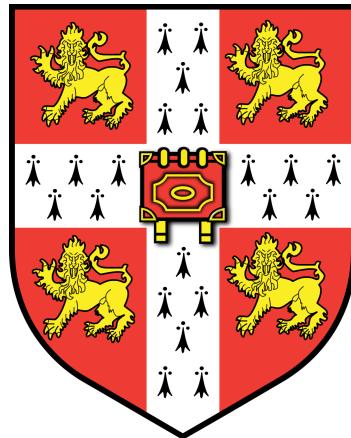


TOWARDS AN ACCURATE BRAIN TRACTOGRAPHY



ELEFTHERIOS GARYFALLIDIS

Wolfson College
University of Cambridge
United Kingdom

May 2012

A dissertation submitted to the University of Cambridge for the degree
of Doctor of Philosophy

Supervisors

DR. IAN NIMMO-SMITH

Medical Research Council
Cognition and Brain Sciences Unit
Cambridge, UK

DR. GUY B. WILLIAMS

Wolfson Brain Imaging Centre
Department of Clinical Neurosciences
University of Cambridge
Cambridge, UK

Dedication

To *my parents*
STAVROS AND AMALIA

for his creativity and her patience.

I needed both to excel.

To *the giants of*
OPEN SOURCE SOFTWARE

who pushed the world forward.

Acknowledgements

This PhD was funded by the Engineering and Physical Sciences Research Council, the Vergottis Foundation, the Board of Graduate Studies of the University of Cambridge, the Medical Research Council and Wolfson College, Cambridge, United Kingdom.

I would like to thank my supervisor, Dr. Ian Nimmo-Smith, for the excellent guidance and advice throughout the entire course of my PhD. He has been an endless source of inspiration and support for my work. I am also thankful to my second supervisor Dr. Guy Williams for his research suggestions and encouragement. The combination of the statistical knowledge from my first supervisor and the knowledge in MR physics from my second supervisor was a key component for the completion of this thesis.

I am grateful to Dr. Matthew Brett who had been such a great friend and promoter of my work. His suggestions have been enormously helpful especially on software development. My great appreciation to my friend and collaborator Dr. Marta Correia who helped with designing and collecting the MR data sets used in this work. The same goes for dearest Mr. John Griffiths and Dr. Virginia Newcombe who were the first users of my work in Cambridge.

A big thank you to Stephan Gerhard with whom I worked to give birth to a new visualization library, to Frank Yeh for his suggestions on tractography algorithms and to Dr. Emanuele Olivetti on Machine Learning approaches.

My great respect to Dr. David Jarvis Alumni Dean of Wolfson College, Professor Mike Proctor and Professor Peter Haynes from the Department of Applied Mathematics and Theoretical Physics who helped me continue my career in research and supported me at the most difficult moments of my life. My gratitude also to Dr. Rik Henson who introduced me to my supervisors at my current department.

Life would be unbearable without friends. Much love and gratitude to my great friends from my homeland: Vassilis Dimitriadis, Dimitris Pritsos, Alexandros Triantafyllidis, Dr. Periklis Akritidis, Iason Oikonomidis, Dr. Nikos Grigoropoulos, Dr. Panagiotis Lavvas and Dr. Vassilis Tsiaras. And the great friends I made here in Cambridge: Dr. Euan Spence, Dr. Ian Charest, Dr. Andreas Georgiou, Dr. Daniel Holland,

Dr. Mina Mprimpari, Dr. Nikos Dikaios, Foivos Karachalios, Yiannis Mattheoudakis, Dr. Ioanna Boulouta, Dr. Johan Carlin, Dr. Annika Linke, Alex Walther, Rowan Grant, Michael Worthington, Dimitra Datsiou, Dr. Pierre De Fouquieres, Dr. Kaoruko Yamazaki and my wonderful house mate Dr. Magda Rapti.

I should not forget my first teachers who helped me develop my scientific abilities and strengthened my courage: Professor Nikolaos Vassilas, Professor Antonis Argyros, Professor Alexandros Tomaras, Stelios Kastridis and Vassilis Valavanis. My great acknowledgement to my college tutor Dr. Giles Yeo who was always available to help with any bureaucratic issues. Finally, I would like to thank my fellow colleagues and developers at NeuroImaging in Python for their hard work and devotion to the principles of free software and open science.

Disclaimer

This dissertation is the result of my own work and contains nothing which is the outcome of work done in collaboration with others, except where stated explicitly.

No part of this dissertation has previously been submitted for any degree or diploma at any institution.

This dissertation does not exceed 60,000 words in length (including tables, footnotes, bibliography and appendices).

Throughout this dissertation the plural pronoun 'we' is used for stylistic reasons and should be taken to refer to either the singular author, the reader and the author or, when stated explicitly, the author and collaborators. The form assumed should be apparent from the context.

Eleftherios Garyfallidis
Cambridge, 14 May 2012

Publications

Conferences

Garyfallidis E, Brett M, Nimmo-Smith I (2010), “*Fast Dimensionality Reduction for Brain Tractography Clustering*”, 16th Annual Meeting of the Organization for Human Brain Mapping.

Garyfallidis E, Brett M, Tsiaras V, Vogiatzis G, Nimmo-Smith I (2010), “*Identification of corresponding tracks in diffusion MRI tractographies*”, 18th Proceedings of the International Society of Magnetic Resonance in Medicine.

Correia MM, Williams GB, Yeh F-C, Nimmo-Smith I, **Garyfallidis E** (2011), “*Robustness of diffusion scalar metrics when estimated with Generalized Q-Sampling Imaging acquisition schemes*”, 19th Proceedings of the International Society of Magnetic Resonance in Medicine.

Garyfallidis E, Brett M, Amirbekian B, Nguyen C, Yeh F-C, Olivetti E, Halchenko Y, Nimmo-Smith I (2011), “*Dipy - a novel software library for diffusion MR and tractography*”, 17th Annual Meeting of the Organization for Human Brain Mapping.

Garyfallidis E, Gerhard S, Avesani P, Nguyen T, Tsias V, Nimmo-Smith I, Olivetti E (2012), “*A software application for real-time, clustering-based exploration of tractographies*”, 18th Annual Meeting of the Organization for Human Brain Mapping.

Olivetti E, Nguyen TB, **Garyfallidis E** (2012), “*The Approximation of the Dissimilarity Projection*”, 2nd IEEE International Workshop on Pattern Recognition in NeuroImaging.

Garyfallidis E, Nimmo-Smith I (2012), “*Cartesian grid q-space reconstruction*”, HARDI Reconstruction Workshop of the 9th IEEE International Symposium on Biomedical Imaging.

Journals

Garyfallidis E, Brett M, Correia M, Williams GB, Nimmo-Smith I (2012), “*QuickBundles for tractography simplification*”, Frontiers in Brain Imaging Methods, submitted.

Chamberlain SR, Hampshire A, Menzies LA, **Garyfallidis E**, Grant JE, Odlaug BL, Craig K, Fineberg N, Sahakian BJ (2010), “*Reduced brain*

white matter integrity in trichotillomania: a diffusion tensor imaging study.",
Archives of General Psychiatry 67(9):965-71.

Outreach

Garyfallidis E, Nimmo-Smith I (2010), "Surfing your Brain Super-Highways". Presented at the 350th celebration of the Royal Society, London, UK. In the presence of Her Majesty Queen Elizabeth II and His Royal Highness the Duke of Edinburgh.

Software

Diffusion Imaging in Python (DIPY): Available at dipy.org.

Free On Shades (FOS): Available at fos.me.

Abstract

The objective of this thesis is to improve on the methods for inferring neural tracts from diffusion weighted magnetic resonance imaging (dMRI). Accordingly, I present improvements to the reconstruction, integration, segmentation and registration modalities of dMRI analysis.

I compare and evaluate different Cartesian-grid q-space dMRI acquisition schemes, using methods based on the Fourier transform of the diffusion signal, with reconstructions by diffusion spectrum imaging or generalised q-ball imaging methods. I propose a new reconstruction method called diffusion nabla imaging (DNI) which works with all these acquisition schemes, using an algorithm that directly approximates the orientation distribution function using the Laplacian of the signal in q-space. DNI has impressive accuracy on low angle crossings.

Most previously published reconstruction methods are closely linked to their own specific track integration method. I have formulated a general, non-inferential, deterministic tractography algorithm (EuDX) which is based on Euler integration and trilinear interpolation, which works with voxel level information about fibre orientations including multiple crossings, and employs a range of stopping criteria. The purpose of this algorithm is to be faithful to the reconstruction results rather than try to correct or enhance them by introducing regional or global considerations.

I have developed an entirely new, fully automatic, linear time, clustering method (QuickBundles) which reduces massive tractographies to just a few bundles. These bundles are characterised by representative tracks which are multi-purpose and can be used for interaction with the data or as the basis for applying higher-complexity clustering methods which would have been impossible or too slow with the full data set. QuickBundles is currently the fastest known tractography clustering algorithm.

After applying QuickBundles to tractographies from different subjects, I show how to use the representative tracks to identify robust landmarks within each subject which I use to directly register the different tractographies together in a highly efficient way. The resulting correspondences provide important evidence for the anatomical plausibility of the derived bundles. I demonstrate how these methods can be used for group analysis, and for atlas creation.

This thesis contributes to the understanding of the diffusion signal in

the context of dMRI acquisitions and builds on this foundation towards a more robust brain tractography which approximates more closely the underlying fibre architecture.

Contents

Title	1
Supervisors	i
Dedication	ii
Acknowledgements	iii
Disclaimer	v
Publications	vi
Abstract	viii
Contents	xii
1 Background	1
1.1 Introduction	1
1.2 Molecular diffusion	1
1.3 Acquisition sequences	2
1.4 Single gradient signal models	6
1.5 Q-space reconstruction	9
1.6 Diffusion Tensor	12
1.7 Orientation Distribution	13
1.8 Tractography	14
1.9 Known problems	21
1.10 Segmentation	22
1.11 Foreword	25
2 Cartesian Lattice	
Q-space Reconstructions	27
2.1 Overview	27
2.2 Theory	28
2.3 Other methods	32
2.4 Diffusion Nabla Imaging	34
2.5 Equatorial Inversion Transform	35
2.6 Implementation	37
2.6.1 Standard EIT	37

2.6.2	Fast EIT	38
2.7	Peak Finding	39
2.8	Spherical Angular Smoothing	40
2.9	Comparisons and Results	42
2.9.1	Multi-fibre Simulations	42
2.9.2	Software Phantoms	47
2.9.3	Results with software phantoms	49
2.9.4	Results with humans	52
2.10	Anisotropy metrics	58
2.10.1	Non-parametric Anisotropy	58
2.10.2	Quantitative Anisotropy	60
2.10.3	Robustness of QA	62
2.11	Discussion and Conclusion	67
3	Tracking with Euler Delta Crossings	69
3.1	Overview	69
3.1.1	Local	70
3.1.2	Global	72
3.1.3	Simulated	74
3.2	The EuDX Algorithm	75
3.3	Results with software phantoms	80
3.4	Results with humans	88
3.5	Conclusion	93
4	Highly Efficient Tractography Clustering	95
4.1	Overview	95
4.2	Track distances and preprocessing	96
4.3	Related Work	98
4.4	Data sets	102
4.5	QuickBundles (QB) Clustering	103
4.5.1	The QB Algorithm	103
4.5.2	Powerful simplifications	108
4.5.3	Complexity and timings	111
4.5.4	Virtual tracks, exemplar tracks and other descriptors.	113
4.6	Comparisons within- and between-subjects	114

4.6.1	Comparison of clusterings	114
4.6.2	Robustness under reordering	116
4.6.3	Measures to compare classifications	117
4.6.4	Bundle Adjacency	122
4.7	Parallel version	123
4.7.1	Algorithm	123
4.7.2	Merging two sets of bundles	124
4.8	Direct applications	124
4.8.1	Rapidly detecting erroneous tracks	124
4.8.2	Alignments, landmarks and atlases	126
4.8.3	QB as input to other learning methods	129
4.8.4	Exemplars vs ROIs vs Masks	132
4.9	Direct Tractography Registration	133
4.10	Bundle Quality Control	137
4.11	Discussion and conclusion	138
5	Conclusion	141
5.1	Summary	141
5.2	Software	142
5.3	Future work	143
5.4	New frontiers	146
A	Appendix	147
A.1	Data and sequences	147
A.2	The Cosine Transform	148
A.3	Fourier Transform of $P(\mathbf{r})r^2$	148
A.4	Radial projection of a symmetric function	149
A.5	The Tensor in GQI	150
A.6	Affinity Propagation	153

1 Background

1.1 Introduction

Diffusion MRI (dMRI) is the principal non-invasive method that provides information about the directional structure of neural tracts found in white matter and the cortex. dMRI acquires one or more T_2 -weighted reference images, and a collection of diffusion-weighted images (see Fig. 1.1) that attenuate the T_2 signal according to the amount of diffusion along prescribed gradient directions [1]. The information is not complete and the tracts cannot be reconstructed in full detail [2]. However, some spatial structures and patterns can be visualised. These are usually represented as trajectories [3, 4] or connectivity maps [5]. The unique new area of study that aims to reconstruct the neural tracts from diffusion data is called diffusion tractography. Other types of tractography are based in staining using for example luxol-fast blue [6] but these can only be used with *in vitro* brains and they lack ease of reproducibility. For non-human brains as for example in macaque there are even *in vivo* methods for tracing down to single axons [7] however, these are not available or recommended for human studies as they are highly invasive.

1.2 Molecular diffusion

Molecular diffusion is a process that occurs incessantly in biological materials, fluids in particular, and accounts for a number of interesting phenomena; the dMRI signal measures the history of random (Brownian) displacements of spin-labelled hydrogen protons (spins) resolved in the direction of a magnetic field gradient. Though the actual probability displacement function of the protons is unaffected by the presence or variation in the magnetic field, the cumulative phase change in the spins reflects the changes in the position-dependent spin frequency induced by the field gradient. Components of the diffusion motion along the direction of the gradient induce such changes. The signal change due to cumulative dephasing is greatest when this coincides with a direction that allows greater random displacements, e.g. because of the orientation of a microstructure within which the proton is moving. It is this link between the directional dependence of the dMRI signal and the orientations

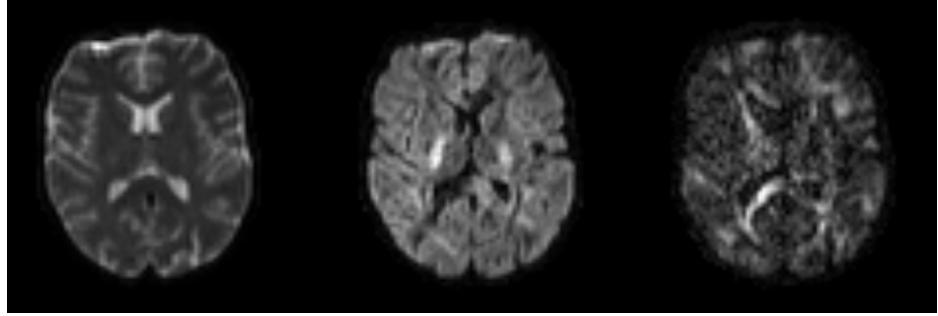


Figure 1.1: Three slices from diffusion data sets gathered with zero gradient strength on the left, medium gradient strength on the middle and high gradient strength on the right.

of the supposed underlying brain fibres that provides the unique insights of diffusion tractography. In dMRI we observe that the protons will move more along the directions of the axons, and move less perpendicular to that direction.

Anisotropy is one of the terms that are very common in diffusion terminology. Anisotropy means that the mean square displacement of the particles is greater along some directions than along others. On the other hand, isotropy means that the mean square displacement is equal in all directions i.e. complete lack of anisotropy. It is this level of anisotropy that is the basis of dMRI as a method of investigation of the structure of biological materials. For a biological interpretation of the signal measured with dMRI see [8], [9] and [1] p. 105.

1.3 Acquisition sequences

MRI data are collected by changing certain magnetic fields on and off in a prescribed sequence, known as pulse sequence [1, 10]. The pulse sequence determines the content, quality, contrast and resolution of the image. MR images primarily reflect the signal from hydrogen nuclei from water and fat concentrations. The hydrogen nuclei possess a magnetic dipole which is often referred to as spin. These dipoles can align themselves with an externally applied magnetic field. The MRI scanner generates a strong, static magnetic field B_0 which is typically measured in Tesla (T). A second magnetic field is applied for only a brief duration and oscillates at radio frequencies; known as the RF pulse [11].

RF pulses are used primarily for excitation and refocusing. In the excitation phase spins will rotate away from their preferred orientation along

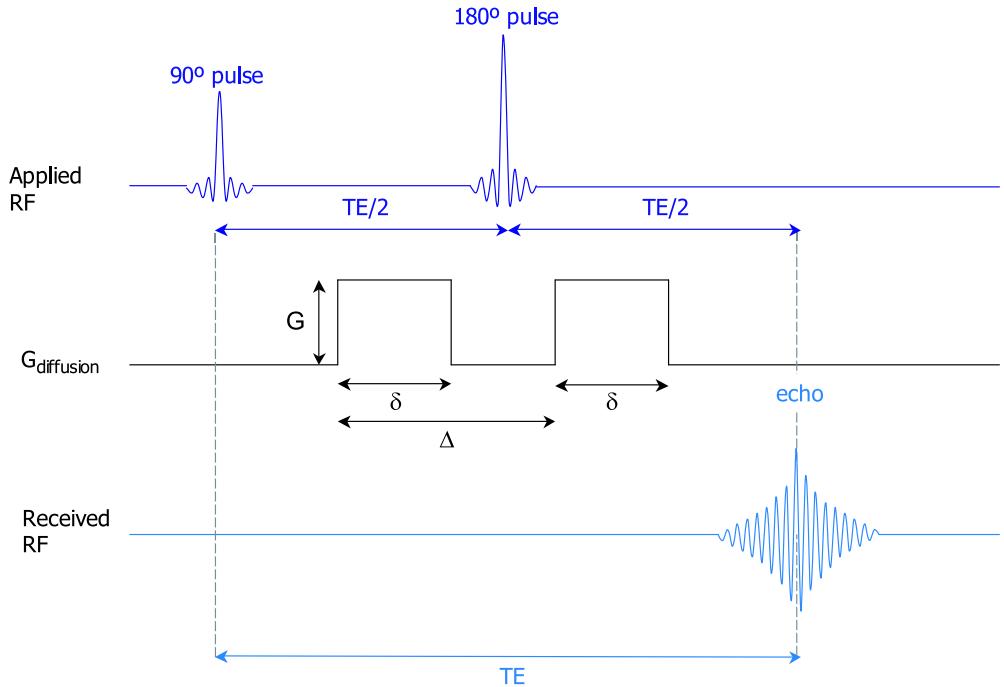


Figure 1.2: Pulsed Gradient Spin Echo (PGSE)

B_0 . Excited spins precess about B_0 at a frequency ν given by the Larmor equation $\nu = \gamma B$ where γ is a constant known as the gyromagnetic ratio. The precessing part that is perpendicular to the direction of B_0 decays exponentially with a time constant T_2 and the spins realign themselves exponentially in the direction of B_0 with a time constant T_1 . T_1 and T_2 vary with tissue but $T_2 < T_1$ for the same tissue type [12]. The generated magnetic field from the coherently precessing spins induces a current in the receiver coils; this current is the signal used to generate MR images and corresponds to image brightness. The more coherent the phase of the precessing spins the higher the brightness in the image pixels. However, with time, spins lose their phase coherence. Signal loss from both T_2 decay and dephasing is called T_2^* signal loss ($T_2^* < T_2$). Often, a second RF pulse is applied at some time $TE/2$ after excitation and flips the spins in the plane perpendicular to B_0 . If the conditions stay the same all spins will be back in phase at a time TE after the excitation pulse. The moment of spin refocus is called a spin echo and creates the measured signal. Acquisition sequences which use a refocusing pulse are called spin echo pulse sequences; and gradient echo sequences otherwise.

The additional magnetic fields generated by an MRI scanner are called magnetic field gradients or simply gradients (G). Including the applied

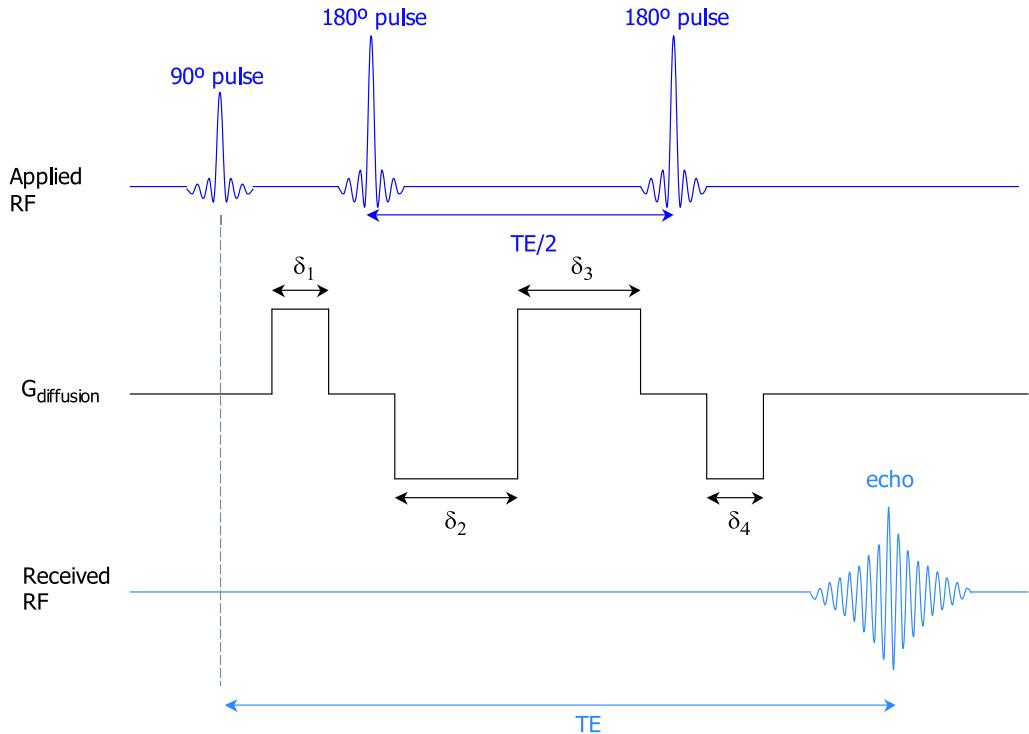


Figure 1.3: Twice-Refocused Spin Echo (TRSE)

gradients the magnetic field in the scanner is given by $B = B_0 + G_x(t)x + G_y(t)y + G_z(t)z$ where x, y and z the three orthogonal directions. Gradients have a special role in diffusion weighting as we will discuss next.

The best known pulse sequence for generating diffusion-weighted images is called Pulsed Gradient Spin Echo method (PGSE), also known as the Stejskal and Tanner method [13]. This has 90° - 180° spin echo pair of RF pulses with one gradient before the second pulse and one equal gradient after the second pulse [12] (see Fig. 1.2). The refocusing is perfect only when the spins do not move between the two pulses. The diffusion weighted contrast acts as an inverse T_2 weighting i.e. tissues with mobile water molecules give lower signal than more solid tissues with smaller mobility.

Eddy currents caused by the onset and offset of the gradients are a problem with PGSE and most recent systems (including Siemens scanners) use Twice-Refocused Spin Echo (TRSE) sequences [14] to reduce these artefacts. Every time the magnetic field gradients switch they generate currents that produce other smaller magnetic fields which disturb the spins. The TRSE sequence is an improvement on the PGSE. This improvement is achieved by the use of another refocusing pulse surrounded by

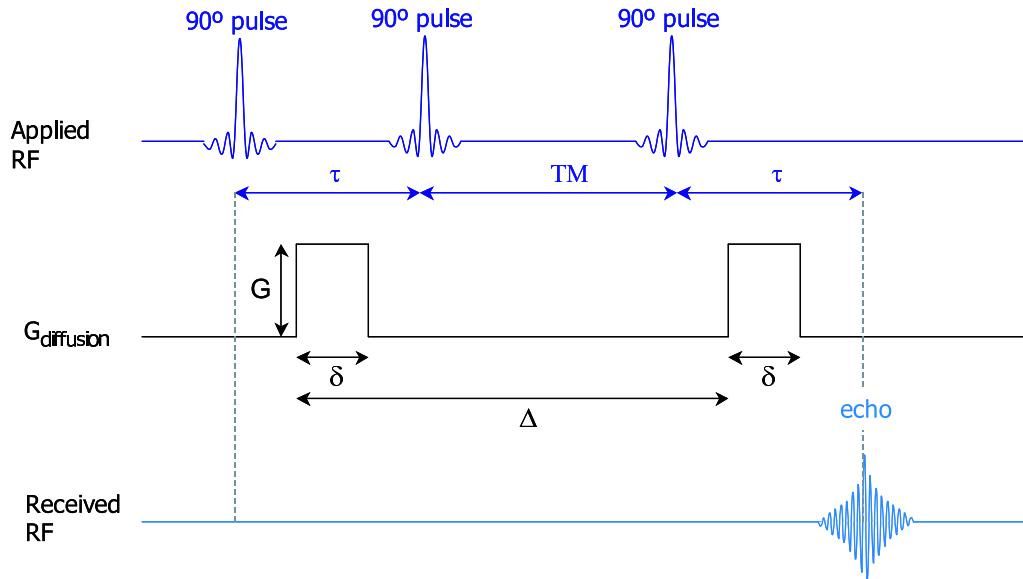


Figure 1.4: STimulated Echo Acquisition Mode (STEAM)

the inverse mirror of the previous diffusion gradients (see Fig. 1.3). By adjusting the timing of the diffusion gradients, eddy currents can be nulled or greatly reduced. This sequence improves the image quality without loss of scanning efficiency i.e. TR duration and it is the standard in most modern MRI scanners.

In the experiments described in this thesis we used a recent (2010) Work In Progress (WIP) protocol from Siemens which uses the STEAM (STimulated Echo Acquisition Mode) sequence [15]. STEAM, is presented in Fig. 1.4 and works in the following way: Three 90° pulses are used to produce a stimulated echo. The first two pulses are separated by a time delay τ . After the same delay τ following the same pulse, a stimulated echo is produced. In order to introduce diffusion weighting into the stimulated echo, two identical diffusion gradient lobes are applied, one during the first and one during the second τ interval. Because the magnetization of the stimulated echo is stored along the longitudinal axis between the second and the third RF pulses, it does not experience any T_2 or T_2^* dephasing during the time interval TM . TM , however, does contribute to the diffusion gradient separation time, Δ . Thus, a high b-value can be obtained without incurring the TE -induced signal loss, as compared to the standard spin echo sequence. The signal amplitude of the stimulated echo is, however, less than that of the corresponding spin-echo sequence with the same TE , because the maximum amplitude of the stimulated echo is

one-half of a spin echo.

STEAM was the preferred sequence for most of the experiments used in this thesis as it gave higher SNR overall which was an advantage crucial at high b-values as the signal in those b-values can be quite low.

1.4 Single gradient signal models

Under the Brownian motion assumption, the diffusion signal strength is described by the following model known as the Stejskal-Tanner [13] formula

$$S_b = S_0 e^{-bD} \quad (1.1)$$

where S_0 is the measured signal when no gradient direction is applied, D is the diffusion coefficient that we wish to measure and b is the b-value – the crucial experimental diffusion weighting parameter which summarises the amount of diffusion sensitising gradient history. D is often referred to as the diffusivity value or apparent diffusivity coefficient (ADC). The units of D are mm²/sec (for water at 37° $D \approx 3 \times 10^{-3} m^2/sec$), and of b are sec/mm², typically in the range of 0–5,000 sec/mm² though some acquisition paradigms can call for very much larger values e.g. greater than 10,000 sec/mm²[16], [3]. In Fig. 1.5 we see the signal decay for different b-values and specific diffusivities. The obvious conclusion here is that signal in areas with high diffusivity, as in the corticospinal fluid (CSF) where water persists, will always be lower than the signal from areas of lower diffusivity as those found in fibrous brain structures, as the corpus callosum. Furthermore, there is lowest diffusivity and lowest signal loss when the gradient direction is into the wall of a fibre; along the fibre has highest diffusivity and highest signal loss. Free water has uniform diffusivity and uniform signal loss towards all gradient directions. Signal decreases exponentially with b . Finally, we can see that the “directional effect”, i.e. the contrast between ‘along’ and ‘transverse’ signal is greatest when $b \sim 1,000$.

The b-value b or *diffusion weighting* is a function of the strength, duration, temporal spacing and timing parameters of the specific paradigm. This function is derived from the Bloch-Torrey equations [17]. In the case of the classical Stejskal-Tanner pulsed gradient spin-echo (PGSE) sequence (see section 1.3), at the time of readout

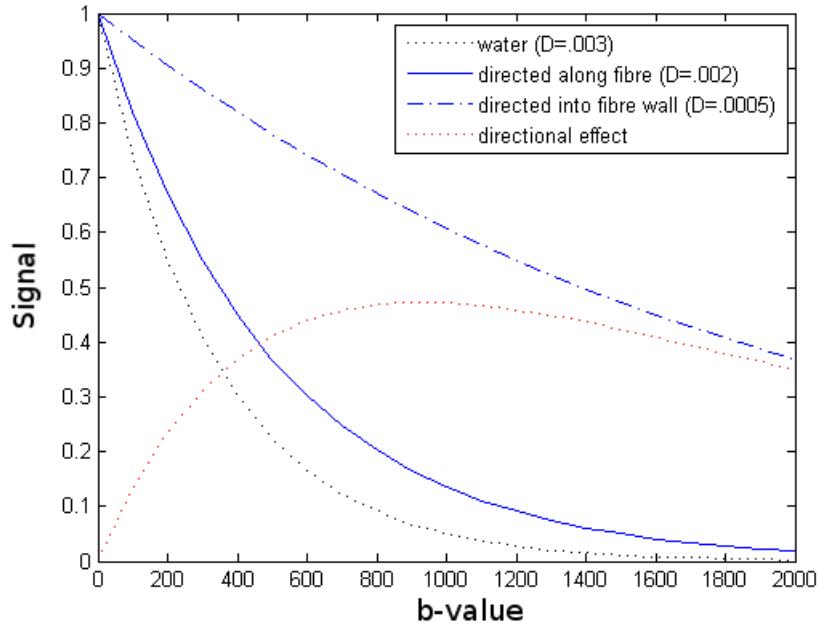


Figure 1.5: Signal as a function of b for various values of D . The directional effect is the difference between signals from gradients along and perpendicular to the fibre.

$$b = \gamma^2 G^2 \delta^2 \left(\Delta - \frac{\delta}{3} \right),$$

where γ is the gyromagnetic ratio, δ denotes the pulse width, G is the gradient amplitude and Δ the centre to centre spacing. γ is a constant which depends on the nucleus, but we can change the other three parameters and in that way control the b-value.

Although the PGSE is useful for expository clarity, in reality as indicated in section 1.3 more complicated but related sequences such as the twice-refocused spin-echo (TRSE) [18, 14] sequence and subsequent refinements such as TRASE [19] are employed as a means of removing the distortion effects from eddy currents resulting from the initial and final ramps of the gradient pulses.

An important point is that we can control the size of b-values by changing the strength and timings of the gradient pulses, and that depending on the b-value, we can expect different amount of signal loss. In Fig. 1.6 we see the directional dependence of the simulated signal of a single fibre oriented at 0° for two b-values. Note that the signal is lowest in the direction of the fibre both at high and low b-values.

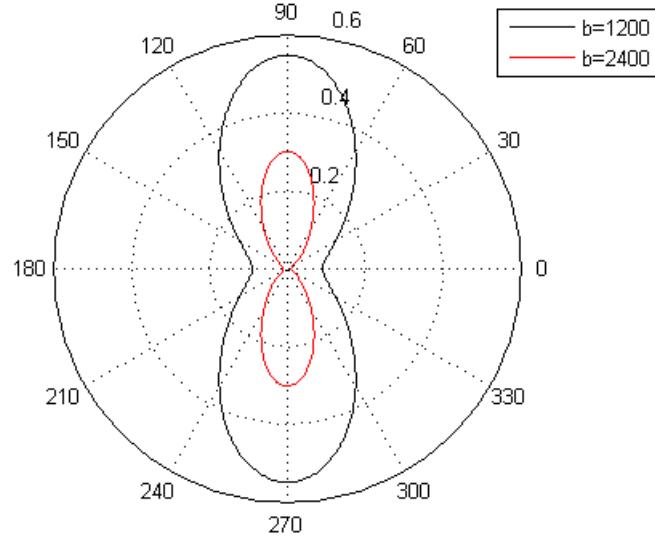


Figure 1.6: Directional dependence of signal for two b -values. Signal is drawn as a function of direction for a Gaussian diffusion function with a horizontal principal direction (0°) and two values of b . D is set to 0.002.

In Eq. 1.1 we assume that the signal can be expressed by a single exponential term (mono-exponential). In fact, there is evidence that this assumption may break down at higher b -values and more complicated models have been proposed in order to deal with this issue. One of these, is the multi-exponential model used by Niendorf et al. [20], Mulkern et al. [21] and Ozarslan et al. [22] which is expressed as

$$S_b = S_0 \sum_i^N f_i e^{-bD_i} \quad (1.2)$$

where N is the number of exponents or compartments, D_i is the i -th diffusion coefficient and f_i is the volume fraction of the i -th compartment.

Another model of higher order is found in diffusion kurtosis imaging (DKI) proposed by Jensen et al. [23]

$$\ln(S_b) = \ln(S_0) - bD + \frac{1}{6}b^2 D^2 K + O(b^3) \quad (1.3)$$

where K is the apparent diffusion kurtosis coefficient.

By putting together these single gradient models (Eq. 1.1, 1.2, 1.3) for every gradient direction we can create systems of equations where we can

fit and identify their unknown parameters. In this thesis, we will try to avoid fitting. We will concentrate on reconstructing the diffusion signal using a non-parametric Fourier-based approach applied to the combined information from many gradient directions.

1.5 Q-space reconstruction

Bloch and Torrey [24] established differential equations governing MR diffusion in non-isotropic magnetic fields by analogy with Fick's Laws [25] for spontaneous dispersion along concentration gradients of inhomogeneous substances. Callaghan [17] showed how these bulk properties can be derived by statistical methods from the collective spin histories of individual protons. When a molecule is at position x_0 , we cannot read exactly where it will be after time t , we can only model a distribution of possible locations. This motion is described by a propagator $P(x; x_0, t)$ which defines the probability of being in x after a time t , starting at x_0 .

Stejskal and Tanner [13] showed that the spin echo magnitude $S(\mathbf{q}, t)$ from a pulsed gradient spin echo (PGSE) experiment (see section 1.3) is directly related to the diffusion propagator by the following (inverse) Fourier relation

$$S(\mathbf{q}, t) = S_0 \int P(\mathbf{r}, t) e^{i\mathbf{q} \cdot \mathbf{r}} d\mathbf{r} \quad (1.4)$$

where S_0 is the signal in the absence of the applied magnetic diffusion gradient \mathbf{g} , \mathbf{r} is the relative spin displacement $\mathbf{x} - \mathbf{x}_0$ at diffusion time t , \mathbf{q} is the spin displacement wave vector. \mathbf{q} is parallel with the applied magnetic gradient \mathbf{g} . With the corresponding direct Fourier transform we can reconstruct the diffusion propagator P by measuring the signal in a number of different directions and gradient magnitudes. Q-space imaging (QSI) and Diffusion Spectrum Imaging (DSI) are the best known methods which reconstruct the diffusion propagator in that way. In Eq. 1.4 the spin density is implicit in S_0 , in a later chapter we will see a more general interpretation of this relationship (see Eq. 2.1).

Q-space is the space defined by the coordinates of the 3D spin displacement wave vectors \mathbf{q} as shown in Eq. 1.4. The vector \mathbf{q} parametrises the space of diffusion acquisitions. It is related to the applied magnetic diffusion gradient \mathbf{g} by the formula $\mathbf{q} = (2\pi)^{-1}\gamma\delta\mathbf{g}$ [26]. Every single vector \mathbf{q} has the same orientation as the direction of diffusion gradient \mathbf{g} and length proportional to the strength g of the gradient field. We have

also $\mathbf{q} = k\sqrt{b}\hat{\mathbf{g}}$, where b is the b-value, $\hat{\mathbf{g}}$ is the unit gradient direction, and k is a multiplication constant which is a function of the timing parameters of the acquisition scheme.

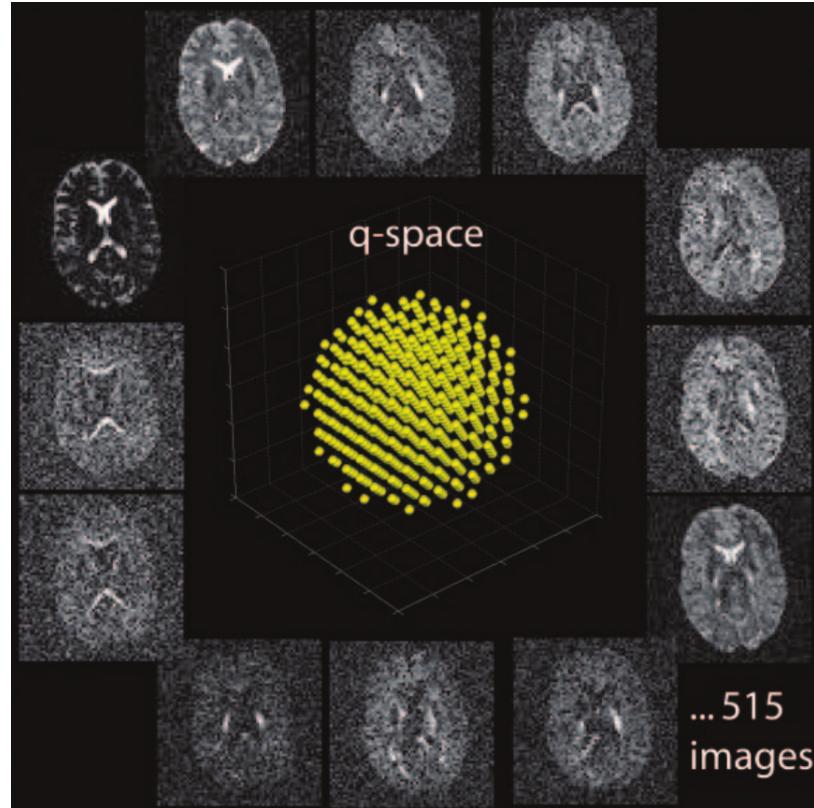


Figure 1.7: One volume is collected for every sampling point in q-space. Picture adapted from Hagmann et al. [27].

Every single point in q-space corresponds to a diffusion weighted image i.e. a 3D brain volume of measured signal for a specific gradient direction and strength (see Fig. 1.7). If for example we have programmed the scanner to apply 60 gradient directions then our data should have 60 diffusion volumes with each volume obtained for a specific gradient. A Diffusion Weighted Image (DWI) is the volume acquired from only one direction gradient. Hence, in the previous example we would gather 60 DWI volumes corresponding to 60 locations (\mathbf{q} -values) in \mathbf{q} -space. An alternative way to think of \mathbf{q} is in mathematical terms as the combination of parameters which produces the inverse Fourier transform relationship between the diffusion signal and the probability displacement distribution. In these terms (see Callaghan [17]) \mathbf{q} is the reciprocal of the probability displacement vector \mathbf{r} , just as in conventional MRI k-space is the reciprocal parametrisation of the space of voxel position \mathbf{v} .

One problem in the diffusion imaging literature is that names for techniques often refer both to a particular type of imaging acquisition, and to a particular method to reconstruct the directional organisation of the voxel. All dMRI acquisition methods acquire data in q-space, and the methods can be categorised by their sampling pattern in q-space.

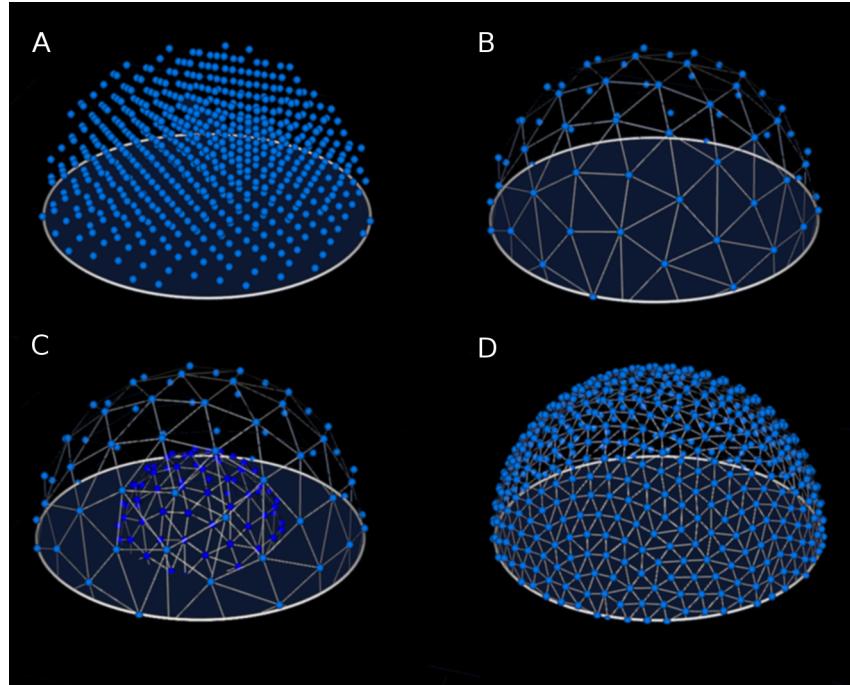


Figure 1.8: A: DSI 604 directions, B: HARDI 65 directions, C: HARDI with 2 shells of 65 directions in each shell, D: QBI 515 directions.

We refer to a method as a *q-space spherical shell* method if it is an acquisition method using a collection of points in q-space that can be thought of as lying on a sphere. Examples include techniques referred to as HARDI (High Angular Diffusion Imaging), Q-ball Imaging and HYDI (Hybrid Diffusion Imaging). A q-space shell method might involve a single shell (see Fig. 1.8B,D) e.g. QBI, or multiple shells (see Fig. 1.8C) e.g. EQBI. In clinical settings, where we can only use a few directions (less than 60 with minimum 6) it is recommended to use the SDT (Single Diffusion Tensor also known as DTI) reconstruction model which is relatively easy to fit as it has only a few parameters.

A *q-space 3D Cartesian grid* method is an acquisition method more easily thought of as a collection of points regularly distributed through a region in q-space. The characteristic example is DSI (diffusion spectrum imaging) (see Fig. 1.8A) and QSI (Q-Space Imaging) [28, 17]. In chapter

2 we will concentrate more on q-space Cartesian grid methods and propose a new reconstruction method which we call DNI (Diffusion Nabla Imaging) which also uses q-space Cartesian grid data sets.

1.6 Diffusion Tensor

Assuming that the diffusion propagator is given by a 3-dimensional Gaussian distribution from Eq. 1.4 we write

$$P(\mathbf{r}, t) = \frac{1}{\sqrt{4\pi t^3 |\mathbf{D}|}} \exp\left(-\frac{\mathbf{r}^T \mathbf{D}^{-1} \mathbf{r}}{4t}\right) \quad (1.5)$$

where \mathbf{D} is known as the diffusion tensor. This Tensor is a 3×3 positive definite symmetric matrix that can be completely described by a centred ellipsoid with 3 principal axes and associated eigenvalues $\lambda_1, \lambda_2, \lambda_3$. The trace of the diffusion Tensor has been found valuable for detecting and evaluating brain ischemia and stroke [29, 30]. Frequently, mean diffusivity (MD) is used instead of the trace defined as

$$MD = \frac{\text{trace}(D)}{3} = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} \quad (1.6)$$

Fractional Anisotropy (FA) is the most common scalar metric used in diffusion imaging which is used to characterise the presence or absence of a preferred direction for diffusion. Like the MD it depends only on the eigenvalues.

$$FA = \frac{1}{\sqrt{2}} \frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \quad (1.7)$$

If FA is equal to 1 that means very anisotropic (infinitely prolonged ellipsoid, a ‘stick’) and if FA is equal to 0 that means completely isotropic (sphere). FA is used in clinical studies to diagnose diseases like stroke and cancer and assess the progress of therapy [31].

Whenever we use FA volumes in our analysis we are implicitly assuming that the propagator of the spin displacements in every voxel has a 3D Gaussian distribution. This assumption is used in most of the diffusion related literature where DTI or Diffusion Tensor Imaging is the prevailing term. Unfortunately, in reality things are much more complicated; inside

our brain the axons are semi-permeable (restriction), the water molecules interact with many different elements in the complex intra fibre fluid, the fibres might cross, kiss, divert or bend inside a voxel or between voxels. Assuming a Gaussian distribution is therefore, a non-trivial approximation. However, FA is still prevalent as it is easy to calculate and it gives similar values across different acquisitions.

1.7 Orientation Distribution

Since one of the primary interests of dMRI is the way that the signal depends on the direction of underlying fibre orientations; it is the orientation information of the diffusion propagator (see Eq. 3.1) that is principally of interest.

One possible approach would be to replace the diffusion probability density function with an isosurface, which is a surface that passes through all points of equal probability density value. For instance, an isosurface of a 3D Gaussian distribution is an ellipsoid. A more commonly used technique that is less sensitive to noise involves the computation of the Orientation Distribution Function (ODF) from the displacement distribution [27, 32, 33]. An ODF may be considered a spherical polar plot whose radius in a given direction is proportional to the integral of the diffusion probability density function in that direction. For ease of visualization, we colour-code the surface according to the diffusion direction $((x, y, z) = (r, g, b)$, where r = red, b = blue, and g = green). An orientation distribution function or isosurface can be plotted for each individual MR imaging voxel (see Fig. 1.9).

The ODF expresses the probability of a spin displacing into a differential solid angle about a possible fibre direction $\hat{\mathbf{u}}$. This is used in order to model and visualise the directional information in diffusion propagator and in simple words it just projects the diffusion function on to the sphere by integrating over the radial coordinate of the diffusion function. The ODF representation symbolised below with ψ sacrifices all the radial information but retains the relevant directional information:

$$\psi(\hat{\mathbf{u}}) = \int_0^\infty P(r\hat{\mathbf{u}})r^2dr \quad (1.8)$$

where $\hat{\mathbf{u}}$ is a unit normal vector, and r is the radial coordinate in the diffusion space. By construction $\psi(\hat{\mathbf{u}})$ is a probability distribution over $\hat{\mathbf{u}}$.

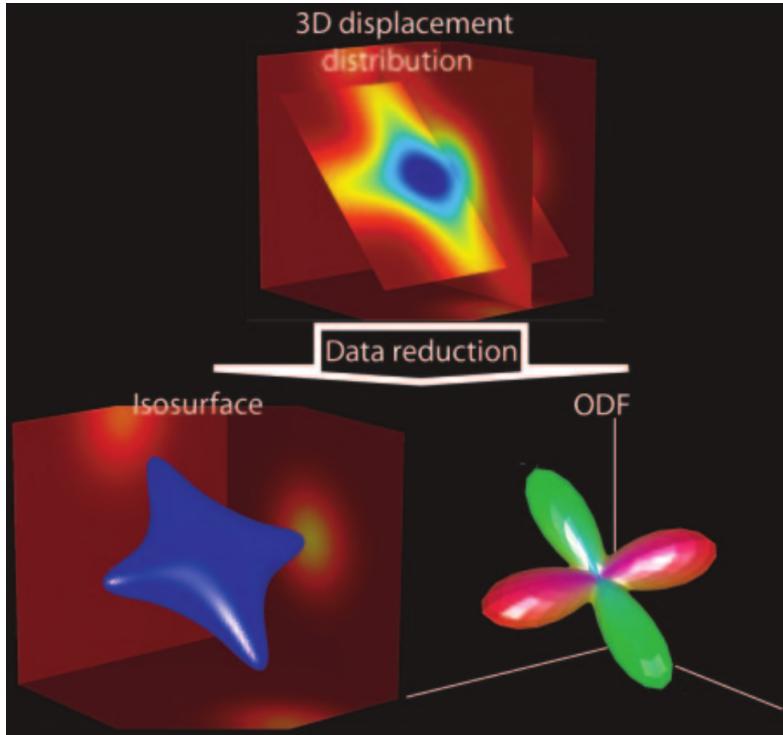


Figure 1.9: Top: The reconstruction of the 3D displacement probability distribution also known as the diffusion propagator or ensemble average propagator (EAP) or diffusion spectrum produced from the inverse Fourier transform of the diffusion signal. Two approaches that may be used to simplify the visual representation of the EAP are shown. Left: the replacement of the displacement distribution with an isosurface. Right: the computation of the commonly used Orientation Distribution Function (ODF). This displacement distribution simulates the crossing of two fibres. In general, the ODF is used essentially to identify the primary directions of the underlying fibres. Picture adapted from Hagmann et al. [27].

The ODF is a function on the sphere. The sphere is usually represented by a discrete spherical grid with evenly distributed points. It is a common procedure to identify the direction of the underlying fibres from the points where the maximum values are found (see Fig. 1.9). This procedure is also known as *peak finding*.

1.8 Tractography

Tractography algorithms integrate local estimates of diffusion direction information. Once we know the orientation of fibres at every voxel, we can join these directions up to reconstruct complete tracks and hence approximate anatomical tracts. In its simplest form, this consists of starting

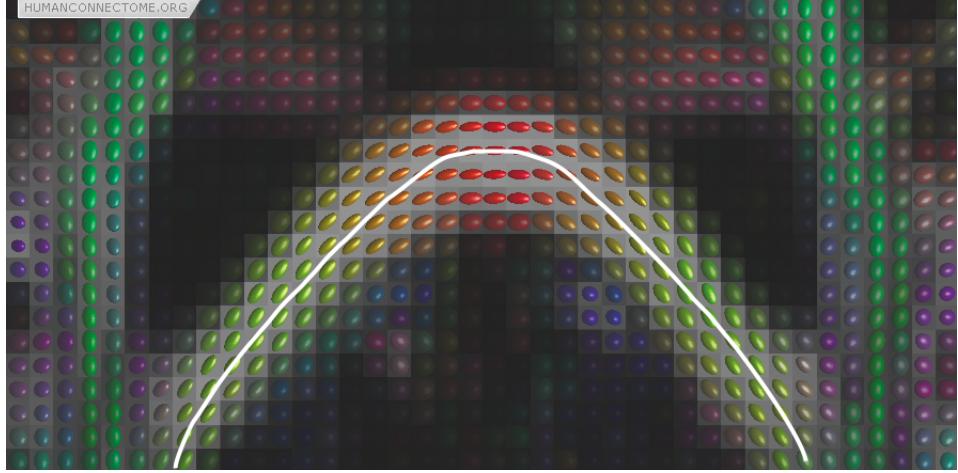


Figure 1.10: The white line shows the track obtained by connecting up a set of voxels based on the direction of the axis of the maximum Tensor eigenvalue and is an example of deterministic tractography. The color is a complementary way of coding the Tensor ellipsoid direction where red denotes left-right, green denotes back-front and blue denotes up-down.

at a seed location and following the preferred direction until we reach a new voxel. We can then change to this voxel’s referred direction and carry on until an entire track is propagated (see Fig. 1.10).

The two best known families of algorithms for track propagation (also known as track integration, tracking or tractography) are *deterministic* [34, 4] and *probabilistic* [35]. A track propagation algorithm belongs to the probabilistic domain if the fibre model that is being used incorporates uncertainty i.e. errors in estimating the orientation of the fibre at every voxel. In the case it does not assume any uncertainty along the path of the track then it belongs to the deterministic domain. In chapter 3 we will give a short overview of many other track propagation methods including global tractography [36].

One of the simplest and earliest deterministic methods is called Fibre Assignment by Continuous Tracking (FACT) [4] (see Fig. 1.11). The FACT algorithm starts through the input of an arbitrary point in the volume and then propagates in both directions i.e. forward and backward. Perhaps the most interesting part with these tracking methods is the way which they decide when to stop tracking. FACT uses a single threshold variable

$$R = \sum_i^s \sum_j^s |\mathbf{e}_i \cdot \mathbf{e}_j| / s(s - 1)$$

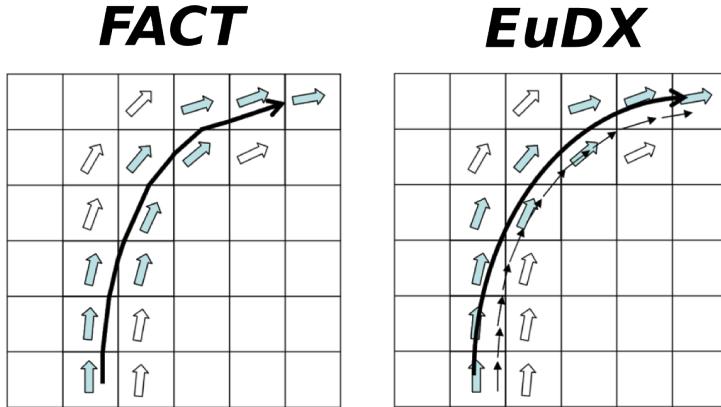


Figure 1.11: Left: FACT propagates with different steps indicated by the entering point on each voxel. Right: Other deterministic methods use constant steps. In this case interpolation of the neighbouring directions is necessary. For example in EuDX trilinear interpolation is used. Picture adapted from Stamatopoulos [37].

where s is the number of neighbouring voxels and \mathbf{e} is the eigenvector corresponding to the highest eigenvalue in each voxel. The simplest case for defining a neighbourhood of a voxel is to use all 26 other adjacent (touching) voxels. Now, let's think of how R will behave in different neighbours. When adjacent fibres are aligned strongly R will be near to 1 as the absolute value of each dot product will come closer to 1 as the normalised vectors become more co-linear. On the other side, R will be smaller in regions without consistency in fibre direction. In voxels with R less than a prespecified threshold e.g. 0.8 the tracks will stop being propagated and FACT will terminate. An important problem with FACT is that it fails to track in areas where there are crossings. In this case, it only tracks one of the major pathways of the crossing area .

Another popular deterministic approach commonly used in the field of fluid dynamics for flow simulation was applied to the field of dMRI by [34] and [38]. The authors made the assumption of a continuous vector field where the track is propagated by the solution of a system of differential equations subject to an initial condition, the position of the seed point. Here the authors propose that a track can be represented by a 3D curve \mathbf{r} parametrised by the arc length s of the track. This is provided by the iterative solution of the differential equation

$$d\mathbf{r}(s)/ds = \mathbf{e}(\mathbf{r}(s)) \quad (1.9)$$

where e is the primary direction of the Tensor or a different model. The solution of this system of Ordinary Differential Equations (ODE) in its simplest case is given by iterative methods like Euler integration bound to an initial condition $\mathbf{r}(0) = \mathbf{r}_0$ where \mathbf{r}_0 is the seed point

$$\mathbf{r}(s_1) \sim \mathbf{r}(s_0) + \alpha \mathbf{e}_1(\mathbf{r}(s_0)) \quad (1.10)$$

where $0 < \alpha \leq 1$ defines the integration step length. In Euler integration we are using the first two terms of the Taylor expansion. Conturo et al. [34] used Euler integration and Basser et al. [38] used a higher order approximation called 4th-order Runge-Kutta scheme.

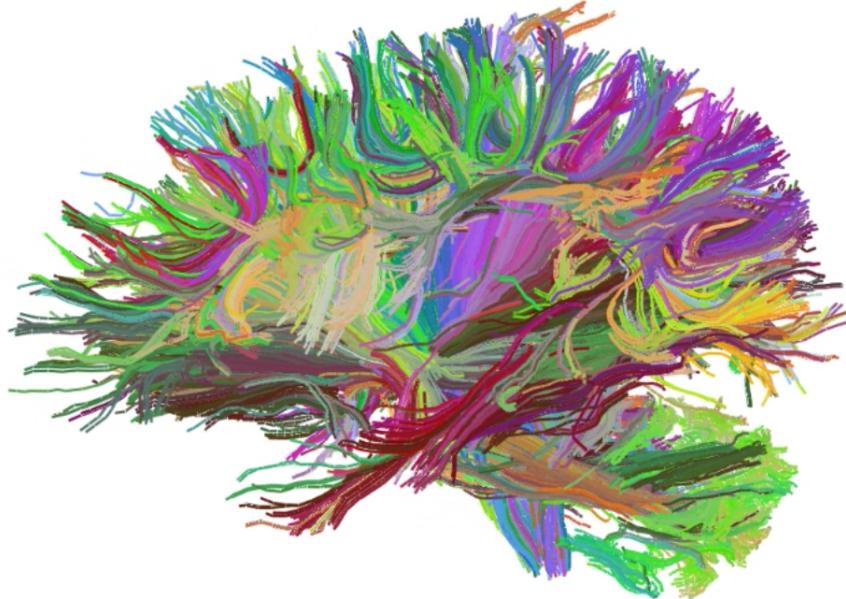


Figure 1.12: Deterministic whole brain tractography based on EuDX, generated using DIPY and visualised using FOS. The colour encodes the orientation of the mid-segment of every track using a colourmap based on Boy's real projective plane immersion [39].

Wedge et al. [3] showed that one could derive the local orientation field of vectors \mathbf{e} from the local maxima of the ODF calculated in each voxel. In that way one could visualise crossing distributions and depict crossing fibres. The authors suggest that diffusion MRI with sufficient signal to noise ratio (SNR) could make tractography a mathematically well-posed problem. However, much longer scanning time is needed in order to reach the necessary resolution and this is often impractical. Our novel tractography algorithm (EuDX) belongs in the deterministic domain and

it is presented in chapter 3 (see Fig. 1.11, 1.12).

In summary, the deterministic algorithms propagate tracks by making a series of discrete locally optimum decisions. These are fast, simple and easy to interpret. Usually, we depict them using tracks (also known as streamlines or polylines). The main disadvantage of deterministic algorithms is that they are vulnerable to local noise.

Probabilistic tractography is meant to deal with this problem of noise and propagate tracks even in regions where the tracking is unclear. This is made possible by assuming that uncertainty exists concerning the orientation of the fibre at each point of the track.

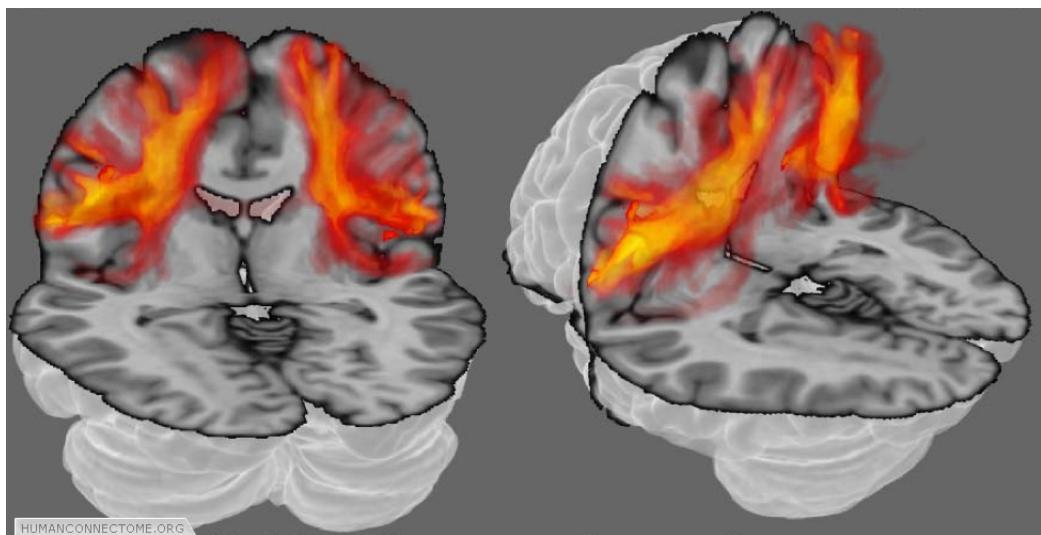


Figure 1.13: The 3D distribution of voxels connected to the seed voxel is called a tractogram.

Imagine a particle in a seed voxel moving in a random fashion with a constant speed within the brain white matter. The transition probability to a neighbouring point depends on the local orientation distribution or underlying model. This yields high transitional probabilities along the main fibre directions. Hence, the particle will move in parallel to the fibre direction with a higher probability than in a perpendicular direction. In this probabilistic method, we start a large number of particles from the same seed point, let the particles move randomly according to the local ODF and count the number of times a voxel is reached by the path of a particle (connectivity values). The random walk is stopped when the particle leaves the white matter volume. Probabilistic tractography is commonly depicted as a tractogram (see Fig. 1.13). In a tractogram, we count and

store the number of tracks that go through each voxel and visualise the entire volume of the stored values.

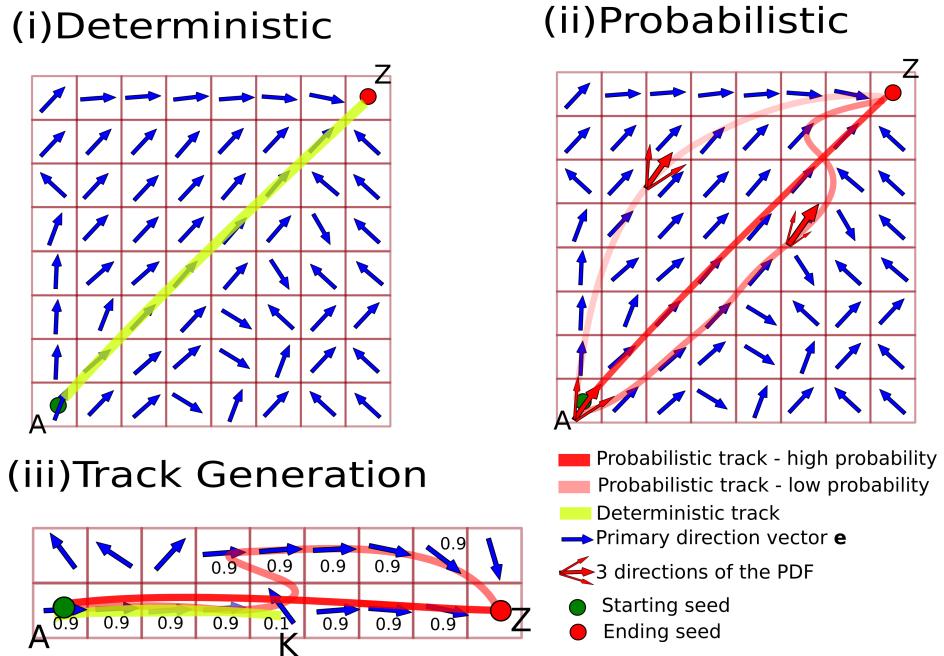


Figure 1.14: A simplified example showing in (i) and (ii) the same data set. (i) The yellow line shows the result of deterministic tractography which is given by a single trajectory and in (ii) is given by connectivity matrix depicting in red the probability of different pathways throughout the hole slice. For the ease of understanding, only 3 possible pathways are depicted. Finally, in (iii) an example is given where it is shown that probabilistic tractography weights more closer connections. However, it can track further deep than deterministic tractography.

We will try to illustrate the difference between the two approaches (deterministic and probabilistic) using a simplified 2D example shown in Fig. 1.14. Generalising afterwards in 3D is straightforward.

Let's imagine that we have a 2D slice where in each pixel we have calculated a vector showing the primary direction for that specific pixel. This vector \mathbf{e} could have been calculated from the Tensor as the principal eigenvector or as a principal direction of a different model e.g. the maximum point of the ODF. Let's now think that we want to find the best track from seed A to seed Z. When using deterministic tractography we are using only the local direction information in every pixel therefore in a direction field as this of Fig. 1.14(i) we will have to use only one track and this is depicted in the diagonal pathway with yellow colour. How-

ever, there are other possible tracks as well in this diagram e.g rather than taking the diagonal we could go first up from A and then right.

Probabilistic tractography aims to identify all the possible tracks by assigning to each one of them a weight. All the weights of all the tracks together sum to 1. This is possible by generating samples from a probability distribution for every pixel. In this toy example shown in Fig. 1.14(ii) the orientation of the blue vectors is represented by a single parameter, angle ω . ω here is a random variable that takes values from a Probability Density Function (PDF). We have many possible directions to move next but with different probabilities. The weights of all directions again sum to 1. After this explanation we can identify in Fig. 1.14(ii) that the most likely track is again the diagonal (with deep red) but there are other possible tracks (with lighter red) that are less likely. In the same diagram we show with 3 combined red arrows some of the many directions that are possible in each point. However, some are more certain than others.

Let's try now to understand how a track is valued as more probable than others. In Fig. 1.14(iii) we have drawn a very simple image with only two pixel rows and we are assuming that the probability of moving along the primary direction (shown with blue arrow) is 0.9 and there is only a secondary direction given by 0.1 (1-0.9) i.e. for ease of understanding, we assume only 2 possible directions. We can see that there is a discontinuity in position K. In that point, an Euler based deterministic approach (without interpolation) has to stop at K (yellow line). The probabilistic method will continue tracking and it will generate two tracks that both reach the target. The probability of each track is calculated by multiplying the probability of a specific direction of each point. Therefore, the shorter and dark red track will have $p_s = 0.9^7 \times 0.1$ and the longer and lighter red track will have $p_l = 0.9^{10}$. It is obvious that $p_s > p_l$ and that the darker red track is more likely to exist according to this method.

This method of multiplying the probabilities at each voxel along the paths has been proposed by many [40],[41], [42], [43],[44],[45] and used in many software packages as well FSL FDT [46, 47, 5] and others. However, there are some problems which are discussed next.

1.9 Known problems

Although the probabilistic methods are able to identify many known tracts, they miss several large tracts such as the visual pathways LGN-MT and callosal MT [48]. The visual pathways are useful test cases for algorithmic development and testing because they diverge and bend significantly. Probabilistic methods have the advantages that they expand the track search space beyond deterministic algorithms and that they can easily expand with complex models supporting crossings (usually not more than 2 crossings). However, they do not compute an accurate probability of brain connections. The phraseology “connection probabilities” or “estimation of global connectivity” or “the probability of the existence of a connection through the data field, between any two distant points” found in [46] can be very misleading because someone might believe that they represent the actual connectivity profile of the subject. For example, we are certain (with probability 1) that LGN (lateral geniculate nucleus) is connected to V1 and V2 (primary visual cortex) in any healthy brain however the estimated connection probability in FDT is much less than 1. In addition, current probabilistic algorithms fail to identify pathways even when they are known to exist or in a few cases they generate pathways even when they do not exist [35, 49, 50]. For example, no connections between left MT+ and the posterior portion of Corpus Callosum were found in PiCo [51] or FDT although it is well established that they do exist.

In 2008, Sherbondy et al. [48] in order to try to deal with the problems explained in the previous paragraph introduced an algorithm that separates the pathway sampling and scoring steps. In that way the scoring does not depend any more on whether we are tracking from seed A to seed B or from B to A; therefore it assumes symmetry when the other probabilistic methods do not. At the same time Sherbondy’s method assumes independence between different tracts i.e. pathway $A \rightarrow B$ is independent from $A \rightarrow C$ or $K \rightarrow L$. This does not happen in most other methods where a pathway depend on other pathways starting from the same seed. Sherbondy’s method was designed to estimate connections that are known to exist. The disadvantage of this method is that it needs a lot of user interaction to add the known tracks and the user needs to be a specialist in white matter anatomy otherwise the results might be biased. Other tools like FSL’s FDT uses waypoint masks to reduce the effect of this

	Deterministic	Probabilistic
Voxel Noise Resistance	Less	More
Non-existing Tracts	Yes	Yes
Execution Time	Fast	Slow
Memory Size	Less	More
Biased on Tract Length	Yes	Yes

Table 1.1: Known problems with deterministic and probabilistic tractography

problem. But these too need to be defined by the user.

A small summary of the comparative strengths of deterministic and probabilistic tractography is given in the qualitative Tab. 1.1 where we can see that although deterministic tractography will most likely stop more frequently at noisy voxels it is much faster to calculate than probabilistic tractography. They both can generate non-existing tracts because of propagation errors or errors in the reconstruction step. For noise related problems e.g. motion and eddy correction and possible solutions see [44, 38, 52, 53, 54, 55, 56], and for methodology and ideas comparing across subjects see [57, 11, 31, 58, 59, 60, 61, 62]. We showed in the previous section that probabilistic tractography usually gives higher weight to shorter pathways. In the deterministic tractography an opposite weighting is required as longer pathways will have higher representation in the datasets and so it is more likely to have more seeds along a long track rather than along a short track. Therefore, a normalization by length both for probabilistic and deterministic tractography is highly recommended.

1.10 Segmentation

The white matter contains pathways known as fibre tracts that connect functional areas of the brain. A diagram of commonly found fibre tracts is sketched in Fig. 1.15 and real fibre tracts are shown in Fig. 1.16. The white matter contains three types of fibre tracts: commissural, association, and projection [63].

Commissural tracts connect related regions of the two cerebral hemispheres. Association fibres connect regions in the same hemisphere. Association fibres come in various sizes: the smallest fibres are completely within the cortex, the medium ones are called u-fibres and connect one gyrus to the next, and the longest association bundles connect different

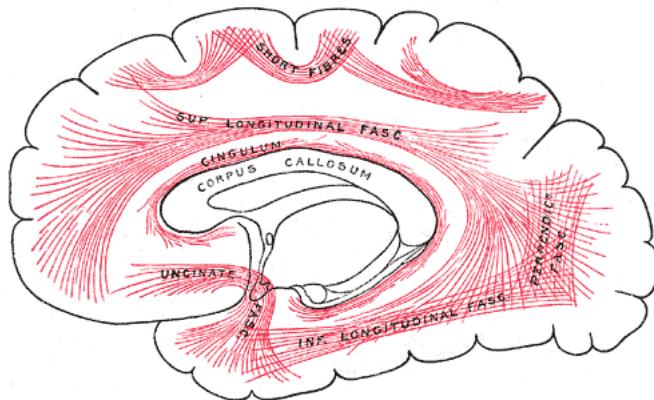


Figure 1.15: Diagram showing principal systems of association fibres in the cerebrum. The white matter fibre tracts are large bundles of axons that interconnect the gray matter processing areas both within and across hemispheres. The association fibres connect fibres from the same hemisphere. Picture from Gray's Anatomy #751 [64].

lobes. Finally, projection fibres connect the cortex and subcortical structures such as the thalamus, basal ganglia, and spinal cord.



Figure 1.16: Fibre pathways are so densely packed in the real brain that a segmentation algorithm of some kind looks like a possible solution for the neurosurgery planning of the future and further understanding of the brain connectivity. Picture from virtual hospital [65].

A major issue with white matter is that it is particularly dense (see Fig. 1.16) i.e. the boundaries between different tracts are very difficult to distinguish, many bundles cross or touch other bundles (see Fig. 1.15) and most bundles diverge into smaller and smaller tracts as they reach

and enter gray matter areas.

Previously it was not possible to automatically create white matter models similar to these anatomical atlas diagrams of fibre tracts in vivo. However, methods are now available to estimate white matter fibre tracts using diffusion MRI. In this thesis we present a method for segmentation of the trajectories estimated from diffusion MRI by automatically grouping them into anatomical regions or to be more accurate regions of similar proximity and shape characteristics. This is the main topic of chapter 4 where we propose QuickBundles a highly efficient algorithm for tractography segmentation.

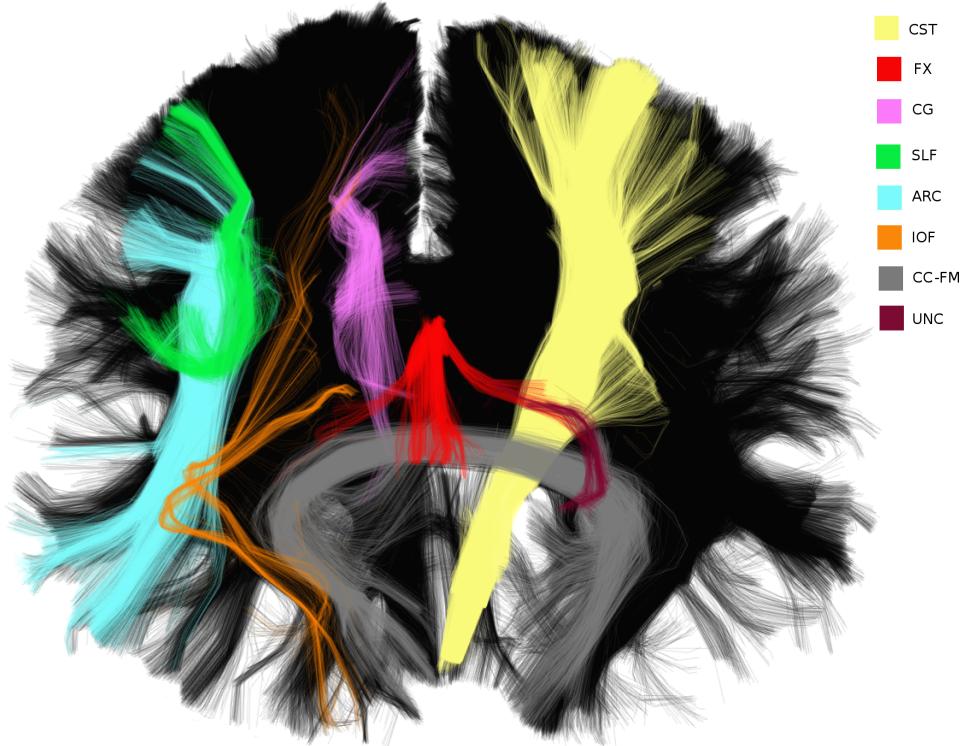


Figure 1.17: An example of a tractography segmentation based on labeling provided by a neuroanatomist. An important aim of this thesis is to automatically find the labels or simplify the work of an expert by clustering tractographies. CST: Corticospinal Tract, FX: Fornix, CG: Cingulum Bundle, SLF: Superior Longitudinal Fasciculus, ARC: Arcuate Fasciculus, IOF: Inferior Occipitofrontal Fasciculus, CC-FM: Corpus Callosum Fornaces Major, UNC: Uncinate Fasciculus.

There are three main goals which should be satisfied by an automatic tractography segmentation algorithm [63]: automatic grouping of like trajectories into regions, region correspondence across subjects, and anatom-

ical labeling of regions. Our ability to perform automatic, subject-specific definition of the white matter fibre tracts has applications in neuroanatomical visualization, neurosurgical planning, and neuroscientific studies of white matter integrity, structure, and variability. An example of a segmentation is shown in Fig. 1.17. The data set and labels used in this figure are from Pittsburgh Brain Competition¹.

1.11 Foreword

In Chapter 2 we concentrate on the problem of establishing multiple fibre directions in single voxels. This is a very important problem as it is considered that at least 30-90% of the brain contains crossings. We focus on comparing methods which acquire data in a Cartesian grid in q-space (see section 1.5) and we contribute with a new method which we call the Equatorial Inversion Transform (EIT). In addition, we show results with a new version of Generalized Q-sampling Imaging (GQI2) which was not yet used with simulated or real data sets. Both methods have impressive accuracy on low angle crossings.

In Chapter 3 we concentrate on the problem of integrating the directional information from voxel to voxel in order to create streamlines (tracks). The streamlines approximate anatomical tracts as we discussed in section 1.8. We contribute with a new deterministic algorithm and show how it can produce different results depending on the underlying reconstruction model and compare it against a probabilistic method (PICo). We show that our algorithm named EuDX overperforms in simulation tests in approximating tract integrity along long distances and gives more uniform results with real data sets against PICo.

In Chapter 4 we concentrate on the problem of simplifying large numbers of streamlines which are common in current tractography analysis procedures. We contribute with an average linear time clustering algorithm (QuickBundles) which groups the streamlines by accounting a distance metric. This metric combines spatial and shape characteristics of the streamlines. QuickBundles provides representative streamlines which can be used for interaction, registration, queries and many other applications. This algorithm opens the way towards accurate segmentation of tractographies. Such an advancement is of great importance as described

¹braincompetition.org

in section 1.10.

In Chapter 5 we conclude the thesis and give some ideas of the work that we plan to do in the near future.

In the next chapters we will use the following definitions. Define a track (or streamline) as a polyline $s = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, where $\mathbf{x} \in \mathbb{R}^3$. Then the entire tractography is defined as $T = \{s_1, \dots, s_m\} \sim \mathfrak{T}$, where usually the number of tracks is $|T| \simeq 2 \times 10^5 - 2 \times 10^6$. For an anatomical (physical) bundle (tract) e.g. Arcuate Fasciculus we use v and for a fibre bundle we use u where $u \subset T$, which approximates v . We think it is very important to always have in mind that we can only approximate a real tract.

2 Cartesian Lattice Q-space Reconstructions

2.1 Overview

Between one to two thirds of imaging voxels in the human brain's white matter are thought to contain multiple fibre bundle crossings [35],[66] in which case the Diffusion Tensor model proposed by Basser et al. [67] breaks down. High Angular Resolution Diffusion Imaging (HARDI) [68], Diffusion Spectrum Imaging (DSI) [69], [33] or Higher Order Tensors [70], [71] and many more reconstruction methods have been proposed to overcome the limitations of the Diffusion Tensor. These methods can be divided into those which need specific acquisition parametrizations, and those which can be used independently of q-space structure. For instance, for Q-ball Imaging [32] sampling needs to be on one or more spherical grids, and in Generalized Q-sampling Imaging (GQI) [72], requires sampling on a Cartesian grid; by contrast DTI can be used independently of q-space structure. A further division considers the level of model assumptions for the diffusion process. Although all methods have some underlying assumptions we generally separate them in model-based and model-free. Model-based methods like the Single Tensor or Multi Tensor require a number of parameters to be fitted. By contrast, in model-free methods fitting is not necessary and the directionality of the underlying tissue can be approximated by some re-parametrization or re-transformation of the signal. The latter is usually more efficient than fitting models with many parameters which typically call for iterative methods.

This chapter presents, evaluates and compares different model-free methods for the reconstruction of orientation distribution functions using diffusion MRI data sampled on a Cartesian lattice in q-space. This non-parametric nature of the algorithms described here allows for the identification of multiple fibre crossings. In addition, a new method is presented named Diffusion Nabla Imaging (DNI) and a family of methods is defined called the Equatorial Inversion Transform (EIT). The EIT is a new way to represent and reconstruct the diffusion signal. Our results show that EIT can perform better or as well as the current state-of-the art methods i.e. DSI and GQI.

2.2 Theory

We start from the classical formulation shown in Eq. 2.1 of joint k-space and q-space imaging described in Callaghan [17], [69] using the narrow pulse gradient spin echo (PGSE) sequence of Tanner and Stejskal

$$RF(\mathbf{k}, \mathbf{q}) = \int \rho(\mathbf{v}) \exp(i2\pi\mathbf{k} \cdot \mathbf{v}) \int P_\Delta(\mathbf{v}, \mathbf{r}) \exp(i2\pi\mathbf{q} \cdot \mathbf{r}) d\mathbf{r} d\mathbf{v} \quad (2.1)$$

Here RF represents the complex RF signal measured at spatial wave vector \mathbf{k} and magnetic gradient wave vector \mathbf{q} , ρ is the local spin density (number of protons per unit volume contributing to the RF signal), Δ is the time between diffusion gradients, P_Δ is the average diffusion propagator (transition probability distribution), \mathbf{v} is the voxel coordinate, and \mathbf{r} is the diffusion displacement.

The k-space reconstruction with the narrow pulse approximation [73] gives us diffusion weighted image data S which reveal the average propagator P_Δ of each voxel

$$S(\mathbf{v}, \mathbf{q}) = \int \rho(\mathbf{v}) P_\Delta(\mathbf{v}, \mathbf{r}) \exp(i2\pi\mathbf{q} \cdot \mathbf{r}) d\mathbf{r} \quad (2.2)$$

For the rest of the chapter we consider each voxel independently and assume intra-voxel spatial homogeneity so we can drop explicit reference to \mathbf{v} and Δ . We note in passing that the shape of P_Δ and hence of the ODF may change with different values of Δ . We will not pursue this matter further here. We can also replace the spin density $\rho(\mathbf{v})$ with S_0 i.e. the measured signal without diffusion weighting $\mathbf{q} = \mathbf{0}$. Therefore we can write

$$S(\mathbf{q}) = S_0 \int P(\mathbf{r}) \exp(i2\pi\mathbf{q} \cdot \mathbf{r}) d\mathbf{r} \quad (2.3)$$

By applying the 3D Fourier transform in Eq. 2.3 we can reconstruct the average propagator also known as the diffusion spectrum [73] or diffusion propagator

$$P(\mathbf{r}) = S_0^{-1} \int S(\mathbf{q}) \exp(-i2\pi\mathbf{q} \cdot \mathbf{r}) d\mathbf{q} \quad (2.4)$$

It was shown by Wedeen et al. [73] that the dMRI signal is positive for any type of spin motion without net flux (i.e. spin displacements due to

thermal molecular agitation) or other random fluxes such as intravoxel incoherent motion. Under this assumption we can replace the complex signal S with its modulus $|S|$ in Eq. 2.4

$$P(\mathbf{r}) = S_0^{-1} \int |S(\mathbf{q})| \exp(-i2\pi\mathbf{q} \cdot \mathbf{r}) d\mathbf{q} \quad (2.5)$$

The modulus of the signal coincides with the output of the standard MRI scanners as dMRI and that simplifies the acquisition procedure. It represents the density of the average relative spin displacement in a voxel. In other words, $P(\mathbf{r})$ is a measure of the probability that a spin in a chosen voxel, during the experimental mixing time Δ , would make a vector displacement \mathbf{r} . We can visualize the propagator for every voxel as a 3D density volume (see Fig. 1.9).

In the classical DSI acquisition, at each location, diffusion-weighted images are acquired for $N = 515$ or fewer values of q-encoding, comprising in q-space the points of a cubic lattice within the sphere of five lattice units in radius. Therefore,

$$\mathbf{q} = \alpha \mathbf{q}_x + \beta \mathbf{q}_y + \gamma \mathbf{q}_z \quad (2.6)$$

with $\alpha, \beta, \gamma \in \mathbb{Z}^+$ and $(\alpha^2 + \beta^2 + \gamma^2)^{1/2} \leq 5$. The signal is premultiplied by a Hanning window before Fourier transform in order to ensure a smooth attenuation of the signal at high q values. Often, to obtain data for the complete grid of 515 q-vectors (which also means that we need to collect 515 diffusion weighted volumes), the overall acquisition time would be too long and a smaller number of unique q-vectors are employed for just a single hemisphere usually between 101 to 257 points [74]. This is valid because the underlying self-diffusion process is symmetric and so the signal is symmetric, therefore the vectors can be mapped on the other hemisphere to create the full q-space.

Since we are mainly interested in the angular structure of the underlying tissue, we further simplify the data by taking the weighted radial summation of $P(\mathbf{r})$

$$\psi_{DSI}(\hat{\mathbf{u}}) = \int_0^\infty P(r\hat{\mathbf{u}}) r^2 dr \quad (2.7)$$

This defines the orientation density function (ODF) for DSI which measures the quantity of diffusion in the direction of the unit vector $\hat{\mathbf{u}}$ where

$$\mathbf{r} = r\hat{\mathbf{u}}.$$

Note at this point that in order to find the ODF we have to first create the diffusion propagator by applying the Fourier transform on the lattice. Yeh et al. [72] proposed a direct way to calculate a slightly different ODF using the Cosine transform.

In order to represent the average propagator in the scale of spin quantity Yeh et al. [72] introduced the *spin density function* Q which is estimated by scaling the average propagator P_Δ with the spin density ρ , i.e. $Q(\mathbf{r}) = \rho P(\mathbf{r}) = S_0 P(\mathbf{r})$. From Eq. 2.3 we obtain

$$S(\mathbf{q}) = \int Q(\mathbf{r}) \exp(i2\pi\mathbf{q} \cdot \mathbf{r}) d\mathbf{r} \quad (2.8)$$

We can apply the Fourier transform again to Eq. 2.8 and obtain

$$Q(\mathbf{r}) = \int S(\mathbf{q}) \exp(-i2\pi\mathbf{q} \cdot \mathbf{r}) d\mathbf{q} \quad (2.9)$$

Because $Q(\mathbf{r})$ is real and $S(\mathbf{q})$ is symmetric (even), i.e. $S(\mathbf{q}) = S(-\mathbf{q})$, we can use directly the Fourier Cosine transform (see section A.2) to calculate

$$Q(\mathbf{r}) = \int S(\mathbf{q}) \cos(2\pi\mathbf{q} \cdot \mathbf{r}) d\mathbf{q} \quad (2.10)$$

and obtain the “spin” orientation distribution function (SDF) ψ_{GQI} from an unweighted truncated radial projection

$$\psi_{GQI}(\hat{\mathbf{u}}) = \int_0^\lambda Q(r\hat{\mathbf{u}}) dr \quad (2.11)$$

$$= \int_0^\lambda \int S(\mathbf{q}) \cos(2\pi r\mathbf{q} \cdot \hat{\mathbf{u}}) d\mathbf{q} dr \quad (2.12)$$

$$= \lambda \int S(\mathbf{q}) \text{sinc}(2\pi\lambda\mathbf{q} \cdot \hat{\mathbf{u}}) d\mathbf{q} \quad (2.13)$$

where λ is a constant called the diffusion sampling length. This parameter acts as a smoothing factor. The higher λ the more detailed the SDF will be but also more noisy. This ODF was used as the basis of the analysis of the GQI method. It provides a simple direct analytical solution of the ODF

which can be written in a simple matrix form

$$\psi_{GQI} = \mathbf{s} \cdot \text{sinc}((6D \cdot G \circ \mathbf{b} \circ \mathbf{1}) \cdot U^T) \lambda / \pi \quad (2.14)$$

where \cdot denotes standard matrix or vector dot product, \circ denotes the Hadamard product, ψ_{GQI} as a M -dimensional vector with components corresponding to the selected directions $\hat{\mathbf{u}}$ on the ODF sphere, \mathbf{s} is a vector with all the signal values, $D=0.00251$ [75] where D is a constant known as the free water diffusion coefficient, G is the $N \times 3$ matrix with the gradient vectors, \mathbf{b} is the $N \times 1$ matrix with the b-values and $\mathbf{1}$ is the $N \times 3$ incidence matrix where all values are equal to 1.

For a similar ODF like the one produced using DSI we need to take the weighted truncated radial projection. This will give us a different “spin” ODF which we symbolize with ψ_{GQI_2}

$$\psi_{GQI_2}(\hat{\mathbf{u}}) = \int_0^\lambda Q(r\hat{\mathbf{u}})r^2 dr \quad (2.15)$$

$$= \lambda^3 \int S(\mathbf{q}) H(2\pi\lambda\mathbf{q} \cdot \hat{\mathbf{u}}) d\mathbf{q} \quad (2.16)$$

$$\text{where } H(x) = \begin{cases} \frac{2\cos(x)}{x^2} + \frac{(x^2-2)\sin(x)}{x^3}, & x \neq 0 \\ 1/3, & x = 0 \end{cases}.$$

This equation can be similarly implemented with a simple matrix transform

$$\psi_{GQI_2} = \mathbf{s} \cdot \text{H}((6D \cdot G \circ \mathbf{b} \circ \mathbf{1}) \cdot U^T) \lambda^3 / \pi$$

and has not to date been published with real or simulated data sets.

The addition of the spin density plays a very important role on normalizing the ODF and providing scalar or vector metrics for the analysis of dMRI data sets. GQI, similarly to DSI, expects the q-vectors to sit on a cubic lattice within a sphere. However, because of the direct analytical formulation of the GQI ODFs; the creation of the volumetric grid with the signal values is not necessary. This makes GQI advantageous on memory and CPU efficiency. Furthermore, no Hanning filter is necessary.

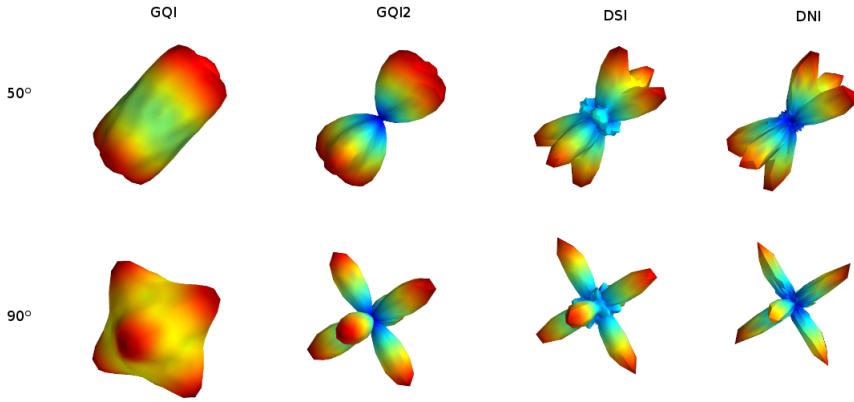


Figure 2.1: Showing the ODFs from two randomly oriented simulated 3-fibre crossings at 50° (top) and 90° angles between each pair of fibres using different Cartesian lattice q-space reconstruction methods.

2.3 Other methods

Pickalov et al. [76] proposed a new method for reconstructing the diffusion propagator by applying an iterative inverse Radon transform on measurements along many radial lines; computing 1D tomographic projections to reconstruct the 3D EAP. This technique measures DW images along a few radial lines of q-space but still requires hundreds of samples to reliably recover the EAP. Currently, to reconstruct the EAP, the state-of-the-art model-free techniques apart from diffusion spectrum imaging are hybrid diffusion imaging (HYDI) [77] and multiple q-shell diffusion propagator imaging (mq-DPI) [78]. HYDI acquires the signal values on five concentric spherical q-space shells, then interpolates onto a cubic grid and applies the standard Fourier transform in the same way as DSI. In mq-DPI the EAP is calculated by solving Laplace's equation for the diffusion signal using a real and symmetric modified spherical harmonic basis. The EAP can be found analytically by the inversion of a linear system using Laplace-Beltrami regularization. In addition, Exact Q-ball imaging (EQBI) [75] provides a different method to calculate the ODF analytically using multiple spherical q-space shells. Similarly, Aganj et al. [79], proposed an analytical solution for the multi-shell case by incorporating a mono-exponential or bi-exponential model (CSA-ODF). Another method for finding a directional distribution on the sphere was proposed by Özarslan et al. [22] called the diffusion orientation transform (DOT). This method calculates a different statistic $P(r_0\hat{u})$, the probability of finding a particle initially at the origin at the point $r_0\hat{u}$, using spherical

harmonics. Not surprisingly there is a relationship connecting CSA with DOT which is

$$\psi_{CSA}(\hat{\mathbf{u}}) = \int_0^\infty DOT(r\hat{\mathbf{u}})r^2 dr \quad (2.17)$$

Jansons et al. [80] proposed a different function on the sphere than the ODFs described above, to be used on data sets acquired on a single spherical q-space shell. They called this spherical function persistent angular structure (PAS). This method has very good angular resolution. It uses the principle of maximum entropy however, it is rather slow as it nonlinear fitting is used in order to identify many parameters. PAS is a statistic on the sphere defined as $PAS(\hat{\mathbf{u}}) = \exp(\lambda_0 + \sum_{j=1}^N \lambda_j \cos(\mathbf{q}_j \cdot k\hat{\mathbf{u}}))$ where λ are the unknown parameters, k is constant and N is the number of DWIs. The relationship $\int PAS(\hat{\mathbf{u}}) \exp(i\mathbf{q}_j \cdot k\hat{\mathbf{u}}) d\hat{\mathbf{u}} = E(\mathbf{q}_j)$ provides the bridge between PAS and the diffusion signal ($E(\mathbf{q})$).

The first publication of using spherical harmonic expansions with diffusivity profiles, which are now quite common in the literature, was by Alexander et al. [81]. Q-ball imaging was introduced by Tuch [32] and a new ODF defined as $\psi(\hat{\mathbf{u}}) = \frac{1}{Z} \int_0^\infty P(r\hat{\mathbf{u}}) dr$ where Z is a normalization constant. It was later provided for Q-Ball imaging a fast and analytical solution using spherical harmonics (SH) and Laplace-Beltrami regularization [82]. Tournier et al. [83], [84] introduced a spherical deconvolution method where first the SH coefficients were estimated, then single fiber ODFs were used as a deconvolution kernel estimated from the real data. Then, the sharper fODF (fiber orientation distribution function) was obtained by a simple linear transformation [85]. Other deconvolution approaches were proposed in [86] and [87].

On Tensor related methods we have the classical Single Tensor [67], Sticks and Ball[35], Multi-Tensor [88], [89] and Higher Rank Tensors [70], [71]. In addition there are also model based methods which try to calculate non-Gaussian properties, for example the Kurtosis Tensor [23], [90] which is used in Diffusion Kurtosis Imaging (DKI).

Finally, new model-based methods are emerging which are trying to calculate statistics like the axonal thickness distribution from dMRI data sets. These are usually based on model free and restricted components; CHARMED [91], [92], AxCaliber [93] and the orientation invariant ActiveAx [94] are some well known methods of this type. Q-space Imaging

(QSI) can be used to identify distributions of axon-diameter too [95].

2.4 Diffusion Nabla Imaging

A new method for the calculation of the real ODF is proposed here. This is based on the theoretical work done by Aganj et al. [96] and Canales-Rodriguez et al. [16] using two important theorems from Fourier Analysis

1. The Fourier transform of $P(\mathbf{r})r^2 = -\nabla^2 E(\mathbf{q})$ where ∇^2 is the Laplacian operator (for proof see section A.3).
2. For a symmetric function $E : \mathbb{R}^3 \rightarrow \mathbb{R}$ and for the arbitrary unit vector $\hat{\mathbf{u}}$ we have $\int_0^\infty E(r\hat{\mathbf{u}})dr = \frac{1}{8\pi^2} \int \int_{\hat{\mathbf{u}}^\perp} E(q) q dq d\phi$ where $\hat{\mathbf{u}}^\perp$ is the plane perpendicular to $\hat{\mathbf{u}}$ (for proof see section A.4).

From Eq. 2.7 we see that the integration is over $P(r\hat{\mathbf{u}})r^2$, therefore we can write

$$\psi_{DNI}(\hat{\mathbf{u}}) = -\frac{1}{8\pi^2} \int_{\hat{\mathbf{u}}^\perp} \int_0^\infty \nabla^2 E(q) q dq d\phi \quad (2.18)$$

where ϕ is the angular rotation component operating on the plane perpendicular to $\hat{\mathbf{u}}$, ∇^2 is the Laplacian operator and $E(q) = S(q)/S_0$ is the normalized diffusion signal. Eq. 2.18 has the advantage that no Fourier transform is necessary. We need however, a way to calculate the Laplacian of the signal. This can be analytically derived for a spherical grid [96] and we propose here that it can be directly calculated in a cubic grid using the standard 3D discrete Laplacian filter which is given by the 3D kernel defined by the following $3 \times 3 \times 3$ array:

$$\left[\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 1 & -6 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right].$$

This is a filter commonly used for image processing. From now on when we use the Laplacian operator in order to measure the directionality of the diffusion signal we will call this reconstruction method Diffusion Nabla Imaging as nabla-squared (∇^2) is the symbol for the Laplacian operator. In Fig. 2.1 we present the ODFs from two randomly oriented simulated 3-fibre crossings at 50° and 90° angles between each other using different grid based reconstruction methods. The parameters used here are for DSI:

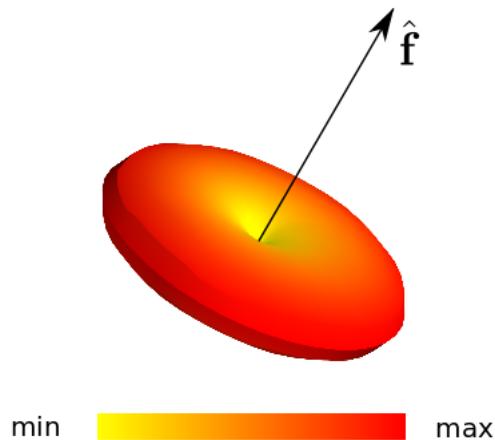


Figure 2.2: The diffusion signal has the beautiful property that it is minimum along the direction of a fibre with unit direction \hat{f} and maximum along the equator defined by the plane perpendicular to that fibre direction. This property is the basic inspiration behind the EIT. In this picture the 3D surface plot of a simulated signal for a spherical grid acquisition with b-value 2,000 is shown using a yellow-red colourmap.

radial sampling range 2.1 – 6 with 0.2 and Hanning filter width 36, GQI: $\lambda=1.2$, GQI2: $\lambda = 3$ and DNI: radial sampling 0 – 5 with 0.2 steps. All methods used the same reconstruction sphere with 642 vertices and 1,280 faces.

2.5 Equatorial Inversion Transform

We propose an important theoretical construction called the Equatorial Inversion Transform (EIT) which creates a general formulation for the interpretation of the directionality of the diffusion signal. This idea is founded on two general properties of the diffusion signal: (a) If we visualize the diffusion signal for a single fibre for all gradient directions we see the generated shape to be lowest towards the direction of the fibre and highest on the plane perpendicular to that direction (see Fig. 2.2). (b) The diffusion signal is additive i.e. $S(\hat{f}_1) + S(\hat{f}_2) = S(\hat{f}_1 + \hat{f}_2)$, where \hat{f}_1, \hat{f}_2 are the unit directions of the fibres. In simple terms the signal of 2-fibre crossing can be decomposed linearly to the signals of the two fibres that create the crossing. The same holds for any number of fibres in a crossing.

These are two very important geometric properties of the signal that we can try to exploit to its limit by calculating equatorial integrals in order to identify the directionality of the signal.

Apart from the visual confirmation, further supporting evidence that equatorial integration is crucial for derivation of directionality can be seen in Eq. 2.18 where an equatorial integral creates a connection between the real ODF and the signal. The Funk-Radon Transform (FRT) used by [32] is another example where equatorial integration is employed using the reconstruction sphere. We will see next that DNI and FRT are just a subset of the EIT.

With EIT the most important goal is to try to identify the orientational variation in the signal in the most accurate way by generating a spherical density. However, it is possible to calculate additionally the classical ODF as defined by Wedeen et al. [73].

The EIT shown in Eq. 2.19 consists of an integration along the equator and along radial lines. A function F of the signal is multiplied by a radial weighting function O . This construction is a generalization of the previous ODFs and it can support successfully many different function families for F and O which can all more or less accurately identify the directional distribution of the signal. More precisely the EIT is defined as

$$\psi_{EIT}(\hat{\mathbf{u}}) = \int_{\hat{\mathbf{u}}^\perp} \int_0^\infty F(E(q))O(q)dq d\phi \quad (2.19)$$

where F could be for example any of the following functions

$$F(E(q)) = \begin{cases} E(q) & (I) \\ -\nabla^2(E(q)) & (II) \\ \nabla^4(E(q)) & (III) \\ \dots \end{cases} \quad (2.20)$$

and O could be for example any of the following functions

$$O(q) = \begin{cases} 1 & (0) \\ q & (1) \\ q^2 & (2) \\ \dots \end{cases} \quad (2.21)$$

In Tab. 2.1 we see that by choosing different functions for F or O we can generate both old and new distribution functions on the sphere. With

F	O	Name	Comment
$-\nabla^2(E(q))$	q	DNI≡EITL	calculates the real ODF without the complications of the FFT
$\nabla^4(E(q))$	q	EITL2	high resolution at low angles
$E(q)$	q	EITS	impressive resolution without any preprocessing of the signal
$E(q)$	1	'QBI'-like	similar to the Funk Radon Transform

Table 2.1: The Equatorial inversion transform (EIT) can be used to explain many other reconstructions algorithms.

$F(E(q)) = -\nabla^2(E(q))$ and $O(q) = q$ we can generate ψ_{DNI} which is theoretically identical to the DSI real ODF(ψ_{DSI}). If $F(E(q)) = E(q)$ and $O(q) = 1$ then this is similar to the Funk Radon Transform (used in Q-ball imaging) but applied to multiple spherical shells. However, we can also try to use different functions like $F(E(q)) = -\nabla^4(E(q))$ and $O(q) = q$ which can potentially increase the amount of directional information beyond that of the standard ODFs. Before starting investigating the realms of EIT we will first give a short overview of other methods commonly found in the literature.

2.6 Implementation

2.6.1 Standard EIT

Eq. 2.18 and 2.19 can be implemented in a standard way by evaluating the 3D signal on the grid multiple times for every direction \hat{u} as shown in Fig. 2.3A. This suggests that if for example we use a reconstruction sphere of 642 vertices and the signal is centered inside a cubic grid of size $17 \times 17 \times 17$ where the radial integration (q) takes place in 30 steps and the equatorial (ϕ) in 63 steps, then we need to interpolate $642 \times 30 \times 30 \simeq 1.2$ million times on the cubic grid. For this reason we invented Fast EIT, a new method that needs an order of magnitude less evaluations.

In this document whenever we use the prefix 's' in front of a method it means it was calculated with the standard EIT algorithm. For example if standard EIT is used for DNI we will write sDNI or sEITL. ('L' stands for 'Laplacian' or 'Nabla'). Of course, sDNI and sEITL are equivalent.

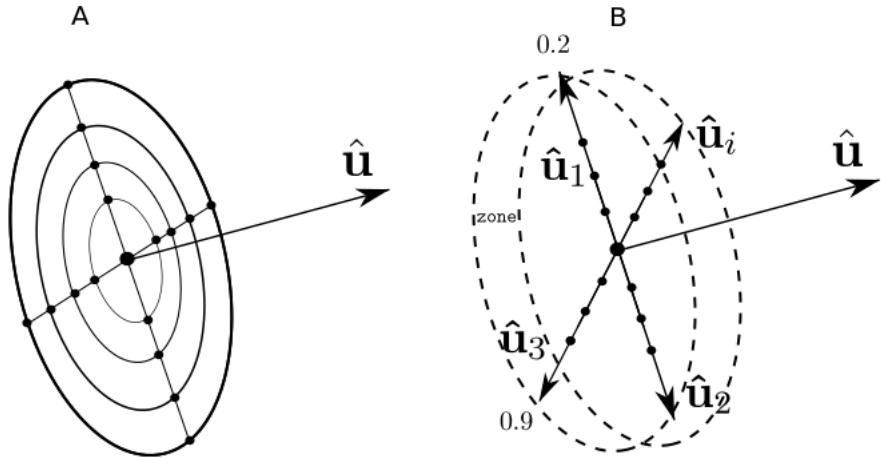


Figure 2.3: A: Standard EIT. B: Fast EIT. Fast EIT is an order of magnitude faster than standard EIT. The key idea here is that we can reduce computations by storing the sum of the radial integrals for every vertex in the reconstruction sphere and then we can also precompute the indices of the vertices that are near the equator of every vertex (inside an equatorial zone).

2.6.2 Fast EIT

A much faster algorithm than the standard EIT is described here. The main idea is that we can store the sum of the radial integrals for every vertex in the reconstruction sphere so we can then precompute the indices of the vertices that are near the equator of every vertex (inside an equatorial zone). This becomes clear in Fig. 2.3B. Following these calculations, the spherical distribution function can be approximated with much less operations. The full algorithm is given in Alg. 1. The input is the vertices $\hat{\mathbf{u}}_i$ of the reconstruction sphere and the normalized signal E . Then, for every point of the reconstruction sphere $\hat{\mathbf{u}}_i$, we save the indices of the vertices j of $\hat{\mathbf{u}}_j$, which are inside an equatorial zone, in list J_i . The width of the equatorial zone z is a constant set empirically to 5° . If a very highly dense reconstruction sphere is used with more than 642 vertices, which is the one we used, then the zone can be smaller. That can potentially increase the angular resolution of the method.

At the next stage we calculate sums along every radius on the direction of $\hat{\mathbf{u}}_i$ in the following way: $\mathbb{B}(\hat{\mathbf{u}}_i) = \sum_{k=0}^n F(E(q_k \hat{\mathbf{u}}_i)) O(q_k \hat{\mathbf{u}}_i)$ and obtain the final EIT ODF as the average of the sums in the equator $\psi_{EIT}(\hat{\mathbf{u}}_i) = \frac{1}{N_i} \sum_{j \in J_i} \mathbb{B}(\hat{\mathbf{u}}_j)$ where F is evaluated with trilinear for example interpolation on the lattice and N_i is the number of indices in J_i .

Algorithm 1 Fast Equatorial Inversion Transform

Input $U = \{\hat{u}_1 \dots \hat{u}_m\}$, E**Output** ψ_{EIT} **For** \hat{u}_i in U **Do**

$$J_i = \{j : |\arccos(\hat{u}_i \cdot \hat{u}_j)| \leq z\}$$

$$\mathbb{B}(\hat{u}_i) = \sum_{k=0}^n F(E(q_k \hat{u}_i)) O(q_k \hat{u}_i)$$

where $F(E(q_k \hat{u}_i))$ is interpolated on the lattice.**EndFor****For** \hat{u}_i in U **Do**

$$\psi_{EIT}(\hat{u}_i) = \frac{1}{N_i} \sum_{j \in J_i} \mathbb{B}(\hat{u}_j)$$

where N_i is the number of indices in J_i .**EndFor**

In section 2.9.1 the standard EITL (sEITL) is compared with fast EITL. In Fig. 2.11 it is shown that the fast EIT has very similar results with the standard EIT. From now on whenever we write EIT we refer to the fast version.

2.7 Peak Finding

After we have generated the ODFs we need to find the peaks (local maxima) from which we can easily approximate the direction of the fibres. Peak finding is not trivial if there are many local maxima in the ODFs or the ODFs are noisy. Here we present an algorithm (see Alg. 2) which reduces the amount of small local variations and returns a number of sorted peaks and their indices in the reconstruction sphere. The input of this algorithm is ψ (ODF) and the faces of a symmetric on the z-axis evenly distributed sphere (see Fig. 2.4C).

We have used a triangulation of the unit sphere (which we refer to simply as 'sphere') obtained by triangular subdivision of a regular icosahedron. It is symmetric over the z-axis, i.e. for each vertex (x, y, z) there is a corresponding vertex $(x, y, -z)$ in the opposite hemisphere. The same sphere was used in [72] for GQI reconstructions. Every face (triangle) corresponds to a list of the 3 indices of the 3 vertices on the sphere. The idea here is that we can travel from face to face and nullify all points on a face which are lower than the higher value of the face. At the end only local maxima will survive the procedure. The algorithm is presented in detail in Alg. 2.

Algorithm 2 Peak Finding with a Symmetric Ordered Sphere

Input ODF ψ , faces Φ
Output peaks P and indices I

```
For face  $\Phi_i$  in  $\Phi$  Do
     $f_0, f_1, f_2 = \Phi_i$ 
     $d_0, d_1, d_2 = \psi[f_0], \psi[f_1], \psi[f_2]$ 
    If  $d_0 \geq d_1$  and  $d_2$  Do
         $P[f_1] = P[f_2] = 0$ 
        continue
    If  $d_1 \geq d_0$  and  $d_2$  Do
         $P[f_0] = P[f_2] = 0$ 
        continue
    If  $d_2 \geq d_0$  and  $d_1$  Do
         $P[f_0] = P[f_1] = 0$ 
        continue
    EndIf
EndFor
```

The sphere we use is of course discrete and therefore it adds some constraints on the angular resolution (worst case $\pm 4.96^\circ$) of the peaks found from the ODF. In addition, the proposed Peak Finding algorithm can reduce slightly more the angular resolution. For example, in Fig. 2.4A, B we show that if point **a** was a local maxima then only points **b** could be alternative local maxima for **a** but none of the unlabeled points could be a second peak. Nevertheless, we found Alg. 2 to be extremely useful and fast. The same algorithm was used also in [72] but it was not documented as such.

2.8 Spherical Angular Smoothing

All current non-parametric dMRI reconstruction algorithms use some type of “smoothing” to reduce the effect of noise in the real data sets. DSI uses a Hanning filter and then avoids sampling from low values in \mathbf{r} -space. In GQI, smoothing is controlled from a scalar parameter; the diffusion sampling length and in spherical harmonic inversion methods [97], [96] the amount of smoothing is controlled by using only a number of the first components of the SH series.

All these approaches smooth and calculate the ODFs simultaneously. Our approach differs in that we propose that the ODF is first calculated and

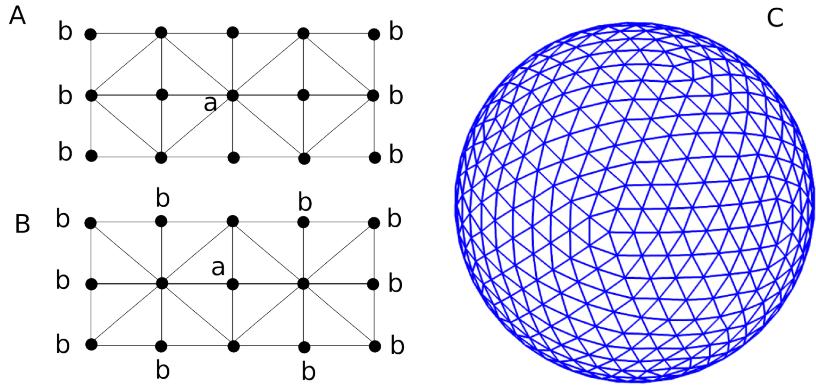


Figure 2.4: A and B: Each point **a** is a local maximum for all its neighboring faces, then only at **b** are other possible local maxima. This simple illustration shows that the triangulation of the sphere is important for the determination of closed peaks and that peaks which belong to the same triangle cannot be determined. C: the sphere used for ODF reconstructions consisting of 642 vertices and 1,280 faces produced by subdivisions of the icosahedron.

then smoothed. For example, using the operator shown below in matrix form

$$W = \exp\left(\frac{U \cdot U^T}{\sigma}\right)$$

where U is the an $N \times 3$ matrix holding the N points of the ODF reconstruction sphere and σ is a smoothing parameter acting like the variance. At the next step we can smooth any $ODF(\psi)$ creating a new $ODF(\psi')$ in the following way

$$\psi' = \psi \cdot \frac{W}{\sum_j W_j} \quad (2.22)$$

where j denotes row indexing, $\sum_j W_j$ acts as a normalization for the angular weighting W , ψ is the initial ODF and ψ' is the smoothed ODF. The advantage of this method is that it is more comprehensive and direct. It also uses information from all directions simultaneously. Similar operators can be constructed that weight more lower or higher peaks. The operator shown here weighs more peaks that are closer in angular distance. In Fig. 2.5 we see the effect of this equation on a simulated triple-fibre crossing; distorted with Gaussian noise with SNR 20 and reconstructed as an EITL density function. The simulation was used using a Sticks and Ball model with diffusivity value 0.0015 and S0=100.

We can easily observe that when we increase the smoothing factor σ ,

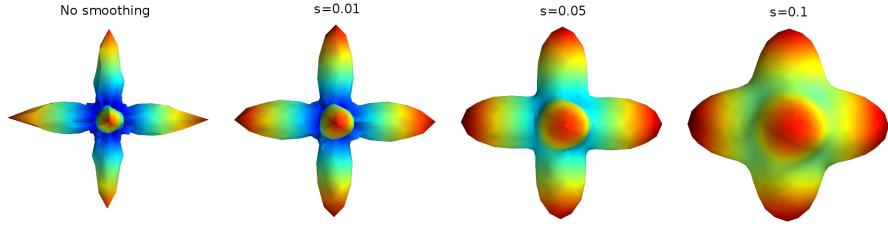


Figure 2.5: An example of spherical angular Gaussian smoothing applied with different smoothing factors on the ODF of a triple-fibre crossing on the left.

small noisy peaks, as seen in the center of the unsmoothed spherical function, can easily be removed. However, with too much smoothing even the strongest peaks can lose their definition. This spherical operator can help to set the trade-off between noise and signal and it can also simplify the peak finding process, i.e. finding the underlying primary fibre directions as this problem is much easier on smooth surfaces.

Finally, decoupling the smoothing from the reconstruction step gives an important advantage: reducing the effect of the noise to our data more strongly and independently. Many spherical operators can be added as plugins independently of the reconstruction phase, and these can work with any function on the sphere (see Eq. 2.22).

2.9 Comparisons and Results

Validation of reconstruction and tractography algorithms is not straightforward due to the lack of relevant gold standards. Simulated voxels and software phantoms are a useful way to overcome this difficulty and test new methods. Following the simulation results, we show results with real human data sets.

2.9.1 Multi-fibre Simulations

For single voxel simulations we used the model proposed in Behrens et al. [35]; the multi-compartment model also known as Sticks and Ball which simulates the diffusion signal as

$$S_i = S_0 \left((1 - \sum_{j=1}^N f_j) \exp(-b_i d) + \sum_{j=1}^N f_j \exp(-b_i d \cos(\theta_{ij})^2) \right) \quad (2.23)$$

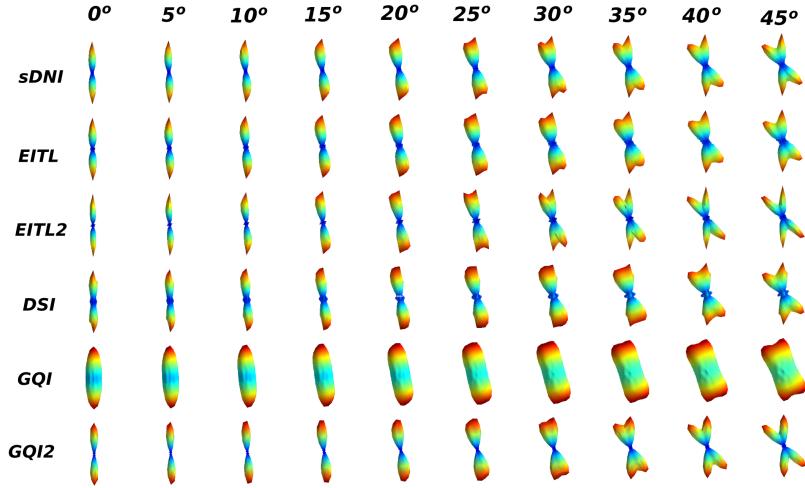


Figure 2.6: Visualizing ODFs created from different reconstruction methods sDNI (sEITL), EITL, EITL2, DSI, GQI and GQI2. We can see that standard DNI (sDNI), EITL and EITL2 can resolve the correct angular fibre directions at lower angles than the other methods. For example see column at angle of 25° .

where θ_{ij} is the angle between gradient direction $\hat{\mathbf{g}}_i$ and fibre (stick) unit direction $\hat{\mathbf{u}}_j$. The amount of representation for every fibre is given by f and d is the diffusivity value for the entire model. A Multi Tensor [89] approach was also created for software phantoms using the formula

$$S_i = S_0 \sum_{j=1}^N \exp(-b\hat{\mathbf{g}}^T D_j \hat{\mathbf{g}}) \quad (2.24)$$

where D_j is the diffusion tensor for every fibre j .

In Fig. 2.6 we present the outcome of an experiment of two crossing fibres using different reconstruction methods: sDNI (sEITL), EITL, EITL2, DSI, GQI and GQI2. These are based on simulations of 2-fibre crossings from 0° to 90° using Eq. 2.23 with diffusivity value of $1.5 \times 10^{-3} \text{ mm}^2/\text{sec}$ and 257 b-values with maximum b-value 11,000. However, all these methods will perform accurately beyond 50° therefore in this figure we present only the lower angles. We can see that sDNI, EITL and EITL2 performed better than the other methods. Especially EITL2 was able to resolve crossing at 25° which is lower than the accuracy resolved in the current state of the art methods. In order to confirm this fascinating result we created

Known	Measured	AS
$(1, 0, 0), (0, 1, 0)$	$(0, 0, 1)$	0
$(1, 0, 0), (0, 1, 0)$	$(0, 1, 0)$	1
$(1, 0, 0), (0, 1, 0)$	$(0, \sqrt{2}/2, \sqrt{2}/2)$	$\sqrt{2}/2$
$(1, 0, 0), (0, 1, 0), (0, 0, 1)$	$(1, 0, 0), (0, 0, 1)$	2

Table 2.2: Examples of angular similarity (AS) behaviour with simple unit vector sets.

a more general experiment with many iterations (of 2-fibre and 3-fibre crossings) which is presented below.

A comparison metric is needed in order to evaluate the new and old reconstruction methods discussed in this chapter. The standard procedure is to calculate the similarity between the measured and simulated ground truth data sets. We want to calculate the angular precision of the ODFs from simulations derived from Eq. 2.23. We define a new similarity metric called Angular Similarity (AS) which computes the cosine distance of the best match between the set of measured fibre orientations and the known set of simulated fibres. This metric will be used to compare 2-fibre and 3-fibre crossings. AS is 0 when there is no match i.e. angular distance is 0, 1 when one fibre is matched (0°), 2 when two fibres are matched and 3 when three fibres are matched. In table 2.2 we show a few examples of AS behaviour with simple unit vector sets.

If our ground truth set consists of $g = [(1, 0, 0), (0, 1, 0)] = [g_0, g_1]$ and the measured set consists of $m = [(0, 0, 1)]$ then AS=0. If the measured set was $m = [(0, \sqrt{2}/2, \sqrt{2}/2)] = [m_0]$ then AS is $\sqrt{2}/2$. This is because according to the AS definition we have $AS(g, m) = \max(|g_0 \cdot m_0|, |g_1 \cdot m_0|)$. Which is equal to $\sqrt{2}/2$.

If $g = [(1, 0, 0), (0, 1, 0)] = [g_0, g_1]$ and $m = g$ then $AS(g, m) = \max(|g_0 \cdot m_0| + |g_1 \cdot m_1|, |g_0 \cdot m_1| + |g_1 \cdot m_0|) = 2$. We created an experiment where we set two fibres at an increasing angle of 2.5° from 0° to 90° and then rotate them uniformly around 200 random axes. This operation produces 7,400 simulated ODFs and the results are shown in Fig. 2.7 and Fig. 2.8 with different signal to noise ratio. For these simulations noise was normally distributed. What we see in the figures is the average angular similarity where the average is calculated from the 200 random orientations for the same angle.

We can easily observe in Fig. 2.7 that EITL2 can resolve more accurately

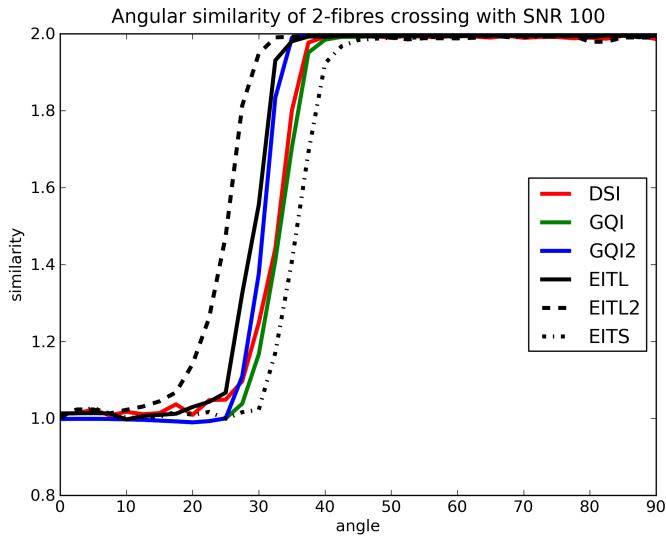


Figure 2.7: Average angular similarity of 2-fibre crossings with SNR 100

fibre crossings at low angles and continues to perform well even at higher angles $> 50^\circ$. EITL performs better than DSI, GQI, GQI2 and EITS at low angles and very well at high angles as well. GQI2 performs better than DSI, GQI, and ETS. It is also impressive that EITS can have such a good performance although it is such a simple operation. In summary we see from the graphs that EITL2 > EITL > GQI2 > DSI > GQI > EITS where $>$ means higher average angular similarity. The same pattern takes place even when we increase the noise level see for example Fig. 2.8. We will see next that the same pattern appears even with 3-fibre crossings and high levels of noise.

We also measured the accuracy in 3-fibre crossings. In this experiment the 3-fibres will always have the same angular distance between each other. That distance will increase from 0° to 90° with steps of 2.3° on average and all 3 fibres will be reoriented 200 times. That gave 8,000 simulated crossings.

The results of the 3-fibre crossings shown in Fig. 2.9 and Fig. 2.10 were very similar to those of the 2-fibre crossings; EITL2 performed better at low angles, showing slightly reduced performance at high angles. EITL performed better with low angles than the rest of the methods, having also high accuracy on larger angles.

These summarization plots give strong evidence that both DNI (EITL) and in general EIT can be used to accurately generate spherical distri-

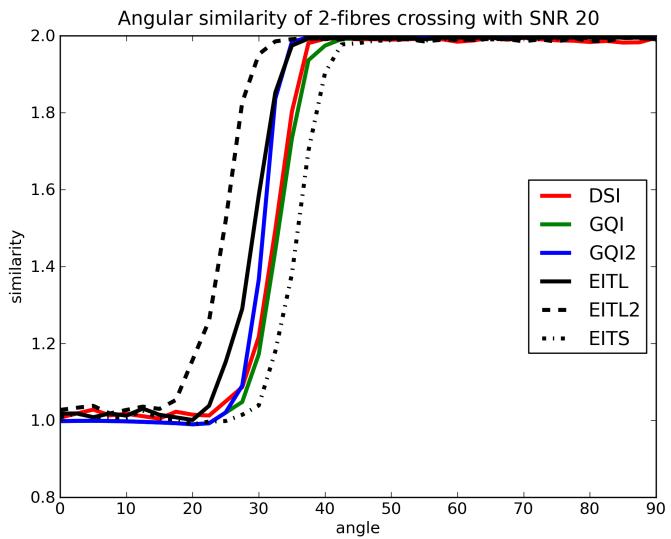


Figure 2.8: Average angular similarity of 2-fibre crossings with SNR 20.

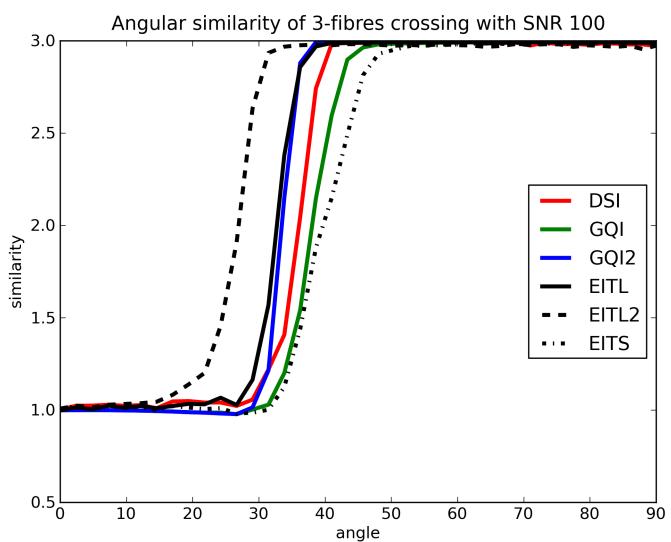


Figure 2.9: Average angular similarity of 3-fibre crossings with SNR 100.

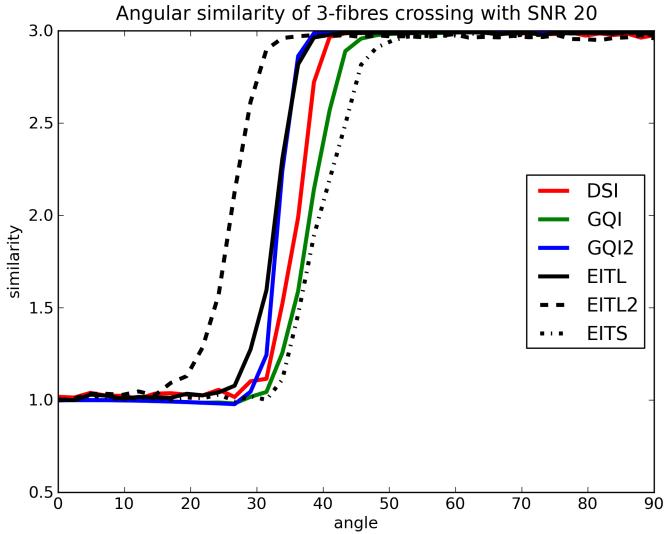


Figure 2.10: Average angular similarity of 3-fibre crossings with SNR 20.

bution functions for the determination of the directional information of the diffusion signal. They performed equivalently or better than the current state-of-the-art grid-based reconstruction methods i.e DSI and GQI. The determination of the fibre directions was not affected considerably by noise.

Furthermore, we can also see that GQI2 can do better than DSI, GQI and that EITS gives results that are very similar to GQI. The parameters used for these simulations were DSI: radial sampling 2.1 – 6, hanning filter width: 36, GQI: $\lambda=1.2$, GQI2: $\lambda = 3$, and EITS, EITL, EITL2 were all calculated with the standard options zonal width ($z = 5^\circ$), grid size $17 \times 17 \times 17$, radial sampling 0 – 5 with 0.1 steps and no further post-processing or smoothing was used. All methods were using the same reconstruction sphere with 642 vertices and 1,280 faces.

In these tests, EIT and fast EIT produced very similar results. For example, a simple test for the 3-fibre case as seen in Fig. 2.11 shows that there is close agreement between the two methods i.e. their results are nearly equivalent. We can therefore conclude that the fast EIT is an acceptable approximation of the standard EIT.

2.9.2 Software Phantoms

A software phantom generation tool was developed which can simulate the diffusion weighted signal for one or more fibres represented by differ-

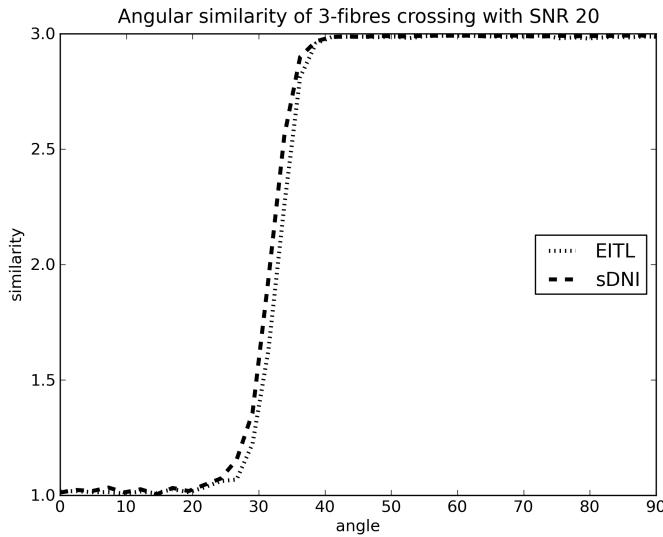


Figure 2.11: This diagram shows that when we compute EITL with the fast or standard method the results are nearly equivalent. The mean angular similarity for the case of 3-fibres crossings is very similar when using standard DNI or fast DNI (EITL).

ent discrete 3D orbital functions. This work is an extension of the phantom developed by Correia et al. [98] which supported only paths with analytically calculated derivatives.

The idea here is that we first create any orbital function $f(t) : \mathbb{R} \rightarrow \mathbb{R}^3$ and calculate numerically its derivatives at small steps Δt . We can then scale it and centre it so that it fits in an image volume of the desired size. We expect that many segments of the discrete function f will fall into every voxel in the volume and that more curved parts of f will have higher representation in the voxel than less curved parts. For every segment, we can find the main direction of the orbit $\mathbf{v} = \frac{f(t+1) - f(t)}{\Delta t}$ and calculate the rotation matrix \mathbf{R} that rotates $\hat{\mathbf{x}} = (1, 0, 0)$ to \mathbf{v} . Then, the signal for each element of the fibre for a given b -value b and a given gradient sampling direction $\hat{\mathbf{g}}$, is given by the following Single Tensor formula

$$\Delta S = S_0 \exp(-b\hat{\mathbf{g}}^T \mathbf{R} \Lambda \mathbf{R}^T \hat{\mathbf{g}}) \quad (2.25)$$

where S_0 is the unattenuated signal of the fibre, and the diffusion tensor is given by

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_{\parallel} & 0 & 0 \\ 0 & \lambda_{\perp} & 0 \\ 0 & 0 & \lambda_{\perp} \end{pmatrix} \quad (2.26)$$

Therefore, the total signal of the voxel for one gradient direction is given by the summations of all the contributions of the K elements in the voxel

$$S_{vox} = \sum_{i=1}^K \Delta S_i \quad (2.27)$$

In addition, we can generate simulations of more than one fibre by generating a single volume for every orbit and then add them all together to create complex configurations in the final volume. This is acceptable, under the assumption that the diffusion is Gaussian in all compartments, because the diffusion signal is additive i.e. the signal of a crossing of two fibres is equal to the sum of the signals of the individual fibres. In this way, we can simulate phantoms with Multi Tensor based diffusion signals as that described in Eq. 2.24. We can increase the thickness of the fibres using a typical smoothing kernel or duplicate the fibres radially. At the end we can add different levels of noise e.g. Rician or Gaussian noise with a prespecified SNR.

The method we use to create these software phantoms offers the opportunity to simulate partial volume effects. If partial volume effects are not desired, we need to normalize by dividing by the number of fibre elements for each voxel. In Fig. 2.12 we can see the volume renderings of two different phantoms created with the method described here. This function is implemented in module `dipy.sims.phantom`.

2.9.3 Results with software phantoms

With the purpose of comparing and visualizing the differences between the reconstruction methods described in this chapter a phantom of two crossing bundles was created. The bundles are crossing at an angle of 90° . The phantom was generated using the method described in the previous section. Here we describe the basic steps: (a) We first represented the first bundle as a discrete straight path starting from point $(-1, -1, 0)$ and ending at point $(1, 1, 0)$ with using 1,000 time steps. (b) We scaled, centred

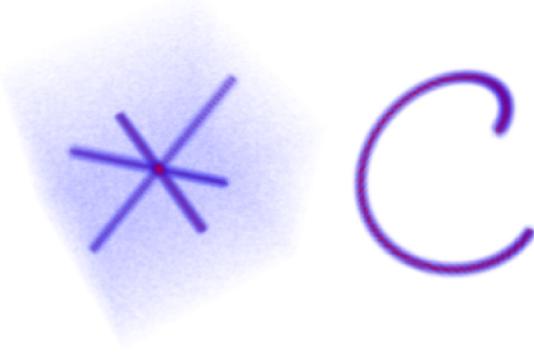


Figure 2.12: Volume renderings of the unattenuated signals of two digital phantoms. On the left 3 fibres intersect on regular angles with Rician noise of SNR=20. On the right a helicoidal fibre is shown clear of noise. For both phantoms $S_0 = 100$ and prolate tensors with eigenvalues $\lambda_{\parallel} = 1.4 \cdot 10^{-3} m^2/sec$ and $\lambda_{\perp} = .35 \cdot 10^{-3} m^2/sec$ were used.

and radially expanded this path so that it fits a volume of size $64 \times 64 \times 64$. This volume corresponds to the diffusion volume without any weighting. (c) We then applied the weightings for all the following volumes corresponding to non-zero b-values. (d) We replicated the same procedure for the other bundle which initially started as an orbit from position $(-1, 1, 0)$ and ended at position $(1, -1, 0)$. (e) We added the two volumes together to create an 'x' shape (see Fig. 2.13, 2.15). (f) We added Rician noise with SNR=5. As in this chapter we concentrate on Cartesian Lattice Q-space acquisitions we generated b-vectors and b-values using a keyhole Cartesian sampling grid [99] with 515 q-vectors. The maximum b-value was 11,538 and the minimum was 0. Two sets of simulation experiments were performed each using a Tensor of different shapes.

In the first experiment shown in Fig. 2.13, 2.14 we used a more anisotropic prolate tensor for the simulation with eigenvalues $\lambda_{\parallel} = 1.4 \times 10^{-3} mm^2/sec$ and $\lambda_{\perp} = 0.1 \times 10^{-3} mm^2/sec$. In the second experiment shown in Fig. 2.15 and Fig. 2.16 we used a much less anisotropic prolate tensor with $\lambda_{\parallel} = 1.7 \times 10^{-3} mm^2/sec$ and $\lambda_{\perp} = 0.3 \times 10^{-3} mm^2/sec$. The values of λ are based in [16]. It is well known that noise has higher effect on less anisotropic areas. We can see this effect by comparing the overlapped FAs of these two figures (2.13, 2.15). We can also see that all six methods (DSI, GQI, GQI2, EITL, EITL2, EITS) can resolve correctly the fibre directions by looking at their spherical distribution functions using a standard

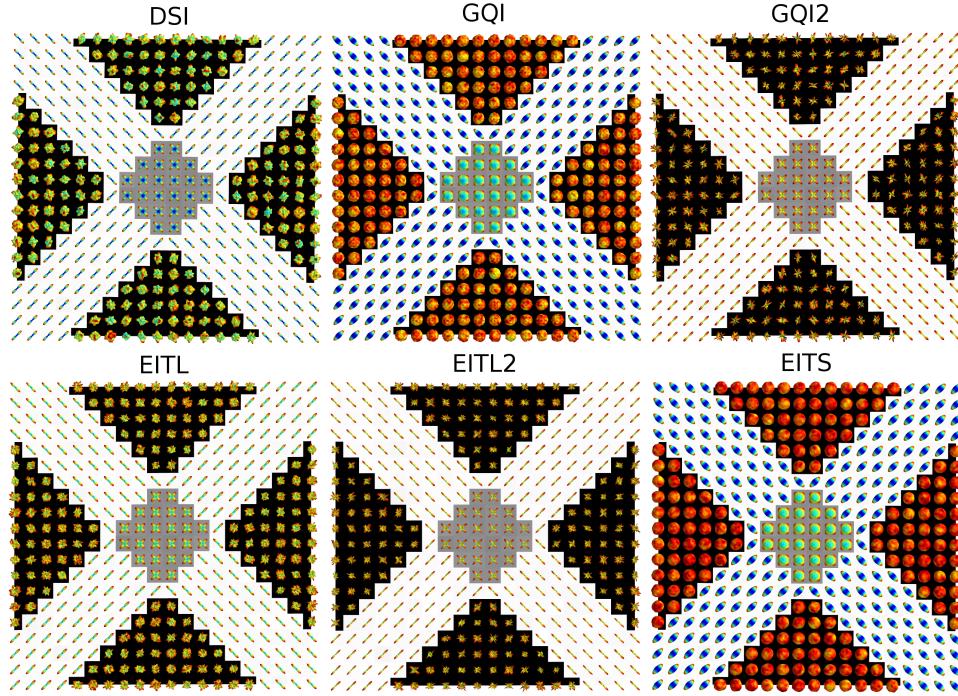


Figure 2.13: Results with an 'x' shape software phantom. Every single tensor compartment had the following eigenvalues $\lambda_{\parallel} = 1.4 \times 10^{-3} \text{ mm}^2/\text{sec}$ and $\lambda_{\perp} = 0.1 \times 10^{-3} \text{ mm}^2/\text{sec}$. Rician noise was added with SNR = 5. GQI is very similar to EITS, GQI2 is very similar to EITL and DSI is very similar to EITL. In Fig. 2.14 the regions at the centers of the phantoms are depicted in higher resolution.

colour map. For visualization purposes all ODFs are shown in relative size as they have been scaled so that their maximum values correspond to 1.

Furthermore, we can easily observe that GQI is mostly similar to EITS, GQI2 is very similar to EITL and DSI is mostly similar to EITL. The fact that DSI ODFs are very similar to those of EITL ODFs is to be expected as the two methods create theoretically the same real ODFs. Remarkably, EITL can create these ODFs without using the Fourier Transform neither using any filter or thresholds in r-space which are necessary in DSI.

Fig. 2.13, 2.15 show that all the different grid-based reconstruction methods can reconstruct correctly the underlying fibre directions even when noise is present. However, we can see that when tensors are less anisotropic the noise has a stronger effect in the resulting spherical distributions. We can also see that GQI & EITS are less sharp than DSI & EITL and these are less sharp than GQI2 & EITL2. Also DSI, GQI2, EITL, EITL2 have much

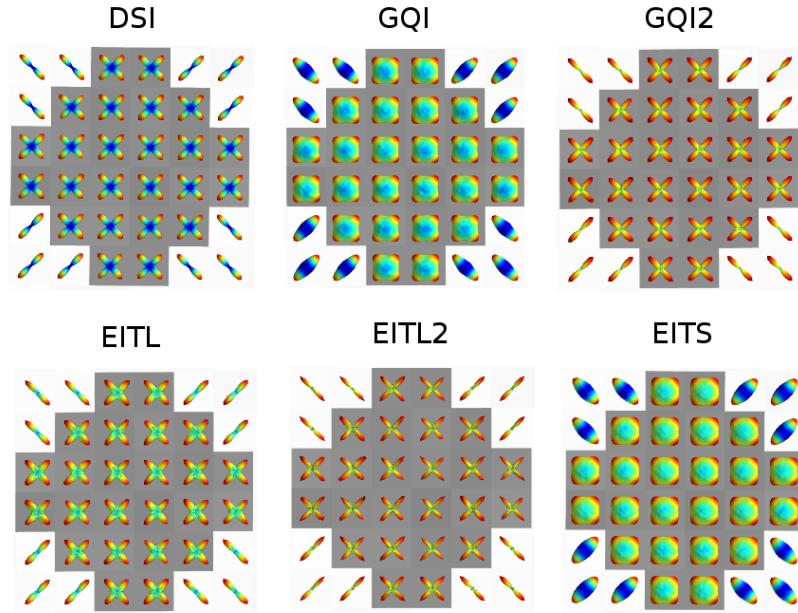


Figure 2.14: Same as Fig. 2.13 showing in higher resolution the spherical distributions in the centers of the phantoms.

lower minima than GQI and EITS.

In the EIT-based reconstruction results shown in Fig. 2.14 and Fig. 2.16 we do not use any amount of smoothing as used in DSI (through hanning filter), GQI, GQI2 (through sampling length) and it is extraordinary that we still obtain such well defined distributions. If we want to apply some weighting/smoothing/denoising in EIT-based methods that is simply possible through the spherical angular smoothing approach described in section 2.8.

The parameters used for these simulations were for DSI: radial sampling 2.1 – 6, hanning filter width: 36 , GQI: $\lambda=1.2$, GQI2: $\lambda = 3$, and EITS, EITL, EITL2 were all calculated with the standard options ($z = \pm 5$) and no further post-processing or smoothing was used. All methods were using the same reconstruction sphere with 642 vertices and 1,280 faces.

2.9.4 Results with humans

We want to compare reconstruction methods on Cartesian grid-based acquisitions first with data sets which are rich on directions and commonly used for DSI processing. For this purpose we used a data set which was available online at cmtk.org from the Diffusion Group at Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland. So, this data set was

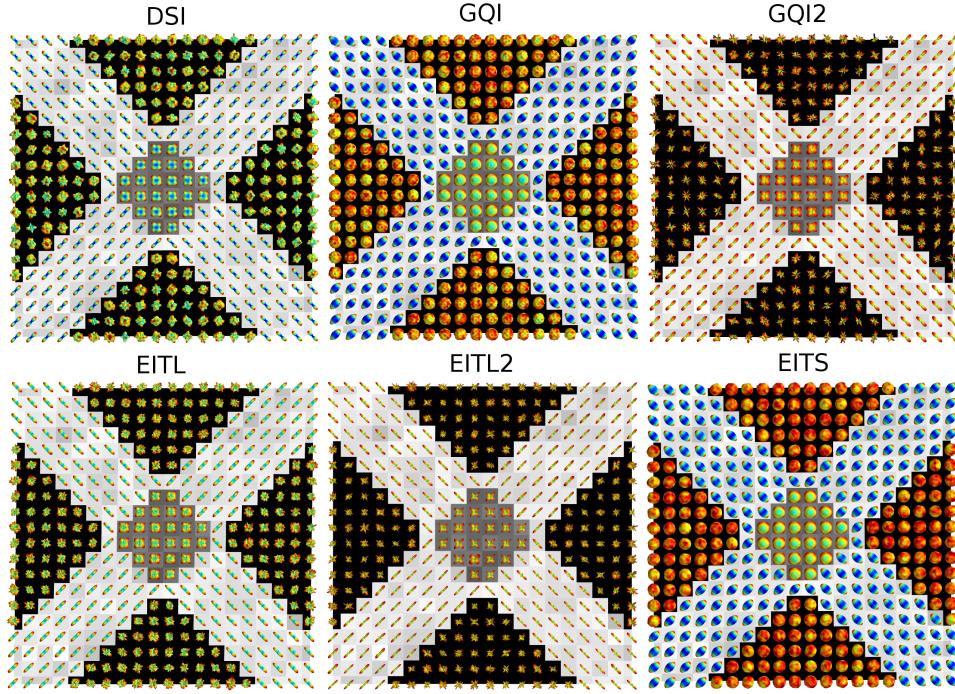


Figure 2.15: Showing the spherical distribution functions (DSI, GQI, GQI2, EITL, EITL2, EITS) of a software phantom generated by two bundles where each bundle contains single tensors along the direction of the phantom. On the crossing area, a dual Tensor effect in every voxel is observed. Every single Tensor compartment had the following eigenvalues $\lambda_{\parallel} = 1.7 \times 10^{-3} \text{ mm}^2/\text{sec}$ and $\lambda_{\perp} = 0.3 \times 10^{-3} \text{ mm}^2/\text{sec}$. Rician noise was added with SNR=5. We also visualize simultaneously the FA for this slice. We can see that in the crossing area (gray background) the FA values drop considerably. However, the ODFs represent precisely the crossing.

obtained from a 3T scanner (TIM Trio, Siemens) with a 32 channels head coil. The field of view was $210 \times 210 \text{ mm}^2$, matrix size 96×96 , and slice thickness 3 mm. 44 slices were acquired and the voxel resolution was $2.2 \times 2.2 \times 3.0 \text{ mm}^3$. A 258-point half grid acquisition scheme with a maximum b-value of 8011 s/mm^2 also known as DSI515 [100] was used. The total acquisition time was 34 min with TR=8200 ms and TE=165 ms.

The parameters used for these simulations were for DSI: radial sampling 2.1 – 6, Hanning filter width: 36, GQI: $\lambda=1.2$, GQI2: $\lambda = 3$, and for EITS, EITL, EITL2 were all calculated using the standard options for zonal width ($z = 5^\circ$) and spherical angular smoothing ($s = 0.05$). All methods were using the same reconstruction sphere 642 vertices and 1,280 faces. The results of this experiment are shown on top of an FA slice of a healthy human in Fig. 2.17 and in higher resolution in Fig. 2.18. It is observed

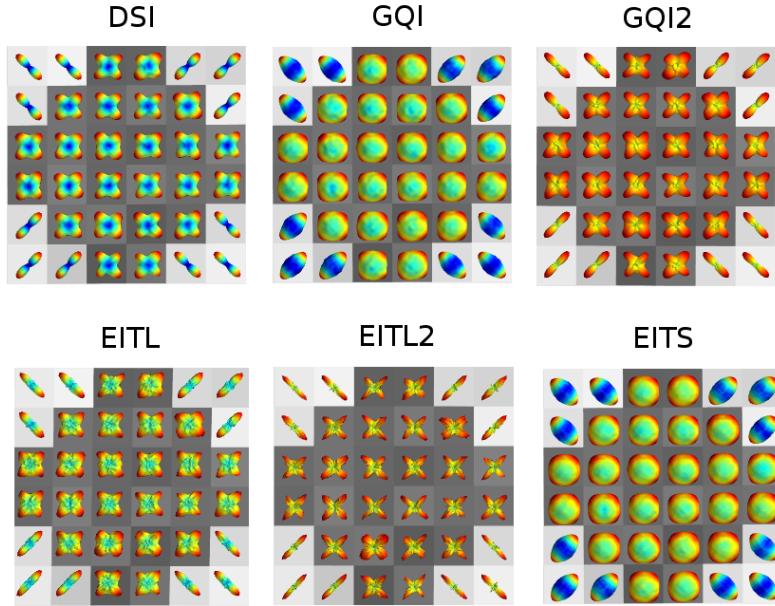


Figure 2.16: A zoomed version of Fig. 2.15 showing the spherical distributions in the centers of the phantoms at higher resolution.

that EITL, EITL2 and EITS can be used for reconstructing these data sets as their results appear very similar to the results given by DSI, GQI and GQI2. We can also easily see that EITL and EITL2 are relatively sharp which can be of an advantage for the purpose of recovering correctly the underlying real fibre directions.

We also tested our results with another human brain data set generated at a 3T scanner (TIM Trio, Siemens) at the Medical Research Council Cognition and Brain Sciences Unit, Cambridge, UK. We used Siemens advanced diffusion work-in-progress sequence, and STEAM [101, 15] as the diffusion preparation method. The field of view was $240 \times 240 \text{ mm}^2$, matrix size 96×96 , and slice thickness 2.5 mm (no gap). 55 slices were acquired to achieve full brain coverage, and the voxel resolution was $2.5 \times 2.5 \times 2.5 \text{ mm}^3$. In this experiment a smaller number of gradient vectors were used. A 102-point half grid acquisition with a maximum b-value of $4,000 \text{ s/mm}^2$ was used. The total acquisition time was only 14 min 21 s with TR=8, 200 ms and TE=69 ms.

In Fig. 2.19 a slice is shown where different parts of white matter are visible with the FA background image. We can clearly see structures like the Corpus Callosum (CC) and Cortical-Spinal Tract (CST) and Centrum Semiovale areas. The ODFs of EITL are shown superimposed on the FA.

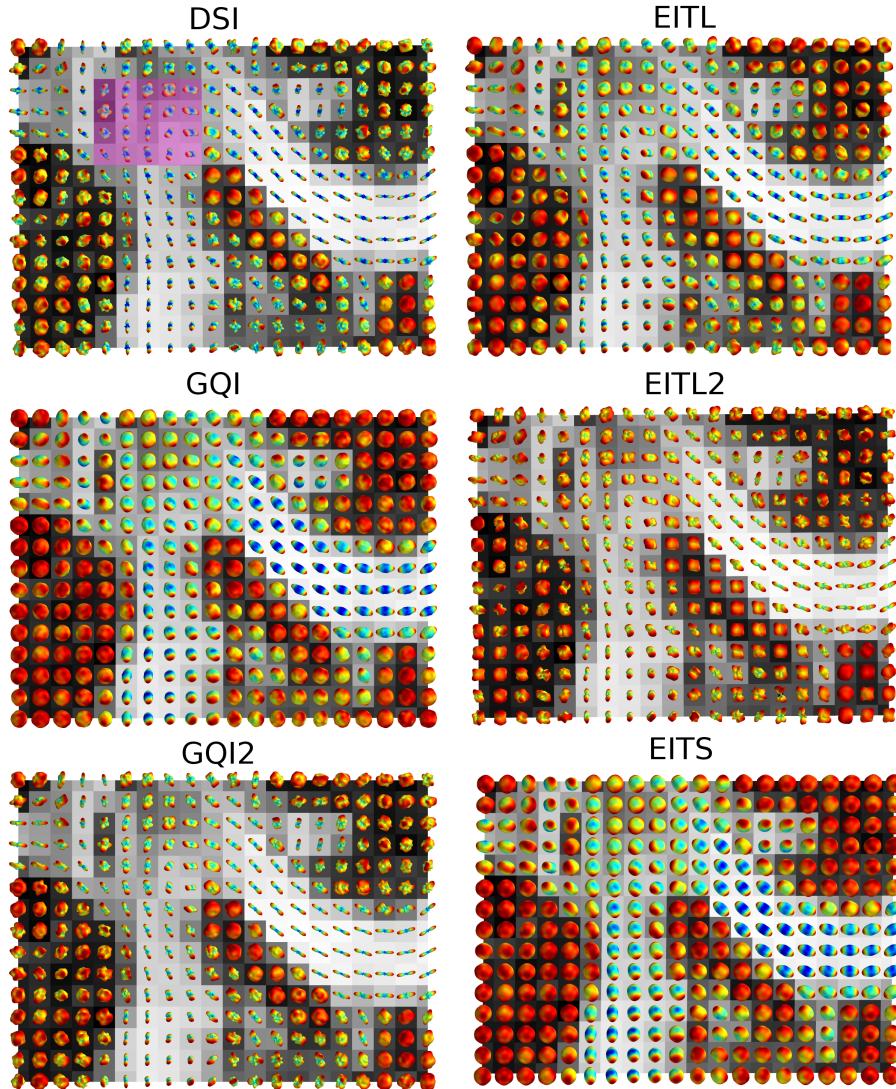


Figure 2.17: Showing the same slice of a human brain reconstructed with 6 different Cartesian grid q-space based methods. The ODFs are visualized on top of the FA slice. A clearer presentation of a region near the left upper corner (with purple shading) is given in Fig. 2.18 for all the 6 methods.

The parameters used for EITL were: a standard zonal width $z = 5^\circ$ and spherical angular smoothing $s = 0.05$ with the same reconstruction sphere (642 vertices, 1,280 faces) as before.

For illustration purposes the upper part of Fig. 2.19 is depicted again in Fig. 2.20, and the region with purple shading from Fig. 2.20 is given at an even higher resolution in Fig. 2.21. We used Mayavi [102], a Python visualization library based on VTK to make the visualizations shown in the figures of this section.

Although less directions were used in this acquisition scheme we ob-

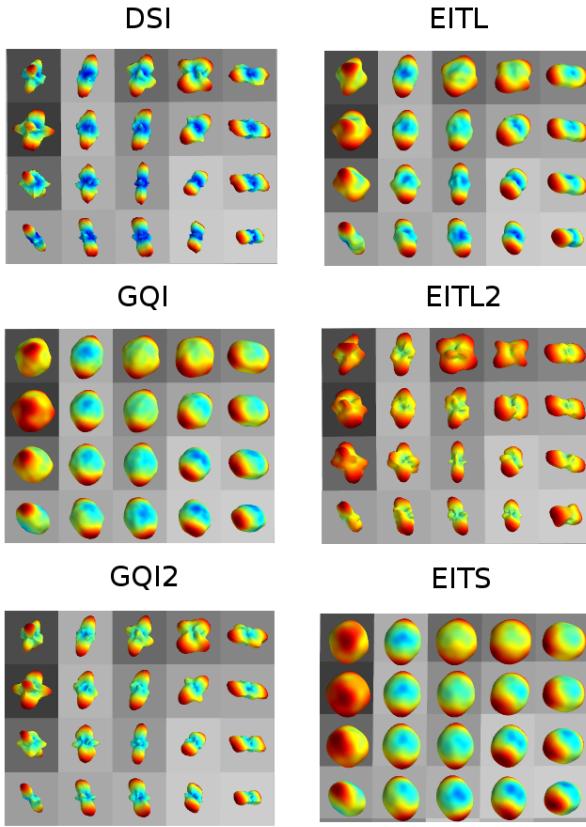


Figure 2.18: The upper-left corners (purple shading region) of the panels of Fig. 2.17 are shown here at higher resolution. These data sets belong to a real human. In contrast with the results shown in simulations (see Fig. 2.14) we applied spherical angular smoothing with $s = 0.05$ for EITL, EITL2 and EITS in order to remove small noisy spikes in the distributions. In agreement with the results of Fig. 2.14, EITS is very similar to GQI. The difference between DSI, GQI2 and EITL, EITL2 is smaller as a result of the application of angular weighting.

tain a similarly accurate depiction of the underlying white matter structure in comparison with that of 258 directions. This gives great hope that we can use grid-based reconstruction methods with half-grid sequences with 100 gradient directions. This was also shown by [74] and [72] who used similar number of directions.

In all the figures with real data sets we can see single fibres as those usually found at the center of CC, and 2 or 3-fibre crossings in the intersection areas of CC with the CST and other bundles.

We can see for example in Fig. 2.21 that the effect of spherical angular smoothing can help alleviate the noise effects and focus our concentration on depicting the major directions which are also of highest concern.

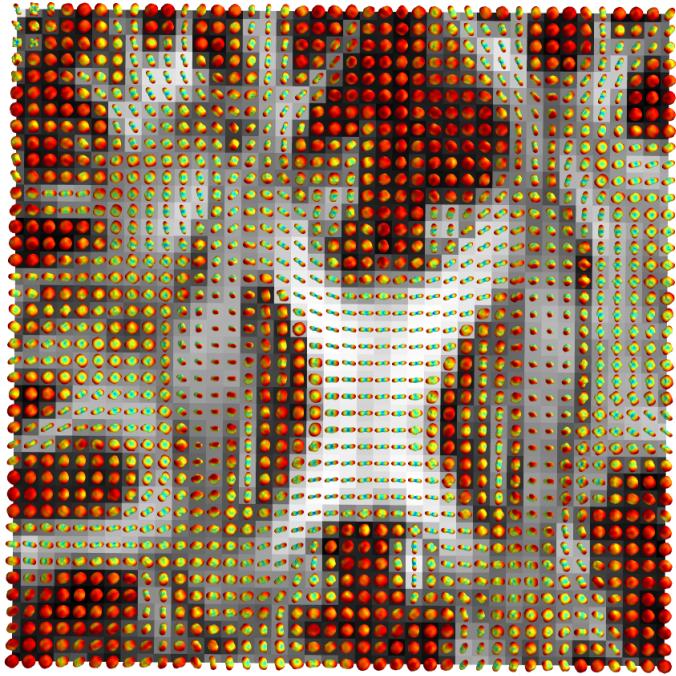


Figure 2.19: EITL ODFs rendered on top of FA of a human brain data set. A small 102-point half grid acquisition with a maximum b-value of $4,000 \text{ s/mm}^2$ was used. Fig. 2.20 and 2.21 are zoomed versions of the same figure. We can see clearly single fibres on the CC and CST areas but also crossing fibres at the Centrum Semiovale and at the areas where big bundles cross. Also the non-white matter areas are evidently more isotropic.

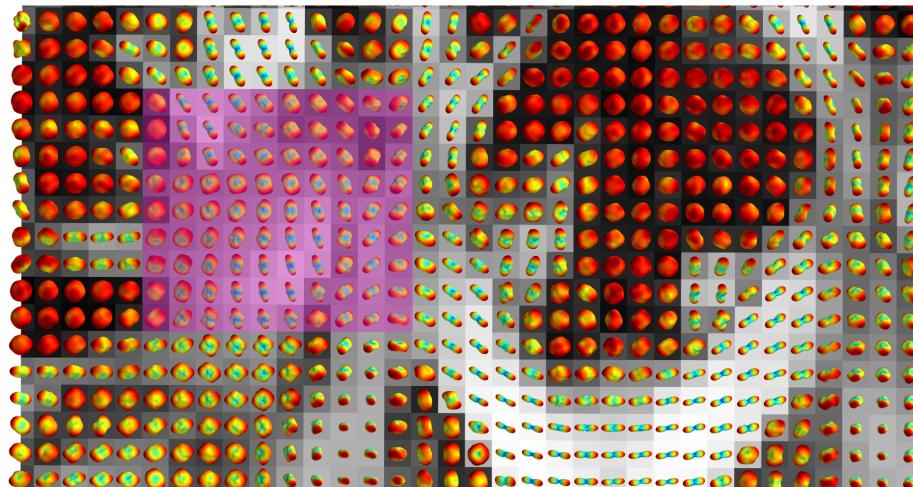


Figure 2.20: The upper part of Fig. 2.19 is shown here at higher resolution. The purple shaded part is given in higher resolution in Fig. 2.21

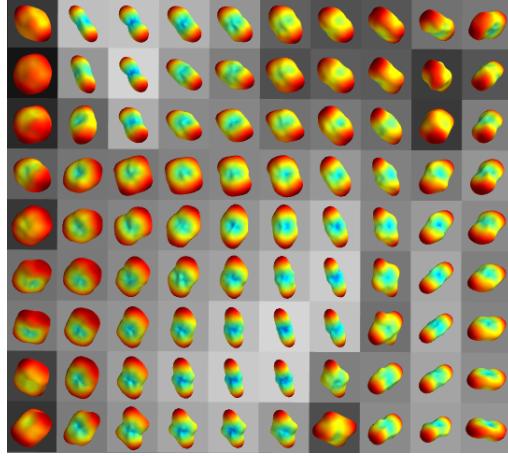


Figure 2.21: EITL ODFs of 1-fibre, 2-fibre and 3-fibre crossings from a real human data set of 101 applied weighted diffusion volume and 1 without weighting (b_0). This picture is a zoomed version of the purple shaded area shown in Fig. 2.20.

2.10 Anisotropy metrics

Until this moment we discussed about density functions on the sphere as a way to represent complex fibre directionality in the voxel. These density functions are represented as multidimensional vectors containing 200 or more dimensions in each voxel and it can be cumbersome to use them directly for subject comparisons or visualization purposes. For this purpose most people use simple scalar summarizing metrics e.g. Tensor-based FA, MD or ODF-based such as the Generalized FA (GFA) [68]. In this section we will show that a similar scalar function like FA can be constructed non-parametrically. We call this NPA which stands for non-parametric anisotropy. We will also start experimenting with metrics that have more than one scalar value and can represent more accurately the directionality in each voxel that is lost with FA, GFA and MD. We will investigate and explain here the realms and robustness of Quantitative Anisotropy which was first introduced by Yeh et al. [72].

2.10.1 Non-parametric Anisotropy

Local voxelwise measures such as fractional anisotropy (FA), apparent diffusivity coefficient (ADC), or mean diffusivity (MD) have been extensively adopted in clinical and applied research practice based on diffusion weighted MR imaging (dMRI). This underlines the need for valid and reliable measures which can indicate the degree of local organisation

of white matter in the brain. The measures listed above are based on the parametric simple diffusion tensor (SDT or DTI) model [67] which works well when there is a single dominant fibre direction. When the local organization is more complex however, the information it provides is not so valid [72, 99]. We show how model-free, alternatives can yield non-parametric anisotropy (NPA). These are constructed from the GQI ODF. We apply exact analytical results which show the form of the GQI-ODF when the single tensor model is correct, and further indicate how the tensor's parameters may be estimated from this model-free approach. We compare the performance of these parametric and non-parametric measures for simulated data.

Simulations were computed for a 102-point grid sampling scheme, with a maximum b-value of $4,000 \text{ s/mm}^2$. The simulated fibre was aligned with the gradient frame of reference, and the diagonal elements of the diffusion tensor, D , were chosen to match typical values for white matter: $\lambda_1 = 1.4 \times 10^{-3} \text{ mm}^2/\text{s}$, and $\lambda_2 = \lambda_3 = 0.35 \times 10^{-3} \text{ mm}^2/\text{s}$. Variable fibre orientation was realised by spatially rotating the simulated fibres at discrete orientations. 100 orientations were used, which spanned uniformly the space of (θ, ϕ) .

In addition to the SDT a two compartment model with an isotropic component was added with volume fraction 0.5 and diffusivity $0.7 \times 10^{-3} \text{ mm}^2/\text{s}$. For each acquisition scheme and fibre type, the “ideal” (noise-free) diffusion weighted signals were calculated according to the SDT model, assuming a constant ideal value of the baseline signal $S_0 = 100$. Complex Gaussian noise was then superimposed upon the ideal signals to provide the complex noise-contaminated signals and their magnitude was then obtained. This results in noisy values with a Rician distribution, which can be scaled in order to set the signal to noise ratio to any desired level. In this study the SNRs were 20, 40, 60, 80 and 100. The GQI ODF and SDT were fitted using DIPY (dipy.org).

The GQI ODF was calculated for a tessellated spherical icosahedron with 362 vertices and 720 faces. Two values (1.2 and 3.5) were used for λ , the diffusion sampling length. Non-parametric FA, NPA, was calculated from the ODF by:

1. Locating the vertex V_1 with maximum GQI ODF value \max_1 .
2. With V_1 as pole, locating the vertex V_2 on the corresponding equato-

rial band of width ± 5 degrees with maximum GQI ODF value \max_2 .

3. Locating a vertex V_3 in the equatorial band at approximately 90 degrees away from V_2 , denoting the GQI ODF value of \max_3 at V_3 .
4. With $\text{npd}_1 = \max_1^2$, $\text{npd}_2 = \max_2^2$, and $\text{npd}_3 = \max_3^2$, non-parametric anisotropy (NPA) was calculated by applying the classical FA formula 1.7 to the 3 values (npd_1 , npd_2 , npd_3).

The rationale for the squared ODF values is based on Tuch's formula (Eq. A.8) for the ODF in the SDT case which implies that the ODF in the 3 principal axes directions of the tensor is proportional to the square root of the corresponding eigenvalue of the tensor. We have further derived an exact formula (see section A.5): $\max_j \propto \sqrt{\lambda_j} [\Phi(cL_\Delta / \sqrt{\lambda_j}) - .5]$ where c is a constant that depends on the acquisition parameters, and Φ is the cumulative distribution function of the standard Gaussian distribution.

The average NPA and FA are presented in Fig. 2.22 and 2.23 for 200 simulations for each noise level, and single fibres with or without an isotropic component and with different diffusion sampling length. We can see that NPA gives very similar results with FA and as expected it is modulated by the degree of smoothing controlled by the value of the diffusion sampling length.

We plan to extend this approach with voxels containing multiple peaks where FA would be unable to give an informative result and also extend it to other types of ODFs. In summary, we have shown that an informative new scalar anisotropy function (NPA) can be calculated without fitting just from the GQI ODF which promises to be a model-free proxy for FA. NPA differs from GFA [32] in that it uses just 3 values of the GQI ODF with a geometric relationship instead of the entire ODF.

2.10.2 Quantitative Anisotropy

Quantitative anisotropy (QA) was first used by Yeh et al. [72] as a way of representing the peaks of the ODF with as few values as possible. This works in the following way: a) we create the ODF, b) we find the peaks using Alg. 2, c) then QA_i is equal to the peak i minus the minimum value for the entire ODF. This is illustrated in Fig. 2.24 where we can see a star-shaped ODF with three peaks (symmetric) (PK). This ODF can be represented just with 3 QA values where for example the highest value will be:

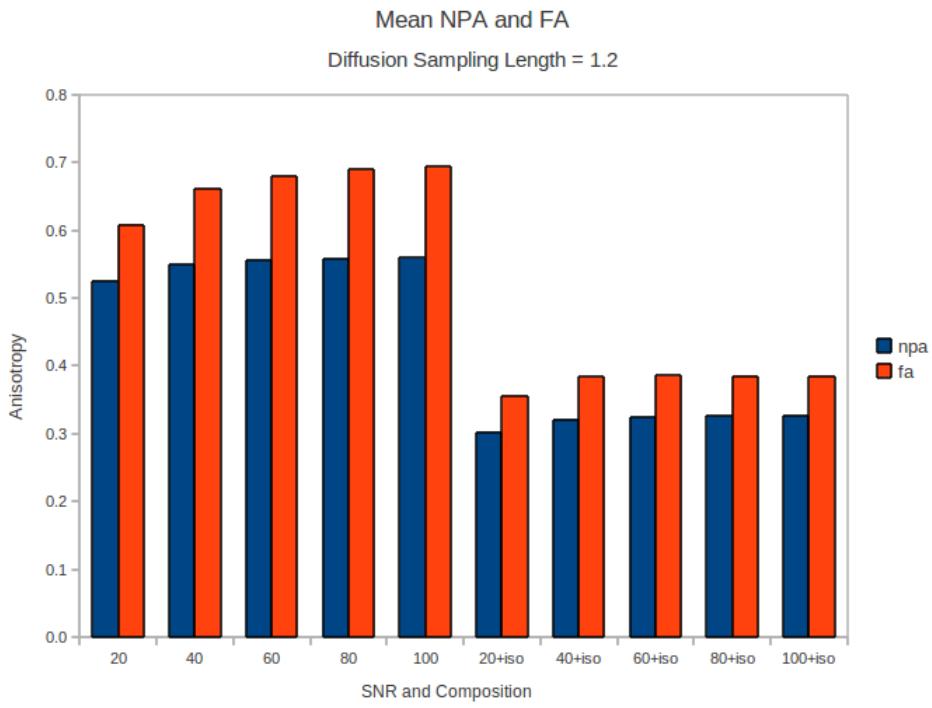


Figure 2.22: Comparison of NPA with FA for single fiber with and without an isotropic compartment at a range of signal to noise ratios.

$QA_0 = \max(\psi_{GQI}) - \min(\psi_{GQI})$, where $\max(\psi_{GQI})$ is the value of the first peak PK_0 , and $QA_2 = PK_2 - \min(\psi_{GQI})$ with $PK_0 \geq PK_1 \geq PK_2$.

QA acts like a differential operator which is higher on anisotropic ODFs and lower on more isotropic. Actually, for a purely isotropic ODF, $QA = 0$. QA can be also easily normalized by the maximum ODF value of all voxels which is usually at the CSF where there is a great amount of water. If this normalization is in effect then we can very easily remove the background noise i.e. non-white matter areas, scalp, skin, muscles etc. just because these will have very low QA values. We can see this interesting property of QA in Fig. 2.25. Of course the most important property of QA is that it can represent crossings and assign a weight for every peak. We will make great use of these weightings in Chapter 3. for the creation of tractographies. Fig. 2.25 was created using DSI Studio ² and the sequence parametrization is the same with the one provided in the experiments of the next section.

²dsi-studio.labsolver.org

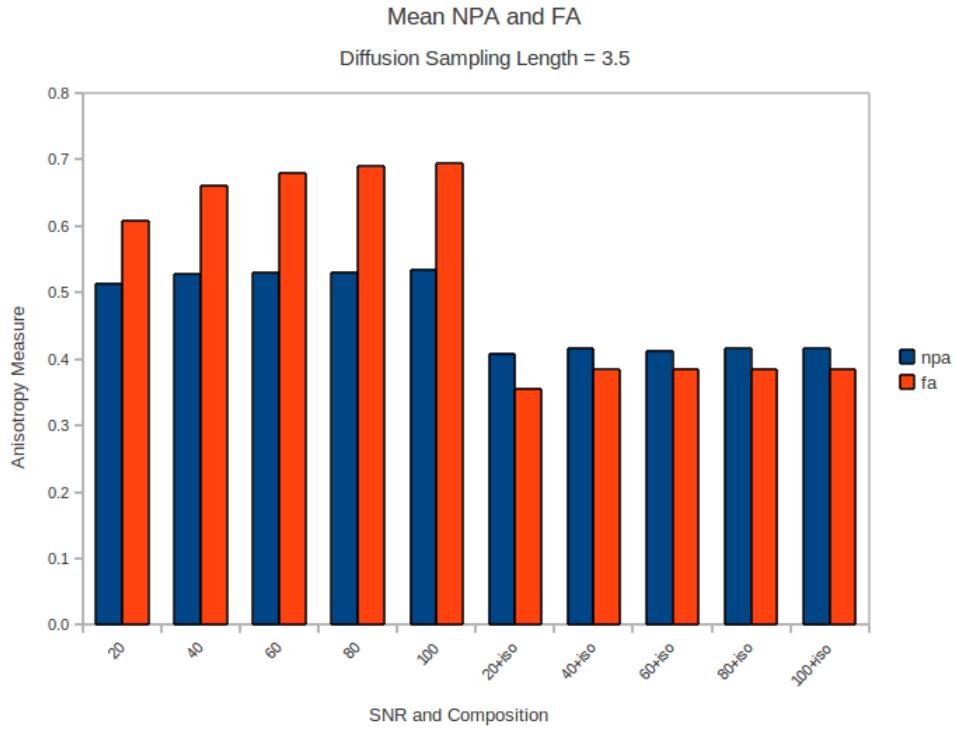


Figure 2.23: As in Fig. 2.22 but with higher diffusion sampling length - less smoothing.

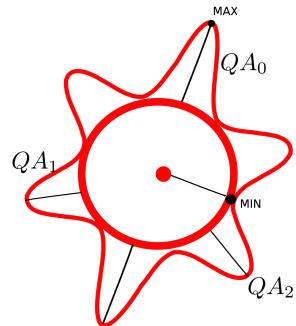


Figure 2.24: QA is calculated from an ODF. The sphere represents the “isotropic” component (minimum value) of a GQI ODF (star) which will be removed from the calculation of QA. QA acts like a differential component with higher values in anisotropic areas and lower in isotropic. The big advantage over FA is that it can represent crossings.

2.10.3 Robustness of QA

GQI was shown to have comparable accuracy to other well established q-space methods when it comes to resolving crossing fibres. In addition, this is achievable with as few as 102 points on a grid sampling scheme,

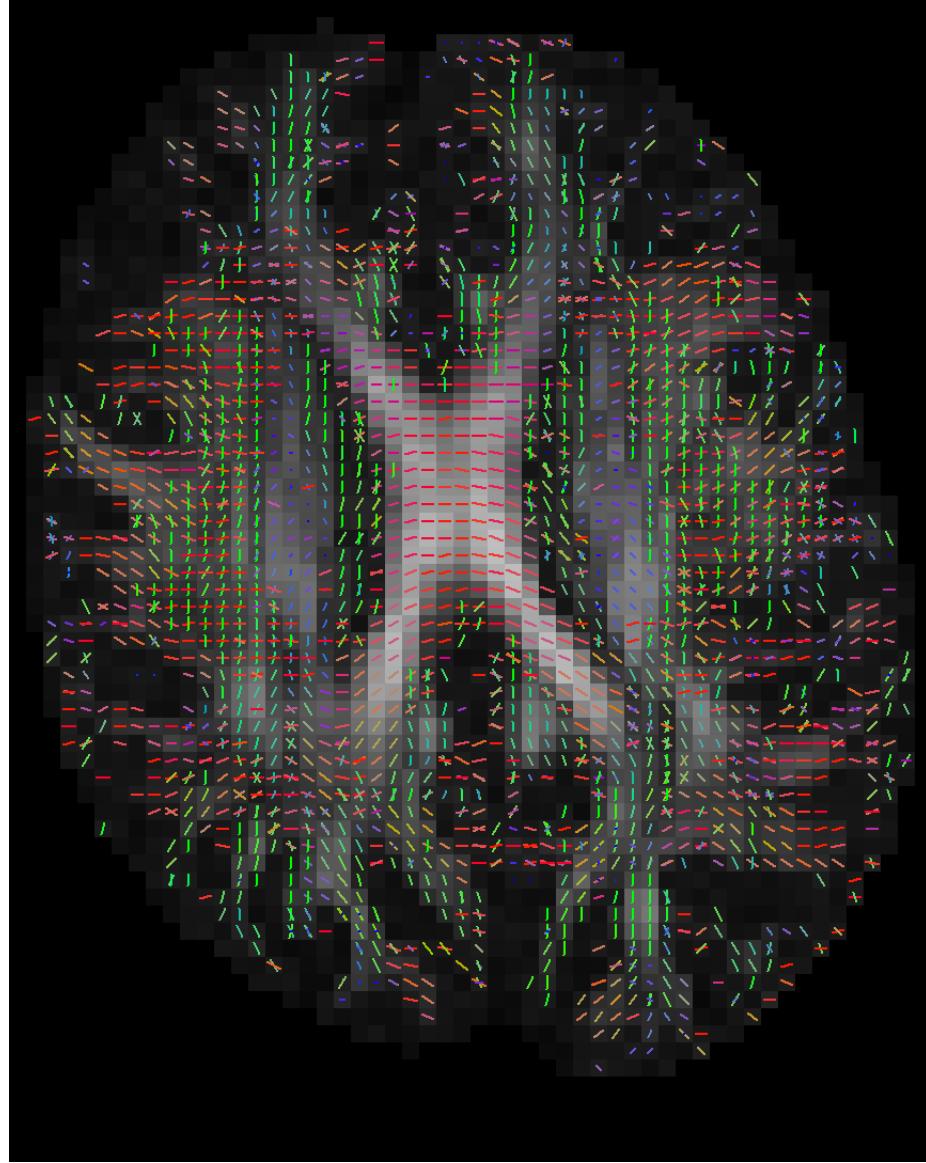


Figure 2.25: Multiple crossings of a real human data set using Quantitative Anisotropy. The first component of QA (QA_0) is also shown in the background.

bringing the total acquisition time down to a clinically acceptable level. Another advantage of GQI is that it is also applicable to a shell sampling scheme. Despite their successes in tractography applications, q-space techniques have until now failed to produce scalar metrics that could replace the ones derived from the diffusion tensor model (e.g. mean diffusivity, MD, and fractional anisotropy, FA) in terms of their multi-subject comparability and specificity to pathology. The data acquired with a grid sampling scheme can still be used to estimate a diffusion tensor and re-

spective scalar parameters, but the effects of the high b-values required for q-space imaging ($> 2,000 \text{ s/mm}^2$) in the accuracy of the resulting DTI parameters has not been well characterized. The authors of GQI have also proposed a new scalar metric called quantitative anisotropy (QA) which was described in the previous sections, but its properties have not been compared to those of FA. In this section we will compare the estimated values of MD, FA and QA_0 (first component of QA) obtained with grid and shell sampling schemes, in terms of their precision and ability to differentiate between different brain fibre populations. Therefore, we will try to test next the hypothesis that QA_0 can be used for subject comparisons.

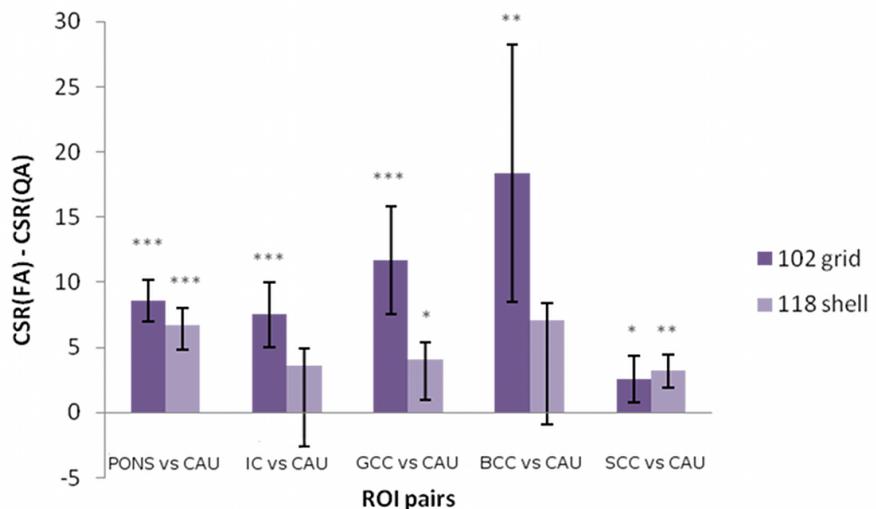


Figure 2.26: Sample results of the paired t-tests comparing CSR (FA) and CSR (QA_0)

Twelve healthy volunteers aged between 18 and 40 were scanned on a 3T scanner (TIM Trio, Siemens), using Siemens advanced diffusion work-in-progress sequence, and STEAM [101, 15] as the diffusion preparation method. The field of view was $240 \times 240 \text{ mm}^2$, matrix size 96×96 , and slice thickness 2.5 mm (no gap). 55 slices were acquired to achieve full brain coverage, and the voxel resolution was $2.5 \times 2.5 \times 2.5 \text{ mm}^3$. Two sampling schemes were considered: a 102-point grid acquisition with a maximum b-value of $4,000 \text{ s/mm}^2$, and a single shell acquisition using 118 non-collinear gradient directions and a b-value of $1,000 \text{ s/mm}^2$ [103]. The two acquisition schemes were matched for total acquisition time in (14 min 37 s), voxel resolution, and bandwidth. FA, MD and QA_0 maps

were then generated for each acquisition scheme and for the 12 volunteers using DIPY [104]. All the FA datasets were non-linearly registered into MNI space using FSL tools, and the same transformation parameters were applied to MD and QA_0 maps. Fourteen ROIs of different brain regions were drawn in MNI space: Putamen (left and right), Caudate (left and right), Thalamus (left and right), Para-sagittal white matter (left and right), Pons, Internal Capsule (left and right), and Genu, Body and Spleenium of the Corpus Callosum. Small cubic ROIs were also constructed by finding the centroid of each anatomical ROI and using it as the centre for a $3 \times 3 \times 3 \text{ mm}^3$ ROI. For each ROI we calculated the mean value for each metric, and the spatial coefficient of variation (CV) within the ROI (see Eq. 2.28).

$$CV_{ROI} = \frac{\sigma_x}{\langle x \rangle} = \frac{N_{voxels} \sqrt{\sum_{x_i \in ROI} (x_i - \langle x \rangle)^2}}{\sqrt{N_{voxels} - 1} \sum_{x_i \in ROI} x_i} \quad (2.28)$$

The coefficient of variation of each ROI mean across subjects was also calculated, as a measure of each metric's comparability between subjects. The contrast-to-scatter ratio (CSR) (calculated for FA in Eq. 2.29) is a good measure of a metric's ability to differentiate between different brain fibre populations [103].

$$CSR(FA) = \frac{mean(FA)_{ROI_1} - mean(FA)_{ROI_2}}{\sqrt{var(FA)_{ROI_1} + var(FA)_{ROI_2}}} \quad (2.29)$$

Combining the left and right versions of each ROI, we have 9 ROIs of different brain populations, which can be used to define 36 pairs of ROIs, and the CSR of all metrics was calculated for each of these pairs. Paired t-tests were then conducted to compare the performance of each metric with the two acquisition schemes, and also to compare FA and QA_0 directly for each acquisition scheme.

The 102 grid sampling scheme produces significantly higher mean FA and QA_0 values than the ones obtained with the 118 shell scheme, while the opposite was observed for MD. The CSR results for FA and QA_0 were not significantly different between the two acquisition schemes, but the 102 grid scheme produces significantly higher CSRs for MD for 26/36 ROI pairs (Fig. 2.26). For MD, no significant difference was found for the CV across subjects, but for FA and QA_0 the 102 scheme produced results

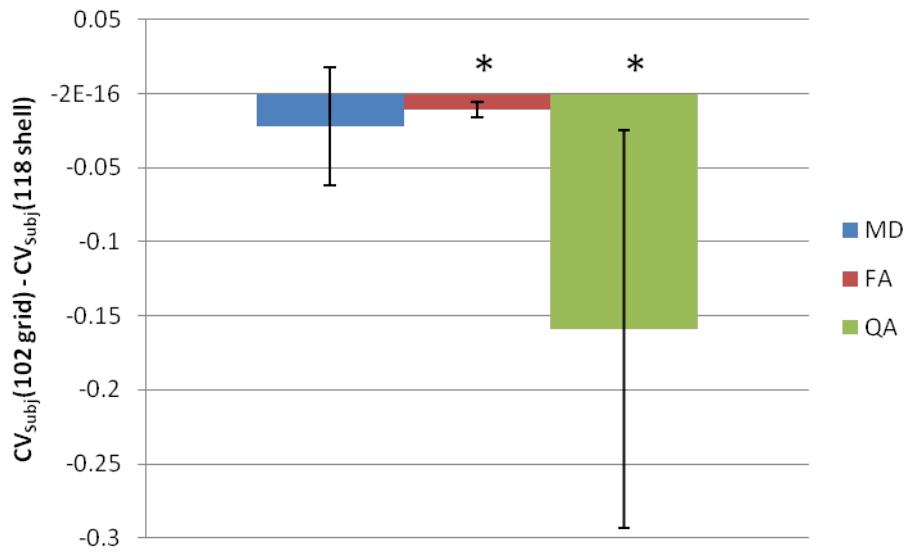


Figure 2.27: Results of the paired t-test comparing the CVs across subjects for MD, FA and QA0.

more comparable across the different volunteers (Fig. 2.27). For FA and MD the 102 scheme showed lower CV within ROIs, especially for white matter, but no difference was found for QA_0 . When comparing FA and QA_0 directly, our results show that FA produces higher CSRs than QA_0 for 23/36 ROI pairs for the 102 grid sampling, and for 19/36 ROI pairs for the 118 scheme. FA also shows lower variation across subjects for both acquisition schemes. Finally, FA shows lower CVs within white matter ROIs, while QA_0 shows less variability for grey matter. The results described and shown above were obtained with the cubic ROIs, but do not differ significantly when the same analysis was applied to larger anatomical ROIs.

Our results indicate that the MD and FA maps generated from a grid sampling scheme designed for GQI are still suitable for analysis, since they do not show poorer performance when compared to a single shell and low b-value acquisition. In fact, the overall results suggest that the 102 grid sampling produces slightly more robust results than the 118 shell acquisition. A previous study [105] has shown that metrics such as MD and FA benefit from the use of multiple b-values, which could explain the better performance of the 102 grid scheme.

2.11 Discussion and Conclusion

Non-parametric methods have the advantage of representing the signal with minimum number of assumptions and without needing any fitting. For many years there has been a trend in science to prefer model-based rather than model free (non-parametric) methods. This is perhaps because model-based can be easier to describe, and allow the use of popular Bayesian approaches more readily. However, there are some crucial issues with fitting: (a) Usually the interesting models have many parameters and that makes fitting very slow. (b) Commonly non-linear fitting is needed and accurate fitting is not trivial. (c) Often the model does not represent precisely the complexity of the real problem. (d) The more complex the model, the more difficult to fit [106], [107], [108].

Non-parametric methods avoid fitting model parameters and that gives them a big advantage. The focus of this chapter was on introducing and developing new non-parametric methods (EIT) or comparing and extending existing ones (GQI2). We showed that a simple, fast and comprehensive transform exists that we call the Equatorial Inversion Transform (EIT). With this transform we showed that we can represent accurately the directional information of the diffusion signal. Furthermore, we showed that there are many different functions (F and O see Eq. 2.19, 2.20, 2.21) which can be used in order to create spherical density functions and use these to find the primary fibre directions. With a correct choose of F and O we can create theoretically the same ODF as the real ODF (DSI ODF). This can be done using EITL which is a type of EIT. Nonetheless, other density functions can be created that can identify the leading fibre directions without being real ODFs but they are still different types of spherical densities. EITL2 and EITS are examples of this last case.

The EIT concept opens new doors for the investigation of dMRI where many new functionals can be invented in the future that emphasise different properties of the signal. We have already illustrated and measured that EIT has the best performance with simulations against the state-of-the-art methods like DSI and GQI and that empirically, EIT gives as good results with real data sets.

The EIT finds the ODF directly without creating the diffusion propagator. If for some purpose the diffusion propagator is still required, then DSI or DPI [78] are favourable. It could be interesting in the future to try and

recover the propagator using ideas from the EIT. However, nearly always the propagator is not needed for the analysis. Furthermore, comparing 4D densities like the propagator is a non-trivial problem, also storing the propagator for every voxel is very inefficient.

We discussed that GQI can be used for creating Quantitative Anisotropy (QA). We observed that QA acts like a differential operator. It bears some similarities with FA as it is maximum on anisotropic and 0 on isotropic voxels. QA assumes that a substantial isotropic part can always be removed from the ODF and that makes it more favourable for spherical functions like those of GQI and EITS. This is in contrast to sharper densities like those of DSI and GQI2 where QA is not as useful because the minimum value of these densities will be usually near 0.

GQI needs a manually set parameter; the diffusion sampling length and in contrast the EIT is fully automatic i.e. we always just used the few default parameters for all experiments. The diffusion sampling length can be slightly different from experiment to experiment. The asset of GQI and GQI2 (which was presented together with GQI but not investigated until today) is that they are fast to compute and have simple analytical solutions. GQI2 seems robust and smooth and it has good performance both with simulations and real data.

It is important to stress that there are similarities between all these methods; DSI is similar to EITL, GQI to EITS, GQI2 to EITL2. In addition, we showed that we can denoise the signal using a Gaussian Spherical Angular method which operates on spherical densities and has a single parameter which is similar to the variance.

Finally, we showed that the first component of QA (highest QA value) can be used for subject comparisons in a similar way to FA. We also showed that NPA could replace FA if we want to calculate anisotropy in a completely geometric way.

The source code for all the methods analyzed in this chapter is available at dipy.org under module `dipy.reconst`.

3 Tracking with Euler Delta Crossings

3.1 Overview

Tractography methods provide tools to resolve major neuronal fibre bundles non-invasively and in-vivo [109]. Since the development of the first tractography algorithms [4, 34] nearly 12 years ago a great number of methods have been published. Often these algorithms depend strictly on the underlying voxel model or acquisition paradigm making it difficult for other researchers to apply their own reconstruction methods and evaluate their data sets.

In this work we designed a purely deterministic method which is fast, accurate and all-inclusive. Most importantly it can have as input model-based or model-free reconstruction algorithms of most known algorithms. We call this algorithm EuDX. Eu stands for Euler integration, D stands for Delta function which is a function that checks for many different stopping criteria and X stands for fibre crossings. EuDX can deal with any number of crossing fibres as long as the reconstruction algorithm supports them. The purpose of this algorithm is to be faithful to the reconstruction results rather than try to correct or enhance them by introducing regional or global considerations which is the topic of other methods reviewed below. Therefore, EuDX serves mainly as a robust method for quickly inspecting different reconstruction results using streamlines. EuDX is noise-friendly i.e. if a voxel is too noisy then EuDX will stop tracking on that voxel. This property is often useful when validating underlying reconstruction models. Branching is also supported by a combination of trilinear interpolation and propagation along multiple peaks per voxel. This method is an extension of the method used by Conturo et al. [34] and Yeh et al. [72] with the additional support for propagation along multiple fibre directions.

In sections 1.8, 1.9 we discussed some of the ideas and the problems behind the most popular propagation methods; deterministic and probabilistic. The focus of this section is to give a more general overview and introduce many more methods.

Most tractography techniques, as pointed out in Sotropoulos thesis [37], can be grouped in three categories: a) local, b) global and c) simulated. Lo-

cal approaches propagate a curve from a starting (seed) point using locally greedy criteria, i.e. tracking sequentially through orientation estimates in adjacent voxels. Global approaches identify the best path between two points of interest, according to some optimization criterion, rather than identifying paths arising from a single point. Simulated approaches comprise of algorithms that simulate the diffusion process or solve the diffusion equation to reconstruct white matter tracks. A detailed literature review is given below.

3.1.1 Local

Deterministic tractography was the first to appear. Tracks (also known as streamlines) are created as trajectories in the form of polylines; orthograde and retrograde along an initial direction at a specific point (seed) in the 3D volume. In FACT [4] tracks are propagated in unequal steps governed by the entry point of the streamline in the voxel (see Fig. 1.11). Euler integration with equal steps was used in Conturo et al. [34] and similarly Runge-Kutta integration was used in Basser et al. [38]. Deterministic approaches usually stop propagating when a low anisotropy region (usually $FA < 0.2$) is found. This is useful in order to avoid propagation within the CSF where anatomical tracts do not exist or within deep gray matter regions where tracking is uncertain. They usually also check for large angular changes (e.g. larger than 90°) between successive steps to avoid unrealistically sharp turns.

Deterministic methods can also be utilized when multiple orientations are estimated in a single voxel (crossing fibres). These orientations can for example be obtained as the principal eigenvectors of multiple Tensors fitted to the data [99], or from the local peaks of the diffusion ODF estimated using DSI [33] and QBI [32] or from the orientations from the fibre ODFs [83]. There are different approaches for propagating across voxels where more than one fibre orientation has been identified. One approach is, upon entering a voxel, to choose the orientation that produces the smallest curvature with the incoming path used in Wedeen et al. [3]. Another approach follows all orientations that do not exceed a curvature threshold, by initiating a new streamline per orientation using in Chao et al. [110] and Descoteaux et al. [111].

An interesting point is that most methods of this category utilize only

the fibre orientation estimates. Tensor deflection tractography (TEND) proposed by Lazar et al. [112] is a FACT variant that uses the whole DTI Tensor rather than just its principal eigenvector to determine the direction of curve propagation.

All the methods described up to this point provide binary connectivity information i.e. a voxel B can be either connected or not connected to the seed S , depending on whether a streamline from S passes through B .

Probabilistic tractography was introduced by Parker et al. [113] and Behrens et al. [46]. Here the standard procedure is to calculate a spatial distribution of tracks arising from a single seed rather than a single track. In each propagation step of each streamline, a random perturbation of the underlying fibre orientation estimate is followed. Perturbations are generated using functions that characterize the uncertainty in the fibre orientation within each voxel. A probabilistic index of connectivity (PICo) is defined between a seed and an arbitrary point as M/N ; where N is the number of all the tracks that start from the seed and M is the number of tracks that traverse the seed and the arbitrary point.

Probabilistic approaches mainly differ in the way that the orientation uncertainty is assessed. Most commonly a Bayesian framework will be used to calculate the posterior probability of the reconstruction model's orientation parameters [46], [35], [114], [115] and [116]. In Behrens et al. [46], [35] Monte Carlo-Markov chain (MCMC) was used to sample the orientation posterior distribution. In Friman et al. [115] the posterior was computed numerically after using Dirac priors. In Zhang et al. [116] particle filtering was used for the same purpose.

Bootstrap tractography is another method which characterizes the uncertainty of the fibre orientation. Pajevic et al. [117] and Lazar et al. [118] are two of the first to apply this method in dMRI. This is a non-parametric approach where a diffusion acquisition is repeated many times creating a large set of images for the same subject. Some images from this set are drawn in random with replacement. This process gives a single bootstrap sample. Drawing many samples will give a distribution for the fibre orientation. The advantage of bootstrap tractography is that no ad-hoc assumptions are made on the noise and it is sensitive to all sources of variability that affect the acquired data set. The disadvantage is that many repeated acquisitions are required; at least 5 for DTI according to O'Gorman et al. [119].

Model-based residual bootstrap offers an alternative, since it requires only a single data acquisition (Chung et al. [120], Berman et al. [121], Harroon et al. [122], Jones et al. [123]). A single bootstrap sample can then be generated by permuting freely the residuals (or just the signs of the residuals using wild bootstrap Jones et al. [123], Whitcher et al. [124]) between all model predicted values. The bootstrap technique was first introduced by Efron [125] in 1979.

Other probabilistic approaches estimate the orientation uncertainty as an empirically defined function. For example, in Parker et al. [113] this is determined using the value of FA. The higher the FA the higher the confidence on the principal eigenvector of the Tensor. In Parker and Alexander [126], Monte-Carlo simulations are used to predict the orientation uncertainty for multiple Tensors and later for PAS [127]. In Descoteaux et al. [128], the fODF was used for the same purpose. Cook et al. [129] used a Watson distribution and Seunarine et al. [130] used a Bingham distribution. The work of Bjornemo et al. [42] can also be classified in the same category who created a regularized stochastic method for probabilistic tractography. This method utilizes the principles of a statistical Monte Carlo method called Sequential Importance Sampling and Resampling (SISR). This technique is similar with particle filters. The disadvantage of the method is that it has strong assumptions for the Single Tensor as the reconstruction model. However, this is often the case with most tracking algorithms.

3.1.2 Global

A limitation of probabilistic tractography is that the probabilistic index of connectivity decreases with distance from the seed point (see section 1.8, 1.9). Another limitation is that it is still sensitive to local noise. Global approaches try to overcome these limitations by being distance-independent and by increasing resistance against noise in a global fashion. These are achieved by finding an optimal path between two voxels, according to a global property [1].

Jbabdi et al. [131] developed a **Global Bayesian** model to derive the posterior probability of connections. The path trajectories represented by splines are compatible with the local fibre orientations in regions with low uncertainty estimates. In regions with high uncertainty, the global connec-

tivity information constrains the local parameter estimation and affects the path sampling.

Front evolution techniques often employ fast marching techniques. The front expands from the seed neighbours to the next neighbouring nodes with speeds determined by the local fibre orientations. As the front propagates, a time of front arrival can be associated with each visited voxel. Once all image voxels have been traversed by the front, paths of connection can be obtained going backwards in the map of front arrival times. Starting from an arbitrary voxel, a gradient descent algorithm can find the fastest route back to the seed. A connectivity index can be associated with each path, representing either the weakest link along the path or the agreement between the path tangents and the underlying vector orientation field (see Parker et al. [132], Tournier et al. [133], Cambell et al. [134], Fletcher et al. [135] and Gigandet et al. [136]).

Graph-based tractography utilises weighted networks (graphs). This type of tractography was presented by Iturria-Medina et al. [137], Zalesky et al. [138], Lifshits et al. [139], Fillard et al. [140] and Sotropoulos et al. [141]. The common concept of these approaches is that each image voxel becomes a node in the graph where the edges of the graph connect pairs of neighbouring voxels or ROIs. The edges are assigned weights that can be representative of any type of structural information. Anatomical paths are then defined as chains with successive elements being neighbouring voxels. The weights of the edges are used to determine the path strength. The strongest path between any image voxel and a seed can then be identified using algorithms that search efficiently the image graph.

Energy Minimization methods [142] try to optimize all tracks from the whole brain volume simultaneously. Each tract is represented as a chain of cylinders, whose position and orientation can change. The method tries to find the set of cylinders that best approximate the underlying white matter bundles. This is achieved by minimizing the overall energy of all cylinders simultaneously, mimicking natural phenomena e.g. the polymerization process which is a process of interacting monomer molecules together in a chemical reaction to form three-dimensional networks. Many standard algorithms e.g. gradient descent are usually employed with this framework; however Gibbs sampling is the most common in tractography. Kreher et al. [36], Reisert et al. [143], Lazar et al. [144] and Fillard et al. [145] showed results using energy minimization. Despite the very promising

results shown by Reisert et al. [143], whose team won the Fibre Cup competition [146] the very high computation time was an important drawback of this framework.

Microstructure Tracking is an exciting new family of algorithms that combine global tractography and direct microstructure estimation using diffusion weighted imaging data. Connectivity via tractography, axon diameter distribution and density estimates are all combined in order to inform one another given the common assumption that microstructural features remain consistent along fibers. MicroTrack [147] is a recent example of this category. Algorithms of this type require their own acquisition schemes, similar to those employed in ActiveAx, developed by Alexander et al. [94].

3.1.3 Simulated

This family of methods take a very different approach from what we have discussed up until this point. They simulate the diffusion of water molecules within the brain tissue or directly solve Fick's second law of diffusion in the entire brain. In this category belong the work of Batchelor et al. [148], Kang et al. [149], Hageman et al. [150] and Hagmann et al. [151].

In Batchelor et al. [148], the diffusion equation is solved using a finite elements approach. Successive diffusion simulations over the entire brain, starting from a seed, are performed in Kang et al. [149]. Tractography by simulating fluid flow through a "pressure" Tensor field is performed in Hageman et al. [150]. The Navier-Stokes equation is solved using a finite elements approach. However, solving a partial differential equation increases execution time. Furthermore, it is not always easy with these approaches to obtain a connectivity map across the whole brain volume and there is usually a large number of parameters to set.

For further understanding of the current map of all the different tractography algorithms see the reviews by Fillard et al. [146], Sotiroopoulos et al. [37] and Jbabdi et al. [152]. In Fillard et al. [146] 10 tractographies were simulated using a novel hardware phantom. Comparing all these different methods which all have different parameters and are based on different underlying models is a difficult process and a ground truth from anatomical studies, digital or hardware phantoms is highly recommended.

3.2 The EuDX Algorithm

We created an algorithm that has many similarities with the classical deterministic methods [4, 34, 38] and with more recent ones as those described in Descoteaux et al. [111] and briefly in Yeh et al. [72]. Our concentration was to create a more general tractography algorithm which can be used with very different families of anisotropic functions and work well with multiple crossing fibres. This algorithm which we call EuDX is applied usually in native space image coordinates and it assumes that the voxel dimensions are equal in all three dimensions i.e. it assumes equal voxel size in all three dimensions. If the provided data do not have isotropic voxel size then a reslicing preprocessing step to isotropic is required.

In order to create tracks we need to provide initially one or more seed points S . These can be chosen randomly or we can specify them explicitly. However, these seed points need to be constrained by the volume's dimensions. Every seed point \mathbf{p}_0 becomes the starting point for the track propagation. For the integration we solve for $\mathbf{p}_t = \mathbf{p}_0 + \int_0^t \mathbf{v}(\mathbf{p}(\mathbf{s}))d\mathbf{s}$ and we perform the integration numerically using Euler's method

$$\mathbf{p}_{n+1} = \mathbf{p}_n + \mathbf{v}(\mathbf{p}_n)\Delta s \quad (3.1)$$

where Δs is the propagation step size which should be at least smaller than the voxel size and \mathbf{v} is the propagation direction. Alternatively, Runge-Kutta of 2nd and 4th order could be used. However, in this document we only experimented with Euler's method.

For EuDX's stopping criteria we can use a standard scalar function like FA but we can also use vector functions like Quantitative Anisotropy (QA) [72] or even the full Orientation Density Function (ODF) [96]. The only constraint for these functions is to be greater or equal to zero everywhere in the volume. In most cases, all these functions try to measure in some way the anisotropy of diffusion in every single voxel so we decided to use the letter \mathcal{A} for the purpose of representing all these different functions applied on the image grid. Therefore, when we write $\mathcal{A}(\mathbf{u}_i) = \alpha_i$ this reads for the peak unit direction of \mathbf{u}_i the peak value was α_i . For the simple case of FA, \mathbf{u} is equal with the eigenvector corresponding to the highest eigenvalue, $\mathcal{A}(\mathbf{e}) = FA$. For QA which can allow for any number of peaks, where usually we constrain it to maximum of 3, we use $\mathcal{A}(\mathbf{u}_i) = QA_i$

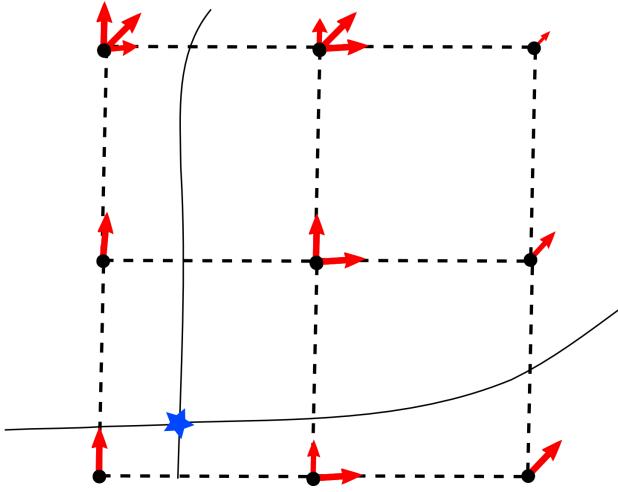


Figure 3.1: In every voxel centre (black dot) there are one or more vectors. These vectors represent peaks where their length is equal to their anisotropy value and the direction is equal to the direction of the peak e.g. calculated from a given ODF. EuDX can track multiple peaks starting from a single seed point (star) if their anisotropy values are higher than a threshold. In that way we can track from the same seed towards different directions and support tracking in crossing areas as it is shown here.

where $i \in [1, 3]$ denotes the current peak. The peak can be characterised by two things: (a) the anisotropy value α_i and (b) the unit direction of the peak \mathbf{u}_i . The concept of tracking with the combination of multiple peaks is presented in Fig. 3.1. In order to reduce storing space, the vector \mathbf{u} can be replaced by an index to the closest vertex of an evenly distributed and dense unit sphere. For generality and simplicity we require this indexing process even for the case of the Single Tensor where the peak direction is only one. Alternatively, we can see this process as a strategy which always maps any representation of the voxel on the sphere.

The EuDX algorithm can be described further in the following way: $\mathcal{A}(\mathbf{u}_i)$ is estimated at every point of the volume. This represents a composite vector field where every point contains the peak directions superimposed to the anisotropy values. We create an empty list of tracks $T = \emptyset$ and then we generate random or prespecified seed points. In more detail: we select a seed point \mathbf{p}_0 and start propagating. We need to remember that the propagation direction can go forward and backward, or better said towards one direction (orthograde) and its opposite direction (retrograde).

For the moment we only propagate towards one direction set by

$$\mathbf{v}(\mathbf{p}_0) = \arg \max_{\mathbf{u}} \mathcal{A}(\mathbf{u}) \quad (3.2)$$

but we need to remember to propagate also towards the opposite direction $-\mathbf{v}(\mathbf{p}_0)$. As \mathcal{A} can have multiple values in each voxel (representing different peaks) we need to remember that when we finish with this track followed by direction \mathbf{v} we will need to propagate towards the direction of second peak, third peak etc. This is necessary if and only if \mathcal{A} gives information for multiple peaks as it is common with QA (see section 2.10.2).

Apart from the direction, we also need to check when to stop tracking. EuDX takes as input a threshold for anisotropy \mathcal{A}_{thr} . If $\mathcal{A}(\mathbf{p}_n) < \mathcal{A}_{thr}$ then EuDX stops propagating. Otherwise it appends the point to the current track. This can be useful for canceling out any seed points which are in the background or in very low anisotropy areas where tracking is not recommended e.g. in the CSF. One important point here is that \mathcal{A}_{thr} depends on the reconstruction method and it will have a different value for every different metric QA, FA, GFA etc. Therefore, we expect EuDX to give different results with different \mathcal{A} functions.

In order to generate a smooth tractography it is recommended to use some kind of interpolation; this is in contrast with FACT which does not use neighbouring information. Here we have been using trilinear interpolation which works in the following way: the seed devides the neighbouring area (constrained by the centers of the neighbouring voxels) in 8 regions in 3D (4 in 2D) and the total contribution of the neighbouring points is added according to the weights w of the antipodal side. The weights express subvolumes in 3D (subareas in 2D). The trilinear interpolation give us the weights w that assign the contribution of the directions of the peaks of the neighbouring voxels to the seed's next direction. It is important to clarify at this point that trilinear interpolation is used only for interpolating the peak directions of the propagation and not the values of the peaks.

We describe here how we use the trilinear interpolation weights in order to find the next propagation direction. (a) We find the nearest direction from the seed's initial direction $\mathbf{v}(\mathbf{p}_n)$ to every peak direction \mathbf{u}_i of every one of the 8 corners of the neighbourhood of the seed. (b) If $\arccos(\mathbf{x}_i, \mathbf{p}_n) \leq \theta_{thr}$ we count the corresponding weight; otherwise we

continue to the next weight. We simultaneously check for the condition $\mathcal{A}(\mathbf{u}) < \mathcal{A}_{thr}$. (c) All the adjacent weights will contribute to the new direction according to the following formula

$$\mathbf{v}'(\mathbf{p}_n) = \sum_m w_m \mathbf{v}(\mathbf{p}_m) \quad (3.3)$$

where $\mathbf{v}(\mathbf{p}_m) = \frac{\mathbf{v}'(\mathbf{p}_m)}{\|\mathbf{v}'(\mathbf{p}_m)\|}$ is normalized. d) The next point is calculated with Eq. 3.1. e) We insert the new point in the track and continue tracking until one of the stopping criteria is met. The next step will be to repeat the a-e steps for the opposite direction of the initial peak direction $\mathbf{v}(\mathbf{p}_o)$ and for the smaller peaks as described above. Finally, we will have to repeat the procedure for the next seed point until all seed points are visited. When all seeds have been visited we will have in our hands the entire tractography T . A formal description of EuDX is given in Alg. 3 and Alg. 4.

Apart from the anisotropic threshold \mathcal{A}_{thr} and angular threshold θ_{thr} other anisotropic criteria are also incorporated in EuDX. These are: a) The total sum of weights TW , which checks there is enough overall neighbouring contribution to continue tracking (default value of 0.5). This is very useful in edges or corners where tracking should stop. b) It is possible for a track to get trapped in a loop and start looping for ever. We can check for that using a maximum length threshold or a threshold for maximum number of points describing a track MNP (default maximum value 1000 points). c) Finally, we need to check that we are always inside the image volume. The 3 dimensions of the volume are hold in variable V .

Algorithm 3 EuDX – All tracks

Input $\mathcal{A}, \mathcal{I}, S, U, \Delta s, V$
 $\mathcal{A}_{thr}, \theta_{thr}, TW, MNP$
Output T

```

 $T \leftarrow \emptyset$ 
For seed in  $S$  Do
    For peak in  $(\mathcal{A}, \mathcal{I})$  Do
        track  $\leftarrow$  EUDX_Core(seed,peak)
        append( $T$ ,track)
    EndFor
EndFor

```

Algorithm 4 EuDX_Core

Input seed, peak, ΔS , V **Output** track

```
track  $\leftarrow \emptyset$ 
delta, i_direction  $\leftarrow$  Initial_Direction(seed)
#propagate orthograde
direction  $\leftarrow$  i_direction
While delta is True Do
    delta, n_direction  $\leftarrow$  New_Direction(direction)
    If delta is False Do
        Break
    EndIf
    point  $\leftarrow$  point +  $\Delta s \times n\_direction$ 
    append(track,point)
    direction  $\leftarrow$  n_direction
EndWhile
delta  $\leftarrow$  True
#propagate retrograde
direction  $\leftarrow -i\_direction$ 
While delta is True Do
    #Same as above
```

In Alg. 3 we see how EuDX creates an entire tractography with tracks grown from seeds S . The core algorithm which is the same for every seed and every peak is given in Alg. 4. In Alg. 3 the input parameters are: \mathcal{A} the 4D volume holding the peak values, \mathcal{I} the 4D volume with the indexed directions of each peak in relation to unit sphere U . U is an array of size $N \times 3$ where N is the number of vertices in the sphere. S is an $M \times 3$ array with the precomputed seeds. Δs is the propagation step size. We also input different stopping thresholds: \mathcal{A}_{thr} defines the lowest possible peak value that allows tracking to continue. θ_{thr} is the maximum allowed angle between the current propagation direction and the next direction. TW checks the overall contribution of the neighbourhood for the next propagation direction and MNP checks that a track does not pass from the same point more than a number of times. MNP checks this condition by counting the number of current points appended in a track as the track grows. Usually, only parameters \mathcal{A} , \mathcal{I} , S and V need to be updated with different datasets; the rest of the parameters can have default values which can be used across all experiments.

The core part of EuDX is given in Alg. 4. This is called for every seed and every peak. The algorithm starts by calling the function Initial_Direction which finds the closest peak direction to follow from the nearby voxels. If the values of these peaks are lower than A_{thr} then it returns False otherwise it returns True and the vector with the initial direction. The variable delta is used as check point that no stopping criteria are met. As long as delta is True we can continue tracking using Euler integration (see Eq. 3.1). The function New_Direction calculates the new direction of the propagation by accumulating the weights created by trilinear interpolation as shown in Eq. 3.3 and finds the nearest direction for the eight neighbouring voxels. It also checks for the total weight threshold TW and if we are inside the volume's boundaries given by V . After this is done we update the points as shown in Eq. 3.1 and repeat the same procedure until one of the stopping criteria is met by checking variable delta. EuDX stems its D letter from that delta function. After tracking towards the orthograde direction has stopped and delta is now False, we repeat the same procedure as before for the retrograde direction in order to create the complete track (see Alg. 4).

3.3 Results with software phantoms

In order to validate the performance of EuDX we created a 3D software phantom with the method described in section 2.9.2 using eigenvalues $\lambda_{\parallel} = 1.7 \times 10^{-3} m^2/sec$ and $\lambda_{\perp} = 0.1 \times 10^{-3} m^2/sec$ which are in the range typically found in the human brain [75]. The software phantom consists of two parts: a) a diagonal orbit and b) an elliptical orbit with axes ratio $\lambda_2/\lambda_1 = 0.6$. Both parts were added together to create a crossing configuration and partial volume effects are assumed negligible. The average thickness of both parts was 5 voxels. The elliptical orbit did not complete a full ellipse in order to avoid creating looping tracks which have no anatomical relevance. We created two experiments using the same data set of size $64 \times 64 \times 64 \times 102$ where the 3 first dimensions give the size of the volume and the last dimension is the number of diffusion weighted volumes (101) plus 1 unweighted b0 volume (see Fig. 3.2(i) lower right corner). We generated b-vectors and b-values by using a keyhole Cartesian sampling grid [99] (see Fig. 1.8); therefore, DSI, GQI, EIT or DTI were all suitable for the signal reconstruction into ODFs or Tensors respectively

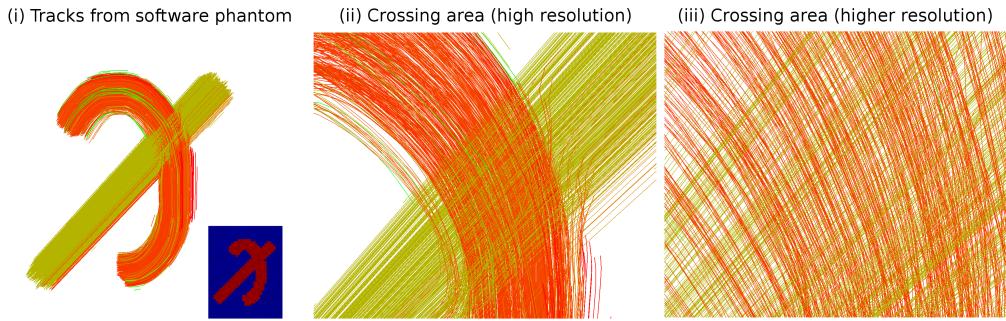


Figure 3.2: (i) Tractography generated by EuDX with DSI as the reconstruction method on a software phantom containing two intersecting bundles. A slice from the b0 volume of the phantom is also shown at the lower right corner. (ii) and (iii) In higher resolution we can see that tracks traveling from the two different bundles cross unimpeded in the intersection area of the two simulated bundles.

for these type of data. In contrast, QBI is not suitable because it assumes a spherical grid in q-space (see section 1.5). For both experiments we used a high SNR of 100 as the main goal was to validate the algorithm on good conditions. We will discuss later the validation process with human data sets where SNR is naturally lower.

In the first experiment (see Fig. 3.2) we generated 200,000 uniformly sampled random seeds in the entire 3D volume of the phantom. We used DSI reconstruction with standard parameters: q-space grid size 16, hanning filter width 32, and radial integration range from 2.1 to 6 at steps of 0.2. As we presented earlier, EuDX expects as input the peaks and the directions of the peaks. For every voxel we used DSI to create the corresponding ODF (sampled on an evenly distributed sphere of 642 vertices and 1,280 faces) and from the ODF we used the peak finding function introduced in section 2.7. We further removed the peaks which had values less than 70% of the highest peak and normalized the rest so that the maximum peak equals to 1. From now on we will call this output function from the ODF; PK (PeAK anisotropy). PK is different from QA because we do not remove the isotropic component neither we normalize with the overall maximum ODF value as it is common practice with QA (see section 2.10.2). We also calculated FA and zeroed the peaks with FA values lower than 0.2. In that way we ensured that there will be no tracking in the background area of the phantom.

For EuDX we used parameters $\mathcal{A}_{thr} = 0.2$, $\Delta s = 0.5$, $\theta_{thr} = 60^\circ$ and

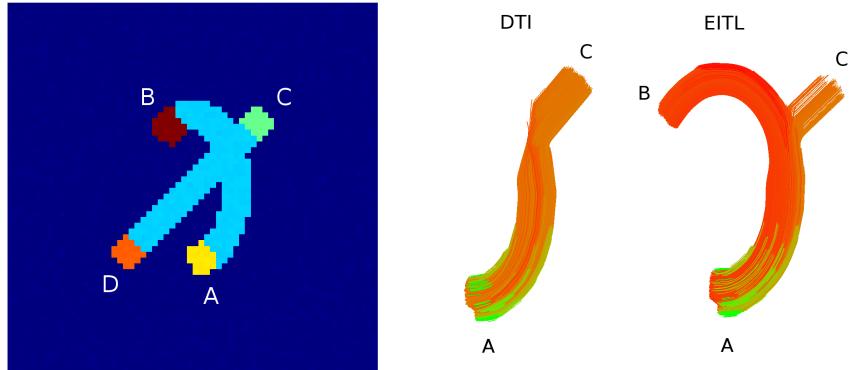


Figure 3.3: On the left panel we see the b_0 volume of the software phantom with 4 ROIs; A, B, C and D presented with different colours. 2,000 seeds were generated inside each ROI and we then measured the amount of tracks which reached the other regions. On the right panel we see that when the DTI (Single Tensor) was used for the reconstruction; the tracking started from A reached only C. However, when EITL was used the tracks reached both areas B and C as it was expected.

$TW = 0.5$. The results are shown in Fig. 3.2. The tracks are colour-coded with their orientation; this is defined as the unit vector connecting the first with the last point of each track. Because we generated random seeds that went everywhere in the phantom some of them had to fall in the crossing area. In accordance with what we discussed in the previous section we know that if a seed falls in a crossing region, EuDX will propagate towards both directions of the crossing if and only if multiple peaks are supported from the underlying reconstruction method. We can confirm this by observing the result at Fig. 3.2. We can see that the crossing area is well represented and that the propagation was successful. We can also observe that not all tracks travel the entire distance from end to end. Otherwise we would expect to see all tracks of the entire elliptic part with one colour and the tracks of the straight orbit all with a different colour because they have different orientation. It seems that this observation had something to do with the discrete nature of the phantom especially near the bundles' boundary areas and of course the functionality of EuDX. For this purpose, we created a different experiment to evaluate this finding.

In the second experiment we generated 2,000 seeds inside specific regions of the bundles (see Fig. 3.3). More precisely, the end-point areas of the phantom denoted with A, B, C and D. The goal was to count the percentage of tracks which reached any of the other end-point areas. We tried this with 4 different reconstruction methods: Single Tensor, EITL,

DSI and GQI. The Single Tensor was fitted using weighted least squares. For EITL we used for radial sampling from 0 to 5 with steps of 0.4 and Gaussian weighting of 0.05. GQI had sampling length of 1.2 and DSI had the same parameters as in the previous experiment. The orientations of the peaks for all the methods were found in a reconstruction sphere of 642 vertices. The same peaking function was also used with the ODFs of EITL, GQI and DSI. For these 3 methods the peaking procedure was the same as described in the previous experiment. The single peak of the DTI reconstruction was found by the eigenvector corresponding to the maximum eigenvalue. EuDX expects the orientation input of the peaks as indices on a unit sphere. Therefore, for the Single Tensor case; for the eigenvector which corresponds to the maximum eigenvalue we found the vertex in the reconstruction sphere with the minimum angular distance. An alternative way would be to calculate the ODF directly from the Tensor in an analytical way. Then we could continue as we do with the other methods i.e. find the peaks in the ODF with our standard peak finding method. We did not follow this approach because theoretically the orientation of the peak of the Tensor ODF is identical with the direction of the eigenvector which corresponds to the maximum eigenvalue and the reconstruction sphere is densely sampled.

The results of this study are summarized in Tab. 3.1. A first general observation is that independently from which end-point ROIs the tracks start there will always be a percentage of tracks which will stop before reaching the other end-points. We can clearly see this phenomenon by observing the last column of each sub-table. The last column that we symbolize with \emptyset holds the percentage of tracks which did not reach any of the other end-points e.g. tracks which started from seeds in region A never reached regions B, C or D. We can observe in these columns that the best case was 3.7% loss from reconstructions using EITL and the worst case was 43.5% loss using DTI.

To simplify comparisons within Tab. 3.1 we can use a maximum standard error of difference (MaxSED) between any two corresponding entries in any pair of rows. $\text{MaxSED} = 100z_{\alpha/2}\sqrt{2p(1-p)/n}$, where $z_{\alpha/2} = \alpha/2$ th percentile of standard normal distribution, $n = 2,000$ is the number of seeds in each row (ROI), and $p = 0.5$ gives the worst case (largest) standard error. p is the probability that tracks from seeds in the row ROI will terminate in the column ROI. If we take $\alpha = 0.0001$, $z_{\alpha/2} = 3.29$,

DTI	A	B	C	D	\emptyset	EITL	A	B	C	D	\emptyset	
	A	—	0%	76.4%	0%	23.6%	A	—	63.4%	8.7%	0%	27.9%
	B	0%	—	0%	59.8%	40.2%	B	65.6%	—	0%	5.6%	28.8%
	C	79.9%	0%	—	0%	20.1%	C	14.5%	0%	—	76.8%	8.7%
	D	0%	56.5%	0%	—	43.5%	D	0%	0.5%	95.8%	—	3.7%
DSI	A	B	C	D	\emptyset	GQI	A	B	C	D	\emptyset	
	A	—	65.3%	9.6%	0.0%	25.1%	A	—	57.8%	8.7%	0.0%	33.5%
	B	72.6%	—	0.0%	5.5%	21.9%	B	67.2%	—	0.0%	10.7%	22.1%
	C	14%	0.0%	—	79.9%	6.1%	C	37.5%	0.0%	—	55.5%	7%
	D	0.0%	10.6%	84.8%	—	4.6%	D	0.0%	22.3%	72.3%	—	5.4%

Table 3.1: Every sub-table shows the percentage of tracks which started from areas A, B, C or D (rows) and reached the other areas A, B, C, D or \emptyset (columns) using EuDX with input from different reconstruction methods DTI, EITL, DSI, GQI. The column \emptyset symbolizes the number of tracks which did not reach any of the A, B, C, D areas. For example, by looking only the first row of each sub-table we easily observe that the crossing area was well represented by EITL, DSI and GQI but not from DTI as it was expected because the Single Tensor cannot resolve crossings. We can also observe by comparing all the rows that in the more curve branch AB of the phantom fewer tracks reached their target than in the diagonal branch CD.

then MaxSED= 5.2%. The same value can be used as a conservative MaxSED in comparisons involving Tab. 3.2 for which the effective n is greater than 2,000. Differences greater than MaxSED are highly significant ($p < 0.0001$).

When DTI is compared with any of EITL, DSI and GQI, each of the connections $A \rightarrow \emptyset$, $B \rightarrow \emptyset$, $C \rightarrow \emptyset$, $D \rightarrow \emptyset$ is significantly more frequent with DTI. Similarly $A \rightarrow B$, and $B \rightarrow A$ are less frequent with DTI, as are $C \rightarrow D$, and $D \rightarrow C$. By contrast $A \rightarrow C$, and $C \rightarrow A$ are more frequent with DTI, as are $B \rightarrow D$, and $D \rightarrow B$.

Between EITL and DSI we note that for tracks starting in B , $B \rightarrow \emptyset$ more frequently with EITL whereas $B \rightarrow A$ less frequently. For tracks starting in D , $D \rightarrow C$ more frequently with EITL than DSI , and $D \rightarrow B$ less frequently with EITL then DSI.

Between EITL and GQI, for tracks starting in A , $A \rightarrow \emptyset$ less often with EITL than with GQI, whereas $A \rightarrow B$ and $A \rightarrow C$ occur more often. However $B \rightarrow \emptyset$ more often for EITL than GQI. By contrast $C \rightarrow D$ occurs more frequently, and $C \rightarrow A$ less frequently. Also $D \rightarrow B$ occurs less.

Finally, comparing DSI with GQI, $A \rightarrow \emptyset$ less often with DSI whereas

$A \rightarrow B$ and $B \rightarrow A$ more often. Similarly $C \rightarrow D$ and $D \rightarrow C$ more often. Correspondingly $D \rightarrow B$ happens more often with GQI.

So, why are there always a number of tracks which will not reach the other end-points? This happens because most of the area in high-dimensional volumes (3D volumes) is concentrated close to the boundaries of the volumes; therefore many random seeds will fall close to the boundaries. In addition, the boundaries are not smooth (see left panel of Fig. 3.3) and for this reason the tracks which will start close to the boundaries will most likely not travel long distances and because the number of seeds which will fall close to the boundaries is not negligible; the number of tracks which will not travel far is not negligible as well. This is something to have in mind when generating tractographies. A naive solution would be to increase the resolution of the volume by interpolation. This approach has also its disadvantages; the diffusion data sets are already very large and increasing the resolution would result in a further increase. Interpolation on the other hand is usually associated with smoothing which will reduce the peak resolution in relation to high b-values where the signal is generally low. For this purpose we decided to not preprocess the phantom any further.

Another general observation is that in the curved branch less tracks reached their target than in the diagonal branch. This is to be expected because the discretization is more asymmetric in the curved branch (see Fig. 3.3-left panel) therefore, more tracks close to the boundaries will stop propagating because they reach background voxels where anisotropy is much lower. This is in agreement with EuDX which is designed to stop in low anisotropic regions as defined by parameter \mathcal{A}_{thr} .

In the right panel of Fig. 3.3 we see that the tracks which started from ROI A, reached only ROI C when we used the peaks from the Single Tensor reconstruction as input to EuDX. However, when we used EITL the tracks which started from A, reached both B and C. The tracks did not reach ROI D because of the angular threshold θ_{thr} . In contrast, tracks which traveled from the right part of the bundle did reach ROI C. This happened because the trilinear interpolation close to the intersection area gave more weight to the directions of the neighbouring voxels and the diverting angle was less than θ_{thr} . This was another confirmation about the behaviour of EuDX in crossing areas. This identified that even if the seeds have not fallen in the intersection region of the two bundles, the tracks

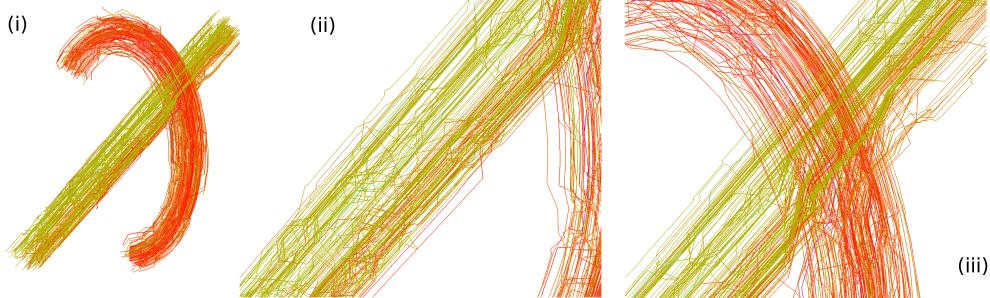


Figure 3.4: Probabilistic tractography performed on the software phantom shown in Fig. 3.2. We use Pico and a Two Tensor model (PMT) to generate the tracks. (i) 1 iteration of the Pico (showing 321 tracks). (ii) A detail of the left bundle. Some of the tracks are diverting on a zig-zag fashion although the phantom at these voxels has only one main direction. (iii) Focus on the crossing region of the phantom. Many of the PMT tracks are successfully propagating towards both pathways.

which will reach the crossing region will follow the branch to which they are closer to in terms of spatial and angular distances. This is possible because of the usage of trilinear interpolation.

Independently from the deterministic tractography described previously we repeated the same experiments with probabilistic tracks generated from the Camino Diffusion MRI Toolkit³. We will call these PMT tracks where PMT stands for Probabilistic MultiTensor Tractography. In order to create the PMT tracks we fitted a Two Tensor model in the data with both Tensors enforced to be prolate. For the fitting method we used the standard non-linear fit provided in Camino with positivity constraints. For this purpose we used the command `modelfit` with options `-model cylcyl nldt_pos`. The probabilistic tractography was generated using Pico PDFs sampled from a Watson distribution (`picopdfs -pdf -watson`). In order to generate the streamlines we used the Camino command `track` which took as input the PDFs from before and a seed file with the mask.

As previously, we performed two experiments; one where the seeds can be anywhere in the simulated bundles and a second where the seeds were constrained in the pre-specified ROIs A, B, C or D (see Fig. 3.3). The results of the first experiment can be seen in Fig. 3.4.

For illustration purposes we see in Fig. 3.4(i) the result with 321 tracks of 1 iteration of the Camino `track` command. In Fig. 3.4(ii) we see a detail of the left bundle. It is obvious that some of the tracks are diverting vigor-

³cmic.cs.ucl.ac.uk/camino

PMT100	A	B	C	D	\emptyset	PMT5000	A	B	C	D	\emptyset
A	–	22%	45.2%	0.1%	32.6%	A	–	22.6%	44.%	0.2%	33.2%
B	40.8%	–	0.6%	11.1%	47.5%	B	40.9%	–	0.7%	9.3%	49.1%
C	28.9%	0.4%	–	51%	19.7%	C	28.2%	0.3%	–	50.2%	21.3%
D	0.1%	3.8%	85.5%	–	10.6%	D	0.3%	4.2%	85.2%	–	10.3%

Table 3.2: In comparison with Tab. 3.1 we observe that PMT performed worse than EuDX with DSI, GQI or EIT. This becomes evident by observing the \emptyset columns of the corresponding tables where EuDX had a lower percentage of tracks which did not reach any of the other end-points. The difference between PMT100 and PMT5000 is in the number of iterations which was 100 and 5,000 respectively. Increasing the number of iterations did not increase considerably the performance of PMT.

ously on a zig-zag fashion although the phantom at these voxels has only one main direction. In Fig. 3.4(iii) we zoom in the crossing region of the phantom where we observe that many of the PMT tracks are successfully propagating towards both pathways through the crossing region.

The results from the second experiment where the seeding takes place only in specific regions of the phantom are summarized in Tab. 3.2. The distinction between PMT100 and PMT5000 is in the number of iterations run from Camino track command. The default value recommended is 5,000 however for this experiment we do not see a significant difference between 100 and 5,000 iterations. In comparison with Tab. 3.1 we observe that PMT performed worse than EuDX with any of DSI, GQI or EIT. This becomes obvious when comparing the \emptyset columns of the corresponding tables where we see that EuDX had a significantly lower percentage of tracks that did not reach any of the other end-points. Further detailed comparison of EITL with PMT shows that the two-way connections $A \rightarrow B$ & $B \rightarrow A$, and $C \rightarrow D$ & $D \rightarrow C$ are stronger with EITL. By contrast $A \rightarrow C$ & $C \rightarrow A$ are stronger with PMT.

This is an important outcome as probabilistic approaches are usually supposed to perform better than deterministic [35], [113]. In this study we have not investigated if PMT performs poorly because of the Two Tensor fit or because of the probabilistic orientation sampling. However, we plan to investigate this further in the future.

3.4 Results with humans

In this section we investigate how EuDX performed with real data sets. There are mainly two ways to validate tractographies with real data sets of healthy humans. The first is to directly ask for feedback from specialist neuroanatomists who have good understanding of the underlying white matter anatomy and the second is to compare against the results of another published tractography method which has already gone through anatomical validation from experts. We investigated both approaches. In order to do that we generated tractographies (see Fig. 3.5B) and asked expert neuroanatomists Professor Luigi Cataneo and Dr. Nivedita Agarwal from the Center for Brain/Mind Science (CiMeC) in Trento, Italy to validate the quality of the data in relationship to their knowledge of anatomy. We then asked them to manually label bundles of their scientific interest in different subjects. We show here two of the many known bundles that they segmented. In Fig. 3.5A we see Arcuate Fasciculus (AC) which is described in the literature as belonging in the language pathways [153], [154], [155] and [156]. In Fig. 3.6A we see the right Corticospinal bundle [157] in another healthy subject.

In order to help the medical practitioners to perform the manual labeling we developed an interactive visualization application (available in [fos.me](#)). This is based on a novel tractography clustering method which is the topic of the next chapter. This application has the capability to create an accessible representation of the initial tractography into bundles of interest (BOIs). In this concept, a bundle is a collection of tracks with similar spatial and shape characteristics. After the BOIs are created, the medical practitioners can select one or more BOIs interactively and hide the parts of the tractography which are not interesting to their investigations. This application is further explained in Chapter 5.3.

The tractographies shown in Fig. 3.5 and at the left panel of Fig. 3.6 were calculated with EuDX parameters $\mathcal{A}_{thr} = 0.02$, $\Delta s = 0.5$, $\theta_{thr} = 60^\circ$, $TW = 0.5$ and with input from QA of GQI reconstruction with sampling length $\lambda = 1.2$. The data sets were generated at a 3T scanner (TIM Trio, Siemens) in Medical Research Council, Cognition and Brain Sciences Unit, Cambridge, UK. We used Siemens advanced diffusion work-in-progress sequence, and STEAM [101, 15] as the diffusion preparation method. The field of view was $240 \times 240 \text{ mm}^2$, matrix size 96×96 , and slice thickness

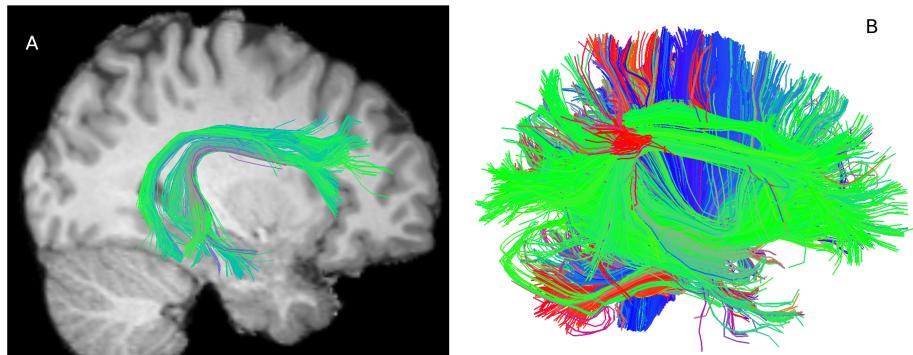


Figure 3.5: A: right arcuate fasciculus generated by EuDX and segmented by expert neuroanatomists. The tracks are in MNI coordinates and we visualize simultaneously the T1 slice ($X=29$) for the same subject. B: the sagittal view of the whole brain tractography of the same subject is shown. For visualization purposes we are depicting only tracks of length from 120 mm to 150 mm.

2.5 mm (no gap). 55 slices were acquired to achieve full brain coverage, and the voxel resolution was $2.5 \times 2.5 \times 2.5 \text{ mm}^3$. A 102-point half grid acquisition with a maximum b-value of $4,000 \text{ s/mm}^2$ was used. The total acquisition time was only 14 min 21 s with TR=8, 200 ms and TE=69 ms.

The tractographies were generated in diffusion native space and then linearly registered in MNI space. For that reason FA volumes were generated from the same data sets using Tensor fitting with weighted least squares after skull stripping with FSL bet. These FA volumes were again in diffusion native space, therefore we used FSL flirt to align them in MNI space. For this purpose a standard FA atlas FMRIB58 from the FSL toolbox was used as the reference image. We then applied the affine transformation matrix from the previous step to the initial tractography to align it to MNI space.

In order to help the neuroanatomists to guide themselves with the segmentation of bundles, we linearly registered structural MRIs (T1 MPRAGE) images from the same subjects to the standard template MNI152. Therefore, the tractography and the T1 image were registered in the same space. This was very useful for the neuroanatomists because they could find known bundles from the regions they connect in the cortex which are better visible in the T1 image rather than in the FA image.

In section 3.3 we discussed with simulation experiments that EuDX was able to propagate correctly in crossing areas (see Fig. 3.2) by gener-

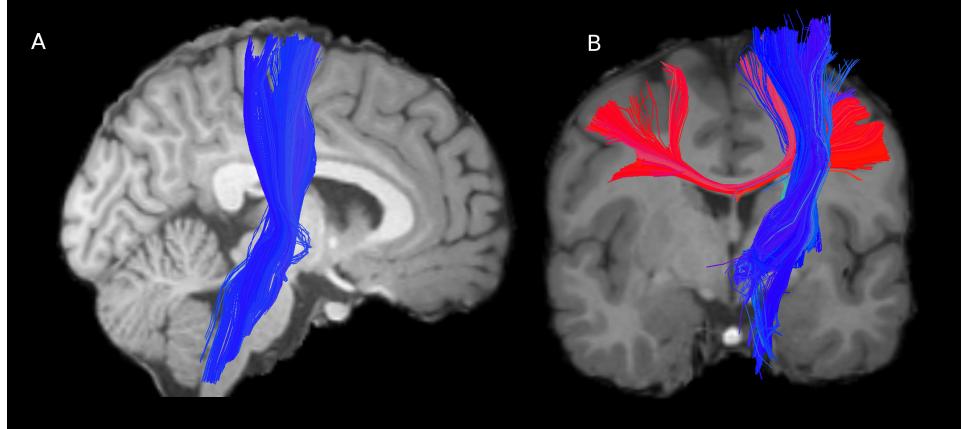


Figure 3.6: A: EuDX tracks from the CST segmented by our expert neuroanatomists. The tracks are linearly registered in MNI standard space and visible is also the T1 slice ($Y=5$) from the same subject. B: The intersection of the BCC with the right CST. This is a confirmation that EuDX can propagate successfully in crossing areas of human brain data. The T1 slice ($Y=-1$) from the same subject is also visible.

ating tracks towards both directions of the crossings. We wanted to confirm if tracking in crossing areas was also robust with human data sets where the noise artefacts are less predictable. This was indeed confirmed by looking at the intersection of two well known bundles: the Body of Corpus Callosum (BCC) and the Corticospinal Tract (CST). In Fig. 3.6B we can see that a part from the BCC shown with red is intersecting the CST bundle shown with blue without being diverted from the CST. If EuDX was not able to propagate in crossing areas then the tracks from BCC would stop in the intersection area or divert towards the direction of the CST.

In the next experiment we visually compared specific bundles across different healthy subjects. We concentrated at a pair of bundles not so often studied in literature: the Cingulum bundles; across 12 healthy subjects (20 – 40 years old). The Cingulum (CG) is an association fibre tract that runs within the Cingulate Gyrus along its entire length. It collects axons from the Cingulate Gyrus that travel immediately dorsal to the Corpus Callosum (CC) and along the ventral face of the Hippocampus, forming a large C-shape trajectory. It carries afferent connections from the Cingulate Gyrus to the entorhinal cortex. Because of its narrow tubular shape, it is often difficult to reveal its entire length by a single data set [11]. CG has the characteristic that both left and right CGs travel in parallel and are very close to each other. Often many tractography algorithms do not rep-

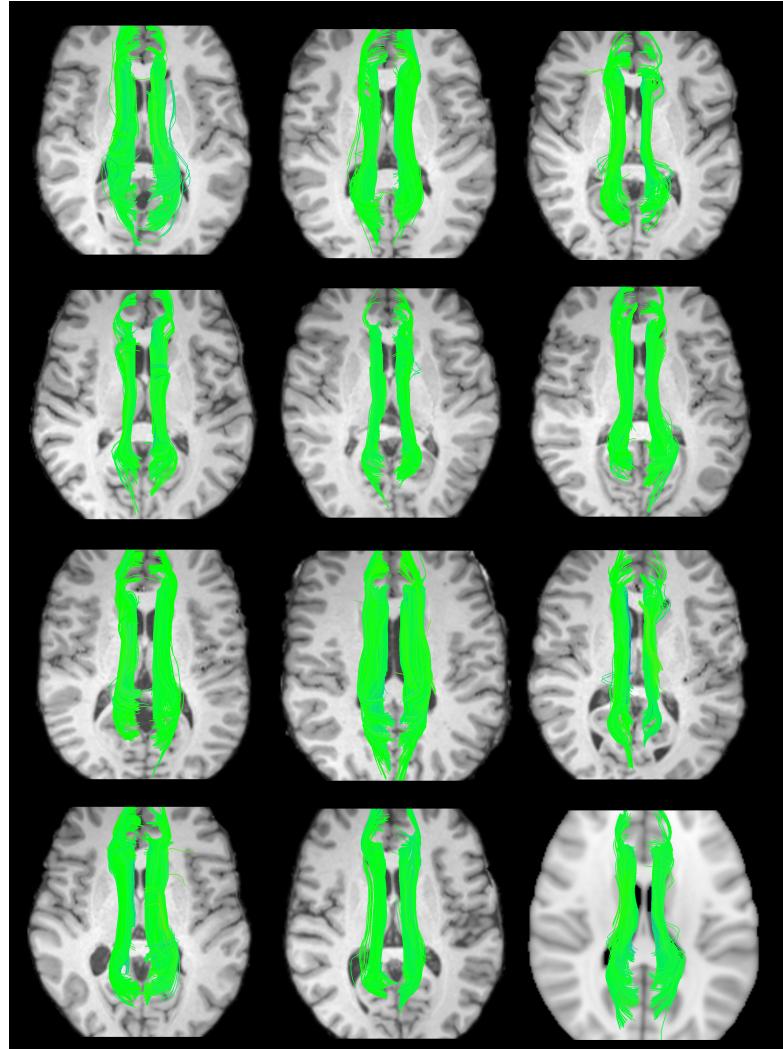


Figure 3.7: The left and right Cingulum (CG) bundles of 12 healthy subjects are presented here. Left and right CGs were selected in MNI space from entire EuDX tractographies of 1 million seeds and GQI as the reconstruction input. We also show the T1 image slice Z=10 for every subject except from last subject whose T1 was not available. For that subject we use the standard MNI template.

resent the Cingulum bundle very well because it is very close to the CC. However, using EuDX we see that both left and right CG bundles were consistent across all 12 subjects (see Fig. 3.7). The cingulum was selected using the interactive tool referenced previously from the entire tractography. The reconstruction parameters and EuDX parameters were the same as before with the exception of the number of initial seeds which was 1 million. Furthermore, the tractographies of all subjects were registered together with the structural images in MNI space. The only exception was with the last subject 12 whose T1 image was not available and it was visualized together with the MNI standard template (MNI152) (see bottom-right corner of Fig. 3.7).

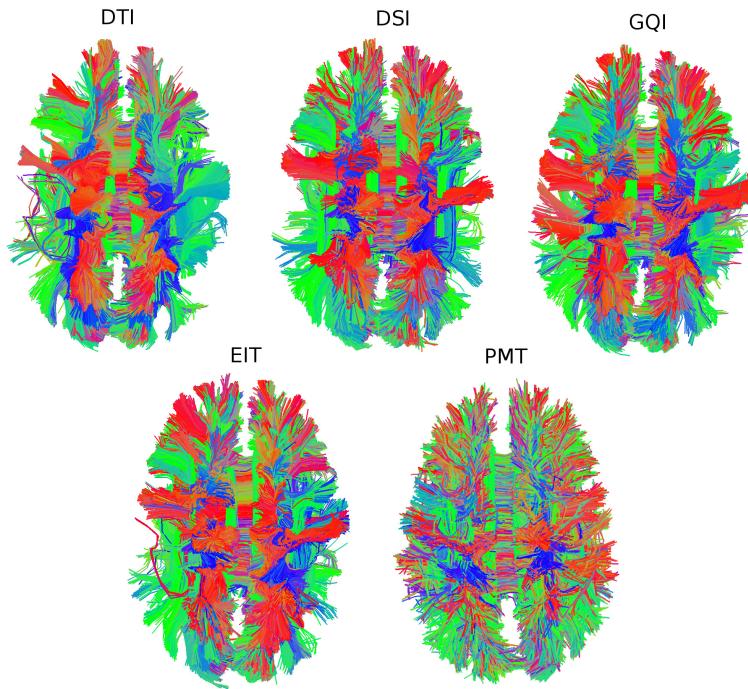


Figure 3.8: 5 tractographies from the same data sets of the same subject. The 3 top and the bottom left are created using our proposed deterministic approach (EuDX). The bottom right is created using Probabilistic tractography (PMT) .

In Fig. 3.8 we show results generated from a single subject. The data had size $96 \times 96 \times 55 \times 102$ and were acquired as described previously in this section. We first removed the scalp using bet and secondly created an FA image for this data. We masked out all voxels with $FA < 0.2$ in order to make sure that seeding will take place in areas of high anisotropy. We

then generated EuDX tracks using DTI, GQI, DSI and EITL reconstructions respectively. We also used the same seeds to generate probabilistic PMT tracks. In Fig. 3.8 we can see that EuDX based tracks are more uniform than the PMT tracks. We can also observe that the DTI tracks do not represent well the SLF regions of the bundles especially on the left hemisphere. Finally, we can observe the GQI, DSI and EIT tracks look the most similar which was expected as they have comparable angular accuracy (see Fig. 2.8). The EuDX parameters used here were $\mathcal{A}_{thr} = 0.2$, $\Delta s = 0.5$, $\theta_{thr} = 60^\circ$, $TW = 0.5$ with input from PK of GQI, DSI and EITL correspondingly. The Single Tensor was fitted using weighted least squares. For the DSI reconstruction we used standard parameters: q-space grid size 16, hanning filter width 32, and radial integration range from 2.1 to 6 at steps of 0.2. For GQI we used sampling length 1.2. For EITL we used for radial sampling from 0 to 5 with steps of 0.4 and Gaussian weighting of 0.05. The running time of EuDX was on average 11 seconds for generating on average 70,000 tracks.

3.5 Conclusion

We showed that EuDX is a fast deterministic tractography method which can be used to propagate in crossing and non-crossing areas in simulations and human subjects. With the help of expert neuroanatomists we confirmed that EuDX can be used to find bundles which are known from anatomy like AC, CST and we investigated how CG looked between different subjects.

Currently the neuroanatomists are continuing inspecting and segmenting more and more bundles and we hope that in the near future we will have a large collection of segmented bundles. These will be used to create better tractography algorithms by prior anatomical knowledge or exploit supervised learning techniques in order to find similar bundles in non-labeled subjects.

The purpose of EuDX is to be faithful to the reconstruction results rather than try to correct or enhance them by introducing regional or global considerations which is the topic of other methods which were reviewed earlier. Therefore, EuDX serves mainly as a robust method for quickly inspecting different reconstruction results using deterministic tractography. However, we observed that EuDX performed better than stan-

dard probabilistic approaches in the simulations. This observation needs further exploration and contradicts what is commonly assumed i.e. that probabilistic tractography will perform better.

We think that most algorithms concentrated on inferring the distributions of the major directions in each voxel. Nevertheless, not much attention was payed to the actual track density and integrity of the bundle. We believe this is an important issue especially for clinical applications.

EuDX will stop tracking on voxels with low anisotropy and will not take into consideration other voxels as it would happen with other probabilistic techniques. This property is often useful when validating underlying reconstruction models. Furthermore we showed that EuDX can take as input many different anisotropic functions: FA and QA or PK which have no restrictions for the underlying model neither to the number of peaks per voxel.

The source code for EuDX can be found in module `dipy.tracking.eudx` which is freely available at dipy.org.

4 Highly Efficient Tractography Clustering

4.1 Overview

Current tractography propagation algorithms can generate very large tractographies which are difficult to interpret and visualize. A clustering of some kind seems to be a solution to simplify the complexity of these datasets and provide a useful segmentation; however most proposed clustering algorithms are very slow and often need to calculate pairwise distances of size $N \times N$ where N is the number of tracks. This amount of comparisons adds a heavy load on clustering algorithms forcing them to be inefficient and therefore impractical for everyday analysis as it is difficult to compute all these distances or even store them in memory. This adds a further overhead to the use of tractography for clinical applications but also introduces a barrier on understanding and interpreting the quality of diffusion data sets. We show in this chapter that a stable, on average linear time clustering algorithm exists. We call this algorithm Quick-Bundles (QB). QB can be used to generate meaningful clusters in seconds with minimum memory consumption. In our approach we do not need to calculate all pairwise distances unlike most of the other existing methods. Furthermore, we can update our clustering online or in parallel. We show that we can generate meaningful clusters of the order of 1,000 times faster than any other available method and that it can be used to segment from a few hundred to many millions of tracks. Moreover our method is multi-purpose; its results can either stand on their own to explore the neuroanatomy directly, or the clustering technique can be used as a precursor tool which reduces the dimensionality of the data, which can then be used as an input to other algorithms of higher order complexity, resulting in their greater efficiency. Beyond the use of this algorithm to simplify tractographies, we show here how it can help identify landmarks, create atlases, and compare and register tractographies.

4.2 Track distances and preprocessing

For clarity we first give brief details of various metrics for distances between tracks as they are integral to an understanding of the track clustering literature. Numerous distance metrics between two trajectories have been proposed in the literature, such as in [158], [159], [160] with the most common being the Hausdorff distance found in [161] and many other studies. We mainly use a very simple symmetric distance proposed in [162] and [163] which we call Minimum average Direct-Flip MDF(s_A, s_B) distance between track s_A and track s_b (see Eq. 4.1). This distance can be applied only when both tracks have the same number of points. Therefore, we assume that an initial downsampling of tracks has been implemented, where all segments on a track have the same length, and all tracks have the same number of segments. Under that assumption MDF is defined as:

$$\begin{aligned} \text{MDF}(s_A, s_B) &= \min(d_{\text{direct}}, d_{\text{flipped}}), \text{ where} \\ d_{\text{direct}}(s_A, s_B) &= \frac{1}{K} \sum_{i=1}^K \|\mathbf{x}_i^A - \mathbf{x}_i^B\|_2 \text{ and} \\ d_{\text{flipped}}(s_A, s_B) &= \frac{1}{K} \sum_{i=1}^K \|\mathbf{x}_i^A - \mathbf{x}_{K-i}^B\|_2 \end{aligned} \quad (4.1)$$

where K is the number of points \mathbf{x}_i on the two tracks A and B .

In some cases it is still valid to use a family of Hausdorff distances which for simplicity we denote as MAM distances – short for Minimum, or Maximum, or Mean, Average Minimum distance (MAM). We mostly use the Mean version of this family, (see Eq. 4.5) but the others are potentially useful as they can weight different properties of the tracks. These distances are slower to compute than MDF but they can work with different number of segments on tracks; a property that is useful for some ap-

plications. The equations below show the formulation of these distances:

$$d_{\text{avg}}(s_A, s_B) = \frac{1}{K_A} \sum_{i=1}^{K_A} d(x_i^A, s_B),$$

$$d_{\min}(s_A, s_B) = \min_{j=1, \dots, K_B} d(x_j^A, s_B), \text{ and} \quad (4.2)$$

$$d_{\max}(s_A, s_B) = \max_{j=1, \dots, K_B} d(x_j^A, s_B) \text{ where} \quad (4.3)$$

$$d(\mathbf{x}, s_B) = \min_{j=1, \dots, K_B} \|\mathbf{x} - \mathbf{x}_j^B\|_2.$$

$$\text{MAM}_{\min}(s_A, s_B) = \min(d_{\text{avg}}(s_A, s_B), d_{\text{avg}}(s_B, s_A)) \quad (4.4)$$

$$\text{MAM}_{\max}(s_A, s_B) = \max(d_{\text{avg}}(s_A, s_B), d_{\text{avg}}(s_B, s_A))$$

$$\text{MAM}_{\text{avg}}(s_A, s_B) = (d_{\text{avg}}(s_A, s_B) + d_{\text{avg}}(s_B, s_A))/2 \quad (4.5)$$

where the number of points K_A and K_B on the two tracks are not necessarily the same. For the same threshold value MAM_{\min} , MAM_{\max} and MAM_{avg} will give different results. For example, MAM_{\min} will bring together more short tracks with long tracks than MAM_{\max} and MAM_{avg} will have an in between effect. Finally, other distances than the average minimum based on the minimum (see Eq. 4.2) or maximum distance (see Eq. 4.3) can be used. However, we have not investigated them in this thesis.

The main advantages of the MDF distance (see Eq. 4.1), are that it is fast to compute, it takes account of track direction issues through consideration of both direct and flipped tracks, and that it is easy to understand how it will behave, from the simplest case of parallel equi-length tracks to the most complicated of very divergent tracks. Another advantage is that it will separate short tracks from long tracks; a track A that is half the length of track B will be relatively poorly matched on MDF to B. We will see later in this chapter that this helps to find broken or erroneous tracks. An asset of having tracks with the same number of points is that we can easily do pairwise calculations on them; for example add two or more tracks together to create a new average track. We will see in the next section that track addition is a key property of our clustering algorithm. Some care should be taken into consideration with the number of points allowed in a track (track downsampling). We always keep the endpoints intact and then downsample in equidistant segments. This means that short tracks will have the same number of points as long tracks. There-

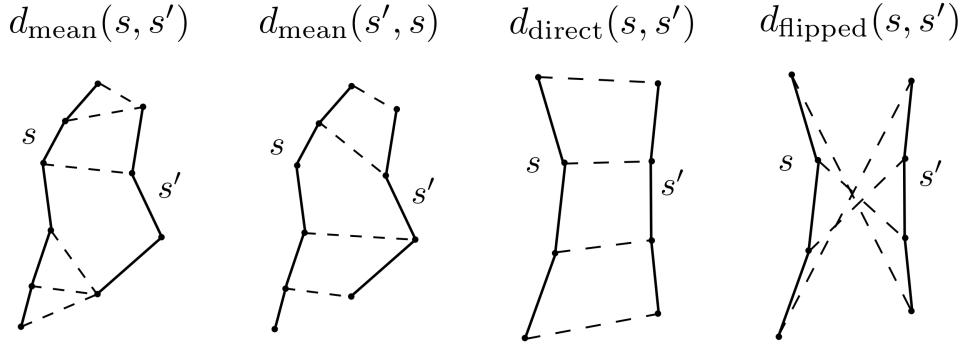


Figure 4.1: Distances used in this work. The main distance used is minimum average direct flip (MDF) distance $MDF = \min(d_{\text{direct}}, d_{\text{flipped}})$ which is a symmetric distance that can deal with the track bi-directionality problem and works on tracks which have the same number of points. Another distance is the mean average distance which is again symmetric but does not need for the tracks to have the same number of points $MAM_{\text{avg}} = (d_{\text{avg}}(s_A, s_B) + d_{\text{avg}}(s_B, s_A))/2$. The components of both distances are shown; with solid lines we draw the tracks, and then with dashed lines we connect the pairs of points of the two tracks whose distances contribute to the overall metrics.

fore, the curvature from the long tracks will be lost relative to the short tracks i.e. the short tracks will have higher resolution. We found empirically that this is not an important issue and that for clustering purposes even downsampling to only 3 points in total could be useful [162]. Depending on the application less or more points can be used.

4.3 Related Work

During the last 10 years there have been numerous efforts from many researchers to address the unsupervised and supervised learning problems of brain tractography. As far as we know all these methods suffer from low efficiency, however they provide many useful ideas which we describe in this section.

Tractography clustering algorithms are rarely compared in the literature. Nonetheless, Moberts et al. [164] are an exception. They evaluated different popular hierarchical clustering methods including a less common one, shared nearest neighbor (SNN), against a gold standard segmentation by physicians. The authors concluded that single-link cluster-

ing with mean average distance was the method which performed best. Wang et al. [165] proposed a nonparametric Bayesian framework using hierarchical Dirichlet processes mixture model (HDPM). This is one of the very few methods not based on distances. In this work a track is modeled as a discrete distribution over a codebook of discretized orientations and voxel regions. The authors explain that calculating pairwise distances is very time consuming and therefore they avoid using them. Their approach automatically learns the number of clusters from data with Dirichlet processes priors but it is still not efficient enough for real time operation. A disadvantage of this method is that the priors do not originate from anatomical knowledge.

Visser et al. [163] used hierarchical clustering and fuzzy c-means together with recombination of subsets of the same tractography to reduce the effect of the large datasets on the distance matrix based on the MDF distance (see section 4.2) [162]. An interesting result with this method was that they could automatically find the different sub-bundles of the Arcuate Fasciculus region in accordance with the supervised labeling described in [166]. The algorithm that we present in this chapter also uses the minimum average flip (MDF) metric as a measure of distance between tracks. Gerig et al. [167] also used hierarchical clustering with a symmetrised version of closest point distances, MAM_{avg} and MAM_{max} (Hausdorff). However, they tested their method with only two bundles: Uncinate Fasciculus and the Corticospinal Tract.

Guevara et al. [168] combined a great number of different algorithms from hierarchical clustering to 3D watershed on track extremities. They first divided the tractography into left-right hemisphere, inter-hemispheric and cerebellum subsets. They then created further subsets of different track length, used hierarchical clustering based on the random voxel parcels, used watershed over extremities and finally used hierarchical clustering to merge the different sub-bundles using the Hausdorff distance (see section 4.2). This work stressed the need to divide the data set between shorter and longer tracks. Tsai et al. [169] used a combination of cluster methods based on minimum spanning trees, locally linear embedding and k-means. They were able to incorporate both local and global structures by changing a few parameters. The main advantage of this method was that it showed a way to merge a chain of neighbouring structures into one cluster. Zhang and Laidlaw [160] used an agglomerative hierarchical

clustering using the same distance as in [170] and later in [171] combined distance-based single linkage hierarchical clustering with expert labeling of specific bundles. Zvitia et al. [172], [173] used adaptive mean shift. This is a clustering algorithm which finds automatically the number of clusters. This is in contrast for example with k-means that the user needs to prespecify the number of clusters. They also used this approach for direct registration of tractographies but only with tractographies from the same subject. El Kouby et al. [174] created a ROI-based connectivity matrix where the i, j th entry of the matrix holds the number of tracks which connect ROI_i to ROI_j . K-means was used afterwards on the rows of the matrix to cluster the tracks. This technique can be used for clustering bundles across subjects.

Brun et al. [175] used the mean and covariance of the track as the feature space and normalized cuts based on a graph theoretic approach for the segmentation. Ding et al. [176] used k-nearest neighbours, another agglomerative approach, applied to corresponding track segments. Corouge et al. [161] used different types of track distances, e.g. Hausdorff distances, and other geometric properties such as torsion and curvature, and in [177] and [178] used Generalized Procrustes Analysis and Principal Components Analysis (PCA) to analyze the shape of bundles.

O'Donnell et al. [179] created a tractographic atlas using spectral embedding and expert anatomical labeling. They then automatically segmented using spectral clustering and expressed the tracks as points in the embedded space to the closest existing atlas clusters. The full affinity matrix was too big to compute, therefore they used the Nystrom approximation: working on a subset and avoid generating the complete affinity/distance matrix. Later in [180] they tried group analysis on prespecified bundles.

Maddah et al. [181] used B-spline representations of tracks referenced to an atlas, and then the tracks were clustered based on the labeled atlas. Later Maddah et al. [182] using a similar track representation (quintic B-splines) calculated a model for each bundle as the average and standard deviation of that parametric representation. In that way they created an atlas which is used as a prior for expectation maximization (EM) clustering of the Corpus Callosum tracks into Witelson subdivisions [183] using population averages. Later in [184] Maddah et al. it is showed that it is possible to combine spatial priors with metrics for the shape of the tracks

in order to guide the clustering process.

Jonasson et al. [185] created a large $N \times N$ co-occurrence matrix, where N is the number of the fibers to cluster. The co-occurrence (affinity) matrix contained the number of times that two fibers share the same voxel. They then used spectral clustering. Jianu et al. [186] presented a new method for visualizing and navigating through tractography data combining dendograms from hierarchical clustering along with 3D- and 2D-embeddings using the approximation that Chalmers [187] introduced for the technique of Eades [188].

Durrleman et al. [189] introduced electrical current models of fibre bundles where a fibre is seen as a set of wires sending information in one direction at constant rate. Currents have good diffeomorphic properties and can be used for registration of bundles as shown in [189] and later in [190]. This methodology does not impose point-to-point or fibre-to-fibre correspondences, however it is sensitive to fibre density and orientation of the bundles and it is computationally expensive.

Leemans and Jones [191] used affinity propagation (section A.6) to cluster the fronto-occipital fibres, Cingulum and Arcuate Fasciculus after reducing the complexity of the data sets using additional frontal and occipital boolean masks on the right cerebrum. Results however were shown on a very small part of the entire tractography where clustering is a much easier problem. Later Malcolm et al. [192] used affinity propagation to cluster a full brain tractography created using filtered tractography and suggested that affinity propagation is not suitable for group clustering.

Ziyan et al. [193] introduced a probabilistic registration and clustering algorithm based on expectation maximization (EM) which creates a sharper atlas from a set of subjects on three bundles: Corpus Callosum, Cingulate and Fornix. This work used an initial spectral clustering [179] to label the bundles and then updated these labels iteratively while performing bundle-wise registration combined using polyaffine integration.

Often, it is useful to use some protocols in order to add prior information to the automated learning process. Protocols to manually label 11 major white matter tracts were described in Wakana et al. [194] using ROIs to include or exclude tracks generated by deterministic tractography. Hua et al. [195] used regions of interest together with probabilistic tractography in order to create probability maps of known fibre bundles.

From this short review we observe two main trends in the literature.

The first and most common one makes use of track distances and calculates distance matrices. The most prevailing approaches here for deciphering the distance matrix are with Hierarchical and Spectral Clustering which are applied only on subsets of the initial tractography. The second trend and least common recommends avoiding track distances because the computation of the distance matrix is memory intensive. In this case, using Dirichlet Processes or Currents or Connectivity based parcelation seem to be some viable solutions. However, clustering is to be applied in clinical usage or to make neuroscientists' analysis more efficient and practical we need algorithms that can provide useful clusters and cluster descriptors in minimum time. None of the papers described in this literature review provide a solution to this issue of efficiency and most of the methods would require from many hours to many days to run on a standard sized data set. The method we propose in this document can provide a solution to this problem and it is an extensive update of our preliminary work described in Garyfallidis et al. [162].

Most authors agree that unsupervised learning with tractographies is a difficult problem as the data sets are very large, dense, cluttered with noisy tracks which could have no anatomic relevance and bundles which are more than often tangled together in many areas. Furthermore, we observe that there is a strong disagreement on the number of clusters (from 10 to 60). Because of the difficulty of the problem an international contest was also organized by SchLab in Pittsburgh University (PBC Brain Connectivity Challenge - IEEE ICDM) in 2009. However, the competition did not conclude to any directly viable solutions. We think that in order to find big clusters a lot of anatomical prior knowledge needs to be introduced in a way that is not yet established. Nevertheless, the clustering that we propose concentrates on reducing the complexity of the data rather than finding bundles with anatomical relevance. We believe this step is more useful at this stage of tractography analysis research.

4.4 Data sets

We experimented with QuickBundles using simulations, 10 human tractographies collected and processed by ourselves, and one tractography with segmented bundles which was available online.

Simulated trajectories. We generated three different bundles of para-

metric paths sampled at 200 points. The tracks were made from different combinations of sinusoidal and helicoidal functions. Each bundle contained 150 tracks. For the red bundle in Fig. 4.4 a pencil of helical tracks all starting at the same point on a cylinder was generated by linearly varying the pitch of the helices; the green bundle was made up from a divergent pencil of rays on a sinusoidally corrugated sheet; the blue bundle was similarly made from a divergent rays on a sinusoidally corrugated sheet, with the rays undergoing sinusoidal modulated lateral bending over a range of amplitudes. The data set contained 450 tracks in total.

Human subjects. We collected data from 10 healthy subjects at the MRC-CBU 3T scanner (TIM Trio, Siemens), using Siemens advanced diffusion work-in-progress sequence, and STEAM [101, 15] as the diffusion preparation method. The field of view was $240 \times 240 \text{ mm}^2$, matrix size 96×96 , and slice thickness 2.5 mm (no gap). 55 slices were acquired to achieve full brain coverage, and the voxel resolution was $2.5 \times 2.5 \times 2.5 \text{ mm}^3$. A 102-point half grid acquisition [72] with a maximum b -value of 4,000 s/mm² was used. The total acquisition time was 14' 21'' with TR=8,200 ms and TE=69 ms. The experiment was approved by the Cambridge Psychology Research Ethics Committee (CPREC).

For the reconstruction of the real data sets we used GQI (formula 2.14) with diffusion sampling length 1.2 and for the tractography propagation we used EuDX (Euler integration with trilinear interpolation, see 3) with 1 million random seeds, angular threshold 60° , total weighting 0.5, propagation step size 0.5 and anisotropy stopping threshold 0.0239 (see Figs. 4.10 and 4.14).

PBC human subjects. We also used a few labeled data sets (see Fig.4.3, 4.5), from the freely available tractography database used in the Pittsburgh Brain Completion Fall 2009 ICDM⁴.

4.5 QuickBundles (QB) Clustering

4.5.1 The QB Algorithm

QB is a surprisingly simple, linear time $O(N)$ (see section 4.5.3), distance based clustering algorithm that we created in order to segment huge trajectory data sets such as those produced by current state-of-the-art tractography generation algorithms [113, 3]. In general, there are very few linear

⁴ braincompetition.org

time clustering algorithms. Just two are well known in the literature of artificial intelligence, machine learning and data mining: CLARANS [196] and BIRCH [197]. QB is different from both of these methods; we will motivate it by describing some aspects of BIRCH as a starting point for the presentation of QB.

BIRCH [198] has two key components: first is relatively simple and involves the use and updating of clusters' descriptors; second is the construction of a tree structure in which the accumulated clusters are held. The latter component is aimed at maintaining efficient searchability of the database while balancing what is kept in memory and what is on disc for very large databases. BIRCH uses clustering descriptors which are either directly available for each item in the data set or are easily computed from them, e.g. squares and products of components; these form specific vectors of a fixed dimension of numerical values. Each cluster in turn has a descriptor which is an aggregate of the properties of the items that belong to it (e.g. the sum or mean of the individual descriptor vectors). Proceeding by a single sweep through the dataset, items are adjoined to clusters on the basis of their proximity to the clusters, subject to a maximum cluster size, or they are added as new leaves into the hierarchical tree structure in which the evolving clusters are held. Updating steps follow which can involve the merging of previously created clusters in a k-means fashion [199, 200].

It is the linear nature of BIRCH combined with the fixed dimensionality of its cluster descriptors that makes it quite fast. However, the further steps involving reorganisation of the accumulated tree do add some major overheads to BIRCH's performance. QB capitalises on these positive features but does not try to create any kind of hierarchical structure for the clusters. Moreover, while items in BIRCH are fixed dimension vectors with no additional structure, in QB each item (track) is a fixed-length ordered sequence of points in \mathbb{R}^3 , and uses metrics and amalgamations which take account of, and preserve, this structure. Furthermore, each item is either added to an existing cluster on the basis of a distance between the cluster descriptor of the item and the descriptors of the current set of clusters or a new cluster is created. Clusters are held in a list which is extended according to need.

The complete QB algorithm is described in formal detail in Alg. 5 and a simple step by step visual example is given in Fig. 4.2. One of the reasons

why QB has on average linear time complexity derives from the structure of the cluster node: we only save the sum of current tracks h in the cluster and the sum is cumulative; moreover there is no recalculation of clusters, the tracks are passed through only once and a track is assigned to one cluster only.

Algorithm 5 QuickBundles

Input tracks $T = \{s_1, \dots, s_i, \dots, s_N\}$, threshold θ
Output clustering $C = \{c_1, \dots, c_k, \dots, c_M\}$ where cluster $c = (I, h, N)$

```

 $c_1 \leftarrow ([1], s_0, 1)$ 
 $C \leftarrow \{c_1\}$  # the first track becomes the first cluster
 $M \leftarrow 1$  # the total number of clusters is 1
For  $i = 2$  to  $N$  Do # all tracks
     $t \leftarrow T_i$ 
     $\text{alld} \leftarrow \text{infinity}(M)$  # distance buffer
     $\text{flip} \leftarrow \text{zeros}(M)$  # flipping check buffer
    For  $k = 1$  to  $M$  Do # all clusters
         $v \leftarrow C_k.h / C_k.n$ 
         $d \leftarrow d_{\text{direct}}(t, v)$ 
         $f \leftarrow d_{\text{flipped}}(t, v)$ 
        If  $f < d$  Then
             $d \leftarrow f$ 
             $\text{flip}_k \leftarrow 1$ 
             $\text{alld}_k \leftarrow d$ 
        EndIf
         $m \leftarrow \min(\text{alld})$ 
         $l \leftarrow \text{argmin}(\text{alld})$ 
        If  $m < \theta$  Then # append in current cluster
            If  $\text{flip}_l = 1$  Then
                 $C_l.h \leftarrow C_l.h + \text{reverse}(t)$ 
            Else
                 $C_l.h \leftarrow t$ 
                 $C_l.n \leftarrow C_l.n + 1$ 
                append( $C_l.I, i$ )
            Else # create new cluster
                 $c_{M+1} \leftarrow ([i], t, 1)$ 
                append( $C, c_{M+1}$ )
                 $M \leftarrow M + 1$ 
            EndIf
        EndFor
    
```

QB creates an online list of cluster nodes. The cluster node is defined as $c = (I, h, n)$ where I is the list of the integer indices of the tracks in that

cluster, h is an $K \times 3$ matrix, the most important descriptor of the cluster, and n is the number of tracks on that cluster. h is a matrix which can be updated online when a track is added to a cluster and is equal to

$$h = \sum_{i=1}^n s_i \quad (4.6)$$

where s_i is the $K \times 3$ matrix representing track i , Σ represents matrix addition, and n is the number of tracks in the cluster. QB assumes that all tracks have the same number of points K , therefore a downsampling of tracks, typically equidistant, is necessary before QB starts. A short summary of the algorithm goes as follows.

Select the first track s_1 and place it in the first cluster $c_1 \leftarrow ([1], s_1, 1)$. For all remaining tracks (i) go to next track s_i ; (ii) calculate MDF distance between this track and virtual tracks of all existing clusters c_k , where a virtual track is defined on the fly as $v = h/n$; (iii) if the minimum MDF distance is smaller than a distance threshold θ add the track to the cluster $c_j \leftarrow (I, h, n)$ with the minimum distance and update $c_j \leftarrow (I \cup [i], h + s, n + 1)$; otherwise create a new cluster $c_{M+1} \leftarrow ([i], s_i, 1)$ and increase the total number of clusters $M \leftarrow M + 1$. The complete algorithm is given in Alg. 5.

Choice of orientation can become an issue when using the MDF distance and adding tracks together. This happens because the diffusion signal is symmetric around the origin. Therefore, the $K \times 3$ track can equivalently have its points ordered $1, \dots, K$ or be flipped with order $K, \dots, 1$; the diffusion signal does not allow us to distinguish between these two directions. A step in QB takes account of the possibility of needing to perform a flip of a track before adding it to a representative track according to which direction produced the MDF value. Though the appropriate orientation (direct or flip) of a track was available in the MDF calculation at the time it entered a cluster, we allow for the possibility this might not be the same later on when the virtual track has evolved so it will need to be recalculated.

One of the reasons why QB has on average linear time complexity derives from the fact that we only save the sum of current tracks in the cluster and this is achieved cumulatively. QB passes through the tracks only once and that a track is assigned to one cluster only. By contrast, if we were using k-means at every iteration we would have to re-assign tracks

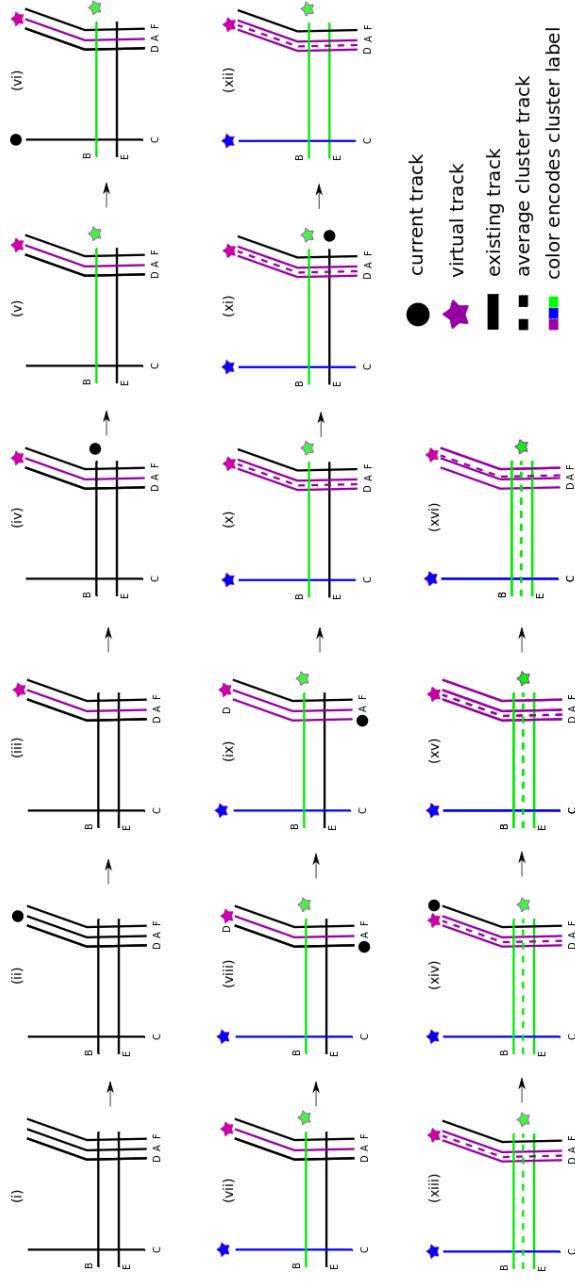


Figure 4.2: Step-by-step description of QB. Panel (i): 6 unclustered tracks (A-F) are presented; the distance threshold used is the MDF distance (Eq. 4.1) between B and E. The algorithm starts and in (ii) track A was selected, so no other clusters exist therefore track A becomes the first cluster (labeled with purple color) and the virtual track of that cluster is identical with A as seen in (iii), next in (iv) track B is selected and we calculate the MDF distance between B and the virtual track of the other clusters. At this moment there is only one cluster to compare so QB calculates MDF(B,virtual-purple) and this is obviously bigger than threshold (that being MDF(B,E)) therefore a new cluster is assigned for B and B becomes the virtual track of that cluster as shown in (v). In (vi) the next track is selected and this is again far away from both purple and blue virtuals therefore another cluster is created and B is the virtual of the blue cluster as shown in (vii). In (viii) track D is the current track and after we have calculated MDF(D,purple), MDF(D,blue) and MDF(D,green) it is obvious that D belongs to the purple cluster as MDF(D,purple) is smaller and lower than threshold as shown in (ix). However, we see in (x) that things change for the purple cluster because the virtual track is not anymore made by only one track but it is the average of D and A shown with dashline. In (xi) E is the current track and will be assigned at the green cluster as shown in (xii) because $MDF(E,virtual\ green) = MDF(E,B)$ = threshold, and in (xiii) we see the updated virtual track for the green cluster which is equal to $(B+E)/2$, where + means track addition. In (xiv) the last track is picked and compared with the virtual tracks of the other 3 clusters; obviously $MDF(F,purple)$ is the only with smaller threshold, therefore F is assigned to the purple cluster in (xv). Finally, in (xvi) the virtual purple track is updated as $(D+A+F)/3$. As there are no more tracks to select, the algorithm stops. We observe that all three clusters have been found and all tracks have been assigned successfully.

to clusters and recalculate averages which is computationally much more intensive.

QB can be extended for specific applications to contain more information about the clusters. For example, we could redefine $c \leftarrow (I, h, n, h^{(2)})$ to obtain second order information and in that way we could calculate the variance of the cluster where

$$h^{(2)} \leftarrow (\sum_{i,j} \mathbf{x}_{ij}^2, \sum_{i,j} \mathbf{y}_{ij}^2, \sum_{i,j} \mathbf{z}_{ij}^2, \sum_{i,j} \mathbf{x}_{ij}\mathbf{y}_{ij}, \sum_{i,j} \mathbf{y}_{ij}\mathbf{z}_{ij}, \sum_{i,j} \mathbf{x}_{ij}\mathbf{z}_{ij})$$

and \mathbf{x}_{ij} , \mathbf{y}_{ij} , \mathbf{z}_{ij} are the coordinates of the j th point of the i th track in the cluster. Although this alternative would be very useful, as even more refined cluster distances could be used which take into account the additional information, this is not addressed in this thesis.

One of the disadvantages of most clustering algorithms is that they give different results with different initial conditions; for example this is recognised with k-means, expectation-maximization [201] and k-centres [202], where it is common practice to try a number of different random initial configurations. The same holds for QB so if there are not distinct clusters such that the distance between any pair of clusters is supra-threshold, then with different permutations of the same tractography we will typically see similar number of clusters but different underlying clusters. We will examine the robustness of QB in this respect in section 4.6.

4.5.2 Powerful simplifications

One of the major benefits of applying QB to tractographies is that it can provide meaningful simplifications and find structures that were previously invisible or difficult to locate because of the high density of the tractography. We used QB for example to cluster the corticospinal tract (CST). This bundle was part of the datasets provided by the Pittsburgh Brain Competition (PBC2009-ICDM) and it was selected by an expert. The result is clearly shown in Fig. 4.3 where every partition is represented by a virtual track. To generate this clustering we used a tight threshold of 10 mm and downsampling to 12 points. We observe that only a few virtual tracks span the full distance from bottom to top and that many tracks are broken (i.e. shorter than what was initially expected) or highly divergent.

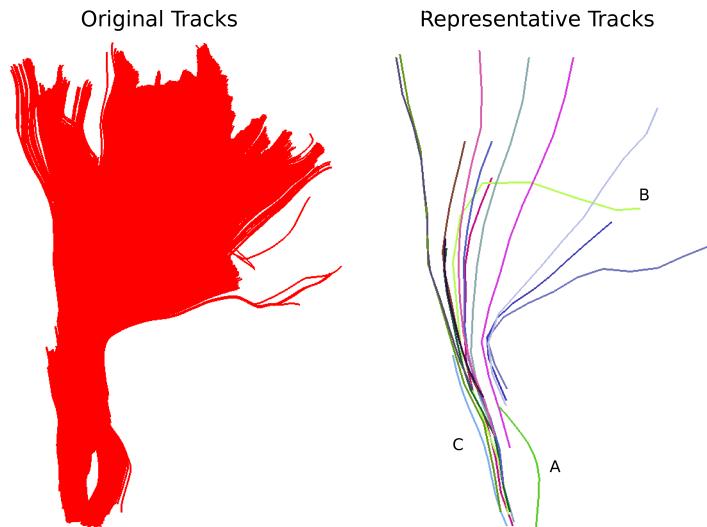


Figure 4.3: Part of the CST bundle (red) consisting of 11,041 tracks labelled by an expert. At first glance it looks as though all tracks have a similar shape, possibly converge towards the bottom, and fan out towards the top. However, this is a misreading caused by the opaque density when all the tracks are visualised. QB can help us see the finer structure of the bundle and identify its elements. On the right hand side we see the 14 QB representative tracks (virtuals) of the CST. We can now clearly see that several parts which looked homogeneous are actually broken bundles e.g. dark green (A), light blue (C), or bundles with very different shape e.g. light green (B). To cluster this bundle took 0.1 seconds.

Another interesting feature of QB is that it can be used to merge or split different structures by changing the distance threshold. This is shown in Fig. 4.4; on the left we see simulated paths made from simple sinusoidal and helicoidal functions packed together. The colour coding is used to distinguish the three different structures. With a lower threshold the three different structures remain separated but when we use a higher threshold the red and blue bundles are represented by only one cluster; represented by a purple virtual.

Similarly, with the simulations shown in Fig. 4.4 we can see the same effect on real tracks, e.g. those of the fornix shown at the left panel of Fig. 4.5. Different number of clusters can be obtained at different thresholds. In that way we can stress thinner or larger sub-bundles inside other bigger bundles.

A full tractography containing 250,000 tracks was clustered using QB with a distance threshold of 10 mm (see Fig. 4.5). We produced a useful re-

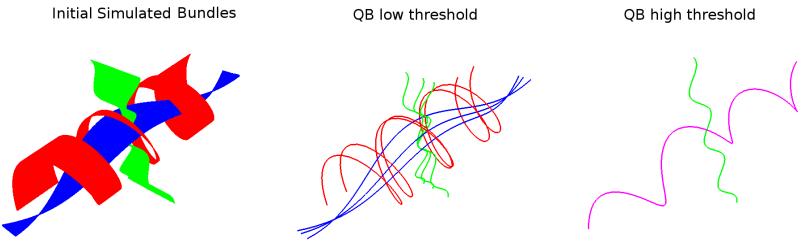


Figure 4.4: Left: 3 bundles of simulated trajectories; red, blue and green consisting of 150 tracks each. All 450 tracks are clustered together using QB. Middle and Right: virtual tracks using thresholds 1 and 8 respectively. At low threshold the underlying structure is reflected in a more detailed representation. At higher threshold, closer bundles merge together. Here the red and blue bundle have merged together in one cluster represented by the purple virtual track.

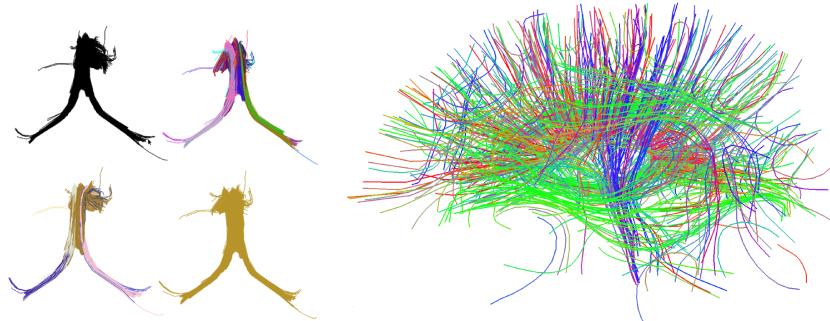


Figure 4.5: Left: QB clustering of the fornix bundle. The original fornix is shown in black (1,076 tracks). All tracks were equidistantly downsampled at 3 points. With a 5 mm threshold QB generates 22 clusters (top right). With 10 mm it generates 7 (bottom left) and with 20 mm the whole fornix is determined by one cluster only (bottom right). The colour encodes cluster label. Right: an example of a full tractography (0.25×10^6 tracks) being clustered using QB with a distance threshold of 10 mm. 763 virtual tracks were produced which is a huge simplification of the initial tractography. Every track shown here represents an entire cluster from 10 to 5,000 tracks each. These can be thought as fast access points to explore the entire data set. The colour here encodes track orientation.

duction of the initial tractography leaving only 763 virtual tracks. Bundles smaller than 10 tracks were removed. Every track shown here represents an entire cluster containing from 10 to 5,000 tracks each. The virtual tracks have a great usage as fast access points to explore the complete tractography (see Fig. 4.5).

4.5.3 Complexity and timings

To apply QB to a data set we need to specify three key parameters: K , the fixed number of downsampled points per track; θ the distance threshold, which controls the heterogeneity of clusters; and N the size of the subset of the tractography on which the clustering will be performed. When θ is higher, fewer more heterogeneous clusters are assembled, and conversely when θ is low, more clusters of greater homogeneity are created.

The complexity of QB is in the best case linear time $\mathcal{O}(N)$ with the number of tracks N and worst case $\mathcal{O}(N^2)$ when every cluster contains only one track. The average case is $\mathcal{O}(MN)$ where M is the number of clusters. However, because M is usually much smaller than N ($M \ll N$) we can neglect M and denote it only as $\mathcal{O}(N)$ as it is common in complexity theory.

We created the following experiment to investigate this claim and we found empirically that the average case is actually $\mathcal{O}(N)$ for tractographies (see Fig. 4.6). In this experiment we timed the duration of QB clustering of tractographies containing from 10^5 to 10^6 tracks, with different initial number of points per track (3, 6, 12 and 18) and different QB thresholds (10, 15, 20, 25 mm). The final factor, not shown explicitly in these diagrams, is the underlying structure of the data which is expressed by the resulting number of clusters. These results were obtained on a single thread of an Intel(R) CPU at 2.50GHz on a standard PC. The results can be seen in Fig. 4.6. We see how the linearity of the QB algorithm with respect to N only reduces slightly even when we use a very low threshold such as 10 mm which can generate many thousand of clusters. This experiment concludes that QB is suitable for fast clustering. Even when the threshold value becomes impressively low (10 mm) the linearity is only slightly disturbed.

Furthermore, the memory usage of QB is $\mathcal{O}(M)$ where M is the number of clusters and because this is usually much smaller than N we consider

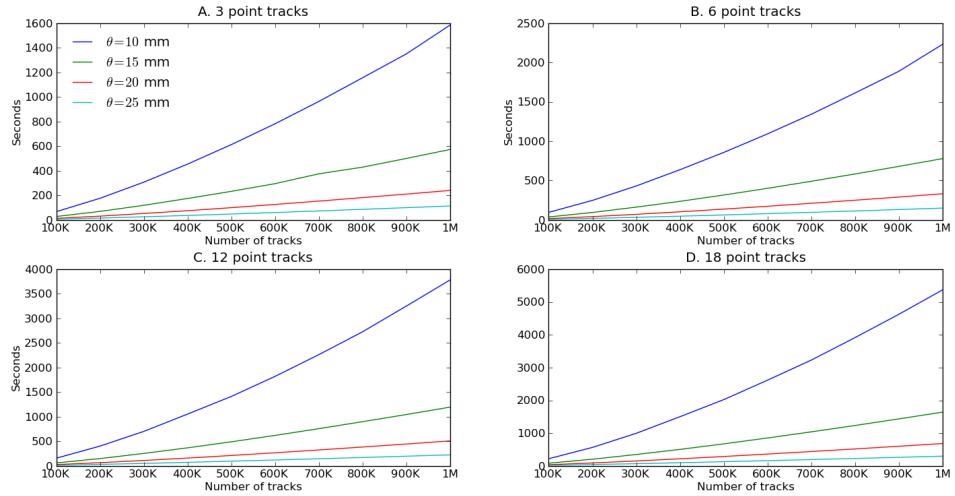


Figure 4.6: Time comparisons of QB using different number of points per track, different distance thresholds and different number of tracks. QB is a very efficient algorithm whose performance is controlled by just three parameters. (1) the initial downsampling K of the tracks exemplified in four sub-diagrams: 3 points (A), 6 points (B) 12 points (C), 18 points (D). (2) the distance threshold θ in millimeters shown in 4 colours: 10 mm (blue), 15 mm (green), 20 mm (red), 25 mm (cyan). We used a full tractography to generate these figures without removing or preselecting any parts. Random subsets of the tractography were chosen with size N from 10^5 to 10^6 (x-axis).

memory consumption to be negligible. Because in QB we store only the indices of the tracks, even for very large tractographies 20 or more clusterings can be stored simultaneously in the RAM of a simple notebook without any problems. Memory efficiency is therefore another feature of QB.

We compared QB with 12 point tracks and distance threshold at $\theta = 10$ mm versus some timings reported from other state of the art methods found in the literature (Tab. 4.1). Unfortunately, timings were very rarely reported because most algorithms were very slow on full data sets. Nonetheless the speedup that QB offers is obviously of great importance and even real-time on data sets of less than 20,000 tracks (see Tab. 4.1). It holds also the prospect of real-time clustering on massive tractographies using standard parallelization techniques (see section 4.7).

Number of tracks (N)	Algorithms	Timings (secs)	QB (secs)	Speedup
1000	Wang et al. [165]	30	0.07	429
60,000	Wang et al. [165]	14,400	14.7	980
400,000	Visser et al. [163]	75,000	160.1	468

Table 4.1: QB run on $K = 12$ point tracks and distance threshold at $\theta = 10$ mm compared with some timings reported from other state of the art methods found in the literature. Timings were very rarely reported until today as most algorithms were very slow on full data sets. Nonetheless, we can observe in this table that the speedup that QB offers is substantial.

4.5.4 Virtual tracks, exemplar tracks and other descriptors.

The virtual tracks created by QB have very nice properties as they represent an average track which can stand as the most important feature of the cluster that they belong to. However, now that we have segmented our tractography into small bundles we can calculate many more potentially important descriptors for the cluster. The Cluster Spread (CS) for instance can be computed for any cluster c as a vector of length K whose j -th component is $\sum_{x \in c} |x_j - v_j|^2 / n$. Here, x_j is the j -th point in the track x in cluster c , v_j is the corresponding point of the virtual track, and n is the size of the cluster. CS provides a profile of the tightness or looseness of the cluster along the length of the virtual track. Many other similar or higher order statistics can be readily computed in an analogous fashion. One of the most useful features is the calculation of exemplars.

Exemplars. Another fruitful idea relating to the virtual track is to identify a corresponding descriptor for the bundle which actually belongs to the tractography. In other words to find an exemplar or medoid track. Virtual tracks do not necessarily coincide with real tracks as they are just the outcome of large amalgamations. There are many strategies for how to select good exemplars for the bundles. A very fast procedure that we use in this work is to find which real track from the cluster is closest (by MDF distance) to the virtual track. We call this exemplar track e_1 such that $e_1 = \arg \min_{x \in C} \text{MDF}(v, x)$. The computational complexity of finding e_1 is linear in cluster size, and that will be very useful if we have created clusterings with clusters containing more than $\sim 5,000$ tracks (depending on system memory).

A different exemplar can be defined as the most similar track among all tracks in the bundle, which we denote by $e_2 = \arg \min_{x \in C} \sum_{y \in C} \text{MDM}(y, x)$, or

if we want to work with tracks with possibly different numbers of points we could instead use $e_3 = \arg \min_{x \in C} \sum_{y \in C} \text{MAM}(y, x)$. Identification of exemplar tracks of type e_2 and e_3 will be efficient only for small bundles of less than $\sim 5,000$ tracks because we need to calculate all pairwise distances in the bundle. Many applications of the exemplars will be discussed later.

In summary, a virtual (centroid) track is the average of all tracks in the cluster. We call it virtual because it doesn't really exist in the real data set and to distinguish it from exemplar (medoid) tracks which are again descriptors of the cluster but are represented by real tracks.

4.6 Comparisons within- and between-subjects

4.6.1 Comparison of clusterings

We have found rather few systematic ways available in the literature to compare different clustering results for tractographies directly, beyond that of [164] who quantified the agreement between a clustering and a ‘gold standard’ tractography labelled by their team. We have used a more symmetrical measure of agreement between two clusterings that do not require a prior labelled data set. It is called Optimised Matched Agreement (OMA). As with the Adjusted Rand Index [164], OMA requires the calculation of the $M \times N$ cross-classification matrix $X = (x_{ij})$ which counts the number of streamlines in the intersection of all pairs of clusters, one from each of the two clusterings. Here $\mathcal{A} = \{A_i : i = 1 \dots M\}$ and $\mathcal{B} = \{B_j : j = 1 \dots N\}$ are the two clusterings, and $x_{ij} = |A_i \cap B_j|$. As there is no *a priori* correspondence or *matching* between the clusters in \mathcal{A} and those in \mathcal{B} , and vice versa, we need to find one empirically. If $j = \pi(i)$ is such a matching then the matched agreement is $\text{MA}(\pi) = \sum_{i=1}^M x_{i,\pi(i)}$. A matching π that yields OMA by maximising $\text{MA}(\pi)$ can be found using the Hungarian Algorithm [203]. The interpretation of the OMA statistic is analogous to that of the well-known Kappa measure of inter-rater agreement [204], with the range 61% to 80% corresponding to a ‘good’ strength of agreement.

As well as the computational overheads in calculating the cross classification matrix, a further fundamental disadvantage of these methods is that they do not work with clusterings of different tractographies. Being able to compare results of clusterings is crucial for creating stable brain

imaging procedures, and therefore it is necessary to develop a way to compare different tractography clusterings or different sets of streamlines from the same subject or different subjects.

Although we recognise that these are difficult problems, we propose the following approach with three novel comparison functions which we call *coverage*, *overlap* and *bundle adjacency* (BA).

If S and T are sets of streamlines, and $\theta > 0$ is selected as a threshold, we say that $s \in S$ has a θ -neighbour in T if $\min_{t \in T} [\text{MDF}(s, t) < \theta]$. We define the coverage of S by T as the fraction of streamlines in S that have a θ -neighbour in T :

$$\text{coverage}(S, T) = |\{s \in S \text{ has a } \theta\text{-neighbour in } T\}| / |S|.$$

Coverage ranges between 0 (when no streamline in S has a close enough neighbour in T) and 1 (when every streamline in S has a neighbour in T).

We define the overlap of T in S as the average number of θ -neighbours in T for streamlines in S :

$$\text{overlap}(S, T) = \sum_{s \in S} |\{t \in T : t \text{ is a } \theta\text{-neighbour of } s\}| / |S|.$$

Overlap can take any non-negative value, with higher values indicating possible redundancy of T in S ; if T has several similar streamlines then this will tend to boost overlap.

BA is a symmetric measure of the similarity of the two sets of streamlines S and T . BA is the average of the θ -coverages of T by S and of S by T :

$$\text{BA}(S, T) = (\text{coverage}(S, T) + \text{coverage}(T, S)) / 2.$$

BA ranges between 0, when no streamlines of S or T have neighbours in the other set, and 1 when they all do.

If S is a good approximation to T then S will have high coverage of T ; if S has low redundancy as an approximation to T then the overlap of S in T will be low; and if S and T are globally similar then BA will be high. More details on BA are presented in section 4.6.4 and there is a detailed explanation of classification measures like MA in section 4.6.3.

4.6.2 Robustness under reordering

One of the disadvantages of most clustering algorithms is that they give different results with different initial conditions; for example this is recognised with k-means, expectation-maximization [201] and k-centers [202] where it is common practice to try a number of different random initial configurations. The same holds for QB so if there are not distinct clusters such that the distance between any pair of clusters is supra-threshold and the diameter of all clusters is sub-threshold, then with different permutations of the same tractography we will typically see similar number of clusters but different underlying clusters. We will examine the robustness of QB in this respect.

As a first step we recorded the numbers of QB clusters in 20 different random orderings of the tractographies of 10 human subjects. We first removed short streamlines shorter than 40 mm and downsampled the streamlines at 12 points. Then we applied QB with threshold at 10 mm. The mean number of clusters was 2645.9 (min 1937.6; max 3857.8; s.d. 653.8). There is therefore a considerable between-subject variation in this metric. By contrast the within-subject variability of the number of clusters across random orderings is rather small, with mean standard deviation 12.7 (min 7.3; max 17.4). This suggests a good level of consistency in the data reduction achieved by QB.

Next we investigated how consistent QB clusterings are when data sets are re-ordered. Twelve different random orderings were generated for each of 10 tractographies and the corresponding QB clusterings were computed with MDF threshold 10 mm. For each subject the 66\$ pairings of QB clusterings were compared using the optimised matched agreements index and then averaged. Across subjects the mean OMA was 74.1% ($\pm 0.39\%$) which can be interpreted as a good level of agreement [204].

As well as checking that QB created sets of centroids with good coverage and overlap statistics, we went on to show that the performance of QB generalises to sets of streamlines different from the training set, and is superior to a random sample of streamlines. We split each of the 10 tractographies randomly into two halves T_1 (training set) and T_2 (test set). The QB clustering at distance threshold 10 mm was derived for T_1 . Denote by C_1 and c_1 the set of centroids and the number of them. Let R_1 be a random subset of T_1 of size c_1 . Using the measures described in previous section

we found that with distance threshold 10 mm the mean coverage (s.d.) of T_1 by C_1 was 99.96% ($\pm 0.007\%$), of T_2 by C_1 was 99.31% ($\pm 0.08\%$) and of T_2 by R_1 was 90.49% ($\pm 0.41\%$). The mean overlap (s.d.) at this threshold of C_1 in T_1 was 2.44 (± 0.08), of C_1 in T_2 was 2.44 (± 0.08), and of R_1 in T_2 was 5.57 (± 0.50).

The same analyses were performed with QB clusterings for distance threshold 20mm and with distance threshold 20mm. (Note that though we have selected the same values here for the two thresholds they do not have to be the same.) We found that with distance threshold 10mm the mean coverage (s.d.) of T_1 by C_1 was 99.99% ($\pm 0.004\%$), of T_2 by C_1 was 99.91% ($\pm 0.02\%$) and of T_2 by R_1 was 95.86% ($\pm 0.62\%$). The mean overlap (s.d.) at this threshold of C_1 in T_1 was 3.54 (± 0.18), of C_1 in T_2 was 3.54 (± 0.18), and of R_1 in T_2 was 6.53 (± 0.93).

We conclude from these analyses that QB has good coverage and overlap properties with respect to the training set and to the test set of streamlines, while an equivalent random selection of streamlines has worse coverage and overlap. Moreover the performance of QB is better with the lower closeness threshold. The poor performance of random subsets is to be expected as they will oversample in denser parts of the tractography space, and undersample in sparser regions.

4.6.3 Measures to compare classifications

Considerable attention has been paid to measuring the performance of one or more classifiers in the context of supervised learning, see for instance [205]. We now outline some of these metrics before applying them to the comparisons we are interested in. Let $\mathcal{A} = \{A_1, A_2, \dots, A_m\}$ and $\mathcal{B} = \{B_1, B_2, \dots, B_n\}$ be two classifications of N items. Let the number of items in A_i and B_j be a_i and b_j , with t_{ij} items in the intersection $A_i \cap B_j$. There are a number of ways for measuring the similarity or dissimilarity of \mathcal{A} and \mathcal{B} . The first two, Gini Purity and Maximum Likelihood Accuracy, are based on ways we might estimate the \mathcal{A} -labels if we just have the \mathcal{B} -labelling, or vice versa.

Purity. Suppose we have a probability distribution $P=(p_1, p_2, \dots, p_m)$ such that the probability that any item has label i is p_i . Not knowing what this for any item is we apply ‘probability matching’ and randomly estimate a label from the set $\{1, 2, \dots, m\}$ by random selection using the same

distribution P . Then, the probability of assigning the correct label is $\sum p_i^2$; this is the Purity of the distribution. The purity of a distribution lies in the range $[\frac{1}{m}, 1]$. The upper limit occurs when P assigns probability 1 to just one label (i.e. a very pure, concentrated distribution); the lower limit occurs when all m labels have equal probability $\frac{1}{m}$. We now extend this to the case when we have some additional information about the item, namely the label that is assigned to it in a different classification \mathcal{B} .

If $P_{\mathcal{A}|B_j}$ is the observed conditional probability distribution ($p_{i|j} = \frac{t_{ij}}{b_j}$, $i = 1, \dots, m$) of \mathcal{A} given B_j , then we define the Purity of \mathcal{A} with respect to \mathcal{B} as $\text{purity}(\mathcal{A}|\mathcal{B}) = \sum_{j=1}^n \frac{b_j}{N} \text{purity}(P_{\mathcal{A}|B_j})$. In terms of the matrix $T = (t_{ij})$ this is the \mathcal{B} -weighted average of the impurities of the rows of T . We similarly define $\text{purity}(\mathcal{B}|\mathcal{A})$ and it is equal to the \mathcal{A} -weighted average of the impurities of the columns of T . In what follows we will use the symmetrised value $\text{purity}(\mathcal{A}, \mathcal{B}) = [\text{purity}(\mathcal{A}|\mathcal{B}) + \text{purity}(\mathcal{B}|\mathcal{A})]/2$.

Maximum probability matching. Another way to estimate a label for each item is to assign it the label with maximum probability $i_{\max} = \arg \max p_i$. The Random Accuracy in this case is $p_{i_{\max}} = \max_i p_i$. When we do this conditional on the \mathcal{B} -label and average over those labels, we get the Maximum Probability Matching of \mathcal{A} conditional on \mathcal{B} ,

$$\text{MPM}(\mathcal{A}|\mathcal{B}) = \sum_{j=1}^n \frac{b_j}{N} \max_i p_{i|\mathcal{B}_j}.$$

We define $\text{MPM}(\mathcal{B}|\mathcal{A})$ similarly, $\text{MPM}(\mathcal{B}|\mathcal{A}) = \sum_{i=1}^m \frac{a_i}{N} \max_j p_{j|\mathcal{A}_i}$. A further simplification is to use the symmetrized value

$$\text{MPM}(\mathcal{A}, \mathcal{B}) = [\text{MPM}(\mathcal{A}|\mathcal{B}) + \text{MPM}(\mathcal{B}|\mathcal{A})]/2.$$

Correctness and completeness (splitting and lumping pairs of items). For the next two metrics the focus moves to comparison of the labels assigned by \mathcal{A} and \mathcal{B} to pairs of items. Differences in the partitions \mathcal{A} and \mathcal{B} are reflected in two ways. Items assigned the same label by \mathcal{A} are said to be split by \mathcal{B} if their \mathcal{B} -labels are not equal; alternatively items assigned different \mathcal{A} -labels are said to be lumped by \mathcal{B} if they are assigned the same \mathcal{B} -label. Note that what is lumped (split) by \mathcal{B} will equally be lumped by \mathcal{A} .

The total number of pairs from N items is $\text{pairs}(\mathcal{A}) = \binom{N}{2} = \frac{N(N-1)}{2}$.

The number of pairs assigned the same \mathcal{A} -labels is together(\mathcal{A}) = $\sum_{i=1}^m \binom{a_i}{2}$.

The number of pairs assigned different labels is apart(\mathcal{A}) = $\text{pairs}(\mathcal{A}) - \text{together}(\mathcal{A})$. This can also be written as $\sum_{1 \leq i \neq i' \leq m} a_i a_{i'}$ which in turn can

be expressed in terms of the cumulative sum of (a_i) which is an efficient way of programming these calculation of sums of all products with unequal subscripts. The number of \mathcal{A} -pairs split by \mathcal{B} is

$$\text{split}(\mathcal{A}|\mathcal{B}) = \sum_{i=1}^m \left(\sum_{1 \leq j \neq j' \leq n} n_{ij} n_{ij'} \right) = \text{lumped}(\mathcal{B}|\mathcal{A}).$$

Similarly,

$$\text{lumped}(\mathcal{A}|\mathcal{B}) = \sum_{j=1}^n \left(\sum_{1 \leq i \neq i' \leq m} n_{ij} n_{i'j} \right) = \text{split}(\mathcal{B}|\mathcal{A}).$$

Completeness and Correctness are defined in terms of these quantities:

$$\text{completeness}(\mathcal{A}|\mathcal{B}) = 1 - \text{split}(\mathcal{A}|\mathcal{B}) / \text{together}(\mathcal{A})$$

and

$$\text{correctness}(\mathcal{A}|\mathcal{B}) = 1 - \text{lumped}(\mathcal{A}|\mathcal{B}) / \text{apart}(\mathcal{A}).$$

Symmetrized measures of completeness and correctness for \mathcal{A} and \mathcal{B} are defined as

$$\text{completeness}(\mathcal{A}, \mathcal{B}) = [\text{completeness}(\mathcal{A}|\mathcal{B}) + \text{completeness}(\mathcal{B}|\mathcal{A})]/2$$

$$\text{correctness}(\mathcal{A}, \mathcal{B}) = [\text{correctness}(\mathcal{A}|\mathcal{B}) + \text{correctness}(\mathcal{B}|\mathcal{A})]/2.$$

For the clusterings encountered in tractography, the number of apart pairs in \mathcal{A} is very high, and only a small percentage (e.g. 0.5%) of these pairs will be lumped by \mathcal{B} . This is because the average cluster size is small by comparison with the number of clusters. As a consequence, the correctness measure is not a particularly useful metric. By contrast, the number of together pairs is modest, and the completeness measure is more sensitive.

Maximum Agreement (κ_{\max}). Our fifth metric is Cohen's κ , which is a well-known measure of agreement between raters on the assignment of

a set of items to a shared classification scheme. It adjusts the agreements (items on which the raters agree) for the number of agreements that might have occurred by chance:

$$\kappa = \frac{p_{\text{agreement}} - p_{\text{chance agreement}}}{1 - p_{\text{chance agreement}}}.$$

This can be simply represented in terms of the overlap matrix $T = (t_{ij})$ by the formula:

$$\kappa(T) = \frac{\sum_{i=1}^M t_{ii}/N - \sum_{i=1}^M r_i c_i / N^2}{1 - \sum_{i=1}^M r_i c_i / N^2},$$

where r_i and c_j represent the row and column totals of T . We have extended T to a square matrix of size $M = \max(m, n)$ by adding, if necessary, rows or columns of zeros. When we adapt this measure to the case of comparing two clusterings we further need to take into account the lack of prior correspondence between the two sets of labels. The κ_{\max} statistic is the result of maximising κ over all possible correspondences:

$$\kappa_{\max} = \max_{\pi} \kappa(T_{\pi}) = \frac{\sum_{i=1}^M t_{i\pi(i)}/N - \sum_{i=1}^M r_i c_{\pi(i)}/N^2}{1 - \sum_{i=1}^M r_i c_{\pi(i)}/N^2},$$

where T_{π} is the matrix T with columns reordered by a permutation π . The principal trouble with the κ_{\max} statistic is that its computation is $O(N!)$ if all permutations are tried. One way out to overcome the problem caused by the size of the search set might be to use a randomised search strategy for instance based on a simulated annealing approach.

Matched Agreement via the Hungarian Method. An alternative is to look for a simpler quantity that might be optimised. One obvious choice is the maximized number of agreements $\mu_{\max} = \sum_{i=1}^M t_{i\pi(i)}$ corresponding to the permutation π ; this is the leading term in the numerator of κ_{\max} . Maximizing the number of agreements amongst all permutations π is a classical combinatorial optimization problem (weighted assignment problem on a bipartite graph) that can be reformulated as a linear programming

Metric	Purity	MPM	Comp	Corr	MA	MK
Mean	70.8	79.2	65.5	99.9	74.1	74.0
Mean S.D	0.51	0.37	1.11	0.02	0.39	0.39

Table 4.2: Mean and mean standard deviation of six classification comparison metrics for 10 different tractographies: Purity, Maximum Probability Matching (MPM), Completeness (Comp), Correctness (Corr), Matched Agreement (MA) and Matched Kappa (MK). For each of 10 tractographies the 66 pairings of QB clusterings for 12 different orderings were evaluated. All are represented as percentages (%). Matched agreements use the Hungarian Algorithm to create a mapping between each pair of clusters; matched kappa evaluates Cohen’s kappa using this same optimal mapping.

problem whose efficient solution by the Hungarian Method [203] is well known.

We have tested various published implementations of the version by Lawler [206] of the Hungarian Method and have found that the one by Carpaneto et al. [207], implemented by them [208] in FORTRAN, is both fast and capable of handling assignment problems of unlimited size.

We calculated the average of each of these comparison metrics for QB clusterings of 12 different orderings for each of 10 tractographies (see Tab. 4.2). A number of observations are worth making. Matched agreement and matched kappa are essentially the same metric (correlation 0.97). Of these two metrics we prefer matched agreement because is both simpler to calculate and understand. Correctness, for the reasons discussed above, is too insensitive to be of use. We would therefore suggest, and on the basis of the mean of the standard deviations across pairings, that maximum probability matching, and matched agreement are suitable metrics for evaluating tractography clusterings. It is also worth noting that maximum probability matching is a simple first approximation to the optimal matching identified by the Hungarian method although it is not necessarily one-to-one.

We have noticed that these metrics are all costly to calculate in terms of time and memory requirements. Therefore, they will not be used further in this study. We instead look at ways to compare clusterings of tractographies that will work when comparing different tractographies either for the same or different subjects. These need to be based on metrics for distances between tracks, whether virtual tracks, exemplar tracks or raw

tracks from the original tractographies. This is the subject of the next section.

4.6.4 Bundle Adjacency

We have found rather few systematic ways to compare different clustering results for tractographies in the literature [164]. Being able to compare results of clusterings is crucial for creating stable brain imaging procedures. It is therefore necessary to develop a way to compare different clusterings of the same subject or different subjects. Although this is a difficult problem, we propose the following solution with a metric which we call bundle adjacency (BA). BA works as follows: let us assume that we have gathered the exemplar tracks from clustering A in $E_A = \{e_1, \dots, e_{|E_A|}\}$ and from clustering B in $E_B = \{e'_1, \dots, e'_{|E_B|}\}$ where $|E|$ denotes the number of exemplar tracks of each clustering E . The size of set E_A does not need to be the same as that of E_B (i.e. both $|E_A| \neq |E_B|$ and $|E_A| = |E_B|$ are acceptable). Next, we calculate all pairwise MDF distances between the two sets and store them in rectangular matrix D_{AB} . The minima of the rows of D_{AB} provide the distance to the nearest track in B of each track in A ($E_{A \rightarrow B}$) and similarly the minima of the columns of D_{AB} the distance to the nearest track in A of each track in B ($E_{B \rightarrow A}$). From these correspondences we only keep those distances that are smaller than a tight threshold θ . Then we define BA (Bundle Adjacency) to be

$$BA = \frac{1}{2} \left(\frac{|E_{A \rightarrow B} \leq \theta|}{|E_A|} + \frac{|E_{B \rightarrow A} \leq \theta|}{|E_B|} \right) \quad (4.7)$$

where $|E_{A \rightarrow B} \leq \theta|$ denotes the number of exemplars from A which had a neighbour in B that is closer than θ and similarly for $|E_{B \rightarrow A} \leq \theta|$ the number of exemplars from B to A which their distance was smaller than θ (see a similar definition of BA in section 4.6.1). In other words, BA is the mean of the fraction of row minima of D_{AB} that are less than θ and the fraction of column minima less than θ . When $BA = 0$ every exemplar from one set was further than θ to all exemplars in the other set. When $BA = 1$ all exemplars from one set had a θ -close neighbour in the other set. This metric is extremely useful especially when comparing tractographies from different subjects because it does not require $|E_A| = |E_B|$ which was a requirement with the metrics proposed in the previous section.

We ran an experiment where we compared BA between pairs of 10 subjects with their tractographies warped in MNI space (see section 4.4). This generated $\binom{10}{2} = 45$ BA values with $\theta = 10$ mm. We performed this experiment twice; first by only keeping the bundles with more than 10 tracks (BA10) and secondly by only keeping the bundles with more than 100 tracks (BA100). The average value for BA10 was 47% and standard deviation 2.6%. As expected BA100 (bigger landmarks) performed better with average value of 53% and standard deviation 4.9%. The difference between BA10 and BA100 is highly significant: Student's $t = 4.692$, $df=88$, $p = 1.97 \times 10^{-5}$, two-sided; and, as a precaution against non-normality of the underlying distributions, Mann-Whitney $U = 530.$, $p = 5.65 \times 10^{-5}$. If we think that the small bundles of size < 100 are more idiosyncratic or possibly more likely to reflect noise in the data, whereas larger bundles are more indicative of substantial structures and landmarks in the tractographies, then we are encouraged to see that on average the virtual tracks of 50% of larger bundles of each tractography lie within 10 mm of those of the other tractographies. This supports the notion that QB can be used to find agreements between different brains by concentrating on the larger (more important) clusters. Further evidence of this is discussed in section 4.8.2.

4.7 Parallel version

4.7.1 Algorithm

QB is a very fast algorithm; however we wanted to make it even more efficient so that for example it is trivial to cluster hundreds of subjects together and use many CPUs or computers simultaneously. This could be used to create an atlas of hundreds of subjects in a few minutes. Therefore, we have extended QB to a parallel version which we call pQuickBundles (pQB). This algorithm works as follows. We first redirect and downsample all tracks. Then we put all tracks together and break them into subsets. For every subset we assign a new thread and set QB to run on that thread. We have therefore many QBs running on different CPUs. Then we collect all individual clusterings and start merging them together. We can pair every two results together and merge them in a binary fashion or just merge all clusterings to the first clustering. We can do merging with many different ways. We present here the most modest but useful attempt.

4.7.2 Merging two sets of bundles

We can merge bundles using exemplar tracks or virtual tracks. We first set a distance threshold θ usually the same as the one we used for the QBs in the previous step. Let's assume now that we have gathered the virtual tracks from clustering A in $V_A = \{v_1, \dots, v_{|V_A|}\}$ and from clustering B in $V_B = \{v'_1, \dots, v'_{|V_B|}\}$ where $|V|$ denotes the number of virtual tracks of each clustering. $|V_A|$ can be different $|V_B|$. (a) For every v'_i in set V_B we find the closest v_j in set V_A and store the distance between these two tracks. Therefore we now have a set of minimum distances from V_B to V_A . The size of this set is equal to $|V_B|$. (b) Finally, we merge those clusters from B whose virtual tracks have minimum distances smaller than θ into the corresponding clusters of A , and if a virtual track in V_B has no sub-threshold neighbour in V_A then its cluster becomes a new cluster in the merged clustering. In that way clusters from the two sets who have very similar features will merge together. If not, new clusters will be created. Using this approach, no information loss will occur from the merge of the two sets of clusters.

4.8 Direct applications

We found that QB has numerous applications from detecting erroneous tracks to creating atlases, finding landmarks and guiding registration algorithms. Here we present just a few of the strategies that can be further pursued.

4.8.1 Rapidly detecting erroneous tracks

It is well known that there are different artifacts seen in tractographies caused by subject motion, poor voxel reconstruction, incorrect tracking and many other reasons. There is no known automatic method to detect these tracks and therefore remove them from the data sets. The idea here is to use QB to speed up the search for erroneous tracks. We will concentrate on tracks that loop one or many times; something that it is considered impossible to happen in nature.

Tracks most likely to be erroneous are those which wind more than one time, like a spiral. We can detect those with the following approach: let us assume that we have a track s and we want to check if it winds: (a) we

perform a singular value decomposition on the centered track $U, \mathbf{d}, V = \text{SVD}(s - \bar{s})$; (b) project the highest singular value \mathbf{d}_0 to the first column of U , U_0 creating the first component of a two dimensional coordinate p_x and the second highest \mathbf{d}_1 to the second column U_1 creating the second coordinate p_y ; and (c) calculate the cumulative winding angle on the 2D plane; d) if the cumulative angle is more than 400° it would mean that the initial track s is winding and therefore needs to be removed (see Fig. 4.7).

Winding tracks can be dangerous when we merge clusters because they could be close to many different clusters of different shape simultaneously. We found that winding tracks often form bundles with many similar tracks. As these are usually long tracks, they will not be removed by filters which remove short tracks. In Fig. 4.8 we show an example where 161 erroneous bundles were automatically detected by our winding method. They all had total winding angle higher than 500° . To cluster the initial tractography not shown here we used QB with threshold 10 mm. This is the first known automatic detection system of outliers and erroneous tracks for tractography data based on more advanced shape characteristics that go beyond simple track length filtering. By calculating the number of winding tracks in the data sets over the total number of tracks we could have an indicator of the quality of the data sets.

We can use QB with a low threshold to reduce the number of tracks while avoiding embedding winding tracks into otherwise ordinary clusters and then run the winding algorithm just on the exemplar tracks of the bundles rather than the entire tractography.

QB can also simplify detection of tracks which are very dissimilar to others and therefore very distant from all other clusters. Usually, when we use a QB threshold of about 10 mm, the tracks will be part of small bundles containing a few tracks and the distance of the bundle they belong to from all other bundles will be much higher than average. This can give us another detection method for outliers. We could find for example which bundles are most distant from all other bundles and remove them from the data sets.

Finally, QB can be used to remove small or broken tracks in an interactive way, for example see Fig. 4.3 where the red large bundle has been merged by an expert and then with QB we can extract the skeleton of the bundle and see which parts create that structure. Without QB it would be too difficult to work out that this bundle consists of many small or diver-

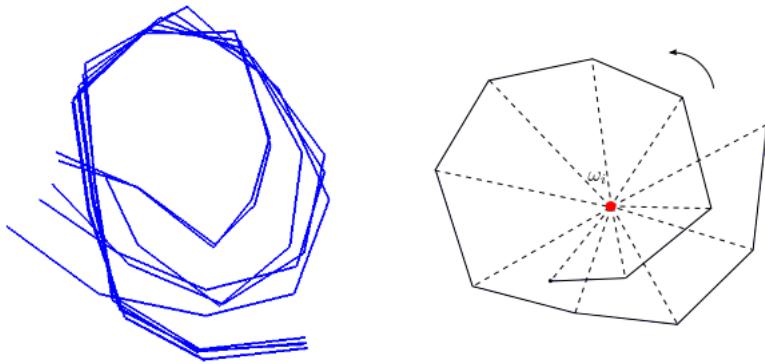


Figure 4.7: Example of detecting a possibly erroneous 3D bundle (on the left) by projecting its exemplar track and counting the winding cumulative angle $\sum_0^N \omega_i$ on the 2D plane as shown on the right, where N is the total number of track segments. Usually bundles with total angle higher than 400° are removed from the data sets as most likely to be erroneous.

gent parts. In this figure both very diverging, small or broken tracks can be identified after the simplification provided by QB.

In summary, we have shown that QB can facilitate a fully automatic, efficient and robust detection system for erroneous tracks in specific bundles or entire tractographies.

4.8.2 Alignments, landmarks and atlases

We have used QB to construct a robust tractographic atlas in MNI space from 10 subjects' data sets. Here we explain the steps we used to achieve that.

Alignment. Tractographies were created using EuDX as described in section 4.4 (see section 1.3 for acquisition details). The tractographies for all subjects were initially in native space and the goal was to warp them in MNI space, using nonlinear registration.

Because the registration of tractographies is generally considered a difficult problem with a non-unique solution we wanted to make sure we are using a known, well established and robust method. We chose therefore, to use fnirt with the same parameters as used with the first steps of TBSS [209]. For that reason, FA volumes were generated from the same data sets using Tensor fitting with weighted least squares after skull stripping with bet and parameters -F -f .2 -g 0. These FA volumes were again in native space therefore we needed to warp them in MNI space.

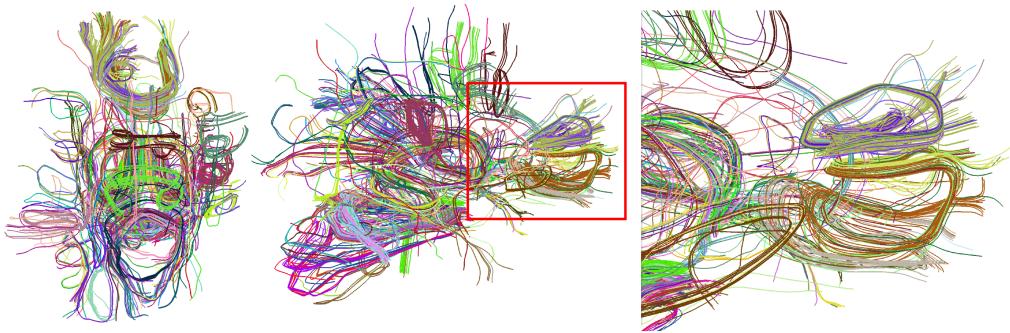


Figure 4.8: Example with erroneous tracks detected on real data sets. Left: the erroneous bundles on their exact position in the data set from the top of the head, Middle: the same from the saggital view. Right: the area surrounded by the red box from the middle slightly rotated and zoomed. The colour encodes different bundle label.

For this purpose, a standard FA template (FMRIB58) from the FSL toolbox was used as the reference volume. However, we wanted primarily to have the displacements which would do a point wise mapping from native space to MNI space and we found this to be technically very difficult with the FSL tools as they assume that these displacements will be applied only on volumetric data and not with point data as those used in tractographies. Finally, after some considerable effort we found a combination of `flirt`, `invwarp`, `fnirtfileutils` and `fnirtfileutils -withaff` which gave us the correct displacements. The code is available in module (`dipy.external.fsl`). It is also important to say that we did not use eddy correction with any of this type of data sets. Eddy correction is unstable with volumes at high b-values because there is not enough signal for guiding a correct registration with the other volumes at lower b-values. It is like trying to match two figures that have no similarities at all. The matching will be certainly poor and error prone.

After creating the displacements for every subject; these were applied to all tractographies in the native space so they are mapped in the MNI space of voxel size $1 \times 1 \times 1 \text{ mm}^3$. Having all tractographies in MNI space is something very useful because we can now compare them against available templates or against each other and calculate different statistics. However this is not where we stop; we proceed to generate a tractographic atlas using QB clusterings.

Tractographic Atlas. For all subjects: (a) load warped tractography, (b) downsample the tracks to have only 12 points, (c) calculate and store QB

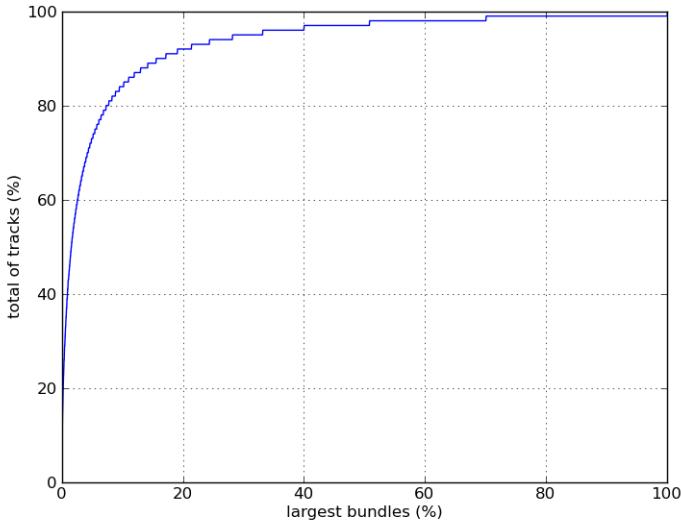


Figure 4.9: 14,520 clusters created by joining the QB clusterings of 10 subjects in MNI space. Most clusters had a few tracks and only few had many. 20% of the largest clusters had more than 90% of the total amount of tracks. The agreement between different subjects which would be useful for a solid atlas with the biggest bundles becoming landmark bundles and the small bundles removed as outliers.

clustering with a 10 mm threshold, (d) merge all clusterings with 10 mm threshold as explained in section 4.7 (merging). When creating an atlas by merging many different subjects the most important issue is what you remove from the atlas as outliers. QB here provides a possible solution for this problem. If we plot the number of tracks for each cluster sorted in ascending order we can see an interesting pattern (see Fig. 4.9). In this diagram we observe that 20% of the largest clusters had more than 90% of the total amount of tracks. This shows that there is much agreement between the biggest bundles of different subjects. We will use this property to create a solid atlas in which we keep the biggest bundles (landmarks) and remove the smallest bundles (outliers).

Finding and Using Landmarks. One can use this atlas or similar atlases created from more subjects in order to select specific structures and study these structures directly in different subjects without using any of the standard ROI based methods.

A simple example is given in Fig. 4.10. In the first row we see a tractographic atlas joined by merging the QB clusterings of 10 healthy subjects as described in the previous section. From these clusters, repre-

sented by their virtual tracks we only keep 196 biggest clusters i.e. those which contain the highest number of tracks, so that we are sure there is enough agreement between the different tractographies. From these we just pick by way of an example 19 virtual tracks which correspond to well known bundle structures in the literature: 1 from Genu of Corpus Callosum (GCC), 3 from the Body of Corpus Callosum (BCC), 1 from the Spleenium (SCC), 1 from the Pons Cerebellar Peduncle (CP), 1 from left Arcuate Fasciculus (ARC-L), 1 from right Arcuate Fasciculus (ARC-R), 1 from left Inferior Occipitofrontal Fasciculus (IFO-L) and 1 from right Inferior Occipitofrontal Fasciculus (IFO-R), 1 from right Fornix (FX-R), 1 from left Fornix (FX-L), 1 from the Optic Radiation (OR), 1 left Cingulum (CGC-L), 1 from right Cingulum (CGC-R), 1 from left Corticospinal tract (CST-L), 1 from right Corticospinal tract (CST-R), 1 from left Uncinate (UNC-L) and 1 from right Uncinate (UNC-R). These 19 tracks are coloured randomly. On the second row we show, for the first 6 of these selected representative tracks, the tracks closer than 20 mm from 3 arbitrarily selected subjects. Similarly, on the third row the tracks closer than 15 mm to the next 7 selected tracks. Finally, on the last row, we bring the tracks from the same 3 subjects which are closer than 18 mm. The colours used for the selected tracks are automatically assigned from the colours of tracks picked from the atlas. We can see significant reliability and continuity both within and between subjects even though we have only selected a very small number of representative tracks. Using a similar procedure we could create a book of bundles for every subject and then compare the subjects at the level of bundles.

4.8.3 QB as input to other learning methods

We found that QB is of great value as an adjunct to many less efficient algorithms e.g. hierarchical clustering, affinity propagation, nearest neighbours, spectral clustering and other unsupervised and supervised learning methods. We present here one example with QB as input to affinity propagation and one with QB as input to hierarchical clustering.

Most clustering algorithms need to calculate all pairwise distances between tracks; meaning that for a medium sized tractography of 250,000 tracks we would need 232 GBytes of RAM with single floating point precision. Something which is not and will not be available soon in personal

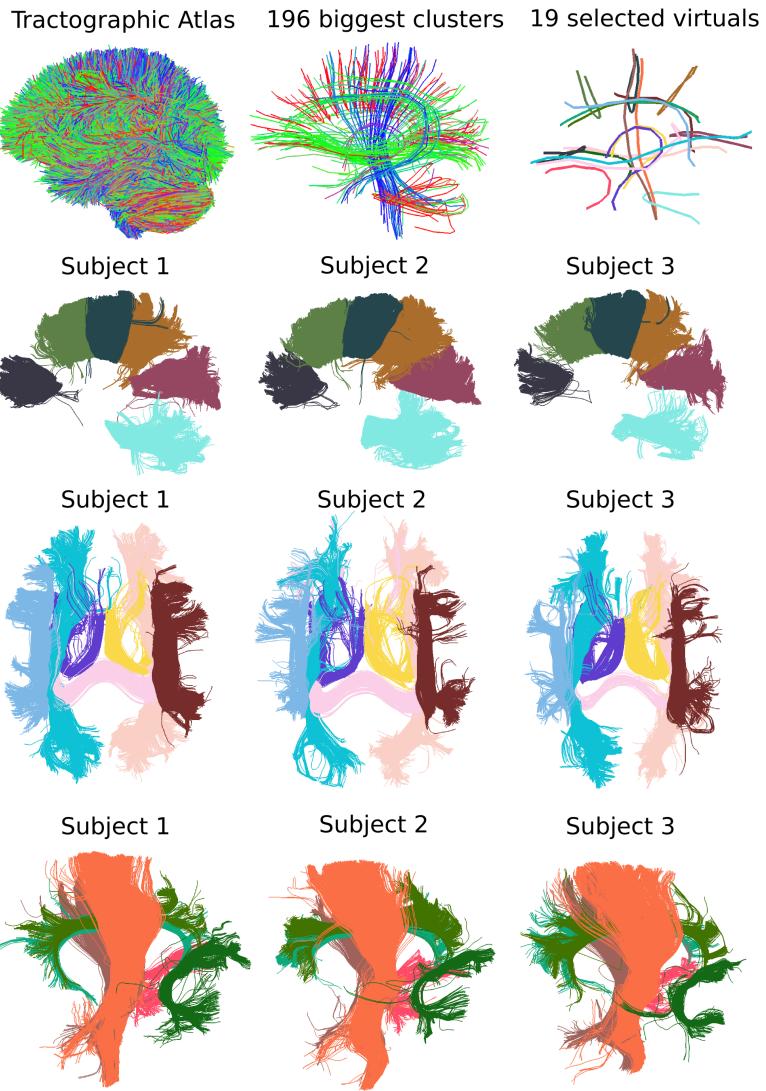


Figure 4.10: A novel way to do comparisons between subjects. Correspondence between different subjects (last 3 rows) and a few landmarks picked from the tractographic atlas generated by merging QB clusterings of 10 subjects (top row). The fact, there is such a level of agreement and continuity on the last 3 rows from such a few skeletal tracks offers a great prospect for implementing new robust ways of statistical comparisons using tractographic data sets.

computers. A naive solution would be to use sparse matrices to approximate the distance matrix; however tractographies are densely packed and produce very dense distance matrices. Therefore, this is not a viable solution. The straightforward solution to this problem is to use QB in order to first segment in small clusters and then use the representatives (i.e. exemplar or virtual tracks) of these clusters with other higher complexity

operations and merge the clusters together in bigger clusters. More precisely we propose to:

1. Cluster using QB as explained in section 4.8.2.
2. Gather virtual tracks.
3. Calculate MDF distance of virtual tracks with themselves.
4. Use any other clustering method to segment this much smaller distance matrix D .

In Fig. 4.11 at the left panel we show a result where we used hierarchical clustering with single linkage for step (4) with a threshold of 20 mm using the package `hcluster` [210]. A known drawback of single linkage is the so-called chaining phenomenon: clusters may be brought together due to single elements being close to each other, even though many of the elements in each cluster may be very distant to each other. Chaining is usually considered as a disadvantage as it is driven by local neighbours. Nevertheless, we can use this property to cluster the corpus callosum (CC) all together (shown with dark red in left top of Fig. 4.11) creating a fully automatic CC detection system. Furthermore, we can use different cutting thresholds on the underlying dendrogram to amalgamate together different structures e.g. see the cingulum bundles in the same panel.

In the right panel of Fig. 4.11 we see the implementation of step (4) using a more recent algorithm: affinity propagation (AP) [211], which was earlier identified by us and [192] for being impossible to be used for group analysis or to cluster entire tractographies of many thousands of tracks. A small outline of how this algorithm works is given in section A.6. In the bottom right panel of Fig. 4.11, we observe how nicely AP, after the simplification provided by QB, has clustered Arcuate, Longitudinal Occipitofrontal Fasciculus and other structures known from the literature. The input of AP was the negative distance matrix $-D$, the preference weights were set to matrix `median($-D$)` and the hierarchical clustering parameter was set to 20 mm.

For hierarchical clustering parts we used the software `hcluster` and for affinity propagation we used the library `scikit-learn`. They are both implemented in Python.

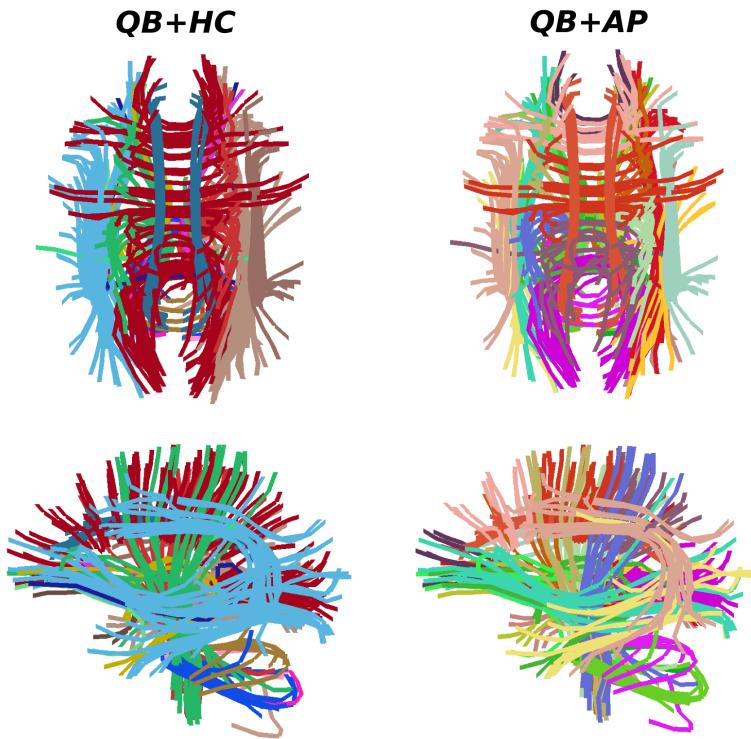


Figure 4.11: Two examples where QB output is used to cluster an entire set of 10 tractographies together and then the result is given as input to hierarchical clustering (HC) using single linkage on the left and to affinity propagation (AP) on the right. Colours encode cluster labels. On the left side we see 19 clusters and on the right 23. QB facilitates significantly the operation of the other two algorithms which would not be able to cluster the entire data sets on current computers. Pay attention at the top left panel where QB+HC have managed to cluster the entire CC as one bundle.

4.8.4 Exemplars vs ROIs vs Masks

Medical practitioners and neuroanatomists often argue that when they use multiple spherical or rectangular masks to select some bundles many tracks are thrown away because they are small and the mask operations cannot get hold of them. Our method provides a solution to this problem as it can identify broken or smaller bundles inside other bigger bundles which are otherwise very difficult or even sometimes impossible to identify visually or with the use of masks. Our method attacks this problem and suggests a very efficient and robust solution which sets the limit for unsupervised clustering of tractographies and facilitates tractography exploration and interpretation. One can now use exemplar tracks as access

points into the full tractography and with a single click on that exemplar track obtain the entire bundle. Therefore, a super-bundle can be created just with a few clicks, based on a selection from exemplar tracks.

In order to create this system we implemented a 3D visualization and interaction system for tractographies based on QB in Python and OpenGL. This project is available online at f os . me.

4.9 Direct Tractography Registration

Direct tractography registration is a recently described problem with only a small number of publications, and as far as we know there are no publicly available solutions. By direct registration we mean that no other information apart from the tractographies themselves is used to guide the registration. This is in contrast to the previous sections where we used FA registration mappings applied to tractographies (see section 4.8.2) which is also most commonly used in the literature along with other Tensor based methods [212].

The current described methodologies on this subject are as follows. Leemans et al. [213] uses the invariance of curvature and torsion under rigid registration along with Procrustes analysis to co-register together different tractographies. Mayer et al. [214] used iterative closest point applied to register pre-selected bundles (bundles of interest - BOI) , [215] and extended it using probabilistic boosting tree classifiers for bundle segmentation in [216]. Durrleman et al. [190] reformulated the tracks as currents and implemented a currents based registration. Zvitia et al. [172], [173], used adaptive mean shift clustering to extract a number of representative fibre-modes. Each fibre mode was assigned to a multivariate Gaussian distribution according to its population thereby leading to a Gaussian Mixture model (GMM) representation for the entire set of fibres. The registration between two fibre sets was treated as the alignment of two GMMs and is performed by maximizing their correlation ratio. A further refinement was added using RANSAC [217] to obtain all 12 affine parameters. Ziyan et al. [218] developed a nonlinear registration algorithm based on the log-Euclidean polyaffine framework [219]. However, we will not classify this approach as a direct tractography registration algorithm as the authors first created scalar volumes from the tracks and next warped the volumes. Therefore, they did not register the tracks in a straight fashion

in their space.

We now describe our algorithm and show that it is efficient and simple to use. In addition, it is completely automatic and provides an evidently robust direct rigid tractography registration algorithm available in seconds. This algorithm could be of great use when comparing healthy versus severely diseased brains e.g. stroke or vegetative state patients when non-rigid registration is not recommended because of severe asymmetries in the diseased brains. The algorithm is based on the robustness of QB to find good representative descriptors.

Here we describe a simple algorithm where 2 tractographies T_A, T_B are brought into alignment in native space. The main steps of this approach are:

1. All tracks with length smaller than 100 mm and longer than 300 mm are removed from the data sets. This reduces the size of tractography to about 1/4 of its initial size (200,000 tracks). (This filtering may have different effects depending on brain size. We have not investigated this question at present.)
2. Both tractographies are equidistantly downsampled so every track contains only 12 points.
3. We run QB with distance threshold at 10 mm for both tractographies.
4. Collect all exemplar tracks from clusters containing more than 0.2% of total number of tracks. Let us assume we have these in E_A and E_B .
5. Calculate all pairwise distances $D = \text{MDF}(E_A, E_B)$ and save them in rectangular matrix D .
6. Create a cost function (optimizer) which will try to minimize the symmetric minimum distance $\text{SMD} = \sum_i \min_j D(i, j) + \sum_j \min_i D(i, j)$.
7. Use modified Powell's method [220] to minimize SMD over rigid rotations of E_B starting with zeroed initial conditions. At each iteration of the optimization, E_B will be transformed by a rigid rotation and SMD will be recalculated. To ensure smooth rotations we use the Rodriguez rotation formula.



Figure 4.12: Two tractographies from different subjects before (left) and after rigid registration (right) using our method.

In Fig. 4.12 we see the result of this algorithm applied to two tractographies – represented with their exemplar tracks – depicted with orange and purple. We can see in the left panel that the orange tractography is misaligned with respect to the purple one, and in the right panel we see their improved alignment after applying our algorithm.

Metric. SMD is proposed here for registration of trajectory data sets, but one could equally use mutual information [221] or the correlation ratio [222] for registration of volumetric data sets. Nonetheless, the advantage of SMD is that it comes from robust landmarks generated by QB which bring together local and global components. Initially, it was not clear if we should use SMD or just the sum of all distances $SD = \sum_{i,j} D(i,j)$. Therefore, we performed an experiment to validate the smoothness and convexity of these two cost functions. We plotted both functions under a single-axis translation or a single-angle rotation of the same tractography as show in Fig. 4.13. From these two diagrams we can see, that although for translations only the SD was entirely convex, with rotations, the SD had stronger local minima which is not a good property for registration. Furthermore, the SMD had steeper gradients towards the global minimum which is a positive indicator for faster convergence.

Experiments. The first large scale experiment took place using the same tractography of a single individual copied and transformed 1,000 times with range of all three angles from -45° to 45° and range of all

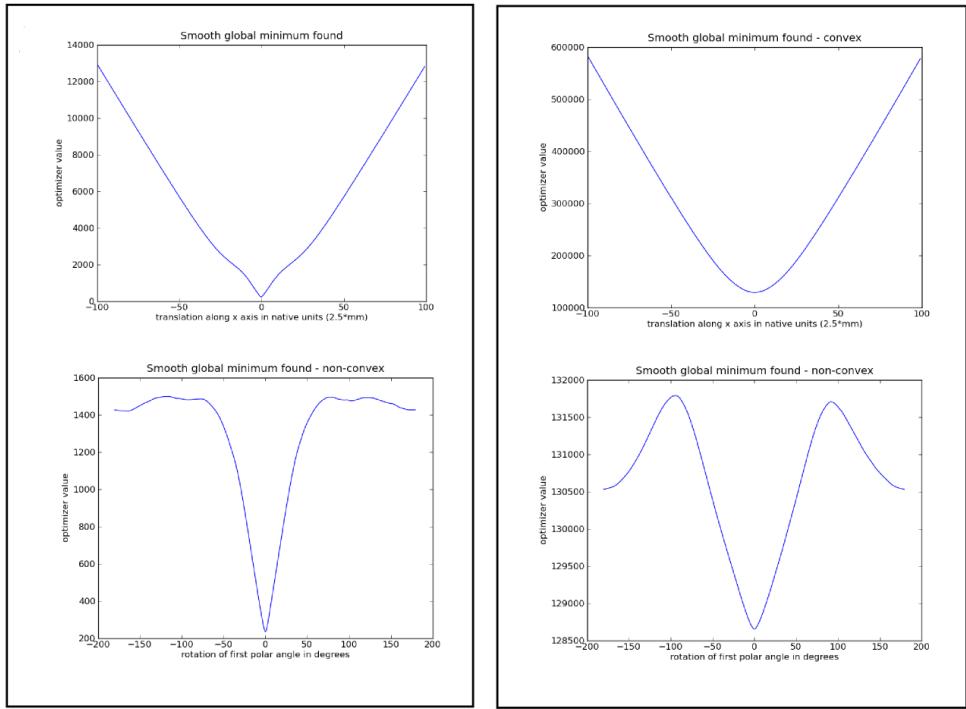


Figure 4.13: Left: The metric *SMD* that we chose to optimize for two copies of the same tractography with the second copy translated (above) and rotated (below). This metric appears to be smooth with a single global minimum and is only slightly non-convex with small local minima. Right: Another possible candidate metric was the *SD*. Although more convex on translations it had stronger local minima with rotations.

x, y, z translations from -113 to 113 mm. Then we registered all transformed tractographies to the static one and calculated all pairwise MDF distances storing them in a square matrix D . We would expect that if the registration was correct then the sum of all diagonal elements of D would be close to 0. This was confirmed with both cost functions used SD and SMD getting close to zero 99.8% of the time; however, SMD was always closer to perfect alignment than SD, having precision of more than 7 decimals. Consequently we chose SMD as a better cost function for direct tractography registration.

We used GQI-based tractographies from 10 subjects and we registered all combinations of pairs $\binom{10}{2} = 45$. Comparing different tractographies is not a trivial problem however, we can use the bundle adjacency (BA) metric explained in section 4.6.4. We are happy to report the mean initial BA was $34.8\% \pm 8.0\%$ and the mean final BA after applying our direct registration method was $48.1\% \pm 6.1\%$. This was a statistically highly sig-

nificant improvement ($t_{\text{paired}}(44) = 11.2$, $p \leq 10^{-13}$). We are planning in the future to compare this registration method against other standard methods which are common in the literature.

4.10 Bundle Quality Control

In many parts of this document we did not consider short tracks. That is perfectly valid because (a) the longer tracks are more likely to be used as useful landmarks when comparing or registering different subjects because it is more likely for them to exist in most subjects, (b) removing short tracks facilitates the usage of distance based clustering (no need for manually setting the distance threshold) and interaction with the tractography, (c) someone would first want to see the overall representation of the tractography and go to the details later. Nonetheless, after having clustered the longer tracks there are many ways to assign the smaller bundles to their closest longer bundles. For this purpose, we recommend the use of different distance from MDF for example the minimum version of MAM referred to as MAM_{\min} (see Eq. 4.4).

Some simple strategies for clustering short fibres are discussed. The first is for unsupervised clustering and the second one is for supervised learning.

1. Cluster the long tracks using QB with distance threshold at 10 mm and then cluster the short tracks (<100 mm) to a lower threshold and assign them to their closest long track bundle from the first clustering using the MAM_{\min} distance.

2. Read the tractography of a single subject, use a tractographic atlas as the one created in section 4.8.2 and pick one or more representative tracks from that atlas. Then, find the closest tracks from the subject to that selected tracks using MDF. Cluster the closest tracks found from the previous step and for each one of these new skeletons find the closest tracks using MAM_{\min} distance. We should now have an amalgamation of shorter and longer fibres in one cluster.

An example of this second strategy is shown in Fig. 4.14. First we selected a single track from Arcuate Fasciculus. Next, we brought all tracks closer than 15 mm using the MDF distance. Then, we cluster the last tracks to 23 virtuals using QB with $\theta = 6.25$ mm. Finally, we bring all tracks with 6 mm (MAM_{\min} distance) from the entire tractography. Using this simple

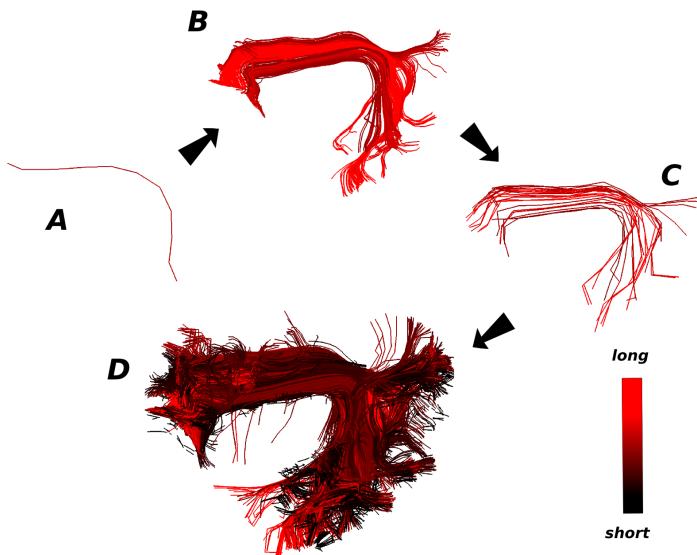


Figure 4.14: A simple and vigorous strategy for handling short and long tracks together by picking a track of interest from one of our atlases. Colourmap encodes track length. A: one selected atlas track, B: 245 subject tracks closer than 15 mm (MDF distance), C: B tracks clustered in 23 virtuals, D: 3,421 tracks closer than 6 mm (MAM distance) from the representatives of B are shown. A great number of short tracks have been brought together along with the tracks in B. In that way we managed to bring together an entire bundle consisting both of long and short fibres by just selecting one track.

strategy we were able to bring together from the entire data set and with minimum effort a bundle that consists of many shorter and longer tracks.

4.11 Discussion and conclusion

In this chapter we presented a novel and powerful algorithm – QuickBundles (QB). This algorithm provides simplifications to the old problem of white matter anatomy packing which has recently attracted much scientific attention; it can also be used for any trajectory clustering problem and it is recommended when large data sets are involved. QB can be used with all types of diffusion MRI tractographies which generate streamlines (e.g. probabilistic or deterministic) and it is independent of the reconstruction model.

In common with mainstream clustering algorithms such as k-means, k-centers and expectation maximization (EM), QB is not a global clustering method. It can give different results under different initial conditions

of the data set when there is no obvious distance threshold which can separate the clusters into meaningful bundles; for example we should expect different clusters under different permutations/orderings of the tracks in a densely packed tractography. However, we found that there is enough agreement even between two clusterings of the same tractography with different orderings. If the clusters are truly separable by distances then there is a global solution independent of orderings. This is often perceivable in smaller subsets of the initial tractography. We empirically found that this problem is minimized even with real data sets when a low distance threshold of about 10 – 20 mm is used.

Furthermore, the output of QB can become input for another recent quick algorithm of quadratic time on average $O(M^2)$ called affinity propagation where now $M \ll N$ therefore, the overall time stays linear on the number of tracks N . Other algorithms previously too slow to be used on the entire tractography can now be used efficiently too e.g. kNN, hierarchical clustering and many others.

We saw that QB is a linear time clustering method based on track distances, which is on average linear time $O(N)$ where N is the number of tracks and with worst case $O(N^2)$ when every track is a singleton cluster itself. QB is the fastest known tractography clustering method and even real-time on tractographies with less than 20,000 tracks (depending on system CPU). We also showed that it uses a negligible amount of memory.

QB is fully automatic and very robust. It gives good agreements even between different subjects and can be used to create tractography atlases at high speed. Additionally, it can be used to explore multiple tractographies and find correspondences between tractographies, create landmarks used for registration or population comparisons.

QB can be used as well for reducing the dimensionality of the data sets at the time of interaction; providing an alternative way to ROIs using BOIs (bundles of interest) or TOIs (tracks of interest). We also showed it can be used to find “hidden” tracks not visible to the user at first instance. Therefore QB opens up the road to create rapid tools for exploring tractographies of any size.

The main concept of this clustering method is that a cluster can be represented by virtual tracks which are used only during cluster comparisons and not updated at every iteration.

A virtual (centroid) track is the average of all tracks in the cluster. We call it virtual because it doesn't need to correspond to an actual track in the real data set, and to distinguish it from exemplar (medoid) tracks which are again descriptors of the cluster but are represented by actual tracks.

The clustering creates a book of bundles/clusters which have easily obtainable descriptors. When clusters are held in a tree structure this permits upwards amalgamations to form bundles out of clusters, and downwards disaggregation to split clusters into finer sub-clusters corresponding to a lower distance threshold. However, we did not touch this hierarchical extension of this algorithm here and mostly concentrated on one level amalgamations.

We worked mostly with long tracks but strategies for short tracks or bundles are straightforward and documented. We also showed an efficient method where QB can speedup finding erroneous bundles or detecting structures of specific characteristics.

We showed results with simulated, single or multiple real subjects and the code for QuickBundles is freely available at dipy.org in module `dipy.segment.quickbundles`.

5 Conclusion

5.1 Summary

Here we present a summary of our main original contributions.

Reconstructing voxels. We first proposed a new reconstruction method called Diffusion Nabla Imaging (DNI) using an algorithm that directly approximates the Orientation Distribution Function using the Laplacian of the signal in q-space. Additionally, we found that a family of transforms exists which is a superset of DNI. We call this the Equatorial Inversion Transform (EIT). We showed that EIT has higher angular accuracy in simulations than the other methods and that it introduces interesting theoretical foundations for the interpretation of the dMRI signal. We compared and evaluated different Cartesian-grid q-space dMRI acquisition schemes, using methods based on the inverse Fourier transform of the diffusion signal, with reconstructions by Diffusion Spectrum Imaging (DSI), Generalized Q-sampling Imaging (GQI) and the EIT. We also compared EIT against GQI2 which had not been applied to simulated or real data until now. We found that GQI2 has similar performance with that of the EIT and it can generate smooth ODFs.

Integrating to tracks. Most previously published reconstruction methods are closely linked to their own specific tracking method. We have formulated a minimal tracking algorithm (EuDX) which is based on Euler integration and trilinear interpolation. This algorithm integrates voxel level information about fibre orientations including multiple crossings, and employs a range of stopping criteria. The purpose of this algorithm is to be faithful to the reconstruction results rather than try to correct or enhance them by introducing regional or global considerations which is the topic of other popular approaches. Interestingly, in the experiments with the software phantoms, EuDX performed better than a popular probabilistic method. With the real data sets it generated more uniform bundles.

Segmenting tracks. The end goal of clustering is to be able to segment tractography into tracts that have biological meaning. This is a difficult problem with no well-defined gold standard. In order to succeed better, we need to be able to compare the results of tractographies, and we need to be able to allow experts in anatomy to interact with the results of the tractography. Unfortunately, most current methods are so slow to com-

pute that it is not practical to compare different methods in reasonable time, and they cannot run fast enough for an expert to interact with them in close to real time. This thesis provides a complete solution to this problem. We developed a surprisingly simple, fully automatic, linear time, clustering method (QuickBundles) which reduces massive tractographies into a few easily accessible bundles. These bundles are characterised by representative tracks which are multi-purpose and can be used for interaction with the data or as the basis for applying higher complexity clustering methods which would have been impossible or too slow with the full data set. QuickBundles is as far as we know the fastest existing tractography clustering algorithm; providing the opportunity for clinical real-time applications.

Registering tracks. After applying QuickBundles to tractographies from different subjects, we showed how to use the representative tracks to identify robust landmarks within each subject which, with similarity metrics which we have introduced, we can use them to directly register the different tractographies together in a highly efficient way. We believe the resulting correspondences provide important evidence for the anatomical plausibility of the derived bundles. We demonstrated how these methods can be used for group analysis, as well as for atlas creation.

5.2 Software

In providing this thesis we tried to do our best to follow state-of-the-art scientific practices. One of the important achievements was to create and distribute two different software libraries DIPY [104] and FOS [223]. DIPY is used for dMRI analysis and FOS concentrates only on the visualization aspects using OpenGL. They are both implemented in the Python programming language. We hope that these projects will add up to the stack of existing OpenSource projects in the Neuroimaging community like Camino, FSL, DSI Studio and SPM. We believe that by providing our code open source and not-for-profit we allow other researchers to test and extend our findings. We believe that this is a factor which can increase the quality of scientific research beyond the standard expectations. Ioannidis et al. [224] showed that many scientists are avoiding to publish negative results and the Neuroimaging community is not an exception. We believe that perhaps a way out of this problem is publishing and sharing code. In

that perspective others can confirm, validate and push forward our current findings with speeds which were impossible in the past.

We have enjoyed writing thousands of lines of code in order to generate the dMRI algorithms or even the figures of this dissertation. Both DIPY and FOS have attracted further developers and scientists from acknowledged universities around the world who contribute today to these platforms.

5.3 Future work

Here we will describe what our future plans are and the research path we wish to take after the completion of this dissertation.

Extending EIT. Currently the EIT expects Cartesian grid-based q-space data like those used in DSI reconstructions. However, it seems that because the EIT integrates directional information radially on spherical shells it will be straightforward to extend it to one or more spherical grid q-space data. Spherical grid data are more commonly available and usually they need less scanner time. In order to make this possible we will need to do further research on spherical interpolations. Perhaps our spherical smoothing functions described in section 2.8 will become handy. However, this needs further examination.

Non-linear Direct Registration. Our direct registration approach allows only for linear transformations. It would be very interesting first to validate our method against volume based registration. Furthermore, we could investigate a log-Euclidean polyaffine framework which allows for smooth non-linear transformations. This will be a beautifully challenging problem as the optimization will be more difficult. Powell’s method works perfectly well with the few parameters needed in linear registration for this approach. However, with the nonlinear deformations many more parameters will need fitting. Perhaps with the aid of robust optimizers like the Particle Swarm Optimizer (PSO) [225], [226] we will be able to provide more accurate registration.

Clinical Applications. We propose to extend our preliminary involvement with clinical research into trichotilomania [227] from fractional anisotropy to track density calculations. We would also like to investigate bundle differences with other disorders such as autism or schizophrenia. Both autism and schizophrenia are considered to have strong relationship

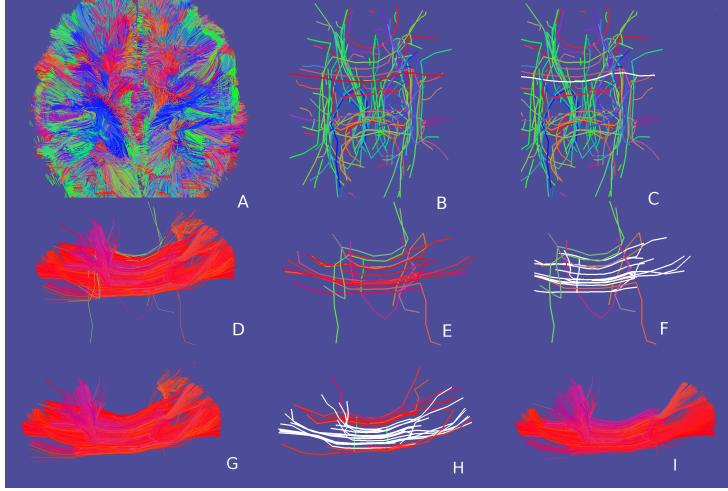


Figure 5.1: An example of how a medical practitioner can use our visualization software to select fibre bundles of their interest.

with defects in white matter architecture caused by disrupted connectivity [228].

Interactive Labeling. We are developing a scientific visualization tool that solves the problem of interacting with tractographies by creating real-time simplifications of the underlying anatomical bundle structures. The process that we propose works recursively: starting from a small number of clusters of streamlines the user decides which clusters to explore. Exploring a cluster means that the application re-clusters its content at a finer grained level in real-time. Of course these representative tracks are provided by QuickBundles which can cluster thousands of tracks in milliseconds.

Our approach starts by providing a first simplified version (see Fig. 5.1B) of the initial full tractography (see Fig. 5.1A). After visually inspecting the simplified tractography (see Fig. 5.1B) the practitioner can interactively select one or more representative tracks (see Fig. 5.1C, white track). When one or more representative tracks are selected the practitioner can see the content of the related clusters (see Fig. 5.1D). In order to explore the detailed structure of the selection the user may ask to re-cluster the selected BOIs into smaller clusters (see Fig. 5.1E). In that way one can further refine his previous selection. After selecting one or more of the small clusters through their representatives (see Fig. 5.1F, white tracks) the user can repeat the visual inspection step (Fig. 5.1G), and the re-clustering step (Fig. 5.1H) as required in order to unveil the local structures (Fig. 5.1I)

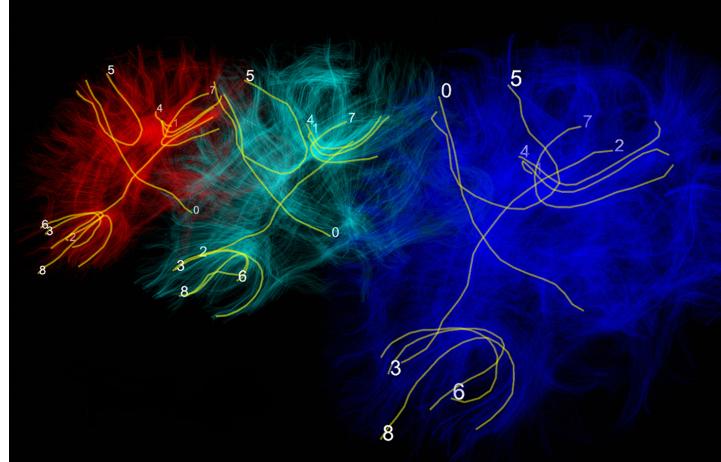


Figure 5.2: Direct track correspondence between different subjects.

which are interesting to their work.

Shape Correspondence. Using multi-brain visualization we were able to have a first investigation of shape correspondence [229] of tracks between different subjects (see Fig. 5.2). We would like to augment this first implementation from track to bundle correspondence. In other words a medical practitioner could select a representative track or a bundle in one subject and see in real-time the corresponding bundles in the other subjects. We imagine this as an amalgamation of the tools presented in Fig. 5.2 and 5.1 which are developed in DIPY and FOS.

Microstructure. Our work up till now has focused on estimating the structure of white matter in the brain from standard diffusion MRI acquisitions. However, these acquisitions provide voxel sizes of about 2 mm. Axon fibres, which are about one micron in diameter, are much smaller than single voxels. Nevertheless, diffusion MRI can provide information on the distribution of microstructural features, such as the fibre orientation, fibre diameter or density, within each voxel. We would also like to extend our work in that area of microanatomy tracking which has recently shown very interesting results [230], [231]. This high resolution domain require model-fitting of many parameters. It would be interesting to investigate if EIT or other non-parametric ideas could help alleviate this problem.

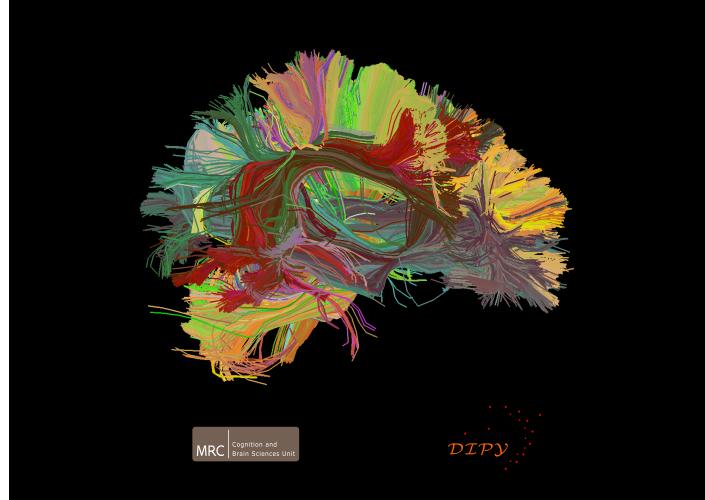


Figure 5.3: Every participant received a picture of his tractography as a gift for their help to our experiments.

5.4 New frontiers

White matter fibre crossings – from the voxel level to the tractography and bundle level – have been a major motivation for this thesis. In March 2012, Wedeen et al. [232] published in Science Magazine a fascinating work that reinforces the importance of these topics. They showed that they can identify in the tractographies of the brains of humans and other animals (in vivo and in vitro) fibre bundles which are in agreement with confocal microscopy and other staining techniques. The authors clarified that a grid-like structure is prevalent in the brain i.e. fibre bundles crossing in more areas than would previously have been expected. Furthermore, they also showed that the bundles curve more vigorously than previously understood. We believe that in this thesis we have significantly enhanced and extended the techniques that were used to establish these groundbreaking results, and have created a framework for them to be applied by the neuroscience community.

In bringing this thesis to a close we would like to thank the participants who took part in our imaging studies. To honour them we created special pictures of their tractographies like the one shown in Fig. 5.3 which were subsequently presented to them.

A Appendix

A.1 Data and sequences

Summary of the acquisition protocols used in this thesis.

101 and 117 gradients at MRC-CBU. 3T scanner (TIM Trio, Siemens) with a 12 channel coil, using Siemens advanced diffusion work-in-progress sequence, and STEAM [101, 15] as the diffusion preparation method. The field of view was $240 \times 240 \text{ mm}^2$, matrix size 96×96 , and slice thickness 2.5 mm (no gap). 55 slices were acquired to achieve full brain coverage, and the voxel resolution was $2.5 \times 2.5 \times 2.5 \text{ mm}^3$. Two sampling schemes were considered: a 102-point half grid acquisition (TR=8, 200 ms, TE=69 ms) with a maximum b-value of $4,000 \text{ s/mm}^2$, and a single shell acquisition using 118 non-collinear gradient directions (TR=7,000 ms, TE=47ms) and a b-value of $1,000 \text{ s/mm}^2$. The two acquisition schemes were matched for total acquisition time (*14 min 37s*), voxel resolution, and bandwidth.

101 gradients at MRC-CBU. 3T scanner (TIM Trio, Siemens) with a 32 channel coil, using Siemens advanced diffusion work-in-progress sequence, and STEAM [101, 15] as the diffusion preparation method. The field of view was $240 \times 240 \text{ mm}^2$, matrix size 96×96 , and slice thickness 2.5 mm (no gap). 55 slices were acquired to achieve full brain coverage, and the voxel resolution was $2.5 \times 2.5 \times 2.5 \text{ mm}^3$. A 102-point half grid acquisition with a maximum b-value of $4,000 \text{ s/mm}^2$ was used. The total acquisition time was *14 min 21s* with TR=8, 200ms and TE=69ms.

257 gradients at EPFL. 3T scanner (TIM Trio, Siemens) with a 32 channel coil. The field of view was $210 \times 210 \text{ mm}^2$, matrix size 96×96 , and slice thickness 3 mm . 44 slices were acquired and the voxel resolution was $2.2 \times 2.2 \times 3.0 \text{ mm}^3$. A 258-point half grid acquisition scheme with a maximum b-value of 8011 s/mm^2 (DSI515) was used. The total acquisition time was *34 min* with TR=8200 ms and TE=165 ms.

A.2 The Cosine Transform

Here we provide a simple derivation of the cosine transform: $\int_0^\infty \cos(st)g(t)dt$ where $g(t)$ defined on $[t, \infty)$.

Let $f(t)$ be an even function, $f(t) = f(-t)$, defined for $-\infty < t < \infty$.

$$\begin{aligned}
F(s) &= \int_{-\infty}^{\infty} f(t)e^{its}dt \\
&= \int_0^{\infty} f(t)e^{its}dt + \int_{-\infty}^0 f(t)e^{its}dt \\
&= \int_0^{\infty} f(t)e^{its}dt - \int_{-\infty}^0 f(-t)e^{-its}dt \\
&= \int_0^{\infty} f(t)e^{its}dt + \int_0^{\infty} f(t)e^{-its}dt \\
&= \int_0^{\infty} f(t)[e^{its} + e^{-its}]dt \\
&= \int_0^{\infty} f(t)[\cos(its) + i\sin(its) + \cos(its) - i\sin(its)]dt \\
&= 2 \int_0^{\infty} f(t)\cos(st)dt
\end{aligned}$$

In the third row above we replaced $t \rightarrow -t$ in the second integral. If we want to express the Fourier transform of a symmetric function as an integral over the whole space we have $F(s) = \int_{-\infty}^{\infty} f(t)\cos(st)dt$.

A.3 Fourier Transform of $P(\mathbf{r})r^2$

From Fourier analysis we know that if $E(\mathbf{q})$ is the Fourier transform function of $P(\mathbf{r})$ then

$$\begin{aligned}
\mathfrak{F}(xP(\mathbf{r})) &= i \frac{\partial E(\mathbf{q})}{\partial \mathbf{q}_x} \\
\mathfrak{F}(x^2P(\mathbf{r})) &= - \frac{\partial^2 E(\mathbf{q})}{\partial \mathbf{q}_x^2}
\end{aligned}$$

where $\mathfrak{F}()$ is the Fourier transform. By writing the second equation for y and z and summing them all together we obtain

$$\mathfrak{F}(r^2 P(\mathbf{r})) = -\frac{\partial^2 E(\mathbf{q})}{\partial \mathbf{q}_x^2} - \frac{\partial^2 E(\mathbf{q})}{\partial \mathbf{q}_y^2} - \frac{\partial^2 E(\mathbf{q})}{\partial \mathbf{q}_z^2} = -\nabla^2 E(\mathbf{q})$$

A.4 Radial projection of a symmetric function

Let $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ be a symmetric function with the 3D Fourier transform function $\hat{f}(\mathbf{q})$ and $\hat{\mathbf{u}}$ be an arbitrary unit vector. We will show that $\int_0^\infty f(r\hat{\mathbf{u}}) dr = \frac{1}{8\pi^2} \int \int_{\hat{\mathbf{u}}^\perp} \hat{f}(\mathbf{q}) q dq d\phi$ where $\hat{\mathbf{u}}^\perp$ is the plane perpendicular to $\hat{\mathbf{u}}$.

Without loss of generality, we align $\hat{\mathbf{u}}$ with the z-axis having $\hat{\mathbf{z}} = \hat{\mathbf{u}}$. Using the Dirac delta function (make use of Lebesgue integral) we can now write

$$\begin{aligned} \int_0^\infty f(r\hat{\mathbf{z}}) dr &= \int_0^\infty f(0, 0, z) dz \\ &= \frac{1}{2} \int \int \int_{\mathbb{R}^3} f(x, y, z) \delta(x) \delta(y) dx dy dz \end{aligned}$$

The factor $1/2$ is required because we need the integral only in the positive half of the z-axis, and the function is symmetric. Let us define $g(x, y, z) \equiv \delta(x) \delta(y)$. For the two functions $f, g : \mathbb{R}^3 \rightarrow \mathbb{R}$ with Fourier transform functions $\hat{f}(\mathbf{q})$ and $\hat{g}(\mathbf{q})$, Parseval's theorem states that

$$\int \int \int_{\mathbb{R}^3} f(x, y, z) g^*(x, y, z) dx dy dz =$$

$$(2\pi)^{-3} \int \int \int_{\mathbb{R}^3} f(q_x, q_y, q_z) \hat{g}^*(q_x, q_y, q_z) dq_x dq_y dq_z$$

Furthermore, $\hat{g}(q_x, q_y, q_z) = 2\pi\delta(q_z)$ and replacing it in the above equations leads to

$$\begin{aligned}
\int_0^\infty f(r\hat{\mathbf{z}}) &= \int \int \int_{\mathbb{R}^3} \frac{1}{2} f(x, y, z) g(x, y, z) dx dy dz \\
&= \frac{1}{2(2\pi)^3} \int \int \int_{\mathbb{R}^3} f(q_x, q_y, q_z) 2\pi \delta(q_z) dq_x dq_y dq_z \\
&= \frac{1}{8\pi^2} \int_{-\infty}^\infty \hat{f}(q_x, q_y, 0) dq_x dq_y
\end{aligned}$$

A.5 The Tensor in GQI

We now apply this formulation under the assumption that the diffusion voxel can be represented by a single tensor model. Eq. 2.2 can be written in the form

$$S(\mathbf{q}) = S_0 \exp(-b\mathbf{q}^T D \mathbf{q}) \quad (\text{A.1})$$

where S_0 is the image when b-value is equal to 0, b is the b-value for a specific direction and D is a 3×3 matrix, known as the diffusion tensor. Then from Eq. 2.9 and A.1 the Fourier transform of S is equal to

$$Q(\mathbf{R}) = \int S_0 \exp(-b\mathbf{q}^T D \mathbf{q}) \exp(-j2\pi \mathbf{q} \cdot \mathbf{R}) d\mathbf{q} \quad (\text{A.2})$$

The same equation in its triple integral form can be written as

$$\begin{aligned}
Q(\mathbf{R}) &= S_0 \iiint \exp\left(-b \sum_{i=1}^3 q_i^2 \lambda_i - j2\pi \sum_{i=1}^3 q_i R_i\right) dq_1 dq_2 dq_3 \\
&= S_0 \iiint \prod_{i=1}^3 \exp\left(-b q_i^2 \lambda_i - j2\pi q_i R_i\right) dq_1 dq_2 dq_3 \\
&= S_0 \prod_{i=1}^3 \int \exp\left(-b q_i^2 \lambda_i - j2\pi q_i R_i\right) dq_i \\
&= S_0 \prod_{i=1}^3 \int \exp\left(-b \lambda_i [q_i^2 + \frac{j2\pi R_i}{b \lambda_i} q_i]\right) dq_i \\
&= S_0 \prod_{i=1}^3 \int \exp\left\{-b \lambda_i [(q_i + \frac{j\pi R_i}{b \lambda_i})^2 + \frac{\pi^2 R_i^2}{b^2 \lambda_i^2}]\right\} \\
&= S_0 \prod_{i=1}^3 \int \exp\left\{-b \lambda_i (q_i + \frac{j\pi R_i}{b \lambda_i})^2\right\} \exp\left\{-\frac{\pi^2 R_i^2}{b \lambda_i}\right\} \quad (\text{A.3})
\end{aligned}$$

In that stage we can make use of the formula $\int \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) dx = 1$. Now we can see that $b\lambda_i = 1/2\sigma^2$ and μ corresponds to $\mu = -jR_i/b\lambda_i$. Therefore, Eq. A.3 can now be written as

$$\begin{aligned} Q(\mathbf{R}) &= S_0 \prod_{i=1}^3 \sqrt{\frac{\pi}{b\lambda_i}} \exp\left(-\frac{\pi^2 R_i^2}{b\lambda_i}\right) \\ &= S_0 \left(\frac{\pi}{b}\right)^{3/2} \frac{1}{\sqrt{\prod_{i=1}^3 \lambda_i}} \exp\left(-\frac{\pi^2}{b} \mathbf{R}^T D^{-1} \mathbf{R}\right) \end{aligned} \quad (\text{A.4})$$

where D is the diffusion tensor. We can replace the displacement vector \mathbf{R} with a scalar value L and a unit vector $\hat{\mathbf{u}}$ i.e. $\mathbf{R} = L\hat{\mathbf{u}}$ and from Eq. A.4 we can replace $\frac{2\pi^2}{b} \hat{\mathbf{u}}^T D^{-1} \hat{\mathbf{u}}$ with k and $S_0 \left(\frac{\pi}{b}\right)^{3/2} \frac{1}{\lambda_1 \lambda_2 \lambda_3}$ with α . Using that last change of variables we can now write

$$\begin{aligned} \psi_Q(\mathbf{r}, \hat{\mathbf{u}}) &= \int_0^{L_\Delta} Q(\mathbf{r}, L\hat{\mathbf{u}}) dL \\ &= \alpha \int_0^{L_\Delta} \exp(-L^2 \frac{k}{2}) dL \end{aligned} \quad (\text{A.5})$$

Setting $m = \sqrt{k}L$ and using the derivation for the error function Eq. A.5 illustrates the remarkable result that we can calculate analytically the spin ODF for Gaussian diffusion using the cumulative distribution function *CDF*.

$$\psi_Q(\hat{\mathbf{u}}) = \frac{\alpha}{\sqrt{k}} \int_0^{\sqrt{k}L_\Delta} e^{-m^2/2} dm \quad (\text{A.6})$$

$$= \alpha \sqrt{\frac{2\pi}{k}} \left[CDF(\sqrt{k}L_\Delta) - \frac{1}{2} \right] \quad (\text{A.7})$$

This can be used as a check to compare the approximated or sampled spin ODF that is derived in [72] with Eq. A.7 for the case of gaussian diffusion.

What is also very interesting is to try to derive what the normalization factor should be for the spin ODF in Eq. A.5. Because calculating a spher-

ical integral from Eq. A.7 seems at the moment very complicated we first work with the simpler gaussian diffusion ODF derived by Tuch [32]

$$\psi_{p_\Delta} = \frac{1}{Z} \sqrt{\frac{\pi\tau}{\mathbf{u}^T D^{-1} \mathbf{u}}} \quad (\text{A.8})$$

$$\frac{Z}{\sqrt{\pi\tau}} = \iint_{S^2} (\mathbf{u}^T D^{-1} \mathbf{u})^{-\frac{1}{2}} d\mathbf{u} \quad (\text{A.9})$$

Let us now define $f_D(\mathbf{u}) = (\mathbf{u}^T D^{-1} \mathbf{u})^{-\frac{1}{2}}$. From [233] (19.31.2) we know that we can calculate the following integral on the entire space (eq. A.10). By expanding it in polar form we can find the surface integral required in Eq. A.9

$$MHG = \iiint f_D(x) e^{-|x|^2} dx \quad (\text{A.10})$$

$$= \int_0^\infty \left[\iint f_D(ru) du \right] e^{-r^2} r^2 dr, \quad x = ru \text{ (polar)} \quad (\text{A.11})$$

However, we know that $f_D(r\mathbf{u}) = ((r\mathbf{u})^T D^{-1} (r\mathbf{u}))^{-\frac{1}{2}} = r^{-1} (\mathbf{u}^T D^{-1} \mathbf{u})^{-\frac{1}{2}}$. Therefore,

$$\begin{aligned} MHG &= \int_0^\infty \left[\iint f_D(u) du \right] e^{-r^2} r dr \\ &= \iint f_D(u) du \int_0^\infty r e^{-r^2} dr \\ &= \frac{1}{2} \iint f_D(u) du \end{aligned}$$

Consequently, $\frac{Z}{\sqrt{\pi\tau}} = 2MHG$ where MHG is the multivariate hypergeometric function with $\mu = -\frac{1}{2}, B = I, n = 3$ and $\lambda_1, \lambda_2, \lambda_3$ the eigenvalues

ues of \mathbf{A} derived from [233] (19.31.2) & (19.16.9). Therefore,

$$\begin{aligned}\frac{Z}{\sqrt{\pi\tau}} &= 2MHG \\ &= \frac{2\pi^{\frac{3}{2}}\Gamma(1)}{\sqrt{\det(I)\Gamma(\frac{3}{2})}} R_{-\frac{1}{2}}\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}; \lambda_1, \lambda_2, \lambda_3\right)\end{aligned}$$

and $R_{-\frac{1}{2}} = \frac{1}{2} \int_0^\infty t^0(t + \lambda_1)^{-\frac{1}{2}}(t + \lambda_2)^{-\frac{1}{2}}(t + \lambda_3)^{-\frac{1}{2}} dt$ with $\alpha = \frac{1}{2}$ and $\alpha' = 1$.

Given $\lambda_1, \lambda_2, \lambda_3$ we can integrate numerically or even possibly analytically. For the isotropic case the integral simplifies to $\frac{1}{2} \int_0^\infty t^0(t + \lambda)^{-\frac{3}{2}} dt = \frac{1}{\sqrt{\lambda}}$ and for the cylindrical case ($\lambda_2 = \lambda_3$) to $\frac{1}{2} \int_0^\infty (t + \lambda_1)^{-\frac{1}{2}}(t + \lambda_2)^{-1} dt$.

A.6 Affinity Propagation

Affinity propagation (AP) is a recent $O(N^2)$ clustering method invented by Frey et al. [234] and Dueck et al. [211] which is inspired by loopy belief propagation [235] and other recent innovations in graphical models and more specifically is an instance of the max-sum algorithm in factor graphs. For the completeness of this thesis and because AP is a relatively new algorithm we give a short description of the AP in this section. AP is an exemplar based clustering method where the center of a cluster is a real data point (exemplar) as in k-medoids, and k-centres rather than an average virtual point as in k-means. AP starts by simultaneously considering all data points as potential exemplars. Each data point is a node in a network and AP recursively transmits real-valued messages along the edges of the network until a good set of exemplars and corresponding clusters emerges.

AP takes as input a collection of similarities between data points, where the similarity $S(i, k)$ indicates how well the data point with index k is suited to be the exemplar for data point i . In order to understand AP we can think just for the moment that we try to cluster 2D data points and each similarity is expressed as the negative Euclidean distance $S(i, k) = -||x_i - x_k||^2$ (see Fig. A.1) therefore S for the moment is the negative complete squared distance matrix. Rather than requiring the number of clusters to be prespecified, AP adds a real number (preference weights) to the diagonal elements of S , one for each data point so that larger values of $S(k, k)$ are more likely to become exemplars. For simplicity, we

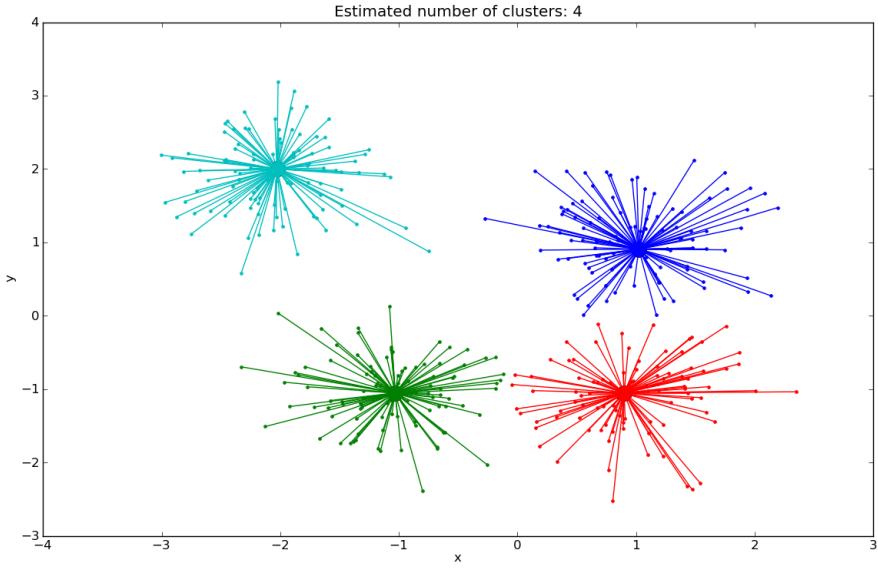


Figure A.1: Simple example of affinity propagation (AP) at work where it can precisely identify 4 different normal distributions with means $(1, 1)$, $(-1, -1)$, $(1, -1)$, $(-2, 2)$ and standard deviation 0.5. You can see the exemplars - most representative actual points - with thicker dots perfectly aligned with the means.

can choose the $\text{median}(S)$ as the common preference weight for all points; in this way we do not enforce any *a priori* information for one point to be an exemplar any more than any other point. For some applications this could be an appropriate requirement. There are two different messages exchanged between points: (1) responsibilities $R(i, k) = S(i, k) - \max_{k': k' \neq k} [S(i, k') + A(i, k')]$ and (2) availabilities which are initially $A(i, k) = 0$ and then equal to

$$\forall i, k : A(i, k) = \begin{cases} \sum_{i': i' \neq i} \max[0, R(i', k)], & \text{for } k = i \\ \min \left[0, r(k, k) + \sum_{i': i' \notin \{i, k\}} \max[0, r(i', k)] \right], & \text{for } k \neq i \end{cases} \quad (\text{A.12})$$

A very interesting fact is the way we get the final exemplars using AP. After the messages have converged, there are two ways you can identify exemplars:

1. For data point i , if $R(i, i) + A(i, i) > 0$, then data point i is an exemplar.

2. For data point i , if $R(i,i) + A(i,i) > R(i,j) + A(i,j)$, for all i not equal to j , then data point i is an exemplar.

Therefore, the availabilities and responsibilities are added to identify exemplars. For point i , the value of k that maximizes $A(i,k) + R(i,k)$ either identifies i as an exemplar if $k = i$, or identifies the data point that is the exemplar for point i . The message passing procedure is terminated either after a fixed number of iterations, or after changes in the messages stay low, or local decisions stay constant; also the messages are damped - combining previous with current message - to avoid numerical oscillations.

Of course, when we need to calculate distances between many points then the distance matrix becomes too big for the available memory. In that case, if we are lucky and the data sets are sparse then we can use AP on sparse matrices. When the data sets are not sparse, as it is the case with tractographies, we need to reduce the dimensionality of the data sets and this is why QB can be very handy. The complete algorithm for AP is given in Alg. 6.

Algorithm 6 Affinity Propagation

Input Similarity/affinity matrix S where the diagonal elements of $S(k, k)$ indicate the a priori preference for k to be chosen as an exemplar

Output Clustering $CAP \leftarrow \{c_1, \dots, c_k, \dots, c_{|CAP|}\}$, where a cluster $c \leftarrow (I, \mathbf{e}, N)$

$\forall i, k : A(i, k) \leftarrow R(i, k) \leftarrow 0$

$S \leftarrow S + n$ # remove degeneracies

$d \leftarrow 0.5$ # set damping factor

$last_iter \leftarrow 100$ # last iteration

For iter = 1 to last_iter **Do**

$R_{old} \leftarrow R$

$\forall i, k : R(i, k) \leftarrow S(i, k) - max_{k' : k' \neq k}[S(i, k') + A(i, k')]$

$R \leftarrow (1 - d)R + d * R_{old}$ # dampen responsibilities

$A_{old} \leftarrow A$

update availabilities

$\forall i, k : A(i, k) \leftarrow \begin{cases} \sum_{i' : i' \neq i} max[0, R(i', k)], & k = i \\ \min \left[0, R(k, k) + \sum_{i' : i' \notin \{i, k\}} [0, R(i', k)] \right], & k \neq i \end{cases}$

$A \leftarrow (1 - d)A + dA_{old}$ # dampen availabilities

$\forall i, I_e \leftarrow argmax S(i, I_d)$ # find indices of exemplars

$I_e(I_d) \leftarrow 1 : size(I_d)$

$L \leftarrow I_d(I_e)$ # assign labels

$CAP \leftarrow \{c_0, \dots, c_k, \dots, c_{|CAP|}\}$ # clustering output

where a cluster $c \leftarrow (I, \mathbf{e}, N)$ holds the AP exemplars \mathbf{e} ,

the indices I of the cluster elements and N the number of elements

References

- [1] H. Johansen-Berg and T. E. J. Behrens, eds., *Diffusion MRI*. Academic Press, 2009.
- [2] M. Denis Le Bihan, C. Poupon, A. Amadon, and F. Lethimonnier, "Artifacts and pitfalls in diffusion MRI," *Journal of Magnetic Resonance Imaging*, vol. 24, pp. 478–488, 2006.
- [3] V. J. Wedeen, R. P. Wang, J. D. Schmahmann, T. Benner, W. Y. Tseng, G. Dai, D. N. Pandya, P. Hagmann, H. D'Arceuil, and A. J. de Crespigny, "Diffusion spectrum magnetic resonance imaging (DSI) tractography of crossing fibers," *Neuroimage*, vol. 41, no. 4, pp. 1267–77, 2008.
- [4] S. Mori, B. J. Crain, V. P. Chacko, and P. C. van Zijl, "Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging," *Ann Neurol*, vol. 45, no. 2, pp. 265–9, 1999.
- [5] T. E. J. Behrens, H. Johansen-Berg, M. W. Woolrich, C. A. M. Wheeler-Kingshott, P. A. Boulby, G. J. Barker, E. L. Sillery, K. Sheehan, O. Ciccarelli, A. J. Thompson, J. M. Brady, and P. M. Matthews, "Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging," *Nature Neuroscience*, vol. 6, no. 7, pp. 750–757, 2003.
- [6] G. Margolis and J. Pickett, "New applications of the Luxol fast blue myelin stain," *Laboratory investigation; a journal of technical methods and pathology*, vol. 5, no. 6, p. 459.
- [7] D. S. Tuch, J. J. Wisco, M. H. Khachaturian, W. Vanduffel, L. B. Ekstrom, and R. Ko, "Q-ball imaging of macaque white matter architecture," *Phil. Trans. R. Soc. B*, vol. 360, no. 1457, pp. 869–879, 2005.
- [8] C. Beaulieu, "The basis of anisotropic water diffusion in the nervous system - a technical review," *NMR Biomed*, vol. 15, no. 7-8, pp. 435–55, 2002.
- [9] G. Zhai, W. Lin, K. P. Wilber, G. Gerig, and J. H. Gilmore, "Comparisons of regional white matter diffusion in healthy neonates and

adults performed with a 3.0-T head-only MR imaging unit," *Radiology*, vol. 229, no. 3, pp. 673–81, 2003.

- [10] S. Mori and J. Zhang, "Principles of diffusion tensor imaging and its applications to basic neuroscience research," *Neuron*, vol. 51, pp. 527–39, 2006.
- [11] S. Mori, S. Wakana, P. Van Zijl, and L. Nagae-Poetscher, *MRI atlas of human white matter*. Am Soc Neuroradiology, 2005.
- [12] D. McRobbie, E. Moore, and M. Graves, *MRI from Picture to Proton*. Cambridge University Press, 2006.
- [13] E. O. Stejskal and J. E. Tanner, "Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient," *The Journal of Chemical Physics*, vol. 42, no. 1, pp. 288–292, 1965.
- [14] T. Reese, O. Heid, R. Weisskoff, and V. Wedeen, "Reduction of eddy-current-induced distortion in diffusion mri using a twice-refocused spin echo," *Magnetic Resonance in Medicine*, vol. 49, no. 1, pp. 177–182, 2003.
- [15] M. A. Bernstein, K. F. King, and X. J. Zhou, *Handbook of MRI Pulse Sequences*. Elsevier Academic Press, 2004.
- [16] E. J. Canales-Rodríguez, L. Melie-García, and Y. Iturria-Medina, "Mathematical description of q-space in spherical coordinates: exact q-ball imaging," *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, vol. 61, no. 6, pp. 1350–67, 2009.
- [17] P. T. Callaghan, *Principles of Nuclear Magnetic Resonance Microscopy*. Oxford University Press, 1991.
- [18] O. Heid, "Eddy current-nulled diffusion weighting," *In Proceedings of the 8th Annual Meeting of ISMRM, Denver*, p. 799, 2000.
- [19] E. J. Auerbach and K. Ugurbil, "Improvement in diffusion mri at 3t and beyond with the twice-refocused adiabatic spin echo (trase) sequence," *In Proceedings of the 12th Annual Meeting of ISMRM*, 2004.

- [20] T. Niendorf, R. M. Dijkhuizen, D. G. Norris, M. van Lookeren Campagne, and K. Nicolay, "Biexponential diffusion attenuation in various states of brain tissue: Implications for diffusion-weighted imaging," *Magnetic Resonance in Medicine*, vol. 36, no. 6, pp. 847–857, 1996.
- [21] R. Mulkern, H. Gudbjartsson, C. Westin, H. Zengigonul, W. Gartner, C. Guttmann, R. Robertson, W. Kyriakos, R. Schwartz, D. Holtzman, *et al.*, "Multi-component apparent diffusion coefficients in human brain," *NMR Biomed*, vol. 12, pp. 51–62, 1999.
- [22] E. Ozarslan, T. Shepherd, B. Vemuri, S. Blackband, and T. Mareci, "Resolution of complex tissue microarchitecture using the diffusion orientation transform (dot)," *NeuroImage*, vol. 31, no. 3, pp. 1086–1103, 2006.
- [23] J. Jensen, J. Helpern, A. Ramani, H. Lu, and K. Kaczynski, "Diffusional kurtosis imaging: The quantification of non-gaussian water diffusion by means of magnetic resonance imaging," *Magnetic Resonance in Medicine*, vol. 53, no. 6, pp. 1432–1440, 2005.
- [24] H. Torrey, "Bloch equations with diffusion terms," *Physical Review*, vol. 104, no. 3, pp. 563–565, 1956.
- [25] A. Einstein, *Investigations on the Theory of the Brownian Movement*. Dover Publications, 1956.
- [26] C. Lenglet, J. Campbell, M. Descoteaux, G. Haro, P. Savadjiev, D. Wassermann, A. Anwander, R. Deriche, G. Pike, G. Sapiro, *et al.*, "Mathematical methods for diffusion mri processing," *Neuroimage*, vol. 45, no. 1, pp. S111–S122, 2009.
- [27] P. Hagmann, L. Jonasson, P. Maeder, J.-P. Thiran, V. J. Wedeen, and R. Meuli, "Understanding diffusion mr imaging techniques: From scalar diffusion-weighted imaging to diffusion tensor imaging and beyond," *Radiographics*, vol. 26, no. 1, pp. 205–223, 2006.
- [28] M. King, J. Houseman, S. Roussel, N. Van Bruggen, S. Williams, and D. Gadian, "q-Space imaging of the brain," *Magnetic Resonance in Medicine*, vol. 32, no. 6, pp. 707–713, 1994.

- [29] M. Moseley, K. Butts, M. Yenari, M. Marks, and A. Crespigny, "Clinical aspects of dwi," *NMR in Biomedicine*, vol. 8, no. 7, pp. 387–396, 1995.
- [30] S. Warach, J. Gaa, B. Siewert, P. Wielopolski, and R. Edelman, "Acute human stroke studied by whole brain echo planar diffusion-weighted magnetic resonance imaging," *Annals of neurology*, vol. 37, no. 2, pp. 231–241, 1995.
- [31] M. Maddah, *Quantitative Analysis of Cerebral White Matter Anatomy from Diffusion MRI*. PhD thesis, Massachusetts Institute of Technology, 2008.
- [32] D. Tuch, "Q-ball imaging," *Magnetic Resonance in Medicine*, vol. 52, no. 6, pp. 1358–1372, 2004.
- [33] V. Wedeen, P. Hagmann, W. Tseng, T. Reese, and R. Weisskoff, "Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging," *Magnetic Resonance in Medicine*, vol. 54, no. 6, pp. 1377–1386, 2005.
- [34] T. E. Conturo, N. F. Lori, T. S. Cull, E. Akbudak, A. Z. Snyder, J. S. Shimony, R. C. McKinstry, H. Burton, and M. E. Raichle, "Tracking neuronal fiber pathways in the living human brain.," *Proc Natl Acad Sci U S A*, vol. 96, no. 18, pp. 10422–7, 1999.
- [35] T. Behrens, H. Johansen-Berg, S. Jbabdi, M. Rushworth, and M. Woolrich, "Probabilistic diffusion tractography with multiple fibre orientations: What can we gain?," *NeuroImage*, vol. 34, no. 1, pp. 144–155, 2007.
- [36] B. W. Kreher, I. Mader, and V. G. Kiselev, "Gibbs tracking: A novel approach for the reconstruction of neuronal pathways," *Proc. Intl. Soc. Mag. Reson. Med.*, vol. 16, p. 425, 2008.
- [37] S. Sotropoulos, "Processing of diffusion mr images of the brain: from crossing fibres to distributed tractography," 2010.
- [38] P. J. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi, "In vivo fiber tractography using DT-MRI data.," *Magn Reson Med*, vol. 44, no. 4, pp. 625–32, 2000.

- [39] C. Demiralp, J. F. Hughes, and D. H. Laidlaw, "Coloring 3D line fields using Boy's real projective plane immersion," *IEEE Trans. on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1457–1463, 2009.
- [40] O. Friman and C. F. Westin, "Uncertainty in white matter fiber tractography," *Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv*, vol. 8, no. Pt 1, pp. 107–14, 2005.
- [41] A. Anwander, M. Tittgemeyer, D. von Cramon, A. Friederici, and T. Knosche, "Connectivity-Based Parcellation of Broca's Area," *Cerebral Cortex*, vol. 17, no. 4, p. 816, 2007.
- [42] M. Björnemo, A. Brun, R. Kikinis, and C.-F. Westin, "Regularized stochastic white matter tractography using diffusion tensor MRI," in *Fifth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'02)*, (Tokyo, Japan), pp. 435–442, 2002.
- [43] P. Hagmann, J. P. Thiran, L. Jonasson, P. Vandergheynst, S. Clarke, P. Maeder, and R. Meuli, "DTI mapping of human brain connectivity: statistical fibre tracking and virtual dissection," *Neuroimage*, vol. 19, no. 3, pp. 545–54, 2003.
- [44] T. Hosey, G. Williams, and R. Ansorge, "Inference of multiple fiber orientations in high angular resolution diffusion imaging," *Magnetic Resonance in Medicine*, vol. 54, no. 6, pp. 1480–1489, 2005.
- [45] M. Perrin, C. Poupon, Y. Cointepas, B. Rieul, N. Golestani, C. Pallier, D. Riviere, A. Constantinesco, D. Le Bihan, and J. Mangin, "Fiber tracking in q-ball fields using regularized particle trajectories," in *Information Processing in Medical Imaging*, pp. 595–634, Springer, 2005.
- [46] T. E. J. Behrens, M. W. Woolrich, M. Jenkinson, H. Johansen-Berg, R. G. Nunes, S. Clare, P. M. Matthews, J. M. Brady, and S. M. Smith, "Characterization and propagation of uncertainty in diffusion-weighted {MR} imaging," *Magnetic Resonance in Medicine*, vol. 50, pp. 1077–1088, 2003.

- [47] T. Behrens and H. Johansen-Berg, "Relating connectional architecture to grey matter function using diffusion imaging," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 360, no. 1457, pp. 903–911, 2005.
- [48] A. J. Sherbondy, R. F. Dougherty, M. Ben-Shachar, S. Napel, and B. A. Wandell, "ConTrack: Finding the most likely pathways between brain regions using diffusion tractography," *J. Vis.*, vol. 8, pp. 1–16, 7 2008.
- [49] S. Jbabdi, M. Woolrich, J. Andersson, and T. Behrens, "A Bayesian framework for global tractography," *Neuroimage*, vol. 37, no. 1, pp. 116–129, 2007.
- [50] A. Sherbondy, D. Akers, R. Dougherty, M. Ben-Shachar, S. Napel, and B. Wandell, "MetroTrac: A metropolis algorithm for probabilistic tractography," *Human Brain Mapping, Florence*, 2006.
- [51] G. J. Parker, H. A. Haroon, and C. A. Wheeler-Kingshott, "A framework for a streamline-based probabilistic index of connectivity (PICo) using a structural interpretation of MRI diffusion measurements," *J Magn Reson Imaging*, vol. 18, no. 2, pp. 242–54, 2003.
- [52] D. Jones, M. Horsfield, and A. Simmons, "Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging," *optimization*, vol. 525, 1999.
- [53] J. C. Haselgrove and J. R. Moore, "Correction for distortion of echo-planar images used to calculate the apparent diffusion coefficient," *Magn Reson Med*, vol. 36, no. 6, pp. 960–4, 1996.
- [54] J. Zhuang, J. Hrabe, A. Kangarlu, D. Xu, R. Bansal, C. A. Branch, and B. S. Peterson, "Correction of eddy-current distortions in diffusion tensor images using the known directions and strengths of diffusion gradients," *J Magn Reson Imaging*, vol. 24, no. 5, pp. 1188–93, 2006.
- [55] J. Andersson and S. Skare, "A model-based method for retrospective correction of geometric distortions in diffusion-weighted EPI," *Neuroimage*, vol. 16, no. 1, pp. 177–199, 2002.

- [56] R. Wirestam, A. Bibic, J. Latt, S. Brockstedt, and F. Stahlberg, "Denoising of complex MRI data by wavelet-domain filtering: Application to high-b-value diffusion-weighted imaging," *Magnetic Resonance in Medicine*, vol. 56, no. 5, 2006.
- [57] R. A. Kanaan, S. S. Shergill, G. J. Barker, M. Catani, V. W. Ng, R. Howard, P. K. McGuire, and D. K. Jones, "Tract-specific anisotropy measurements in diffusion tensor imaging," *Psychiatry Res*, vol. 146, no. 1, pp. 73–82, 2006.
- [58] R. Deriche and O. Faugeras, "2-D curve matching using high curvature points: application to stereo vision," in *Pattern Recognition, 1990. Proceedings., 10th International Conference on*, vol. 1, 1990.
- [59] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor, "A minimum description length approach to statistical shape modeling," *IEEE Trans Med Imaging*, vol. 21, no. 5, pp. 525–37, 2002.
- [60] L. J. O'Donnell, C. F. Westin, and A. J. Golby, "Tract-based morphometry for white matter group analysis," *Neuroimage*, vol. 45, no. 3, pp. 832–44, 2009.
- [61] R. F. Dougherty, M. Ben-Shachar, R. Bammer, A. a. Brewer, and B. a. Wandell, "Functional organization of human occipital-callosal fiber tracts," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, pp. 7350–5, May 2005.
- [62] J. Dauguet, S. Peled, V. Berezovskii, T. Delzescaux, S. K. Warfield, R. Born, and C. F. Westin, "Comparison of fiber tracts derived from in-vivo DTI tractography with 3D histological neural tract tracer reconstruction on a macaque brain," *Neuroimage*, vol. 37, no. 2, pp. 530–8, 2007.
- [63] L. O'Donnell, *Cerebral white matter analysis using diffusion imaging*. PhD thesis, 2006.
- [64] H. Gray, "Gray's Anatomy," *Lea and Febiger, 20th u.s. edition edition, 1918. Online edition: bartelby.com*, 2000.
- [65] T. Williams, N. Gluhbegovic, and J. Jew, *The human brain: dissections of the real brain*. University of Iowa, 1999.

- [66] J.-D. Tournier, C.-H. Yeh, F. Calamante, K.-H. Cho, A. Connelly, and C.-P. Lin, "Resolving crossing fibres using constrained spherical deconvolution: validation using diffusion-weighted imaging phantom data.," *NeuroImage*, vol. 42, no. 2, pp. 617–25, 2008.
- [67] P. J. Basser, J. Mattiello, and D. LeBihan, "MR diffusion tensor spectroscopy and imaging," *Biophysical Journal*, vol. 66, pp. 259–267, 1994.
- [68] D. S. Tuch, T. G. Reese, M. R. Wiegell, N. Makris, J. W. Belliveau, and V. J. Wedeen, "High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity," *Magnetic Resonance in Medicine*, vol. 48, no. 4, pp. 577–582, 2002.
- [69] P. Callaghan, C. Eccles, and Y. Xia, "NMR microscopy of dynamic displacements: k-space and q-space imaging," *Journal of Physics E: Scientific Instruments*, vol. 21, p. 820, 1988.
- [70] E. Ozarslan and T. Mareci, "Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging," *Magnetic Resonance in Medicine*, vol. 50, no. 5, pp. 955–965, 2003.
- [71] A. Barmpoutis, M. Hwang, D. Howland, J. Forder, and B. Vemuri, "Regularized positive-definite fourth order tensor field estimation from DW-MRI," *NeuroImage*, vol. 45, no. 1, pp. S153–S162, 2009.
- [72] F. Yeh, V. Wedeen, and W. Tseng, "Generalized Q-sampling imaging," *IEEE Transactions on Medical Imaging*, vol. 29, no. 9, pp. 1626–1635, 2010.
- [73] V. J. Wedeen, R. P. Wang, J. D. Schmahmann, T. Benner, W. Y. Tseng, G. Dai, D. N. Pandya, P. Hagmann, H. D'Arceuil, and A. J. de Crespigny, "Diffusion spectrum magnetic resonance imaging (DSI) tractography of crossing fibers.," *Neuroimage*, vol. 41, no. 4, pp. 1267–77, 2008.
- [74] L. W. Kuo, J. H. Chen, V. J. Wedeen, and W. Y. Tseng, "Optimization of diffusion spectrum imaging and q-ball imaging on clinical MRI system," *Neuroimage*, vol. vol, pp. 41pp7–18.

- [75] E. Canales-Rodríguez, L. Melie-García, and Y. Iturria-Medina, “Mathematical description of q-space in spherical coordinates: Exact q-ball imaging,” *Magnetic Resonance in Medicine*, vol. 61, no. 6, pp. 1350–1367, 2009.
- [76] V. Pickalov and P. Basser, “3D tomographic reconstruction of the average propagator from MRI data,” in *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006*, pp. 710–713, 2006.
- [77] Y. Wu and A. Alexander, “Hybrid diffusion imaging,” *NeuroImage*, vol. 36, no. 3, pp. 617–629, 2007.
- [78] M. Descoteaux, R. Deriche, D. Le Bihan, J. Mangin, and C. Poupon, “Multiple q-shell diffusion propagator imaging,” *Medical image analysis*, vol. 15, no. 4, pp. 603–621, 2011.
- [79] I. Aganj, C. Lenglet, N. Jahanshad, E. Yacoub, N. Harel, P. Thompson, and G. Sapiro, “A hough transform global probabilistic approach to multiple-subject diffusion mri tractography,” *Medical image analysis*, vol. 15, no. 4, pp. 414–425, 2011.
- [80] K. M. Jansons and D. C. Alexander, “Persistent angular structure: new insights from diffusion magnetic resonance imaging data,” *Inverse Problems*, vol. 19, no. 5, pp. 1031–1046, 2003.
- [81] D. Alexander, G. Barker, and S. Arridge, “Detection and modeling of non-gaussian apparent diffusion coefficient profiles in human brain data,” *Magnetic Resonance in Medicine*, vol. 48, no. 2, pp. 331–340, 2002.
- [82] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche, “Regularized, fast, and robust analytical q-ball imaging,” *Magnetic Resonance in Medicine*, vol. 58, no. 3, pp. 497–510, 2007.
- [83] J. Tournier, F. Calamante, D. Gadian, A. Connelly, *et al.*, “Direct estimation of the fiber orientation density function from diffusion-weighted mri data using spherical deconvolution,” *NeuroImage*, vol. 23, no. 3, pp. 1176–1185, 2004.
- [84] J. D. Tournier, F. Calamante, and A. Connelly, “Robust determination of the fibre orientation distribution in diffusion MRI: non-

negativity constrained super-resolved spherical deconvolution," *NeuroImage*, vol. 35, no. 4, pp. 1459–72, 2007.

- [85] M. Descoteaux, R. Deriche, and A. Anwander, "Deterministic and probabilistic q-ball tractography: from diffusion to sharp fiber distribution," *Research Report 6273, INRIA Sophia Antipolis*, 2007.
- [86] K. Sakaie and M. Lowe, "An objective method for regularization of fiber orientation distributions derived from diffusion-weighted mri," *NeuroImage*, vol. 34, no. 1, pp. 169–176, 2007.
- [87] F. Yeh, V. Wedeen, and W. Tseng, "Estimation of fiber orientation and spin density distribution by diffusion deconvolution," *Neuroimage*, vol. 55, no. 3, pp. 1054–1062, 2011.
- [88] O. Pasternak, Y. Assaf, N. Intrator, and N. Sochen, "Variational multiple-tensor fitting of fiber-ambiguous diffusion-weighted magnetic resonance imaging voxels," *Magnetic resonance imaging*, vol. 26, no. 8, pp. 1133–1144, 2008.
- [89] C. Liu, R. Bammer, B. Acar, and M. Moseley, "Characterizing non-gaussian diffusion by using generalized diffusion tensors," *Magnetic Resonance in Medicine*, vol. 51, no. 5, pp. 924–937, 2004.
- [90] H. Lu, J. Jensen, A. Ramani, and J. Helpern, "Three-dimensional characterization of non-gaussian water diffusion in humans using diffusion kurtosis imaging," *NMR in Biomedicine*, vol. 19, no. 2, pp. 236–247, 2006.
- [91] Y. Assaf and P. Basser, "Composite hindered and restricted model of diffusion (charmed) mr imaging of the human brain," *Neuroimage*, vol. 27, no. 1, pp. 48–58, 2005.
- [92] Y. Assaf, R. Freidlin, G. Rohde, and P. Basser, "New modeling and experimental framework to characterize hindered and restricted water diffusion in brain white matter," *Magnetic Resonance in Medicine*, vol. 52, no. 5, pp. 965–978, 2004.
- [93] Y. Assaf, T. Blumenfeld-Katzir, Y. Yovel, and P. Basser, "AxCaliber: a method for measuring axon diameter distribution from diffusion MRI," *Magnetic Resonance in Medicine*, vol. 59, no. 6, pp. 1347–1354, 2008.

- [94] D. Alexander, P. Hubbard, M. Hall, E. Moore, M. Ptito, G. Parker, and T. Dyrby, "Orientationally invariant indices of axon diameter and density from diffusion MRI," *NeuroImage*, vol. 52, no. 4, pp. 1374–1389, 2010.
- [95] H. Ong and F. Wehrli, "Quantifying axon diameter and intracellular volume fraction in excised mouse spinal cord with q-space imaging," *Neuroimage*, vol. 51, no. 4, pp. 1360–1366, 2010.
- [96] I. Aganj, C. Lenglet, G. Sapiro, E. Yacoub, K. Ugurbil, and N. Harel, "Reconstruction of the orientation distribution function in single- and multiple-shell q-ball imaging within constant solid angle," *Magnetic Resonance in Medicine*, vol. 64, no. 2, pp. 554–566, 2010.
- [97] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche, "Regularized, fast, and robust analytical q-ball imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 3, pp. 497–510, 2007.
- [98] M. M. Correia, *Development of Methods for the Acquisition and Analysis of Diffusion Weighted MRI Data*. PhD thesis, University of Cambridge, 2009.
- [99] D. Tuch, *Diffusion MRI of complex tissue structure*. PhD thesis, Massachusetts Institute of Technology, Division of Health Sciences and Technology, 2002.
- [100] V. J. Wedeen, R. P. Wang, J. D. Schmahmann, T. Benner, W. Y. Tseng, G. Dai, D. N. Pandya, P. Hagmann, H. D'Arceuil, and A. J. de Crespigny, "Diffusion spectrum magnetic resonance imaging (DSI) tractography of crossing fibers.,," *Neuroimage*, vol. 41, no. 4, pp. 1267–77, 2008.
- [101] K. Merboldt, W. Hanicke, H. Bruhn, M. Gyngell, and J. Frahm, "Diffusion imaging of the human brain in vivo using high-speed STEAM MRI," *Magnetic Resonance in Medicine*, vol. 23, no. 1, pp. 179–192, 1992.
- [102] P. Ramachandran and G. Varoquaux, "Mayavi: 3d visualization of scientific data," *Computing in Science & Engineering*, vol. 13, no. 2, pp. 40–51, 2011.

- [103] M. Correia, G. Williams, F.-C. Yeh, I. Nimmo-Smith, and E. Garyfallidis, "Robustness of diffusion scalar metrics when estimated with generalized q-sampling imaging acquisition schemes," in *Proceedings of the 19th Annual Meeting of the International Society of Magnetic Resonance in Medicine*, 2011.
- [104] E. Garyfallidis, M. Brett, B. Amirbekian, C. Nguyen, F. Yeh, Y. Halchenko, and I. Nimmo-Smith, "Dipy - a novel software library for diffusion MR and tractography," in *17th Annual Meeting of the Organization for Human Brain Mapping*, 2011.
- [105] M. Correia, T. Carpenter, and G. Williams, "Looking for the optimal DTI acquisition scheme given a maximum scan time: are more b-values a waste of time?," *Magnetic resonance imaging*, vol. 27, no. 2, pp. 163–175, 2009.
- [106] J. Rice, *Mathematical statistics and data analysis*. Thomson Learning, 2006.
- [107] P. Lee, *Bayesian statistics*. Arnold London, UK, 1997.
- [108] D. Montgomery, E. Peck, G. Vining, and J. Vining, *Introduction to linear regression analysis*, vol. 3. Wiley New York, 2001.
- [109] M. Catani, R. J. Howard, S. Pajevic, and D. K. Jones, "Virtual in vivo interactive dissection of white matter fasciculi in the human brain," *NeuroImage*, vol. 17, pp. 77–94, 2002.
- [110] Y. Chao, J. Chen, K. Cho, C. Yeh, K. Chou, and C. Lin, "A multiple streamline approach to high angular resolution diffusion tractography," *Medical engineering and physics*, vol. 30, no. 8, pp. 989–996, 2008.
- [111] M. Descoteaux, R. Deriche, T. R. Knösche, and A. Anwander, "Deterministic and probabilistic tractography based on complex fibre orientation distributions.,," *IEEE transactions on medical imaging*, vol. 28, pp. 269–86, Feb. 2009.
- [112] M. Lazar, D. M. Weinstein, J. S. Tsuruda, K. M. Hasan, K. Arfanakis, M. E. Meyerand, B. Badie, H. A. Rowley, V. Haughton, A. Field, and A. L. Alexander, "White matter tractography using diffusion tensor deflection.,," *Hum Brain Mapp*, vol. 18, no. 4, pp. 306–21, 2003.

- [113] G. J. M. Parker, H. A. Haroon, and C. A. M. Wheeler-Kingshott, "A framework for a streamline-based probabilistic index of connectivity (PiCo) using a structural interpretation of MRI diffusion measurements," *Journal of Magnetic Resonance Imaging*, vol. 18, no. 2, pp. 242–54, 2003.
- [114] T. Hosey, G. Williams, and R. Ansorge, "Inference of multiple fiber orientations in high angular resolution diffusion imaging," *Magnetic Resonance in Medicine*, vol. 54, no. 6, pp. 1480–1489, 2005.
- [115] O. Friman, G. Farneback, and C. Westin, "A Bayesian approach for stochastic white matter tractography," *Medical Imaging, IEEE Transactions on*, vol. 25, no. 8, pp. 965–978, 2006.
- [116] F. Zhang, E. Hancock, C. Goodlett, and G. Gerig, "Probabilistic white matter fiber tracking using particle filtering and von mises-fisher sampling," *Medical image analysis*, vol. 13, no. 1, pp. 5–18, 2009.
- [117] S. Pajevic and P. Basser, "Parametric and non-parametric statistical analysis of DT-MRI data," *Journal of Magnetic Resonance*, vol. 161, no. 1, pp. 1–14, 2003.
- [118] M. Lazar and A. Alexander, "Bootstrap white matter tractography (BOOT-TRAC)," *NeuroImage*, vol. 24, no. 2, pp. 524–532, 2005.
- [119] R. O'Gorman and D. Jones, "Just how much data need to be collected for reliable bootstrap DT-MRI?," *Magnetic resonance in medicine*, vol. 56, no. 4, pp. 884–890, 2006.
- [120] S. Chung, Y. Lu, and R. Henry, "Comparison of bootstrap approaches for estimation of uncertainties of dti parameters," *Neuroimage*, vol. 33, no. 2, pp. 531–541, 2006.
- [121] J. Berman, S. Chung, P. Mukherjee, C. Hess, E. Han, and R. Henry, "Probabilistic streamline q-ball tractography using the residual bootstrap," *Neuroimage*, vol. 39, no. 1, pp. 215–222, 2008.
- [122] H. Haroon, D. Morris, K. Embleton, D. Alexander, and G. Parker, "Using the model-based residual bootstrap to quantify uncertainty in fiber orientations from *q*-ball analysis," *Medical Imaging, IEEE Transactions on*, vol. 28, no. 4, pp. 535–550, 2009.

- [123] D. Jones, "Tractography gone wild: probabilistic fibre tracking using the wild bootstrap with diffusion tensor MRI," *Medical Imaging, IEEE Transactions on*, vol. 27, no. 9, pp. 1268–1274, 2008.
- [124] B. Whitcher, D. Tuch, J. Wisco, A. Sorensen, and L. Wang, "Using the wild bootstrap to quantify uncertainty in diffusion tensor imaging," *Human brain mapping*, vol. 29, no. 3, pp. 346–362, 2008.
- [125] B. Efron, "Bootstrap methods: another look at the jackknife," *The annals of Statistics*, vol. 7, no. 1, pp. 1–26, 1979.
- [126] G. Parker and D. Alexander, "Probabilistic monte carlo based mapping of cerebral connections utilising whole-brain crossing fibre information," in *Information Processing in Medical Imaging*, pp. 684–695, Springer, 2003.
- [127] G. Parker and D. Alexander, "Probabilistic anatomical connectivity derived from the microscopic persistent angular structure of cerebral tissue," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 360, no. 1457, p. 893, 2005.
- [128] M. Descoteaux, R. Deriche, T. Knoesche, and A. Anwander, "Deterministic and probabilistic tractography based on complex fibre orientation distributions," *IEEE Trans Med Imaging*, vol. 28, no. 2, pp. 269–86, 2009.
- [129] P. Cook, D. Alexander, and G. Parker, "Modelling noise-induced fibre-orientation error in diffusion-tensor MRI," in *Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on*, pp. 332–335, IEEE, 2004.
- [130] K. Seunarine, P. Cook, M. Hall, K. Embleton, G. Parker, and D. Alexander, "Exploiting peak anisotropy for tracking through complex structures," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pp. 1–8, IEEE, 2007.
- [131] S. Jbabdi, M. Woolrich, J. Andersson, and T. Behrens, "A Bayesian framework for global tractography," *Neuroimage*, vol. 37, no. 1, pp. 116–129, 2007.

- [132] G. Parker, C. Wheeler-Kingshott, and G. Barker, "Estimating distributed anatomical connectivity using fast marching methods and diffusion tensor imaging," *Medical Imaging, IEEE Transactions on*, vol. 21, no. 5, pp. 505–512, 2002.
- [133] J. Tournier, F. Calamante, D. Gadian, A. Connelly, *et al.*, "Diffusion-weighted magnetic resonance imaging fibre tracking using a front evolution algorithm," *NeuroImage*, vol. 20, no. 1, pp. 276–288, 2003.
- [134] J. Campbell, K. Siddiqi, V. Rymar, A. Sadikot, and G. Pike, "Flow-based fiber tracking with diffusion tensor and q-ball data: validation and comparison to principal diffusion direction techniques," *NeuroImage*, vol. 27, no. 4, pp. 725–736, 2005.
- [135] P. Fletcher, R. Tao, W. Jeong, and R. Whitaker, "A volumetric approach to quantifying region-to-region white matter connectivity in diffusion tensor MRI," in *Information Processing in Medical Imaging*, pp. 346–358, Springer, 2007.
- [136] X. Gigandet, *Global brain connectivity analysis by diffusion MR tractography: algorithms, validation and applications*. PhD thesis, École Polytechnique Fédérale de Lausanne, 2009.
- [137] Y. Iturria-Medina, E. Canales-Rodríguez, L. Melie-García, P. Valdés-Hernández, E. Martínez-Montes, Y. Alemán-Gómez, and J. M. Sánchez-Bornot, "Characterizing brain anatomical connections using diffusion weighted MRI and graph theory," *Neuroimage*, vol. 36, no. 3, pp. 645–660, 2007.
- [138] A. Zalesky, "DT-MRI fiber tracking: a shortest paths approach," *Medical Imaging, IEEE Transactions on*, vol. 27, no. 10, pp. 1458–1471, 2008.
- [139] S. Lifshits, A. Tamir, and Y. Assaf, "Combinatorial fiber-tracking of the human brain," *NeuroImage*, vol. 48, no. 3, pp. 532–540, 2009.
- [140] P. Fillard, C. Poupon, and J. Mangin, "A novel global tractography algorithm based on an adaptive spin glass model," *Medical Image Computing and Computer-Assisted Intervention*, pp. 927–934, 2009.

- [141] S. Sotropoulos, L. Bai, P. Morgan, C. Constantinescu, and C. Tench, “Brain tractography using Q-ball imaging and graph theory: Improved connectivities through fibre crossings via a model-based approach,” *NeuroImage*, vol. 49, no. 3, pp. 2444–2456, 2010.
- [142] B. Kreher, I. Mader, and V. Kiselev, “Gibbs tracking: a novel approach for the reconstruction of neuronal pathways,” *Magnetic Resonance in Medicine*, vol. 60, no. 4, pp. 953–963, 2008.
- [143] M. Reisert, I. Mader, C. Anastasopoulos, M. Weigel, S. Schnell, and V. Kiselev, “Global fiber reconstruction becomes practical,” *Neuroimage*, vol. 54, no. 2, pp. 955–962, 2011.
- [144] M. Lazar and A. Alexander, “White matter tractography using random vector (RAVE) perturbation,” in *Proceedings of ISMRM Annual Meeting, Honolulu*, 2002.
- [145] P. Fillard, C. Poupon, and J. Mangin, “Spin Tracking: A Novel Global Tractography Algorithm,” *NeuroImage*, vol. 47, pp. S127–S127, 2009.
- [146] P. Fillard, M. Descoteaux, A. Goh, S. Gouttard, B. Jeurissen, J. Malcolm, A. Ramirez-Manzanares, M. Reisert, K. Sakaie, F. Tensaouti, *et al.*, “Quantitative evaluation of 10 tractography algorithms on a realistic diffusion mr phantom,” *Neuroimage*, vol. 56, no. 1, pp. 220–234, 2011.
- [147] A. Sherbondy, M. Rowe, and D. Alexander, “Microtrack: An algorithm for concurrent projectome and microstructure estimation,” *Medical Image Computing and Computer-Assisted Intervention*, pp. 183–190, 2010.
- [148] P. Batchelor, D. Hill, D. Atkinson, and F. Calamante, “Study of connectivity in the brain using the full diffusion tensor from MRI,” in *Information Processing in Medical Imaging*, pp. 121–133, Springer, 2001.
- [149] N. Kang, J. Zhang, E. Carlson, and D. Gembri, “White matter fiber tractography via anisotropic diffusion simulation in the human brain,” *Medical Imaging, IEEE Transactions on*, vol. 24, no. 9, pp. 1127–1137, 2005.

- [150] N. Hageman, A. Toga, K. Narr, and D. Shattuck, “A diffusion tensor imaging tractography algorithm based on navier–stokes fluid mechanics,” *Medical Imaging, IEEE Transactions on*, vol. 28, no. 3, pp. 348–360, 2009.
- [151] P. Hagmann, J. Thiran, L. Jonasson, P. Vandergheynst, S. Clarke, P. Maeder, and R. Meuli, “DTI mapping of human brain connectivity: statistical fibre tracking and virtual dissection,” *Neuroimage*, vol. 19, no. 3, pp. 545–554, 2003.
- [152] S. Jbabdi, M. W. Woolrich, J. L. Andersson, and T. E. Behrens, “A Bayesian framework for global tractography,” *NeuroImage*, vol. 37, pp. 116–129, 2007.
- [153] S. Frey, J. Campbell, G. Pike, and M. Petrides, “Dissociating the human language pathways with high angular resolution diffusion fiber tractography,” *The Journal of Neuroscience*, vol. 28, no. 45, pp. 11435–11444, 2008.
- [154] J. Rilling, M. Glasser, T. Preuss, X. Ma, T. Zhao, X. Hu, and T. Behrens, “The evolution of the arcuate fasciculus revealed with comparative DTI,” *Nature Neuroscience*, vol. 11, no. 4, pp. 426–428, 2008.
- [155] M. Glasser and J. Rilling, “DTI tractography of the human brain’s language pathways,” *Cerebral Cortex*, vol. 18, no. 11, pp. 2471–2482, 2008.
- [156] N. Makris, D. Kennedy, S. McInerney, A. Sorensen, R. Wang, V. Caviness Jr, and D. Pandya, “Segmentation of subcomponents within the superior longitudinal fascicle in humans: a quantitative, in vivo, DT-MRI study,” *Cerebral Cortex*, vol. 15, no. 6, pp. 854–869, 2005.
- [157] T. Verstynen, K. Jarbo, S. Pathak, and W. Schneider, “In vivo mapping of microstructural somatotopies in the human corticospinal pathways,” *Journal of Neurophysiology*, vol. 105, no. 1, pp. 336–346, 2011.

- [158] Z. Ding, J. Gore, and A. Anderson, “Classification and quantification of neuronal fiber pathways using diffusion tensor MRI,” *Magn. Reson. Med.*, vol. 49, pp. 716–721, 2003.
- [159] M. Maddah, W. M. Wells, 3rd, S. K. Warfield, C. F. Westin, and W. E. Grimson, “Probabilistic clustering and quantitative analysis of white matter fiber tracts.,” *Inf Process Med Imaging*, vol. 20, pp. 372–83, 2007.
- [160] S. Zhang and D. Laidlaw, “DTI fiber clustering and cross-subject cluster analysis,” in *In proceedings of the Int. Soc. Magn. Reson. Med*, 2005.
- [161] I. Corouge, S. Gouttard, and G. Gerig, “Towards a shape model of white matter fiber bundles using diffusion tensor MRI,” in *International Symposium on Biomedical Imaging*, pp. 344–347, 2004.
- [162] E. Garyfallidis, M. Brett, and I. Nimmo-Smith, “Fast Dimensionality Reduction for Brain Tractography Clustering,” *16th Annual Meeting of the Organization for Human Brain Mapping*, 2010.
- [163] E. Visser, E. H. J. Nijhuis, J. K. Buitelaar, and M. P. Zwiers, “Partition-based mass clustering of tractography streamlines.,” *NeuroImage*, vol. 54, no. 1, p. 303.
- [164] B. Mobergss, A. Vilanova, and J. van Wijk, “Evaluation of fiber clustering methods for diffusion tensor imaging,” in *IEEE Visualization*, pp. 65–72, 2005.
- [165] X. Wang, W. Grimson, and C. Westin, “Tractography segmentation using a hierarchical Dirichlet processes mixture model,” *NeuroImage*, 2010.
- [166] M. Catani, D. Jones, *et al.*, “Perisylvian language networks of the human brain,” *Annals of neurology*, vol. 57, no. 1, pp. 8–16, 2005.
- [167] G. Gerig, S. Gouttard, and I. Corouge, “Analysis of brain white matter via fiber tract modeling,” in *Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE*, vol. 2, pp. 4421–4424, IEEE, 2004.

- [168] P. Guevara, C. Poupon, D. Rivière, Y. Cointepas, M. Descoteaux, B. Thirion, and J.-F. Mangin, “Robust clustering of massive tractography datasets.,” *NeuroImage*, vol. 54, pp. 1975–1993, Oct. 2010.
- [169] A. Tsai, C. Westin, A. Hero, and A. Willsky, “Fiber tract clustering on manifolds with dual rooted-graphs,” in *Computer Vision and Pattern Recognition, 2007. CVPR’07. IEEE Conference on*, pp. 1–6, IEEE, 2007.
- [170] S. Zhang, C. Demiralp, and D. H. Laidlaw, “Visualizing diffusion tensor mr images using streamtubes and streamsurfaces,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 9, pp. 454–462, 2003.
- [171] S. Zhang, S. Correia, and D. Laidlaw, “Identifying White-Matter Fiber Bundles in DTI Data Using an Automated Proximity-Based Fiber Clustering Method,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 5, p. 1044, 2008.
- [172] O. Zvitia, A. Mayer, and H. Greenspan, “Adaptive mean-shift registration of white matter tractographies,” in *5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro.*, pp. 692–695, 2008.
- [173] O. Zvitia, A. Mayer, R. Shadmi, S. Miron, and H. K. Greenspan, “Co-registration of white matter tractographies by adaptive-mean-shift and Gaussian mixture modeling.” *IEEE transactions on medical imaging*, vol. 29, pp. 132–45, Jan. 2010.
- [174] V. El Kouby, Y. Cointepas, C. Poupon, D. Rivière, N. Golestani, J. B. Poline, D. Le Bihan, and J. F. Mangin, “MR diffusion-based inference of a fiber bundle model from a population of subjects.,” *International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 8, pp. 196–204, Jan. 2005.
- [175] A. Brun, H. Knutsson, H. Park, M. Shenton, and C. Westin, “Clustering fiber traces using normalized cuts,” *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2004*, pp. 368–375, 2004.
- [176] Z. Ding, J. C. Gore, and A. W. Anderson, “Classification and quantification of neuronal fiber pathways using diffusion tensor MRI,” *Magnetic Resonance in Medicine*, vol. 49, no. 4, pp. 716–21, 2003.

- [177] I. Corouge, S. Gouttard, and G. Gerig, “A statistical shape model of individual fiber tracts extracted from diffusion Tensor MRI,” *Medical Image Computing and Computer Assisted Intervention*, pp. 671–679, 2004.
- [178] I. Corouge, P. T. Fletcher, S. Joshi, S. Gouttard, and G. Gerig, “Fiber tract-oriented statistics for quantitative diffusion tensor MRI analysis,” *Medical image analysis*, vol. 10, no. 5, pp. 786–98, 2006.
- [179] L. J. O’Donnell and C. F. Westin, “Automatic tractography segmentation using a high-dimensional white matter atlas,” *IEEE Trans Med Imaging*, vol. 26, no. 11, pp. 1562–75, 2007.
- [180] L. O’Donnell, C. Westin, and A. Golby, “Tract-based morphometry for white matter group analysis,” *NeuroImage*, vol. 45, no. 3, pp. 832–844, 2009.
- [181] M. Maddah, A. U. Mewes, S. Haker, W. E. Grimson, and S. K. Warfield, “Automated atlas-based clustering of white matter fiber tracts from DT-MRI,” *MICCAI*, vol. 8, no. 1, pp. 188–95, 2005.
- [182] M. Maddah, W. Grimson, and S. Warfield, “Statistical modeling and em clustering of white matter fiber tracts,” in *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*, pp. 53–56, IEEE, 2006.
- [183] S. Witelson, “Hand and sex differences in the isthmus and genu of the human corpus callosum,” *Brain*, vol. 112, no. 3, p. 799, 1989.
- [184] M. Maddah, L. Zollei, W. Grimson, C. Westin, and W. Wells, “A mathematical framework for incorporating anatomical knowledge in DT-MRI analysis,” in *5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 105–108, IEEE, 2008.
- [185] L. Jonasson, P. Hagmann, J. P. Thiran, and V. J. Wedeen, “Fiber tracts of high angular resolution diffusion MRI are easily segmented with spectral clustering,” in *ISMRM*, 2005.
- [186] R. Jianu, C. Demiralp, and D. Laidlaw, “Exploring 3D DTI fiber tracts with linked 2D representations,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 15, no. 6, pp. 1449–1456, 2009.

- [187] M. Chalmers, “A linear iteration time layout algorithm for visualising high-dimensional data,” in *Visualization'96. Proceedings.*, pp. 127–131, IEEE, 1996.
- [188] P. Eades, “A heuristic for graph drawing,” *Congressus numerantium*, vol. 42, pp. 149–160, 1984.
- [189] S. Durrleman, P. Fillard, X. Pennec, A. Trouvé, and N. Ayache, “A statistical model of white matter fiber bundles based on currents,” *Proceedings of the 21st Conference on Information Processing in Medical Imaging*, vol. 21, pp. 114–25, Jan. 2009.
- [190] S. Durrleman, P. Fillard, X. Pennec, A. Trouvé, and N. Ayache, “Registration, atlas estimation and variability analysis of white matter fiber bundles modeled as currents,” *NeuroImage*, vol. 55, no. 3, pp. 1073–1090, 2011.
- [191] A. Leemans and D. Jones, “A new approach to fully automated fiber tract clustering using affinity propagation,” in *ISMRM*, vol. 17, p. 856, 2009.
- [192] J. Malcolm, M. Shenton, and Y. Rathi, “Filtered Tractography: State estimation in a constrained subspace,” *Proceedings of MICCAI (Diffusion Modeling and Fiber Cup)*, pp. 122–133, 2009.
- [193] U. Ziyan, M. Sabuncu, W. Grimson, and C. Westin, “Consistency clustering: a robust algorithm for group-wise registration, segmentation and automatic atlas construction in diffusion MRI,” *International Journal of Computer Vision*, vol. 85, no. 3, pp. 279–290, 2009.
- [194] S. Wakana, A. Caprihan, M. M. Panzenboeck, J. H. Fallon, M. Perry, R. L. Gollub, K. Hua, J. Zhang, H. Jiang, P. Dubey, A. Blitz, P. van Zijl, and S. Mori, “Reproducibility of quantitative tractography methods applied to cerebral white matter,” *NeuroImage*, vol. 36, no. 1, pp. 630–644, 2007.
- [195] K. Hua, J. Zhang, S. Wakana, H. Jiang, X. Li, D. S. Reich, P. A. Calabresi, J. J. Pekar, P. C. van Zijl, and S. Mori, “Tract probability maps in stereotaxic spaces: Analyses of white matter anatomy and tract-specific quantification,” *NeuroImage*, vol. 39, no. 1, pp. 336–347, 2008.

- [196] R. Ng and J. Han, "CLARANS: A method for clustering objects for spatial data mining," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1003–1016, 2002.
- [197] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: A new data clustering algorithm and its applications," *Data Mining and Knowledge Discovery*, vol. 1, no. 2, pp. 141–182, 1997.
- [198] J. Kogan, *Introduction to clustering large and high-dimensional data*. Cambridge Univ Pr, 2007.
- [199] H. Steinhaus, "Sur la division des corp materiels en parties," *Bull. Acad. Polon. Sci*, vol. 1, pp. 801–804, 1956.
- [200] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 281–297, California, USA, 1967.
- [201] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, pp. 1–38, 1977.
- [202] T. Gonzalez, "Clustering to minimize the maximum intercluster distance," *Theoretical Computer Science*, vol. 38, pp. 293–306, 1985.
- [203] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.
- [204] D. G. Altman, *Practical statistics for medical research*. London: Chapman and Hall, 1995.
- [205] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 2004.
- [206] E. L. Lawler, *Combinatorial optimization: networks and matroids*. Mineola, N.Y.: Dover Publications, 2001.
- [207] G. Carpaneto, S. Martello, and P. Toth, "Algorithms and codes for the assignment problem," *Annals of Operations Research*, vol. 13, pp. 191–223, 1988.

- [208] G. Carpaneto, S. Martello, and P. Toth, “Linear assignment problem freeware.” www.assignmentproblems.com/APC_APS.htm, downloaded June 2011.
- [209] S. M. Smith, M. Jenkinson, H. Johansen-Berg, D. Rueckert, T. E. Nichols, C. E. Mackay, K. E. Watkins, O. Ciccarelli, Z. Cader, P. M. Matthews, and T. E. Behrens, “Tract-based spatial statistics: Voxel-wise analysis of multi-subject diffusion data,” *NeuroImage*, vol. 31, pp. 1487–1505, 2006.
- [210] D. Eads, “hcluster: Hierarchical clustering for scipy.” scipy-cluster.googlecode.com, 2008.
- [211] D. Dueck, *Affinity propagation: Clustering data by passing messages*. PhD thesis, University of Toronto, 2009.
- [212] A. Goh and R. Vidal, “Algebraic methods for direct and feature based registration of diffusion tensor images,” *Computer Vision–ECCV 2006*, pp. 514–525, 2006.
- [213] A. Leemans, J. Sijbers, S. De Backer, E. Vandervliet, and P. Parizel, “Multiscale white matter fiber tract coregistration: A new feature-based approach to align diffusion tensor data,” *Magnetic Resonance in Medicine*, vol. 55, no. 6, pp. 1414–1423, 2006.
- [214] A. Mayer and H. Greenspan, “Bundles of interest based registration of White Matter tractographies,” in *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pp. 919–922, IEEE, 2008.
- [215] A. Mayer and H. Greenspan, “Direct registration of white matter tractographies with application to atlas construction,” in *MICCAI 2007 Workshop Statistical Registration PairWise and GroupWise Alignment and Atlas Formation*, 2007.
- [216] A. Mayer, G. Zimmerman-Moreno, R. Shadmi, A. Batikoff, and H. Greenspan, “A supervised framework for the registration and segmentation of white matter fiber tracts,” *IEEE Transactions on Medical Imaging*, vol. 30, no. 1, pp. 131–145, 2011.

- [217] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [218] U. Ziyan, M. R. Sabuncu, L. J. O'Donnell, and C. F. Westin, "Nonlinear registration of diffusion MR images based on fiber bundles," *International Conference in Medical Image Computing and Computer-Assisted Intervention*, vol. 10, no. Pt 1, pp. 351–8, 2007.
- [219] V. Arsigny, O. Commowick, N. Ayache, and X. Pennec, "A Fast and Log-Euclidean Polyaffine Framework for Locally Linear Registration," *Journal of Mathematical Imaging and Vision*, vol. 33, pp. 222–238, Jan. 2009.
- [220] R. Fletcher, "Practical methods of optimization," *John and Sons, Chichester*, 1987.
- [221] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 187–198, 1997.
- [222] A. Roche, G. Malandain, X. Pennec, and N. Ayache, "The correlation ratio as a new similarity measure for multimodal image registration," *International Conference in Medical Image Computing and Computer-Assisted Intervention*, p. 1115, 1998.
- [223] E. Garyfallidis, S. Gerhard, P. Avesani, T. Nguyen, V. Tsiaras, Y. Halchenko, I. Nimmo-Smith, and E. Olivetti, "A software application for real-time, clustering-based exploration of tractographies," in *18th Annual Meeting of the Organization for Human Brain Mapping*, 2012.
- [224] J. Ioannidis, "Why most published research findings are false," *PLoS medicine*, vol. 2, no. 8, p. e124, 2005.
- [225] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Neural Networks, 1995. Proceedings., IEEE International Conference on*, vol. 4, pp. 1942–1948, IEEE, 1995.

- [226] I. Oikonomidis, N. Kyriazis, and A. Argyros, “Full dof tracking of a hand interacting with an object by modeling occlusions and physical constraints,” in *Computer Vision (ICCV), 2011 IEEE International Conference on*, pp. 2088–2095, IEEE, 2011.
- [227] S. Chamberlain, A. Hampshire, L. Menzies, E. Garyfallidis, J. Grant, B. Odlaug, K. Craig, N. Fineberg, and B. Sahakian, “Reduced Brain White Matter Integrity in Trichotillomania: A Diffusion Tensor Imaging Study,” *Archives of General Psychiatry*, vol. 67, no. 9, p. 965, 2010.
- [228] A. Zalesky, A. Fornito, M. Seal, L. Cocchi, C. Westin, E. Bullmore, G. Egan, and C. Pantelis, “Disrupted axonal fiber connectivity in schizophrenia,” *Biological psychiatry*, vol. 69, no. 1, pp. 80–89, 2011.
- [229] E. Garyfallidis, M. Brett, M. Tsiaras, G. Vogiatzis, and I. Nimmo-Smith, “Identification of corresponding tracks in diffusion MRI tractographies,” in *18th Proceedings of the International Society of Magnetic Resonance in Medicine*, 2010.
- [230] E. Panagiotaki, T. Schneider, B. Siow, M. Hall, M. Lythgoe, and D. Alexander, “Compartment models of the diffusion mr signal in brain white matter: a taxonomy and comparison.,” *NeuroImage*, vol. 59, no. 3, p. 2241, 2012.
- [231] D. Alexander, “A general framework for experiment design in diffusion MRI and its application in measuring direct tissue-microstructure features,” *Magnetic Resonance in Medicine*, vol. 60, no. 2, pp. 439–448, 2008.
- [232] V. Wedeen, D. Rosene, R. Wang, G. Dai, F. Mortazavi, P. Hagmann, J. Kaas, and W. Tseng, “The geometric structure of the brain fiber pathways,” *Science*, vol. 335, no. 6076, pp. 1628–1634, 2012.
- [233] F. Olver, D. Lozier, R. Boisvert, and C. Clark, “NIST handbook of mathematical functions,” 2010.
- [234] B. Frey and D. Dueck, “Clustering by passing messages between data points,” *Science*, vol. 315, no. 5814, p. 972, 2007.
- [235] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.