaussian diffusion and a restricted diffusion part which is non-Gaussian diffusion.

Mixture of tensor model is widely used to generate synthetic data for evaluation due to many quantities have closed forms in this model. Please see Appendix 5.8 for the closed forms, and see Appendix A for synthetic data generation using mixture of tensor model.

Summary of Advantages and Limitations:

- ✔ It is a natural extension of DTI model from Gaussian case to mixture of Gaussian case. It can detect the crossing fibers, compared to DTI.
- ✔ It provides closed forms for the EAP many useful features of the EAP, which makes it widely used in synthetic data generation. See Appendix 5.8 and Appendix A.
- ✘ The model selection for number of tensors is an open problem and seems to be set arbitrarily, although some papers [Behrens et al., 2007] proposed some possible ways.
- ✘ The minimization process depends on the initial point and it takes a long time.
- ✘ The radial decay of the mixture of tensor model is close to, but NOT, the Gaussian function. Consider the number of tensors is $K = 2$ in Fig. 4.8, along a given direction one component decays fast and the other one decays slowly. For large b value, the component with slow decay dominates the signal.

Generalized DTI (GDTI)

In Generalized Diffusion Tensor Imaging (GDTI) model [Liu et al., 2003, 2004], the signal is represented as

$$\begin{aligned} E(\mathbf{q}) &= \exp\left(\sum_{l=2}^{L} i^l D^{(l)}_{i_1 i_2 \ldots i_l} b^{(l)}_{i_1 i_2 \ldots i_l}\right) \\ &= \exp\left(\sum_{l=2}^{L} (2\pi i)^l \left(\Delta - \frac{l-1}{l+1}\delta\right) D^{(l)}_{i_1 i_2 \ldots i_l} q_{i_1} q_{i_2} \cdots q_{i_l}\right) \end{aligned} \tag{4.34}$$

where

$$b^{(l)}_{i_1 i_2 \ldots i_l} = \gamma^l G_{i_1} G_{i_2} \cdots G_{i_l} \delta^l \left(\Delta - \frac{l-1}{l+1} \delta \right)$$

is the generalized b value, and we use the Einstein summation convention, i.e. $D^{(l)}_{i_1 i_2 \ldots i_l} q_{i_1} q_{i_2} \cdots q_{i_l}$ means $\sum_{i_1=1}^{3} \sum_{i_2=1}^{3} \cdots \sum_{i_l=1}^{3} D^{(l)}_{i_1 i_2 \ldots i_l} q_{i_1} q_{i_2} \cdots q_{i_l}$. When $L = 2$, GDTI becomes the DTI model in Eq. (4.25). The generalized diffusion coefficients $D^{(l)}_{i_1 i_2 \ldots i_l}$ can be estimated by the least square fitting the samples of $\ln E(\mathbf{q})$. Since $E(\mathbf{q})$, $P(\mathbf{R})$ are related by Fourier transform and $P(\mathbf{R})$ is a PDF, $E(\mathbf{q})$ can be seen as the characteristic function of $P(\mathbf{R})$ [Liu et al., 2004]. The characteristic function can be represented by the cumulants $Q^{(l)}_{i_1 i_2 \ldots i_l}$:

$$E(\mathbf{q}) = \exp\left(\sum_{l=0}^{L} \frac{(-2\pi i)^l}{l!} Q^{(l)}_{i_1 i_2 \ldots i_l} q_{i_1} \cdots q_{i_l} \right) \tag{4.35}$$

Thus we have the relation between the cumulants and diffusion coefficients as

$$Q^{(l)}_{i_1 i_2 \ldots i_l} = (-1)^l l! D^{(l)}_{i_1 i_2 \ldots i_l} \left(\Delta - \frac{l-1}{l+1} \delta \right) \tag{4.36}$$

Then based on the property of probabilists' Hermite polynomial $He_n(x)$, we have the closed form for the EAP as the famous Gram-Charlier A series [Liu et al., 2004]

$$P(\mathbf{R}) = N(\mathbf{R}|Q^{(2)}_{i_1 i_2}) \left(1 + \sum_{l=3}^{L} \frac{Q^{(l)}_{i_1 i_2 \ldots i_l}}{l!} He^{(l)}_{i_1 i_2 \ldots i_l}(\mathbf{R}) \right) \tag{4.37}$$

where

$$N(\mathbf{R}|Q^{(l)}_{i_1 i_2}) = \frac{1}{2\pi Q^{(2)}_{i_1 i_2}} \exp\left(-\frac{\mathbf{R}^T \mathbf{R}}{2 Q^{(2)}_{i_1 i_2}} \right)$$

is the Gaussian distribution with zero mean and covariance $Q^{(2)}_{i_1 i_2}$, and $He^{(l)}_{i_1 i_2 \ldots i_l}(\mathbf{R})$ is the l-order probabilists' Hermite polynomial defined as

$$\begin{aligned} He^{(l)}_{i_1 i_2 \ldots i_l}(\mathbf{R}) &= (-1)^l \exp\left(-\frac{\mathbf{R}^T \mathbf{R}}{2} \right) \left(\frac{\partial}{\partial R_{i_1}} \frac{\partial}{\partial R_{i_2}} \cdots \frac{\partial}{\partial R_{i_l}} \right) \exp\left(\frac{\mathbf{R}^T \mathbf{R}}{2} \right) \\ &= He_{n_1}(R_x) He_{n_2}(R_x) He_{n_3}(R_z) \end{aligned} \tag{4.38}$$

where $n_j = \sum_{k=1}^{l} \delta(i_k = j)$.

Summary of Advantages and Limitations:

- ✔ GDTI is model-free. It uses a 3D basis to model the ADC such that it works for multiple shell sampling in Fig. 4.5(d).
- ✔ It provides closed form EAP estimation.
- ✔ It estimates the cumulants of EAP, which can be a set of useful scalar indices of EAP.
- ✘ It models the ADC using the polynomial basis, which is not orthogonal. It's well known that the basis matrix has large condition number when using high order polynomial basis to fit the function.

✘ Although theoretically the ADC can be modeled as infinite terms, in practice a truncated order L is needed in Eq. (4.34). However, it was proved in [Marcinkiewicz, 1939] that the Gaussian distribution is the only distribution which has a finite number of non-zero cumulants. Thus a truncation order L only results in a reasonable PDF if the EAP is Gaussian and $L = 2$ in this case. For other cases, the estimated EAP and cumulants are theoretically problematic. Moreover, estimation of the PDF from its cumulants is known to be very problematic [Blinnikov and Moessner, 1998; Ghosh et al., 2010].

High Order Tensor Model, ADC Based Model

High Order Tensor (HOT) model [Özarslan and Mareci, 2003; Özarslan et al., 2005] assumes the diffusion signal is represented as

$$E(\mathbf{q}) = \exp(-4\pi^2\tau q^2 D(\mathbf{u})), \tag{4.39}$$

Sometimes HOT is also called as GDTI in papers, however we call it HOT to distinguish HOT in [Özarslan and Mareci, 2003; Özarslan et al., 2005] and GDTI in [Liu et al., 2003, 2004]. In HOT model, the ADC is independent of radial part q, and is represented as the homogeneous polynomial of $\mathbf{u}$ with order L, i.e.

$$D(\mathbf{u}) = \sum_{i_1=1}^{3}\sum_{i_2=1}^{3}\cdots\sum_{i_L=1}^{3} D_{i_1 i_2 \ldots i_L} u_{i_1} u_{i_2} \cdots u_{i_L} = \sum_{n_1+n_2+n_3=L} D_{n_1 n_2 n_3} u_1^{n_1} u_2^{n_2} u_3^{n_3} \tag{4.40}$$

where $\mathbf{u} = (u_1, u_2, u_3)^T \in \mathbb{S}^2$, L is even because $D(\mathbf{u}) = -D(\mathbf{u})$ when L is odd and negative diffusion coefficients are non-physical. $\{u_1^{n_1} u_2^{n_2} u_3^{n_3}\}_{n_1+n_2+n_3=L}$ is the **homogeneous polynomial basis** restricted in $\mathbb{S}^2$, which is also called the **High Order Tensor (HOT) basis** in dMRI domain. See Section 2.4 for the homogeneous polynomial basis. When $L = 2$, HOT model is just the DTI model in Eq. (4.25), which means HOT is a kind of generalization of DTI model.

Note in HOT, the diffusion signal decays as a mono-exponential function, which is called as **mono-exponential decay assumption**. With this assumption, if we know $E(q_0\mathbf{u})$ with radius $q_0 > 0$ in one direction, we know all signal $E(q\mathbf{u})$ in the same direction, i.e.

$$E(q\mathbf{u}) = E(q_0\mathbf{u})^{q^2/q_0^2} \tag{4.41}$$

Compared to GDTI which is model-free method, HOT is model-based. Mono-exponential decay assumption is not satisfied in real decay of signal [Kuchel et al., 1997], but can be an good approximation of the signal, especially when the b value is around $1500 s/mm^2$ [Özarslan et al., 2006]. This assumption will be used later to derive the analytical formulae in Diffusion Orientation Transform (DOT) [Özarslan et al., 2006] and exact Q-Ball Imaging (QBI) [Canales-Rodrıguez et al., 2009; Aganj et al., 2010b; Tristán-Vega et al., 2010].

Historically people used both High Order Tensor basis [Özarslan and Mareci, 2003; Özarslan et al., 2005] and Spherical Harmonic (SH) basis [Frank, 2002; Descoteaux et al., 2006] to estimate ADC from measured signal. See Section 2.5.2 for the SH basis. The following theorem shows the equivalence of these two bases, whose proof mainly detailed in Chapter 2.

Theorem 4.1. *The following three bases form the same function space in* $\mathbb{S}^2$.

- *HOT basis, i.e. homogeneous polynomial basis restricted in* $\mathbb{S}^2$, $\left\{u_1^{n_1}u_2^{n_2}u_3^{n_3}\right\}_{n_1+n_2+n_3=L}$ *with even order* L.
- *Polynomial basis restricted in* $\mathbb{S}^2$ *with the order no more than* L, $\left\{u_1^{n_1}u_2^{n_2}u_3^{n_3}\right\}_{0\le l=n_1+n_2+n_3\le L}$ *with even* l *and* L.
- *SH basis with even order no more than* L, *i.e.* $\left\{Y_l^m(\mathbf{u})\right\}_{0\le l\le L}$, *where* Y_l^m *is the real symmetric spherical harmonic basis defined in Eq.* (2.28).

Proof. This proof is based on Proposition 2.1, Corollary 2.1 and Theorem 2.5. With even order L, Proposition 2.1 demonstrates the polynomial basis restricted in $\mathbb{S}^2$ with the order no more than L forms the same space with the homogeneous polynomial basis, i.e. HOT basis, with order L. With even order L, Corollary 2.1 and Theorem 2.5 demonstrate HOT basis with even L and SH basis with even order no more than L form the same space. □

HOT model uses single shell data in Fig. 4.5(c), which is a kind of sHARDI method. Normally the coefficients of HOT basis or SH basis are estimated via simple least square method. Since ADC $D(\mathbf{u})$ is non-negative definite, some methods estimate these coefficients by considering the non-negativity constraint [Barmpoutis et al., 2007, 2009; Ghosh et al., 2008a, 2009; Qi et al., 2010]. The estimated coefficients of ADC can be used in classification of the isotropic diffusion and anisotropic diffusion in single direction and multiple directions [Frank, 2002; Chen et al., 2004, 2005]. Some scalar indices like trace and variance can be defined by the ADC profile $D(\mathbf{u})$ [Özarslan et al., 2005].

Based on Theorem 4.1, the SH basis with $L \le 2$ forms the same space as the HOT basis with $L = 2$, which is the function space represented by quadratic form $\mathbf{u}^T\mathbf{D}\mathbf{u}$. This quadratic form has only one maximum which can not represent more than one fiber directions, which is the main limitation in DTI model. So in HARDI literatures, the maximal order of SH basis or the order of HOT basis must be higher than 4. Normally 4 or 6 is used in practice.

HOT model [Özarslan and Mareci, 2003; Özarslan et al., 2005] represents ADC using HOT basis. However ADC modeling like HOT method has its intrinsic and fatal limitation, i.e. both the maxima and the minima of ADC profile $D(\mathbf{u})$ are inconsistent with the fiber directions when $L > 2$ [von dem Hagen and Henkelman, 2002]. Only when $L = 2$, HOT model reduces to DTI model, then the minima of $D(\mathbf{u}) = \mathbf{u}^T\mathbf{D}\mathbf{u}$ are the maxima of EAP profile $P(R_0\mathbf{r})$ for any given R_0 which are consistent with fiber directions. For HOT model with order than $L > 2$, the maxima and minima of $D(\mathbf{u})$ correspond with fiber directions with a more complex way, which is still not known in dMRI if we do not perform Fourier transform to obtain EAP numerically or analytically [Özarslan et al., 2006], or perform spherical deconvolution by assuming a fiber model [Weldeselassie et al., 2010; Jiao et al., 2011]. Spherical Deconvolution method will be discussed in subsection 4.3.6. Fig. 4.8 demonstrates the ADC $D(\mathbf{u})$ for the synthetic data generated from mixture of tensor model with crossing angle of 90°. It shows that

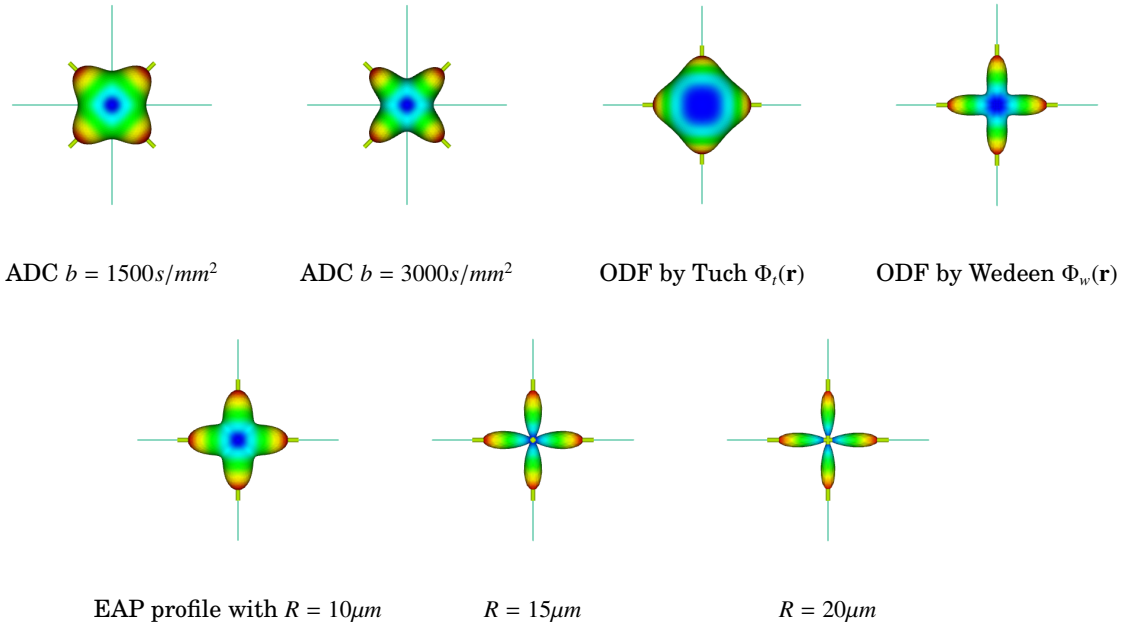


Figure 4.8: Fiber directions and ADC profiles with different b values, two kinds of ODFs, EAP profiles with different radius R. The data was generated from mixture of tensor model with two tensors which have the eigenvalues $[1.7, 0.3, 0.3] \times 10^{-3} mm^2/s$ and crossing angle of $90°$. We set $\tau = \frac{1}{4\pi^2}$ such that $b = q^2$. The long sticks with blue color along x-axis and y-axis are the fiber directions, i.e. the eigenvectors with the largest eigenvalues. The short sticks with yellow color are the detected maxima of the spherical functions. Note there is a coincidence that the minima of ADC agree with the fiber directions in the case of experiment. However, in general case the minima and maxima of ADC have a complex relation with fiber directions.

- the maxima of ADC do not agree with fiber directions.
- even in this simple mixture of tensor model, the ADC D is actually dependent on b value, i.e. the mono-exponential decay assumption is violated. For the data with different b values, the ADC is determined by $D = -\frac{1}{b} \ln E(\mathbf{q})$, which means D is dependent on b if $E(\mathbf{q}) = \sum_{i=1}^{K} \exp(-b\mathbf{u}^T \mathbf{D}_i \mathbf{u})$.
- there is a coincidence that the minima of ADC agree with the fiber directions in this specific case of mixture of tensor model with $90°$. However, in general case the minima and maxima of ADC have a complex relation with fiber directions.

Summary of Advantages and Limitations:

✔ ADC modeling methods like HOT [Özarslan and Mareci, 2003; Özarslan et al., 2005] generalizes the DTI model and avoids the Gaussian propagator assumption in DTI.

✔ The estimated ADC can be used to define some scalar indices [Frank, 2002; Chen et al., 2004, 2005; Özarslan et al., 2005] for classification of different diffusion in the surrounding tissue in voxels.

✔ HOT model normally needs single shell sampling with around 60 DWI samples. See Fig. 4.5(c).

✘ For the data from different b values, HOT model obtains different ADC, which means HOT can not be used in multiple shell sampling.

✘ The maxima and minima of ADC do not agree with fiber directions when HOT basis or SH basis is used with order $L \geq 4$. When $L = 2$ the minima of ADC can be used as the fiber directions in DTI model.

✘ Compared to GDTI which is model-free, HOT assumes mono-exponential decay of $E(\mathbf{q})$ which is not satisfied in real diffusion. Even in synthetic data from mixture of tensor model, the mono-exponential decay is violated.

4.3.2 Diffusion Spectrum Imaging (DSI)

Because EAP $P(\mathbf{R})$ is related with measured diffusion signal $E(\mathbf{q})$ in Fourier transform in Eq. (4.24) when narrow pulse assumption $\delta \ll \Delta$ is satisfied, a straightforward idea is to estimate $P(\mathbf{R})$ using fast Fourier transform from exhaustive signal samples [Callaghan, 1991; Tuch, 2002; Wedeen et al., 2000, 2005]. This technique is called as Diffusion Spectrum Imaging (DSI).

In practice, only limited number of samples are obtained and $\delta \ll \Delta$ is not always satisfied. [Wedeen et al., 2005] used 515 DWI images in a Cartesian sampling lattice in $\mathbf{q}$-space and the signal in $\mathbf{q}$-space was premultiplied by a Hanning window to obtain smooth attenuation of the signal at high $\mathbf{q}$ values. See Fig. 4.5(b) for the sketch map of the Cartesian sampling. Please note that in practice the Cartesian samples inside a given ball, not a given cube, are used. Then the samples inside the cube and outside the ball can be obtained by extrapolation. Thus interpolation and extrapolation are normally performed on given signal samples $\{E(\mathbf{q}_i)\}$ before numerical Fourier transform. In [Wedeen et al., 2005], δ is close to Δ which violates the narrow pulse assumption. But the results in [Wedeen et al., 2005] are still exciting and show clearly some crossing fibers, which means that even though the narrow pulse assumption is violated, the Fourier transform can still obtain meaningful EAPs.

[Wedeen et al., 2000, 2005] visualized the **EAP profile**, or called **iso-surface of EAP**, which is the EAP with given radius R_0, i.e.

$$P(R_0\mathbf{r}) = P(R\mathbf{r})|_{R=R_0} \tag{4.42}$$

The maxima of EAP profile were used to describe fiber directions later in many HARDI works [Özarslan et al., 2006, 2009; Assemlal et al., 2009a; Descoteaux et al., 2010]. See Fig. 4.8 for the EAP profile with different radius R. The larger the radius R, the sharper the EAP profile is. However, EAP profile with large R has more estimation error. Thus normally $R = 15\mu m$ is used in EAP profile to detect the fiber directions [Özarslan et al., 2006; Descoteaux et al., 2010].

[Wedeen et al., 2000, 2005] also proposed another important feature of EAP, i.e. the Orientation Distribution Function (ODF), defined as

$$\Phi_w(\mathbf{r}) \overset{\text{def}}{=} \int_0^\infty P(\mathbf{R})R^2 \mathrm{d}R \tag{4.43}$$

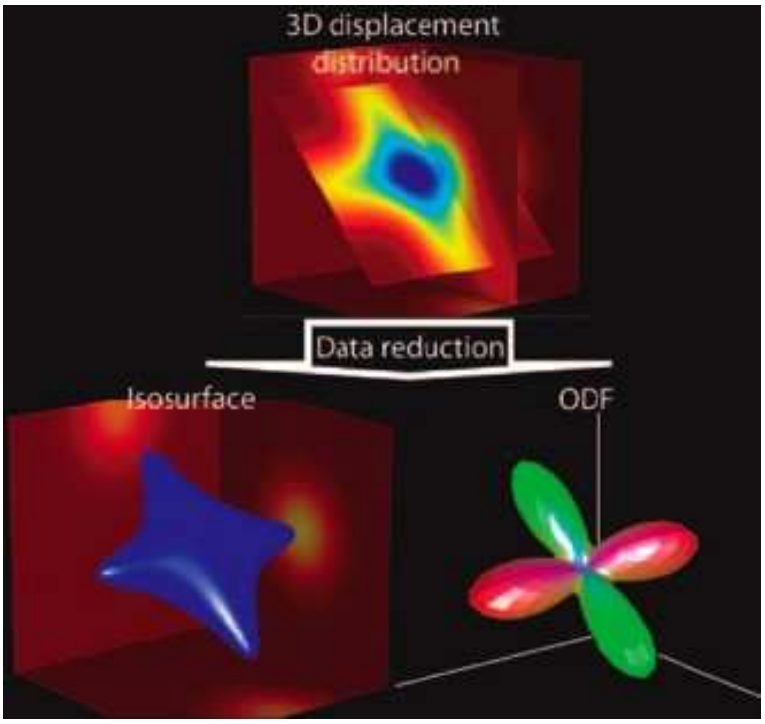


Figure 4.9: EAP in 3D **R**-space, and its two features, i.e. EAP profile (or called iso-surface of EAP) and ODF. The figure is from [Hagmann et al., 2006].

It is called as **ODF by Wedeen**, denoted by $\Phi_w(\mathbf{r})$. $\Phi_w(\mathbf{r})$ is the marginal distribution of EAP $P(\mathbf{R})$, so the integration of $\Phi_w(\mathbf{r})$ over $\mathbb{S}^2$ is naturally 1, which means it does not need artificial normalization factor Z. [Wedeen et al., 2000, 2005] proposed to first estimate EAP via numerical Fourier transform, then estimate the ODF in Eq. (4.43) by numerical integration. Historically there are several kinds of ODFs which can be seen in the following of this section. Like the EAP profile, the maxima of ODFs are also normally assumed to be the directions of underlying fibers. Please see Fig. 4.9 and Fig. 4.8 for EAP in 3D space and its two features, i.e. EAP profile and ODF.

Summary of Advantages and Limitations:

✔ Compared to DTI, DSI is a model-free method which avoids Gaussian assumption.

✘ DSI needs a dense Cartesian sampling in **q**-space with many DWI images and a very large range of b value which takes a long time and makes it impractical for normal scanner. For example, [Wedeen et al., 2005] used more than 500 DWI images and b value up to $17000 s/mm^2$.

✘ DSI uses numerical Fourier transform and an artificial interpolation and extrapolation step, which makes the final results have much numerical error.

4.3.3 Hybrid Diffusion Imaging (HYDI)

Hybrid Diffusion Imaging (HYDI) proposes to measure data in **q**-space with multiple shell sampling [Wu and Alexander, 2007; Wu et al., 2008]. See Fig. 4.5(d) for the sketch map of such sparse sampling. The HYDI data in the shell with low b values can be modeled by DTI. The HYDI data in the shell with high b values can be modeled by Q-Ball Imaging and other sHARDI methods. The whole HYDI data set can be used in DSI after re-griding data from multiple shell to Cartesian lattice.

HYDI in [Wu and Alexander, 2007; Wu et al., 2008] proposed two useful scalar features of EAP, i.e. the Return-To-Origin probability (RTO) and the Mean Squared Displacement (MSD). RTO denoted by Po is the EAP value when $R = 0$, i.e.

$$Po = P(\mathbf{R})|_{R=0} = P(0) = \int_{\mathbb{R}^3} E(\mathbf{q})\mathrm{d}\mathbf{q} \tag{4.44}$$

RTO is the probability of water molecules that minimally diffuse within the diffusion time Δ. RTO map can be used in tissue segmentation and some other applications [Wu and Alexander, 2007]. MSD is the variance of the EAP, i.e.

$$MSD = \int_{\mathbb{R}^3} P(\mathbf{R})\mathbf{R}^T\mathbf{R}\mathrm{d}\mathbf{R} \tag{4.45}$$

[Wu and Alexander, 2007; Wu et al., 2008] estimate these two scalar indices using numerical integration. [Wu et al., 2008] also demonstrated that the ODF by Tuch $\Phi_t(\mathbf{r})$ in Eq. (4.47) is proportional to the integration of $E(\mathbf{q})$ in the orthogonal plane

$$\Pi_{\mathbf{r}} = \{q\mathbf{u} : \ \mathbf{u}^T\mathbf{r} = 0\} \tag{4.46}$$

It is an important relation between the ODF by Tuch in **R**-space and the signal $E(\mathbf{q})$ in **q**-space, and it is used in exact QBI to estimate ODFs analytically in [Canales-Rodrıguez et al., 2009; Aganj et al., 2010b; Tristán-Vega et al., 2009; Tristán-Vega et al., 2010]. See the following subsection 4.3.4 for more details on QBI and ODFs.

Summary of Advantages and Limitations:

- ✔ The first contribution of HYDI is the idea of multiple shell sampling in **q**-space.
- ✔ HYDI proposed some useful scalar indices.
- ✔ HYDI proposed the plane integration of $E(\mathbf{q})$ to estimate the ODF by Tuch. However a numerical way was used to estimate the plane integration in HYDI.
- ✘ HYDI proposed to use QBI for the data with high b values. However, for the data in different shells QBI obtains different results, which makes the results inconsistent.
- ✘ HYDI tried to use different models for the data with different b values, i.e. DTI for data with low b values, QBI for data with high b values and DSI for whole data set. However, different models/methods may obtain different results which are hard to be explained.

- ✘ DSI is used in the whole data set after a re-griding of data from multiple shell sampling to Cartesian sampling, which brings much numerical error.
- ✘ The estimation of scalar indices also suffers from the numerical error introduced by *ad hoc* interpolation.

4.3.4 Q-Ball Imaging (QBI)

Q-Ball Imaging (QBI) is the most widely used HARDI method. DSI needs a dense Cartesian sampling with a large range of b value, which makes it impractical [Wedeen et al., 2000, 2005]. QBI was proposed to estimate the several kinds of ODFs, not EAP, from single shell sampling demonstrated in Fig. 4.5(c), rather than Cartesian sampling inside a given ball used in DSI in 4.5(b).

Original Q-Ball Imaging

QBI was first proposed by Dr. Tuch in [Tuch, 2002, 2004] in a numerical way and then was improved by an analytical way based on Spherical Harmonic basis in [Anderson, 2005; Hess et al., 2006; Descoteaux et al., 2007].

Instead of estimation of EAP, Dr. Tuch proposed to estimate a kind of ODF defined as

$$\Phi_t(\mathbf{r}) \overset{\text{def}}{=} \frac{1}{Z}\int_0^\infty P(R\mathbf{r})\mathrm{d}R \tag{4.47}$$

where Z is the normalization factor which makes $\int_{\mathbb{S}^2}\Phi_t(\mathbf{r})\mathrm{d}\mathbf{r} = 1$. This ODF is called as **ODF by Tuch** and denoted by $\Phi_t(\mathbf{r})$. Note $\Phi_t(\mathbf{r})$ is different from the ODF $\Phi_w(\mathbf{r})$ defined in Eq. (4.43). $\Phi_w(\mathbf{r})$ is the marginal PDF of EAP which does not need artificial normalization factor, however, $\Phi_t(\mathbf{r})$ needs the normalization factor Z to make it as a PDF.

In DSI, $\Phi_w(\mathbf{r})$ is estimated from pre-estimated EAP via a numerical integration. Dr. Tuch proposed to estimate $\Phi_t(\mathbf{r})$ directly from samples of $E(\mathbf{q})$ in single shell data based on Funk-Radon Transform (FRT). See Fig. 4.5(c) for the sketch map of single shell sampling. For single shell data with $b = 4\pi^2\tau q_0^2$, the FRT of $E(\mathbf{q})$ [Tuch, 2004] in direction $\mathbf{r}$ is the **circle integration** in the orthogonal plane, i.e.

$$\mathrm{FRT}\{E(q_0\mathbf{u})\}(\mathbf{r}) = \int_{\Pi_\mathbf{r}} E(q\mathbf{u})\delta(q-q_0)q\mathrm{d}q\mathrm{d}\mathbf{u} = q_0\int_{\mathbf{u}\in\mathbb{S}^2} E(q_0\mathbf{u})\delta(\mathbf{u}^T\mathbf{r})\mathrm{d}\mathbf{u} \tag{4.48}$$

where $\Pi_\mathbf{r}$ is defined in Eq. (4.46).

$$\begin{aligned}
\mathrm{FRT}\{E(q_0\mathbf{u})\}(\mathbf{r}') &= q_0\int_{\mathbf{u}\in\mathbb{S}^2} E(q_0\mathbf{u})\delta(\mathbf{u}^T\mathbf{r}')\mathrm{d}\mathbf{u} \\
&= q_0\int_{\mathbf{u}\in\mathbb{S}^2}\left(\int_{\mathbb{R}^3} P(\mathbf{R})\exp(-2i\pi Rq_0\mathbf{u}^T\mathbf{r})\mathrm{d}\mathbf{R}\right)\delta(\mathbf{u}^T\mathbf{r}')\mathrm{d}\mathbf{u} \\
&= q_0\int_{\mathbb{R}^3} P(\mathbf{R})\underbrace{\left(\int_{\mathbb{S}^2}\exp(-2i\pi Rq_0\mathbf{u}^T\mathbf{r})\delta(\mathbf{u}^T\mathbf{r}')\mathrm{d}\mathbf{u}\right)}_{\mathcal{K}(\mathbf{r},\mathbf{r}')}\mathrm{d}\mathbf{R}
\end{aligned} \tag{4.49}$$

Without loss of generality, we assume $\mathbf{r}' = (0,0,1)^T$ is the z-axis, and $\mathbf{u} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)^T$, $\mathbf{R} = R(\sin\theta_{\mathbf{r}}\cos\phi_{\mathbf{r}}, \sin\theta_{\mathbf{r}}\sin\phi_{\mathbf{r}}, \cos\theta_{\mathbf{r}})^T = (R_x, R_y, R_z)^T$ in Cartesian coordinate. Then the kernel $\mathcal{K}(\mathbf{r}, \mathbf{r}')$ can be solved as

$$\begin{aligned}
\mathcal{K}(\mathbf{r}, \mathbf{r}') &= \int_0^{\pi}\int_0^{2\pi} \delta(\cos\theta)\exp\left(-2i\pi q_0\left(R_x\sin\theta\cos\phi + R_y\sin\theta\sin\phi + R_z\cos\theta\right)\right)\sin\theta \mathrm{d}\phi\mathrm{d}\theta \\
&= \int_{-1}^{1}\delta(t)\exp\left(-2\pi i q_0 R_z t\right)\int_0^{2\pi}\exp\left(-2\pi i q_0\left(R_x\sqrt{1-t^2}\cos\phi + R_y\sqrt{1-t^2}\sin\phi\right)\right)\mathrm{d}\phi\mathrm{d}t \qquad (4.50) \\
&= \int_0^{2\pi}\exp\left(-2\pi i q_0(R_x\cos\phi + R_y\sin\phi)\right)\mathrm{d}\phi \qquad (4.51) \\
&= \int_0^{2\pi}\exp(2\pi i q_0 R\sin\theta_{\mathbf{r}}\sin\phi)\mathrm{d}\phi \qquad (4.52) \\
&= 2\pi J_0(2\pi q_0 R\sin\theta_{\mathbf{r}}) = 2\pi J_0\left(2\pi q_0 R\sqrt{1-(\mathbf{r}^T\mathbf{r}')^2}\right) \qquad (4.53)
\end{aligned}$$

After setting $t = \cos\theta$, we have Eq. (4.50). Considering the property of delta function, i.e. $\int_{-1}^{1}\delta(t)f(t)\mathrm{d}t = f(0)$, we have Eq. (4.51). Using the summation property of cosine function, we obtain Eq. (4.52). Considering the Bessel integral in Eq. (2.43), we have the final kernel function in Eq. (4.53), where $J_0(\cdot)$ is the Bessel function of the first kind with order 0. See Section 2.5.3 for more information on Bessel function. Put Eq. (4.53) into Eq. (4.49) and use the cylinder coordinate $\mathbf{R} = (r_{\mathbf{R}}, \phi_{\mathbf{R}}, z_{\mathbf{R}})^T$ where z-axis is along $\mathbf{r}'$, then the FRT of $E(q_0\mathbf{u})$ is

$$\begin{aligned}
\mathrm{FRT}\{E(q_0\mathbf{u})\}(\mathbf{r}') &= 2\pi q_0\int_{\mathbb{R}^3} P(\mathbf{R})J_0\left(2\pi q_0 R\sqrt{1-(\mathbf{r}^T\mathbf{r}')^2}\right)\mathrm{d}\mathbf{R} \\
&= 2\pi q_0\int_{-\infty}^{\infty}\int_0^{2\pi}\int_0^{\infty} P(r_{\mathbf{R}}, \phi_{\mathbf{R}}, z_{\mathbf{R}})J_0(2\pi q_0 r_{\mathbf{R}})r_{\mathbf{R}}\mathrm{d}r_{\mathbf{R}}\mathrm{d}\theta_{\mathbf{R}}\mathrm{d}z_{\mathbf{R}} \qquad (4.54)
\end{aligned}$$

However based on the definition of the ODF in Eq. (4.47), the ODF can be written as

$$\begin{aligned}
\Phi_t(\mathbf{r}') &= \frac{1}{2Z}\int_{-\infty}^{\infty} P(r_{\mathbf{R}}, \phi_{\mathbf{R}}, z_{\mathbf{R}})\mathrm{d}z_{\mathbf{R}} \\
&= \frac{1}{2Z}\int_{-\infty}^{\infty}\int_0^{2\pi}\int_0^{\infty} P(r_{\mathbf{R}}, \phi_{\mathbf{R}}, z_{\mathbf{R}})\delta(r_{\mathbf{R}})\delta(\phi_{\mathbf{R}})r_{\mathbf{R}}\mathrm{d}r_{\mathbf{R}}\mathrm{d}\theta_{\mathbf{R}}\mathrm{d}z_{\mathbf{R}} \qquad (4.55)
\end{aligned}$$

Comparing Eq. (4.55) and Eq. (4.54), the estimation of $\Phi_t(\mathbf{r})$ through FRT is inspirit to approximate the delta function using Bessel function $\frac{a}{2}J_0(ax)$. As q_0 increases, $2\pi q_0 J_0(2\pi q_0 r_{\mathbf{R}})$ will be more close to delta function which will increase the accuracy of this approximation, because the lobes of J_0 become more concentrated around origin point. However, the signal has smaller values for larger q_0 which results in low Signal-to-Noise Ratio (SNR). Thus there is a trade-off for q_0 between approximation accuracy and SNR. Normally QBI works suggest the data with b values around $3000 s/mm^2$ [Tuch, 2004; Descoteaux et al., 2007].

Dr. Tuch proposed to estimate the circle integration in FRT using numerical integration [Tuch, 2004]. The points in the circle in the orthogonal plane need to be interpolated by sampling points in its neighborhood through spherical radial basis function. Then the summation of these points in the circle was used to approximate the circle integration. The numerical QBI was later replaced by analytical QBI based on the representation of $E(\mathbf{q})$ using SH basis. There were several groups which independently proposed the same analytical QBI using SHs. [Anderson, 2005] obtained

the analytical solution by considering the rotation property of SHs. [Hess et al., 2006] used addition theorem and rotation property of SHs, and considered Tikhonov regularization in least square estimation. [Descoteaux et al., 2007] applied 3D Funk-Hecke theorem to find the analytical solution of FRT. [Descoteaux et al., 2007] also proposed a simple and useful Laplace-Beltrami regularization scheme in least square estimation, which was shown to outperform the simple Tikhonov regularization and later became very popular for general least square estimation of spherical functions in HARDI domain. See the addition theorem and Funk-Hecke theorem, and their relation with plane wave expansion theorem in Section 2.5.2 and Appendix 5.7.

Analytical QBI represents the signal $E(\mathbf{q})$ as SH basis, i.e.

$$E(q_0\mathbf{u}) = \sum_{l=0}^{L}\sum_{m=-l}^{l} c_{lm} Y_l^m(\mathbf{u}) \tag{4.56}$$

where $Y_l^m(\mathbf{u})$ is the symmetric real spherical harmonic with order l and degree m. See Section 2.5.2 for more information on SHs. The coefficients $\{c_{lm}\}$ are normally estimated from signal samples by minimizing a least square cost function with Laplace-Beltrami regularization in Eq. (4.59) [Descoteaux et al., 2007].

$$\|B_M\boldsymbol{c} - \mathbf{E}\|^2 + \boldsymbol{c}^T\Lambda\boldsymbol{c} \tag{4.57}$$

where $\boldsymbol{c} = (c_{00}, \cdots, c_{LL})^T$ is the coefficient vector with $(L+1)(L+2)/2$ elements, $E = (E_1, \cdots, E_{N_s})^T$ is the signal vector with N_s samples, B_M is the $N_s \times (L+1)(L+2)/2$ basis matrix generated by SHs, and Λ is the diagonal matrix with elements $\Lambda_{lm} = \lambda l^2(l+1)^2$. $\boldsymbol{c}^T\Lambda\boldsymbol{c}$ is the Laplace-Beltrami regularization term, because based on Eq. (2.32), we have

$$\begin{aligned}\lambda\|\Delta_b\mathbf{E}(\mathbf{q})\|^2 &= \lambda\int_{\mathbb{S}^2}\sum_{l=0}^{L}\sum_{m=-l}^{l}\left(c_{lm}\Delta_b Y_l^m(\mathbf{u})\right)\sum_{l'=0}^{L}\sum_{m'=-l'}^{l'}\left(c_{l'm'}\Delta_b Y_{l'}^{m'}(\mathbf{u})\right)\mathrm{d}\mathbf{u}\\ &= \lambda\sum_{l=0}^{L}\sum_{m=-l}^{l} l^2(l+1)^2 c_{lm}^2 = \boldsymbol{c}^T\Lambda\boldsymbol{c}\end{aligned} \tag{4.58}$$

where Δ_b is the Laplace-Beltrami operator defined in Eq. (2.19). The least square problem has the closed form solution as

$$\boldsymbol{c} = (B_M^T B_M + \Lambda)^{-1} B_M^T \mathbf{E} \tag{4.59}$$

Based on Funk-Hecke theorem 2.7 or Corollary 2.4, the estimated ODF from FRT of $E(\mathbf{q})$ can be analytically obtained from the estimated $\{c_{lm}\}$ in Eq. (4.60).

$$\widetilde{\Phi}_t(\mathbf{r}) = \frac{1}{Z}\mathrm{FRT}\{E(q_0\mathbf{u})\}(\mathbf{r}) = \frac{1}{Z}\sum_{l=0}^{L}\sum_{m=-l}^{l} 2\pi P_l(0) c_{lm} Y_l^m(\mathbf{u}) \tag{4.60}$$

where $P_l(0)$ is the Legendre polynomial of order l evaluated at 0.

Note in practice the ODF by Tuch $\Phi_t(\mathbf{r})$ in Eq. (4.47) is much smooth. The peaks of the ODF are only a little higher than the baseline values. Dr. Tuch proposed a min-max normalization method for visualization of $\Phi_t(\mathbf{r})$ to enhance the peaks of ODFs. Min-max normalization is a linear scaling to transform the ODF values into $[0, 1]$, and

it has been a commonly used way to visualize $\Phi_t(\mathbf{r})$ in literature. However, the min-max normalization also enhances the peaks of the ODFs in the area with isotropic diffusion. Throughout this thesis in order to better compare ODFs and EAPs, we visualize $\Phi_t(\mathbf{r})$ using its real values without min-max normalization, except for the visualization of the phantom data in Appendix B which is known to be more isotropic than clinical data.

Dr. Tuch also proposed a useful scalar index, named Generalized Fractional Anisotropy (FA), to describe the anisotropy of the ODFs, which can be seen as a generalization of previous FA in DTI model.

$$\mathrm{GFA}\{\Phi_t(\mathbf{r})\} \stackrel{\mathrm{def}}{=} \sqrt{\frac{N\sum_{i=1}^{N}(\Phi_t(\mathbf{r}_i)-\langle\Phi_t(\mathbf{r})\rangle)^2}{(n-1)\sum_{i=1}^{N}\Phi_t(\mathbf{r}_i)^2}} \tag{4.61}$$

where $\langle\Phi_t(\mathbf{r})\rangle$ is the mean of $\Phi_t(\mathbf{r})$. If the ODF is represented by SH basis with coefficients $\{c_{lm}\}$, the GFA can be represented by

$$\mathrm{GFA}\{\Phi_t(\mathbf{r})\} = \frac{\|\Phi_t(\mathbf{r})-\langle\Phi_t(\mathbf{r})\rangle\|}{\|\Phi_t(\mathbf{r})\|} = \sqrt{1-\frac{c_{00}^2}{\sum_{l=0}^{L}\sum_{m=-l}^{l}c_{lm}^2}} \tag{4.62}$$

because of the orthogonality of SHs.

Summary of Advantages and Limitations:

- ✔ Analytical QBI is now the most widely used HARDI method, because it only needs single shell data to estimate the ODF whose maxima demonstrate the fiber directions, and it is very easy to be implemented.
- ✔ Compared to ADC based modeling like HOT method, the maxima of ODFs agree with the fiber directions.
- ✔ GFA was proposed in QBI, and it is now a standard index to describe the anisotropy of spherical function in HARDI.
- ✘ FRT has the intrinsic blurring effect due to the Bessel function as we have discussed. Thus the estimated ODF from FRT has intrinsic modeling error.
- ✘ The estimated ODF by Tuch from FRT is much smooth so that it needs artificial sharping techniques to obtain better results. That is because of two reasons.
  (a) The ODF by Tuch $\Phi_t(\mathbf{r})$ defined in Eq. (4.47) is not a good choice, because the ODF by Wedeen $\Phi_w(\mathbf{r})$ defined in Eq. (4.43) is theoretically more sharper than $\Phi_t(\mathbf{r})$.
  (b) The estimation via FRT has intrinsic blurring effect which obtains smooth ODFs.

  The sharping techniques include the min-max normalization for visualization, and the spherical deconvolution technique to estimate the so-called fiber Orientation Distribution Function (fODF). See Section 4.3.6 for spherical deconvolution and the difference between fODF and diffusion ODF.

✘ Many papers in HARDI refer QBI as a model-free method because it represents spherical signal using SHs which forms an orthonormal basis in $\mathbb{S}^2$. However, QBI actually assumes the radial part of $E(\mathbf{q})$ as a delta function which is unrealistic. The burring effect from FRT is the direct consequence from this assumption of radial decay. So we consider QBI as a model-based method with strong assumption of radial decay of signal.

✘ QBI can not be used in multiple shell data, because the data from different b values obtain different ODFs from FRT. Note [Khachaturian et al., 2007] proposed a method to generalize QBI from single shell data to the data sampled in two shells with both low b value and high b value. Two coefficient sets are estimated from the two single shell data set respectively. Then in each dimension of the coefficient vector an *ad hoc* process is used to choose the maximal coefficient. The fused coefficient vector is set as the final result. The ad hoc method to combine data from different b values seems inconvincible.

Exact Q-Ball Imaging

The ODF by Tuch is approximated by circle integration in original QBI, which has intrinsic limitations as we have discussed above. Exact QBI was proposed by several groups independently [Wu et al., 2008; Canales-Rodrıguez et al., 2009; Aganj et al., 2010b; Tristán-Vega et al., 2009; Tristán-Vega et al., 2010] to estimate ODFs through a plane integration, not a circle integration.

Based on the famous **projection-slice theorem** in Fourier transform[4], the projection of $P(R\mathbf{r})$ along direction $\mathbf{r}$, i.e. the radial integration, equals to the integration of $E(\mathbf{q})$ in the orthogonal plane $\Pi_{\mathbf{r}}$. This can be seen easily from the following proposition.

Proposition 4.1. *Let $f(\mathbf{q})$ be a symmetric function defined in $\mathbb{R}^3$ and $\mathcal{F}_{3D}\{f(\mathbf{q})\}(\mathbf{R}) = \hat{f}(\mathbf{R})$, then*

$$\int_0^\infty \hat{f}(R\mathbf{r})\mathrm{d}R = \frac{1}{2}\int_{\Pi_{\mathbf{r}}} f(\mathbf{q})\mathrm{d}\mathbf{q} = \frac{1}{2}\int_{\mathbb{R}^3} f(\mathbf{q})\delta(\mathbf{q}^T\mathbf{r})\mathrm{d}\mathbf{q} \tag{4.63}$$

where $\Pi_{\mathbf{r}}$ is the plane orthogonal to $\mathbf{r}$ defined in Eq. (4.46).

Proof. Since $f(\mathbf{q}) = f(-\mathbf{q})$, we have $\hat{f}(\mathbf{R}) = \hat{f}(-\mathbf{R})$ based on the definition of Fourier transform. Then

$$\begin{aligned}
\int_0^\infty \hat{f}(R\mathbf{r})\mathrm{d}R &= \frac{1}{2}\int_{-\infty}^\infty \hat{f}(R\mathbf{r})\mathrm{d}R = \frac{1}{2}\int_{-\infty}^\infty \left(\int_{\mathbb{R}^3} f(\mathbf{q})\exp(-2\pi i\mathbf{q}^T\mathbf{R})\mathrm{d}\mathbf{q}\right)\mathrm{d}R \\
&= \frac{1}{2}\int_{\mathbb{R}^3} f(\mathbf{q})\left(\int_{-\infty}^\infty \exp(-2\pi iR\mathbf{q}^T\mathbf{r})\mathrm{d}R\right)\mathrm{d}\mathbf{q} \\
&= \frac{1}{2}\int_{\mathbb{R}^3} f(\mathbf{q})\delta(\mathbf{q}^T\mathbf{r})\mathrm{d}\mathbf{q} = \frac{1}{2}\int_{\Pi_{\mathbf{r}}} f(\mathbf{q})\mathrm{d}\mathbf{q}
\end{aligned}$$

□

[4] http://en.wikipedia.org/wiki/Projection-slice_theorem

Thus we have the following corollary which is a straightforward result of the above proposition and has been used to estimate both ODF by Tuch and ODF by Wedeen [Wu et al., 2008; Canales-Rodrıguez et al., 2009; Aganj et al., 2010b; Tristán-Vega et al., 2009; Tristán-Vega et al., 2010; Cheng et al., 2010a].

Corollary 4.1. *The ODF by Tuch* $\Phi_t(\mathbf{r})$ *and ODF by Wedeen* $\Phi_w(\mathbf{r})$ *can be written as the plane integration of in the plane* $\Pi_\mathbf{r}$ *which is orthogonal to* $\mathbf{r}$*, i.e.*

$$\Phi_t(\mathbf{r}) = \frac{1}{Z}\int_{\Pi_\mathbf{r}} E(\mathbf{q})\mathrm{d}\mathbf{q}, \quad \Phi_w(\mathbf{r}) = \frac{1}{4\pi} - \frac{1}{8\pi^2}\int_{\Pi_\mathbf{r}} \frac{1}{q}\Delta_b E(\mathbf{q})\mathrm{d}\mathbf{q} \tag{4.64}$$

Proof. Considering Proposition 4.1, $\mathcal{F}\{E(\mathbf{q})\}(\mathbf{R}) = P(\mathbf{R})$ and $\mathcal{F}\{\Delta E(\mathbf{q})\} = -4\pi^2R^2P(\mathbf{R})$, where Δ is the Laplace operator, we have

$$\Phi_t(\mathbf{r}) = \frac{1}{Z}\int_{\Pi_\mathbf{r}} E(\mathbf{q})\mathrm{d}\mathbf{q}, \quad \Phi_w(\mathbf{r}) = -\frac{1}{8\pi^2}\int_{\Pi_\mathbf{r}} \Delta E(\mathbf{q})\mathrm{d}\mathbf{q} \tag{4.65}$$

The Laplace operator in 3D can be separated as

$$\Delta = \frac{1}{q^2}\frac{\partial}{\partial q}\Big(q^2\frac{\partial}{\partial q}\Big) + \frac{1}{q^2}\Delta_b, \quad \Delta_b = \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2} + \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\Big(\sin\theta\frac{\partial}{\partial\theta}\Big)$$

where Δ_b is the Laplace-Beltrami operator in $\mathbb{S}^2$. Then

$$\Phi_w(\mathbf{r}) = \underbrace{-\frac{1}{8\pi^2}\int_{\Pi_\mathbf{r}} \frac{1}{q^2}\frac{\partial}{\partial q}\Big(q^2\frac{\partial E(\mathbf{q})}{\partial q}\Big)\mathrm{d}\mathbf{q}}_{I} - \frac{1}{8\pi^2}\int_{\Pi_\mathbf{r}} \frac{1}{q}\Delta_b E(\mathbf{q})\mathrm{d}\mathbf{q} \tag{4.66}$$

$$I = -\frac{1}{8\pi^2}\int_0^{2\pi}\int_0^{\infty} \frac{1}{q}\frac{\partial}{\partial q}\Big(q^2\frac{\partial E(\mathbf{q})}{\partial q}\Big)\mathrm{d}q\mathrm{d}\phi = -\frac{1}{8\pi^2}\int_0^{2\pi}\Big(E(q\mathbf{u}) + q\frac{\partial E(q\mathbf{u})}{\partial q}\Big)\Big|_0^{\infty}\mathrm{d}\phi = \frac{1}{4\pi}$$

Note in above formula, we use the prior of $E(\mathbf{q})$ that for given direction $\mathbf{u}$, $\mathbf{E}(0\mathbf{u}) = 1$, $\lim_{q\to\infty} E(q\mathbf{u}) = 0$ and the assumption of $\lim_{q\to\infty} q\frac{\partial E(q\mathbf{u})}{\partial q} = 0$. □

Note the assumption of $\lim_{q\to\infty} q\frac{\partial E(q\mathbf{u})}{\partial q} = 0$ used in the above corollary is satisfied in DTI model, mixture of tensor model and SPFI and SHORE discussed later.

[Wu et al., 2008] noticed the projection-slice theorem and used it to estimate $\Phi_t(\mathbf{r})$ in a numerical way based on interpolation of $E(\mathbf{q})$ from multiple shell data. [Canales-Rodrıguez et al., 2009] estimated $\Phi_t(\mathbf{r})$ based on the plane integration corollary and mono-exponential decay assumption which has been introduced in subsection 4.3.1. Under the mono-exponential decay assumption in Eq. (4.39), the $\Phi_t(\mathbf{r})$ in Eq. (4.64) can be rewritten as

$$\begin{aligned}\Phi_t(\mathbf{r}) &= \frac{1}{Z}\int_{\mathbb{S}^2}\Big(\int_0^{\infty}\exp(-4\pi^2\tau q^2 D(\mathbf{u}))q\mathrm{d}q\Big)\delta(\mathbf{u}^T\mathbf{r})\mathrm{d}\mathbf{u} \\ &= \frac{1}{8\pi^2\tau Z}\int_{\mathbb{S}^2}\frac{1}{D(\mathbf{u})}\delta(\mathbf{u}^T\mathbf{r})\mathrm{d}\mathbf{u} = -\frac{q_0}{2Z}\int_{\mathbb{S}^2}\frac{1}{\ln E(q_0\mathbf{u})}\delta(\mathbf{u}^T\mathbf{r})\mathrm{d}\mathbf{u}\end{aligned} \tag{4.67}$$

Thus $\Phi_t(\mathbf{r})$ can be estimated from FRT of a function $-\frac{1}{\ln E(q_0\mathbf{u})}$ of single shell data $E(q_0\mathbf{u})$ with $q = q_0$.

$\Phi_w(\mathbf{r})$ has been proposed in DSI by Wedeen [Wedeen et al., 2005], where it was estimated from numerical radial integration of a pre-estimated EAP from DSI. [Aganj et al., 2009, 2010b] proved the above corollary for the estimation of $\Phi_w(\mathbf{r})$. [Aganj et al., 2009, 2010b] also proved that if $E(\mathbf{q})$ follows the mon-exponential decay assumption, we have

$$\Phi_w(\mathbf{r}) = \frac{1}{4\pi} + \frac{1}{16\pi^2}\int_{\mathbb{S}^2} \Delta_b \ln\left(-\ln E(q_0\mathbf{u})\right)\delta(\mathbf{u}^T\mathbf{r})\mathrm{d}\mathbf{u} \tag{4.68}$$

By representing $\ln\left(-\ln E(q_0\mathbf{u})\right) = \sum_{l=0}^{L}\sum_{m=-l}^{l} c_{lm}Y_l^m(\mathbf{u})$, and considering $\Delta_b Y_l^m(\mathbf{u}) = -l(l+1)Y_l^m(\mathbf{u})$ in Eq. (2.32) and Corollary 2.4, we have

$$\Phi_w(\mathbf{r}) = \frac{1}{4\pi} - \frac{1}{8\pi}l(l+1)P_l(0)c_{lm}Y_l^m(\mathbf{u})$$

The coefficients $\{c_{lm}\}$ can be estimated through a least square fitting from the samples of $\ln\left(-\ln E(q_0\mathbf{u})\right)$. Note based on the above formula the first coefficient is $c_{00} = \frac{1}{\sqrt{4\pi}}$, then the integration of the estimated ODF is $\int_{\mathbb{S}^2} c_{00}Y_0^0(\mathbf{u})\mathrm{d}\mathbf{u} = 1$. Thus the estimated $\Phi_w(\mathbf{r})$ is naturally normalized, which agrees with the definition.

Dr. Tristán-Vega also proposed two ways to estimate $\Phi_w(\mathbf{r})$ from single shell data $E(q_0\mathbf{u})$ in [Tristán-Vega et al., 2009; Tristán-Vega et al., 2010].

- [Tristán-Vega et al., 2009] noticed that two ODFs can be considered as a plane integration of $E(\mathbf{q})$ as shown in Eq. (4.65). Since QBI by Tuch has been successfully used in HARDI domain to approximate $\Phi_t(\mathbf{r})$ using FRT of $E(\mathbf{q})$, [Tristán-Vega et al., 2009] estimated $\Phi_w(\mathbf{r})$ using the FRT of $-\frac{1}{8\pi^2}\Delta E(\mathbf{q})$ with a specific assumption to solve the first integration term I in Eq. (4.66). This estimator has its intrinsic limitation from FRT because it assumes radial decay is a delta function such that the plane integration can be replaced by circle integration. The estimated ODF needs an artificial normalization, although $\Phi_w(\mathbf{r})$ is naturally normalized by definition. Our experiments demonstrated that the estimator in [Tristán-Vega et al., 2009] works well when q is small as shown in the paper. However, it obtains noisy results with many negative values when q is large, e.g. $b = 3000s/mm^2$. This modeling error is actually introduced by assumption of delta function in radial decay, which will be revisited and explained in subsection 5.2.2.
- [Tristán-Vega et al., 2010] claimed that it proposed an estimator of $\Phi_w(\mathbf{r})$ with mono-exponential decay assumption only in a local sense around the single shell with $q = q_0$, not the whole plane $\Phi_\mathbf{r}$ as proposed in [Aganj et al., 2010b]. It replaced the plane integration of $\frac{1}{q}\Delta_b E(\mathbf{q})$ in Eq. (4.1) as an integration in the disk with the given radius q_0, while it keeps the first term $\frac{1}{4\pi}$. Based on the so called Stokes' theorem, the disk integration is solved by circle integration after estimation of a vector field in the circle. However, based on our analysis in Appendix 4.7, the disk integration is actually solved based on the mono-exponential decay assumption inside the disk, not just around the boundary of the disk. Moreover the Stokes' theorem in [Tristán-Vega et al., 2010] is equivalent to separation disk integration into the radial integration and circular integration used in other papers [Canales-Rodrıguez et al., 2009; Aganj et al., 2010b; Cheng et al., 2010a].

Besides the two ODFs $\Phi_t(\mathbf{r})$ and $\Phi_w(\mathbf{r})$, [Canales-Rodrıguez et al., 2009] proposed a general form of ODF as

$$\Phi_k(\mathbf{r}) \overset{\text{def}}{=} \frac{1}{Z}\int_0^\infty P(\mathbf{R})R^k \mathrm{d}R \tag{4.69}$$

where Z is the normalization factor which makes $\int_{\mathbb{S}^2}\Phi_k(\mathbf{r})\mathrm{d}\mathbf{r} = 1$. Then $\Phi_0(\mathbf{r})$ is $\Phi_t(\mathbf{r})$ when $k = 0$, and $\Phi_2(\mathbf{r})$ is $\Phi_w(\mathbf{r})$ when $k = 2$. Like $\Phi_0(\mathbf{r})$, Φ_k needs a normalization factor Z to make it a PDF. Only when $k = 2$, $\Phi_k(\mathbf{r}) = \Phi_w(\mathbf{r})$ is naturally a PDF without artificial normalization factor, i.e. $Z = 1$.

Note the EAP profile $P(R\mathbf{r})$ normally has better angular resolution when R is larger, which can be seen from the synthetic data experiment in Fig. 4.8. Theoretically, $k = 1$ means the same weight is considered for EAP profile $P(R\mathbf{r})$ with large R and small R. While $k > 1$ gives more weight for $P(R\mathbf{r})$ with large R, which means the $\Phi_k(\mathbf{r})$ is sharper when k is larger. However, EAP profile $P(R\mathbf{r})$ with large R value has more estimation error. Thus normally only the ODFs with $k = 0$ and $k = 2$ are considered, and the ODF for other k are not used in literature. The experiments showed that although $\Phi_w(\mathbf{r})$ is sharper than $\Phi_t(\mathbf{r})$ with better angular resolution, $\Phi_t(\mathbf{r})$ is more robust than $\Phi_w(\mathbf{r})$ [Aganj et al., 2010b; Tristán-Vega et al., 2010; Cheng et al., 2010a].

The ODFs Φ_t in Eq. (4.47) and Φ_k in Eq. (4.69) need a normalization factor Z to make them as PDFs in $\mathbb{S}^2$. Many papers in HARDI set $Z = \sum_{j=1}\Phi_k(\mathbf{r}_j)$ such that the summation of ODF over some given direction samples $\{\mathbf{r}_j\}_{j=1}^{N_s}$ in $\mathbb{S}^2$ is 1 [Tuch, 2004; Khachaturian et al., 2007], which is actually problematic in theory and may bring some errors in some situations. See Appendix 4.6 for the reason and two correct ODF normalization ways.

Note like the original QBI, the exact QBI methods based on mono-exponential decay assumption can not work for multiple shell data. [Aganj et al., 2010b] proposed to generalize the mono-exponential decay assumption to multi-exponential decay assumption, i.e.

$$E(\mathbf{q}) = \sum_{i=1}^{K} w_i \exp(-4\pi^2\tau q^2 D_i(\mathbf{u})) = \sum_{i=1}^{K} w_i E_i(\mathbf{u})^b, \quad E_i(\mathbf{u}) = \exp(-D_i(\mathbf{u})) \tag{4.70}$$

$\{E_i(\mathbf{u})\}$ and $\{w_i\}$ can be estimated by numerical optimization such as trust region algorithm. An analytical solution exists when $K = 2$, the data were obtained from three b values, and $0, b_1, b_2, b_3$ is an arithmetic progress.

Summary of Advantages and Limitations:

- ✔ The linear analytical solution is fast and easy to be implemented.
- ✔ Compared to the delta function assumption of radial decay of $E(q\mathbf{u})$ used in original QBI, exact QBI considers mono-exponential decay of $E(q\mathbf{u})$ to obtain analytical results, which becomes more and more popular in sHARDI methods. Mono-exponential decay assumption seems better than delta function assumption. It at least satisfies the prior $E(0) = 1$ and can be seen as ADC modeling which is a kind of generalization of DTI. See Table 4.1 for the summary of different estimators in original QBI under delta function decay assumption and exact QBI under mono-exponential decay assumption.

✔ $\Phi_w(\mathbf{r})$ has better angular resolution than $\Phi_t(\mathbf{r})$ normally used in original QBI, although normally $\Phi_t(\mathbf{r})$ is more robust to noise. $\Phi_w(\mathbf{r})$ normally does not need the artificial normalization, sharpening technique like min-max normalization or spherical deconvolution. That is the main reason why $\Phi_w(\mathbf{r})$ is now more popular than $\Phi_t(\mathbf{r})$ in sHARDI methods.

✘ Although mono-exponential decay assumption is better than delta function assumption, it is still a strong and unrealistic assumption of radial decay. Even the signal generalized by a simple mixture of tensor model does not follow this assumption as shown in Fig. 4.8.

✘ Exact QBI is a kind of sHARDI method. It obtains different results for the data from different shells. [Aganj et al., 2010b] proposed an extension of the model based on multi-exponential decay assumption. However it is impractical for general multiple shell sampling which needs a numerical optimization to estimate each mono-exponential part. An analytical solution only exists when two multi-exponential parts are considered for the data sampled from three b values which are arithmetic progress.

4.3.5 Diffusion Orientation Transform (DOT)

Diffusion Orientation Transform (DOT) was proposed by Dr. Özarslan in [Özarslan et al., 2006] to estimate the EAP profile $P(R\mathbf{r})$ from single shell data under the mono-exponential decay assumption in Eq. (4.39). Thus it can be seen as an estimator of EAP in exact QBI methods. Consider the plane wave equation in Eq. (4.71)

$$\cos(2\pi\mathbf{q}^T\mathbf{R}) = 4\pi\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(-1)^{l/2}j_l(2\pi qR)Y_l^m(\mathbf{u})Y_l^m(\mathbf{r}) \tag{4.71}$$

where $j_l(\cdot)$ is the l order spherical Bessel function described in Section 2.5.3. See Theorem 2.8 and Eq. (2.53) for the plane wave equation. Then the EAP can be represented as

$$P(R\mathbf{r}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\left(\int_{\mathbb{S}^2}Y_l^m(\mathbf{u})I_l(R,\mathbf{u})\mathrm{d}\mathbf{u}\right)Y_l^m(\mathbf{r}) \tag{4.72}$$

$$I_l(R,\mathbf{u}) = 4\pi(-1)^{l/2}\int_0^{\infty}E(\mathbf{q})j_l(2\pi qR)q^2dq = \frac{R^l\Gamma(0.5l+1.5)\,{}_1F_1(0.5l+1.5,l+1.5,-\frac{R^2}{4\tau D(\mathbf{u})})}{(-1)^{l/2}2^{l+1}\pi^{0.5}(D(\mathbf{u})\tau)^{0.5l+1.5}\Gamma(l+1.5)} \tag{4.73}$$

For fixed $R = R_0$ and given direction samples $\{u_j\}$, samples $\{I_l(R_0,\mathbf{u}_j)\}$ can be calculated analytically via Eq. (4.73) from the samples of ADC $\{D(\mathbf{u}_j)\}$. Then a least square fitting can be used to obtain the coefficients of $P(R_0\mathbf{r})$ under SH basis from $\{I_l(R_0,\mathbf{u}_j)\}$.

[Özarslan et al., 2006] validated the mono-exponential decay assumption through synthetic data generated from cylinder model [Söderman and Jönsson, 1995]. It showed that signal decay can be approximated well as a mono-exponential function around $b = 1500s/mm^2$. For the b value large than $3000s/mm^2$, the mono-exponential decay assumption is not well satisfied, and the data with large b value has low SNR. Thus $1500s/mm^2$ seems to be the optimal b value for DOT.

Note like original QBI and exact QBI, DOT can not handle multiple shell data, because the data in different shell obtains different EAP profile. [Özarslan et al.,

	$\Phi_t(\mathbf{r})$	$\Phi_w(\mathbf{r})$	$P(\mathbf{R})$
original QBI $\delta(q-q_0)$	[Tuch, 2004] [Descoteaux et al., 2007]	[Tristán-Vega et al., 2009] subsection 5.2.2	subsection 5.2.2
exact QBI $E(\mathbf{q}) = \exp(-4\pi^2\tau D(\mathbf{u}))$	[Canales-Rodrıguez et al., 2009]	[Aganj et al., 2010b] [Tristán-Vega et al., 2010]	DOT [Özarslan et al., 2006]
SPFI	[Assemlal et al., 2009a; Cheng et al., 2010a] proposition 5.3	proposition 5.4	[Cheng et al., 2010b] proposition 5.2
SHORE	proposition 5.7 [Özarslan et al., 2009]		proposition 5.6

Table 4.1: HARDI methods with different assumptions for estimation of ODFs and EAP. Original QBI and exact QBI are sHARDI methods. They need respectively delta function assumption and mono-exponential decay assumption for signal decay. DOT can be seen as a kind of exact QBI method, because it is an estimator of EAP from single shell data based on mono-exponential decay. SPFI and SHORE consider radial basis for the radial decay. They are model-free.

2006] proposed to extend the mono-exponential model to multi-exponential model as shown in Eq. (4.70). It fits the multi-exponential function in each direction $\mathbf{u}$ which needs many samples along $\mathbf{u}$. Thus the multi-exponential assumption is impractical and so far to our knowledge there is no work to implement the multi-exponential version of DOT in real data.

DOT will be revisited in subsection 5.2.2, and theoretically and experimentally compared with other HARDI methods.

Summary of Advantages and Limitations:

✔ The linear analytical solution is fast and easy to be implemented.

✔ Compared to other ADC based methods like HOT model, DOT estimates the EAP profile with given radius R_0 from single shell data. The maxima of EAP profile normally can be used as the fiber directions. See Fig. 4.8.

✔ It works well with typical value $b = 1500s/mm^2$ compared to typical b value $3000s/mm^2$ in original QBI.

✘ DOT can not handle multiple shell data. The data from different shells obtain inconsistent results. The extension to multi-exponential model is impractical.

✘ In DOT, there is no full representation for EAP $P(R\mathbf{r})$ in $\mathbb{R}^3$. The results from DOT are the coefficients of the EAP profile $P(R_0\mathbf{r})$ with the given radius R_0 under SH basis. Even if one EAP profile $P(R_1\mathbf{r})$ is estimated, the estimation process needs to be performed again for EAP profile $P(R_2\mathbf{r})$ with different radius R_2.

4.3.6 Spherical Deconvolution (SD)

Spherical Deconvolution (SD) methods generalize the mixture model from discrete case to continuous case. In previous mixture of tensor model, $E(\mathbf{q})$ is assumed to

be generated from K tensors in Eq. (4.32). [Tournier et al., 2004, 2007] proposed to consider the continuous mixture model as

$$E(q\mathbf{u}) = \int_{\mathbb{S}^2} \Phi_f(\mathbf{r}) R(\mathbf{r}^T\mathbf{u}) \mathrm{d}\mathbf{r} \tag{4.74}$$

where $\Phi_f(\mathbf{r})$ is called as the fiber ODF (fODF) which needs to be estimated and $R(\mathbf{r}^T\mathbf{u})$ is the typical signal generated from one fiber. The spherical deconvolution is a model-based method because it assumes the typical signal $R(\mathbf{r}^T\mathbf{u})$ and linear combination in the convolution. Mixture of tensor model is suffering from the model selection of the number of tensors and local minima of cost function. However, SD can be solved analytically by considering the Funk-Hecke theorem and representing the $E(q\mathbf{u})$ and $R(\mathbf{r}^T\mathbf{u})$ using SHs [Descoteaux et al., 2008a]. The continuous weighting function $\Phi_f(\mathbf{r})$ avoids the limitation of mixture of tensor model in discrete case.

Note the SD method can be also used in some EAP features generated from signal. For example, consider Φ_t estimated from FRT, then based on the linearity of FRT we have

$$\Phi_t(\mathbf{r}) = \mathrm{FRT}\{E(q\mathbf{u})\} = \int_{\mathbb{S}^2} \Phi_f(\boldsymbol{w}) \mathrm{FRT}\{R(\boldsymbol{w}^T\mathbf{u})\} \mathrm{d}\boldsymbol{w} = \int_{\mathbb{S}^2} \Phi_f(\boldsymbol{w}) \Phi_t^R(\mathbf{r}^T\boldsymbol{w}) \mathrm{d}\boldsymbol{w} \tag{4.75}$$

Thus, if we use FRT to estimate $\Phi_t(\mathbf{r})$, the SD performed on $E(q\mathbf{u})$ is equivalent with SD performed on estimated $\Phi_t(\mathbf{r})$ [Descoteaux et al., 2008a]. Since $\Phi_t(\mathbf{r})$ estimated from FRT is normally very smooth. SD becomes a good option to obtain the sharpened fiber ODF $\Phi_f(\mathbf{r})$.

Compared to fiber ODF $\Phi_f(\mathbf{r})$, the ODFs Φ_t, Φ_w and Φ_k defined as the radial integration of EAP is normally called as the diffusion ODFs. Note that throughout this thesis, ODF means the diffusion ODF unless otherwise specified, because we only consider the estimation of diffusion ODFs, and once we assume a kernel model for single fiber, fiber ODF can be estimated from diffusion ODF easily.

The SD method also can be used as a general continuous weighting method in a specific manifold. Then many methods can be seen as some kinds of SD methods [Jian and Vemuri, 2007; Jian et al., 2007].

Summary of Advantages and Limitations:

- ✔ SD obtains more sharper ODFs, which is useful in fiber tracking. Note the diffusion ODFs sometime merge their maxima if two fiber directions are close. The sharpened ODFs obtained by SD sometimes can separate these merged maxima.
- ✔ Normally SD on sphere can be solved analytically using SH basis. However, it is not easy to do SD in a general manifold [Jian and Vemuri, 2007].
- ✘ SD is a model-based method. How to choose the kernel based on some priors is an open question.
- ✘ Although compared to diffusion ODFs, fiber ODFs from SD can obtain more maxima, we still do not know if added maxima are because of real fiber directions or noise. Even in the area with isotropic diffusion, SD can make sharper ODFs, which is unrealistic.

4.3.7 Diffusion Propagator Imaging (DPI)

Diffusion Propagator Imaging (DPI) was proposed to model the signal $E(\mathbf{q})$ as the solution of Laplace's equation [Descoteaux et al., 2009, 2010]. Note the original description of DPI in [Descoteaux et al., 2009] has some problems due to the integration of Bessel function, thus we use the formulation in [Descoteaux et al., 2010]. In DPI, the signal is assumed to be

$$E(q\mathbf{u}) = \sum_{l=0}^{L} \sum_{m=-l}^{l} \left(\frac{c_{lm}}{q^{l+1}} + d_{lm} q^l \right) Y_l^m(\mathbf{u}) \tag{4.76}$$

Then the EAP is estimated from incomplete 3D integration inside the ball with a given radius q_{max}, because the complete integration in $\mathbb{R}^3$ does not converge.

$$P(R_0\mathbf{r}) \approx \frac{1}{Z} \int_0^{q_{max}} \int_{\mathbb{S}^2} E(q\mathbf{u}) q^2 \exp\left(-2\pi i q R_0 \mathbf{u}^T \mathbf{r}\right) \mathrm{d}q\mathrm{d}\mathbf{u} = \frac{1}{Z} \sum_{l=0}^{L} \sum_{m=-l}^{l} \left(p_{0l}(R_0) c_{lm} + p_{1l}(R_0) d_{lm}\right) Y_l^m(\mathbf{r}) \tag{4.77}$$

$$p_{0l}(R_0) = \frac{(-1)^{l/2}}{R_0^{3/2}} \left(\frac{2^l R_0^{l-1/2} \pi^{l-1}}{(2l-1)!!} - \frac{J_{l-1/2}(2\pi q_{max} R_0)}{q_{max}^{l-1/2}} \right)$$

$$p_{1l}(R_0) = (-1)^{l/2} q_{max}^{l+3/2} R_0^{-3/2} J_{l+3/2}(2\pi q_{max} R_0)$$

where $(n-1)!! = (n-1) \cdot (n-3) \cdots 3 \cdot 1$, $J_n(x)$ is the Bessel function of order n, Z is the normalization factor to make $\int_{\mathbb{R}^3} P(\mathbf{R})\mathrm{d}\mathbf{R} = 1$, q_{max} is the maximum q value used in DPI acquisition. See subsection 2.5.3 for Bessel function and its properties. The coefficients $\{c_{lm}\}$ and $\{d_{lm}\}$ can be calculated from DWI samples via a standard least square estimation.

After obtaining the coefficients, DPI also proposed several EAP features analytically from incomplete radial integration, such as two ODFs in Eq. (4.78) and Eq. (4.79), and RTO in Eq. (4.80).

$$\Phi_t(\mathbf{r}) = \frac{1}{Z} \int_0^{R_{max}} P(R\mathbf{r}) \mathrm{d}R = \frac{1}{Z} \sum_{l=0}^{L} \sum_{m=-l}^{l} (-1)^{l/2} \left(t_{0l} c_{lm} - t_{1l} c_{lm} + t_{2l} d_{lm}\right) Y_l^m(\mathbf{r}) \tag{4.78}$$

$$t_{0l} = \frac{2^l \pi^{l-1}}{(2l-1)!!} \left(\frac{R_{max}^{l-1}}{l-1} \right), \quad t_{1l} = \frac{\pi (l+1)!!}{2 q_{max}^{l-1} (l/2+1)!}, \quad t_{2l} = \frac{(-1)^{l/2} q_{max}^2 (l-1)!!}{2^{l/2+2} (2\pi)^{3/2} (l/2+1)!}$$

$$\Phi_w(\mathbf{r}) = \frac{1}{Z} \int_0^{R_{max}} P(R\mathbf{r}) R^2 \mathrm{d}R = \frac{1}{Z} \sum_{l=0}^{L} \sum_{m=-l}^{l} (-1)^{l/2} \left(m_{0l} c_{lm} - m_{1l} c_{lm} + m_{2l} d_{lm}\right) Y_l^m(\mathbf{r}) \tag{4.79}$$

$$m_{0l} = \frac{2^l \pi^{l-1}}{(2l-1)!!} \left(\frac{R_{max}^{l+1}}{l+1} \right), \quad m_{1l} = \frac{(l-1)!!}{2\pi q_{max}^{l+1} 2^{l/2} (l/2-1)!}, \quad m_{2l} = \frac{(l+1)!!}{2\pi 2^{l/2} (l/2)!}$$

$$Po = \int_0^{q_{max}} \int_{\mathbb{S}^2} E(\mathbf{q}) q^2 \mathrm{d}q\mathrm{d}\mathbf{u} = q_{max}^2 \left(c_{00} + d_{00} \frac{2 q_{max}}{3} \right) \sqrt{\pi} \tag{4.80}$$

Summary of Advantages and Limitations:

- ✔ DPI can be seen as a generalization of QBI method to handle multiple shell data, although this generalization has many problems which will be revisited in subsection 5.2.2.
- ✘ DPI is a model-based method, because it assumes $\Delta E(\mathbf{q}) = 0$.
- ✘ Compared to DOT based on the mono-exponential decay assumption, the normalization factor Z is needed because the integration of $E(\mathbf{q})$ is incomplete. In DOT, the estimated EAP is naturally normalized, because the mono-exponential decay assumption in DOT satisfies $E(0) = 1 = \int_{\mathbb{R}^3} P(\mathbf{R})d\mathbf{R}$.
- ✘ The ODFs are obtained from incomplete radial integration of the pre-estimated EAP. Thus the estimated ODFs suffer from two incomplete integrations, one for EAP estimation, the other one for ODF estimation.
- ✘ The introduction of q_{max} makes all estimated EAP and its features are function of q_{max}. We still do not know how to choose q_{max} generally. Although we can set it as the maximal q values of the acquired DWI data, it is actually not a good choice. See subsection 5.3.1 for the discussion of the problems brought by q_{max}.
- ✘ The DPI model does not satisfy some priors of signal $E(\mathbf{q})$ which brings intrinsic modeling errors. It also can not represent an isotropic Gaussian signal. Please see subsection 5.2.2 and 5.3.1 for more details.

4.3.8 Simple Harmonic Oscillator Reconstruction and Estimation (SHORE)

Simple Harmonic Oscillator Reconstruction and Estimation (SHORE) was proposed by Dr. Özarslan in [Özarslan et al., 2008] for 1D signal and in [Özarslan et al., 2009] for 3D signal.

SHO-1D Basis

[Özarslan et al., 2008] proposed to represent diffusion signal $E(q)$ in 1D as

$$E(q) = \sum_{n=0}^{N-1} a_n \psi_n(x)(q,u), \quad \text{where } \psi_n(q,u) = i^n (2^n n!)^{-1/2} e^{-2\pi^2 q^2 u^2} H_n(2\pi u q) \tag{4.81}$$

where $H_n(x)$ is the n order physicists' Hermite polynomial, and u is a characteristic length which scales the decay of the basis. Please refer subsection 2.5.4 for physicists' Hermite polynomial and its properties. The scale u was proposed to be estimated from the first few samples of $E(q)$, and the coefficients $\{a_n\}$ can estimated from signal samples via a standard least square fitting. As we assume $E(q)$ is symmetric, only even order n is used. Note $\{\psi_n(q,u)\}$ is an orthogonal basis, while the basis is not normalized to unit norm. We may consider the normalized version given in Eq. (2.60) which is the solution of 1D quantum mechanical harmonic oscillator problem, and is called as Simple Harmonic Oscillator in 1D (SHO-1D) basis in this thesis.

Based on the property of Hermite polynomial in Eq. (2.57), the EAP in 1D that is the Fourier transform of $E(q)$ can be analytically obtained as

$$P(x) = \sum_{n=0}^{N-1} \frac{(-i)^n}{\sqrt{2\pi u}} a_n \psi_n(x, (2\pi u)^{-1}) \tag{4.82}$$

Then RTO is simply $P(0)$, and the m-th order moment of EAP can be given by

$$\langle x^m \rangle = \int_{-\infty}^{\infty} x^m P(x) \mathrm{d}x = u^m \sum_{k=0,2,\ldots}^{N-1} \frac{(k+m-1)!!}{k!} \sum_{l=0,2,\ldots}^{N-k-1} (-1)^{l/2} \frac{\sqrt{2^{k-l}(k+l)!}}{(l/2)!} a_{k+l} \tag{4.83}$$

Note that GDTI considers product of probabilists' Hermite polynomials in 3D space, while the SHO-1D basis uses physicists' Hermite polynomials in 1D space. It is possible to consider the product of SHO-1D basis as a 3D basis, called SHO-1D3 basis in Definition 5.2. Please refer Theorem 5.2 for the equivalence between SHO-1D3 basis and SHO-3D basis which will be introduced in next subsection.

Summary of Advantages and Limitations:

- ✔ Since SHO-1D basis is complete, SHORE method in 1D is model-free.
- ✔ The linear analytical solutions for EAP and its features are very fast and easy to be implemented.
- ✔ The Gaussian-Hermite function considers the priors of signal which is a Gaussian like signal.
- ✘ Since the diffusion signal in real data is in 3D space, SHO-1D basis can not handle 3D signals.

SHO-3D Basis

[Özarslan et al., 2009] proposed to represent diffusion signal $E(q)$ in 3D as

$$E(\mathbf{q}) = \sum_{N=0}^{N_{max}} \sum_{\substack{l+2j=N+2 \\ j\geq 1, l\geq 0}} \sum_{m=-l}^{l} A_{jlm} \psi_{jlm}(\mathbf{q}, u) \tag{4.84}$$

$$\psi_{jlm}(\mathbf{q}, u) = i^{-l} \sqrt{4\pi} \left(2\pi^2 u^2 q^2\right)^{l/2} \exp\left(-2\pi^2 u^2 q^2\right) L_{j-1}^{l+1/2}\left(2\pi^2 u^2 q^2\right) Y_l^m(\mathbf{u}) \tag{4.85}$$

Note that [Özarslan et al., 2009] referred that the function set $\{\psi_{jlm}\}$ in Eq. (4.85) is an orthogonal basis. However, one can verify that these functions are not orthogonal, that is because the wrong scale is used in $\exp(\cdot)$. The solution of 3D quantum mechanical harmonic oscillator problem has the form in Eq. (2.69), which is called as the Simple Harmonic Oscillator in 3D (SHO-3D) basis. Actually this basis has been already used in the computation of the molecular electron orbitals and molecular docking [Ritchie and Kemp, 2000; Huzinaga, 1965], which we will discuss later. The SHO-3D basis is equivalent with the product of SHO-1D basis in 3D space. See Theorem 5.2.

[Özarslan et al., 2009] claimed that the EAP can be represented by the same form in Eq. (4.85) although it did not give the explicit formula. We will give the analytical formula based on the SHO-3D basis in Eq. (2.69). The orientation-dependent radial moments, i.e. the ODFs in Eq. (4.69), is given as

$$\Phi_{k+2}(\mathbf{r}) = \int_0^\infty P(R\mathbf{r})R^{k+2} = \sum_{jlm} A_{jlm} Y_l^m(\mathbf{r}) u^k \frac{2^{k/2}\Gamma(j+(l+k+1)/2)}{\pi(j-1)!} \tag{4.86}$$

Note that we did not verify the above formula for ODFs, because we think there may be some problems in the basis in [Özarslan et al., 2009]. The ODFs in SHORE-3D method will be revisited in subsection 5.2.2 using the corrected SHO-3D basis in Eq. (2.69). See Table 4.1 for an overview.

Summary of Advantages and Limitations:

- ✔ SHORE is model-free, since the SHO-1D basis and SHO-3D basis are both complete basis in their domains.
- ✔ The linear analytical solutions are very fast.
- ✘ Some formulae have some problems. For example, the proposed formula for SHO-3D basis in Eq. (4.85) is not orthogonal basis, which conflicts with the claim that SHO-3D basis are orthogonal. It can be easily improved by using the correct orthonormal basis in Eq. (2.69). We will revisit this topic in subsection 5.2.2.

4.3.9 Spherical Polar Fourier Imaging (SPFI)

Spherical Polar Fourier Imaging (SPFI) was first proposed by Dr. Assemlal in [Assemlal et al., 2008, 2009a], where the diffusion signal $E(\mathbf{q})$ is represented by Spherical Polar Fourier (SPF) basis. SPF basis is a 3D orthonormal basis with SHs in spherical part and Gaussian–Laguerre functions in radial part. This basis is motivated by SHO-3D basis used in the computation of the molecular electron orbitals and molecular docking [Ritchie and Kemp, 2000; Huzinaga, 1965]. However SPF basis is different from the SHO-3D basis. We will compare SPF basis and SHO basis theoretically and experimentally in Chapter 5.

After we estimate the coefficients of diffusion signal under SPF basis, EAP and its various features, e.g. ODFs, RTO, can be obtained in a numerical way from an inner proposed by Dr. Assemlal [Assemlal et al., 2008, 2009a], which will be described in details in subsection 5.1.1. In Chapter 5, we propose an analytical way to obtain the EAP and its various features from the pre-estimated SPF coefficients [Cheng et al., 2010b,a]. We also propose some improvements in the estimation of the coefficients of SPF basis. See Table 4.1 and Table 5.1 for an overview.

4.4 METRICS AND FRAMEWORKS TO PROCESS PDF-VALUED DATA

As we have demonstrated in subsection 4.1.2 and 4.1.3, the data in dMRI field is in 6D space. In every voxel $\boldsymbol{x}$, there is an EAP $P(\mathbf{R})$ defined $\mathbb{R}^3$. The whole EAP field is $P(\mathbf{R}, \boldsymbol{x})$ in 6D space, which is different from the conventional MRI data in 3D space. Thus we need to study how to process such a PDF field.

4.4.1 Metrics and Frameworks for Tensor Computing

Euclidean Metric

Under the assumption of Gaussian propagator, the EAP field becomes the tensor field. Since the tensor space is a convex set in symmetric matrix space, we can define the distance between two tensors as the distance between two symmetric matrices, i.e. the Euclidean distance between 6D vectors.

$$d_{\text{Euc}}(\mathbf{D}_1, \mathbf{D}_2) = \|\mathbf{D}_1 - \mathbf{D}_2\| = \sqrt{\text{Trace}\left((\mathbf{D}_1 - \mathbf{D}_2)^T(\mathbf{D}_1 - \mathbf{D}_2)\right)} \tag{4.87}$$

In this case, the Euclidean metric in $\mathbb{R}^6$ is used. The Euclidean distance is rotational invariant. Based on this distance, the mean of K tensors is

$$\overline{\mathbf{D}} = \frac{1}{K}\sum_{i=1}^{K}\mathbf{D}_i \tag{4.88}$$

J-divergence

The Kullback-Leibler (KL) divergence between two Gaussian propagators is

$$d_{\text{KL}}(\mathbf{D}_1, \mathbf{D}_2) = \int_{\mathbb{R}^3} N(\mathbf{R}|2\tau\mathbf{D}_1)\ln\frac{N(\mathbf{R}|2\tau\mathbf{D}_1)}{N(\mathbf{R}|2\tau\mathbf{D}_2)}\mathrm{d}\mathbf{R} = \frac{1}{2}\left(\ln\frac{|\mathbf{D}_2|}{|\mathbf{D}_1|} + \text{Trace}\left(\mathbf{D}_2^{-1}\mathbf{D}_1\right) - 3\right) \tag{4.89}$$

It is affine invariant, i.e. $d_{KL}(A^T\mathbf{D}_1A, A^T\mathbf{D}_2A) = d_{KL}(\mathbf{D}_1, \mathbf{D}_2)$, $\forall|A| \neq 0$. Since the KL divergence is not symmetric, we can use its symmetric version called J-divergence and its square root distance as [Wang and Vemuri, 2005]:

$$d_J(\mathbf{D}_1, \mathbf{D}_2) = \sqrt{\frac{1}{2}\left(d_{KL}(\mathbf{D}_1, \mathbf{D}_2) + d_{KL}(\mathbf{D}_2, \mathbf{D}_1)\right)} = \frac{1}{2}\sqrt{\text{Trace}\left(\mathbf{D}_1^{-1}\mathbf{D}_2 + \mathbf{D}_2^{-1}\mathbf{D}_1\right) - 6} \tag{4.90}$$

J-divergence is also affine invariant, because KL divergence is affine invariant. Although J-divergence is symmetric, it does not satisfy the triangle inequality, which means it is a divergence, not a distance. [Wang and Vemuri, 2005] showed that the mean of tensors $\{\mathbf{D}_i\}_{i=1}^K$, i.e. the minimizer of $\sum_{i=1}^K d_J(\mathbf{D}, \mathbf{D}_i)^2$, is

$$\overline{\mathbf{D}} = \mathbf{B}^{-\frac{1}{2}}\left(\mathbf{B}^{\frac{1}{2}}\mathbf{A}\mathbf{B}^{\frac{1}{2}}\right)^{\frac{1}{2}}\mathbf{B}^{-\frac{1}{2}} \tag{4.91}$$

where

$$\mathbf{A} = \sum_{i=1}^{K}\mathbf{D}_i, \quad \mathbf{B} = \sum_{i=1}^{K}\mathbf{D}_i^{-1}$$

Riemannian Metric and Log-Euclidean Metric

The Riemannian metric for Gaussian distribution family has been proposed in 1980s [Atkinson and Mitchell, 1981; Skovgaard, 1984] based on Rao's seminal work on information geometry [Rao, 1945]. However, the Riemannian metric was only known in dMRI domain after 2005, until a number of works [Pennec et al., 2006; Moakher, 2005; Batchelor et al., 2005; Lenglet et al., 2006b; Fletcher and Joshi, 2007] which introduced Riemannian metric in dMRI. The Riemannian metric can be introduced based on the analysis of Gaussian distribution family in Eq. (4.92) [Lenglet et al., 2006b] where $\Sigma = 2\tau\mathbf{D}$, or based on the analysis of affine invariance [Pennec et al., 2006; Fletcher and Joshi, 2007; Fletcher, 2004; Moakher, 2005].

$$\left\{ N(\mathbf{R}|\Sigma) = \frac{1}{(2\pi)^{3/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{R}^T\Sigma^{-1}\mathbf{R}\right) : \quad \Sigma \in Sym_3^+ \right\} \tag{4.92}$$

The Riemannian distance between two tensors is

$$d_{\text{Rie}}(\mathbf{D}_1, \mathbf{D}_2) = \sqrt{\frac{1}{2}\sum_{i=1}^{3}(\ln \lambda_i)^2} \tag{4.93}$$

where $\{\lambda_i\}_{i=1}^3$ are the eigenvalues of matrix[5] $\mathbf{D}_1^{-1}\mathbf{D}_2$. This Riemannian distance is affine invariant. Please see Example 3.10 for more details on the Riemannian metric of Gaussian distribution family. For a set of tensors $\{\mathbf{D}_i\}_{i=1}^K$, the Riemannian mean has no closed form. It needs a gradient descent method to find a local minimum of the following cost function.

$$\min_{\mathbf{D}} d_{\text{Rie}}(\mathbf{D}, \mathbf{D}_i)^2 \tag{4.94}$$

Note the cost function is convex and the Riemannian mean of tensors uniquely exist because the sectional curvature in tensor space is non-positive definite. See Theorem 3.1. Thus the estimated local minimum is also the global minimum.

The Log-Euclidean metric for tensors is defined based on the isometry of tensor space and symmetric matrix space [Arsigny et al., 2006; Fillard et al., 2007]. The matrix logarithm can be seen as the isometry, then the distance between tensors is defined by the distance between their logarithms, i.e.

$$d_{\text{LogEuc}}(\mathbf{D}_1, \mathbf{D}_2) = \|\ln \mathbf{D}_1 - \ln \mathbf{D}_2\| = \sqrt{\text{Trace}\left((\ln \mathbf{D}_1 - \ln \mathbf{D}_2)^T(\ln \mathbf{D}_1 - \ln \mathbf{D}_2)\right)} \tag{4.95}$$

Then the Log-Euclidean mean of tensors is

$$\overline{\mathbf{D}} = \sum_{i=1}^{K} \ln \mathbf{D}_i \tag{4.96}$$

Log-Euclidean distance is different from Riemannian distance, but in practice, these two distances for tensors in human brain are qualitatively/visually similar. Thus Log-Euclidean distance can be seen as an approximation of Riemannian distance. When computing the Riemannian mean, it is fast and convenient to set the initialization point as the Log-Euclidean mean.

[5] Some papers use $\mathbf{D}_2^{-1}\mathbf{D}_1$ or $\mathbf{D}_1^{-\frac{1}{2}}\mathbf{D}_2\mathbf{D}_1^{-\frac{1}{2}}$ here. All these descriptions are OK, because it can be proved that the final results are the same.

The Geodesic Anisotropy (GA) of a tensor $\mathbf{D}$ is defined as the geodesic distance between $\mathbf{D}$ and its nearest isotropic tensor [Fletcher, 2004; Batchelor et al., 2005], i.e.

$$\mathrm{GA}(\mathbf{D}) = d_{\mathrm{Rie}}(\mathbf{D}, |\mathbf{D}|^{\frac{1}{3}}\mathbf{I}) = d_{\mathrm{Rie}}(|\mathbf{D}|^{-\frac{1}{3}}\mathbf{D}, \mathbf{I}) = \sqrt{\sum_{i=1}^{3}\left(\ln\lambda_i - \frac{1}{3}\ln(\lambda_1\lambda_2\lambda_3)\right)^2} \tag{4.97}$$

The Riemannian framework and Log-Euclidean framework of tensors can be used in DTI or other domains like computer vision [Li et al., 2008]. In dMRI, these two frameworks can be used in tensor estimation [Lenglet et al., 2006b; Fillard et al., 2007], regularization [Pennec et al., 2006], segmentation [Lenglet et al., 2006a], Principal Component Analysis (PGA) [Fletcher et al., 2004], statistical test [Schwartzman, 2006], etc.

4.4.2 Metrics and Frameworks for ODF and EAP Computing

Since DTI model based on Gaussian assumption has its intrinsic limitations, we do not assume Gaussian propagator in HARDI. The ODF and EAP are PDFs respectively defined in $\mathbb{S}^2$ and $\mathbb{R}^3$. Historically there are several choices proposed for the distance between ODFs. [Tuch, 2004] used KL divergence to measure the difference of ODFs, which is calculated from ODF samples. [Descoteaux and Deriche, 2009; Wassermann et al., 2008] considered the Euclidean distance between ODFs in ODF segmentation, which can be represented by coefficients of SH basis. Assume $\Phi(\mathbf{r}) = \sum_{l=0}^{L}\sum_{m=-l}^{l} c_{lm}Y_l^m(\mathbf{r})$ and $\Phi'(\mathbf{r}) = \sum_{l=0}^{L}\sum_{m=-l}^{l} c'_{lm}Y_l^m(\mathbf{r})$, then the Euclidean distance between them is

$$\|\Phi(\mathbf{r}) - \Phi'(\mathbf{r})\| = \sqrt{\int_{\mathbb{S}^2}\left(\sum_{lm}(c_{lm} - c'_{lm})Y_l^m(\mathbf{r})\right)^2 \mathrm{d}\mathbf{r}} = \sqrt{\sum_{l=0}^{L}\sum_{m=-l}^{l}(c_{lm} - c'_{lm})^2} \tag{4.98}$$

[McGraw et al., 2006] proposed to represent the ODF as a mixture of Von Mises–Fisher distributions.

$$\Phi(\mathbf{r}) = \sum_{i=1}^{K} w_i M(\mathbf{r}|\boldsymbol{\mu}_i, \kappa_i), \quad M(\mathbf{r}|\boldsymbol{\mu}_i, \kappa_i) = \frac{\kappa_i}{2\pi(e^{\kappa_i} - e^{-\kappa_i})}\exp\left(\kappa_i\boldsymbol{\mu}_i^T\mathbf{r}\right) \tag{4.99}$$

The parameters $\{(w_i, \boldsymbol{\mu}_i, \kappa_i)\}_{i=1}^{K}$ need to be estimated through minimizing a nonlinear least square cost function. Since The Von Mises-Fisher distribution $M(\mathbf{r}|\boldsymbol{\mu}, \kappa)$ is not symmetric, both $M(\mathbf{r}|\boldsymbol{\mu}, \kappa)$ and $M(\mathbf{r}| - \boldsymbol{\mu}, \kappa)$ need to be considered in the cost function. The barrier functions for the positive $\{w_i\}$ and $\{\kappa_i\}$ are also needed in the cost function. Moreover, the model selection for parameter K needs to be set manually. Thus the estimation of the parameters are complex and unstable. After the parameters are obtained, since Von Mises-Fisher distribution has been well studied in information geometry, the distance between ODFs can be defined as the Riemannian distance in the mixture of Von Mises-Fisher distributions. Thus the Riemannian distance for ODFs is model-based and can not be used for general ODFs. To our knowledge, this is the first work for the Riemannian framework for ODFs, although it has many limitations.

[Goh et al., 2009a, 2011] and our previous work [Cheng et al., 2009a] proposed a model-free intrinsic Riemannian framework for ODFs based on square root parameterization. [Goh et al., 2009a, 2011] considered the continuous ODF as its histogram, and directly applied the Riemannian framework for multinomial distribution family to ODFs. However, our work considered the orthonormal basis representation, which is essentially more general and efficient than the formulation of histogram. Actually the histogram formulation in [Goh et al., 2009a, 2011] can be seen as a specific case of our basis representation formulation when the piecewise constant basis is considered. Moreover, our formulation can be used to EAPs by considering a 3D basis, e.g. SPF basis. While the histogram formulation is hard to be used in EAPs since EAP is defined in unbounded $\mathbb{R}^3$. To our knowledge, our previous work in [Cheng et al., 2011a] is the first work for the Riemannian framework on general EAPs, which is a natural generalization of previous Riemannian framework on Gaussian EAPs in DTI. Please refer Chapter 6 for more details and comparisons.

4.5 Summary

We have presented some basic concepts and background knowledge of MRI and dMRI for Chapter 5 and 6. We have reviewed some state-of-the-art reconstruction methods to estimate tensors, ODFs, EAPs, and other meaningful features from the measured diffusion signals, and have reviewed some metrics and frameworks used to process tensor/ODF/EAP valued images. Since DTI has its intrinsic limitations due to the Gaussian assumption, in this thesis we focus on ODF and EAP estimation and processing in HARDI methods. HARDI methods can be categorized into two classes: single shell HARDI (sHARDI) methods and multiple shell HARDI (mHARDI) methods. sHARDI methods like QBI cannot be used in multiple shell or sparse sampled data. These sHARDI methods always need some assumptions in radial decay which cannot be assumed in a local sense. See Table 4.1 and Appendix 4.7. We also proposed the correct way for ODF normalization. See Appendix 4.6, which is useful for the Riemannian framework of ODFs. If a complete 3D basis is used, mHARDI methods are indeed model-free, and they naturally combine the benefits of the high SNR of the data with the low b values and high angular resolution of the data at high b values. mHARDI methods are trade-off between a DSI acquisition and a single shell acquisition used in sHARDI methods. With hardware improvements of current scanners, mHARDI acquisitions will become more and more practical and popular. In CCM group, we mainly focus on neuroscience applications based on the scalar indices or brain network analysis in DTI model [Gong et al., 2005; Lin et al., 2006; Yu et al., 2008; Shu et al., 2009; Li et al., 2009; Wang et al., 2011; Li et al., 2012]. In Odyssée/Athena group, DTI and the Riemannian framework for tensors have been well studied by Dr. Lenglet [Lenglet, 2006; Lenglet et al., 2006a,b], and QBI and its related methods have been well studied by Dr. Descoteaux [Descoteaux, 2008; Descoteaux et al., 2007, 2008a]. In this thesis, we focus on mHARDI reconstruction and process. Analytical SPFI in Chapter 5 can be seen as a generalization of analytical QBI by Dr. Descoteaux and others, and the Riemannian framework for ODFs and EAPs in Chapter 6 is a generalization of the Riemannian framework for tensors by Dr. Lenglet and others. The scalar indices developed in mHARDI methods in Chapter 5 and 6 can be used as biomarkers in neuroscience applications, and the proposed

EAP/ODF estimations can improve the fiber tracking results for brain network analysis.

4.6 APPENDIX A: ODF NORMALIZATION

An important issue for ODFs is that the estimated ODF by Tuch $\Phi_t(\mathbf{r})$ and $\Phi_k(\mathbf{r})$ in Eq. (4.69) need to be normalized so that the integration in $\mathbb{S}^2$ is 1 [Tuch, 2004; Descoteaux et al., 2007]. Please note that the normalization is **NOT** to let the summation of ODF samples at given evenly distributed directions $\{\mathbf{r}_j\}_{j=1}^{N_s}$ be 1, because

$$1 = \int_{\mathbb{S}^2} \Phi_k(\mathbf{r})\mathrm{d}\mathbf{r} \simeq \sum_{j=1}^{N_s} \Phi_k(\mathbf{r}_j) S_j \neq \sum_j \Phi_k(\mathbf{r}_j)$$

where S_j is the area element for the histogram bins. Since the N_s samples are evenly distributed in $\mathbb{S}^2$, $\{S_j\}$ are close to a constant, i.e.

$$S_j \simeq \frac{4\pi}{N_s}, \quad 1 \leq j \leq N_s$$

and the ODF should be normalized such that

$$\sum_{j=1}^{N_s} \Phi_k(\mathbf{r}_j) = \frac{N_s}{4\pi}$$

So the normalization factor is

$$Z = \frac{4\pi}{N_s} \sum_{j=1}^{N_s} \Phi_k(\mathbf{r}_j) \tag{4.100}$$

Then the integration of new normalized ODF is

$$\int_{\mathbb{S}^2} \frac{1}{Z} \Phi_t(\mathbf{r})\mathrm{d}\mathbf{r} \simeq \left(\frac{N_s}{4\pi \sum_{j=1}^{N_s} \Phi_k(\mathbf{r}_j)} \right) \sum_{j=1}^{N_s} \Phi_k(\mathbf{r}_j) S_j \simeq 1$$

This is the discrete normalization. Note that in many papers $Z = \sum_{j=1}^{N_s} \Phi_k(\mathbf{r}_j)$ which only misses a constant factor $\frac{4\pi}{N_s}$. Thus normally this inappropriate normalization does not affect the final results based on the shape of ODFs. But the Riemannian framework in Chapter 6 needs the ODF/EAP correctly normalized. Otherwise the ODF/EAP is not a PDF.

We suggest ODF normalization in continuous domain which is better and independent of samples. When the estimated ODF is represented by SH basis, we have

$$\Phi_k(\mathbf{r}) = \sum_{lm} c_{lm} Y_l^m(\mathbf{r}), \quad 1 = \int_{\mathbb{S}^2} \Phi_k(\mathbf{r})\mathrm{d}\mathbf{r} = c_{00}\sqrt{4\pi}$$

So the coefficient for Y_0^0 should be normalized such that

$$c_{00} = \frac{1}{\sqrt{4\pi}}$$

which means the normalization factor is

$$Z = \sqrt{4\pi}c_{00} \tag{4.101}$$

Note the original ODF by Wedeen $\Phi_w(\mathbf{r})$ is naturally normalized in its estimation by exact QBI [Canales-Rodrıguez et al., 2009; Aganj et al., 2010b; Tristán-Vega et al., 2010], or by SPFI in Chapter 5, [Cheng et al., 2011c], and Square Root Parameterized Estimation (SRPE) in Chapter 6. Thus no extra normalization is needed for $\Phi_w(\mathbf{r})$ for these estimation methods. However the estimation methods in [Descoteaux et al., 2010; Tristán-Vega et al., 2009] for $\Phi_w(\mathbf{r})$ need an artificial normalization due to the incomplete integration.

4.7 Appendix B: Is mono-exponential decay assumption only needed locally for single shell data?

As we have shown in Corollary 4.1 that the two kinds of ODFs $\Phi_t(\mathbf{r})$ and $\Phi_w(\mathbf{r})$ can be represented as the plane integration of $E(\mathbf{q})$ and $-\frac{1}{8\pi^2}\frac{1}{q}\Delta_b E(\mathbf{q})$. To estimate the ODFs and EAPs from single shell data, historically there are two popular assumptions for signal decay $E(q\mathbf{u})$, i.e. delta function decay assumption in original QBI [Tuch, 2004; Anderson, 2005; Hess et al., 2006; Descoteaux et al., 2007] and mono-exponential decay assumption in exact QBI [Wu et al., 2008; Canales-Rodrıguez et al., 2009; Aganj et al., 2010b; Tristán-Vega et al., 2009; Tristán-Vega et al., 2010]. Compared to the delta function decay assumption, mono-exponential decay assumption seems better and becomes more and more popular in HARDI domain.

[Tristán-Vega et al., 2010] claimed that $\Phi_t(\mathbf{r})$ and $\Phi_w(\mathbf{r})$ can be estimated from single shell data with the mono-exponential assumption in a local sense in an environment of shell $q = q_0$. However, we think for single shell data the radial assumption is needed in whole plane, not in a local sense, such that the plane integration can be solved. Please see the following analysis in details. Note [Tristán-Vega et al., 2010] proposed the estimators for both $\Phi_t(\mathbf{r})$ and $\Phi_w(\mathbf{r})$. Our analysis for $\Phi_w(\mathbf{r})$ is given as follows which can be also used in the estimator of $\Phi_t(\mathbf{r})$.

Formulation based on Stokes' theorem in [Tristán-Vega et al., 2010]. The idea in [Tristán-Vega et al., 2010] is to replace the plane integration in Eq. (4.64) with the integration inside the disk with the given radius q_0, denoted by $\Pi_{\mathbf{r},q_0}$, i.e.

$$\Phi_w(\mathbf{r}) = \frac{1}{4\pi} - \frac{1}{8\pi^2}\int_{\Pi_\mathbf{r}} \frac{\Delta_b E(\mathbf{q})}{q}\mathrm{d}\mathbf{q} \simeq \frac{1}{4\pi} - \frac{1}{8\pi^2}\int_{\Pi_{\mathbf{r},q_0}} \frac{\Delta_b E(\mathbf{q})}{q}\mathrm{d}\mathbf{q}$$

Note in [Tristán-Vega et al., 2010] the second term is approximated by the disk integration, while the first term is $\frac{1}{4\pi}$ which is still plane integration. Then considering the Stokes' theorem, the disk integration of $E(q, \frac{\pi}{2}, \upsilon)$ equals the circle integration of some specific function $\widetilde{F_\upsilon^\Phi}(q, \frac{\pi}{2}, \upsilon)$, i.e.

$$\Phi_w(\mathbf{r}) \simeq \frac{1}{4\pi} - \frac{1}{8\pi^2}\int_{\Pi_{\mathbf{r},q_0}} \frac{\Delta_b E(\mathbf{q})}{q}\mathrm{d}\mathbf{q} = \frac{1}{4\pi} - \frac{1}{8\pi^2}\int_0^{2\pi} \widetilde{F_\upsilon^\Phi}(q_0, \frac{\pi}{2}, \upsilon) q_0 \mathrm{d}\upsilon \tag{4.102}$$

where $\widetilde{F_{\upsilon}^{\Phi}}(q,\xi,\upsilon)$ is given in Eq.(C.2) in [Tristán-Vega et al., 2010].

$$\widetilde{F_{\upsilon}^{\Phi}}(q,\xi,\upsilon) = \frac{-1}{8\pi^2 q}\int_0^q \frac{1}{u}\widetilde{\Delta_b E}(u,\xi,\upsilon)du \simeq \frac{-1}{8\pi^2 q}\int_0^q \frac{\Delta_b \exp(-4\pi^2\tau u^2 \widetilde{D}(q_0,\xi,\upsilon))}{u}du \tag{C.2}$$

Note $\widetilde{E}(u,\xi,\upsilon)$ is the assumed signal which was said to follow mono-exponential decay in a local environment of q_0, i.e. the Eq. (C.1) in [Tristán-Vega et al., 2010].

$$\widetilde{E}(q,\xi,\upsilon)|_{q\simeq q_0} \simeq \exp(-4\pi^2\tau q^2 \widetilde{D}(q_0,\xi,\upsilon)) \tag{C.1}$$

However, the integration in Eq. (C.2) actually assumes $E(q,\xi,\upsilon)$ is mono-exponential decay in $[0,q_0]$, not in a local sense in Eq. (C.1). To approximate the plane integration with disk integration, it also assumes $E(q,\xi,\upsilon)=0$ when $q>q_0$.

Formulation based on standard radial integration and spherical integration. Actually the Stokes' theorem used in the paper is just two integrals (radial integral and spherical integral), which is in spirit the same with other papers on plane integration [Canales-Rodrıguez et al., 2009; Aganj et al., 2010b; Cheng et al., 2010a].

$$\begin{aligned}
\Phi_w(\mathbf{r}) &\simeq \frac{1}{4\pi} + \frac{-1}{8\pi^2}\iint_{\Pi_{\mathbf{r},q_0}} \frac{1}{q^2}\Delta_b E(\mathbf{q})\mathrm{d}\mathbf{q} \\
&= \frac{1}{4\pi} + \frac{-1}{8\pi^2}\int_0^{q_0}\int_0^{2\pi} \frac{1}{q^2}\widetilde{\Delta_b E}(q,\frac{\pi}{2},\upsilon)q\mathrm{d}\upsilon\mathrm{d}q \\
&= \frac{1}{4\pi} + \int_0^{2\pi} \underbrace{\left(\frac{-1}{8\pi^2}\int_0^{q_0}\frac{1}{q}\widetilde{\Delta_b E}(q,\frac{\pi}{2},\upsilon)\mathrm{d}q\right)}_{q_0\widetilde{F_{\upsilon}^{\Phi}}(q_0,\frac{\pi}{2},\upsilon)} d\upsilon
\end{aligned} \tag{4.103}$$

Compared the above formula to Eq. (C.2) and Eq. (4.102), the function $\widetilde{F_{\upsilon}^{\Phi}}(q_0,\frac{\pi}{2},\upsilon)$ obtained by Stokes' theorem is actually the radial integration of $E(\mathbf{q})$. So the so called Stokes' theorem is indeed not necessary to be mentioned and we can get exactly the same results without this theorem, just using the standard way to separate the plane integration into radial integration and spherical integration.

Counterexample. Actually we can give some counterexamples where the signal follows the mono-exponential decay only in a local sense. For a Gaussian propagator parameterized by tensor $\mathbf{D}$, i.e. $P(\mathbf{R}) = N(\mathbf{R}|2\tau\mathbf{D})$, the ODF Φ_w is

$$\Phi_w(\mathbf{r}) = \frac{1}{4\pi\sqrt{|\mathbf{D}|}}\frac{1}{(\mathbf{r}^T\mathbf{D}^{-1}\mathbf{r})^{3/2}}$$

See Appendix 5.8 for the derivations. This is for plan integral. For the disk integral, we can still get the analytical formula, denoted by $\Phi_w(\mathbf{r},q_0)$, which is a analytical function of q_0. Then if we have lots of DWI samples, the estimation result proposed in [Tristán-Vega et al., 2010] will approximate to $\Phi_w(\mathbf{r},q_0)$, because the data follows mono-exponential decay.

Let's assume the tensor $\mathbf{D}$ above does not change during $[q_0/2, \infty]$, and changes arbitrarily in $[0, q_0/2]$. For example,

$$\mathbf{D}(q) = \begin{cases} \mathbf{D}_0 & \text{if } q \in [q_0/2, \infty) \\ \frac{2q}{q_0}\mathbf{D}_0 + (1 - \frac{2q}{q_0})\mathbf{D}_{iso} & \text{if } q \in [0, q_0/2) \end{cases} \qquad \mathbf{D}_{iso} = \frac{1}{3}\,\text{Trace}(\mathbf{D})\mathbf{I}$$

where $\mathbf{I}$ is the identity matrix. It definitely satisfies the assumption of mono-exponential decay around q_0. In this case $\mathbf{D}(q_0) = \mathbf{D}_0$, $\mathbf{D}(0) = \mathbf{D}_{iso}$, $\mathbf{D}(q)$ is continuous in $[0, \infty)$. Let $\Phi'_w(\mathbf{r}, q_0)$ denote the ground truth for disk integration in this case, which obviously does not equal to $\Phi_w(\mathbf{r}, q_0)$ in Gaussian case. Now if we have lots of DWIs without noise, does the estimated ODF $\widetilde{\Phi'_w}(\mathbf{r})$ tend to the ground truth $\Phi'_w(\mathbf{r}, q_0)$? If the assumption is correct, the answer will be yes. However, the answer is NO. Because in this non-Gaussian case $E(\mathbf{q})$ around q_0 is the same as $E(\mathbf{q})$ in Gaussian case, $\widetilde{\Phi'_w}(\mathbf{r})$ will be $\widetilde{\Phi'_w}(\mathbf{r}, q_0) = \Phi_w(\mathbf{r}, q_0)$ from the proposed method in [Tristán-Vega et al., 2010], however in the non-Gaussian case $\Phi'_w(\mathbf{r}, q_0) \neq \Phi_w(\mathbf{r}, q_0)$.

Explanation of Results. As we have shown above, the assumption used in [Tristán-Vega et al., 2010] is actually that $E(\mathbf{q})$ follows the mono-exponential decay inside the disk with radius q_0 and $E(\mathbf{q}) = 0$ outside the disk. Note the results of disk integration in [Tristán-Vega et al., 2010] slightly outperformed the plane integration in [Aganj et al., 2010b]. If there is no noise, $E(\mathbf{q}) = 0$ for $q > q_0$ brings modeling error. However in practice the estimated $E(\mathbf{q})$ from limited number of DWIs with noise may be more noisy in the area with higher q value, thus forcing $E(\mathbf{q}) = 0$ when $q > q_0$ may increase the robustness of the estimation. However, the final results from disk integration are dependent on the noise level and the q_0 value. If q_0 is small, the estimated ODF may be much smooth due to the incomplete integration. If the SNR is high enough, the modeling error of forcing $E(\mathbf{q}) = 0$ outside the disk may have more effect than the gained robustness.

Part II

Contributions

CHAPTER 5

ANALYTICAL SPHERICAL POLAR FOURIER IMAGING

"Fourier is a mathematical poem."

– William Thomson

"I have tried to avoid long numerical computations..."

– David Hilbert

Contents

OVERVIEW

Based on the analysis in section 4.3, the sHARDI methods like QBI, exact QBI and DOT always make strong assumptions in radial part of the diffusion signal, and represent the signal or a specific function of the signal in $\mathbb{S}^2$ by symmetric real spherical harmonic (SH) basis. In most published works, QBI was referred as an model-free method, however QBI assumes a delta function decay of the signal. Exact QBI and DOT assume mono-exponential decay which can not be assumed only in a local sense as discussed in Appendix 4.7. The assumptions on radial decay conflicts with experimental MR signal, e.g. the radial decay of diffusion signal from erythrocytes in Fig. 5.1(a). For the data in different shell, sHARDI methods obtain different results, which means they difficultly handle multiple shell data. In order to avoid the strong assumption in such kind of method, we may use 3D basis with both radial basis and spherical basis to represent signals. In this chapter, we introduce such a method using 3D basis, named Spherical Polar Fourier Imaging (SPFI). SPFI represents the signal with Spherical Polar Fourier (SPF) basis and obtains analytically EAP and its various features like ODFs. Our work shows that it is easy to find an analytical ODF and EAP estimation method, and we propose some possible estimation methods in this chapter. To elucidate the relation, similarities and differences between methods, we propose a general Analytical Fourier Transform in Spherical Coordinate (AFT-SC) framework to incorporate sHARDI and mHARDI methods. We also propose some criteria for evaluating different basis in different method.

Organization of this chapter:

First, the previous works by Dr. Assemlal on SPFI were reviewed in subsection 5.1.1, which includes the SPF basis, the least square estimation with regularization, the numerical SPFI to obtain EAP and its features. Second, the analytical forms for EAP, two kinds of ODFs and three scalar indices are given in subsection 5.2.2. Proposition 5.1 proposed the analytical form of Fourier dual SPF basis (dSPF) which is the Fourier transform of SPF basis. Proposition 5.2 shows that EAP is represented under dSPF basis with the same coefficients of signal under SPF basis. Proposition 5.3 and 5.4 give the analytical form for two kinds of ODFs. Some implementation issues are discussed in section 5.1.3 which includes three independent steps in Table 5.1. Two ways for scale estimation are proposed based on the typical ADC value and fitting the GHOT model in Theorem 5.1. The prior $E(0) = 1$ is proposed to be considered in estimation, which improves largely the estimation results. Proposition 5.5 shows the effect of diffusion time τ for the final results. In section 5.2.2, we propose a general framework, named Analytical Fourier Transform in Spherical Coordinate (AFT-SC), which incorporates most widely used sHARDI and mHARDI methods, i.e. QBI, exact QBI, DOT, DPI, SHORE and SPFI. See Table 5.2. For QBI, we propose another novel derivation of ODF in analytical QBI. We also propose the EAP in QBI and demonstrate why EAP and ODF by Wedeen in QBI model are impractical. For DOT, exact QBI and DPI, we proposed several variants to avoid their limitations. For SHORE, we propose the correct form of SHO-3D basis and the analytical forms of EAP and ODFs in proposition 5.6 and 5.7. Section 5.3 proposes some criteria for theoretically comparing different basis in different method. These criteria include completeness, representability, separation of spherical and radial parts, orthogonality. See Table 5.3. More comparisons between SHO-3D basis and SPF basis are given

in Lemma 5.2, Theorem 5.2 and 5.3. The sHARDI and mHARDI methods are validated and compared by synthetic data, phantom data and real data in section 5.4.

5.1 Analytical Spherical Polar Fourier Imaging (SPFI)

5.1.1 Assemlal's work: Spherical Polar Fourier Basis, Least Square Estimation, numerical SPFI

Spherical Polar Fourier Basis

Spherical Polar Fourier Imaging (SPFI) was first proposed by Dr. Assemlal in [Assemlal et al., 2008, 2009a; Assemlal, 2010]. It represents the diffusion signal $E(\mathbf{q})$ with Spherical Polar Fourier basis denoted by $B_{nlm}^{\mathrm{SPF}}(\mathbf{q}|\zeta)$ in Eq. (5.1), where $Y_l^m(\mathbf{u})$ is the symmetric real SHs with even order l, degree m, $G_n(q|\zeta)$ is the Gaussian–Laguerre function defined in Eq. (5.2), and ζ is the scale parameter.

$$E(\mathbf{q}) = \sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm} B_{nlm}^{\mathrm{SPF}}(\mathbf{q}|\zeta) \qquad B_{nlm}^{\mathrm{SPF}}(\mathbf{q}|\zeta) = G_n(q|\zeta) Y_l^m(\mathbf{u}) \tag{5.1}$$

$$G_n(q|\zeta) = \kappa_n(\zeta)\exp\left(-\frac{q^2}{2\zeta}\right) L_n^{1/2}\left(\frac{q^2}{\zeta}\right) \qquad \kappa_n(\zeta) = \left[\frac{2}{\zeta^{3/2}}\frac{n!}{\Gamma(n+3/2)}\right]^{1/2} \tag{5.2}$$

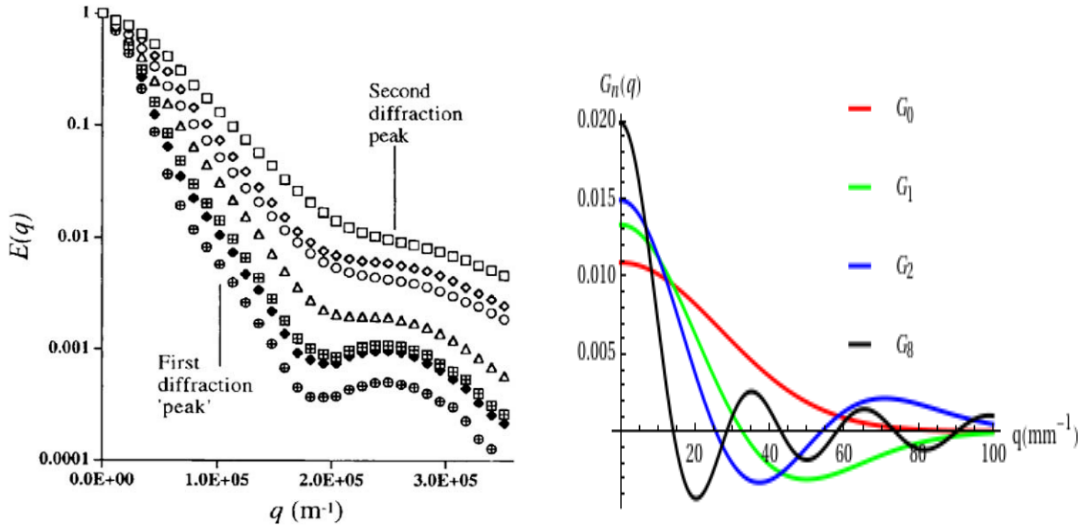
For a given ζ, $\{G_n(q|\zeta)\}$ forms an orthonormal basis in $\mathbb{L}^2([0,\infty), q^2)$, i.e.

$$\int_0^\infty G_n(q|\zeta) G_{n'}(q|\zeta) q^2 \mathrm{d}q = \delta_{nn'} \tag{5.3}$$

because $\kappa_n(\zeta) L_n^{1/2}(\frac{q^2}{\zeta})$ is an orthonormal basis in $\mathbb{L}^2([0,\infty), q^2 e^{-\frac{q^2}{\zeta}})$. Thus the SPF basis function set $\{G_n(q|\zeta)Y_l^m(\mathbf{u})\}$ forms an orthonormal basis in $\mathbb{L}^2(\mathbb{R}^3, 1)$. So $E(\mathbf{q})$ can be always represented well by SPF basis with large enough N and L. If an appropriate ζ is chosen, N can be chosen as a small integer. An inappropriate ζ needs a large N to represent $E(\mathbf{q})$. ζ needs to be set by experience or estimated from data. How to set ζ will be discussed in Section 5.1.3. Fig. 5.1(b) shows the first several basis functions $\{G_n(q|\zeta)\}$ when an empirical value of $\zeta = 714.29 mm^{-2}$ is used. The low order basis functions exhibit Gaussian attenuation behavior, and the higher order basis functions capture oscillating components of the diffusion signal. Fig. 5.1(b) demonstrates the radial basis is appropriate to represent the diffusion signal decay which is a Gaussian like function shown in Fig. 5.1(a).

The SPF basis $B_{nlm}^{\mathrm{SPF}}(\mathbf{q}|\zeta)$ in Eq. (5.1) is actually the SHO-3D basis in Eq. (2.69) when $l = 0$ in radial part. The representation of $E(\mathbf{q})$ using SPF basis and SHO-3D basis will be theoretically analyzed in Section 5.3.

Compared to SHs normally used in QBI and other sHARDI methods, SPF basis is a orthonormal basis in whole $\mathbb{R}^3$ by considering radial basis $G_n(q)$, which avoids strong assumptions in radial decay of $E(\mathbf{q})$. That is the main merit of representing $E(\mathbf{q})$ using SPF basis.



(a) Diffusion signal decay curves for water in suspensions of human erythrocytes as a function of q value at different hematocrit [Kuchel et al., 1997].

(b) The radial parts of several SPF basis functions with scale parameter $\zeta = 714.29mm^{-2}$.

Figure 5.1: Comparison of an experimental plot of MR signal decay (a) and the radial part of SPF basis (b).

Least Square Estimation with Regularization

Because SPF basis is a 3D orthonormal basis, if we know the continuous representation of $E(\mathbf{q})$ in whole $\mathbf{q}$-space, then the coefficients $\{a_{nlm}\}$ will be

$$a_{nlm} = \langle E(\mathbf{q}), B_{nlm}^{\text{SPF}}(\mathbf{q}|\zeta)\rangle = \int_{\mathbb{R}^3} E(\mathbf{q})G_n(q|\zeta)Y_l^m(\mathbf{u})\mathrm{d}\mathbf{q}$$

which is the solution of the following continuous least square estimation problem:

$$\min_{\{a_{nlm}\}} \int_{\mathbb{R}^3} \left(E(\mathbf{q}) - \sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm}G_n(q|\zeta)Y_l^m(\mathbf{u})\right)^2 \mathrm{d}\mathbf{q} \tag{5.4}$$

If we know exhaustive number of samples of $E(\mathbf{q})$, this inner product can be approximated by numerical integration. Then the approximated signal will be

$$\tilde{E}(\mathbf{q}) = \sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm}G_n(q|\zeta)Y_l^m(\mathbf{u}) \tag{5.5}$$

Then the approximation error, i.e. MSE, $\int_{\mathbb{R}^3}\left(E(\mathbf{q}) - \tilde{E}(\mathbf{q})\right)^2\mathrm{d}\mathbf{q}$ will converge to zero as N and L increase. So N and L should be as large as possible.

However in practice we do not know the continuous representation of $E(\mathbf{q})$, and we do not have exhaustive number of samples. Our goal is to estimate $\{a_{nlm}\}$ from limited number, say N_s, of samples $\{E_i\}_{i=1}^{N_s}$. The samples $\{E_i\}$ are normally contaminated by Rician noise and other acquisition noise, which makes the estimation more difficult. In

practice we consider the following classical discrete least square minimization problem with given N and L:

$$\min_{\{a_{nlm}\}} \sum_{i=1}^{N_s} \left(E_i - \sum_{n=0}^{N} \sum_{l=0}^{L} \sum_{m=-l}^{l} a_{nlm} G_n(q_i|\zeta) Y_l^m(\mathbf{u}_i) \right)^2$$

The above problem can be written in matrix formulation. Denote the coefficient vector by $\mathbf{A} = (a_{000}, \cdots, a_{NLL})^T$, the signal sample vector by $\mathbf{E} = (E_1, \cdots, E_{N_s})^T$, the basis matrix by M_{SPF} in Eq. (5.6) which is a N_s by $(N+1)(L+1)(L+2)/2$ matrix.

$$M_{SPF} = \begin{pmatrix} G_0(q_1|\zeta) Y_0^0(\mathbf{u}_1) & \cdots & G_N(q_1|\zeta) Y_L^L(\mathbf{u}_1) \\ \vdots & \ddots & \vdots \\ G_0(q_{N_s}|\zeta) Y_0^0(\mathbf{u}_{N_s}) & \cdots & G_N(q_{N_s}|\zeta) Y_L^L(\mathbf{u}_{N_s}) \end{pmatrix}_{N_s \times (N+1)(L+1)(L+2)/2} \tag{5.6}$$

Then the solution of the discrete least square minimization is

$$\mathbf{A} = (M_{SPF}^T M_{SPF})^{-1} M_{SPF}^T \mathbf{E}$$

This kind of least square has been already used extensively in sHARDI methods [Alexander et al., 2002; Frank, 2002; Özarslan and Mareci, 2003].

Since the DWI data are quite noisy especially at high b values, it is better to add a regularization term in least square estimation. Laplace-Beltrami regularization has been successfully used in QBI which shows significant improvement than no regularization and simple Tikhonov regularization [Descoteaux et al., 2007]. It has been reported that Laplace-Beltrami regularization is also very useful in other methods on spherical data [Descoteaux et al., 2006; Prckovska et al., 2008; Tristán-Vega et al., 2010]. Thus Dr. Assemlal propose a similar regularization term as $\mathbf{A}^T \Lambda \mathbf{A}$, where Λ is a $(N+1)(L+1)(L+2)/2$ by $(N+1)(L+1)(L+2)/2$ diagonal matrix with the diagonal element as $\Lambda_{nlm} = \lambda_l l^2(l+1)^2 + \lambda_n n^2 (n+1)^2$.

$$\mathbf{\Lambda} = \begin{pmatrix} \Lambda_{000} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Lambda_{NLL} \end{pmatrix}, \quad \Lambda_{nlm} = \lambda_l l^2(l+1)^2 + \lambda_n n^2 (n+1)^2 \tag{5.7}$$

The regularization has two parts wighted by λ_l and λ_n. The spherical part $\lambda_l l^2(l+1)^2$ is just the Laplace-Beltrami regularization in spherical basis $\{Y_l^m(\mathbf{u})\}$, which penalizes high frequency of spherical basis. The radial part $\lambda_m n^2(n+1)^2$ penalizes high order of radial basis which is analogous to Laplace-Beltrami regularization in spherical part.

Thus the final cost function in least square estimation with regularization is

$$\min_{\{a_{nlm}\}} \sum_{i=1}^{N_s} \left(E_i - \sum_{n=0}^{N} \sum_{l=0}^{L} \sum_{m=-l}^{l} a_{nlm} G_n(q_i|\zeta) Y_l^m(\mathbf{u}_i) \right)^2 + \sum_{n=0}^{N} \sum_{l=0}^{L} \sum_{m=-l}^{l} a_{nlm}^2 \Lambda_{nlm} = \min_{\mathbf{A}} \|\mathbf{E} - M_{SPF}\mathbf{A}\|^2 + \mathbf{A}^T \Lambda \mathbf{A} \tag{5.8}$$

Then the solution is

$$\mathbf{A} = (M_{SPF}^T M_{SPF} + \Lambda)^{-1} M_{SPF}^T \mathbf{E} \tag{5.9}$$

Dr. Assemlal showed in [Assemlal et al., 2009a] that diffusion signal $E(\mathbf{q})$ can be reconstructed well by the least square estimation with the only first several SPF basis functions if appropriate ζ, λ_l and λ_n are set.

Dr. Assemlal also proposed another nonlinear robust estimation which considers the Rician noise [Assemlal et al., 2009a]. In this chapter, we are only interested in the least square method with regularization, because least square method is also used in QBI [Anderson, 2005; Hess et al., 2006; Descoteaux et al., 2007], exact QBI [Canales-Rodrıguez et al., 2009; Aganj et al., 2010b; Tristán-Vega et al., 2009; Tristán-Vega et al., 2010], DOT [Özarslan et al., 2006] and DPI [Descoteaux et al., 2009, 2010]. It will be a fair comparison in experimental part if all methods use the least square estimation. Moreover, we suggest that the Rician correction can be performed directly on the DWI data as a pre-processing step [Descoteaux et al., 2008b; Tristán-Vega and Aja-Fernández, 2010].

Numerical SPFI

After the coefficient $\mathbf{A}$ is estimated, we have a continuous representation of the signal in Eq. (5.5). However, since the diffusion signal $E(\mathbf{q})$ in $\mathbf{q}$-space does not explicitly shows much meaningful biological information on surrounding tissues, we need to extract EAP or its feature from $E(\mathbf{q})$.

Dr. Assemlal proposed an inner product method to estimate various EAP features [Assemlal et al., 2009a; Assemlal, 2010]. He represented the EAP feature $\mathcal{G}(\mathbf{k})$ in a general form as the inner product between EAP $P(\mathbf{R})$ and a given $H_{\mathbf{k}}(\mathbf{R})$, where $\mathbf{k} \in \mathcal{M}$ and the domain $\mathcal{M}$ can be $\mathbb{R}^1$, $\mathbb{S}^2$ or $\mathbb{R}^3$.

$$\mathcal{G}(\mathbf{k}) = \int_{\mathbb{R}^3} P(\mathbf{R}) H_{\mathbf{k}}(\mathbf{R}) \mathrm{d}\mathbf{R} \tag{5.10}$$

various EAP features can be represented in this form. For example, the ODF by Tuch can be represented as

$$\Phi_t(\mathbf{k}) = \int_0^\infty P(R\mathbf{k}) \mathrm{d}R = \int_{\mathbb{R}^3} P(R\mathbf{r}) \delta(1 - \mathbf{k}^T\mathbf{r}) \mathrm{d}\mathbf{R}$$

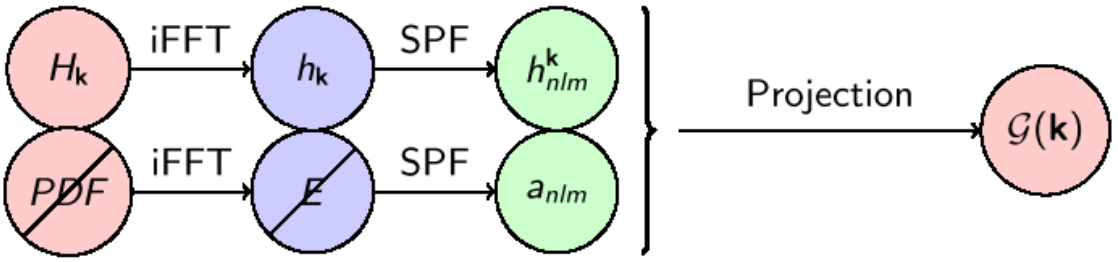
Thus $H_{\mathbf{k}}(\mathbf{R}) = \delta(1 - \mathbf{kr})$ for $\Phi_t(\mathbf{k})$. The Fourier transform of the kernel $H_{\mathbf{k}}(\mathbf{R})$ in $\mathbf{R}$-space is $h_{\mathbf{k}}(\mathbf{q})$ in $\mathbf{q}$-space. If we can calculate the SPF coefficients $\{h^{\mathbf{k}}_{nlm}\}$ of $h_{\mathbf{k}}(\mathbf{q})$, then based on the property of Fourier transform and the orthonormality of SPF basis, we have

$$\mathcal{G}(\mathbf{k}) = \int_{\mathbb{R}^3} P(\mathbf{R}) H_{\mathbf{k}}(\mathbf{R}) \mathrm{d}\mathbf{R} = \int_{\mathbb{R}^3} E(\mathbf{q}) h_{\mathbf{k}}(\mathbf{q}) \mathrm{d}\mathbf{q} = \sum_{n=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_{nlm} h^{\mathbf{k}}_{nlm} \tag{5.11}$$

Note the summation is up to infinity, because normally the kernel with delta functions cannot be represented by finite SPF basis. However, if $E(\mathbf{q})$ is represented by SPF with N and L, the summation is up to N and L based on the orthonormality of SPF basis. The whole process of this inner product method can be summarized in Fig. 5.2.

Since the coefficients $\{h^{\mathbf{k}}_{nlm}\}$ need to be calculated for each $\mathbf{k} \in \mathcal{M}$, this method is time consuming. For $\mathbf{k} \in \mathbb{S}^2$ or $\mathbb{R}^3$, Dr. Assemlal proposed to calculate the coefficients in one direction, e.g. z-axis, then the coefficients in other directions can be analytically calculated from the coefficients this fixed direction based on the rotation property of spherical harmonics [Assemlal et al., 2009a]. This implementation significantly accelerates the estimation process.

Figure 5.2: Overview of the inner product method for the computation of an EAP feature $\mathcal{G}(\mathbf{k})$, where iFFT means inverse Fast Fourier Transform, which can also be solved analytically in some kinds of kernels [Assemlal et al., 2009a; Assemlal, 2010]. This figure is from the slides by Dr. Assemlal for his thesis defense.



In order to calculate the coefficients $h^{\mathbf{k}}_{nlm}$ of the kernel $h_{\mathbf{k}}(\mathbf{q})$ in a fixed direction, [Assemlal et al., 2009a] proposed to use numerical integration, while [Assemlal, 2010] analytically found the coefficients of some kinds of kernels, e.g. RTO, exact ODF by Tuch $\Phi_t(\mathbf{r})$, ODF by Tuch using FRT. However these formulae of the analytical kernels in [Assemlal, 2010] have some problems which makes the final EAP features are incorrect. Please see the following subsection for the details.

This inner product method proposed by Dr. Assemlal has the following drawbacks.

- For the EAP $P(\mathbf{R})$, this method does not give a continuous mathematical formula to represent $P(\mathbf{R})$ in $\mathbf{R}$-space. The coefficients of the EAP profile $P(R_0\mathbf{r})$ with fixed $R = R_0$ and fixed direction $\mathbf{r} = \mathbf{k}$ still need to be calculated in numerical integration. After we estimated EAP profile $P(R_0\mathbf{r})$, for another radius $R = R_1$, the coefficients of the EAP profile $P(R_1\mathbf{r})$ need to be calculated again via the numerical integration.
- For ODFs, the analytical results of exact ODF by Tuch $\Phi_t(\mathbf{r})$ and the original ODF by Tuch using FRT in [Assemlal, 2010] are problematic, and [Assemlal, 2010] did not give the analytical result for ODF by Wedeen $\Phi_w(\mathbf{r})$. Actually the kernel $h_{\mathbf{k}}(\mathbf{q})$ of $\Phi_w(\mathbf{r})$ is so complex that $\Phi_w(\mathbf{r})$ can not be represented as an simple inner product as $\int_{\mathbb{R}^3} E(\mathbf{q})h_{\mathbf{k}}(\mathbf{q})\mathrm{d}\mathbf{q}$. See the following subsection for details.
- In general, the kernel $h_{\mathbf{k}}(\mathbf{q})$ is not easy to be analytically obtained from $H_{\mathbf{k}}(\mathbf{R})$, which is the main reason why some important analytical formulae of $h_{\mathbf{k}}(\mathbf{q})$ and its SPF coefficients are wrong in [Assemlal, 2010].

5.1.2 Analytical Spherical Polar Fourier Imaging

After the SPF coefficients $\{a_{nlm}\}$ for $E(\mathbf{q})$ are estimated, the next step is to estimate the EAP and its features from the pre-estimated $\{a_{nlm}\}$. Dr. Assemlal proposed a good least square method in the estimation of SPF coefficients, and an inner product method in the estimation of the EAP and its features, where the coefficients for kernels need to be calculated numerically.

In this subsection, we propose the analytical results of the EAP and its features from pre-estimated SPF coefficients.

Fourier dual SPF (dSPF) basis and Analytical Form of EAP profile

Theoretically, since $E(\mathbf{q})$ is fully represented by $\{a_{nlm}\}$ and is related with $P(\mathbf{R})$ by Fourier transform, the EAP $P(\mathbf{R})$ and its various features are still fully determined by $\{a_{nlm}\}$. We proved that there are some linear, analytical transforms to obtain the EAP $P(\mathbf{R})$ and its various features from the pre-estimated $\{a_{nlm}\}$. Our method is close in spirit to the methods in DOT [Özarslan et al., 2006] and DPI [Descoteaux et al., 2009]. Adding a strong assumption (in DOT) or choosing a good representation of $E(\mathbf{q})$ (in DPI, SHORE and SPFI) will dramatically simplify the Fourier transform in Eq. (4.24). By considering the plane wave equation in Eq. (5.12), which is Eq. (2.53) with $\mathbf{q}$ vector and $\mathbf{R}$ vector,

$$\cos(2\pi\mathbf{q}^T\mathbf{R}) = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (-1)^{l/2} j_l(2\pi qR) Y_l^m(\mathbf{u}) Y_l^m(\mathbf{r}), \tag{5.12}$$

we first propose analytically the Fourier dual SPF (dSPF) basis which is the Fourier transform of SPF basis.

Proposition 5.1 (Fourier dual SPF basis). *The Fourier transform of $B_{nlm}^{SPF}(\mathbf{q}|\zeta)$, named as Fourier dual Spherical Polar Fourier (dSPF) basis, is $B_{nlm}^{dSPF}(\mathbf{R}|\zeta) = F_{nl}(R|\zeta)Y_l^m(\mathbf{r})$, i.e.*

$$B_{nlm}^{dSPF}(\mathbf{R}|\zeta) \overset{\text{def}}{=} \mathcal{F}\{G_n(q|\zeta)Y_l^m(\mathbf{u})\} = F_{nl}(R|\zeta)Y_l^m(\mathbf{r}), \quad \mathbf{R} = R\mathbf{r} \tag{5.13}$$

where

$$F_{nl}(R|\zeta) = \frac{\zeta^{0.5l+1.5}\pi^{l+1.5}R^l\kappa_n(\zeta)}{(-1)^{l/2}\Gamma(l+1.5)} \sum_{i=0}^{n} \binom{n+0.5}{n-i} \frac{(-1)^i}{i!} 2^{0.5l+i+1.5}\Gamma(0.5l+i+1.5)\,{}_1F_1\left(\frac{2i+l+3}{2}; l+\frac{3}{2}; -2\pi^2R^2\zeta\right) \tag{5.14}$$

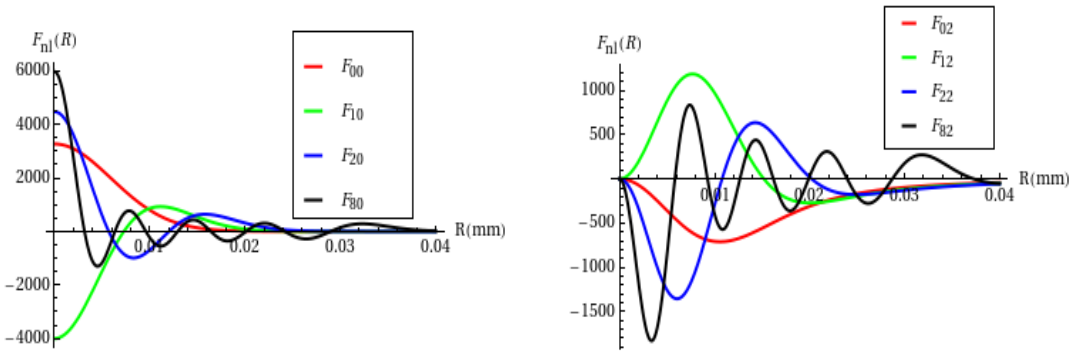
The dSPF basis is an orthonormal basis in real symmetric function subspace in $\mathbb{L}^2(\mathbb{R}^3, 1)$.

The proof can be found in Appendix 5.6. Although the radial part of dSPF basis $F_{nl}(R|\zeta)$ looks complex, it also has a Gaussian-like decay as the radial part of SPF basis. Please see Fig. 5.3 for $F_{nl}(R|\zeta)$ with typical scale $\zeta = 714.29mm^{-2}$. Based on the above proposition, the EAP is represented by dSPF basis with the same coefficients $\{a_{nlm}\}$.

Proposition 5.2 (Analytical EAP represented by dSPF basis). *If $E(\mathbf{q})$ is represented by SPF basis as $E(\mathbf{q}) = \sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm}G_n(q|\zeta)Y_l^m(\mathbf{u})$, then the EAP $P(\mathbf{R})$ is analytically obtained by dSPF basis as*

$$P(\mathbf{R}) = \sum_{n=0}^{N} \sum_{l=0}^{L} \sum_{m=-l}^{l} a_{nlm}F_{nl}(R|\zeta)Y_l^m(\mathbf{r}), \quad \mathbf{R} = R\mathbf{r} \tag{5.15}$$

where $F_{nl}(R|\zeta)$ is in Eq. (5.14).



(a) $n = 0, 1, 2, 8$, $l = 0$ (b) $n = 0, 1, 2, 8$, $l = 2$

Figure 5.3: The radial parts of several dSPF basis functions with scale parameter $\zeta = 714.29mm^{-2}$.

Proof. This proposition is a straightforward result of Proposition 5.1, because

$$\begin{aligned}\mathscr{F}\{E(\mathbf{q})\} &= \mathscr{F}\{\sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm}G_n(q|\zeta)Y_l^m(\mathbf{u})\} \\ &= \sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm}\mathscr{F}\{G_n(q|\zeta)Y_l^m(\mathbf{u})\} \\ &= \sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm}F_{nl}(R|\zeta)Y_l^m(\mathbf{r})\end{aligned}$$

□

Thus we have a continuous representation of $P(\mathbf{R})$ represented by dSPF basis with the coefficients $\{a_{nlm}\}$. The EAP profile $P(R_0\mathbf{r})$ with radius R_0 can be represented as

$$P(R_0\mathbf{r}) = \sum_{l=0}^{L}\sum_{m=-l}^{l}\left(\sum_{n=0}^{N} a_{nlm}F_{nl}(R_0|\zeta)\right)Y_l^m(\mathbf{r}) \tag{5.16}$$

which is represented analytically by SHs with the coefficient in Eq. (5.17), denoted by $c_{lm}^P(R_0)$.

$$c_{lm}^P(R_0) = \sum_{n=0}^{N} a_{nlm}F_{nl}(R_0|\zeta) \tag{5.17}$$

Now we obtain the linear transform from $\{a_{nlm}\}$ to $\{c_{lm}\}$ for EAP profile $P(R_0\mathbf{r})$ with given radius $R = R_0$, which could be implemented as an matrix multiplication. Please note the following important differences between SPFI and DOT. Here our transformation is independent with the data, since $f_{nl}(R_0)$ is just dependent on ζ and R_0. Once we give a radius R_0 and the basis scale ζ, we have the transform. While in DOT, the transform is dependent on the ADC value of data. Similarly with the appendix in [Özarslan et al., 2006], here the confluent hypergeometric function ${}_1F_1$ could also be analytically separated into some more simple terms. However in practice we implemented ${}_1F_1$ in C++ with some scientific library, e.g. GSL[1] or std::tr1[2]. If the scale

[1] http://www.gnu.org/software/gsl/
[2] http://en.wikipedia.org/wiki/C++_Technical_Report_1

parameter ζ is set as the same value, the transform matrix just needs to be calculated only once for all voxels. While in DOT, the transform that contains ${}_1F_1$ needs to be calculated in each voxel. In SPFI, $P(R_0\mathbf{r})$ and $P(R_1\mathbf{r})$ with different radii R_0 and R_1 shares the same coefficients $\{a_{nlm}\}$, thus we just need to estimate $\{a_{nlm}\}$ once for different EAP profile with different radius. While in DOT, even if we have estimated the EAP profile $P(R_0\mathbf{r})$, we need to perform the estimation again for another EAP profile $P(R_1\mathbf{r})$ with another radius R_1.

An important similarity with DOT is that if we just choose $N = 0$ in radial part, our transform will be the DOT, which could be seen from Eq. (5.17) and Eq. (5.14). That is understandable because the order 0 of the radial basis follows mono-exponential decay used in DOT. However, in SPFI we should use $N \geq 1$ to describe anisotropic decay, because the SPF basis with $N = 0$ in Eq. (5.1) only can represent isotropic EAP. We will explain it in two aspects later in estimation of $\phi_w(\mathbf{r})$ in the following subsection and the estimation of $\{a_{nlm}\}$ with $E(0) = 1$ consideration in Section 5.1.3.

Analytical Forms of Orientation Distribution Functions

In this subsection we would like to deduce the elegant analytical solutions for these two kinds of ODFs from the estimated SPF coefficients $\{a_{nlm}\}$. It has been shown in subsection 4.3.4 that the radial integral of $P(R\mathbf{r})$ along direction $\mathbf{r}$ in $\mathbf{R}$-space is equivalent to the integration in the plane $\Pi_{\mathbf{r}}$ in $\mathbf{q}$-space which is orthogonal to $\mathbf{r}$ [Aganj et al., 2010b; Tristán-Vega et al., 2010; Canales-Rodrıguez et al., 2009]. See Proposition 4.1 and Eq. (4.64), where Δ_b is the Laplace-Beltrami operator.

We also give the results of the integration in a given disk $\Pi_{\mathbf{r},C}$ with the radius C. In Eq. (4.64), the integration $\Phi_t(\mathbf{r})$ gives the same weight for $E(\mathbf{q})$ with large q and for $E(\mathbf{q})$ with small q. However, if we have the data with several b values, the error of estimated signal $\tilde{E}(\mathbf{q})$ will be small if q is between these b values and will be large if q is larger than all b values. Thus if an approximate C is given, the disk integrations $\Phi_t(\mathbf{r}, C)$ and $\Phi_w(\mathbf{r}, C)$ may have better angular resolution than plane integrations $\Phi_t(\mathbf{r})$ and $\Phi_w(\mathbf{r})$ [Tristán-Vega et al., 2010].

The following two propositions give two kinds of estimations by plane integration and disk integration respectively for $\Phi_t(\mathbf{r})$ and $\Phi_w(\mathbf{r})$. Their proofs are in Appendix 5.6.

Proposition 5.3 (Analytical $\Phi_t(\mathbf{r})$ represented by SH basis). *If $E(\mathbf{q})$ is represented by SPF basis as $E(\mathbf{q}) = \sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm}G_n(q|\zeta)Y_l^m(\mathbf{u})$, then the ODF by Tuch $\Phi_t(\mathbf{r})$ is analytically obtained by SH basis as*

$$\Phi_t(\mathbf{r}) = \sum_{l=0}^{L}\sum_{m=-l}^{l} c_{lm}^{\Phi_t} Y_l^m(\mathbf{r}), \quad c_{lm}^{\Phi_t} = \sum_{n=0}^{N}\sum_{i=0}^{n} \kappa_n(\zeta)\binom{i-0.5}{i}(-1)^{n-i}P_l(0)a_{nlm} \tag{5.18}$$

The disk integration with radius C for $\Phi_t(\mathbf{r})$ is analytically given as

$$\Phi_t(\mathbf{r}, C) = \sum_{l=0}^{L}\sum_{m=-l}^{l} c_{lm}^{\Phi_t}(C) Y_l^m(\mathbf{r}), \quad c_{lm}^{\Phi_t}(C) = \sum_{n=0}^{N}\sum_{i=0}^{n} \kappa_n(\zeta)\binom{n+0.5}{n-i}\frac{(-2)^i}{i!}\gamma(i+1, 0.5C^2/\zeta)P_l(0)a_{nlm} \tag{5.19}$$

Proposition 5.4 (Analytical $\Phi_w(\mathbf{r})$ represented by SH basis). *If $E(\mathbf{q})$ is represented by SPF basis as $E(\mathbf{q}) = \sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm} G_n(q|\zeta) Y_l^m(\mathbf{u})$, then the ODF by Wedeen $\Phi_w(\mathbf{r})$ is analytically obtained by SH basis as*

$$\Phi_w(\mathbf{r}) = \sum_{l=0}^{L}\sum_{m=-l}^{l} c_{lm}^{\Phi_w} Y_l^m(\mathbf{r}), \quad c_{lm}^{\Phi_w} = \frac{1}{\sqrt{4\pi}}\delta_{lm}^{00} - \frac{1}{8\pi}\sum_{n=1}^{N}\sum_{i=1}^{n}(-1)^i \kappa_n(\zeta)\binom{n+0.5}{n-i}\frac{2^i}{i} P_l(0)(-l)(l+1)a_{nlm} \tag{5.20}$$

The disk integration with radius C for $\Phi_w(\mathbf{r})$ is analytically given as

$$\Phi_w(\mathbf{r}, C) = \sum_{l=0}^{L}\sum_{m=-l}^{l} c_{lm}^{\Phi_w}(C) Y_l^m(\mathbf{r}) \tag{5.21}$$

$$c_{lm}^{\Phi_w}(C) = \frac{1}{\sqrt{4\pi}}\delta_{lm}^{00} - \frac{1}{8\pi}\sum_{n=1}^{N}\sum_{i=1}^{n}(-1)^i \kappa_n(\zeta)\binom{n+0.5}{n-i}\frac{2^i}{i!}\gamma(i, 0.5C^2/\zeta) P_l(0)(-l)(l+1)a_{nlm}$$

Now we have two estimators for $\Phi_w(\mathbf{r})$. One is the integration in the whole plane, which is similar with [Aganj et al., 2010b], and the other one is the integration in a given disk, which is similar with [Tristán-Vega et al., 2010]. However, the mono-exponential decay model was assumed during the disk in [Tristán-Vega et al., 2010] as discussed in Appendix 4.7, and in the whole plane in [Aganj et al., 2010b], so that the radial integration can be estimated from single shell data in the circle with radius $q = q_0$. While our method does not need any assumption on the data and it can handle the data in different shells thanks to the radial basis in SPF basis.

Moreover please note three important points in the formulae above.

- *First*, we obtain the exponential integral $E_1(x)$ in the derivation process, but it is negligible. While in [Tristán-Vega et al., 2010], it is indispensable.
- *Second*, Eq. (5.21) and Eq. (5.20) tell us that at least order 1 of $G_n(q|\zeta)$ is needed to represent an anisotropic ODF. That is true because if we just use the radial basis of order zero, i.e. $n = 0$, it is easily seen that the estimated signal is just an isotropic one, which means the estimated ODF is isotropic. Thus we need to use at least $n = 1$ to obtain a reasonable results, which explains why we need to set $N \geq 1$. In order to estimate the coefficients well, it is better to use the data from more than one shell, although our methods can be performed in single shell data. This will be revisited later.
- *Third*, for single shell data, the radius C can be naturally chosen as the q_0 value of the sampled shell [Tristán-Vega et al., 2010]. However, in multiple shell data, how to choose an approximate C for $\Phi_w(\mathbf{r}, C)$ and $\Phi_t(\mathbf{r}, C)$ is still unknown and out of the scope of this thesis. Since the improvement of $\Phi_w(\mathbf{r}, C)$ in [Tristán-Vega et al., 2010] over $\Phi_w(\mathbf{r})$ in [Aganj et al., 2010b] is very subtle [Tristán-Vega et al., 2010], in this thesis we just consider the $\Phi_w(\mathbf{r})$ and $\Phi_t(\mathbf{r})$, not $\Phi_w(\mathbf{r}, C)$, $\Phi_t(\mathbf{r}, C)$.

Analytical Forms of Scalar Indices

Return-To-Origin (RTO) probability. In SFPI, we can easily estimate Return-To-Origin (RTO) probability from Eq. (5.1), or by setting $R_0 = 0$ in Eq. (5.16). These

two ways are equivalent. Considering Eq. (2.83) [Gradshteyn and Ryzhik, 2007] and $\int_{\mathbb{S}^2} Y_l^m(\mathbf{u})\mathrm{d}\mathbf{u} = \sqrt{4\pi}\delta_{lm}^{00}$, we have

$$\begin{aligned} Po &= \int_{\mathbb{R}^3} E(\mathbf{q})\mathrm{d}\mathbf{q} = \int_{\mathbb{R}^3} \sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm}G_n(q|\zeta)Y_l^m(\mathbf{u})q^2\mathrm{d}q\mathrm{d}\mathbf{u} \\ &= \sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm}\left\{\int_0^\infty G_n(q|\zeta)q^2\mathrm{d}q\right\}\left\{\int_{\mathbb{S}^2} Y_l^m(\mathbf{u})\mathrm{d}\mathbf{u}\right\} \\ &= \sqrt{8\pi}\sum_{n=0}^{N}(-1)^n\kappa_n(\zeta)\zeta^{1.5}\frac{\Gamma(n+1.5)}{n!}a_{n00} \\ &= 4\sqrt{\pi}\zeta^{\frac{3}{4}}\sum_{n=0}^{N}(-1)^n\sqrt{\frac{\Gamma(n+1.5)}{n!}}a_{n00} \end{aligned} \tag{5.22}$$

Note if we set $R_0 = 0$ in Eq. (5.16), we obtain the EAP profile $F_{nl}(0|\zeta)Y_l^m(\mathbf{u})$ which is a spherical function represented by SH basis. However, it is easy to verify that

$$F_{nl}(0|\zeta) = 0, \quad \text{if } l > 0$$

which means this EAP profile at $R = 0$ is an isotropic spherical function whose value is the same in Eq. (5.22).

Mean Squared Displacement (MSD). Mean Squared Displacement (MSD) defined in Eq. (4.45) [Wu and Alexander, 2007] is used to describe the average diffusivity. Since $E(\mathbf{q})$ and $P(\mathbf{R})$ are related by Fourier transform in Eq. (4.24), $-\frac{1}{4\pi^2}\Delta E(\mathbf{q})$ and $R^2P(\mathbf{R})$ are also related by Fourier transform. By considering $E(\mathbf{q}) = \sum_{nlm} a_{nlm}G_n(q|\zeta)Y_l^m(\mathbf{u})$ we have

$$\begin{aligned} MSD &= \int_{\mathbb{R}^3} P(\mathbf{R})R^2\mathrm{d}\mathbf{R} = -\frac{1}{4\pi^2}\Delta E(\mathbf{q})\Big|_{\mathbf{q}=0} \\ &= -\frac{1}{4\pi^2}\Big(\frac{1}{q}\frac{\partial^2(qE(\mathbf{q}))}{\partial q^2} + \frac{1}{q^2}\nabla_b E(\mathbf{q})\Big)\Big|_{\mathbf{q}=0} \qquad (5.23) \\ &= -\frac{1}{4\pi^2}\Big(\sum_{nlm} a_{nlm}\Big(\frac{2}{q}G_n'(q|\zeta)Y_l^m(\mathbf{u}) + G_n''(q|\zeta)Y_l^m(\mathbf{u}) + \frac{1}{q^2}G_n(q|\zeta)(-l)(l+1)Y_l^m(\mathbf{u})\Big)\Big)\Big|_{\mathbf{q}=0} \qquad (5.24) \\ &= -\frac{1}{4\pi^2}\Big(\sum_{nlm}(3+\frac{1}{2}(-l)(l+1))a_{nlm}G_n''(q|\zeta)Y_l^m(\mathbf{u})\Big)\Big|_{\mathbf{q}=0} \qquad (5.25) \\ &\simeq \frac{3}{8\pi^{2.5}}\sum_{n=0}^{N} a_{n00}\frac{\kappa_n(\zeta)}{\zeta}(2L_{n-1}^{3/2}(0) + L_n^{1/2}(0)) \qquad (5.26) \end{aligned}$$

From Eq. (5.23) to Eq. (5.24), we use the properties of SH, $\nabla_b Y_l^m(\mathbf{u}) = (-l)(l+1)Y_l^m(\mathbf{u})$. From Eq. (5.24) to Eq. (5.25), we consider $E(0) = 1$, $G_n'(0|\zeta) = 0$ and keep performing the L'Hôpital's rule. Please note that in Eq. (5.25) we actually obtain a spherical function for $q = 0, \forall \mathbf{u} \in \mathbb{S}^2$, while MSD is just a scalar value. If the SPF basis represents $E(\mathbf{q})$ without numerical error, the spherical function should be isotropic with the same value for $\mathbf{u} \in \mathbb{S}^2$. However, our experiments showed that this spherical function is not always isotropic in practice, probably because the proposed MSD estimator uses the second derivative which is not very stable with the estimation error of $\{a_{nlm}\}$. So we use the mean of the spherical function as an approximation of MSD, which is shown in Eq. (5.26), where we denote $L_{-1}^{3/2}(0) = 0$.

Generalized Fractional Anisotropy (GFA) for EAPs. In dMRI, the data in each voxel is normally a coefficient vector which has some complex biological meaning. It is useful to extract some scalar measurements from the complex coefficient vectors. In clinical study, various brain pathologies may be best detected by measuring the diffusion anisotropy. In DTI, the Fractional Anisotropic (FA) of tensor $\mathbf{D}$ is defined as

$$\frac{\sqrt{3}\|\mathbf{D} - \frac{1}{3}\operatorname{Trace}(\mathbf{D})\mathbf{I}\|}{\sqrt{2}\|\mathbf{D}\|}$$

where $\mathbf{I}$ is the identity matrix. FA is actually normalized ℓ_2 distance between the given tensor $\mathbf{D}$ and the **nearest isotropic tensor** $\frac{1}{3}\operatorname{Trace}(\mathbf{D})\mathbf{I}$. In QBI [Tuch, 2004], FA was generalized from tensor to ODF. For a given ODF $\phi(\mathbf{u})$ GFA is defined as $\frac{std(\phi(\mathbf{u}))}{rms(\phi(\mathbf{u}))}$ [Tuch, 2004]. In discrete case, it is given in Eq. (4.61). If we use SH to represent $\phi(\mathbf{u})$, i.e. $\phi(\mathbf{u}) = \sum_{l=0}^{L}\sum_{m=-l}^{l} c_{lm} Y_l^m(\mathbf{u})$, then the GFA is

$$\sqrt{1 - \frac{c_{00}^2}{\sum_{l=0}^{L}\sum_{m=-l}^{l} c_{lm}^2}}$$

which is actually still the normalized ℓ_2 distance between the given ODF $\phi(\mathbf{u})$ and the **nearest isotropic ODF** $c_{00}Y_0^0(\mathbf{u})$. Please note that the GFA for ODFs actually could be directly used for any function defined in $\mathbb{S}^2$. For example, GFA can be used for EAP profiles [Cheng et al., 2010b]. Thus we call it as **GFA for spherical functions**. However, to our knowledge so far there is no counterpart for EAPs which are functions in $\mathbb{R}^3$. Although we can calculate GFA for given EAP profile with radius R_0, this kind of GFA can not describe the anisotropy in whole $\mathbf{R}$-space. Moreover, it is not unique for a given EAP, because different radius R_0 has different GFA. Here we proposed GFA for EAPs, which is defined as the normalized ℓ_2 distance between the given EAP $P(\mathbf{R})$ to **the nearest isotropic EAP**. In general case, the EAP $P(\mathbf{R})$ is isotropic if and only if $P(\mathbf{R}) = F(R)Y_0^0(\mathbf{r})$, where $F(R)$ can be any nonnegative definite function. Thus for any general EAP $P(\mathbf{R})$, its nearest EAP is

$$P_{\text{iso}}(R\mathbf{r}) = \left(\int_{\mathbb{S}^2} P(R\mathbf{r})Y_0^0(\mathbf{r})\mathrm{d}\mathbf{r}\right) Y_0^0(\mathbf{r}) = \frac{1}{4\pi}\int_{\mathbb{S}^2} P(R\mathbf{r})\mathrm{d}\mathbf{r}$$

which is the solution of minimization

$$\min_{F(R)} \|P(\mathbf{R}) - F(R)Y_0^0(\mathbf{u})\|^2$$

Note that the integration of $P_{\text{iso}}(R\mathbf{r})$ is one and $P_{\text{iso}}(R\mathbf{r}) \geq 0$ because $P(R\mathbf{r}) \geq 0$. So $P_{\text{iso}}(R\mathbf{r})$ is indeed an EAP. Then the GFA for EAP $P(\mathbf{R})$ is defined as

$$GFA = \sqrt{1 - \frac{\left\|\left(\int_{\mathbb{S}^2} P(R\mathbf{r})Y_0^0(\mathbf{r})\mathrm{d}\mathbf{r}\right) Y_0^0(\mathbf{r})\right\|^2}{\|P(\mathbf{R})\|^2}} \tag{5.27}$$

where $\|\cdot\|$ is the standard ℓ_2 norm in $\mathbb{L}^2(\mathbb{R}^3, 1)$. The GFA for EAPs is naturally normalized in $[0, 1]$. Mathematically all isotropic tensors/ODFs/EAPs form a linear space (rigorously a convex space because of some biological constraints), the GFA for tensors/ODFs/EAPs is defined as the normalized projection distance. Please note that this concept of GFA for EAPs can have different implementation in different

HARDI methods. That is because in different HARDI method the representation of the nearest isotropic EAP is different. In SPFI, for any given EAP represented as $P(\mathbf{R}) = \sum_{n=0}^{N} \sum_{l=0}^{L} \sum_{m=-l}^{l} a_{nlm} B_{nlm}^{\mathrm{dSPF}}(\mathbf{R})$, its nearest isotropic EAP is

$$\sum_{n=0}^{N} a_{n00} B_{n00}^{\mathrm{dSPF}}(\mathbf{R})$$

So in SPFI the GFA for EAPs is defined as the normalized ℓ_2 distance as

$$GFA = \sqrt{1 - \frac{\sum_{n=0}^{N} a_{n00}^2}{\sum_{n=0}^{N} \sum_{l=0}^{L} \sum_{m=-l}^{l} a_{nlm}^2}} \tag{5.28}$$

5.1.3 Implementation Issues

Three Steps

The implementation of SPFI includes three steps, where we have demonstrated above the analytical forms for the third step.

1. The first step is to estimate the scale parameter ζ in the SPF basis used in the following steps.
2. The second step is to estimate SPF coefficients $\{a_{nlm}\}$ from the signal samples $\{E(\mathbf{q}_i)\}$. These coefficients are also the coefficients of EAP $P(\mathbf{R})$ under dSPF basis. This is the estimation step with estimation error.
3. The third step is the linear analytical transforms proposed above from $\{a_{nlm}\}$ to $\{c_{lm}\}$ of EAP profile and ODFs, and to scalar indices, which is actually independent of the first two steps. The whole estimation error is just from the first two steps, since the third step is analytical and compact.

Note it is possible to incorporate the first two steps into one step which estimates both $\{a_{nlm}\}$ and ζ simultaneously from the signal samples $\{E(\mathbf{q}_i)\}$. However we separate it into two steps, which has good efficiency and property.

Estimation of Scale Parameter

The SPF basis $\{B_{nlm}^{\mathrm{SPF}}(\mathbf{q}|\zeta)\}$ forms an orthonormal basis in the real symmetric function subspace in $\mathbb{L}^2(\mathbb{R}^3, 1)$. If N and L are large enough, we can use SPF to represent the signal with any given representation MSE, no matter how the ζ is. However, for a given representation MSE, an appropriate ζ can let N and L be chosen as small numbers as possible. In practice, the N and L can not be chosen as very large numbers due to the limited number of signal samples with noise. Thus the scale parameter ζ is crucial for estimation results. There are several possible ways to set ζ.

- [Assemlal et al., 2008, 2009a] proposed an experimental strategy for ζ, which is dependent on the radial truncation order N. However, we think the scale parameter should be an intrinsic property of the signal which is just dependent on the signal, not on the basis order.

- [Assemlal et al., 2009b; Assemlal, 2010] considered to set ζ based on the estimated ADC, which we also considered in [Cheng et al., 2010a,b]. Consider an isotropic Gaussian signal as $E(\mathbf{q}) = \exp(-4\pi^2\tau q^2 D)$. Then if we set

$$\zeta = \frac{1}{8\pi^2\tau D} \tag{5.29}$$

$E(\mathbf{q}) = \exp(-\frac{q^2}{2\zeta}) \propto G_0(q|\zeta)$, then we can use only the first basis $B_{000}^{\text{SPF}}(\mathbf{q}|\zeta) = G_0(q|\zeta)Y_0^0(\mathbf{u})$ to represent it. So the idea is to fit the signal with an isotropic tensor model, then set ζ based on the ADC D. Note this setting is not optimal, because it minimizes the following cost function

$$\min_{\zeta} \|E(\mathbf{q}) - a_{000}G_0(q|\zeta)Y_0^0(\mathbf{u})\|^2 = \min_{\zeta} \int_{\mathbb{R}^3} \left(E(\mathbf{q}) - \exp(-\frac{q^2}{2\zeta})\right)^2 \mathrm{d}\mathbf{q} \tag{5.30}$$

where $a_{000}Y_0^0(\mathbf{u}) = 1$ because $E(0) = 1$. However if $E(\mathbf{q})$ can be represented without error by

$$E(\mathbf{q}) = \sum_{n=0}^{N_*} \sum_{l=0}^{L_*} \sum_{m=-l}^{l} a_{nlm}^* G_n(q|\zeta_*)Y_l^m(\mathbf{u}) \tag{5.31}$$

and $a_{n00} \neq 0$ for at least one $n > 0$, then the solution of Eq. (5.30) is not ζ_*. That is because we do not consider infinite terms in Eq. (5.30).

- We proposed in [Cheng et al., 2010a,b] to set the scale ζ as an typical ADC in brain, i.e.

$$D_0 = 0.7 \times 10^{-3} mm^2/s, \quad \zeta = \frac{1}{8\pi^2\tau D_0} \tag{5.32}$$

Then if $\tau = \frac{1}{4\pi^2} = 0.02533s$ the typical scale ζ is

$$\zeta = \frac{1}{8\pi^2\tau D_0} = 714.29 mm^{-2} \tag{5.33}$$

which has been used in Fig. 5.1(b) and will be used in Chapter 6. If ζ is set by Eq. (5.32), every voxel shares the same scale and the same basis set, which is more efficient and makes Riemannian framework in Chapter 6 computable.

- Another possible solution is to minimize the following cost function with respect to both a_{nlm} and ζ for chosen N and L.

$$\min_{\{a_{nlm}\},\zeta} \left\| E(\mathbf{q}) - \sum_{n=0}^{N} \sum_{l=0}^{L} \sum_{m=-l}^{l} a_{nlm}G_{nlm}(q|\zeta)Y_l^m(\mathbf{u}) \right\|^2 \tag{5.34}$$

This is to incorporate the first step into the second step. The cost function can be minimized via an iterative process.

1. Set an initialization for ζ, which can be set from a typical D_0 or from the fitting of the isotropic tensor model.
2. Loop the iteration process until convergence. When ζ is fixed, the coefficients of $\{a_{nlm}\}$ are analytically represented by matrix multiplication. When $\{a_{nlm}\}$ are fixed, ζ can be updated by a classical gradient descent method.

The final solution of ζ is not guaranteed to be the global minimum, and even the global minimum of above minimization is still not guaranteed to be the true value ζ_* unless N_* in Eq. (5.31) is less than N. Moreover, the estimation is time consuming since the iterative process needs to be performed in each voxel.

In order to estimate the scale value, we first define the Spherical Polar Non-Polynomial (SPNP) basis and Generalized High Order Tensor (GHOT) model.

Definition 5.1 (Spherical Polar Non-Polynomial Basis and Generalized High Order Tensor Model). *Spherical Polar Non-Polynomial (SPNP) basis is defined as*

$$B_{nlm}^{SPNP}(\mathbf{q}|\zeta) = \exp\left(-\frac{q^2}{2\zeta}\right)\left(\frac{q^2}{\zeta}\right)^n Y_l^m(\mathbf{u}) \tag{5.35}$$

Generalized High Order Tensor (GHOT) model is defined as

$$E(\mathbf{q}) = \exp\left(-\sum_{n=1}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} b_{nlm}\left(\frac{q^2}{\zeta}\right)^n Y_l^m(\mathbf{u})\right) \tag{5.36}$$

*where $\{Y_l^m(\mathbf{u})\}$ are the real symmetric SHs, ζ is the scale parameter. For given diffusion time τ, the **pseudo-ADC** of GHOT model denoted by D_p is defined as the coefficient of $4\pi^2\tau q^2$, i.e.*

$$D_p = \frac{b_{100}}{8\pi^{5/2}\tau\zeta} \tag{5.37}$$

Please note the following things for SPNP basis and GHOT model.

- $\{B_{nlm}^{\mathrm{SPNP}}(\mathbf{q}|\zeta)\}$ is a basis of the symmetric real function subspace in $\mathbb{L}^2(\mathbb{R}^3, 1)$, because $\left\{\exp\left(-\frac{q^2}{2\zeta}\right)\left(\frac{q^2}{\zeta}\right)^n\right\}$ is a basis in $\mathbb{L}^2([0,\infty), q^2)$ and $\{Y_l^m(\mathbf{u})\}$ is an orthonormal basis in symmetric real function subspace of $\mathbb{L}^2(\mathbb{S}^2, 1)$. Thus for any given $E(\mathbf{q})$, the representation error converges to zero as N and L increase.
- For given N, the function space spanned by polynomial basis $\{q^{2n}\}_{0\le n\le N}$ restricted in $[0,\infty)$ is the same space spanned by $\{L_n^{1/2}(\frac{q^2}{\zeta})\}_{0\le n\le N}$, $\forall\zeta$. Thus for given N and L, the SPNP basis set $\{B_{nlm}^{\mathrm{SPNP}}(\mathbf{q}|\zeta)\}_{0\le n\le N, 0\le l\le L}$ forms the same space as the SPF basis set $\{B_{nlm}^{\mathrm{SPF}}(\mathbf{q}|\zeta)\}_{0\le n\le N, 0\le l\le L}$.
- In GHOT model, the index n start from 1, because the prior $E(0) = 1$ means $b_{0lm} = 0$ if n starts from 0. When $N = 1$, GHOT becomes the HOT model with tensor order L, because the space spanned by HOT basis with order L is the same space spanned by SHs with orders no more than L. See Theorem 4.1. When $N = 1$ and $L = 2$, GHOT model is the DTI model.
- In GHOT model, the ADC is represented by the basis $\left\{\left(\frac{q^2}{\zeta}\right)^n Y_l^m(\mathbf{u})\right\}$ that is a basis of $\mathbb{L}^2(\mathbb{R}^3, 1)$. Thus GHOT representation is model-free, which can represent any $E(\mathbf{q}) \in \mathbb{L}^2(\mathbb{R}^3, 1)$. The pseudo-ADC is the isotropic quadratic part of ADC.
- Actually the scale parameter ζ can be put into $\{b_{nlm}\}$ in both SPNP basis and GHOT model. ζ is only set for reducing the numerical error when estimating the coefficients of polynomial basis. For the estimation of the coefficients from DWI samples, the scale ζ is set as $0.5q_{max}^2$ in practice, where q_{max} is the maximal q value of the measured $E(\mathbf{q})$.

With the definition of GHOT model, we have the following theorem which is useful to set the scale parameter ζ in SPF basis. See Appendix 5.6 for the proof.

Theorem 5.1 (Set the scale in SPF basis via GHOT model). *When using SPF basis to fit the diffusion signal attenuation represented by GHOT model as* $E(\mathbf{q}) = \exp\left(-\sum_{n=1}^{\infty}\sum_{l=0}^{\infty}\sum_{m=-l}^{l} b_{nlm}\left(\frac{q^2}{\zeta_1}\right)^n Y_l^m(\mathbf{u})\right)$*, if* $b_{100} > 0$*, the optimal scale parameter* ζ *of SPF basis for any given* L *and large enough* N *is*

$$\zeta = \frac{\zeta_1\sqrt{\pi}}{b_{100}} = \frac{1}{8\pi^2\tau D_p} \tag{5.38}$$

In other words, define the representation MSE for any given L *and* N *as*

$$MSE_{NL}(\zeta) = \left\| E(\mathbf{q}) - \sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm} G_n(q|\zeta) Y_l^m(\mathbf{u}) \right\|^2 = \int_{\mathbb{R}^3} \left(E(\mathbf{q}) - \sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm} G_n(q|\zeta) Y_l^m(\mathbf{u}) \right)^2 \mathrm{d}\mathbf{q} \tag{5.39}$$

then the limit

$$\lim_{N\to\infty} MSE_{NL}(\zeta) \tag{5.40}$$

obtains its global minimal value when $\zeta = \frac{\zeta_1\sqrt{\pi}}{b_{100}}$.

Note the global optimal value in above theorem is for infinite radial terms. That is what we want as discussed above on the limitations of other solutions. Based on this theorem, we can first fit the signal samples using GHOT model with $\zeta_1 = 0.5q_{max}^2$ in practice, then set ζ based on the first coefficient b_{100}. Note although

$$\zeta = \frac{\zeta_1\sqrt{\pi}}{b_{100}} = \frac{0.5q_{max}^2\sqrt{\pi}}{b_{100}}$$

ζ is actually independent of ζ_1, because ζ_1 is used to reduce the condition number of the basis matrix for large N and $\frac{b_{100}}{\zeta_1}$ is the real coefficient for the basis $q^2Y_0^0(\mathbf{u})$. Like HOT model the estimated coefficients $\{b_{nlm}\}$ in GHOT model are dependent on order N and L in Eq. (5.36). The estimated ADC is sensitive to the order N and L. However we do not need to estimate ADC well, what we need is only b_{100} for scale ζ in SPF basis. Fitting GHOT model can be formalized to a least square problem as

$$\min_{\boldsymbol{b}} \|\ln\mathbf{E} + M\boldsymbol{b}\|^2 \tag{5.41}$$

where ln is element-wise logorithmic operator performed in $\mathbf{E}$, M is the basis matrix generated from basis $\left\{\left(\frac{q^2}{\zeta_1}\right)^n Y_l^m(\mathbf{u})\right\}$, and $\boldsymbol{b} = (b_{100}, \cdots, b_{NLL})^T$. Thus

$$\boldsymbol{b} = (M^TM)^{-1}M^T\ln\mathbf{E} \tag{5.42}$$

Note the setting of ζ based on GHOT model fitting is in spirit an ADC modeling, which is a generalization of previous setting based on isotropic tensor fitting. If signal samples are obtained from multiple shells and in each shell the samples are evenly distributed in $\mathbb{S}^2$, then the scale ζ by the least square fitting in Eq. (5.41) with N and L is equivalent with the least square fitting with $N' = N$ and $L' = 0$, which is just the isotropic Gaussian function fitting in Eq. (5.30) when $N = 1$. That is because based on the orthogonality of SH basis, the columns of basis matrix M are orthonormal to each other. So the isotropic Gaussian function fitting is a specific case of the GHOT model fitting.

Least Square Estimation with the prior $E(0) = 1$

The estimation of $\{a_{nlm}\}$ is done in the second step, which has several choices. [Assemlal et al., 2009a] proposed two methods to estimate $\{a_{nlm}\}$, a linear least square fitting with regularization in the radial and spherical parts, and a non-linear variational optimization process, which considers the Rician noise and spatial information. We suggest that the Rician noise removal could be performed directly on the DWI data as a pre-processing step [Tristán-Vega and Aja-Fernández, 2010; Descoteaux et al., 2008b], although in our experiments, in order to perform fair comparisons of methods we did not perform any Rician noise removal. In this chapter we choose the least square method in the second step which is known to be faster. Note even in such a simple least square estimation, there is an important issue which needs to be considered carefully.

In order to reduce the estimation error in this step, we need to consider the prior of $E(0) = 1$ in the estimation. When $E(\mathbf{q})$ is represented by SPF basis in Eq. (5.1), we have

$$E(0) = \sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm} G_n(0) Y_l^m(\mathbf{u}) = \sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm} \kappa_n(\zeta) Y_l^m(\mathbf{u}) \tag{5.43}$$

Note that $E(0)$ in the above formula is a spherical function, not a scalar value. That means the function under SPF representation is not continuous when $\mathbf{q} = 0$, unless we add some constraints in $\{a_{nlm}\}$ such that $E(0\mathbf{u})$ is an isotropic spherical function. Moreover normally when we have N_s DWI images, the basis matrix in Eq. (5.6) is generated from these N_s samples, then the solution of least square problem in Eq. (5.8) has no guarantee that the estimated signal satisfies the prior $E(0) = 1$.

In our previous studies [Cheng et al., 2010a,b, 2011b,c; Caruyer et al., 2011; Merlet et al., 2011], we considered an artificial sampled shell at $q = 0$, such that the least square problem in Eq. (5.8) becomes

$$\min_{\mathbf{A}} \|\mathbf{E} - M_{SPF}\mathbf{A}\|^2 + w^2\|\mathbf{E}^{(0)} - M_{SPF}^{(0)}\mathbf{A}\|^2 + \mathbf{A}^T\Lambda\mathbf{A}$$

where $\mathbf{E}^{(0)}$ is an artificial data vector $(1, 1, \cdots, 1)^T$ with dimension $N_s^{(0)}$. $M_{SPF}^{(0)}$ is generated in Eq. (5.6) from $N_s^{(0)}$ directions $\{\mathbf{u}_j\}_{j=1}^{N_s^{(0)}}$ and $q = 0$, w is a weight between the term with real DWI samples and artificial samples in the shell with $q = 0$. Then the new least square problem is equivalent with

$$\min_{\mathbf{A}} \|\mathbf{E}' - M'_{SPF}\mathbf{A}\|^2 + \mathbf{A}^T\Lambda\mathbf{A}, \quad \mathbf{E}' = \begin{pmatrix}\mathbf{E}\\ w\mathbf{E}^{(0)}\end{pmatrix}, M'_{SPF} = \begin{pmatrix}M_{SPF}\\ wM_{SPF}^{(0)}\end{pmatrix} \tag{5.44}$$

Then the final solution is

$$\mathbf{A} = (M'^T_{SPF}M'_{SPF} + \Lambda)^{-1}M'^T_{SPF}\mathbf{E}'$$

In practice, we choose $\{\mathbf{u}_j\}_{j=1}^{N_s^{(0)}}$ as 81 evenly distributed vertices of a tessellated icosahedron in a hemisphere. Normally $w = 1$ works well in most cases. If the number of DWI data are large and SNR is very low, w needs to be increased. In practice, $w \in [1, 5]$ works well. Note w can not be too large, otherwise the condition number of M'_B will be large.

Here we propose a new way to consider the constraint $E(0) = 1$. Based on Eq. (5.43), $E(0) = 1$ means

$$\sum_{n=0}^{N} a_{nlm} G_n(0) = \sqrt{4\pi}\delta_l^0, \quad 0 \le l \le L, -l \le m \le l \tag{5.45}$$

which are $(L+1)(L+1)/2$ equations that should be satisfied by $(N+1)(L+1)(L+1)/2$ unknowns $\{a_{nlm}\}$. If $N = 0$, $\mathbf{A}$ can be solved by Eq. (5.45) such that

$$a_{000} = \frac{\sqrt{4\pi}}{G_0(0)}, \quad a_{0lm} = 0, \forall 0 \le l \le L, -l \le m \le l$$

So the estimated signal is

$$\widetilde{E}(\mathbf{q}) = \frac{\sqrt{4\pi}}{G_0(0)} G_0(q) Y_0^0(\mathbf{u}) = \frac{G_0(q)}{G_0(0)}$$

which represents only the isotropic signal. It is another explanation why we need $N \ge 1$.

In order to satisfy Eq. (5.45), we can consider $\{a_{0lm}\}_{l \le L}$ are dependent on $\{a_{nlm}\}_{0<n\le N, l\le L}$, i.e.

$$a_{0lm} = \frac{1}{G_0(0)} \left(\sqrt{4\pi}\delta_l^0 - \sum_{n=1}^{N} a_{nlm} G_n(0) \right), \quad 0 \le l \le L, -l \le m \le l \tag{5.46}$$

Then we have

$$\sum_{n=1}^{N} \sum_{l=0}^{L} \sum_{m=-l}^{l} a_{nlm} Y_l^m(\mathbf{u}) \left(G_n(q) - \frac{G_n(0)}{G_0(0)} G_0(q) \right) = E(\mathbf{q}) - \frac{G_0(q)}{G_0(0)} \tag{5.47}$$

So the least square solution of $\mathbf{A}' = (a_{100}, \cdots, a_{NLL})^T$ is

$$\mathbf{A}' = (M_{SPF}'^T M_{SPF}' + \Lambda')^{-1} M_{SPF}'^T \mathbf{E}' \tag{5.48}$$

where

$$M_{SPF}' = \begin{pmatrix} \left(G_1(q_1|\zeta) - \frac{G_1(0|\zeta)}{G_0(0|\zeta)} G_0(q_1|\zeta)\right) Y_0^0(\mathbf{u}_1) & \cdots & \left(G_N(q_1|\zeta) - \frac{G_N(0|\zeta)}{G_0(0|\zeta)} G_0(q_1|\zeta)\right) Y_L^L(\mathbf{u}_1) \\ \vdots & \ddots & \vdots \\ \left(G_1(q_{N_s}|\zeta) - \frac{G_1(0|\zeta)}{G_0(0|\zeta)} G_0(q_{N_s}|\zeta)\right) Y_0^0(\mathbf{u}_{N_s}) & \cdots & \left(G_N(q_{N_s}|\zeta) - \frac{G_N(0|\zeta)}{G_0(0|\zeta)} G_0(q_{N_s}|\zeta)\right) Y_L^L(\mathbf{u}_{N_s}) \end{pmatrix} \tag{5.49}$$

$$\mathbf{E}' = \begin{pmatrix} E(\mathbf{q}_1) - \frac{G_0(q_1)}{G_0(0)} \\ \vdots \\ E(\mathbf{q}_{N_S}) - \frac{G_0(q_{N_s})}{G_0(0)} \end{pmatrix}, \quad \Lambda' = \begin{pmatrix} \Lambda'_{100} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Lambda'_{NLL} \end{pmatrix}, \quad \Lambda'_{nlm} = \lambda_l l^2 (l+1)^2 + \lambda_n n^2 (n+1)^2 \tag{5.50}$$

Note that here we use the regularization proposed by Dr. Assemlal shown in subsection 5.1.1. Compared to the original least square method in Eq. (5.9), the regularization here is only considered for the coefficients $\{a_{nlm}\}_{0<n\le N}$. Please note that although we can set any set of $(L+1)(L+2)/2$ coefficients dependent on the other coefficients, we choose $\{a_{nlm}\}_{0<n\le N, l\le L}$ because we would like to keep the high order coefficients with large regularization Λ_{nlm} in estimation. Note that $E(\mathbf{q}) - \frac{G_0(q)}{G_0(0)} =$

$E(\mathbf{q}) - \exp(-\frac{q^2}{2\zeta})$ is the signal removed the approximated isotropic Gaussian part, and $\left(G_n(q) - \frac{G_n(0)}{G_0(0)}G_0(q)\right) Y_l^m(\mathbf{u})$ is the basis $G_n(q)Y_l^m(\mathbf{u})$ removed the isotropic Gaussian part. After we obtain $\mathbf{A}'$, another part of coefficients

$$\mathbf{A}_0 = (a_{000}, \cdots, a_{0LL})^T$$

can be calculated from Eq. (5.46). Then the final coefficient vector which naturally satisfies $E(0) = 1$ is

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_0 \\ \mathbf{A}' \end{pmatrix} \tag{5.51}$$

Please refer Table 5.1 for the whole estimation process in analytical SPFI, which includes three steps.

1. The first step is to estimate scale parameter ζ by two ways. One is based on the typical value from typical ADC in brain, and the other one is based on fitting the signal with the GHOT model.

2. The second step is the least square estimation of coefficient $\mathbf{A}$ by considering the prior $E(0) = 1$ and the radial and spherical regularizations.

3. The third step is to analytically obtain EAP and its various features from the estimated $\mathbf{A}$.

Note these three steps are actually independent of each other. So other methods are possible to be considered in scale estimation and $\mathbf{A}$ estimation. For example, we have proposed to use ℓ_1 norm minimization to estimate the coefficient $\mathbf{A}$ in the second step [Cheng et al., 2011c], which can be seen as the compressive sensing EAP reconstruction.

In the first step, if we set the scale parameter based on typical ADC value, then the scale ζ is shared in each voxel. In this case the basis matrix M'_{SPF} and $(M'^T_{SPF}M'_{SPF} + \Lambda')^{-1}$ in Eq. (5.48) only need to be calculated once for all voxels. The analytical linear transforms in the third step also only need to be calculated once for all voxels. Thus the whole estimation process is very fast. Moreover, if the shared scale ζ is used, the following operations on EAP field can be performed only on coefficient vector $\mathbf{A}$, which is needed in Riemannian framework in Chapter 6. Otherwise we need to consider the effect of basis in the operation of EAP field, which is normally complex and time consuming. However, using adaptive scale in different voxel can obtain better estimation results. When using adaptive scale, the basis matrix is different for different voxel, so is the linear transform in the third step. Thus it takes more time using adaptive scale. But it is still very fast, if least square method is used. The consideration of $E(0) = 1$ makes a set of coefficients $\{a_{0lm}\}_{l\leq L}$ out of the estimation process, which accelerates the method. Moreover, it is possible to accelerate the method by reusing the pre-calculated basis matrix with close scale. Although the scale is different for different voxel when we set scale by fitting the GHOT model, a good estimation of scale normally obtains smooth scale map with a small range. Then the range can be separated into small partitions, and the voxels with the scale values inside the same partition can be considered approximately to share the same scale and same basis matrix.

Table 5.1: SPFI via least square estimation with scale estimation and $E(0)$ consideration.

Input:
$\mathbf{V}$: X x Y x Z x N_s diffusion weighted MRI volume
$\mathbf{E}_{xyz}$: N_s x 1 diffusion weighted signal vector at voxel (x, y, z)
N, L : the order of Gaussian-Laguerre function and the order of SH basis in SPF basis, $N \geq 1$, $L \geq 4$
λ_n, λ_l: regularization weights in radial and spherical parts
N', L' : the order of polynomial basis and order of SH basis in GHOT model, $N' \geq 1$, $L' \geq 0$

Output:
$\mathbf{A}_{xyz}$: $N_{\mathbf{A}} = (N+1)(L+1)(L+2)/2$ dimensional coefficient vector of $E(\mathbf{q})$ under SPF basis at voxel (x, y, z)
EAP and its features: $P(\mathbf{R})$, $P(R_0\mathbf{r})$, $\Phi_t(\mathbf{r})$, $\Phi_w(\mathbf{r})$, Po, MSD, GFA

1. Scale ζ Estimation based on typical value or GHOT model:
a) based on typical value: $\zeta = \frac{1}{8\pi^2\tau D_0}$, $D_0 = 0.7 \times 10^{-3} mm^2/s$
b) M: basis matrix generated from basis $\left\{\left(\frac{q^2}{0.5q_{max}^2}\right)^n Y_l^m(\mathbf{u})\right\}$
$\boldsymbol{b} = (M^T M)^{-1} M^T \ln \mathbf{E}$, $\boldsymbol{b} = (b_{100}, \cdots, b_{N'L'L'})^T$ is the coefficient vector
$$\zeta = \frac{0.5q_{max}^2 \sqrt{\pi}}{b_{100}}$$

2. Least Square Estimation of coefficient vector A:
M'_{SPF}= Eq. (5.49): N_s x $N_{\mathbf{A}'}$ matrix from SPF basis, $N_{\mathbf{A}'} = N(L+1)(L+2)/2$
$$\mathbf{E}' = \begin{pmatrix} E(\mathbf{q}_1) - \frac{G_0(q_1)}{G_0(0)} \\ \vdots \\ E(\mathbf{q}_{N_S}) - \frac{G_0(q_{N_S})}{G_0(0)} \end{pmatrix} : N_s \text{ x } 1 \text{ dimensional data vector, } G_n(q) = \text{Eq. (5.2)}$$
$$\Lambda' = \begin{pmatrix} \Lambda'_{100} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Lambda'_{NLL} \end{pmatrix} : \begin{array}{l} N_{\mathbf{A}'} \text{ x } N_{\mathbf{A}'} \text{regularization matrix} \\ \Lambda'_{nlm} = \lambda_l l^2(l+1)^2 + \lambda_n n^2(n+1)^2 \end{array}$$
$$\mathbf{A}' = (M'^T_{SPF} M'_{SPF} + \Lambda')^{-1} M'^T_{SPF} \mathbf{E}'$$
$$\mathbf{A}_0 = (a_{000}, \cdots, a_{0LL})^T, \quad a_{0lm} = \frac{1}{G_0(0)} \left(\sqrt{4\pi}\delta_l^0 - \textstyle\sum_{n=1}^N a_{nlm} G_n(0) \right)$$
$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_0 \\ \mathbf{A}' \end{pmatrix}$$

3. Analytical Calculation of EAP and its features from A:

$$P(R\mathbf{r}) = \sum_{n=0}^{N} \sum_{l=0}^{L} \sum_{m=-l}^{l} a_{nlm} F_{nl}(R) Y_l^m(\mathbf{r}), \quad F_{nl}(R) = \text{Eq. (5.14)}$$

$$P(R_0\mathbf{r}) = \sum_{l=0}^{L} \sum_{m=-l}^{l} c_{lm}^P(R_0) Y_l^m(\mathbf{r}), \quad c_{lm}^P(R_0) = \sum_{n=0}^{N} a_{nlm} F_{nl}(R_0)$$

$$\Phi_t(\mathbf{r}) = \sum_{l=0}^{L} \sum_{m=-l}^{l} c_{lm}^{\Phi_t} Y_l^m(\mathbf{r}), \quad c_{lm}^{\Phi_t} = \text{Eq. (5.18)}, \quad \Phi_w(\mathbf{r}) = \sum_{l=0}^{L} \sum_{m=-l}^{l} c_{lm}^{\Phi_w} Y_l^m(\mathbf{r}), \quad c_{lm}^{\Phi_w} = \text{Eq. (5.20)}$$

$$Po = \text{Eq. (5.22)}, \quad GFA = \text{Eq. (5.28)}, \quad MSD = \text{Eq. (5.26)}$$

The effect of the diffusion time τ

Diffusion time τ is used to determine q values from b values. Normally when we have a data set, only b values and gradients are known. The diffusion time τ is not stated in most of time unless the people who acquired the data tell us additionally. Fortunately many methods do not need diffusion time, e.g. DTI, QBI, etc. However, in SPFI we use the q values explicitly which are calculated based on $b = 4\pi^2\tau q^2$. Then how SPFI works for the data without knowing the diffusion time τ? For example, the public phantom data used in Fiber Cup in MICCAI 2009 described in Appendix B [Fillard et al., 2011; Poupon et al., 2008] did not report the diffusion time τ. We may assume a diffusion time τ_1 and estimate the EAP and its features under the assumption $\tau = \tau_1$. Thus it is necessary to study how the estimated results change when τ changes and discuss how we set τ in SPFI. For different diffusion time τ the estimated EAP and its features from SPFI satisfy the following proposition whose proof can be found in Appendix 5.6.

Proposition 5.5 (Invariance for different diffusion time τ). *For a given data set with known b values and gradients $\{\mathbf{u}_j\}_{j=1}^{N_s}$ and unknown diffusion time τ, if ζ is set in Eq. (5.38) by fitting GHOT model, then the estimated $\Phi_t(\mathbf{r})$, $\Phi_t(\mathbf{r}, C)$, $\Phi_w(\mathbf{r})$, $\Phi_w(\mathbf{r}, C)$, RTO, MSD, GFA are invariant for different diffusion time τ. If we denote the EAP estimated from τ by $P(R\mathbf{r}, \tau)$, then*

$$P(R\mathbf{r}, \tau_2) = \left(\frac{\tau_1}{\tau_2}\right)^{3/2} P(\sqrt{\frac{\tau_1}{\tau_2}}R\mathbf{r}, \tau_1) \tag{5.52}$$

Note that the proof of this proposition is independent of the least square estimation of $\{a_{nlm}\}$. For other estimation of $\{a_{nlm}\}$, this proposition also holds. The proposition shows the axis of estimated EAP is scaled due to the change of τ. Since the ODFs are defined as the normalized radial integration of EAP, they are invariant for different τ. Based on the definition of RTO, MSE, GFA, they are also scale invariant. This is analogous to DTI, where the tensor $\mathbf{D}$, ODFs and scalar indices are all independent of τ, while the EAP in DTI is $P(\mathbf{R}) = N(\mathbf{R}|2\tau\mathbf{D})$ which is scaled by τ. See Appendix 5.8. In DTI, Eq. (5.52) also holds, i.e.

$$N(\mathbf{R}|2\tau_2\mathbf{D}) = \left(\frac{\tau_1}{\tau_2}\right)^{3/2} N(\sqrt{\frac{\tau_1}{\tau_2}}\mathbf{R}|2\tau_1\mathbf{D})$$

Note if we set ζ in SPFI based on isotropic tensor model or based on a typical ADC D_0 as we have discussed, this proposition also holds, because in these cases $\tau\zeta$ are constant in the estimation.

Thus although we use q values explicitly in SPFI, it was proved that SPFI only needs b values and gradients to obtain the EAP and its features. The shape of the estimated EAP profile (NOT EAP profile itself) is independent of the diffusion time. Other EAP features are invariant to the change of τ. In practice, if we do not know τ, we can set it as $\tau = \frac{1}{4\pi^2} = 0.02533s$ such that $b = q^2$. Throughout this thesis, we always set $\tau = 0.02533s$ in experiments because we do not know the exact τ of data sets.

Parameter Selection in Practice

In this subsection, we would like to discuss how to choose the parameters in the estimation process shown in Table 5.1. The practical process includes three parts.

- **Determine N', L' in GHOT model based on pseudo-ADC map for scale estimation.** In the first step, if the scale parameter ζ is set by fitting the GHOT model with order N' and L', then ζ is dependent on N' and L' due to the limited number of samples in discrete least square estimation. Note that the fitting of GHOT model is very fast because the basis matrix in GHOT model is shared in each voxel. Thus in practice for real data, we can perform scale estimation several times efficiently with different N', L'. Then we obtain a set of pseudo-ADC maps for different N' and L', where there is a pseudo-ADC value in each voxel. A good scale estimation should obtain smooth pseudo-ADC map with good quality. Thus we can set N', L' by checking the quality of pseudo-ADC map. Normally $N' = 1$, $L' = 4$ works in most cases. If the samples are obtained from multiple shells and evenly distributed in each shell, then we can set $L' = 0$. However, N' should be as large as possible if the pseudo-ADC map has good quality. For a large number of DWI data with high SNR, N', L' can be increased.
- **Determine N, L, λ_n, λ_l in SPF coefficient estimation based on GFA, MSD and RTO maps.** In the second step, if least square estimation is used, the estimation is fast, although different voxel may use different scale. Thus we can perform the estimation several times with different N and L. Then a set of GFA, MSD and RTO maps can be obtained from a set of possible N and L. Again, a good coefficient $\mathbf{A}$ estimation means a good quality of these scalar index map. Thus we can determine an appropriate pair of N and L by checking the quality of these scalar maps. In practice, if least square estimation is used for coefficient $\mathbf{A}$ estimation, N can be $1, 2$ and L can be $4, 6$. Normally $(N, L) = (1, 4)$ works in most cases, especially when scale ζ is estimated well. However, N, L should be as large as possible if the scalar maps keep good quality as N, L increase. λ_n and λ_l are dependent on sampling scheme, noise level, selection order N, L, and the scale ζ. So there is no general optimal values for λ_l and λ_n. Based on our experiments $\lambda_n = \lambda_l = 10^{-8}$ can be used in most cases, even for low SNR=10. For SNR higher than 10, λ_n and λ_l can be decreased.
- **Determine radius R_0 for EAP profiles based on GFA for spherical functions.** In the third step, another parameter R_0 is needed for EAP profile $P(R_0\mathbf{r})$. Since the third step for EAP profiles is just an analytical linear transform independent of the first two steps, the process is very fast in this step. So we can perform this transform several times with different radius R_0. For each $P(R_0\mathbf{r})$, GFA for spherical functions, i.e. GFA for ODFs proposed in [Tuch, 2004], can be calculated for $P(R_0\mathbf{r})$. Note that since $P(R_0\mathbf{r})$ is analytically represented by SH basis, its GFA can be analytically obtained by Eq. (4.62), which is very fast. Again, an appropriate R_0 results in a GFA map with good quality. Thus we can set R_0 by checking the quality of GFA map for $P(R_0\mathbf{r})$.

5.2 Analytical Fourier Transform in Spherical Coordinate (AFT-SC)

In this section, we would like to theoretically compare different sHARDI methods including QBI, exact QBI, DOT and mHARDI methods including SPFI, SHORE and DPI in the same framework. The three sHARDI methods have some intrinsic modeling errors or need some unrealistic assumptions as we have discussed in Section 4.3. Moreover they have difficulty to deal with signals from different shells. Note we will discuss exact QBI in DOT, because they use the same model, i.e. mono-exponential assumption, as discussed in Section 4.3. See Table 4.1. The three mHARDI methods all represent diffusion signal with some basis functions in spherical coordinate and use plane wave expansion to analytically solve the Fourier transform. To our knowledge, there is no theoretical analysis and practical comparison among these sHARDI and mHARDI methods. In this section, we propose the Analytical Fourier Transform in Spherical Coordinate (AFT-SC) framework and analyze these methods and their possible variants in this framework.

5.2.1 AFT-SC Framework

The central idea in QBI, exact QBI, DOT, DPI, SPFI and SHORE is to fit the $E(\mathbf{q})$ with some function sets and find the analytical relation between signal $E(\mathbf{q})$ and EAP $P(\mathbf{R})$ or ODFs $\Phi_k(\mathbf{r})$. Here we present them in the same framework, named AFT-SC. Assume $E(\mathbf{q})$ can be represented in Eq. (5.53) as a linear combination of functions $\{B_k(q\mathbf{u})\}$, where the basis function $B_k(q\mathbf{u}) = R_k(q)Q_k(\mathbf{u})$ separates radial part and spherical part.

$$E(\mathbf{q}) = \sum_{k=0}^{K} a_k B_k(q\mathbf{u}) = \sum_{k=1}^{K} a_k R_k(q) Q_k(\mathbf{u}) \tag{5.53}$$

By considering the well known plane wave formula in Eq. (5.12), $P(\mathbf{R})$ can be represented by dual basis $\{D_k(\mathbf{R})\}$ with the same coefficients $\{a_k\}$, and $D_k(\mathbf{R})$ could be separated into radial integration $F_{kl}(R)$ and spherical integration $T_{klm}(\mathbf{r})$ in Eq. (5.55).

$$\begin{aligned} P(\mathbf{R}) &= 4\pi \int_{\mathbb{R}^3} \sum_{k=0}^{K} a_k R_k(q) Q_k(\mathbf{u}) \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (-1)^{l/2} j_l(2\pi q R) Y_l^m(\mathbf{u}) Y_l^m(\mathbf{r}) \mathrm{d}\mathbf{q} \\ &= \sum_{k=0}^{K} a_k \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left(4\pi(-1)^{l/2} \int_0^{\infty} R_k(q) j_l(2\pi q R) q^2 \mathrm{d}q \right) \left(\left\{ \int_{\mathbb{S}^2} Q_k(\mathbf{u}) Y_l^m(\mathbf{u}) d\mathbf{u} \right\} Y_l^m(\mathbf{r}) \right) \\ &= \sum_{k=0}^{K} a_k D_k(\mathbf{R}) = \sum_{k=0}^{K} \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_k F_{kl}(R) T_{klm}(\mathbf{r}) \end{aligned} \tag{5.54}$$

$$F_{kl}(R) = 4\pi(-1)^{l/2} \int_0^{\infty} R_k(q) j_l(2\pi q R) q^2 \mathrm{d}q \quad T_{klm}(\mathbf{r}) = \left\{ \int_{\mathbb{S}^2} Q_k(\mathbf{u}) Y_l^m(\mathbf{u}) \mathrm{d}\mathbf{u} \right\} Y_l^m(\mathbf{r}) \tag{5.55}$$

Thus there is an analytical EAP estimation once $F_{kl}(R)$ and $T_{klm}(\mathbf{r})$ have closed forms. Since SH is the orthonormal basis which has been widely used in many domains such as dMRI and graphics and $E(\mathbf{q})$ and $P(\mathbf{R})$ are both antipodally symmetric, it is reasonable to choose $Q_k(\mathbf{u})$ as the real Spherical Harmonic with even l [Descoteaux et al., 2007], which is denoted by $Y_l^m(\mathbf{u})$ and defined in Eq. (2.28). Then we have $T_{klm}(\mathbf{r}) = Y_l^m(\mathbf{r})$, because $\int_{\mathbb{S}^2} Y_{l'}^{m'}(\mathbf{q}) Y_l^m(\mathbf{u}) = \delta_{ll'}^{mm'}$. Then different methods can be obtained

Table 5.2: Several kinds of methods in Analytical Fourier Transforms in Spherical Coordinate (AFT-SC) framework.

method	$R_k(q)$	$Q_k(\mathbf{u})$	$F_{kl}(R) = 4\pi(-i)^l \int_0^\infty R_k(q) j_l(2\pi qR) q^2 dq$	$T_{klm}(\mathbf{r})$
QBI	$R(q) = \delta(q - q_0)$	$Y_l^m(\mathbf{u})$	$F_l(R) = 4\pi(-1)^{l/2} j_l(2\pi q_0 R) q_0^2$	$Y_l^m(\mathbf{r})$
SHORE	$R_{nl}(q) = G_{nl}(q\|\zeta)$	$Y_l^m(\mathbf{u})$	$F_{nl}(R\|\zeta) = (-1)^n G_{nl}(R\|\frac{1}{4\pi^2\zeta})$	$Y_l^m(\mathbf{r})$
SPFI	$R_n(q) = G_n(q\|\zeta)$	$Y_l^m(\mathbf{u})$	Eq. (5.14)	$Y_l^m(\mathbf{r})$
DPI_0	$R_{0l}(q) = (\frac{q}{\sqrt{\zeta}})^l$ $R_{1l}(q) = (\frac{q}{\sqrt{\zeta}})^{-l-1}$	$Y_l^m(\mathbf{u})$	$F_{0l} = (-1)^{l/2} q_{max}^{l+1.5} \zeta^{-0.5l} R^{-1.5} J_{l+1.5}(2\pi q_{max} R)$ $F_{1l} = (-1)^{l/2} R^{-1.5} \zeta^{0.5l+0.5} (\frac{(\pi R)^{l-0.5}}{\Gamma(l+0.5)} - \frac{J_{l-0.5}(2\pi q_{max} R)}{q_{max}^{l-0.5}})$	$Y_l^m(\mathbf{r})$
DPI_1	$R_l(q) = (\frac{q^2}{\zeta})^{l/2} \exp(-\frac{q^2}{2\zeta})$	$Y_l^m(\mathbf{u})$	$F_l(R) = 2^{l+1.5} \zeta^{0.5l+1.5} \pi^{l+1.5} R^l \exp(-2\zeta\pi^2 R^2)$	$Y_l^m(\mathbf{r})$
DOT_1	$R_n(q) = (\frac{q}{\sqrt{\zeta}})^{2n}$	$Y_l^m(\mathbf{u})$	$\frac{q_{max}^{2n+l+3} \pi^{l+1.5} R^l \Gamma(1.5+0.5l+n)_1 F_2(1.5+0.5l+n; 1.5+l, 2.5+0.5l+n; -\pi^2 q_{max} R)}{(-1)^{l/2} \zeta^n \Gamma(1.5+l) \Gamma(2.5+0.5l+n)}$	$Y_l^m(\mathbf{r})$
DOT_2 (SPNP)	$R_n(q) = (\frac{q^2}{\zeta})^n \exp(-\frac{q^2}{2\zeta})$	$Y_l^m(\mathbf{u})$	$\frac{2^{n+0.5l+1.5} \zeta^{0.5l+1.5} \pi^{l+1.5} R^l \Gamma(n+0.5l+1.5)_1 F_1(n+0.5l+1.5, l+1.5, -2\zeta\pi^2 R^2)}{(-1)^{l/2} \Gamma(l+1.5)}$	$Y_l^m(\mathbf{r})$

by choosing different radial functions. See table 5.2 for an overview and we will introduce them one by one in the following.

Note that similarly with SPFI, in AFT-SC framework $E(0)$ is still a spherical function, i.e.

$$E(0) = \sum_{k=1}^{K} a_k R_k(0) Q_k(\mathbf{u})$$

Thus we also need to consider the prior $E(0) = 1$ in estimation. Otherwise, there is no guarantee that the estimated $E(\mathbf{q})$ satisfies this prior.

5.2.2 HARDI methods in AFT-SC framework

QBI

As we discussed in subsection 4.3.4, the signal in QBI [Tuch, 2004; Descoteaux et al., 2007; Assemlal et al., 2009a; Hess et al., 2006] is assumed as

$$E(q\mathbf{u}) = \sum_{l=0}^{L} \sum_{m=-l}^{l} a_{lm} \delta(q - q_0) Y_l^m(\mathbf{u})$$

where $q = q_0$ is the q value for the given shell. The analytical form of ODF could be obtained from several ways [Descoteaux et al., 2007; Assemlal et al., 2009a; Hess et al., 2006]. In AFT-SC,

$$F_{kl}(R) = 4\pi(-1)^{l/2} \int_0^\infty \delta(q - q_0) j_l(2\pi qR) q^2 dq = 4\pi(-1)^{l/2} j_l(2\pi q_0 R) q_0^2 \tag{5.56}$$

So the EAP in QBI can be obtained as

$$P(\mathbf{R}) = \sum_{l=0}^{L} \sum_{m=-l}^{l} a_{lm} 4\pi(-1)^{l/2} j_l(2\pi q_0 R) q_0^2 Y_l^m(\mathbf{r}) \tag{5.57}$$

Then the ODF by Tuch $\Phi_t(\mathbf{r})$ in QBI is

$$\begin{aligned}\Phi_t(\mathbf{r}) &= \frac{1}{Z}\int_0^\infty P(R\mathbf{r})\mathrm{d}R = \frac{2q_0}{Z}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{lm}(-1)^{l/2}Y_l^m(\mathbf{r})\int_0^\infty j_l(x)\mathrm{d}x \\ &= \frac{\sqrt{\pi}q_0}{Z}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{lm}\frac{(-1)^{l/2}\Gamma(l/2+1/2)}{\Gamma(l/2+1)}Y_l^m(\mathbf{r}) \\ &= \frac{q_0}{Z}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{lm}2\pi P_l(0)Y_l^m(\mathbf{r}) \end{aligned} \tag{5.58}$$

In the above derivations, we consider

$$\int_0^\infty j_l(x)dx = \frac{\sqrt{\pi}\Gamma(0.5l+0.5)}{2\Gamma(0.5l+1)} = (-1)^{l/2}\frac{\pi}{2}P_l(0)$$

based on the definition of $j_l(x)$ in Eq. (2.44), the property of Bessel function in Eq. (2.86), and $P_l(0)$ in Eq. (2.39).

The ODF formula is the same as the one in [Descoteaux et al., 2007; Assemlal et al., 2009a; Hess et al., 2006], but the derivation is more simple and straightforward. Compared to the addition theorem and Funk-Hecke theorem used in [Descoteaux et al., 2007; Assemlal et al., 2009a; Hess et al., 2006], we use plane wave theorem to prove the solution of $\Phi_t(\mathbf{r})$. The plane wave theorem seems to be more fundamental. As proved in Theorem 5.4 and Theorem 5.5 in Appendix 5.7, Funk-Hecke theorem 2.7 and addition theorem 2.6 are equivalent, and they can be proved by plane wave expansion theorem 2.8.

Compared to previous works [Descoteaux et al., 2007; Assemlal et al., 2009a; Hess et al., 2006], in AFT-SC we can obtain the new form for the EAP in Eq. (5.57) as well as ODF in QBI in Eq. (5.58). To our knowledge, the EAP formula has not been proposed before in QBI. However, the EAP in QBI seems useless, because it has many negative values when b is large. That can be seen from the mean of spherical function $P(R\mathbf{r})$, i.e.

$$\int_{\mathbb{S}^2} P(R\mathbf{r})\mathrm{d}\mathbf{r} = a_{00}q_0^2(4\pi)^{3/2}j_0(2\pi q_0 R)$$

Note normally a_{00} is positive, because it is related with the mean of $E(q_0\mathbf{u})$. Fig. 5.4 shows that as q value increases $j_0(2\pi qR)$ has more and more negative values in R-axis, especially for the large R values.

One may think that we can obtain $\Phi_w(\mathbf{r})$ from the EAP in Eq. (5.57), i.e.

$$\Phi_w(\mathbf{r}) = \int_0^\infty P(R\mathbf{r})R^2\mathrm{d}R = \frac{1}{2\pi^2 q_0}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{lm}(-1)^{l/2}Y_l^m(\mathbf{r})\int_0^\infty j_l(x)x^2\mathrm{d}x \tag{5.59}$$

However the integration $\int_0^\infty j_l(x)x^2\mathrm{d}x$ does not converge, which means any estimator of $\Phi_w(\mathbf{r})$ based on delta function assumption fails. Note the estimator in [Tristán-Vega et al., 2009] works for low b value, probably because it adds additional assumption. For large b value, the estimator in [Tristán-Vega et al., 2009] fails with many negative values in practice.

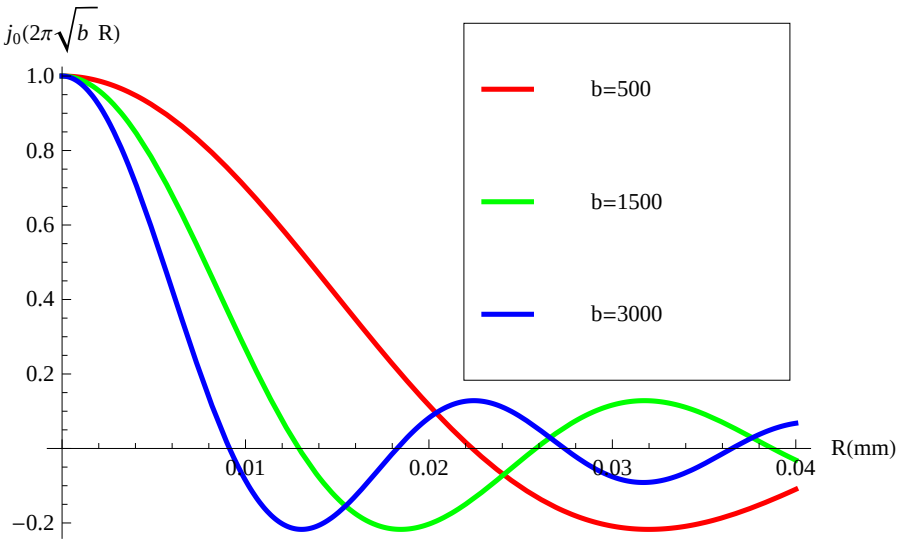


Figure 5.4: $j_0(2\pi\sqrt{b}R)$ under different b values. We choose τ such that $b = q^2$. It can be seen that as b increases $j_0(2\pi\sqrt{b}R)$ has more negative values in R-axis.

SHORE

We have introduced SHORE method for signals both in 1D and in 3D cases in subsection 4.3.8. The SHO-1D basis in Eq. (2.60) can not handle the signal in real data in 3D case. As we has discussed, the functions defined in Eq. (4.85) are not orthogonal because of a small mistake in the exponential part in Eq. (4.85) [Özarslan et al., 2009]. So in this subsection we would like to study only the corrected SHO-3D basis in Eq. (2.69).

SHO-3D Basis. SHORE represents $E(\mathbf{q})$ as a linear combination of SHO-3D basis in Eq. (5.60). SHO-3D basis is denoted by $\{B_{nlm}^{\mathrm{SHO3}}\}$, which is the solution of 3D quantum mechanical harmonic oscillator problem [Özarslan et al., 2009] as shown in subsection 2.5.5.

$$E(\mathbf{q}) = \sum_{n=0}^{N}\sum_{l=0}^{2n}\sum_{m=-l}^{l} a_{nlm} B_{nlm}^{\mathrm{SHO3}}(\mathbf{q}|\zeta), \quad B_{nlm}^{\mathrm{SHO3}}(\mathbf{q}|\zeta) = G_{nl}(q|\zeta)Y_l^m(\mathbf{u}) \tag{5.60}$$

$$G_{nl}(q|\zeta) = \kappa_{nl}(\zeta)\left(\frac{q^2}{\zeta}\right)^{l/2}\exp\left(-\frac{q^2}{2\zeta}\right)L_{n-l/2}^{l+1/2}\left(\frac{q^2}{\zeta}\right) \qquad \kappa_{nl}(\zeta) = \left[\frac{2}{\zeta^{3/2}}\frac{(n-l/2)!}{\Gamma(n+l/2+3/2)}\right]^{1/2} \tag{5.61}$$

The radial part of SHO-3D basis with the scale ζ is shown in $G_{nl}(q|\zeta)$ Eq. (5.61), where $L_n^{\alpha}(x)$ is the generalized Laguerre polynomial [Ritchie and Kemp, 2000; Huzinaga, 1965; Özarslan et al., 2009]. Please note the difference between the N in Eq. (5.60) and N_{max} in Eq. (4.84) [Özarslan et al., 2009]. N_{max} in Eq. (4.84) denotes the order of polynomial in radial part which is $2N$ in our notation. Actually this basis has been already used in the computation of the molecular electron orbitals and molecular docking [Ritchie and Kemp, 2000; Huzinaga, 1965; Assemlal et al., 2009a], where $n - l/2$ is replaced by $n - l$. Here we use $n - l/2$ such that the order of polynomial term in B_{nlm}^{SHO3} is $2(n - l/2 + 2 \cdot l/2) = 2n$, and the maximal order of polynomial term is $2N$.

Note that $G_{n0}(q|\zeta)$ in Eq. (5.61) is just the radial function $G_n(q|\zeta)$ in Eq. (5.2). Thus SHO-3D basis $\{B^{\text{SHO3}}_{nlm}(\mathbf{q}|\zeta) = G_{nl}(q|\zeta)Y^m_l(\mathbf{u})\}$ and SPF basis $\{B^{\text{SPF}}_{nlm}(\mathbf{q}|\zeta) = G_n(q|\zeta)Y^m_l(\mathbf{u})\}$ are very similar. However their Fourier dual basis are quite different. See proposition 5.6 for the Fourier dual basis of SHO-3D basis.

EAP and its features. Like SPFI, the following propositions demonstrate that in SHORE EAP and its features can be analytically calculated from the pre-estimated coefficients $\{a_{nlm}\}$.

Proposition 5.6 (EAP in SHORE). *The Fourier dual basis of SHO-3D basis* $B^{SHO3}_{nlm}(\mathbf{q}|\zeta)$ *is* $(-1)^n B^{SHO3}_{nlm}(\mathbf{R}|\frac{1}{4\pi^2\zeta})$*, i.e.*

$$\mathscr{F}_{3D}\{G_{nl}(q|\zeta)Y^m_l(\mathbf{u})\}(\mathbf{R}) = (-1)^n G_{nl}(R|\frac{1}{4\pi^2\zeta})Y^m_l(\mathbf{r}) \tag{5.62}$$

If the signal is represented by SHO-3D basis in Eq. (5.60)*, the EAP is analytically obtained as*

$$P(\mathbf{R}) = \sum_{n=0}^{N}\sum_{l=0}^{2n}\sum_{m=-l}^{l} a_{nlm}(-1)^n G_{nl}(R|\frac{1}{4\pi^2\zeta})Y^m_l(\mathbf{r}) \tag{5.63}$$

where G_{nl} *is defined in Eq.* (5.61)*.*

Proof. Based on Eq. (5.61), the radial integration in Eq. (5.55) is

$$\begin{aligned}
F_{nl}(R) &= 4\pi(-1)^{l/2}\int_0^\infty G_{nl}(q|\zeta)j_l(2\pi qR)q^2\mathrm{d}q \\
&= 4\pi(-1)^{l/2}\kappa_{nl}(\zeta)\int_0^\infty (\frac{q^2}{\zeta})^{l/2}\exp\left(-\frac{q^2}{2\zeta}\right)L^{l+1/2}_{n-l/2}(\frac{q^2}{\zeta})j_l(2\pi qR)q^2\mathrm{d}q \\
&= 4\pi(-1)^{l/2}\left(\frac{2(n-l/2)!}{\Gamma(n+l/2+3/2)}\right)^{1/2}\zeta^{-l/2-3/4}\frac{1}{2\sqrt{R}}\int_0^\infty q^{l+\frac{2}{3}}e^{-\frac{q^2}{\zeta}}L^{l+1/2}_{n-l/2}(\frac{q^2}{\zeta})J_{l+\frac{1}{2}}(2\pi Rq)\mathrm{d}q
\end{aligned}$$

Considering the property of Laguerre polynomial in Eq. (2.81) [Gradshteyn and Ryzhik, 2007], we have

$$F_{nl}(R) = (-1)^n G_{nl}(R|\frac{1}{4\pi^2\zeta})$$

Considering the spherical part $Y^m_l(\mathbf{r})$, the Fourier dual SHO-3D basis is given in Eq. (5.62). Then considering the linearity of Fourier transform, the EAP can be represented by the dual basis in Eq. (5.63). □

Proposition 5.7 (ODFs in SHORE). *If the signal is represented by SHO-3D basis in Eq.* (5.60)*, the ODF* $\Phi_k(\mathbf{r})$ *defined in Eq.* (4.69) *is analytically obtained as*

$$\Phi_k(\mathbf{r}) = \frac{1}{Z}\sum_{l=0}^{2N}\sum_{m=-l}^{l} c_{lm}Y^m_l(\mathbf{r}) \tag{5.64}$$

$$c_{lm} = (4\pi^2\zeta)^{-\frac{k+1}{2}}\sum_{n=l/2}^{N}\sum_{j=0}^{n-l/2} a_{nlm}\frac{(-1)^{n+j}}{j!}\binom{n+l/2+1/2}{n-l/2-j}2^{\frac{k+l-1}{2}+j}\Gamma(\frac{k+l-1}{2}+j) \tag{5.65}$$

Proof. Based on Proposition 5.6, the EAP can be represented in Eq. (5.63). Then the ODF can be written as

$$\Phi_k(\mathbf{r}) = \frac{1}{Z}\int_0^\infty P(R\mathbf{r})R^k \mathrm{d}R = \frac{1}{Z}\sum_{n=0}^{N}\sum_{l=0}^{2n}\sum_{m=-l}^{l}(-1)^n a_{nlm}Y_l^m(\mathbf{r})\underbrace{\left(\int_0^\infty G_{nl}(R|\frac{1}{4\pi^2\zeta})R^k\mathrm{d}R\right)}_{I_{nlk}} \tag{5.66}$$

$$= \frac{1}{Z}\sum_{l=0}^{2N}\sum_{m=-l}^{l}Y_l^m(\mathbf{r})\underbrace{\left(\sum_{n=l/2}^{N}(-1)^n a_{nlm}I_{nlk}\right)}_{c_{lm}} \tag{5.67}$$

where

$$\begin{aligned}
I_{nlk} &= \int_0^\infty R^k(4\pi^2\zeta R^2)^{l/2}e^{-2\pi^2\zeta R^2}L_{n-l/2}^{l+1/2}(4\pi^2\zeta R^2)\mathrm{d}R \\
&= (4\pi^2\zeta)^{-\frac{k+1}{2}}\frac{1}{2}\int_0^\infty e^{-\frac{x}{2}}x^{\frac{k+l-1}{2}}L_{n-l/2}^{l+1/2}(x)\mathrm{d}x \\
&= (4\pi^2\zeta)^{-\frac{k+1}{2}}\frac{1}{2}\int_0^\infty e^{-\frac{x}{2}}x^{\frac{k+l-1}{2}}\sum_{j=0}^{n-l/2}(-1)^j\binom{n+l/2+1/2}{n-l/2-j}\frac{x^j}{j!}\mathrm{d}x \\
&= (4\pi^2\zeta)^{-\frac{k+1}{2}}\sum_{j=0}^{n-l/2}\frac{(-1)^j}{j!}\binom{n+l/2+1/2}{n-l/2-j}2^{\frac{k+l-1}{2}+j}\Gamma(\frac{k+l-1}{2}+j)
\end{aligned}$$

where we use the property of Laguerre polynomial in Eq. (2.63) and the definition of Gamma function. Then put I_{nlk} back to Eq. (5.67), which proves Eq. (5.64). □

Eq. (5.64) shows the SH coefficients $\{c_{lm}\}$ of ODFs can be obtained by a linear transform of $\{a_{nlm}\}$. For $\Phi_w(\mathbf{r})$, i.e. $k = 2$, the normalization factor is $Z = 1$ in Eq. (5.64).

Similarly with SPFI, the scalar indices RTO, MSD and GFA can be also calculated from the pre-estimated $\{a_{nlm}\}$. It is easy to see that the formula for GFA is given as

$$GFA = \sqrt{1 - \frac{\sum_{n=0}^{N}a_{n00}^2}{\sum_{n=0}^{N}\sum_{l=0}^{2n}\sum_{m=-l}^{l}a_{nlm}^2}} \tag{5.68}$$

The other two indices can be also obtained easily.

Least Square Estimation with $E(0) = 1$ consideration. Similarly with SPFI, SHORE can consider $E(0) = 1$ in the estimation process. $E(0) = 1$ means

$$E(0) = \sum_{n=0}^{N}\sum_{l=0}^{2n}\sum_{m=-l}^{l}a_{nlm}G_{nl}(0)Y_l^m(\mathbf{u}) = 1 = \sqrt{4\pi}Y_0^0(\mathbf{u})$$

Because $G_{nl}(0) = 0, \forall l > 0$, we have

$$\sum_{n=0}^{N}a_{n00}G_{n0}(0) = \sqrt{4\pi}$$

It is just one equation, compared to a set of equations in Eq. (5.45). Then we can assume a_{000} is dependent on other coefficients, i.e.

$$a_{000} = \frac{1}{G_{00}(0)}\left(\sqrt{4\pi} - \sum_{n=1}^{N} a_{n00}G_{n0}(0)\right) \tag{5.69}$$

$$\sum_{n=1}^{N}\sum_{l=0}^{2n}\sum_{m=-l}^{l} a_{nlm}\left(G_{nl}(q) - \frac{G_{n0}(0)\delta_l^0}{G_{00}(0)}G_{00}(q)\right)Y_l^m(\mathbf{u}) = E(\mathbf{q}) - \frac{G_{00}(q)}{G_{00}(0)} \tag{5.70}$$

Then the coefficients $\{a_{nlm}\}_{n>0}$ can be estimated from a least square method, and the final coefficient vector is the combination of these two parts. The whole process is similar with the process of SPFI in Table 5.1.

SPFI

SPFI also belongs to the AFT-SC framework, because it represents $E(\mathbf{q})$ using SPF basis in Eq. (5.1) and EAP $P(\mathbf{R})$ using dSPF basis in Eq. (5.15). As we have shown above, SPF basis and SHO-3D basis are very similar, while their Fourier dual bases are quite different. We will discuss the differences and similarities between SPF and SHO-3D basis in details in the Section 5.3.

$\mathbf{DPI_0}$ and $\mathbf{DPI_1}$

DPI in [Descoteaux et al., 2010], called DPI_0 here, assumes the signal $E(\mathbf{q})$ is the solution of 3D Laplace equation, i.e.

$$E(\mathbf{q}) = \sum_{n=0}^{1}\sum_{l=0}^{L}\sum_{m=-l}^{l} c_{nlm}R_{nl}(q)Y_l^m(\mathbf{u}) \tag{5.71}$$

$$R_{0l}(q) = (\frac{q}{\sqrt{\zeta}})^l, \quad R_{1l}(q) = (\frac{q}{\sqrt{\zeta}})^{-l-1} \tag{5.72}$$

Please note that here we introduce the scale parameter ζ in DPI_0 motivated by SHORE and SPFI. It is the same as the original DPI in [Descoteaux et al., 2009, 2010] if $\zeta = 1$. If $\zeta \neq 1$, it is equivalent with the original one but more numerically stable when an appropriate ζ is chosen. For the original DPI, one need to choose carefully the unit for numerical stability in least square estimation as suggested in [Descoteaux et al., 2010]. Here we can choose a scale ζ so that $q/\sqrt{\zeta}$ is independent with unit. Experimentally we choose $\zeta = 0.5q_{max}^2$ for good numerical stability, where q_{max} is the maximum q value for DWI signals. In DPI_0, the radial integration $F_{nl}(R|\zeta)$ can not be analytically solved, and [Descoteaux et al., 2010] introduced q_{max} and approximate $F_{nl}(R|\zeta)$ using the integration from 0 to q_{max}, which can be seen as follows.

$$F_{0l} = (-1)^{l/2}q_{max}^{l+1.5}\zeta^{-0.5l}R^{-1.5}J_{l+1.5}(2\pi q_{max}R) \tag{5.73}$$

$$F_{1l} = (-1)^{l/2}R^{-1.5}\zeta^{0.5l+0.5}(\frac{(\pi R)^{l-0.5}}{\Gamma(l+0.5)} - \frac{J_{l-0.5}(2\pi q_{max}R)}{q_{max}^{l-0.5}}) \tag{5.74}$$

Note that in DPI_0 the estimated signal in Eq. (5.71) does not satisfy the prior $E(0) = 1$. Actually, when q tends to zero, the irregular term $R_1l(q)$ tends to infinity,

which means $E(0)$ has no definition in DPI_0. Moreover, DPI_0 considers polynomial decay in radial part, which does not satisfy the prior of Gaussian-like decay. Thus we propose DPI_1 which only uses regular terms in DPI_0 and considers Gaussian-like decay, i.e.

$$E(\mathbf{q}) = \sum_{l=0}^{L} \sum_{m=-l}^{l} c_{lm} R_l(q) Y_l^m(\mathbf{u}), \quad R_l(q) = (\frac{q}{\sqrt{\zeta}})^l \exp\left(-\frac{q^2}{2\zeta}\right) \tag{5.75}$$

Note that the basis in DPI_1 $(\frac{q}{\sqrt{\zeta}})^l \exp\left(-\frac{q^2}{2\zeta}\right) Y_l^m(\mathbf{u})$ is proportional to SHO-3D basis $G_{nl}(q)Y_l^m(\mathbf{u})$ with $n = \frac{l}{2}$. Thus the radial integration of DPI_1 basis can be solved by considering the result for radial integration of SHO-3D basis. Since SHO-3D basis is complete and DPI_1 basis is just a part of SHO-3D basis, the basis in DPI_1 is not complete. We will discuss it later.

DOT_0 (exact QBI), DOT_1 and DOT_2 (SPNP)

The original DOT, called DOT_0 here, assumes $E(\mathbf{q})$ follows mono-exponential decay [Özarslan et al., 2006], i.e.

$$E(\mathbf{q}) = \exp(-4\pi^2 \tau q^2 D(\mathbf{u}))$$

Actually DOT_0 cannot be contained in AFT-SC framework because $E(\mathbf{q})$ in DOT_0 cannot be separated into radial part and spherical part Eq. (5.53). [Özarslan et al., 2006] analytically solved the radial integration in Eq. (5.76) for the given samples $\{E(\mathbf{q}_i)\}$. Then the inner product between $I_l(R, \mathbf{u})$ and $Y_l^m(\mathbf{u})$ in the spherical integration Eq. (5.77) was solved numerically using least square fitting for the samples $\{I_l(R, \mathbf{u}_i)\}$ obtained in Eq. (5.76).

$$I_l(R, \mathbf{u}) = 4\pi(-1)^{l/2} \int_0^\infty E(\mathbf{q}) j_l(2\pi q R) q^2 dq = \frac{R^l \Gamma(0.5l + 1.5)\, {}_1F_1(0.5l + 1.5, l + 1.5, -\frac{R^2}{4\tau D(\mathbf{u})})}{(-1)^{l/2} 2^{l+1} \pi^{0.5} (D(\mathbf{u})\tau)^{0.5l+1.5} \Gamma(l + 1.5)} \tag{5.76}$$

$$P(\mathbf{R}) = \left\{ \int_{\mathbb{S}^2} I_l(R, \mathbf{u}) Y_l^m(\mathbf{u}) d\mathbf{u} \right\} Y_l^m(\mathbf{r}) \tag{5.77}$$

Please note there is no full representation for $E(\mathbf{q})$ and $P(\mathbf{R})$, because one needs to re-estimate $P(\mathbf{R})$ in different R. Note that the mono-exponential decay model in DOT_0 can be used to analytically estimate the ODF by Wedeen $\Phi_w(\mathbf{r})$. See subsection 4.3.4.

Although DOT_0 is not contained in AFT-SC, it is still possible to explain DOT_0 in AFT-SC framework. We can represent

$$D(\mathbf{u}) = \sum_{lm} b_l^m Y_l^m(\mathbf{u}) \tag{5.78}$$

and expand the signal as

$$E(\mathbf{q}) = \exp(-4\pi^2 \tau q^2 D(\mathbf{u})) = \sum_{n=0}^{\infty} \frac{(-4\pi^2 \tau q^2)^n}{n!} (\sum_{lm} b_l^m Y_l^m(\mathbf{u}))^n = \sum_{nlm} a_{nlm} q^{2n} Y_l^m(\mathbf{u})$$

where a_{nlm} can be analytically obtained from b_l^m by expanding the product of two SHs in terms of SHs itself based on the integration in Eq. (2.35). We call the modified representation of DOT as DOT_1, which separates spherical part and radial part. DOT_1 is

equivalent with DOT_0 if $\{b_l^m\}$ is estimated from DWI samples and $\{a_{nlm}\}$ is calculated from $\{b_l^m\}$. But alternatively we can fit the signal directly with the representation $E(\mathbf{q}) = \sum_{nlm} a_{nlm}(\frac{q^2}{\zeta})^n Y_l^m(\mathbf{u})$, where ζ is the fixed scale parameter. In this case DOT_1 is not DOT any more and more similar with DPI. It does not assume mono-exponential decay and works for multiple shell data. Similarly with DPI there is no analytical form for radial integration. Thus we introduce q_{max} for the integration from 0 to q_{max} [Gradshteyn and Ryzhik, 2007]. See table 5.2, where ${}_1F_2$ is the generalized hypergeometric function.

Note that when the ADC is represented in Eq. (5.78), the signal is represented in GHOT model in Eq. (5.36) with $N = 1$. Thus we have another expansion of $E(\mathbf{q})$ in DOT as

$$\begin{aligned} E(\mathbf{q}) &= \exp(-4\pi^2\tau q^2 D(\mathbf{u})) = \exp(-4\pi^2\tau b_0^0 Y_0^0 q^2)\exp\left(\sum_{lm,l\neq 0} 4\pi^2\tau q^2 b_l^m Y_l^m(\mathbf{u})\right) \\ &= \exp\left(-\frac{q^2}{2\zeta}\right)\left(\sum_{n=0}^{\infty}\frac{(-4\pi^2\tau q^2)^n}{n!}\left(\sum_{lm,l\neq 0} b_l^m Y_l^m(\mathbf{u})\right)^n\right) \\ &= \exp\left(-\frac{q^2}{2\zeta}\right)\sum_{nlm} a_{nlm}\left(\frac{q^2}{\zeta}\right)^n Y_l^m(\mathbf{u}) \end{aligned}$$

where $\zeta = \frac{1}{8\pi^{5/2}\tau b_0^0}$. Now we represent $E(\mathbf{q})$ in SPNP model in Eq. (5.35), which separates spherical part and radial part and can be called as DOT_2. DOT_2 is equivalent with DOT_0 if $\{b_l^m\}$ is estimated from DWI samples and $\{a_{nlm}\}$ is calculated from $\{b_l^m\}$. In this sense, DOT_2 (SPNP) model is a generalization of HOT model in DOT. Also we can fit the signal directly using SPNP basis.

5.3 THEORETICAL COMPARISONS

Based on AFT-SC framework, it seems to be an easy job to deduce an analytical EAP reconstruction and we can have many analytical EAP estimation including the listed methods in Table 5.2. Then which one is better? We propose some criteria for evaluation.

5.3.1 Some Criteria for Evaluation

Completeness. In AFT-SC, $E(\mathbf{q})$ is represented by a linear combination of some basis functions $\{B_i(\mathbf{q})\}$ in Eq. (5.53). $\{B_i(\mathbf{q})\}$ is complete if it can represent any symmetric square integrable $E(\mathbf{q})$ in 3D space. Completeness means more samples we have, better reconstruction we get. If all DWI samples are known in whole $\mathbf{q}$ space, the samples will be fitted without any modeling error.

QBI assumes $E(\mathbf{q})$ exists only on the sphere $\mathbb{S}^2$, which means the radial decay is a delta function. DOT_0 and exact QBI assume mono-exponential decay. So QBI and DOT_0 are not complete in $\mathbb{R}^3$. SHO-3D basis is complete because it is a combination of the eigenfunctions in several Sturm-Liouville equations when solving the SHO-3D

equation in Eq. (2.68). SPF basis is complete because $\{Y_l^m(\mathbf{u})\}$ is complete in $\mathbb{S}^2$ and $L_n^{1/2}(x)$ is complete in $[0,\infty)$ with the weight $x^{1/2}\exp(-x)$.

DPI_0 is not complete because not every function satisfies Laplace's equation. Based on Weierstrass theorem the continuous $E(\mathbf{q})$ in the ball with a given radius can be uniformly approximated by a polynomial function, i.e.

$$E(\mathbf{q}) = \sum_{ijk} a_{ijk} q_x^i q_y^j q_z^k$$

Because $E(\mathbf{q})$ is symmetric, $i+j+k$ is even, say $2n$. Then by representing every monomial via spherical coordinate, we have

$$E(\mathbf{q}) = \sum_{nlm} a_{nlm} (\frac{q}{\sqrt{\zeta}})^{2n} Y_l^m(\mathbf{u})$$

which means the basis in DOT_1 is complete for $E(\mathbf{q})$ inside a ball. However for both DPI and DOT_1, since only finite terms are used, the represented $E(\mathbf{q})$ tends to ∞ as q increases. It contradicts with the fact that $E(\mathbf{q})$ tends to 0 when q increases, which means DPI_0 and DOT_1 have intrinsic modeling error in the region with large q. So the q_{max} is needed for an incomplete integration instead of the complete integration. However, if q_{max} is chosen as the maximal q values of DWI signals like [Descoteaux et al., 2010], the estimated EAPs are less anisotropic because of incomplete integration, which means DPI_0 and DOT_1 cannot work well with DWIs with only small b values. If q_{max} is chosen as a large value, the estimated EAPs are likely noisy because of the modeling error in the area with large q value.

Representability. We know three priors (P1, P2 and P3) of the diffusion signal $E(\mathbf{q})$.

- **P1**: $E(0) = 1$ because $\int_{\mathbb{R}^3} P(\mathbf{R})\text{d}\mathbf{R} = 1$.
- **P2**: $E(\mathbf{q})$ tends to 0 when q tends to ∞.
- **P3**: $E(\mathbf{q})$ radially decays like (but NOT) a Gaussian function.

Please note that the estimated $P(\mathbf{R})$ is ***globally*** affected by $E(\mathbf{q})$ in whole $\mathbf{q}$ space because Fourier Transform is a global transform. So even though one method can fit the given samples of the DWI well, it does not mean the estimated EAP is good. For the signal fitting in existing works, the given samples have been well considered in the estimation process. The results will be better if the model can also consider these priors. The model which satisfies these priors has good results, even if it is not complete, e.g. mixture of tensor model.

QBI satisfies none of the priors. In practice we find the EAP profile in QBI has many negative values especially when $b > 1500 s/mm^2$ and $R > 10\mu m$, and the EAP profile in QBI is very smooth if b or R is small. See Fig. 5.4 for the reason, which is because delta function decay assumption in QBI is lack of representability. DOT_0 automatically considers the first two priors, while it assumes mono-exponential decay. SHORE and SPFI consider the second and third in their model. For $E(0) = 1$, SHORE, SPFI and DOT_1 can consider the prior $E(0) = 1$ into estimation process as

demonstrated in subsection 5.1.3. While DPI cannot consider $E(0) = 1$ in estimation because $E(0)$ does not exist in DPI model due to the irregular term in its radial basis. Moreover, in DPI as well as DOT_1 $E(\mathbf{q})$ tends to ∞ when q increases due to the regular term in its radial basis, and infinite radial terms are needed if one uses polynomials to approximate a Gauss-like decay. So even though DPI may fit the given DWI signals well as shown in [Descoteaux et al., 2010], the estimated EAP is problematic. Motivated by SHORE and SPFI, we can avoid the problem by adding exponential term into the basis function of DPI and DOT_1 and ignoring irregular terms in DPI. Then we have two new methods, DPI_1 where

$$E(\mathbf{q}) = \sum_{lm} a_{lm} (\frac{q^2}{\zeta})^{l/2} \exp(-\frac{q^2}{2\zeta}) Y_l^m(\mathbf{u})$$

and DOT_2 using SPNP basis in Eq. (5.35), where

$$E(\mathbf{q}) = \sum_{nlm} a_{nlm} (\frac{q^2}{\zeta})^{n} \exp(-\frac{q^2}{2\zeta}) Y_l^m(\mathbf{u})$$

Similarly with SHORE and SPFI, they both satisfy all three priors. DOT_2 is equivalent with SPFI because after Gram-Schmidt orthonormalization in radial part, DOT_2 (SPNP) will become SPFI. See Lemma 5.2. Thus DOT_2 is complete. While DPI_1 is not, because DPI_1 now assumes $E(\mathbf{q})$ is a harmonic polynomial multiplied by a Gaussian, or because the basis in DPI_1 is a part of SHO-3D basis. DOT_2 and DPI_1 both have analytical EAP forms based on the previous derivations for SPFI and SHORE and we do not need q_{max} any more. See table 5.2 for the analytical formulae.

Separation information between spherical and radial parts. Please note the important difference between the basis functions in SHORE, SPFI and DPI_0. DPI_0 is lack of representability partially because it coupled radial information and spherical information together. For example DPI_0 cannot represent an isotropic Gaussian function, because isotropic function forces $l = 0$ in SHs in spherical part, then in radial part $R_{00} = 1$ and $R_{10} = (\frac{q}{\sqrt{\zeta}})^{-1}$ which contradicts with Gaussian-like function. In SPFI the radial part and spherical part are completely separated, which allows one chooses higher order in spherical part but low order in radial part. While in SHORE, $n-l/2 \geq 0$ is forced in Eq. (5.60), which means for a given l in spherical part, the corresponding order of power in radial part is $2n \geq l$. In SPFI, the minimal order for anisotropy diffusion is $N = 1$, $L = 4$, which means it contains SH of order $0, 2, 4$ and power $0, 2$ in radial part. In SHORE, the minimal order is $N = 2$ ($N_{max} = 4$ in Eq. (4.84) [Özarslan et al., 2009]), which means it contains SH of order $0, 2, 4$ and power $0, 2, 4$ in radial part. Besides it was proved in theorem 5.3 that the function space spanned by SHO-3D basis $\{G_{nl} Y_l^m\}_{n \leq N}$ with order $n \leq N$ is a subspace of the space spanned by SPF basis $\{G_{n0} Y_l^m\}_{n \leq N, l \leq 2N}$ with radial order $n \leq N$ and spherical order $l \leq 2N$. Thus if order N is enough to represent the signal in SHORE, radial order N and spherical order $2N$ is also enough for representation in SPF basis. However if limited samples with low SNR are given for estimating the coefficients, a truncated basis is needed to avoid overfitting. Then SHO-3D basis with higher power order in radial part may have some overfitting effects, and the effect will be enhanced when larger N (6 or 8) is used. Please see the subsection 5.3.2 for more theoretical comparisons.

Orthogonality and stability. In approximation theory, complete orthonormal basis is preferred because monomial basis is known to have poor numerical stability and the coefficients under orthonormal basis are independent with the basis order chosen in Least Square approximation if all samples are known. For example, if exhaustive samples of diffusion signal are known in a single shell, the coefficients of SH basis are independent with the chosen order in least square fitting, while the coefficients of HOT basis is dependent on the chosen order [Özarslan and Mareci, 2003], although these two bases are both complete in $\mathbb{S}^2$. The bases in DOT_1, DOT_2 and DPI_0, DPI_0 are not orthogonal. SHORE and SPFI use orthonormal bases in $\mathbb{R}^3$ while QBI and DOT_0 use orthonormal bases only in $\mathbb{S}^2$. When orthonormal basis $\{B_k(\mathbf{q})\}$ is used to represent $E(\mathbf{q})$, its Fourier dual basis $\{D_k(\mathbf{R})\}$ that represents $P(\mathbf{R})$ is still orthonormal because of the Parseval's theorem.

Single shell and multiple shells. These mHARDI methods can work for multiple shell data. However, when only single shell data are given, these methods are unstable in general. Let's take SPFI as an example. In SPFI, for a given q_0, the ratio between two basis functions

$$\frac{G_n(q_0)Y_l^m(\mathbf{u})}{G_{n'}(q_0)Y_l^m(\mathbf{u})}$$

is a constant independent with $\mathbf{u}$, which means these two basis functions with same l and m but different n are undistinguishable for single shell data. When least square is used to estimate the coefficients, the basis matrix is rank deficient with very large conditional number. To solve this problem, we can consider the constraint $E(0) = 1$. In SPFI, we have shown that when considering this constraint, $\{a_{0lm}\}_{l\leq L}$ is dependent on $\{a_{nlm}\}_{0<n\leq N,l\leq L}$ and only the basis function set $\{G_n(q)Y_l^m(\mathbf{u})\}_{0<n\leq N,l\leq L}$ is used in estimation. So SPFI is stable for single shell data if $N = 1$ with only $\{G_0(q)Y_l^m\}_{l\leq L}$ used in estimation. If $N \geq 2$, it needs to add some regularization to make the estimation stable. However SHORE is not stable for single shell data, because the constraint $E(0) = 1$ only means a_{000} is redundant, and we can not separate the basis $G_{nl}(\mathbf{u})Y_l^m(\mathbf{u})$ and $G_{n'l}Y_l^m(\mathbf{u})$. DPI_0 can not consider the constraint $E(0) = 1$, so it is unstable. It is possible to add some regularization in SHORE and DPI_0 such that the estimation results are meaningful. The same analysis can be performed in DPI_1, DOT_1 and DOT_2. DPI_1 works with single shell data without regularization because its radial part is fully coherent with spherical part. DOT_1 and DOT_2 need some regularization for single shell data, because $E(0) = 1$ makes only one basis redundant in estimation like SHORE.

Summary. SHORE and SPFI use 3D complete orthonormal basis and have the best representability. SHORE requires higher orders in radial part than SPFI which completely separate the information between spherical and radial part. SPFI is well appropriate for both single and multiple shell data, while SHORE and DPI_0 need some regularization to obtain stable results from single shell data. QBI and DPI_0 are lack of representability, because of the delta function decay assumption in QBI and regular and irregular monomials in radial basis of DPI_0. DPI_0 works only for the data with large b values due to the effect of q_{max}. Please see Table 5.3 for more information.

Table 5.3: Some criteria for evaluating methods. P1, P2 and P3 are three priors. Note that although some mHARDI methods are unstable for single shell data, it is possible to add some regularization to obtain meaningful results.

method	$R_k(q)$	$Q_k(\mathbf{u})$	Completeness	P1	P2	P3	orthogonal	single shell	Separation
QBI	$R(q)=\delta(q-q_0)$	$Y_l^m(\mathbf{u})$	in $\mathbb{S}^2$	No	No	No	in $\mathbb{S}^2$	Yes	Yes
SHORE	$R_{nl}(q)=G_{nl}(q\|\zeta)$	$Y_l^m(\mathbf{u})$	Yes	Yes	Yes	Yes	Yes	No	No
SPFI	$R_n(q)=G_n(q\|\zeta)$	$Y_l^m(\mathbf{u})$	Yes	Yes	Yes	Yes	Yes	Yes, if $N=1$	Yes
DPI_0	$R_{0l}(q)=(\frac{q}{\sqrt{\zeta}})^l$ $R_{1l}(q)=(\frac{q}{\sqrt{\zeta}})^{-l-1}$	$Y_l^m(\mathbf{u})$	No	No	No	No	No	No	No
DPI_1	$R_l(q)=(\frac{q^2}{\zeta})^{l/2}\exp(-\frac{q^2}{2\zeta})$	$Y_l^m(\mathbf{u})$	No	Yes	Yes	Yes	No	Yes	No
DOT_1	$R_n(q)=(\frac{q^2}{\zeta})^n$	$Y_l^m(\mathbf{u})$	in a ball	Yes	No	No	No	No	Yes
DOT_2 (SPNP)	$R_n(q)=(\frac{q^2}{\zeta})^n\exp(-\frac{q^2}{2\zeta})$	$Y_l^m(\mathbf{u})$	Yes	Yes	Yes	Yes	No	No	Yes
DOT_0	—	—	in $\mathbb{S}^2$	Yes	Yes	No	in $\mathbb{S}^2$	Yes	Yes

5.3.2 More Comparisons between SPF basis and SHO basis

In this subsection, we would like to analyze theoretically more similarities and differences between SPF basis, SHO-1D basis and SHO-3D basis. See the following analysis.

Polynomial basis and SHO basis

First SHO-1D basis is equivalent with the Gaussian polynomial basis in 1D.

Lemma 5.1. *For any given integer N, SHO-1D basis in Eq. (2.60) with $0 \le n \le N$ and the Gaussian polynomial basis function set $\{e^{-\frac{q^2}{2\zeta}}(\frac{q}{\sqrt{\zeta}})^n\}_{n\le N}$ forms the same function space.*

Proof. It is straightforward because $\{q^n\}_{n\le N}$ and $\{H_n(\frac{x}{\sqrt{\zeta}})\}_{n\le N}$ are two bases of polynomial space $\mathcal{P}_N(\mathbb{R}^1)$. □

Definition 5.2 (SHO-1D basis and product of SHO-1D basis in 3D). *SHO-1D basis is defined in Eq. (2.60) and denoted by $B_n^{SHO1}(x)$. The product of SHO-1D basis in 3D space with variable $\mathbf{q}=(q_x,q_y,q_z)^T$ is called as the product of SHO-1D basis in 3D space (SHO-1D3) basis and defined as*

$$\begin{aligned} B_{n_1n_2n_3}^{SHO1}(\mathbf{q}|\zeta) &\overset{\text{def}}{=} B_{n_1}^{SHO1}(q_x|\zeta)B_{n_1}^{SHO1}(q_y|\zeta)B_{n_1}^{SHO1}(q_z|\zeta) \\ &= 2^{-\frac{n_1+n_2+n_3}{2}}(n_1!n_2!n_3!)^{-\frac{1}{2}}(\pi\zeta)^{-\frac{3}{4}}e^{\frac{\|\mathbf{x}\|^2}{2\zeta}}H_{n_1}\left(\frac{q_x}{\sqrt{\zeta}}\right)H_{n_2}\left(\frac{q_y}{\sqrt{\zeta}}\right)H_{n_3}\left(\frac{q_z}{\sqrt{\zeta}}\right) \end{aligned} \tag{5.79}$$

where we only consider the even order $n_1+n_2+n_3=n$, because $E(\mathbf{q})$ is symmetric.

Then based on the above definition and lemma, the SHO-1D3 basis is also equivalent with Gaussian polynomial basis in 3D.

Definition 5.3 (Spherical Polar Polynomial basis and Generalized DTI model). *Spherical Polar Polynomial (SPP) basis is defined as*

$$B_{nlm}^{SPP}(\mathbf{q}|\zeta) = \exp\left(-\frac{q^2}{2\zeta}\right)\left(\frac{q^2}{\zeta}\right)^n Y_l^m(\mathbf{u}), \quad n \geq l/2 \tag{5.80}$$

where n is integer and $n \geq l/2$. GDTI model is defined as

$$E(\mathbf{q}) = \exp\left(-\sum_{n=2}^{N}\sum_{n_1+n_2+n_3=n} D_{n_1n_2n_3}\left(\frac{q_x}{\sqrt{\zeta}}\right)^{n_1}\left(\frac{q_y}{\sqrt{\zeta}}\right)^{n_2}\left(\frac{q_z}{\sqrt{\zeta}}\right)^{n_3}\right) \tag{5.81}$$

which has been used in GDTI in [Liu et al., 2003, 2004]. Note that n starts from 2 in GDTI, because $E(0) = 1$ means $D_{000} = 0$.

Theorem 5.2. *For any given integer N, the following three bases are equivalent, i.e. they form the same function space.*

- *SHO-1D3 basis $\{B_{n_1n_2n_3}^{SHO1}(\mathbf{q}|\zeta)\}_{n_1+n_2+n_3=n\leq 2N}$ defined in Eq. (5.79) with even order $0 \leq n \leq 2N$.*
- *SPP basis $\{B_{nlm}^{SPP}(\mathbf{q}|\zeta)\}_{n\leq N}$ defined in Eq. (5.80) with order $0 \leq n \leq N$.*
- *SHO-3D basis $\{B_{nlm}^{SHO3}(\mathbf{q})\}_{n\leq N}$ defined in Eq. (5.60) with order $0 \leq n \leq N$.*

Proof. Because of Eq. (2.63) and Eq. (2.67), $\left\{\left(\frac{q^2}{\zeta}\right)^{n-l/2}\right\}_{n\leq N}$ and $\left\{L_{n-l/2}^{l+1/2}\left(\frac{q^2}{\zeta}\right)\right\}_{n\leq N}$ are equivalent. Thus SPP basis $\{B_{nlm}^{\mathrm{SPP}}(\mathbf{q}|\zeta)\}_{n\leq N}$ and SHO-3D basis $\{B_{nlm}^{\mathrm{SHO3}}(\mathbf{q})\}_{n\leq N}$ are equivalent. Thus we only need to prove SHO-1D3 basis is equivalent with SPP basis.

Because of Lemma 5.1, SHO-1D3 basis $\{B_{n_1n_2n_3}^{\mathrm{SHO1}}(\mathbf{q}|\zeta)\}_{n_1+n_2+n_3=n\leq 2N}$ is equivalent with Gaussian polynomial basis $\{B_{n_1n_2n_3}^{\mathrm{GP}}(\mathbf{q}|\zeta)\}_{n_1+n_2+n_3=n\leq 2N}$, where $B_{n_1n_2n_3}^{\mathrm{GP}}(\mathbf{q}|\zeta)$ is defined as

$$B_{n_1n_2n_3}^{\mathrm{GP}}(\mathbf{q}|\zeta) = \exp\left(-\frac{q^2}{2\zeta}\right)\left(\frac{q_x}{\sqrt{\zeta}}\right)^{n_1}\left(\frac{q_y}{\sqrt{\zeta}}\right)^{n_2}\left(\frac{q_z}{\sqrt{\zeta}}\right)^{n_3} \tag{5.82}$$

Then based on the harmonic decomposition theorem 2.4 and Theorem 2.5, $B_{n_1n_2n_3}^{\mathrm{GP}}(\mathbf{q}|\zeta)$ can be separated as

$$B_{n_1n_2n_3}^{\mathrm{GP}}(\mathbf{q}|\zeta) = \sum_{j=0}^{\frac{n}{2}}\left(\frac{q^2}{\zeta}\right)^j \sum_{m=-(n-2j)}^{n-2j} a_{jm}\left(\frac{q^2}{\zeta}\right)^{n/2-j} Y_{n/2-j}^m(\mathbf{u}) = \sum_{l=0}^{n}\sum_{m=-l}^{l} a_{lm}\left(\frac{q^2}{\zeta}\right)^{n/2} Y_l^m(\mathbf{u})$$

where $n = n_1 + n_2 + n_3$ is even. Thus Gaussian polynomial basis $\{B_{n_1n_2n_3}^{\mathrm{GP}}(\mathbf{q}|\zeta)\}_{n_1+n_2+n_3=n\leq 2N}$ is equivalent with SPP basis $\{B_{nlm}^{\mathrm{SPP}}(\mathbf{q}|\zeta)\}_{n\leq N}$. Then SHO-1D3 basis $\{B_{n_1n_2n_3}^{\mathrm{SHO1}}(\mathbf{q}|\zeta)\}_{n_1+n_2+n_3=n\leq 2N}$ is also equivalent with SPP basis $\{B_{nlm}^{\mathrm{SPP}}(\mathbf{q}|\zeta)\}_{n\leq N}$. □

The above theorem shows that SHO-3D basis and the triple product of SHO-1D basis are equivalent. That means the methods which fit exhaustive signal samples using these two bases will obtain theoretically the same results, although numerically the results are slightly different. Note that GDTI [Liu et al., 2004] represents EAP using the products of probabilists' Hermite polynomials in Eq. (4.37), and SHO-1D3 represents EAP using products of Physicists' Hermite polynomials. Thus GDTI and

SHORE are theoretically equivalent. However, the estimation of diffusion coefficients in GDTI in Eq. (4.34) is based on fitting $\ln E(\mathbf{q})$, which may bring more numerical error than fitting $E(\mathbf{q})$ directly. That is because $E(\mathbf{q}) \leq 1$ and $E(\mathbf{q})$ decays as q increase, which makes the logarithm operation emphasize more the samples of $E(\mathbf{q})$ for large q value. However, $E(\mathbf{q})$ for large q value has lower SNR. Thus SHORE and SPFI based on fitting $E(\mathbf{q})$ are better than GDTI based on fitting $\ln E(\mathbf{q})$. Moreover, the closed form in GDTI is based on the estimation of cumulants, which is theoretically problematic due to the truncation [Marcinkiewicz, 1939]. Last, the estimation of the PDF from its cumulants is known to be very problematic [Blinnikov and Moessner, 1998; Ghosh et al., 2010].

Non-polynomial basis and SPF basis

Lemma 5.2. *For any given integer N and even integer L, SPF basis $\{B_{nlm}^{SPF}(\mathbf{q}|\zeta)\}_{n\leq N,l\leq L}$ defined in Eq. (5.1) and SPNP basis $\{B_{nlm}^{SPNP}(\mathbf{q}|\zeta)\}_{n\leq N,l\leq L}$ defined in Eq. (5.35) are equivalent, i.e. they form the same function space.*

Proof. It is straightforward, because the space spanned by $\left\{L_n^{1/2}\left(\frac{q^2}{\zeta}\right)\right\}_{n\leq N}$ is the same space spanned by $\left\{\left(\frac{q^2}{\zeta}\right)^n\right\}_{n\leq N}$. □

The space spanned by SHO-3D basis and SPF basis and the MSE

Based on Theorem 5.2, SHO-3D basis is essentially the SPP basis in Eq. (5.80). While Lemma 5.2 shows that SPF basis is essentially the SPNP basis in Eq. (5.35). Compared to SPP basis, SPNP basis has some non-polynomial term with $n < l/2$, which is essentially the difference between SHO-3D basis and SPF basis. This difference brings the following theorem.

Theorem 5.3. *For any given integer N, if we denote the space spanned by SPF basis $\{B_{nlm}^{SPF}(\mathbf{q}|\zeta)\}_{n\leq N,l\leq 2N}$ as $Span\{B_{nlm}^{SPF}(\mathbf{q}|\zeta)\}_{n\leq N,l\leq 2N}$, the space spanned by SHO-3D basis as $Span\{B_{nlm}^{SHO3}(\mathbf{q}|\zeta)\}_{n\leq N}$, then we have*

$$Span\{B_{nlm}^{SHO3}(\mathbf{q}|\zeta)\}_{n\leq N} \subset Span\{B_{nlm}^{SPF}(\mathbf{q}|\zeta)\}_{n\leq N,l\leq 2N} \tag{5.83}$$

For any finite N', we have

$$Span\{B_{nlm}^{SHO3}(\mathbf{q}|\zeta)\}_{n\leq N'} \not\supseteq Span\{B_{nlm}^{SPF}(\mathbf{q}|\zeta)\}_{n\leq N,l\leq 2N} \tag{5.84}$$

Proof. Based on the definition of SHO-3D basis in 5.60, $G_{nl}(q)$ is a Gaussian polynomial function with polynomial order $2n$. Thus $G_{nl}(q)$ can be represented as

$$G_{nl}(q|\zeta) = \exp\left(-\frac{q^2}{2\zeta}\right)\sum_{j=0}^{n} a_{jl}L_j^{1/2}(q|\zeta)$$

where a_{jl} is constant which can be obtained based on Eq. (2.63) and Eq. (2.67). Then the SHO-3D basis $B^{\text{SHO3}}_{nlm}(\mathbf{q}|\zeta)$ can be written as

$$B^{\text{SHO3}}_{nlm}(\mathbf{q}|\zeta) = G_{nl}(q)Y^m_l(\mathbf{u}) = \exp\left(-\frac{q^2}{2\zeta}\right)\sum_{j=0}^{n} a_{jl}L^{1/2}_j(q|\zeta)Y^m_l(\mathbf{u})$$

$$= \sum_{j=0}^{n}\frac{a_{jl}}{\kappa_n(\zeta)}B^{\text{SPF}}_{jlm}(\mathbf{q}|\zeta)$$

It proves Eq. (5.83).

The basis set $\{B^{\text{SPF}}_{nlm}(\mathbf{q}|\zeta)\}_{n\le N, l\le 2N}$ can be separated into two parts, i.e.

$$\{B^{\text{SPF}}_{nlm}(\mathbf{q}|\zeta)\}_{n<l/2\le N}, \quad \{B^{\text{SPF}}_{nlm}(\mathbf{q}|\zeta)\}_{l/2\le n\le N}$$

$B^{\text{SPF}}_{nlm}(\mathbf{q}|\zeta)$ with $n \ge l/2$ can be represented by SPP basis $\{B^{SPP}_{jlm}(\mathbf{q}|\zeta)\}_{j\le n}$. Then based on Theorem 5.2, $B^{\text{SPF}}_{nlm}(\mathbf{q}|\zeta)$ with $n \ge l/2$ can be represented by SHO-3D basis $\{B^{SHO3}_{jlm}(\mathbf{q}|\zeta)\}_{j\le n}$. For the SPF basis $\{B^{\text{SPF}}_{nlm}\}_{n<l/2}$, let's assume that it can be represented by SHO-3D basis with finite order N', i.e.

$$B^{\text{SPF}}_{nlm}(\mathbf{q}) = \sum_{j=0}^{N'} a_j B^{\text{SHO3}}_{jlm}(\mathbf{q}),\ n < l/2$$

Note that the summation is only over j because of the orthogonality of SH basis. Because of the equivalence of SPF basis and SPNP basis in Lemma 5.2 and the equivalence of SHO-3D basis and SPP basis in Theorem 5.2, we have

$$B^{\text{SPNP}}_{n'lm}(\mathbf{q}) = \sum_{n=0}^{n'} b_n B^{\text{SPF}}_{nlm}(\mathbf{q}), \quad B^{\text{SHO3}}_{jlm}(\mathbf{q}) = \sum_{j'=0}^{j} c_{j'} B^{\text{SPP}}_{j'lm}(\mathbf{q})$$

Then SPNP can be represented linearly by SPP basis with finite N' as

$$B^{\text{SPNP}}_{n'lm}(\mathbf{q}) = \sum_{n=0}^{n'}\sum_{j=0}^{N'}\sum_{j'=0}^{j} b_n a_j c_{j'} B^{\text{SPP}}_{j'lm}(\mathbf{q})$$

This is equivalent with that $x^{n'}$ can be represented by basis $\{x^{j'}\}_{j'>l}$ with finite order N'. However, for fixed l and $n' < l$, that is not true. So with any finite order N', SHO-3D basis $\{B^{\text{SHO3}}_{nlm}(\mathbf{q}|\zeta)\}_{n\le N'}$ can not represent $B^{\text{SPF}}_{nlm}(\mathbf{q}|\zeta)$ with $n < l/2$, which proves Eq. (5.84). □

This theorem shows that when we use SHO-3D basis $\{B^{\text{SHO3}}_{nlm}(\mathbf{q}|\zeta)\}_{n\le N}$ and SPF basis $\{B^{\text{SPF}}_{nlm}(\mathbf{q}|\zeta)\}_{n\le N, l\le 2N}$ to represent signal, SPF basis obtain lower representation MSE, i.e. $\|E(\mathbf{q}) - \sum_{n=0}^{N}\sum_{l=0}^{2N}\sum_{m=-l}^{l} a_{nlm}B^{\text{SPF}}_{nlm}(\mathbf{q}|\zeta)\|^2 \le \|E(\mathbf{q}) - \sum_{n=0}^{N}\sum_{l=0}^{2n}\sum_{m=-l}^{l} a_{nlm}B^{\text{SHO3}}_{nlm}(\mathbf{q}|\zeta)\|^2$. However, please note that SPF has more number of basis functions because it also contains non-polynomial terms. Please note that the non-polynomial terms is important in HARDI model. For example, HOT model used in DOT and exact QBI is essentially a non-polynomial model, where the ADC is represented by low order in radial part, but high order in spherical part. SPFI can be seen as a generalization of HOT model, because SPF basis is equivalent with SPNP basis which is used in DOT_2.

Please note the difference when representing GDTI model and GHOT model via SPF basis and SHO-3D basis.

- For GDTI model, the signal $E(\mathbf{q})$ is

$$E(\mathbf{q}) = \exp\left(-\sum_{n=2}^{N}\sum_{n_1+n_2+n_3=n} D_{n_1n_2n_3}\left(\frac{q_x}{\sqrt{\zeta_1}}\right)^{n_1}\left(\frac{q_y}{\sqrt{\zeta_1}}\right)^{n_2}\left(\frac{q_z}{\sqrt{\zeta_1}}\right)^{n_3}\right)$$
$$= \exp\left(-\frac{q^2}{2\zeta}\right)\exp\left(\frac{1}{2\zeta}q^2 - \sum_{n=2}^{N}\sum_{n_1+n_2+n_3=n} D_{n_1n_2n_3}\left(\frac{q_x}{\sqrt{\zeta_1}}\right)^{n_1}\left(\frac{q_y}{\sqrt{\zeta_1}}\right)^{n_2}\left(\frac{q_z}{\sqrt{\zeta_1}}\right)^{n_3}\right)$$

  where the scale ζ is the well chosen to be close to the optimal one based on the coefficient of q^2, i.e. $\zeta = \frac{\zeta_1}{2(D_{200}+D_{020}+D_{002})}$. Its truncated Taylor expansion with maximal order N denoted by $E_N(\mathbf{q})$ is given as

$$E_N(\mathbf{q}) = \exp\left(-\frac{q^2}{2\zeta}\right)\sum_{n=0}^{N}\sum_{n_1+n_2+n_3=n} d_{n_1n_2n_3}q_x^{n_1}q_y^{n_2}q_z^{n_3}$$

  which contains the main energy of $E(\mathbf{q})$. In the above formula, $E_N(\mathbf{q})$ is represented by SHO-1D3 basis, thus it is in the space $\text{Span}\{B_{nlm}^{\text{SHO3}}(\mathbf{q}|\zeta)\}_{n\le N} \subset \text{Span}\{B_{nlm}^{\text{SPF}}(\mathbf{q}|\zeta)\}_{n\le N, l\le 2N}$. So the representation MSE of $E(\mathbf{q})$ by SHO-3D basis and SPF basis are similar for GDTI model, although SPF basis still has less MSE due to the residual part $E(\mathbf{q}) - E_N(\mathbf{q})$. Based on the linearity of the basis representation, the MSE when representing the mixture of GDTI model by SHO-3D basis and SPF basis are also similar. As a special case, the MSE when representing mixture of tensor model by these two bases are also similar, although SPF basis still have less MSE.

- For GHOT model, the signal $E(\mathbf{q})$ is

$$E(\mathbf{q}) = \exp\left(-\sum_{n=1}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} b_{nlm}\left(\frac{q^2}{\zeta}\right)^n Y_l^m(\mathbf{u})\right)$$
$$= \exp\left(-\frac{q^2}{2\zeta}\right)\exp\left(\frac{1}{2\zeta}q^2 - \sum_{n=1}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} b_{nlm}\left(\frac{q^2}{\zeta}\right)^n Y_l^m(\mathbf{u})\right)$$

  where the scale ζ is the well chosen to be close to the optimal one based on the coefficient of q^2, i.e. $\zeta = \frac{\zeta_1\sqrt{\pi}}{b_{100}}$. Its truncated Taylor expansion with maximal radial order N and spherical order L denoted by $E_{NL}(\mathbf{q})$ is given as

$$E_{NL}(\mathbf{q}) = \exp\left(-\frac{q^2}{2\zeta}\right)\sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} q^{2n}Y_l^m(\mathbf{u})$$

  $E_{NL}(\mathbf{q})$ is represented by SPNP basis, thus it is in the space $\text{Span}\{B_{nlm}^{\text{SPF}}(\mathbf{q}|\zeta)\}_{n\le N, l\le L}$. While $E_{NL}(\mathbf{q})$ can not be represented by SHO-3D basis with any finite order N'. So the representation MSE of $E(\mathbf{q})$ by SHO-3D basis and SPF basis are quite different for GHOT model, and SPF basis still has less MSE based on Theorem 5.3. Based on the linearity of the basis representation, the MSE when representing the mixture of GDTI model by SHO-3D basis and SPF basis are also quite different. As a special case, the MSE when representing mixture of HOT model by these two bases are also different, and SPF basis still have less MSE.

Based on the above analysis, it is unfair for SPF basis if the synthetic data is generated by mixture of tensor model, while it is unfair for SHO-3D basis if the synthetic data is generated by mixture of HOT model. The mixture of tensor model essentially assumes the signal is in space spanned by SHO-1D3 (SPP and SHO-3D) basis, which is a subset of the space spanned by SPNP (SPF) basis. Thus in the following experiments, we prefer Söderman cylinder model described in Appendix A [Özarslan et al., 2006; Söderman and Jönsson, 1995], which do not assume the space spanned by SPP (SHO-1D3 and SHO-3D) basis or the space spanned by SPNP (SPF) basis.

5.4 Experimental Comparisons

Although we have proposed several variants of DOT and DPI, we only compare QBI, exact QBI, DOT_0 (orignal DOT), DPI_0 (original DPI), SPFI and SHORE in experiments, because they are widely used in dMRI domain. Theoretically, exact QBI and DOT can be seen as the ODF and EAP estimation from the same mono-exponential decay model, which has been discussed in subsection 5.2.2. Please note that SHORE uses quadratic programming with nonnegative constraints to estimate the coefficients [Özarslan et al., 2009]. Actually, all these six methods can use convex optimization with constraints for a better reconstruction. However, to perform a fair comparison, we implement SHORE and SPFI via least square with regularization λ_l in spherical part and λ_n in radial part and the scale ζ is set by fitting the signal to the GHOT model with order N' and L'. Least square with Laplace-Beltrami regularization is used for QBI, DOT and DPI [Descoteaux et al., 2007; Özarslan et al., 2006; Descoteaux et al., 2010].

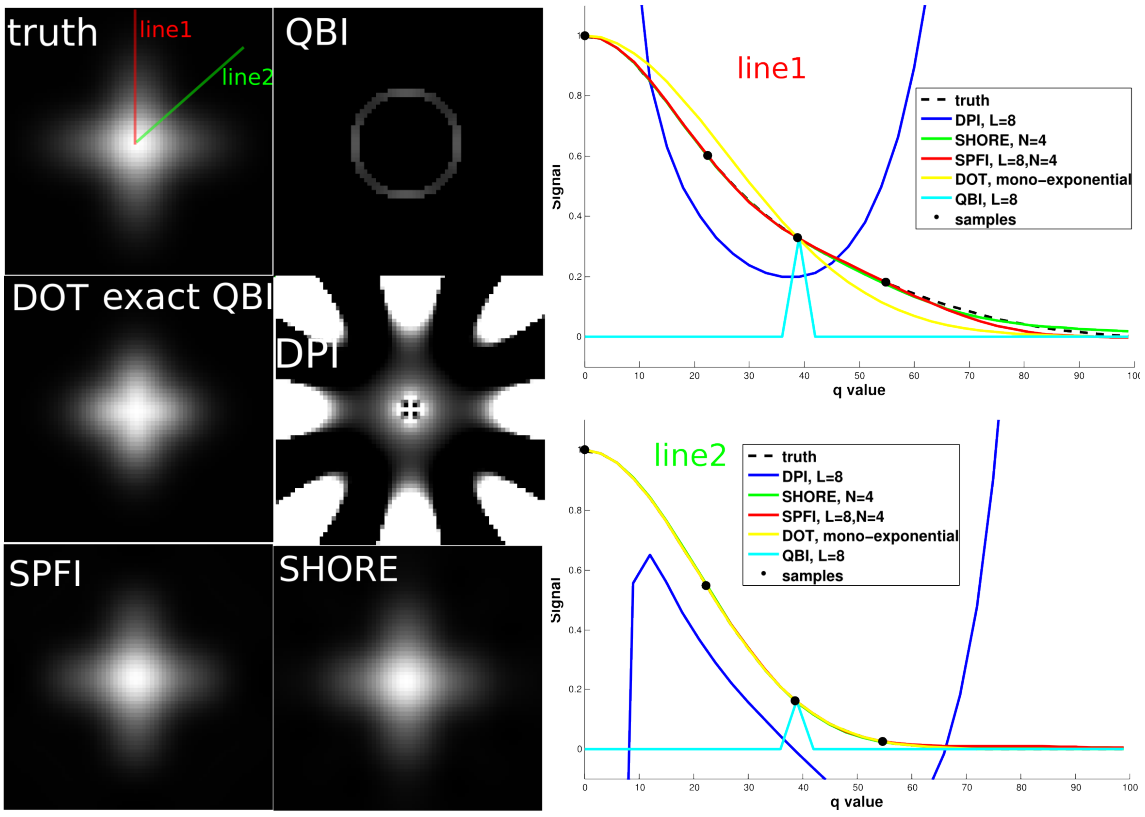
5.4.1 Synthetic data

Noise-free experiment

We generate synthetic data using Söderman cylinder model described in Appendix A [Özarslan et al., 2006; Söderman and Jönsson, 1995]. The parameters are set as the same in [Özarslan et al., 2006]: length $L = 5mm$, radius $\rho = 5\mu m$, free diffusion coefficient $D_0 = 2.02 \times 10^{-3} mm^2/s$, $\Delta/\delta = 20.8/2.4ms$. DWI data were generated in 3 shells with b value $500/1500/3000 s/mm^2$, 60 evenly distributed samples per shell.

In the noise free experiment, the data was generated from two fibers crossing with 90° along x-axis and y-axis. The ground truth signal in x-y plane was visualized in Fig. 5.5, which also showed the reconstructed signals in the six methods. Note that the reconstructed signals in exact QBI and DOT are the same because they are both based on mono-exponential decay assumption. $L = 8$ and $\lambda_l = 10^{-9}$ were set for QBI, exact QBI, DOT and DPI, $L = 8$, $N = 4$ and $\lambda_l = \lambda_n = 10^{-9}$ for SPFI and SHORE. The scale parameters ζ in SPFI and SHORE were set by fitting the GHOT model with $N' = 1$ and $L' = 4$. For DOT and exact QBI, the signal samples on single shell of $b = 1500 s/mm^2$ suggested in [Özarslan et al., 2006] were used to extrapolate the signal in other positions based on mono-exponential decay assumption. The single shell samples were also used to estimate the coefficients for QBI. Please note in QBI, the signal inside and outside the q-ball was forced to zero as we have discussed. Three

Figure 5.5: Fitting synthetic noise-free signal by six methods. The left side shows the ground truth signal and the reconstructed signals from six methods in x-y plane, where exact QBI and DOT are in the same extrapolate model for diffusion signal. Please note the origin point in DPI is singular, so we set it as the mean of the 4 points in its neighborhood. The right side shows the signal values along two lines (Line1 and Line2).



shell samples were used in DPI, SPFI and SHORE. We also plotted the values along given two lines where Line1 is along y-axis and Line2 has 45° azimuth. Along Line1, the value curves obtained from SPFI and SHORE are much closer to the ground truth and better than DOT. Along Line2, SPFI, SHORE and DOT obtained very similar results with the ground truth. DPI gave the worst results where the estimated signal tends to infinity as q increases, the origin point is singular, and the signal near the origin has unacceptable very large absolute values. As we have discussed, the reconstructed signal from DPI is a polynomial along each direction, not a Gaussian-like decay. In details, the regular term in the basis function in DPI makes the signal tend to infinity as q increases, and the irregular term makes a singular point when $q = 0$. QBI only considers the signal in the q-ball and does not extrapolate signal in **q**-space, which is inappropriate for EAP estimation since Fourier transform is a ***global*** transform.

Experiment with Rician noise

The performance on detecting fiber directions of the six methods was compared with the synthetic data corrupted by Rician noise. The noise was added for 1000 trials with SNR=10. EAP profiles at $15\mu m$ were estimated from DOT, DPI, SPFI and

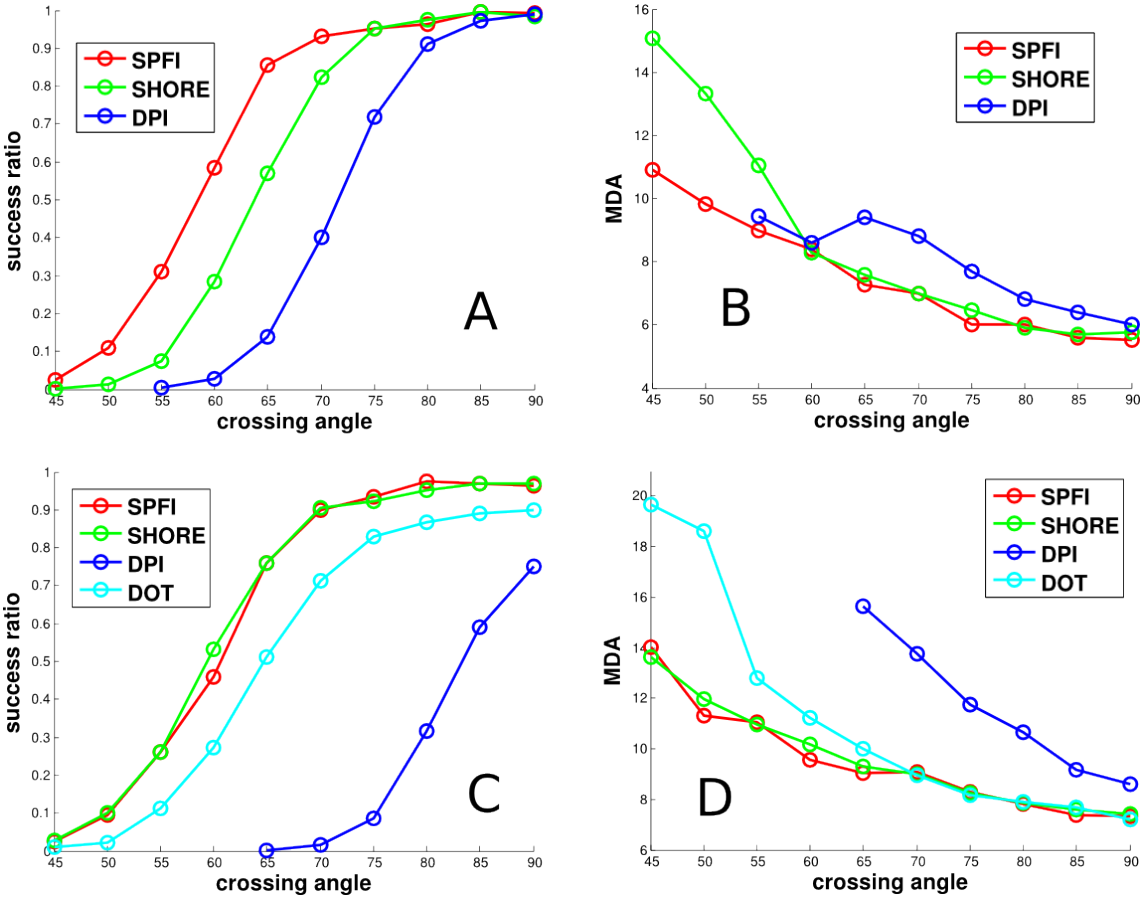
SHORE, and two kinds of ODFs Φ_t and Φ_w were estimated from QBI [Descoteaux et al., 2007], exact QBI [Aganj et al., 2010b], SPFI and SHORE. Note that we did not compare EAPs from QBI, because EAP in QBI has many negative values especially when $R > 10\mu m$ as we have discussed in subsection 5.2.2. The maxima of EAP profiles or ODFs were detected as described in Appendix A. The successful ratio to detect two maxima was recorded. The mean difference of angle (MDA) was calculated from the successful trials. An truncated basis is needed for the data corrupted by noise to avoid overfitting. For DOT, QBI, exact QBI and DPI we set $L = 4$ and $\lambda_l = 0.006$ suggested in [Descoteaux et al., 2007]. For SPFI, we set $L = 4$, $N = 1$ and $\lambda_l = \lambda_n = 1e-8$ in practice. For SHORE, please note that $N = 3$ ($N_{max} = 6$ in [Özarslan et al., 2009]) used in [Özarslan et al., 2009] does not work well when SNR= 10. Thus we set $N = 2$ ($N_{max} = 4$) and $\lambda_l = \lambda_n = 1e-8$ for SHORE similarly with SPFI. The scale parameter ζ in SPFI and SHORE was set by fitting the GHOT model with $N' = 1$ and $L' = 4$.

Fig. 5.6 shows the results of EAP profiles, where (A) and (B) are the success ratio and MDA of the EAP profiles from SPFI, SHORE and DPI using 3 shells, (C) and (D) are EAP results from SPFI, SHORE, DPI and DOT using only single shell at $b = 1500 s/mm^2$. We use the shell at $b = 1500 s/mm^2$ because the data at $3000 s/mm^2$ has relatively lower SNR and 1500 was suggested for DOT in [Özarslan et al., 2006]. It is clear in Fig. 5.6 that SPFI has best results among mHARDI methods using 3 shell data, while it has similar results with SHORE in single shell data. SPFI and SHORE obtain better results even for single shell data than other methods. DOT is better than DPI in single shell data, probably because DPI is unstable in single shell data and q_{max} calculated from $b = 1500$ is relatively small considering the maximal b value 8000 was used in [Descoteaux et al., 2010]. Note that SPFI and SHORE obtain better results using 3 shell data than using only single shell data.

Fig. 5.7 shows the results of two kinds of ODFs, where (A) and (B) are the success ratio and MDA of the ODFs from SPFI and SHORE using 3 shells, (C) and (D) are for ODFs from SPFI, SHORE, QBI and exact QBI using only single shell at $b = 1500 s/mm^2$. (A) and (B) showed that with 3 shell data Φ_w estimated from SHORE has higher success ratio for large crossing angle and lower success ratio for small crossing angle than Φ_w estimated from SPFI. For SPFI, Φ_w has better angular resolution than Φ_t, i.e. higher success ratio for small crossing angle, while Φ_t is more robust to noise for large crossing angle. The results of SPFI agree with the results of Φ_t from QBI and Φ_w from exact QBI (C,D) using single shell data [Tristán-Vega et al., 2010]. However an interesting phenomenon is that for SHORE, Φ_w has lower success ratio for small crossing angle and higher success ratio for large crossing angle than Φ_t. Φ_w estimated from SHORE has lowest MDA than others. For the ODFs from single shell data in (C) and (D), Φ_t estimated from SPFI has the lowest MDA, and highest success ratio for large crossing angle. However, for small crossing angle Φ_t from SPFI only has higher success ratio than Φ_t from QBI. Φ_w from SPFI has overall highest MDA and lowest success ratio for large crossing angle. Φ_t and Φ_w from SHORE have similar results of both success ratio and MDA.

Comparing the results of EAP profiles in Fig. 5.6 and the results of ODFs in Fig. 5.7, the EAP profile estimated from SPFI using 3 shell data has the best results. The EAP profiles from SPFI and SHORE using single shell data has the second best results. Overall the EAP profiles have better results than ODFs, which agrees previous study on the comparison between Φ_t from QBI and EAP profile from DOT in [Prckovska et al., 2008].

Figure 5.6: Evaluation of EAP estimation methods by synthetic data with Rician noise. A, B: success ratio and MDA respectively for the EAP profiles estimated from SPFI, SHORE and DPI using the data from three shells; C, D: success ratio and MDA respectively for the EAP profiles estimated from SPFI, SHORE, DPI and DOT using in single shell with $b = 1500$;

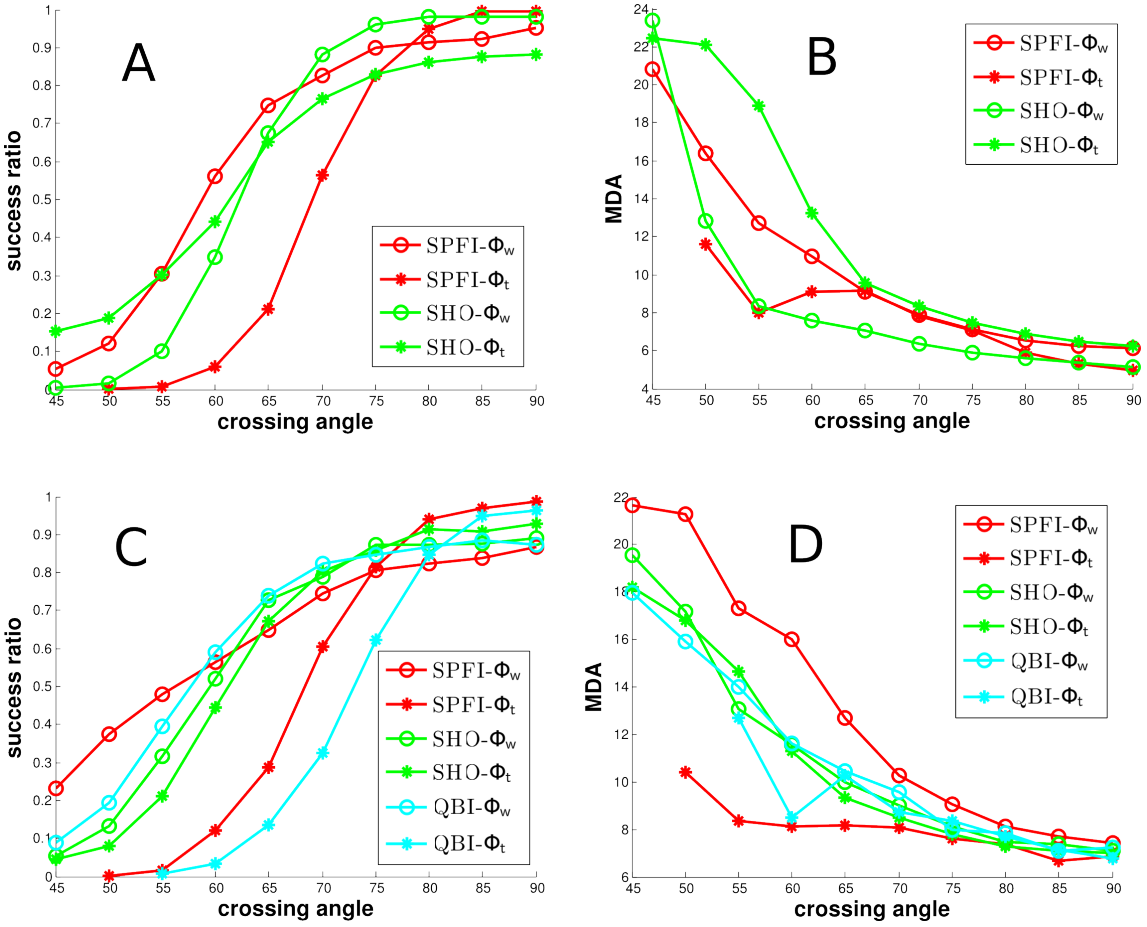


5.4.2 Phantom data

The six methods were performed on the public phantom data with 3 shells ($500/1500/2000 s/mm^2$) which was used in Fiber Cup in MICCAI 2009 [Poupon et al., 2008; Fillard et al., 2011]. This data is now a standard test data for evaluating both estimation methods and fiber tracking algorithms. The ground truth of fiber directions was shown in Fig. B.1. Note we set the unknown diffusion time $\tau = \frac{1}{4\pi^2}$ for calculating q values from b values. Please see Appendix B for more details and challenges on this data.

In our previous studies [Cheng et al., 2010a,b], we set the scale ζ based on typical ADC value, and estimated the EAP profiles at $15\mu m$ with small regularization λ_l and λ_n. In this subsection, we would like to compare the robustness of different methods without regularization. For QBI, exact QBI, DOT and DPI, we set $L = 4$ and $\lambda_l = 0$. We set $L = 4$, $N = 1$ for SPFI, $L = 4$, $N = 2$ for SHORE, $\lambda_l = \lambda_n = 0$ for both SPFI and SHORE. Note that as we have analyzed in subsection 5.3.1, SPFI can work with single shell data without any regularization, while SHORE and DPI do not work for single shell data if no regularization. Thus for single shell data we considered a little regularization with $\lambda_n = \lambda_l = 10^{-9}$ for SHORE, $\lambda_l = 10^{-8}$ for DPI, and we still use

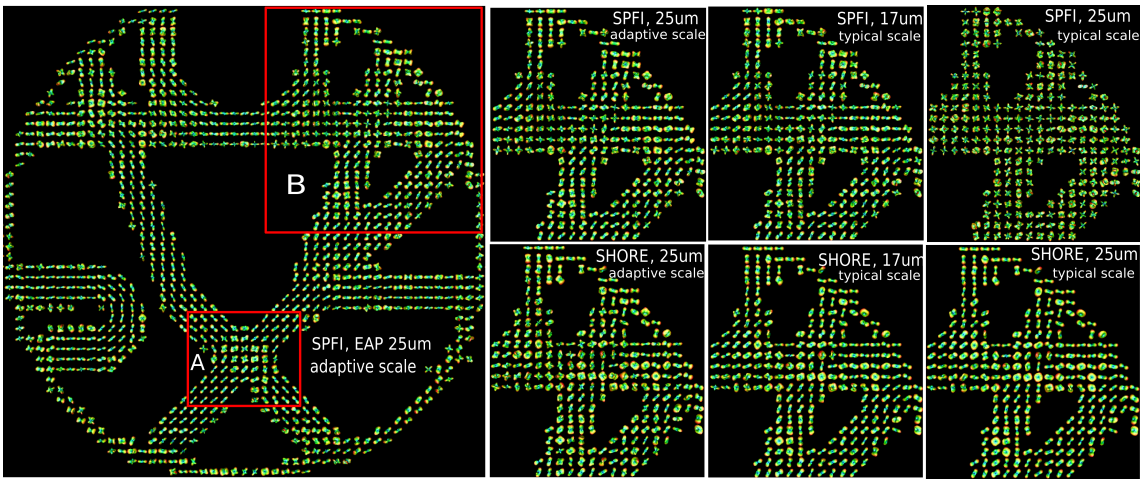
Figure 5.7: Evaluation of ODF estimation methods by synthetic data with Rician noise. A, B: success ratio and MDA respectively for two kinds of ODFs estimated from SPFI, SHORE and DPI using the data from three shells; C, D: success ratio and MDA respectively for two kinds of ODFs estimated from SPFI, SHORE, DPI and DOT using in single shell with $b = 1500$;



$\lambda_n = \lambda_l = 0$ for SPFI because it is stable for single shell data when $N = 1$ as we have analyzed.

Fig. 5.8 shows the effect of scale parameter set by two ways in SPFI and SHORE. One is to set ζ as the typical scale based on the typical ADC in Eq. (5.32). The other one is to set adaptive ζ by fitting DWI signal with GHOT model with $N' = 1$, $L' = 4$, i.e. HOT model. The left part of Fig. 5.8 is the whole field of view of the estimated EAP profiles based on the adaptive scale ζ from GHOT model fitting. Two regions A and B are enlarged for visualization in the following figures on this data set. The right part of Fig. 5.8 shows the comparison of the EAP fields under two scale selection methods. For the typical scale set by typical ADC $D_0 = 0.7\times10^{-3} mm^2/s$, SPFI works well at $15\mu m$ and $17\mu m$, which has been shown in [Cheng et al., 2010b]. For the radius $25\mu m$, the results of SPFI are very noisy. While SHORE works for both radii $17\mu m$ and $25\mu m$, although the results are not so good. When using adaptive scale, the EAP profiles from SPFI at $25\mu m$ are largely improved and have better quality than EAP profiles from SHORE which seem to remain the estimation quality as the case of typical scale. The pseudo-ADC map obtained by adaptive scale estimation has pseudo-ADC values around $1.6\times10^{-3} mm^2/s$ which is about twice of the typical ADC D_0. That is the reason why these two scale selection obtains different results. Normally EAPs are estimated

Figure 5.8: Comparison on the scale selection by two ways in phantom data. Left part: whole field of view of the EAP profiles with $R_0 = 25\mu m$ estimated based on adaptive scale, where two regions are used in are enlarged for visualization in the following figures on this data set. Right part: the EAP profiles in region B estimated by SPFI and SHORE with two scale selection ways.

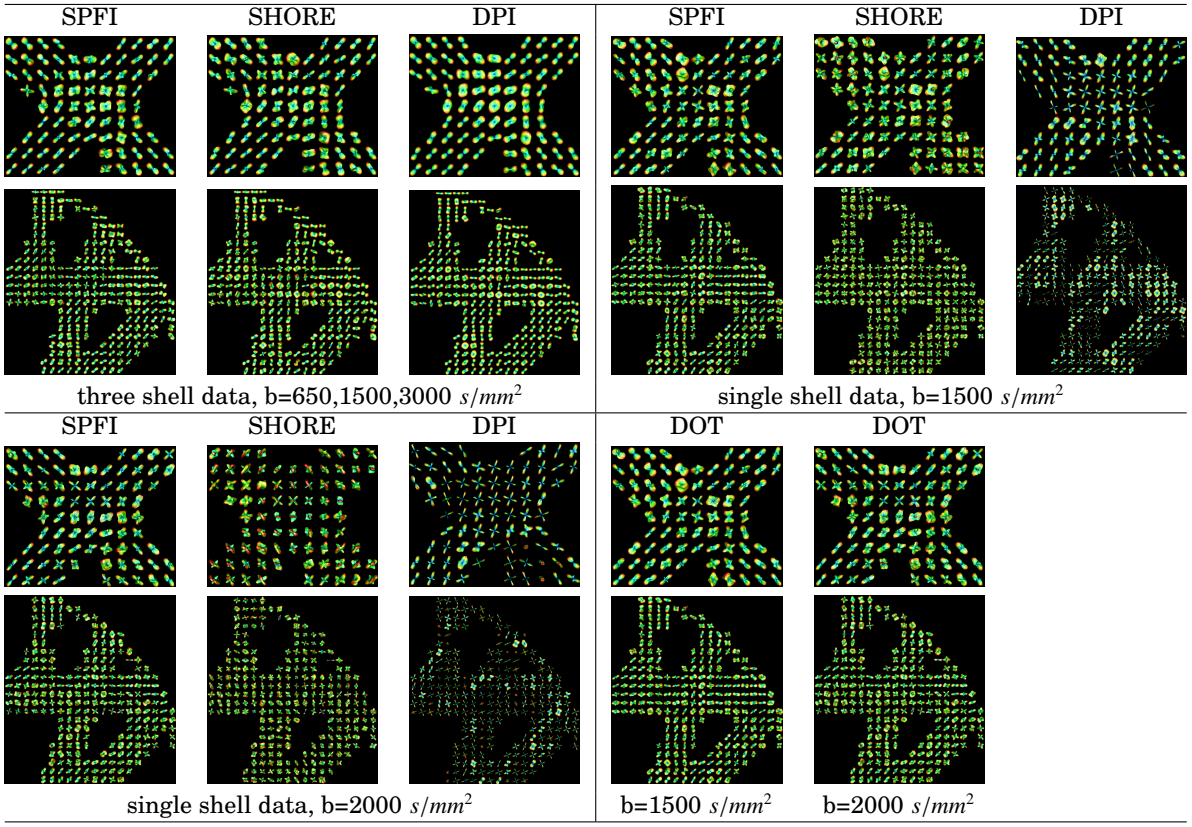


around $15\mu m$ and EAPs at $R_0 > 20\mu m$ are likely noisy [Cheng et al., 2010b; Özarslan et al., 2006]. However, please note that the phantom data has very low anisotropy so that we need to perform min-max normalization for visualization. Thus we believe for this data it is reasonable that EAP profile is less anisotropic around $15\mu m$ and a larger radius is needed. Moreover, in the whole field of view for EAPs at $25\mu m$ in Fig. 5.8 the crossing areas are obvious and other areas are also very clean. So in the following results on this data in this chapter, we always use the adaptive scale by fitting the GHOT model and estimate the EAP profiles at radius $R_0 = 25\mu m$.

Fig. 5.9 shows the results of EAP profiles estimated from 3 shell data by SPFI, SHORE, DPI respectively and the results from single shell data by SPFI, SHORE, DPI and DOT. It is clear to see from Fig. 5.9 that for mHARDI methods, e.g. SPFI, SHORE and DPI, multiple shell data obtains better results than single shell data. The results from the single shell data with $b = 1500 s/mm^2$ is better than the results from the data with $b = 2000 s/mm^2$. That is understandable. The data with higher b value has better angular resolution but lower SNR, while the data with lower b value has worse angular resolution but higher SNR. Thus mHARDI methods with multiple shell data which combines both merits obtain better results. The EAP profiles estimated by DPI from single shell data have many negative values which are removed in visualization. DPI for single shell data has the worst among all methods for single shell data, probably because its basis matrix has very large condition number for single shell data. For 3 shell data, the EAP profiles by DPI are very smooth, probably because the incomplete integration introduced by q_{max} which is small in this data. SPFI used in 3 shell data obtains the best results in both regions. SPFI and DOT in single shell data with $b = 1500 s/mm^2$ obtain the second best results. It is interesting to see that for single shell data with $b = 1500 s/mm^2$, SPFI has similar but slightly more noisy results with DOT.

Fig. 5.10 shows the results of two kinds of ODFs estimated by SPFI, SHORE, QBI and exact QBI from single shell and three shell data. For three shell data, Φ_w es-

Figure 5.9: EAP profiles estimated by SPFI, SHORE, DPI and DOT from phantom data in single shell and three shells.



timated by SPFI seems sharper but slightly more noisy than Φ_w by SHORE, which agrees with the ODF results of synthetic experiment in Fig. 5.7(A). For single shell data with $b = 1500 s/mm^2$, Φ_w estimated by SPFI and exact QBI are similar and they are slightly sharper but more noisy than Φ_w estimated by SHORE. An interesting thing is the ODF by Tuch Φ_t estimated by SPFI, SHORE and QBI from single shell data are much similar. That is because the estimated Φ_t from these three methods in this data are much smooth and the difference between them are slightly visible before min-max normalization. However, min-max normalization enhanced the directions and discarded these small difference of shape. For single shell data with $b = 2000 s/mm^2$, Φ_w by all methods are too noisy. Φ_t by three methods are very similar.

Compared to ODFs, the EAPs estimate by SPFI from 3 shell data and EAPs estimated by SPFI and DOT from single shell data with $b = 1500 s/mm^2$ seem to be more clean and sharper. This agrees with the previous results in synthetic data. Note that SHORE did not obtain good results in both single shell data and 3 shell data as it did in synthetic experiments. That is probably because we use a larger radius $R_0 = 25\mu m$ and the SNR is higher than the SNR in synthetic experiments.

5.4.3 Real Monkey data

These six methods were also performed in a real monkey data with three shells ($b = 500, 1500, 3000 s/mm^2$). Each shell has the same thirty evenly distributed sam-

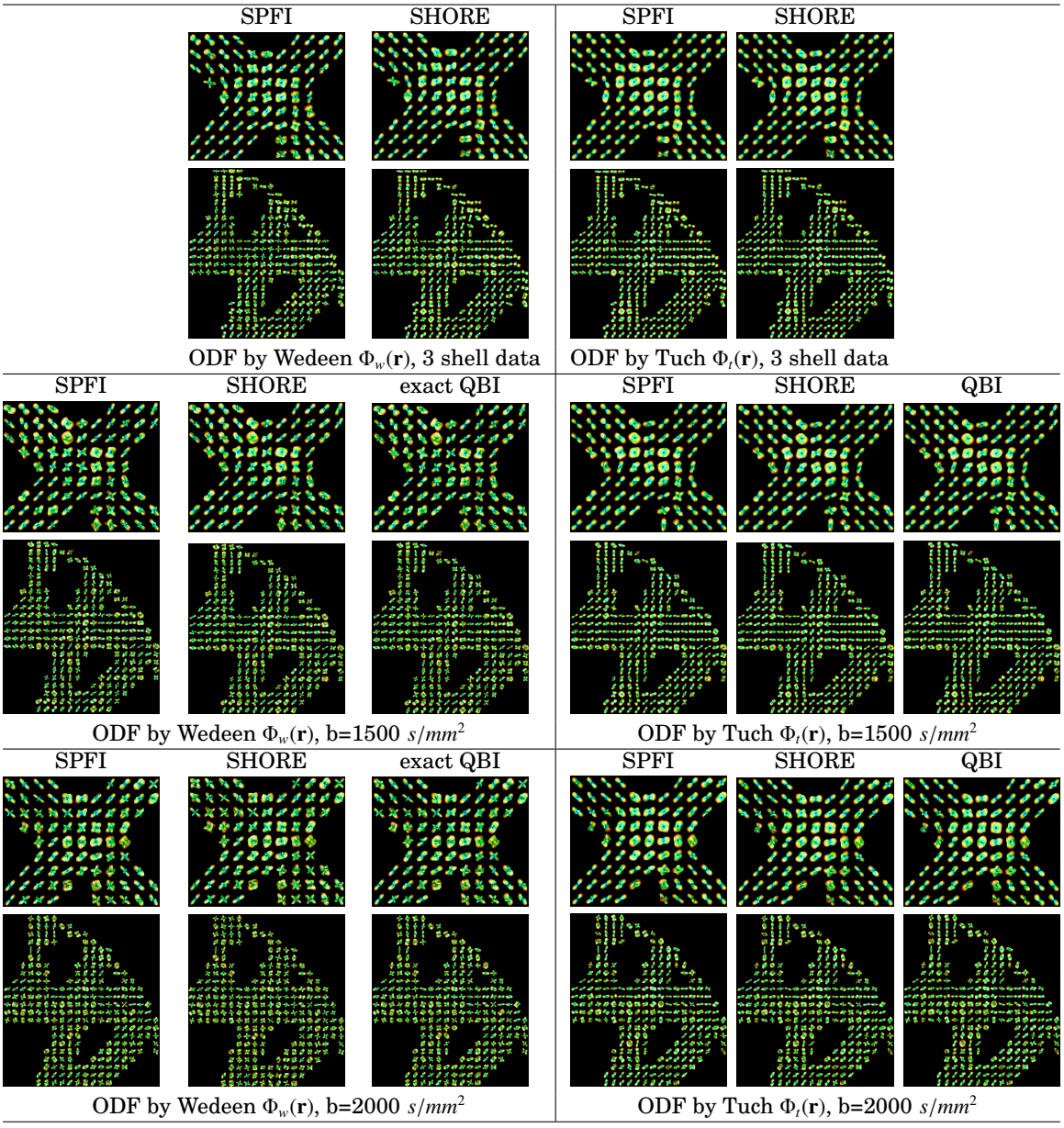
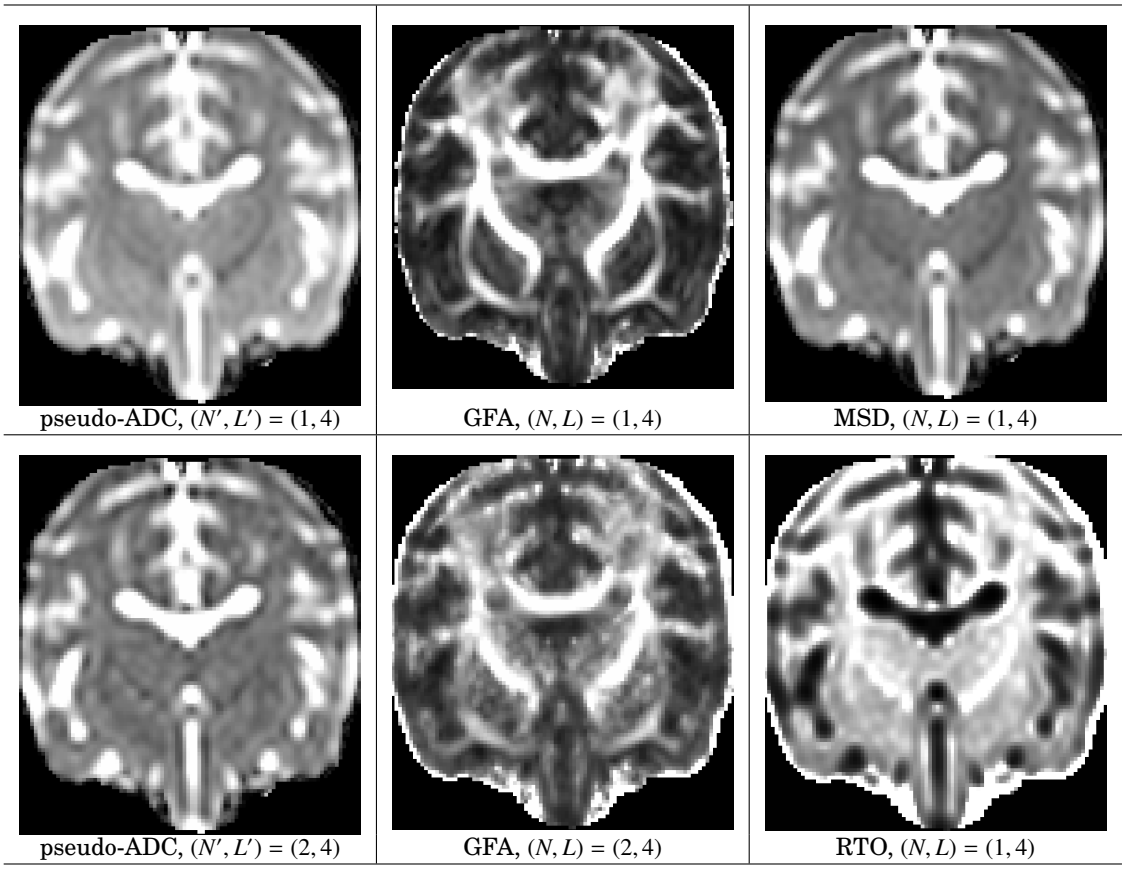


Figure 5.10: Two kinds of ODFs estimated by SPFI, SHORE, QBI and exact QBI from phantom data in single shell and three shells.

ples. Please refer Appendix B for more details on this data set. In our previous studies [Cheng et al., 2010a,b], we set the scale ζ based on typical ADC value, and estimated EAPs/ODFs with small regularization. In this subsection, we would like to demonstrate the parameter selection based on scalar maps, and compare different methods without regularization. The same parameters were used in methods in this data as the parameters in phantom data.

Fig. 5.11 shows the parameter selection process. The pseudo-ADC map with $(N', L') = (1, 4)$ ranges from 5.6×10^{-6} to about $1.4 \times 10^{-3} mm^2/s$ which is twice of the typical ADC. The pseudo-ADC map with $(N', L') = (2, 4)$ ranges from -1.1×10^{-3} to about $3.3 \times 10^{-3} mm^2/s$ which is five times of the typical ADC. Note that the pseudo-ADC map with $(N', L') = (2, 4)$ has large range with negative values, which is not

Figure 5.11: Scalar maps for parameter selection in real data. The first column includes the pseudo-ADC map under $(N', L') = (1, 4)$ and $(N', L') = (2, 4)$. $(N', L') = (1, 4)$ is used to generate the scalar maps in the second and third columns because of its better quality. The second column includes the GFA maps under $(N, L) = (1, 4)$ and $(N, L) = (2, 4)$. The third column includes the MSD map and RTO map under $(N, L) = (1, 4)$. $(N, L) = (1, 4)$ is used in EAP/ODF estimation and scalar map estimation in the third column because of its better quality.



appropriate. Pseudo-ADC map with $(N, L) = (1, 4)$ has better quality with normal range in brain. Thus we fix $(N', L') = (1, 4)$ to obtain scale ζ for this data. It can be seen that the pseudo-ADC map has similar contrast with ADC map (i.e. MD map) in tensor model in Fig. 4.7. The Cerebrospinal fluid (CSF) has high pseudo-ADC value. In GFA maps, the anisotropic areas have high intensity, which means it is indeed the generalized version of previous FA for tensors and GFA for ODFs. The CSF areas have high intensity in MSD map and low intensity in RTO map. That is because CSF areas have high diffusivity although the anisotropy of these regions is low. The GFA map with $(N, L) = (1, 4)$ has better contrast and is smoother than the GFA map with $(N, L) = (2, 4)$. Thus we fix $(N, L) = (1, 4)$ for the estimation of ODF/EAP fields and other scalar maps. See the whole filed of view of the estimated ODFs and EAP profiles at $15\mu m$ in Fig. 5.12, where a square region is marked in EAP field and is enlarged for visualization of ODF and EAP results in Fig. 5.13, 5.14 and 5.15. Comparing the estimated EAP field and two kinds of ODF field shown in Fig. 5.12, an obvious advantage for EAP profile and ODF by Wedeen Φ_w over the ODF by Tuch Φ_t is that we do not do artificial min-max normalization for EAP profile and Φ_w because they are sharp enough compared to Φ_t. Moreover, the estimated EAP profile and Φ_w are isotropic in grey matter areas and anisotropic in white matter areas. If we perform min-max

Figure 5.12: Whole field of view of the estimated EAP profiles at $15\mu m$ and two kinds of ODFs with $(N', L') = (1, 4)$, $(N, L) = (1, 4)$ in real data. The glyphs of EAP profiles and ODFs are colored by the GFA for spherical functions which is also used to set the background color. The square region marked in EAP field is enlarged for visualization of ODF/EAP glyphs in the Fig. 5.13, 5.14, and 5.15.

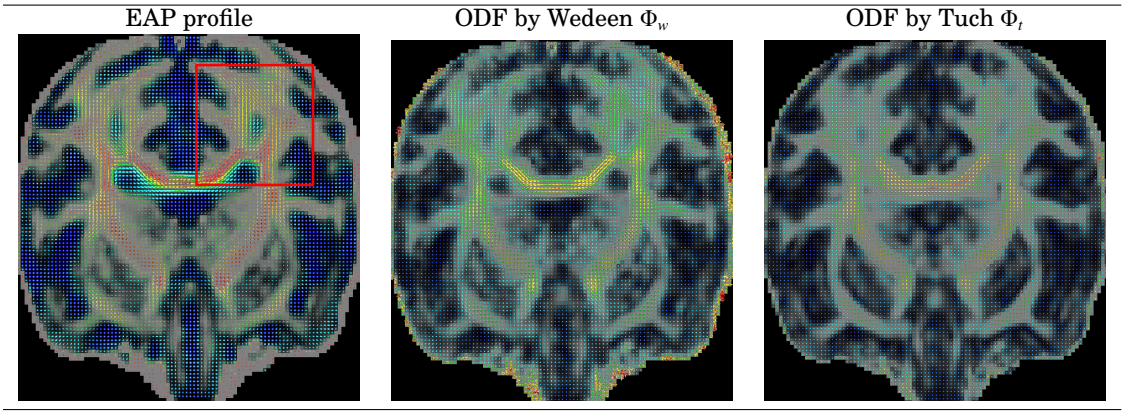
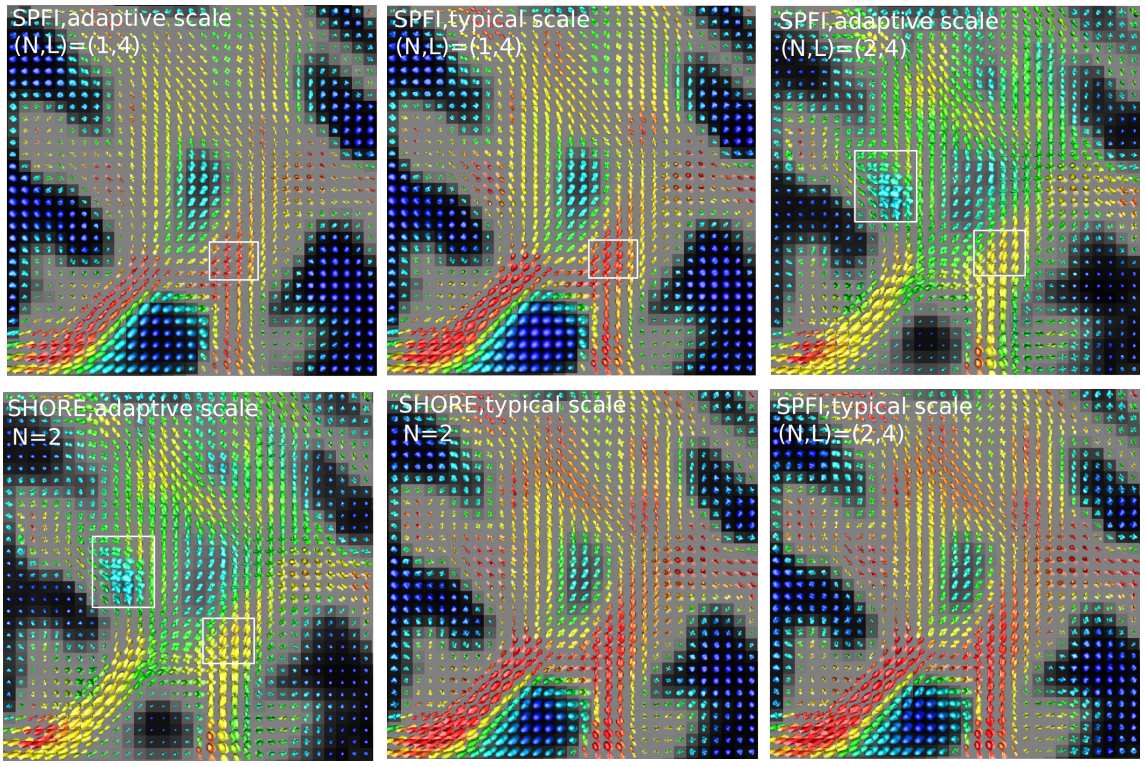


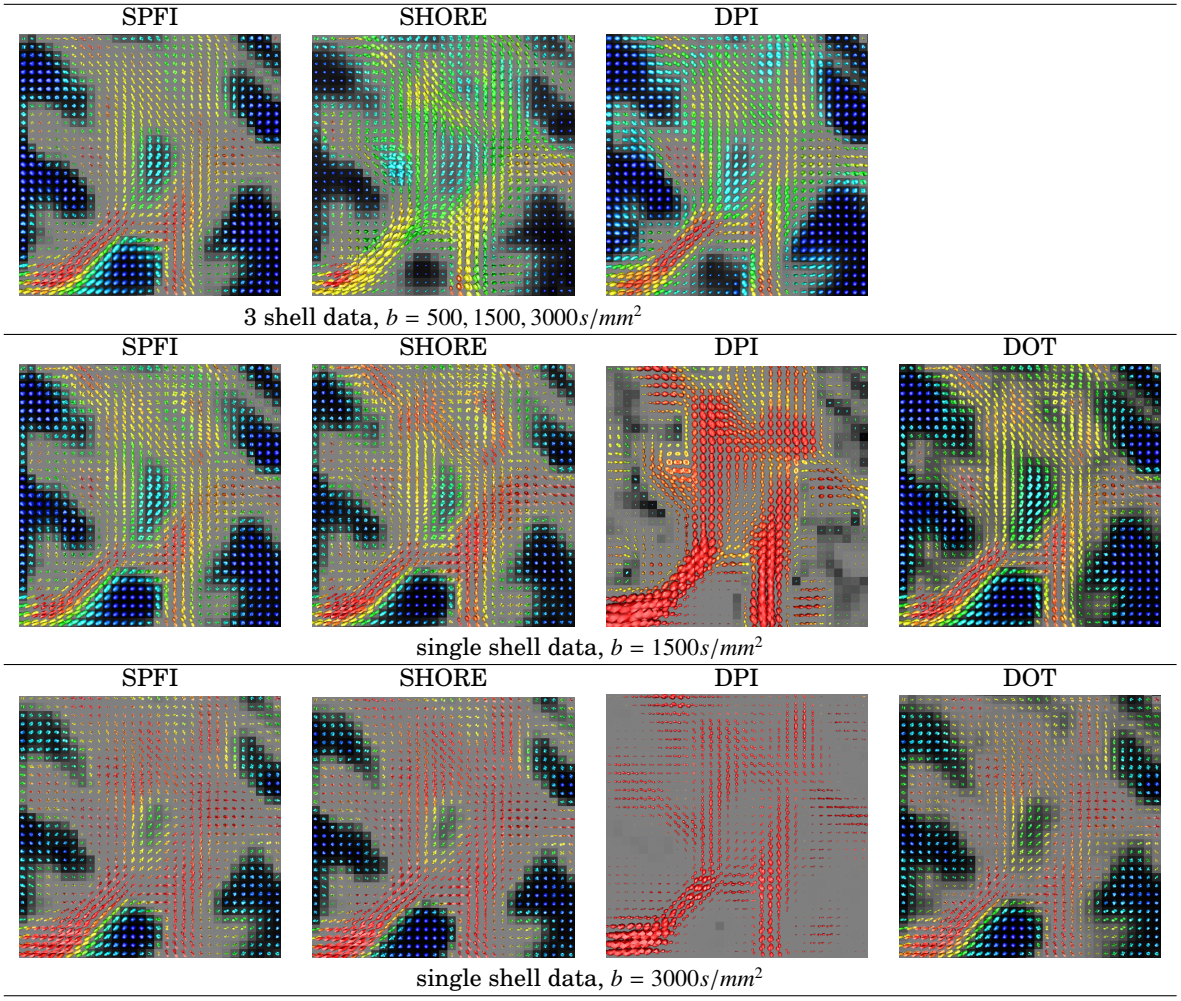
Figure 5.13: EAP profiles in real data estimated from SHORE and SPFI under different order and different scale selection. GFA for spherical functions is used to set the color of glyphs and the background.



normalization, the grey matter areas become more anisotropic which may bring some errors in the following tractography.

Fig. 5.13 shows the EAP profiles by SPFI and SHORE with the scale ζ set by two ways, i.e. the typical scale and the adaptive scale by fitting GHOT model with $(N', L') = (1, 4)$. When $(N, L) = (1, 4)$ is used in SPFI, the EAP profiles with adaptive scale are sharper in the marked square area than the results based on typical

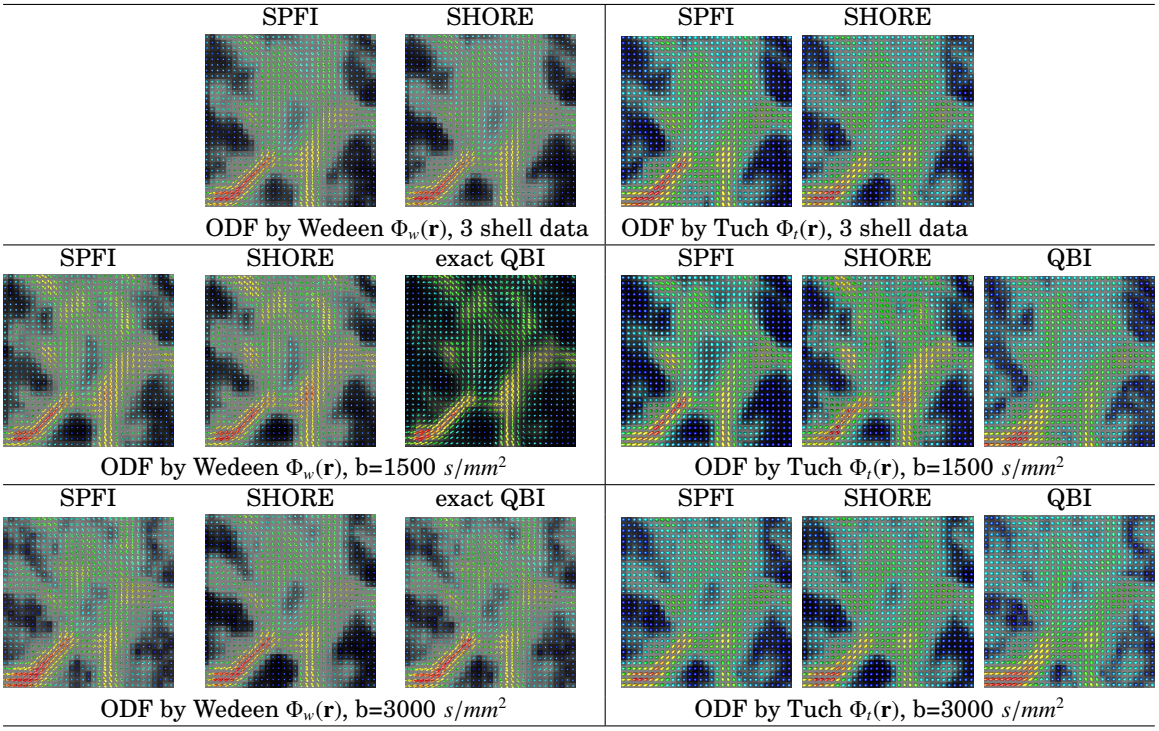
Figure 5.14: EAP profiles at $15\mu m$ estimated by SPFI, SHORE, DPI and DOT from real data in single shell and three shells. GFA for spherical functions is used to set the color of glyphs and the background.



scale. However, when minimal order $N = 2$ needed for anisotropic diffusion is used in SHORE, the results from adaptive scale seem to be worse than the results from typical scale. See the marked square region. An interesting thing is that the results from SPFI with $(N, L) = (2, 4)$ and adaptive scale are very similar with the results from SHORE with $N = 2$ and adaptive scale. The results from SPFI with $(N, L) = (2, 4)$ and typical scale are similar with the results from SHORE with $N = 2$ and typical scale. This phenomenon can be explained by Theorem 5.3. However, the results from SPFI with $(N, L) = (1, 4)$ and adaptive scale are better than the results from SHORE with $N = 2$ and adaptive scale. The results in this figure showed that in some cases, we need to use low order in radial part compared to the high order in spherical part. While SHORE forces $n > l/2$, which may bring some overfitting effect. SPFI completely separates the radial part and spherical part, which gives more freedom to select different orders in two parts. Note that since adaptive scale is well estimated when low order $N = 1$ is used, low order in radial part may be enough for a good estimation in this data.

Fig. 5.14 shows the EAP profiles at $15\mu m$ estimated from both multiple shell data

Figure 5.15: Two kinds of ODFs estimated by SPFI, SHORE, QBI and exact QBI from monkey data in single shell and three shells.



and single shell data using SPFI, SHORE, DPI and DOT. With three shell data, SPFI obtains the best results, while SHORE has a little overfitting. The advantage of SPFI over SHORE is just to use lower order in radial part, because SPFI with $(N, L) = (2, 4)$ obtains similar results by SHORE with $N = 2$ as we have discussed above. EAPs in DPI are also less anisotropic, because we did not do min-max normalization like [Descoteaux et al., 2010] and the maximal b value for q_{max} used here is 3000, while 8000 is used in [Descoteaux et al., 2010]. For single shell data, in all methods the results from data with $b = 1500$ are better than the results from data with $b = 2000$. Once again $b = 1500 s/mm^2$ seems the optimal b value for estimation on single shell data [Özarslan et al., 2006]. DPI fails in single shell data probably due to its unstablility in single shell data. SPFI and SHORE obtain similar results in single shell data, and their results are more sharper than DOT, which agrees with the synthetic experiment in Fig. 5.6(C).

Fig. 5.15 shows the ODFs Φ_w and Φ_t estimated by SPFI, SHORE, QBI and exact QBI from both three shell data and single shell data. Both ODFs by SPFI and SHORE from three shell data seem to be very similar. For single shell data, Φ_w by SHORE is more robust to noise than Φ_w by SPFI and exact QBI, because it obtains cleaner isotropic glyphs in grey matter areas, especially when $b = 3000 s/mm^2$. This agrees with the results of synthetic experiment in Fig. 5.7. However, for single shell data, Φ_t estimated by SPFI is more robust and anisotropic than Φ_t estimated by SHORE and QBI, which also agrees with the experiment in 5.7.

Comparing the estimated EAP profiles and two kinds of ODFs, the EAP profiles generally have better angular resolution. EAP profile estimated by SPFI with three shell data is the best one, and the EAP profile estimated by SPFI and SHORE from

single shell data with $b = 1500s/mm^2$ are the second results. EAP profiles by DOT takes the third position.

5.5 SUMMARY

In this chapter, based on the SPF basis and least square estimation proposed by Dr. Assemlal, we have proposed the analytical Spherical Polar Fourier Imaging (SPFI), which analytically estimates EAP, ODFs, MSD, GFA, RTO by representing signal with SPF basis. The implementation of SPFI includes three steps, i.e. scale estimation, SPF coefficients estimation, analytical linear transforms for EAP and its various features. See Table 5.1 for the process. In scale estimation, we proposed two ways. One is to set scale ζ based on typical ADC value in Eq. (5.32). The other one is to fit the signal with GHOT model and set scale parameter by the pseudo-ADC in GHOT model, which was proved in Theorem 5.1 to be the optimal scale for infinite radial terms. In least square estimation of SPF coefficients, we proposed to consider the prior $E(0) = 1$ to improve the estimation results. In analytical linear transforms for EAP and its features, we proposed in proposition 5.1 the closed form of Fourier dual SPF (dSPF) basis that is the Fourier transform of SPF basis. Then EAP can be represented by dSPF basis with the same coefficients of signal under SPF basis as shown in proposition 5.2. Proposition 5.3 and 5.4 proved that two kinds of ODFs Φ_t and Φ_w can be analytically represented by SH basis with the coefficients obtained from linear transforms of the pre-estimated SPF coefficients. The effect of diffusion time and parameter selection were also discussed in subsection 5.1.3.

Our derivation of EAP is based on plane wave expansion in Eq. (5.12) which seems more fundamental than addition theorem and Funk-Hecke theorem used in analytical QBI derivation. See the relation of these three theorems in Appendix 5.7.

The SPF basis can be seen as a generalization of SH basis from $\mathbb{S}^2$ to $\mathbb{R}^3$ by considering radial basis. In this sense, SPFI can be seen as a generalization of QBI or DOT. Actually our work showed that it is an easy job for analytical ODF and EAP estimation. We proposed a general framework named Analytical Fourier Transform in Spherical Coordinate (AFT-SC) to incorporate and compare most widely used sHARDI and mHARDI methods, e.g. QBI, exact QBI, DOT, DPI, SHORE, SPFI. In this framework, we also proposed several variants of DPI and DOT which are analytical ODF/EAP estimation methods and can be seen as bridges to demonstrate the relation of various methods. Please refer Table 5.2 for methods in AFT-SC framework. QBI can be explained in AFT-SC framework by considering radial decay as a delta function. We derived the closed form for EAP in QBI and analyzed why it obtains many negative values for large b value. We also explained why the estimation of Φ_w in QBI fails. SHORE is similar with SPFI. We derived the closed form for EAP and ODFs, GFA for SHORE. For DPI and DOT, we theoretically analyzed their limitations and proposed several variants to avoid the limitations. Those variants also serve as bridges to demonstrate the relations between different methods. For example DOT_2 uses SPNP basis which is proved to be equivalent with SPF basis in Lemma 5.2.

In order to evaluate the different basis function used in different methods, we proposed several criteria for evaluation. Please refer Table 5.3 in details. A good basis

should be orthonormal such that it is model-free, and it also should satisfy the listed three priors. A good fitting of the given DWI samples is just a necessary condition, not a sufficient condition for a good EAP/ODF estimation. DPI has bad performance partially due to that it uses the basis that does not satisfy the priors, and the introduction of q_{max} for incomplete integration is another big problem. SPF basis and SHO-3D basis are better than other bases in these criteria except the separation of spherical and radial parts. SHO-3D basis partially separates these two parts since it forces higher order in radial part than spherical part, while SPF basis separate them completely. Separating these two parts can let us give different orders in different part. Normally low order is needed in radial part, but high order is needed in spherical part. We also analyzed the instability of mHARDI methods in single shell data. DPI and SHORE are unstable which need regularization to obtain meaningful results from single shell data. However, SPFI with order $N = 1$ is stable in single shell data which can work without any regularization.

Since SHO-3D basis and SPF basis are similar, we theoretically compared these two bases. Essentially, SHO-3D basis uses Gaussian polynomial basis in radial part, while SPF basis considers Gaussian non-polynomial basis in radial part. Their relation can be seen in Theorem 5.3. SPF basis completely separates the spherical part and radial part, such that we can give low order in radial part and obtain robust results, which has been validated in experiments.

We now summarize the contributions of this chapter.

1. Based on the previous work on SPF basis and least square estimation by Dr. Assemlal, we proposed analytical SPFI. Please refer Table 5.1 for the method which includes three independent steps. The main contribution is the analytical linear transforms in the third step.

   (a) In the first step, we proposed two ways form scale parameter estimation. The first way is to set scale ζ based on the typical ADC value D_0 in Eq. (5.32). The second way is to set scale by fitting the GHOT model, which was proved to be optimal for infinite radial terms in Theorem 5.1. In the first way, the scale is shared in every voxel, which largely accelerates the whole estimation process. Then the following operations on the estimated EAP field can be performed on coefficients because EAPs in different voxels are represented by the same dSPF basis. In the second way, the scale is adaptive in each voxel, which normally obtains better estimation results. However, the estimation process takes more time because the basis matrix is different in each voxel, and the following operations on EAP filed need to consider the basis.

   (b) In the second step for SPF coefficients estimation, we proposed to consider $E(0) = 1$ in the estimation process. It has several advantages. First, it improves largely the results because the estimated signal satisfies the prior $E(0) = 1$. Second, it makes mHARDI methods more stable for single shell data, because it reduces the number of unknown variables in estimation. Third, since it reduces the number of variables, it accelerates the estimation process.

   (c) The third step is the analytical linear transforms for EAP and ODFs in proposition 5.2, 5.3 and 5.3. In proposition 5.1, we proposed the ana-

lytic form of Fourier dual SPF (dSPF) basis which is the Fourier transform of SPF basis. dSPF basis is proved to be an orthonormal basis. The three scalar indices, RTO, MSD, GFA are given in Eq. (5.22), Eq. (5.26) and Eq. (5.28). Note that the GFA in Eq. (5.27) is for EAPs defined in $\mathbb{R}^3$, which is the generalization of GFA for ODFs and EAP profiles defined in $\mathbb{S}^2$ and FA for tensors defined in Sym_3^+. With SPF basis representation, Eq. (5.27) can be analytically written as Eq. (5.28). Compared to numerical SPFI in [Assemlal et al., 2009a; Assemlal, 2010], the proposed analytical linear transforms avoid the numerical errors.

(d) We analyzed parameter selection by scalar maps in subsection 5.1.3 and the effect of diffusion time in proposition 5.5.

2. We proposed Analytical Fourier Transform in Spherical Coordinate (AFT-SC) framework, which incorporates most widely used sHARDI and mHARDI methods and demonstrates well their relations. Based on AFT-SC framework, an analytical ODF and EAP estimation seems to be a very easy job. We demonstrated some variants of methods which can be used as the analytical estimations and avoid the limitations of the original methods. See Table 5.2 for the methods in AFT-SC framework.

   (a) QBI considers delta function as its radial basis. We derived the EAP in QBI, and discussed why EAP in QBI and Φ_w in QBI are both impractical. We derived the analytical QBI for Φ_t in a new way based on plane wave expansion. This is different from previous derivations of analytical QBI based on the addition theorem and Funk-Hecke theorem. Plane wave expansion can derive the other two theorems which are proved to be equivalent. See Appendix 5.7. Thus plane wave expansion theorem seems to be more fundamental than the other two theorems.

   (b) We derived correct form for SHORE, which uses SHO-3D basis. We gave the analytical solution for EAP, ODFs, and other indices in SHORE. We also showed how to consider $E(0) = 1$ in SHROE.

   (c) DPI considers regular and irregular polynomial terms as its radial basis. The regular terms make the estimated signal tends to infinity as q increases and irregular terms make original point singular. We proposed DPI_1 to avoid the limitations in DPI. The basis in DPI_1 is a specific part of SHO-3D basis.

   (d) DOT can not be used in multiple shell data. We proposed DOT_1 and DOT_2 to generalize the original DOT for multiple shell data. DOT_2 uses SPNP basis which is proved to be equivalent with SPF basis in Lemma 5.2. Thus SPFI can be seen as a generalization of DOT.

3. We proposed several criteria for evaluation of different basis in different methods. See Table 5.3.

   (a) The basis should be complete such that the method is model-free.

   (b) The basis is better to be orthogonal. There are several advantages for estimation with orthogonal basis.

   (c) The basis should satisfy the three priors of $E(\mathbf{q})$ in estimation, i.e. the prior $E(0) = 1$, $E(\mathbf{q})$ tends to zero as q increase, and the radial decay of $E(\mathbf{q})$ is like

(but NOT) a Gaussian function. DPI does not satisfy these priors because of the regular and irregular polynomial terms, which is the main reason why DPI fails in experiments. Another reason is the q_{max} for incomplete integration. The limitations of DPI can be avoided by removing irregular terms and considering the product of the regular terms and Gaussian function, which results in DPI_1. The basis of DPI_1 is a subset of SHO-3D basis. In this sense, SHORE can be seen a generalization of DPI.

(d) The bases in mHARDI methods are generally unstable for single shell data so that the regularization is always needed. However, SPF basis is stable for single shell data by considering $E(0) = 1$ and $N = 1$, which works without any regularization.

(e) The basis is better to separate completely the radial part and spherical part, such that we can give different order in different part. SHO-3D basis forces $n \geq l/2$ which may result in overfitting. SPF basis separate these two parts completely.

4. We theoretically compared the difference between SHO-3D basis and SPF basis. Theorem 5.2 shows the equivalence between SHO-3D basis, SHO-1D3 basis and SPP basis. Lemma 5.2 shows the equivalence between SPF basis and SPNP basis. Essentially SHO-3D basis is the Gaussian polynomial basis, while SPF basis is the Gaussian non-polynomial basis. SPF basis contains two parts of basis. One is the polynomial part and the other one is non-polynomial part. Theorem 5.3 demonstrates that the function space spanned by SHO-3D basis with order N is a subset of the function space spanned by SPF basis with order $(N, 2N)$. The space spanned by SPF basis can not be represented by SHO-3D basis with any finite order.

5. The experiments showed that EAP profile generally has better performance than ODFs, and the proposed scale estimation largely improves the results of SPFI. For multiple shell data, SPFI obtains better EAP estimation than others. For single shell data, SPFI, SHORE and DOT with $b = 1500 s/mm^2$ have better EAP estimation than others. The experiment also demonstrated that the better performance of SPFI over SHORE is probably because SPFI can give lower order in radial part to avoid overfitting.

5.6 APPENDIX A: PROOFS ON ANALYTICAL SPFI

Proof of Proposition 5.1 on analytical form of dSPF basis. Based on the plane wave equation in Eq. (5.12)

$$\begin{aligned} B_{nlm}^{\text{dSPF}}(\mathbf{R}|\zeta) &\overset{\text{def}}{=} \mathscr{F}\{G_n(q|\zeta)Y_l^m(\mathbf{u})\}(\mathbf{R}) = \int_{\mathbb{R}^3} G_n(q|\zeta)Y_l^m(\mathbf{u})\cos(2\pi i\mathbf{q}^T\mathbf{R})\mathrm{d}\mathbf{q} \\ &= 4\pi\int G_n(q|\zeta)Y_l^m(\mathbf{u})\left\{\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'}(-1)^{l'/2}j_{l'}(2\pi qR)Y_{l'}^{m'}(\mathbf{u})Y_{l'}^{m'}(\mathbf{r})\right\}\mathrm{d}\mathbf{q} \\ &= 4\pi(-1)^{l/2}\underbrace{\left\{\int_0^{\infty} j_l(2\pi qR)G_n(q|\zeta)q^2\mathrm{d}q\right\}}_{I_{ln}(R)}Y_l^m(\mathbf{r}) \end{aligned} \tag{5.85}$$

where we use the orthonormal property of SHs, i.e. $\int_{\mathbb{S}^2} Y_l^m(\mathbf{u}) Y_{l'}^{m'}(\mathbf{u}) d\mathbf{u} = \delta_{ll'}\delta_{mm'}$, and define $I_{ln}(R) = \int_0^\infty j_l(2\pi qR) G_n(q|\zeta) q^2 dq$. Considering Eq. (2.63) in Eq. (5.2), i.e.

$$L_n^{1/2}(x) = \sum_{i=0}^{n} l_n^i x^i, \quad l_n^i = (-1)^i \binom{n+0.5}{n-i} \frac{1}{i!}$$

we have

$$I_{ln}(R) = \frac{\kappa_n(\zeta)\zeta^{1.25}}{2\sqrt{R}} \sum_{i=0}^{n} l_n^i \int_0^\infty x^{2i+1.5} J_{l+0.5}(2\pi R\sqrt{\zeta}x) \exp(-0.5x^2) dx \tag{5.86}$$

Then consider the property of Bessel function in Eq. (2.82) [Gradshteyn and Ryzhik, 2007]. In Eq. (5.86) $\alpha = 0.5, \beta = 2\pi R\sqrt{\zeta}, \nu = l + 0.5, \mu = 2i + 1.5$, then we have

$$I_{ln}(R) = \frac{\kappa_n(\zeta)\zeta^{0.5l+1.5}\pi^{l+0.5}R^l}{\Gamma(l+1.5)} \sum_{i=0}^{n} l_n^i 2^{0.5l+i-0.5} \Gamma(0.5l+i+1.5)\, {}_1F_1\left(\frac{2i+l+3}{2}; l+\frac{3}{2}; -2\pi^2R^2\zeta\right)$$

Put it into Eq. (5.85), we proved Eq. (5.13).

Because of the orthogonality of SPF basis and Parseval's theorem, we have

$$\delta_{nlm}^{n'l'm'} = \langle B_{nlm}^{\text{SPF}}(\mathbf{q}), B_{n'l'm'}^{\text{SPF}}(\mathbf{q})\rangle = \langle B_{nlm}^{\text{dSPF}}(\mathbf{R}), B_{n'l'm'}^{\text{dSPF}}(\mathbf{q})\rangle$$

Thus $\{B_{nlm}^{\text{dSPF}}(\mathbf{R})\}$ is an orthonormal function set. $\forall f(\mathbf{R}) \in \mathbb{L}^2(\mathbb{R}^3, 1)$, $\hat{f}(\mathbf{q}) = \mathscr{F}\{f(\mathbf{R})\}(\mathbf{q})$ is also in $\mathbb{L}^2(\mathbb{R}^3, 1)$. So the MSE for given N and L is

$$\left\| f(\mathbf{R}) - \sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} \left\langle f(\mathbf{R}), B_{nlm}^{\text{dSPF}}(\mathbf{R})\right\rangle B_{nlm}^{\text{dSPF}}(\mathbf{R}) \right\|^2 = \left\| \hat{f}(\mathbf{q}) - \sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} \left\langle \hat{f}(\mathbf{q}), B_{nlm}^{\text{SPF}}(\mathbf{q})\right\rangle B_{nlm}^{\text{SPF}}(\mathbf{q}) \right\|^2$$

which means the MSE converges to zero as N and L increase, because SPF is orthonormal basis. So dSPF basis is also an orthonormal basis in $\mathbb{L}^2(\mathbb{R}^3, 1)$. □

Proof of Proposition 5.3 on analytical form of $\Phi_t(\mathbf{r})$. Put Eq. (5.1) into Eq. (4.64), we can easily obtain the solution.

$$\begin{aligned}
\Phi_t(\mathbf{r}) &= \frac{1}{Z}\int_{\Pi_{\mathbf{r}}} \sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm} G_n(q|\zeta) Y_l^m(\mathbf{u}) q\mathrm{d}q\mathrm{d}\mathbf{u} \\
&= \frac{1}{Z}\sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm} \left(\int_{\mathbb{S}^2} Y_l^m(\mathbf{u})\delta(\mathbf{r}^T\mathbf{u})\mathrm{d}\mathbf{u}\right)\left(\int_0^\infty G_n(q|\zeta) q\mathrm{d}q\right) \\
&= \frac{1}{Z}\sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm} \left(2\pi P_l(0) Y_l^m(\mathbf{r})\right)\left(\frac{\kappa_n(\zeta)\zeta}{2}\int_0^\infty \exp(-\frac{x}{2}) L_n^{1/2}(x)\mathrm{d}x\right) \qquad (5.87) \\
&= \frac{2\pi\zeta}{Z}\sum_{l=0}^{L}\sum_{m=-l}^{l}\left(\sum_{n=0}^{N}\sum_{i=0}^{n} \kappa_n(\zeta)\binom{i-0.5}{i}(-1)^{n-i} P_l(0) a_{nlm}\right) Y_l^m(\mathbf{r}) \qquad (5.88)
\end{aligned}$$

where $P_l(0)$ is the Legendre polynomial of order l at 0. We obtain Eq. (5.87) because SH is the eigenfunction of the FRT in Corollary 2.4 [Descoteaux et al., 2007]. From Eq. (5.87) to Eq. (5.88), we use the property of Laguerre polynomial

in Eq. (2.84) [Gradshteyn and Ryzhik, 2007]. Thus here we have a linear transformation from the coefficients a_{nlm} of $E(\mathbf{q})$ to the coefficients $c_{lm}^{\Phi_t}$ in Eq. (5.18). Please note that the Dr. Assemlal in [Assemlal, 2010, pp. 122] gave a solution for $\Phi_t(\mathbf{r})$. However, the integrand there was wrong because of wrong volume element. Here we give the right analytical formulae.

Note that $\Phi_t(\mathbf{r})$ can also be proved by the integration of $P(R\mathbf{r})$, i.e.

$$\Phi_t(\mathbf{r}) = \frac{1}{Z}\left(\int_0^\infty F_{nl}(R|\zeta)\mathrm{d}R\right) Y_l^m(\mathbf{r})$$

Considering Eq. (5.14) and Eq. (2.85), we can obtain the same $c_{lm}^{\Phi_t}$ in Eq. (5.18).

Considering Eq. (2.80), the lower incomplete gamma function in Eq. (2.79), and

$$L_n^{1/2}(x) = \sum_{i=0}^{n} l_n^i x^i, \quad l_n^i = (-1)^i \binom{n+0.5}{n-i}\frac{1}{i!}$$

we have

$$\begin{aligned}\Phi_t(\mathbf{r}, C) &= \frac{2\pi\zeta}{Z}\sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm}\left(P_l(0)Y_l^m(\mathbf{r})\right)\left(\frac{\kappa_n(\zeta)}{2}\int_0^{C^2/\zeta}\exp(-\frac{x}{2})L_n^{1/2}(x)\mathrm{d}x\right) \\ &= \frac{2\pi\zeta}{Z}\sum_{l=0}^{L}\sum_{m=-l}^{l}\left(\sum_{n=0}^{N}\sum_{i=0}^{n}\kappa_n(\zeta)\binom{n+0.5}{n-i}\frac{(-2)^i}{i!}\gamma(i+1, 0.5C^2/\zeta)P_l(0)a_{nlm}\right)Y_l^m(\mathbf{r})\end{aligned} \tag{5.89}$$

Thus we have the coefficients $c_{lm}^{\Phi_t}(C)$ under SH basis in Eq. (5.19).

Note although Eq. (5.18) and Eq. (5.19) seem to be different, it can be verified that

$$\Phi_t(\mathbf{r}) = \lim_{C\to\infty}\Phi_t(\mathbf{r}, C), \quad c_{lm}^{\Phi_t} = \lim_{C\to\infty} c_{lm}^{\Phi_t}(C)$$

□

Proof of Proposition 5.4 on analytical form of $\Phi_w(\mathbf{r})$. Similarly, put Eq. (5.1) into Eq. (4.64), then we can obtain the analytical expression for $\Phi_w(\mathbf{r})$.

$$\Phi_w(\mathbf{r}) = \frac{1}{4\pi} - \frac{1}{8\pi^2}\sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm}\left(\int_0^{2\pi}\Delta_b Y_l^m(\mathbf{u})\delta(\mathbf{r}^T\mathbf{u})d\mathbf{u}\right)\left(\int_0^\infty \frac{G_n(q|\zeta)}{q}dq\right)$$

However, we can not solve it just like what we did for $\Phi_t(\mathbf{r})$, because the division by q introduces a pole. It is a little hard to find the analytical solution for $\Phi_w(\mathbf{r})$, and [Assemlal, 2010; Assemlal et al., 2009a] did not give any solution for that. We solve this problem by considering $E(0) = 1$, which is a true fact for any DWI data. That means, for our basis, the following identity holds for all $\mathbf{u} \in \mathbb{S}^2$.

$$1 = E(0) = \sum_{nlm} a_{nlm}G_n(0)Y_l^m(\mathbf{u}) = \sum_{nlm} a_{nlm}\kappa_n(\zeta)Y_l^m(\mathbf{u}) = 1, \quad \forall\mathbf{u}\in\mathbb{S}^2 \tag{5.90}$$

Also keep in mind that a constant addition inside Δ_b does not change the final result. First we consider the integral inside a given disk $\Pi_{\mathbf{r},C}$ whose radius is C, instead of in the whole plane in Eq. (4.64), then we have

$$\begin{aligned}\Phi_w(\mathbf{r},C) &= \frac{1}{4\pi} - \frac{1}{8\pi^2}\int_{\Pi_{\mathbf{r},C}} \frac{\Delta_b(E(\mathbf{q})-E(0))}{q}\mathrm{d}q\mathrm{d}\mathbf{u} \\ &= \frac{1}{4\pi} - \frac{1}{8\pi^2}\sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm}\left(\int_{\mathbb{S}^2}\Delta_b\left(\int_0^C \frac{G_n(q|\zeta)-G_n(0|\zeta)}{q}dq\right)Y_l^m(\mathbf{u})\delta(\mathbf{r}^T\mathbf{u})\right)\mathrm{d}\mathbf{u} \\ &= \frac{1}{4\pi} - \frac{1}{8\pi^2}\sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm}\kappa_n(\zeta)\left(\int_{\mathbb{S}^2}\Delta_b I_n(C)Y_l^m(\mathbf{r})\delta(\mathbf{r}^T\mathbf{u})\mathrm{d}\mathbf{u}\right) \end{aligned} \tag{5.91}$$

Now there is no pole! For $I_n(C)$ we have

$$\begin{aligned} I_n(C) &= \int_0^C \frac{G_n(q|\zeta)-G_n(0|\zeta)}{\kappa_n(\zeta)q}\mathrm{d}q \\ &= \frac{1}{2}\int_0^{C^2/\zeta}\left(\frac{\exp(-x/2)-1}{x} + \sum_{i=1}^{n} l_n^i x^{i-1}\exp(-\frac{x}{2})\right)\mathrm{d}x \\ &= \underbrace{0.5(-\gamma - E_1(0.5C^2/\zeta) - \log(0.5C^2/\zeta))}_{I^1(C)} + \underbrace{0.5\sum_{i=1}^{n} l_n^i 2^i\gamma(i,0.5C^2/\zeta)}_{I_n^2(C)} \end{aligned} \tag{5.92}$$

where $\gamma \simeq 0.5772$ is the Euler–Mascheroni constant, $E_1(x) = \int_x^\infty \frac{\exp(-t)}{t}dt$ is the exponential integral. Although there are two parts in $I_n(C)$, and $I^1(C)$ is independent of n and it tends to infinity as C increase, it actually has no contribution for $\Phi_w(\mathbf{r},C)$, because

$$\sum_{nlm} a_{nlm}\kappa_n(\zeta)Y_l^m(\mathbf{u})I^1(C) = I^1(C)$$

is a constant inside Δ_b. Then considering $Y_l^m(\mathbf{u})$ is the eigenfunction of FRT and Δ_b, we have the analytical result for $\Phi_w(\mathbf{r},C)$ in Eq. (5.93).

$$\Phi_w(\mathbf{r},C) = \frac{1}{4\pi} - \frac{1}{8\pi}\sum_{n=1}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm}\kappa_n(\zeta)\sum_{i=1}^{n} l_n^i 2^i\gamma(i,0.5C^2/\zeta)P_l(0)(-l)(l+1)Y_l^m(\mathbf{r}) \tag{5.93}$$

Considering $Y_0^0(\mathbf{u}) = \frac{1}{\sqrt{4\pi}}$, we have the coefficients $c_{lm}^{\Phi_w}(C)$ in Eq. (5.21). Let C tends to infinity and consider $\lim_{C\to\infty}\gamma(i,0.5C^2/\xi) = (i-1)!$, then we have

$$\Phi_w(\mathbf{r}) = \lim_{C\to\infty}\Phi_w(\mathbf{r},C), \quad c_{lm}^{\Phi_w} = \lim_{C\to\infty} c_{lm}^{\Phi_w}(C)$$

and coefficients $c_{lm}^{\Phi_w}$ are shown in Eq. (5.20). □

Proof of Theorem 5.1 on scale selection for SPF basis. Rewrite $E(\mathbf{q})$ as

$$E(\mathbf{q}) = \exp\left(-\frac{b_{100}}{\zeta_1}q^2 Y_0^0\right)\exp\left(-\sum_{\substack{nlm\\(n,l,m)\neq(1,0,0)}}^{\infty} b_{nlm}\left(\frac{q^2}{\zeta_1}\right)^n Y_l^m(\mathbf{u})\right)$$

$$= \exp\left(-\frac{b_{100}q^2}{2\zeta_1\sqrt{\pi}}\right)\sum_{k=0}^{\infty}\frac{1}{k!}\left(-\sum_{\substack{nlm\\(n,l,m)\neq(1,0,0)}}^{\infty} b_{nlm}\left(\frac{q^2}{\zeta_1}\right)^n Y_l^m(\mathbf{u})\right)^k \tag{5.94}$$

$$= \exp\left(-\frac{q^2}{2\zeta_*}\right)\left(1+\sum_{\substack{nlm\\(n,l,m)\neq(1,0,0)}}^{\infty} c_{nlm}\left(\frac{q^2}{\zeta_*}\right)^n Y_l^m(\mathbf{u})\right) \tag{5.95}$$

In Eq. (5.94), we use $Y_0^0(\mathbf{u}) = \frac{1}{2\sqrt{\pi}}$ and $\exp(x) = \sum_{k=0}^{\infty}\frac{x^k}{k!}$. In Eq. (5.95), we set $\zeta_* = \frac{\zeta_1\sqrt{\pi}}{b_{100}}$, and put scale ζ_1 into the coefficients. Note when we set $\zeta = \zeta_*$ as the scale parameter in SPF basis, the first term, which is a Gaussian function and normally has the largest energy in DWI signal, is presented by the first SPF basis. If $\zeta \neq \zeta_*$, we need infinite radial terms to represent the Gaussian function $\exp(-\frac{q^2}{2\zeta_*})$, which results in the leak of energy. See the following analysis.

Let $c_{000} = 1$ and $c_{100} = 0$, then Eq. (5.95) is separated into three parts $E_1(\mathbf{q})$, $E_2(\mathbf{q})$ and $E_3(\mathbf{q})$.

$$E(\mathbf{q}) = \underbrace{\sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} c_{nlm}B_{nlm}^{\mathrm{SPNP}}(\mathbf{q}|\zeta_*)}_{E_1(\mathbf{q})} + \underbrace{\sum_{\substack{nlm\\ n>N,\text{ and } l\leq L}}^{\infty} c_{nlm}B_{nlm}^{\mathrm{SPNP}}(\mathbf{q}|\zeta_*)}_{E_2(\mathbf{q})} + \underbrace{\sum_{\substack{nlm\\ l>L}}^{\infty} c_{nlm}B_{nlm}^{\mathrm{SPNP}}(\mathbf{q}|\zeta_*)}_{E_3(\mathbf{q})} \tag{5.96}$$

Note E_3 is orthogonal to E_1, E_2 and $\{B_{nlm}^{\mathrm{SPF}}(\mathbf{q}|\zeta)\}_{l\leq L}$, because the orthogonality of SHs. Thus

$$MSE_{NL}(\zeta) = \|E_3(\mathbf{q})\|^2 + \sum_{l=0}^{L}\sum_{m=-l}^{l}\sum_{n=N+1}^{\infty}\langle E_1(\mathbf{q}) + E_2(\mathbf{q}), B_{nlm}^{\mathrm{SPF}}(\mathbf{q}|\zeta)\rangle^2$$

$$= \|E_3(\mathbf{q})\|^2 + \sum_{l=0}^{L}\sum_{m=-l}^{l}\sum_{n=N+1}^{\infty}\Big(\langle E_1(\mathbf{q}), B_{nlm}^{\mathrm{SPF}}(\mathbf{q}|\zeta)\rangle^2 + \langle E_2(\mathbf{q}), B_{nlm}^{\mathrm{SPF}}(\mathbf{q}|\zeta)\rangle^2$$

$$+ 2\langle E_1(\mathbf{q}), B_{nlm}^{\mathrm{SPF}}(\mathbf{q}|\zeta)\rangle\langle E_2(\mathbf{q}), B_{nlm}^{\mathrm{SPF}}(\mathbf{q}|\zeta)\rangle\Big)$$

where $\|E_3(\mathbf{q})\|^2$ is independent of ζ, and

$$\langle E_1(\mathbf{q}), B_{nlm}^{\mathrm{SPF}}(\mathbf{q}|\zeta)\rangle = \sum_{n_1=0}^{N} c_{n_1lm}\langle B_{n_1lm}^{\mathrm{SPNP}}(\mathbf{q}|\zeta_*), B_{nlm}^{\mathrm{SPF}}(\mathbf{q}|\zeta)\rangle$$

$$\langle E_2(\mathbf{q}), B_{nlm}^{\mathrm{SPF}}(\mathbf{q}|\zeta)\rangle = \sum_{n_2=N+1}^{\infty} c_{n_2lm}\langle B_{n_2lm}^{\mathrm{SPNP}}(\mathbf{q}|\zeta_*), B_{nlm}^{\mathrm{SPF}}(\mathbf{q}|\zeta)\rangle$$

Note because $\|E(\mathbf{q})\|^2$ is finite, we have

$$\lim_{N\to\infty}\sum_{l=0}^{L}\sum_{m=-l}^{l}\sum_{n=N+1}^{\infty}\langle E_2(\mathbf{q}), B_{nlm}^{\mathrm{SPF}}(\mathbf{q}|\zeta)\rangle^2 < \lim_{N\to\infty}\|E_2(\mathbf{q})\|^2 = 0$$

So we have

$$\lim_{N\to\infty} MSE_{NL}(\zeta) = \|E_3(\mathbf{q})\|^2 + \sum_{l=0}^{L}\sum_{m=-l}^{l}\sum_{n=N+1}^{\infty}\langle E_1(\mathbf{q}), B_{nlm}^{\mathrm{SPF}}(\mathbf{q}|\zeta)\rangle^2 \tag{5.97}$$

If $\zeta = \zeta_*$, the linear independent SPNP basis set $\left\{B_{nlm}^{\mathrm{SPNP}}(\mathbf{q}|\zeta_*)\right\}_{0\le n\le N, 0\le|m|\le L}$ and the SPF basis set $\{B_{nlm}^{\mathrm{SPF}}(\mathbf{q}|\zeta)\}_{0\le n\le N, 0\le|m|\le L}$ are two bases of the same function space. That means

$$\sum_{l=0}^{L}\sum_{m=-l}^{l}\sum_{n=N+1}^{\infty}\langle E_1(\mathbf{q}), B_{nlm}^{\mathrm{SPF}}(\mathbf{q}|\zeta)\rangle^2 \begin{cases} = 0 & \text{if } \zeta = \zeta_* \\ > 0 & \text{if } \zeta \neq \zeta_* \end{cases}$$

It is positive if $\zeta \neq \zeta_*$, because $\exp\left(-\frac{q^2}{2\zeta_*}\right)$ can be separated as

$$\exp\left(-\frac{q^2}{2\zeta_*}\right) = \exp\left(-\frac{q^2}{2\zeta}\right)\exp\left(q^2\frac{\zeta_* - \zeta}{2\zeta\zeta_*}\right) = \exp\left(-\frac{q^2}{2\zeta}\right)\sum_{n=0}^{\infty}\left(\frac{\zeta_* - \zeta}{2\zeta\zeta_*}\right)^n q^{2n}$$

which makes E_1 be represented by infinite SPF basis $\{B^{\mathrm{SPF}}(\mathbf{q}|\zeta)\}$. Then considering Eq. (5.97), we have

$$\lim_{N\to\infty} MSE_{NL}(\zeta) \ge \lim_{N\to\infty} MSE_{NL}(\zeta_*) = \|E_3(\mathbf{q})\|^2$$

which proves the theorem. □

Proof of Proposition 5.5 on the invariance of diffusion time τ. Since we only know b values and gradients $\{\mathbf{u}_j\}_{j=1}^{N_s}$, we need to rewrite Eq. (5.36) with respect to b values, i.e.

$$E(q\mathbf{u}) = \exp\left(-\sum_{n=1}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} b_{nlm}\left(\frac{b}{4\pi^2\tau\zeta_1}\right)^n Y_l^m(\mathbf{u})\right) \tag{5.98}$$

Then the solution under a certain estimation for $\boldsymbol{b} = (b_{100}, \cdots, b_{NLL})^T$ will obtain the coefficient of $bY_0^0(\mathbf{u})$ as

$$\frac{b_{100}}{4\pi^2\tau\zeta_1}$$

Note that no matter which estimation is used, the estimated coefficient is only dependent on DWI samples $\{E(\mathbf{q}_j)\}_{j=1}^{N_s}$, b values $\{b_j\}_{j=1}^{N_s}$ and gradients $\{\mathbf{u}\}_{j=1}^{N_s}$, which means

$$T = \frac{b_{100}}{\sqrt{\pi}\tau\zeta_1}$$

is theoretically invariant under changes of τ and ζ_1. However, an inappropriate ζ_1 may result in a large condition number of the basis matrix. So T is theoretically invariant but numerically changes if an inappropriate unit of b and a very large N are used, which is the reason why we set $\zeta_1 = 0.5q_{max}^2$ such that $\frac{b}{4\pi^2\tau\zeta_1} \in (0, 2)$ and we have a good numerical accuracy.

The scale parameter ζ calculated by b_{100} is

$$\zeta = \frac{\zeta_1\sqrt{\pi}}{b_{100}} = \frac{1}{T\tau} \tag{5.99}$$

Now we rewrite the SPF expansion in Eq. (5.1) as

$$E(\mathbf{q}) = \sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm}\kappa_n(\zeta)\exp\left(-\frac{b}{8\pi^2\tau\zeta}\right)L_n^{1/2}\left(\frac{b}{4\pi^2\tau\zeta}\right)Y_l^m(\mathbf{u})$$
$$= \sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm}\kappa_n(\zeta)\exp\left(-\frac{b}{8\pi^2 T}\right)L_n^{1/2}\left(\frac{b}{4\pi^2 T}\right)Y_l^m(\mathbf{u})$$

So no matter which estimation method is used, $a_{nlm}\kappa_n(\zeta)$ is invariant under the change of τ. Or $a_{nlm}\zeta^{-\frac{3}{4}}$ is invariant, considering the definition of $\kappa_n(\zeta)$ in Eq. (5.2). Then based on Eq. (5.18), Eq. (5.19), Eq. (5.20), Eq. (5.21), the estimated ODFs $\Phi_t(\mathbf{r})$, $\Phi_t(\mathbf{r}, C)$, $\Phi_w(\mathbf{r})$, $\Phi_w(\mathbf{r}, C)$ are invariant to the change of τ. Based on Eq. (5.22), Eq. (5.26) and Eq. (5.28), RTO, MSD and GFA are also invariant.

For EAP, based on the definition of $F_{nl}(R)$ in Eq. (5.14), we have

$$F_{nl}(R|\zeta) = \frac{\tau^{-1.5}\pi^{l+1.5}\left(\frac{R}{\sqrt{\tau}}\right)^l\kappa_n(\zeta)}{T^{0.5l+1.5}(-1)^{l/2}\Gamma(l+1.5)}\sum_{i=0}^{n}\binom{n+0.5}{n-i}\frac{(-1)^i}{i!}2^{0.5l+i+1.5}$$
$$\Gamma(0.5l+i+1.5)\,{}_1F_1\left(\frac{2i+l+3}{2}; l+\frac{3}{2}; -\frac{2\pi^2}{T}\left(\frac{R}{\sqrt{\tau}}\right)^2\right)$$

Considering $\zeta = \frac{1}{T\tau}$ and $a_{nlm}\kappa_n(\zeta)$ are invariant, we have

$$P(R\mathbf{r}, \tau_2) = \sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{nlm}F_{nl}(R|\zeta)Y_l^m(\mathbf{r}) = \left(\frac{\tau_1}{\tau_2}\right)^{3/2}P\left(\sqrt{\frac{\tau_1}{\tau_2}}R\mathbf{r}, \tau_1\right) \tag{5.100}$$

where $P(R\mathbf{r}, \tau_1)$ and $P(R\mathbf{r}, \tau_2)$ are the EAPs estimation from τ_1 and τ_2 respectively. So the axis of EAP will be scaled when τ changes. □

5.7 APPENDIX B: PROOFS ON THE RELATION OF THREE THEOREMS ON SPHERICAL HARMONICS

Theorem 5.4. *Funk-Hecke theorem 2.7 and addition theorem 2.6 are equivalent.*

Proof. The proof has two parts.

1. *From Funk-Hecke theorem 2.7 to addition theorem 2.6.* Let $g(\boldsymbol{u}^T\boldsymbol{v}) = P_l(\boldsymbol{u}^T\boldsymbol{v})$. For a given $\boldsymbol{v}$, $P_l(\boldsymbol{u}^T\boldsymbol{v})$ is a spherical function of $\boldsymbol{u}$. So $P_l(\boldsymbol{u}^T\boldsymbol{v})$ can be written in Laplace series as

$$P_l(\boldsymbol{u}^T\boldsymbol{v}) = \sum_{l'=0}^{\infty}\sum_{m=-l'}^{l'} c_{l'}^m y_{l'}^m(\boldsymbol{u}) \quad c_{l'}^m = \int_{\mathbb{S}^2} P_l(\boldsymbol{u}^T\boldsymbol{v})y_{l'}^m(\boldsymbol{u})\mathrm{d}\boldsymbol{u}$$

Based on Funk-Hecke theorem 2.7, we have

$$c_{l'}^m = \int_{\mathbb{S}^2} P_l(\boldsymbol{u}^T\boldsymbol{v})y_{l'}^m(\boldsymbol{u})\mathrm{d}\boldsymbol{u} = 2\pi y_{l'}^m(\boldsymbol{v})\int_{-1}^{1} P_l(t)P_{l'}(t)\mathrm{d}t = \frac{4\pi}{2l+1}y_{l'}^m(\boldsymbol{v})\delta_{ll'}$$

So $P_l(\boldsymbol{u}^T\boldsymbol{v}) = \sum_{m=-l}^{l}\frac{4\pi}{2l+1}y_l^m(\boldsymbol{v})y_l^m(\boldsymbol{u})$, which is the addition theorem 2.6.

2. *From addition theorem 2.6 to Funk-Hecke theorem 2.7.* $\forall \boldsymbol{u}, \boldsymbol{v} \in \mathbb{S}^2$ and $g(t) \in \mathbb{L}^2([-1,1],1)$, we have

$$g(\boldsymbol{u}^T\boldsymbol{v}) = \sum_{l'=0}^{\infty} c_{l'} P_{l'}(\boldsymbol{u}^T\boldsymbol{v}) \quad c_{l'} = \frac{2l'+1}{2}\int_{-1}^{1} g(t)P_{l'}(t)\mathrm{d}t$$

Based on addition theorem 2.6, we have

$$g(\boldsymbol{u}^T\boldsymbol{v}) = 2\pi \sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'} \Big(\int_{-1}^{1} g(t)P_{l'}(t)\mathrm{d}t\Big) y_{l'}^{m'}(\boldsymbol{u})\overline{y_{l'}^{m'}(\boldsymbol{v})}$$

After performing inner product with $y_l^m(\boldsymbol{v})$ in two sides of above equation, we have

$$\int_{\mathbb{S}^2} g(\boldsymbol{u}^T\boldsymbol{v}) y_l^m(\boldsymbol{v})\mathrm{d}\boldsymbol{v} = \Big(2\pi\int_{-1}^{1} g(t)P_l(t)\mathrm{d}t\Big) y_l^m(\boldsymbol{u})$$

which is the Funk-Hecke theorem in (2.36).

□

Theorem 5.5. *Funk-Hecke theorm 2.7 and addition theorem 2.6 can be proved by plane wave expansion theorem 2.8.*

Proof. Because the equivalence between Funk-Hecke theorm 2.7 and addition theorem 2.6, we only need to prove addition theorem 2.6 from plane wave expansion theorem 2.8, which is as follows.

In Eq. (2.52), let $\boldsymbol{x} = (R,\theta,\phi)$ and $\boldsymbol{k} = (1,\theta,\phi)$, then integrate (θ,ϕ) over $\mathbb{S}^2$, we have

$$e^{\pm iR} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} (\pm i)^l j_l(R) = \sum_{l=0}^{\infty} (\pm i)^l (2l+1) j_l(R) \tag{5.101}$$

In Section 2.5.3, we have already shown that the plane wave $e^{i\boldsymbol{x}^T\boldsymbol{k}}$ is a solution of Helmholtz equation in Eq. (2.48), and the general solution is given in Eq. (2.50). If $\boldsymbol{x} = (R,\theta,\phi)$ and $\boldsymbol{k} = (k,0,0)$, then $\boldsymbol{x}^T\boldsymbol{k} = Rk\cos(\theta)$ which is independent of ϕ. If we assume the general form of the solution that is independent of ϕ as $g(\boldsymbol{x}) = F(R)\Theta(\theta)$, then after some derivation based on separation of variables, we obtain the general solution independent of ϕ as

$$g(R,\theta) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} (a_l^m j_l(kR) + b_l^m y_l(kR)) P_l(\cos(\theta))$$

Then $b_l^m = 0$ since $y_l(Rk)$ is singular at $R = 0$, and we have

$$e^{iRk\cos(\theta)} = \sum_{l=0}^{\infty} a_l j_l(kR) P_l(\cos(\theta))$$

Let $k = 1$, $\theta = 0$ (or $\theta = \pi$), and consider $P_l(0) = 1$, $P_l(-x) = (-l)^l P_l(x)$, $\forall l$, we have

$$e^{\pm iR} = \sum_{l=0}^{\infty} (\pm 1)^l a_l j_l(R) \tag{5.102}$$

Comparing Eq. (5.102) with Eq. (5.101), we have $a_l = 2l + 1$ and

$$e^{\pm iR\cos(\theta)} = \sum_{l=0}^{\infty} (\pm i)^l (2l+1) j_l(R) P_l(\cos(\theta)) \tag{5.103}$$

For two 3D vectors $\boldsymbol{x} = R_{\boldsymbol{x}}\boldsymbol{u}$ and $\boldsymbol{k} = R_{\boldsymbol{k}}\boldsymbol{v}$, where $\boldsymbol{u}$ and $\boldsymbol{v} \in \mathbb{S}^2$, we have

$$e^{\pm i\boldsymbol{x}^T\boldsymbol{k}} = e^{\pm iR_{\boldsymbol{x}}R_{\boldsymbol{k}}\boldsymbol{u}^T\boldsymbol{v}} = \sum_{l=0}^{\infty} (\pm i)^l (2l+1) j_l(R_{\boldsymbol{x}}R_{\boldsymbol{k}}) P_l(\boldsymbol{u}^T\boldsymbol{v}) \tag{5.104}$$

Since $P_l(\boldsymbol{u}^T\boldsymbol{v})$ is a spherical function of $\boldsymbol{v}$ and $\boldsymbol{u}$, we can set it as

$$P_l(\boldsymbol{u}^T\boldsymbol{v}) = \sum_{l_1,m_1} \sum_{l_2,m_2} c_{ll_1l_2}^{m_1m_2} y_{l_1}^{m_1}(\boldsymbol{u}) \overline{y_{l_2}^{m_2}(\boldsymbol{v})}$$

where $\sum_{l_1,m_1} \stackrel{\text{def}}{=} \sum_{l_1=0}^{\infty} \sum_{m_1=-l_1}^{l_1}$. Then by combining the above two formulae and comparing them with Eq. (2.52), we have the equation

$$\sum_{l,m} 4\pi(\pm i)^l j_l(R_{\boldsymbol{x}}R_{\boldsymbol{k}}) y_l^m(\boldsymbol{u}) \overline{y_l^m(\boldsymbol{v})} = \sum_{l'=0}^{\infty} \sum_{l_1,m_1} \sum_{l_2,m_2} (\pm i)^{l'} (2l'+1) j_{l'}(R_{\boldsymbol{x}}R_{\boldsymbol{k}}) c_{l'l_1l_2}^{m_1m_2} y_{l_1}^{m_1}(\boldsymbol{u}) \overline{y_{l_2}^{m_2}(\boldsymbol{v})}$$

Then by performing inner products with $y_{l_1}^{m_1}(\boldsymbol{v})$ and $\overline{y_{l_2}^{m_2}(\boldsymbol{u})}$ in both sides, we have

$$4\pi(\pm i)^l j_l(R_{\boldsymbol{x}}R_{\boldsymbol{k}}) \delta_{ll_1}^{mm_1} \delta_{ll_2}^{mm_2} = \sum_{l'=0}^{\infty} (\pm i)^{l'} (2l'+1) j_{l'}(R_{\boldsymbol{x}}R_{\boldsymbol{k}}) c_{l'l_1l_2}^{m_1m_2}$$

The above equality holds $\forall\ R_{\boldsymbol{x}}R_{\boldsymbol{k}}$. So we have

$$c_{ll_1l_2}^{m_1m_2} = \delta_{ll_1}^{mm_1} \delta_{ll_2}^{mm_2} \frac{4\pi}{2l+1}$$

then

$$P_l(\boldsymbol{u}^T\boldsymbol{v}) = \frac{4\pi}{2l+1} \sum_{m=-l}^{l} y_l^m(\boldsymbol{u}) \overline{y_l^m(\boldsymbol{v})}$$

which proves the addition theorem 2.6. □

5.8 APPENDIX C: EXACT EAP AND ITS FEATURES IN MIXTURE OF TENSOR MODEL

In this Appendix, we would like to deduce the exact EAP and its features in Mixture of Tensor Model where the diffusion signal attenuation is given in Eq. (4.32), i.e.

$$E(\mathbf{q}) = \sum_{i=1}^{K} w_i \exp(-4\pi^2\tau q^2 \mathbf{u}^T \mathbf{D}_i \mathbf{u}) = \sum_{i=1}^{K} w_i \exp(-4\pi^2\tau \mathbf{q}^T \mathbf{D}_i \mathbf{q}) \tag{5.105}$$

where $\sum_{i=1}^{K} w_i = 1$, $w_i \geq 0$. The EAP and its features for free diffusion is a specific case when $w_i = \delta_{i1}$, because Fourier transform is linear. Thus we can first deduce the EAP and its features in tensor model where $E(\mathbf{q}) = \exp(-4\pi^2\tau\mathbf{q}^2\mathbf{D}\mathbf{q})$, then easily generate it to mixture of tensor model.

EAP: The EAP $P(\mathbf{R})$ is the Fourier transform of $E(\mathbf{q})$ in Eq. (4.24). Considering $\mathscr{F}\{\exp(-ax^2)\} = \sqrt{\frac{\pi}{a}}\exp(-\frac{\pi^2}{a}\xi^2)$ and $D = \sum_{i=1}^{3}\lambda_i \boldsymbol{v}_i\boldsymbol{v}_i^T$, where $\{\lambda_i\}$ and $\{\boldsymbol{v}_i\}$ are eigenvalues and eigenvectors, we have

$$
\begin{aligned}
P(\mathbf{R}) &= \mathscr{F}_{3D}\{\exp(-4\pi^2\tau\mathbf{q}^T\mathbf{D}\mathbf{q})\} \\
&= \mathscr{F}_{3D}\left\{\exp\left(-4\pi^2\tau(\lambda_1 q_1^2+\lambda_2 q_2^2+\lambda_3 q_3^2)\right)\right\} && (5.106)\\
&= \mathscr{F}_{3D}\left\{\exp\left(-4\pi^2\tau\lambda_1 q_1^2\right)\exp\left(-4\pi^2\tau\lambda_2 q_2^2\right)\exp\left(-4\pi^2\tau\lambda_3 q_3^2\right)\right\} \\
&= \mathscr{F}_{1D}\left\{\exp\left(-4\pi^2\tau\lambda_1 q_1^2\right)\right\}\mathscr{F}_{1D}\left\{\exp\left(-4\pi^2\tau\lambda_2 q_2^2\right)\right\}\mathscr{F}_{1D}\left\{\exp\left(-4\pi^2\tau\lambda_3 q_3^2\right)\right\} && (5.107)\\
&= \frac{1}{\sqrt{4\pi\tau\lambda_1}}\exp\left(-\frac{R_1^2}{4\tau\lambda_1}\right)\frac{1}{\sqrt{4\pi\tau\lambda_2}}\exp\left(-\frac{R_2^2}{4\tau\lambda_2}\right)\frac{1}{\sqrt{4\pi\tau\lambda_3}}\exp\left(-\frac{R_3^2}{4\tau\lambda_3}\right) \\
&= \frac{1}{\sqrt{(4\pi\tau)^3\lambda_1\lambda_2\lambda_3}}\exp\left(-\frac{1}{4\tau}\left(\frac{R_1^2}{\lambda_1}+\frac{R_2^2}{\lambda_2}+\frac{R_3^2}{\lambda_3}\right)\right) \\
&= \frac{1}{\sqrt{(4\pi\tau)^3|\mathbf{D}|}}\exp\left(\frac{-\mathbf{R}^T\mathbf{D}^{-1}\mathbf{R}}{4\tau}\right) = N(\mathbf{R}|2\tau\mathbf{D}) && (5.108)
\end{aligned}
$$

In Eq. (5.106) we set $q_i = \mathbf{q}^T\boldsymbol{v}_i$. In Eq. (5.107) we set $R_i = \mathbf{R}^T\boldsymbol{v}_i$, and separate 3D Fourier transform into three 1D Fourier transforms, by considering $\{\boldsymbol{v}_i\}$ is an orthonormal basis in $\mathbb{R}^3$. In Eq. (5.108) we use $\mathbf{D}^{-1} = \sum_{i=1}^{3}\lambda_i^{-1}\boldsymbol{v}\boldsymbol{v}^T$ and $|\mathbf{D}| = \lambda_1\lambda_2\lambda_3$. Then $P(\mathbf{R})$ for mixture tensor model is

$$
P(\mathbf{R}) = \sum_{i=1}^{K} w_i N(\mathbf{R}|2\tau\mathbf{D}_i) = \sum_{i=1}^{K}\frac{w_i}{\sqrt{(4\pi\tau)^3|\mathbf{D}_i|}}\exp\left(\frac{-\mathbf{R}^T\mathbf{D}_i^{-1}\mathbf{R}}{4\tau}\right) \tag{5.109}
$$

Return-To-Origin (RTO) Probability: It is straightforward that the RTO value for mixture of tensor model is

$$
Po = P(0) = \sum_{i=1}^{K}\frac{w_i}{\sqrt{(4\pi\tau)^3|\mathbf{D}_i|}} \tag{5.110}
$$

Mean-Squared Displacement (MSD): The covariance matrix of the EAP for mixture of tensor model is

$$
E_{\mathbb{R}^3}(\mathbf{R}\mathbf{R}^T) = \sum_{i=1}^{K} w_i\int_{\mathbb{R}^3}\mathbf{R}\mathbf{R}^T N(\mathbf{R}|2\tau\mathbf{D}_i)\mathrm{d}\mathbf{R} = \sum_{i=1}^{K} 2w_i\tau\mathbf{D}_i
$$

So the MSD value for mixture of tensor model is

$$
MSD = \int_{\mathbb{R}^3} P(\mathbf{R})\mathbf{R}^T\mathbf{R}\mathrm{d}\mathbf{R} = \mathrm{Trace}\left(E_{\mathbb{R}^3}(\mathbf{R}\mathbf{R}^T)\right) = \sum_{i=1}^{K} 2w_i\tau\,\mathrm{Trace}(\mathbf{D}_i) \tag{5.111}
$$

ODFs: Considering Eq. (2.80), the ODF $\Phi_k(\mathbf{r})$ defined in Eq. (4.69) is

$$\begin{aligned}
\Phi_k(\mathbf{r}) &= \frac{1}{Z}\int_0^\infty \sum_{i=1}^K w_i N(\mathbf{R}|2\tau\mathbf{D})R^k \mathrm{d}R \\
&= \frac{1}{Z}\sum_{i=1}^K w_i \frac{1}{\sqrt{(4\pi\tau)^3|\mathbf{D}_i|}}\int_0^\infty \exp\left(-R^2\frac{\mathbf{r}^T\mathbf{D}_i^{-1}\mathbf{r}}{4\tau}\right)R^k\mathrm{d}R \\
&= \frac{1}{2\sqrt{(4\pi\tau)^3}Z}\sum_{i=1}^K w_i|\mathbf{D}_i|^{-\frac{1}{2}}\left(\frac{4\tau}{\mathbf{r}^T\mathbf{D}_i^{-1}\mathbf{r}}\right)^{\frac{k+1}{2}}\Gamma(\frac{k+1}{2}) \\
&= \frac{2^{k-3}\tau^{\frac{k-2}{2}}\Gamma(\frac{k+1}{2})}{Z\pi^{\frac{3}{2}}}\sum_{i=1}^K \frac{w_i}{|\mathbf{D}_i|^{\frac{1}{2}}\left(\mathbf{r}^T\mathbf{D}_i^{-1}\mathbf{r}\right)^{\frac{k+1}{2}}} \qquad (5.112) \\
&= \begin{cases} \frac{1}{4\pi}\sum_{i=1}^K \frac{w_i}{|\mathbf{D}_i|^{\frac{1}{2}}\left(\mathbf{r}^T\mathbf{D}_i^{-1}\mathbf{r}\right)^{\frac{3}{2}}} & \text{if } k=2 \\ \frac{1}{Z}\sum_{i=1}^K \frac{w_i}{|\mathbf{D}_i|^{\frac{1}{2}}\left(\mathbf{r}^T\mathbf{D}_i^{-1}\mathbf{r}\right)^{\frac{k+1}{2}}} & \text{if } k\neq 2 \end{cases} \qquad (5.113)
\end{aligned}$$

Note the normalization factor in Eq. (5.113) when $k \neq 2$ is different from Z in Eq. (5.112). The Z factor is set such that the integration of ODF is 1. In practice it can be set in a discrete way or in a continuous way. See Appendix 4.6.

For tensor model, Eq. (5.113) can be more simple as

$$\Phi_k(\mathbf{r}) = \begin{cases} \frac{1}{4\pi|\mathbf{D}|^{\frac{1}{2}}}\frac{1}{\left(\mathbf{r}^T\mathbf{D}^{-1}\mathbf{r}\right)^{\frac{3}{2}}} & \text{if } k=2 \\ \frac{1}{Z}\frac{1}{\left(\mathbf{r}^T\mathbf{D}^{-1}\mathbf{r}\right)^{\frac{k+1}{2}}} & \text{if } k\neq 2 \end{cases} \qquad (5.114)$$

Please note that because of Eq. (5.114), some papers, e.g. [Descoteaux, 2008, pp. 142], consider the ODF by Tuch $\Phi_t(\mathbf{r})$ in mixture of tensor model is

$$\Phi_t(\mathbf{r}) = \sum_{i=1}^K \frac{w_i}{Z}\frac{1}{\left(\mathbf{r}^T\mathbf{D}_i^{-1}\mathbf{r}\right)^{\frac{1}{2}}}$$

which is actually wrong. Note this is not a linear relation, because different term with $\mathbf{D}_i$ may have different Z_i. The above formula is correct only if every Gaussian term has the same tensor, i.e. $\mathbf{D}_i = \mathbf{D}$, $\forall i$.

CHAPTER 6

RIEMANNIAN FRAMEWORK FOR ODFS AND EAPS

"We lay down a fundamental principle of generalization by abstraction: 'The existence of analogies between central features of various theories implies the existence of a general theory which underlies the particular theories and unifies them with respect to those central features....' "

– E. H. Moore

Contents

Overview

Fisher information metric has been constructed for probability distribution family in Information Geometry theory and it has been successfully applied for tensor computing in DTI on tensor estimation, filtering, registration, statistical analysis, etc. However, to our knowledge, existing works in HARDI mainly focus on ODF/EAP estimation, and there are only a few works on how to process the estimated ODF data [McGraw et al., 2006; Goh et al., 2009a, 2011] and no work for EAP data processing so far. Please see Section 4.4 for related works on the metrics for ODFs and EAPs. In this chapter, we propose a general state-of-the-art Riemannian framework as a mathematical tool to process PDF valued data, e.g. ODFs and EAPs, by representing the square root of the PDF, called wavefunction based on quantum mechanics, as a linear combination of some orthonormal basis functions. This formulation avoids the Gaussian assumption in the DTI. The proposed Riemannian framework is the natural extension of previous Riemannian framework for tensors.

Organization of this chapter:
In Section 6.1, we present the general formulation of the Riemannian framework for parametric family based on square root parametrization and orthonormal basis representation in Section 6.1.1. The closed forms of the geodesic distance, exponential map and logarithmic map are presented in Section 6.1.2. We demonstrate how to choose an appropriate basis for ODFs and EAPs in Section 6.1.3, analyze theoretically the properties of the statistical manifold which is a convex subset of a high dimensional unit sphere in Section 6.1.4, and prove the existence and uniqueness of weighted Riemannian mean and median in the manifold in Section 6.1.7. geodesic anisotropy (GA), Rényi entropy, and Log-Euclidean framework, Affine-Euclidean framework are also proposed in Section 6.1.5 and Section 6.1.6. The Riemannian metric is proved to be diffeomorphism invariant in Section 6.1.8, and is compared with Euclidean metric in Section 6.1.9. Section 6.2.1 proposes to estimate wavefunction from pre-estimated ODFs/EAPs, and Section 6.2.1 proposes a novel nonnegative definite ODF and EAP estimation method, named as Square Root Parameterized Estimation (SRPE). Section 6.2 also lists some potential applications. e.g. nonnegative definite ODF/EAP estimation, interpolation, PGA, filtering, atlas estimation. The experimental results in Section 6.3 validates the proposed frameworks.

6.1 General Riemannian Framework Based On Square Root Parametrization and Orthonormal Basis Representation

In this section, we introduce a general Riemannian framework for arbitrary PDFs whose square root, called **wavefunction**, can be represented by arbitrary orthonormal basis, and we apply it to ODFs/EAPs in dMRI.

6.1.1 Parametric Family, Wavefunction

In quantum mechanics, the square root of the probability of finding the subject at a certain time and position is called as wavefunction. Analogously, we call the square root of the PDF $p(\boldsymbol{x})$ defined in χ as the **wavefunction** denoted by $\psi(\boldsymbol{x})$, i.e. $\psi(\boldsymbol{x}) = \sqrt{p(\boldsymbol{x})}$. Let $\{B_i(\boldsymbol{x})\}_{i=1}^{\infty}$ be a given orthonormal basis in $\mathbb{L}^2(\chi, 1)$ which can sparsely represent $\psi(\boldsymbol{x})$, then $\psi(\boldsymbol{x})$ can be represented by finite linear combination of $\{B_i(\boldsymbol{x})\}$, i.e.

$$\psi(\boldsymbol{x}|\boldsymbol{c}) = \sqrt{p(\boldsymbol{x})} = \sum_{i=1}^{K} c_i B_i(\boldsymbol{x})$$

Please note that the assumption here is very weak, because in practice we always can choose a large enough K to represent $\psi(\boldsymbol{x})$. The Parametric Family (PF) is given in Eq. (6.1), where PF_K means PF with the K parameters. All PDFs will be contained in PF_K if K is large enough thanks to the orthonormal basis representation.

$$PF_K = \left\{ p(\boldsymbol{x}|\boldsymbol{c}) = \psi(\boldsymbol{x}|\boldsymbol{c})^2 : \begin{array}{l} \psi(\boldsymbol{x}|\boldsymbol{c}) = \sum_{i=1}^{K} c_i B_i(\boldsymbol{x}) \geq 0, \\ \int p(\boldsymbol{x}|\boldsymbol{c}) \mathrm{d}\boldsymbol{x} = \sum_{i=1}^{K} c_i^2 = 1 \end{array} \right\} \tag{6.1}$$

PF_K in Eq. (6.1) shows that the Parameter Space (PS) denoted by PS_k is a subset of $\mathbb{S}^{K-1}$.

$$PS_K = \left\{ \boldsymbol{c} \ : \ \|\boldsymbol{c}\| = \sum_{i=1}^{K} c_i^2 = 1, \ \sum_{i=1}^{K} c_i B_i(\boldsymbol{x}) \geq 0, \ \forall \boldsymbol{x} \in \chi \right\} \tag{6.2}$$

Normally we need to choose a good basis to sparsely represent continuous function $\psi(\boldsymbol{x})$ so that K can be as small as possible.

If $\{A_i\}_{i=1}^{K}$ are K given disjoint subsets with the areas of $\{S_i\}$ in χ and $\bigcup_{i=1}^{K} A_i = \chi$, we have the orthonormal basis defined by $\{A_i\}$ in Eq. (6.3), where $\delta(\boldsymbol{x} \in A_i)$ is 1 if $\boldsymbol{x} \in A_i$ and 0 otherwise. This basis is called as the **piecewise constant basis**.

$$\left\{ B_i(\boldsymbol{x}) = \frac{1}{\sqrt{S_i}} \delta(\boldsymbol{x} \in A_i) \right\} \tag{6.3}$$

Then if we assume the PDF $p(\boldsymbol{x})$ is piecewise constant in each A_i, i.e.

$$p(\boldsymbol{x}) = \sum_{i=1}^{K} p_i \frac{1}{S_i} \delta(\boldsymbol{x} \in A_i) \tag{6.4}$$

then PF_K represented by this basis is given in Eq. (6.5), where $p_i = c_i^2$ and

$$PF_K = \left\{ p(\boldsymbol{x}|\boldsymbol{c}) = \psi(\boldsymbol{x}|\boldsymbol{c})^2 : \begin{array}{l} \psi(\boldsymbol{x}|\boldsymbol{c}) = \sum_{i=1}^{K} c_i \frac{1}{\sqrt{S_i}} \delta(\boldsymbol{x} \in A_i), \\ c_i \geq 0, \ \sum_{i=1}^{K} c_i^2 = 1 \end{array} \right\} \tag{6.5}$$

Eq. (6.5) is equivalent to the multinomial distribution family shown in Eq. (3.55).

Although PF_K and PS_K in Eq. (6.1) are similar with the multinomial distribution in Eq. (3.55), there exist some important differences.

1. We consider $p(\boldsymbol{x})$ as a continuous PDF in Eq. (6.1), while multinomial distribution is a discrete PDF in Eq. (3.55) (or piecewise constant in the continuous space χ in Eq. (6.5)).

2. Our formulation is more general, because the multinomial distribution can be seen as the specific case of our formulation when the orthonormal basis is chosen as the piecewise constant basis in Eq. (6.3).

PS_K in Eq. (6.2) is in a high dimensional space for large K, so it is impossible to visualize it. Fig. 3.3 shows the statistical manifold of multinomial distribution family $PS_3 \subset \mathbb{R}^3$, where $p(\boldsymbol{x})$ can be represented by three piecewise constant orthonormal basis functions, i.e. $p(\boldsymbol{x}) = \sum_{i=1}^{3} p_i \frac{1}{S_i} \delta(\boldsymbol{x} \in A_i)$. The left side of Fig. 3.3 shows the PS_3 parameterized by $\{p_i\}$, which is an equilateral triangle with vertices $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$, and the right side shows the PS_3 parameterized by $\{c_i\}$ which is the positive orthant of $\mathbb{S}^2$. In order to better understand the material in this chapter, we put Fig. 3.3 into Fig. 6.1(A) and (B).

Please note that [Goh et al., 2009a, 2011] directly used the results of multinomial distribution in Section 3.3.2 for ODFs, while we consider a different orthonormal basis representation in Eq. (6.1). If we define $\{A_i\}_{i=1}^K$ as K disjoint subsets of $\mathbb{S}^2$, then Eq. (6.5) is the PF formulation for ODFs used in [Goh et al., 2009a, 2011], which used the histogram (multinomial distribution shown in Example 3.11) with many samples to approximate continuous ODFs. Fig. 3.3 (or Fig. 6.1(A,B)) is the simplified case $K = 3$ for PF_K. Thus we have shown here that our formulation is more general than [Goh et al., 2009a, 2011].

6.1.2 Riemannian Metric, Exponential Map and Logarithmic Map

Based on PF_K in Eq. (6.1) and the definition in Eq. (3.40), we have the Fisher information metric as

$$\begin{aligned} g_{ij}(\boldsymbol{c}) &= 4 \int_{\chi} \frac{\partial_i \sqrt{p(\boldsymbol{x}|\boldsymbol{c})}}{\partial c_i} \frac{\partial_j \sqrt{p(\boldsymbol{x}|\boldsymbol{c})}}{\partial c_j} \mathrm{d}\boldsymbol{x} \\ &= 4 \int_{\chi} \frac{\partial_i \left(\sum_{m=1}^{K} c_m B_m(\boldsymbol{x}) \right)}{\partial c_i} \frac{\partial_j \left(\sum_{m=1}^{K} c_m B_m(\boldsymbol{x}) \right)}{\partial c_j} \mathrm{d}\boldsymbol{x} \\ &= 4 \int_{\chi} B_i(\boldsymbol{x}) B_j(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = 4\delta_{ij} \end{aligned} \tag{6.6}$$

which is the same as the metric of multinomial distributions in Eq. (3.57). The constant 4 in the metric can be ignored because a constant proportion change on metric does not affect the final results of the computational methods based on metric. Thus we can set $g_{ij} = \delta_{ij}$. Then PS_K is a subset of the unit sphere $\mathbb{S}^{K-1}$ and the Fisher information metric is the Euclidean metric induced from $\mathbb{R}^K$ by the constraint on $\mathbb{S}^{K-1}$. The geodesic connecting two points $p(\boldsymbol{x}|\boldsymbol{c})$ and $p(\boldsymbol{x}|\boldsymbol{c}')$ is $p(\boldsymbol{x}|\boldsymbol{c}(t))$ where $\boldsymbol{c}(t)$ is in the great circle connecting $\boldsymbol{c}$ and $\boldsymbol{c}'$ in $\mathbb{S}^{K-1}$, based on Example 3.7 on the geodesic of sphere. Then the geodesic distance is the angle between $\boldsymbol{c}$ and $\boldsymbol{c}'$, i.e. $\arccos(\boldsymbol{c}^T \boldsymbol{c}') = \arccos(\int_{\chi} \psi(\boldsymbol{x}|\boldsymbol{c}) \psi(\boldsymbol{x}|\boldsymbol{c}') \mathrm{d}\boldsymbol{x})$, where $\int_{\chi} \psi(\boldsymbol{x}|\boldsymbol{c}) \psi(\boldsymbol{x}|\boldsymbol{c}') \mathrm{d}\boldsymbol{x}$ is the so called **Bhattacharyya coefficient** [Bhattacharyya, 1943] that describes the amount of overlap between two distributions. Let $\boldsymbol{v}_{\boldsymbol{c}}$ denote the tangent vector at $\boldsymbol{c}$ towards $\boldsymbol{c}'$, then the geodesic, exponential map, logarithmic map all have closed forms based on the geometry of sphere shown in Example 3.5, 3.7, 3.8. See Fig. 3.3 for the sketch map of PS_3.

$$\text{Geodesic: } \gamma(t) : p(\boldsymbol{x}|\boldsymbol{c}(t)), \quad \text{where } \boldsymbol{c}(t) = Exp_{\boldsymbol{c}}(tLog_{\boldsymbol{c}}(\boldsymbol{c}')) \tag{6.7}$$

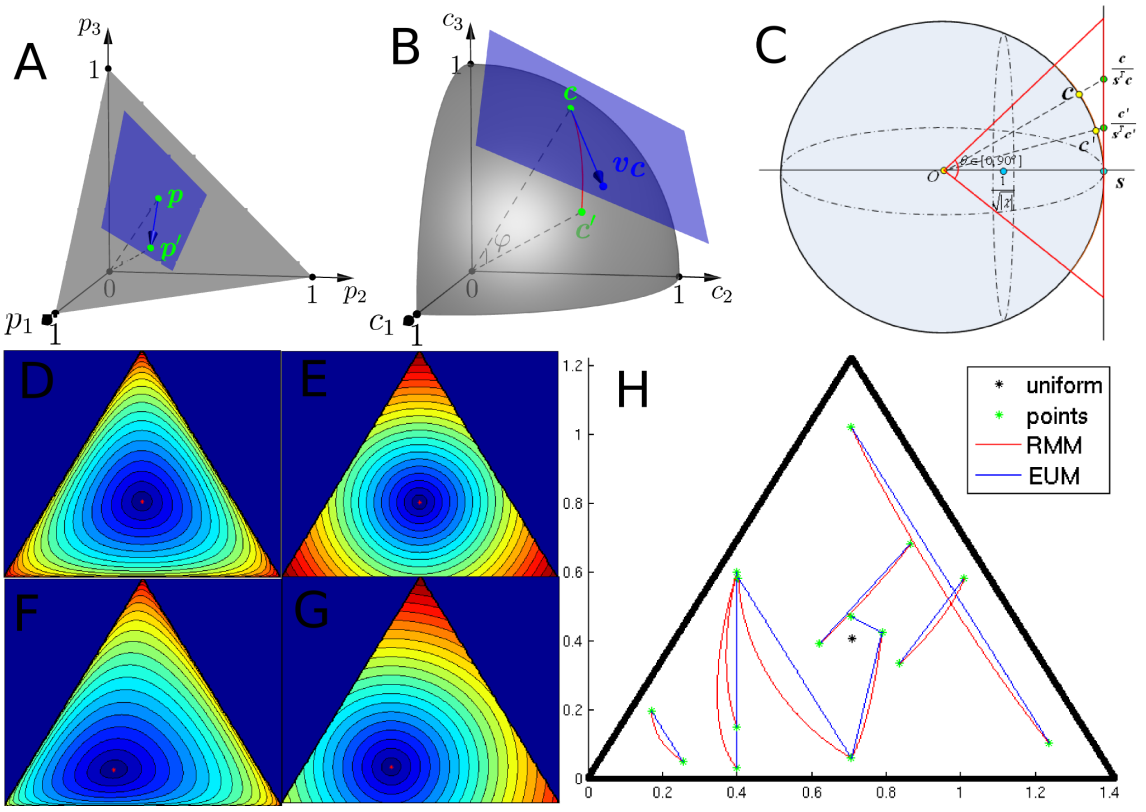


Figure 6.1: Parameter space PS_K, its properties, and comparisons between Riemannian metric and Euclidean metric. (A) and (B): PS_3 represented by $\{p_i\}$ and $\{c_i\}$; (C): properties of PS_K; (D) and (F): isolines of Riemannian metric for given red point; (E) and (G): isolines of Euclidean metric for given red point; (H): geodesics of Riemannian metric (RMM, red) and Euclidean metric (EUM, blue) for given two green points.

$$\text{Exp: } Exp_{\boldsymbol{c}}(\boldsymbol{v_c}) = \boldsymbol{c}' = \boldsymbol{c}\cos\varphi + \frac{\boldsymbol{v_c}}{\|\boldsymbol{v_c}\|}\sin\varphi, \text{ where } \varphi = \|\boldsymbol{v_c}\| \tag{6.8}$$

$$\text{Log: } Log_{\boldsymbol{c}}(\boldsymbol{c}') = \boldsymbol{v_c} = \frac{\boldsymbol{c}' - \boldsymbol{c}\cos\varphi}{\|\boldsymbol{c}' - \boldsymbol{c}\cos\varphi\|}\varphi, \text{ where } \varphi = \arccos(\boldsymbol{c}^T\boldsymbol{c}') \tag{6.9}$$

In Eq. (6.1), we do not assume a specific orthonormal basis $\{B_i(\boldsymbol{x})\}_{i=1}^{\infty}$. For a given function, it may be represented by two orthonormal basis. We have the following theorem for different basis.

Theorem 6.1 (Invariance Under Different Basis). *For different orthonormal basis, the geodesic $\gamma(t)$ in Eq. (6.7) is invariant and the exponential map in Eq. (6.8) and logarithmic map in Eq. (6.9) are invariant under a change of basis matrix.*

Proof. Let's assume there is another orthonormal basis $\{W_\alpha(\boldsymbol{x})\}_{\alpha=1}^{\infty}$ in χ and the wavefunction we focus can be represented as $\psi(\boldsymbol{x}|\boldsymbol{b}) = \sum_{\alpha=1}^{N} b_\alpha W_\alpha(\boldsymbol{x})$ as well as $\psi(\boldsymbol{x}|\boldsymbol{c}) = \sum_{i=1}^{K} c_i B_i(\boldsymbol{x})$. Also assume $B_i(\boldsymbol{x}) = \sum_{\alpha=1}^{N} M_\alpha^i W_\alpha(\boldsymbol{x})$, where $M = [M_\alpha^i]_{N\times K}$ is the change of basis orthogonal matrix. Then we have

$$\boldsymbol{b} = M\boldsymbol{c}, \quad \delta_{ij} = \int_\chi B_i(\boldsymbol{x})B_j(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \sum_{\alpha=1}^{N} M_\alpha^i M_\alpha^j$$

So $M^T M = I_{K\times K}$, where $I_{K\times K}$ is the $K \times K$ identity matrix. For the given two PDFs $p(\boldsymbol{x})$ and $p'(\boldsymbol{x})$, they have two kinds of coordinates, i.e. $\boldsymbol{c}$ and $\boldsymbol{c}'$, $\boldsymbol{b}$ and $\boldsymbol{b}'$. Then we obtain

$$\boldsymbol{b}^T\boldsymbol{b}' = \boldsymbol{c}^T M^T M \boldsymbol{c}' = \boldsymbol{c}^T\boldsymbol{c}'$$

The geodesic distance does not change under the different bases, i.e. $d(p(\boldsymbol{x}), p'(\boldsymbol{x})) = \arccos(\int_\chi \sqrt{p(\boldsymbol{x})p'(\boldsymbol{x})}\mathrm{d}\boldsymbol{x}) = \arccos(\boldsymbol{c}^T\boldsymbol{c}') = \arccos(\boldsymbol{b}^T\boldsymbol{b}')$. Considering $\boldsymbol{v_b} = M\boldsymbol{v_c}$ which is a tangent vector at $\boldsymbol{b}$ towards $\boldsymbol{b}'$, for the coordinate of $\boldsymbol{b}$ we have the new forms for geodesic, exponential map and logarithmic map as

$$\text{Geodesic: } \gamma(t) : p(\boldsymbol{x}|\boldsymbol{b}(t)), \quad \text{where } \boldsymbol{b}(t) = Exp_{\boldsymbol{b}}(tLog_{\boldsymbol{b}}(\boldsymbol{b}'))$$

$$\text{Exp: } Exp_{\boldsymbol{b}}(\boldsymbol{v_b}) = M\boldsymbol{c}\cos\varphi + \frac{M\boldsymbol{v_c}}{\|M\boldsymbol{v_c}\|}\sin\varphi = M\boldsymbol{c}' = \boldsymbol{b}', \quad \text{where } \varphi = \|\boldsymbol{v_b}\| = \|\boldsymbol{v_c}\|$$

$$\text{Log: } Log_{\boldsymbol{b}}(\boldsymbol{b}') = \frac{M\boldsymbol{c}' - M\boldsymbol{c}\cos\varphi}{\|M\boldsymbol{c}' - M\boldsymbol{c}\cos\varphi\|}\varphi = M\boldsymbol{v_c} = \boldsymbol{v_b}, \quad \text{where } \varphi = \arccos(\boldsymbol{b}^T\boldsymbol{b}') = \arccos(\boldsymbol{c}^T\boldsymbol{c}')$$

□

So different orthonormal basis obtain equivalent Riemannian framework. The final results of the following Riemannian operations are numerically similar and theoretically the same if the representation error is negligible, i.e. $\psi(\boldsymbol{x})$ is in both spaces spanned by these two bases. That is why here we consider the general formulation using orthonormal basis in Eq. (6.1), instead of the specific formulation using piecewise constant basis in [Goh et al., 2009a, 2011].

6.1.3 Orthonormal Basis Chosen for ODFs and EAPs

As we have shown, different orthonormal basis obtains equivalent Riemannian framework. However choosing a good orthonormal basis to sparsely represent ODFs/EAPs can largely reduce the approximation error and accelerate the following Riemannian operations.

For ODFs defined in the unite sphere of **R**-space, i.e. $\chi = \mathbb{S}^2$, we choose the orthonormal basis as the Spherical Harmonics (SHs) which has been used in many works on ODF estimation [Tuch, 2004; Hess et al., 2006; Descoteaux et al., 2007; Canales-Rodrıguez et al., 2009; Aganj et al., 2010b; Tristán-Vega et al., 2010]. Eq. (2.27) is the complex SH of order l and degree m. We use its real version with even order in Eq. (2.28) for real antipodally symmetric ODFs/EAPs.

For EAPs defined in **R**-space, i.e. $\chi = \mathbb{R}^3$, we have chosen the orthonormal basis as Fourier dual Spherical Polar Fourier (dSPF) basis in Chapter 5 and [Cheng et al., 2011c]. SPFI actually provides two orthonormal bases, i.e. SPF basis denoted by $\{B_{nlm}\}$ and dSPF basis denoted by $\{D_{nlm}\}$ [Assemlal et al., 2009a; Cheng et al., 2010a,b; Merlet et al., 2011]. Recall the notation $\mathbf{q} = q\mathbf{u}$ in **q**-space and $\mathbf{R} = R\mathbf{r}$ in **R**-space. These two bases are related by Fourier transform in Eq. (6.13), where $L_n^{1/2}(\cdot)$ is generalized Laguerre polynomial shown in Eq. (2.63), ${}_1F_1(\cdot;\cdot;\cdot)$ is confluent hypergeometric function shown in Eq. (2.71), $\Gamma(\cdot)$ is the Gamma function shown in 2.75, and ζ is the given scale value. If EAP $P(\mathbf{R})$ in **R**-space is represented by one basis, DWI signal $E(\mathbf{q})$ in **q**-space is analytically represented by the other basis. In this chapter, we choose SPF basis

for $P(\mathbf{R})$ in Eq. (6.10) because it is convenient for nonnegative definite EAP/ODF estimation, which will be discussed later in 6.2.1 in details. In Eq. (6.10), the coordinate is $\boldsymbol{c} = (c_{000}, ..., c_{NLL})^T$. Another feasible option is to use Simple Harmonic Oscillator (SHO) basis [Özarslan et al., 2009; Cheng et al., 2011b].

$$P(R\mathbf{r}) = (\psi(R\mathbf{r}))^2 = \left(\sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} c_{nlm}B_{nlm}(\mathbf{R})\right)^2 \tag{6.10}$$

$$B_{nlm}(\mathbf{R}) = R_n(R)Y_l^m(\mathbf{r}) \tag{6.11}$$

$$R_n(R) = \kappa_n(\zeta)\exp\left(-\frac{R^2}{2\zeta}\right)L_n^{1/2}(\frac{R^2}{\zeta}), \quad \kappa_n(\zeta) = \left[\frac{2}{\zeta^{3/2}}\frac{n!}{\Gamma(n+3/2)}\right]^{1/2} \tag{6.12}$$

$$D_{nlm}(\mathbf{q}) = \int_{\mathbb{R}^3} B_{nlm}(\mathbf{R})e^{-i\mathbf{q}^T\mathbf{R}}\mathrm{d}\mathbf{R} = F_{nl}(q)Y_l^m(\mathbf{u}) \tag{6.13}$$

$$F_{nl}(q) = \frac{\zeta^{0.5l+1.5}\pi^{l+1.5}q^l\kappa_n(\zeta)}{\Gamma(l+1.5)}\sum_{i=0}^{n}\binom{n+0.5}{n-i}\frac{(-1)^{i+l/2}}{i!}2^{0.5l+i+1.5}$$
$$\Gamma(0.5l+i+1.5)_1F_1(\frac{2i+l+3}{2}; l+\frac{3}{2}; -2\pi^2q^2\zeta) \tag{6.14}$$

Please note that piecewise constant basis used in [Goh et al., 2009a, 2011] is not appropriate since it needs many samples to represent ODFs/EAPs that are not piecewise constant. It is also worth mentioning that [Goh et al., 2009a, 2011] do not consider the area $\{S_i\}$ for spherical bins. [Goh et al., 2009a, 2011] use samples of the wavefunction $\{\psi_i = \frac{c_i}{\sqrt{S_i}}\}$, not the coefficient $\{c_i\}$ in Eq. (6.5). However for two PDFs $p^{(1)}(\boldsymbol{x})$ and $p^{(2)}(\boldsymbol{x})$, we have

$$d(p^{(1)}(\boldsymbol{x}), p^{(2)}(\boldsymbol{x})) = \arccos(\int_{\chi}\psi^{(1)}(\boldsymbol{x})\psi^{(2)}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}) \simeq \arccos(\sum_i \psi_i^{(1)}\psi_i^{(2)}S_i) \neq \arccos(\sum_i \psi_i^{(1)}\psi_i^{(2)})$$

So the formulation in [Goh et al., 2009a, 2011] needs a ***dense uniform sampling*** on χ.

- The sampling needs to be ***dense*** enough such that the integration can be approximated by a weighted summation.
- The sampling needs to be ***uniform*** with the same area S_i for each bin, such that the weights are the same and the constant weight can be ignored in the Riemannian framework.

The dense uniform sampling may be possible for ODFs, but impossible for EAPs defined in unbounded $\mathbb{R}^3$. Compared to [Goh et al., 2009a, 2011], we do not need a specific sampling because when $\psi^{(1)}(\boldsymbol{x})$ and $\psi^{(2)}(\boldsymbol{x})$ can be represented by linear combination of some orthonormal basis with coefficient $\boldsymbol{c}^{(1)}$ and $\boldsymbol{c}^{(2)}$, we have

$$d(p^{(1)}(\boldsymbol{x}), p^{(2)}(\boldsymbol{x})) = \arccos(\int_{\chi}\psi^{(1)}(\boldsymbol{x})\psi^{(2)}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}) = \arccos(\sum_i c_i^{(1)}c_i^{(2)})$$

For ODFs, [Goh et al., 2009a, 2011] use 162 discrete evenly distributed samples to represent the ODF, which is still a tradeoff between computational overhead and

accuracy. We use SH basis with order 4, 6 or 8 (15, 28 or 45 coefficients) which was shown to be enough for ODF representation [Descoteaux et al., 2007]. For EAPs, [Goh et al., 2009a, 2011] may need thousands of bins in $\mathbb{R}^3$ to approximate EAPs, while SPF, dSPF or SHO bases only need dozens of coefficients to represent EAPs. Besides orthonormal basis representation has been widely used in dMRI community, our formulation using orthonormal basis can take advantage of previous works like QBI, SPFI and SHORE.

6.1.4 Properties of Parameter Space

Based on PF_K in Eq. (6.1), the Parameter Space PS_k is a subset of $\mathbb{S}^{K-1}$. If we choose the specific piecewise constant basis defined from disjoint subsets, PS_K in Eq. (6.5) is the positive orthant of the unit sphere $\mathbb{S}^{K-1}$ as discussed in Section 6.1.1. Here we explore the properties of general PS_K in Eq. (6.1) without assuming a specific basis.

Property-1: PS_K ***is closed***. Eq. (6.1) shows the boundary of PS_K is $\{\boldsymbol{c} \mid \sum_{i=1}^{K} c_i^2 = 1, \exists C \subset \chi,\ s.t.\ \forall \boldsymbol{x} \in C,\ \sum_{i=1}^{K} c_i B_i(x) = 0,\ \forall \boldsymbol{x} \notin C,\ \sum_{i=1}^{K} c_i B_i(\boldsymbol{x}) > 0\}$. Therefore, PS_K is closed by definition. Please note that the boundary of PS_K in coordinate $\boldsymbol{c}$ is dependent on the chosen basis.

Property-2: PS_K ***is a geodesically convex subset of sphere*** $\mathbb{S}^{K-1}$**.** For given $\boldsymbol{c}$, $\boldsymbol{c}' \in PS_K$, let $\boldsymbol{c}(t)$ be the geodesic between them, $t \in [0,1]$. Then $\forall t \in [0,1]$, $\exists \lambda \in [0,1]$, s.t. $\boldsymbol{c}(t) = \frac{\lambda \boldsymbol{c} + (1-\lambda)\boldsymbol{c}'}{\|\lambda \boldsymbol{c} + (1-\lambda)\boldsymbol{c}'\|}$, because $\boldsymbol{c}(t)$ is on the great circle (geodesic) and $\lambda \boldsymbol{c} + (1-\lambda)\boldsymbol{c}'$ is on the chord between $\boldsymbol{c}$ and $\boldsymbol{c}'$. So we have

$$p(\boldsymbol{x}|\boldsymbol{c}(t)) = \sum_{i=1}^{K} c_i(t) B(\boldsymbol{x}) = \frac{\lambda \sqrt{p}(\boldsymbol{x}|\boldsymbol{c}) + (1-\lambda) p(\boldsymbol{x}|\boldsymbol{c}')}{\|\lambda \boldsymbol{c} + (1-\lambda)\boldsymbol{c}'\|} \geq 0$$

which means PS_K is geodesically convex.

Property-3: PS_K ***is contained in a convex cone with angle*** $90°$**.** $\forall \boldsymbol{c}, \boldsymbol{c}' \in PS_K$, we have

$$\int_{\chi} \sqrt{p(\boldsymbol{x}|\boldsymbol{c})}\sqrt{p(\boldsymbol{x}|\boldsymbol{c}')} \mathrm{d}\boldsymbol{x} = \sum_{i=1}^{K} c_i c_i' = \cos(\angle(\boldsymbol{c}, \boldsymbol{c}')) \in [0,1]$$

where $\angle(\boldsymbol{c}, \boldsymbol{c}')$ is the angle between $\boldsymbol{c}$ and $\boldsymbol{c}'$. That means the angle between any two points in PS_K is no more than 90^o. See Fig. 6.1(C).

Property-4: If χ ***is bounded and*** $p(\boldsymbol{x}) < 1, \forall \boldsymbol{x} \in \chi$**, the projection of any** $\boldsymbol{c} \in PS_K$ ***on the coordinate*** $\boldsymbol{u}$ ***of the uniform PDF should be more than*** $\frac{1}{\sqrt{|\chi|}}$**, i.e.** $\boldsymbol{c}^T \boldsymbol{u} > \frac{1}{\sqrt{|\chi|}}$**, where** $|\chi| = \int_{\chi} \mathrm{d}\boldsymbol{x}$ ***is the volume of*** χ**.** If χ is bounded, the uniform PDF denoted by $U(\boldsymbol{x})$ is $\frac{1}{|\chi|}$, and $\sqrt{U}(\boldsymbol{x}) = \frac{1}{\sqrt{|\chi|}} = \sum_{i=1}^{K} u_i B_i(\boldsymbol{x})$, then $\forall \boldsymbol{c} \in PS_K$, if $p(\boldsymbol{x}) < 1$, we have

$$\boldsymbol{c}^T \boldsymbol{u} = \cos(\boldsymbol{c}, \boldsymbol{u}) = \int_{\chi} \sqrt{p(\boldsymbol{x}|\boldsymbol{c})} \frac{1}{\sqrt{|\chi|}} \mathrm{d}\boldsymbol{x} > \frac{1}{\sqrt{|\chi|}} \int_{\chi} p(\boldsymbol{x}|\boldsymbol{c}) \mathrm{d}\boldsymbol{x} = \frac{1}{\sqrt{|\chi|}}$$

That means the projection of $\boldsymbol{c}$ on $\boldsymbol{u}$ should be more than $\frac{1}{\sqrt{|\chi|}}$ if $p(\boldsymbol{x}) < 1$. See Fig. 6.1(C). Please note that although $\int_{\chi} p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = 1$, $p(\boldsymbol{x})$ may be more than 1 in some areas since $p(\boldsymbol{x})$ is a continuous PDF defined in χ.

Please note two differences between ODFs and EAPs for Property-4.

- The first one is about the uniform distribution. For ODFs, $\chi = \mathbb{S}^2$ is a bounded space and $|\chi| = 4\pi$. The **uniform distribution** in PS is $U(\mathbf{r}) = \frac{1}{4\pi}$, which is also the unique **isotropic ODF**. When SH basis is used, the coordinate $\boldsymbol{u} = (1, 0, ..., 0)^T$, considering $Y_0^0(\mathbf{r}) = \frac{1}{\sqrt{4\pi}}$. For EAPs, $\chi = \mathbb{R}^3$ is a unbounded space. Thus the uniform distribution $U(\mathbf{R})$ actually does not exist. Of course we can choose a support big enough to define $U(\mathbf{R})$ only in the support, However this kind of uniform distribution is much far from normal EAPs in brain.
- The second difference is about the probability values. Normally the typical ODFs are always less than 1. So the projection of $\boldsymbol{c}$ on u is less than $\frac{1}{\sqrt{4\pi}}$. However, EAPs are normally much larger than 1 in the area $R < 30\mu m$ that we focus, because the diffusion time τ is very small.

6.1.5 Geodesic Anisotropy, Rényi Entropy

Geodesic Anisotropy (GA) [1] [Batchelor et al., 2005; Fletcher, 2004] based on Riemannian metric for tensors is an alternative anisotropy measurement to the well known Fractional Anisotropy (FA). GA for tensors is defined as the Riemannian distance from the given tensor to its nearest isotropic tensor [Fletcher, 2004; Batchelor et al., 2005]. Please note that the isotropic tensor is not unique because every tensor whose eigenvalues are the same is isotropic. Analogously we define GA for ODFs/EAPs as the Riemannian distance from the given ODF/EAP with the coordinate $\boldsymbol{c}$ to its nearest isotropic ODF/EAP with the coordinate $\boldsymbol{s}_{\boldsymbol{c}}$, i.e. $GA(\boldsymbol{c}) = \arccos(\boldsymbol{c}^T \boldsymbol{s}_{\boldsymbol{c}})$. Please note an important difference between ODFs and EAPs. The isotropic ODF is the uniform ODF uniquely defined by $\psi(\mathbf{r}) = \frac{1}{4\pi}$ and it has the coordinate $\boldsymbol{s}_{\boldsymbol{c}} = \boldsymbol{u} = (1, 0, ..., 0)^T$ under SH basis. $\cos(GA(\boldsymbol{c})) = \boldsymbol{c}^T \boldsymbol{u}$ is the projection of $\boldsymbol{c}$ onto $\boldsymbol{u}$ in above Property-4. However, the uniform EAP has no definition, and the isotropic EAP is not unique because $P(R\mathbf{r})$ is isotropic if $P(R\mathbf{r}) = F(R)$, $\forall \mathbf{r} \in \mathbb{S}^2$. The following theoretical result can be used to obtain analytic closed form of the isotropic EAP.

Theorem 6.2 (Wavefunction of the Nearest Isotropic EAP). *For any EAP $P(\mathbf{R}) = \psi(\mathbf{R})^2$, its nearest isotropic EAP is $P_{iso}(\mathbf{R}) = (\psi_{S_{iso}}(\mathbf{R}))^2$, where $\psi_{S_{iso}}(\mathbf{R})$ is the normalized version of the isotropic part of $\psi(\mathbf{R})$, i.e.*

$$\psi_{S_{iso}}(\mathbf{R}) = \frac{\psi_{L_{iso}}(\mathbf{R})}{\|\psi_{L_{iso}}(\mathbf{R})\|}, \quad \psi_{L_{iso}}(\mathbf{R}) = \left(\int_{\mathbb{S}^2} \psi(\mathbf{R}) Y_0^0(\mathbf{r}) d\mathbf{r} \right) Y_0^0(\mathbf{r}) \tag{6.15}$$

and $\|\psi_{L_{iso}}(\mathbf{R})\| = \sqrt{\int_{\mathbb{R}^3} \psi_{L_{iso}}(\mathbf{R})^2 d\mathbf{R}}$ is the standard ℓ_2 norm in function space $\mathbb{L}^2(\mathbb{R}^3, 1)$.

See Appendix 6.5 for the proof which is independent of the basis. When SPF basis is used, the isotropic part of the coordinate of $P(\mathbf{R}|\{c_{nlm}\})$ has analytical form as $\{c_{nlm}\delta_{lm}^{00}\}$, then the nearest EAP has the coordinate as

$$s_{\boldsymbol{c},nlm} = \frac{c_{nlm}\delta_{lm}^{00}}{\sqrt{\sum_{n=0}^{N} c_{n00}^2}} \tag{6.16}$$

[1] We call GA as Geometric Anisotropy in [Cheng et al., 2009a]. Here we follow the original name in [Batchelor et al., 2005; Fletcher, 2004].

GA for EAPs is the Riemannian counterpart of GFA for EAPs defined based on Euclidean distance in Eq. (5.27). Please note that GA based on SPF basis is rotational invariant, because of the rotation property of Spherical Harmonics.

Rényi entropy with order $\frac{1}{2}$ is defined as [A. Rényi, 1960]

$$H_{\frac{1}{2}}(\boldsymbol{c}) = 2\log\left(\int_{\chi}\sqrt{p(\boldsymbol{x}|\boldsymbol{c})}\mathrm{d}\boldsymbol{x}\right) \tag{6.17}$$

If χ is bounded, the uniform distribution exists and it is $U(\boldsymbol{x}) = \frac{1}{|\chi|}$, we have

$$H_{\frac{1}{2}}(\boldsymbol{c}) = \log(|\chi|(\boldsymbol{c}^T\boldsymbol{u})^2) = 2\log(\cos(GA(\boldsymbol{c}))) + \log(|\chi|) \tag{6.18}$$

For ODFs, $H_{\frac{1}{2}}(\boldsymbol{c}) = \log(4\pi(\boldsymbol{c}^T\boldsymbol{u})^2) \in (0, \log(4\pi)]$, because normally $\boldsymbol{c}^T\boldsymbol{u} > \frac{1}{\sqrt{4\pi}}$ as we have discussed in Property-4 in Section 6.1.4. However, the experiment in our previous paper [Cheng et al., 2009a] showed that the Rényi entropy for ODFs do not have good contrast, which actually demonstrates again that ODFs we focus in practice are only in a small space inside general PDF function space shown in 3.35. For EAPs, the uniform EAP has no definition, but by substituting Eq. (6.10) into the definition of $H_{\frac{1}{2}}$ in Eq. (6.17) and considering the result in Eq. (5.22) or directly Eq. (2.83), we can still obtain the analytical closed form as

$$\begin{aligned} H_{\frac{1}{2}}(\boldsymbol{c}) &= 2\log\left(\int_{\mathbb{R}^3}\sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l}c_{nlm}R_n(R)Y_l^m(\mathbf{r})\mathrm{d}\mathbf{R}\right) \\ &= 2\log\left(4\sqrt{\pi}\zeta^{\frac{3}{4}}\sum_{n=0}^{N}(-1)^n\sqrt{\frac{\Gamma(n+1.5)}{n!}}c_{n00}\right) \end{aligned} \tag{6.19}$$

However this analytical form has nothing related with GA, and $H_{\frac{1}{2}}$ for EAPs is always negative in practice because $P(\mathbf{R}|\boldsymbol{c})$ is much more than 1. For example, let's consider Gaussian distribution in $\mathbb{R}^1$, i.e. $N(0,\sigma)$. Its Rényi entropy is

$$H_{\frac{1}{2}} = 2\log\left(\int_{-\infty}^{\infty}\sqrt{N(x|0,\sigma)}\mathrm{d}x\right) = 2\log\left(2^{\frac{3}{4}}\pi^{\frac{1}{4}}\sqrt{\sigma}\right)$$

Thus if $\sigma < 2^{\frac{-3}{2}}\pi^{\frac{-1}{2}}$, $H_{\frac{1}{2}} < 0$.

So considering the drawbacks of Rényi entropy, it seems that GA is more useful than Rényi entropy for both ODFs and EAPs.

6.1.6 Log-Euclidean Framework and Affine-Euclidean Framework

Log-Euclidean framework for tensors is proposed to work in Euclidean space by projecting every tensor onto the tangent space of a fixed isotropic tensor (identity matrix) using logarithmic map [Arsigny et al., 2006]. It is an approximation of the Riemannian framework for tensors. Analogously, we can fix an isotropic ODF/EAP with coordinate $\boldsymbol{s}$ and project all ODFs/EAPs onto its tangent space using a given diffeomorphism, i.e.

$$F:\ PS_K \subset \mathbb{S}^{K-1} \rightarrow F(PS_K) \subset T_{\boldsymbol{s}}$$

Then the metric, exponential map and logarithmic map for PS can be induced from the Euclidean space by F, so that we can work in an Euclidean space, which will approximate the above Riemannian framework. Recall in Section 3.1 the pushforward and pullback maps denoted by F_* and F^* respectively, then

$$\text{Distance}: \ d_{PS_K}(p(\cdot|\boldsymbol{c}), p(\cdot|\boldsymbol{c}')) = d_{Euc}(F(\boldsymbol{c}), F(\boldsymbol{c}')) = \|F(\boldsymbol{c}) - F(\boldsymbol{c}')\| \tag{6.20}$$

$$\text{Geodesic}: \ Exp_{\boldsymbol{c}}(t\boldsymbol{v}_{\boldsymbol{c}}) = F^{-1}(F(\boldsymbol{c}) + tF_*\boldsymbol{v}_{\boldsymbol{c}}), \quad t \in [0,1] \tag{6.21}$$

$$\text{Exp}: \ Exp_{\boldsymbol{c}}(\boldsymbol{v}_{\boldsymbol{c}}) = \boldsymbol{c}' = F^{-1}(F(\boldsymbol{c}) + F_*\boldsymbol{v}_{\boldsymbol{c}}) \tag{6.22}$$

$$\text{Log}: \ Log_{\boldsymbol{c}}(\boldsymbol{c}') = \boldsymbol{v}_{\boldsymbol{c}} = F^*(F(\boldsymbol{c}') - F(\boldsymbol{c})) \tag{6.23}$$

If $F(\boldsymbol{c}) = Log_{\boldsymbol{s}}(\boldsymbol{c})$, then we have the **Log-Euclidean framework** like the one for tensors. If $F(\boldsymbol{c}) = \frac{\boldsymbol{c}}{\boldsymbol{c}^T\boldsymbol{s}} - \boldsymbol{s}$, we have the **Affine-Euclidean framework**, which is very popular in Statistical Shape Analysis [Dryden and Mardia, 1998; Fletcher, 2004]. See Fig. 6.1(C) for a sketch map.

For ODFs, $\{s_{lm} = \delta^{00}_{lm}\}$ represents the unique isotropic ODF when SH is used. For EAPs, we choose $\{s_{nlm} = \delta^{000}_{nlm}\}$ since it represents a typical isotropic Gaussian distribution based on Eq. (6.10). Thus the Log-Euclidean framework and Affine-Euclidean framework can be obtained analytically based on SH and SPF basis representation for ODFs and EAPs.

6.1.7 Weighted Mean, Weighted Median and Principal Geodesic Analysis (PGA)

Given N PDFs $\{p(\boldsymbol{x}|\boldsymbol{c}^{(i)})\}_{i=1}^N$ with the coordinates $\{\boldsymbol{c}^{(i)}\}_{i=1}^N$ in PS_K and the weight vector $\boldsymbol{w} = (w_1, w_2, ..., w_N)^T$ with $\sum_{i=1}^N w_i = 1$, $w_i \in [0,1]$, the **weighted Riemannian mean (Fréchet mean)** $\boldsymbol{\mu}_{\boldsymbol{w}}$ is defined as the minimizer of the weighted sum of squared distances [Fréchet, 1948; Buss and Fillmore, 2001; Karcher, 1977; Kendall, 1990; Batchelor et al., 2005; Pennec et al., 2006; Lenglet et al., 2006b; Fletcher, 2004; Cheng et al., 2009a]. The **weighted Riemannian median** $\boldsymbol{m}_{\boldsymbol{w}}$ is defined as the minimizer of the weighted sum of distances [Fletcher et al., 2009; Cheng et al., 2010c]. They are generalized versions of Euclidean mean and median. Please see Eq. (6.24) or see Section 3.2 for more details. Riemannian median was shown to be more robust than Riemannian mean, which makes it more appropriate for atlas estimation [Fletcher et al., 2009].

$$\boldsymbol{\mu}_{\boldsymbol{w}} \overset{\text{def}}{=} \arg\min_{\boldsymbol{c} \in PS_K} \sum_{i=1}^N w_i d(\boldsymbol{c}, \boldsymbol{c}^{(i)})^2 \qquad \boldsymbol{m}_{\boldsymbol{w}} \overset{\text{def}}{=} \arg\min_{\boldsymbol{c} \in PS_K} \sum_{i=1}^N w_i d(\boldsymbol{c}, \boldsymbol{c}^{(i)}) \tag{6.24}$$

For Log-Euclidean and Affine-Euclidean frameworks, the mean has closed form as

$$\boldsymbol{\mu}_{\boldsymbol{w}} = F^{-1}\left(\sum_{i=1}^N w_i F(\boldsymbol{c}^{(i)})\right) \tag{6.25}$$

The following theorems show the existence, uniqueness of weighted Riemannian mean and median and show that the probability values of weighted mean and median are not less than the N PDFs used in the mean and median estimation. Please see Appendix 6.6 for the proofs.

Algorithm 6.1: Weighted Riemannian Mean

Input: $\boldsymbol{c}^{(1)}, ..., \boldsymbol{c}^{(N)} \in PS_K$, $\boldsymbol{w} = (w_1, ..., w_N)^T$
Output: the *Weighted Mean* $\boldsymbol{\mu}_{\boldsymbol{w}}$.
begin
  initialization: $\boldsymbol{\mu}_{\boldsymbol{w}} \leftarrow \frac{\sum_{i=1}^N w_i \boldsymbol{c}^{(i)}}{\|\sum_{i=1}^N w_i \boldsymbol{c}^{(i)}\|}$; // normalized Euclidean mean
  repeat
    $\boldsymbol{v}_{\boldsymbol{\mu}_{\boldsymbol{w}}} \leftarrow \sum_{i=1}^N w_i Log_{\boldsymbol{\mu}_{\boldsymbol{w}}}(\boldsymbol{c}^{(i)})$;
    $\boldsymbol{\mu}_{\boldsymbol{w}} \leftarrow Exp_{\boldsymbol{\mu}_{\boldsymbol{w}}}(\boldsymbol{v}_{\boldsymbol{\mu}_{\boldsymbol{w}}})$; // step size is 1
  until $\|\boldsymbol{v}_{\boldsymbol{\mu}_w}\| > \varepsilon$;
end

Theorem 6.3 (Existence and Uniqueness of Riemannian mean and median). *Weighted Riemannian mean and weighted Riemannian median uniquely exist in PS_K, and they can be efficiently estimated from Algorithm 6.1 and Algorithm 6.2.*

Theorem 6.4 (Reasonable Riemannian mean and median of PDFs). *Weighted Riemannian mean $p(\boldsymbol{x}|\boldsymbol{\mu}_{\boldsymbol{w}})$ and median $p(\boldsymbol{x}|\boldsymbol{m}_{\boldsymbol{w}})$ in Eq. (6.24) satisfies $p(\boldsymbol{x}|\boldsymbol{\mu}_{\boldsymbol{w}}) \geq \min\{p(\boldsymbol{x}|\boldsymbol{c}^{(i)})\}_{i=1}^N$ and $p(\boldsymbol{x}|\boldsymbol{m}_{\boldsymbol{w}}) \geq \min\{p(\boldsymbol{x}|\boldsymbol{c}^{(i)})\}_{i=1}^N$, $\forall \boldsymbol{x} \in \chi$.*

Please note that the weighted Riemannian mean/median may be greater than the PDFs in some measurable set of χ. See Lemma 6.3 in Appendix 6.6 for the theoretical analysis on the maximum of the projection. Let's take PF in 3D space in Eq. (3.53) as an example, the Riemannian mean of two discrete PDFs $\boldsymbol{p}^{(1)} = (\frac{1}{2}, \frac{1}{2}, 0)^T$ and $\boldsymbol{p}^{(2)} = (\frac{1}{2}, 0, \frac{1}{2})^T$ is $(\frac{2}{3}, \frac{1}{6}, \frac{1}{6})^T$, while the Euclidean mean is $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})^T$. See Fig. 6.1(A,B,H) for the sketch map. Riemannian mean considers the probability vector as a whole unity, while Euclidean mean considers the each dimension independently.

We think the Riemannian mean is more reasonable. Let's imagine we are guessing the color of a ball which may be red, yellow or blue. $\boldsymbol{p} = (p_1, p_2, p_3)^T$ is the probability for the three colors. Two consultants gave us two different forecasts $\boldsymbol{p}^{(1)}$ and $\boldsymbol{p}^{(2)}$. These two consultants reach agreement on red which is more trustable, but they have very different opinions on yellow and blue. Considering the error of forecast may be larger in the other two colors, it is reasonable to give more weighting to red color. Another example in Example 3.10 is that the mean Gaussian PDF of two 1D Gaussian PDFs with different expectations and the same variance as $N(-1, 1)$ and $N(1, 1)$ is $N(0, 1)$ based on Mahalanobis distance [Atkinson and Mitchell, 1981], which has larger probability value around origin point. This is different from the Gaussian distributions with the same expectation but different covariances. See Theorem 6.6 in Section 6.2.2 on tensor interpolation for the details. Furthermore the experiments in [Mio et al., 2005; Read, 1999] also showed larger values in weighted mean of PDFs are reasonable, although [Mio et al., 2005] considered a different computational framework in information geometry theory and [Read, 1999] proposed a total different method from information geometry theory.

Principal Geodesic Analysis (PGA) was proposed to generalize the PCA from Euclidean space to Riemannian manifold [Fletcher et al., 2004]. After obtaining the Riemannian mean μ, PGA can find some principal components by performing eigen-decomposition of covariance matrix in tangent space. See Section 3.2.2 for the details

Algorithm 6.2: Weighted Riemannian Median

Input: $\boldsymbol{c}^{(1)}, ..., \boldsymbol{c}^{(N)} \in PS_K$, $\boldsymbol{w} = (w_1, ..., w_N)^T$
Output: the *Weighted Median* $\boldsymbol{m_w}$.
begin
  initialization: $\boldsymbol{m_w} \leftarrow \frac{\sum_{i=1}^{N} w_i \boldsymbol{c}^{(i)}}{\|\sum_{i=1}^{N} w_i \boldsymbol{c}^{(i)}\|}$; `// normalized Euclidean mean`
  repeat
    $\boldsymbol{v}_{\boldsymbol{m_w}} \leftarrow \sum_{i=1}^{N} \frac{w_i / d(\boldsymbol{m_w}, \boldsymbol{c}^{(i)})}{\sum_{j=1}^{N} w_j / d(\boldsymbol{m_w}, \boldsymbol{c}^{(j)})} Log_{\boldsymbol{m_w}}(\boldsymbol{c}^{(i)})$;
    $\boldsymbol{m_w} \leftarrow Exp_{\boldsymbol{m_w}}(\boldsymbol{v}_{\boldsymbol{m_w}})$; `// step size is 1`
  until $\|\boldsymbol{v}_{\boldsymbol{m_w}}\| > \varepsilon$;
end

on PGA. PGA has been used for shape analysis [Fletcher et al., 2004], tensor processing [Fletcher and Joshi, 2007] and ODF processing [Goh et al., 2009a, 2011]. However, [Goh et al., 2009a, 2011] performed PGA in a high dimensional space based on piecewise constant basis. By considering a better basis like SH or SPF basis, PGA can be done efficiently in a low dimensional space and obtain the same final results as we have shown in Theorem 6.1.

6.1.8 Diffeomorphism Invariance

The Riemannian metric for tensors is affine-invariant. The proposed Riemannian metric and the metric for tensors are both Fisher information metric, so it is possible to explore the connection between them.

Theorem 6.5 (Diffeomorphism Invariant Metric). *Fisher information metric is diffeomorphism invariant by definition.*

Proof. Let $p_{\boldsymbol{x}}(\boldsymbol{x}|\boldsymbol{c})$ denote a PDF on domain χ, $h : \chi \mapsto \chi$ is a diffeomorphism. $p_{\boldsymbol{y}}(\boldsymbol{y}|\boldsymbol{c})$ is the PDF under h, i.e. $y = h(x)$. Then

$$p_{\boldsymbol{y}}(\boldsymbol{y}) = |\nabla h^{-1}(\boldsymbol{y})| p_{\boldsymbol{x}}(h^{-1}(\boldsymbol{y}))$$

By considering $\mathrm{d}\boldsymbol{y} = |\nabla h(\boldsymbol{y})| \mathrm{d}\boldsymbol{x}$, we have

$$g_{ij} = \int_{\chi} \frac{\partial \sqrt{p_{\boldsymbol{y}}(\boldsymbol{y}|\boldsymbol{c})}}{\partial c_i} \frac{\partial \sqrt{p_{\boldsymbol{y}}(\boldsymbol{y}|\boldsymbol{c})}}{\partial c_j} \mathrm{d}\boldsymbol{y} = \int_{\chi} \frac{\partial \sqrt{p_{\boldsymbol{x}}(\boldsymbol{x}|\boldsymbol{c})}}{\partial c_i} \frac{\partial \sqrt{p_{\boldsymbol{x}}(\boldsymbol{x}|\boldsymbol{c})}}{\partial c_j} \mathrm{d}\boldsymbol{x}$$

which proves the metric g_{ij} is diffeomorphism invariant. The Riemannian distance between any given two PDFs $p_{\boldsymbol{x}}(\boldsymbol{x}|\boldsymbol{c})$ and $p_{\boldsymbol{x}}(\boldsymbol{x}|\boldsymbol{c}')$ is also diffeomorphism invariant.

$$\begin{aligned} d(p_{\boldsymbol{y}}(\boldsymbol{y}|\boldsymbol{c}), p_{\boldsymbol{y}}(\boldsymbol{y}|\boldsymbol{c}')) &= \arccos\left(\int_{\chi} \sqrt{p_{\boldsymbol{y}}(\boldsymbol{y}|\boldsymbol{c}) p_{\boldsymbol{y}}(\boldsymbol{y}|\boldsymbol{c}')} \mathrm{d}\boldsymbol{y}\right) \\ &= \arccos\left(\int_{\chi} \sqrt{p_{\boldsymbol{x}}(\boldsymbol{x}|\boldsymbol{c}) p_{\boldsymbol{x}}(\boldsymbol{x}|\boldsymbol{c}')} \mathrm{d}\boldsymbol{x}\right) = d(p_{\boldsymbol{x}}(\boldsymbol{x}|\boldsymbol{c}), p_{\boldsymbol{x}}(\boldsymbol{x}|\boldsymbol{c}')) \end{aligned}$$

□

This proof does not need any orthonormal basis representation. If $(p_{\boldsymbol{x}}(\boldsymbol{x}|\boldsymbol{c}) = (\sum_i c_i B_i(\boldsymbol{x}))^2$ and $\{B_i(\boldsymbol{x})\}$ is an orthonormal basis set, then based on change of variable it can be proved that $\{\sqrt{|\nabla h^{-1}(\boldsymbol{y})|}B_i(h^{-1}(\boldsymbol{y}))\}$ is another orthonormal basis set, and $p_{\boldsymbol{y}}(\boldsymbol{y})$ can be presented under the new basis with the same coefficients $\boldsymbol{c}$ as $p_{\boldsymbol{x}}(\boldsymbol{x})$. Then the distance between two distributions is invariant because the coefficients are invariant. Diffeomorphism invariance may be useful in registration between PDF data.

Moreover, if $p_{\boldsymbol{x}}(\boldsymbol{x}|\Sigma)$ is constrained as a Gaussian distribution parameterized by tensor Σ, $p_{\boldsymbol{y}}(\boldsymbol{y}|\Sigma)$ is still a Gaussian distribution if and only if h is an affine transform, i.e. $h(\boldsymbol{x}) = A\boldsymbol{x}$, A is a nonsingular matrix. So the proposed diffeomorphism invariant metric reduces to previous Affine-invariant metric if constraining the PDFs as Gaussian distributions. In this sense, the proposed diffeomorphism invariant metric for ODFs/EAPs is indeed a ***natural extension*** of previous affine-invariant metric for tensors.

It will be interesting to see the Riemannian distance between Gaussian distributions under the proposed metric. For two Gaussian distributions $N(\boldsymbol{x}|\Sigma_1)$ and $N(\boldsymbol{x}|\Sigma_2)$ where $\boldsymbol{x} \in \mathbb{R}^d$, the affine-invariant distance is $\frac{1}{\sqrt{2}}\|\log(\Sigma_1^{-1}\Sigma_2)\|$. See Example 3.10 for details. The diffeomorphism invariant distance is

$$\begin{aligned} d(N(\boldsymbol{x}|\Sigma_1), N(\boldsymbol{x}|\Sigma_2)) &= \arccos\left(\int_{\mathbb{R}^d} \sqrt{N(\boldsymbol{x}|\Sigma_1)N(\boldsymbol{x}|\Sigma_2)}\mathrm{d}\boldsymbol{x}\right) \\ &= \arccos \frac{|2(\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}|^{\frac{1}{2}}}{|\Sigma_1|^{\frac{1}{4}}|\Sigma_2|^{\frac{1}{4}}} \end{aligned} \tag{6.26}$$

Please note that the distributions on the geodesic between $N(\boldsymbol{x}|\Sigma_1)$ and $N(\boldsymbol{x}|\Sigma_2)$ are no longer Gaussian, which is different from the affine-invariant framework. See the interpolation in 1 dimension in Section 6.3.1.

6.1.9 Theoretical Comparison Between Riemannian Metric and Euclidean Metric

Since ODF and EAP are both PDFs, the convex combination of PDFs are still a PDF. That means the Euclidean metric could be used for ODFs and EAPs, which is analogous to the Euclidean metric for tensors. The Euclidean distance between two distributions $p(\boldsymbol{x})$ and $p'(\boldsymbol{x})$ is

$$\sqrt{\int_\chi (p(\boldsymbol{x}) - p'(\boldsymbol{x}))^2 \mathrm{d}\boldsymbol{x}}$$

Then the Euclidean mean is given as $wp(\boldsymbol{x}) + w'p(\boldsymbol{x})$, where w and w' are weights. More efficiently, we can represent ODF and EAP directly by a linear combination of orthonormal basis [Descoteaux et al., 2007; Aganj et al., 2010b; Cheng et al., 2010a,b; Merlet et al., 2011], then the Euclidean framework can be constructed on the coefficients, which has been used in ODF estimation, interpolation, segmentation, registration etc. [Descoteaux et al., 2008a; Wassermann et al., 2008; Lenglet et al., 2009], just like Euclidean framework for tensors. Based on Euclidean metric, the original PCA can be used directly to analyze ODF/EAP field, instead of PGA based on Riemannian metric. Then what is the difference between the Euclidean framework and Riemannian framework for ODFs and EAPs?

First, we have proved that the proposed Riemannian metric is diffeomorphism invariant, it is easy to show in the same way that Euclidean metric is not. Second we give a visual comparison based on the example in 3D case in Eq. (6.5). In this case, the PS_3 is demonstrated as an equilateral triangle in Fig. 6.1(A) by coordinate $\boldsymbol{p}$, as well as the positive orthant of $\mathbb{S}^2$ by coordinate $\boldsymbol{c}$. The uniform distribution $U(\boldsymbol{x}|\boldsymbol{u})$ could be represented by $\boldsymbol{p} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ in Fig. 6.1(A) and $\boldsymbol{c} = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$ in Fig. 6.1(B), where $\boldsymbol{c}$ is the coordinate in Riemannian framework. For any given point $p_0(\boldsymbol{x})$, we calculate the distances between $p_0(\boldsymbol{x})$ and other points in this triangle and display the isolines of distances in Fig. 6.1(D,F) using Riemannian metric and in Fig. 6.1(E,G) using Euclidean metric. We conclude that the isolines based on Riemannian metric more consider the boundary of PS_k, while the isolines based on Euclidean metric do not. For ODFs, the uniform distribution, i.e. the red point in Fig. 6.1(D,E), is the same as the isotropic distribution. So the isolines of distances in Fig. 6.1(D) are actually isolines of the GA for ODFs. The isolines in Fig. 6.1(E) are the isolines of generalized FA for ODFs [Tuch, 2004]. For EAPs, that is not true because the uniform distribution is not the isotropic EAP as we discussed above. For any given points $p(\boldsymbol{x})$ and $p'(\boldsymbol{x})$, the geodesics based on Riemannian metric and Euclidean metric are shown in Fig. 6.1(H). We conclude that if these two points are close to the uniform distribution, then two geodesics from Riemannian metric and Euclidean metric will be close.

6.2 APPLICATIONS

6.2.1 Riemannian Coordinate Estimation and Nonnegative Definite EAP/ODF estimation

In Section 6.1, we always assume the Riemannian coordinate $\boldsymbol{c}$ for ODFs and EAPs is already known. However, existing works only focus on ODF/EAP estimation and there is no work on how to estimate the wavefunction. Here we propose the following two methods to estimate the wavefunction represented by orthonormal basis.

Wavefunction estimation from given ODFs/EAPs

Assume we obtain the precomputed ODF or EAP denoted by $p(\boldsymbol{x})$ from some methods, e.g. SPFI in Chapter 5 [Cheng et al., 2010a,b] or ℓ_1-SPFI in [Cheng et al., 2011c]. Then we have the continuous expression of wavefunction, $\psi(\boldsymbol{x}) = \sqrt{p(\boldsymbol{x})}$, and the coefficients $\{c_i\}$ are the inner products between $\psi(\boldsymbol{x})$ and $\{B_i(\boldsymbol{x})\}$. What we do is to estimate the inner products by numerical integration. Discrete samples are generated for the numerical integration such that the numerical integration of $\int_\chi B_i(\boldsymbol{x})B_j(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$ is close to δ_{ij} i.e.

$$\delta_{ij} = \int_\chi B_i(\boldsymbol{x})B_j(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \simeq \sum_k B_i(\boldsymbol{x}_k)B_j(\boldsymbol{x}_k)S_k$$

where S_k is the area (or volume) element for sample $\boldsymbol{x}_k$. Our previous works on Riemannian framework [Cheng et al., 2009a, 2011a] considered least square fitting with the generated samples, which is equivalent to numerical inner product if we have

exhaustive samples and incorporate the area element into the least square. In the following we will given some specific details for ODFs and EAPs.

For ODFs, the sampling gradients $\{\mathbf{r}_j\}$ are chosen from icosahedral tessellation with order 2 (or 3) in hemisphere, considering ODF is antipodally symmetric. We use SH basis with order 8 in practice to represent wavefunctions. An important issue for ODFs is that the estimated ODF by Tuch $\Phi_t(\mathbf{r})$ needs to be normalized so that the integration in $\mathbb{S}^2$ is 1 [Tuch, 2004; Descoteaux et al., 2007]. Please note that the normalization is **NOT** to let the summation of ODF samples be 1. See Appendix 4.6 for the reason and the correct ODF normalization. Because many works in HARDI represent ODF using SHs, we use continuous ODF normalization shown in Appendix 4.6. If we can ensure that the estimated ODF is correctly normalized and the estimated ODF is nonnegative, the estimated coefficient $\boldsymbol{c}$ from the least square fitting automatically satisfy $\|\boldsymbol{c}\| = 1$, because

$$\delta_{lm}^{l'm'} = \int Y_l^m(\mathbf{r}) Y_{l'}^{m'}(\mathbf{r}) \simeq \sum_j Y_l^m(\mathbf{r}_j) Y_{l'}^{m'}(\mathbf{r}_j) S_j$$

and

$$1 = \int_{\mathbb{S}^2} \psi(\mathbf{r})^2 \mathrm{d}\mathbf{r} = \int_{\mathbb{S}^2} (\sum_{lm} c_{lm} Y_l^m(\mathbf{r}))^2 \mathrm{d}\mathbf{r} \simeq \sum_{lm} \sum_{l'm'} c_{lm} c_{l'm'} \sum_j Y_l^m(\mathbf{r}_j) Y_{l'}^{m'}(\mathbf{r}_j) S_j = \|\boldsymbol{c}\|$$

Numerically $\|\boldsymbol{c}\|$ is in $[0.99, 1.01]$ in practice.

For EAPs, the inner product between $\psi(\mathbf{R})$ and SPF basis are approximated by numerical integration within $[0, 50]\mu m$ in spherical coordinate system. We use SPF basis with $N = 4$ and $L = 8$ in Eq. (6.11) in practice to represent wavefunctions.

Please note some important issues for EAPs. *First*, although any EAP estimation can be used to estimate the EAP, we need to choose the methods which can ensure that the integration of the EAP is 1, i.e. $E(0) = 1$. Here we use SPFI in Chapter 5 [Cheng et al., 2010b,a] which considers $E(0) = 1$ into estimation process. *Second*, we need to choose an appropriate scale ζ in SPF basis such that the basis can represent the EAP with as limited coefficients as possible. We propose to set ζ based on the typical isotropic Gaussian distribution

$$N(\mathbf{R}|2\tau D_0) = \frac{1}{\sqrt{(4\pi\tau D_0)^3}} \exp(-\frac{R^2}{4\tau D_0})$$

with typical diffusivity (variance) in human brain as

$$D_0 = 0.7 \times 10^{-3} mm/s^2$$

- In Chapter 5 [Cheng et al., 2010a,b, 2011c], we use SPF basis to represent $E(\mathbf{q})$ and set ζ by two ways. One is based on the typical ADC D_0, i.e. $\zeta = \frac{1}{8\pi^2\tau D_0}$, such that SPF basis can sparsely represent $E(\mathbf{q}) = \exp(-4\pi^2\tau q^2 D_0)$ only using $B_{000}(\mathbf{q}) = R_0(q) Y_0^0(\mathbf{u})$. The other one is based on fitting the signal with GHOT model in Eq. (5.36).
- In [Merlet et al., 2011], we use SPF basis to represent $P(\mathbf{R})$ and set ζ such that $N(\mathbf{R}|2\tau D_0)$ can be sparsely represented by only $B_{000}(\mathbf{R}) = R_0(R) Y_0^0(\mathbf{r})$, i.e. $\zeta = 2\tau D_0$.

- In this chapter, we use SPF basis to represent $\sqrt{P(\mathbf{R})}$ and we set ζ such that $\sqrt{N(\mathbf{R}|2\tau D_0)} \propto \exp(-\frac{R^2}{8\tau D_0})$ can be sparsely represented by only $B_{000}(\mathbf{R})$, i.e.

$$\zeta = 4\tau D_0 \tag{6.27}$$

  Note that different scale ζ means different orthonormal basis set $\{R_n(R|\zeta)Y_l^m(\mathbf{r})\}$. Actually we can adaptively set ζ for different voxels when estimating ODFs/EAPs in the subsection 6.2.1. We can adaptively estimate the isotropic tensor with ADC D_{iso} from signal $\{E_i\}_{i=1}^{N_s}$ in each voxel, which can be done by a standard least square estimation in DTI. Then we adaptively set $\zeta = 4\tau D_{iso}$. The isotropic EAP $N(\mathbf{R}|2\tau D_{iso})$ may better approximate the signal samples than the EAP $N(\mathbf{R}|2\tau D_0)$ provided by D_0. In this way ζ is adaptively set for each voxel. However, if we want to work on the estimated coefficients for the following operations on EAP field, we need to force each voxel to share the scale ζ such that the EAP in every voxel is represented by the same basis. If different ζ is used in different voxel, the Fisher information metric in Eq. (6.6) will no longer be $4\delta_{ij}$.

Please note that the results in the Riemannian framework are robust to the parameter ζ, because different ζ results in different orthonormal basis sets and we have proved that the final results of the Riemannian framework are invariant under different orthonormal basis, if the wavefunction can be represented by both basis sets. However, for an inappropriate ζ, we may need a large number of basis functions to represent the wavefunction, which is inefficient for the following Riemannian operations.

Please note that compared to [Goh et al., 2009a, 2011] which directly uses the N_s samples of wavefunction, our method needs an extra numerical inner product (or least square fitting) to obtain the coefficients of the wavefunction. However, the additional computation load can be ignored compared to the computational time it saves in the following Riemannian operation working in lower dimensional space. Moreover, the histogram formulation in [Goh et al., 2009a, 2011] cannot be generalized to EAPs using these N_s multiple shell samples, because these samples contained in different bins have different areas, while [Goh et al., 2009a, 2011] implicitly assumes the same area for the bins as discussed in 6.1.3.

The drawback of our method is that the wavefunction estimation is totally based on the precomputed ODFs/EAPs, which is also a drawback of [Goh et al., 2009a, 2011] If the estimated ODFs/EAPs have some negative values, the square root of these negative values do not exist. In practice, normally ODF values estimated from QBI (or exact QBI) and SPFI are positive in most of time. However, the EAP values estimated from SPFI always has some negative values for large R. In practice, we need to force the negative values as 0 for ODF/EAP samples.

Square Root Parameterized Estimation (SRPE): a nonnegative definite EAP/ODF estimation method

We propose a novel Square Root Parameterized Estimation (SRPE) method to estimate the wavefunction $\psi(\mathbf{R}|\boldsymbol{c})$ of EAP directly from the DWI signals. When $\psi(\mathbf{R})$ is estimated, the EAP $P(\mathbf{R})$ is given as $\psi(\mathbf{R})^2$ in Eq. (6.10) which is nonnegative definite

in $\mathbb{R}^3$. Because ODF is defined as the radial integration of EAP, the ODF is also nonnegative definite. Thus any wavefunction estimation method like the proposed SRPE is also a nonnegative definite ODF and EAP estimation method. Please note that standard QBI and SPFI in Chapter 5 and [Cheng et al., 2011c] do not consider nonnegative properties of ODF and EAP [Descoteaux et al., 2007; Cheng et al., 2010b]. Some existing works add nonnegative constraint in some discrete samples in $\mathbb{S}^2$ for ODFs [Goh et al., 2009b] or in $\mathbb{R}^3$ for EAPs [Özarslan et al., 2009]. However, the discrete constraint only can ensure the ODF/EAP is nonnegative at the given samples, while the ODF/EAP may be negative in other samples. Since the ODF is smooth enough and $\mathbb{S}^2$ is bounded [Goh et al., 2009b], in practice the estimated ODF is nonnegative when adding nonnegative constraint in a small number of discrete samples. However, it is impractical for EAPs to add the constraint in exhaustive samples in unbounded $\mathbb{R}^3$. SRPE naturally ensures ODF and EAP are nonnegative in whole $\mathbb{S}^2$ and $\mathbb{R}^3$.

SRPE maximizes the posteriori probability in Eq. (6.28), where $\{E_i\}_{i=1}^{N_s}$ is the given DWI signal samples for given $\mathbf{q}$ values $\{\mathbf{q}_i\}_{i=1}^{N_s}$, $\mathcal{L}(\{E_i\}|\psi(\mathbf{R}|\boldsymbol{c}))$ is the likelihood probability function, $\mathcal{L}(\{E_i\}|\psi(\mathbf{R}|\boldsymbol{c})) = \prod_{i=1}^{N_s}\mathcal{L}(E_i|\psi(\mathbf{R}|\boldsymbol{c}))$ and $p_1(\psi(\mathbf{R}|\boldsymbol{c}))$ is the prior for the wavefunction.

$$\begin{aligned}\boldsymbol{c} &= \arg\max_{\boldsymbol{c}} p(\psi(\mathbf{R}|\boldsymbol{c})|\{E_i\}) \\ &= \arg\max_{\boldsymbol{c}} \ln\mathcal{L}(\{E_i\}|\psi(\mathbf{R}|\boldsymbol{c})) + \ln p_1(\psi(\mathbf{R}|\boldsymbol{c})) \\ &= \arg\max_{\boldsymbol{c}} \sum_{i=1}^{N_s}\ln\mathcal{L}(E_i|\psi(\mathbf{R}|\boldsymbol{c})) + \ln p_1(\psi(\mathbf{R}|\boldsymbol{c}))\end{aligned} \tag{6.28}$$

Please note that for given $\boldsymbol{c}$, the expected signal $E(\mathbf{q}|\boldsymbol{c})$ can be analytically obtained from Fourier transform. See next paragraph for more details. Thus $\mathcal{L}(E_i|\psi(\mathbf{R}|\boldsymbol{c})) = \mathcal{L}(E_i|E(\mathbf{q}_i|\boldsymbol{c}))$ is a measurement of the similarity between E_i and $E(\mathbf{q}_i|\boldsymbol{c})$. If we consider $\mathcal{L}(E_i|E(\mathbf{q}_i|\boldsymbol{c}))$ as a Gaussian distribution, i.e.

$$\ln\mathcal{L}(E_i|E(\mathbf{q}_i|\boldsymbol{c})) = -(E(\mathbf{q}_i|\boldsymbol{c}) - E_i)^2$$

which results in a least square cost function. It is also possible to consider the $\mathcal{L}(E_i|E(\mathbf{q}_i|\boldsymbol{c}))$ as Rician distribution as [Assemlal et al., 2009a] did in SPFI. In this chapter we just consider $\mathcal{L}(E_i|E(\mathbf{q}_i|\boldsymbol{c}))$ as a Gaussian distribution to demonstrate our method, because the comparison between different likelihood functions is not our focus in this chapter. The prior $p_1(\psi(\mathbf{R}|\boldsymbol{c}))$ can be considered as a regularization. We consider the prior used in SPFI in Chapter 5 that $\psi(\mathbf{R}|\boldsymbol{c})$ is smooth both in radial and spherical part [Assemlal et al., 2009a; Cheng et al., 2010a,b].

$$p_1(\psi(\mathbf{R}|\boldsymbol{c})) = \exp(-\boldsymbol{c}^T\Lambda\boldsymbol{c})$$

where Λ is diagonal matrix with two parts of regularization weights

$$\Lambda_{nlm} = \lambda_n n^2(n+1)^2 + \lambda_l l^2(l+1)^2$$

Thus the SRPE method reduces to the minimization problem

$$\boldsymbol{c} = \arg\min_{\boldsymbol{c}} M(\boldsymbol{c}) = \arg\min_{\boldsymbol{c}} \frac{1}{2}\sum_{i=1}^{N_s}(E(\mathbf{q}_i|\boldsymbol{c}) - E_i)^2 + \frac{1}{2}\boldsymbol{c}^T\Lambda\boldsymbol{c} \tag{6.29}$$

Eq. (6.29) is similar with the least square estimation in SPFI [Assemlal et al., 2009a; Cheng et al., 2010a,b], where $E(\mathbf{q})$ is assumed to be represented linearly by SPF basis and the minimization can be done through matrix multiplication. But in Eq. (6.29) $E(\mathbf{q}|\boldsymbol{c})$ is determined by $\boldsymbol{c}$ in a quadratic way, not linearly.

We now show how to obtain $E(\mathbf{q}|\boldsymbol{c})$ from given $\boldsymbol{c}$ in details. Let $\psi(\mathbf{R}|\boldsymbol{c}) = \sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} c_{nlm}B_{nlm}(\mathbf{R})$, then $P(\mathbf{R}) = \psi(\mathbf{R})^2$ in Eq. (6.10) which is parameterized by its square root, and the expected signal $E(\mathbf{q})$ determined by $\psi(\mathbf{R})$ is expressed in Eq. (6.30).

$$E(\mathbf{q}|\boldsymbol{c}) = \int_{\mathbb{R}^3} \Big(\sum_{n=0}^{N}\sum_{l=0}^{L}\sum_{m=-l}^{l} c_{nlm}B_{nlm}(\mathbf{R}) \Big)^2 e^{-i2\pi\mathbf{q}^T\mathbf{R}} \mathrm{d}\mathbf{R} \tag{6.30}$$

Consider the plane wave equation in Eq. (6.31) shown in Theorem 2.8 and Appendix 5.7,

$$e^{\pm 2\pi i \mathbf{q}\cdot\mathbf{R}} = 4\pi \sum_{\alpha=0}^{\infty}\sum_{\beta=-\alpha}^{\alpha} (\pm i)^{\alpha} j_{\alpha}(2\pi qR) Y_{\alpha}^{\beta}(\mathbf{u}) Y_{\alpha}^{\beta}(\mathbf{r}), \tag{6.31}$$

where $j_\alpha(x)$ is the α-th order spherical Bessel function, then we have

$$E(\mathbf{q}|\boldsymbol{c}) = \sum_{nlm}\sum_{n'l'm'}\sum_{\alpha\beta} 4\pi(-1)^{\frac{\alpha}{2}} c_{nlm}c_{n'l'm'} I_{nn'\alpha}(q) Q_{ll'\alpha}^{mm'\beta} Y_{\alpha}^{\beta}(\mathbf{u}) = \boldsymbol{c}^T \mathbf{K}(\mathbf{q}|\zeta)\boldsymbol{c} \tag{6.32}$$

$$\mathbf{K}_{nlm}^{n'l'm'}(\mathbf{q}|\zeta) = \sum_{\alpha=0}^{2L}\sum_{\beta=-\alpha}^{\alpha} 4\pi(-1)^{\frac{\alpha}{2}} I_{nn'\alpha}(q) Q_{ll'\alpha}^{mm'\beta} Y_{\alpha}^{\beta}(\mathbf{u}) \tag{6.33}$$

where the integration in $\mathbb{R}^3$ in Eq. (6.30) can be separated into radial integration in Eq. (6.34) and spherical integration in Eq. (6.35).

$$I_{nn'\alpha}(q) = \int_0^{\infty} R_n(R) R_{n'}(R) j_{\alpha}(2\pi qR) R^2 \mathrm{d}R \tag{6.34}$$

$$Q_{ll'\alpha}^{mm'\beta} = \int_{\mathbb{S}^2} Y_l^m(\mathbf{r}) Y_{l'}^{m'}(\mathbf{r}) Y_{\alpha}^{\beta}(\mathbf{r}) \mathrm{d}\mathbf{r} \tag{6.35}$$

Both the radial integration and the spherical integration are independent of data and can be analytically obtained.

The spherical integration $Q_{ll'\alpha}^{mm'\beta}$ in Eq. (6.35) is the integration of three real SHs, which has been shown in Eq. (2.35) in Section 2.5.2. It can be computed from $W_{ll'\alpha}^{mm'\beta}$ in Eq. (2.34) which is the integration of product of three complex SHs. Please refer Section 2.5.2 for more details. Please note that in Eq. (6.32), $\sum_{\alpha\beta}$ means $\sum_{\alpha=0}^{2L}\sum_{\beta=-\alpha}^{\alpha}$, because if $\alpha > 2L$, $\alpha > 2L \geq l + l'$ violates the triangle inequality and $Q_{ll'\alpha}^{mm'\beta} = 0$.

The radial integration in Eq. (6.34) can be also analytically solved. Put Eq. (6.12) into Eq. (6.34), then we obtain Eq. (6.36), where $J_{\alpha+0.5}(x) = \sqrt{\frac{2x}{\pi}} j_\alpha(x)$ is the Bessel function of the first kind.

$$I_{nn'\alpha}(q) = \kappa_n(\zeta)\kappa_{n'}(\zeta)\frac{\zeta^{1.25}}{2\sqrt{q}} \int_0^{\infty} x^{1.5} \exp(-x^2) L_n^{0.5}(x^2) L_{n'}^{0.5}(x^2) J_{\alpha+0.5}(2\pi qx\sqrt{\zeta}) \mathrm{d}x \tag{6.36}$$

Similarly with our previous works on SPFI in [Cheng et al., 2010a,b], we consider $L_n^{0.5}(x) = \sum_{i=0}^{n} l_n^i x^i$, $l_n^i = (-1)^i \binom{n+0.5}{n-i} \frac{1}{i!}$ shown in Eq. (2.63), then

$$L_n^{0.5}(x^2) L_{n'}^{0.5}(x^2) = \sum_{i=0}^{n+n'} h_{nn'}^i x^{2i}, \quad \text{where } h_{nn'}^i = \sum_{j=0}^{\min(n,i)} l_n^j l_{n'}^{i-j} \delta(i-j \leq n') \tag{6.37}$$

Then the radial integration becomes

$$I_{nn'\alpha}(q) = \kappa_n(\zeta)\kappa_{n'}(\zeta)\frac{\zeta^{1.25}}{2\sqrt{q}}\sum_{i=0}^{n+n'} h_{nn'}^i \int_0^\infty x^{2i+1.5}\exp(-x^2)J_{\alpha+0.5}(2\pi q x\sqrt{\zeta})\mathrm{d}x \tag{6.38}$$

Based on the property of Bessel function given in Eq. (2.82) [Gradshteyn and Ryzhik, 2007], we have the final result as

$$I_{nn'\alpha}(q) = \kappa_n(\zeta)\kappa_{n'}(\zeta)\frac{\zeta^{0.5\alpha+1.5}\pi^{\alpha+0.5}q^\alpha}{4\Gamma(\alpha+1.5)}\sum_{i=0}^{n+n'} h_{nn'}^i \Gamma(\frac{1}{2}\alpha+i+\frac{3}{2})_1F_1(\frac{2i+\alpha+3}{2};\alpha+\frac{3}{2};-\pi^2q^2\zeta) \tag{6.39}$$

The final quadratic relation $E(\mathbf{q}|\boldsymbol{c}) = \boldsymbol{c}^T\mathbf{K}(\mathbf{q}|\zeta)\boldsymbol{c}$ is quite compact, where the kernel $\mathbf{K}(\mathbf{q}|\zeta)$ is a $N(L+1)(L+2)/2$ dimensional symmetric matrix for each fixed $\mathbf{q}$ and ζ, and $\mathbf{K}(\mathbf{q}|\zeta)$ is independent of data $E(\mathbf{q})$.

Let's back to the question in Section 6.1.3 why we choose SPF basis $\{B_{nlm}\}$ instead of dSPF basis in $\mathbf{R}$-space to represent EAP. That is because it is easy to obtain the analytic form of the radial integration in Eq. (6.34) from SPF basis. If we use dSPF basis $\{D_{nlm}\}$ to represent EAP like our previous work in Chapter 5, [Cheng et al., 2011c] and [Cheng et al., 2011a], the analytic form is hard to deduce, although a numerical integration is also possible.

For the minimization of $M(\boldsymbol{c})$ in Eq. (6.29), we can first set the initial $\boldsymbol{c}^{(0)}$ as $(1,0,...,0)^T$, which is a typical isotropic Gaussian as we discussed above, then perform gradient descent on PS. For a given $\boldsymbol{c}$, let $\boldsymbol{v}$ be a tangent vector in the tangent space of $\boldsymbol{c}$ and $\boldsymbol{c}(t)$ is the geodesic determined by $\boldsymbol{c}$ and $\boldsymbol{v}$. Then $\boldsymbol{c}(0) = \boldsymbol{c}$, $\boldsymbol{v} = \frac{d\boldsymbol{c}(t)}{dt}|_{t=0}$, and $\frac{dM(\boldsymbol{c}(t))}{dt}|_{t=0} = \langle\nabla M(\boldsymbol{c}), \boldsymbol{v}\rangle_{\boldsymbol{c}}$ based on the definition of directional derivative on manifold in Section 3.1.2. On the other hand, because $\boldsymbol{v}^T\boldsymbol{c} = 0$, we have

$$\frac{dM(\boldsymbol{c}(t))}{dt}\Big|_{t=0} = (\frac{\partial M(\boldsymbol{c})}{\partial \boldsymbol{c}})^T\frac{d\boldsymbol{c}(t)}{dt}\Big|_{t=0} = \left(\frac{\partial M(\boldsymbol{c})}{\partial \boldsymbol{c}} - \left(\boldsymbol{c}^T\frac{\partial M(\boldsymbol{c})}{\partial \boldsymbol{c}}\right)\boldsymbol{c}\right)^T\boldsymbol{v}$$

Thus we have the gradient of $M(\boldsymbol{c})$ in Eq. (6.40), where $\frac{\partial M(\boldsymbol{c})}{\partial \boldsymbol{c}}$ is given in Eq. (6.41).

$$\nabla M(\boldsymbol{c}) = \frac{\partial M(\boldsymbol{c})}{\partial \boldsymbol{c}} - \left(\boldsymbol{c}^T\frac{\partial M(\boldsymbol{c})}{\partial \boldsymbol{c}}\right)\boldsymbol{c} \tag{6.40}$$

$$\frac{\partial M(\boldsymbol{c})}{\partial \boldsymbol{c}} = \sum_{i=1}^{N_s} 2\left(\boldsymbol{c}^T\mathbf{K}(\mathbf{q}_i|\zeta)\boldsymbol{c} - E_i\right)\mathbf{K}(\mathbf{q}_i|\zeta)\boldsymbol{c} + \Lambda\boldsymbol{c} \tag{6.41}$$

The gradient descent method is given in Eq. (6.42), where $\boldsymbol{c}^{(k)}$ means $\boldsymbol{c}$ in k-th step, and dt is the step size estimated from standard line search method. See Algorithm 6.3 for the details, where $dt_0 = 0.1$ experimentally, considering $\boldsymbol{c}$ is in the unit sphere.

$$\boldsymbol{c}^{(k+1)} = Exp_{\boldsymbol{c}^{(k)}}(-dt\frac{\nabla M(\boldsymbol{c}^{(k)})}{\|\nabla M(\boldsymbol{c}^{(k)})\|}), \quad \text{if} \quad \|\nabla M(\boldsymbol{c}^{(k)})\| > \varepsilon_1 \tag{6.42}$$

After we obtain $\boldsymbol{c}$ from DWI samples $\{E_i\}$, the EAP $P(\mathbf{R}) = (\psi(\mathbf{R}))^2 = (\sum_{nlm} c_{nlm}B_{nlm}(\mathbf{R}))^2$ is nonnegative definite in $\mathbb{R}^3$, and the integration of $P(\mathbf{R})$ in $\mathbb{R}^3$ is

Algorithm 6.3: Wavefunction Estimation

Input: DWI samples $\{E_i\}$.
Output: Coefficient $\boldsymbol{c}$ of $\psi(\mathbf{R}|\boldsymbol{c})$.
begin
  initialization: $\boldsymbol{c}^{(0)} = (1, 0, ..., 0)^T$, $k = 0$; `// typical isotropic Gaussian`
  repeat
    calculate $\boldsymbol{v} = \nabla M(\boldsymbol{c}^{(k)})$ in Eq. (6.40) ;
    if $\|\boldsymbol{v}\| < \varepsilon_1$ **then** break;
    choose step size $dt \in (0, dt_0]$ via line search;
    $\boldsymbol{c}^{(k+1)} = Exp_{\boldsymbol{c}^{(k)}}(-dt\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|})$;
    $k \leftarrow k + 1$;
  until $\frac{M(\boldsymbol{c}^{(k-1)}) - M(\boldsymbol{c}^{(k)})}{M(\boldsymbol{c}^{(k-1)})} < \varepsilon_2$;
end

naturally one because $\boldsymbol{c}$ is in PS. For given radius R_0, the EAP profile which normally is used to detect fiber directions is given as

$$P(R_0\mathbf{r}) = \sum_{nlm}\sum_{n'l'm'} c_{nlm}c_{n'l'm'}R_n(R_0)R_{n'}(R_0)Y_l^m(\mathbf{r})Y_{l'}^{m'}(\mathbf{r}) \tag{6.43}$$

It also can be linearly represented by SHs.

$$P(R_0\mathbf{r}) = \sum_{\alpha=0}^{2L}\sum_{\beta=-\alpha}^{\alpha}\left(\sum_{nlm}\sum_{n'l'm'} c_{nlm}c_{n'l'm'}R_n(R_0)R_{n'}(R_0)Q_{nn'\alpha}^{mm'\beta}\right)Y_\alpha^\beta(\mathbf{r}) \tag{6.44}$$

The two kinds of ODFs $\Phi_t(\mathbf{r})$ and $\Phi_w(\mathbf{r})$ are also nonnegative definite in $\mathbb{S}^2$ since they are radial integration of $P(\mathbf{R})$. $\Phi_t(\mathbf{r})$ and $\Phi_w(\mathbf{r})$ can be analytically obtained from wavefunction. For ODF by Tuch $\Phi_t(\mathbf{r})$, we have

$$\begin{aligned}\Phi_t(\mathbf{r}) &= \frac{1}{Z}\int_0^\infty (\sum_{nlm} c_{nlm}B_{nlm}(\mathbf{R}))^2 \mathrm{d}R \\ &= \frac{1}{Z}\sum_{nlm}\sum_{n'l'm'}\underbrace{\left(\int_0^\infty R_n(R)R_{n'}(R)\mathrm{d}R\right)}_{I_{nn'}^{(0)}} c_{nlm}c_{n'l'm'}Y_l^m(\mathbf{r})Y_{l'}^{m'}(\mathbf{r})\end{aligned} \tag{6.45}$$

$$\begin{aligned}I_{nn'}^{(0)} &= \int_0^\infty R_n(R)R_{n'}(R)\mathrm{d}R \\ &= \kappa_n(\zeta)\kappa_{n'}(\zeta)\frac{\sqrt{\zeta}}{2}\int_0^\infty \exp(-x)\sum_{i=0}^{n+n'} h_{nn'}^i x^{i-0.5}\mathrm{d}x \qquad (6.46)\\ &= \kappa_n(\zeta)\kappa_{n'}(\zeta)\frac{\sqrt{\zeta}}{2}\sum_{i=0}^{n+n'} h_{nn'}^i\Gamma(i+\frac{1}{2}) \qquad (6.47)\end{aligned}$$

In Eq. (6.46), we use $L_n^{0.5}(x)L_{n'}^{0.5}(x) = \sum_{i=0}^{n+n'} h_{nn'}^i x^i$ in Eq. (6.37). Then put Eq. (6.47) into Eq. (6.45), we have

$$\Phi_t(\mathbf{r}) = \frac{\sqrt{\zeta}}{2Z}\sum_{nlm}\sum_{n'l'm'}\kappa_n(\zeta)\kappa_{n'}(\zeta)c_{nlm}c_{n'l'm'}Y_l^m(\mathbf{r})Y_{l'}^{m'}(\mathbf{r})\sum_{i=0}^{n+n'} h_{nn'}^i\Gamma(i+\frac{1}{2}) \tag{6.48}$$

$\Phi_t(\mathbf{r})$ now is represented by products of SHs. It is also possible to represent $\phi_t(\mathbf{r})$ linearly by SHs.

$$\Phi_t(\mathbf{r}) = \frac{\sqrt{\zeta}}{2Z}\sum_{\alpha=0}^{2L}\sum_{\beta=-\alpha}^{\alpha}\Big(\sum_{nlm}\sum_{n'l'm'}\kappa_n(\zeta)\kappa_{n'}(\zeta)c_{nlm}c_{n'l'm'}Q_{ll'\alpha}^{mm'\beta}\sum_{i=0}^{n+n'}h_{nn'}^{i}\Gamma(i+\frac{1}{2})\Big)Y_\alpha^\beta(\mathbf{r}) \tag{6.49}$$

For ODF by Wedeen $\Phi_w(\mathbf{r})$, we have

$$\begin{aligned}\Phi_w(\mathbf{r}) &= \int_0^\infty (\sum_{nlm} c_{nlm}B_{nlm}(\mathbf{R}))^2R^2\mathrm{d}R \\ &= \sum_{nlm}\sum_{n'l'm'}\Big(\int_0^\infty R_n(R)R_{n'}(R)R^2\mathrm{d}R\Big)c_{nlm}c_{n'l'm'}Y_l^m(\mathbf{r})Y_{l'}^{m'}(\mathbf{r}) &(6.50)\\ &= \sum_{nlm}\sum_{l'm'}c_{nlm}c_{nl'm'}Y_l^m(\mathbf{r})Y_{l'}^{m'}(\mathbf{r}) &(6.51)\\ &= \sum_{\alpha=0}^{2L}\sum_{\beta=-\alpha}^{\alpha}\Big(\sum_{nlm}\sum_{l'm'}c_{nlm}c_{nl'm'}Q_{ll'\alpha}^{mm'\beta}\Big)Y_\alpha^\beta(\mathbf{r}) &(6.52)\end{aligned}$$

From Eq. (6.50) to Eq. (6.51), we use the orthogonality of radial part basis, i.e. $\int_0^\infty R_n(R)R_{n'}(R)R^2\mathrm{d}R = \delta_{nn'}$. From Eq. (6.51), it is clear that $\int_{\mathbb{S}^2}\Phi_w(\mathbf{r})\mathrm{d}\mathbf{r} = \sum_{nlm}c_{nlm}^2 = 1$, considering orthogonality of SHs. So the proposed estimator of $\Phi_w(\mathbf{r})$ is indeed the marginal EAP.

In the proposed method, we choose $\zeta = 4\tau D_0$ as we discussed above, which makes typical isotropic $P(\mathbf{R})$ and $E(\mathbf{q})$ are represented only by $B_{000}(\mathbf{R})$ and $D_{000}(\mathbf{q})$ respectively. Please note that sometimes for real data we do not know the diffusion time τ, and only $\{b_i\}$ values and gradients $\{\mathbf{u}_i\}$ are known. Thus we cannot calculate accurately $\{q_i\}$ from $\{b_i\}$.

Considering $b = 4\pi^2\tau q^2$, Eq. (6.39) can be represented by b as

$$I_{nn'\alpha}(b) = \kappa_n(\zeta)\kappa_{n'}(\zeta)\frac{\zeta^{0.5\alpha+1.5}\pi^{0.5}b^{0.5\alpha}}{2^{\alpha+2}\tau^{0.5\alpha}\Gamma(\alpha+1.5)}\sum_{i=0}^{n+n'}h_{nn'}^{i}\Gamma(\frac{1}{2}\alpha+i+\frac{3}{2})_1F_1(\frac{2i+\alpha+3}{2};\alpha+\frac{3}{2};-\frac{b\zeta}{4\tau})$$

Considering $b = 4\pi^2\tau q^2$ and $\kappa_n(\zeta)$ in Eq. (6.12), if we set $\zeta = 4\tau D_0$ as discussed in Section 6.2.1, Eq. (6.39) can be represented by b in Eq. (6.53).

$$\begin{aligned}I_{nn'\alpha}(b) = &\Big[\frac{n!n'!}{\Gamma(n+3/2)\Gamma(n'+3/2)}\Big]^{1/2}\frac{D_0^{0.5\alpha}\pi^{0.5}b^{0.5\alpha}}{2\Gamma(\alpha+1.5)} \\ &\sum_{i=0}^{n+n'}h_{nn'}^{i}\Gamma(\frac{1}{2}\alpha+i+\frac{3}{2})_1F_1(\frac{2i+\alpha+3}{2};\alpha+\frac{3}{2};-bD_0)\end{aligned} \tag{6.53}$$

It is clear to see that $I_{nn'\alpha}(b)$ only depends on D_0 and is independent with τ when $\zeta = 4\tau D_0$. Based on the above analysis, $I_{nn'\alpha}(q)$ in Eq. (6.34) only depends on D_0 if we choose $\zeta = 4\tau D_0$, so different τ obtains the same $\boldsymbol{c}$ when D_0 is fixed. But different τ will give different scale ζ in R-axis for $P(\mathbf{R})$ in Eq. (6.10). Two kinds of ODFs are invariant for different τ based on Eq. (6.49) and Eq. (6.52). Thus if we do not know τ for real data, we always assume $4\pi^2\tau = 1$, i.e. $\tau = 0.02533s$, which is also used in the following experiments.

We have proposed a method to estimate wavefunction for EAPs. However it is impossible to estimate the wavefunction for ODFs directly from DWI signals, if we do not add some assumptions on $E(\mathbf{q})$. That is because different $E(\mathbf{q})$ may obtain the same ODF, which means we cannot represent $E(\mathbf{q})$ using wavefunction of ODFs as what we did for EAPs in Eq. (6.32). Of course we can make some assumptions on $E(\mathbf{q})$ such that there is one to one map between $E(\mathbf{q})$ and ODF. In details, $\Phi_t(\mathbf{r})$ in QBI uniquely determines signal shell DWI signal [Descoteaux et al., 2007] by Funk-Radon transform. $\Phi_w(\mathbf{r})$ in exact QBI uniquely determines the SH coefficients with $L \geq 2$ of signal shell DWI signal by assuming mono-exponential decay on $E(\mathbf{q})$ [Canales-Rodrıguez et al., 2009; Aganj et al., 2010b]. However we prefer model-free method without assumptions. So we suggest to first estimate the nonnegative definite ODF using the proposed SRPE method, then estimate the wavefunction of ODF from pre-computed ODFs using numerical integration (or least square fitting) shown in Section 6.2.1.

6.2.2 ODF/EAP Interpolation

Like Riemannian framework for tensors, a straightforward application of weighted Riemannian mean is ODF/EAP interpolation, which is an important step in fiber tracking and registration. On a regular lattice in $\mathbb{R}^d$, the interpolated scalar valued function $f(\boldsymbol{s})$ at spatial position $\boldsymbol{s}$ is normally a linear combination of N_s samples in its neighborhood, i.e.

$$f(\boldsymbol{s}) = \sum_i w_i(\boldsymbol{s}) f(\boldsymbol{s}^{(i)})$$

The weighting function $\{w_i(\boldsymbol{s})\}$ are normally determined by the spatial distance between $\boldsymbol{s}^{(i)}$ and $\boldsymbol{s}$, and

$$w_i(\boldsymbol{s}^{(j)}) = \delta_i^j$$

such that $f(\boldsymbol{s}^{(i)})$ is invariant under interpolation operation. Moreover, the weighting functions are normally positive and normalized such that $\sum_i w_i(\boldsymbol{s}) = 1$ and the linear combination is weighted mean. This kind of interpolation is called as **mean value interpolation**. When $\{f(\boldsymbol{s}_i)\}$ are samples of manifold valued function, the weighted mean is

$$f(\boldsymbol{s}) = \arg\min_f \sum_{i=1}^{N_s} w_i(\boldsymbol{s}) d(f(\boldsymbol{s}^{(i)}), f)^2$$

which has no closed form and needs to be solved using Algorithm 6.1. Borrowing from the notation in [Buss and Fillmore, 2001], we denote the weighted Riemannian mean by

$$\widetilde{\sum_{i=1}^{N_s}} w_i(\boldsymbol{s}) f(\boldsymbol{s}^{(i)}) \stackrel{\text{def}}{=} w_1(\boldsymbol{s}) f(\boldsymbol{s}^{(1)}) \oplus w_2(\boldsymbol{s}) f(\boldsymbol{s}^{(2)}) \oplus \cdots \oplus w_N(\boldsymbol{s}) f(\boldsymbol{s}^{(N_s)})$$

While weighted Euclidean mean is $\sum_{i=1}^{N} w_i(\boldsymbol{s}) f(\boldsymbol{s}_i) = w_1(\boldsymbol{s}) f(\boldsymbol{s}^{(1)}) + \cdots + w_N(\boldsymbol{s}) f(\boldsymbol{s}^{(N_s)})$.

In this chapter, we just demonstrate a simple **Lagrange interpolation** which is normally used in MRI data processing. In 1 dimension (1D) case, $w_i(s) = \prod_{j=1, j\neq i}^{N_s} \frac{s - s^{(j)}}{s^{(i)} - s^{(j)}}$, where N_s samples are used in s-axis. If $N_s = 2$, $f(s) = tf(s^{(1)}) \oplus (1-t) f(s^{(2)})$ is the weighted mean, where $t = \frac{s - s^{(1)}}{s^{(2)} - s^{(1)}}$. When t changes in $[0, 1]$, $f(s)$ shows the geodesic connecting $f(s^{(1)})$ and $f(s^{(2)})$, which has the analytical closed form as

$f(s) = Exp_{f(s^{(1)})}(tLog_{f(s^{(1)})}(f(s^{(2)})))$. There is no closed form for $N_s > 2$ in 1D or in high dimension. It is straightforward to generalize interpolation from 1D to high dimension. For example, $f(\boldsymbol{s})$ in 2 dimension where $\boldsymbol{s} = (s_1, s_2)^T$ is shown in Eq. (6.54).

$$f(\boldsymbol{s}) = \widetilde{\sum_{i=1}^{N_{s1}}} \widetilde{\sum_{j=1}^{N_{s2}}} w_i^{(1)}(s_1) w_j^{(2)}(s_2) f(\boldsymbol{s}^{(i,j)})|_{\boldsymbol{s}^{(i,j)}=(s_1^{(i)}, s_2^{(j)})^T} \tag{6.54}$$

$$w_i^{(1)}(s_1) = \prod_{l=1, l\neq i}^{N_{s1}} \frac{s_1 - s_1^{(l)}}{s_1^{(i)} - s_1^{(l)}}, \quad w_j^{(2)}(s_2) = \prod_{m=1, m\neq j}^{N_{s2}} \frac{s_2 - s_2^{(m)}}{s_2^{(j)} - s_2^{(m)}} \tag{6.55}$$

Lagrange interpolation is a polynomial interpolation method whose weighting functions are normally not normalized. When 2 samples are used in each dimension, it reduces to **linear interpolation** which is a kind of **mean value interpolation** with normalized weighting functions. Linear interpolation is closely related to geodesics and was used to compare different metric for tensors [Pennec et al., 2006].

An important issue on linear interpolation which is concerned in dMRI field is the so called **swelling effect**. In linear interpolation of tensors with Euclidean metric, the interpolated tensors sometimes have larger determinant than the fixed tensors in endpoints. This phenomenon is called as swelling effect [Pennec et al., 2006]. However, to our knowledge there is no definition of swelling effect for general PDF valued data interpolation so far. [Goh et al., 2011] discussed the swelling effect for ODF interpolation via **Shannon entropy** which is $\frac{1}{2}\ln((2\pi e)^3|\Sigma|)$ for Gaussian distribution $N(\mathbf{R}|\Sigma)$, and concluded the swelling effect occurs in both Riemannian and Euclidean interpolation of ODFs. However we think the increase of Shannon entropy may be necessary for PDF data interpolation in some cases. For example, when interpolating two discrete PDFs $\boldsymbol{p}^{(1)} = (1,0,0)^T$ and $\boldsymbol{p}^{(2)} = (0,1,0)^T$ which have the Shannon entropy of zero, it is reasonable that the interpolated mean PDF is $\boldsymbol{p} = (\frac{1}{2}, \frac{1}{2}, 0)^T$ which is the common result from both Riemannian and Euclidean interpolation in PDF space and has larger Shannon entropy than $\boldsymbol{p}^{(1)}$ and $\boldsymbol{p}^{(2)}$. See Fig. 6.1(A,B,H) for the sketch map. When interpolating two Gaussian EAPs with the same tensor shape but different directions, the interpolated Gaussian distributions from Riemannian tensor interpolation in tensor space have the same Shannon entropy but no crossing glyph, while the interpolated EAPs from both Riemannian and Euclidean interpolation have larger entropy but crossing glyphs, which is more reasonable.

In this chapter we propose the definition of **reasonable mean value interpolation** of the spatial interpolation on PDF valued function $p(\boldsymbol{s}, \boldsymbol{x})$. The so called swelling effect for tensor interpolation can be considered as a specific phenomenon in **unreasonable mean value interpolation** of Gaussian PDFs. Generally a PDF field can be considered as a PDF valued function defined in spatial space as well as a scalar function defined in both spatial and probability space, i.e. $p(\boldsymbol{s}, \boldsymbol{x}) : \mathbb{R}^d \times \chi \to \mathbb{R}$, where $\mathbb{R}^d$ is the d dimensional spatial space and χ is the measure space for PDFs which is $\mathbb{S}^2$ for ODFs and $\mathbb{R}^3$ for EAPs. $p_{\boldsymbol{s}}(\boldsymbol{x}) \overset{\text{def}}{=} p(\boldsymbol{s}, \boldsymbol{x})$ is the PDF defined at spatial position $\boldsymbol{s}$.

Definition 6.1. *For a PDF valued function $p(\boldsymbol{s}, \boldsymbol{x})$, a mean value interpolation based on N_s samples $\{p_{\boldsymbol{s}^{(i)}}(\boldsymbol{x})\}_{i=1}^{N_s}$ at position $\{\boldsymbol{s}^{(i)}\}_{i=1}^{N_s}$ is **reasonable** if the interpolated PDF $p_{\boldsymbol{s}}(\boldsymbol{x})$ at position $\boldsymbol{s}$ satisfies*

$$p_{\boldsymbol{s}}(\boldsymbol{x}) \geq \min\{p_{\boldsymbol{s}^{(i)}}(\boldsymbol{x})\}_{i=1}^{N}, \quad \forall \boldsymbol{x} \in \chi \tag{6.56}$$

It is straightforward that the mean value interpolation of scalar valued function $f(\boldsymbol{s})$ is reasonable because $\sum_{i=1}^{N_s} w_i(\boldsymbol{s}) f(\boldsymbol{s}^{(i)}) \geq \min\{f(\boldsymbol{s}^{(i)})\}_{i=1}^{N_s}$ when $\sum_i w_i(\boldsymbol{s}) = 1$ and $w_i \geq 0$. For the Euclidean interpolation on Gaussian distribution parameterized by tensors $N_{\boldsymbol{s}}(\mathbf{R}|\Sigma)$, $\Sigma = \sum_{i=1}^{N_s} w_i \Sigma_i$ where $\{\Sigma_i\}_{i=1}^{N_s}$ are covariance tensors of given samples $\{N_{\boldsymbol{s}^{(i)}}(\mathbf{R}|\Sigma_i)\}_{i=1}^{N_s}$. The return-to-origin probability (RTO) [Wu and Alexander, 2007; Cheng et al., 2010b] is

$$Po = N(\boldsymbol{0}|\Sigma) = \frac{1}{\sqrt{(2\pi)^3|\Sigma|}}$$

where $\Sigma = 2\tau\mathbf{D}$. Thus the larger determinant of the tensor Σ (or $\mathbf{D}$) is, the smaller RTO of EAP. When swelling effect occurs, $|\Sigma| > \max\{|\Sigma_i|\}_{i=1}^{N_s}$ [Pennec et al., 2006], which means $N_{\boldsymbol{s}}(\boldsymbol{0}|\Sigma) < \min\{N_{\boldsymbol{s}^{(i)}}(\boldsymbol{0}|\Sigma_i)\}_{i=1}^{N_s}$. So Euclidean interpolation on tensors is unreasonable. It is actually interpolation on matrices, not EAPs. On the other hand, if the mean value interpolation is reasonable, i.e. $N_{\boldsymbol{s}}(\mathbf{R}|\Sigma) \geq \min\{N_{\boldsymbol{s}^{(i)}}(\mathbf{R}|\Sigma_i)\}_{i=1}^{N_s}$, $\forall \mathbf{R} \in \mathbb{R}^3$, then there is no swelling effect. So **swelling effect** is actually a specific phenomenon of **unreasonable mean value interpolation** of Gaussian PDF field. For Riemannian interpolation on tensors, Riemannian and Euclidean interpolation on general EAPs, we have the theorem 6.6 whose proof is given in Appendix 6.6.

Theorem 6.6 (Reasonable Interpolations). *Riemannian and Euclidean mean value interpolations on general EAPs/ODFs are both reasonable. Riemannian mean value interpolation on tensors is reasonable, and more precisely*

$$\max\{N_{\boldsymbol{s}^{(i)}}(\mathbf{R}|\Sigma_i)\}_{i=1}^{N_s} \geq N_{\boldsymbol{s}}(\mathbf{R}|\Sigma) \geq \min\{N_{\boldsymbol{s}^{(i)}}(\mathbf{R}|\Sigma_i)\}_{i=1}^{N_s}, \quad \forall \mathbf{R} \in \mathbb{R}^3 \tag{6.57}$$

Please note two important differences between tensor based interpolation and interpolation on EAP space.

- EAP mean value interpolation with Euclidean metric is totally different from tensor interpolation with Euclidean metric. For given Gaussian distributions $\{N_{\boldsymbol{s}^{(i)}}(\mathbf{R}|\Sigma_i)\}_{i=1}^{N_s}$, the interpolated EAP is $P(\mathbf{R}) = \sum_{i=1}^{N_s} w_i N_{\boldsymbol{s}^{(i)}}(\mathbf{R}|\Sigma_i)$ which is no longer Gaussian.
- EAP mean value interpolation with Riemannian metric may have larger probability values than EAPs in the endpoints, which is reasonable as discussed in Section 6.1.7. However, EAP mean value interpolation with Euclidean metric always obtain the EAPs whose probability values are between the values of EAPs in endpoints. The EAPs from Riemannian interpolation on tensors, called tensor-EAPs, also always have the probability values between the EAPs in the endpoints.

6.2.3 ODF/EAP filtering

Another straightforward application of Riemannian metric is the filtering in ODF/EAP field, which can improve the ODF/EAP estimation by considering the information in a spatial neighborhood. Since weighted mean/median could be calculated efficiently, it is possible to develop some filtering algorithms on ODF/EAP field. Actually all filtering algorithms in [Pennec et al., 2006; Fillard et al., 2007] for tensor field can be generalized into ODF/EAP field. For example, [Goh et al., 2009a,

2011] proposed anisotropic filtering for ODF field. Here we just demonstrate a simple Gaussian smoothing method since it is widely used in MRI field and does not need an iterative process. In this method the filtered EAP $p_{\boldsymbol{s}}(\boldsymbol{x}|\boldsymbol{c})$ in each position $\boldsymbol{s}$ is the weighted mean $p_{\boldsymbol{s}}(\boldsymbol{x}|\boldsymbol{\mu}_{\boldsymbol{w}})$ or median $p_{\boldsymbol{s}}(\boldsymbol{x}|\boldsymbol{m}_{\boldsymbol{w}})$ of the ODFs/EAPs in a given neighborhood $\{p_{\boldsymbol{s}}(\boldsymbol{x}|\boldsymbol{c}^{(i)})\}_{i\in N(\boldsymbol{s})}$, where $N(\boldsymbol{s})$ is in the spatial neighborhood of $\boldsymbol{s}$. The weighting functions $w_i(\boldsymbol{s}^{(i)})$ can be chosen as spatial Gaussian kernel $\exp(-\frac{d(\boldsymbol{s},\boldsymbol{s}^{(i)})}{\sigma_{\boldsymbol{s}}})$ in traditional Gaussian filtering, or Gaussian kernel $\exp(-\frac{d(\boldsymbol{s},\boldsymbol{s}^{(i)})}{\sigma_{\boldsymbol{s}}}\frac{d(p_{\boldsymbol{s}}(\boldsymbol{x}|\boldsymbol{c}),p_{\boldsymbol{s}^{(i)}}(\boldsymbol{x}|\boldsymbol{c}^{(i)})}{\sigma_{\boldsymbol{c}}})$ dependent on both spatial distance and ODF/EAP distance in **bilateral filtering** [Tomasi and Manduchi, 1998], where $\sigma_{\boldsymbol{s}}$ and $\sigma_{\boldsymbol{c}}$ are standard deviation in spatial and PDF domain and need to be chosen by users.

6.2.4 ODF/EAP Atlas Estimation

The weighted Riemannian mean and median can be used in atlas estimation. The measure of robustness of an estimator is the **breakdown point** which is the fraction of the data that can be completely corrupted without affecting the boundedness of the estimator. Compared with the mean, the median whose breakdown point is larger is a much more robust estimator which was shown in [Fletcher et al., 2009]. This property makes the median more appropriate than the mean in atlas estimation.

6.3 Experiments

In this section, we validate the Riemannian framework and its applications in synthetic data, phantom data and real monkey data.

6.3.1 Synthetic Data

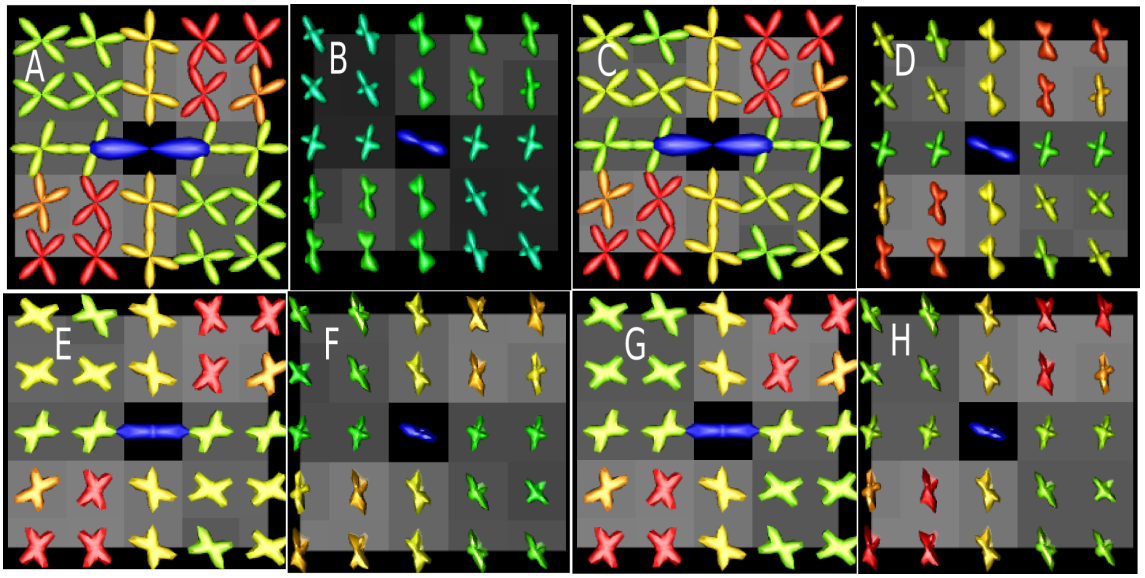
Demonstration for Diffeomorphism Invariance

We have proved in Section 6.1.8 that the proposed Riemannian metric for ODFs/EAPs is diffeomorphism invariant. In this section we give an example to demonstrate this property via an affine transform. See Fig. 6.2. Mixture of tensor model

$$P(\mathbf{R}) = w_1 N(\mathbf{R}|2\tau\mathbf{D}_1) + w_2 N(\mathbf{R}|2\tau\mathbf{D}_2)$$

was used to generate the synthetic data. See Appendix A. $\mathbf{D}_1$ and $\mathbf{D}_2$ have the same eigenvalues $[1.7, 0.3, 0.3]\times 10^{-3} mm^2/s$ but different directions. Based on this model, we can analytically obtain many samples of ODFs by Wedeen [Wedeen et al., 2005] and EAPs. Then the SH coefficients and SPF coefficients for the wavefunctions of ODFs and EAPs can be obtained via numerical inner product with $L = 8$, $N = 4$. Please note that we did not estimate ODF/EAP from limited number of DWI signal samples with noise because we would like to avoid the estimation error which is not our focus in this experiment. We compute the Riemannian distance and Euclidean distances between the EAP/ODF in the center where $w_1 = 1$, $w_2 = 0$ and the EAPs/ODFs in other

Figure 6.2: (A,C): original EAP profiles colored by Euclidean and Riemannian distances; (B,D): transformed EAP profiles colored by Euclidean and Riemannian distances; (E,G): original ODF colored by Euclidean and Riemannian distances; (F,H): transformed ODF colored by Euclidean and Riemannian distances.



voxels where $w_1 = w_2 = 0.5$. Afterwards a given affine transform A was performed on the mixture model, i.e.

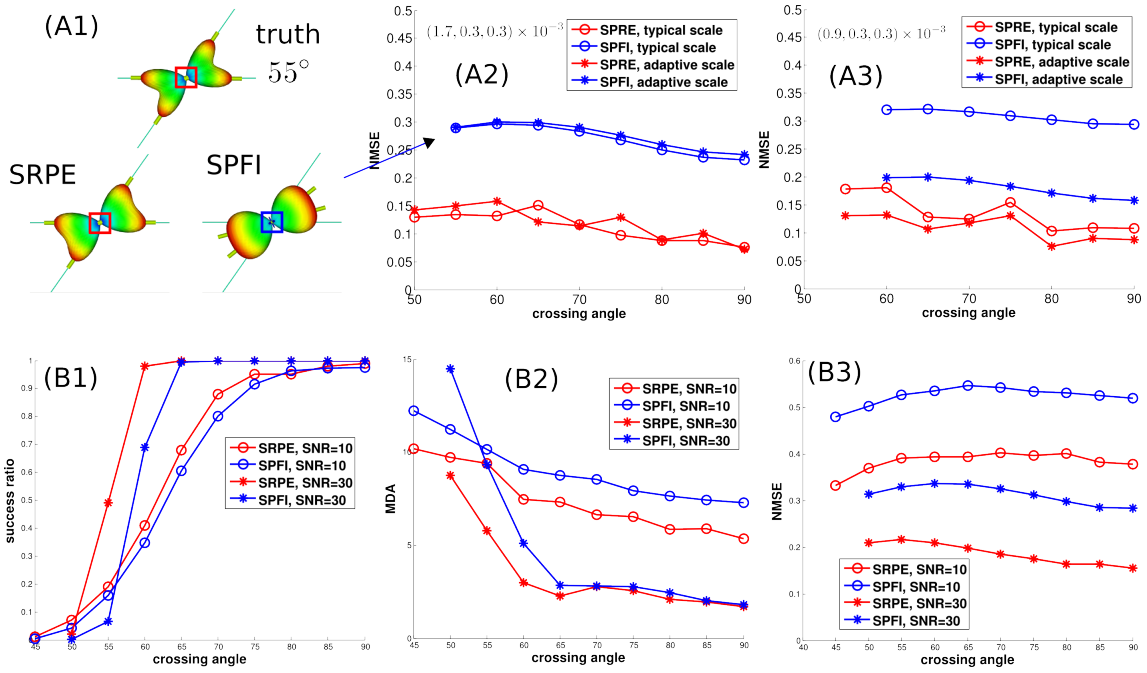
$$(A \circ P)(\mathbf{R}) = w_1 N(\mathbf{R}|2\tau A^T \mathbf{D}_1 A) + w_2 N(\mathbf{R}|2\tau A^T \mathbf{D}_2 A)$$

With this transformed model, we again estimated the Riemannian coordinates for ODFs and EAPs via numerical inner product, calculated the distance maps and visualized glyphs colored by distances. In Fig. 6.2, (A,C) are the original EAP profiles at $15\mu m$ colored by Euclidean distance and Riemannian distance respectively, and (E,G) are the original ODF fields colored by Euclidean and Riemannian distance. The transformed EAPs are shown in Fig. 6.2(B,D) and transformed ODFs in (F,H). The distance maps were used to color the glyphs and were set as the background. It is clear that the Riemannian distance is invariant under the transform. As we know, the Riemannian distance between two Gaussian distributions is invariant under an affine transformation, while Euclidean distance is not. However, Riemannian framework for tensors does not work for non-Gaussian distributions, e.g. mixture of Gaussian model. This experiment demonstrates that Riemannian metric for EAPs and ODFs is diffeomorphism invariant, which is a natural extension of affine-invariant metric for tensors, while Euclidean metric is not.

Nonnegative ODF and EAP estimation via SRPE

In practice we found that the ODFs estimated by many HARDI methods normally have only a small number of negative values close to zero when the SNR is very low, however even with high SNR, the negative values are serious for the estimated EAPs especially for large radius R. Thus we focus on EAP estimation in this subsection.

Figure 6.3: A1: ground truth EAP and estimated EAPs from two methods, where the EAP by SPFI has negative values in the blue square. The long thin sticks and short thick sticks are the ground truth directions and the detected maxima respectively. A2, A3: NMSE in noise free experiment for two tensor configurations T_1, T_2. B1, B2, B3: success ratio, MDA and the mean of NMSE in the experiments with $SNR = 10, 30$.



The synthetic data were generated from mixture of tensor model [Descoteaux et al., 2007] where two tensors cross with a given angle in $[45°, 90°]$. Three shells (b=500,1500,3000 s/mm^2) were used, 60 samples per shell. EAP profiles with radius $R_0 = 15\mu m$ were estimated by SPFI and SRPE. The Normalized Mean Square Error (NMSE) between the ground truth EAP profile $P(R_0\mathbf{r})$ and the estimated EAP profile $\widetilde{P}(R_0\mathbf{r})$ is defined in Eq. (A.8). In the noise-free experiment, the signal was generated from two tensor configurations with eigenvalues $T_1 = (1.7, 0.3, 0.3) \times 10^{-3} mm^2/s$ and $T_2 = (0.9, 0.3, 0.3) \times 10^{-3} mm^2/s$. We set $N = 2$, $L = 4$, $\lambda_l = \lambda_n = 0$ and considered both typical scale and adaptive scale for SPFI and SRPE . Fig. 6.3(A2,A3) recorded the NMSE when two maxima were detected. SRPE generally obtains lower NMSE and has better angular resolution than SPFI. The adaptive scale obtains lower NMSE in two methods when T_2 is used, which is because the ADC in tensor T_1 is much close to the typical D_0, while the ADC in T_2 is not. Note that when T_2 is used, the NMSE obtained from SRPE with typical scale is just slightly higher than the NMSE obtained from SRPE with adaptive scale, which means the typical scale is acceptable in SRPE. Fig. 6.3(A1) shows the ground truth EAP and the estimated EAPs by two methods when T_1 and crossing angle of 55° are used. SRPE has better angular resolution and avoids the negative values around the original point in the EAP by SPFI. Note that the EAP profile estimated by SPFI in (A1) has more than 20% negative values showed in the blue square, although only 1% points are negative and have absolute values larger than one tenth of the maximal value of the EAP profile. In the experiment with Rician noise, T_1 and adaptive scale were used. We set $\lambda_l = \lambda_n = 10^{-8}$ for SPFI suggested in [Cheng et al., 2010b]. Since the coefficient $\boldsymbol{c}$ in SRPE has different range ($\|\boldsymbol{c}\| = 1$) from coefficients in SPFI, in order to perform a fair comparison, we still set

$\lambda_l = \lambda_n = 0$ for SRPE without any regularization. The estimation was performed for 1000 trials with $SNR = 10, 30$, where the success ratio was recored when two maxima were detected, the Mean Difference of Angle (MDA) was calculated in the successful trials, and the mean of NMSE was calculated over all trials. See Fig. 6.3(B1,B2,B3). It is clear that SRPE generally has higher success ratio, lower MDA and lower NMSE than SPFI.

Interpolation and PGA

We demonstrate the Lagrange interpolation of ODFs/EAPs in 1D and 2D field, and compare the results from Riemannian metric and Euclidean metric.

Fig. 6.4 shows the 1D interpolated ODFs and EAP profiles at $15\mu m$ from given two Gaussian distributions parameterized by different configurations of tensors. We have two ways to interpolate ODFs/EAPs between Gaussian distributions.

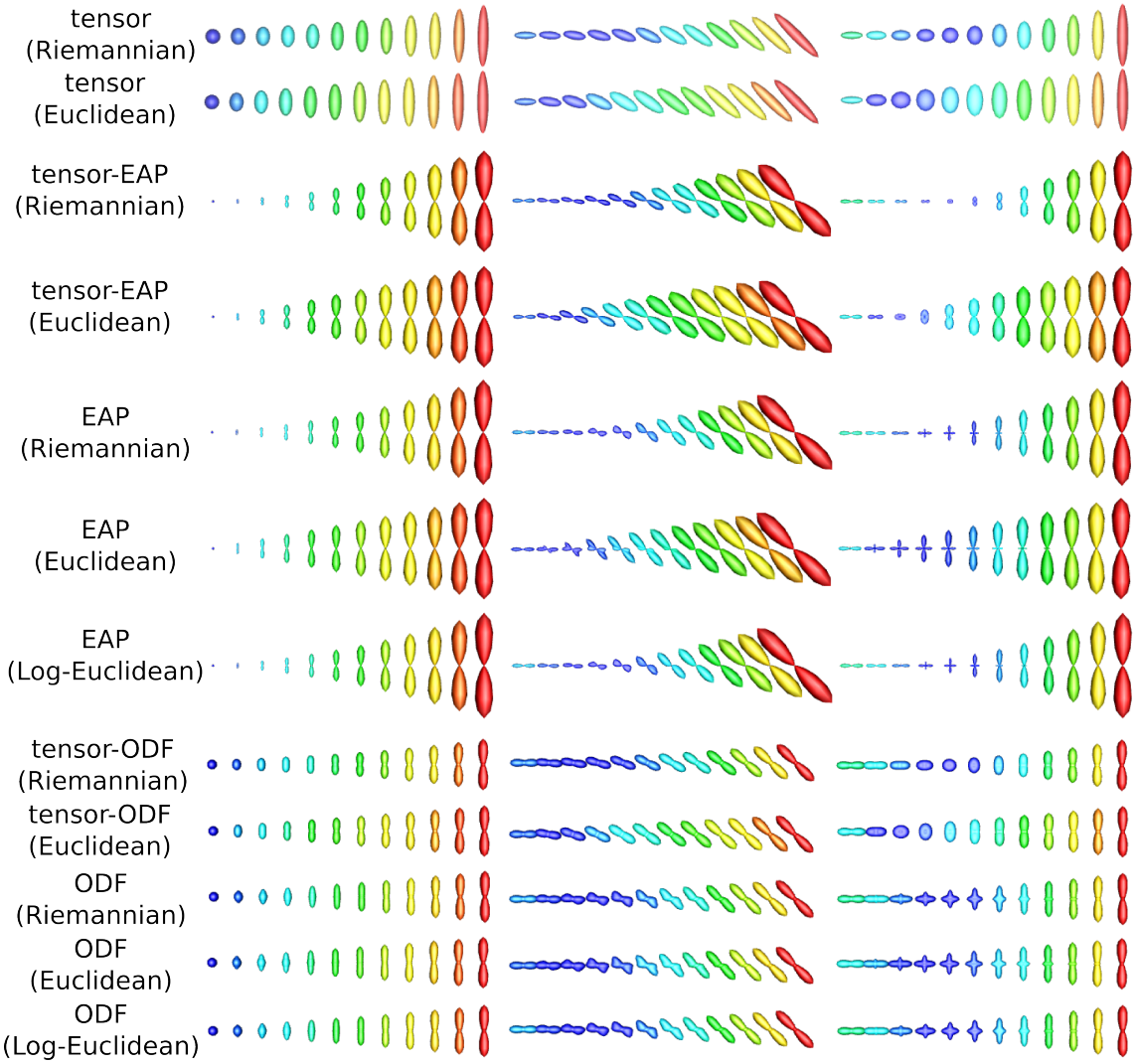
1. The first way is to first interpolate tensors between these two tensors in endpoints with Riemannian metric or Euclidean metric, then obtain the ODFs/EAPs analytically from the interpolated tensors based on the analytical solution of Gaussian distribution. We call them **tensor-ODFs** and **tensor-EAPs**.

2. The second way is to generate the ODFs/EAPs analytically from two fixed tensors, then interpolate ODFs/EAPs using the weighted mean of ODFs/EAPs with Euclidean, Riemannian and Log-Euclidean metrics respectively.

The coefficients of SH basis and SPF basis were estimated with $L = 8$ for ODFs and $L = 8$, $N = 4$ for EAPs. The first column in Fig. 6.4 shows the results between one isotropic Gaussian with eigenvalues of $[0.4, 0.4, 0.4] \times 10^{-3} mm^2/s$ and one anisotropic Gaussian with the eigenvalues of $[1.7, 0.3, 0.3] \times 10^{-3} mm^2/s$. The ODFs/EAPs from Euclidean and Riemannian interpolations are similar with the interpolated tensor-ODFs/EAPs from Euclidean and Riemannian interpolations respectively. The second and third columns give the results from two anisotropic Gaussian distributions with the eigenvalues of $[1.7, 0.3, 0.3] \times 10^{-3} mm^2/s$, $[0.6, 0.2, 0.2] \times 10^{-3} mm^2/s$ and with angular difference of $45°$, $90°$. Fig. 6.5 also shows EAP glyphs at $15\mu m$ and the change of RTO values for the interpolation of two Gaussian distributions which have the eigenvalues of $[1.7, 0.3, 0.3] \times 10^{-3} mm^2/s$, $[1.2, 0.3, 0.3] \times 10^{-3} mm^2/s$ and crossing angle of $90°$.

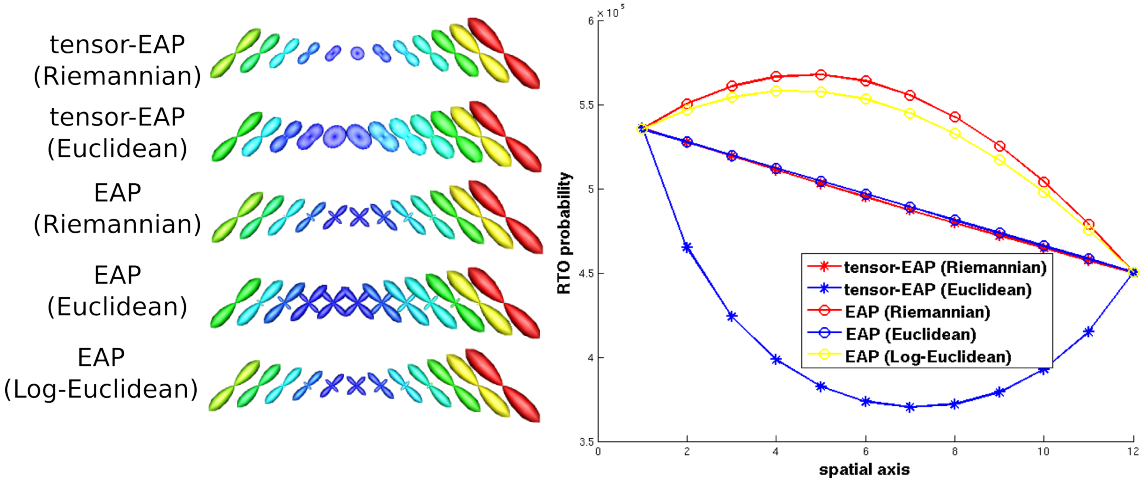
In 2D case, ODFs/EAPs were interpolated from 4 given ODFs/EAPs in 4 corners in Fig. 6.6. Tensor model with eigenvalues of $[0.6, 0.2, 0.2] \times 10^{-3} mm^2/s$ was used to generate ODFs and EAPs in top right and bottom left voxels. The same eigenvalues were used in mixture of two tensors to generate the ODF and EAP in top left voxel. The ODF and EAP in the bottom right voxel were generated from the tensor with eigenvalues of $[1.7, 0.3, 0.3] \times 10^{-3} mm^2/s$. Fig. 6.6 visualizes the interpolated ODFs and EAP profiles at $15\mu m$ from Riemannian and Euclidean interpolations. The glyphs were colored by GA values which were also set as the background grey intensity. Fig. 6.6 also shows the first two principal components (PCs) from PGA for the ODF and EAP fields from Riemannian interpolation, and the first two PCs from original PCA for the results from Euclidean interpolation.

Figure 6.4: Three configurations of tensors in endpoints were used in 1D interpolations of tensors/ODFs/EAPs with different metrics. Tensors and tensor-ODFs/EAPs were colored by GA of tensors, and ODF/EAP glyphs were colored by GA of ODFs/EAPs.



It can be seen in Fig. 6.4 that tensor-ODFs/EAPs cannot generate crossing glyphs, while ODF/EAP interpolation can. It is interesting that the first column in Fig. 6.4 shows that the tensor-EAPs from Riemannian interpolation of tensors are similar with EAPs from Riemannian interpolation of EAPs, and tensor-EAPs from Euclidean interpolation are similar with EAPs from Euclidean interpolation. However, please note the interpolated tensor-EAPs are still Gaussian, while the interpolated EAPs are no longer Gaussian. Fig. 6.5 shows the change of RTO probability in 1D interpolation. The swelling effect occurs in tensor-EAPs from Euclidean interpolation which is unreasonable. EAPs from Riemannian and Log-Euclidean interpolations obtain larger RTO than the EAPs in endpoints, which is possible as we have shown in Section 6.1.7 and 6.2.2. Although EAPs from Riemannian and Log-Euclidean interpolations are very similar when $R = 15\mu m$, they are slightly different for RTO values. EAPs from Euclidean interpolation and tensor-EAPs from Riemannian interpolation on tensors have similar RTO values, although EAP interpolation has crossing glyphs

Figure 6.5: The interpolation results and the change of RTO probability for 1D interpolation of two Gaussian EAPs.

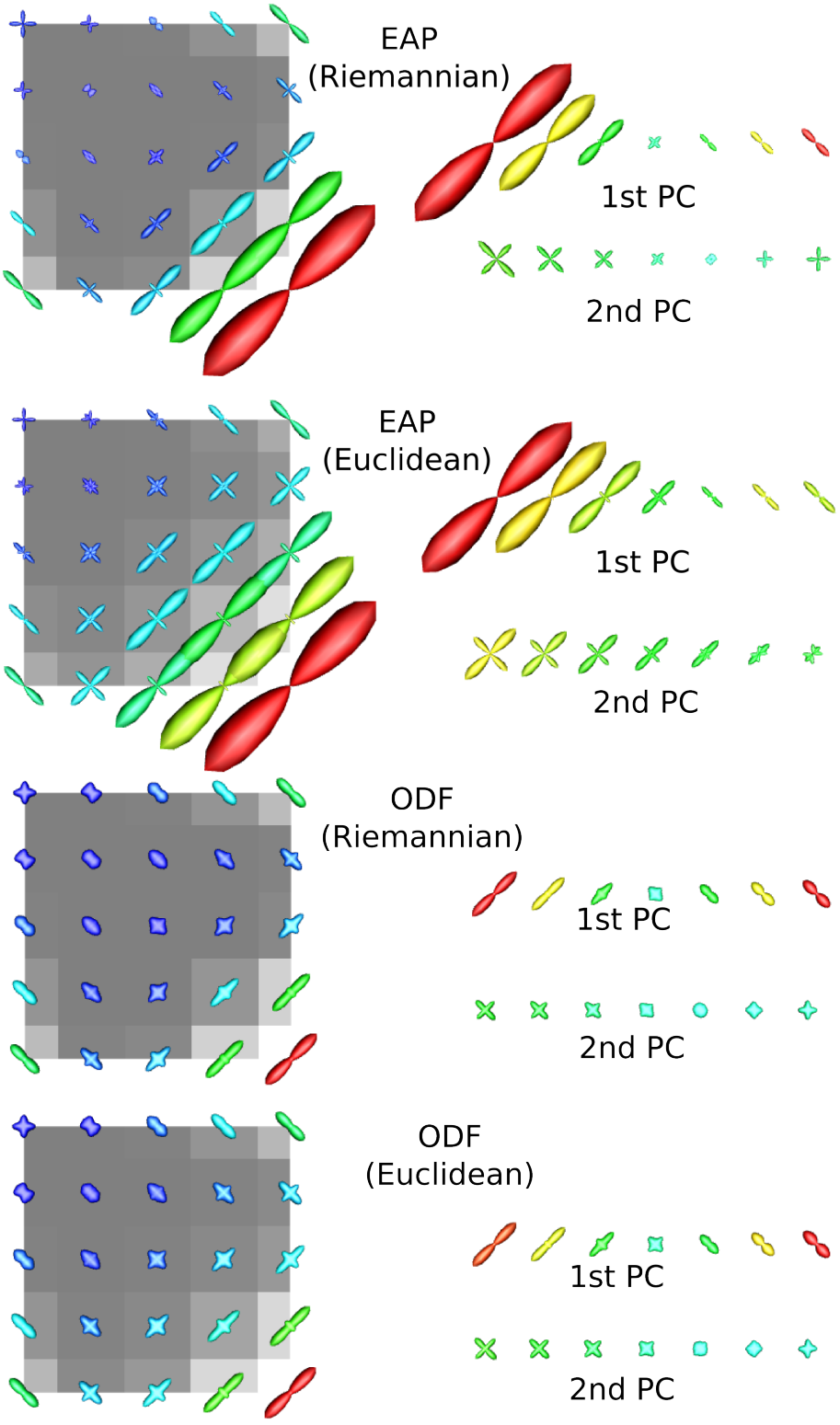


when $R = 15\mu m$, which is totally different from tensor-EAPs. In the top left area of Fig. 6.6, the EAP interpolation with Euclidean metric gives some unreasonable glyphs with many local maxima, while Riemannian metric gives more reasonable results. The PCs of EAPs from Riemannian interpolation show clearly the 3 types of EAPs in 4 corners, while the PCs of EAPs from Euclidean interpolation miss the EAP type in the top left voxel. Euclidean and Riemannian interpolation give similar results for ODF interpolation. PCA and PGA also give similar results for ODFs. Fig. 6.4, 6.5 and 6.6 demonstrate the three important phenomena.

- First, the ODF interpolation results with Riemannian and Euclidean metrics have only a little difference, while the EAP interpolation results with two metrics are quite different.
- Second, ODF/EAP and tensor-ODF/EAP interpolation with Euclidean metric give more weighting for the distribution with large values than the interpolation with Riemannian metric. The phenomenon is more evident in tensor-EAP and EAP interpolation.
- Third, ODFs/EAPs from Riemannian interpolation are very similar with ODFs/EAPs from Log-Euclidean interpolation. Although We did not show the results from Affine-Euclidean interpolation, it is also very similar with Riemannian interpolation.

We would like to explain why the Riemannian metric and Euclidean metric give more different results for EAPs than for ODFs. Please remind the previous discussions in Section 6.1.9 that if two points are near the uniform distribution, the geodesics of Riemannian metric and Euclidean metric will be close. For ODFs, the uniform distribution is also the isotropic ODF. The ODFs in gray matter and CSF are normally close to isotropic ODF. Moreover we find in practice that even in white matter the two kinds of ODFs with or without solid angle defined in [Tuch, 2004; Descoteaux et al., 2007] and [Wedeen et al., 2005; Aganj et al., 2010b] are also very close to the isotropic ODF. If we use SH to represent $\psi(\mathbf{r})$, $\mathbf{r} \in S^2$ for ODFs, the isotropic

Figure 6.6: Comparison between Riemannian and Euclidean metrics in 2D ODF/EAP interpolation. Comparison between PGA and PCA in 2D ODF/EAP field. PGA and PCA were used in the ODF/EAP results from Riemannian and Euclidean interpolations respectively. GA was used to color the glyphs and was set in the background.



ODF is $(1, 0, ..., 0)^T$, i.e. $c_1 = 1$. For a given ODF with the coefficients $\boldsymbol{c}$, c_1 is the projection onto isotropic ODF, which is more than $\frac{1}{\sqrt{4\pi}}$ theoretically as shown in Section 6.1.4. However experimentally c_1 from a typical tensor in our brain is always close to 1, normally more than 0.95. That means the ODFs we focus in human brain are actually always in a small space around the uniform distribution, compared to the large function space in Eq. (6.1). However, that is different for EAPs because the EAPs we study are far from the uniform distribution as we have discussed in Section 6.1.4. This can explain why significant difference occurs between two metrics in interpolation of tensors and EAPs, but only a little difference in ODF interpolation.

Filtering

In this subsection, we would like to compare different metrics in Gaussian filtering. Section 6.2.3 shows some potential filtering methods in ODF/EAP field. The weighting functions in the given neighborhood $N(\boldsymbol{s})$ at spatial position $\boldsymbol{s}$ may be dependent on both spatial distance and PDF distance in bilateral filtering. However, it is difficult to set $\sigma_{\boldsymbol{c}}$ for PDF distance in a fair comparison, because different metrics have different range of distances. Thus we just compare different metrics in traditional Gaussian filtering where the weighting functions are only dependent on spatial distance and the same $\sigma_{\boldsymbol{s}}$ can be used in all metrics.

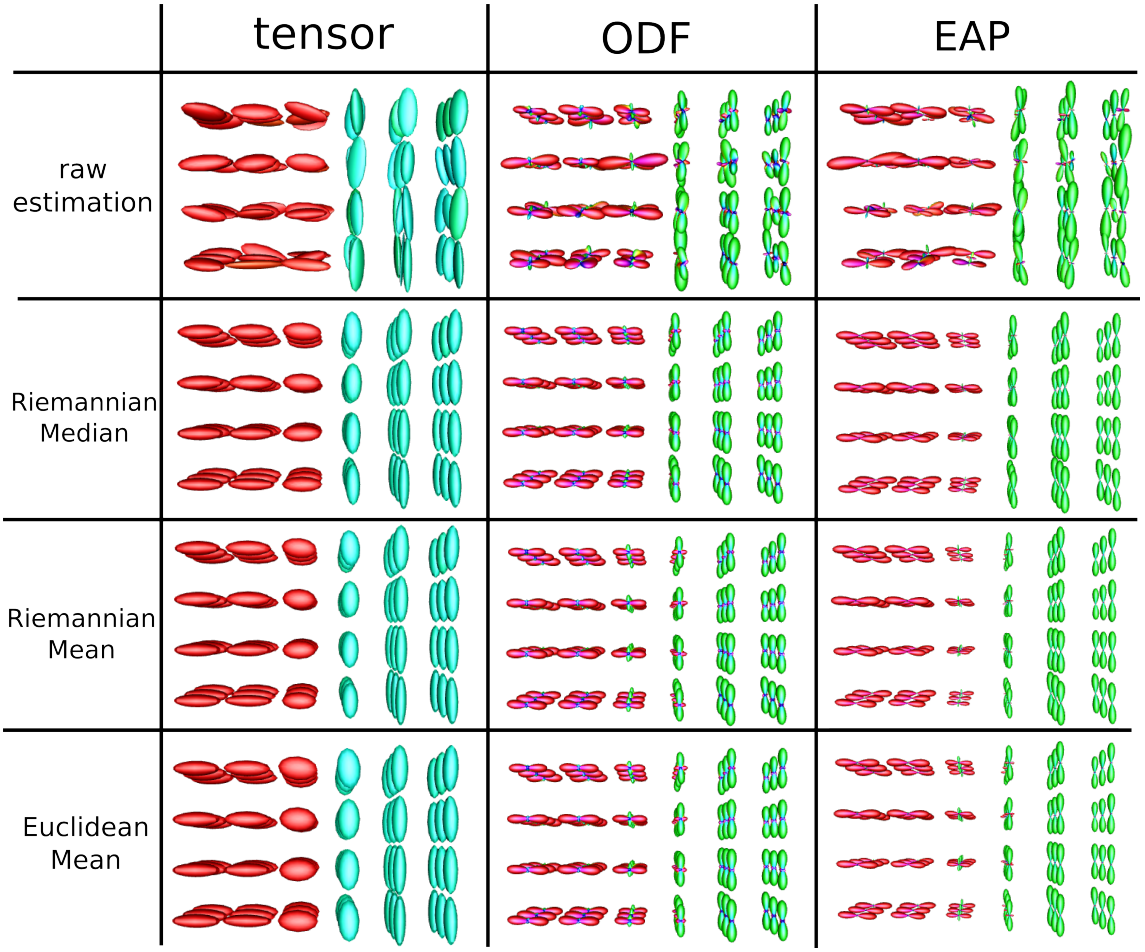
The ground truth signals at $b = 1500 s/mm^2$ were generated from a tensor field where tensors in two regions have the same eigenvalues $[1.7, 0.3, 0.3] \times 10^{-3} mm^2/s$ but orthogonal principal directions. The signals were corrupted by Rician noise with $SNR = 5$, where SNR is defined as $\frac{1}{\sigma}$ and σ is the deviation of complex Gaussian noise [Descoteaux et al., 2007]. Then the single shell noise corrupted signals were used to estimate tensors, ODFs and EAP profiles at $R = 15\mu m$. The tensors were estimated via standard least square fitting [Basser et al., 1994], and the ODFs/EAPs were estimated from the proposed SRPE in 6.2.1 with $L = 4$, $N = 2$, $\lambda_l = \lambda_n = 0$. Please note two issues.

- First, we just used limited samples and did not add any regularization in ODF/EAP estimation because we would like to obtain noisy estimation results for filtering experiments.
- Second, the noisy results in raw estimation come from estimation error of limited samples and Rician noise, not the Gaussian noise in tangent space used in [Pennec et al., 2006; Fillard et al., 2007; Goh et al., 2011] which cannot be found in practical estimation.

The noisy tensor, ODF and EAP fields were listed in the first row of Fig. 6.7. Then Riemannian median, Riemannian mean and Euclidean mean were used in spatial Gaussian filtering, where the common spatial deviation was chosen as $\sigma_{\boldsymbol{s}} = 0.8$.

All three metrics work well inside the regions, and the differences are mainly in the boundary of these two regions. Compared to Riemannian median and mean, the Euclidean mean filtering on tensor field obtain swelling glyphs with less anisotropy around the boundary, which agrees the results in [Pennec et al., 2006; Fillard et al., 2007]. As we discussed above in 6.2.2, the swelling of tensors means decrease of

Figure 6.7: Filtering on tensor, ODF and EAP fields. The first row is the noisy glyphs of raw estimation. The following three rows are filtering results based on Riemannian median, Riemannian mean and Euclidean mean respectively. The glyphs were colored by directions.



probability values. It agrees that the Euclidean mean filtering on EAP field obtains smaller glyphs around the boundary. Euclidean mean filtering on EAP field also gives more interchange for the two regions, and the glyphs around the boundary have more obvious crossing, which decrease the anisotropy. Riemannian median filtering gives the least interchange for the two regions which preserve the discontinuity. The Euclidean mean and Riemannian mean obtain similar results in ODF field, which agrees the discussion above on interpolation.

Robustness of Riemannian Median

In this subsection, we would like to compare the robustness of different estimators, i.e. Riemannian median, mean and Euclidean mean. With the data generation shown in last experiment, we generate tensor, ODF and EAP data with 10 samples from tensor model with eigenvalues $[1.7, 0.3, 0.3] \times 10^{-3} mm^2/s$ and $SNR = 5$ at $b = 1500 s/mm^2$. Then some outliers with different directions were also generated from the same tensor model with $SNR = 5$ but different direction at $b = 1500 s/mm^2$. The first row in

Figure 6.8: Test the robustness of Riemannian median, Riemannian mean and Euclidean mean. First row: 5 of the data from raw estimation. Second row: 5 of outliers. From the third to the fifth row: estimated results from Riemannian median, Riemannian mean and Euclidean mean. Sixth row: the number of outliers in 10 samples.

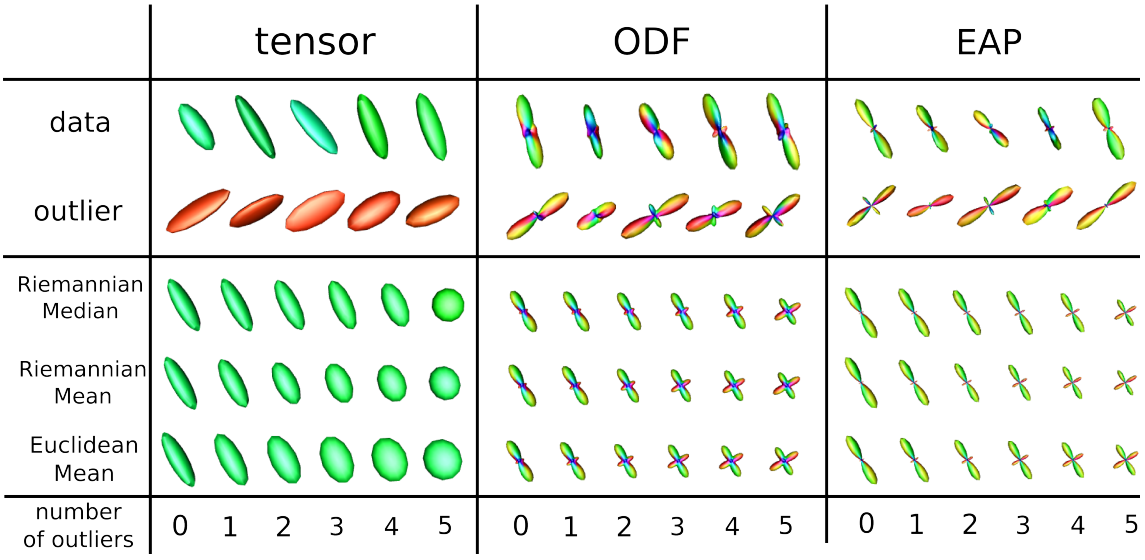


Fig. 6.8 shows the 5 samples of tensors, ODFs and EAPs. The second row shows 5 of the outliers. Then Riemannian median, Riemannian mean and Euclidean mean were estimated from the 10 samples, some of which were replaced by outliers. The results from three estimators were shown from the third to fifth rows in Fig. 6.8, and the number of outliers were listed in the last row. We can see that the Riemannian median is more robust than Riemannian mean and Euclidean mean, which agrees [Fletcher et al., 2009].

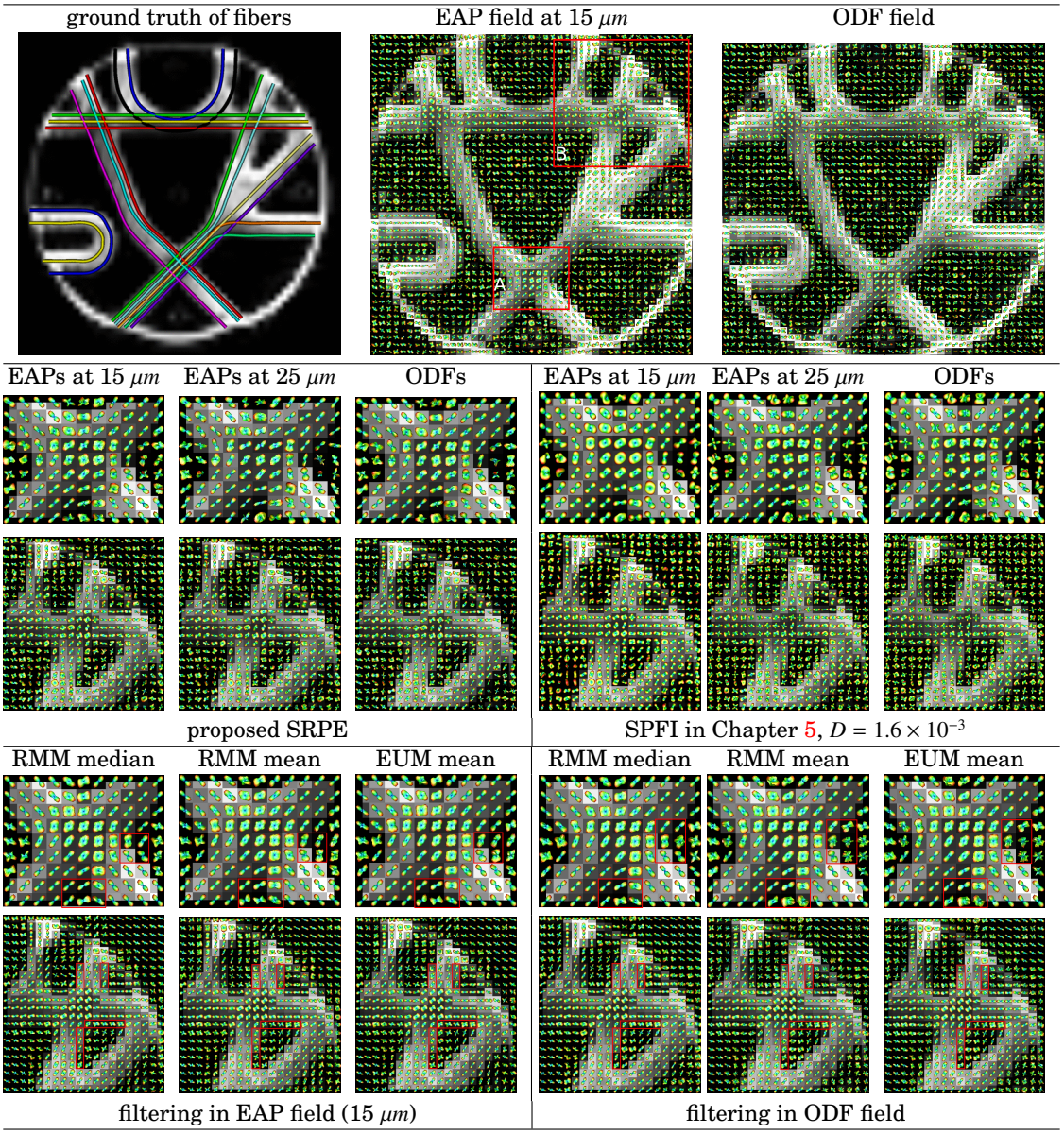
6.3.2 Phantom Data

We performed the proposed SRPE method and Gaussian filtering in a public phantom data (http://lnao.lixium.fr/spip.php?rubrique79). See Appendix B.1 for the details of the data information and preprocessing.

Nonnegative Definite ODF/EAP Estimation

We first obtained the mean DWI data from two repetitions. Then we estimated ODFs and EAPs from the proposed SRPE and SPFI in our previous work [Cheng et al., 2010a,b]. It is possible to consider adaptive scale in SRPE for better estimation, however it will bring troubles in the following Riemannian operations on EAP field because the coefficients in different voxels are under different basis set. Thus we first estimate the mean diffusivity of all voxels in white matter, which is about $D = 1.6 \times 10^{-3} mm^2/s$. Then we use this mean diffusivity value to set the scale $4\tau D = 1.6211 \times 10^{-4}$ in SRPE and the scale ($\frac{1}{8\pi^2\tau D} = 312.5$) in SPFI for all voxels. We set $L = 4$, $N = 1$ for both SRPE and SPFI, and $\lambda_l = \lambda_n = 0$ in order to evaluate the robustness to noise. Fig. 6.9 shows in the first row the ground truth of fiber directions and the estimated

Figure 6.9: Phantom data results. Comparison between SRPE and SPFI. Comparison between different metric in filtering of ODF/EAP field. The reference DWI data without diffusion gradient ($b = 0$) was set as the background in all subfigures. First row: the estimated EAP and ODF glyphs. Second row: the results from two methods in two enlarged regions. Third row: Gaussian filtering results of EAP and ODF fields, where RMM means Riemannian metric and EUM means Euclidean metric. The glyphs in all subfigures were colored by the GFA values of the spherical functions, i.e. EAP profiles and ODFs.



EAP field at 15 μm and ODF field. The reference DWI data without diffusion gradient ($b = 0$) was set as the background color. Since the anisotropy of the phantom data is much low compared to the normal human brains [Fillard et al., 2011], we performed min-max normalization in visualization of all glyphs. The second row visualizes the results by SRPE and SPFI in two enlarged regions. Please note that normally people only visualize the EAPs at 15 μm, because a large R value results in noisy results with many peaks [Özarslan et al., 2006; Cheng et al., 2010b]. However, we found that the proposed SRPE works well even with a large R value. Compared to SPFI which has

very different results for $R = 15\mu m$ and $R = 25\mu m$, SRPE is not very sensitive to R. See the second row for the glyphs of EAPs at 25 μm. It is probably because SRPE naturally considers the nonnegative definite constraint. Note that in Chapter 5, we consider adaptive scale to improve the results, while here we consider the nonnegative definite constraint to improve the results.

ODF/EAP Filtering

We then perform Gaussian filtering on the estimated EAP field at 15 μm and ODF field with three different strategies, i.e. Riemannian median, Riemannian mean and Euclidean mean. The standard deviation of spatial Gaussian kernel was set as 0.8 of the voxel size. Fig. 6.9 shows the smoothed EAP and ODF fields using different strategies. It can be seen that all filtering methods can obtain clearer results compared to the raw estimation both for ODFs and EAPs. For EAPs, Riemannian mean and Euclidean mean obtain similar results where the boundary of the light areas with fibers are more contaminated by glyphs in the dark areas, compared to the result by Riemannian median. See the regions marked in red. For ODFs, Riemannian mean and Euclidean mean obtain visually the same results, which agrees with previous discussions and results in synthetic data. The Riemannian median has better results than Riemannian and Euclidean mean. This experiment shows that Riemannian median is more appropriate in ODF/EAP filtering which can preserve the discontinuity of different regions.

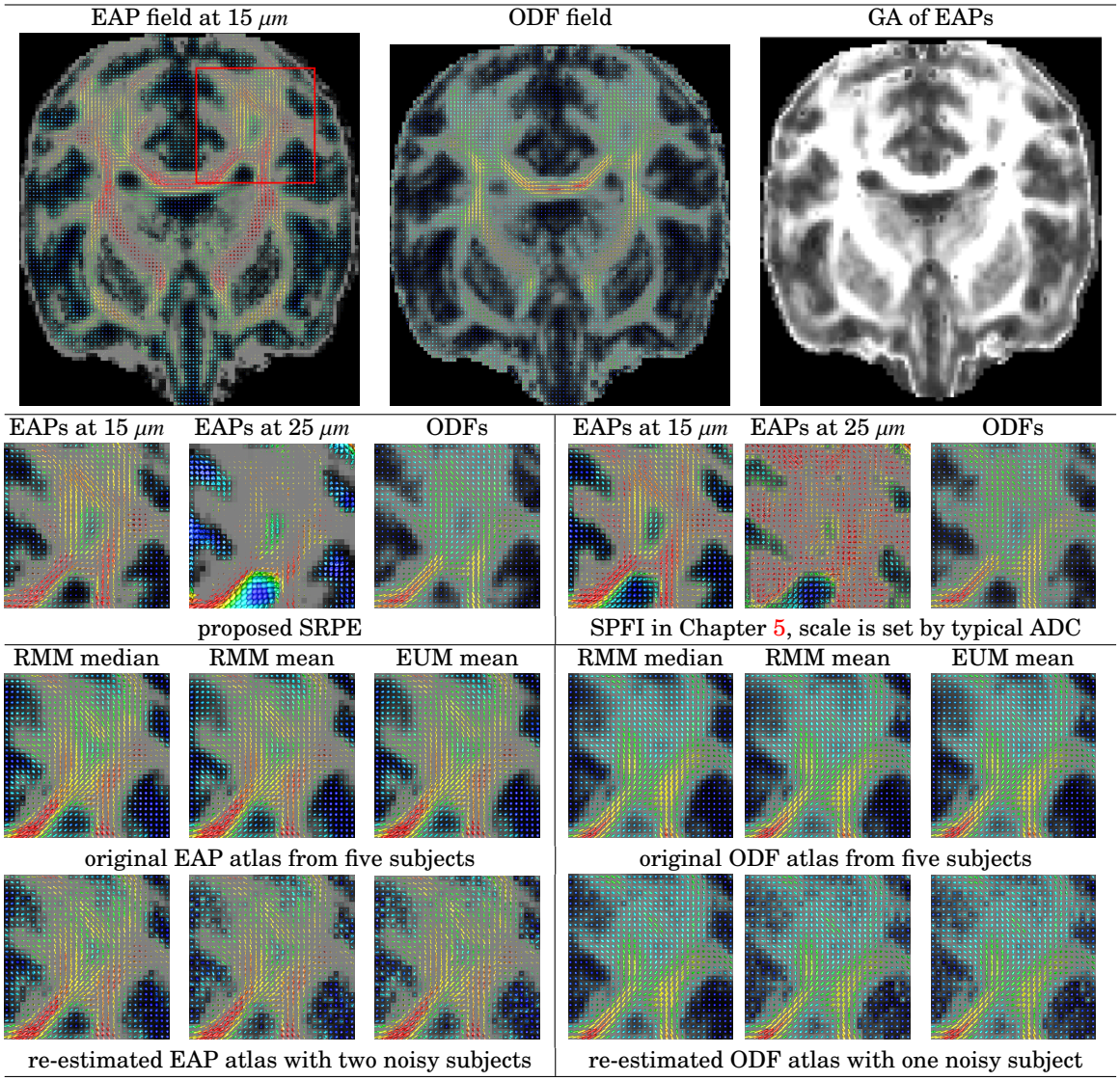
6.3.3 Real Monkey Data

Finally we apply the proposed Riemannian framework methods to a real monkey data. Please refer Appendix B.2 for the data description and preprocessing.

Nonnegative Definite ODF/EAP Estimation

For the first subject of the data, We first estimated the EAPs and ODFs from the DWI data using the proposed SRPE method. In practice, we set $L = 4$, $N = 2$ and we found the results in SRPE almost do not change when λ_l and λ_n range in $[0, 1e-6]$, which agrees the results in the above phantom data. So we set $\lambda_l = \lambda_n = 0$. The first row of Fig. 6.10 shows the estimated EAP and ODF fields, as well as the GA map calculated from the EAP field. Similarly with FA for tensors and GFA for ODFs, the GA map for EAPs has large values in white matter and low values in grey matter and Cerebrospinal fluid (CSF). It may be useful in clinical study. Then we compared the results from SRPE to the results from SPFI in [Cheng et al., 2010a,b]. $L = 4$, $N = 2$ and $\lambda_l = 5e-9$, $\lambda_n = 1e-9$ were set in SPFI which were also used in [Cheng et al., 2010a,b]. Note that we set scale parameter in SPFI shared in each voxel based on the typical ADC D_0 in Eq. (5.32). Fig. 6.10 lists in the second row the estimated EAP fields at $15\mu m$ and $25\mu m$ and ODF fields in an enlarged region. Both SRPE and SPFI obtain anisotropic glyphs in white matter and isotropic glyphs in grey matter and CSF. So they normally do not need an min-max normalization like QBI [Tuch, 2004; Descoteaux et al., 2007] to artificially make anisotropic glyphs. Please note that artificial normalization on glyphs will obtain noisy glyphs in the

Figure 6.10: Monkey data results. First row: estimated EAP, ODF fields and GA map from the first subject. Second row: estimated EAPs and ODFs from two methods in a enlarged region. Third row: original EAP and ODF atlases from 5 subjects via three ways. Fourth row: re-estimated EAP and ODF atlases from 4 original subjects and 1 noisy subject. RMM means Riemannian metric and EUM means Euclidean metric. The glyphs in all subfigures were colored by the GFA values of the spherical functions, i.e. EAP profiles and ODFs.



area with isotropic diffusion like grey matter and CSF, which brings some troubles in the following tractography. Compared to the noisy EAP glyphs at $25\mu m$ in SPFI, the results in SRPE are much cleaner and the CSF area remains isotropic. It is probably because we naturally add the nonnegative constrain.

Atlas Estimation

Registration is crucial in atlas estimation. Since so far there is no common registration method for ODF/EAP data and it is also not our focus here, we just perform a naive way to align the ODF/EAP data using FSL. All DWIs from 5 subjects were

aligned via affine registration to a template made by non-diffusion weighted images. The affine transform was used to re-orientate the gradients for each subject via Finite Strain method [Alexander et al., 2001]. Then we estimated EAPs and ODFs for all five subjects respectively, and calculate the ODF/EAP atlas from the five estimated ODF/EAP fields. The original EAP and ODF atlases are obtained as the Riemannian median, Riemannian mean and Euclidean mean of five EAP and ODF fields, which are shown in the third row of Fig. 6.10. Then we would like to test the robustness of the atlases using different metrics. Rician noise was added with $SNR = 10$ to the DWI data of two subjects, and the noise corrupted ODF/EAP field was re-estimated. The ODF/EAP atlas from three metrics were also re-estimated from the two noisy fields and the other three original ODF/EAP fields. The fourth row of Fig. 6.10 shows that the re-estimated ODF/EAP atlas from three metrics. Riemannian median results in the best atlases similar with the original atlas for both ODFs and EAPs, which validated its robustness. While Riemannian mean and Euclidean mean obtained noisy ODF/EAP atlases. Please note that for ODFs, Riemannian mean and Euclidean mean obtain very similar results. For EAPs, the atlases from Riemannian mean and Euclidean mean are slightly different in the data set.

6.4 SUMMARY

In this chapter, we proposed an intrinsic diffeomorphism invariant Riemannian framework as a mathematical tool for both ODFs and EAPs processing in dMRI, which is a natural extension of previous affine invariant Riemannian framework for tensors. It is based on orthonormal basis representation of the wavefunction. We almost generalized all concepts in the Riemannian framework from tensors to ODFs and EAPs, such as Geodesic Anisotropy (GA), Log-Euclidean framework, affine invariance, swelling effect, etc. Based on this Riemannian framework, we proposed some potential applications, such as nonnegative ODF/EAP estimation via SRPE, ODF/EAP interpolation, filtering, atlas estimation. It can also be used in other applications, not just limited to the applications we have demonstrated in this chapter. Like the Riemannian framework for tensors, the Riemannian framework for ODFs and EAPs is a set of the general mathematical tools which can be used for design of general algorithms.

Compared to the mixture model formulation of Riemannian framework in [McGraw et al., 2006], our formulation is model-free and needs very weak assumption that the wavefunction of the ODF/EAP we focus is contained in the linear span of chosen basis. Compared to the histogram formulation in [Goh et al., 2009a, 2011], our formulation considers orthonormal basis representation which is more efficient for the following Riemannian operations, because we work in the space of limited number of coefficients which has low dimension compared to the histogram. The analysis on piecewise constant basis in Section 6.1.1 and Theorem 6.1 demonstrate that the formulation in [Goh et al., 2009a, 2011] can be seen as a specific case of our formulation when the orthonormal basis is chosen as the piecewise constant basis. As we have shown in this chapter as well as chapter 5 and [Cheng et al., 2011c], the EAPs and ODFs we study in human brains are only in a small space, compared to the general PDF function space shown in 3.35. We have some priors that the DWI signals and ODFs/EAPs are smooth enough such that they can be sparsely represented by

some orthonormal bases.

We now summarize the contributions of this chapter.

1. We proposed the intrinsic model-free Riemannian framework based on wavefunction representation by orthonormal basis. This Riemannian framework can be used in arbitrary PDFs under arbitrary orthonormal basis representation. Theorem 6.1 shows that the Riemannian framework under different orthonormal basis representation obtains theoretically the same and numerically similar final results. In dMRI, this Riemannian framework avoids Gaussian assumption in DTI and can be used to process both ODF and EAP data.

2. In this framework, the geodesic, exponential map and logarithmic map have closed forms. We proposed four properties of the parameter space PS, and noticed the similarities and differences between the space of EAPs and the space of ODFs.

3. The weighted Riemannian mean and median were proved to uniquely exist in the parameter space PS and can be estimated efficiently by gradient descent method.

4. Two efficient frameworks, Affine-Euclidean framework and Log-Euclidean framework, are proposed to approximate the original Riemannian framework and fast process data.

5. We generalize Geodesic Anisotropy from tensors to general ODFs/EAPs, which is defined as the geodesic distance from the given ODF/EAP to its nearest ODF/EAP. Theorem 6.2 showed that GA can be calculated analytically from the given Riemannian coordinate. We also showed how to calculate Rényi entropy from given wavefunction. However, Rényi entropy seems less useful for ODFs and EAPs based on the experiments in [Cheng et al., 2009a] and the analysis in Section 6.1.5.

6. We proposed a nonnegative definite ODF/EAP estimation called Square Root Parameterized Estimation (SRPE), which can guarantee the estimated ODFs and EAPs are nonnegative definite in their domains, i.e. $\mathbb{S}^2$ for ODFs and $\mathbb{R}^3$ for EAPs. While the previous methods [Goh et al., 2009b; Özarslan et al., 2009] only add nonnegative constraint in some given samples. The experiments showed that SRPE works well even in large R value, compared to SPFI in Chapter 5. However, compared to SPFI, SRPE has no closed form solution, which needs a gradient descent method. Compared to ℓ_1-SPFI [Cheng et al., 2011c], the cost function in SRPE is not convex, which means the solution of SRPE is a local minimum. But, when we choose the typical isotropic EAP as the initialization, the estimated local minimum from SRPE normally gives a good result. It once again demonstrates that the EAPs we study in practice are in a small space inside the general probability space.

7. We theoretically analyzed the property of weighted Riemannian mean and median. We proved in Theorem 6.4 that the probability value of the weighted mean/median PDF is not less than the probability value of at least one sampled PDF (ODF, EAP, or Gaussian distribution in DTI) which is used to generate the weighted mean/median. This property directly generalize the so called

swelling effect to *unreasonable mean value interpolation* for general PDFs like ODFs and EAPs. For Gaussian distributions parameterized by tensors, Theorem 6.6 shows the probability value of the weighted mean/median Gaussian distribution is between the minimal value and maximal value of the sampled Gaussian distributions which are used to generate the weighted mean/median.

8. We analyzed theoretically and experimentally the differences between Riemannian framework for tensors and for ODFs and EAPs, the differences between Riemannian metric and Euclidean metric. The Riemannian metric and Euclidean metric seem to obtain similar results for ODFs, but different results for tensors and EAPs.

9. We proposed to use Riemannian median in ODF/EAP atlas estimation. The experiments showed that the Riemannian median is more robust than Riemannian mean and Euclidean mean.

10. The filtering experiments demonstrated that the Riemannian median is more appropriate in filtering.

6.5 APPENDIX A: PROOF ON THE WAVEFUNCTION OF THE NEAREST ISOTROPIC EAP

Proof of Theorem 6.2 on the Wavefunction of the Nearest Isotropic EAP.
The wavefunction of an isotropic EAP can be represented as

$$\psi_{iso}(R\mathbf{r}) = F(R)Y_0^0(\mathbf{r}), \quad \text{where } F(R) \geq 0, \forall R \geq 0.$$

It is in the linear function space $\int_{\mathbb{S}^2} \psi(\mathbf{R})Y_l^m(\mathbf{r})\mathrm{d}\mathbf{r} = 0, \forall l > 0$, denoted by L_{iso}. Please see Fig. 6.11 for the sketch map. It is also in the nonnegative half-plane $\int_{\mathbb{R}^3} \psi(\mathbf{R})\mathrm{d}\mathbf{R} \geq 0$, denoted by P^+. Let S_P denote the high dimensional sphere $\int_{\mathbb{R}^3} \psi(\mathbf{R})^2\mathrm{d}\mathbf{R} = 1$. Then the space of the wavefunctions of isotropic EAPs, denoted by P_{iso}, is

$$P_{iso} = P^+ \cap L_{iso} \cap S_P$$

For any wavefunction $\psi(\mathbf{R})$, its projection onto L_{iso}, denoted by $\psi_{L_{iso}}(\mathbf{R})$, is

$$\psi_{L_{iso}}(\mathbf{R}) = \left(\int_{\mathbb{S}^2} \psi(\mathbf{R})Y_0^0(\mathbf{r})\mathrm{d}\mathbf{r} \right) Y_0^0(\mathbf{r})$$

The normalized version of $\psi_{L_{iso}}(\mathbf{R})$ is $\psi_{S_{iso}}(\mathbf{R}) = \frac{\psi_{L_{iso}}(\mathbf{R})}{\|\psi_{L_{iso}}(\mathbf{R})\|}$. Since $\psi(\mathbf{R}) \geq 0$ and $Y_0^0 = \frac{1}{\sqrt{4\pi}} > 0$, we have $\psi_{S_{iso}}(\mathbf{R}) \in P_{iso}$. The Riemannian distance, i.e. the length of the arc, between $\psi(\mathbf{R})$ and $\psi_{S_{iso}}(\mathbf{R})$ is

$$d_{arc}(\psi(\mathbf{R}), \psi_{S_{iso}}(\mathbf{R})) = \arccos\left(\int_{\mathbb{R}^3} \psi(\mathbf{R})\psi_{S_{iso}}(\mathbf{R})\mathrm{d}\mathbf{R} \right)$$

and the distance of the chord between two wavefunctions is

$$d_{chord}(\psi(\mathbf{R}), \psi_{S_{iso}}(\mathbf{R})) = \|\psi(\mathbf{R}) - \psi_{S_{iso}}(\mathbf{R})\| = \sqrt{2 - 2\cos(d_{arc}(\psi(\mathbf{R}), \psi_{S_{iso}}(\mathbf{R})))}$$

which is a monotonic function of $d_{arc}(\psi(\mathbf{R}), \psi_{S_{iso}}(\mathbf{R}))$. For any isotropic wavefunction $\psi'(\mathbf{R}) \in P_{iso}$ and $\psi'(\mathbf{R}) \neq \psi_{S_{iso}}(\mathbf{R})$, because of triangle inequality we have

$$\|\psi'(\mathbf{R}) - \psi_{L_{iso}}(\mathbf{R})\| > 1 - \|\psi_{L_{iso}}(\mathbf{R})\| = \|\psi_{S_{iso}}(\mathbf{R}) - \psi_{L_{iso}}(\mathbf{R})\|$$

Then

$$\begin{aligned} d_{chord}(\psi'(\mathbf{R}), \psi(\mathbf{R}))^2 &= \|\psi'(\mathbf{R}) - \psi_{L_{iso}}(\mathbf{R})\|^2 + \|\psi(\mathbf{R}) - \psi_{L_{iso}}(\mathbf{R})\|^2 \\ &> \|\psi_{S_{iso}}(\mathbf{R}) - \psi_{L_{iso}}(\mathbf{R})\|^2 + \|\psi(\mathbf{R}) - \psi_{L_{iso}}(\mathbf{R})\|^2 \\ &= d_{chord}(\psi_{S_{iso}}(\mathbf{R}), \psi(\mathbf{R}))^2 \end{aligned}$$

So $d_{arc}(\psi'(\mathbf{R}), \psi(\mathbf{R})) > d_{arc}(\psi_{S_{iso}}(\mathbf{R}), \psi(\mathbf{R}))$. □

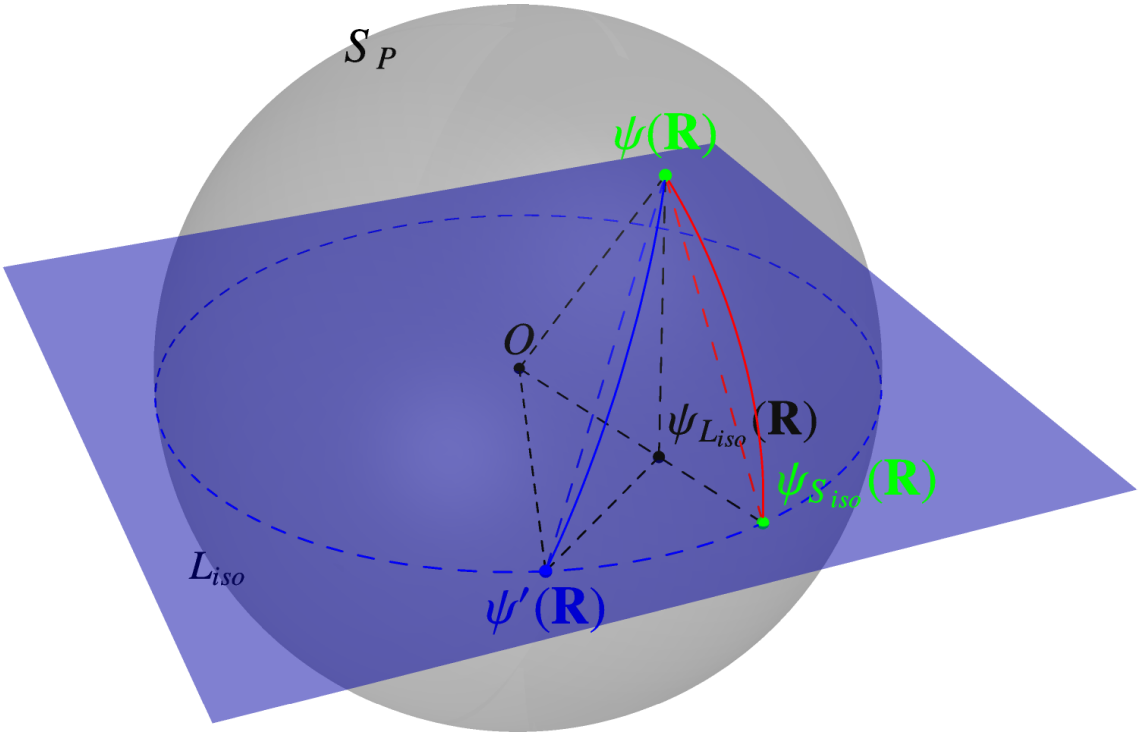


Figure 6.11: Sketch map for the wavefunction of the nearest EAP.

6.6 Appendix B: Proofs on Weighted Riemannian Mean and Median

Proof of Theorem 6.3. Theorem 3.2 [Buss and Fillmore, 2001] proved that the weighted Riemannian mean on a hemisphere exists uniquely if at least one point in the interior of the hemisphere, which also can be proved directly based on Theorem 3.1. [Buss and Fillmore, 2001] also proposed Algorithm 6.1 to estimate the weighted mean. So the weighted Riemannian mean uniquely exists in our parameter space because PS_K is contained in a convex cone with angle 90^o as we have shown in Section 6.1.4. Please note normally for gradient descent method, we need to choose a step size γ_k in each step k for convergence. However based on the theoretical results

in [Buss and Fillmore, 2001], $\gamma_k = 1$ ensures the convergence of Algorithm 6.1. Normalized Euclidean mean is chosen as the initialization, which makes the methods converge fast.

Because PS_K is a convex set in $\mathbb{S}^{K-1}$ and contained in a convex cone with angle 90^o, the sectional curvature of the PS_K is 1 and $diam(U) < \frac{\pi}{2}$. So Theorem 3.3 with $\Delta = 1$ proves the uniqueness and existence of Riemannian median. [Fletcher et al., 2009] used a gradient descent method to calculate the Riemannian median as listed in Algorithm 6.2. Theorem 3.5 [Fletcher et al., 2009] also proved that the step size $\gamma_k \in [0, 2]$ can ensure convergence when sectional curvatures of the manifold are nonnegative and $\text{diam}(U) \cdot \sqrt{\Delta} < \frac{\pi}{2}$. Thus $\gamma_k = 1$ is fixed in in Algorithm 6.2. Normalized Euclidean mean is chosen as the initialization for acceleration, considering Euclidean mean is an approximation of Riemannian median. In practice, Algorithm 6.1 and 6.2 converge in about 10 steps, if the normalized Euclidean mean is set as the initialization. □

Lemma 6.1. *If the Riemannian distance function in Eq. (6.24) is convex, the following two statements are equivalent:*

(a) $\boldsymbol{\mu_w}$ *(or* $\boldsymbol{m_w}$*) is the weighted Riemannian mean (or median) estimated from* $\{\boldsymbol{c}^{(i)}\}_{i=1}^N$ *and* $\boldsymbol{w}$*;*

(b) $\boldsymbol{0}$ *is the weighted Euclidean mean (or median) estimated from* $\{Log_{\boldsymbol{\mu_w}}(\boldsymbol{c}^{(i)})\}_{i=1}^N$ *(or* $\{Log_{\boldsymbol{m_w}}(\boldsymbol{c}^{(i)})\}_{i=1}^N$*) and* $\boldsymbol{w}$ *in Euclidean space.*

Proof. Consider the following two distance function

$$f_{Rie}(\boldsymbol{c}) = \sum_{i=1}^N w_i d(\boldsymbol{c}, \boldsymbol{c}^{(i)})^2, \quad f_{Euc}(\boldsymbol{t}) = \sum_{i=1}^N w_i \|Log_{\boldsymbol{\mu_w}}(\boldsymbol{c}^{(i)}) - \boldsymbol{t}\|^2$$

When $d(\boldsymbol{c}, \boldsymbol{c}^{(i)})$ is convex with respect to $\boldsymbol{c}$, $\boldsymbol{\mu_w}$ uniquely exists such that $\nabla_{\boldsymbol{c}} f_{Rie}(\boldsymbol{\mu_w}) = -2\sum_{i=1}^N w_i Log_{\boldsymbol{\mu_w}}(\boldsymbol{c}^{(i)}) = \boldsymbol{0}$, which is equivalent to $\nabla_{\boldsymbol{t}} f_{Euc}(\boldsymbol{0}) = -2\sum_{i=1}^N w_i Log_{\boldsymbol{\mu_w}}(\boldsymbol{c}^{(i)}) = \boldsymbol{0}$. The same proof can be used in Riemannian median. Thus (a) and (b) are equivalent. □

Lemma 6.2. *For given N points* $\{\boldsymbol{c}^{(i)}\}_{i=1}^N$ *in* PS_K *in Eq. (6.2) and weighting vector* $\boldsymbol{w}$*, weighted Riemannian mean* $\boldsymbol{\mu_w}$ *and median* $\boldsymbol{m_w}$ *defined in Eq. (6.24) are contained in the convex hull C of* $\{\boldsymbol{c}^{(i)}\}_{i=1}^N$*, where C is the intersection of all the hemispheres that contain* $\{\boldsymbol{c}^{(i)}\}_{i=1}^N$ *and is equivalent to the set of all weighted Riemannian mean with all possible* $\boldsymbol{w}$*.*

Proof. Theorem 7 in [Buss and Fillmore, 2001] proved that the convex hull C of $\{\boldsymbol{c}^{(i)}\}_{i=1}^N$ is equivalent to the set of all weighted Riemannian mean with all possible $\boldsymbol{w}$. Thus C contains $\boldsymbol{\mu_w}$. We would like to prove that C also contains $\boldsymbol{m_w}$. By projecting $\{\boldsymbol{c}^{(i)}\}_{i=1}^N$ onto the tangent space of $\boldsymbol{m_w}$ denoted by $T_{\boldsymbol{m_w}}$. Let C' denote the convex hull which contains all $\{Log_{\boldsymbol{\mu_w}}(\boldsymbol{c}^{(i)})\}_{i=1}^N$ in $T_{\boldsymbol{m_w}}$. Please note C' is not the image of C. If $\boldsymbol{m_w}$ is not in C, there exists a hemisphere H with $\boldsymbol{m_w}$ in its boundary which does not contain any $\boldsymbol{c}^{(i)}$. Then the closed half-plane in $T_{\boldsymbol{m_w}}$ with the image of the boundary of H as its boundary contains $\boldsymbol{0}$ in its boundary while it does not contain any $Log_{\boldsymbol{\mu_w}}(\boldsymbol{c}^{(i)})$. That means $\boldsymbol{0}$ is not in the convex hull of $\{Log_{\boldsymbol{\mu_w}}(\boldsymbol{c}^{(i)})\}_{i=1}^N$. Then let $\boldsymbol{t_0}$ be the projection of $\boldsymbol{0}$ onto the convex hull. It is clear that $\|\boldsymbol{t_0} - Log_{\boldsymbol{\mu_w}}(\boldsymbol{c}^{(i)})\| < \|\boldsymbol{0} - Log_{\boldsymbol{\mu_w}}(\boldsymbol{c}^{(i)})\|$, for

$i \in \{1, 2, ..., N\}$, which contradicts with the result in Lemma 6.1 that $\boldsymbol{0}$ is the weighted Euclidean median which minimizes the function $\sum_{i=1}^{N} \|t - Log_{\boldsymbol{m_w}}(\boldsymbol{c}^{(i)})\|$. □

Lemma 6.3. *Let $\boldsymbol{b}$ be any given vector in $\mathbb{R}^K$. Let $\boldsymbol{b}^\perp$ denote the plane $\boldsymbol{b}^T\boldsymbol{c} = 0$, and $\boldsymbol{b}^+$ denote the half-plane $\boldsymbol{b}^T\boldsymbol{c} > 0$. For any given geometrically convex hull $C \subset \mathbb{S}^{K-1} \cap \boldsymbol{b}^+$, the function $f(\boldsymbol{c}) = \boldsymbol{b}^T\boldsymbol{c}$ is strictly concave in C. For any given geodesic $C_{arc} \in C$, let $\boldsymbol{c}^*$ be the projection of $\boldsymbol{b}$ onto the great circle $\mathbb{S}^1$ which contains C_{arc}, where the arc from $\boldsymbol{b}$ to $\boldsymbol{c}^*$ is orthogonal to the great circle at $\boldsymbol{c}^*$. Then $f(\boldsymbol{c})$ takes its maximum at $\boldsymbol{c}^*$ if $\boldsymbol{c}^* \in C_{arc}$, or at one of the endpoints of C_{arc} if $\boldsymbol{c}^* \notin C_{arc}$.*

Proof. Let $\boldsymbol{c}_{arc}^{(1)}$ and $\boldsymbol{c}_{arc}^{(2)}$ be two endpoints of C_{arc}. The geodesic C_{arc} can be represented as

$$\boldsymbol{c}(t) = \boldsymbol{c}^{(1)}\cos(t) + \boldsymbol{v}\sin(t) \quad \text{where} \quad \boldsymbol{v} = \frac{Log_{\boldsymbol{c}_{arc}^{(1)}}(\boldsymbol{c}_{arc}^{(2)})}{\|Log_{\boldsymbol{c}_{arc}^{(1)}}(\boldsymbol{c}_{arc}^{(2)})\|}, \; t \in [0, \|Log_{\boldsymbol{c}_{arc}^{(1)}}(\boldsymbol{c}_{arc}^{(2)})\|]$$

Thus we have

$$\begin{aligned} f(\boldsymbol{c}(t)) &= \boldsymbol{b}^T\boldsymbol{c}^{(1)}\cos(t) + \boldsymbol{b}^T\boldsymbol{v}\sin(t) \\ \frac{\partial f(\boldsymbol{c}(t))}{\partial t} &= \boldsymbol{b}^T\boldsymbol{c}^{(1)}\cos(t + \frac{\pi}{2}) + \boldsymbol{b}^T\boldsymbol{v}\sin(t + \frac{\pi}{2}) \\ \frac{\partial^2 f(\boldsymbol{c}(t))}{\partial t^2} &= -\boldsymbol{b}^T\boldsymbol{c}^{(1)}\cos(t) - \boldsymbol{b}^T\boldsymbol{v}\sin(t) = -\boldsymbol{b}^T\boldsymbol{c}(t) < 0 \end{aligned}$$

Note $\frac{\partial^2 f(\boldsymbol{c}(t))}{\partial t^2} = -\boldsymbol{b}^T\boldsymbol{c}(t) < 0$ because $C_{arc} \subset \mathbb{S}^{K-1} \cap \boldsymbol{b}^+$. That means $f(\boldsymbol{c}(t))$ is strictly concave in C_{arc}. Since C_{arc} is any given geodesic in C, $f(\boldsymbol{c}(t))$ is strictly concave in C. Assume $f(\boldsymbol{c}(t))$ has the maximum at $\boldsymbol{c}^* = \boldsymbol{c}(t^*)$ in the great circle $\mathbb{S}^1$ that contains C_{arc}, then

$$\frac{\partial f(\boldsymbol{c}(t))}{\partial t}|_{t=t^*} = \boldsymbol{b}^T\boldsymbol{c}^{(1)}\cos(t^* + \frac{\pi}{2}) + \boldsymbol{b}^T\boldsymbol{v}\sin(t^* + \frac{\pi}{2}) = 0$$

So $\boldsymbol{c}(t^* + \frac{\pi}{2})$ is orthogonal to $\boldsymbol{b}$. Then $\boldsymbol{c}(t^* + \frac{\pi}{2})$ is orthogonal to the plane determined by $\boldsymbol{b}$ and $\boldsymbol{c}(t^*)$. Let $\boldsymbol{v}_{\boldsymbol{c}^*}^{arc}$ be the tangent vector for $\mathbb{S}^1$ at $\boldsymbol{c}(t^*)$ and let $\boldsymbol{v}_{\boldsymbol{c}^*}^{\boldsymbol{b}}$ be the tangent vector for the arc from $\boldsymbol{b}$ to $\boldsymbol{c}^*$ at $\boldsymbol{c}(t^*)$. Then $\boldsymbol{c}(t + \frac{\pi}{2})$ is parallel to $\boldsymbol{v}_{\boldsymbol{c}^*}^{arc}$ and orthogonal to $\boldsymbol{v}_{\boldsymbol{c}^*}^{\boldsymbol{b}}$. So $\boldsymbol{c}^*$ is the projection of $\boldsymbol{b}$ onto the submanifold $\mathbb{S}^1$. Then the maximum value is at $\boldsymbol{c}^*$ if $\boldsymbol{c}^* \in C_{arc}$, or at one of the endpoints of C_{arc} if $\boldsymbol{c}^* \notin C_{arc}$ since $f(\boldsymbol{c})$ is concave in C_{arc}. □

Proof of Theorem 6.4. Let C be the intersection of all the hemispheres that contain $\{\boldsymbol{c}^{(i)}\}_{i=1}^N$. Then $\boldsymbol{\mu_w}$ and $\boldsymbol{m_w}$ are in C based on Lemma 6.2. For any given $\boldsymbol{x} \in \chi$,

$$\sqrt{p(\boldsymbol{x}|\boldsymbol{c}^{(i)})} = \sum_j c_j^{(i)} B_j(\boldsymbol{x}) = \boldsymbol{b}^T\boldsymbol{c}^{(i)} \geq 0, \quad \text{where} \quad \boldsymbol{b} = (B_1(\boldsymbol{x}), ..., B_K(\boldsymbol{x}))^T$$

Thus C is in the closed half-plane $\boldsymbol{b}^T\boldsymbol{c} \geq 0$, i.e. $\boldsymbol{b}^+ \cup \boldsymbol{b}^\perp$. Then based on Lemma 6.3, $\sqrt{p(\boldsymbol{x}|\boldsymbol{c})} = \boldsymbol{b}^T\boldsymbol{c}$ is strictly concave with respect to $\boldsymbol{c}$ in $C \cap \boldsymbol{b}^+$, and $\sqrt{p(\boldsymbol{x}|\boldsymbol{c})} = 0$ in $C \cap \boldsymbol{b}^\perp$, which means $\sqrt{p(\boldsymbol{x}|\boldsymbol{c})} = \boldsymbol{b}^T\boldsymbol{c} \geq \min\{\sqrt{p(\boldsymbol{x}|\boldsymbol{c}^{(i)})}\}_{i=1}^N$. So $p(\boldsymbol{x}|\boldsymbol{\mu_w}) \geq \min\{p(\boldsymbol{x}|\boldsymbol{c}^{(i)})\}_{i=1}^N$, and $p(\boldsymbol{x}|\boldsymbol{m_w}) \geq \min\{p(\boldsymbol{x}|\boldsymbol{c}^{(i)})\}_{i=1}^N$. □

Proof of Theorem 6.6. Riemannian mean value interpolation on general EAPs/ODFs is reasonable, which is straightforward based on Theorem 6.4. Euclidean mean value interpolation on general EAPs/ODFs is also reasonable because $p_{\boldsymbol{s}}(\boldsymbol{x}|\boldsymbol{c}) = \sum_i w_i p_{\boldsymbol{s}^{(i)}}(\boldsymbol{x}|\boldsymbol{c}^{(i)}) \geq \min\{p_{\boldsymbol{s}^{(i)}}(\boldsymbol{x})\}_{i=1}^{N}$.

We will prove the result for Riemannian mean value interpolation for given Gaussian EAPs $\{N_{\boldsymbol{s}^{(i)}}(\mathbf{R}|\Sigma_i)\}_{i=1}^{N_s}$ and weighting vector $\boldsymbol{w}$, where $N(\mathbf{R}|\Sigma)$ is the Gaussian distribution with zero mean and covariance $\Sigma = 2\tau\mathbf{D}$ in Eq. (4.27). Let C be the geodesically convex hull of $\{\Sigma_i\}_{i=1}^{N_s}$, then $\boldsymbol{\mu}_{\boldsymbol{w}} \in C$. Let $\Sigma_0 \in C$ and $\Sigma(t)$ be a geodesic, $\Sigma(0) = \Sigma_0$, then

$$\ln N(\mathbf{R}|\Sigma(t)) = -\frac{3}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma(t)| - \frac{1}{2}\mathbf{R}^T\Sigma(t)^{-1}\mathbf{R}$$

and

$$\frac{\partial \ln N(\mathbf{R}|\Sigma(t))}{\partial t} = -\frac{1}{2}\mathrm{Trace}(\Sigma(t)^{-1}\Sigma(t)') + \frac{1}{2}\mathrm{Trace}(\Sigma(t)^{-1}\mathbf{R}\mathbf{R}^T\Sigma(t)^{-1}\Sigma(t)')$$

Considering the inner product $\langle A, B\rangle_{\Sigma(t)} = \mathrm{Trace}(\Sigma(t)^{-1}A\Sigma(t)^{-1}B)$ shown in Eq. (3.46), we have

$$\nabla_{\Sigma} \ln N(\mathbf{R}|\Sigma) = -\frac{1}{2}\Sigma + \frac{1}{2}\mathbf{R}\mathbf{R}^T$$

and

$$\nabla_{\Sigma}^2 \ln N(\mathbf{R}|\Sigma) = -\frac{1}{2}\boldsymbol{I}$$

which is negative definite. So $\ln N(\mathbf{R}|\Sigma)$ is concave in Sym_3^+. So the minima of $\ln N(\mathbf{R}|\Sigma)$ in C are in the boundary. That is

$$N(\mathbf{R}|\boldsymbol{\mu}_{\boldsymbol{w}}) \geq \min\{N(\mathbf{R}|\boldsymbol{c}^{(i)})\}_{i=1}^{N}$$

We also have

$$N(\mathbf{R}|\boldsymbol{\mu}_{\boldsymbol{w}}) \leq \max\{N(\mathbf{R}|\boldsymbol{c}^{(i)})\}_{i=1}^{N}$$

because the global maximum satisfying $\nabla_{\Sigma} \ln N(\mathbf{R}|\Sigma) = 0$ is obtained when $\Sigma = \mathbf{R}\mathbf{R}^T$ that is in the boundary of Sym_3^+, not in C. □

Part III

Conclusion

CHAPTER 7

CONCLUSION

"Every end is just a new beginning. "

Contributions

In this thesis, after introduction of some background knowledge on dMRI and necessary mathematical materials, we have first proposed the analytical solution for EAP, ODFs and several scalar indices in SPFI. SPFI considers complete 3D basis with SH basis in spherical part and Gaussian-Laguerre function in radial part, which can be seen as a generalization of QBI with only SH basis in spherical part. Then we proposed a novel AFT-SC framework to incorporate and compare sHARDI and mHARDI methods. We proposed some useful criteria for evaluating different basis functions in different methods. Next, we proposed an intrinsic diffeomorphism invariant Riemannian framework for both ODF and EAP computing in dMRI, which is a natural extension of previous Riemannian framework for tensors. The Riemannian framework was implemented for ODFs using SH basis and was implemented for EAPs using SPF basis. We generalized previous Log-Euclidean framework, GA from tensors to ODFs and EAPs. We also proposed several possible applications based on the Riemannian framework. In the interpolation of ODF/EAP field, we proposed the definition of reasonable mean value interpolation of general PDF data, which is a generalization of previous concept of swelling effect when interpolating tensor data.

Throughout the thesis, we have enumerated our major contributions at the end of chapters. In summary, the important and original contributions of the thesis are:

1. Analytical linear transforms for EAP, ODFs, GFA, MSD, RTO, from the estimated SPF coefficients of the signal. This is the main contribution we have done for SPFI.
   (a) Proposition 5.1 and 5.2 demonstrated that when the signal is represented by SPF basis, the EAP can be analytically represented by the same coefficients under dSPF basis. dSPF basis was proved to be an orthonormal basis.
   (b) Proposition 5.3 and 5.4 showed that when the signal is represented by SPF basis, the two kinds of ODFs can be analytically represented by SH basis and the coefficients of SH basis can be obtained from a linear transform.

(c) Subsection 5.1.2 proposed the analytical form for RTO, MSD and GFA. The proposed GFA for EAPs is defined as the normalized ℓ_2 norm from the EAP to its nearest EAP, which is a generalization of previous FA for tensors and GFA for spherical functions like ODFs and EAP profiles.

2. Implementation of SPFI with scalar estimation and consideration of the prior $E(0) = 1$.

   (a) An appropriate scale is needed in SPF basis in estimation. We proposed two ways for scale estimation. One is to set scale based on the typical ADC, which makes every voxel share the same scale. It makes the following operations on estimated EAP field more computable. The other one is to set scale by fitting the GHOT model, which was proved in Theorem 5.1 to be the optimal if the signal is represented by GHOT model. It has better performance in EAP/ODF estimation, but the following operations on the estimated EAP field need to consider the basis effect because different voxel use different basis with different scale.

   (b) We proposed to consider $E(0) = 1$ in the estimation, which can largely improve the estimation results. It removes some coefficients in estimation because they are dependent on other coefficients. Thus it can accelerate the estimation.

   (c) The least square estimation can be seen in Table 5.1.

   (d) We analyzed the effect of diffusion time τ in proposition 5.5. The estimated ODFs and scalar indices are independent of τ, and the estimated EAP is scaled by τ.

   (e) We proposed parameter selection via the scalar maps in subsection 5.1.3.

3. We proposed the general AFT-SC framework for analytical ODF and EAP estimation.

   (a) Many existing methods can be seen as a method in AFT-SC framework by considering different basis in spherical part or radial part. We demonstrate it is possible and easy to propose a number of analytical ODF and EAP estimation methods in AFT-SC framework. See Table 5.2.

   (b) With AFT-SC framework, EAP can be derived in QBI. We proposed a new way for analytical $\Phi_t(\mathbf{r})$ estimation in QBI, which is based on plane wave expansion theorem. We theoretically analyzed why EAP and $\Phi_w(\mathbf{r})$ are impractical in QBI model. In Appendix 5.7, we demonstrated the relation between plane wave expansion theorem and the previous theorems used in analytical QBI derivation, i.e. addition theorem and Funk-Hecke theorem.

   (c) In Proposition 5.6 and 5.7 we proposed analytical forms for EAP and ODFs in SHORE.

   (d) We theoretically analyzed the limitations of DPI and DOT, and proposed several possible variants to overcome the limitations. These variants also can be seen as bridges between methods. For example, DPI_1 uses a part of SHO-3D basis, and DOT_2 uses the SPNP basis which is proved to be equivalent with SPF basis.

4. Since we have shown that with AFT-SC framework we can propose as many bases as we want for analytical ODF and EAP estimation, we would like to know which basis is the best one. Thus we theoretically and experimentally compared many existing HARDI methods in AFT-SC framework, including QBI, exact QBI, DOT, DPI, SHORE, SPFI and several DPI and DOT variants. We proposed some criteria for theoretical comparisons. See Table 5.3.

   (a) We proposed three priors for evaluating the basis functions. The basis is better to have a Gaussian term such that it satisfies these priors.

   (b) The basis function set used in HARDI methods is better to be an orthonormal and complete basis.

   (c) mHARDI methods are unstable in single shell data. Thus a little regularization is needed for mHARDI methods applied in single shell data. Only SPFI with $N = 1$ in radial part is stable in single shell data, which works without regularization.

   (d) The basis is better to separate spherical part and radial part, such that we can give different order in different part. The experiments showed that it is better to give lower order in radial part if limited number of noisy samples are given.

   (e) We theoretically analyzed the SHO-3D, SHO-1D3, SPNP, SPP, and SPF basis.

5. We proposed a diffeomorphism invariant Riemannian framework for ODF and EAP processing based on the wavefunction representation by orthonormal basis.

   (a) The proposed Riemannian framework can be used in arbitrary PDFs under arbitrary orthonormal basis representation. This framework was proved to be invariant under different basis in Theorem 6.1.

   (b) The Riemannian metric was proved to be diffeomorphism invariant. When constraining PDFs as Gaussian distributions, the proposed Riemannian metric becomes affine-invariant metric, which means the proposed diffeomorphism invariant metric is a natural generalization of previous affine-invariant metric.

   (c) The geodesic, exponential map and logarithmic map have closed form.

   (d) We proposed several properties of the parameter space PS and discussed the similarities and differences between ODF space and EAP space.

   (e) The weighted Riemannian mean and median were proved to uniquely exist in the manifold, and they can be estimated efficiently via a gradient descent method.

   (f) We generalized the previous Log-Euclidean framework from tensors to ODFs and EAPs, which was demonstrated in experiments as an approximation of original Riemannian framework. We also proposed the Affine-Euclidean framework.

   (g) We generalized the GA from tensors to ODFs/EAPs, which is the Riemannian distance from the given ODF/EAP to its nearest ODF/EAP. GA for tensors/ODFs/EAPs is the Riemannian counterpart of GFA for tensors/ODFs/EAPs. We proposed the closed form of GA under SPF basis.

(h) We proposed a novel Square Root Parameterized Estimation (SRPE) for nonnegative definite ODF and EAP estimation and wavefuntion estimation. SRPE can guarantee the estimated ODFs and EAPs are nonnegative definite in their domains, i.e. $\mathbb{S}^2$ for ODFs and $\mathbb{R}^3$ for EAPs.

(i) We analyzed the property of the weighted Riemannian mean and median in Theorem 6.4. The probability value of the weighted mean/median PDF is not less than the probability value of at least one sampled PDF. This general property is applicable in Gaussian distributions in DTI, and general ODFs and EAPs in HARDI. We defined the concept of *reasonable mean value interpolation* on general PDFs. The *swelling effect* when interpolating tensor field can be seen as specific case of *unreasonable mean value interpolation* of Gaussian distributions.

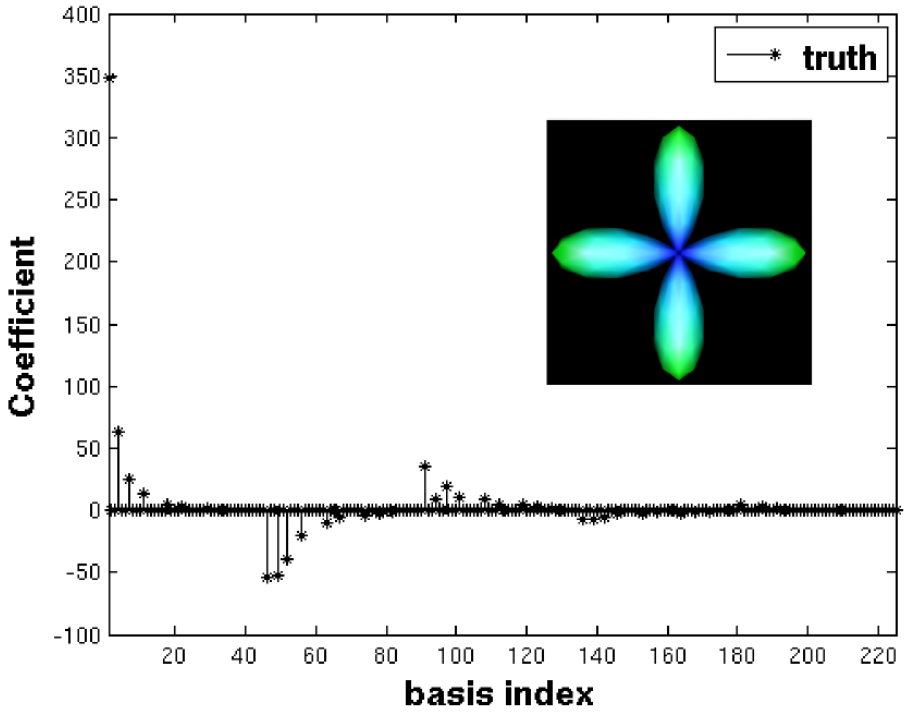
(j) We theoretically and experimentally analyzed the similarities and differences between Riemannian frameworks for tensors and for ODFs and EAPs, between Riemannian metric and Euclidean metric. Riemannian metric and Euclidean metric seem to obtain similar results for ODFs, but different results for tensors and EAPs. That is because the ODFs in human brain are all contained in a small subset in the general PDF space, which can be seen from that the first SH coefficient of wavefunction of ODF is experimentally in $[0.95, 1]$.

(k) The experiments showed that Riemannian median is more robust and obtains better performance in smoothing and atlas estimation than Riemannian mean and Euclidean mean.

Perspectives

During my study for this thesis, I first proposed the Riemannian framework for ODFs in [Cheng et al., 2009a, 2010c] based on my previous work in mean shift functional detection in fMRI [Cheng et al., 2009b]. Since the Riemannian framework can be used in arbitrary PDFs under arbitrary orthonormal basis, I was looking for a good orthonormal basis in $\mathbb{R}^3$ for EAPs. Fortunately, I met Dr. Assemlal in MICCAI 2009, who introduced me his work on SPF basis. I was very happy that the basis was just there when I was looking for it. However, when I tried to implement the SPFI described in [Assemlal et al., 2009a], I found the numerical inner product has many limitations. Then I proposed analytical ODF and EAP estimation in [Cheng et al., 2010b,a], where I set the scale using the typical ADC and I introduced the artificial shell for the prior $E(0) = 1$. Next, I proposed the Riemannian framework for EAPs in [Cheng et al., 2011a] by applying the SPF basis in the Riemannian framework. Based on my experiences in SPFI, I found there exist many basis functions which can be used in analytical ODF and EAP estimation. Thus I proposed AFT-SC framework and some criteria in [Cheng et al., 2011b] to incorporate these basis and find their relations, limitations and advantages. Since I have two bases in SPFI, i.e. SPF basis and dSPF basis, I applied the duality of these two bases in compressive sensing EAP estimation [Cheng et al., 2011c; Merlet et al., 2011], which largely accelerate the estimation by avoiding the numerical Fourier transform. I also collaborate with my colleague Dr. Caruyer for the sampling scheme used in mHARDI methods [Caruyer et al., 2011].

Figure 7.1: Sparsity of diffusion signal under SPF basis representation. The signal is generated by mixture of tensor model with $[\lambda_1, \lambda_2, \lambda_3] = [1.7, 0.3, 0.3] \times 10^{-3} mm^2/s$ and crossing angle of $90°$. It shows the coefficients of each basis with $N = 4, L = 8$. The glyph is the EAP profile at $15\mu m$.



Although SPFI can be used in single shell data with 30 gradients, it has better performance in multiple shell data with about 100 gradients. Although most data only has single b value, I believe the data with multiple b value will be more practical and popular as the development of scanners and acquisition methods. Compressive sensing (CS) technique can be used to reduce the number of measurements. Although we did not list the CS related methods in the thesis, we have already had some early results on CS EAP estimation in [Cheng et al., 2011c; Merlet et al., 2011], where we have shown the sparsity of the diffusion signal and EAP under SPF and dSPF bases, the difference between CS EAP estimation and CS ODF estimation, and the advantages of ℓ_1-SPFI over ℓ_2-SPFI . See Fig. 7.1 for the sparse representation of the diffusion signal generated by mixture of tensor model with $[\lambda_1, \lambda_2, \lambda_3] = [1.7, 0.3, 0.3] \times 10^{-3} mm^2/s$ and crossing angle of $90°$. We will continue working on this topic.

sHARDI methods normally need the samples evenly distributed in $\mathbb{S}^2$. mHARDI methods give more freedom for sampling scheme. We now agree multiple shell sampling has better reconstruction than single shell sampling. However we still do not know what is a good multiple shell sampling scheme. SPFI can be used to evaluation different sampling scheme [Assemlal et al., 2009b; Caruyer et al., 2011]. Although during this thesis we used the same evenly distributed samples for each shell, our work in [Caruyer et al., 2011] showed that it is better to consider interlaced sampling for different shells. So far different sampling schemes are evaluated based on synthetic experiments and SPFI reconstruction. However, there is still no theoretical result on what is the best sampling scheme under a certain condition.

SHO-1D basis has been used to characterize the compartment size distribution

and apparent axon diameters in [Özarslan et al., 2011, 2012]. It is possible to estimate the axon diameters based on SPF basis or SHO-3D basis for signals in 3D space.

In the Riemannian framework for EAPs, we used SPF basis or dSPF basis to represent the wavefunctions, where we needed to set the scale for the basis. It showed that the adaptive scale based on fitting GHOT model has better reconstruction results. However, different scale for different voxel means different basis is used in different voxel, which brings troubles in the following Riemannian operations. Thus a new Riemannian framework which considers different scale is needed for EAPs. A possible way is to represent the basis with different scale using the basis with the same scale, then the methods in Chapter 6 can be used.

The Riemannian metric is diffeomorphism invariant, which may be useful in registration of PDF valued images. Thus registration of ODF/EAP images using Riemannian metric can be another possible topic [Du et al., 2011].

Segmentation with Riemannian metric is also a topic [Descoteaux and Deriche, 2009]. We can test the differences between Riemannian metric and Euclidean metric on tensor/ODF/EAP segmentation.

Since the EAP and ODF fields estimated by SPFI and SRPE have good quality, we can test existing tractography methods in the estimated EAP and ODF field [Descoteaux et al., 2008a], or we can propose some new tractography methods.

In clinical part, CCM group has performed anatomical network analysis through DTI tractography to study different neuroscience problems [Li et al., 2009, 2012; Wang et al., 2011]. We believe the ODF/EAP estimation proposed in this thesis can be used to generate better anatomical network through ODF/EAP tractography for neuroscience study. For a long term perspective, based on the advanced ODF/EAP estimation methods, we would like to devise novel registration and tractography methods and other related methods to analyze the complex anatomical information in human brain.

Part IV

Appendices

APPENDIX **A**

SYNTHETIC DATA GENERATION AND EVALUATION OF ESTIMATION METHODS

Mixture of tensor model and mixture of cylinder model. In order to evaluate the performance of the ODF and EAP estimation methods in dMRI domain, we need to apply these methods to some synthetic data sets with known ground truths. In this thesis, we generate synthetic diffusion data $E(\mathbf{q})$ from the mixture of a given model, i.e.

$$E(\mathbf{q}) = \sum_{k=1}^{K} w_i E^{(i)}(\mathbf{q}) \tag{A.1}$$

where $\sum_{k=1}^{K} w_i = 1$, $w_i \geq 0$ and $E^{(i)}(\mathbf{q})$ is the diffusion signal of the i-th fiber, K is the number of fibers. Based on some biological priors, $K = 1, 2, 3$ for at most 3 fibers in one voxel, and normally $K = 2$ is used to generate the data for crossing fibers. In this thesis, each term of this mixture model, i.e. $E^{(i)}(\mathbf{q})$, is generated from the following two models.

- Tensor model with the assumption of free diffusion, where the signal for each fiber direction $E^{(i)}(\mathbf{q})$ is assumed to be represented by a tensor $\mathbf{D}_i$, i.e.

  $$E^{(i)}(\mathbf{q}) = \exp(-4\pi^2\tau q^2 \mathbf{u}^T \mathbf{D}_i \mathbf{u}) \tag{A.2}$$

  See subsection 4.3.1 and Appendix 5.8 for more details on the mixture of tensor model. In this model, we choose $\tau = \frac{1}{4\pi^2} = 0.02533s$ such that $b = q^2$.

- Söderman cylinder model with the assumption of restricted diffusion inside a cylinder. Based on [Söderman and Jönsson, 1995] the signal for each fiber $E^{(i)}(\mathbf{q})$ is represented as

$$\begin{aligned} E^{(i)}(\mathbf{q}) = & \sum_{n=0}^{\infty}\sum_{k=1}^{\infty}\sum_{m=0}^{\infty} \frac{2K_{nm}\rho^2(2\pi q\rho)^4 \sin^2(2\nu)\gamma_{km}^2}{\left[(n\pi\rho/L)^2 - (2\pi q\rho\cos\nu)^2\right]^2} \\ & \times \frac{[1-(-1)^n\cos(2\pi qL\cos\nu)][J'_m(2\pi q\rho\sin\nu)]^2}{L^2\left[\gamma_{km}^2 - (2\pi q\rho\sin\nu)^2\right]^2(\gamma_{km}^2 - m^2)} \\ & \times \exp\left(-\left[\left(\frac{\gamma_{km}}{\rho}\right)^2 + \left(\frac{n\pi}{L}\right)^2\right] D_0\Delta\right) \end{aligned} \tag{A.3}$$

where J_m is the m-th order Bessel function, γ_{km} is the k-th solution of equation $J'_m(\gamma) = 0$ with the convention $\gamma_{10} = 0$, and $K_{nm} = \delta^0_n\delta^0_m + 2[(1-\delta^0_n) + (1-\delta^0_m)]$, ν is the angle between the cylinder direction and the applied diffusion gradient $\mathbf{u}$. In this thesis, we fix the parameters as the same as [Özarslan et al., 2006], i.e. cylinder length $L = 5mm$, radius $\rho = 5\mu m$, free diffusion coefficient $D_0 = 2.02 \times 10^{-3} mm^2/s$, $\Delta/\delta = 20.8/2.4ms$, and the diffusion time $\tau = \Delta - \frac{1}{3}\delta$, $b = 4\pi^2\tau q^2$.

Mixture of tensor model is widely used to generate synthetic data for evaluation in dMRI domain, because it has a clear relation with tensor model in DTI and the EAP and its features all have closed form. See Appendix 5.8. The isotropic signal is generated from this model with $K = 1$ and $\mathbf{D} = D_o\mathbf{I}$, where

$$D_0 = 0.7 \times 10^{-3} mm^2/s \tag{A.4}$$

is a typical ADC value in human brain, and $\mathbf{I}$ is the identity matrix. The signal for crossing fibers is normally generated with $K = 2$, $w_1 = w_2 = 0.5$, and $\mathbf{D}_i$ is chosen as a typical anisotropic tensor with eigenvalues $[\lambda_1, \lambda_2, \lambda_3] = [1.7, 0.3, 0.3] \times 10^{-3} mm^2/s$.

Although many papers use mixture of tensor in synthetic data experiments, this model actually favors the methods which assume mixture of Gaussian propagator implicitly. For example, when comparing mixture of tensor method in subsection 4.3.1 with other methods like QBI, it is biased to generate synthetic data using mixture of tensor model.

Compared to the mixture of tensor model, mixture of Söderman cylinder model does not assume Gaussian prior for single fiber, which makes the evaluation more fair. However, there are no closed forms for EAP and ODFs in this model, which means the ground truths of EAP and ODFs are needed to be calculated from numerical Fourier transform and numerical integration. Thus it takes more time when using Söderman cylinder model. Fortunately the ground truths of fiber directions are always known in this model, which can be used for evaluation.

Signal-to-Noise Ratio. The evaluation process can be done with noise or without noise. Normally the noise is assumed to be Rician noise in DWI data. After we generate the ground truth signal $E(\mathbf{q})$ from the mixture of tensor model or mixture of Söderman cylinder model described above, Signal-to-Noise Ratio (SNR) is defined as the ratio of maximal signal intensity of $E(0) = 1$ to the standard deviation σ of complex Gaussian noise, i.e.

$$\text{SNR} = \frac{1}{\sigma} \tag{A.5}$$

Then the Rician noise corrupted signal is

$$\tilde{E}(\mathbf{q}) = \sqrt{(E(\mathbf{q}) + s_1)^2 + s_2} \tag{A.6}$$

where s_1 and s_2 are random number generated from Gaussian distribution $N(0|\sigma^2)$. Note that since $E(q\mathbf{u})$ decays as q increase and $E(0) = 1$, the real SNR, i.e. $\frac{E(\mathbf{q})}{\sigma}$ decreases as q increase. See Fig. 4.3 for DWIs with different q values.

Evaluation of estimation results. There are two ways to evaluate the difference between the estimated EAPs/ODFs and the ground truths. One is to compare the local maxima of ODF/EAP profiles with the real fiber directions used to generate the DWI data. See Fig. 4.8 for the fiber directions and the local maxima of ODF/EAP profiles. The other one is to compare the estimated ODF/EAP values with the true values.

In the first way, assume the true fiber directions are $\{\boldsymbol{d}_i\}_{i=1}^K$, and the detected directions with local maximal values of ODF/EAP profiles are $\{\mathbf{u}_i\}_{i=1}^N$. How to detect maxima from spherical functions, e.g. ODF/EAP profiles, will be discussed later. There are two possibilities between these two direction sets, i.e. $N = K$ and $N \neq K$. In this thesis, we think that $N \neq K$ means a unsuccessful estimation, and $N = K$ means a successful estimation. If we perform the estimation N_t times from the signal corrupted by the Rician noise, and there are N_1 successful estimation, then the **success ratio** is defined as the frequency of successful estimation, i.e. $\frac{N_1}{N_t}$.

In a successful experiment with true fiber directions $\{\boldsymbol{d}_i\}_{i=1}^K$ and detected directions $\{\mathbf{u}_i\}_{i=1}^K$ for local maximal values, we need to reorder $\{\mathbf{u}_i\}_{i=1}^K$ such that $\mathbf{u}_i$ is the nearest direction from $\boldsymbol{d}_i$ in $\{\mathbf{u}_i\}_{i=1}^K$. Then the Difference of Angle (DA) can be calculated as

$$\mathrm{DA} = \frac{1}{K}\sum_{i=1}^{K} \arccos(\boldsymbol{d}_i^T \mathbf{u}_i) \tag{A.7}$$

In noise-free experiments, DA is enough for evaluation. In experiments with Rician noise and total N_t times, the Mean Difference of Angle (MDA) is defined as the mean of DA over only the N_1 successful estimations.

Note that some papers do not consider the success ratio. After obtaining the true directions $\{\boldsymbol{d}_i\}_{i=1}^K$ and the estimated directions $\{\mathbf{u}_i\}_{i=1}^N$, K nearest directions $\{\mathbf{u}_{j_i}\}_{i=1}^K$ to $\{\boldsymbol{d}_i\}_{i=1}^K$ are extracted from $\{\mathbf{u}_i\}_{i=1}^N$, where $\mathbf{u}_{j_i}$ is the nearest direction to $\boldsymbol{d}_i$ in $\{\mathbf{u}_i\}_{i=1}^N$. Then DA and MDA are calculated based on the extracted K directions. This evaluation process is actually biased. It favors the estimation methods which can detect more directions. For example, the estimated EAP profile with $N = 10$ maxima has lower DA and MDA with higher probability than the estimated EAP profile with the same number of maxima, say $N = 2$, as the ground truth of EAP profile.

The second way for evaluation is to compare the estimated ODF/EAP values with the true values. The estimated ODFs and EAP profiles are spherical functions. Thus we define the Normalized Mean Squared Error (NMSE) for EAP profile as

$$\mathrm{NMSE} = \sqrt{\frac{\int_{\mathbb{S}^2} \left(\tilde{P}(R_0\mathbf{r}) - P(R_0\mathbf{r})\right)^2 \mathrm{d}\mathbf{r}}{\int_{\mathbb{S}^2} P(R_0\mathbf{r})^2 \mathrm{d}\mathbf{r}}} \tag{A.8}$$

NMSE for ODFs can be defined in the same way. The integration in $\mathbb{S}^2$ can be approximated by numerical integration using a summation of the values at a given direction set. NMSE was used in the evaluation of ℓ_1-SPFI [Cheng et al., 2011c].

Note that some papers define NMSE as

$$NMSE = \sqrt{\frac{\sum_j \left(\tilde{E}(\mathbf{q}_j) - E(\mathbf{q}_j)\right)^2}{\sum_j E(\mathbf{q}_j)^2}}$$

where $\{\mathbf{q}_j\}$ are the given DWI samples in $\mathbf{q}$-space. Actually this definition of NMSE is biased, because our goal is to reconstruct DWI signal in whole $\mathbf{q}$-space to estimate the ODFs or EAP profiles, not to fit the given samples. The above definition of NMSE can be used as a cost function in estimation. But it is not enough for evaluation, because the Fourier transform is a ***global*** transform, and fitting the given samples well does not mean a good estimation as we have discussed in subsection 5.3.1. Actually, NMSE represented by $E(\mathbf{q})$ should be defined as

$$NMSE = \sqrt{\frac{\int_{\mathbb{R}^3} \left(\tilde{E}(\mathbf{q}) - E(\mathbf{q})\right)^2 \mathrm{d}\mathbf{q}}{\int_{\mathbb{R}^3} E(\mathbf{q})^2 \mathrm{d}\mathbf{q}}} \tag{A.9}$$

Please note the above integration is over whole $\mathbf{q}$ space, not just in the limited given DWI samples. Thus Eq. (A.9) is impractical, and Eq. (A.8) is preferred.

Maxima detection of spherical functions. The maxima of ODFs and EAP profiles are normally used as fiber directions. How to detect the maxima from these spherical functions with good accuracy and efficiency are very important for evaluation of estimation methods and for the fiber tracking.

Many papers detect the maxima of spherical function through a discrete mesh search [Özarslan et al., 2006; Descoteaux, 2008]. With a given mesh in a hemisphere, e.g. a set of evenly distributed vertices from a tessellated icosahedron, the local maxima are detected from the values at these vertices. However this method does not search the maxima in continuous $\mathbb{S}^2$, and the minimal angle between adjacent vertices is the resolution of this method. Thus normally the detected maxima need to be refined by a finer mesh, which is time consuming.

Because of the equivalence between polynomial basis, homogeneous polynomial basis and SH basis shown in Theorem 4.1, we can represent these spherical function using homogeneous polynomial basis, i.e. HOT basis. Then the maxima detection becomes the problem of finding the solution of polynomial equations. One may think it will be better to find the analytical solution directly from the coefficients under the given basis. Unfortunately even in 1D space, the polynomial equation with order more than 5 has no analytical form of solution[1]. Thus numerical optimization is necessarily needed in maxima detection. For the spherical function represented by HOT basis or SH basis with order no more than 4, the maxima detection was proved to be an 1D search problem by using the analytical forms of solution for the polynomial with order no more than 4 [Aganj et al., 2010a]. [Ghosh et al., 2008b, 2011; Ghosh, 2011] proposed to analytically break the possible regions for extrema and find them numerically. It actually solves much harder problem than maxima detection, since it finds all extrema. Moreover some useful scalar indices can be derived from the detected extrema.

In this thesis, we use a hybrid way to find the maxima, which is very fast. The process is as follows.

1. First, we perform a mesh search on the spherical function to find some local maxima, by using a coarse mesh, e.g. 321 directions in a hemisphere from an icosahedral tessellation.

[1] http://en.wikipedia.org/wiki/Polynomial#Solving_polynomial_equations

2. Second, perform a threshold to remove the local maxima whose values are less than a half of the maximal value among these 321 values.

3. Third, set the remaining maxima as initialization points, and perform gradient ascent method to update the maxima in continuous space. When the spherical function is represented by SH basis, the derivatives of the function with respect to θ and ϕ are obtained by the derivatives of SH basis.

$$\frac{\partial}{\partial\theta}y_l^m(\theta,\phi) = \begin{cases} \frac{m}{\tan\theta}y_l^m(\theta,\phi) + e^{-i\phi}\sqrt{\frac{\Gamma(1+l-m)\Gamma(2+l+m)}{\Gamma(l-m)\Gamma(1+l+m)}}y_l^{1+m}(\theta,\phi) & \text{if } m \neq l \\ \frac{l}{\tan\theta}y_l^m(\theta,\phi) & \text{if } m = l \end{cases} \tag{A.10}$$

$$\frac{\partial}{\partial\phi}y_l^m(\theta,\phi) = imy_l^m(\theta,\phi) \tag{A.11}$$

The above formulae are derivatives of complex SH basis, which directly result in the derivations of the symmetric real SH basis based on Definition 2.4.

APPENDIX B

REAL DATA SETS

In this thesis, we use the following two real data sets.

- Public phantom data from LNAO.
- Monkey data from Xuanwu Hospital, Capital Medical University.

We express our great gratitudes towards the people who provide us the data sets.

B.1 PUBLIC PHANTOM DATA

Source: The public phantom data from LNAO has different configurations of fibers and was used to evaluate fiber tracking algorithms in Fiber Cup of MICCAI 2009 [Poupon et al., 2008; Fillard et al., 2011].

Data Descriptions: It has two data sets with spatial resolution of $3mm$ and $6mm$, which can be downloaded from http://lnao.lixium.fr/spip.php?article112. We choose the data set with spatial resolution of $3mm$, which was also used by 9 works among 10 candidates in Fiber Cup [Fillard et al., 2011]. The data set with the $3mm$ isotropic acquisition has three slices, three b values, $b = 650/1500/2000 s/mm^2$, with repetition time TR=$5s$ and TE=$77/94/102ms$. In each b value, the data has two repetitions of one image without diffusion sensitization and 64 diffusion weighted images with 64 uniformly distributed gradient directions. Other parameters: field of view FOV=$19.2cm$, matrix 64×64, slice thickness TH=$3mm$. The diffusion time τ is unknown for this data, so we set it as $\tau = \frac{1}{4\pi^2}$ such that $b = q^2$. The relation of estimation results and different τ can be seen in the analysis in Proposition 5.5 and subsection 6.2.1 The ground truth of fiber directions can be seen in Fig. B.1. Please refer [Fillard et al., 2011] for more parameters about this data and refer [Poupon et al., 2008] for the technique used to generate the data.

Preprocessing: Throughout this thesis, we calculate the mean of DWI images from twice acquisitions in each fixed b value and direction. Thus we obtain 64 DWI images for each b value, total 192 DWI images for three b values $b = 650/1500/2000 s/mm^2$. No further preprocessing was performed.

Figure B.1: The ground truth of fiber directions in phantom data.

Challenges: This phantom data has two main challenges for reconstruction methods and the following fiber tracking algorithms.

- First, the maximal b value of this data is $2000 s/mm^2$, while many methods may need higher b values. However we think it is an advantage if a reconstruction method can detect the fiber directions well only using small b values. Note that for real data, larger b value means lower SNR. It is still impractical for clinical study to obtain data with b value larger than $3000 s/mm^2$.

- Second, this phantom data is known to be more isotropic than real human data. The estimated ODFs, EAP profiles in many papers were performed min-max normalization for visualization. Otherwise, the anisotropy is not very clear in visualization.

B.2 REAL MONKEY DATA

Source: The real monkey data is obtained with the help of Dr. Wen Qin and Prof. Chunshui Yu from Xuanwu Hospital, Capital Medical University, and Dr. Yonghui Li from CCM, LIAMA, CASIA. Many thanks for providing us the data.

Data Descriptions: The data was acquired from 3T Siemens MR scanner using a twice-refocused spin echo (TRSE) sequence, where TE/TR/matrix=$120ms/6000ms/128\times 128$. The data has five subjects with three b values, $b = 500/1500/3000 s/mm^2$, the same 30 diffusion gradients in each b value. Note that we do not know exact diffusion time τ for this data, so we set it as $\tau = \frac{1}{4\pi^2}$ such that $b = q^2$.

Preprocessing: We preformed eddy current correction on all DWI images using FSL (http://www.fmrib.ox.ac.uk/fsl/).

APPENDIX C

PUBLICATIONS OF THE AUTHOR

Journal Papers

- **Jian Cheng**, Aurobrata Ghosh, Rachid Deriche, Tianzi Jiang, "An Intrinsic Diffeomorphism Invariant Riemannian Framework for Probability Density Function Computing in diffusion MRI", *IEEE Transaction on Medical Imaging* (in revision)
- **Jian Cheng**, Aurobrata Ghosh, Tianzi Jiang, Rachid Deriche, "Spherical Polar Fourier Imaging", *IEEE transaction on Medical Imaging* (to be submitted)
- Nianming Zuo, **Jian Cheng**, Tianzi Jiang, "Diffusion Magnetic Resonance Imaging for Brainnetome: A Critical Review", *Neuroscience Bulletin*, 2012 (accepted).

Conference Papers

- **Jian Cheng**, Tianzi Jiang, Rachid Deriche, "Nonnegative Definite EAP and ODF Estimation via a Unified Multi-Shell HARDI Reconstruction", *in Proceedings of 15th International Conference on Medical Image Computing and Computer Assisted Intervention (**MICCAI'12**)*, Nice, October 1-5, 2012. (**MICCAI Student Travel Award**)
- **Jian Cheng**, Aurobrata Ghosh, Tianzi Jiang, Rachid Deriche, "Diffeomorphism Invariant Riemannian Framework for Ensemble Average Propagator Computing", *in Proceedings of 14th International Conference on Medical Image Computing and Computer Assisted Intervention (**MICCAI'11**)*, Toronto, September 18-22, 2011. (**MICCAI Student Travel Award**)
- **Jian Cheng**, Tianzi Jiang, Rachid Deriche, "Theoretical Analysis and Practical Insights on EAP Estimation via a Unified HARDI Framework", *Computational Diffusion MRI workshop (**CDMRI'11**)*, in conjunction with the MICCAI'11, Toronto, September 18-22, 2011. (**Oral**)
- **Jian Cheng**, Sylvain Merlet, Aurobrata Ghosh, Emmanuel Caruyer, Tianzi Jiang, Rachid Deriche, "Compressive Sensing Ensemble Average Propagator Estimation via L1 Spherical Polar Fourier Imaging", *Computational Diffusion MRI workshop (**CDMRI'11**)*, in conjunction with the MICCAI'11, Toronto, September 18-22, 2011. (**Oral**)

- Emmanuel Caruyer, **Jian Cheng**, Christophe Lenglet, Guillermo Sapiro, Tianzi Jiang, Rachid Deriche, "Optimal Design of Multiple Q-shells experiments for Diffusion MRI", *Computational Diffusion MRI workshop (**CDMRI'11**)*, in conjunction with the MICCAI'11, Toronto, September 18-22, 2011. (**Oral**)

- Sylvain Merlet, **Jian Cheng**, Aurobrata Ghosh, Rachid Deriche, "Spherical Polar Fourier EAP and ODF Reconstruction via Compressed Sensing in Diffusion MRI", *in Proceedings of 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro (**ISBI'11**)*, Chicago, United States, March 30 - April 2, 2011.

- **Jian Cheng**, Aurobrata Ghosh, Tianzi Jiang, Rachid Deriche, "Model-free and Analytical EAP Reconstruction via Spherical Polar Fourier Diffusion MRI", *in Proceedings of 13th International Conference on Medical Image Computing and Computer Assisted Intervention (**MICCAI'10**)*, Beijing, September 20-24, 2010.

- **Jian Cheng**, Aurobrata Ghosh, Rachid Deriche, Tianzi Jiang, "Model-Free, Regularized, Fast, and Robust Analytical Orientation Distribution Function Estimation", *in Proceedings of 13th International Conference on Medical Image Computing and Computer Assisted Intervention (**MICCAI'10**)*, Beijing, September 20-24, 2010.

- **Jian Cheng**, Aurobrata Ghosh, Tianzi Jiang, Rachid Deriche, "A Riemannian Framework for Orientation Distribution Function Computing", *in Proceedings of 12th International Conference on Medical Image Computing and Computer Assisted Intervention (**MICCAI'09**)*, London, September 20-24, 2009.

- **Jian Cheng**, Feng Shi, Kun Wang, Ming Song, Jiefeng Jiang, Lijuan Xu, Tianzi Jiang, "Nonparametric Mean Shift Functional Detection on Functional Space for Task and Resting-state fMRI", *Workshop on fMRI data analysis: statistical modeling and detection issues in intra- and inter-subject functional MRI data analysis*, in conjunction with the MICCAI 2009, London, September 20-24, 2009. (**Oral**)

- Xi Li, Weiming Hu, Zhongfei Zhang, Xiaoqin Zhang, Mingliang Zhu, **Jian Cheng**, "Visual tracking via incremental Log-Euclidean Riemannian subspace learning", *in Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (**CVPR'08**)*, Anchorage, Alaska. June 24-26, 2008.

Conference Abstracts

- **Jian Cheng**, Aurobrata Ghosh, Tianzi Jiang, and Rachid Deriche, "A Riemannian Framework for Ensemble Average Propagator Computing", *19th Scientific Meeting and Exhibition of the **ISMRM***, Montréal, Canada, May 7-13, 2011.

- **Jian Cheng**, Sylvain Merlet, Aurobrata Ghosh, Emmanuel Caruyer, Tianzi Jiang, and Rachid Deriche, "Compressive Sensing Ensemble Average Propagator Estimation via L1 Spherical Polar Fourier Imaging", *19th Scientific Meeting and Exhibition of the **ISMRM***, Montréal, Canada, May 7-13, 2011.

- **Jian Cheng**, Aurobrata Ghosh, Tianzi Jiang, Rachid Deriche, "Fast, model-free, analytical diffusion PDF profile estimation from the DWI signals", *in Proceedings of the Sixteenth Annual Meeting of the Organization for Human Brain Mapping (**OHBM'10**)*, Barcelona, Span, June 6-10, 2010

- **Jian Cheng**, Aurobrata Ghosh, Tianzi Jiang, Rachid Deriche, "Fast, Model-Free, Analytical ODF Reconstruction from the Q-Space Signal", *in Proceedings of the Sixteenth Annual Meeting of the Organization for Human Brain Mapping (**OHBM'10**)*, Barcelona, Span, June 6-10, 2010.

- **Jian Cheng**, Aurobrata Ghosh, Tianzi Jiang, Rachid Deriche, "Riemannian Median and Its Applications for Orientation Distribution Function Computing", *in Proceedings of 18th Scientific Meeting and Exhibition of the **ISMRM***, Stockholm, Sweden, May 1-7, 2010.

Bibliography

A. Rényi: 1960, 'On Measures of Entropy and Information'. *Proc. Fourth Berkeley Symp. Math., Statistics, and Probability* **15**, 547–561. 185

Afsari, B.: 2011, 'Riemannian Lp center of mass: Existence, uniqueness, and convexity'. *Proceedings of the American Mathematical Society* **139**(2), 655–673. 52

Aganj, I., C. Lenglet, and G. Sapiro: 2009, 'ODF reconstruction in q-ball imaging with solid angle consideration'. In: *ISBI*. 89

Aganj, I., C. Lenglet, and G. Sapiro: 2010a, 'ODF maxima extraction in spherical harmonic representation via analytical search space reduction'. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010* pp. 84–91. 234

Aganj, I., C. Lenglet, G. Sapiro, E. Yacoub, K. Ugurbil, and N. Harel: 2010b, 'Reconstruction of the orientation distribution function in single and multiple shell q-ball imaging within constant solid angle'. *Magnetic Resonance in Medicine* **2**, 554–566. 12, 33, 77, 82, 87, 88, 89, 90, 91, 92, 103, 104, 105, 114, 118, 119, 151, 181, 189, 198, 206

Alexander, D. and G. Barker: 2005, 'Optimal imaging parameters for fiber-orientation estimation in diffusion MRI'. *NeuroImage* **27**, 357–367. 72

Alexander, D., G. Barker, and S. Arridge: 2002, 'detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data'. *Magnetic Resonance in Medicine* **48**, 331–340. 113

Alexander, D. C., C. Pierpaoli, P. J. Basser, and J. C. Gee: 2001, 'Spatial Transformations of Diffusion Tensor Magnetic Resonance Images'. *IEEE Transactions On Medical Imaging* **20**, 1131–1139. 214

Amari, S.: 1985, *Differential-geometrical methods in statistic*. Springer. 36, 53, 54, 57

Amari, S. and H. Nagaoka: 2000, *Methods of Information Geometry*. USA: American Mathematical Society. 36, 53

Anderson, A.: 2005, 'Measurement of fiber orientation distributions using high angular resolution diffusion imaging'. *Magnetic Resonance in Medicine* **54**(5), 1194–1206. 2, 33, 83, 84, 103, 114

Andrews, G., R. Askey, and R. Roy: 1999, *Special functions*. Cambridge University Press. 12, 25

Arfken, G., H. Weber, and H. Weber: 2005, *Mathematical methods for physicists*. Academic press New York, 6th edition edition. 12, 13, 19, 21, 24, 25, 27, 30

Arsigny, V., P. Fillard, X. Pennec, and N. Ayache: 2006, 'Log-Euclidean metrics for fast and simple calculus on diffusion tensors'. *Magnetic Resonance in Medicine* **56**, 411–421. 99, 185

Askey, R. and J. Wilson: 1985, *Some basic hypergeometric orthogonal polynomials that generalize Jacobi polynomials*, Vol. 319. Amer Mathematical Society. 12

Assaf, Y., R. Freidlin, G. Rohde, and P. Basser: 2004, 'New modeling and experimental framework to characterize hindered and restricted water diffusion in brain white matter'. *Magnetic Resonance in Medicine* **52**(5), 965–978. 74, 75

Assemlal, H., D. Tschumperlé, and L. Brun: 2008, 'Efficient computation of pdf-based characteristics from diffusion mr signal'. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008* pp. 70–78. 97, 111, 122

Assemlal, H., D. Tschumperlé, L. Brun, and K. Siddiqi: 2011, 'Recent advances in diffusion MRI modeling: Angular and radial reconstruction'. *Medical Image Analysis*. 62

Assemlal, H.-E.: 2010, 'Diffusion MR image analysis for the estimation of tissues local architecture'. Ph.D. thesis, the University of Caen. 111, 114, 115, 123, 163, 166

Assemlal, H.-E., D. Tschumperlé, and L. Brun: 2009a, 'Efficient and robust computation of PDF features from diffusion MR signal'. *Medical Image Analysis* **13**, 715–729. 74, 80, 92, 97, 111, 113, 114, 115, 122, 126, 133, 134, 135, 163, 166, 181, 193, 194, 226

Assemlal, H.-E., D. Tschumperlé, and L. Brun: 2009b, 'Evaluation of q-Space Sampling Strategies for the Diffusion Magnetic Resonance Imaging'. In: *Medical Image Computing and Computer Assisted Intervention (MICCAI)*. 123, 227

Atkinson, C. and A. Mitchell: 1981, 'Rao's distance measure'. *Sankhyā: The Indian Journal of Statistics, Series A* **43**(3), 345–365. 55, 56, 58, 99, 187

Axler, S., P. Bourdon, and W. Ramey: 2001, *Harmonic function theory*, Vol. 137. Springer Verlag. 12, 20

Bar-Shir, A., L. Avram, E. Ozarslan, P. Basser, and Y. Cohen: 2008, 'The effect of the diffusion time and pulse gradient duration ratio on the diffraction pattern and the structural information estimated from q-space diffusion MR: experiments and simulations'. *Journal of Magnetic Resonance* **194**(2), 230–236. 70

Barmpoutis, A., M. Hwang, D. Howland, J. Forder, and B. Vemuri: 2009, 'Regularized positive-definite fourth order tensor field estimation from DW-MRI'. *NeuroImage* **45**(1), S153–S162. 78

Barmpoutis, A., B. Jian, and B. C. Vemuri: 2007, 'Symmetric Positive 4th Order Tensor & Their Estimation from Diffusion Weighted MRI'. In: *IPMI*, Vol. 66. pp. 308–319. 78

Basser, P. J., J. Mattiello, and D. LeBihan: 1994, 'MR Diffusion Tensor Spectroscropy and Imaging'. *Biophysical Journal* **66**, 259–267. 1, 67, 69, 70, 72, 208

Batchelor, P., M. Moakher, D. Atkinson, F. Calamante, and A. Connelly: 2005, 'A Rigorous Framework for Diffusion Tensor Calculus'. *Magnetic Resonance in Medicine* pp. 221–225. 56, 58, 99, 100, 184, 186

Behrens, T., H. Berg, S. Jbabdi, M. Rushworth, and M. Woolrich: 2007, 'Probabilistic diffusion tractography with multiple fibre orientations: What can we gain?'. *Neuroimage* **34**(1), 144–155. 2, 75

Bernstein, M. A., K. F. King, and X. J. Zhou (eds.): 2004, *Handbook of MRI Pulse Sequences*. Elsevier Academic Press. 66, 67

Bhattacharyya, A.: 1943, 'On a measure of divergence between two statistical populations defined by their probability distributions'. *Bull. Calcutta Math. Soc* **35**(99-109), 4. 179

Bihan, D. L., E. Breton, D. Lallemand, P. Grenier, E. Cabanis, and M. Laval-Jeantet: 1986, 'MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders'. *Radiology* **161**, 401–407. 1, 67, 69

Blinnikov, S. and R. Moessner: 1998, 'Expansions for nearly Gaussian distributions'. *Astronomy and Astrophysics Supplement Series* **130**(1), 193–205. 77, 146

Boothby, W.: 1986, *An introduction to differentiable manifolds and Riemannian geometry*, Vol. 120. Academic Press. 36

Braunstein, S. and C. Caves: 1994, 'Statistical distance and the geometry of quantum states'. *Physical review letters* **72**(22), 3439–3443. 53

Brody, D. and L. Hughston: 1998, 'Statistical geometry in quantum mechanics'. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **454**(1977), 2445. 53

Buss, S. R. and J. P. Fillmore: 2001, 'Spherical Averages and Applications to Spherical Splines and Interpolation'. *ACM Transactions on Graphics* **20**, 95–126. 48, 50, 186, 198, 217, 218

Byron, F. and R. Fuller: 1992, *Mathematics of classical and quantum physics*. Dover Pubns. 12, 15, 23

Callaghan, P. T.: 1991, *Principles of nuclear magnetic resonance microscopy*. Oxford University Press. 68, 80

Canales-Rodrıguez, E. J., L. Melie-Garcıa, and Y. Iturria-Medina: 2009, 'Mathematical description of q-space in spherical coordinates: Exact q-ball imaging'. *Magnetic Resonance In Medicine* **61**, 1350–1367. 33, 77, 82, 87, 88, 89, 90, 92, 103, 104, 114, 118, 181, 198

Caruyer, E., J. Cheng, C. Lenglet, G. Sapiro, T. Jiang, R. Deriche, et al.: 2011, 'Optimal Design of Multiple Q-shells experiments for Diffusion MRI'. In: *Computational Diffusion MRI - MICCAI Workshop*. 126, 226, 227

Chefd'hotel, C., D. Tschumperlé, R. Deriche, and O. Faugeras: 2004, 'Regularizing Flows for Constrained Matrix-Valued Images'. *Journal of Mathematical Imaging and Vision* **20**(1-2), 147–162. 72

Chen, Y., W. Guo, Q. Zeng, X. Yan, F. Huang, H. Zhang, G. He, B. Vemuri, and Y. Liu: 2004, 'Estimation, smoothing, and characterization of apparent diffusion coefficient profiles from high angular resolution dwi'. In: *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, Vol. 1. pp. I–588. 78, 79

Chen, Y., W. Guo, Q. Zeng, X. Yan, M. Rao, and Y. Liu: 2005, 'Apparent diffusion coefficient approximation and diffusion anisotropy characterization in DWI'. In: *Information Processing in Medical Imaging*. pp. 246–257. 78, 79

Cheng, J., A. Ghosh, R. Deriche, and T. Jiang: 2010a, 'Model-Free, Regularized, Fast, and Robust Analytical Orientation Distribution Function Estimation'. In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, Vol. 6361. pp. 648–656. 88, 89, 90, 92, 97, 104, 123, 126, 152, 156, 181, 189, 190, 191, 193, 194, 210, 212, 226

Cheng, J., A. Ghosh, T. Jiang, and R. Deriche: 2009a, 'A Riemannian Framework for Orientation Distribution Function Computing'. In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, Vol. 5761. pp. 911–918. 47, 59, 101, 184, 185, 186, 190, 215, 226

Cheng, J., A. Ghosh, T. Jiang, and R. Deriche: 2010b, 'Model-free and Analytical EAP Reconstruction via Spherical Polar Fourier Diffusion MRI'. In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, Vol. 6361. pp. 590–597. 92, 97, 121, 123, 126, 152, 153, 154, 156, 181, 189, 190, 191, 193, 194, 200, 203, 210, 211, 212, 226

Cheng, J., A. Ghosh, T. Jiang, and R. Deriche: 2010c, 'Riemannian Median and Its Applications for Orientation Distribution Function Computing'. In: *ISMRM*. 186, 226

Cheng, J., A. Ghosh, T. Jiang, and R. Deriche: 2011a, 'Diffeomorphism Invariant Riemannian Framework for Ensemble Average Propagator Computing'. In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, Vol. 6892 of *LNCS*. pp. 98–106, Springer Berlin / Heidelberg. 47, 59, 101, 190, 195, 226

Cheng, J., T. Jiang, and R. Deriche: 2011b, 'Theoretical Analysis and Practical Insights on EAP Estimation via a Unified HARDI Framework'. In: *Computational Diffusion MRI - MICCAI Workshop*. 126, 182, 226

Cheng, J., S. Merlet, E. Caruyer, A. Ghosh, T. Jiang, R. Deriche, et al.: 2011c, 'Compressive Sensing Ensemble Average Propagator Estimation via L1 Spherical Polar Fourier Imaging'. In: *Computational Diffusion MRI - MICCAI Workshop*. 103, 126, 128, 181, 190, 191, 193, 195, 214, 215, 226, 227, 233

Cheng, J., F. Shi, K. Wang, M. Song, J. Jiang, L. Xu, and T. Jiang: 2009b, 'Non-parametric Mean Shift Functional Detection in the Functional Space for Task and Resting-state fMRI'. In: *Workshop on fMRI data analysis: statistical modeling and detection issues in intra- and inter-subject functional MRI data analysis in conjunction with the MICCAI*. 47, 59, 226

Chikuse, Y.: 2003, *Statistics on special manifolds*, Vol. 174. Springer Verlag. 47

Conway, J.: 1990, *A course in functional analysis*, Vol. 96. Springer. 12

Courant, R. and D. Hilbert: 1989, *Methods of mathematical physics*, Vol. 1. Wiley. 13

Descoteaux, M.: 2008, 'High Angular Resolution Diffusion MRI: from Local Estimation to Segmentation and Tractograpghy'. Ph.D. thesis, INRIA Sophia Antipolis. 3, 62, 69, 72, 101, 174, 234

Descoteaux, M., E. Angelino, S. Fitzgibbons, and R. Deriche: 2006, 'Apparent Diffusion Coefficients from High Angular Resolution Diffusion Images: Estimation and Application'. *Magnetic Resonance in Medicine* **56**, 395–410. 77, 113

Descoteaux, M., E. Angelino, S. Fitzgibbons, and R. Deriche: 2007, 'Regularized, Fast and Robust Analytical Q-ball Imaging'. *Magnetic Resonance in Medicine* **58**, 497–510. 3, 12, 26, 33, 83, 84, 85, 92, 101, 102, 103, 113, 114, 132, 133, 134, 149, 151, 165, 181, 183, 189, 191, 193, 198, 203, 206, 208, 212

Descoteaux, M. and R. Deriche: 2009, 'High Angular Resolution Diffusion MRI Segmentation Using Region-Based Statistical Surface Evolution'. *Journal of Mathematical Imaging in Vision* **33**, 239–252. 100, 228

Descoteaux, M., R. Deriche, D. Bihan, J. Mangin, and C. Poupon: 2010, 'Multiple q-Shell Diffusion Propagator Imaging'. *Medical Image Analysis*. 74, 80, 94, 103, 114, 138, 141, 142, 149, 151, 160

Descoteaux, M., R. Deriche, D. L. Bihan, J.-F. Mangin, and C. Poupon: 2009, 'Diffusion Propagator Imaging: Using Laplace's Equation and Multiple Shell Acquisitions to Reconstruct the Diffusion Propagator'. In: *IPMI*. 94, 114, 116, 138

Descoteaux, M., R. Deriche, T. R. Knösche, and A. Anwander: 2008a, 'Deterministic and Probabilistic Tractography Based on Complex Fiber Orientation Distributions'. *IEEE Transactions in Medical Imaging* **28**, 269–286. 3, 93, 101, 189, 228

Descoteaux, M., N. Wiest-Daessle, S. Prima, C. Barillot, and R. Deriche: 2008b, 'Impact of Rician Adapted Non-local Means Filtering on HARDI'. In: *MICCAI*. 114, 126

Do Carmo, M.: 1992, *Riemannian geometry*. Birkhauser. 36, 45

Douek, P., R. Turner, J. Pekar, N. Patronas, and D. LeBihan: 1991, 'MR color mapping of myelin fiber orientation'. *Journal of Computer Assisted Tomography* **15**, 923–929. 67

Dryden, I. and K. Mardia: 1998, *Statistical Shape Analysis*, Vol. 4. John Wiley & Sons Chichester, UK. 47, 186

Du, J., A. Goh, and A. Qiu: 2011, 'Diffeomorphic Metric Mapping of High Angular Resolution Diffusion Imaging based on Riemannian Structure of Orientation Distribution Functions'. *Arxiv preprint arXiv:1107.4763*. 228

Einstein, A.: 1956, *Investigations on the Theory of the Brownian Movement*. Dover Pubns. 63

Fick, D.: 1855, 'V. On liquid diffusion'. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **10**(63), 30–39. 63

Fillard, P., M. Descoteaux, A. Goh, S. Gouttard, B. Jeurissen, J. Malcolm, A. Ramirez-Manzanares, M. Reisert, K. Sakaie, F. Tensaouti, et al.: 2011, 'Quantitative evaluation of 10 tractography algorithms on a realistic diffusion MR phantom'. *NeuroImage*. 130, 152, 211, 237

Fillard, P., X. Pennec, V. Arsigny, and N. Ayache: 2007, 'Clinical DT-MRI Estimation, Smoothing, and Fiber Tracking With Log-Euclidean Metrics'. *IEEE Transactions On Medical Imaging* **26**, 1472–1482. 1, 72, 99, 100, 200, 208

Fletcher, P. and S. Joshi: 2007, 'Riemannian geometry for the statistical analysis of diffusion tensor data'. *Signal Processing* **87**, 250–262. 51, 56, 58, 99, 188

Fletcher, P., C. Lu, S. Pizer, and S. Joshi: 2004, 'Principal geodesic analysis for the study of nonlinear statistics of shape'. *Medical Imaging, IEEE Transactions on* **23**(8), 995–1005. 1, 50, 51, 100, 187, 188

Fletcher, P. T.: 2004, 'Statistical Variability in Nonlinear Spaces Application to Shape Analysis and DT-MRI'. Ph.D. thesis, University of North Carolina. 47, 48, 51, 99, 100, 184, 186

Fletcher, P. T., S. Venkatasubramanian, and S. Joshi: 2009, 'The geometric median on Riemannian manifolds with application to robust atlas estimation'. *NeuroImage* **45**, S143–S152. 47, 51, 52, 186, 201, 210, 218

Frank, L. R.: 2002, 'Characterization of Anisotropy in High Angular Resolution Dffusion-Weighted MRI'. *Magnetic Resonance in Medicine* **47**, 1083–1099. 33, 77, 78, 79, 113

Fréchet, M.: 1948, 'Les éléments aléatoires de nature quelconque dans un espace distancié'. *H. Poincaré,(10)* pp. 215–310. 48, 186

Ghosh, A.: 2011, 'High Order Models in Diffusion MRI and Applications'. Ph.D. thesis, INRIA Sophia Antipolis. 62, 234

Ghosh, A., R. Deriche, and M. Moakher: 2009, 'Ternary quartic approach for positive 4th order diffusion tensors revisited'. In: *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*. pp. 618–621. 78

Ghosh, A., M. Descoteaux, and R. Deriche: 2008a, 'Riemannian Framework for Estimating Symmetric Positive Definite 4th Oder Diffusion Tensors'. In: *MICCAI*, Vol. 66. pp. 308–319. 78

Ghosh, A., E. Ozarslan, R. Deriche, et al.: 2010, 'Challenges in Reconstructing the Propagator via a Cumulant Expansion of the One-Dimensional qspace MR Signal'. In: *ISMRM*. 77, 146

Ghosh, A., E. Tsigaridas, M. Descoteaux, P. Comon, B. Mourrain, and R. Deriche: 2008b, 'A polynomial based approach to extract the maxima of an antipodally symmetric spherical function and its application to extract fiber directions from the orientation distribution function in diffusion MRI'. In: *MICCAI*, Vol. 66. pp. 308–319. 234

Ghosh, A., D. Wassermann, and R. Deriche: 2011, 'A polynomial approach for maxima extraction and its application to tractography in HARDI'. In: *Information Processing in Medical Imaging*. pp. 723–734. 234

Goh, A., C. Lenglet, P. Thompson, and R. Vidal: 2009a, 'A nonparametric Riemannian framework for processing high angular resolution diffusion images (HARDI)'. In: *CVPR*. 101, 177, 179, 181, 182, 183, 188, 192, 200, 214

Goh, A., C. Lenglet, P. Thompson, and R. Vidal: 2009b, 'Estimating orientation distribution functions with probability density constraints and spatial regularity'. In: *MICCAI*. pp. 877–885, Springer. 193, 215

Goh, A., C. Lenglet, P. Thompson, and R. Vidal: 2011, 'A nonparametric Riemannian framework for processing high angular resolution diffusion images and its applications to ODF-based morphometry'. *NeuroImage*. 4, 57, 59, 101, 177, 179, 181, 182, 183, 188, 192, 199, 201, 208, 214

Gong, G., T. Jiang, C. Zhu, Y. Zang, Y. He, S. Xie, and J. Xiao: 2005, 'Side and handedness effects on the cingulum from diffusion tensor imaging'. *Neuroreport* **16**(15), 1701. 1, 101

Gradshteyn, I. and I. Ryzhik: 2007, *Table of Integrals, Series, and Products*. Elsevier, 7th edition edition. 32, 120, 136, 140, 165, 166, 195

Haberman, R.: 1987, *Elementary applied partial differential equations: with Fourier series and boundary value problems*. Prentice Hall, second edition edition. 12

Hagmann, P., L. Jonasson, P. Maeder, J.-P. Thiran, V. J. Wedeen, and R. Meuli: 2006, 'Understanding Diffusion MR Imaging Techniques: From Scalar Diffusion-weighted Imaging to Diffusion Tensor Imaging and Beyond'. *RadioGraphics* **26**, S205–S223. 62, 69, 81

Hess, C. P., P. Mukherjee, E. T. Han, D. Xu, and D. B. Vigneron: 2006, 'Q-Ball Reconstruction of Multimodal Fiber Orientations Using The Spherical Harmonic Basis'. *Magnetic Resonance In Medicine* **56**, 104–117. 2, 33, 83, 85, 103, 114, 133, 134, 181

Hosey, T., G. Williams, and R. Ansorge: 2005, 'Inference of multiple fiber orientations in high angular resolution diffusion imaging'. *Magnetic Resonance in Medicine* **54**(6), 1480–1489. 74, 75

Huettel, S. A., A. W. Song, and G. Mcgarthy: 2009, *Functional magnetic resonance imaging*. Sinauer Associates, second edition edition. 62, 65

Huzinaga, S.: 1965, 'Gaussian-Type Functions for Polyatomic Systems. I'. *The Journal of Chemical Physics* **42**, 1293. 96, 97, 135

Jian, B. and B. C. Vemuri: 2007, 'A Unified Computational Framework for Deconvolution to Reconstruct Multiple Fibers from Diffusion Weighted MRI'. *IEEE Transactions on Medical Imaging* **26**, 1464–1471. 93

Jian, B., B. C. Vemuri, E. Özarslan, P. R. Carney, and T. H. Marecid: 2007, 'A novel tensor distribution model for the diffusion-weighted MR signal'. *NeuroImage* **37**, 164–176. 93

Jiao, F., Y. Gur, C. Johnson, and S. Joshi: 2011, 'Detection of crossing white matter fibers with high-order tensors and rank-k decompositions'. In: *Information Processing in Medical Imaging*. pp. 538–549. 78

Johansen-Berg, H. and T. E. Behrens: 2009, *Diffusion MRI: From quantitative measurement to In vivo neuroanatomy*. Elsevier. 62, 64, 73

Jones, D. K., M. A. Horsfield, and A. Simmons: 1999, 'Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging'. *Magnetic Resonance in Medicine* **42**, 515–525. 72

Karcher, H.: 1977, 'Riemannian center of mass and mollifier smoothing'. *Communications on pure and applied mathematics* **30**(5), 509–541. 47, 48, 49, 51, 52, 186

Kendall, W.: 1990, 'Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence'. *Proceedings of the London Mathematical Society* **3**(2), 371. 47, 48, 49, 52, 186

Khachaturian, M. H., J. J. Wisco, and D. S. Tuch: 2007, 'Boosting the Sampling Efficiency of q-Ball Imaging Using Multiple Wavevector Fusion'. *Magnetic Resonance in Medicine* **57**, 289–296. 87, 90

Koay, C. G., L.-C. Chang, J. D. Carew, C. Pierpaoli, and P. J. Basser: 2006, 'a unifying theoretical and algorithmic framework for least squares methods of estimation in diffusion tensor imaging'. *Journal of Magnetic Resonance* **182**, 115–125. 72

Kreher, B., J. Schneider, I. Mader, E. Martin, J. Hennig, and K. Il'yasov: 2005, 'Multitensor approach for analysis and tracking of complex fiber configurations'. *Magnetic Resonance in Medicine* **54**(5), 1216–1225. 74

Kuchel, P., A. Coy, and P. Stilbs: 1997, 'NMR "diffusion-diffraction" of water revealing alignment of erythrocytes in a magnetic field and their dimensions and membrane transport characteristics'. *Magnetic Resonance in Medicine* **37**(5), 637–643. 77, 112

Lafferty, J. and G. Lebanon: 2006, 'diffusion kernels on statistical manifolds'. *Journal of Machine Learning Research* **6**(1), 129. 53

Le Bihan, D. et al.: 2003, 'Looking into the functional architecture of the brain with diffusion MRI'. *Nature Reviews Neuroscience* **4**(6), 469–480. 62

Lebanon, G.: 2006, 'Metric Learning for Text Documents'. *IEEE Transactions on Pattern Analysis And Machine Intelligence* **28**, 497–508. 53, 57

LeBihan, D., E. Breton, D. Lallemand, P. Grenier, E. Cabanis, and M. L. Jeantet: 1986, 'MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders'. *Radiology* **61**, 401–407. 67

Lenglet, C.: 2006, 'Geometric and Variational Methods for Diffusion Tensor MRI Processing'. Ph.D. thesis, INRIA Sophia Antipolis. 1, 62, 101

Lenglet, C., J. Campbell, M. Descoteaux, G. Haro, P. Savadjiev, D. Wassermann, A. Anwander, R. Deriche, G. Pike, G. Sapiro, K. Siddiqi, and P. Thompson: 2009, 'Mathematical Methods for Diffusion MRI Processing'. *NeuroImage* **45**, S111–S122. 189

Lenglet, C., M. Rousson, and R. Deriche: 2006a, 'DTI segmentation by statistical surface evolution'. *IEEE Transactions On Medical Imaging* **25**, 685–700. 1, 100, 101

Lenglet, C., M. Rousson, and R. Deriche: 2006b, 'Statistics on the Manifold of Multivariate Normal Distributions Theory and Application to Diffusion Tensor MRI Processingy'. *Journal of Mathematical Imaging and Vision* **25(3)**, 423–444. 1, 56, 58, 99, 100, 101, 186

Li, J., Y. Liu, W. Qin, J. Jiang, Z. Qiu, J. Xu, C. Yu, and T. Jiang: 2012, 'Age of Onset of Blindness Affects Brain Anatomical Networks Constructed Using Diffusion Tensor Tractography'. *Cerebral Cortex*. 1, 101, 228

Li, X., W. Hu, Z. Zhang, X. Zhang, M. Zhu, and J. Cheng: 2008, 'Visual tracking via incremental log-Euclidean Riemannian subspace learning'. In: *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. pp. 1–8. 47, 100

Li, Y., Y. Liu, J. Li, W. Qin, K. Li, C. Yu, and T. Jiang: 2009, 'Brain anatomical network and intelligence'. *PLoS computational biology* **5**(5), e1000395. 1, 101, 228

Lin, F., C. Yu, T. Jiang, K. Li, C. Zhu, W. Zhu, W. Qin, Y. Duan, Y. Xuan, H. Sun, et al.: 2006, 'Discriminative analysis of relapsing neuromyelitis optica and relapsing–remitting multiple sclerosis based on two-dimensional histogram from diffusion tensor imaging'. *NeuroImage* **31**(2), 543–549. 1, 101

Liu, C., R. Bammer, B. Acar, and M. E. Moseley: 2004, 'Characterizing Non-Gaussian Diffusion by Using Generalized Diffusion Tensors'. *Magnetic Resonance In Medicine* **51**, 925–937. 74, 75, 76, 77, 145

Liu, C., R. Bammer, and M. E. Moseley: 2003, 'Generalized diffusion tensor imaging (gdti): A method for characterizing and imaging diffusion anisotropy caused by non-Gaussian diffusion'. *Israel Journal of Chemistry* **43**, 145–154. 75, 77, 145

Mahalanobis, P.: 1936, 'On the generalized distance in statistics'. In: *Proceedings of the National Institute of Science, Calcutta*, Vol. 12. p. 49. 53

Mair, R., P. Sen, M. Hürlimann, S. Patz, D. Cory, and R. Walsworth: 2002, 'The narrow pulse approximation and long length scale determination in xenon gas diffusion NMR studies of model porous media'. *Journal of Magnetic Resonance* **156**(2), 202–212. 70

Marcinkiewicz, J.: 1939, 'Sur une propriété de la loi de Gauss'. *Mathematische Zeitschrift* **44**(1), 612–618. 77, 146

Mardia, K. and P. Jupp: 2000, *Directional statistics*. John Wiley & Sons Inc. 47

Maybank, S.: 2004, 'detection of image structures using the Fisher information and the Rao metric'. *IEEE transactions on pattern analysis and machine intelligence* pp. 1579–1589. 53

McGraw, T., B. Vemuri, R. Yezierski, and T. Mareci: 2006, 'Segmentation of High Angular Resolution Diffusion MRI Modeled as a Field of von Mises-Fisher Mixtures'. In: *ECCV*. 100, 177, 214

Merboldt, K., W. Hänicke, and J. Frahm: 1985, 'Self-diffusion NMR imaging using stimulated echoes'. *Journal of magnetic resonance* **64**(3), 479–486. 1

Merlet, S., J. Cheng, A. Ghosh, and R. Deriche: 2011, 'Spherical Polar Fourier EAP and ODF Reconstruction via Compressed Sensing in Diffusion MRI'. In: *ISBI*. 126, 181, 189, 191, 226, 227

Mio, W., D. Badlyans, and X. Liu: 2005, 'A computational approach to Fisher information geometry with applications to image analysis'. In: *Energy Minimization Methods in Computer Vision and Pattern Recognition*. pp. 18–33. 187

Moakher, M.: 2005, 'A differential geometric approach to the geometric mean of symmetric positive-definite matrices'. *SIAM Journal on Matrix Analysis and Applications* **26**(3), 735–747. 56, 58, 99

Mori, S.: 2007, *Introduction to Diffusion Tensor Imaging*. Elsevier. 72

Moseley, M., Y. Cohen, J. Mintorovitch, J. Kucharczyk, J. Tsuruda, P. Weinstein, and D. Norman: 1990, 'Evidence of Anisotropic Self-Diffusion'. *Radiology* **176**, 439–445. 67

Özarslan, E., C. Koay, and P. Basser: 2008, 'Simple harmonic oscillator based estimation and reconstruction for one-dimensional q-space mr'. In: *International Society for Magnetic Resonance in Medicine (ISMRM)*. p. 35. 95

Özarslan, E., C. Koay, T. Shepherd, S. Blackband, and P. Basser: 2009, 'Simple harmonic oscillator based reconstruction and estimation for three-dimensional q-space mri'. In: *International Society for Magnetic Resonance in Medicine (ISMRM)*. p. 1396. 74, 80, 92, 95, 96, 97, 135, 142, 149, 151, 182, 193, 215

Özarslan, E. and T. H. Mareci: 2003, 'Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging'. *Magnetic Resonance in Medicine* **50**, 955–965. 74, 77, 78, 79, 113, 143

Özarslan, E., N. Shemesh, C. Koay, Y. Cohen, and P. Basser: 2011, 'Nuclear magnetic resonance characterization of general compartment size distributions'. *New Journal of Physics* **13**, 015010. 228

Özarslan, E., T. Shepherd, C. Koay, S. Blackband, and P. Basser: 2012, 'Temporal scaling characteristics of diffusion as a new MRI contrast: Findings in rat hippocampus'. *NeuroImage*. 228

Özarslan, E., T. M. Shepherd, B. C. Vemuri, S. J. Blackband, and T. H. Mareci: 2006, 'Resolution of complex tissue microarchitecture using the diffusion orientation transform (DOT)'. *NeuroImage* **31**, 1086–1103. 12, 33, 77, 78, 80, 91, 92, 114, 116, 117, 139, 149, 151, 154, 160, 211, 232, 234

Özarslan, E., B. Vemuri, and T. Mareci: 2005, 'Generalized scalar measures for diffusion MRI using trace, variance, and entropy'. *Magnetic Resonance in Medicine* **53**(4), 866–876. 77, 78, 79

Pelletier, B.: 2005, 'Kernel density estimation on Riemannian manifolds'. *Statistics & probability letters* **73**(3), 297–304. 48

Pennec, X.: 2006, 'Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements'. *Journal of Mathematical Imaging and Vision* **25**(1), 127–154. 47, 48

Pennec, X., P. Fillard, and N. Ayache: 2006, 'A Riemannian Framework for Tensor Computing'. *International Journal of Computer Vision* **66**, 41–66. 1, 50, 56, 58, 99, 100, 186, 199, 200, 208

Pierpaoli, C. and P. Basser: 1996, 'Toward a Quantitative Assessment of Diffusion Anisotropy'. *Magnetic Resonance in Medicine* **36**, 893–906. 1, 72

Poupon, C.: 1999, 'Détection des faisceaux de fibres de la substance blanche pour l'étude de la connectivité anatomique cérébrale.'. Ph.D. thesis, Ecole Nationale Supérieure des Télécommunications. 72

Poupon, C., B. Rieul, I. Kezele, M. Perrin, F. Poupon, and J. Mangin: 2008, 'New diffusion phantoms dedicated to the study and validation of high-angular-resolution diffusion imaging (HARDI) models'. *Magnetic Resonance In Medicine* **60(6)**, 1276–1283. 130, 152, 237

Prckovska, V., A. Roebroeck, W. Pullens, A. Vilanova, and B. ter Haar Romeny: 2008, 'Optimal acquisition schemes in high angular resolution diffusion weighted imaging'. In: *MICCAI*, Vol. 5242. pp. 9–17. 113, 151

Qi, L., G. Yu, and E. Wu: 2010, 'Higher order positive semi-definite diffusion tensor imaging'. *SIAM Journal on Imaging Sciences* **3**(3), 416–433. 78

Rao, C.: 1945, 'Information and the accuracy attainable in the estimation of statistical parameters'. *Bull. Calcutta Math. Soc.* **37**, 81–91. 53, 54, 99

Read, A.: 1999, 'Linear interpolation of histograms'. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **425**(1-2), 357–360. 187

Ritchie, D. and G. Kemp: 2000, 'Protein docking using spherical polar Fourier correlations'. *Proteins: Structure, Function, and Bioinformatics* **39**(2), 178–194. 96, 97, 135

Schwartzman, A.: 2006, 'Random Ellipsoids And False Discovery Rates: Statistics For Diffusion Tensor Imaging Data'. Ph.D. thesis, Stanford. 1, 100

Shu, N., J. Li, K. Li, C. Yu, and T. Jiang: 2009, 'Abnormal diffusion of cerebral white matter in early blindness'. *Human brain mapping* **30**(1), 220–227. 1, 101

Skovgaard, L.: 1984, 'A Riemannian geometry of the multivariate normal model'. *Scandinavian journal of statistics* pp. 211–223. 55, 56, 58, 99

Söderman, O. and B. Jönsson: 1995, 'Restricted diffusion in cylindrical geometry'. *Journal of magnetic resonance. Series A* **117**(1), 94–97. 91, 149, 231

Spivak, M.: 1999, *A comprehensive introduction to differential geometry*, Vol. 1. Publish or perish Berkeley, 3rd edition edition. 36, 43, 44

Srivastava, A., I. Jermyn, and S. Joshi: 2007, 'Riemannian analysis of probability density functions with applications in vision'. In: *2007 IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1–8. 53, 57

Stejskal, E. and J. Tanner: 1965, 'Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient'. *The Journal of Chemical Physics* **42**, 288–292. 66, 67, 68

Subbarao, R. and P. Meer: 2009, 'Nonlinear mean shift over riemannian manifolds'. *International journal of computer vision* **84**(1), 1–20. 47

Taylor, D. and M. Bushell: 1985, 'The spatial mapping of translational diffusion coefficients by the NMR imaging technique'. *Physics in medicine and biology* **30**, 345. 1

Tomasi, C. and R. Manduchi: 1998, 'Bilateral filtering for gray and color images'. In: *ICCV*. pp. 839–846. 201

Tournier, J., F. Calamante, A. Connelly, et al.: 2007, 'Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution'. *NeuroImage* **35**(4), 1459–1472. 33, 93

Tournier, J.-D., F. Calamante, D. Gadian, and A. Connelly: 2004, 'direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution'. *NeuroImage* **23**, 1176–1185. 93

Tristán-Vega, A. and S. Aja-Fernández: 2010, 'DWI filtering using joint information for DTI and HARDI'. *Medical Image Analysis* **14**, 205–218. 114, 126

Tristán-Vega, A., C. Westin, and S. Aja-Fernández: 2009, 'Estimation of fiber orientation probability density functions in high angular resolution diffusion imaging'. *NeuroImage* **47**(2), 638–650. 33, 82, 87, 88, 89, 92, 103, 114, 134

Tristán-Vega, A., C.-F. Westin, and S. Aja-Fernández: 2010, 'a new methodology for the estimation of fiber populations in the white matter of the brain with the Funk-Radon transform'. *NeuroImage* **49**, 1301–1315. 33, 77, 82, 87, 88, 89, 90, 92, 103, 104, 105, 113, 114, 118, 119, 151, 181

Tschumperlé, D. and R. Deriche: 2003, 'Variational Frameworks for DT-MRI Estimation, Regularization and Visualization'. In: *ICCV*. 72

Tuch, D., R. Weisskoff, J. Belliveau, and V. Wedeen: 1999, 'High angular resolution diffusion imaging of the human brain'. In: *Proceedings of the 7th Annual Meeting of ISMRM*. 74

Tuch, D. S.: 2002, 'Diffusion MRI of Complex Tissue Structure'. Ph.D. thesis, MIT. 62, 80, 83

Tuch, D. S.: 2004, 'Q-ball imaging'. *Magnetic Resonance in Medicine* **52**, 1358–1372. 2, 83, 84, 90, 92, 100, 102, 103, 121, 131, 133, 181, 190, 191, 206, 212

Tuch, D. S., T. G. Reese, M. R. Wiegell, N. Makris, J. W. Belliveau, and V. J. Wedeen: 2002, 'High Angular Resolution Diffusion Imaging Reveals Intravoxel White Matter Fiber Heterogeneity'. *Magnetic Resonance in Medicine* **48**, 577–582. 2, 74

von dem Hagen, E. A. and R. M. Henkelman: 2002, 'Orientational Diffusion Reflects Fiber Structure Within a Voxel'. *Magnetic Resonance in Medicine* **48**, 454–459. 78

Wang, Q., T. Su, Y. Zhou, K. Chou, I. Chen, T. Jiang, and C. Lin: 2011, 'Anatomical insights into disrupted small-world networks in schizophrenia'. *Neuroimage*. 1, 101, 228

Wang, Z. and B. C. Vemuri: 2005, 'DTI Segmentation Using an Information Theoretic Tensor Dissimilarity Measure'. *IEEE Transactions on Medical Imaging* **24**, 1267–1277. 98

Wassermann, D., M. Descoteaux, and R. Deriche: 2008, 'Diffusion maps clustering for magnetic resonance Q-Ball imaging segmentation'. *International Journal of Biomedical Imaging* pp. 1–12. 100, 189

Wedeen, V., T. Reese, D. Tuch, M. Weigel, J.-G. Dou, R. Weiskoff, and D. Chessler: 2000, 'Mapping fiber orientation spectra in cerebral white matter with Fourier transform diffusion MRI'. In: *Proceedings of the International Society of Magnetic Resonance in Medicine*. p. 82. 80, 81, 83

Wedeen, V. J., P. Hagmann, W.-Y. I. Tseng, T. G. Reese, and R. M. Weisskoff: 2005, 'Mapping Complex Tissue Architecture With Diffusion Spectrum Magnetic Resonance Imaging'. *Magnetic Resonance In Medicine* **54**, 1377–1386. 80, 81, 83, 89, 201, 206

Weldeselassie, Y., A. Barmpoutis, and M. Atkins: 2010, 'Symmetric Positive-Definite Cartesian Tensor Orientation Distribution Functions (CT-ODF)'. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*. 78

Westin, C.-F., S. Maier, H. Mamata, A. Nabavi, F. Jolesz, and R. Kikinis: 2002, 'Processing and visualization for diffusion tensor MRI'. *Medical Image Analysis* **6**, 93–108. 73

Wootters, W.: 1981, 'Statistical distance and Hilbert space'. *Physical Review D* **23**(2), 357. 53, 57

Wu, Y.-C. and A. L. Alexander: 2007, 'Hybrid diffusion imaging'. *NeuroImage* **36**, 617–629. 82, 120, 200

Wu, Y.-C., A. S. Field, and A. L. Alexander: 2008, 'Computation of Diffusion Function Measures in q -Space Using Magnetic Resonance Hybrid Diffusion Imaging'. *IEEE Transactions On Medical Imaging* **27**, 858–865. 82, 87, 88, 103

Yu, C., J. Li, Y. Liu, W. Qin, Y. Li, N. Shu, T. Jiang, and K. Li: 2008, 'White matter tract integrity and intelligence in patients with mental retardation and healthy adults'. *Neuroimage* **40**(4), 1533–1541. 1, 101