Geometric and variational methods for diffusion tensor MRI processing

Christophe Lenglet

To cite this version:

Christophe Lenglet. Geometric and variational methods for diffusion tensor MRI processing. Human-Computer Interaction [cs.HC]. Université de Nice Sophia Antipolis, 2006. English. NNT : . tel-00457463

HAL Id: tel-00457463
https://theses.hal.science/tel-00457463

Submitted on 17 Feb 2010

PhD THESIS

prepared at
INRIA Sophia Antipolis

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

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

DOCTOR OF SCIENCE
Specialized in Control, Signal and Image Processing

Geometric and Variational Methods for Diffusion Tensor MRI Processing

Christophe Lenglet

Advisers	Pr. Rachid Deriche	INRIA Sophia Antipolis, France
	Pr. Olivier Faugeras	INRIA Sophia Antipolis, France
Reviewers	Pr. Peter Basser	NICHD, USA
	Pr. Guillermo Sapiro	University of Minnesota, USA
	Pr. Carl-Fredrik Westin	Harvard University, USA
Examiners	Pr. Nicholas Ayache	INRIA Sophia Antipolis, France
	Dr. Maher Moakher	ENIT Tunis, Tunisia
Invited members	Pr. Stéphane Lehéricy	Hospital La Pitié-Salpêtrière, France
	Dr. Jean-Philippe Thirion	QuantifiCare S.A. Sophia Antipolis, France

UNIVERSITÉ NICE-SOPHIA ANTIPOLIS - UFR Sciences

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

THÈSE

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

présentée et soutenue par

Christophe LENGLET

Méthodes Géométriques et Variationnelles pour le Traitement d'IRM du Tenseur de Diffusion

Thèse dirigée par Pr. Rachid DERICHE et Pr. Olivier FAUGERAS

Soutenue le 12 décembre 2006

Composition du jury:

Rapporteurs	Pr. Peter BASSER	NICHD, USA
	Pr. Guillermo SAPIRO	University of Minnesota, USA
	Pr. Carl-Fredrik WESTIN	Harvard University, USA
Examinateurs	Pr. Nicholas AYACHE	INRIA Sophia Antipolis, France
	Dr. Maher MOAKHER	ENIT Tunis, Tunisie
Membres invités	Pr. Stéphane LEHÉRICY	Hôpital La Pitié-Salpêtrière, France
	Dr. Jean-Philippe THIRION	QuantifiCare S.A. Sophia Antipolis, France

Abstract

This thesis deals with the development of new processing tools for diffusion tensor Magnetic Resonance Imaging (MRI). This recent MRI technique is the unique non invasive method currently available to explore the microstructure of biological tissues like the human brain. It is thus of utmost importance to acquire a better understanding of the brain mechanisms and to improve the diagnosis of neurological disorders. However, because of the complexity of the data, this imaging modality raises a large amount of mathematical and computational challenges. We introduce new algorithms relying on Riemannian geometry, partial differential equations and front propagation techniques to process diffusion tensor MRI. The first part of this work is theoretical. After a few reminders about the human nervous system, MRI and important notions of differential geometry, we study the space of multivariate normal distributions. The introduction of a Riemannian structure on that space allows us to define statistics and intrinsic numerical schemes that will constitute the core of the algorithms proposed in the second part. The properties of that space are important for diffusion tensor MRI since tensors can be seen as covariance matrices of normal laws modeling the diffusion of water molecules at each voxel of the acquired volume. The second part of this thesis is methodological. We start with the introduction of original approaches for the estimation and regularization of diffusion tensor MRI. We then show how to evaluate the degree of connectivity between cortical areas. Next, we introduce a statistical surface evolution framework for the segmentation of those images. Finally, we propose a non-rigid registration method. The last part of this thesis is dedicated to the application of our tools to two important neuroscience problems: the analysis of the connections between the cerebral cortex and the basal ganglia, implicated in motor tasks, and the study of the anatomo-functional network of the human visual cortex. This work was done in collaboration with the Center for Magnetic Resonance Research of the University of Minnesota (Minneapolis), the centre IRMf of the hospital la Timone (Marseille) and the neuroradiology center of the hospital La Pitié-Salpêtrière (Paris).

Résumé

Cette thèse est consacrée au développement d'outils de traitement pour l'Imagerie par Résonance Magnétique (IRM) du tenseur de diffusion. Cette technique d'IRM récente est l'unique moyen non invasif disponible pour explorer la microstructure de tissus biologiques tels que ceux du cerveau humain. Elle est donc d'une grande importance pour mieux comprendre le fonctionnement du cerveau ou pour améliorer le diagnostic de pathologies neurologiques. Cependant, de part la complexité des images produites, elle soulève de nombreux problèmes mathématiques et computationnels. Nous proposons des méthodes basées sur la géométrie Riemannienne, les équations aux dérivées partielles et les techniques de propagation de front pour le traitement de ces images. La première partie de ce travail est théorique. Après quelques rappels sur le système nerveux de l'homme, l'IRM et sur des éléments de géométrie différentielle, nous étudions l'espace des lois normales multivariées. L'introduction d'une structure Riemannienne sur cet espace nous permet de définir des statistiques et des schémas numériques intrinsèques qui sont à la base des algorithmes proposés dans la seconde partie. Les propriétés de cet espace sont importantes pour l'IRM du tenseur de diffusion car ces tenseurs sont les matrices de covariance de lois normales modélisant la diffusion des molécules d'eau en chaque voxel du milieu imagé. La seconde partie de cette thèse est méthodologique. Nous commençons par introduire des approches originales pour l'estimation et la régularisation de l'IRM du tenseur de diffusion. Nous montrons ensuite comment évaluer le degré de connectivité entre aires corticales. Puis nous introduisons un modèle statistique d'évolution de surface permettant de segmenter ces images. Finalement, nous proposons une méthode de recalage non-rigide. La dernière partie de cette thèse est consacrée à l'application de nos outils à deux problèmes d'intérêt en neurosciences : l'analyse des connexions entre le cortex cérébral et les noyaux gris centraux, impliquées dans des tâches motrices et l'étude du réseau anatomo-fonctionnel du cortex visuel humain. Ces travaux ont été réalisés en collaboration avec le Center for Magnetic Resonance Research de l'Université du Minnesota (Minneapolis), le centre IRMf de l'hôpital de la Timone (Marseille) et le centre de neuroradiologie de l'hôpital La Pitié-Salpêtrière (Paris).

Acknowledgments

This thesis was performed in the Odyssée lab at INRIA Sophia-Antipolis, under the joint supervision of Rachid Deriche and Olivier Faugeras. It was funded by the INRIA Odyssée lab and the région Provence-Alpes-Côte-d'Azur.

First, I would like to thank Olivier Faugeras for having welcomed me in his research team. I also want to express my deep gratitude to Rachid Deriche and Olivier Faugeras, my advisers, for their trust and invaluable advice. Their experience and complementary points of view on my work encouraged me throughout the last three years.

Next, I would also like to express my deep thanks to Peter Basser, Guillermo Sapiro and Carl-Fredrik Westin for having accepted to review my manuscript despite the important amount of work this required. I am also extremely grateful to Nicholas Ayache, Maher Moakher, Stéphane Lehéricy and Jean-Philippe Thirion for their participation in my jury.

Many people have widely and in various ways contributed to different parts of the work presented in this manuscript and I want to acknowledge them here: Mikaël Rousson, Emmanuel Prados, Nicolas Wotawa, Stéphane Lehéricy, Jean-Philippe Pons, Maxime Descoteaux, Théo Papadopoulo, Carlos Castaño-Moraga, Jean-Luc Anton, Muriel Roth, Kaleem Siddiqi, Jennifer Campbell, Alfred Anwander, David Tschumperlé, Maureen Clerc, Pierre Kornprobst, Thierry Viéville and Marie-Cécile Lafont.

I also want to thank Kamil Ugurbil and Guillermo Sapiro who welcomed me many times at the Center for Magnetic Resonance Research and the Electrical and Computer Engineering department of the University of Minnesota. This collaboration was supported in part by grant NSF-0404617 and the INRIA Direction des Relations Européennes et Internationales (DREI) under the US-France (INRIA) Cooperative Research program. It was also partly funded by grants NIH R21-RR019771, NIH-RR008079, the MIND Institute and the Keck foundation. The French National Project ACI Obs-Cerv, the INRIA-FQNRT program and the

PAI PROCOPE respectively supported the collaborations with the Centre IRMf de Marseille (Hospital La Timone), the McConnell Brain Imaging Center (McGill University) and the Max Planck Institute for Human Cognitive and Brain Sciences.

Finally, the Odyssée lab is a great place to work and escape from work. I will never forget those last three years and all my friends: Frank, Max, Sylvain, Beubeu, Nico, Appo, Sandrine, Maria-José, Adrien, Jérôme, Pete, Mat, Elaine, Shaun, Demian and Jonathan.

Contents

CHAPTER 1

INTRODUCTION

CONTEXT

The development of Magnetic Resonance Imaging (MRI), and its application to neuroimaging, have brought in the last 20 years great hopes to explore, in a non-invasive manner, the anatomy and activity of the human brain. As the magnetic field strength of scanners increases (up to 9.4 Tesla today for human), as new acquisition techniques like parallel imaging appear and, as new and powerful processing tools are developed, we are getting closer to a better undertanding of the most complex structure of the human body. These improvements are also crucial for the rapid and accurate diagnosis of neurological disorders.

This thesis is dedicated to the development of new processing tools for diffusion tensor MRI, also known as Diffusion Tensor Imaging (DTI). Diffusion MRI is a technique introduced in the middle of the 80's by Le Bihan *et al.* [34, 35], Merboldt *et al.* [205] and Taylor *et al.* [278]. It provides a very sensitive probe of biological tissues architecture. Anatomical MRI enables us to distinguish and classify grey matter, white matter and cerebrospinal fluid. However, with this contrast, white matter retains a homogeneous aspect, preventing any observation of neural fibers and thus of neuronal connectivity. In order to access the neural fibers bundle architecture, anatomists used to perform cerebral dissection [87] and strychnine or other chemical markers neuronography [251], [266]. More recently neural pathways tracking based on local injection of chemical markers, and subsequent observation of the induced propagation yielded high-quality connectivity mapping in the cat and monkey cerebral cortex [317], [266]. As of today, diffusion MRI is the unique non-invasive technique capable of quantifying the anisotropic diffusion of water molecules in tissues like the human brain and muscles.

The diffusion phenomenon is the macroscopic physical process resulting from the random walk of water molecules. All microscopic particles permanently undergo a translation and rotation motion due to their thermal energy and also known as Brownian motion. This microscopic phenomenon, related to the bombardment by molecules obeying a Maxwellian velocity distribution, was formalized by A. Einstein in 1905 and identified as depending on temperature and viscosity of the domain, as well as particles size. Diffusion shows, at a larger scale, how molecules tend to move from low concentration areas to high concentration areas. The key concept that is of primary importance for diffusion MRI is that diffusion in biological tissues reflects their architecture at a microscopic scale. This is due to the fact that molecular motion is favored in directions aligned with fiber bundles and hindered in the orthogonal direction. Measuring, at each voxel, that very same motion along a certain number of sampling directions provides an exquisite insight into the local orientation of fibers and is known as Diffusion Weighted MRI.

Shortly after the first acquisitions of images characterizing the anisotropic diffusion of water molecules in vivo [214], [224], Basser *et al.* [25, 24, 23] proposed the rigorous formalism of the diffusion tensor model. It features an analytic means to precisely describe the three-dimensional nature of anisotropy in tissues. The diffusion tensor model encapsulates the averaged diffusion properties of water molecules inside a voxel with typical extent of 1 to 3mm into the 3×3 covariance matrix of a Gaussian distribution.

DTI has now proved to be extremely useful to study the normal and pathological human brain [37, 98]. It has been applied, for instance, to investigate the brain development [292, 101], the effects of aging [276], cerebral anatomo-functional networks [170, 134] as well as the structure of the thalamus and various fiber bundles [312, 299]. DTI is also particularly relevant to a wide range of clinical studies related, for example, to brain ischemia [2] or stroke detection [274], multiple sclerosis [132], Alzheimer's [257] or Parkinson's disease [264], schizophrenia [10], neurosurgery [277], tumor growth modeling [148, 75] ...etc

DTI raises a large amount of mathematical and computational challenges because of the complexity of the data and the need for adequate models. In this thesis, we address many different theoretical and computational issues related to the processing of diffusion tensor images. We also apply some of the proposed techniques to investigate the anatomo-functional architecture of the human motor and visual systems.

ORGANIZATION AND CONTRIBUTIONS OF THIS THESIS

This manuscript is organized in three parts, which reflect the different types of contributions of the thesis. They are threefold: theoretical, methodological and applied.

- Theoretical contributions deal with the introduction of a novel point of view to compute statistics on the manifold of multivariate Gaussian distributions. This will be at the core of many of our processing tools for DTI, since diffusion tensors can be seen as the parameters of three-dimensional Gaussian laws. We

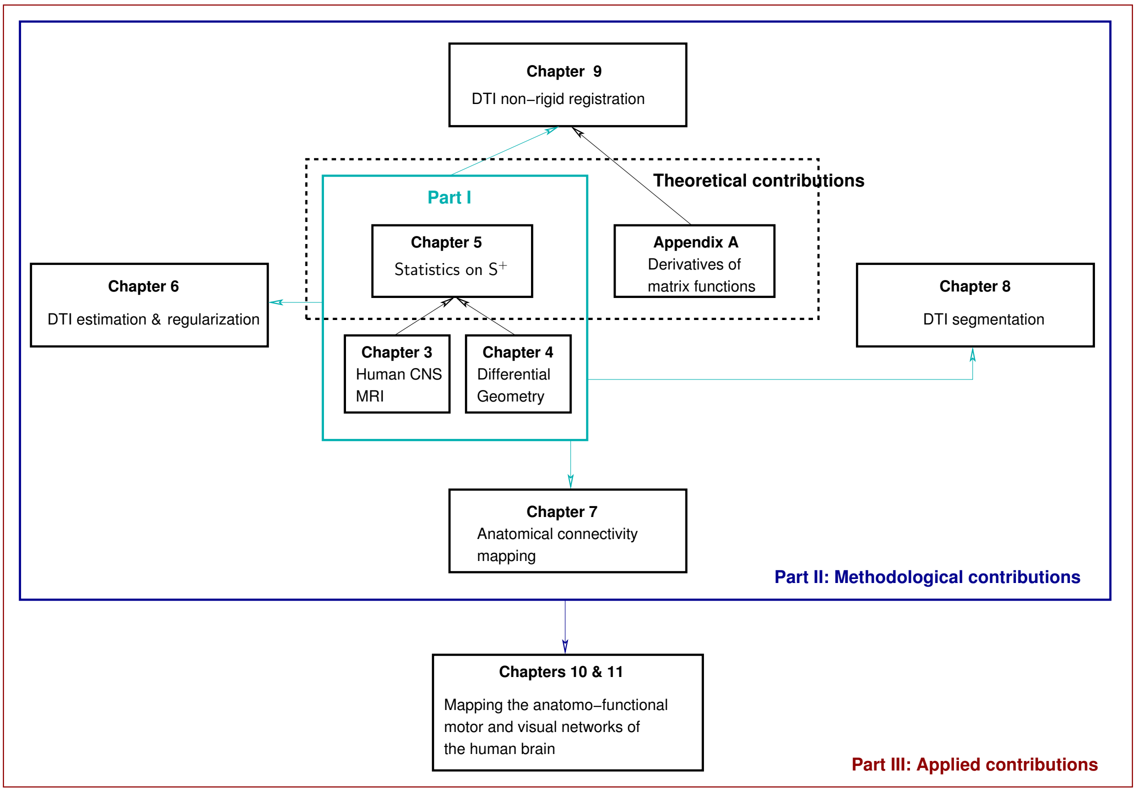


Figure 1.1: Graphical depiction of the manuscript organization

also address the problem of the fast computation of directional derivatives of matrix functions, which is a key element of our registration algorithm.

- Methodological contributions are related to the introduction of original techniques for the estimation, regularization, segmentation and registration of diffusion tensor images as well as for the mapping of the anatomical connectivity of the human brain from DTI.

- Applied contributions are concerned with the use and validation of our connectivity mapping techniques to study the human brain visual and motor systems.

Figure 1.1 provides a graphical depiction of the organization of this thesis.

Part I: Background

Chapter 3 is an introduction to the architecture of the human central nervous system and, in particular, the human brain. It also introduces the principles of Magnetic Resonance Imaging, with a particular emphasis on diffusion MRI. This chapter provides the basic knowledge necessary to understand the biological and physical basis of the diffusion MR signal.

Chapter 4 is meant to be a reminder of Riemannian geometry. After describing the notion of differentiable manifold, we introduce the concepts of metric, geodesic, connection and curvature. Finally, we show how to apply them to manifold-valued datasets. This should give to anyone, not familiar with Riemannian geometry, enough information to follow the developments in chapter 5.

Chapter 5 takes a differential geometrical point of view to analyze the space of multivariate Gaussian distributions. It relies on the properties of the underlying parameters space, endowed with a Riemannian metric, and allows us to overcome many problems encountered when using the Euclidean metric. We will see that this space is highly relevant for the study of diffusion tensor images because those tensors are nothing but the covariance matrices of three-dimensional Gaussian laws. We derive and experiment with original methods to compute the mean and covariance matrix of a set of diffusion tensors. We also show how to approximate a Gaussian law on the set of diffusion tensors. This chapter is at the core of the algorithms proposed in chapters 6, 8 and 9.

Part II: Methods

Chapter 6 addresses two of the most basic processing tasks for Diffusion Weighted Images (DWI), namely the estimation of diffusion tensors and the regularization of the resulting matrix-valued images. We show that it is possible to formulate these problems within the geometrical framework proposed in chapter 5 and thus to naturally enforce the positiveness of the computed or regularized diffusion tensors. The anisotropic regularization algorithm is a joint work with Carlos Castaño-Moraga (Center for Technology in Medicine, University of Las Palmas de Gran Canaria, Tafira, Spain).

Chapter 7 presents two different front propagation techniques, respectively based on level set and fast marching methods, to estimate and quantify the likelihood of anatomical connections in the human brain. This chapter uses the fact that a Riemannian metric can be derived from the diffusion tensor and used to compute geodesic distance between cerebral regions. We want to stress here that this metric (called “diffusion metric”) has absolutely no relation with the metric described in chapter 5. Loosely speaking, this last metric is a similarity measure between features (tensors), whereas the metric of chapter 7 is a spatial distance (in $\mathbb{R}^3$). Our goal in this chapter is to approximate neural fibers as shortest paths, ie. geodesics, associated with this diffusion metric. To achieve this goal, a key step of the algorithm is the computation of the geodesic distance function. It is obtained by solving the so-called anisotropic Eikonal equation. This is much more tricky than solving the classical Eikonal equation and requires specific numerical schemes. We adopt an historical point of view in this chapter. We start with the level set (or dynamic) formulation of the anisotropic Eikonal equation, which we developed first because of its easier numerical implementation. However, this approach has many drawbacks which led us to the Fast Marching (or stationary) formulation of the problem. We also show how to quantify the likelihood of each estimated fiber with the two methods. The control interpertation of the stationary formulation yields a very natural connectivity index that can be used to generate connectivity maps. Chapters 10 and 11 respectively use the dynamic and stationary formulation of the problem. The development of the theory and implementation of the Fast Marching method is a joint work with Emmanuel Prados (INRIA Rhône-Alpes, Grenoble, France) and Jean-Philippe Pons (CERTIS Laboratory, Ecole Nationale des Ponts et Chaussées, Marne-la-Vallée, France).

Chapter 8 focuses on the problem of the segmentation of diffusion tensor images. We propose a statistical surface evolution method to estimate the optimal partitioning according to two hypothesis: (1) Diffusion tensors exhibit different Gaussian distributions within the structure of interest and the background. (2) The surface evolution should stop wherever there exists a high variation in the diffusion tensor image (ie. when different fiber bundles meet). The goal of this chapter is thus to automatically extract a subpart of a DTI volume, corresponding to a fiber bundle of interest. We demonstrate through extensive numerical experimentation that the statistics derived from the Riemannian metric introduced in chapter 5 outperform the statistics associated to the Euclidean or Kullback-Leibler metrics. This is a joint work with Mikaël Rousson (Siemens Corporate Research, Princeton, NJ, USA).

Chapter 9 describes our last methodological contribution, namely a variational framework for the dense non-rigid registration of diffusion tensor images. We propose a matching energy that aims to minimize the difference in the local statistical content (means and covariance matrices) of two diffusion tensor images through a gradient descent procedure. The result of the algorithm is a dense vector field that can be used to warp one image onto the other. Of course, the statistics are computed as proposed in chapter 5. A large part of this chapter is essentially mathematical and the computations are a bit tedious. We provide, as a proof of concept, different numerical examples that illustrate the feasibility of the method.

Part III: Applications

Chapter 10 is an application of the level set based connectivity mapping technique introduced in chapter 7. Once again, this study relies on the dynamic formulation of the anisotropic Eikonal equation because of historical reasons. It was initiated about two years ago and it is in fact because of its high computational overhead that our work on the stationary formulation was started. This chapter demonstrates that it is possible, with this method, to identify different cortico-striatal networks involved in the acquisition and storage of motor skills. This is a joint work with Stéphane Lehéricy (Center for Magnetic Resonance Research, University of Minnesota, Minneapolis, USA / La Pitié-Salpêtrière Hospital, Paris, France).

Chapter 11 studies the anatomical connections between the various visual areas of the human brain. We not only reproduce previous findings on the topology of fibers linking the LGN to V1 or homologous regions of the visual cortex but also bring new insights on the connectivity between the human MT complex and the retinotopic areas. This study has been made possible by the development of our Fast Marching method because the number and extent of the regions of interest simply prohibited the computation of all the required geodesic distance functions with the level set approach. This is a joint work with Nicolas Wotawa (INRIA Sophia-Antipolis, France).

Appendices

Appendix A introduces an original formulation for the computation of directional derivatives of matrix functions. It is based on the spectral decomposition of matrices and we will show that it yields a fast algorithm. The case of the matrix logarithm is studied since it is a fundamental building block of the registration method presented in chapter 9. This is a joint work with Théo Papadopoulo (INRIA Sophia-Antipolis, France).

Appendix B provides all the details for the computation of the gradient of the matching term $\mathcal{J}_{\mathrm{AC}}(h, Dh)$ for the registration problem of chapter 9.

Appendix C presents the publications of the author.

Software contributions

Finally, we would like to point out that all the algorithms presented in this manuscript are now available upon request as an extension of the Brainvisa[1] software platform for visualization and analysis of multi-modality brain data and as a toolbox for SPM5[2]. This integration work has been done in collaboration with Demian Wassermann (INRIA Sophia-Antipolis, France / Computer Science Department, University of Buenos Aires, Argentina) and Maxime Descoteaux (INRIA Sophia-Antipolis, France).

---

[1] `http://brainvisa.info`

[2] `http://www.fil.ion.ucl.ac.uk/spm/software/spm5/`

CHAPTER 2

INTRODUCTION (FRANÇAIS)

CONTEXTE

Le développement de l'Imagerie par Résonance Magnétique (IRM), et ses applications en neuroimagerie, a apporté de grands espoirs au cours des 20 dernières années pour l'exploration non-invasive de l'anatomie et de l'activité du cerveau humain. Avec la croissance de la puissance des champs magnétiques (jusqu'à 9.4 Tesla actuellement pour l'homme), avec l'arrivée de nouvelles techniques d'acquisition comme l'imagerie parallèle et, avec le développement de nouveaux et puissants outils de traitement, notre compréhension de la structure la plus complexe du corps humain s'améliore progressivement. Ces avancées sont cruciales pour le diagnostic rapide et aussi précis que possible des pathologies neurologiques.

Cette thèse est dédiée au développement de nouveaux algorithmes de traitement pour l'IRM du Tenseur de Diffusion, également connu sous le nom d'Imagerie du Tenseur de Diffusion (ITD). L'IRM de diffusion est une technique introduite au milieu des années 80 par Le Bihan *et al.* [34, 35], Merboldt *et al.* [205] et Taylor *et al.* [278]. Sensible au mouvement des molécules d'eau, elle permet d'inférer des informations sur l'architecture des tissus biologiques étudiés. Les IRM anatomiques nous permettent en effet de distinguer et classifier la matière grise, blanche et le liquide céphalo-rachidien. Cependant, avec ce contraste, la matière blanche conserve un aspect homogène, empêchant toute observation des fibres nerveuses et donc de la connectivité cérébrale. Afin d'accéder à la configuration des faisceaux de fibres, les anatomistes réalisaient des dissections du cerveau [87], des neuronographies à base de strychnine ou autres marqueurs chimiques [251], [266]. Plus récemment, l'estimation des connexions neuronales par injection de marqueurs chimiques et observation de leur propagation a permis d'obtenir des cartes de connectivité de haute qualité du cortex cérébral chez le chat et le singe [317], [266].

Actuellement, l'IRM de diffusion est la seule technique non-invasive permettant de quantifier l'anisotropie de la diffusion des molécules d'eau dans des tissus comme le cerveau ou les muscles de l'homme. Le phénomène de diffusion est le processus physique macroscopique résultant du mouvement aléatoire des molécules d'eau. Toutes les particules microscopiques subissent en effet en permanence un mouvement de translation et de rotation dû à leur énergie thermique, et également connu sous le nom de mouvement Brownien. Ce phénomène microscopique, dont l'origine est l'incessant bombardement moléculaire de particules obéissant à une distribution de vitesse Maxwellienne, a été formalisé en 1905 par A. Einstein et caractérisé comme dépendant de la température, de la viscosité du domaine et de la taille des particules. La diffusion traduit, à une échelle plus grande, la tendance des molécules à migrer des régions à basse concentration vers les régions à forte concentration. Le concept clé, essentiel pour l'IRM de diffusion, est que tout processus de diffusion dans un milieu biologique reflète directement l'architecture de ce dernier à une échelle microscopique. Ceci s'explique par le fait que le mouvement moléculaire est favorisé dans les directions alignées avec les faisceaux de fibres et contraint dans les directions orthogonales. La mesure, à chaque voxel, de ce mouvement selon un certain nombre de directions fournit de précieuses informations sur l'orientation locale des fibres. Les données ainsi obtenues constituent des IRM pondérées en diffusion.

Peu après les premières acquisitions d'images caractérisant la diffusion anisotrope des molécules d'eau in vivo [214], [224], Basser *et al.* [25, 24, 23] ont proposé le formalisme rigoureux du modèle du tenseur de diffusion. Celui-ci fournit une forme analytique décrivant précisément la nature tridimensionnelle de l'anisotropie des tissus. Le modèle du tenseur de diffusion encapsule dans la matrice de covariance 3×3 d'une distribution Gaussienne les propriétés de diffusion moyenne des molécules d'eau au sein d'un voxel dont la taille est, typiquement, de 1 à 3 mm.

Il est maintenant admis que l'ITD constitue un outil extrêmement utile pour étudier le cerveau humain sain ou pathologique [37, 98]. Cette modalité a d'ores et déjà été appliquée, par exemple, pour mieux comprendre le développement cérébral [292, 101], l'effet du vieillissement [276], l'architecture des réseaux anatomo-fonctionnels cérébraux [170, 134] ainsi que la structure du thalamus et de différents faisceaux de fibres [312, 299]. L'ITD est aussi particulièrement pertinente pour un grand nombre d'études cliniques s'intéressant par exemple aux ischémies cérébrales [2] ou à la détection d'accidents vasculaires cérébraux [274], la sclérose en plaques [132], les maladies d'Alzheimer [257] ou de Parkinson citescherfler-schocke-etal:06, la schizophrénie [10], la neurochirurgie [277], la modélisation de la croissance de tumeurs [148, 75] ...etc.

L'ITD, de part la complexité des données et le besoin de modèles adéquats, soulève un grand nombre de questions mathématiques et computationnelles. Dans cette thèse, nous nous attaquons à de nombreux problèmes théoriques et computationnels, reliés au traitement d'IRM du tenseur de diffusion. Nous appliquons également certains des outils proposés à l'analyse de l'architecture anatomo-fonctionnelle des systèmes moteurs et visuels humains.

ORGANISATION ET CONTRIBUTIONS DE CETTE THÈSE

Ce manuscrit est organisé en trois parties qui reflètent les différents types de contributions de cette thèse : théoriques, méthodologiques et appliquées.

- Nos contributions théoriques concernent l'introduction d'un point de vue original pour le calcul de statistiques sur la variété des distributions Gaussiennes multivariées. Ceci sera à la base d'un certain nombre de nos outils de traitement pour l'ITD, étant donné que les tenseurs de diffusion peuvent être considérés comme les paramètres de lois Gaussiennes tridimensionnelles. Nous nous intéressons également au problème du calcul rapide de dérivées directionnelles de fonctions matricielles. Ceci est un élément clé de notre algorithme de recalage.

- Nos contributions méthodologiques consistent en l'introduction de techniques originales pour l'estimation, la régularisation, segmentation et le recalage d'images du tenseur de diffusion ainsi que pour l'analyse de la connectivité anatomique du cerveau humain par ITD.

- Nos contributions appliquées sont liées à l'utilisation et la validation de nos techniques d'analyse de la connectivité à travers l'étude des systèmes moteurs et visuels du cerveau humain.

La figure 2.1 fournit une description graphique de l'organisation de cette thèse.

Partie I : Préliminaires et Théorie

Chapitre 3 - Le Système Nerveux Humain et l'Imagerie par Résonance Magnétique : Le cerveau humain, selon le point de vue connexionniste, est organisé en régions de traitement distinctes et connectées par un réseau de relais anatomiques. Les unités de traitement assurent les fonctions cognitives primaires tandis que les tâches cognitives de plus haut niveau émergent d'une coordination globale entre ces unités. Le signal neural est analysé par le cortex cérébral et transmis aux diverses aires du cerveau via la matière blanche. Celle-ci doit son nom à l'apparence blanche des axones myélinisés. Les faisceaux de la matière blanche sont, en général, classifiés en fibres commissurales, associatives et projectives, en fonction des régions qu'elles relient. Comparé à notre compréhension de l'architecture cérébrale chez des animaux comme le chat ou la souris, où l'utilisation de procédés invasifs est possible [266], notre connaissance de l'organisation du cerveau humain est assez pauvre.

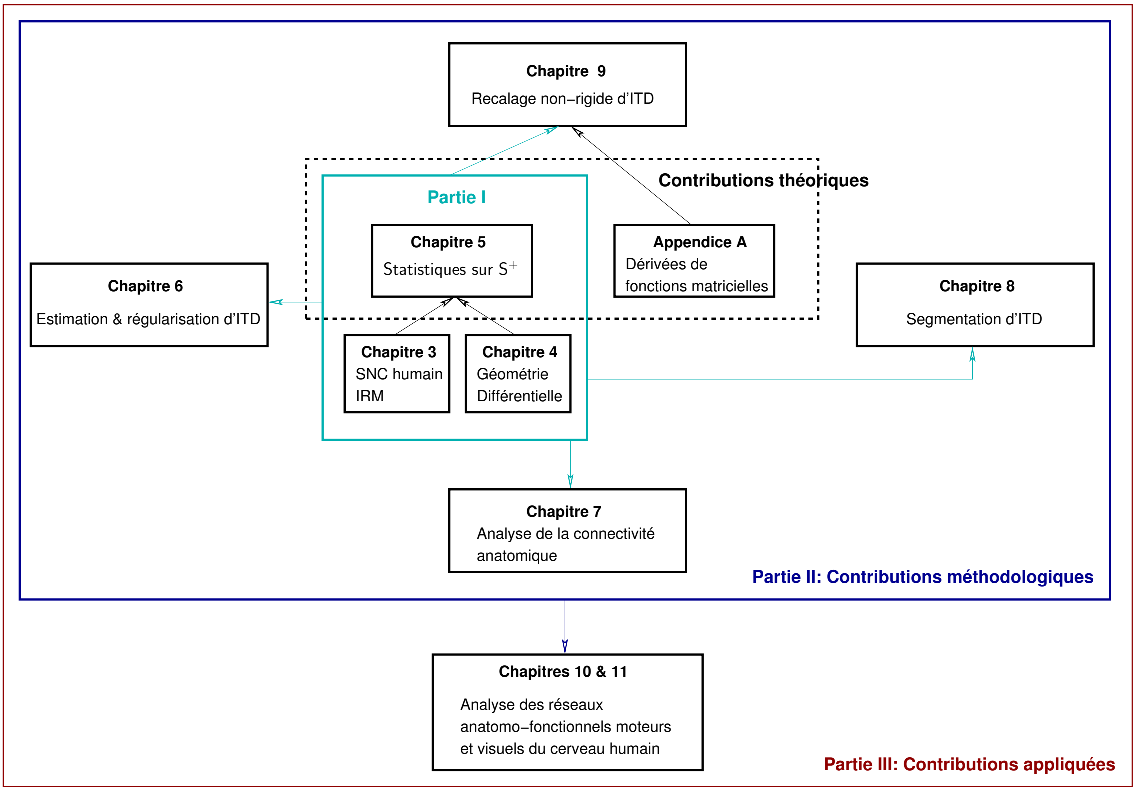


Figure 2.1: Description graphique de l'organisation de ce manuscrit

Les techniques de neuroimagerie comme l'Imagerie par Résonance Magnétique, et en particulier l'IRM de diffusion, ont ouvert de nouveaux horizons pour accéder à cette précieuse information. L'IRM est, de loin, l'application la plus connue de la Résonance Magnétique Nucléaire (RMN) pour le diagnostic médical. Cependant, la RMN est aussi largement utilisée en chimie dans le cadre de la spectroscopie par RMN, i.e. l'étude de la composition atomique d'échantillons. La RMN fut décrite simultanément par Felix Bloch [39], à l'université de Stanford, et par Edward Mills Purcell [252], à l'université de Harvard en 1946. En 1952, ils reçurent le Prix Nobel de Physique pour leur découverte. Le principe de base de la RMN est le suivant : Après avoir aligné un noyau magnétique comme l'hydrogène-1 avec un champ magnétique externe puissant, la réponse de ce noyau à une perturbation de son alignement par un champ électromagnétique lui est propre. En 1950, quatre années après cette découverte, Herman Carr proposa de créer les premières images MR unidimensionnelles en introduisant un gradient dans le champ magnétique. En 1971, Raymond Damadian démontra que les temps de relaxation T1 et T2 de tissus tumoraux étaient significativement plus longs que ceux de tissus sains, ce qui ouvrit de grands espoirs pour améliorer les moyens de détection des cancers. Peu après la découverte de Bloch et Purcell, Hahn publia un article séminal [137] sur l'écho de spin RMN dans lequel il nota que le mouvement Brownien des spins réduisait l'amplitude du signal observé en présence d'un champ magnétique inhomogène. Il

s'agit d'une notion fondamentale pour comprendre l'IRM de diffusion.
Dès 1973, Paul Lauterbur proposa une méthode [171], basée sur des gradients de champs magnétiques, pour reconstruire des IRM bidimensionnelles. Peter Mansfield [203] prolongea ces travaux et, en étudiant les propriétés mathématiques du signal RMN, proposa une nouvelle technique d'acquisition ultra-rapide, connue sous le nom de technique écho-planaire. En 2003, ils reçurent le Prix Nobel de Physiologie et de Médecine pour leur découvertes sur l'IRM. Actuellement, de nombreux types d'images peuvent être obtenus grâce à un scanner IRM. Elles nous renseignent sur l'anatomie, l'activité fonctionnelle, les propriétés de diffusion des molécules d'eau, le flux sanguin, la distribution de divers métabolites ou la structure des vaisseaux sanguins.
Ce chapitre est divisé en deux parties. Nous commençons par une rapide description du système nerveux humain, et en particulier de la matière grise et la matière blanche. Nous introduisons ensuite la technique d'IRM et montrons comment elle peut être exploitée pour accéder à la connectivité anatomique cérébrale.

Chapitre 4 - Notions de Géométrie Différentielle : La géométrie différentielle, et en particulier la géométrie Riemannienne, joue un rôle important dans cette thèse et ce chapitre donne un aperçu des notions importantes de ce domaine. Développée par Bernhard Riemann au cours du 19ème siècle, la géométrie Riemannienne étudie les variétés lisses équipées d'une métrique Riemannienne. Ces deux concepts fondamentaux sont expliqués dans ce chapitre. Une métrique Riemannienne fournit une information locale sur les angles, la longueur des courbes et les volumes, ce qui rend la géométrie Riemannienne particulièrement adaptée pour l'étude des courbes et surfaces.
Après la description d'une variété différentiable, nous introduisons la notion de métrique et de géodésique. Ensuite, nous discutons les concepts de connexion et de transport parallèle qui seront importants pour le chapitre 9. Enfin, nous définissons la courbure d'une variété car cette notion sera nécessaire pour le chapitre 5. Nous montrons comment appliquer toutes ces notions à des images prenant valeurs dans une variété Riemannienne, pour le calcul de statistiques et de gradients spatiaux.

Chapitre 5 - La Variété des Distributions Normales Multivariées : Ce chapitre est dédié à l'analyse statistique de l'espace des distributions normales multivariées, avec une application au traitement des images du tenseur de diffusion. La définition de structures différentielles géométriques pour les modèles statistiques commença en 1936 avec le travail de Mahalanobis [200] sur les distributions normales multivariées à matrice de covariance fixe. Depuis lors, plusieurs auteurs en géométrie de l'information [199], [9] et références associées, et physique [57] ont contribué à la description de ces géométries. Rao [253] exprima l'un des résultats fondamentaux en 1945 en montrant qu'il était possible d'utiliser la matrice d'information de Fisher comme une métrique Riemannienne entre densités de probabilité paramétrées. L'information de Fisher est une mesure populaire de la quantité d'information portée par les réalisations d'une variable aléatoire à propos des paramètres inconnus de la densité sous-jacente. Elle est souvent utilisée pour dériver des estimateurs au maximum de vraisemblance pour les paramètres de densités. En 1982, Burbea et Rao [45] ont proposé une approche unifiée de la dérivation de métriques dans des espaces de densités de probabilité. Ils ont introduit la notion de ϕ-fonctionnelle dont la matrice Hessienne selon une direction de l'espace tangent de l'espace des paramètres est prise comme métrique. Prolongeant le travail séminal de Rao [253] et un théorème de Jensen (1976, communication privée dans [15]), Atkinson et Mitchell obtinrent des expressions en forme close pour les distances géodésiques entre éléments de familles de distributions connues telles que les densités normales multivariées à moyenne fixe.

Dans ce chapitre, nous nous focalisons tout d'abord sur les propriétés géométriques de ces distributions particulières et utilisons les résultats présentés dans [272], [44], [48] et [117] pour proposer un nouveau cadre d'analyse statistique des densités normales multivariées. Nous montrons ensuite comment approximer une loi normale sur la variété des densités normales et dérivons les schémas numériques adaptés. Finalement, nous décrivons un algorithme simple pour générer des densités normales aléatoires suivant une loi normale imposée. Nous montrons également comment réaliser l'interpolation d'images du tenseur diffusion et illustrons ces techniques par des expériences numériques.

Partie II : Méthodes

Chapitre 6 - Des Images Pondérées en Diffusion aux Images du Tenseur de Diffusion : Ce chapitre s'attaque à deux des traitements les plus fondamentaux pour les Images Pondérées en Diffusion (IPD) : l'estimation des tenseurs de diffusion et la régularisation des images à valeur matricielle ainsi obtenues. Nous montrons comment formuler ces problèmes dans le cadre géométrique proposé au chapitre 5 et ainsi, comment naturellement préserver les propriétés des tenseurs de diffusion.

Comme nous le décrivons au chapitre 3, l'ITD estime, depuis un ensemble d'images pondérées en diffusion (dépendantes de la direction et de l'amplitude des gradients de diffusion), la matrice de covariance Σ du mouvement Brownien des molécules d'eau à chaque voxel du volume d'acquisition $\Omega \subset \mathbb{R}^3$. En d'autres termes, l'ITD approxime la densité de probabilité modélisant le mouvement des molécules d'eau par une distribution normale tridimensionnelle de moyenne nulle $\overline{r} \in \mathbb{R}^3$. L'estimation du champ de matrices 3×3 symétriques et définies positives est réalisée grâce à l'équation de Stejskal-Tanner [275] pour la diffusion anisotrope. Cette équation relie l'atténuation du signal de résonance magnétique au tenseur de diffusion et aux paramètres de la séquence d'acquisition de la façon suivante :

$$S_i(x) = S_0(x) \exp\left(-b g_i^T \Sigma(x) g_i\right) \quad \forall x \in \Omega, i = 1, ..., N \tag{2.1}$$

$g_i = (g_i^1, g_i^2, g_i^3)^T, \quad i = 1, ..., N$ sont les directions des gradients de diffusion correspondants à chaque image pondérée en diffusion S_i et b est le facteur de pondération en diffusion. De plus, une image de référence S_0, sans gradient de diffusion, doit être acquise [27], [157] , [140]. De nombreuses techniques ont déjà été implémentées pour estimer les tenseurs de diffusion $\Sigma(x)$ depuis un ensemble d'IPD (au nombre minimum de 6 puisqu'un tenseur a 6 degrés de liberté) et nous référons le lecteur aux articles suivants pour plus de détails [310], [202], [305], [285] [60], [217] [167, 168]. Nous appuyons également sur les divers travaux, tels que [139] ou [227], qui ont étudié l'influence du choix des gradients de diffusion sur la qualité des tenseurs estimés. De manière générale, il est judicieux d'utiliser plus de 6 gradients de diffusion de façon à minimiser la propagation du bruit des IPD aux composantes des tenseurs. Comme démontré dans [156], un minimum de 30 directions est nécessaire pour réaliser une estimation robuste du coefficient de diffusion apparente, de l'anisotropie fractionnaire (voir section 6.1.4) et des vecteurs propres du tenseur.

Nous commençons par rappeler la méthode classique des moindres carrés et introduisons ensuite une approche robuste et performante pour l'estimation des tenseurs. Les méthodes de régularisation et de filtrage pour les champs de tenseurs sont largement étudiées dans la littérature, en particulier dans le domaine de l'ITD. Par exemple, [285] s'attaque à ce problème en résolvant une équation aux dérivées partielles afin de régulariser des champs bruités de tenseurs de diffusion, i.e. des matrices $n \times n$ symétriques et définies positives. Une approche complémentaire, présentée dans [63], fournit une interprétation géométrique des flots de fonctions à valeur matricielle. Elle aboutit à des schémas numériques basés sur l'application exponentielle et préservant naturellement les propriétés des tenseurs. Dans [308], les auteurs généralisent les processus de diffusion anisotropes et non-linéaires aux données à valeur matricielle. Plus récemment, [80] a proposé une méthode variationnelle restaurant la direction principale des tenseurs tout en utilisant le résultat de ce processus pour régulariser les valeurs propres via un processus de diffusion anisotrope. D'autres travaux ont couplé le processus de régularisation avec l'estimation des tenseurs depuis les images pondérées en diffusion. Par exemple, [305] s'appuie sur une formulation variationnelle contrainte nécessitant la minimisation d'un terme de régularisation. Ce dernier est basé sur des normes L^p et est sujet à des contraintes non-linéaires sur les données pour obtenir des tenseurs définis positifs. Récemment, [311] a introduit des techniques de régularisation basées sur la convolution normalisée ou les champs aléatoires de Markov.

Nous nous appuyons sur les résultats mathématiques présentés au chapitre 5 pour développer un algorithme de filtrage anisotrope contrôlé par l'amplitude du gradient spatial du champ de tenseur. Une analyse détaillée des performances de notre

approche est ensuite réalisée et des résultats quantitatifs et qualitatifs obtenus sur des données synthétiques, par comparaison avec [308], démontrent la supériorité de cette méthode.
L'algorithme de régularisation anisotrope est un travail réalisé en collaboration avec Carlos Castaño-Moraga (Center for Technology in Medicine, University of Las Palmas de Gran Canaria, Tafira, Spain).

Chapitre 7 - Analyse de la Connectivité Anatomique du Cerveau Humain : Dans ce chapitre, nous introduisons une approche originale pour l'analyse de la connectivité de la matière blanche cérébrale à partir d'ITD. Notre méthode s'appuie sur une modélisation globale du volume IRM $\Omega \subset \mathbb{R}^3$ comme une variété Riemannienne dont la métrique dérive directement du tenseur de diffusion. Il est important de bien distinguer la métrique utilisée dans ce chapitre de la métrique obtenue à partir de la matrice d'information de Fisher et utilisée dans les chapitres précédents pour comparer les tenseurs de diffusion. Dans la suite, ces tenseurs sont utilisés pour obtenir une “métrique de diffusion” (voir [83]) et pour mesurer des distances physiques tridimensionnelles entre différents voxels d'une image du tenseur de diffusion. Le concept clé de ce chapitre est la notion de distance géodésique puisqu'elle va nous permettre de trouver les chemins optimaux de la matière blanche et ainsi d'estimer les faisceaux de fibres nerveuses. La fonction distance est la solution de deux problèmes théoriquement équivalents mais, en pratique, extrêmement différents. Ils dérivent tous deux de la notion de propagation de front et constituent des formulations équivalentes de l'équation Eikonale :

- Un problème de valeur initiale de la forme $\frac{\partial \psi}{\partial t} + F|\nabla \psi| = 0$. F est la vitesse de propagation du front et ce point de vue est intrinsèquement *dynamique*. Il sera implémenté en ayant recours au cadre des ensembles de niveaux.

- Un problème de frontière de la forme $|\nabla \phi| = \frac{1}{F}$ qui est, au contraire, intrinsèquement *stationnaire*. Il sera implémenté en ayant recours à la méthode dite de cheminement rapide.

Comme nous le montrons dans ce chapitre, ces deux approches ont des propriétés très différentes qui les rendent plus ou moins adéquates pour notre problème et plus ou moins computationnellement efficaces. La formulation dynamique est assez aisée à implémenter mais présente de nombreux inconvénients. Au contraire, la formulation stationnaire repose sur la théorie du contrôle et s'avère beaucoup plus délicate à mettre en oeuvre. Cependant, elle a de nombreux avantages (comme la rapidité et la robustesse) qui la rendent plus appropriée pour notre problème d'analyse de la connectivité cérébrale. Une fois la distance géodésique calculée, il est nécessaire d'appliquer un algorithme de rétro-propagation sur le gradient de cette distance pour retrouver les chemins optimaux. Comme nous le détaillons dans ce chapitre, l'approche stationnaire produit naturellement ce champ de vecteur sans avoir à

différencier la fonction distance, d'où sa robustesse. Finalement, nous présentons différentes mesures de connectivité possibles, reflétant le degré de connectivité entre régions du cerveau. Une mesure naturelle est dérivée depuis la formulation stationnaire du problème. Nous illustrons ces notions sur des données synthétiques et réelles.
Après avoir introduit la notion de "métrique de diffusion" et détaillé les formulations dynamique et statique de l'équation Eikonale, nous montrons comment les statistiques, le long des géodésiques, de notre mesure de connectivité permettent de détecter avec une certaine confiance les connexions de la matière blanche.
Le développement de la théorie et l'implémentation de la méthode de cheminement rapide est un travail réalisé en collaboration avec Emmanuel Prados (INRIA Rhône-Alpes, Grenoble, France) et Jean-Philippe Pons (CERTIS Laboratory, Ecole Nationale des Ponts et Chaussées, Marne-la-Vallée, France).

Chapitre 8 - Segmentation d'Images du Tenseur de Diffusion : L'ITD est donc précieuse pour identifier les connexions neuronales du cerveau humain. Cependant, la plupart des techniques d'estimation de cette connectivité travaillent à "l'échelle" de la fibre. En d'autres termes, elles ne prennent pas en compte la cohérence globale qui existe entre fibres d'un même faisceau. Le travail récent de Corouge *et al.* [79] propose de grouper et d'aligner les fibres de façon à rendre possible une analyse statistique des propriétés géométriques et physiologiques des faisceaux. Un travail similaire d'O'Donnell *et al.* [219] a été récemment introduit pour grouper ces fibres. Ces méthodes reposent sur l'extraction d'un ensemble de courbes depuis des images du tenseur de diffusion par la méthode proposée dans [212] et qui est sensible au bruit et instable dans les régions de croisements de fibres. Pour ces raisons, nous proposons de réaliser directement la segmentation d'ITD de façon à extraire les faisceaux de fibres. Alors que de nombreuses techniques ont été proposées pour classifier la matière grise, blanche et le liquide céphalo-rachidien depuis des images anatomiques (voir [322] par exemple), la littérature adressant la segmentation des structures de la matière blanche à partir d'ITD est encore récente. Zhukov *et al.* [326] ont défini une mesure d'anisotropie invariante afin de guider l'évolution d'une surface et d'isoler les régions fortement anisotropes du cerveau. La réduction du tenseur à une valeur scalaire est susceptible de résulter en de relativement faibles capacités de discrimination, ce qui peut engendrer des erreurs de segmentation. Wiegell *et al.* [312], Feddern *et al.* [111, 112], Rousson *et al.* [260], Wang *et al.* [304] et [303], Lenglet *et al.* [185] et Jonasson *et al.* [154] utilisent ou proposent d'autres mesures de dissimilarité entre tenseurs de diffusion. Dans [312], [304] et [260], les auteurs utilisent la norme de Frobenius de la différence de tenseurs (i.e. la distance Euclidienne). Un algorithme de type "k-means" avec une contrainte de cohérence spatiale et un modèle de contours actifs avec un terme de régularité sont respectivement utilisés par les deux premières méthodes ([312],

[304]) pour segmenter diverses structures cérébrales telles que les noyaux du thalamus ou le corps calleux. La troisième méthode [260] s'appuie sur une évolution de surface utilisant des statistiques de régions. Dans [304], une généralisation des contours actifs basés-région aux images à valeur matricielle est proposée. Cependant, elle est restreinte au cas bidimensionnel et donc d'intérêt limité pour nos images tridimensionnelles. Dans [111, 112], des équations aux dérivées partielles basées sur le mouvement par courbure moyenne, les "self-snakes" et contours actifs géodésiques sont étendus aux images bidimensionnelles et tridimensionnelles à valeur matricielle grâce à une généralisation de la notion de tenseur de structure aux données à valeur matricielle. Cette méthode repose toujours sur la métrique Euclidienne entre tenseurs. Elle est appliquée à la régularisation et segmentation d'ITD. Dans [154], les auteurs introduisent une mesure géométrique de dissimilarité en calculant le "produit scalaire" normalisé de deux tenseurs, ce qui peut être interprété comme une mesure de recouvrement. Enfin, les méthodes décrites dans [303] et [185] utilisent la divergence de Kullback-Leibler symétrisée pour dériver une mesure de dissimilarité affine invariante entre tenseurs de diffusion.

Nous montrons dans ce chapitre que la définition d'une mesure de dissimilarité et de statistiques entre tenseurs est une tâche non triviale qui doit être adressée avec précaution. Nous affirmons et démontrons que l'utilisation des concepts introduits au chapitre 5 permet d'améliorer la qualité des résultats de segmentation obtenus avec d'autres mesures de dissimilarité telles que la distance Euclidienne ou la divergence de Kullback-Leibler. Le but principal de ce chapitre est de prouver que le choix de cette métrique a d'importantes conséquences sur les statistiques de tenseurs et donc sur les résultats de segmentation. Nous définissons un cadre variationnel pour estimer la segmentation optimale d'une image du tenseur de diffusion. Nous supposons, d'une part, que les tenseurs de diffusion suivent une distribution Gaussienne dans les différentes partitions et que, d'autre part, le gradient spatial d'une ITD permet de détecter les interfaces entre les différentes structures. Nous validons et comparons les résultats obtenus sur divers exemples synthétiques, un fantôme biologique et des ITD de cerveaux humains.

Ce travail a été réalisé en collaboration avec Mikaël Rousson (Siemens Corporate Research, Princeton, NJ, USA).

Chapitre 9 - Recalage Non-rigide d'Images du Tenseur de Diffusion : Ce chapitre traite du problème de l'estimation des déformations géométriques entre deux images du tenseur de diffusion. Ce problème a déjà été extensivement adressé dans le cas scalaire [110]. Il est résolu en minimisant un critère d'erreur par rapport au champ de déformation tout en tenant compte de deux sources d'information a priori, à savoir les propriétés des valeurs des images dont on estime les déformations et les contraintes sur les déformations géométriques possibles. Dans notre cas, "intensités" signifie tenseurs de diffusion. Comme nous l'avons déjà mentionné précédemment, le choix d'une mesure de dissimilarité entre ces intensités est une tâche difficile. Le schéma numérique de descente de gradient, utilisé pour minimiser le critère d'erreur, peut s'avérer très complexe.

Les travaux antérieurs dans ce domaine ont été initiés par Alexander *et al.* [4] grâce à l'extension des techniques de recalage multi-résolution aux images du tenseur de diffusion et grâce à l'introduction de diverses mesures de dissimilarité pour ces images [5]. Dans [263] et [262], Les auteurs proposent de recaler des images tridimensionnelles scalaires, vectorielles ou matricielles en alignant les régions fortement structurées et propageant le champ de déplacement ainsi estimé à l'ensemble du volume. D'autres approches telles que [155], [130], [229] et [255] s'appuient sur un ou plusieurs invariants des tenseurs de diffusion comme les valeurs propres, des mesures d'anisotropie, le coefficient de diffusion apparente ou même les composantes du tenseur pour effectuer le recalage. Quand plusieurs caractéristiques sont utilisées, ce qui est souvent le cas, des méthodes de recalage multi-spectrales comme l'algorithme des démons [131] sont utilisées. Dans [321, 319] puis [320], les auteurs ont proposé une technique de recalage affine par morceaux basée sur la norme L^2 de profiles de diffusion. Ils ont aussi étudié le problème de la réorientation des tenseurs, soulevé par Alexander *et al.* dans [6]. Récemment, Cao *et al* [52] ont appliqué le cadre des Transformations Métriques Difféomorphiques pour les Grandes Déformations (TMDGD) aux ITD. Enfin, Leemans [177] a introduit une technique de recalage affine multi-spectrale basée sur l'information mutuelle, ainsi qu'une méthode rigide originale utilisant la courbure et torsion des fibres. Bien que différents du problème de recalage d'ITD, nous mentionnons quelques travaux récents [11, 236] qui font usage des métriques Riemannienne ou Log-Euclidienne, dans un contexte de recalage d'images scalaires, pour caractériser les propriétés des difféomorphismes estimés.

Nous étendons [110] à des images à valeur matricielle et proposons un cadre variationnel original pour le recalage dense non-rigide d'ITD, basé sur la structure Riemannienne de l'espace des lois normales multivariées. L'existence d'expressions en forme close pour les géodésiques et les symboles de Christoffel nous permet de définir des statistiques et de réaliser le transport parallèle de vecteurs tangents. Notre énergie d'appariement vise à minimiser les différences entre statistiques locales de deux images du tenseur de diffusion. Ce chapitre est essentiellement une étude mathématique du problème de recalage dont l'implémentation est très délicate. Nous illustrons la faisabilité de l'approche sur des exemples synthétiques bidimensionnels.

Partie III : Applications

Chapitre 10 - Analyse des Connexions Cortico-striatales Impliquées dans l'Apprentissage Moteur : Dans ce court chapitre, nous décrivons une application de la méthode d'analyse de la connectivité présentée au chapitre 7, et reposant sur le cadre des ensembles de niveaux, pour l'étude du réseau cortico-striatal de l'homme. Cette étude utilise la formulation dynamique de l'équation Eikonale anisotrope pour des raisons "historiques". Elle a été initiée il y a environ deux ans et c'est en fait à cause de son coût computationnel important que nous avons ensuite travaillé sur la formulation stationnaire du problème.

Comme rapporté par Lehéricy *et al.* dans [178], il est possible de parcelliser les connexions du striatum de l'homme grâce aux algorithmes de suivi de fibres déterministes basés sur la méthode présentée dans [26]. Les études invasives chez l'animal ont montré en effet que les fibres projectives de la matière blanche entre le cortex et le striatum sont organisées en un ensemble discret de circuits [8]. Chaque circuit est responsable d'une fonction précise comme la préparation et l'exécution du mouvement, la prise de décision, la planification ou l'apprentissage. Ces circuits transmettent donc des informations de natures différentes (sensorimotrice, associative ou limbique). Dans [178], les auteurs ont présenté la première étude chez l'homme de la connectivité des noyaux gris centraux avec le cortex. Ils ont pu mettre en évidence l'existence de connexions distinctes pour les compartiments postérieur (sensorimoteur), antérieur (associatif) et ventral (limbique) du striatum avec les aires corticales. Dans ce chapitre, nous montrons qu'il est possible, avec notre méthode d'analyse de la connectivité, d'identifier différents réseaux cortico-striataux impliqués dans l'acquisition et le stockage de connaissances motrices.

Ce travail a été réalisé en collaboration avec Stéphane Lehéricy (Center for Magnetic Resonance Research, University of Minnesota, Minneapolis, USA / Hôpital La Pitié-Salpêtrière, Paris, France).

Chapitre 11 - Analyse des Connexions Anatomiques du Cortex Visuel Humain : L'analyse de la connectivité anatomique du système visuel de l'homme par ITD a été réalisée par différents groupes de recherche, avec des protocoles et méthodes variées. En utilisant une technique de suivi de fibres classique avec interpolation du champ de tenseurs [212], les auteurs de [78] ont pu reconstruire divers faisceaux dont ceux du système visuel. Ils ont mis en évidence une certaine topologie des fibres du splénium du corps calleux ainsi que des connexions géniculo-occipitales. Pour ces dernières, les fibres médiales/latérales du noyau géniculé latéral (NGL) aboutissent respectivement dans la partie supérieure ou inférieure du cortex occipital. En utilisant une technique similaire de suivi de fibres, [55] a pu identifier différents faisceaux de fibres visuelles dont les faisceaux occipito-frontaux et occipito-temporaux. Dans une étude plus récente [56], les auteurs sont parvenus à mettre en évidence: (i) le faisceaux optique allant du chiasme au NGL; (ii) les radiations optiques du NGL vers le cortex occipital, que l'on peut subdiviser en une partie ventro-temporale arrivant dans la partie inférieure de la fissure calcarine et en une partie dorsale aboutissant dans la partie supérieure de la fissure calcarine; (iii) le faisceaux du splénium connectant les pôles occipitaux; (iv) des fibres courtes en U occipito-temporales; (v) le faisceaux inférieur longitudinal. Malgré ces résultats très intéressants, aucune information fonctionnelle n'a été utilisée pour identifier les régions d'intérêt. Ciccarelli *et al.* ont utilisé la technique de Suivi de Fibres par Cheminement Rapide (SFCR) proposée dans [231] pour analyser trois faisceaux de fibres : le faisceau pyramidal, les radiations callosales antérieures et les radiations optiques [72, 70]. Récemment, ils ont appliqué cette technique pour étudier les changements dans les radiations optiques de patients atteints d'une pathologie spécifique du nerf optique (névrite optique) [71]. La méthode SFCR a été partiellement validée, dans une étude combinée sur les macaques et l'homme, en détectant le faisceau cortico-spinal et les radiations optiques chez les deux espèces [233].

Actuellement, seules quelques études, comme [164], ont combiné IRM fonctionnelle et ITD pour étudier le cortex visuel. Dans les travaux mentionnés ci-dessus, [78] a utilisé des cartes d'activation IRMf pour grossièrement localiser le NGL et le cortex occipital visuel. [309] a montré que l'anisotropie fractionnaire était plus basse dans le cortex occipital activité que dans les radiations optiques, ce qui est cohérent avec la relative isotropie de la matière grise, par comparaison à la matière blanche [243]. Grâce à une méthode de suivi de fibres probabiliste, [280] a étendu ce travail en montrant une corrélation entre le degré d'activation IRMf du cortex visuel et les valeurs moyennes d'anisotropie fractionnaire le long des fibres optiques. Cependant, l'identification des aires visuelles considérées n'a pas été réalisée par IRMf. A notre connaissance, l'étude la plus complète et précise est décrite dans [99]. Les auteurs ont combiné une méthode de suivi de fibre déterministe avec l'identification fonctionnelle des aires rétinotopiques occipitales pour estimer les fibres passant par le splénium. Ils ont pu mettre en évidence que les fibres de la partie dorsale, respectivement ventrale, du cortex se projettent dans la partie dorsale, respectivement ventrale, du splénium. Ceci est en accord avec des études par autoradiographie chez le macaque. Ils ont également constaté qu'un gradient de la fovéa vers la périphérie se retrouve dans la direction antérieure-postérieure au niveau du splénium.

Dans ce chapitre, nous avons utilisé la méthode d'analyse de la connectivité présentée au chapitre 7, et reposant sur la méthode de cheminement rapide, pour étudier les connexions présentes entre les diverses aires visuelles du cerveau humain. Ces aires ont été identifiées individu par individu, par IRMf. Nous avons tout d'abord validé notre méthode en reproduisant des résultats de la littérature sur la topologie des fibres reliant le NGL à V1 ou connectant des aires homologues du cortex visuel. Nous proposons ensuite des résultats intéressants sur les connexions du complexe MT avec les régions rétinotopiques.

Ce travail a été réalisé en collaboration avec Nicolas Wotawa (INRIA Sophia-Antipolis, France).

Appendices

Appendice A - Dérivées Directionnelles de Fonctions Matricielles : Cet appendice introduit une formulation originale pour le calcul de dérivées directionnelles de fonctions matricielles. Cette méthode repose sur la décomposition spectrale des matrices et est particulièrement rapide. Le cas du logarithme est étudié puisqu'il constitue un élément fondamental de la méthode de recalage présentée au chapitre 9.
Ce travail a été réalisé en collaboration avec Théo Papadopoulo (INRIA Sophia-Antipolis, France).

Appendice B - Détails sur la Première Variation de $\mathcal{J}_{\text{AC}}(h, Dh)$: Cet appendice présente les détails nécessaires au calcul du gradient du terme $\mathcal{J}_{\text{AC}}(h, Dh)$ de notre algorithme de recalage du chapitre 9.

Appendice C - Publications de l'Auteur : Cet appendice présente les publications de l'auteur.

Contributions logicielles

Enfin, nous souhaitons souligner que tous les algorithmes présentés dans ce manuscrit sont désormais disponibles, sur demande, comme extension de la plateforme logicielle Brainvisa[1] pour la visualisation et l'analyse d'images cérébrales provenant de différentes modalités. Ils sont également disponibles sous la forme d'une extension pour SPM5[2].
Ce travail d'intégration a été réalisé en collaboration avec Demian Wassermann (INRIA Sophia-Antipolis, France / Computer Science Department, University of Buenos Aires, Argentina) et Maxime Descoteaux (INRIA Sophia-Antipolis, France).

---

[1] `http://brainvisa.info`

[2] `http://www.fil.ion.ucl.ac.uk/spm/software/spm5/`

Part I

Background

CHAPTER 3

THE HUMAN NERVOUS SYSTEM AND MAGNETIC RESONANCE IMAGING

Contents

3.1 INTRODUCTION

The human brain, according to the connectionist point of view, is organized into distinct processing regions interconnected by a network of anatomical relays. The processing units handle the execution of primary cognitive functions and higher cognitive tasks arise from a global coordination between those processing units. Neural signal is processed in the cerebral cortex and transmitted to various regions of the brain through the white matter, so called because of the white appearance of the myelinated axons. The white matter pathways are generally categorized into commissural, association and projection fibers, depending on the areas they connect. Compared to our understanding of neural circuitry in animals such as cat or mouse where the use of invasive tracers is possible [266], our knowledge of the human brain organization is relatively poor. Brain imaging techniques like Magnetic Resonance Imaging (MRI), and in particular diffusion MRI, shall allow us to access this critical information.

Organization of this chapter: This chapter is divided in two parts. We first start, in section 3.2, by a quick description of the human nervous system with a particular emphasis on the cerebral grey and white matters. We then introduce the MRI technique in section 3.3 and how it can be used to infer anatomical connectivity (diffusion MRI).

Keywords: human nervous system, brain, cortex, white matter, basal ganglia, nuclear magnetic resonance (NMR), magnetic resonance imaging (MRI), diffusion MRI

3.2 THE HUMAN NERVOUS SYSTEM

The nervous system plays a major role in the muscular control, regulation of sensory inputs and organs monitoring. It can be divided into the peripheral nervous system (PNS) and the central nervous system (CNS) and its principal building blocks are the neurons and the nerves.

The peripheral nervous system is an extension of the central nervous system, dedicated to the control of the limbs and various organs. It consists of two type of neurons: (i) the motor neurons which connect the CNS to muscles and glands to control their activity, and (ii) the sensory neurons which inform the CNS of the stimuli recorded by our senses. It can also be subdivided into two functionally distinct parts: (i) the sensory-somatic nervous system, whose nerves carry information from and to the sensory organs and the muscles (i.e. the external environment) and (ii) the autonomic nervous system, involved in the regulation of vital functions such as breathing, blood circulation, digestion or hormones secretion (i.e. the internal environment). Finally, sensory-somatic nervous system (SSNS) and autonomic nervous system (ANS) can each be divided into two subparts.

The former (SSNS) essentially supports conscious responses and comprises 12 pairs of cranial nerves and 31 pairs of spinal nerves. All spinal are mixed, meaning they contain both sensory and motor neurons. Some cranial nerves are also mixed, like the ones involved in taste, but they are otherwise typically either sensory or motor. Pairs involved in smell, vision, hearing and balance are only sensory neurons. Those involved in eyelid/eyeball control, tongue control, head/shoulders control and swallowing are exclusively motor neurons.

The later (ANS) essentially supports unconscious responses and can be divided into three subsystems: (i) The sympathetic nervous system regulates the activity of smooth muscle, cardiac muscle and glands. It generally increases their activity. (ii) On the contrary, the parasympathetic nervous system targets the same organs but lowers their activity. It works to bring the body back to its normal level of activity after a modification of metabolism induced by the sympathetic nervous system. (iii) Finally, the enteric nervous system is dedicated to the control of the viscera.

This thesis focuses on the use of diffusion tensor MRI to investigate the architecture of the CNS and, more precisely, of the brain. We thus now concentrate on the description of the CNS. For a more detailed view of the brain structures, various atlases and books are now available. We refer the interested reader to the atlas by Duvernoy [104].

3.2.1 The Central Nervous System

The central nervous system is composed of the brain and the spinal cord. It is the largest part of the nervous system and is protected by the meninges. The brain is itself composed by a lower part, the brainstem (figure 3.1), and an upper part, the prosencephalon. The brainstem connects the prosencephalon to the spinal cord and equally comprises an upper and a lower segment. The medulla oblongata, or lower part, controls unconscious activity of muscles and glands involved in breathing, heart contraction, salivation ...etc. Just above the medulla, the pons constitutes the upper part of the brainstem. and connects the two hemispheres of the cerebellum. The prosencephalon, or forebrain, is composed of two main units. One is known as the diencephalon, is located in the midline of the brain and contains the thalamus and the hypothalamus. The other is called the telencephalon (or cerebrum, figure 3.2) and holds the lateral ventricles, the basal ganglia and the cerebral cortex.

We now focus our description on parts of the cerebrum. Before describing the main facts about grey and white matter tissues, we give some details on their major building block: the neuron.

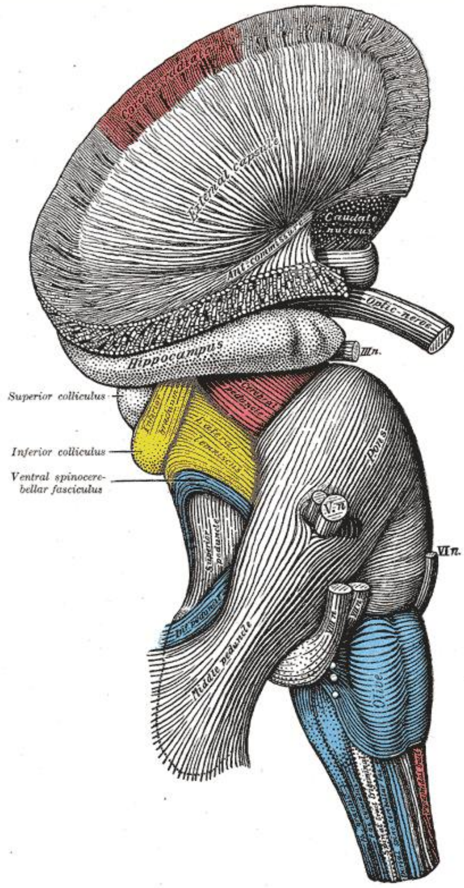


Figure 3.1: Superficial dissection of the brainstem. From 20th U.S. edition of Gray's Anatomy of the Human Body (public domain)

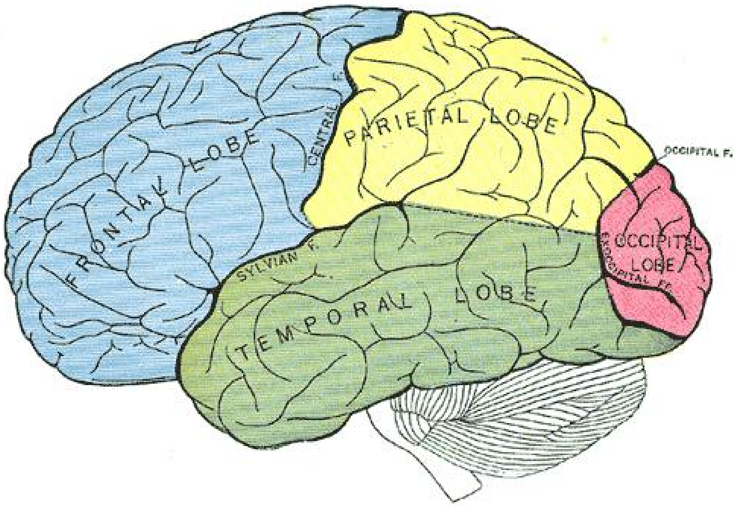


Figure 3.2: The cerebrum, or telencephalon, and its four lobes. From 20th U.S. edition of Gray's Anatomy of the Human Body (public domain)

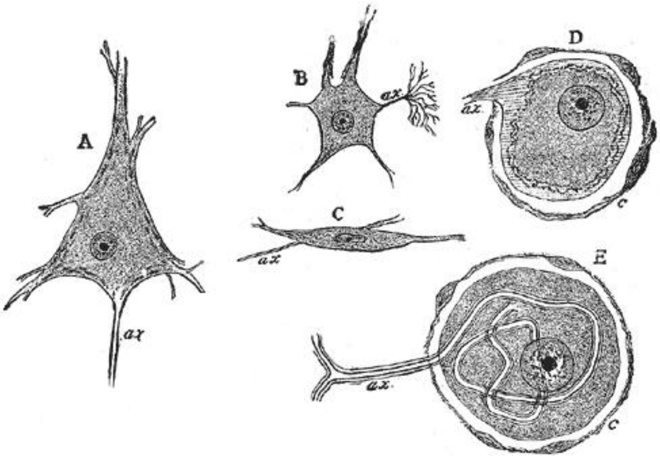


Figure 3.3: Various forms of nerve cells. (A) Pyramidal cell. (B) Small multipolar cell, in which the axon quickly divides into numerous branches. (C) Small fusiform cell. (D) and (E) Ganglion cells (E shows T-shaped division of axon). ax. Axon. c. Capsule. From 20th U.S. edition of Gray's Anatomy of the Human Body (public domain)

3.2.2 The Neuron

The human nervous system is made of about 100 billion nerve cells, or neurons. One important characteristic of neurons is that they have excitable membranes which allow them to generate and propagate electrical signals to process and transmit the neural information. Neurons exhibit a great diversity in shape and size (figure 3.3) but present a common structure. Neurons can receive electrical stimulations from other neurons on their soma, through their multiple dentrites. They can integrate this information and propagate it to more or less distant locations of the cerebrum by an extension called axon. Nerve signal communication is performed at specialized loci called synapses. At those locations, the axon of a presynaptic nerve cell encounters a dentrite or the soma of a postsynaptic cell. The vast majority of the synapses found in the human brain are chemical synapses, i.e. the information transfer is achieved by specific molecules known as neurotransmitters. Each neuron has on average 1000 synaptic connections with other neurons. This yields about 100 trillion of connections within a human brain. All these synapses result in an impressively dense and complex network between functional areas, which can be understood as aggregates of nerve cells' soma and dendrites. They are essentially located in the grey matter while the underlying wiring constitutes the white matter.

3.2.3 Organization of the Grey Matter

The grey matter essentially forms the outer part of the cerebrum, some nuclei within the brain, as well as deep parts of the spinal cord. It is made of neurons and their unmyelinated fibers.

The cerebral cortex

The cerebral cortex is the most important structure of the grey matter and plays a major role in functions such as memory, attention and language. It constitutes the outermost layers of the cerebrum and is highly folded to increase its surface in the limited volume of the skull. This folding process creates grooves on the surface of the brain called sulci and ridges called gyri. About two thirds of the cortical surface is buried in those sulci.
The two hemispheres of the brain are separated by a prominent central fissure. Each hemisphere of the cerebral cortex is made of four lobes (figure 3.2): The frontal lobes, located in front of the central sulcus, play an important role in reasoning, planning, language, memory and motor control. Parietal lobes, behind the central sulcus, are more important to integrate sensory information and process some visuospatial stimuli. The temporal lobes, on the most lateral parts of the cortex and below the lateral fissure, are involved in auditory processing as well as language and vision related functions. Finally, the occipital lobes occupy the rearmost part of the cortex and are the processing center of visual stimuli. In addition to these four lobes, neurologists consider an internal lobe, called the limbic system, which lies along the medial part of the cortex and the insular cortex buried within the lateral sulcus (aka. the Sylvian fissure).

The thickness of the cerebral cortex ranges from 2 to 4 mm and is organized in various layers (usually six) that are tangent to the cortical surface. These layers can be distinguished histologically, functionally and through the connectivity pattern they exhibit with each other. Moreover, neuroanatomists noticed that the neurons distribution and size are not homogeneous across the cortex. This led some of them to propose a parcellation of the cortex into distinct zones, or cortical areas with coherent cells structure. The most famous ones are the cytoarchitectonic maps of Brodmann [41] based on microscopical studies of local cellular and laminar structure.

Subcortical structures

The cortex is not the only grey matter part of the brain. The basal ganglia (figure 3.4) are the striatum, the internal/external segments of the globus pallidus (GPi/e), the subthalamic nucleus (STN) and the substantia nigra (SN). They are composed of grey matter and involved in motor and learning functions. It is also the case for the thalamus which comprises many different pairs of nuclei, such as the pulvinar or the lateral/medial geniculate nuclei (L/MGN), most of which project to the cortex. Some thalamic nuclei are sensory relays, i.e. nuclei that receive signals from sensory receptors, process them, and then transmit them to the appropriate areas of sensory cortex. For example, the LGN, MGN, and the ventral posterior nuclei (VPN) are important relay stations respectively in the visual, auditory, and somatosensory

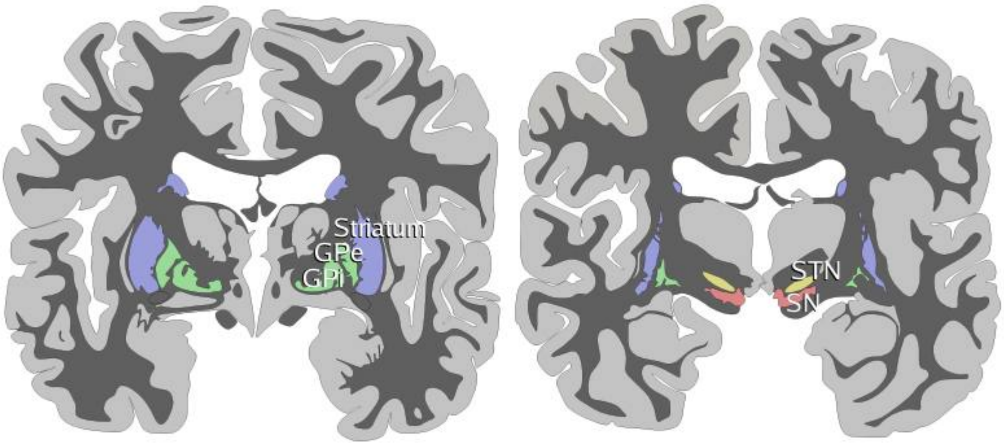


Figure 3.4: Two schematic drawings of coronal sections of human brain labelling the basal ganglia. Blue=striatum, green=globus pallidus (external and internal segments), yellow=subthalamic nucleus, red=substantia nigra (pars reticulata and pars compacta). Image by Andrew Gillies under the Gnu Free Documentation License 1.2

systems.

Connections between the cortex, the thalamus and the basal ganglia are as follows: The striatum is the principal entry gate to the basal ganglia for the cerebral cortex (essentially motor and prefrontal cortical areas). Then, within the basal ganglia, there are mainly two pathways back to the cortex and through the thalamus:

- The direct pathway: striatum / GPi-SN / thalamus / cortex
- The indirect pathway: striatum / GPe / STN / GPi-SN / thalamus / cortex

In the following section, we describe the principal types of connections existing between regions of the grey matter.

3.2.4 Organization of the White Matter

White matter is composed of axonal nerve fibers, covered by a myelin sheath giving its distinctive color. It is found in the inner layer of the cortex, the optic nerves, the central and lower areas of the brain and surrounding the central shaft of grey matter in the spinal cord. The white matter axons can be distributed diffusely or concentrated in bundles, also referred to as tracts or fiber pathways. Three main types of neural tracts are found in the white matter:

- The **Projection tracts** establish connections between the cerebral cortex and subcortical structures. Two types of projection tracts can be distinguished: ascending tracts and descending tracts. Ascending tracts carry sensory informa-

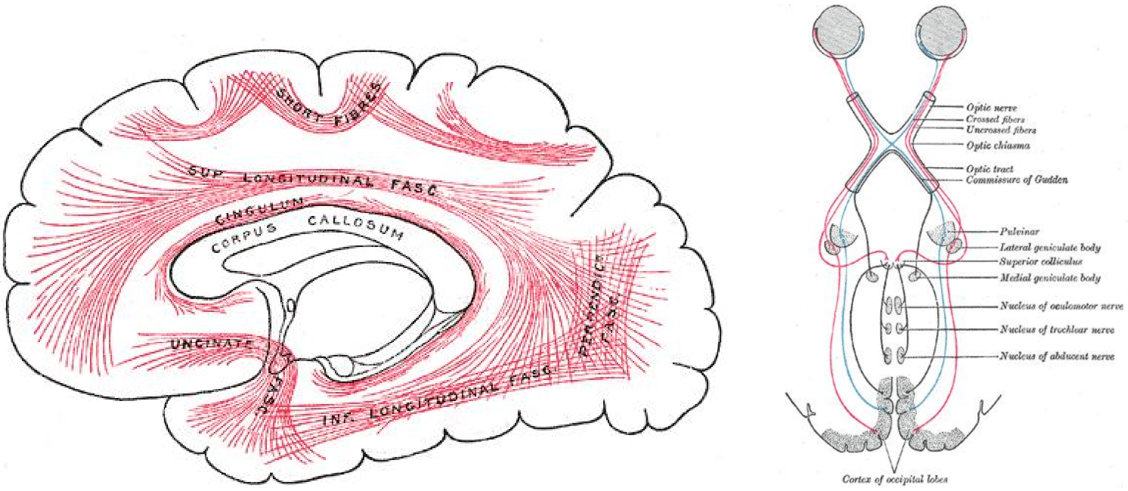


Figure 3.5: [Left] Diagram showing principal systems of association fibers in the cerebrum. [Right] Scheme showing central connections of the optic nerves and optic tracts. From 20th U.S. edition of Gray's Anatomy of the Human Body (public domain)

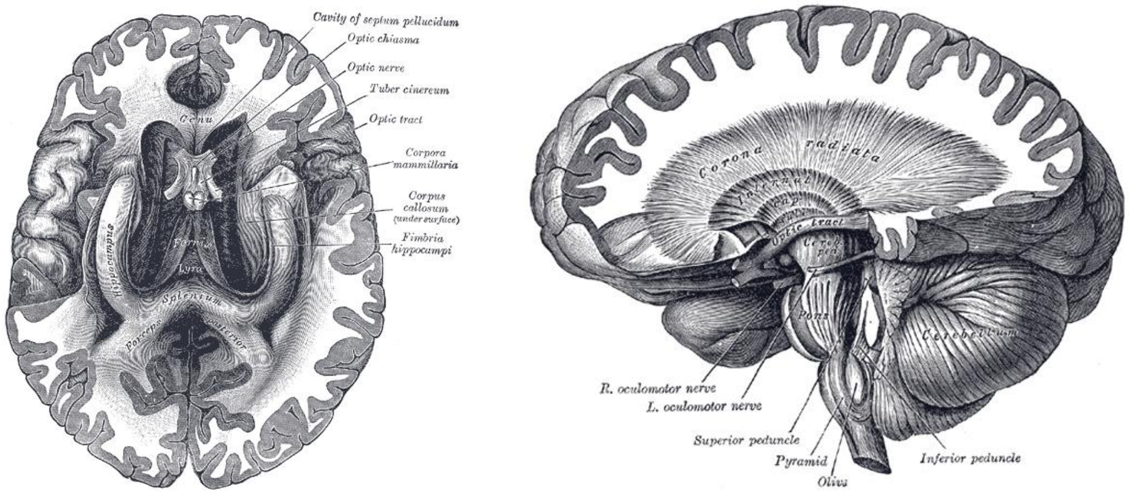


Figure 3.6: [Left] The fornix and corpus callosum from below. (From a specimen in the Department of Human Anatomy of the University of Oxford.) [Right] Dissection showing the course of the cerebrospinal fibers. (E. B. Jamieson). From 20th U.S. edition of Gray's Anatomy of the Human Body (public domain)

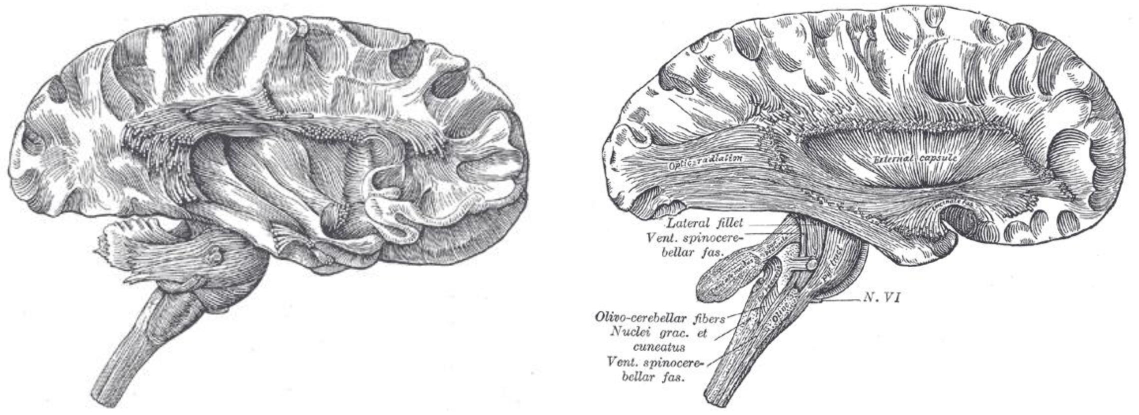


Figure 3.7: Dissection of cortex and brain-stem showing association fibers after removal of the superficial grey matter. From 20th U.S. edition of Gray's Anatomy of the Human Body (public domain)

tion from different parts of the body to the cerebral cortex. All sensory information, except olfactory, end up in the primary sensory cortex by the means of the thalamo-cortical fibers. The thalamus receives the somesthetic, gustatory, visual and auditory stimuli through these ascending pathways. Descending tracts carry motor commands from the motor cortex down to the muscles and glands through the lower brain structures and the spinal cord. They reach structures like the thalamus, the red nucleus, the medulla and serve muscles of the torso, extremities, facial and neck region. The cerebrospinal fibers are one example, see figure 3.6 [Right].

- The **Association tracts** are the communication paths between different cortical areas within a given hemisphere (see examples on figure 3.5 [Left] and figure 3.7). They can be divided into two categories: short and long association tracts. Short association tracts build up connections between regions of a given lobe. The smallest link adjacent cortical zones separated by a sulcus, hence their name of U-shaped fibers. Long association fibers establish connections between different cerebral lobes and often form a bundle macroscopically visible (see for instance the uncinate fasciculus on figure 3.7 [Right].
- The **Commissural tracts** are bundles of axons connecting a region in one hemisphere to another region of the opposite hemisphere. The corpus callosum (figure 3.6 [Left]) is the most important of the commissural tracts and can be broken down into four parts: The rostrum (anterior most part) and the genu (anterior curvature) are made up of fibers connecting the anterior and ventral parts of the frontal lobes. The corpus (large middle portion) links posterior portions of the frontal lobes as well as the parietal lobes. Finally, the splenium (caudal curvature) enables communications between the temporal and occipital lobes.

In the next section, we will introduce the foundation of MRI and diffusion MRI. We will show that diffusion MRI constitutes an amazing and non-invasive means to investigate the three-dimensional architecture of the human brain white matter that has just been described.

3.3 MAGNETIC RESONANCE IMAGING

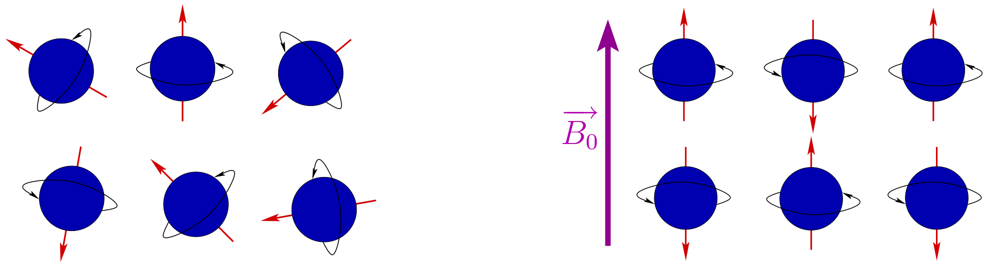
Magnetic Resonance Imaging (MRI) is by far the most popular application of Nuclear Magnetic Resonance (NMR), for medical diagnosis. However, NMR is also widely used in chemistry to perform NMR spectroscopy, ie. to study the atomic composition of a given sample.

NMR was simultaneously described by Felix Bloch [39] at Stanford University and by Edward Mills Purcell [252] at Harvard University in 1946. In 1952, they both received the Nobel Prize in Physics for their discovery. The basic principle behind NMR is that, after aligning a magnetic nucleus like hydrogen-1 with a very strong external magnetic field, its response to a perturbation of the alignment by an electromagnetic field is characteristic. Four year after this discovery, in 1950, Herman Carr, proposed to create the first one-dimensional MR images by introducing a gradient in the magnetic field. In 1971, it was shown by Raymond Damadian that T1 and T2 relaxation times of tumoral tissues are significantly longer than for the corresponding normal tissues, hence opening great hopes for cancer diagnosis. Shortly after Bloch and Purcell discovery, Hahn published his seminal paper [137] on the NMR spin echo in which he noted that the random thermal motion of the spins would reduce the amplitude of the observed signal in the presence of a magnetic field inhomogeneity. This is a fundamental notion to understand diffusion MRI.

As soon as 1973, Paul Lauterbur proposed a method [171], based on gradients of magnetic fields, to reconstruct two dimensional MR images. Peter Mansfield [203] further developed the use of magnetic fields gradients and, by studying the mathematical properties of the MRI signal, proposed a new ultrafast acquisition technique known as the echo-planar technique. In 2003, they jointly received the Nobel prize in Physiology and Medicine for their discoveries on MRI.

MRI thus allows to acquire non-invasively 3D images at high spatial resolution. Various modalities can be obtained with the same device such as detailed anatomy (structural MRI), functional activity (functional MRI), water-molecules diffusion (diffusion weighted MRI), blood flow measurements (perfusion MRI), distribution of various metabolites (MR Spectroscopy) and also blood vessels (MR Angiography).

The first part of this section briefly exposes the basic principles of MR imaging. We then give the outlines of diffusion MRI.



a) Random spins directions b) Spins aligned with $\overrightarrow{B_0}$

Figure 3.8: Random directions of spins in the absence of an external magnetic field (a) and aligned spins in the presence of an external magnetic field $\overrightarrow{B_0}$ (b). Note that the actual spin rotation around $\overrightarrow{B_0}$ occurs within a cone around $\overrightarrow{B_0}$.

3.3.1 MRI Principles

Physical model

Atoms are made of electrons, which hold a negative charge and rotate around a nucleus. The nucleus can be divided in neutrons (not charged) and protons (charged positively) and rotates around itself. MRI is based on this rotation motion. Some nuclei have the property to align with a magnetic field if their mass number is odd, i.e. if the sum of numbers of protons and neutrons is odd. This is named angular moment or spin. Among others, ^1H atoms, which represent 99.89% of naturally found hydrogens atoms and are widely represented in biological systems, have a spin. MRI is thus particularly relevant to study the structure of biological tissues such as the human brain.

Spin nuclei being positively charged, their motion induces a magnetic field. Conversely, the resulting magnetic moment can be oriented by the application of a magnetic field. This reciprocity is largely used in MRI. From a macroscopical point of view, no resulting field can be observed directly since each spin has its own, independent, random orientation (figure 3.8 (a). However, when placed in a powerful external magnetic field $\overrightarrow{B_0}$, the spin directions align parallel to this field (figure 3.8 (b). More precisely, each spin rotates within a cone around $\overrightarrow{B_0}$. This is called the spin precession. The frequency of rotation, called the Larmor frequency, is related to the magnetic field $\overrightarrow{B_0}$ through the gyromagnetic ratio γ by the following equation:

$$\omega_0 = \gamma \|\overrightarrow{B_0}\| \tag{3.1}$$

γ depends on the nucleus. Hydrogen, for instance, has a gyromagnetic ratio $\gamma = 42.57$MHz/T. This corresponds to a rotation frequency of $f_{^1H} = 63.86$MHz in a 1.5 Tesla magnetic field. Because of this rotation motion, a spin can be modeled by a small magnetic dipole with moment $\overrightarrow{m}$ verifying

$$\frac{d\overrightarrow{m}}{dt} = \gamma \overrightarrow{m} \times \overrightarrow{B_0}$$

It is actually cumbersome to understand MRI at a microscopic scale. It is convenient, at a macroscopic level, to replace the individual spin by a single magnetization vector representing the spin of all the particles in a voxel (about 2.10^{15} protons/mm^3). The net resulting magnetization $\overrightarrow{M}$ is the sum of all the elementary moments and, by making the assumption of a uniform distribution of the dipoles orientations in a given voxel, we simply end up with a $\overrightarrow{M} = \overrightarrow{0}$.

However, under the action of a static magnetic field $\overrightarrow{B_0}$ (from 1 and up to more than 9.4 Tesla), particles get aligned in the direction of that field and induce a magnetization parallel to $\overrightarrow{B_0}$ at equilibrium. In that state, the amplitude of $\overrightarrow{M}$ represents only a small fraction of what it would have been if all the particles were aligned in the same direction. Actually, by the laws of thermal dynamics, the number of spins following the orientation imposed by $\overrightarrow{B_0}$ (low energy state, called spin-up) slightly outnumbers the amount of spins anti-parallel to the outer field (high energy state, called spin-down). The difference is small and given by the ratio:

$$\frac{N^-}{N^+} = \exp\left(-\frac{E}{kT}\right)$$

where N^- and N^+ are respectively the number of spins in the upper and lower states, k is the Boltzmann constant and T the temperature in Kelvin. Applying the Boltzmann relation, one can estimate that, at the ambient temperature and within a 1.5 T field, there is a difference of 10 in favor of low energy protons among a total of 1 million protons.

The net magnetization $\overrightarrow{M}$ can be decomposed into two components (figure 3.9):

- A longitudinal component $\overrightarrow{M_z}$, i.e. parallel to $\overrightarrow{B_0}$
- A transverse component $\overrightarrow{M_{xy}}$, orthogonal to $\overrightarrow{B_0}$

At equilibrium, after a sufficient exposition time to $\overrightarrow{B_0}$, the transverse component $\overrightarrow{M_{xy}}$ vanishes. All the individual spins are indeed precessing, but they are all out of phase with each other.

Excitation phase

By applying an oscillating electromagnetic (radio-frequency) pulse toward the area of the body to be examined, it is possible to compensate the difference in the number of atoms between the two energy states. The idea is to use a much weaker field than $\overrightarrow{B_0}$ at the Larmor frequency of the targeted nuclei and to apply it through a rotating reference frame orthogonal to $\overrightarrow{B_0}$. It causes the particles in that area to absorb the energy required to make them spin in a different direction and move from the lower energy state towards the higher.

The exposure to the radio-frequency pulse causes the net magnetization to spiral

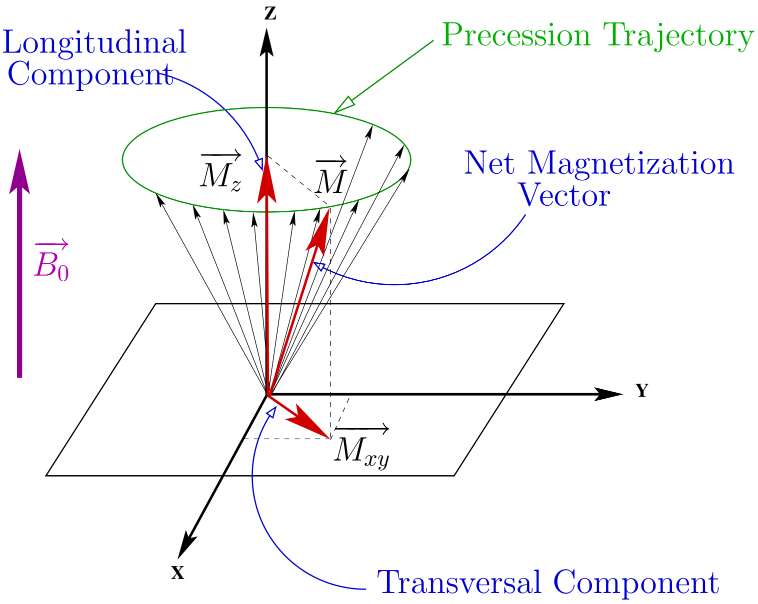


Figure 3.9: The net magnetization vector $\overrightarrow{M}$, decomposed into a longitudinal component $\overrightarrow{M_z}$ and a transverse component $\overrightarrow{M_{xy}}$.

away from $\overrightarrow{B_0}$. $\overrightarrow{M}$ rotates away from the longitudinal position in an amount proportional to the duration of the pulse. It is even possible to flip the original direction of $\overrightarrow{M}$. A pulse of 90 degrees would zero out the longitudinal component of $\overrightarrow{M}$ (figure 3.10) while a 180 degrees pulse, or ”inversion pulse”, completely inverts the longitudinal component through an excess of antiparallel spins.

The net magnetization also starts to dephase since different particles experience a slightly different magnetic field. This is usually referred to as phase coherence. All the magnetic moments are in phase in their precession motion. The MRI signal is acquired by measuring a current induced in the plane where the radio-frequency pulse was applied. The frequency of this current is the Larmor frequency of the nucleus and its amplitude is directly linked to the amount of magnetization in that plane.

Relaxation phase

By removing the radio-frequency pulse, particles begin to return to their initial energy state, aligned with the external field, from the higher to the lower. This is associated with a loss of stored excess energy to surrounding particles which can be detected by the coil of the MRI scanner. We can then observe two different types of relaxation processes:

- T1 weighted images follow the evolution of the increasing longitudinal component of $\overrightarrow{M}$
- T2 weighted images follow the evolution of the decreasing transversal component of $\overrightarrow{M}$

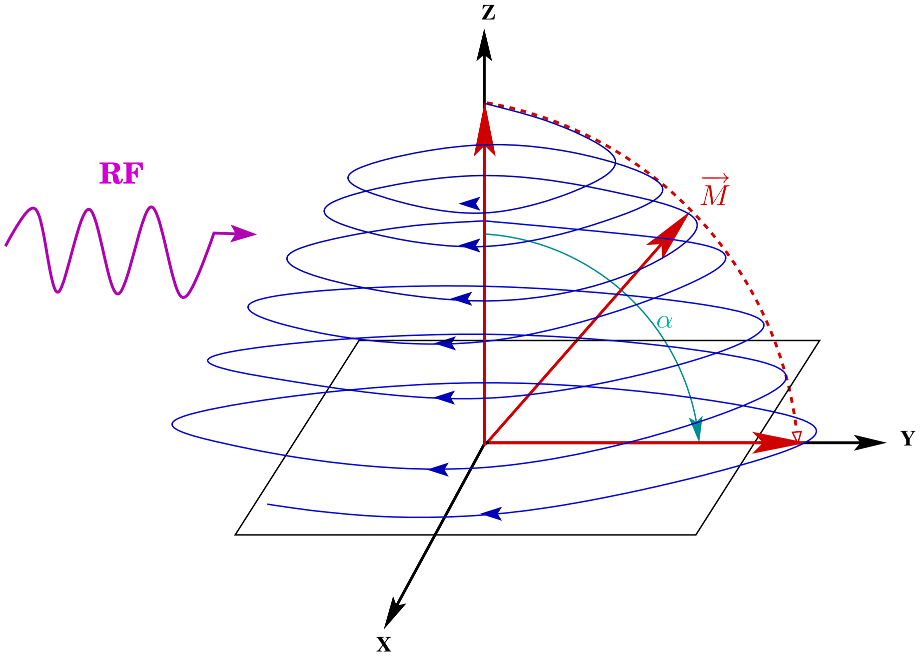


Figure 3.10: Excitation phase: the energy given by the RF pulse flips the net magnetization vector $\overrightarrow{M}$ of an angle α (here $\alpha = 90°$).

In clinical MRI, the radio-frequency pulse is typically chosen to coincide with the Larmor frequency of the hydrogen nucleus. The energy release during relaxation is thus an estimate of the number of protons or, in other words, the amount of water.

Spin lattice relaxation (T1): The spin lattice relaxation is based on the energy exchange between protons and surrounding molecules. This energy dissipation is characterized by the restoration of the longitudinal component to its equilibrium value. This recovery process is modeled by an exponential function characterized by a time constant T1, the period for the longitudinal magnetization to recover 63% of its equilibrium value (figure 3.11). For a 90-degree excitation pulse, we have:

$$M_z = M_0(1 - \exp\left(-\frac{t}{T_1}\right))$$

The recovery process is considered as finished after 5 T1 periods.

Spin-spin relaxation (T2): Spin-spin relaxation refers to the loss of net magnetization in the transverse plane related to protons dephasing. Spins do not only give up their energy to surrounding lattice molecules but also to other neighboring nonexcited spins. This process is also modeled by an exponential function characterized by another time constant T2, which corresponds to the period for the transversal component to loose 63% of its value just after the RF pulse:

$$M_{xy} = M_0 \exp\left(-\frac{t}{T2}\right)$$

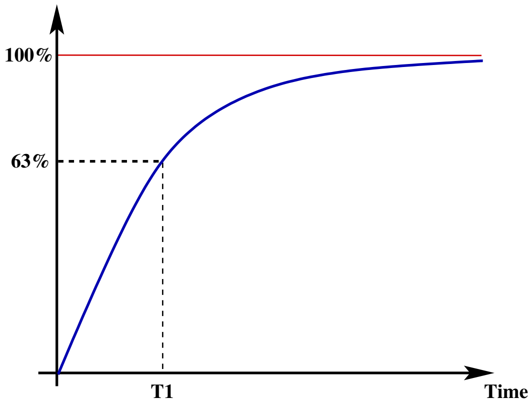


Figure 3.11: Spin lattice relaxation describes the longitudinal component recovery as a function of time and is characterized by the T1 constant.

This dephasing is actually further increased by local magnetic field inhomogeneities, since the Larmor frequency will also be nonuniform throughout the region. A time constant slightly different to T2, T2*, is therefore used. The transverse component induces a current in a coil, known as Free Induction Decay (FID). The T2* constant can be evaluated through the convex envelop of the FID curve (figure 3.12).

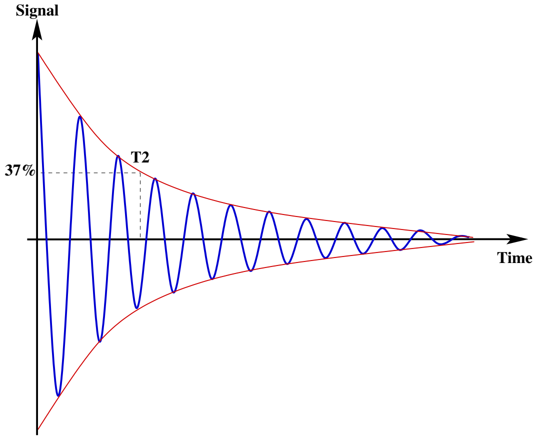


Figure 3.12: Spin-spin relaxation describes the exponential decrease of the transversal component as a function of time and is characterized by the T2 constant.

The different biological tissues are characterized by respective T1 and T2 values, as shown in table 3.1. The intensities of MR images comes from these values.

Image construction through pulse sequence

A pulse sequence is a series of RF pulses and/or magnetic field gradients applied to a sample to produce a specific form of MR signal. It is indeed possible to encode and thus recover the MR signal from specific regions in the volume of interest by means of RF and linear gradients applied along the 3 spatial directions.

Tissue	T1 (ms)	T2 (ms)
CSF	2200-2400	500-1400
Grey matter	920	100
White matter	780	90
Fat	240-250	60-80
Blood (deoxygenated)	1350	50
Blood (oxygenated)	1350	200
Muscles	860-900	50

Table 3.1: Approximate T1 and T2 values (ms) in various tissues at 1.5T.

Figure 3.13 illustrates a basic pulse sequence. A first gradient $\mathbf{G}_z$ in the $\overrightarrow{B_0}$ magnetic field direction results in a linear intensity variation of the magnetic field that can be used to select a slice. In this case, a slice is a plane orthogonal to $\overrightarrow{B_0}$ with a typical thickness of 1-10mm. Based on relashionship (3.1), the spins of a given slice are hence characterized by a specific Larmor frequency. After the RF pulse at the frequency related to the target slice, two transient gradients are applied to encode the x and y dimensions in the slice plane. A first gradient $\mathbf{G}_y$ in the y direction induces a phase shift related to the position along the y axis: this is the *phase encoding*. A second gradient $\mathbf{G}_x$ in the remaining x direction is applied, leading to a precession frequency variation along the x axis: this is the *frequency encoding*. This process actually performs an acquisition of the plane data in the frequency space (or *k-space*). For each selected slice, an inverse Fourier transform finally maps these data back into the 2D spatial domain.

A pulse sequence is first characterized by the delay between two similar RF pulses, called the Repetition Time (TR). The other parameters of interest depend on the actual sequence. Indeed, different pulse sequences were developed to measure the relaxation times. For instance, Gradient Echo simply repeats the Free Induction Decay described above and allows to sample T2*. Most sequences often comprise additional RF pulses following the slice selection one, to partially refocus the transverse magnetization and produce an echo, leading to a more reliable measure. Spin-Echo is the application of a 90 degree pulse followed by a 180 degree pulse after a time TE/2. This second pulse, which refocuses the transverse magnetization and results in an echo at time TE (Echo Time), removes local field inhomogeneities dephasing, hence allowing to directly measure the T2 decay. On the other hand, Inversion Recovery, which relies on a 180 degree pulse followed after a time TI (Inversion Time) by a 90 degree pulse, enhances the T1 weighting. The choice of the specific pulse sequence parameters (TR, TE, TI,...) finally determines the image contrast. Two distinct tissues may for instance have similar T1 values but distinct

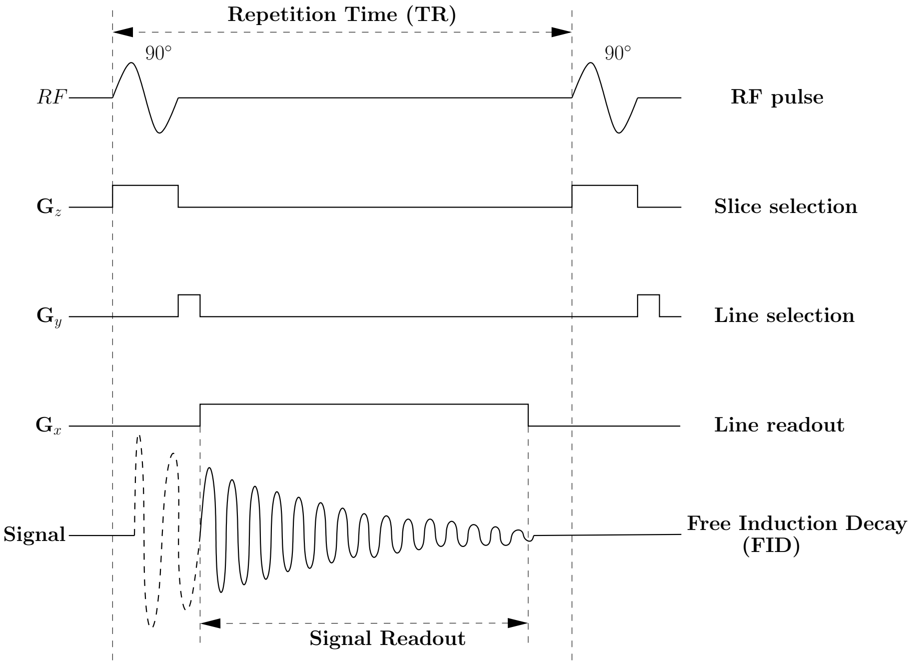


Figure 3.13: A simplified MRI pulse sequence timing diagram.

T2 values, so the choice of the sequence depends on the information of interest.

The straightforward application of a given pulse sequence allows to get a static image contrasting different tissues. However, based on the same principles, it is possible to indirectly image dynamic processes such as oxygen flow or the motion of water molecules. In the next section, we introduce the basic principles of diffusion MRI.

3.3.2 Diffusion MRI

Diffusion MRI is the unique non-invasive technique that allows to probe and quantify the diffusion of water molecules in the body. By modeling the local anisotropy of this diffusion process, it becomes possible to indirectly infer the architecture and properties of tissues such as the brain white matter.

Physical principles of Diffusion Tensor Imaging (DTI)

Above the absolute zero temperature, water molecules freely move and collide with each other in an isotropic medium according to Brownian motion (figure 3.14) [42]. At a macroscopical scale, this phenomenon yields a diffusion process. In an isotropic medium, the diffusion coefficient D was related by Einstein [105] to the root mean square of the diffusion distance:

$$D = \frac{1}{6\tau}\langle \mathbf{R}^T\mathbf{R}\rangle \tag{3.2}$$

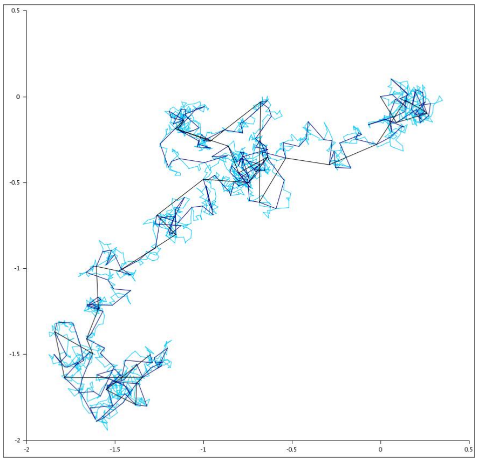


Figure 3.14: An image of Brownian motion, done with three different step sizes. The hierarchical structure is clearly visible. More saturated colors represent smaller step sizes. Image under the Gnu Free Documentation License 1.2

In this expression, τ is the diffusion time and R is the net displacement vector $\mathbf{R} = \mathbf{r} - \mathbf{r}_0$, with $\mathbf{r}_0$ the original position of a particle and $\mathbf{r}$, its position after the time τ. $\langle\rangle$ denotes an ensemble average.
The scalar constant D, known as the diffusion coefficient, measures the molecules mobility in the isotropic case and depends on the molecule-type and the medium properties but not on the direction. For example, at normal brain temperature, 68% of the water molecules diffuse in 50ms in a sphere of 17 μm diameter.

In anisotropic biological tissues, water molecules mobility is constrained by obstacles formed by surrounding structures, such as the axonal membranes in the brain. Moreover, it is known that the myelin sheath can modulate the anisotropy of the diffusion while the microtubules and neurofilaments do not modify it [29]. In this case, the scalar diffusion coefficient D must be replaced by a bilinear operator $\mathbf{D}$. Einstein relation 3.2 can be generalized be considering the covariance matrix of the net displacement vector R

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_{xx} & \mathbf{D}_{xy} & \mathbf{D}_{xz} \\ \mathbf{D}_{xy} & \mathbf{D}_{yy} & \mathbf{D}_{yz} \\ \mathbf{D}_{xz} & \mathbf{D}_{yz} & \mathbf{D}_{zz} \end{pmatrix} = \frac{1}{6\tau}\langle \mathbf{R}\mathbf{R}^T \rangle \tag{3.3}$$

It was proposed in 1994 by Basser *et al.* [24] to use this second order symmetric and positive-definite tensor to model the intrinsic diffusion properties of biological tissues. The diffusion coefficient d related to any direction $\overrightarrow{u} \in \mathbb{R}^3$ is given by:

$$d = \overrightarrow{u}^T \mathbf{D} \overrightarrow{u}$$

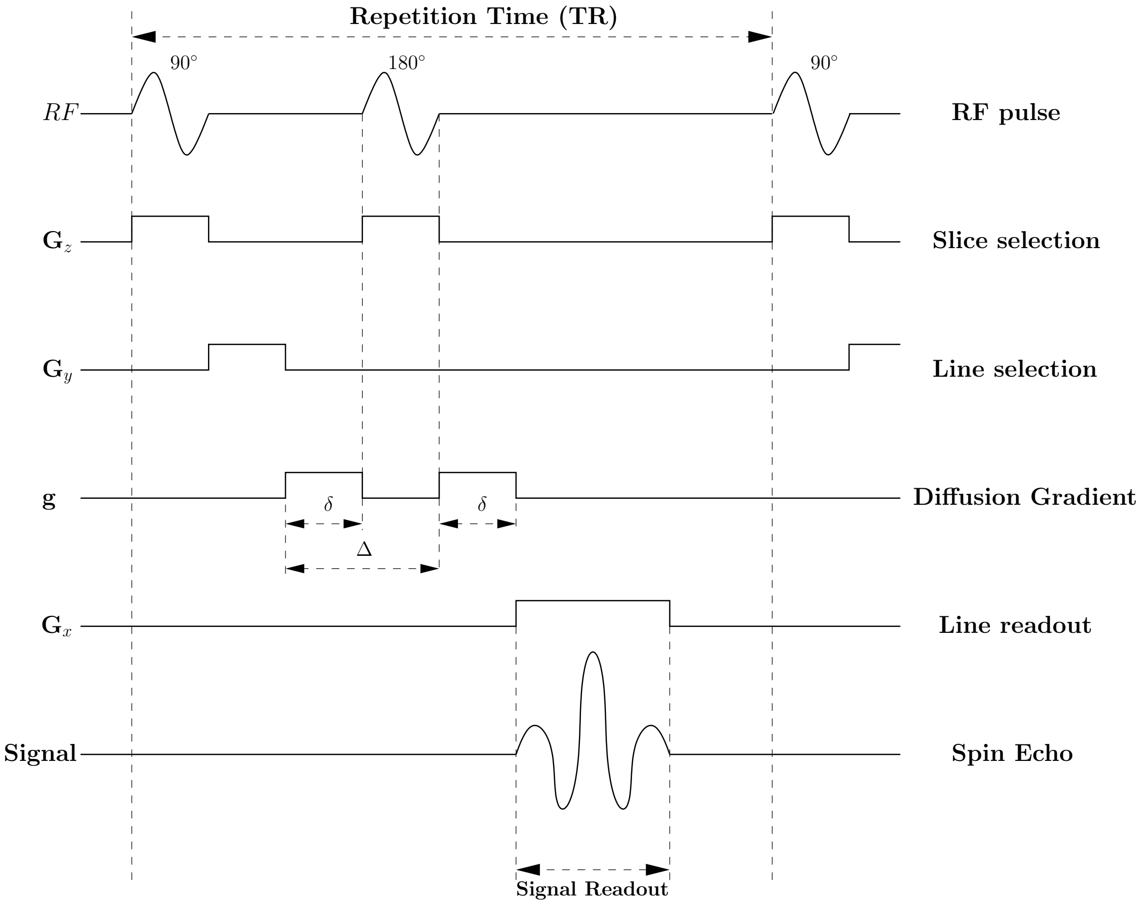


Figure 3.15: Stejskal-Tanner imaging sequence.

It possible to introduce a Gaussian model for water molecules free diffusion. The probability to find a molecule, initially at position $\mathbf{r}_0$, at $\mathbf{r}$ after a delay τ is given by:

$$p(\mathbf{r}|\mathbf{r}_0,\tau) = \frac{1}{\sqrt{(4\pi\tau)^3|\mathbf{D}|}} \exp\left(-\frac{(\mathbf{r}-\mathbf{r}_0)^T\mathbf{D}^{-1}(\mathbf{r}-\mathbf{r}_0)}{4\tau}\right) \tag{3.4}$$

One of the first problems encountered in Diffusion Tensor Imaging (DTI) is to estimate the 6 independent parameters of $\mathbf{D}$. This can be achieved with a minimum of 6 diffusion weighted images (DWI), each measuring a $T2$ signal attenuation related to the diffusion coefficient in a specific direction $\overrightarrow{g_i} = \frac{\overrightarrow{\mathbf{g}_i}}{|\overrightarrow{\mathbf{g}_i}|}$, $i = 1, ..., N$, plus one reference image acquired without any diffusion weighting. The diffusion weighted images can be obtained with an appropriate imaging sequence using diffusion gradients $\overrightarrow{\mathbf{g}_i}$.

Imaging sequence

To measure water molecules diffusion in a given direction $g_i, \quad i = 1, ..., N$ (for the sake of clarity, we note $\mathbf{g}_i = \overrightarrow{\mathbf{g}_i}$ in the remainder), the Stejskal-Tanner imaging sequence [275] is used (figure 3.15). This sequence uses two gradient pulses $\mathbf{g}(t)$ in the direction g, of duration time δ, to control the diffusion weighting. They are placed before and after a 180 degrees refocusing pulse. More specifically, a first 90 degrees RF is applied to flip the magnetization in the transverse plane. The first gradient pulse then causes a phase shift ϕ_1 of the spins whose position is now a function of time $r(t)$:

$$\phi_1(t) = \gamma \int_0^{\delta} \mathbf{g}(t)^T \mathbf{r}(t) dt \tag{3.5}$$

Spin position is in fact assumed to stay constant during time δ. Finally, the 180 degrees pulse combined with the second gradient pulse induces another phase shift

$$\phi_2(t) = -\gamma \int_{\Delta}^{\Delta+\delta} \mathbf{g}(t)^T \mathbf{r}(t) dt \tag{3.6}$$

It is applied after a time Δ separating the two gradient pulses. This pulse cancels the phase shift ϕ_1 only for static spins. On the other hand, spins under Brownian motion during the time period Δ separating the two pulses undergo different phase shifts by the two gradient pulses, resulting in a $T2$ signal attenuation [58].

Figure 3.16 shows examples of diffusion weighted images acquired with two different directions $\mathrm{g}(t)$. It illustrates the direction specific attenuation related to white matter fibers orientation. By assuming the pulses to be infinitely narrow (see [288] for instance), equations 3.5 and 3.6 can be rewritten to yield a net phase shift

$$\phi = \phi_1 + \phi_2 = \gamma\delta\mathbf{g}^T(\mathbf{r}(\mathbf{0}) - \mathbf{r}(\boldsymbol{\Delta})) = \gamma\delta\mathbf{g}^T\mathbf{R}$$

where $\mathbf{R}$ denotes the spin displacement between the two pulses. For the remaining of this section, it is convenient to introduce the displacement reciprocal vector $\mathbf{q} = \gamma\delta\mathrm{g}$ [288].

The signal attenuation can be modeled by the following equation [141]

$$S(\mathbf{q}, \tau) = S_0 \langle \exp{(i\phi)} \rangle \tag{3.7}$$

where S_0 is the reference signal without diffusion gradient. This expression can be rewritten as follows:

$$S(\mathbf{q}, \tau) = S_0 \int_{\mathbb{R}^3} p(\mathbf{r}|\mathbf{r}_0, \tau) \exp{\left(i\mathbf{q}^T\mathbf{R}\right)} dr \tag{3.8}$$

where $p(\mathbf{r}|\mathbf{r}_0, \tau)$ is the so-called ensemble-average diffusion propagator (EAP) [161, 288]. It is easy to see in equation 3.8 that the ratio $\frac{S(\mathbf{q},\tau)}{S_0}$ is nothing but the Fourier transform of the EAP. This is a key observation that is at the core of q-space or diffusion displacement imaging [46] since it potentially gives access to the complex diffusion profile of water molecules at each voxel. However, the actual computation of the inverse Fourier transform of $S(\mathbf{q}, \tau)$ is difficult in practice and has given rise to many acquisition and computational techniques to approximate the EAP.

Diffusion Spectrum Imaging (DSI) was proposed by Tuch *et al.* [289, 192] and

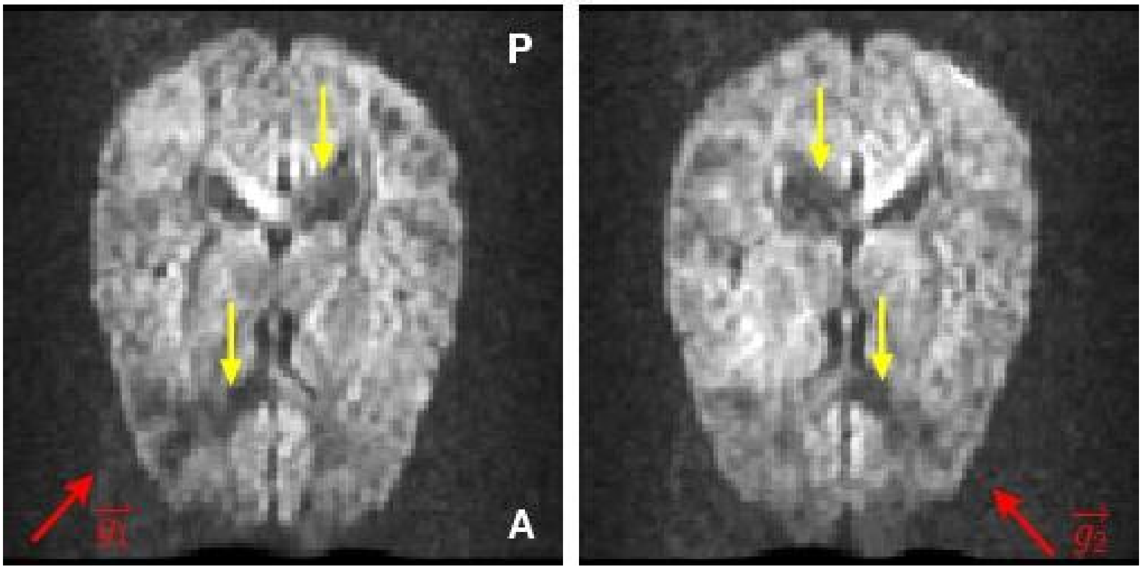


Figure 3.16: Axial slice of diffusion-weighted images (DWI) with two different diffusion gradient directions (red arrows). MR signal attenuation is found in regions having fibers mostly aligned with diffusion gradient direction (yellow arrows).

based on the sampling of a three-dimensional Cartesian grid (of typical size $11 \times 11 \times 11$) at each voxel. The subsequent 3D inverse Fourier transform of the modulus of the diffusion signal yields the EAP. The major drawback of this technique is its extremely high acquisition time. In order to alleviate this constraint, Tuch proposed to sample the q-space only on a shell since we are, in fact, only interested in the angular information of the EAP to differentiate multiple fiber orientations within a given voxel. He showed [287] that it was indeed possible to reconstruct the Orientation Distribution Function (ODF), ie. the radial projection of the EAP,

$$\psi(u) = \int_0^\infty p(\rho u | \mathbf{r}_0, \tau) d\rho$$

by working directly on the sphere and thus bypassing the 3D grid sampling necessary for DSI. Many techniques have been proposed to compute ODFs from High Angular Resolution Imaging (HARDI) [118, 290, 225, 50, 90, 91]. HARDI typically requires the acquisition of 30 to several hundreds diffusion weighted images with different non collinear diffusion gradients g_i to be able to clearly discriminate multiple diffusion directions.

If we make the assumption of free diffusion, the probability density function $p(\mathbf{r}|\mathbf{r}_0, \tau)$ can be written as

$$p(\mathbf{r}|\mathbf{r}_0, \tau) = \frac{1}{\sqrt{(4\pi\tau D)^3}} \exp \frac{|\mathbf{r} - \mathbf{r}_0|^2}{4\tau D}$$

for isotropic media and it becomes as in equation 3.4 for anisotropic media. Using these expressions of the EAP yields simple expressions for the signal $S(\mathbf{q}, \tau)$, ie. respectively [275]

$$S(\mathbf{q}_i, \tau) = S_0 \exp\left(-bD\right)$$

(it is independent on the direction g_i) and

$$S(\mathbf{q}_i, \tau) = S_0 \exp\left(-b g_i^T \mathbf{D} g_i\right)$$

where b is the diffusion weighting factor depending on scanner parameters and proposed by Le Bihan *et al.* [36]:

$$b = \gamma^2 \delta^2 |\mathbf{g}|^2 \left(\Delta - \frac{\delta}{3}\right)$$

We recall that |g| is the magnitude of the diffusion gradient pulse, δ its duration and Δ the time separating two pulses (see figure 3.15).

Hence, signal attenuation, i.e. signal sensitivity to water molecules diffusion, is stronger if the diffusion coefficient $g_i^T \mathbf{D} g_i$ is important. Note also the importance of the b factor that has to be appropriately tuned with respect to $g_i^T \mathbf{D} g_i$ to avoid either a very low signal attenuation if b is too small or a poor SNR if b is too high. A typical value is $b = 1000 s.mm^{-2}$.

For the purpose of DTI, images are collected with one or more b factor(s) and at least 6 independent gradient directions g_i and one reference image S_0. The diffusion tensor $\mathbf{D}$ can then be estimated at each voxel using the $S(\mathbf{q}_i, \tau)$ and S_0. The classical method to derive the tensors uses least squares technique, but various alternative methods have been proposed. We will come back to this particular point in chapter 6. We finally end-up with a diffusion tensor image, i.e. a 3D image with 6 parameters describing the local tensor $\mathbf{D}$ at each voxel. From the eigenvalue decomposition of $\mathbf{D}$, one can visualize the diffusion in each voxel by a diffusion ellipsoid: the directions of the main axes are given by the eigenvectors of $\mathbf{D}$ and their lengths are proportional to the square root of their respective eigenvalues. If all the eigenvalues are of the same magnitude, the ellipsoid will be spherical, while if one of the eigenvalues is much greater than the others, it will be more elongated. More details can be found for instance in [27, 310]. Figure 3.17 illustrates the corresponding ellipsoids field in an axial slice. The blue (respectively red) color refers to elongated anisotropic (resp. spherical isotropic) ellipsoids.

3.4 CONCLUSION

We have presented, in the first section of this chapter, an overview of the

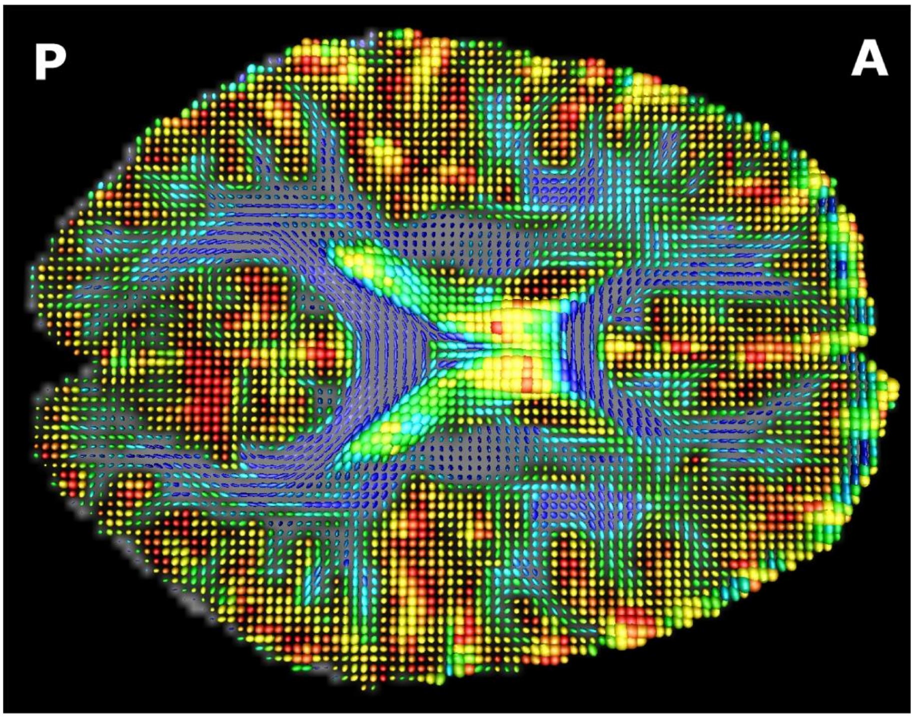


Figure 3.17: Axial slice of a Diffusion Tensor image. At each voxel, an ellipsoid represents the estimated diffusion tensor.

human nervous system with a particular emphasis on the cerebrum and, more specifically, on the grey and white matters. The goal of this thesis was indeed to develop new processing tools for a recent neuroimaging modality named DTI and, in the next chapters, we will propose efficient numerical algorithms to recover the complex architecture of the brain white matter networks from DTI. The objective of the second section of this chapter was to recall basic notions on magnetic resonance imaging and, in particular, on diffusion MRI. The following chapter is meant to be an introduction to differential geometry since we will heavily rely on concepts from this area on subsequent parts of this manuscript.

CHAPTER 4

NOTIONS OF DIFFERENTIAL GEOMETRY

Contents

4.1 INTRODUCTION

Differential geometry, and in particular Riemannian geometry, will play a central role in this thesis and we hereafter give an overview of important notions of this field. In section 4.2, after the description of a differentiable manifold, we introduce the notion of metric and geodesic. Next, we discuss the concepts of connection and parallel transport which will be important in chapter 9. Finally, we talk about curvature since it will be useful in chapter 5. Section 4.3 shows how to apply all these notions to manifold-valued data to compute statistics and spatial gradient. We refer the reader to [33], [96, 97], [32] or [176] for in-depth studies of differentiable and Riemannian manifolds.

Keywords: manifold, metric, geodesic, connection, Christoffel symbols, covariant derivative, curvature

4.2 RIEMANNIAN GEOMETRY BASICS

Developed by Bernhard Riemann in the nineteenth century, Riemannian geometry studies smooth manifolds equipped with a Riemannian metric. These two fundamental concepts will be explained in the following. A Riemannian metric gives local information on angles, length of curves and volumes which make Riemannian geometry particularly suitable to work on curves and surfaces.

4.2.1 Differential Manifold

Manifolds are generalization of linear spaces like $\mathbb{R}^n$ that only locally look like $\mathbb{R}^n$. Their global structure can however be much more complicated. A formal definition is the following:

Definition 4.2.1.1. *An n-manifold $\mathcal{M}$ is a Hausdorff space, ie. a topological space in which points can be separated by neighborhoods, where every points has a neighborhood that is homeomorphically mapped onto an open Euclidean n-dimensional ball $\mathcal{B} = \{x \in \mathbb{R}^n / |x| < 1\}$.*

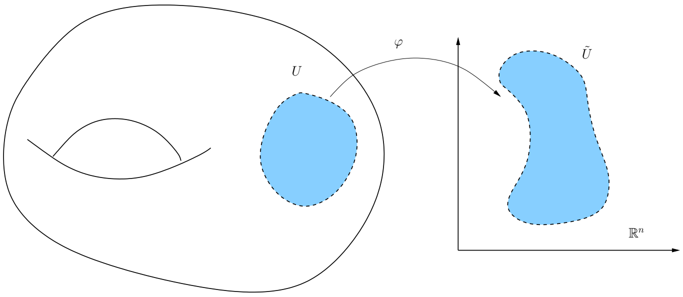


Figure 4.1: A coordinate chart (U, φ)

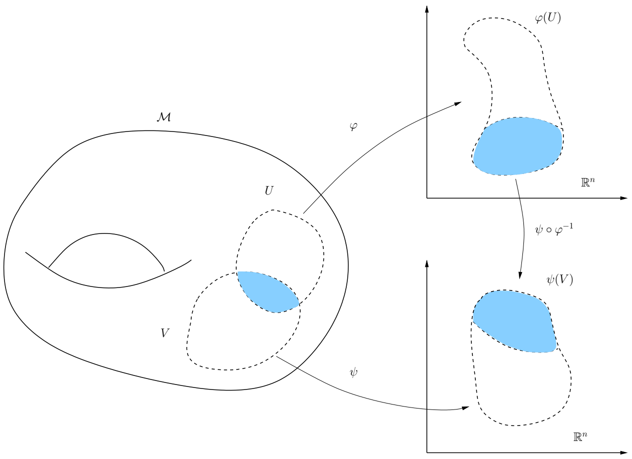


Figure 4.2: A transition map

In general, manifolds have additional properties, like differentiability, that reflects some degree of consistency between maps and overlapping maps. Differentiability is necessary to perform calculus on manifolds like we do on $\mathbb{R}^n$. To understand the notion of differentiable manifold, we need to introduce the concept of coordinate chart.

Definition 4.2.1.2. *A coordinate chart (or just chart) on an n-manifold $\mathcal{M}$ is a pair (U, φ) where U is an open subset of $\mathcal{M}$ and $\varphi : U \to \tilde{U}$ is a homeomorphism from U onto an open subset $\tilde{U} = \varphi(U)$ of $\mathbb{R}^n$.*

U is called a coordinate domain and, if $\varphi(U)$ is an open ball in $\mathbb{R}^n$, it is called a coordinate ball. Moreover, φ is called a local coordinate map such that for any point $p \in \mathcal{M}$, $\varphi(p) = \mathbf{x} = (\mathbf{x}^1, ..., \mathbf{x}^n)^T$ defines the local coordinates of p on U.

Charts have to be at least compatible. That means that, for two charts (U, φ) and (V, ψ) with $U \cap V$ non empty, the map taking a point from $\varphi(U \cap V)$ into $\psi(U \cap V)$ must be a homeomorphism. This map, defined as the composition $\psi \circ \varphi^{-1}$ is called a transition map (or change of coordinates) (figure 4.2). We call atlas a collection of coordinate charts whose union of coordinate domains entirely covers the manifold $\mathcal{M}$ (figure 4.3). A differentiable manifold is then simply a manifold with a differentiable atlas, in other words with a chart whose any transition map is differentiable. Here is a more formal definition:

Definition 4.2.1.3. *A differentiable manifold of dimension n is a set $\mathcal{M}$ and a family of coordinate charts (the atlas) $(U_\alpha, \varphi_\alpha)$ with $U_\alpha \subset \mathcal{M}$ and $\varphi_\alpha : \mathcal{M} \to \mathbb{R}^n$ such that:*

- $\bigcup_\alpha U_\alpha = \mathcal{M}$
- *for any α,β with $U_\alpha \cap U_\beta = W \neq \emptyset$, the sets $\varphi_\alpha(W)$ and $\varphi_\beta(W)$ are open sets in $\mathbb{R}^n$ and the mappings $\varphi_\beta \circ \varphi_\alpha^{-1}$ are differentiable.*

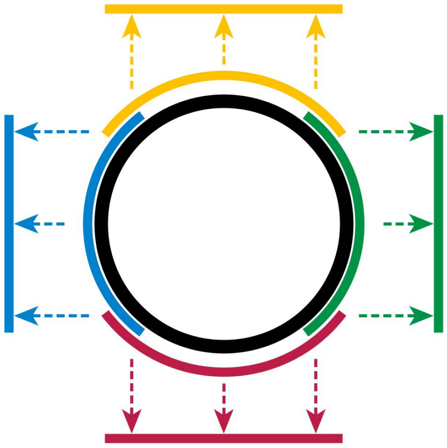


Figure 4.3: An example of atlas: Each of the four charts maps part of the circle to an open interval, and together, they cover the whole circle. (Image under the Creative Commons Attribution ShareAlike 2.5 License)

If the transition maps are C^∞ (ie. infinitely differentiable or smooth), we have a smooth atlas and thus a smooth manifold. Riemannian geometry studies smooth manifolds with Riemannian metrics and we now introduce the notions of metric and geodesic.

4.2.2 Metric, Geodesic, Distance

Metric

For an n-dimensional manifold $\mathcal{M}$, a Riemannian metric is a collection of inner products $\langle .,. \rangle_p$ defined for every point p of $\mathcal{M}$. These inner products are defined on the tangent space $T_p\mathcal{M}$ of $\mathcal{M}$ at p and provide a natural way to measure the lengths of vectors tangent to $\mathcal{M}$ at location p.

First, let us define the notions of tangent vector and tangent space. A tangent space, $T_p\mathcal{M}$ is simply an n-dimensional real vector space, attached to p, which contains all the possible *directions* of curves passing through p. The notion of *direction* can be clarified by thinking of it as a direction of derivation along a curve. Let $\gamma : [-\epsilon, \epsilon] \rightarrow \mathbb{R}^n$ be a differentiable curve in $\mathbb{R}^n$ such that $\gamma(0) = p$ and $\gamma(t) = \left(\mathbf{x}^1(t), ..., \mathbf{x}^n(t)\right)^T \in \mathbb{R}^n$, $\forall t \in [-\epsilon, \epsilon]$. Then

$$\frac{d\gamma(0)}{dt} = \dot{\gamma}(0) = \left(\dot{\mathbf{x}}^1(0), ..., \dot{\mathbf{x}}^n(0)\right)^T \in \mathbb{R}^n$$

can be used to define the directional derivative of a function $f : \mathcal{M} \rightarrow \mathbb{R}$ in the direction $\dot{\gamma}(0)$.

Definition 4.2.2.1. *A differentiable function* $\gamma : [-\epsilon, \epsilon] \rightarrow \mathcal{M}$ *is called a differentiable curve in* $\mathcal{M}$. *Let* $\gamma(0) = p \in \mathbb{M}$ *and let* $\mathcal{D}$ *be the set of functions* f *on* $\mathcal{M}$ *defined in a*

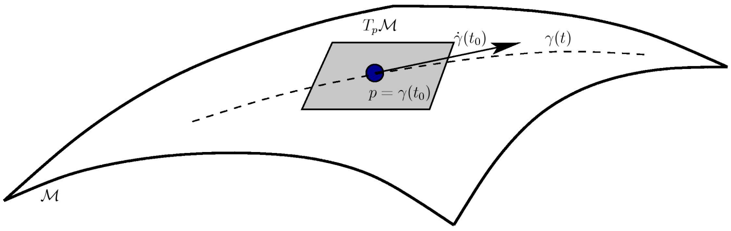


Figure 4.4: Tangent space at p of a Riemannian manifold $\mathcal{M}$

neighborhood of p *and differentiable at* p*.* f *can be restricted to* γ *and its directional derivative with respect to* $\dot\gamma(0)$ *is given by*

$$\left.\frac{d(f\circ\gamma)}{dt}\right|_{t=0} = \sum_i \left.\frac{\partial f}{\partial \mathbf{x}^i}\right|_{t=0} \left.\frac{d\mathbf{x}^i}{dt}\right|_{t=0} = \left(\sum_i \dot{\mathbf{x}}^i(0)\frac{\partial}{\partial \mathbf{x}_i}\right) f = \dot\gamma(0) f \tag{4.1}$$

The tangent vector to the curve γ *at* $t = 0$ *is the function* $\dot\gamma(0) : \mathcal{D} \to \mathbb{R}^n$ *given by equation 4.1, for all functions* $f \in \mathcal{D}$*.*

A tangent vector at p is simply the tangent vector at $t = 0$ of any curve $\gamma : [-\epsilon, \epsilon] \to \mathcal{M}$ starting at p. The set of all tangent vectors at p defines the tangent space of $\mathcal{M}$ at p, $T_p\mathcal{M}$ (figure 4.4). Given a local coordinate map φ, it is also possible to determine a basis in the tangent space $T_p\mathcal{M}$ denoted by $\left(\frac{\partial}{\partial \mathbf{x}_1}, \cdots, \frac{\partial}{\partial \mathbf{x}_n}\right)$ (and also written $(\partial_1, \cdots, \partial_n)$).

Any element of the tangent space can hence be expressed in the form $\sum_i \mathbf{x}^i \frac{\partial}{\partial \mathbf{x}_i}$ and the inner products $\langle \frac{\partial}{\partial \mathbf{x}_i}, \frac{\partial}{\partial \mathbf{x}_j} \rangle_p$ define an $n \times n$ symmetric, bilinear and positive-definite form $G = g_{ij}$ known as the *local representation of the Riemannian metric*. The inner product of two tangent vectors u and v of $T_p\mathcal{M}$ is then expressed as

$$\langle u, v\rangle_p = u^T G v$$

(the reference to the location p is usually discarded in the notation g_{ij}). Here is a more formal definition of the Riemannian metric:

Definition 4.2.2.2. *A Riemannian metric on a differentiable manifold* $\mathcal{M}$ *is a correspondence associating a symmetric, bilinear and positive-definite form* $\langle\, , \,\rangle_p$ *on* $T_p\mathcal{M}$ *to each point* $p \in \mathcal{M}$*. This form varies differentiably in the sense that if* $\varphi : U \subset \mathcal{M} \to \tilde{U} \subset \mathbb{R}^n$ *is a coordinate map defined in a neighborhood of* p *with* $\varphi(q) = \left(\mathbf{x}^1, ..., \mathbf{x}^n\right)^T \in \mathbb{R}^n$ *and* $\frac{\partial}{\partial \mathbf{x}_i}(q) = d\varphi_q^{-1}(0, ..., 1, ...0)$ *then* $\langle \frac{\partial}{\partial \mathbf{x}_i}(q), \frac{\partial}{\partial \mathbf{x}_j}(q)\rangle_q = g_{ij}\left(\mathbf{x}^1, ..., \mathbf{x}^n\right)$ *is a differentiable function on* $\tilde{U}$ *(for all* q *in the neighborhood of* p*).*

Geodesic and Distance

Equipped with these notions we can now define the concept of geodesic on a Riemannian manifold $\mathcal{M}$. It is the equivalent of straight line in Euclidean spaces and defined as the locally length-minimizing piecewise smooth curve $\gamma : I \subset \mathbb{R} \mapsto \mathcal{M}$ characterized by the fact that it is autoparallel, or in other words, that the field of tangent vectors $\dot{\gamma}(t)$ stays parallel along $\gamma(t)$ (this point will be detailed in the next paragraph). It is equivalent to say that, in local coordinates notations, a curve is a geodesic if and only if it is the solution of the n second order Euler-Lagrange equations:

$$\frac{d^2\mathbf{x}^k(t)}{dt^2} + \sum_{i,j=1}^{n} \Gamma_{ij}^k \frac{d\mathbf{x}^i(t)}{dt}\frac{d\mathbf{x}^j(t)}{dt} = 0 \quad \forall k = 1, ..., n \tag{4.2}$$

where the Γ_{ij}^k are the so-called Christoffel symbols of the second kind, introduced in the next paragraph.

The tangent vector $\dot{\gamma}(t) = \frac{d\gamma(t)}{dt} \in T_{\gamma(t)}\mathcal{M}$ defines the instantaneous speed of the curve and its norm $|\dot{\gamma}(t)| = \langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)}^{1/2}$ is the instantaneous velocity. Integrating $|\dot{\gamma}(t)|$ along γ yields its length between the two endpoints $p_1 = \gamma(t_1)$ and $p_2 = \gamma(t_2)$, which we note $\mathcal{L}_\gamma(p_1, p_2)$:

$$\mathcal{L}_\gamma(p_1, p_2) = \int_{t_1}^{t_2} |\dot{\gamma}(t)|_{\gamma(t)} dt = \int_{t_1}^{t_2} \sqrt{\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)}} dt$$

The geodesic distance between any two points p_1 and p_2 is defined as the infinimum, over all the possible curves joining p_1 and p_2, of the length of those curves:

$$\mathcal{D}(p_1, p_2) = \inf_\gamma \{\mathcal{L}_\gamma(p_1, p_2) : p_1 = \gamma(t_1), p_2 = \gamma(t_2)\}$$

Finally, taking $I = [0, 1]$ for simplicity, it is possible to show, under certain conditions, that a geodesic $\gamma : [0, 1] \mapsto \mathcal{M}$ is uniquely defined by its starting point $\gamma(0)$ and its initial velocity $\dot{\gamma}(0) \in T_{\gamma(0)}\mathcal{M}$. The endpoint $\gamma(1)$ can be easily computed by applying the exponential map at $\gamma(0)$ to $\dot{\gamma}(0)$: $\gamma(1) = \exp_{\gamma(0)}(\dot{\gamma}(0))$. A detailed presentation of this map can be found in [97]. The inverse map, known as the logarithm map of $\gamma(1)$ at $\gamma(0)$: $\log_{\gamma(0)}(\gamma(1))$, yields the unique tangent vector $\dot{\gamma}(0)$ if we know the two endpoints of the curve.
Moreover, it can be proved that

$$\mathcal{D}(\gamma(0), \gamma(1)) = \langle \dot{\gamma}(0), \dot{\gamma}(0) \rangle_{\gamma(0)}^{1/2}$$

In the following chapters, we will often use the fact that the velocity $\dot{\gamma}(0)$ can be computed from the gradient of the squared geodesic distance with respect to $\gamma(0)$. In other words, we have

$$\dot{\gamma}(0) = -\nabla_{\gamma(0)} \mathcal{D}^2(\gamma(0), \gamma(1)) \tag{4.3}$$

Using these concepts, we will be able to define the notions of mean and covariance matrix on a Riemannian manifold in section 4.3.

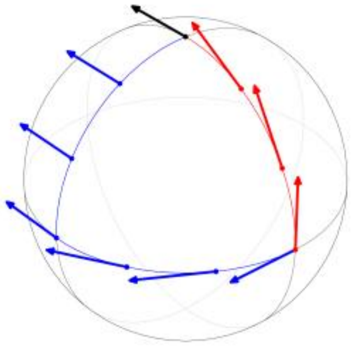


Figure 4.5: Transporting a vector along two different curves with the same initial and terminal point on a manifold gives different results. The pictures illustrates the use of the Levi-Civita connection for parallel transport along the red an blue curves on the sphere. (Image by Florian Jung in the public domain)

4.2.3 Affine Connection

Before introducing the notion of curvature, we need to precise the notion of connection. It is a crucial concept in geometry since it allows to transport quantities along curves in a consistent manner and, ultimately, to compare local geometries defined at different locations of a manifold [191]. In other words, the connection makes it possible to map any tangent space $T_{p_1}\mathcal{M}$ at p_1 onto another tangent space $T_{p_2}\mathcal{M}$ at p_2. The need for the definition of such a mapping arises from the following observation: As depicted on figure 4.5, let us imagine that we want to transport a given vector (the black arrow), in a parallel manner, from its original location p_1 to a different point p_2 of a manifold (the sphere here). In general, the parallel transport procedure is dependent on the choice of coordinate system, which is not desirable. The Levi-Civita parallelism [190], introduced in 1917, allows to solve this inconsistency by simply rotating the sphere so that p_1 follows the curve of interest without axis rolling.
If this procedure is applied along two different curves, joining p_1 and p_2 (the red and blue arrows on figure 4.5), the vectors resulting from both parallel transports will be different. This reflects the curvature of the sphere. We now detail the notion of covariant derivative, parallel transport and Levi-Civita connection.

Covariant derivative and parallel transport

The dependence of parallel transport on the choice of coordinate system directly comes from the fact that the classical directional derivative does not behave well under changes of the coordinate system. In fact, if $\mathcal{M}$ is a surface in $\mathbb{R}^3$, $\mathcal{C} : I \subset \mathbb{R} \to \mathcal{M}$ is a curve in $\mathcal{M}$ and $V : I \to \mathbb{R}^3$ is a vector field tangent to $\mathcal{M}$, the quantity $\frac{dV}{dt}(t)$, $\forall t \in I$ is not, in general, in the tangent space $T_{\mathcal{C}(t)}\mathcal{M}$.

It is possible to solve this problem, ie. to make the differentiation intrinsic, by generalizing the directional derivative so that it behaves well under changes of the coordinate system. It is achieved by considering the orthogonal projection $\frac{DV}{dt}(t)$ of $\frac{dV}{dt}(t)$ on the tangent space $T_{\mathcal{C}(t)}\mathcal{M}$. $\frac{DV}{dt}(t)$ is called the covariant derivative.
$\frac{DV}{dt}(t)$, which is also frequently noted $\nabla_X V$, with $X = \dot{\mathcal{C}}(t)$ is independent on the chosen coordinate system. In fact, it can be expressed as a modification of the directional derivative by the operator known as the Christoffel symbols Γ^k_{ij}:

$$(\nabla_X V)^k = X^i \partial_i V^k + \Gamma^k_{ij} X^i V^j$$

where $X = X^i \partial_i$ and $V = V^i \partial_i$. The concept of parallelism can then be expressed quite naturally:

Definition 4.2.3.1. *Let $\mathcal{M}$ be a differentiable manifold. A vector field V along a curve $\mathcal{C} : I \to \mathcal{M}$ is called parallel when $\frac{DV}{dt}(t) = \nabla_{\dot{\mathcal{C}}(t)} V(t) = 0$ for all $t \in I$.*

Moreover, we have the following

Proposition 4.2.3.1. *Let $\mathcal{M}$ be a differentiable manifold. Let $\mathcal{C} : I \to \mathcal{M}$ be a differentiable curve on $\mathcal{M}$ and V_0 a tangent vector in $T_{\mathcal{C}(t_0)}\mathcal{M}$ at $\mathcal{C}(t_0)$, $t_0 \in I$. There exists a unique parallel vector field V along $\mathcal{C}$ with $V(t_0) = V_0$ and we call $V(t)$ the parallel transport of $V(t_0)$ along $\mathcal{C}$.*

To be parallel along a curve $\mathcal{C}$, a vector field V must verify the system of n differential equations in $V^k(t)$:

$$\left(\nabla_{\dot{\mathcal{C}}(t)} V(t)\right)^k = \frac{dV^k}{dt}(t) + \Gamma^k_{ij} \dot{\mathcal{C}}^i(t) V^j(t) = 0 \quad k = 1, ..., n$$

It is interesting to note that, by definition, the geodesics are the curves whose velocity field is parallel. Parallel transport will be important for the registration technique presented in chapter 9.

We can note that, like the directional derivative, $\nabla_{\dot{\mathcal{C}}(t_0)} V$ depends on $V(t_0)$ around $t = t_0$ but only depends on the value of $\dot{\mathcal{C}}(t)$ at t_0. An affine connection can directly be defined from the covariant derivative. We have the following

Definition 4.2.3.2. *Let $\Xi(\mathcal{M})$ denote the space of vector fields of class C^∞ on $\mathcal{M}$. An affine connection ∇ on $\mathcal{M}$ is a mapping $(X, Y) \to \nabla_X Y$*

$$\nabla : \Xi(\mathcal{M}) \times \Xi(\mathcal{M}) \to \Xi(\mathcal{M})$$

with the properties: For all functions of class C^∞ and vector fields X, Y, Z on $\mathcal{M}$,

1. $\nabla_{fX+gY} Z = f\nabla_X Z + g\nabla_X Z$
2. $\nabla_X (Y + Z) = \nabla_X Y + \nabla_X Z$

3. $\nabla_X(fY) = f\nabla_X Y + (\partial_X f)Y$

The so-called Levi-Civita connection of a Riemannian manifold is symmetric and parallel transports tangent vectors along a curve while preserving the inner product of this vector along the curve (we say it is compatible with the metric). It is defined in each local chart by the n^3 Christoffel symbols and can thus be directly obtained from the metric, hence the name metric connection:

$$\nabla_{ij}^k = \Gamma_{ij}^k = g^{kl}\Gamma_{ijl} = \frac{1}{2}g^{kl}\left(\frac{\partial g_{jl}}{\partial \sigma_i} + \frac{\partial g_{il}}{\partial \sigma_j} - \frac{\partial g_{ij}}{\partial \sigma_k}\right) \; i,j,k,l = 1,...,n \tag{4.4}$$

We now briefly discuss the notion of curvature of Riemannian manifolds.

4.2.4 Curvature

The Riemann curvature tensor

The notion of curvature for Riemannian manifolds of dimension at least 3 cannot be fully described by a scalar quantity at each point p of the manifold. Riemann introduced the curvature tensor, in terms of the Levi-Civita connection. It can be constructed from the metric tensor and its first and second derivatives. We have the

Definition 4.2.4.1. *Let $\mathcal{M}$ be a differentiable manifold with an affine connection ∇. The Riemann curvature tensor R is a correspondence that associates to every pair $(X,Y) \in \Xi(\mathcal{M}) \times \Xi(\mathcal{M})$ a mapping $R(X,Y) : \Xi(\mathcal{M}) \to \Xi(\mathcal{M})$ such that*

$$R(X,Y)Z = \nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z + \nabla_{[X,Y]} Z$$

where $Z \in \Xi(\mathcal{M})$ and $[\,,\,]$ denotes the Lie bracket.

That is a quite formal definition, which makes it difficult to "feel" what R actually measures. By taking X and Y to be some elements of a local coordinate basis, ie. respectively ∂_i and ∂_j, we obtain:

$$R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z$$

In other words, the curvature tensor is a measure of the covariant derivative non-commutativity.

In local coordinates $\left(\mathrm{x}^1, ..., \mathrm{x}^n\right)^T$, the curvature tensor can be expressed through the Christoffel symbols as follows (see [97] page 92):

$$R_{ijk}^l = \partial_j \Gamma_{ki}^l - \partial_k \Gamma_{ji}^l + \Gamma_{jm}^l \Gamma_{ki}^m - \Gamma_{km}^l \Gamma_{ji}^m$$

The sectional curvature

Sectional curvature also completely describes the curvature of a Riemannian manifold. A sectional curvature $\kappa(\mathcal{E}, p)$ is associated to any two-dimensional subset of vectors $\mathcal{E}$ in the tangent space at p, $T_p\mathcal{M}$. This 2-plane is called a section. It is defined as the Gaussian curvature of that hypersurface made of the set of points reached by the geodesics starting at p in all the directions described by the plane $\mathcal{E}$. It can be expressed as

$$\kappa(\mathcal{E}, p) = \frac{\langle R(u,v)v, u\rangle}{|u|^2|v|^2 - \langle u, v\rangle^2}$$

for any $u, v \in \mathcal{E}$. Although this definition may seem a little tedious, we will see in the next chapter that it is possible, in some very particular cases, to actually get closed-form expressions for $\kappa(\mathcal{E}, p)$.

The Ricci and scalar curvatures

Respectively up to dimension 3 and 2, those quantities completely describe the curvature of a Riemannian manifold. For manifolds of dimension equal to or greater than 4 they become insufficient. However, they play an important role as we will see, for instance in chapter 5, when we will define the notion of normal distribution over the space of multivariate normal distributions.

The Ricci curvature tensor $\mathcal{R}$ can be thought of as a way to measure how much n-dimensional volumes in regions of an n-dimensional manifold differ from the volumes of equivalent regions in $\mathbb{R}^n$. It can be defined as the contraction of the Riemann curvature tensor

$$\mathcal{R}_{ij} = R^k_{ijk} = R_{ijkl}g^{kl} \tag{4.5}$$

where g^{kl} denotes the inverse of the metric. Finally, the scalar curvature, or Ricci scalar $S \in \mathbb{R}$, is the simplest way to characterize the curvature of a manifold. It is given by the trace of the Ricci curvature tensor

$$S = g^{ij}R_{ij}$$

4.3 MANIFOLD-VALUED DATA

In this section, we consider an image $I : \Omega \subset \mathbb{R}^3 \mapsto \mathcal{M}$ and would like to define the notions of statistics and spatial gradient, while taking into account the properties of the space $\mathcal{M}$ where I takes its values. They will play a central role for instance in the variational formulations (equations 8.16, 8.18 and 9.3) of the segmentation and registration problems to be detailed in chapters 8 and 9. We also show how to compute the spatial gradient of a tensor field which will be useful to introduce a boundary term in our segmentation energy (equation 8.17).

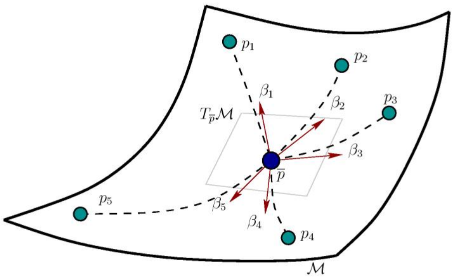


Figure 4.6: Definition of the covariance matrix $\Lambda_{\overline{p}}$

4.3.1 Statistics

As defined by Fréchet in [119] and used by Pennec in [235], the empirical mean of a set of N random elements $\{p_i\},\ i = 1, ..., N$ of $\mathcal{M}$, such as diffusion tensors as we will see in the next chapter, is defined as the minimizer $p = \overline{p}$ of the variance $\sigma_p^2\left(\{p_i\}\right)$ of the p_i with respect to p:

$$\sigma_p^2\left(\{p_i\}\right) = \mathbb{E}\left[\mathcal{D}^2(p, p_i)\right] = \frac{1}{N}\sum_{i=1}^{N}\mathcal{D}^2(p, p_i) \tag{4.6}$$

Here the set $\{p_i\}$ must be seen as the values taken by I in some neighborhood of a pixel $x \in \mathbb{R}^3$. The empirical covariance matrix of the set $\{p_i\}$, with respect to the mean $\overline{p}$ is defined as the expected value of the quantity $\beta_i.\beta_i^T$ and denoted by $\Lambda_{\overline{p}}$. As depicted on figure 4.6, β_i is the initial velocity $\dot{\gamma}_i(0)$ of the i^{th} geodesic joining $\gamma_i(0) = \overline{p}$ to p_i (see equation 4.3) and expressed in local coordinates, i.e. it is taken to be the n-dimensional vector of coordinates $\varphi\left(\dot{\gamma}_i(0)\right) \in \mathbb{R}^n$ and not the tangent vector $\dot{\gamma}_i(0)$ itself. The dot product then boils down to a simple Euclidean dot product and we have:

$$\Lambda_{\overline{p}} = \frac{1}{N}\sum_{i=1}^{N}\varphi\left(\beta_i\right)\varphi\left(\beta_i\right)^T \text{ with } \beta_i = -\nabla_{\overline{p}}\mathcal{D}^2(\overline{p}, p_i) \tag{4.7}$$

where φ is the coordinate chart introduced in section 4.2.2 and also used, for instance, in section 8.2.1. In this section, we will apply these definitions to the Euclidean, Kullback-Leibler and geodesic probability metrics in order to approximate Gaussian distributions of diffusion tensors based on these various dissimilarity measures. We point out, for the sake of clarity, that the expression $-\nabla_{\overline{p}}\mathcal{D}^2(\overline{p}, p_i)$ has nothing to do with the covariant derivative previously introduced. It is the simple gradient of a scalar-valued function with respect to a point of interest.

4.3.2 Spatial Gradient

We recall that we are interested in images I associating to each location of a regular sampling Ω of $\mathbb{R}^3$ an element of a Riemannian manifold $\mathcal{M}$. The spatial gradient of I can be estimated from the gradient of the squared distance as:

$$\nabla_{\mathrm{s}e_k}I(x) = -\frac{\mathrm{s}}{|e_k|}\left(\nabla_{I(x)}\mathcal{D}^2\left(I(x), I(x + \mathrm{s}e_k)\right)\right) \quad \forall x \in \Omega$$

where the $e_k,\ k = 1, 2, 3$ denote the canonical basis of $\mathbb{R}^3$ and are used to access the neighbors of $I(x)$ on the grid Ω. s is either $+1$ or -1 and denotes the forward and backward approximation of the gradient. $\nabla_{-e_1} I(x)$ is, for example, the initial tangent vector of the geodesic joining $I(x)$ and its neighborhood $I(x - |e_1|(1, 0, 0)^T)$. $\nabla_{\mathrm{s}e_k} I(x)$ is a tangent vector living in $T_{I(x)}\mathcal{M}$.
It is then straightforward to compute the squared norm of the gradient at location x as:

$$\begin{aligned} |\nabla I(x)|^2 &= \frac{1}{2}\sum_{k=1}^{3}\sum_{\mathrm{s}=\pm 1} |\nabla_{\mathrm{s}e_k} I(x)|^2_{I(x)} \\ &= \frac{1}{2}\sum_{k=1}^{3}\sum_{\mathrm{s}=\pm 1} \mathcal{D}^2(I(x), I(x + \mathrm{s}e_k)) \end{aligned}$$

where the $\frac{1}{2}$ factor arises from the fact that we use $3 \times 3 \times 3$ neighborhoods (ie. central finite differences).

We purposely wrote the various quantities of interest in terms of the distance $\mathcal{D}$. This will allow us, in the chapter 8, to investigate different metrics and their influences on the computed statistics and gradient norm.

4.4 CONCLUSION

In this chapter, we recalled basic notions for Riemannian geometry that will be at the core the second part of this thesis, where processing tools for diffusion tensor images will be introduced. These concepts will also play a fundamental role in the next chapter where we study the structure of the space of multivariate normal distributions from an information geometry point of view.

CHAPTER 5

THE MANIFOLD OF MULTIVARIATE NORMAL DISTRIBUTIONS

Contents

OVERVIEW

This chapter is dedicated to the statistical analysis of the space of multivariate normal distributions with an application to the processing of Diffusion Tensor Images (DTI). It relies on the differential geometrical properties of the underlying parameters space, endowed with a Riemannian metric, as well as on recent works that led to the generalization of the normal law on Riemannian manifolds. We review the geometrical properties of the space of multivariate normal distributions with zero mean vector and focus on an original characterization of the mean, covariance matrix and generalized normal law on that manifold. We extensively address the derivation of accurate and efficient numerical schemes to estimate these statistical parameters. A major application of the present work is related to the analysis and processing of DTI datasets and we show promising results on synthetic and real examples.

Keywords: multivariate normal distribution, symmetric positive-definite matrix, information geometry, Riemannian geometry, Fisher information matrix, geodesics, geodesic distance, Ricci tensor, curvature, statistics, mean, covariance matrix, diffusion tensor magnetic resonance imaging

5.1 INTRODUCTION

The definition of differential geometrical structures for statistical models started in 1936 with the work of Mahalanobis [200] on multivariate normal distributions of fixed covariance matrix. Since then, several authors in information geometry [199], [9] and references therein, and physics [57] have contributed to the description of those geometries. Rao [253] expressed one of the fundamental results in 1945 by showing that it was possible to use the Fisher information matrix as a Riemannian metric between parameterized probability density functions (*pdf*). The Fisher information is a popular measure of the amount of *information* carried by the realizations of a random variable about the unknown parameters of the underlying probability density. This is classically used to derive maximum likelihood estimators of density parameters. In 1982, Burbea and Rao [45] proposed a unified approach to the derivation of metrics in *pdf*s spaces. They introduced the notion of ϕ-functional whose Hessian in a direction of the tangent space of the parameters space is taken as the metric. Following the pioneering work of Rao [253], and a theorem by Jensen (1976, private communication in [15]), Atkinson and Mitchell obtained closed-form expressions for the geodesic distances between elements of well-known families of distributions such as multivariate normal *pdf*s of fixed mean. We focus, in this chapter, on the geometrical properties of those particular distributions and make use of results stated in [272], [44], [48] and [117] to propose a novel framework for the statistical analysis of a set of multivariate normal *pdf*s.

In [235], the author generalized the notion of normal law to random samples of primitives belonging to an n-dimensional Riemannian manifold $\mathcal{M}$. In that framework, and under certain technical hypothesis, it is possible to define the mean, as proposed by Karcher [160] and Kendall [163], as well as the covariance matrix of a subset of $\mathcal{M}$. Using an information minimization approach, Pennec approximated the normal model by the usual Gaussian in the tangent space $T_{\overline{\mathrm{x}}}\mathcal{M}$ at the mean value $\overline{\mathrm{x}} \in \mathcal{M}$. We propose to combine the Riemannian characterization of the multivariate normal model, based on the Fisher information matrix, with this notion of generalized normal law on manifolds to study the statistical properties of diffusion tensor images. We also derive original, accurate and efficient computational tools to process these images.

We recall that these tensors can be estimated from the acquisition of diffusion weighted images (DWI) in several non-collinear sampling directions as well as a $T2$-weighted image. Diffusion tensors must be understood as parameters of normal distributions and, as such, perfectly fit the model we are about to develop.

Previous works related to DTI processing:

Preliminary results were presented in [187] and [186]. Other works related to the statistical analysis or filtering of DTI datasets have been carried out by Basser *et al.* [21], Pennec *et al.* [237, 238] and Fletcher and Joshi [116]. In [21], a symmetric positive-definite fourth-order tensor is used to encode the variability of a set of diffusion tensors but the geometry of the parameters space is not taken into account as in the other works. Barbaresco *et al.* used similar ideas in [18] for purposes related to the anisotropic regularization of normal or Gamma law parameters and of radar or Doppler data. Independently to our contribution, [116] first analyzed the space of symmetric, positive definite matrices from a Lie groups perspective and showed that this space does not form a vector space but can rather be regarded as the Riemannian symmetric space $GL^{+}(3)/SO(3)$. The authors used these ideas to develop efficient and elegant methods for computing statistics and modes of variation of diffusion tensor data. Interestingly, by extending the symmetry group to $SL(3)$, [195] studied a slightly different metric which turns out to exhibit an additional term in one direction. This extension was also recently formulated by [206] in terms of a weighted Fisher information matrix. They also generalized this approach to a wider class of elliptical densities. More recently, [238] also developed a nice and elegant computational framework for tensor processing with a particular emphasis on interpolation, regularization and restoration of noisy tensor fields. Fillard *et al.* [114] reported an interesting application of these tools to study the variability of the human brain.
Arsigny *et al.* [12, 14, 13] also proposed an original log-Euclidean metric that yields an efficient computational framework for tensor data processing. Fillard *et al.* [113] successfully relied on this metric to propose efficient regularization and estimation

algorithms for diffusion tensor images. Schwartzman [265] described a similar approach to derive statistical tests and false discovery rates for diffusion tensor images. Adding to these works where similar tools have been developed in order to study the statistical variability of diffusion tensor images, one can also refer to our original point of view based on the information geometry, recently developed in [186], and detailed in this chapter. It is thus very interesting to note that comparable results were obtained through different means while studying the statistical variability of diffusion tensor images.

Contributions of this chapter:

In this chapter, we derive and experiment with original methods to compute the mean and covariance matrix of a set of multivariate normal distributions. We also show how to compute and use the Ricci curvature tensor in order to accurately approximate a normal law on the manifold $\mathcal{M}$ of multivariate normal distributions. In the following chapters, we will successfully apply these numerical schemes to tackle, in an original manner, important processing tasks for diffusion tensor datasets.

Organization of this chapter:

Section 5.2 reviews necessary material related to the Riemannian geometry of the multivariate normal model. Section 5.3 introduces the theoretical basis and the numerical schemes that lead to an approximated normal law on the manifold described in section 5.2. In section 5.4, we first provide a simple algorithm to generate random normal distributions following a given normal law. We also demonstrate how to perform a correct spatial interpolation of diffusion tensor images. For these two applications, numerical experiments are conducted to illustrate their respective performance.

5.2 GEOMETRY OF THE MULTIVARIATE NORMAL MODEL

We hereafter review important notions that led to the metrization of probability density functions spaces and apply them to the multivariate normal model with fixed zero mean. This yields a characterization of the connection and curvature of that space as well as the definition of a distance between normal distributions under certain regularity conditions.

5.2.1 Metrization of the Space of Probability Density Functions

We start by a general characterization of the space of probability density functions, together with the characterization of its possible metrics. Let $L^1(\mathcal{X},\mu)$ denote the space of integrable μ-measurable real functions defined over the space $\mathcal{X} \subset \mathbb{R}^m$, e.g.:

$$L^1(\mathcal{X},\mu) = \{\mathscr{P} : \|\mathscr{P}\|_\mu = \int_{\mathcal{X}} |\mathscr{P}(r)| d\mu(r) < \infty\}$$

We are interested in the subset $\mathcal{P}$ of L^1_+ such that:

$$\mathcal{P} = \{\mathscr{P} \in L^1_+ : \|p\|_\mu = 1\}$$

$$\text{with} \quad L^1_+ = \{\mathscr{P} \in L^1(\mathcal{X}, \mu) : \mathscr{P}(r) \geq 0 \text{ for } \mu\text{-almost all } x \in \mathcal{X}\}$$

Let ϕ be a continuous real function on $I_\phi \subset \mathbb{R}^+$ and $\mathcal{F}_\phi(\mathcal{X}, \mu)$ the set of μ-measurable functions p defined over $\mathcal{X}$ and taking values in I_ϕ. The ϕ-entropy functional introduced in [45] is defined as:

$$H_\phi(\mathscr{P}) = -\int_{\mathcal{X}} \phi(\mathscr{P}(r)) d\mu(r) \quad \forall \mathscr{P} \in L^1_\phi = L^1_+ \cap \mathcal{F}_\phi$$

The second order differential of the entropy functional H_ϕ at $\mathscr{P}$ in the direction of $f \in L^1_\phi$ is given by:

$$d^2 H_\phi(\mathscr{P}; f) = -\int_{\mathcal{X}} \phi''(\mathscr{P}(r))(f(r))^2 d\mu(r)$$

We now introduce the set of parameters $\{\mathrm{x} : \mathrm{x} = (\mathrm{x}^1, ..., \mathrm{x}^n) \in \mathcal{O} \subset \mathbb{R}^n\}$. This set defines a manifold $\mathcal{M}$ in $\mathbb{R}^n$ and we consider the subset $\mathcal{F}_\mathcal{M}$ of $\mathcal{P}_\phi = \mathcal{P} \cap \mathcal{F}_\phi$:

$$\mathcal{F}_\mathcal{M} = \{\mathscr{P}(r|\mathrm{x}) \in \mathcal{P}_\phi : x \in \mathcal{X}, \mathrm{x} \in \mathcal{M}\}$$

$\mathcal{F}_\mathcal{M}$ is the family of probability density functions of the random variable $x \in \mathcal{X}$ parameterized by the n-dimensional vector x. We wish to quantify the second variation of the entropy functional H_ϕ in the direction $d\mathscr{P}(.|\mathrm{x})$ of the tangent space $T\mathcal{M}$, with:

$$d\mathscr{P}(.|\mathrm{x}) = \sum_{i=1}^{n} \frac{\partial \mathscr{P}(.|\mathrm{x})}{\partial \mathrm{x}^i} d\mathrm{x}^i$$

denoting the first order approximation of the difference between the densities associated with the parameters x and x + dx. Hence the second variation of H_ϕ at x writes:

$$d^2 H_\phi(\mathscr{P}(.|\mathrm{x}); d\mathscr{P}(.|\mathrm{x})) = -\int_{\mathcal{X}} \phi''(\mathscr{P}(r|\mathrm{x}))(d\mathscr{P}(r|\mathrm{x}))^2 d\mu(r)$$

Under the assumption that ϕ is convex in I_ϕ, we set:

$$ds^2_\phi(\mathrm{x}) = -d^2 H_\phi(\mathscr{P}(.|\mathrm{x}); d\mathscr{P}(.|\mathrm{x})) = \sum_{i,j=1}^{n} g^{(\phi)}_{ij}(\mathrm{x}) d\mathrm{x}^i d\mathrm{x}^j$$

with

$$g^{(\phi)}_{ij}(\mathrm{x}) = \int_{\mathcal{X}} \phi''(\mathscr{P}(r|\mathrm{x})) \frac{\partial \mathscr{P}(r|\mathrm{x})}{\partial \mathrm{x}^i} \frac{\partial \mathscr{P}(r|\mathrm{x})}{\partial \mathrm{x}^j} d\mu(r)$$

$g^{(\phi)}_{ij}$ defines a positive-definite form on the tangent space $T\mathcal{M}$ and thus gives a Riemannian metric on $\mathcal{M}$, an $n \times n$ matrix known as the ϕ-entropy metric.
The line element $ds = (ds^2_\phi(\mathrm{x}))^{1/2}$ is easily seen to be invariant under transformation of x. Consequently, $g^{(\phi)}_{ij}(\mathrm{x})$ is a second order covariant symmetric tensor.

Various possible choices for the entropy function ϕ have been proposed. We concentrate, in this chapter, on the Shannon entropy associated with $\phi(\mathscr{P}) = \mathscr{P}\log\mathscr{P}$, $\forall \mathscr{P} \in \mathcal{F}_{\mathcal{M}}$. Then,

$$H(\mathscr{P}) = -\int_{\mathcal{X}} \mathscr{P}(r|\mathbf{x})\log\mathscr{P}(r|\mathbf{x})d\mu(r)$$

and the components of the metric tensor g_{ij}, known as the Fisher information matrix, become:

$$g_{ij}(\mathbf{x}) = \int_{\mathcal{X}} \frac{\partial\log\mathscr{P}(r|\mathbf{x})}{\partial\mathbf{x}^i}\frac{\partial\log\mathscr{P}(r|\mathbf{x})}{\partial\mathbf{x}^j}\mathscr{P}(r|\mathbf{x})d\mu(r) = \mathbb{E}\left[\frac{\partial\log\mathscr{P}(r|\mathbf{x})}{\partial\mathbf{x}^i}\frac{\partial\log\mathscr{P}(r|\mathbf{x})}{\partial\mathbf{x}^j}\right]$$

It is interesting to note that, in this case, the squared line element $ds^2_\phi(\mathbf{x})$ coincides with the variance of the relative difference between $\mathscr{P}(.|\mathbf{x})$ and $\mathscr{P}(.|\mathbf{x}+d\mathbf{x})$. Indeed, we have:

$$\frac{d\mathscr{P}(.|\mathbf{x})}{\mathscr{P}(.|\mathbf{x})} = \frac{\mathscr{P}(.|\mathbf{x}+d\mathbf{x}) - \mathscr{P}(.|\mathbf{x})}{\mathscr{P}(.|\mathbf{x})} = \sum_{i=1}^{n}\frac{\partial\log\mathscr{P}(.|\mathbf{x})}{\partial\mathbf{x}^i}d\mathbf{x}^i$$

It follows that the expected value of the relative difference is zero since

$$\int_{\mathcal{X}}\left(\frac{\partial\log\mathscr{P}(r|\mathbf{x})}{\partial\mathbf{x}^i}d\mathbf{x}^i\right)\mathscr{P}(r|\mathbf{x})d\mu(r) = \frac{\partial}{\partial\mathbf{x}^i}\left(\int_{\mathcal{X}}\mathscr{P}(r|\mathbf{x})d\mu(r)\right)d\mathbf{x}^i = 0$$

but that its variance does not vanish and defines a positive-definite quadratic form $ds^2(\mathbf{x}) = \sum_{i,j=1}^{n} g_{ij}d\mathbf{x}^i d\mathbf{x}^j$, based on the Fisher information matrix, that can be used as a metric. We exploit that result in the next section to describe the differential geometrical properties of the space of multivariate normal distributions with fixed zero mean.

5.2.2 Geometrical Properties of the Multivariate Normal Model

Our ultimate goal being to define statistics between multivariate normal distributions and to apply it to diffusion tensor data, in other words 3-variate normal distributions with zero mean, we identify the space of parameters $\mathcal{M} = \{\mathbf{x} : \mathbf{x} = (\mathbf{x}^1, ..., \mathbf{x}^n) \in \mathcal{O} \subset \mathbb{R}^n\}$ with the manifold $S^+(m,\mathbb{R})$ and endow it with the information metric g_{ij}, $i,j = 1,...,n$ previously introduced.

$S^+(m,\mathbb{R})$ denotes the set of $m \times m$ ($m = 3$ in our case) real symmetric positive-definite matrices. Its elements are used to describe the covariance matrices of the zero mean normal distributions. Through the local coordinates chart φ that associates to each $\Sigma \in S^+(m,\mathbb{R})$ its components σ_{kl}, $k \le l$, $k,l = 1,...,m$, we see that $S^+(m,\mathbb{R})$ is isomorphic to $\mathbb{R}^n$ with $n = \frac{1}{2}m(m+1)$. Hence, from now on, the elements of parameter space $\mathcal{M}$ will simply be the components of the covariance matrices Σ and linearly accessed through φ, with $\mathbf{x}^i = \sigma_i = \sigma_{kl}$ with $i = 1,...,n$, $k \le l = 1,...,m$ and $\mathbf{x}^1 = \sigma_{11}, \mathbf{x}^2 = \sigma_{12}, ..., \mathbf{x}^6 = \sigma_{33}$.

We denote by $\frac{\partial}{\partial \mathrm{x}^i} = E_i$ the canonical basis of the tangent space $TS^+(m,\mathbb{R}) = S(m,\mathbb{R})$ (e.g. the space of vector fields). We equally denote by E_i^*, $i = 1,...,n$ the dual basis of the cotangent space $T^*S^+(m,\mathbb{R}) = S^*(m,\mathbb{R})$ (e.g. the space of differential forms). The tangent space $S(m,\mathbb{R})$ coincides with the space of $m \times m$ symmetric matrices and the basis is given by:

$$E_i = E_{kl} = \begin{cases} 1_{kk} & ,\ k = l \\ (1_{kl} + 1_{lk}) & ,\ k \neq l \end{cases} \qquad E_i^* = E_{kl}^* = \begin{cases} 1_{kk} & ,\ k = l \\ \frac{1}{2}(1_{kl} + 1_{lk}) & ,\ k \neq l \end{cases}$$

where 1_{kl} stands for the $m \times m$ matrix with 1 at row k and column l and 0 everywhere else. For the clarity of expressions, we will drop the references to m and $\mathbb{R}$ in $S^+(m,\mathbb{R})$ when no ambiguity is possible.

As detailed by the authors of [272], [44], [107], [48] and [117], we can characterize S^+ as a Riemannian manifold for which closed form expressions are available for the metric g, the Christoffel symbols and the associated affine connection, the Riemann curvature tensor, the solution of the geodesic equations as well as for the geodesic distance (also known as Rao's distance). Those constitute all the fundamental mathematical tools that we need to derive algorithms for the mean and covariance matrix of elements of S^+ and express a generalized normal law on this manifold.

Metric Tensor, Affine Connection and Curvature Tensors

The proofs of the theorems stated in this section are available in [271].

The metric tensor

The metric tensor for S^+, derived from the Fisher information matrix presented in section 5.2.1, is given by the following theorem:

Theorem 5.2.2.1. *The Riemannian metric for the space $S^+(m,\mathbb{R})$ of multivariate normal distributions with zero mean is given, $\forall \Sigma \in S^+(m,\mathbb{R})$ by the twice covariant tensor:*

$$g_{ij} = g(E_i, E_j) = \langle E_i, E_j \rangle_\Sigma = \frac{1}{2}\mathrm{tr}(\Sigma^{-1} E_i \Sigma^{-1} E_j) \quad i,j = 1,...,n \tag{5.1}$$

In practice, this means that for any vectors $A, B \in S$, their inner product relative to Σ is $\langle A, B \rangle_\Sigma = \frac{1}{2}\mathrm{tr}(\Sigma^{-1} A \Sigma^{-1} B)$. In particular, the distance between two infinitesimally close elements Σ and $\Sigma + d\Sigma$ of S^+, with respect to Σ, is:

$$\|d\Sigma\|_\Sigma = \sqrt{\frac{1}{2}\mathrm{tr}((\Sigma^{-1} d\Sigma)^2)}$$

It is very informative, at this stage, to look at the well-known Kullback-Leibler divergence $\mathcal{D}_{kl}$, or relative entropy, that we also use for the segmentation task that will be addressed in chapter 8. In [185], we used it as a measure of dissimilarity between probability density functions, as often done in the information theory community. However, it can be shown by a second order Taylor expansion of the relative entropy between two infinitesimally close *pdf*s parameterized by Σ and $\Sigma + d\Sigma$ (assuming that $\mathcal{D}_{kl}$ is twice differentiable around $\mathscr{P}(r|\Sigma)$) that:

$$\mathcal{D}_{kl}(\mathscr{P}(r|\Sigma), \mathscr{P}(r|\Sigma + d\Sigma)) = \int_{\mathcal{X}} \mathscr{P}(r|\Sigma) \log \frac{\mathscr{P}(r|\Sigma)}{\mathscr{P}(r|\Sigma + d\Sigma)} d\mu(r)$$

$$= \frac{1}{2} \int_{\mathcal{X}} \left(\frac{1}{p^2(r|\Sigma)} \frac{\partial \mathscr{P}(r|\Sigma)}{\partial \sigma_i} \frac{\partial \mathscr{P}(r|\Sigma)}{\partial \sigma_j} - \frac{1}{\mathscr{P}(r|\Sigma)} \frac{\partial^2 \mathscr{P}(r|\Sigma)}{\partial \sigma_i \partial \sigma_j} \right) \mathscr{P}(r|\Sigma) d\sigma_i d\sigma_j d\mu(r)$$

which can be shown to reduce to:

$$\mathcal{D}_{kl}(\mathscr{P}(r|\Sigma), \mathscr{P}(r|\Sigma + d\Sigma)) = \frac{1}{2} \mathbb{E} \left[\frac{\partial \log \mathscr{P}(r|\Sigma)}{\partial \sigma_i} \frac{\partial \log \mathscr{P}(r|\Sigma)}{\partial \sigma_j} \right] d\sigma_i d\sigma_j$$

if the partial derivatives with respect to σ_i and σ_j commute with the integral.
As a consequence, the relative entropy simply equals half of the squared line element ds^2 and it coincides with the geodesic distance for infinitesimal distances. Computing it between distant *pdf*s would yield a result very different from the geodesic distance.

The choice of the affine connection

We now would like to define the two fundamental tensors in Riemannian geometry for the manifold S^+, known as the Riemann and Ricci curvature tensor and denoted, like in section 4.2.4 by R and $\mathcal{R}$. The latter will play an important role in the expression of the generalized normal law on S^+.
But before being able to define these elements, we have to choose a Riemannian connection ∇. We recall that ∇ allows us to map any tangent space $T_{\Sigma_1} S^+$ at Σ_1 to the tangent space $T_{\Sigma_2} S^+$ at Σ_2. The canonical affine connection on a Riemannian manifold is known as the Levi-Civita connection (or covariant derivative, see section 4.2.3). An important property of this connection is that it is compatible with the metric.
Using the local coordinates, the Christoffel symbols can also be expressed in terms of the elements of the canonical and dual basis $\{E_i\}_{i=1,\ldots,n}$ and $\{E_i^*\}_{i=1,\ldots,n}$.

$$\Gamma(E_i, E_j; E_k^*) = E_k^* . (\nabla^F_{E_i} E_j) \tag{5.2}$$

By the fact that (see lemma 2.3 in [271]):

$$\frac{\partial g(E_i, E_j)}{\partial \sigma_k} = -\frac{1}{2} \text{tr}(\Sigma^{-1} E_k \Sigma^{-1} E_i \Sigma^{-1} E_j) - \frac{1}{2} \text{tr}(\Sigma^{-1} E_i \Sigma^{-1} E_k \Sigma^{-1} E_j)$$

the following result can be proved from equation 4.4:

$$\Gamma(E_i, E_j; E_k^*) = -\frac{1}{2} \text{tr}(E_i \Sigma^{-1} E_j E_k^*) - \frac{1}{2} \text{tr}(E_j \Sigma^{-1} E_i E_k^*) \tag{5.3}$$

It is then possible to use this result in order to derive the expression of the unique affine connection (Levi-Civita) ∇^F associated with the Fisher information metric from equation 5.2. However, other connections have been proposed and we still need to make a choice since this will greatly influence the curvature properties of the manifold. Amari [9] has indeed introduced a one-parameter family of affine connections, known as the α-connections, in order to better represent the intrinsic properties of the family of probability distributions. The α-connections are defined in the following manner:

$$\langle \nabla^{\alpha}_{E_i} E_j, E_k \rangle_{\Sigma} = \langle \nabla^{F}_{E_i} E_j, E_k \rangle_{\Sigma} + \alpha T(E_i, E_j, E_k)$$

where we remind that $\langle ., . \rangle_{\Sigma}$ denotes the inner product and the third-order symmetric tensor T_{ijk} is defined as

$$T_{ijk} = \mathbb{E}\left[\frac{\partial \log \mathscr{P}(r|\mathbf{x})}{\partial \sigma_i} \frac{\partial \log \mathscr{P}(r|\mathbf{x})}{\partial \sigma_j} \frac{\partial \log \mathscr{P}(r|\mathbf{x})}{\partial \sigma_k}\right]$$

Obviously, we see that the 0-connection boils down to the Levi-Civita connection. We recall that it is the only one to be compatible with the metric, in other words, the only one by which the parallel transport of a vector does not affect its length. The α-connections are not compatible with the metric for $\alpha \neq 0$.

Moreover, it was stated in [44] that, for any exponential family (a special type of distributions in which the multivariate normal model can be recast), the α-Riemann curvature tensor writes:

$$R^{\alpha}_{ijkl} = (1 - \alpha^2) R^{F}_{ijkl}$$

thus giving an α-curvature to the multivariate normal model distinct from the curvature induced by the Levi-Civita connection for all $\alpha \neq 0$. Because of their non-compatibility with the metric and their induced 0-curvature, the ± 1-connections (respectively known as Efron and David connections) do not seem to be good candidates for the following derivations of statistics on the manifold of multivariate normal distributions. We will indeed require our space to exhibit a non-positive sectional curvature in order to ensure the existence and uniqueness of the Riemannian barycenter. For this reason, we will work with the classical Levi-Civita connection in the remaining developments. We have to notice, however, that α-connections with $\alpha \neq \pm 1$ will have to be investigated.

The curvature tensors

The Riemann curvature tensor for S^+ (see definition in section 4.2.4), derived from the Fisher information metric presented in section 5.2.1, is given by the following theorem (see [271]):

Theorem 5.2.2.2. *The Riemann curvature tensor derived from the Fisher information metric and the classical Levi-Civita affine connection in $S^+(m, \mathbb{R})$ is given by:*

$$\begin{aligned} R^{F}_{ijkl} = R^{F}(E_i, E_j, E_k, E_l) &= \frac{1}{4}\mathbf{tr}(E_j \Sigma^{-1} E_i \Sigma^{-1} E_k \Sigma^{-1} E_l \Sigma^{-1}) \\ &- \frac{1}{4}\mathbf{tr}(E_i \Sigma^{-1} E_j \Sigma^{-1} E_k \Sigma^{-1} E_l \Sigma^{-1}) \end{aligned}$$

where E_i, E_j, E_k and E_l denote the elements of the canonical basis of vector fields and $\Sigma \in S^+(m, \mathbb{R})$

The important point is that we can now compute the sectional curvature κ of the manifold S^+ (see details in section 4.2.4) and verify that it is actually non-constant and, more importantly, non-positive.

It can be shown that, if we denote by $\rho_{rs}^2 = \frac{\sigma_{rs}^2}{\sigma_{rr}\sigma_{ss}}$ the correlation coefficient between the components σ_{rs}, $r \leq s$, $r, s = 1, ..., m$ of the covariance matrices Σ, the sectional curvature at Σ is given, for $r \neq s$, by

1. $\kappa(\mathcal{E}, \Sigma) = -\frac{\rho_{rs}^2}{1+\rho_{rs}^2}$ if $\mathcal{E} = \text{span}(E_{rr}, E_{ss})$
2. $\kappa(\mathcal{E}, \Sigma) = -\frac{1}{2}$ if $\mathcal{E} = \text{span}(E_{rr}, E_{rs})$

which indeed depends on Σ and is non-positive.

Finally, we recall that the Ricci curvature tensor $\mathcal{R}$ is defined as the contraction of the Riemann curvature tensor (see details in section 4.2.4), which can be obtained through symbolic computations with a software like Maple$^{\text{TM}}$ by Maplesoft.

Summarizing everything, we have the expression for R and g. A quite interesting point is that, by comparison of the Ricci tensor with the metric tensor, we can also deduce that the space of zero mean multivariate normal distributions is, in general, not an Einstein manifold. It is indeed a space of non-constant non-positive sectional curvature for which there does not exist a constant L such that

$$\mathcal{R}_{ij} = L g_{ij}$$

Now that we have defined the metric, connection and curvature in S^+, we can characterize the geodesics of that manifold.

Geodesics and Geodesic Distance between Normal Distributions

The geodesic distance $\mathcal{D}$ induced by the Riemannian metric g, derived from the Fisher information matrix, was investigated for some parametric distributions in [15], [45], [221], [44], [107] and references therein. More recently, Calvo and Oller derived an explicit solution of the geodesic equations for the general multivariate normal model in [48].

With our current notations, we recall that, if $\Sigma : t \mapsto \Sigma(t) \in \mathcal{M}$, $\forall t \in [t_1, t_2] \subset \mathbb{R}$ denotes a curve segment in $\mathcal{M}$ between two parameterized distributions $\mathscr{P}(.|\Sigma_1)$ and $\mathscr{P}(.|\Sigma_2)$, its length is expressed as:

$$\begin{aligned} \mathcal{L}_\Sigma(\mathscr{P}(.|\Sigma_1), \mathscr{P}(.|\Sigma_2)) &= \int_{t1}^{t2} \left(\langle \dot{\Sigma}(t), \dot{\Sigma}(t) \rangle_{\Sigma(t)} \right)^{1/2} dt \\ &= \int_{t1}^{t2} \left(\sum_{i,j=1}^{n} g_{ij}(\Sigma(t)) \frac{d\sigma_i(t)}{dt} \frac{d\sigma_j(t)}{dt} \right)^{1/2} dt \end{aligned}$$

Solving the Euler-Lagrange equations 4.2 and evaluating the geodesic distance $\mathcal{D}$ constitute, in general, a difficult task. However, for the multivariate normal model with zero mean, it can be proved that those equations reduce to:

$$\frac{d^2\Sigma(t)}{dt^2} - \frac{d\Sigma(t)}{dt}\Sigma(t)^{-1}\frac{d\Sigma(t)}{dt} = 0 \tag{5.4}$$

Proof. This is straightforward from the use of equation 5.3 in the general geodesic equation 4.2. □

It is interesting to note that the closed-form expression for the geodesic curves $\Sigma(t)$, $t \in [t_1, t_2] \subset \mathbb{R}$ and the geodesic distance have been independently derived by several authors:

In [271] and [211], the geodesics equation were obtained respectively by solving equation 5.4 and by identifying $S^+(m, \mathbb{R})$ with the quotient space $GL^+(m, \mathbb{R})/SO(m, \mathbb{R})$. In this last case, it is easy to recast the expression of the geodesics from any point $\Sigma(0)$ of the space into the simpler configuration $\Sigma(0) = \mathbf{I}$, because of the invariance by congruence transformation of $GL^+(m, \mathbb{R})$.

Burbea [44] addressed this problem by using the properties of the group of automorphisms of S^+ onto itself. Calvo and Oller, in [48], proposed a more general solution on the basis of the information geometry described in [272], [44] and [107]. They derived an explicit expression of the geodesics for the multivariate normal model with non-constant mean vector.

Regarding the geodesic distance, it seems to have been derived for the first time in 1976 by Jensen (private communication in [15]) for multivariate distributions of fixed mean, and then again in the independent work [271]. Another distinct work by Förstner and Moonen [117] proposed the same distance measure with a similar point of view as the one used in [211]. However, the geodesic equations were not solved in [117]. Other references related to the information geometry approach can be found in [44] and [47]. Motivated by medical image processing tasks, Fletcher and Joshi [116], Lenglet *et al.* [187, 186] and Pennec *et al.* [238] have recently used these results to derive statistical and filtering tools on tensor fields.

As stated for example in [211], the geodesic starting from $\Sigma(t_1) \in S^+$ in the direction $\dot{\Sigma}(t_1) = \Sigma(t_1)^{1/2} X \Sigma(t_1)^{1/2}$ with $\dot{\Sigma}(t_1), X \in S = TS^+$ is given by:

$$\Sigma(t) = \Sigma(t_1)^{1/2} \exp\left((t - t_1)X\right) \Sigma(t_1)^{1/2} \in S^+, \forall t \in [t_1, t_2] \tag{5.5}$$

where the matrix square root is well-defined since it always applies to symmetric positive-definite matrices.

We recall that the geodesic distance $\mathcal{D}$ between any two elements Σ_1 and Σ_2 of S^+ is the length of the minimizing geodesic between Σ_1 and Σ_2. It is given by the following theorem, whose original proof is available in an appendix of [15] but different versions can also be found in [271] and [117].

Theorem 5.2.2.3. *(S.T. Jensen, 1976)*
Consider the family of multivariate normal distributions with common mean vector but different covariance matrices. The geodesic distance between two members of the family with covariance matrices Σ_1 and Σ_2 is given by

$$\mathcal{D}(\Sigma_1,\Sigma_2) = \sqrt{\frac{1}{2}\mathbf{tr}(\log^2(\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2}))} = \sqrt{\frac{1}{2}\sum_{i=1}^{m}\log^2(\eta_i)}$$

where η_i denote the m eigenvalues of the matrix $\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2} \in S^+$.

Properties of $\mathcal{D}$:
$\mathcal{D}$ is indeed a distance on S^+ and exhibits nice properties that we hereafter summarize (see [117] for details and proofs):

[P1] Positivity: $\mathcal{D}(\Sigma_1,\Sigma_2) \geq 0,\ \mathcal{D}(\Sigma_1,\Sigma_2) = 0 \Leftrightarrow \Sigma_1 = \Sigma_2$

[P2] Symmetry: $\mathcal{D}(\Sigma_1,\Sigma_2) = \mathcal{D}(\Sigma_2,\Sigma_1)$

[P3] Triangle inequality: $\mathcal{D}(\Sigma_1,\Sigma_3) \leq \mathcal{D}(\Sigma_1,\Sigma_2) + \mathcal{D}(\Sigma_2,\Sigma_3)$

[P4] Invariance under congruence transformations: $\forall P \in GL^+(m,\mathbb{R})$
$\mathcal{D}(\Sigma_1,\Sigma_2) = \mathcal{D}(P\Sigma_1P^T, P\Sigma_2P^T)$

[P5] Invariance under inversion: $\mathcal{D}(\Sigma_1,\Sigma_2) = \mathcal{D}(\Sigma_1^{-1},\Sigma_2^{-1})$

We must note however that no complete proof of [P3] was given in the work by Förstner and Moonen. Now that we have setup all the mathematical tools that we need, we define important statistical parameters on the manifold S^+ and provide efficient numerical schemes to compute them. This will lead to the definition of the generalized normal law on S^+.

5.3 STATISTICS ON NORMAL DISTRIBUTIONS

5.3.1 Intrinsic Mean

Definition

We propose a new gradient descent algorithm for the computation of the intrinsic mean distribution of a set of multivariate normal distributions with zero mean vector. It relies on the classical definition of the Riemannian center of mass and uses the geodesics equation to derive a manifold-constrained numerical integrator and thus ensures that each step forward of the gradient descent stays within the space S^+. As we will show in the numerical experiments, this method is very efficient and usually converges in just a few iterations.
We seek to estimate the empirical mean as proposed by Fréchet [119], Karcher [160], Pennec [235] or Moakher [210].

Definition 5.3.1.1. *(Empirical Riemannian mean)*
The normal distribution $\mathscr{P}(.|\overline{\Sigma})$ with zero mean vector, parameterized by $\overline{\Sigma} \in S^+(m,\mathbb{R})$, and defined as the empirical mean of N distributions $\mathscr{P}(.|\Sigma_k),\ k = 1,...,N$, achieves a local minimum of the objective function $\lambda^2 : S^+(m,\mathbb{R}) \to \mathbb{R}^+$ known as the empirical variance and defined as:

$$\lambda^2(\Sigma_1,...,\Sigma_N) = \frac{1}{N}\sum_{k=1}^{N}\mathcal{D}^2(\Sigma_k,\overline{\Sigma}) = \mathbb{E}[\mathcal{D}^2(\Sigma_k,\overline{\Sigma})] \tag{5.6}$$

Karcher proved in [160] that such a mean, also known as the Riemannian barycenter, exists and is unique for manifolds of non-positive sectional curvature. This was shown to be the case for S^+ in the previous section so that we can be assured to always find a solution, given that the numerical gradient descent is carefully designed and does not get stuck in some local minimum.

In order to derive this gradient descent algorithm, we rely on the following remarks (see [64] and references therein for more details): We want to derive a flow evolving an initial guess $\overline{\Sigma}(0)$ toward the mean of a set of N elements of S^+. If we denote by $\overline{\Sigma}(s),\ s \in [0,\infty)$ the family of solutions for:

$$\frac{\partial\overline{\Sigma}(s)}{\partial s} = \mathcal{V}(\overline{\Sigma}(s))$$

where $\mathcal{V}$ denotes the velocity (e.g. the tangent vector) driving the evolution, we have the equivalence:

$$\begin{aligned} & \overline{\Sigma}(s) && \in S^+(m,\mathbb{R}),\ \forall s > 0 \\ \Leftrightarrow\ & \overline{\Sigma}(0) && \in S^+(m,\mathbb{R}) \text{ and } \mathcal{V}(\overline{\Sigma}(s)) \in T_{\overline{\Sigma}(s)}S^+(m,\mathbb{R}) = S(m,\mathbb{R}),\ \forall s > 0 \end{aligned}$$

In other words, we are guaranteed to stay in S^+ as long as $\overline{\Sigma}(0)$ does and that the velocity belongs to the space of real symmetric matrices. We shall see that the artificial time-step ds will be directly related to the geodesic parameter t through the definition of the exponential map.

We identify the velocity $\mathcal{V}$ with the opposite of the gradient of the objective function $\lambda^2(\Sigma_1,...,\Sigma_N)$. This was shown to be (see [211]):

$$\nabla\lambda^2(\Sigma_1,...,\Sigma_N) = \frac{\overline{\Sigma}(s)}{N}\sum_{k=1}^{N}\log(\Sigma_k^{-1}\overline{\Sigma}(s)) \tag{5.7}$$

Hence the evolution:

$$\frac{\partial\overline{\Sigma}(s)}{\partial s} = -\frac{\overline{\Sigma}(s)}{N}\sum_{k=1}^{N}\log(\Sigma_k^{-1}\overline{\Sigma}(s)) \tag{5.8}$$

Numerical Implementation

As mentioned in [64], the corresponding numerical implementation has to be dealt with carefully and we have to build a step-forward operator $\mathbf{K}_{ds}$ such that the discrete flow:

$$\Sigma_{l+1} = \mathbf{K}_{ds}(\Sigma_l),\ \Sigma(0) \in S^+$$

provides an intrinsic, or ”consistent”, approximation of the evolution equation 5.8 (l denotes the discrete version of the continuous family parameter s and we have dropped the overline for the clarity of the expressions). As we have previously pointed out, a closed-form expression for the geodesics of S^+ does exist. Hence, all we need to do is to make sure that we perform our gradient descent along the geodesics of the manifold. If we were to simply move in the direction given by $-ds\nabla\lambda^2$, we would have to check, at every single step, that the current estimate of the mean lies within S^+ and reproject it if needed. Instead, we have the following proposition:

Proposition 5.3.1.1. *For any $\Sigma(s) \in S^+(m, \mathbb{R})$, $s \in [0, \infty)$ and any tangent vector $\mathcal{V} = -\nabla\lambda^2 \in S(m, \mathbb{R})$, an intrinsic (or consistent) approximation of the flow 5.8 can be obtained by using the step-forward operator:*

$$\mathbf{K}_{ds}(\Sigma_l) = \Sigma_l^{1/2} \exp\left(-ds\Sigma_l^{-1/2}\nabla\lambda^2\Sigma_l^{-1/2}\right)\Sigma_l^{1/2} \tag{5.9}$$

Proof. The geodesic starting from $\Sigma(s)$ and pointing in the direction of interest $\mathcal{V} = \Sigma(s)^{1/2}X\Sigma(s)^{1/2}$, $X \in S(m, \mathbb{R})$ is given by:

$$\Sigma(s + ds) = \Sigma(s)^{1/2} \exp(dsX)\Sigma(s)^{1/2} \quad \forall ds \in [0, 1]$$

If $\mathcal{V}$ is now identified with the opposite of the gradient of any objective functional λ^2 (which is in $S(m, \mathbb{R})$), we obtain:

$$\begin{aligned} X &= -\Sigma(s)^{-1/2}\nabla\lambda^2\Sigma(s)^{-1/2} \quad \text{and} \\ \Sigma(s + ds) &= \Sigma(s)^{1/2} \exp\left(-ds\Sigma(s)^{-1/2}\nabla\lambda^2\Sigma(s)^{-1/2}\right)\Sigma(s)^{1/2} \end{aligned}$$

Identifying $\Sigma(s)$ with the current estimate of the solution of the gradient descent Σ_l yields the result. □

It is natural to ask how this numerical scheme compares to the gradient descent associated with the following step-forward operator:

$$\mathbf{K}_{ds}(\Sigma_l) = \Sigma_l - ds\nabla\lambda^2 \tag{5.10}$$

We have the following proposition:

Proposition 5.3.1.2. *The extrinsic step-forward operator 5.10 is a first order approximation of the operator of Proposition 5.3.1.1.*

Proof. Recalling that the matrix exponential is defined by the following power series:

$$\exp(A) = \sum_{n=0}^{\infty} \frac{A^n}{n!} = \mathbf{I} + A + \frac{AA}{2} + \frac{AAA}{6} + ... \quad \forall A \in S(m, \mathbb{R})$$

A first order expansion of 5.9 yields:

$$\begin{aligned} \mathbf{K}_{ds}(\Sigma_l) &= \Sigma_l^{1/2}(\mathbf{I} - ds\Sigma_l^{-1/2}\nabla\lambda^2\Sigma_l^{-1/2})\Sigma_l^{1/2} \\ &= \Sigma_l - ds\nabla\lambda^2 \end{aligned}$$

$\square$

The exponential map is defined from $[0,1]$ onto S^+. The optimal time step ds is then equal to 1 and we have checked that the gradient descent is indeed stable for any value of ds between 0 and 1. The optimal step-forward operator is then expressed as follows:

$$\mathbf{K}(\Sigma_l) = \Sigma_l^{1/2} \exp(-\Sigma_l^{-1/2}\nabla\lambda^2\Sigma_l^{-1/2})\Sigma_l^{1/2}$$

This will be illustrated in the section dedicated to the numerical experiments.

Conclusion

We now come back to the derivation of a numerical algorithm to estimate the Riemannian barycenter $\overline{\Sigma}$ of a set of parameterized multivariate normal distributions with zero mean vector. We make use of the explicit expression of the gradient $\nabla\lambda^2$ given in 5.7. This yields the simple step-forward operator:

$$\mathbf{K}_{ds}(\overline{\Sigma}_l) = \overline{\Sigma}_l^{1/2} \exp\left(-ds\,\frac{1}{N}\overline{\Sigma}_l^{1/2}\left(\sum_{k=1}^{N} \log(\Sigma_k^{-1}\overline{\Sigma}_l)\right)\overline{\Sigma}_l^{-1/2}\right)\overline{\Sigma}_l^{1/2} \tag{5.11}$$

whose associated flow converges toward the barycenter for any initial guess $\overline{\Sigma}_0$.
At this stage, we can verify that our result is consistent with the gradient descent proposed in [238]. Indeed, using the fact that $\forall A, B \in GL(m,\mathbb{R})$, $\log\left(A^{-1}BA\right) = A^{-1}\log(B)A$ and $\log\left(A^{-1}\right) = -\log(A)$, we have:

$$\begin{aligned} & \overline{\Sigma}_l^{1/2} \exp\left(-ds\,\tfrac{1}{N}\overline{\Sigma}_l^{1/2}\left(\textstyle\sum_{k=1}^{N}\log(\Sigma_k^{-1}\overline{\Sigma}_l)\right)\overline{\Sigma}_l^{-1/2}\right)\overline{\Sigma}_l^{1/2} \\ = \;& \overline{\Sigma}_l^{1/2} \exp\left(-ds\,\tfrac{1}{N}\left(\textstyle\sum_{k=1}^{N}\log\left(\overline{\Sigma}_l^{1/2}\Sigma_k^{-1}\overline{\Sigma}_l\overline{\Sigma}_l^{-1/2}\right)\right)\right)\overline{\Sigma}_l^{1/2} \\ = \;& \overline{\Sigma}_l^{1/2} \exp\left(ds\,\tfrac{1}{N}\left(\textstyle\sum_{k=1}^{N}\log\left(\overline{\Sigma}_l^{-1/2}\Sigma_k\overline{\Sigma}_l^{-1/2}\right)\right)\right)\overline{\Sigma}_l^{1/2} \end{aligned}$$

which coincides with the result in the above mentioned work. We describe this procedure in the Algorithm 1.

Algorithm 1 Riemannian estimation of the mean tensor

```
Require: $\{\Sigma_i\} \in S^+(3)$, $i = 1, ...N$ and $nit$, the number of iterations
Ensure: $\overline{\Sigma}$, the mean tensor
$M \leftarrow \mathbb{I}$
for $k = 1$ to $nit$ do
    $V \leftarrow \mathbb{O}$ {$3 \times 3$ zero matrix}
    for $i = 1$ to $N$ do
        $V \leftarrow \log\left(\Sigma_i^{-1} M\right)$
    $V = \frac{1}{N} M V$
    $M \leftarrow M^{1/2} \exp\left(-M^{-1/2} V M^{-1/2}\right) M^{1/2}$
Return $M$
```

5.3.2 Intrinsic Covariance Matrix and Principal Modes

We propose a new algorithm for the computation of the intrinsic empirical covariance matrix of a set of N multivariate normal distributions with zero mean vector. We follow [62], and references therein, where this problem was addressed in the infinite dimensional case of a set of planar closed curves.

The works of [116] and [238] are closely related. They derived the expression of the Riemannian logarithmic map by considering S^+ as a homogeneous space and using the invariance property of the metric under congruence transformation. In [187], we addressed this problem with an information geometric approach. As it will become clear in the following, we use the gradient of the squared geodesic distance as the initial velocity of the unique geodesic connecting two elements of S^+. The estimation of the covariance matrix becomes then computationally easier.

Our objective is to derive an intrinsic numerical scheme for the estimation of the covariance matrix Λ relative to the empirical mean $\overline{\Sigma}$ of N normal distributions and using the explicit solution of the geodesic distance. As we consider the 6-dimensional manifold of parameterized normal distributions with zero mean vector, we will naturally end up with $\Lambda \in S^+(6, \mathbb{R})$ acting on the space $S(6, \mathbb{R})$.

We associate to each of the N normal distributions $\mathscr{P}(.|\Sigma_k)$ the unique tangent vector $\beta_k \in S(m, \mathbb{R})$ (seen as an element of $\mathbb{R}^n$) such that the empirical mean $\overline{\Sigma}$ is mapped onto Σ_k by the exponential map $\exp_{\overline{\Sigma}}(\beta_k) = \overline{\Sigma}^{1/2} \exp\left(\overline{\Sigma}^{-1/2} \beta_k \overline{\Sigma}^{-1/2}\right) \overline{\Sigma}^{1/2}$ (see Figure 5.1). We then have the following definition:

Definition 5.3.2.1. *Given N elements of $S^+(m, \mathbb{R})$ and a mean value $\overline{\Sigma}$, the empirical covariance matrix relative to $\overline{\Sigma}$ is defined as:*

$$\Lambda_{\overline{\Sigma}} = \frac{1}{N} \sum_{k=1}^{N} \beta_k \beta_k^T$$

We identify the β_k with the opposite of the gradient of the squared geodesic dis-

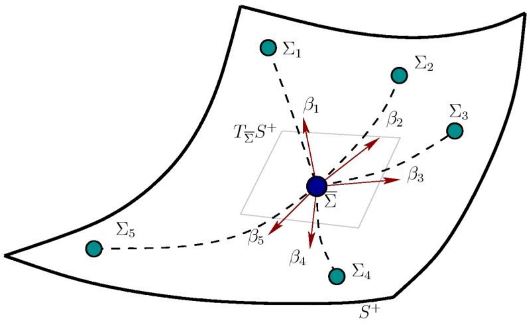


Figure 5.1: Depiction of the velocity field β_k at the empirical mean $\overline{\Sigma}$ (The black dotted lines represent the geodesics between each pair $(\Sigma_k, \overline{\Sigma})$ and the red arrows are the associated initial tangent vectors β_k)

tance function $\nabla\mathcal{D}^2(\Sigma_k, \overline{\Sigma})$. As detailed in [211], we have:

$$\nabla\mathcal{D}^2(\Sigma_k, \overline{\Sigma}) = \overline{\Sigma}\log(\Sigma_k^{-1}\overline{\Sigma})$$

Once we have computed Λ, it is fairly easy and instructive, in order to further understand the structure of our subset of normal distributions, to compute the eigenvalues and eigenvectors of the covariance matrix. Fletcher and Joshi [115] generalized the notion of Principal Component Analysis to Lie groups by seeking geodesic submanifolds, by analogy with the lower-dimensional linear subspaces of PCA, that maximize the projected variance of the data. They successfully applied this method to diffusion tensor datasets in [116]. This actually amounts to characterizing the tangent space at the mean element $\overline{\Sigma}$ through classical PCA in order to construct an orthogonal basis of tangent vectors v_k, $k = 1, ..., d \leq n$ that can be used to generate l-dimensional subspaces $V_l = \text{span}(v_1, ..., v_l)$, $l \leq n$ that maximize the projected variance. The v_k are defined as the set of eigenvectors of the covariance matrix Λ. However, since we have the geodesics equation in S^+, we have a closed-form expression to generate elements of any geodesic submanifold H_k defined as the exponential mapping of the V_k.

Indeed, we can define any linear combination of the v_k, $v = \sum_{k=1}^{d} \alpha_k v_k \in S(m, \mathbb{R})$ and then compute the unique element C of $S^+(m, \mathbb{R})$ reached by following the geodesic emanating from $\overline{\Sigma}$ in the direction v and computed as:

$$C = \overline{\Sigma}^{1/2} \exp\left(\overline{\Sigma}^{-1/2}\left(\sum_{k=1}^{d} \alpha_k v_k\right)\overline{\Sigma}^{-1/2}\right)\overline{\Sigma}^{1/2} \tag{5.12}$$

It is straightforward to apply our definition of the covariance matrix Λ to compute those principal modes. In the next section, we will see how the covariance matrix must be modified, because of the manifold curvature, to yield the concentration matrix and an approximated continuous Gaussian *pdf* on S^+.

5.3.3 A Generalized Normal Law on $S^+(3)$

Our last contribution uses the various quantities derived up to that point and fully exploits the information they provide in order to derive the expression of a normal law on S^+. We proceed by using the appropriate quantities in the generalization of the normal distribution to Riemannian manifolds proposed in [235] for sufficiently concentrated probability density functions, e.g. for small covariance matrices. The main idea of this work was to correctly deal with the curvature of the manifold since it modifies the definition of the concentration matrix γ. In Euclidean space, this matrix is simply the inverse of the covariance matrix Λ. On a Riemannian manifold, we must incorporate the contribution of the Ricci curvature tensor $\mathcal{R}$. We have the following theorem :

Theorem 5.3.3.1. *The generalized normal distribution in $S^+(m, \mathbb{R})$ for a covariance matrix Λ of small variance $\sigma^2 = \mathrm{tr}(\Lambda)$ is of the form:*

$$p(\Sigma|\overline{\Sigma}, \Lambda) = \frac{1 + O(\sigma^3) + \epsilon(\frac{\sigma}{\xi})}{\sqrt{(2\pi)^{m(m+1)/2}|\Lambda|}} \exp \frac{-\beta^T \gamma \beta}{2} \quad \forall \Sigma \in S^+(m, \mathbb{R})$$

$\overline{\Sigma}$ is computed as in section 5.3.1.
β is defined as $-\nabla \mathcal{D}^2(\Sigma, \overline{\Sigma}) = -\overline{\Sigma} \log(\Sigma^{-1}\overline{\Sigma})$ and expressed in vector form.
The concentration matrix is $\gamma = \Lambda^{-1} - \mathcal{R}/3 + O(\sigma) + \epsilon(\frac{\sigma}{\xi})$, with Λ defined as in section 5.3.2 and $\mathcal{R}$ as in section 5.2.2 (both are computed at location $\overline{\Sigma}$).
ξ is the injectivity radius at $\overline{\Sigma}$ and ϵ is such that $\lim_{0^+} x^{-\omega}\epsilon(x) = 0 \; \forall \omega \in \mathbb{R}^+$.

It should be clear now that, in section 5.3.2, the contribution of the Ricci curvature tensor was not taken into account and the concentration matrix was identified with the inverse of the empirical covariance matrix Λ. When approximating a continuous probability density function such as $p(\Sigma|\overline{\Sigma}, \Lambda)$ in the theorem above, equation 4.5 can be used to compute the tensor $\mathcal{R}$ and, hence, to modify the covariance matrix in order to reflect the local curvature properties of the manifold.

5.4 Numerical Experiments

We now would like to propose two interesting illustrations of the theory we just developed. The first one is a direct application of the notions presented in section 5.3 and shows how to consistently generate a set of random covariance matrices so that they follow our generalized normal law. The second illustration deals with the issue of the interpolation of sparsely sampled S^+-valued data.

5.4.1 Generation of Normally Distributed Random Tensors

Background

Figure 5.1 is useful to understand the situation and the problem we have to solve. It is basically the opposite of the covariance matrix estimation problem. Let us consider that we lie in $S^+(m,\mathbb{R})$ at $\overline{\Sigma}$ and that we want to *shoot* along the geodesics of the space to reach the random, normally distributed, symmetric positive-definite matrices $\Sigma_k,\ k=1,...,N$. Knowing the 6×6 covariance matrix Λ that we want to impose, all we need to do is to randomly choose the *shooting* directions, e.g. the tangent vectors $\beta_k \in S(m,\mathbb{R})$. In other words, we must be able to draw random samples of the β_k, seen as elements of $\mathbb{R}^{m(m+1)/2}$ with zero mean and covariance matrix Λ. Any random element $\Sigma_k \in S^+$ is then obtained by applying the exponential map at $\overline{\Sigma}$ in a given direction β_k. In practice however, there is no need to do so since we have the following relation between β_k and Σ_k: $\beta_k = -\overline{\Sigma}\log(\Sigma_k^{-1}\overline{\Sigma})$. Hence, the random covariance matrix is readily obtained as:

$$\Sigma_k = \left(\exp(-\overline{\Sigma}^{-1}\beta_k)\overline{\Sigma}^{-1}\right)^{-1} \tag{5.13}$$

We now discuss how to compute the random vector β_k. As described for example in [94], it is always possible to produce a random vector according to an imposed covariance matrix Λ and a zero mean value. Indeed, if Z is a random vector of $\mathbb{R}^n$, $n = m(m+1)/2$ with independent and identically distributed components $Z_1,...,Z_n$ of zero mean and unit variance, the vector that we seek is $\beta_k = HZ$, such that:

$$\begin{aligned} \mathbb{E}[\beta_k] &= H\mathbb{E}[Z] = 0 \\ \mathbb{E}[\beta_k\beta_k^T] &= \mathbb{E}[HZZ^TH^T] = H\mathbb{E}[ZZ^T]H^T = HH^T = \Lambda \end{aligned}$$

Putting everything together, the three steps to generate a random element of S^+ with imposed mean and covariance matrix are:

```
Algorithm 2 Generation of Gaussian noise in $S^+(3)$
Require: $\overline{\Sigma}$ and $\Lambda$, mean tensor and covariance matrix
Ensure: $\overline{\Sigma}_i$, $N$ normally distributed elements of $S^+(3)$
for $i = 1$ to $N$ do
    $\Lambda = HH^T$ {Cholesky decomposition of the covariance matrix}
    Create a random vector $Z \in \mathbb{R}^6$, with zero mean and unit variance
    From $\beta_i = \varphi^{-1}(HZ) \in S(3)$
    $\Sigma_i \leftarrow \left(\exp\left(-\overline{\Sigma}^{-1}\beta_i\right)\overline{\Sigma}^{-1}\right)^{-1}$
```

We illustrate this method with the following example and use it to demonstrate the performance of the numerical schemes described in section 5.3.

Results

We have generated a set S of up to 100000 covariance matrices parameterizing 3-variate normal distributions and following the statistical distribution $p(\Sigma|\overline{\Sigma},\Lambda)$ as defined in theorem 5.3.3.1. To that end, we have randomly chosen the mean tensor $\overline{\Sigma}$

Elements	10	100	1000	10000	100000
$\mathcal{D}\left(\overline{\Sigma},\widetilde{\Sigma}\right)$	0.5787	0.1110	0.0267	0.0114	0.0028
$\mathcal{D}\left(\Lambda,\widetilde{\Lambda}\right)$	2.1959	0.5518	0.1484	0.0490	0.0208

Table 5.1: Geodesic distances $\mathcal{D}\left(\overline{\Sigma},\widetilde{\Sigma}\right)$ and $\mathcal{D}\left(\Lambda,\widetilde{\Lambda}\right)$ in terms of the number of elements

and the covariance matrix Λ while ensuring to verify the main hypothesis of theorem 5.3.3.1, namely a moderate variance $\mathrm{tr}(\Lambda) = 1.0$. Those two quantities are given below. It is then straightforward to use them in the above algorithm to generate our set $\mathcal{S}$. Once this synthetic dataset was created, we applied the numerical schemes presented in the previous section to estimate the mean and covariance matrix. The results of these computations with 100000 elements are denoted by $\widetilde{\Sigma}$ and $\widetilde{\Lambda}$. As we can notice, they are very close to the expected values.

$$\overline{\Sigma} = \begin{pmatrix} 0.90324 & 0.12560 & -0.3106 \\ 0.12560 & 0.74092 & 0.20922 \\ -0.3106 & 0.20922 & 1.25043 \end{pmatrix} \widetilde{\Sigma} = \begin{pmatrix} 0.9040 & 0.1254 & -0.3111 \\ 0.1254 & 0.7401 & 0.2107 \\ -0.3111 & 0.2107 & 1.2495 \end{pmatrix}$$

$$\Lambda = \begin{pmatrix} 0.3956 & -0.0538 & -0.0204 & 0.1725 & -0.1387 & -0.0698 \\ -0.0538 & 0.0551 & -0.0140 & -0.0013 & -0.0041 & 0.0455 \\ -0.0204 & -0.0140 & 0.1236 & -0.0256 & -0.0550 & 0.0330 \\ 0.1725 & -0.0013 & -0.0256 & 0.1436 & -0.1035 & 0.0220 \\ -0.1387 & -0.0041 & -0.0550 & -0.1035 & 0.1430 & -0.0526 \\ -0.0698 & 0.0455 & 0.0330 & 0.0220 & -0.0526 & 0.1391 \end{pmatrix}$$

$$\widetilde{\Lambda} = \begin{pmatrix} 0.3954 & -0.0541 & -0.0209 & 0.1714 & -0.1375 & -0.0712 \\ -0.0541 & 0.0549 & -0.0143 & -0.0013 & -0.0036 & 0.0452 \\ -0.0209 & -0.0143 & 0.1253 & -0.0258 & -0.0555 & 0.0330 \\ 0.1714 & -0.0013 & -0.0258 & 0.1425 & -0.1021 & 0.0213 \\ -0.1375 & -0.0036 & -0.0555 & -0.1021 & 0.1415 & -0.0514 \\ -0.0712 & 0.0452 & 0.0330 & 0.0213 & -0.0514 & 0.1383 \end{pmatrix}$$

Table 5.1 shows how well the set $\mathcal{S}$ fits the imposed Gaussian distribution as the number of elements increases. As expected, the error greatly decreases with the size of the random set. One last point of interest before moving to the next application is the performance of the gradient descent algorithm, as already pointed out in [187]. By running it repeatedly with different initial guesses, we have ensured that it was not affected by that parameter since it converged every time in no more than 5 iterations. This was highly reproducible and tested on the synthetic dataset $\mathcal{S}$. In practice, we thus always start from the identity. We also present, in table 5.2, the evolution of the time of convergence (on a Pentium IV at 2.4GHz) in terms of the number of elements in $\mathcal{S}$. This is, of course, $O(N)$.

Elements	10	100	500	1000	5000	10000	50000
Time (in s)	0.0665	0.4143	1.9685	3.8459	19.5005	39.4221	193.7162

Table 5.2: Mean estimation convergence time in terms of the number of elements

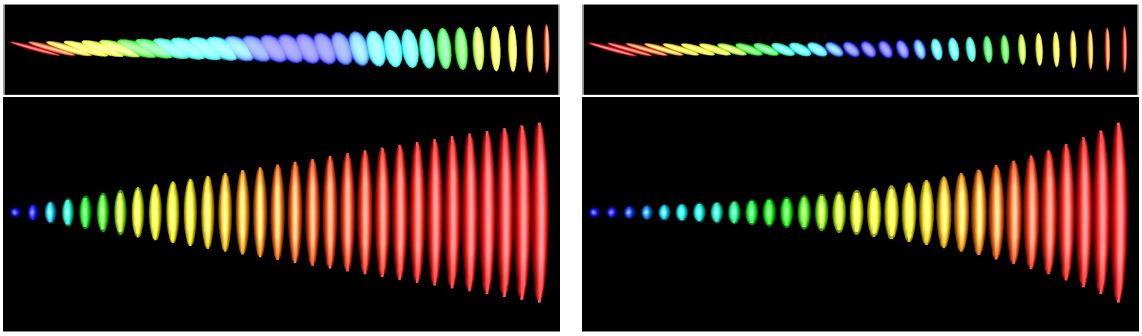


Figure 5.2: Linear Euclidean (left) / Riemannian (right) interpolation examples between two elements of S^+

5.4.2 S^+-valued Data Interpolation

We hereafter apply the above theory to the interpolation of S^+-valued images like those obtained by DTI. We show that moving from the Euclidean to the Riemannian metric has a deep impact on the interpolation results. We experiment on simple synthetic data in one dimension and then show how the procedure can easily be extended to data defined on a regular sampling of a domain $\Omega \subset \mathbb{R}^n$, with n typically being 2 or 3. Many refined approaches have already been proposed in the literature such as [226], [54], [198] and [208] but few have investigated the importance of the choice of the metric [184, 238, 20].

We start by considering the simple problem of the linear interpolation between two elements of S^+, as shown in figure 5.2. In the top row, the interpolation is performed between tensors Σ_1 and Σ_2 with identical eigenvalues but different orientations. In the bottom row, we want to evolve from a perfectly isotropic tensor to a very anisotropic one. In figure 5.2, the color code represents the anisotropy (low is blue and high is red). The interpolation can be done in a Euclidean fashion with

$$\Sigma(t) = (1-t)\Sigma_1 + t\Sigma_2 \quad \forall t \in [0,1]$$

or by resorting to the Riemannian metric and following the geodesic joining Σ_1 to Σ_2 in S^+:

$$\Sigma(t) = \Sigma_1^{1/2} \exp\left(-t\Sigma_1^{1/2} \log\left(\Sigma_2^{-1}\Sigma_1\right)\Sigma_1^{-1/2}\right)\Sigma_1^{1/2}$$

It is quite obvious, from figure 5.2 and figure 5.3 that the interpolation obtained with the Riemannian metric is more natural. Indeed, the tensors shape and anisotropy evolve in a much more consistent manner with this metric than in the Euclidean case (see tensors shape and color code evolution in figure 5.2 and graph in figure 5.3).

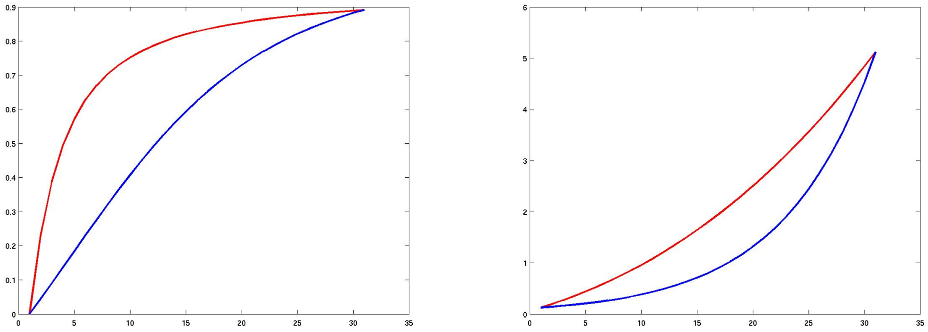


Figure 5.3: Evolution of the anisotropy (left) and determinant (right) for the example depicted in figure 5.2 (bottom row) under Euclidean (red) and Riemannian (blue) interpolation

This is certainly desirable when dealing with human brain DTI datasets since using the Euclidean metric could result, for instance, in artificially low/high anisotropy measures in resampled datasets.

It is straightforward to extend this one-dimensional interpolation procedure to two- or three-dimensional datasets. We only detail the bilinear interpolation procedure in the following, but trilinear interpolation can be achieved in a very similar way and more advanced techniques, like bicubic or spline interpolation, could also be used (see [20] for instance).

For a two-dimensional image $I : \Omega \subset \mathbb{R}^2 \to \mathcal{M}$ such that $I(x) \in \mathcal{M}, \forall x = (x_1, x_2) \in \Omega$, bilinear interpolation at location $x = (x_1, x_2)$ can be done if we know the value of I at four neighboring locations $p^1 = (p_1^1, p_2^1)$, $p^2 = (p_1^1, p_2^2)$, $p^3 = (p_1^3, p_2^1)$ and $p^4 = (p_1^3, p_2^2)$. Successive linear interpolations in both directions yield:

$$\begin{aligned} I(x) \simeq \; & \frac{(p_1^3 - x_1)(p_2^2 - x_2)}{(p_1^3 - p_1^1)(p_2^2 - p_2^1)} I(p^1) \quad + \quad \frac{(x_1 - p_1^1)(p_2^2 - x_2)}{(p_1^3 - p_1^1)(p_2^2 - p_2^1)} I(p^3) \; + \\ & \frac{(p_1^3 - x_1)(x_2 - p_2^1)}{(p_1^3 - p_1^1)(p_2^2 - p_2^1)} I(p^2) \quad + \quad \frac{(x_1 - p_1^1)(x_2 - p_2^1)}{(p_1^3 - p_1^1)(p_2^2 - p_2^1)} I(p^4) \end{aligned} \tag{5.14}$$

In other words, $I(x)$ is a simple weighted average of the values of I at locations p^1, p^2, p^3 and p^4. As we will detail in section 6.2.1, it is easy to extend the numerical scheme proposed in section 5.3 for the computation of the average of a set of elements $\{\Sigma_i\}$, $i = 1, ..., N$ of S^+ to compute a weighted average. This is done by replacing the uniform weights $\frac{1}{N}$ by any set of scalars ω_i such that $\sum_{i=1}^{N} \omega_i = 1$. $I(x)$ can thus be approximated by evaluating the quantity $\sum_{i=1}^{4} \omega_i I(p^i)$ with ω_i readily obtained from equation 5.14.

5.5 Conclusion

We have presented a geometric approach to the statistical analysis of multivariate normal distributions with zero mean vector. We have developed novel algorithms for the estimation of the mean and covariance matrix. We have also described how to compute the Ricci tensor for the space of zero-mean multivariate normal distributions. All these contributions have been used in order to derive and fully characterize a generalized normal law on the space of zero-mean multivariate normal distributions. Finally, we illustrated these concepts by applying them to the generation of random elements of S^+ and to the interpolation of S^+-valued data.

Part II

Methods

CHAPTER 6

FROM DIFFUSION WEIGHTED IMAGES TO DTI

Contents

OVERVIEW

In this chapter, we will address two of the most basic processing tasks for diffusion weighted images (DWI), namely the estimation of diffusion tensors and the regularization of the resulting matrix-valued images. Estimation is of course necessary and must be carefully addressed in order to enforce the properties of diffusion tensors. Otherwise, this may yield, for example, inconsistent directional or anisotropy information. Regularization, on the other side, can be useful to enhance tractography results for instance, when dealing with particularly noisy datasets.

Keywords: diffusion tensor, estimation, M-estimators, weighted average, anisotropic regularization, nonlinear diffusion

6.1 DIFFUSION TENSOR ESTIMATION

As already quickly described in chapter 3, DTI evaluates, from a set of diffusion weighted images, the covariance matrix Σ of the water molecules Brownian motion at each voxel of the acquisition volume $\Omega \subset \mathbb{R}^3$. In other words, it approximates the probability density function modeling the motion of water molecules by a 3-variate normal distribution of zero mean vector $\overline{r} \in \mathbb{R}^3$.

The estimation of the field of 3×3 symmetric positive definite matrices $\Sigma : \Omega \to S^+(3)$ is performed by using the Stejskal-Tanner equation [275] for anisotropic diffusion. We recall that this equation relates the magnetic resonance signal attenuation to the diffusion tensor and the sequence parameters:

$$S_i(x) = S_0(x) \exp\left(-b g_i^T \Sigma(x) g_i\right) \quad \forall x \in \Omega, i = 1, ..., N \tag{6.1}$$

$g_i = (g_i^1, g_i^2, g_i^3)^T, \quad i = 1, ..., N$ are the unit and non-collinear gradient directions corresponding to each diffusion weighted image S_i and b is the diffusion weighting factor. Moreover a reference image S_0, without diffusion gradient, must be acquired [27], [157] , [140]. Many approaches have already been derived to estimate the diffusion tensors $\Sigma(x)$ from a set of DWI (at least 6 since a tensor has 6 degrees of freedom) and references can be found for example in [310], [202], [305], [285] [60], [217] [167, 168]. We must also point out that various works, such as [139] or [227], have addressed the issue of the optimal acquisition scheme for DTI. It is in general a good idea to use much more than 6 diffusion gradients in order to minimize the propagation of noise from the DW images on the tensor elements. As demonstrated in [156], at least 30 gradient directions are necessary to achieve a robust estimation of the apparent diffusion coefficient, fractional anisotropy (see section 6.1.4) and tensor orientations. In the following, we will first review the classical least squares approach and then introduce a novel method, based on the geometry of $S^+(3)$, for tensor estimation. We will show that the latter clearly outperforms the former.

6.1.1 Classical Least Squares Estimation Procedure

The classical technique for tensor components estimation relies on a least squares procedure. The basic idea is to rewrite equation 6.1 as a linear system which can then be efficiently solved.
First, taking the logarithm in equation 6.1 yields

$$-\frac{1}{b}\ln\left(\frac{S_i(x)}{S_0(x)}\right) = g_i^T \Sigma(x) g_i$$

Then, noticing that $g_i^T \Sigma(x) g_i$ can be rewritten as the Euclidean inner product $\langle G_i, \varphi(\Sigma) \rangle$ with

$$G_i = (g_i^1 g_i^1,\, 2g_i^1 g_i^2,\, 2g_i^1 g_i^3,\, g_i^2 g_i^2,\, 2g_i^2 g_i^3,\, g_i^3 g_i^3)^T \in \mathbb{R}^6$$

and

$$\varphi(\Sigma(x)) = (\sigma_{11},\, \sigma_{12},\, \sigma_{13},\, \sigma_{22},\, \sigma_{23})^T \in \mathbb{R}^6$$

as in section 5.2.1, we finally end up with the system

$$\underbrace{\begin{bmatrix} G_1^T \\ \vdots \\ G_N^T \end{bmatrix}}_{\mathbf{G}} \varphi(\Sigma(x)) = \underbrace{\begin{bmatrix} -\frac{1}{b}\ln\left(\frac{S_1(x)}{S_0(x)}\right) \\ \vdots \\ -\frac{1}{b}\ln\left(\frac{S_N(x)}{S_0(x)}\right) \end{bmatrix}}_{\mathbf{Y}(x)} \tag{6.2}$$

In general, N is larger than 6 and this over-determined system can be solved by resorting to the linear least squares method (see [250] for instance):

$$\varphi(\Sigma(x)) = \left(\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{G}^T\right)\mathbf{Y}(x)$$

$\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{G}^T$ being the Moore-Penrose pseudoinverse of $\mathbf{G}$.

This technique is very fast (no more than a few minutes even on large datasets) but does not constrain the tensors to have positive eigenvalues. We now describe a method, based on intrinsic numerical schemes for $S^+(3)$, which naturally produces diffusion tensors with the right mathematical properties and, hence, the right physical meaning (ie. negative eigenvalues are impossible).

6.1.2 Fitting Tensors on $S^+(3)$

We now seek to minimize the following general objective function at each $x \in \Omega$ by searching for the optimal $\Sigma \in S^+(3)$ (we note Σ instead of $\Sigma(x)$ for the sake of clarity):

$$\mathscr{E}(S_0, ..., S_N) = \sum_{i=1}^{N} \psi\left(\frac{1}{b}\ln\left(\frac{S_i}{S_0}\right) + g_i^T \Sigma g_i\right) \tag{6.3}$$

where $\psi : \mathbb{R} \rightarrow \mathbb{R}$ is a real-valued functional that, in the M-estimators framework, tries to reduce the effect of outliers by replacing the classical squared residual $\psi(r_i) = r_i^2/2$ by another function. M-estimators, or "maximum likelihood estimators", always assume some particular distribution for the data of interest. If the data do not follow this distribution or present many outliers, M-estimators can become inefficient. It was proposed by Huber [144] to choose ψ, or its derivative ρ, so as to exhibit an unbiased and efficient behavior when the data do fit the assumed model and a reasonable behavior when it follows a different but close distribution. ψ must have a unique minimum at zero and should be chosen to be less increasing than the square function (see for example [323]).

Following [285], we will use intrinsic numerical schemes on $S^+(3)$, similar to the gradient descent proposed in section 5.3 to estimate the empirical mean, for our tensor estimation purpose. We will not incorporate any smoothness constraint on the tensor field but the major advantage of this approach is to naturally evolve on $S^+(3)$. Compared to simple least squares, this ensures the symmetry and positive definiteness of each diffusion tensor. Moreover, the combination of robust regression methods with this intrinsic gradient descent enables us to propose a more reliable and efficient estimation technique.

In order to minimize energy 6.3 through a gradient descent, we follow exactly the same idea than for the mean and thus need to compute the gradient of $\mathscr{E}$. We recall that, given a smooth function $f : \Sigma \in S^+(3) \mapsto f(\Sigma) \in \mathbb{R}$, its derivative in the direction v at Σ:

$$\mathfrak{D}f(\Sigma)v = \lim_{s \rightarrow 0} \frac{f(\Sigma + sv) - f(\Sigma)}{s}$$

and the inner product $\langle ., . \rangle_\Sigma$, the gradient ∇f exists and is unique by the Riesz-Fréchet theorem. It is defined by the relationship:

$$\mathfrak{D}f(\Sigma)v = \left.\frac{df(\Sigma(t))}{dt}\right|_{t=0} = \left.\langle \nabla f(\Sigma(t)), v \rangle_{\Sigma(t)}\right|_{t=0}$$

where $\Sigma(t)$ is the unique geodesic starting from $\Sigma(0) = \Sigma$ in the direction $v = \dot{\Sigma}(0)$. With the residual $r_i(\Sigma(t)) = \frac{1}{b} \ln\left(\frac{S_i}{S_0}\right) + g_i^T \Sigma(t) g_i$ we have by the chain rule:

$$\begin{aligned}
\left.\frac{d\psi(r_i(\Sigma(t)))}{dt}\right|_{t=0} &= \left.\frac{d\psi(r_i)}{dr_i}\right|_{t=0} \left.\left(\sum_{k,l=1}^{3} \frac{\partial r_i(\Sigma(t))}{\partial \Sigma_{kl}(t)} \frac{d\Sigma_{kl}(t)}{dt} \right)\right|_{t=0} \\
&= \frac{d\psi(r_i)}{dr_i} \mathbf{tr} \left.\left(\frac{dr_i(\Sigma(t))}{d\Sigma(t)} \left(\frac{d\Sigma(t)}{dt} \right)^T \right)\right|_{t=0} \\
&= \psi'(r_i) \mathbf{tr} \left(g_i g_i^T \dot{\Sigma}(0) \right) \\
&= \psi'(r_i) \mathbf{tr} \left(\Sigma^{-1} \Sigma g_i g_i^T \Sigma^T \Sigma^{-T} \dot{\Sigma}(0) \right) \\
&= \psi'(r_i) \mathbf{tr} \left(\Sigma^{-1} \Sigma g_i \left(\Sigma g_i \right)^T \Sigma^{-1} \dot{\Sigma}(0) \right)
\end{aligned}$$

The last line uses the symmetry of Σ, ie. $\Sigma^{-T} = \Sigma^{-1}$. Hence, because of equation 5.1,

$$\nabla\psi(r_i(\Sigma)) = \psi'(r_i(\Sigma))\Sigma g_i \left(\Sigma g_i\right)^T \in S(3)$$

and finally:

$$\nabla\mathscr{E} = \sum_{i=1}^{N} \psi'(r_i(\Sigma))\Sigma g_i \left(\Sigma g_i\right)^T$$

We can use the intrinsic step-forward integrator of proposition 5.3.1.1 to obtain the gradient descent (which depends on the function ψ) to be carried out at each voxel x of the volume Ω:

$$\begin{aligned} \Sigma_{l+1} &= \Sigma_l^{1/2} \exp\left(-ds\Sigma_l^{-1/2}\nabla\mathscr{E}\Sigma_l^{-1/2}\right)\Sigma_l^{1/2} \\ &= \Sigma_l^{1/2} \exp\left(-ds\Sigma_l^{-1/2}\left(\sum_{i=1}^{N}\psi'(r_i(\Sigma_l))\Sigma_l g_i \left(\Sigma_l g_i\right)^T\right)\Sigma_l^{-1/2}\right)\Sigma_l^{1/2} \end{aligned}$$

We now discuss some numerical issues and show that this methods yields better results on synthetic and real datasets.

6.1.3 Numerical Experiments

The numerical scheme given above has been implemented for various well-known functions ψ, namely the Cauchy, Fair, Huber and Tukey M-estimators, and their associated influence function $\psi'(r)$. In practice however, we use Huber's function

$$\psi(r) = \begin{cases} \frac{r^2}{2} & \text{for} \quad |r| \le k \\ k(|r| - \frac{k}{2}) & \text{for} \quad |r| > k \end{cases}$$

since this estimator is very satisfactory with the classical tuning constant $k = 1.2107$ which allows to achieve an asymptotic efficiency of 95% on the standard normal distribution.

The combination of the M-estimators with our intrinsic gradient descent has shown to be far less sensitive to noise than the classical least squares and to naturally avoid the estimation of non-positive definite tensors. By applying a least squares estimation procedure on various DTI datasets, we usually obtained a few thousands negative definite or positive semidefinite tensors (see [167] for a recent investigation on this matter). This is obviously always corrected by our method. We however need to point out the higher computational overhead of the gradient descent, by comparison to least squares. Our method, because of the need to perform a gradient descent at voxel of the image (this can take up to a few hundreds iterations), typically requires a computation time of one to two hours for volumes of dimensions $128 \times 128 \times 64$.

To prove the robustness of our method, we have generated a synthetic dataset

	NPD tensors	Mean	Variance	Minimum	Maximum
Least Squares	225	12.8906	40.4182	0.00198837	227.899
Gradient Descent	0	0.587561	5.78362	0.00409459	143.845

Table 6.1: Comparison of the performance of least squares and gradient descent tensors estimation: [COL. 1] Number of non-positive definite tensors, [COL. 2-5] Statistics of the geodesic distance between estimated and ground-truth tensor fields

consisting of a 50×50 tensor field divided in 2 regions: the background and a 20×20 square area centered at pixel $(25, 25)$. Each region was assigned with a distinct mean tensor and covariance matrix. We used the method proposed in section 5.4.1 to randomly choose the tensors belonging to each region. They follow 2 different normal distributions and thus span a wide range of tensor configurations, which is desirable to consistently evaluate the performance of our estimator. The tensor field is presented on figure 6.1 (Left) where each of the 3×3 matrix is depicted as a 3D ellipsoid. We created an artificial set of diffusion weighted images from this tensor field by using the Stejskal-Tanner equation 6.1. S_0 was taken to be constant and equal to 10 everywhere while the diffusion weighting factor b was set to $1 s.mm^{-2}$. Finally, 12 gradient directions g_i were used and given by the vertices of the icosahedron. The images were corrupted by a Gaussian noise with variance 0.5^2 (images intensity amplitude was approximately $[2, 10]$).

We performed the estimation of the diffusion tensors by least squares and by our gradient descent algorithm with time step $ds = 0.2$ and 600 iterations. It was then possible to compare the accuracy of the reconstructed tensor fields by computing the geodesic distance between the tensor estimates and the ground-truth data at each voxel. Table 6.1 summarizes the results and is a striking proof of the gradient descent approach superiority.

We also applied our approach to a real set of diffusion weighted images. They were acquired at the Center for Magnetic Resonance Research at the University of Minnesota, on a 3 Tesla Siemens Magnetom Trio whole-body clinical scanner. As for the synthetic dataset, measurements were made along 12 gradient directions. We used a classical diffusion weighting factor $b = 1000 s.mm^{-2}$, TE$= 92ms$ and TR$= 1690ms$. The images were obtained on 64 evenly spaced axial planes with 128×128 pixels per slice. Voxel size is $2 \times 2 \times 2mm$. An axial slice of the obtained tensor field, after 1500 iterations with $ds = 0.2$, is shown on figure 6.1 (Right). We have observed particularly significant differences between the two estimation methods in highly anisotropic regions, such as the corpus callosum, where the least square method could lead to negative definite or positive semidefinite tensors. In such cases, the nonpositive eigenvalues are artificially set to a very small non-null

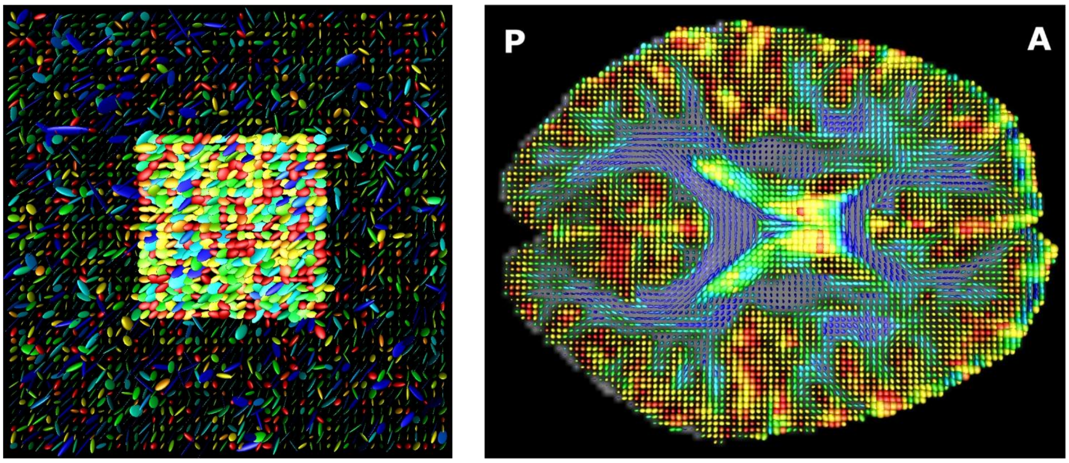


Figure 6.1: Diffusion Tensor Estimation: (Left) Synthetic tensor field used to compare our method to least squares estimation, (Right) Estimation of diffusion tensors from real DWI by gradient descent (Red: low anisotropy / Blue: high anisotropy / A: anterior / P: posterior)

value, leading to unreliable, highly planar (with one negative or null eigenvalue) or linear (with two negative or null eigenvalues) anisotropic tensors and possibly to numerical instabilities in subsequent tensor based computations. This situation, naturally, never occurs with the intrinsic gradient descent method as the solution necessarily belongs to $S^+(3)$. Figure 6.2 illustrates these differences with a close up of an axial slice containing the splenium of the corpus callosum.

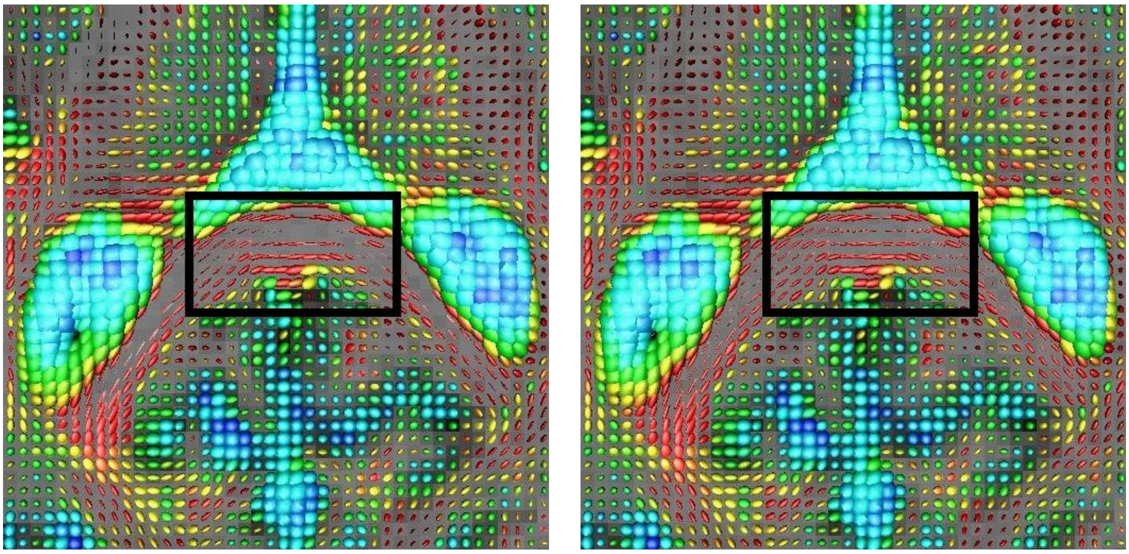
6.1.4 From Diffusion Tensor to Tissues Properties

We quickly recall some of the important quantities that can be derived from diffusion tensors to characterize the microstructure of the tissues of interest. First of all, a tensor Σ can be diagonalized, since it is symmetric and positive definite, such that

$$\Sigma = UDU^T$$

where U is an orthogonal matrix whose columns coincide with the eigenvectors $\mathbf{u}_1, \mathbf{u}_2$ and $\mathbf{u}_3$ of Σ and D is a diagonal matrix composed by the eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$ of Σ. Eigenvectors $\mathbf{u}_i$ provide information on the local orientation of tissues. For instance, the major eigenvector $\mathbf{u}_1$, also called principal diffusion vector, is widely used to generate color coded images where the red, green and blue components of the RGB image are given by the three elements of each vector $\mathbf{u}_1$. From the eigenvalues, it is possible to derive several scalar quantities like, among others, the mean diffusivity (MD), fractional anisotropy (FA), linear anisotropy (LA), planar anisotropy (PA) and spherical anisotropy (SA) [22, 27, 242, 310] (see examples in figure 6.3):

- The mean diffusivity is independent of the orientation of diffusion, it is an overall evaluation of the mean-squared displacement of water molecules. It is com-



Least squares method Riemannian method

Figure 6.2: Estimation of Diffusion Tensors: comparison between classical least squares (Left) and gradient descent in $S^+(3)$ (Right). (Blue: low anisotropy; Red: high anisotropy). Notice the difficulty to represent tensors in the middle of the corpus callosum with the least square approach, suggesting degenerate “needle-shaped” tensors in this region.

puted from the trace of the diffusion tensor:

$$MD = \frac{\text{tr}(\Sigma)}{3} = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}$$

- The fractional anisotropy is certainly the most widely used anisotropy measure. It is based on the variance of the eigenvalues and expressed as

$$FA = \frac{\sqrt{3\left((\lambda_1 - \lambda_2)^2 + (\lambda_1 - \lambda_3)^2 + (\lambda_2 - \lambda_3)^2\right)}}{\sqrt{2\left(\lambda_1^2 + \lambda_2^2 + \lambda_3^3\right)}}$$

- Westin *et al.* [310] introduced the notions of linear, planar and spherical anisotropy which respectively measure the tendency of a tensor to be elongated, oblate or spherical. They come from the fact that any tensor Σ can be expressed in a basis composed by a linear tensor $\mathbf{L} = \mathbf{u}_1\mathbf{u}_1^T$, a planar tensor $\mathbf{P} = \mathbf{u}_1\mathbf{u}_1^T + \mathbf{u}_2\mathbf{u}_2^T$ and a spherical tensor $\mathbf{S} = \mathbf{u}_1\mathbf{u}_1^T + \mathbf{u}_2\mathbf{u}_2^T + \mathbf{u}_3\mathbf{u}_3^T$, in other words $\Sigma = c_l\mathbf{L} + c_p\mathbf{P} + c_s\mathbf{S}$ with $c_l = LA = \frac{\lambda_1-\lambda_2}{\lambda_1+\lambda_2+\lambda_3}$, $c_p = PA = \frac{2(\lambda_2-\lambda_3)}{\lambda_1+\lambda_2+\lambda_3}$ and $c_s = SA = \frac{3\lambda_3}{\lambda_1+\lambda_2+\lambda_3}$.

6.1.5 Conclusion

We have demonstrated that it was possible to naturally enforce the positivity constrain in the tensor estimation procedure by working with adequate numerical tools

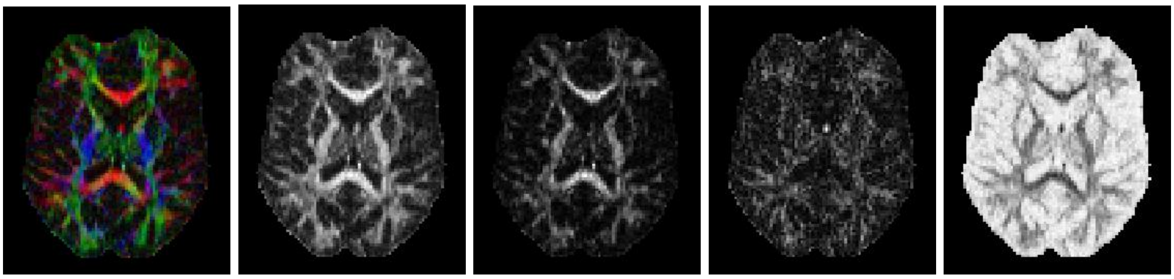


Figure 6.3: From left to right: Principal diffusion vector $\mathbf{u}_1$, fractional anisotropy, linear anisotropy, planar anisotropy and spherical anisotropy

on the manifold $S^+(3)$. Our experiments illustrated the advantage of the method presented in section 6.1.2 over classical least squares. However, working with tensor fields may also require some regularization to reduce the noise arising, for instance, from the acquisition process. In the next section, we focus on the contribution of the mathematical tools, developed in chapter 5, to the anisotropic filtering and regularization of tensor fields. To validate our approach we present promising results on both, synthetic and real DTI data.

6.2 REGULARIZATION OF DIFFUSION TENSOR IMAGES

Regularization and filtering schemes of tensor fields are widely studied in the literature, especially in the domain of DTI. As an example, [285] deals with the problem of regularizing noisy fields of diffusion tensors, considered as symmetric and positive definite $n \times n$ matrices, through a PDE-based scheme and a spectral decomposition. A complementary work is that presented in [63] which provides a geometric interpretation of constrained flows for matrix-valued functions. This yields, through the use of exponential maps, suitable numerical schemes that are also constraints preserving. Another approach presented in [308] provides a generalization of anisotropic and nonlinear diffusion process to matrix-valued data. More recently, [80] proposes a scheme, based on variational methods, restoring the main direction of the tensors and using the resulting direction to regularize the eigenvalues by an anisotropic diffusion process. However, tensor eigenvalues tend to regularize faster than the associated eigenvectors. This phenomenon is known as the *eigenvalue swelling effect* for long regularization time, as shown in [284], while noise removal is not quite significant for short regularization time. Other works try to couple the regularization with the tensors estimation process from diffusion-weighted images. For example, [305] presents a constrained variational principle which involves the minimization of a regularization term, based on L^p norms and subject to a nonlinear constraint on the data to obtain positive definite tensors. Recently, [311] proposes techniques based on normalized convolution or Markov random fields for the regularization of tensor fields.

In this section, we will rely on the properties of the space of multivariate normal distributions, as presented in chapter 5.

Contributions of this section:

In the following, we use the mathematical framework presented in chapter 5 to take into account the particular geometry of the set of symmetric, positive definite matrices in the smoothing process. We introduce an anisotropic filtering algorithm controlled by the magnitude of the spatial gradient of the tensor field. On the contrary of [116], where a statistical analysis of diffusion tensors was proposed but without addressing the problem of smoothing, or [238], where the regularization task was addressed with a PDE point of view, our smoothing method, developed in parallel to [238] and based on our previous work developed earlier in [187], only relies on simple and local anisotropic averaging. Adding to that, our method is favorably compared to a state-of-the-art approach [308]. A detailed analysis of the performances of our approach is performed and qualitative and quantitative results obtained on noisy and synthetic data show that our approach outperforms the one proposed in [308]. The emphasize of this section is on the clear and efficient application of the tools presented in chapter 5 to solve the important problem of anisotropically smoothing a set of noisy tensor data using the right concepts and tools, extending previous PDE-based approaches such as [285].

Organization of this section:

Section 6.2.1 shows how the tools introduced in chapters 4 and 5 can be used to compute local averages and spatial gradient of diffusion tensor images. Section 6.2.2 introduces the filtering process. Finally, section 6.2.3 presents and discusses numerical experiments. We will show that our Riemannian anisotropic filtering method yields better results on both synthetic and real DTI datasets when compared to other approaches such as the nonlinear diffusion proposed in [308].

6.2.1 Local Average and Spatial Gradient of Diffusion Tensor Fields

We hereafter present how to compute the anisotropic local average and spatial gradient of a diffusion tensor image. The first paragraph is a direct extension of section 5.3.1 and also useful for the interpolation task described in section 5.4.2. The second paragraph is the application of section 4.3.2 to the particular case of the manifold of interest here, namely $S^+(3)$.

Weighted intrinsic mean

In the same way the intrinsic mean was defined in section 5.3.1, we can also calculate a weighted intrinsic mean which ponderates a set of N normal distributions

$\Sigma_1, \Sigma_2, \dots, \Sigma_N \in S^+(3)$ with the weights $\omega_1, \omega_2, \dots, \omega_N, \omega_i \in \mathbb{R}^+$ In this case, the normal distribution parameterized by $\hat{\Sigma}^w \in S^+(3)$ and defined as the weighted empirical mean of N distributions $\Sigma_1, \Sigma_2, \dots, \Sigma_N$, achieves a minimum of the weighted sum of squared distances defined by

$$\mu^w(\hat{\Sigma}^w, \Sigma_1, \Sigma_2, \dots, \Sigma_N) = \frac{\sum_{k=1}^N \omega_k \mathcal{D}^2(\hat{\Sigma}^w, \Sigma_k)}{\sum_{k=1}^N \omega_k} \tag{6.4}$$

Following the same steps as in section 5.3.1, the following evolution equation for the weighted intrinsic mean is readily obtained:

$$\hat{\Sigma}^w_{l+1} = \hat{\Sigma}_l^{w^{1/2}} \exp\left(-dt \frac{\hat{\Sigma}_l^{w^{1/2}} \left(\sum_{k=1}^N \omega_k \log(\Sigma_k^{-1} \hat{\Sigma}_l^w) \right) \hat{\Sigma}_l^{w^{-1/2}}}{\sum_{k=1}^N \omega_k} \right) \hat{\Sigma}_l^{w^{1/2}} \tag{6.5}$$

In practice, for the sake of efficiency and stability, it is important to implement this evolution by carefully checking the amplitude and direction of the velocity. This will be detailed in section 6.2.2.

Spatial gradient of diffusion tensor fields

In this section, we show why the notions introduced in section 4.3.2 are true for the manifold of interest S^+. We recall, from sections 4.3.2 and 5.3, that it is possible to estimate the magnitude of the gradient of a diffusion tensor field $\Sigma(x) : \Omega \subset \mathbb{R}^3 \mapsto S^+(3)$ at $x \in \Omega$ through the sum of squared geodesic distances between tensors in orthogonal directions on a discrete grid, as indicated by the following expression:

$$|\nabla \Sigma(x)|^2 \simeq \sum_{k=1}^3 \mathcal{D}^2(\Sigma(x), \Sigma(x \pm e_k)) \tag{6.6}$$

where e_k denotes the elements of the canonical basis in $\mathbb{R}^3$. To derive equation 6.6, we use the explicit formulation of the geodesic presented in section 5.2, which allows us to calculate the geodesic starting at $\Sigma_0 = \Sigma(x)$ in the direction $V = \dot{\Sigma}(0)$ as:

$$\Sigma(t) = \Sigma_0^{1/2} \exp(t \Sigma_0^{-1/2} V \Sigma_0^{-1/2}) \Sigma_0^{1/2} \tag{6.7}$$

Hence, as we know that $\Sigma(1) = \Sigma(x \pm e_k) = \Sigma_1$, we obtain that:

$$V = \Sigma_0^{1/2} \log(\Sigma_0^{-1/2} \Sigma_1 \Sigma_0^{-1/2}) \Sigma_0^{1/2} = \dot{\Sigma}(0) \tag{6.8}$$

It can be shown that this quantity is actually equivalent to the opposite of the gradient of the squared geodesic distance $\nabla \mathcal{D}^2(\Sigma_0, \Sigma_1)$ whose expression was given by Moakher [211]:

$$\nabla \mathcal{D}^2(\Sigma_0, \Sigma_1) = \Sigma_0 \log\left(\Sigma_1^{-1} \Sigma_0\right)$$

Indeed, it is easy to see that $\Sigma_0^{1/2} \log(\Sigma_0^{-1/2} \Sigma_1 \Sigma_0^{-1/2}) \Sigma_0^{1/2}$ can be rewritten as $\Sigma_0 \Sigma_0^{-1/2} \log(\Sigma_0^{-1/2} \Sigma_1 \Sigma_0^{-1/2}) \Sigma_0^{1/2}$. Using the following property:

$$M_1^{-1} \log(M_2) M_1 = \log(M_1^{-1} M_2 M_1), \ \forall M_1, M_2 \in GL(m) \tag{6.9}$$

we get

$$V = \Sigma_0 \log(\Sigma_0^{-1}\Sigma_1) = -\Sigma_0 \log(\Sigma_1^{-1}\Sigma_0)$$

The symmetric matrix V can thus be used to approximate the spatial directional derivative of a tensor field. To better understand why this is correct, we can consider the Euclidean distance between tensors

$$\mathcal{D}_E^2(\Sigma_0, \Sigma_1) = |\Sigma_0 - \Sigma_1|_F^2 = \mathrm{tr}\left((\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_1)^T\right)$$

for which, using the fact that $\nabla \mathrm{tr}(XY) = Y^T$, $\forall X, Y \in GL(m)$, we have $\nabla \mathcal{D}_E^2(\Sigma_0, \Sigma_1) = \Sigma_0 - \Sigma_1$. In other words, this corresponds to the *difference* tangent vector, usually used in finite difference schemes to approximate a spatial gradient. V generalizes this notion by taking into account the Riemannian structure of the space $S^+(3)$.

All we need to prove now is that the magnitude of the tangent vector V, taking into account the Riemannian metric of $S^+(3)$, is equal to the squared geodesic distance between Σ_0 and Σ_1, so that equation 6.6 is true.

Noting that V is, by definition, the logarithm of Σ_1 at Σ_0, $\log_{\Sigma_0}(\Sigma_1) = \dot{\Sigma}_0$, we know this is true since one of its properties is:

$$\langle \dot{\Sigma}_0, \dot{\Sigma}_0 \rangle_{\Sigma_0} = \mathcal{D}^2(\Sigma_0, \Sigma_1)$$

Indeed, we have

$$\langle V, V \rangle_{\Sigma_0} = \frac{1}{2}\mathrm{tr}\left(\left(-\Sigma_0^{-1}\Sigma_0 \log(\Sigma_1^{-1}\Sigma_0)\right)^2\right) = \frac{1}{2}\mathrm{tr}\left(\left(\log(\Sigma_0^{-1}\Sigma_1)\right)^2\right)$$

and since the matrices $\Sigma_0^{-1}\Sigma_1$ and $\Sigma_0^{-1/2}\Sigma_1\Sigma_0^{-1/2}$ are similar, we have $\langle V, V \rangle_{\Sigma_0} = \frac{1}{2}\mathrm{tr}\left(\log^2\left(\Sigma_0^{-1/2}\Sigma_1\Sigma_0^{-1/2}\right)\right) = \mathcal{D}^2(\Sigma_0, \Sigma_1)$. Hence, to obtain the gradient magnitude within the Riemannian framework, we simply calculate the geodesic distances in orthogonal directions following equation 6.6.

Let us now compare the differences between the gradient magnitude in a Euclidean sense and our definition in the Riemannian framework. For the sake of simplicity, we are going to study the case of tensors in a two-dimensional neighborhood $\mathcal{N}$ containing a boundary like that shown in figure 6.4, where white dots represent tensors $T_1 = \lambda_1 \mathbf{v_1}\mathbf{v_1}^T + \lambda_2 \mathbf{v_2}\mathbf{v_2}^T$ and black dots represent tensors $T_2 = k_1\lambda_1 \mathbf{v_1}\mathbf{v_1}^T + k_2\lambda_2 \mathbf{v_2}\mathbf{v_2}^T$, with $\lambda_i, (i = 1, 2)$ being the eigenvalues of T_1, associated to eigenvectors $\mathbf{v_i}$ and some constants k_i. Using equation 6.6 and the definition of the geodesic distance in equation 9.4, it is not hard to see that:

$$|\nabla \mathcal{N}|^2 = \mathcal{D}^2(T_1, T_2) = \frac{1}{2}(\log^2(k_1) + \log^2(k_2)) \tag{6.10}$$

while the magnitude of the Euclidean gradient is:

$$|\nabla \mathcal{N}|_E^2 = |T_1 - T_2|_F^2 = \lambda_1^2(1 - k_1)^2 + \lambda_2^2(1 - k_2)^2 \tag{6.11}$$

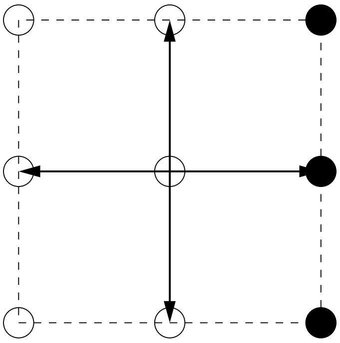


Figure 6.4: 3×3 Neighborhood $\mathcal{N}$ of a tensor field. White dots represent tensors T_1 whereas black dots are tensors T_2

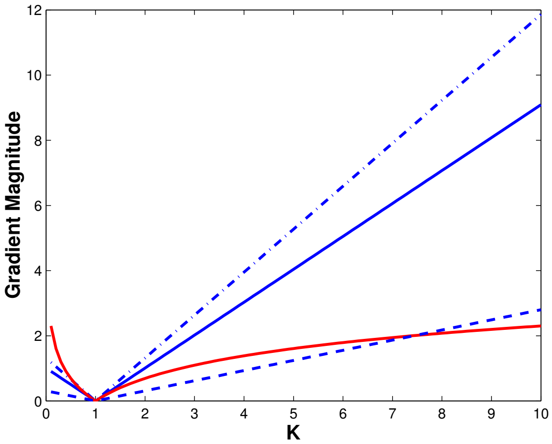


Figure 6.5: Gradient magnitudes as a function of parameter K. Red Line: Riemannian Gradient Magnitude. Blue Lines: Euclidean Gradient Magnitude for different eigenvalues Dashed Line: $\lambda_1 = 0.22, \lambda_2 = 0.22$. Solid Line: $\lambda_1 = 0.94, \lambda_2 = 0.33$. Dashed-Dotted Line: $\lambda_1 = 1.05, \lambda_2 = 0.80$.

It is remarkable that the Riemannian gradient magnitude is independent of the eigenvalues since it only depends on the constant factors, that is to say, the magnitude is scale-independent. Figure 6.5 presents both gradient magnitudes for different values of $k_1 = k_2 = K$ and different values of λ_1, λ_2. While the Riemannian gradient is not affected by different values of λ_1, λ_2, the slope of the Euclidean gradient magnitude is determined by the square root of the sum of the squared eigenvalues, showing that the larger the eigenvalues, the steepest the slope represented by the blue lines in figure 6.5.

Let us now study the different behaviors when we have, in our neighborhood $\mathcal{N}$, the same tensors but rotated by an angle α. The white tensors are now defined as $T_1 = \lambda_1 \mathbf{v_1} \mathbf{v_1}^T + \lambda_2 \mathbf{v_2} \mathbf{v_2}^T$ while $T_2 = P^T T_1 P$, with P an orthogonal unitary rotation matrix. Again, using equations 6.6 and 9.4 it is not difficult to derive the following expression for the Riemannian gradient magnitude:

$$|\nabla\mathcal{N}|^2 = \mathcal{D}^2(T_1, T_2) = \frac{1}{2}(\log^2(A^+) + \log^2(A^-)) \tag{6.12}$$

where A^+ and A^- are, respectively

$$A^{\pm} = \frac{(2\lambda_1\lambda_2 + \sin^2(\alpha)(\lambda_1 - \lambda_2)^2 \pm \sin(\alpha)(\lambda_1 - \lambda_2)\sqrt{\sin^2(\alpha)(\lambda_1 - \lambda_2)^2 + 4\lambda_1\lambda_2})}{2\lambda_1\lambda_2} \tag{6.13}$$

On the other hand, the Euclidean gradient magnitude is:

$$|\nabla\mathcal{N}|_E^2 = |T_1 - T_2|_F^2 = 2\sin^2(\alpha)(\lambda_1 - \lambda_2)^2 \tag{6.14}$$

Again, the dependence of the Riemannian gradient magnitude on eigenvalues is less important than that of the Euclidean gradient magnitude. Figure 6.6 shows different responses of the gradient magnitudes as a function of the angle $\alpha \in [0, \pi]$. On the left, we show the gradient magnitudes for anisotropic tensors, for which eigenvalues ratio is large. In that case, for $\lambda_1 = 0.99, \lambda_2 = 0.2$ it can be seen that the response of the Riemannian gradient (solid red line) is larger than the response of the Euclidean gradient. On the right, we show the gradient magnitudes with more isotropic tensors, with eigenvalues ratio close to one ($\lambda_1 \simeq \lambda_2$). In this case, the Riemannian gradient magnitude provides a smaller response (solid red line) than the Euclidean gradient magnitude (solid blue line). However, as the Euclidean gradient magnitude is proportional to the squared eigenvalues, a problem may arise with isotropic tensors if the eigenvalues are too large, as the dashed lines show. In that case, with $\lambda_1 = 19, \lambda_2 = 17.5$, the Euclidean gradient magnitude is much bigger than the Riemannian one which, on the contrary, stays almost identical to its value for small λ_i.

6.2.2 Anisotropic Filtering

We now make use of the concepts previously presented to develop our Riemannian anisotropic smoothing algorithm. It is detailed in the first paragraph of this section. We also review, in the second paragraph, another filtering algorithm based on the generalization of PDEs to tensor fields. These methods will be quantitatively compared in the next section.

Riemannian anisotropic smoothing

In this section we develop a boundary preserving smoothing algorithm. In practice, we simply use the step-forward operator of equation 6.5 to compute local weighted averages. The anisotropic behavior is introduced by weighting each sample, within a local neighborhood, by a function that depends on the Riemannian gradient magnitude, as developed in section 6.2.1. This function is chosen so that, in homogeneous regions, the weights are constant and the tensors are isotropically averaged. On the

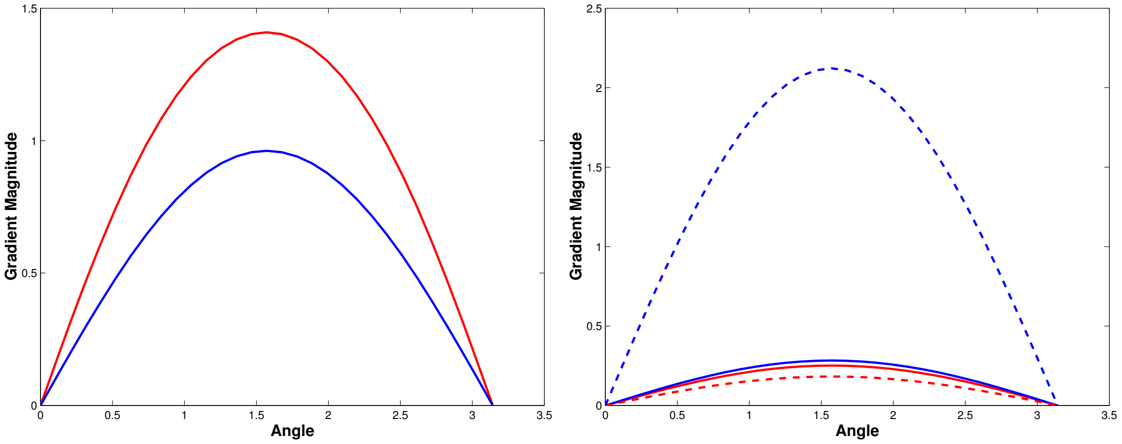


Figure 6.6: Gradient magnitudes as a function of rotation angle α. Red Lines: Riemannian Gradient Magnitude. Blue Lines: Euclidean Gradient Magnitude. Left Image: Response for anisotropic tensors. Right Image: Response for more isotropic tensors.

contrary, when lying on an edge of the image, we would like that only samples on that boundary, and not those across, contribute to the local averaging. To achieve this goal and avoid mixing structures of the image, one possible choice for the weighting function is $\omega_k = \epsilon + |\nabla\Sigma(x)|^2$.

A major advantage of this approach is that a straightforward C++ implementation yields a quite computationally efficient algorithm since, to regularize a $50 \times 50 \times 50$ volume of 3×3 tensors, using a $3 \times 3 \times 3$ averaging neighborhood, we obtain a processing time of about 8 minutes on a 1.7 GHz Pentium M CPU with 1 Gb of RAM. Moreover, it is easy to automatically detect the convergence of the gradient descent, detailed in equation 6.5, by checking the evolution speed

$$\frac{\Sigma_l^w \left(\sum_{k=1}^N \omega_k \log(\Sigma_k^{-1} \hat{\Sigma}_l^w) \right)}{\sum_{k=1}^N \omega_k}$$

and stopping whenever a given norm (Frobenius, Riemannian ...etc for instance) of this symmetric matrix has reached a certain threshold (10^{-6} in practice). Hence not only do we ensure the convergence of the weighted mean but we also discard the need for a parameter such as the number of iterations. Instead of setting a fixed number of iterations, often too large (for the full volume) we can decide (for each voxel) whether the gradient descent has converged or not.

In order to avoid any oscillating behavior in the gradient descent, we verify at each iteration that the velocity does not change sign. If this is the case, it means that we are close to the minimum but that the time step is too large, hence generating oscillations around that minimum. Whenever this situation occurs, we simply come back to the previous state of the gradient descent and adapt the time step by dividing it by 2 until convergence. This may be repeated while approaching the minimum, if necessary.

Nonlinear diffusion of tensor fields

In this section, we focus on nonlinear diffusion filters, as presented in [308] and studied, in the context of local structure tensor estimation, in [43]. For scalar images, diffusion filters provide a family of images $\{u(x,t) \mid t > 0\}$ for an initial image $u(x,0)$ by solving the following partial differential equation (PDE)

$$\partial_t u(x) = \triangle u(x) \tag{6.15}$$

which is equivalent to a convolution with a Gaussian kernel with standard deviation $\sigma = \sqrt{2t}$. The goal of nonlinear diffusion filters is to reduce smoothing across boundaries [239]. To achieve that goal, it is necessary to introduce in equation 6.15 a function g, called diffusivity function, which correlates the amount of smoothing with the gradient magnitude of the image, as follows:

$$\partial_t u(x) = div(g(|\nabla u|)\nabla u(x)) \tag{6.16}$$

Several diffusivity functions have been proposed in the literature with different impact on the resulting image [239, 307, 61]. The total variation flow, for instance, (Equation 6.17) has been widely used since it is theoretically well-founded:

$$g(|\nabla u|) = \frac{1}{|\nabla u|} \tag{6.17}$$

This scheme can also be generalized for a tensor field $T(x) = (u(x)_{i,j}) : \Omega \mapsto S^+(3)$ as proposed in [308], which leads to the following PDE:

$$\partial_t u(x)_{i,j} = div(g(\sum_{k,l=1}^{m} |\nabla u_{k,l}|^2)\nabla u(x)_{i,j}) \tag{6.18}$$

To deal with digital images, whether tensor- or scalar-valued images, we have to discretize the PDE and build an iterative algorithm in order to obtain a filtered version of the initial digital image. The final results mainly depends on the number of iterations n and the time step δt between two consecutive iterations. The main drawback of this approach is that there is no criterion to decide the optimal number of iterations and/or time step, and different solutions may be obtained. However, it can be easily shown that tensors evolving under equation 6.18 stay symmetric positive definite if the initial value $T(0) = (u(0)_{i,j})$ is symmetric positive definite [43].

6.2.3 Numerical Experiments

In this section, we present various numerical experiments and compare the regularization methods previously detailed on synthetic and real DTI data.

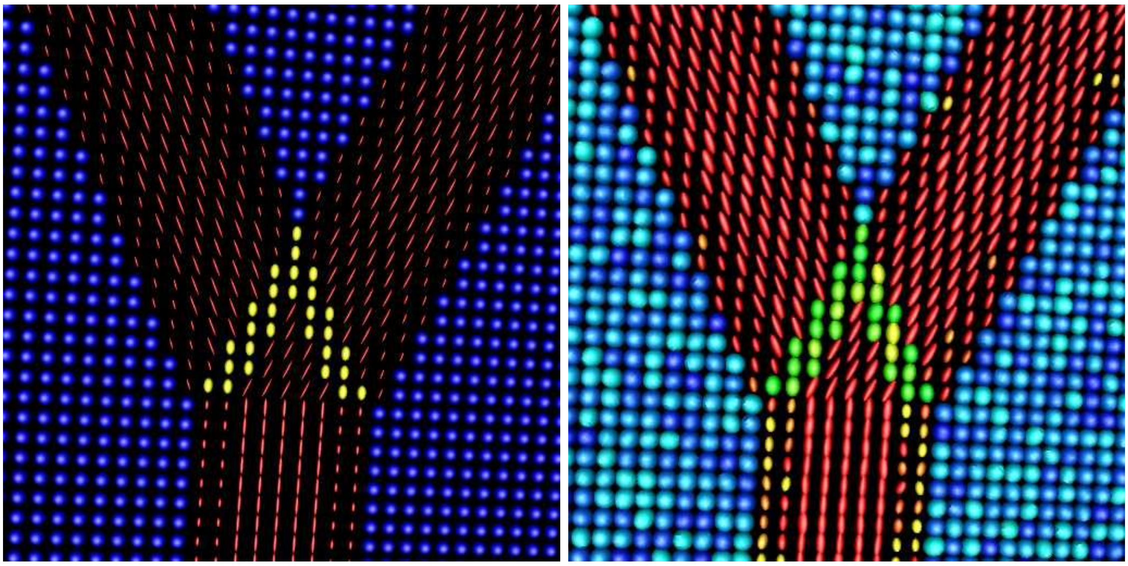


Figure 6.7: Left: Original synthetic dataset. Right: Noisy image. / Color code: Blue = low FA and Red = high FA

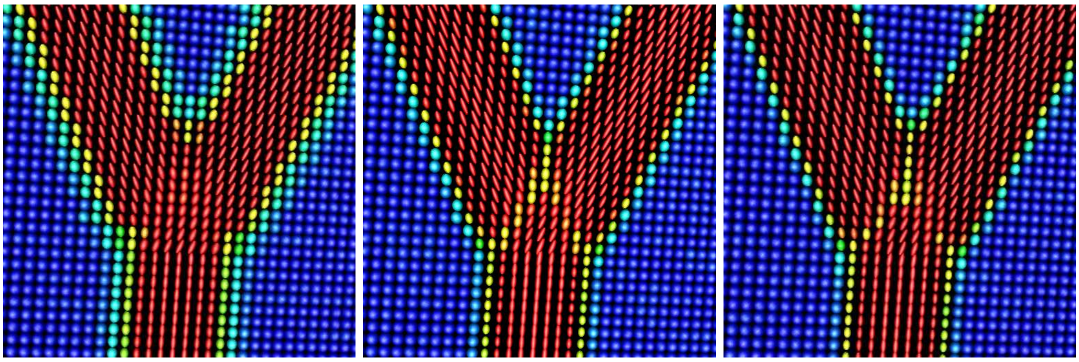


Figure 6.8: Results on denoising synthetic data. Left: Gaussian convolution (Equation 6.15). Middle: Nonlinear diffusion (Equation 6.18). Right: Anisotropic Riemannian filtering (section 6.2.2).

Synthetic Data

Experiment 1

In order to check the performance of our approach we generate a $50 \times 50 \times 50$ synthetic field of 3×3 tensors which roughly simulates a bifurcation of two fibers. In figure 6.7, we display a partial view of one slice in our volume, without noise on the left and with a low level of noise on the right. For this image, as well as any other tensor field presented in this section, the color code is related to the fractional anisotropy (FA) of the tensors with: Blue = low FA / Red = high FA. To generate the noise we use a generalization of the Gaussian distribution for samples belonging to $S^+(3)$. By using the algorithm proposed in [184] (see section 5.4.1), we can easily generate a set of random positive definite tensors with the desired mean and

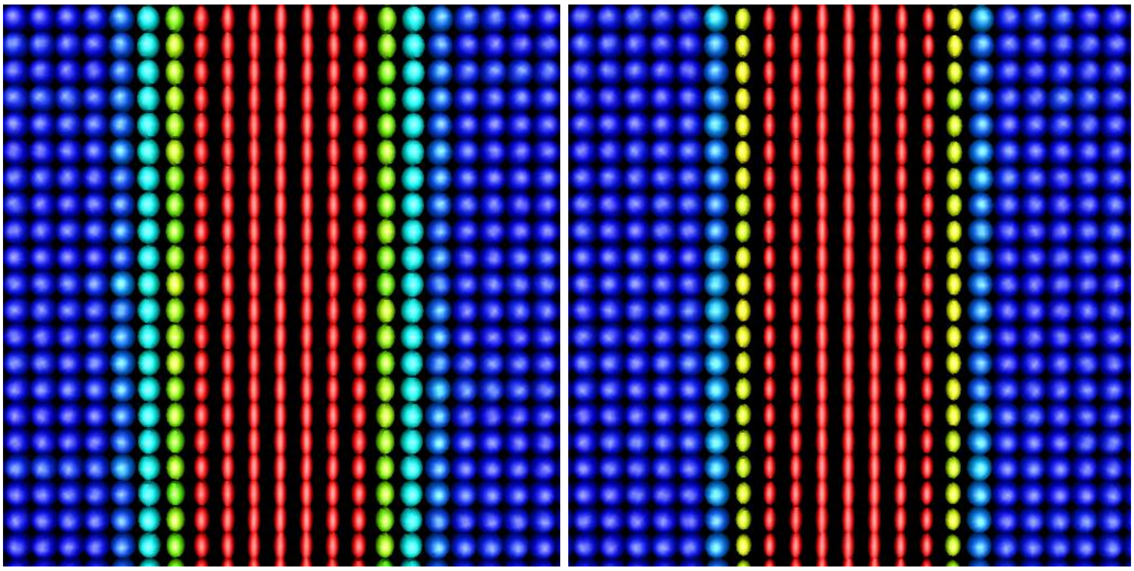


Figure 6.9: Detail of boundaries. Left: Gaussian filtering (Equation 6.15). Right: Anisotropic Riemannian filtering (section 6.2.2).

covariance matrix. This is a much more satisfying approach than, as usually done, simply building symmetric matrices with independent and identically normally distributed components and then enforcing their positiveness since this leaves no grasp on the actual distribution of the tensors. Moreover, the algorithm proposed in [184] is consistent with the parametric model for noise in DTI proposed in [21]. In this work, the authors proved that, assuming that the magnitude diffusion weighted images are Rician distributed, noise in diffusion tensor data within a voxel follows a 6-dimensional Normal distribution.

We then try to recover the original image from the noisy version. Figures 6.8 compares the outputs of the filtering schemes presented in this section from a qualitative point of view. The image on the left was obtained by convolving each component of the tensor with a Gaussian kernel ($\sigma = 1.5$). This is equivalent to the diffusion process presented in equation 6.15. In the middle we can see the best output we have obtained with the nonlinear diffusion process presented in equation 6.18 with time step 0.01 and 36 iterations. The result on the right uses our anisotropic Riemannian approach presented in section 6.2.2. We must point out here that the optimization of the diffusion time for the nonlinear diffusion scheme is clearly both critical and difficult to adjust. It is definitely a major limitation of this last approach. From figure 6.8, it is obvious that boundaries are better preserved by anisotropic methods than by the isotropic approach. To more clearly emphasize these different behaviors, figure 6.9 shows a detail of the output of the Gaussian filtering scheme and of the anisotropic Riemannian approach.

	Mean Error	Std Deviation	Max	Min
Anisotropic Riemannian Filtering	0.4216	0.1816	1.7076	0.1327
Nonlinear Diffusion	0.4240	0.1590	1.6562	0.1617
Gaussian Euclidean Filtering	0.7223	0.1471	1.7776	0.3827

Table 6.2: Statistics of the error measured following equation 6.19 for experiment 1 using the different regularization schemes proposed in this article.

	Mean Error	Std Deviation	Max	Min
Anisotropic Riemannian Filtering	0.5218	0.0712	1.0770	0.1651
Nonlinear Diffusion	0.5283	0.0409	0.9957	0.2044
Gaussian Euclidean Filtering	1.0542	0.0830	1.2246	0.5092

Table 6.3: Statistics of the error measured following equation 6.20 for experiment 1 using the different regularization schemes proposed in this article.

In order to quantify and compare the performances of the nonlinear diffusion with our Riemannian anisotropic smoothing method, we propose to measure the error at each voxel by using the squared geodesic distance between tensors:

$$Error(x) = \mathcal{D}^2(\Sigma(x), \hat{\Sigma}(x)) \tag{6.19}$$

where $\Sigma(x)$ is the original tensor field and $\hat{\Sigma}(x)$ is the filtered one. Table 6.2 shows some statistics on the error, where it can be seen that the Riemannian approach improves the behavior of the Euclidean counterparts. Since using the geodesic error metric may bias the comparison in favor of our approach, another error measure is also used in table 6.3. It shows the statistics of the Frobenius norm of the difference between $\Sigma(x)$ and $\hat{\Sigma}(x)$:

$$Error(x) = \operatorname{tr}\left((\Sigma(x) - \hat{\Sigma}(x))(\Sigma(x) - \hat{\Sigma}(x))^T\right) \tag{6.20}$$

From those last results, it is also clear that our approach provides the most accurate results.

Experiment 2

We now would like to evaluate the behavior of the different algorithms for a higher amount of noise. To that end, we generate a $32 \times 32 \times 32$ volume with tensors oriented along the vertical direction and introduce a $16 \times 16 \times 16$ cube with tensors oriented in the orthogonal direction, as shown in the top left image of figure 6.10. Then, we add a high level of Gaussian noise, resulting in the noisy tensor field in the top right image of figure 6.10. The bottom right image corresponds to our proposed anisotropic Riemannian filtering approach while the image on the left is the output of the nonlinear diffusion. Differences arise when comparing these two images. First of all, noise is better removed with the Riemannian approach than with the

	Mean Error	Std. Deviation	Max	Min
Anisotropic Riemannian Filtering	0.2572	0.1313	1.3065	0.0300
Nonlinear Diffusion	0.8284	0.1209	3.2551	0.2630

Table 6.4: Statistics of the error measured following equation 6.19 for experiment 2.

	Mean Error	Std. Deviation	Max	Min
Anisotropic Riemannian Filtering	0.6296	0.5166	4.4222	0.0600
Nonlinear Diffusion	2.1677	0.7970	91.2742	0.7367

Table 6.5: Statistics of the error measured following equation 6.20 for experiment 2.

nonlinear diffusion, where some misoriented tensors remain even after 536 iterations with a time step of 0.01, which are the parameters that provide the optimal response. Furthermore, the well-known swelling effect, due to a faster regularization of the eigenvalues, is observed for the nonlinear diffusion, whereas it is not noticeable for our Riemannian approach. This can be observed by looking at the colors of the tensors: First, we point out that the original tensors are blue because they are all identical, thus have the same FA (0.77), and our visualization software assigns the color associated to the lowest value in that case. But most importantly, we can see that the regularized tensor field obtained with our approach is more anisotropic (tensors are yellow and FA is around 0.75) than that obtained with the nonlinear diffusion (tensors are green and FA is around 0.60).

From a quantitative point of view, we measure the error between the original and the regularized images following equation 6.19 and equation 6.20. As shown in table 6.4, mean squared geodesic distance is much lower for the anisotropic Riemannian approach than for the nonlinear diffusion. Table 6.5 shows the error statistics based on the Frobenius norm. It is also obvious that the accuracy of the Riemannian approach is much better than that achieved with the nonlinear diffusion.

Real DTI Data

For experiments with real data, diffusion weighted images were acquired on a 3 Tesla MEDSPEC 30/80 AVANCE (Bruker) at the Centre IRMf de Marseille, France, using a quadrature bird-cage head coil. We used 12 gradient directions and a b-value of 1000 s/mm^2. Acquisitions were repeated 8 times for each direction to ensure a good signal-to-noise ratio. The sequence parameters were chosen as follows: $\Delta = 38.5$ms, $\delta = 21.6$ms, TE$= 79$ms, TR$= 10000$ms and voxel size was $2 \times 2 \times 2$ mm^3. Diffusion tensors were estimated by the robust gradient descent algorithm preserving their symmetry and positive definiteness, as presented in section 6.1.2 (see also [184]).

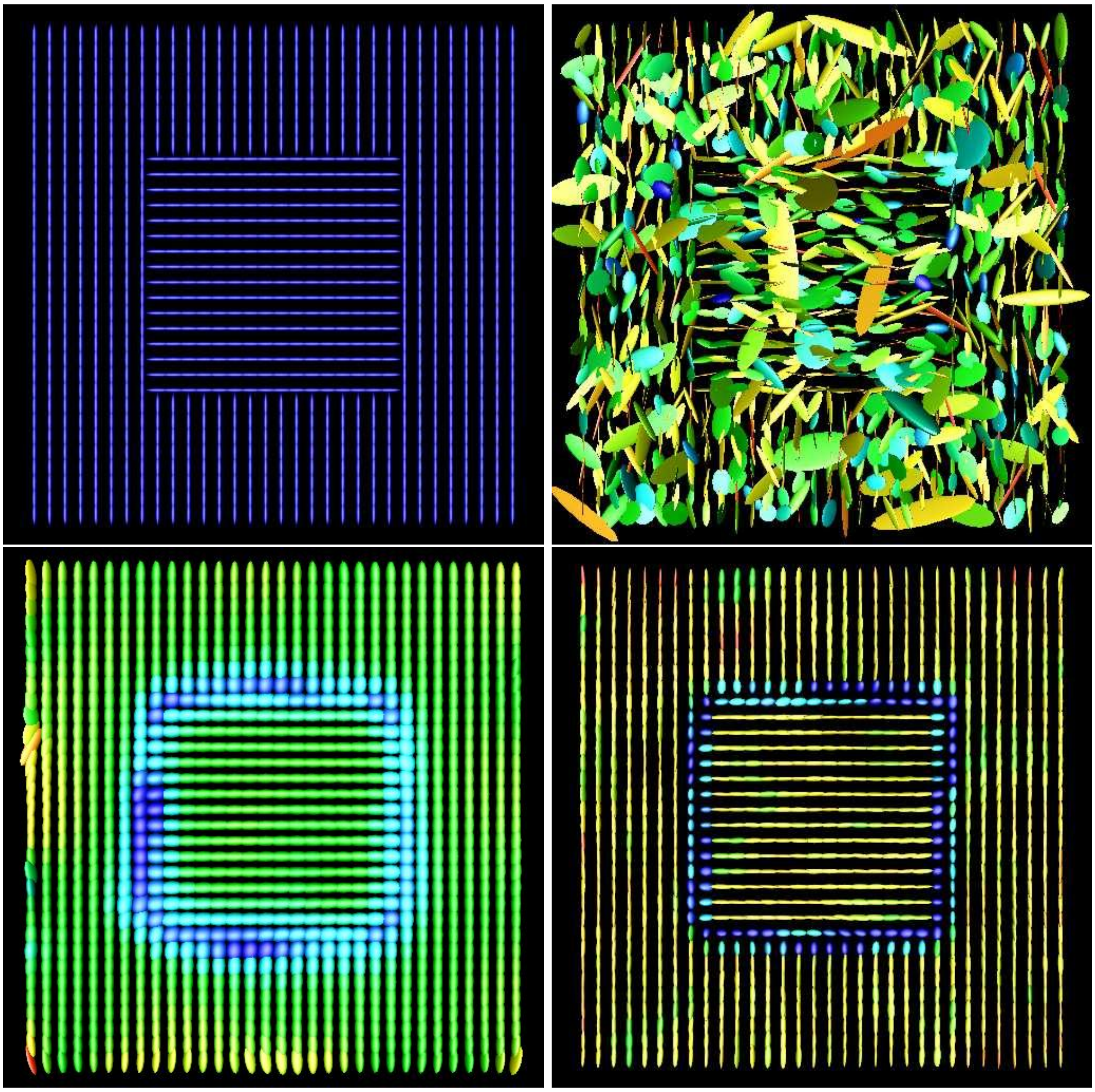


Figure 6.10: Results in presence of a high level of noise. Top Left: Original Image. Top Right: Noisy Image. Bottom Left: Nonlinear diffusion (Equation 6.18). Bottom Right: Anisotropic Riemannian filtering (section 6.2.2).

Figure 6.11 (top left) shows the estimated diffusion tensors. On the top right, we display the regularized image using the Riemannian filtering approach, while bottom images are regularized using nonlinear diffusion, both with 10 iterations, but different time steps: 0.001 on the left and 0.01 on the right. If we analyze the different structures found on this axial slice, we can see that tensors orientation within the splenium of the corpus callosum (CC(S) in the image) and in the genu of the corpus callosum (CC(G) in the image) is more coherent with our Riemannian filtering scheme. Anisotropy in these areas is also better preserved than with the nonlinear diffusion case, which yields blurred areas most likely because of the properties of the Euclidean gradient. In addition, the ventricles areas (VE in the image), since they

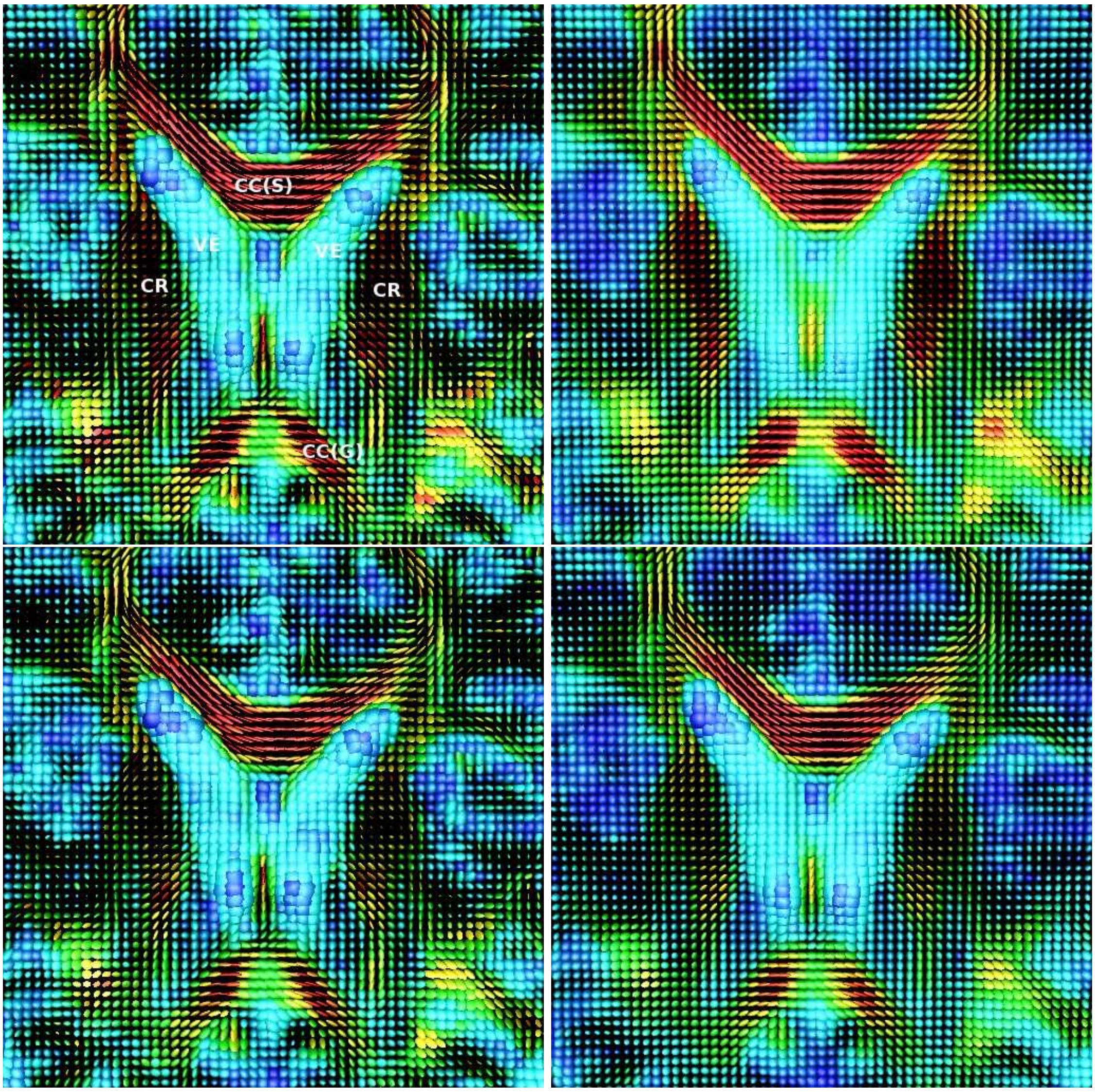


Figure 6.11: Results on real DTI data. Top Left: Original DTI data. CC(S): Splenium of the Corpus Callosum. CC(G): Genu of the Corpus Callosum. VE: Ventricles. CR: Corona Radiata. Top Right: anisotropic Riemannian filtering. Bottom Left: nonlinear diffusion (time step 0.001, 10 iter.). Bottom Right: nonlinear diffusion (time step 0.01, 10 iter.)

are mainly homogeneous structures, are better regularized with our approach, as inhomogeneities do not disappear with nonlinear diffusion. Finally, the corona radiata (CR in the image) is well preserved in our approach while it is completely smoothed away from the image with longer diffusion time. In figure 6.12 we show a complete human brain volume where fiber orientation is color coded as follows: Red: Right-Left / Green: Anterior-Posterior / Blue: Inferior-Superior. Original data is shown at the top of the image, while the filtered version is shown at the bottom. An axial slice from that volume is presented in figure 6.13, where we compare original data (on the

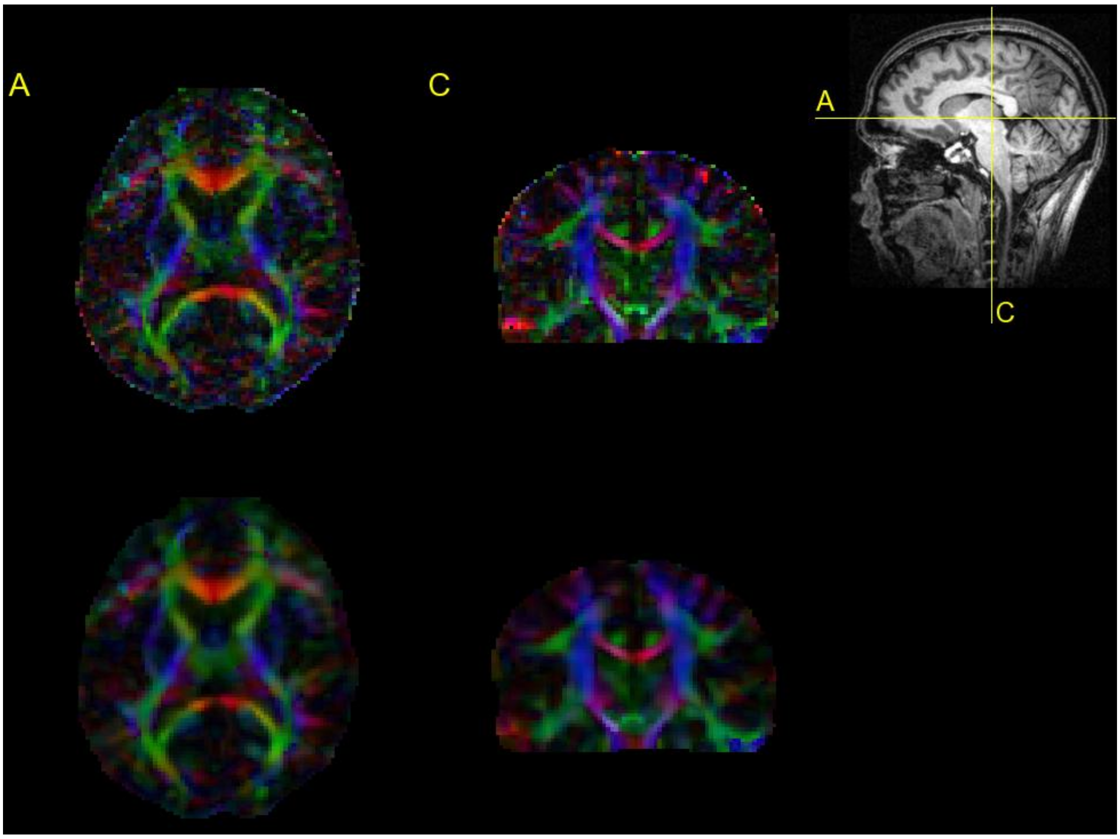


Figure 6.12: Real data volume of a human brain.

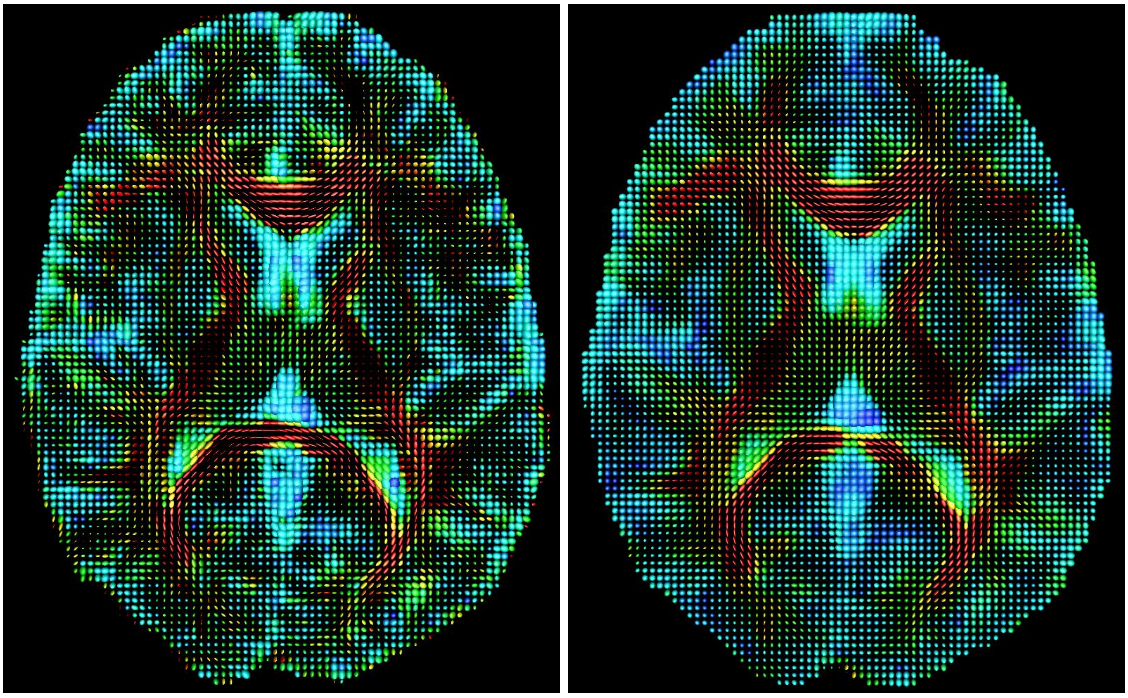


Figure 6.13: Raw tensor field (left) compared to regularized tensor field (right)

left) with the filtered version (on the right) using our approach. It is clear tensors belonging to particular fiber bundles are more coherent while the interfaces between the various tracts have been preserved.

6.2.4 Conclusion

In this section, we have presented a novel differential geometrical approach for the anisotropic regularization of tensor fields, seen as fields of multivariate normal distributions. The anisotropic behavior is introduced through the gradient magnitude, simply computed by using the geodesic distance between distributions. This gradient magnitude, contrary to the Euclidean one, benefits from the affine invariance of the underlying distance. It is not sensitive to the scale of the tensors which allows to better discriminate tissues interfaces and regularize misoriented tensors because of the noise as shown in our numerical experiments. Our filtering scheme is also compared to nonlinear diffusion of matrix-valued data to point out its added value over the PDE-based approaches and to show that our approach can easily solve some of their limitations, yielding better results on synthetic and real DTI data. Regularization may be important to enhance the results of subsequent processing like tractography or segmentation. In the next chapter, we will address one of the most important application of DTI: the mapping of anatomical connections in the human cerebral white matter. We will tackle this task by recasting it into a front propagation problem in $\mathbb{R}^3$ equipped with metric derived from the diffusion tensor itself.

CHAPTER 7

MAPPING THE ANATOMICAL CONNECTIVITY OF THE HUMAN BRAIN

Contents

OVERVIEW

In this chapter, we introduce an original approach for the cerebral white matter connectivity mapping from DTI. Our method relies on a global modeling of the acquired MRI volume $\Omega \subset \mathbb{R}^3$ as a Riemannian manifold $\mathcal{M}$ whose metric g directly derives from the diffusion tensor. We want to make it clear that the metric g used in this chapter has absolutely nothing to do with the metric derived from the Fisher information matrix and used, up to now, to compare diffusion tensors. In the following, these tensors will be used to derive a "diffusion metric" (see [83]) and to measure physical three-dimensional distances between different locations of a brain diffusion tensor image. The key concept of this chapter is the notion of geodesic distance that will allow us to find optimal paths in the white matter, approximating neural fiber bundles. The geodesic distance function can be seen as the solution of two theoretically equivalent but, in practice, very different problems. They both derive from front propagation modeling and are equivalent formulations of the so-called Eikonal equation:

- An initial value problem of the form $\frac{\partial \psi}{\partial t} + F|\nabla \psi| = 0$. F is the front propagation speed and this point of view is intrinsically *dynamic*. It will be implemented within the level set framework.
- A boundary value problem of the form $|\nabla \phi| = \frac{1}{F}$ which is, on the contrary, intrinsically *stationary*. It will be implemented by using the fast marching method.

As we will show, these two approaches have very different properties which make them more or less adequate for our problem and more or less computationally efficient. The dynamic formulation is quite easy to implement but has many drawbacks. On the contrary, the stationary formulation relies on control theory and is much more tedious to implement. However it has many virtue (like computational time and robustness) which makes it more suitable for our connectivity mapping problem. Once the geodesic distance has been computed, it is necessary to resort to a back-propagation algorithm on the gradient of this distance to recover shortest paths. As we will see, the stationary approach naturally yields this vector field without having to differentiate the distance function, hence its robustness. Finally, we will present different possible measures of connectivity, reflecting the degree of connectivity between different regions of the brain. A natural connectivity measure will be derived from the stationary formulation of the problem. We will illustrate these notions on synthetic and real DTI datasets.

Keywords: Brownian motion, diffusion process, control theory, partial differential equations, Riemannian manifold, Hamilton-Jacobi-Bellman equations, level set, fast marching method, anisotropic Eikonal equation, intrinsic distance function, brain connectivity mapping

7.1 INTRODUCTION

We recall that diffusion tensor (DT) imaging models the probability density function of the three-dimensional molecular motion, at each voxel of a DT image, by a local Gaussian process whose covariance matrix is precisely given by the diffusion tensor. Among other applications, diffusion tensor imaging (DTI) is extremely useful to estimate the anatomical connectivity of the human brain.

Following [212], various local approaches, often named deterministic tractography, have been proposed to tackle this problem. They are based on line propagation techniques and rely on the fact that the eigenvector of the diffusion tensor associated to the major eigenvalue, provides a relatively accurate estimate of the fibers' orientation at each voxel. These methods (see for instance [158],[26], [297] , [213], [174]) may be refined to incorporate some natural constraints such as regularity or local uncertainty and avoid being stopped in regions of low anisotropy. All these efforts aim to overcome the intrinsic ambiguity of diffusion tensor data arising from partial volume effects at locations of fiber crossings [3]. They provide relatively accurate models of the white matter macroscopic bundles.

More recent work can be divided into approaches based on Bayesian models, diffusion simulation and geometric methods, the latter being essentially based on front-propagation techniques. These methods are both more robust to noise and partial volume effects than deterministic tractography, and naturally yield probability/scalar measures which can be used to evaluate the degree of connectivity between voxels.

In [31], [230], [232], [38], [122, 121] stochastic tractography algorithms were introduced by modeling the uncertainty of the local fiber orientation. Through uncertainty propagation, they can also provide a powerful means to evaluate the probability of connection between points of the white matter. However, the intrinsic drawback of these methods is their computational complexity since it is necessary to resort to Markov Chain Monte Carlo methods or, as in [121], to evaluate probability density functions at enough locations of the space of interest. Other probabilistic approaches like [169], [173], [201], [136], [172] use perturbation methods, random walk ... etc to assess the uncertainty in tractography.

Diffusion simulation approaches like [28], [218], [68], [159], [318] or [135] use the full diffusion tensor to simulate a diffusion process or a fluid flow. The resulting concentration or flow maps can be used to evaluate some degree of connectivity between regions of interest.

Finally, geometric methods use either level set (see [223]) methods [218], [183], Fast Marching Methods (see [286, 268]) [234, 233], [246, 247] or iterative sweeping techniques [147] to evolve a front on the basis of the diffusion tensor directional information. As described in [281] and [50], it is possible to adapt the level set based front propagation technique to take advantage of the information provided by high angular resolution diffusion MRI (HARDI). In [241, 240], the authors also proposed to use curve evolution techniques to recover fibers from HARDI data. As we will see in the following, level set approaches to connectivity mapping tend to be somewhat inefficient since, even with a narrow-band implementation, the number of points where the evolution speed has to be evaluated greatly increases as the surface grows. We will show that this class of methods is also prone to interpolation errors at the boundary of the domain. For our brain connectivity problem, this may lead to erroneous connections in highly convoluted areas and will lead us to consider another approach, namely the stationary formulation of the Eikonal equation.

Contributions of this chapter:

Our contribution is threefold:

1. First, we describe how to use the diffusion tensor as a Riemannian metric. We clarify the link between Brownian motion and diffusion MRI and expose how the knowledge of the diffusion properties of water molecules on a manifold $\mathcal{M} = (\mathbb{R}^3, g)$ is sufficient to infer its geometry.

2. Next, we present the dynamic formulation of the anisotropic Eikonal equation (within the level set framework), show how to solve it and to estimate neural fiber bundles by back-propagation.

3. Finally, we demonstrate that we can greatly improve the computational time and the robustness of the previous method by recasting the intrinsic geodesic distance computation into a stationary problem (with a fast marching implementation). In fact, it is possible to solve, quickly and simultaneously, for the geodesic distance, the optimal vector field (optimal dynamics) corresponding to the geodesics velocities and the statistics, along those curves, of a local connectivity measure. To our knowledge, the proposed GCM (for "Geodesic Connectivity Mapping") algorithm is faster than any other existing method.
   Also, contrary to other approaches, we simply solve the anisotropic Eikonal equation and do not resort to any anisotropy related parameter to constrain the front propagation. Moreover, our algorithm can naturally work within a mask of the white matter (accurately obtained by segmentation of a high-resolution anatomical MRI for instance). As we will show, this is crucial for the applications of interest since we must strictly respect the geometry of the cortical foldings or white matter / cerebrospinal fluid (CSF) interface to recover meaningful connections.
   For a region of interest x_0 (i.e. a point of the white matter), our GCM method generates statistics of a local connectivity measure along the geodesics linking

x_0 to other locations of the brain. This can be used to discriminate likely and unlikely connections.

Organization of this chapter:

Section 7.2 justifies the use of the diffusion tensor as a metric. Next, section 7.3 details the dynamic formulation of our geodesic distance computation problem and proposes a numerical scheme within the level set framework. Section 7.4 reformulates the anisotropic Eikonal equation as an optimal control problem which leads, in section 7.5, to a fast numerical algorithm for the distance function approximation. We finally introduce a general local connectivity measure whose statistics along the optimal paths may be used to evaluate the degree of connectivity of any pair of voxels. All those quantities can be computed simultaneously in a Fast Marching framework, directly yielding the connectivity maps. We illustrate our techniques by showing results on real and synthetic datasets in section 7.6.

7.2 WHITE MATTER AS A RIEMANNIAN MANIFOLD

Diffusion tensor, as thermal or electrical conductivity tensors, belongs to the broader class of general effective property tensors and is defined as the proportionality term between an averaged generalized intensity B and an averaged generalized flux F. In our particular case of interest, B is the water molecules concentration gradient ∇C and F is the mass flux J such that Fick's (first) law holds: $J = -\mathbf{D}\nabla C$. By considering the conservation of mass, the general diffusion equation (Fick's second law) is readily obtained:

$$\frac{\partial C}{\partial t} = \nabla.(\mathbf{D}\nabla C) \tag{7.1}$$

In an isotropic and homogeneous medium, equation 7.1 boils down to the heat equation ($\mathbf{D}$ is the identity matrix) whose solution is well-known: a Gaussian distribution. In anisotropic tissues (like the cerebral white matter), water molecules motion varies in direction depending on obstacles (such as axonal membranes). We recall that the symmetric and positive definite tensor $\mathbf{D}$ has been related [105] to the root mean square of the diffusion displacement $\mathbf{R}$ during time τ by $\mathbf{D} = \frac{1}{6\tau}\langle \mathbf{R}\mathbf{R}^T\rangle$ (see section 3.3.2) and is thus an effective means to characterize the diffusion properties of anisotropic media.

For an unbounded anisotropic homogeneous medium, the minimal fundamental solution of equation 7.1 with initial concentration $\lim_{t\to 0} p(\mathbf{r}|\mathbf{r}_0, t) = \delta(\mathbf{r} - \mathbf{r}_0)$ is also well-known and expressed as:

$$p(\mathbf{r}|\mathbf{r}_0, \tau) = \frac{1}{\sqrt{(4\pi\tau)^3|\mathbf{D}|}} \exp\left(-\frac{(\mathbf{r}-\mathbf{r}_0)^T\mathbf{D}^{-1}(\mathbf{r}-\mathbf{r}_0)}{4\tau}\right)$$

However, when dealing with anisotropic and inhomogeneous media (like the cerebral white matter), explicit derivation of p becomes non-trivial since the tensor $\mathbf{D}$ depends on its spatial location $x \in \Omega$ and will be noted D_x now on. It turns out that, in fact, this explicit derivation is not necessary. It is indeed possible to reformulate equation 7.1 by considering the diffusion process to occur on a Riemannian manifold $\mathcal{M}$ and not a linear space as up to now. The basic idea is to rely on the metric (varying from point to point) to naturally take into account the space anisotropy and inhomogeneity (ie. the spatial variation of D_x). Intuitively, we transform an anisotropic and inhomogeneous diffusion on a linear space $\mathbb{R}^n$ into an isotropic and homogeneous diffusion on a nonlinear space $\mathcal{M}$.

The counterpart of the Laplacian on a Riemannian manifold is classically defined as the divergence of the gradient. It is known as the Laplace-Beltrami operator which, for a scalar function f, writes:

$$\Delta_{\mathcal{M}} f = \text{div}(\text{grad} f)$$

The intrinsic divergence and gradient operators grad and div are defined as follows in local coordinates:

$$(\text{grad} f)^i = g^{ij} \frac{\partial f}{\partial \mathrm{x}^j} \quad \forall i, j = 1, ..., n \tag{7.2}$$

and

$$\text{div} X = \frac{1}{\sqrt{|\det(g)|}} \frac{\partial}{\partial \mathrm{x}^i} \left(\sqrt{|\det(g)|} X^i \right) \tag{7.3}$$

where g^{ij} denotes the components of the inverse of the metric g, X is a vector field on $\mathcal{M}$ and Einstein notation has been used. Putting things together, the Laplace-Beltrami operator can be written in local coordinates as:

$$\Delta_{\mathcal{M}} f = \frac{1}{\sqrt{|\det(g)|}} \frac{\partial}{\partial \mathrm{x}^i} \left(\sqrt{|\det(g)|} g^{ij} \frac{\partial f}{\partial \mathrm{x}^j} \right) = \frac{1}{\sqrt{|\det(g)|}} \frac{\partial}{\partial \mathrm{x}^i} \left(\sqrt{|\det(g)|} G^{-1} \nabla f \right)$$

Comparing this expression with equation 7.1, we can see that the diffusion tensor **D** plays the same role than the inverse of the Riemannian metric G^{-1}. This relation is actually at the basis of many works on stochastic processes on Riemannian manifolds such as [85] or [83] and has been proposed, in the context of DTI, by O'Donnell *et al.* [218]. In the following, we will use the inverse of the diffusion tensor (note that inversion will not always be required in practice) as a Riemannian metric for the cerebral white matter.

7.3 A LEVEL SET FORMULATION FOR THE GEODESIC DISTANCE (THE DYNAMIC PERSPECTIVE)

We are now concerned with the computation of the distance function u from a closed, non-empty subset K of the 3-dimensional Riemannian manifold $\mathcal{M} = (\mathbb{R}^3, g)$. In the remaining, K will be restricted to a single point x_0 or to several voxels (ie. a region of interest). We will formulate everything in term of K since considering the distance to a larger subset of $\mathcal{M}$ can be interesting.

7.3.1 Distance on Riemannian Manifolds

Let us now further discuss the notion of distance function on a Riemannian manifold. Given two points $x, y \in \mathcal{M}$, we consider all the piecewise differentiable curves joining x to y. For a Riemannian manifold, such curves do exist and we have the

Definition 7.3.1.1. *The distance $\phi(x, y)$ is defined as the infinimum of the lengths of the $\mathcal{C}^1$ curves starting at x and ending at y.*

The length $\mathcal{L}$ of a curve is defined by equation 4.2.2. Moreover, we also have (see [204]) the

Proposition 7.3.1.1. *If $x_0 \in \mathcal{M}$, the function $u : \mathcal{M} \to \mathbb{R}$, called distance function and given by $u(x) = \phi(x, x_0)$, is continuous on $\mathcal{M}$ but in general it is not everywhere differentiable.*

We adopt the notation $u(x)$ for the clarity of expressions but recall that u depends of course on x_0. We consider a general Hamilton-Jacobi partial differential equation with Dirichlet boundary conditions

$$\begin{cases} H(x, \nabla u(x)) = 0 & \text{on } \mathcal{M} \setminus K \\ u(x) = u_0(x) & \text{when } x \in K \end{cases}$$

where u_0 is a continuous real function on K and the Hamiltonian $H : \mathcal{M} \times T^\star\mathcal{M} \to \mathbb{R}$ is a continuous real function. We make the assumption that $H(x, .)$ is convex and we set $u_0(x) = 0 \; \forall x \in K$.

We denote by $|v|$ the magnitude of a tangent vector v, defined as $\sqrt{g(v, v)}$. In matrix notation, by forming $G = \{g_{ij}\}$ the metric tensor, this writes $\sqrt{v^T G v}$. Then, by setting $H(x, p) = |p| - 1$, we will work on the following theorem (for details on viscosity solutions on a Riemannian manifold, we refer to [204])

Theorem 7.3.1.1. *The distance function u is the unique viscosity solution of the Hamilton-Jacobi problem*

$$\begin{cases} |\mathbf{grad}\, u| = 1 & \textit{on } \mathcal{M} \setminus K \\ u(x) = 0 & \textit{when } x \in K \end{cases} \tag{7.4}$$

in the class of bounded uniformly continuous functions.

This is the well-known Eikonal equation on the Riemannian manifold $\mathcal{M}$. The viscosity solution u at $x \in \mathcal{M}$ is the minimum time $t \geq 0$ for any curve γ to reach a point $\gamma(t) \in K$ starting at x with the conditions $\gamma(0) = x$ and $|\frac{d\gamma}{dt}| \leq 1$. u is the value function of the minimum arrival time problem. This will enable us to solve equation 7.4 as a dynamic problem and thus to take advantage of the great flexibility of level set methods. On the basis of [222], [267], [283] and [66], we reformulate equation 7.4

by considering u as the zero level set of a function ψ and requiring that the evolution of ψ generates u so that

$$\{\psi(x,t)=0\} \Leftrightarrow \{u(x)=t\} \tag{7.5}$$

Osher ([222]) showed by using Theorem 5.2 from [66] that, under the hypothesis that the Hamiltonian H is independent of u, the level set generated by 7.5 is a viscosity solution of 7.4 if ψ is the viscosity solution of

$$\begin{cases} \frac{\partial\psi}{\partial t} + F(t,x,\nabla\psi(t,x)) = 0 \quad \forall t>0 \\ \psi(x,0)=\psi_0(x) \end{cases} \tag{7.6}$$

provided that the speed F is strictly positive and does not change sign. This is typically the case for our anisotropic Eikonal equation where

$$F(t,x,\nabla\psi) = H(t,x,\nabla\psi)+1 = |\text{grad}\psi|.$$

To find our solution, all we need to do is thus to evolve $\psi(x,t)$ while tracking, for all $x\in\Omega$, the time $\bar{t}$ when it changes sign. Now we have to solve 7.6 with $F(t,x,\nabla\psi) = |\text{grad}\psi|$. We recall that for any function $f\in\mathbb{F}$, where $\mathbb{F}$ denotes the ring of smooth functions on $\mathcal{M}$, the metric tensor G and its inverse define isomorphisms between tangent vectors (in $T\mathcal{M}$) and 1-forms (in $T^\star\mathcal{M}$). In particular, we have seen earlier that the gradient operator is defined as grad$f = G^{-1}\nabla f$ where ∇f denotes the first-order differential of f. It directly follows that

$$|\text{grad}\psi| = \sqrt{g(\text{grad}\psi,\text{grad}\psi)} = \left(g_{ij}\frac{\partial\psi}{\partial \mathbf{x}^l}g^{li}\frac{\partial\psi}{\partial\mathbf{x}^k}g^{kj}\right)^{1/2} = \left(\frac{\partial\psi}{\partial\mathbf{x}^k}\frac{\partial\psi}{\partial\mathbf{x}^l}g^{kl}\right)^{1/2} \tag{7.7}$$

and we now present the numerical schemes used to estimate geodesics as well as the viscosity solution of

$$\psi_t + |\text{grad}\psi| = 0 \tag{7.8}$$

7.3.2 Numerical Schemes

Numerical approximation of the hyperbolic term in 7.8 is now carefully reviewed on the well-known basis of available schemes for hyperbolic conservative laws. We seek a three-dimensional numerical flux approximating the continuous flux $|\text{grad}\psi|^2$ and that is consistent and monotonous so that it satisfies the usual jump and entropy conditions and converges towards the unique viscosity solution of interest. References can be found in [189]. On the basis of the Engquist-Osher flux [267] and the approach by Kimmel-Amir-Bruckstein for level set distance computation on 2D manifolds [165], we propose the following numerical flux for our quadratic Hamiltonian $\nabla\psi^T G^{-1}\nabla\psi$:

$$|\text{grad}\psi|^2 = \sum_{i=1}^{3} g^{ii}(\max(\nabla^-_{\mathbf{x}^i}\psi,0)^2 + \min(\nabla^+_{\mathbf{x}^i}\psi,0)^2) + \\ \sum_{\substack{i,j=1\\ i\neq j}}^{3} g^{ij}\text{minmod}(\nabla^+_{\mathbf{x}^i}\psi,\nabla^-_{\mathbf{x}^i}\psi)\text{minmod}(\nabla^+_{\mathbf{x}^j}\psi,\nabla^-_{\mathbf{x}^j}\psi)$$

where the $\nabla^{\pm}_{\mathbf{x}^i}\psi$ are the forward/backward approximations of the gradient in x_i. First of all, we point out that, because of equation 7.7, we only have to use the inverse of the metric, ie. the diffusion tensor, and not the metric itself. Consequently, we do not have to invert the diffusion tensor, which could have been a source of instability of the numerical scheme. Moreover, we have also experimented with higher order approximation schemes in order to increase the accuracy of the method. This is done by introducing WENO schemes in our numerical flux instead of the upwind gradients. WENO schemes are based on ENO (essentially non-oscillatory) schemes, which were first introduced by Harten, Osher, Engquist and Chakravarthy in [138] in the form of cell averages. They basically use a polynomial approximation of the derivatives and avoid oscillations when a shock is detected. WENO schemes of Liu, Osher and Chan [194], instead of approximating the numerical flux using the best candidate stencil, use a weighted convex combination of all the candidates stencils. We now quickly describe a very well-known method that speeds up the estimation of the distance function u.

Narrow Band Method: In order to overcome the high computational overhead of the front propagation approach ($\mathcal{O}(N^3)$ if N denotes the number of grid points along a side), the narrow band method relies on the fact that it is sufficient to compute the level set function only in a small neighborhood (at a distance δ) around its zero level set because only this zero level set is physically meaningful (representing the interface). It was introduced by Chopp in [67]. In that case, the complexity drops to $\mathcal{O}(kN^2)$ in three dimensions, where k is the number of voxels in the band. In practice, at an iteration n, only the points in the band are updated and other points are kept intact. When the front moves near to the edge of the band, the calculation is stopped and a new band is initialized with the zero level set interface boundary at the center.

We finally describe the method used for the computation of geodesics, in order to approximate paths of diffusion on $\mathcal{M}$ eventually corresponding to neural fibers tracts. Geodesics are classically obtained by performing back-propagation from a given point x to the source x_0, in the direction provided by the opposite of the distance function gradient. Our problem of interest thus consists in starting from a given voxel of the white matter and in computing the optimal pathway in term of the distance u until x_0 is reached. The simplest way to infer a geodesic path from u is to backtrace on the manifold itself by solving the ordinary differential equation

$$\frac{d\gamma(t)}{dt} = -\mathbf{grad}\, u = -G^{-1}\nabla u \tag{7.9}$$

where γ is the sought geodesic parameterized by t. $\mathbf{grad}\, u$ involves the Riemannian metric as well as the differential of the distance function ∇u, evaluated with appropriate finite difference schemes. We have experimented with Euler, 2^{nd} and 4^{th} order

Runge-Kutta integration methods with sensibly better results obtained by higher order Runge-Kutta schemes but noticeable computational overheads. When integrating within a voxel, trilinear interpolation of the vector field $\operatorname{grad} u$ is performed by using the 8 available values in the neighborhood (ie. the corners of the voxel of interest).

While a given geodesic is estimated, we also compute on-the-fly statistics of a confidence measure $\mathcal{C}$ along this curve. Typically, as proposed in [234] we take $\mathcal{C}$ to be the absolute value of the Euclidean inner product between the normalized geodesic velocity and the principal diffusion vector of the interpolated diffusion tensor field at location $\gamma(t)$:

$$\left| \langle \frac{\dot{\gamma}(t)}{|\dot{\gamma}(t)|_E}, \mathbf{u}^1 \rangle \right| \tag{7.10}$$

This quantity reflects how well the local orientation of the geodesic matches the directional information provided by the diffusion tensor image. In practice, we can use the minimum of this quantity along γ as a “worst case” value (as close to 1 as possible), its average (as close to 1 as possible) and standard deviation (as low as possible) in order to discriminate geodesics that are likely to represent true neural bundles from false connections.

Numerical experiments for the distance function computation and the geodesics estimation will be presented in section 7.6. An application of this method to quantify the likelihood of connection between the putamen and the motor areas of the human brain will be presented in chapter 10. As we will see in the following, the method that was just described, though quite easy to implement, presents some drawbacks in term of efficiency and robustness and they will be detailled in section 7.6. We will now introduce the counterpart of the dynamic formulation of the Eikonal equation, namely its stationary formulation. This approach is based on notions from control theory, which we introduce in the next section before proposing the associated algorithm.

7.4 FROM GEOMETRY TO CONTROL THEORY

As in the previous section and in [218] [183], the white matter is interpreted as a Riemannian manifold and the inverse of the diffusion tensor D_x^{-1}, at location x, provides the Riemannian metric, which in turn determines white matter fibers as geodesic paths.

7.4.1 Overview

We remind the basic definition of geodesics for convenience [97] and to clarify the notations.

Definition 7.4.1.1. *Let $\mathcal{M}$ be a Riemannian manifold. Let $x, y \in \mathcal{M}$. The geodesic connecting x to y is the curve γ_0 which minimizes the arc length, i.e.*

$$\gamma_0 = \arg\min_{\gamma \in \Gamma x,y} \int_0^{T_{xy}} |\gamma'(t)|_R dt$$

where $\Gamma x, y$ is the set of curves $\gamma : [0, T_{xy}] \to \mathcal{M}$ such that $\gamma(0) = x$, $\gamma(T_{xy}) = y$ and $|\gamma'(t)|_R = 1$, with $|.|_R$ the norm associated to the Riemannian metric.

In section 7.2, we showed that the appropriate metric to our problem is the one associated to the norm $|.|_R$ defined by $|x|_R = \sqrt{x^T D_x^{-1} x}$, where D_x is the symmetric positive definite 3×3-matrix given by the measured diffusion tensor. Let us also denote by A_x the (symmetric positive definite) square root matrix of D_x and with $|.|_E$ the Euclidean norm. Let us note that we trivially have

$$|x|_R = |A_x^{-1} x|_E$$

Here, rather than interpreting the problem in terms of Riemannian geometry, we adopt an optimal control point of view. The two interpretations are equivalent, but focus on different aspects of the problem. In the Riemannian setting, the emphasis is on the description of the geometry and in particular on the geodesics. In the optimal control interpretation, the emphasis is on the optimal control which coincides here with the intrinsic gradient of the distance function and with the vector field tangent to the geodesics. Also, from the computational point of view, the intrinsic gradients are much more important and relevant than the geodesics by themself (see next sections). In particular, the intrinsic gradients are fundamental for numerically estimating the connectivity measure. Moreover the geodesics can be directly computed from them.

Another advantage of the optimal control interpretation is the relaxation of the dependency with respect to the specific geometry of the manifold. In the control interpretation, all the objects are governed by the Euclidean metric.

Finally, the white matter is an open subset of $\mathbb{R}^3$. If one forsakes the specific metric attached to the white matter, that is to say, if one interprets the metric in another way (i.e. not as a metric), then the representation of the white matter as a manifold is unnecessary. Also, in the optimal control interpretation, we do not need to deal with manifold [298, 146]. We will come back to the advantages of this interpretation at the end of this section.

7.4.2 Optimal Control Problems

In this paragraph, we briefly state the determinist continuous optimal control problems on open subset of $\mathbb{R}^n$ [19]. We also describe the links between the introduced mathematical objects and the brain connectivity mapping problem.

The problem is the following: we want that a particle reaches a target. We can control the particle thanks to a “control”. The problem consists then in finding the sequence of controls which minimizes a certain cost (which depends on the position of the particle and on the choice of the value of the control).

Let a domain Ω be a subset of $\mathbb{R}^n$. Ω is the space in which the particle can move. We consider a target $\mathcal{T} \subset \Omega$. Specifically in our problem, the domain $\Omega \subset \mathbb{R}^3$ represents the white matter; the target $\mathcal{T}$ is the point of interest x_0, origin of the distance function.

We consider the set $\mathcal{A}$ (compact subset of $\mathbb{R}^m$) of admissible controls a ($a \in \mathcal{A}$). We call *control function*, a function $\alpha(.) : \Omega \to \mathcal{A}$. In the following, the functions $\xi(.) : \mathbb{R}^+ \to \Omega : t \mapsto \xi(t)$ are candidate trajectories for the particle. The control functions $\alpha(.)$ control the dynamics of the trajectories of the particles. This control is done through a vector field

$$f : \Omega \times \mathcal{A} \to \mathbb{R}^n : (x, a) \mapsto f(x, a).$$

f is called *dynamics*. Mathematically, we pose

$$\xi'(t) = f(\xi(t), \alpha(\xi(t))),\ t > 0.$$

Under some regularity assumptions, to each control function α and $x \in \Omega$, we can associate a single trajectory $\xi_{x,\alpha}(t) \in \Omega$ following the dynamics $\xi'(t) = f(\xi(t), \alpha(\xi(t))),\ t > 0$ and s.t. $\xi(0) = x$, imposed by the control α, see [19].

Now, let us define a (local) *cost*

$$l : \Omega \times \mathcal{A} \to \mathbb{R} : (x, a) \mapsto l(x, a).$$

To each control function α and $x \in \Omega$, we can then associate a *global cost*: the integral of the (local) cost along the associated trajectory $\xi_{x,\alpha}(.)$, i.e.

$$\int_0^{T_{x,\mathcal{T},\alpha}} l(\xi_{x,\alpha}(t), \alpha(\xi_{x,\alpha}(t)))dt, \tag{7.11}$$

where $T_{x,\mathcal{T},\alpha}$ is the first time for which the trajectory $\xi_{x,\alpha}$ (controlled by the dynamics f) reaches the target $\mathcal{T}$. Under some regularity assumptions, one can prove that there exists a control function α^* (the *optimal control*) such that for all x, the global cost (7.11) is minimal [19]:

$$\alpha^* \overset{\text{def}}{=} \arg\max_{\alpha} \int_0^{T_{x,\mathcal{T},\alpha}} l(\xi_{x,\alpha}(t), \alpha(\xi_{x,\alpha}(t)))dt, \tag{7.12}$$

We then denote the *optimal trajectory* starting from x by

$$\xi_x^* \overset{\text{def}}{=} \xi_{x,\alpha^*},$$

and the *optimal dynamics* at x by

$$f_x^* \overset{\text{def}}{=} f(x, \alpha^*(x)).$$

In our specific problem, the optimal trajectories coincide with the white matter fibers[1] and the optimal controls and optimal dynamics give the direction of the fibers;

---

[1] when the start point x and the target point x_0 of the white matter are connected by a fiber.

the other candidate trajectories and the other control values describe the spectrum of all the curves connecting x to x_0 and staying in the white matter as well as their tangent vector fields.

In the optimal control problems, the goal is then to characterize and compute this optimal control α^* which minimizes the cost (7.11) (f_x^* being immediately deduced from α^*).

7.4.3 Geodesics and the Optimal Control Framework

Now, let us consider the above optimal control problem with

$$l(x, a) = 1.$$

The problem consists then in finding the control function α^* s.t. for all x in Ω and for all α

$$\int_0^{T_{x,x_0,\alpha^*}} l(\xi_{x,\alpha^*}(t), \alpha^*(\xi_{x,\alpha^*}(t)))dt \leq \int_0^{T_{x,x_0,\alpha}} l(\xi_{x,\alpha}(t), \alpha(\xi_{x,\alpha}(t)))dt,$$

then s.t.

$$T_{x,x_0,\alpha^*} \leq T_{x,x_0,\alpha},$$

where $T_{x,x_0,\alpha}$ is the first time for which the trajectory $\xi_{x,\alpha}$ (controlled by the dynamics f) reaches the target[2] x_0. $T_{x,x_0,\alpha} = +\infty$ if the trajectory does not reach x_0. In other words, misusing the notations[3], α^* is

$$\alpha^* = \arg\min_{\alpha(.)} \left\{ \int_0^{T_{x,x_0,\alpha}} 1dt \right\} = \arg\min_{\alpha(.)} \left\{ T_{x,x_0,\alpha} \right\}. \tag{7.13}$$

Furthermore let $\mathcal{A}$ be the unit Riemannian sphere $\mathcal{S}_R(0,1)$ associated to A_x, i.e.

$$\mathcal{A} = \mathcal{S}_R(0,1) = \{A_x^T b \mid b \in \mathcal{S}_E(0,1)\},$$

$\mathcal{S}_E(0,1)$ being the Euclidean unit sphere, and let

$$f(x, a) = -a$$

i.e. the dynamics is equal to the control (up to the sign). Then for all t, $\xi'(t) = -\alpha(\xi(t))$ covers the unit Riemannian sphere when α covers the set of the functions $\Omega \to \mathcal{A}$.

Let us denote $T_{\xi,\mathcal{T}}$ the first time t for which the trajectory ξ reaches the target $\mathcal{T}$, i.e. for which $\xi(t) \in \mathcal{T}$. We have then $T_{x,x_0,\alpha} = T_{\xi_{x,\alpha},\{x_0\}}$. So, minimizing the cost

$$\int_0^{T_{x,x_0,\alpha}} 1dt$$

for $\alpha : \Omega \to \mathcal{A}$ is equivalent to minimize

$$\int_0^{T_{\xi,x_0}} 1dt$$

---

[2] for simplicity, we have fixed $\mathcal{T} = \{x_0\}$.

[3] α^* minimizes (7.13) for all $x \in \Omega$. Let us remind that we can prove that α^* always exists [19].

for ξ such that $|\xi'(t)|_R = 1$. In other words, in this case, the optimal trajectories ξ_x^* correspond with the geodesics considered in [183] (see Definition 7.4.1.1).

7.4.4 Advantages of the Optimal Control Interpretation

The control interpretation has distinct advantages: All the objects of interest live in $\mathbb{R}^3$ (instead of a manifold), and are governed by the Euclidean metric, hence the interpretation is independent of the geometric structure. As an illustration of this benefit, in order to estimate the direction of the geodesics, we proposed in section 7.3 to compute the gradient of the distance function on the manifold, which requires some care in order to take into account the geometry imposed by the metric and is a challenging task when working on an irregular domain such as the brain white matter. In the control formalism the interpretation is rather direct: *the tangent of the geodesics is in fact the optimal dynamics* f_x^* (since the geodesic corresponds to the optimal trajectories). Also, the optimal dynamics f_x^* coincides with the optimal control (up to its sign), which is the direct outcome of our algorithm.

The control framework reveals the fact that the *value function* V defined by the $\min$ of (7.11)

$$V(x) = \min_{\alpha} \left\{ \int_0^{T_{x,x_0,\alpha}} l(\xi(t), \alpha(\xi(t)))dt \right\} \tag{7.14}$$

is the viscosity solution [193, 19] of the partial differential equation (PDE)

$$\sup_{a \in \mathcal{A}} \{ -f(x,a) \cdot \nabla u(x) - l(x,a) \} = 0, \tag{7.15}$$

verifying $u(x_0) = 0$ and complemented by state constraints on the boundary of the domain $\partial\Omega$ [273, 249]. This results is quite general and it applies to any optimal control problem as the one described in section 7.4.2. In our specific case, for $f(x,a) = -a$, $l(x,a) = 1$ and $\mathcal{A} = \mathcal{S}_R$ (see section 7.4.3), equation (7.15) coincides with the explicit equations:

$$|A_x \nabla u|_E = 1 \tag{7.16}$$

i.e.

$$|\nabla u|_R = 1. \tag{7.17}$$

Also, we logically associate with this PDE the explicit Hamiltonian

$$\begin{aligned} H_{AEik}(x,p) &= |A_x p|_E - 1 \qquad (7.18) \\ &= |p|_R - 1. \qquad (7.19) \end{aligned}$$

Proof. By the Cauchy-Schwarz inequality, we have

$$|A_x p|_E = \sup_{b \in \mathcal{S}_E(0,1)} (A_x p) \cdot b = \sup_{b \in \mathcal{S}_E(0,1)} p \cdot (A_x^T b).$$

Then

$$|A_x p|_E = \sup_{a \in \mathcal{S}_R(0,1)} p \cdot a.$$

So

$$\sup_{a \in \mathcal{S}_R(0,1)} \{a \cdot p - 1\} = |A_x p|_E - 1.$$

$\square$

In other respects, in our application, we have

$$V(x) = \min_{\alpha} \left\{ \int_0^{T_{x,x_0,\alpha}} 1dt \right\} = \min_{\alpha} T_{x,x_0,\alpha}. \tag{7.20}$$

Since the control values stay in the Riemannian unit sphere $S_R(0,1)$, we tautologically have

$$|a|_R = 1.$$

Since we have moreover

$$|a|_R = |f(\xi(t), a)|_R = |\xi'(t)|_R,$$

then all the candidate trajectories have a constant speed of norm 1 for the Riemannian metric. The covered distance coincides then with the time required for the trajectories to reach the target. This was at the basis of the dynamic formulation in section 7.3. The value function $V(x)$ defined by equation (7.20) is then naturally the distance function considered in the Riemannian context. Thus, one recovers the fact that the (anisotropic) distance function is the viscosity solution of the Anisotropic Eikonal equation (7.17).

The control framework [19] also reveals that $f(x, \alpha_x^*) = -\nabla H(x, \nabla u(x))$ where H is the Hamiltonian associated to the PDE (7.15) and ∇u is the gradient of its solution. This trick is quite useful and efficient in practice for computing the optimal dynamics.

More globally, let us remind that the Riemannian geometry focuses on the description of the geometry and in particular on the geodesics, when the main concern of the optimal control framework is the optimal control which here coincides with the optimal dynamics (up to its sign). Also, as we will see in the next sections, from the computational point of view, the notion of optimal dynamics is a fundamental key. In particular, this notion is essential for designing an efficient numerical scheme. In fact, the control formulation of the problem directly yields our numerical method, which we report in sections 7.5.2 and 7.5.3. Finally, the notion of optimal dynamics is also at the basis of our connectivity measure, from the theoretical point of view as well as from the computational point of view.
For practical purposes, we will adopt either interpretation (Riemannian geometry versus optimal control) depending on the situation and exploit their complementary benefits.

7.4.5 Connectivity Measures

We would like to precise and improve the connectivity measure $\mathcal{C}$ introduced in section 7.3.2. We start by pointing out that, for a fixed point x_0 and any point x, the geodesic γ_x (associated to the metric given by the tensors) connecting x to x_0 always exits. If x is connected to x_0 by a white matter fiber then the associated geodesic γ_x coincides with the fiber. Nevertheless, for any x, the associated geodesic γ_x does not necessarily coincide with a fiber. Also, in order to reach our goal (reconstruction of the white matter fibers) we then need to be able to trace the geodesics and to evaluate if a point is potentially connected to x_0.

In this section, we propose a well-founded score to measure the expectation that a given geodesic truly represents the connection of a point x with x_0. By computing statistical maps of this measure for all points x in the brain, we can then determine which points are likely to be connected to x_0 and then trace the fibers. In section 7.5 we propose an original numerical scheme based on Fast Marching Methods (FMM) to efficiently compute these maps.

Let us fix a point of interest $x_0 \in \overline{\Omega}$ and let us consider the PDE/control/Riemannian problem associated with DTI. In section 7.4, we show that, $\forall x \in \Omega$, the optimal dynamics f_x^* coincides with the derivatives of the geodesics $\gamma'(t)$ at x and that they are in the Riemannian unit ball $B_R(0,1)$ which is also the set $\{A_x q,\ q \in B_E(0,1)\}$. So, for a fixed point x (and a fixed tensor D_x), the larger the Euclidean norm of f_x^*, the more confident we are in the local direction of the geodesic. Following this idea, we then define a general (local) confidence measure:

$$\mathcal{C}(x) = \sqrt{f_x^{*T} D_x^{\alpha} f_x^*},$$

α being in $\mathbb{R}$. In addition to being intuitive, this measure inherits the robustness to noise of the optimal dynamics. It also exploits the full information provided by the diffusion tensor. Finally, it does not penalize any direction in case of isotropy. Let us now discuss the possible values of α: if $\alpha = -1$, we get $\mathcal{C}(x) = 1,\ \forall x \in \overline{\Omega}$. This simply means that, when we use the Riemannian metric given by the inverse of the diffusion tensor, all the geodesics are equivalent. On the contrary, when $\alpha = 0$, we have $\mathcal{C}(x) = |f_x^*|_E$ and we claim that it is a natural local measure of connectivity since this measures the *speed* of propagation at x. Finally, when $\alpha \to \infty$, this boils down to considering the alignment of the optimal dynamics with the local major eigenvector. This was used in section 7.3.2 and [234] but it may be sensitive to isotropic areas where, by definition, the major eigenvector is undefined.

From this local connectivity measure, we can define global information from its statistics (mean and standard deviation) along the optimal trajectory:

$$\mu(x) = \langle \mathcal{C}(x) \rangle = \frac{1}{\tau_x^*} \int_0^{\tau_x^*} \mathcal{C}(\xi_x^*(t))dt,$$

$$\sigma(x) = \sqrt{\langle \mathcal{C}(x)^2 \rangle - \langle \mathcal{C}(x) \rangle^2}.$$

where τ_x^* is the length of the optimal trajectory ξ_x^*. We should point out that, since $|\xi_x^{*\,\prime}|_R = 1$, this length (i.e. the geodesic distance between the curve endpoints x_0 and x) coincides with the arrival time T_{x,x_0} introduced in section 7.4.

A point x connected to x_0 by a white matter fiber will ideally have a large value for $\mu(x)$ and a small standard variation $\sigma(x)$. The choice of using the mean instead of just integrating along the trajectories allows the comparison of two points x and y which are located at different distance from x_0, i.e. s.t. $\tau_x^* \neq \tau_y^*$. Although the mean value of the connectivity may be sufficient to discriminate likely fibers, the variance of this quantity may also be of great help since an ideal fiber would exhibit a high coherence of $\mathcal{C}(x)$ along its trajectory.

Remark 1.
To compute the optimal dynamics, we need to compute also the geodesic distance, which in fact is equal to τ_x^*. In practice, we just need to compute

$$\mathcal{R}(x) = \int_0^{\tau_x^*} \mathcal{C}(\xi_x^*(t))dt,$$

and

$$\mathcal{S}(x) = \int_0^{\tau_x^*} \mathcal{C}(\xi_x^*(t))^2 dt\,.$$

The values of $\mu(x)$ and $\sigma(x)$ are then derived immediately by using the value τ_x^*.

7.5 A FAST NUMERICAL ALGORITHM (THE STATIONARY PERSPECTIVE)

We now describe a algorithm which can, in one single pass over the volume Ω, compute all the quantities of interest.

7.5.1 Related Work and Contributions

To the best of our knowledge, there is no algorithm to compute *directly* the geodesics or a fiber connectivity confidence map to a point x from DTI data. All the methods recovering white matter fibers by using some front propagation technique proceed by implementing successively the following four steps:

1. Computation of the distance function to x;
2. Extraction of the gradients of the distance function;

3. Estimation of the optimal dynamics from the gradients of the distance function;
4. Tracing of the geodesics from the computed directions. This step needs in particular an interpolation of the derivatives of the geodesics.

Some slight variants are proposed in the literature (see [76, 166] and references therein). For example, in the particular case of the isotropic Eikonal equation (where the optimal dynamics coincide with the gradient of the distance function), [166] suggests not to compute the gradients for all voxels and later interpolate them, but rather to directly compute the interpolated gradients from the distance function.
We wish to emphasize that the explicit tracing of the geodesics is a prerequisite to all the previous methods for computing connectivity confidence measures which in fact consist in the integration of a local criterion along the entire geodesic during the geodesics tracing step. Thus, the estimation of a complete map of connectivity measures needs to explicitly trace all the geodesics starting from all the points of the map. This approach is rather computationally intensive.

The numerical method we propose here for computing the confidence measures does not need to trace any geodesic. The confidence measure map is a direct output of our algorithm. It simultaneously and consistently computes the (geodesic) distance function, the optimal dynamics and the confidence measures.

The methods of the type "Fast Marching" [286, 268, 269, 248] are "one-pass" methods allowing to solve numerically partial differential equations of the type (7.15). Based on a causality principle, the Fast Marching Methods (FMM) stand in contrast to iterative methods (see for example [261, 282] and more specifically [147] in our field) which iteratively update the approximations of the solution by using paths that do not depend on the data. The idea of the FMM consists in computing the solution of the PDE as a front propagates along the optimal trajectories. Moreover, it has been recently shown in [315] that using a quantization of the priorities in the marching computation reduces the original complexity of FMM from $\mathcal{O}(N \log N)$ to $\mathcal{O}(N)$. Our algorithm extends the classical FMM [286, 268, 269, 248] by computing and returning, in addition, the optimal dynamics and the connectivity confidence measures. The consistency of our results relies on the fact that for all the computations we use the same (optimal) simplex.

Remark 2.
1) All the quantities we compute are essential: The optimal dynamics is necessary in order to trace the geodesics, which in turn is useful for the visualization of the fibers. Even if the result of the computation of the (geodesic) distance is not required for tracing the geodesic, it is essential for obtaining the final measures (expectation and standard deviation) we use in practice to estimate the connectivity confidence.

2) Our method is a "one pass method" based on front propagation. An important consequence is that we do not need to wait for the complete computation of the distance function on the whole domain to be able to exploit it for computing the connectivity measures. Also, if at any time the process stops, all the values already computed are valid approximations, unlike other iterative techniques.

In the sequel, we describe our global algorithm and then the implementation of each specific step.

7.5.2 Global Algorithm

As in the classical Fast Marching Method [268, 269, 248], the grid points are divided into the three classes: *Accepted*, *Considered*, *Far*. Below U, f, R and S are respectively the approximations of the (geodesic) distance function, the optimal dynamics f_x^*, $\mathcal{R}$ and $\mathcal{S}$ (defined in section 7.4.5). x_0 is the interest point. The algorithm is then the following:

Algorithm 3 Fast Marching algorithm for the computation of U, f, R and S

```
1: Start with all the grid points in Far.
2: Move $x_0$ and the grid points on the boundary $\partial\Omega$ to Accepted. Set $U(x_0) = 0$ and $U(x) = +\infty$ (FLT_MAX in practice) for all $x \in \partial\Omega$.
3: Move all the grid points adjacent to the Accepted points into Considered and for such points $x$, evaluate $U(x)$ by using the update scheme (7.21) and modify the associated optimal dynamics to $f(x)$; see section 7.5.3.
4: Find the Considered point $\tilde{x}$ with the smallest value $U(x)$. Move $\tilde{x}$ from Considered to Accepted. Compute and assign $R(\tilde{x})$ and $S(\tilde{x})$, see section 7.5.4.
5: Move from Far into Considered, all the Far points which are adjacent to $\tilde{x}$.
6: Re-evaluate $U(x)$ and the associated dynamics $f(x)$ for all the Considered points adjacent to $\tilde{x}$, see section 7.5.3.
7: If the set Considered points is not empty, return to step 4.
```

7.5.3 Distance and Optimal Dynamics Computation

Here, we focus on the implementation of the updating step returning the approximation of the distance function and the optimal dynamics. Following [248], we use the scheme

$$S(\rho, x, t, u) = \sup_{a \in A}\{-f(x,a) \cdot P_{s_1(x,a),..,s_n(x,a)} - l(x,a)\} \tag{7.21}$$

where $[P_{s_1,..,s_n}]_i = \frac{t-u(x+s_ih_ie_i)}{-s_ih_i}$, $s_i(x,a) = sign\,(f_i(x,a))$, h_i denotes the grid size in the i^{th} direction and $\{e_i\}$ is the canonical basis of $\mathbb{R}^n$. In our case, $n = 3$.

Basically, this scheme is obtained by replacing ∇u by $P_{s_1,..,s_n}$ in equation (7.15) and by choosing the simplex (i.e. $(s_1, .., s_n)$) which contains the dynamics of the optimal control. Moreover, we take advantage of this in order to obtain *simultaneously and consistently* the approximations of the geodesic distance function and of the optimal dynamics.

Separation and choice of the good simplex

Let us fix $x \in \Omega$. The updating step consists in computing the value we want to assign to $U(x)$ from the values $U(x \pm s_i h_i e_i)$. The update value for $U(x)$ is the solution of the equation $S(\rho, x, t, u) = 0$ (equation in t), i.e.

$$\max_{s \in \{\pm 1\}^n} \sup_{a \in A_s} \{-f(x,a) \cdot P_{x,s,U}(t) - l(x,a)\} = 0$$

where we note $s = (s_1, ..s_n) \in \{\pm 1\}^n$, $[P_{x,s,U}(t)]_i = \frac{t - U(x + s_i h_i e_i)}{-s_i h_i}$ and

$$A_s = \{a \in A \mid \forall i = 1..n, s_i(x,a) = s_i\} .$$

Now, for all $s \in \{\pm 1\}^n$, let us denote

$$G_s(t) = \sup_{a \in A_s} \{-f(x,a) \cdot P_{x,s,U}(t) - l(x,a)\} \tag{7.22}$$

and t_s, the solution of the equation (in t) $G_s(t) = 0$. Since $G_s(t)$ is increasing with respect to t, the solution of $\max_s G_s(t) = 0$ is $t_0 = \min_{s \in \{\pm 1\}^n} t_s$. Hence, the implementation of the update step is reduced to the computation of the 2^n solutions t_s and, finally, to the choice of the smallest one. We thus choose here the "good" simplex. We call it the *optimal simplex*. In other respects, when we compute t_s, we also compute $f_s = f(x, a_s)$, where $a_s \in A_s$ is the optimal control of (7.22) (see subsection 7.5.3). We can then associate to t_0 the optimal dynamics $f_0 = f_s$ where s is the optimal simplex.

Computation of t_s and of the associated dynamics

Now let us fix $s = (s_1, ..s_n) \in \{\pm 1\}^n$. If we denote

$$g_s(a,t) = -f(x,a) \cdot P_{x,s,U}(t) - l(x,a) ,$$

then t_s is the solution of

$$\sup\nolimits_{a \in A_s} g_s(a,t) = 0. \tag{7.23}$$

By continuity of $f(x,.)$, A_s is a closed subset of $\mathbb{R}^m$. Let a_s in A_s be the *optimal control* of (7.23). We then have two cases:

1. $\forall k \in [1..n],\ f_k(x, a_s) \neq 0$ (in other words $a_s \in Interior(A_s)$): One can prove that this is equivalent to: t_s *is the solution of the equation* $H(x, P_{x,s,U}(t)) = 0$ *with the associated optimal control*[4] *in* A_s.

---

[4] i.e. the optimal control of $\sup_{a \in A}\{-f(x,a) \cdot P_{x,s,U}(t) - l(x,a)\} = 0$. This optimal control can be anywhere in the whole set A.

2. $\exists k \in [1..n]$ such that $f_k(x, a_s) = 0$: In this case, $t_s = \min t_s^i$, where for each $i \in [1..n]$, t_s^i is the solution of the equation in t: $\sup_{a \in A_s^{i+}} g_s(a,t) = 0$ where $A_s^{i+} = A_s^k \cap A_s$ and $A_s^k = \{a \in A \mid f_i(x,a) = 0\}$.

In practice, we first compute the roots t_s of $H(x, P_{x,s,U}(t)) = 0$. We then test if the optimal control $a_s \in A_s$ (basically it is in A). To do that, we just have to estimate

$$f(x, a_s) = -\nabla H(x, P_{x,s,U}(t_s))$$

and to verify that $\forall k, sign\,(f_k(x,a_s)) = s_k$. In particular, this test does not require the knowledge of the optimal control a_i and directly provides the associated optimal dynamics. If all the signs are correct, we have found our solution and we stop here. Otherwise, we have to compute the solutions t_s^i. To achieve this goal, we can make

$$H_s^i(x,p) \overset{\text{def}}{=} \sup_{a \in A_s^k} \{-f(x,a) \cdot p - l(X,a)\} = 0$$

explicit and deal with H_s^i in the same way we have dealt with H. To make H_s^i explicit we use the Legendre Transform [248].

Details for the 3D-anisotropic Eikonal equation

Here we detail the successive Hamiltonians necessary for the implementation of the method described in the previous section in the case of the 3D-anisotropic Eikonal equation.

Let us remind that for any Hamiltonian $H(x,p)$, we call the *Legendre Transform* the function H^* defined by

$$H^*(x,a) = sup_{p \in Dom(H(x,.))}\{p \cdot a - H(x,p) \leq +\infty\}$$

see for example [106, 244]. For simplicity, we denote below the Hamiltonian H_{aEik} by H. We have

$$H_i(x,p) \overset{\text{def}}{=} \sup_{\substack{a \in \mathbb{R}^3 \\ a_i = 0}} \{a \cdot p - H^*(x,a)\} = \tilde{p_t^i}^T \, ([\![D_x^{-1}]\!]^i)^{-1} \, \tilde{p_t^i} - 1,$$

$$H_{ij}(x,p) \overset{\text{def}}{=} \sup_{\substack{a \in \mathbb{R}^3 \\ a_i = 0, a_j = 0, i \neq j}} \{a \cdot p - H^*(x,a)\} = \frac{{p_{tk}}^2}{(D_x^{-1})_{k,k}} - 1,$$

where $\tilde{p_t^i} = (p_{t1}, ..., p_{ti-1}, p_{ti+1}, ..., p_{tn})$, and $[\![D_x^{-1}]\!]^i$ is the matrix D_x^{-1} without the i^{th} row and i^{th} column. In practice our method boils down to resolving basic second order equations and to testing some signs.

7.5.4 Connectivity Measures Computation

In this section we detail how to compute the connectivity measure $\mathcal{R}(\tilde{x})$ at the step 4 of our global algorithm. At this stage, we already know the optimal dynamics $f^*_{\tilde{x}}$, the optimal simplex $(\tilde{x}, x_1, x_2, x_3)$ (we denote $x_i = \tilde{x} + s_i(\tilde{x})h_i e_i$ where $s_i(\tilde{x})$ is the sign of the i^{th} component of $f^*_{\tilde{x}}$ and $h_1 \times h_2 \times h_3$ is the size of the voxels) and the values $\mathcal{R}(x_i)$ for $i = 1..3$.

Let y be the intersection of the optimal trajectory with the front. By assuming that the trajectory is locally affine, we have: $y = \tilde{x} + \tau f^*_{\tilde{x}}$ where τ is the time for the trajectory to reach the front, see figure 7.1(a). As in [245], we can prove that

$$\tau = 1/\sum_{i=1..3} q_i$$

where q_i is the absolute value of the i^{th} component of $f^*_{\tilde{x}}$ divided by h_i. By assuming that $\mathcal{R}$ is locally affine, we have [245]

$$\mathcal{R}(y) = \sum_{i=1}^{3} \tau q_i \mathcal{R}(x_i).$$

Thus by noting that

$$\mathcal{R}(\tilde{x}) = \mathcal{R}(y) + \int_0^\tau \mathcal{C}(\xi^*_{\tilde{x}}(t))dt,$$

we obtain

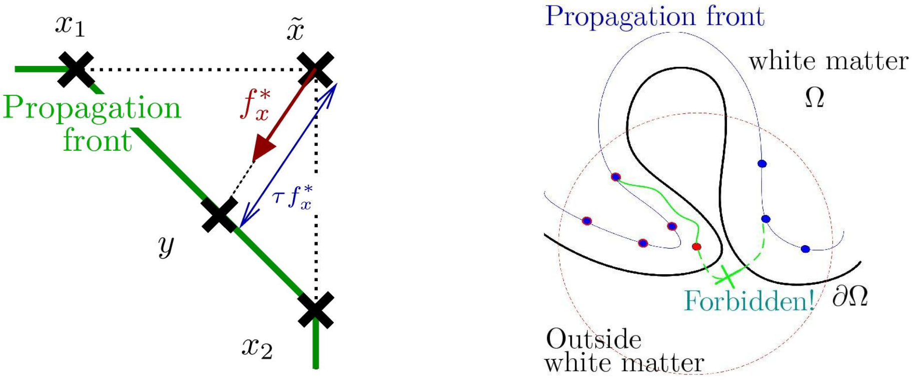
$$\mathcal{R}(\tilde{x}) \simeq \textstyle\sum_{i=1}^{3} \tau q_i \mathcal{R}(x_i) + \tau \mathcal{C}(\tilde{x}). \tag{7.24}$$

Remark 3.
1) The approximation of $\mathcal{S}(\tilde{x})$ required for the computation of the standard deviation $\sigma(\tilde{x})$ is obtained exactly in the same way. We just have to replace $\mathcal{C}$ by $\mathcal{C}^2$ in equation (7.24).

2) This scheme can also be obtained by discretizing the equation $\langle \nabla_E \mathcal{R}(x), f^*_x \rangle_E = \mathcal{C}(x)$ (obtained by evaluating $\lim_{\varepsilon \to 0} \frac{\mathcal{R}(x+\varepsilon f^*_x) - \mathcal{R}(x)}{\varepsilon}$) and by slightly modifying the scheme proposed by [1] for solving a similar equation.

7.6 EXPERIMENTAL RESULTS

In this section, we first present some results illustrating the computation of the distance function and geodesics by the method proposed in section 7.3. We then emphasize its possible limitations. Finally, we show that the algorithm introduced in the previous section can efficiently solve those issues.



(a) (b)

Figure 7.1: (a) Approximation of the geodesic and localization of y, (b) Depiction of the topological problem in a convoluted area of the white matter

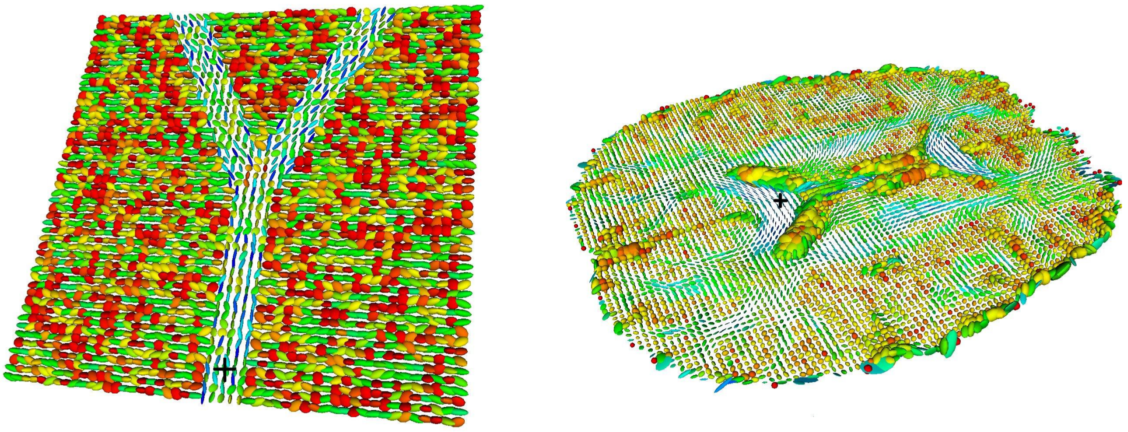


Figure 7.2: Synthetic (left) and human brain (right) DTI datasets

7.6.1 The Level Set Perspective

We will consider synthetic and real DTI datasets to illustrate and quantify the quality of the estimated distance functions with upwind or fifth order WENO finite differences schemes. Our criterion will be the a posteriori evaluated map $|\text{grad}u|$ which must be equal to 1 everywhere except at the origin x_0 since the distance function is not differentiable at this location. Figures 7.2 and 7.3 present a synthetic and a real DTI dataset and the associated distance functions. The synthetic tensor field is composed by Y pattern with tensors aligned along the main orientations of this pattern. The background is made of tensors following the direction orthogonal to the main branch of the Y. Gaussian noise was added to the initial piecewise constant dataset. We used the algorithm proposed in section 5.4.1. The origin x_0 was chosen to be at the lower part of the main branch of the Y (black cross). The resulting distance function is shown in figure 7.3 (left) where we represent one axial slice of the volume as an elevation surface to emphasize the variations of the function. We

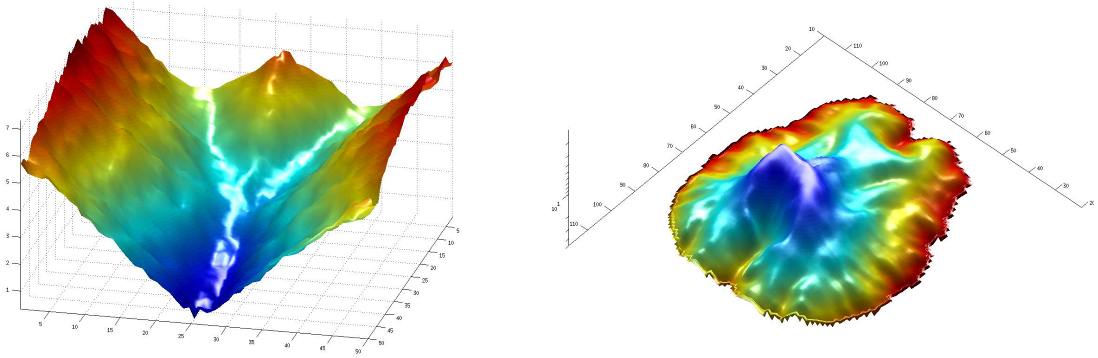


Figure 7.3: Axial slices of distance functions computed for the synthetic (left) and real (right) datasets of figure 7.2 (Colormap: blue is small and red is large)

can notice that the distance stays small within the Y pattern and rapidly increases as it reaches the background.

The human brain DTI dataset was acquired at CEA-SHFJ/Orsay, France and provided to us by J.F. Mangin and J.B Poline. We focused on the posterior part of the corpus callosum, known as the genu. The origin x_0 of the distance was chosen in the middle of the genu (black cross) and the resulting distance function is presented in figure 7.3 (right) as an elevation surface (inverted for the sake of clarity). As expected, from neuroanatomical knowledge, the distance functions stays small along the commissural radiations towards the occipital areas of the brain and increases otherwise. By initiating the back-propagation procedure of section 7.3.2 in the visual area for instance (see chapter 11 for details on that point), or more generally in regions of the occipital part of the brain, we could recover the geodesics of figure 7.4, which agree with the well-known shape of neural fibers in these regions.

Finally, we compared the accuracy of the distance computation that can be achieved with upwind or WENO schemes. We used the previous human brain DTI dataset but resorted to a more complex synthetic example. This dataset is made of three intersecting cylinders oriented along the x, y and z axis of the volume. Tensors in each cylinder are aligned with the main direction of the cylinder and the background is made of small isotropic tensors. This results in perfectly isotropic tensors at the intersection of the three cylinders, surrounded by planar tensors in the area where only two cylinders cross each others. The origin was chosen at one of the extremity of one cylinder. Our numerical scheme 7.9 behaves fairly well on the two datasets, as shown in table 7.1 where a sensible improvement is noticeable when using fifth order WENO schemes.

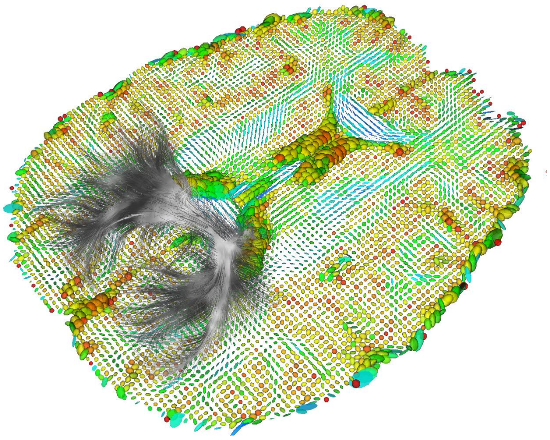


Figure 7.4: Geodesics recovered from distance function of figure 7.3 (right) by starting from regions of the occipital areas

DataSet	Scheme	Mean	Std. Dev	Maximum
Synthetic	**Upwind**	0.9854	0.123657	4.50625
Synthetic	**WENO**	0.977078	0.116855	2.0871
Real DTI	**Upwind**	0.994332	0.116326	4.80079
Real DTI	**WENO**	0.973351	0.110364	3.72567

Table 7.1: Statistics on $|\text{grad}u|$ for synthetic and real DTI datasets

7.6.2 Challenging Computational Issues

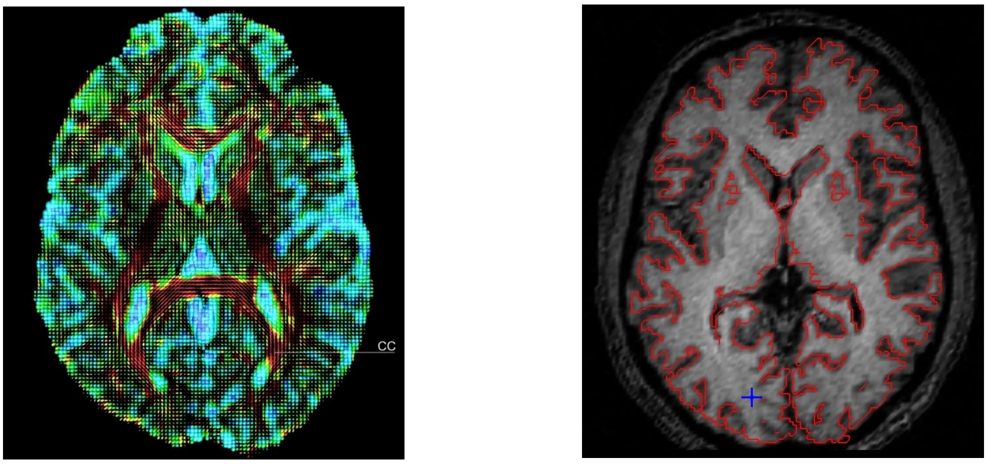
The nature of the problem we are trying to solve raises two major computational difficulties which, to our knowledge, are not very well dealt with in the literature. They may also be an issue, in certain circumstances, for the dynamic formulation of the anisotropic Eikonal equation.

Handling the white matter convoluted geometry

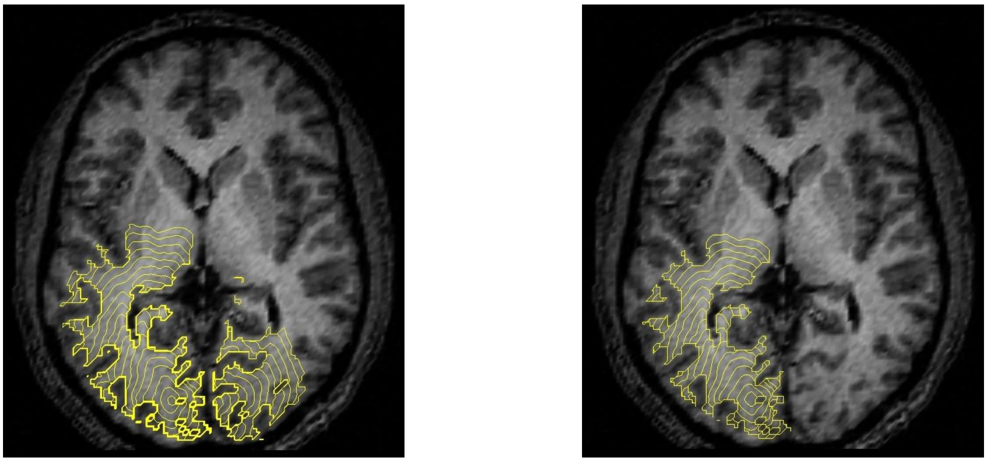
First of all, as presented in figure 7.5 and detailed in figure 7.1(b), solving the anisotropic Eikonal equation within a convoluted domain such as the brain white matter is necessary and complicated. Indeed, the connections we are looking for are defined between cortical areas or between cortical areas and the basal ganglia. In other words, we are essentially interested in pathways linking together parts of the domain boundary.

In figure 7.5, the geodesic distance to the blue cross in image (b) (i.e. x_0) was computed, for the DTI data presented in image (a) and within the mask outlined in red in image (b). Its isovalues (in the range $[0, 1500]$) are depicted by the yellow lines in images (c) and (d). With the level set implementation, the front diffuses through the cerebrospinal fluid and directly connects the right hemisphere. This is anatomically incorrect since the fibers starting from the blue cross (located in the V1 visual area) go through the corpus callosum (CC) to reach the other hemisphere. With the stationary formulation, we can correctly estimate the distance since, by definition, the Fast Marching implementation ignores all the locations outside the mask.

This kind of difficulty is also encountered with the Ordered Upwind Method (OUM) recently proposed by Sethian and Vladimirsky [269]. The OUM is a numerical method of type FMM which uses enlarged neighborhoods. The more anisotropic the tensor, the larger the neighborhood. In addition to increasing the computation time, Sethian and Vladimirsky's method explicitly authorizes this type of topological error by allowing the trajectories to step outside the mask and to directly connect any nearby voxel located on the front. Figure 7.1(b) illustrates this potential problem. The scheme presented in section 7.5 only uses nearest neighbors (six nearest neighbors in 3D). It is thus not prone to this problem and always respects the topology of the mask.



(a) DTI axial slice (Anisotropy color code: blue=low/red=high) (b) White matter segmentation



(c) Level set algorithm [183] (d) Fast marching algorithm [246]

Figure 7.5: Topological inconsistency in the occipital cortex.

Robust estimation of the optimal dynamics

The second issue is related to the robustness of the optimal dynamics (i.e. the geodesics tangent vectors) computation. Indeed, all the existing methods need to explicitly compute the derivatives of the distance function. This is well-known to be sensitive to noise, especially on the boundaries where the discretization of the differential needs to be adapted. We present, in figure 7.6, a comparison of the vector fields obtained by the method proposed in section 7.3 (top row) and by the approach of section 7.5 (bottom row) on a 3D synthetic DTI dataset (see image (a)). The origin of the distance function is located at the center of region B (see image (b)).

7.6.3 The Fast Marching Perspective

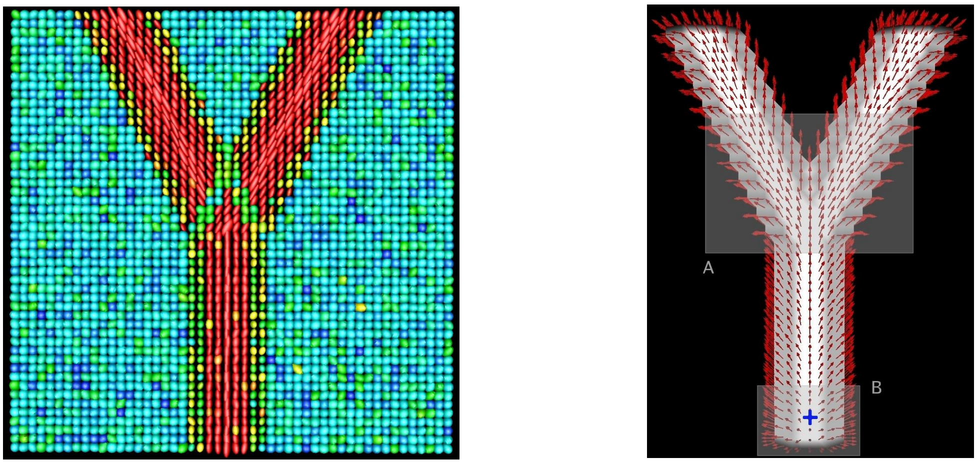
In the following, we illustrate our Fast Marching method by computing the quantities μ and σ, introduced in section 7.4.5, as well as the geodesics associated to the highest connectivity measure. This is done on the synthetic tensor field of figure 7.6 as well as in the splenium (posterior part) of the corpus callosum for the real dataset of figure 7.5.

Data acquisition

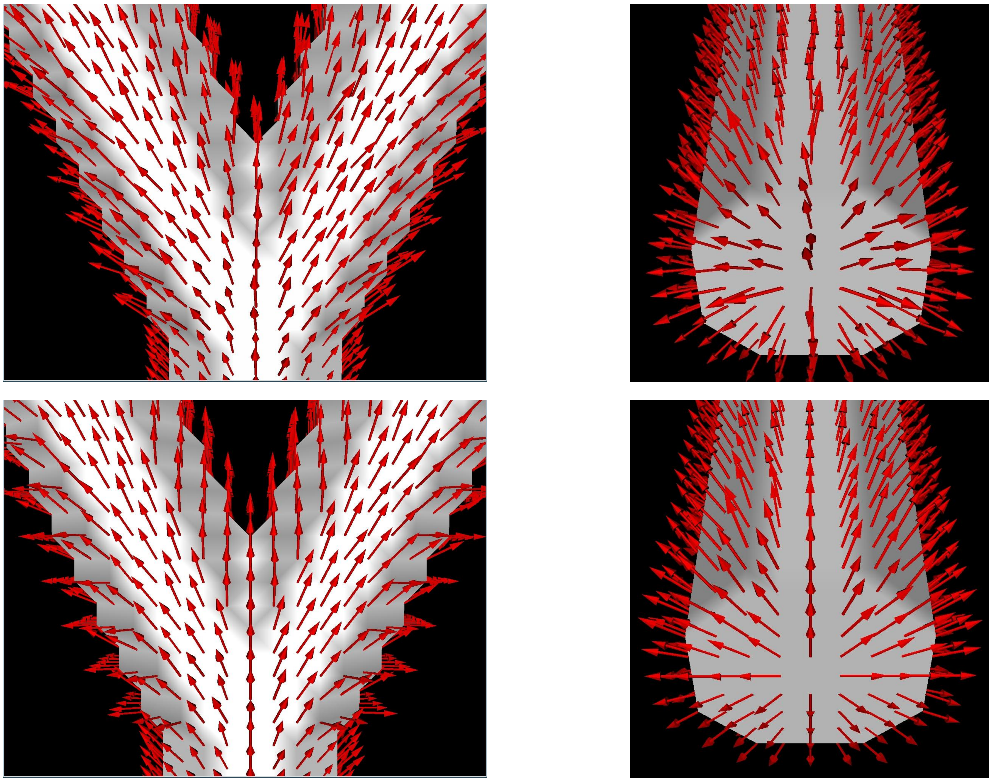
Diffusion weighted images were acquired on a 3 Tesla MEDSPEC 30/80 AVANCE (Bruker) at the Centre IRMf de Marseille, France, using a quadrature bird-cage head coil. We used 12 diffusion gradient directions and a b factor of 1000 s/mm^2. Acquisitions were repeated 8 times for each direction in order to ensure a good signal-to-noise ratio. Voxel size was $2 \times 2 \times 2$ mm^3 and diffusion tensors were estimated by the robust gradient descent algorithm proposed in section 6.1.2 [186]. An axial slice of the resulting DT image is presented in figure 7.5(a).

Computational efficiency

PDE methods for brain connectivity mapping such as [218, 234, 183, 147, 50] have the great advantage to yield connectivity information for a point of interest x_0 to the rest of the brain by exploiting the full information of the diffusion tensor. They are however in general quite time consuming and must be iteratively applied to all the voxels of functional regions of interest, which can contain hundreds or thousands of points. By comparison with the methods presented in [147] and [183], our Fast Marching algorithm achieves a dramatic improvement in computational speed. For the geodesic distance computation, Jackowski *et al.* reported a convergence time of about 7 minutes for their iterative sweeping method for a $128 \times 128 \times 40$ DTI dataset on a 1.7 GHz Intel Pentium Xeon with 1.5 Gb of RAM. Our level set formulation required about 20 minutes for a $128 \times 128 \times 58$ DTI dataset on a 1.7 GHz Intel Pentium M with 1 Gb of RAM.



(a) DTI axial slice (Anisotropy color code: blue=low/red=high) (b) Optimal dynamics



(c) Region A (d) Region B

Figure 7.6: Optimal dynamics estimation by differentiation of the distance [(c-d) Top] and by the control theory formulation [(c-d) Bottom].

The computation of the geodesics, together with the evaluation of the statistics of $\mathcal{C}(x)$, is itself a time-consuming task since for each curve, we need to explicitly propagate through the tangent vectors field using, for instance, a 4^{th} order Runge-Kutta integration scheme. In [147], no time is given for the computation of the $14,952$ fibers of interest. However, on our data and for $135,029$ voxels inside the white matter mask, it took approximately 30 minutes on the same computer than the one used for the distance computation. All these computations (distance, vector field and connectivity measures) take about 7 seconds with our stationary/control theory formulation.

Performance of the connectivity measure

We now demonstrate how the statistics of the quantity $\mathcal{C}(x)$ can be used to evaluate the degree of connectivity of pairs of voxels. First of all, we use the synthetic dataset of figure 7.6. The point of interest x_0 is again located at the center of region B (see image (b)). Figure 7.7 (left) presents an axial slice of the thresholded map μ which is consistent with the DT image since we can see that μ is higher along the centerline of the Y shape where the tensors are more anisotropic. Moreover, the right branch is clearly more connected to the origin. This is due to the asymmetry imposed by the tensor field in the diverging region (see figure 7.6 (a)). In figure 7.7 (right), we show the geodesics computed from the 873 voxels with values of μ in the range $[1.5, 1.67]$, i.e. the top 10% most likely connected voxels. Finally, we consider the real dataset of figure 7.8. The origin is located in the middle of the splenium of the corpus callosum. A first threshold is applied on the map σ in order to keep only coherent fibers. This yields a binary mask (threshold value: 0.0056) which is applied to the map μ. As previously, we then threshold this map to preserve only the top 10% most likely connected voxels, with values of μ in the range $[0.0335, 0.0380]$. This yields 2561 fibers that are consistent with neuro-anatomical knowledge.

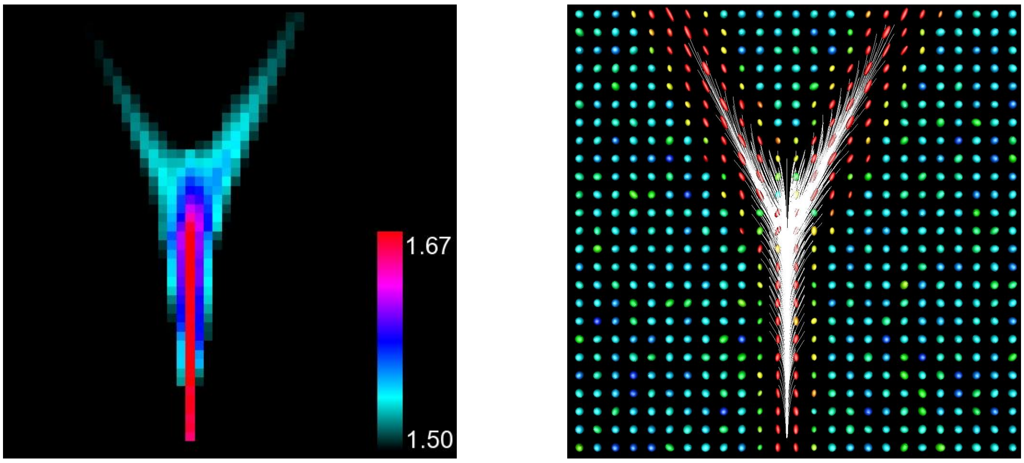


Figure 7.7: Synthetic dataset: (left) Axial slice of the map μ, (right) Most likely connections.

7.7 CONCLUSION

We have proposed a novel global approach to white matter connectivity

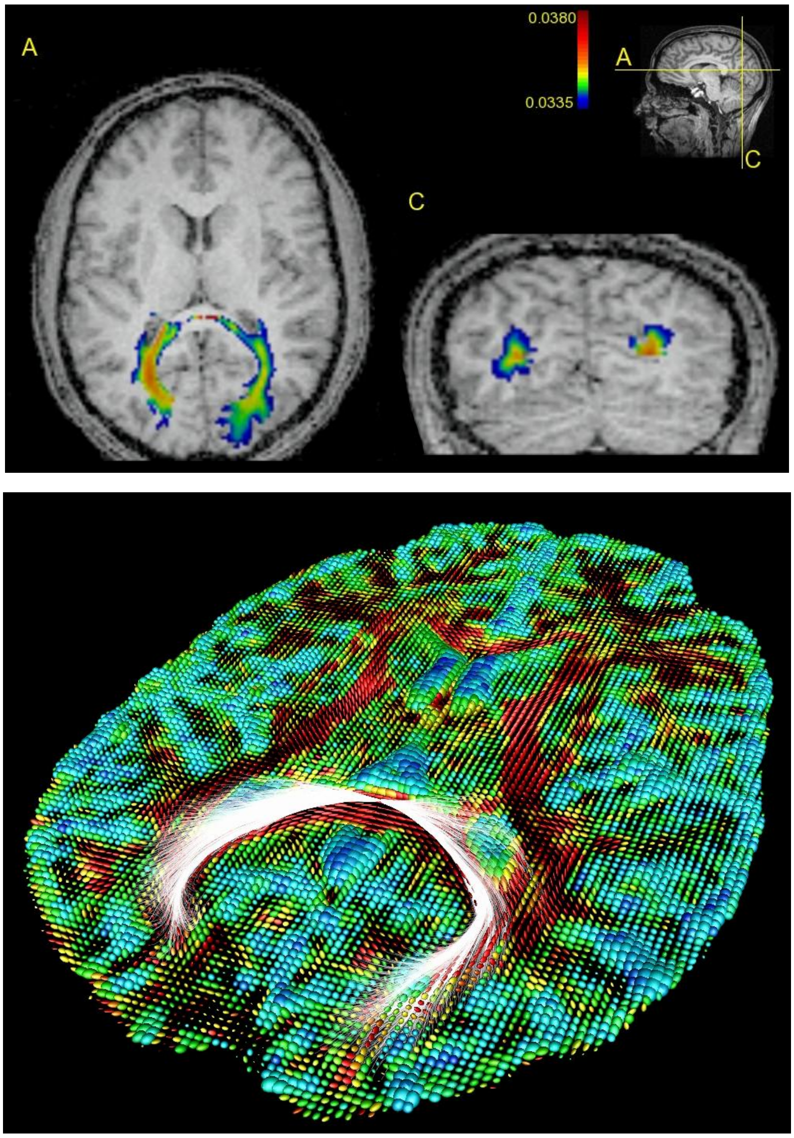


Figure 7.8: Real dataset: [Top] Axial and coronal slices of the map μ, [Bottom] Most likely connections (Anisotropy color code: blue=low/red=high).

mapping. It relies on the fact that probing and measuring a diffusion process on a manifold $\mathcal{M}$ provides enough information to infer the geometry of $\mathcal{M}$ and compute its geodesics, corresponding to diffusion pathways. We then introduced a dynamic formulation of the anisotropic Eikonal equation and a method to estimate neural fibers. Exploiting both an optimal control and the Riemannian interpretation, we achieved a number of improvements over existing methods. We proposed a fast algorithm that reduces CPU time by 2 or 3 orders of magnitude relatively to existing work. We have introduced a general local connectivity measure and experimentally demonstrated its relevance on real data sets. Our algorithm is numerically stable and efficient, since it simultaneously computes the distance function, the optimal dynamics and the statistics of our local connectivity measure from the DT images. Finally we showed that our method overcomes some numerical limitations that cause existing algorithms to fail in highly convoluted regions. The C++ implementation of our GCM algorithm will be soon freely distributed on the web. In the next chapter, we address the issue of the segmentation of diffusion tensor images. This is a complementary processing task to fiber tracking where we seek to recover, at once, entire fiber bundles.

CHAPTER 8

SEGMENTATION OF DIFFUSION TENSOR IMAGES

Contents

OVERVIEW

In this chapter, we address the problem of the segmentation of cerebral white matter structures from diffusion tensor images (DTI). As we will show in this chapter, the definition of a dissimilarity measure and statistics between such quantities is a non trivial task which must be tackled carefully. We claim and demonstrate that, by using the theoretically well-founded differential geometrical properties of the manifold of multivariate normal distributions, as exposed in chapter 5, it is possible to improve the quality of the segmentation results obtained with other dissimilarity measures such as the Euclidean distance or the Kullback-Leibler divergence. The main goal of this chapter is to prove that the choice of the probability metric, i.e. the dissimilarity measure, has a deep impact on the tensor statistics and, hence, on the achieved results. We introduce a variational formulation, in the level-set framework, to estimate the optimal segmentation of a diffusion tensor image according to the following hypothesis: Diffusion tensors exhibit a Gaussian distribution in the different partitions. We must also respect the geometric constraints imposed by the interfaces existing among the cerebral structures and detected by the gradient of the diffusion tensor image. We show how to express all the statistical quantities for the different probability metrics. We validate and compare the results obtained on various synthetic data-sets, a biological rat spinal cord phantom and human brain DT images.

Keywords: diffusion tensor MRI, segmentation, probability metric, Riemannian geometry, information geometry, Fisher information matrix, Kullback-Leibler divergence, level-set

8.1 INTRODUCTION

As presented in the previous chapter, diffusion tensor imaging is extremely useful in order to identify the neural connectivity patterns of the human brain [212], [30], [51], [183]. However, most of the existing techniques addressing this last issue work on a fiber-wise basis. In other words, they do not take into account the global coherence that exists among fibers of a given tract. Recent work by Corouge *et al.* [79] has proposed to cluster and align fibers by local shape parameterization so that a statistical analysis of the tract geometrical and physiological properties can be carried out. A similar work by O'Donnell *et al.* [219] has been recently proposed to cluster fibers. These methods rely on the extraction of a set of streamlines from diffusion tensor images by the method proposed in [212] which is known to be sensible to noise and unreliable in areas of fiber crossings.

For these reasons, we propose to directly perform the segmentation of diffusion tensor images in order to extract neural fiber bundles. While many techniques have been proposed to classify the gray matter, white matter and cerebrospinal fluid from T1-weighted MR images (see [322] for example), the literature addressing the segmentation of white matter structures from DTI is still new. We hereafter draw a quick state of the art of the diffusion tensor images segmentation problem:

Zhukov *et al.* [326] defined an invariant anisotropy measure in order to drive the evolution of a level-set and isolate strongly anisotropic regions of the brain. The reduction of the full tensor to a single scalar value can result in a relatively low discrimination capability, potentially yielding the segmentation of mixed structures. Alternatively, Wiegell *et al.* [312], Feddern *et al.* [111, 112], Rousson *et al.* [260], Wang *et al.* [304] and [303], Lenglet *et al.* [185] and Jonasson *et al.* [154] use or propose different measures of dissimilarity between diffusion tensors. In [312], [304] and [260], the authors use the Frobenius norm of the difference of tensors (i.e. the Euclidean distance). A k-means algorithm with a spatial coherence constraint and an active contours model with a regularity term were respectively used by the first two methods ([312], [304]) to perform the segmentation of different cerebral structures such as the thalamus nuclei or the corpus callosum. The third method [260] used a region-based surface propagation. In [304], a generalization of the region-based active contours to matrix-valued images is proposed. However, it is restricted to the two-dimensional case and obviously limited when it comes to three-dimensional brain data. In [111, 112], partial differential equations based on mean curvature motion, self-snakes and geodesic active contours models are extended to two-dimensional and three-dimensional tensor-valued images by generalizing the notion of structure tensor to matrix-valued data. This method still relies on the Euclidean metric between tensors. The authors apply this framework to the regularization and segmentation of diffusion tensor images. In [154], the authors introduce a geometric measure of dissimilarity by computing the normalized tensor “scalar product” of two tensors, which can be interpreted as a measure of overlap. Finally, the methods exposed in [303] and [185] rely on the symmetrized Kullback-Leibler divergence to derive an affine invariant dissimilarity measure between diffusion tensors.

Contributions of this chapter:

Our contributions are threefold:

- First, we recast the DTI segmentation problem into a unified statistical surface evolution framework. We also make use of the tensor field gradient to detect boundaries between various structures of the white matter. This framework can be implemented with different probability metrics. This is done for the Euclidean distance, Kullback-Leibler divergence and geodesic distance on the manifold of multivariate normal distributions.
- The second contribution is related to the use of the rigorous differential geometrical framework, as presented in [188], rooted in the information geometry and used to express a Gaussian law between diffusion tensors. We overcome

the classical hypothesis considering covariance matrices as a linear space and define relevant statistics to model the distribution of diffusion tensors. To that end, we also extend the methods proposed in [303] and [185] by showing how to compute the covariance matrix, associated to the Kullback-Leibler divergence, of a set of tensors.

- Finally, we demonstrate that the properties of the geodesic distance lead to its superiority, for our segmentation task, over the other two dissimilarity measures. This is achieved by presenting results on both synthetic and real data-sets as well as on a biological phantom, for which only this method succeeds by comparison with the ground truth or neuroanatomical knowledge.

Organization of this chapter:

Section 8.2 recalls how to approximate a Gaussian distribution between diffusion tensors and how to evaluate the norm of a tensor field spatial gradient, needed for the implementation of the boundary term. These three quantities are derived for the three dissimilarity measures of interest. Section 8.3 sets up the Bayesian formulation of the segmentation problem that will be used throughout this chapter. Section 8.4 presents and discusses experimental results on synthetic data-sets, a biological phantom and human brain DT images .

8.2 DTI STATISTICS AND GRADIENT

In this section, we would like to investigate the various possible definitions of a Gaussian distribution between diffusion tensors as well as the possible expressions for the norm of a diffusion tensor image spatial gradient, depending on the chosen metric. We denote such an image by $\Sigma : \Omega \mapsto S^+(3)$ so that for all $x \in \Omega$, $\Sigma(x)$ is a diffusion tensor belonging to $S^+(3)$, the space of 3×3 real, symmetric, positive-definite matrices. Ω is a bounded and regular region of interest, i.e. the acquisition grid which is a subset of $\mathbb{R}^3$.

8.2.1 Derivations of statistics and gradient norms

As described in chapter 5, the manifold $\mathcal{M}$ of three-dimensional normal distributions with zero mean can be identified with the manifold $S^+(3)$ of 3×3 real, symmetric, positive-definite matrices which provides a natural means of parameterizing those probability density functions. Ultimately, we will use the fact that the Fisher information matrix corresponds to the Riemannian metric on this manifold (see [107] for example) and induces a geodesic distance $\mathcal{D}_g$. However, other distances between parameterized normal distributions (i.e. between covariance matrices and, hence, diffusion tensors) have been introduced. We will first use the Euclidean distance $\mathcal{D}_e$, then exploit the properties of the symmetrized Kullback-Leibler divergence $\mathcal{D}_j$, also known as the J-divergence [149], and finally describe the geometry of $S^+(3)$ equipped with a metric derived from the Fisher information matrix.

Euclidean probability metric

We consider $S^+(3)$ with the simple Euclidean metric. In this case, the dissimilarity measure between diffusion tensors is given by the Frobenius norm of the difference such that for all $A, B \in S^+(3)$, we have

$$\mathcal{D}_e(A,B) = |A-B|_F = \sqrt{\operatorname{tr}\left((A-B)(A-B)^T\right)} \tag{8.1}$$

where tr denotes the trace operator. Using the fact that $\nabla_X \operatorname{tr}(XY) = Y^T$ for $X, Y \in GL(n)$, it is easy to see that:

$$\nabla_A \mathcal{D}_e^2(A,B) = A - B \tag{8.2}$$

In other words, we find that the gradient of the squared Euclidean distance corresponds to the usual *difference* tangent vector. This is a symmetric matrix which can be used to compute the 6×6 covariance matrix 4.7 of a set of N diffusion tensors. Plugging equation 8.1 into equation 4.6, the empirical mean diffusion tensor is estimated as:

$$\overline{\Sigma}_e = \frac{1}{N}\sum_{i=1}^{N} \Sigma_i$$

where we denote by Σ_i the tensor located at voxel x_i in Ω. The associated covariance matrix is obtained as:

$$\Lambda_e = \frac{1}{N}\sum_{i=1}^{N} \varphi\left(\Sigma_i - \overline{\Sigma}_e\right)\varphi\left(\Sigma_i - \overline{\Sigma}_e\right)^T$$

The map $\varphi : S^+(3) \mapsto \mathbb{R}^6$ associates to each symmetric matrix $\beta_i = \Sigma_i - \overline{\Sigma}_e$ its 6 independent components. In this Euclidean setting, we can define a Gaussian distribution between diffusion tensors with the probability function:

$$p_e\left(\Sigma_i | \overline{\Sigma}_e, \Lambda_e\right) = \frac{1}{\sqrt{(2\pi)^6 |\Lambda_e|}} \exp\left(-\frac{\varphi(\beta_i)^T \Lambda_e^{-1} \varphi(\beta_i)}{2}\right) \tag{8.3}$$

with $\beta_i = \Sigma_i - \overline{\Sigma}_e$. We will use this expression, in the Euclidean case, for the probability distributions $p_{in/out}$ in equation 8.16 of section 8.3.3. Finally, the squared norm of the spatial gradient of a diffusion tensor image $\Sigma : \Omega \mapsto S^+(3)$ is given by

$$|\nabla\Sigma(x)|_e^2 = \frac{1}{2}\sum_{k=1}^{3}\sum_{\mathrm{s}=\pm1} \operatorname{tr}\left((\Sigma(x) - \Sigma(x + \mathrm{s}e_k))(\Sigma(x) - \Sigma(x + \mathrm{s}e_k))^T\right) \tag{8.4}$$

and is used in the distribution p_b defined by equation 8.17 for the Euclidean case.

J-divergence probability metric

We now adopt a more information-theoretic point of view and consider another dissimilarity measure between Gaussian probability densities known as the Kullback-Leibler divergence $\mathcal{D}_{kl}$ or *relative entropy*. This probability metric has the desirable property of being invariant under affine transformation of the density parameters, hence it is invariant under congruence transformations such that

$$\mathcal{D}_{kl}\left(A,B\right)=\mathcal{D}_{kl}\left(XAX^{T},XBX^{T}\right),\ \forall A,B\in S^{+}(3),X\in GL(3) \tag{8.5}$$

This property does not hold for the Euclidean distance previously introduced. The Kullback-Leibler divergence is defined for parametric as well as non-parametric densities. In equation 8.5, A and B actually stand for the covariance matrices of three-dimensional normal distributions $\mathscr{P}\left(r|A\right)$ and $\mathscr{P}\left(r|B\right)$ with zero mean and we have:

$$\mathcal{D}_{kl}\left(A,B\right)=\int_{\mathbb{R}^3}\mathscr{P}\left(r|A\right)\log\frac{\mathscr{P}\left(r|A\right)}{\mathscr{P}\left(r|B\right)}\,dr$$

We recall that diffusion tensors are indeed the parameters of Gaussian distributions $\mathscr{P}$ modeling the local displacement r of water molecules.
It turns out however that the Kullback-Leibler divergence is not symmetric and hence not a true metric. We will use, as in [303], its symmetrized version, or J-divergence:

$$\frac{1}{2}\int_{\mathbb{R}^3}\mathscr{P}\left(r|A\right)\log\frac{\mathscr{P}\left(r|A\right)}{\mathscr{P}\left(r|B\right)}+\mathscr{P}\left(r|B\right)\log\frac{\mathscr{P}\left(r|B\right)}{\mathscr{P}\left(r|A\right)}\,dr$$

As we will see in the next section, the J-divergence is closely related to the squared geodesic distance on $S^{+}(3)$ induced by the Fisher information matrix but only coincides with the latter for special probability densities. Hence it is natural to define:

$$\mathcal{D}_{j}\left(A,B\right)=\sqrt{\frac{1}{2}\left(\mathcal{D}_{kl}\left(A,B\right)+\mathcal{D}_{kl}\left(B,A\right)\right)}$$

As stated in [316] and used in [303], the expression of this distance is particularly simple when $\mathscr{P}$ is a three-dimensional Gaussian density:

$$\mathcal{D}_{j}\left(A,B\right)=\sqrt{\frac{1}{4}\mathrm{tr}\left(A^{-1}B+B^{-1}A\right)-6} \tag{8.6}$$

We have the following proposition:

Proposition 8.2.1.1. *The gradient of the squared distance $\mathcal{D}_j^2$ between three-dimensional normal distributions parameterized by their covariance matrix $A,B\in S^{+}(3)$ is*

$$\nabla_{A}\mathcal{D}_j^2\left(A,B\right)=\frac{1}{4}\left(B^{-1}-A^{-1}BA^{-1}\right) \tag{8.7}$$

Proof. This comes from the fact that $\nabla_{A}\mathrm{tr}\left(B^{-1}A\right)=\nabla_{A}\mathrm{tr}\left(A^{T}B^{-T}\right)=B^{-T}=B^{-1}$ and that $\nabla_{A}\mathrm{tr}\left(A^{-1}B\right)=-\left(A^{-1}BA^{-1}\right)^{T}=-A^{-1}BA^{-1}$ □

From this result, we are able to compute the covariance matrix 4.7 of a set of diffusion tensors. We just need to define the empirical mean diffusion tensor 4.6 associated to the distance $\mathcal{D}_j$ 8.6. This was already proposed in [303] as the following theorem:

Theorem 8.2.1.1. *The empirical mean diffusion tensor of a set of* N *tensors* $\{\Sigma_i\},\ i = 1, ...N$ *is given by*

$$\overline{\Sigma}_j = \underset{\Sigma \in S^+(3)}{\operatorname{Arg\,min}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{D}_j^2(\Sigma, \Sigma_i) = V^{-1/2} \left(U^{1/2} V U^{1/2}\right)^{1/2} V^{-1/2}$$

with $U = \frac{1}{N}\sum_{i=1}^{N} \Sigma_i$ *and* $V = \frac{1}{N}\sum_{i=1}^{N} \Sigma_i^{-1}$

The associated covariance matrix is obtained as:

$$\Lambda_j = \frac{1}{N} \sum_{i=1}^{N} \varphi\left(\beta_i\right) \varphi\left(\beta_i\right)^T$$

where, once again, the map φ associates to each symmetric matrix $\beta_i = -\frac{1}{4}\left(\Sigma_i^{-1} - \overline{\Sigma}_j^{-1} \Sigma_i \overline{\Sigma}_j^{-1}\right)$ its 6 independent components. In this information-theoretic setting, we now define a Gaussian distribution between diffusion tensors with the probability function:

$$p_j\left(\Sigma_i | \overline{\Sigma}_j, \Lambda_j\right) = \frac{1}{\sqrt{(2\pi)^6 |\Lambda_j|}} \exp\left(-\frac{\varphi\left(\beta_i\right)^T \Lambda_j^{-1} \varphi\left(\beta_i\right)}{2}\right) \tag{8.8}$$

with $\beta_i = -\frac{1}{4}\left(\Sigma_i^{-1} - \overline{\Sigma}_j^{-1} \Sigma_i \overline{\Sigma}_j^{-1}\right)$. We will use this expression, in the J-divergence case, for the probability distributions $p_{in/out}$ in equation 8.16 of section 8.3.3. Finally, we can easily obtain the squared norm of the spatial gradient of a DT image Σ as

$$|\nabla \Sigma(x)|_j^2 = \frac{1}{2} \sum_{k=1}^{3} \sum_{\mathrm{s}=\pm 1} \left(\frac{1}{4} \operatorname{tr}\left(\Sigma(x)^{-1} \Sigma(x + \mathrm{s} e_k) + \Sigma(x) \Sigma(x + \mathrm{s} e_k)^{-1}\right) - 6\right) \tag{8.9}$$

and use it in the distribution p_b of equation 8.17 for the J-divergence case.

Geodesic probability metric

We introduce, as in chapter 5, a last dissimilarity measure between diffusion tensors, which we claim to be more natural and powerful for the comparison of three-dimensional normal distributions. Its superiority will be demonstrated through the numerical experiments presented in section 8.4.

Using the Fisher information matrix, we already showed that the Riemannian metric for the space of three-dimensional normal distributions with zero mean, $S^+(3)$ is given, for all $A \in S^+(3)$ by:

$$g_{ij} = \langle \partial_i, \partial_j \rangle_A = \frac{1}{2} \operatorname{tr}\left(A^{-1} \partial_i A^{-1} \partial_j\right) \quad i, j = 1, ..., 6$$

Below are two examples of the metric tensor G, respectively computed for $A_1 = \mathbb{I}$ and $A_2 = \operatorname{diag}\left(\sigma_1^2, \sigma_2^2, \sigma_3^2\right)$ with $\mathbb{I}$ and $\operatorname{diag}()$ denoting the identity and diagonal matrices. They correspond to a locally isotropic diffusion process and to the more general case of an anisotropic diffusion, with variances σ_1^2, σ_2^2 and σ_3^2, whose principal axes coincide with the coordinate frame of the image:

$$G_{A_1} = \begin{pmatrix} 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 \end{pmatrix}$$

$$G_{A_2} = \begin{pmatrix} \frac{1}{2\sigma_1^4} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_1^2\sigma_2^2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sigma_1^2\sigma_3^2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2\sigma_2^4} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sigma_2^2\sigma_3^2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2\sigma_3^4} \end{pmatrix}$$

It is obvious from these examples that the second, third and fifth diagonal terms of the metric tensor receive contributions from cross-terms of the diffusion variances. Hence the factor $1/2$ in the first, fourth and sixth diagonal terms.

We recall that the associated geodesic distance $\mathcal{D}_g$ between any two elements A and B is given by the following expression:

$$\mathcal{D}_g(A, B) = \sqrt{\frac{1}{2}\mathbf{tr}(\log^2(A^{-1/2}BA^{-1/2}))} = \sqrt{\frac{1}{2}\sum_{i=1}^{3}\log^2(\eta_i)} \tag{8.10}$$

where η_i denote the 3 eigenvalues of the matrix $A^{-1/2}BA^{-1/2}$.
Apart from being a true distance, hence being positive, symmetric and verifying the triangle inequality (see [117] although no complete proof of the triangle inequality was provided by the authors), this distance is also invariant under congruence transformation (i.e. affine invariant) as well as under inversion.

It is interesting, at this stage, to study the relationship between this geodesic distance and the J-divergence. As summarized in [16], given suitable technical conditions on two *nearby* densities $\mathscr{P}(r|A)$ and $\mathscr{P}(r|A + dA)$, the zeroth and first order terms of a Taylor expansion of the Kullback-Leibler divergence around $\mathscr{P}(r|A)$ vanish. Assuming second-order differentiability of $\mathcal{D}_{kl}$, a second order expansion of $\mathcal{D}_{kl}\left(A, A + dA\right)$ yields:

$$\frac{1}{2}\int_{\mathbb{R}^3}\left(\frac{1}{\mathscr{P}^2(r|A)}\frac{\partial\mathscr{P}(r|A)}{\partial\mathbf{x}^i}\frac{\partial\mathscr{P}(r|A)}{\partial\mathbf{x}^j} - \frac{1}{\mathscr{P}(r|A)}\frac{\partial^2\mathscr{P}(r|A)}{\partial\mathbf{x}^i\partial\mathbf{x}^j}\right)\mathscr{P}(r|A)d\mathbf{x}^i\,d\mathbf{x}^j\,dr$$

which can be shown to reduce to

$$\mathcal{D}_{kl}\left(A, A+dA\right) = \frac{1}{2}\mathbb{E}\left[\frac{\partial \log \mathscr{P}(r|A)}{\partial \mathbf{x}^i}\frac{\partial \log \mathscr{P}(r|A)}{\partial \mathbf{x}^j}\right] d\mathbf{x}^i d\mathbf{x}^j$$

(if the partial derivatives commute with the integral) and which is precisely half of the squared geodesic distance between A and $A + dA$. Consequently it is easy to see that the J-divergence coincides, up to the second order, with half of the squared geodesic distance between two nearby diffusion tensors. Whenever the tensors are not *infinitesimally close*, the two distances become inconsistent. This is another reason supporting our claim that diffusion tensors statistics based on the geodesic distance should improve the quality of DTI segmentation results.

It was shown in [211] that the gradient of the squared geodesic distance writes:

$$\nabla_A \mathcal{D}_g^2\left(A, B\right) = A \log\left(B^{-1}A\right) \tag{8.11}$$

Based on this result and on the method for the computation of the mean tensor in our Riemannian setting (section 5.3.1), we will be able to estimate the covariance matrix 4.7 of a set of diffusion tensors $\{\Sigma_i\}$, $i = 1, ..., N$ and, finally, approximate a Gaussian distribution on $S^+(3)$. As presented in section 5.3.1 and [186], a closed-form expression for the empirical mean 4.6 cannot be obtained but a gradient descent algorithm was proposed. It estimates a quantity, known as the Riemannian barycenter, which exists and is unique for manifolds of non-positive sectional curvature (see [160]) like $S^+(3)$. The algorithm is based on the evolution of an initial guess of the mean (like the identify matrix $\mathbb{I}$) along the geodesics of $S^+(3)$ (equation 9.20) with a velocity given by the gradient of the variance, i.e. a tangent vector V such as

$$V = -\frac{1}{N}\sum_{i=1}^{N} \nabla_M \mathcal{D}_g^2\left(M, \Sigma_i\right) = -\frac{1}{N} M \sum_{i=1}^{N} \log\left(\Sigma_i^{-1} M\right)$$

where M denotes the evolving mean tensor. The associated covariance matrix is obtained as:

$$\Lambda_g = \frac{1}{N}\sum_{i=1}^{N} \varphi\left(\beta_i\right)\varphi\left(\beta_i\right)^T$$

where $\beta_i = -\overline{\Sigma}_g \log\left(\Sigma_i^{-1}\overline{\Sigma}_g\right)$ and φ associates to each β_i its 6 independent components. The notion of Gaussian distribution was generalized to random samples of primitives belonging to a Riemannian manifold in [235] where more details can be found regarding this particular point. From this work, we have proposed in section 5.3 and [186] a definition of the Gaussian law between diffusion tensors which can be approximated as follows for a covariance matrix Λ_g of small variance $\sigma^2 = \text{tr}(\Lambda_g)$:

$$p_g(\Sigma_i|\overline{\Sigma}_g, \Lambda_g) \simeq \frac{1}{\sqrt{(2\pi)^6 |\Lambda_g|}} \exp \frac{-\varphi\left(\beta_i\right)^T \Gamma \varphi\left(\beta_i\right)}{2} \tag{8.12}$$

where β_i is defined as $\beta_i = -\overline{\Sigma}_g \log\left(\Sigma_i^{-1}\overline{\Sigma}_g\right)$ and the concentration matrix is $\Gamma \simeq \Lambda_g^{-1} - \mathcal{R}/3$, with $\mathcal{R}$ the Ricci tensor at the mean $\overline{\Sigma}_g$. The computation of the Ricci tensor $\mathcal{R}$ can be performed on the basis of closed-form expressions for the metric and the Riemann tensor provided in [271] and simply involving traces of matrix products (see section 5.2.2). As we will point out in section 8.4, our numerical experiments have shown that the Ricci tensor exhibits a difference of at least 2 orders of magnitude with the inverse of the covariance matrix. Hence we can approximate Γ by Λ_g^{-1}.

We will use p_g, in the geodesic case, for the probability distributions $p_{in/out}$ in equation 8.16 of section 8.3.3. Finally, the squared norm of the spatial gradient of a DT image can be estimated as follows:

$$|\nabla\Sigma(x)|_g^2 = \frac{1}{2}\sum_{k=1}^{3}\sum_{\mathrm{s}\pm 1}\left(\frac{1}{2}\mathrm{tr}\left(\log^2\left(\Sigma(x)^{-1/2}\Sigma(x+\mathrm{s}e_k)\Sigma(x)^{-1/2}\right)\right)\right) \tag{8.13}$$

and subsequently used in the distribution p_b of equation 8.17 in the geodesic case.

Summary and numerical examples

We summarize, in table 8.1, the expressions of the squared distance, its gradient, and of the mean tensor for the Euclidean, J-divergence and geodesic cases. The evaluation of the squared distance and its gradient for the matrices A_1 and B_1 respectively given below shows a good coherence (although the Euclidean distance is quite larger than the other two) and, more importantly, illustrates the fact that the J-divergence accurately approximates half of the squared geodesic distance when the tensors are relatively close:

$$\begin{pmatrix} 0.9878 & -0.0527 & 0.0050 \\ -0.0527 & 1.0112 & -0.0372 \\ 0.0050 & -0.0372 & 1.0391 \end{pmatrix}, \quad \begin{pmatrix} 1.0384 & -0.0012 & 0.0107 \\ -0.0012 & 1.0056 & -0.0060 \\ 0.0107 & -0.0060 & 1.0233 \end{pmatrix}$$

$$\mathcal{D}_e^2(A_1, B_1) = 0.010158, \quad \nabla_{A_1}\mathcal{D}_e^2(A_1, B_1) = \begin{pmatrix} -0.0506 & -0.0515 & -0.0057 \\ -0.0515 & 0.0056 & -0.0312 \\ -0.0057 & -0.0312 & 0.0158 \end{pmatrix}$$

$$\mathcal{D}_j^2(A_1, B_1) = 0.002526, \quad \nabla_{A_1}\mathcal{D}_j^2(A_1, B_1) = \begin{pmatrix} -0.0274 & -0.0266 & -0.0040 \\ -0.0266 & -0.0002 & -0.0147 \\ -0.0040 & -0.0147 & 0.0066 \end{pmatrix}$$

$$\mathcal{D}_g^2(A_1, B_1) = 0.005050, \quad \nabla_{A_1}\mathcal{D}_g^2(A_1, B_1) = \begin{pmatrix} -0.0480 & -0.0503 & -0.0048 \\ -0.0503 & 0.0074 & -0.0314 \\ -0.0048 & -0.0314 & 0.0164 \end{pmatrix}$$

On the contrary, if we consider the matrices A_2 and B_2, which are much more dif-

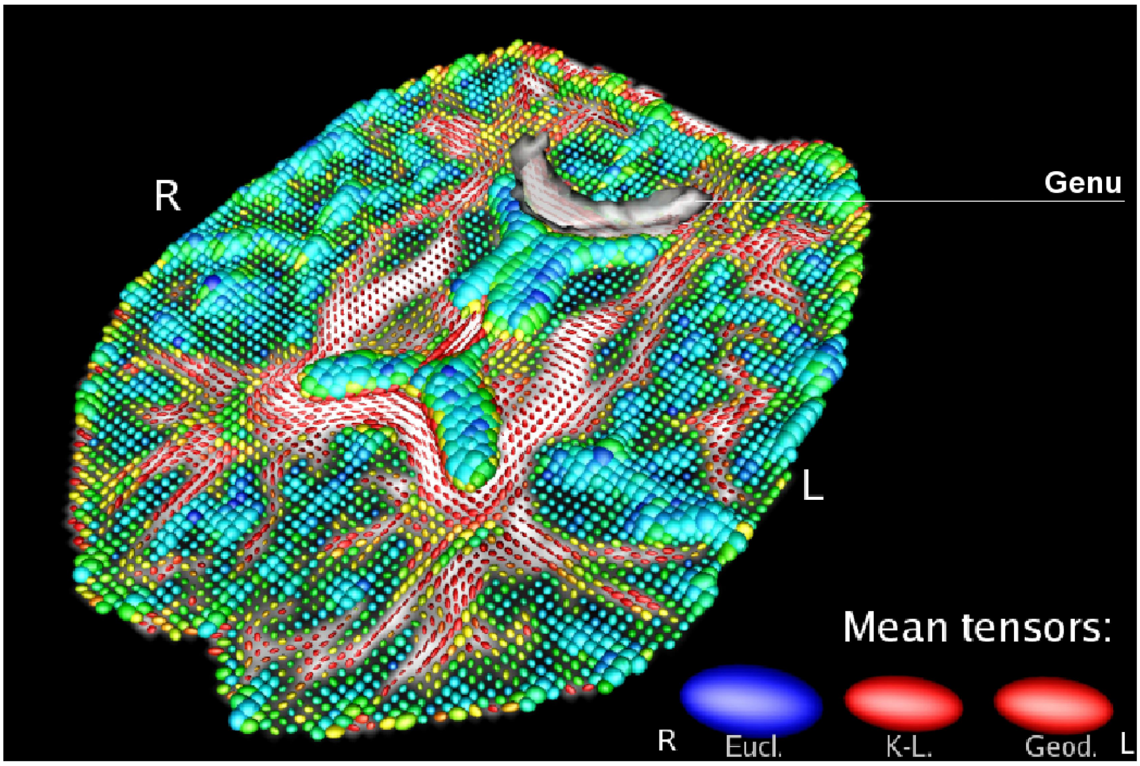


Figure 8.1: Statistics in the genu of the corpus callosum (R:right, L:left)

ferent than A_1 and B_1, we find out that the J-divergence becomes sensibly different from half of the squared geodesic distance:

$$\begin{pmatrix} 1.0696 & -0.0563 & 0.4035 \\ -0.0563 & 0.5621 & 0.1068 \\ 0.4035 & 0.1068 & 1.4086 \end{pmatrix}, \quad \begin{pmatrix} 1.2813 & 0.2320 & 0.0327 \\ 0.2320 & 1.2782 & 0.1965 \\ 0.0327 & 0.1965 & 0.9392 \end{pmatrix}$$

$$\mathcal{D}_e^2(A_2, B_2) = 1.111446, \quad \nabla_{A_2}\mathcal{D}_e^2(A_2, B_2) = \begin{pmatrix} -0.2117 & -0.2883 & 0.3708 \\ -0.2883 & -0.7160 & -0.0897 \\ 0.3708 & -0.0897 & 0.4695 \end{pmatrix}$$

$$\mathcal{D}_j^2(A_2, B_2) = 0.329119, \quad \nabla_{A_2}\mathcal{D}_j^2(A_2, B_2) = \begin{pmatrix} -0.2029 & -0.2875 & 0.1765 \\ -0.2875 & -0.8811 & 0.0783 \\ 0.1765 & 0.0783 & 0.0880 \end{pmatrix}$$

$$\mathcal{D}_g^2(A_2, B_2) = 0.621560, \quad \nabla_{A_2}\mathcal{D}_g^2(A_2, B_2) = \begin{pmatrix} -0.0648 & -0.1598 & 0.4483 \\ -0.1598 & -0.4424 & -0.0799 \\ 0.4483 & -0.0799 & 0.6295 \end{pmatrix}$$

Now, in order to compare the statistics derived from each distance, we have manually segmented the genu of the corpus callosum on a DTI data-set used in the last section of this chapter. This is a well-known region of the brain white matter (figure 8.1) where fibers are essentially aligned in a right-left fashion, i.e. along the x_1 axis on an

Distance	Euclidean	J-divergence	Geodesic
$\mathcal{D}^2(A,B)$	$\mathrm{tr}\left((A-B)(A-B)^T\right)$	$\frac{1}{4}\left(\mathrm{tr}\left(A^{-1}B+B^{-1}A\right)-6\right)$	$\frac{1}{2}\mathrm{tr}\left(\log^2\left(A^{-1/2}BA^{-1/2}\right)\right)$
$\nabla_A\mathcal{D}^2(A,B)$	$A-B$	$\frac{1}{4}(B^{-1}-A^{-1}BA^{-1})$	$A\log\left(B^{-1}A\right)$
$\overline{\Sigma}$	$\frac{1}{N}\sum_{i=1}^N\Sigma_i$	$V^{-1/2}\left(U^{1/2}VU^{1/2}\right)^{1/2}V^{-1/2}$ with $U=\frac{1}{N}\sum_{i=1}^N\Sigma_i$ and $V=\frac{1}{N}\sum_{i=1}^N\Sigma_i^{-1}$	Algorithm 1

Table 8.1: Squared distances and their gradient for $A, B \in S^+(3)$, mean tensor.

axial slice. Consequently, the tensors in this region are very anisotropic with a major eigenvector close to $(1,0,0)^T$. This resulted in a set $\{\Sigma_i\}$ of $N = 614$ tensors. The ellipsoids presented in the bottom-right corner of figure 8.1 represent the mean tensor respectively computed, from left to right, with the Euclidean distance, J-divergence and geodesic distance (the color encodes their relative fractional anisotropy). Visually, we can see that the Euclidean mean is somehow more oblate than the other two ellipsoids. This can be explained by the fact that Euclidean averaging is blind to the spectral components of the tensors (eigenvalues and eigenvectors) and has a tendency to mix them. We now present the estimated statistics for each distance (We scaled by a factor 2 the values obtained for the J-divergence to make the comparisons easier).

Euclidean probability metric:

$$\overline{\Sigma}_e = \begin{pmatrix} 2.6923 & -0.1334 & 0.0347 \\ -0.1334 & 1.3947 & 0.0526 \\ 0.0347 & 0.0526 & 1.1228 \end{pmatrix}, \quad \mathrm{tr}\left(\Lambda_e\right) = 3.0615$$

$$\Lambda_e = \begin{pmatrix} 2.8685 & -0.2261 & 0.0589 & 0.6690 & 0.0892 & 0.2083 \\ -0.2261 & 0.0178 & -0.0046 & -0.0527 & -0.0070 & -0.0164 \\ 0.0589 & -0.0046 & 0.0012 & 0.0137 & 0.0018 & 0.0042 \\ 0.6690 & -0.0527 & 0.0137 & 0.1560 & 0.0208 & 0.0485 \\ 0.0892 & -0.0070 & 0.0018 & 0.0208 & 0.0027 & 0.0064 \\ 0.2083 & -0.0164 & 0.0042 & 0.0485 & 0.0064 & 0.0151 \end{pmatrix}$$

***J*-divergence probability metric:**

$$\overline{\Sigma}_j = \begin{pmatrix} 2.2901 & -0.1063 & 0.0296 \\ -0.1063 & 1.0833 & 0.0455 \\ 0.0296 & 0.0455 & 0.8775 \end{pmatrix}, \quad \mathrm{tr}\left(\Lambda_j\right) = 1.5161$$

$$\Lambda_j = \begin{pmatrix} 0.0369 & 0.0075 & -0.0015 & 0.0139 & -0.0105 & 0.0816 \\ 0.0075 & 0.0692 & 0.0035 & 0.0137 & -0.0051 & 0.0236 \\ -0.0015 & 0.0035 & 0.0413 & -0.0135 & 0.0019 & -0.0142 \\ 0.0139 & 0.0137 & -0.0135 & 0.4958 & -0.0405 & 0.5147 \\ -0.0105 & -0.0051 & 0.0019 & -0.0405 & 0.0432 & -0.0552 \\ 0.0816 & 0.0236 & -0.0142 & 0.5147 & -0.0552 & 0.8296 \end{pmatrix}$$

Geodesic probability metric:

$$\overline{\Sigma}_g = \begin{pmatrix} 2.3296 & -0.1088 & 0.0312 \\ -0.1088 & 1.1102 & 0.0523 \\ 0.0312 & 0.0523 & 0.8912 \end{pmatrix}, \quad \mathrm{tr}\,(\Lambda_g) = 2.0370$$

$$\Lambda_g = \begin{pmatrix} 0.7706 & -0.0297 & 0.0207 & -0.0003 & -0.0267 & 0.1970 \\ -0.0297 & 0.3156 & 0.0431 & -0.0461 & -0.0016 & -0.0041 \\ 0.0207 & 0.0431 & 0.1332 & -0.0113 & -0.0086 & 0.0073 \\ -0.0003 & -0.0461 & -0.0113 & 0.4592 & 0.0094 & 0.3010 \\ -0.0267 & -0.0016 & -0.0086 & 0.0094 & 0.0276 & 0.0068 \\ 0.1970 & -0.0041 & 0.0073 & 0.3010 & 0.0068 & 0.3306 \end{pmatrix}$$

It is clear that there are important differences between these three approaches. They are hard to interpret though on such a simple example but their effect on the segmentation results will be outlined in the section 8.4.

In the next section, we set up a unified Bayesian formulation of the segmentation problem that will be used throughout this chapter. It relies on the different possible estimates of the mean $\overline{\Sigma}$ and covariance matrix Λ (equation 4.7) to evaluate the likelihood of a diffusion tensor to belong to a given subset of the DTI data-set. This will be used in equation 8.16. We recall that we will consider 3 different cases associated to the Euclidean distance 8.1, the J-divergence 8.6 and the geodesic distance 8.10. Within these 3 different frameworks, we have shown how to approximate a Gaussian distribution between diffusion tensors (see equations 8.3, 8.8 and 8.12) by using the information provided by the gradient of the squared geodesic distance (see equations 8.2, 8.7 and 8.11). We will also exploit the information provided by the norm of the tensor field spatial gradient (see equations 8.4, 8.9 and 8.13) to localize the boundaries between structures of the brain white matter and avoid mixing them through the boundary term 8.17 in our energy 8.18.

8.3 SEGMENTATION BY SURFACE EVOLUTION

We recall that our goal is to compute the optimal 3D surface separating an anatomical structure of interest from the rest of a DTI data-set. The statistical surface evolution, as developed in [258], is a well-suited framework for our segmentation problem. We hereafter summarize the important notions of this technique.

8.3.1 Bayesian formulation for image partitioning

Following general works on image segmentation [175], [325], [17], [228], we seek the optimal partition of the image domain Ω by maximizing the *a posteriori* frame partition probability $p(\mathcal{P}(\Omega)\,|\,\Sigma)$ for the observed diffusion tensor image Σ. The Bayes rule allows to express this probability as:

$$p(\mathcal{P}(\Omega)\,|\,\Sigma) \propto p(\Sigma\,|\,\mathcal{P}(\Omega))p(\mathcal{P}(\Omega)) \tag{8.14}$$

This formulation yields a separation of the image-based cues from the geometric properties of the boundary given by $\mathcal{P}(\Omega)$. While being valid for any number of regions, we restrict this formulation to binary partitions: the structure of interest and the background. The image partition can be represented as the zero-crossing of a level-set function ϕ [88], [89], [223],[59]. Noting $\mathcal{B}$ the interface between the two regions Ω_{in} and Ω_{out}, ϕ is constructed as the signed distance function to $\mathcal{B}$:

$$\begin{cases} \phi(x) = 0, & \text{if } x \in \mathcal{B} \\ \phi(x) = \mathscr{D}(x,\mathcal{B}), & \text{if } x \in \Omega_{in} \\ \phi(x) = -\mathscr{D}(x,\mathcal{B}), & \text{if } x \in \Omega_{out} \end{cases}$$

where $\mathscr{D}(x,\mathcal{B})$ stands for the Euclidean distance between x and $\mathcal{B}$. Hence, the optimal partition is obtained by maximizing: $p(\phi|\Sigma) \propto p(\Sigma|\phi)p(\phi)$. At this stage, these two terms still need to be defined. For this purpose, several assumptions on the structure of interest need to be introduced. In the following, a smoothness constraint is imposed with the term $p(\phi)$ while $p(\Sigma|\phi)$ expresses the likelihood of the diffusion tensors to be inside, outside or on the boundary of the structure. This yields an optimization criterion similar to the *Geodesic Active Regions* presented in [228].

8.3.2 Smoothness constraint

The second term of equation 8.14 expresses the probability of the interface to represent the structure of interest and can be used to introduce prior shape knowledge. For the segmentation of diffusion tensor images, we have no high level prior information but we can use this term to impose shape regularity. Such a constraint can be obtained by favoring structures with a smaller surface $|\mathcal{B}|$ with $p(\phi) \propto \exp(-\nu|\mathcal{B}|)$. This can be expressed with ϕ by introducing the Dirac function [324]:

$$p(\phi) \propto \exp\left(-\nu \int_\Omega \delta(\phi)|\nabla\phi(x)|\,dx\right) \tag{8.15}$$

8.3.3 Data term

To further specify the image term $p(\Sigma|\phi)$, we introduce some hypothesis. First, for a given level-set ϕ, we can classify the voxels into three classes: inside, outside or on the boundary. Then, we can define the probability density functions of a diffusion tensor for each class: p_{in}, p_{out} and p_b. Assuming the diffusion tensors to be independent and identically distributed realizations of the corresponding random process, the data term is given by:

$$p(\Sigma|\phi) = \prod_{x\in\Omega_{in}} p_{in}(\Sigma(x)) \;\;.\; \prod_{x\in\Omega_{out}} p_{out}(\Sigma(x)) \;\;.\; \prod_{x\in\mathcal{B}} p_b(\Sigma(x)) \tag{8.16}$$

This gives two different types of probability distributions: region-based with $p_{in/out}$ and boundary-based with p_b. p_{in} and p_{out} are given by the Gaussian distributions on tensors introduced in section 8.2.1 p_e (equation 8.3) p_j, (equation 8.8) and p_g (equation 8.12). The parameters of these laws may be known a priori but in the absence of such information, they are introduced as unknown parameters.

Regarding p_b, the probability should be close to one for high gradients of the diffusion tensors field and around zero for small variations. This leads to:

$$p_b(\Sigma(x)) \propto \exp\left(-g_\alpha(|\nabla\Sigma(x)|)\right) \tag{8.17}$$

with $g_\alpha(u) = 1/(1+u^\alpha)$. This type of boundary term is the basis of several works referred to as active contours [53] and, often, $\alpha = 1$ or 2 is chosen. For the sake of readability, we will use the short notation $g_\alpha(\Sigma(x))$. $|\nabla\Sigma(x)|$ will be computed by using equation 8.4 for the Euclidean case, equation 8.9 for the J-divergence case, or equation 8.13 for the geodesic case.

8.3.4 Energy formulation

Maximizing the *a posteriori* segmentation probability is equivalent to minimizing its negative logarithm. Integrating the regularity constraint 8.15 and the image term 8.16, we end up with the following energy:

$$\begin{aligned} E(\phi, \overline{\Sigma}_{in/out}, \Lambda_{in/out}) = & \\ & \nu \int_\Omega \delta(\phi)|\nabla\phi|\, dx + \int_\Omega \delta(\phi)|\nabla\phi| g_\alpha(\Sigma(x))\, dx \\ & - \int_{\Omega_{in}} \log p(\Sigma(x)|\overline{\Sigma}_{in}, \Lambda_{in}) dx - \int_{\Omega_{out}} \log p(\Sigma(x)|\overline{\Sigma}_{out}, \Lambda_{out}) dx \end{aligned} \tag{8.18}$$

The boundary term of this energy corresponds to the Geodesic Active Contours [53] and naturally includes a regularization[1] on the interface. Following [162], [259], an alternate minimization is employed to perform the optimization for the two types of unknown parameters. For given statistical parameters, the Euler-Lagrange equations are computed to derive the implicit front evolution:

$$\frac{\partial\phi}{\partial t} = \delta(\phi)\left((\nu + g_\alpha(\Sigma))\,\mathrm{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right) + \frac{\nabla\phi}{|\nabla\phi|}\cdot\nabla g_\alpha(\Sigma) + \log\frac{p(\Sigma|\overline{\Sigma}_{in}, \Lambda_{in})}{p(\Sigma|\overline{\Sigma}_{out}, \Lambda_{out})}\right) \tag{8.19}$$

while the statistics can be updated after each evolution of ϕ from their empirical estimates, as described in section 8.2.1. More details on this level-set based optimization can be found in [59], [259], where different applications were considered.

The right-hand side of equation 8.19, between parenthesis, corresponds to the magnitude of the velocity used to deform each point of the evolving surface $\mathcal{B}$ along its normal at that point. The purpose of the next section will be to evaluate the influence

---

[1] The regularity term 8.15 could be included in p_b by replacing g_α by $g_{\alpha,\nu} = \nu + g_\alpha$.

of the choice of the density function p, which can be taken in its Euclidean version p_e (equation 8.3), J-divergence version p_j (equation 8.8) or geodesic version p_g (equation 8.12). We will describe several numerical experiments in order to evaluate the respective performances of each probability metrics for our DTI segmentation task. We demonstrate that the Riemannian statistical tools presented in section 8.2.1 achieve the best results.

8.4 RESULTS AND VALIDATION

We begin our numerical experiments with three different synthetic data-sets of increasing complexity in order to emphasize the respective virtue of the Euclidean, Kullback-Leibler and geodesic probability metrics. We then apply our algorithm to a biological rat spinal cord phantom. Finally, we consider real DTI data-sets on which we perform the segmentation of the corpus callosum.

In practice, there are a few important points that must be carefully addressed when implementing and running our segmentation algorithm: When dealing with real DTI data, we use a mask of the brain so that tensors statistics of Ω_{out} are not corrupted by the signal from the outside of the brain. Regarding the initialization, we noticed and will demonstrate that our method is very robust. We will show that the geodesic distance is indeed the only metric capable of representing, through the associated Gaussian distribution, a smoothly varying tensor field with relatively high variability. Next, there are two parameters that have to be chosen: The first one is the value of ν in equation 8.15. It constrains the smoothness of the surface and is usually set in the range 1 to 10. The second parameter arises from the very definition of the Gaussian distribution on $S^+(3)$ presented in section 8.2.1. The main hypothesis for this definition to be valid is that the trace of the covariance matrix Λ_g should be small and this means that we restrict ourselves to concentrated distributions. Hence, we set a threshold for the variance which, whenever reached, induces the end of the update for the statistical parameters. We then let the surface evolve while using a fixed mean and covariance matrix to model the distributions of the tensors in Ω_{in}/Ω_{out}. The threshold is chosen in the range $[0.01, 0.1]$ for tensors with components around 1.0. We noticed that the variance, after a few iterations of increase at the very beginning of the algorithm, keeps decreasing as the segmentation process converges. Consequently, a careful selection of this parameter is not critical. Finally, we improved the computational efficiency of the method using the geodesic distance by noticing and experimentally verifying that, in equation 8.12, the term involving the 6×6 Ricci tensor $\mathcal{R}/3$ can be neglected since we have found, in our numerical experiments, a difference of at least 2 orders of magnitude between Λ_g^{-1} and $\mathcal{R}/3$.

Regarding the computational cost of the method, we should point out that it is fairly efficient since the results presented on figure 8.18 and 8.19 were respectively obtained, on $128 \times 128 \times 64$ images, in 5 and 10 minutes on a 1.7 GHz Pentium M processor with 1 Gb of RAM.

8.4.1 Synthetic examples

Each of the three synthetic data-sets consists of a $40 \times 40 \times 40$ 3D tensor field with a main pattern and a background. The tensors follow the shape of the pattern so that, as the shape becomes more twisted, the tensors variability increases and makes it more difficult for the algorithm to recover the entire shape. The regularity factor ν is set to 1 for all the experiments. The initialization is done by the means of one or two spheres (see figures below). Finally, the mean tensor and covariance matrix used for the generation of noise in all the experiments are the following:

$$\overline{\Sigma} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\Lambda = \begin{pmatrix} 0.0885 & -0.0568 & -0.0260 & 0.0119 & -0.0394 & 0.0035 \\ -0.0568 & 0.0701 & 0.0039 & -0.0070 & 0.0122 & -0.0112 \\ -0.0260 & 0.0039 & 0.0183 & -0.0023 & 0.0218 & 0.0095 \\ 0.0119 & -0.0070 & -0.0023 & 0.0078 & -0.0113 & 0.0010 \\ -0.0394 & 0.0122 & 0.0218 & -0.0113 & 0.0416 & 0.0118 \\ 0.0035 & -0.0112 & 0.0095 & 0.0010 & 0.0118 & 0.0160 \end{pmatrix}$$

The generation of random tensors, i.e. Gaussian noise in $S^+(3)$, is usually addressed by simply building symmetric matrices with i.i.d. normally distributed components and then enforcing their positivity. As discussed in section 5.4.1, the main drawback of this approach is that it leaves no grasp on the actual distribution of tensors. We proposed in this section and in [186], to use the equation 8.12, to generate random tensors with a known mean $\overline{\Sigma}$ and covariance matrix Λ. We use this method in the following and recall that it is fairly simple since all we need to do is to randomly choose the initial velocities $\{\beta_i\}, i = 1, ..., N$ of the geodesics in $S^+(3)$ joining the imposed mean tensor $\overline{\Sigma}$ to the random elements Σ_i. In practice, this operation is performed in local coordinates so that we only need to draw random samples of the $\varphi\left(\beta_i\right) \in \mathbb{R}^6$ with zero mean and covariance matrix Λ. The Σ_i are easily obtained by using the expression $\beta_i = -\overline{\Sigma} \log\left(\Sigma_i^{-1}\overline{\Sigma}\right)$ (table 8.1).

The Y tensor field

We start with a simple example composed by a *diverging* tensor field and a background of isotropic tensors (figure 8.2). Within the Y shape, tensors fractional anisotropy decreases as we get away from the center-line. This example is relatively simple since the tensors variability stays low and the segmentation procedure succeeds with the three probability metrics. One important difference must be noted

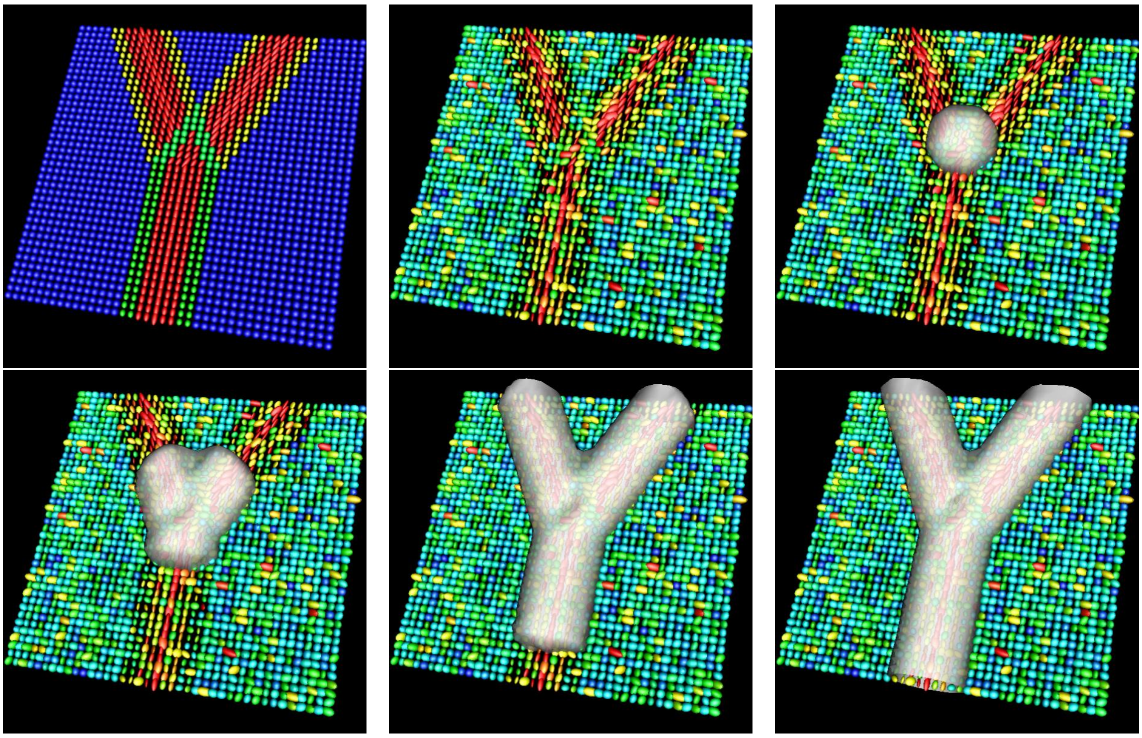


Figure 8.2: Segmentation of the Y tensor field. From left to right: Axial slice of the original and noisy data-set, evolution of the segmentation (color map indicates FA)

though: By comparison with the Euclidean distance, which requires 45 iterations to segment the Y structure, the process converges faster when the J-divergence is used (30 iterations), and relatively faster with the geodesic distance (28 iterations). This is easily explained by the fact that the velocity of the evolving surface, at location x of the image Σ, is directly related to the likelihood of tensor $\Sigma(x)$ to belong to Ω_{in} or Ω_{out}. It is hence a first argument in favor of our claim that the geodesic probability metrics yields more adequate tensor statistics.

The *torus* tensor field

Next, we consider another example where the tensors follow the tangent of the centerline of a torus (figure 8.3) and share the same eigenvalues. This yields a higher orientational variability of the tensors. A direct consequence of this increased variability is the failure of the segmentation process when we use the Euclidean probability metric. The evolution is presented on figure 8.5. The initial sphere is setup so that it covers half of the torus and contains the part of the background situated 'inside' the torus. The surface evolution falls into a local minimum and is unable to recover the desired shape. On the contrary, the J-divergence and geodesic distance behave consistently and succeed to segment the complete torus (figure 8.6). We notice, as in the previous example, that the segmentation using the geodesic distance converges faster (20 iterations) than the one relying on the J-divergence (27 iterations). The result presented

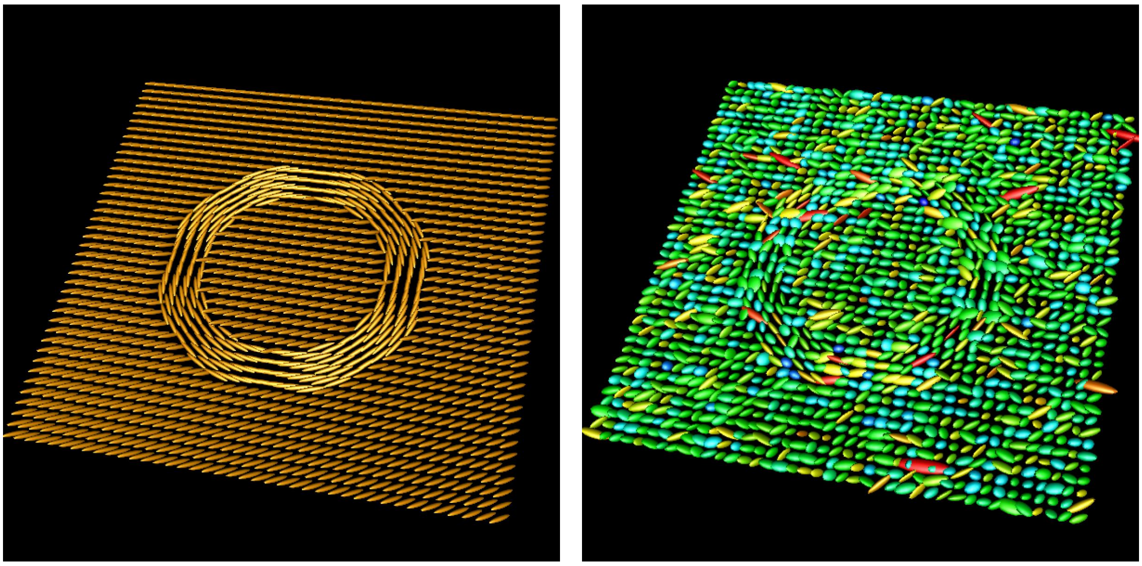


Figure 8.3: Axial slice of the original and noisy *torus* tensor field

for the Euclidean metric on figure 8.5 is the final state after 600 iterations.

The *helix* tensor field

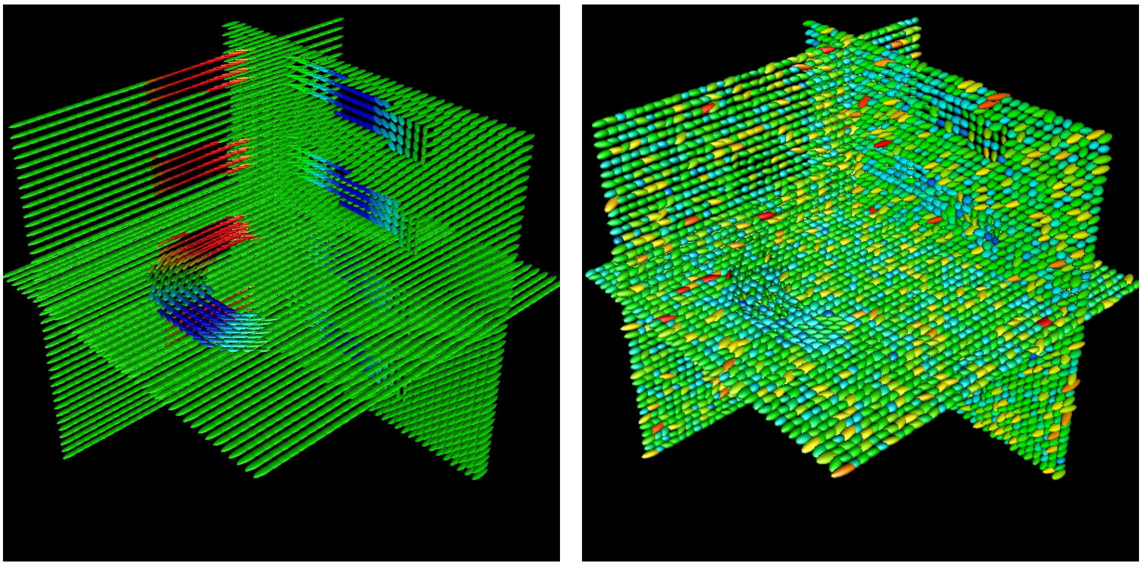


Figure 8.4: Axial, coronal and sagittal slices of the original and noisy *helix* tensor field

The last synthetic data-set that we consider is the *helix* tensor field presented on figure 8.4. It is composed of a background with anisotropic tensors aligned on the x_1 axis of the 3D field and an helix containing tensors oriented along the tangent of its center-line. The fractional anisotropy of the helix tensors varies around each spire. Moreover, the tensors orientation spans a broader range of possibilities than

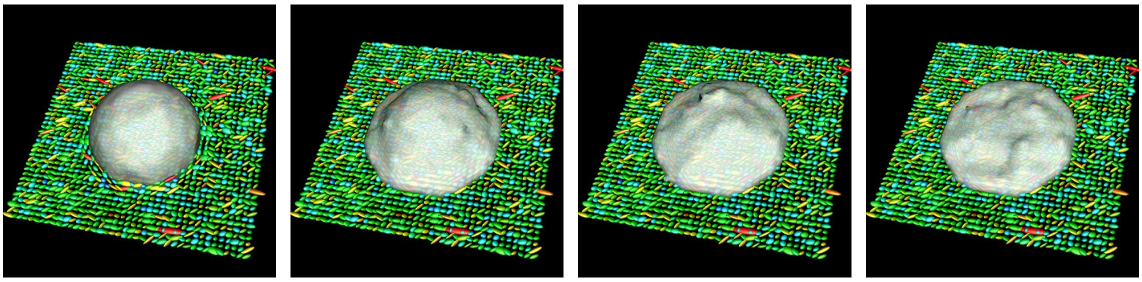


Figure 8.5: Failure of the *torus* segmentation with the Euclidean distance (right: final state after 600 iterations)

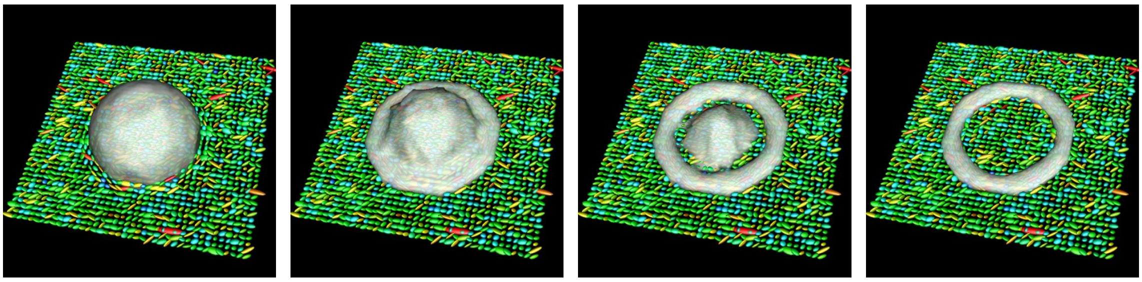


Figure 8.6: Successful segmentations of the *torus* with the J-divergence and geodesic distances (right: final state after 27 iterations with the J-divergence or 20 iterations with the geodesic distance. The evolutions are similar.)

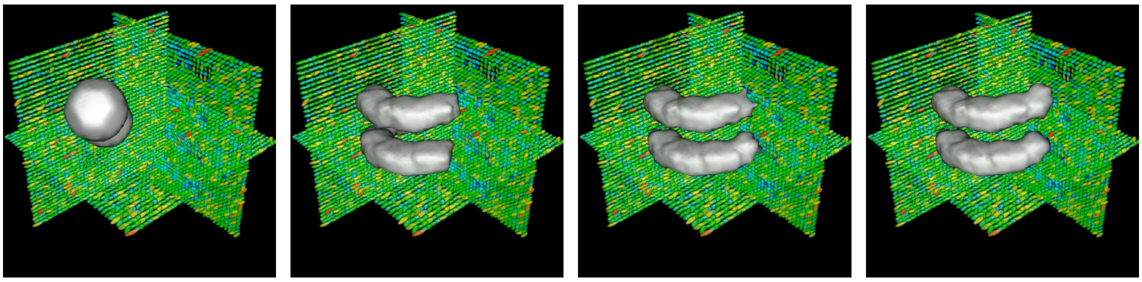


Figure 8.7: Failure of the *helix* segmentation with the Euclidean distance

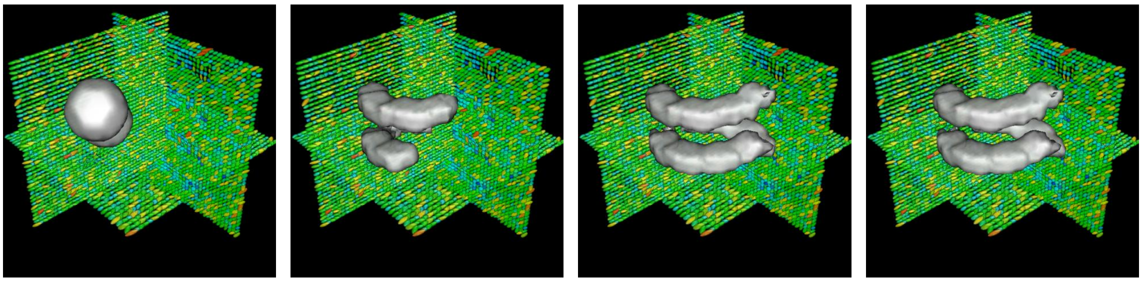


Figure 8.8: Failure of the *helix* segmentation with the J-divergence

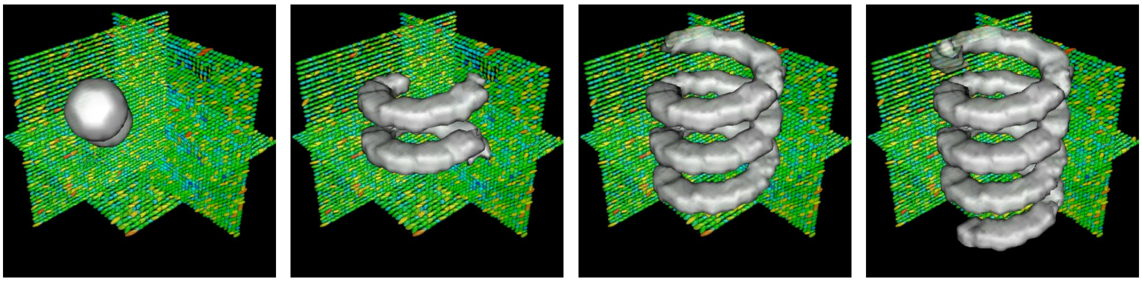


Figure 8.9: Successful segmentation of the *helix* with the geodesic distance

in the *torus* case since it changes along the x_1, x_2 and x_3 axes. This is certainly an example on which it is desirable for our segmentation algorithm to succeed since this tensors variation pattern is fairly realistic and may be found in real DT images. As a matter of fact, only the statistics computed with the geodesic distance enable our segmentation framework to achieve a correct extraction of the helix. The initialization consists of 2 small spheres overlapping the helix and the background. As we can see on figure 8.7 and 8.8, the surface evolution quickly stops when it uses the Euclidean distance or the J-divergence, even though the latter propagates further than the former as we could have expected. The local minima are respectively reached after 130 and 80 iterations. Using the geodesic distance, the complete helix is recovered (figure 8.9) after 300 iterations.

This last example undoubtedly demonstrates the superiority of the Riemannian framework over the statistics derived from the Euclidean or Kullback-Leibler dissimilarity measures.

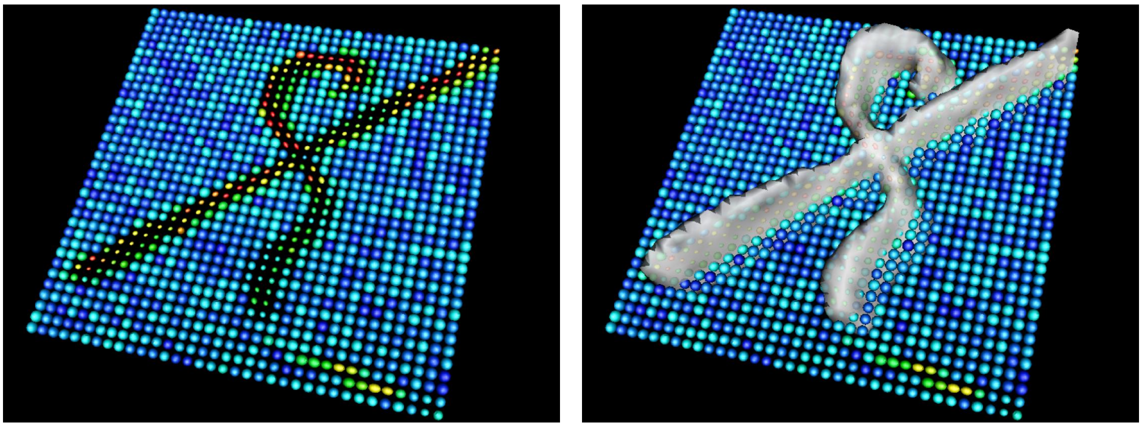


Figure 8.10: Segmentation of the rat spinal cords phantom - Axial slice of the data-set (left) and final segmentation using the geodesic distance (right)

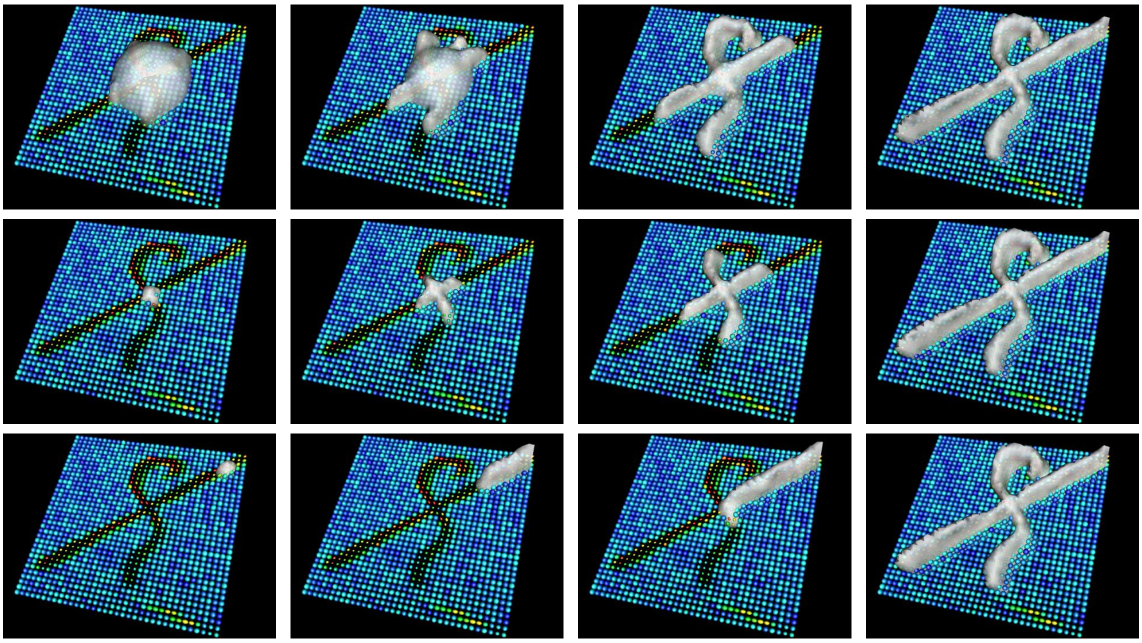


Figure 8.11: Segmentation of the rat spinal cords phantom with the geodesic distance and a large sphere initialization (1st row), a small sphere initialization (2nd row) and initialization at one end of a cord (3rd row).

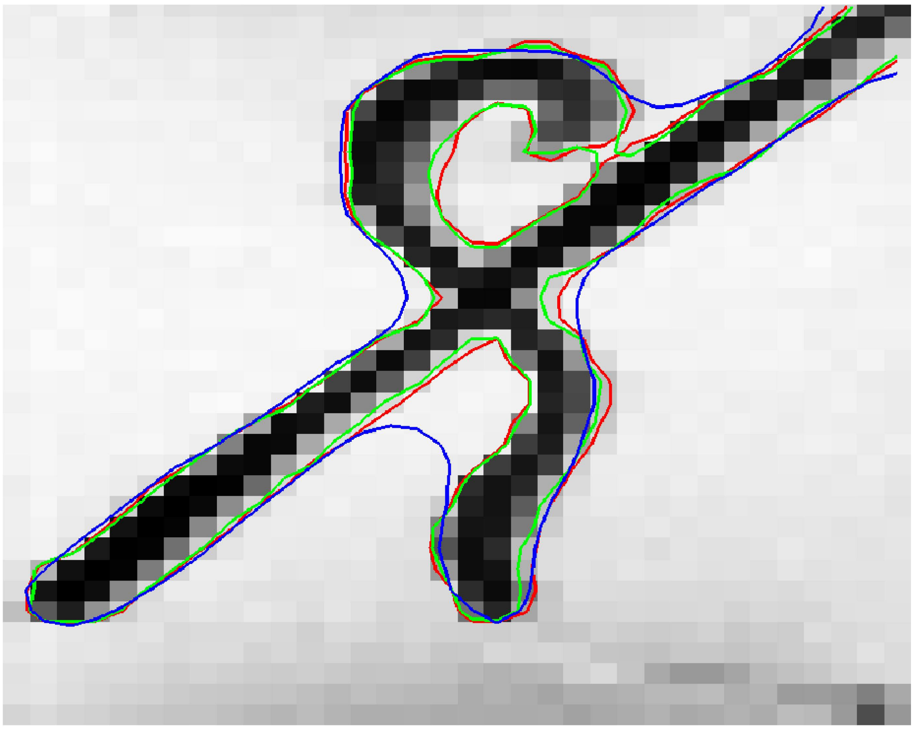


Figure 8.12: Comparison of the rat spinal cords phantom segmentation results with the 3 distances (colors as on figure 8.17)

8.4.2 Biological phantom data-set

We tested our algorithm on a biological phantom produced by J. Campbell *et al.* at the McConnell Brain Imaging Center and Montreal Neurological Institute [49, 50] and was created from two excised Sprague-Dawley rat spinal cords embedded in 2% agar. Diffusion weighted images (DWI) were acquired along 90 gradient directions with a b factor of 1000 s.mm^{-2} on a 1.5 Tesla Siemens Sonata scanner using a knee coil. Diffusion tensors are computed by using the method proposed in section 6.1.2. An axial slice of the resulting DT image is presented on figure 8.10 together with a 3D surface modeling the spinal cords.

This data-set is well suited to evaluate the robustness to the initialization of our segmentation framework as well as to demonstrate the importance of the Riemannian framework to achieve good segmentation results.
Figure 8.11 illustrates the evolution of the segmentation process, using the geodesic distance, for 3 very different initializations: One large sphere and one small sphere centered at the cord crossing, and one small sphere placed at one end of a cord. These three examples yield the same final result, thus experimentally showing the non-dependence of our method on the initialization. Finally, figure 8.12 displays, on top of the Apparent Diffusion Coefficient image, the three final segmentation results obtained by using the Euclidean distance (blue), J-divergence (green) and geodesic distance (red). We can see that the most accurate result is obtained with the latter. Especially, it is interesting to note that, in the upper right part of the image where the two cords are very close to each other, only the geodesic distance is able to distinguish between the two structures. This is another example of the better properties of the Riemannian statistics to model the distribution of the diffusion tensors.
In the next section, we will show that the Riemannian statistical approach also performs better on human brain diffusion tensor images.

8.4.3 Real DTI data-sets

Method and tensors estimation

Diffusion weighted images were acquired at the Center for Magnetic Resonance Research, University of Minnesota, on a 3 Tesla Siemens Magnetom Trio whole-body clinical scanner. Measurements were made along 12 gradient directions. Acquisition parameters were: b factor = 1000 s.mm^{-2}, TE = 92 ms and TR = 1.7 s. The images were obtained on 64 evenly spaced axial planes with 128×128 pixels per slice. The voxel size is $2 \times 2 \times 2$ mm. As for the biological rat spinal cord phantom, diffusion tensors are computed by using the method proposed in [184]. An example of the resulting DT image is presented on figure 8.13. It uses a red-green-blue color scheme to encode the tracts orientation. Following [300] and [150], we indicate the names of major fiber

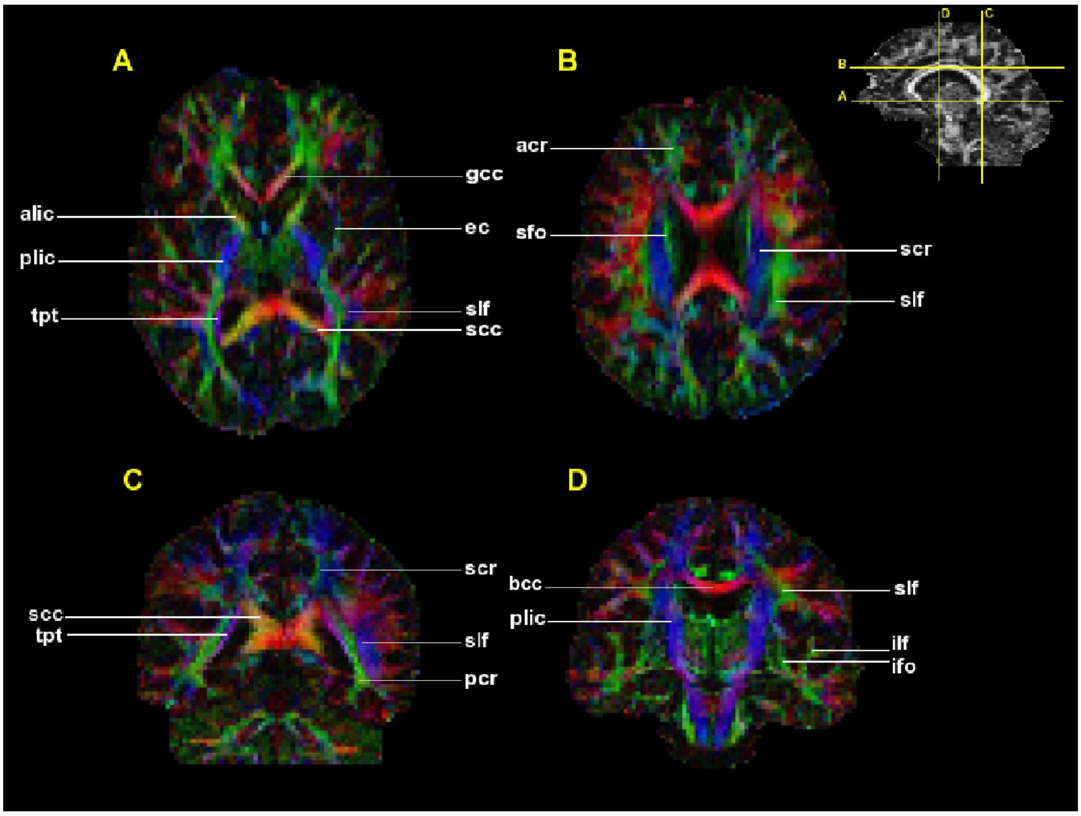


Figure 8.13: Axial (A,B) and coronal (C,D) DT images color maps and major fiber bundles. *acr* = anterior region of the corona radiata, *alic* = anterior limb of the internal capsule, *bcc* = body of the corpus callosum, *ec* = external capsule, *gcc* = genu of the corpus callosum, *ilf* = inferior longitudinal fasciculus, *ifo* = inferior fronto-occipital fasciculus, *pcr* = posterior region of the corona radiata, *plic* = posterior limb of the internal capsule, *scc* = splenium of the corpus callosum, *scr* = superior region of the corona radiata, *sfo* = superior fronto-occipital fasciculus, *slf* = superior longitudinal fasciculus, *tpt* = tapetum

bundles.

Performance of the probability metrics

In order to further compare the performance of the three probability metrics, within our segmentation framework, we have experimented with the extraction of the corpus callosum from a given DTI data-set. This important structure corresponds to the so-called callosal radiations which connect homologous areas of each hemisphere. It can be roughly divided into three main parts known as the genu (*gcc*), body (*bcc*) and splenium (*scc*). The genu radiates into the prefrontal, orbital and inferior premotor areas to form the forceps minor. The body of the corpus callosum radiates into the premotor, motor and supplementary motor cortical areas. Finally, the splenium radiates into the inferior/superior temporal, occipital and posterior parietal regions

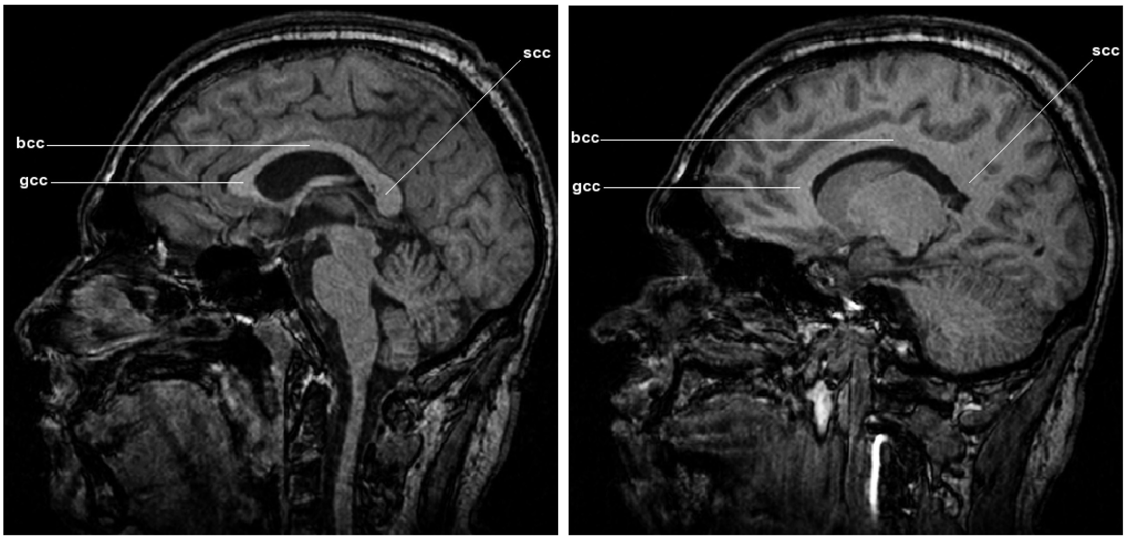


Figure 8.14: Corpus callosum on a midsagittal (left) and sagittal (right) slice from a T1 MRI

to form the forceps major.
It turns out that, near the midsagittal plane, all the fibers follow the same right-left orientation pattern making it quite easy to extract this structure from anatomical MRI (see figure 8.14 [left]). This has been used in group studies [216] to investigate architectural variability of the corpus callosum in relation with pathologies like schizophrenia. However, as we can see on figure 8.14 [right], once we get away from the midsagittal planes the callosal radiations quickly merge within the white matter and cannot be segmented anymore. We show that our Riemannian segmentation framework is able to provide more accurate segmentations of the corpus callosum.

The initialization is obtained either by a quick and approximate delineation of the genu and splenium on only 2 axial slices (figure 8.15) or by a simple sphere of radius 8 voxels centered in the middle of the body of the corpus callosum. In both cases, results are identical and presented on figure 8.17 and 8.18. It is obvious that there is a clear improvement of the segmentation quality (especially in the region of the splenium, figure 8.17) when moving from the Euclidean distance to the J-divergence and it is much better when the statistics are computed with the geodesic distance.

The splenium of the corpus callosum is almost entirely recovered by the Riemannian approach while it is barely visible with the Euclidean method and only partially extracted when using the J-divergence. We noticed moreover than the Euclidean approach has a tendency to misclassify some tensors from the ventricles. This means that the statistics are not enough discriminant and even take over the boundary term at some locations. The geodesic distance definitely yields the best

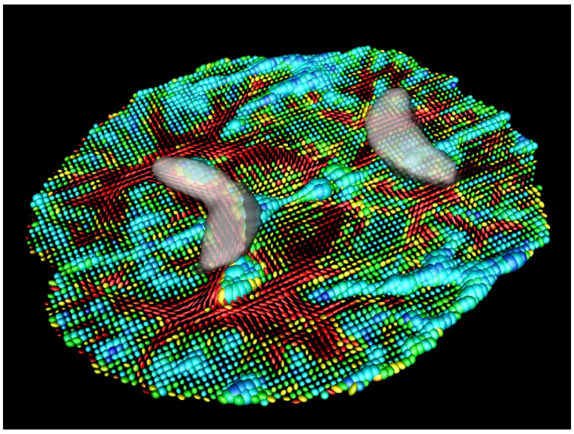


Figure 8.15: Initialization of the corpus callosum segmentation by an approximate delineation of the genu and splenium

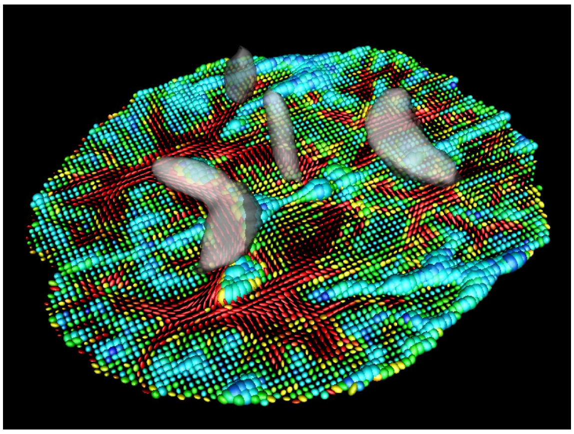


Figure 8.16: Initialization of the corpus callosum segmentation with added tensors from the superior region of the corona radiata

results.

Multiple fiber bundles segmentation

We conclude our numerical experiments on human brain DTI by trying to also recover fibers from the corona radiata, which is known to merge with the corpus callosum. The initialization is presented on figure 8.16 and is meant to include some tensors from the superior part of the corona radiata (*scr*). To that end we simply added tensors of the *scr* on 2 coronal slices. It turns out that, with the Euclidean distance and J-divergence, these new tensors quickly disappear from the segmentation and the final results are the same as those presented on figure 8.18. This is not suprising and proves that the associated statistics do not constitute accurate descriptors of the tensors distribution. On the other side, the statistics computed with the geodesic distance make it possible to perform the desired segmentation, as presented on figure 8.19. We believe that this is an interesting result since the superior part of the corona radiata is partially recovered. But more importantly, fiber tracts which are known to mingle with the callosal radiations are also segmented. It is indeed well-known that the corpus callosum merges with association and projection fibers as its gets toward the cortex. We can see on figure 8.19 that the tapetum, the posterior region of the corona radiata and a part of the superior longitudinal fasciculus are extracted since they fuse with the splenium of the corpus callosum. The posterior limb of the internal capsule (essentially the corticospinal tract) is equally segmented since it intersects with the corpus callosum and with the superior longitudinal fasciculus in the region of the centrum semiovale. All these results contribute to clearly validate our claim that the proposed Riemannian framework achieves the best segmentation results.

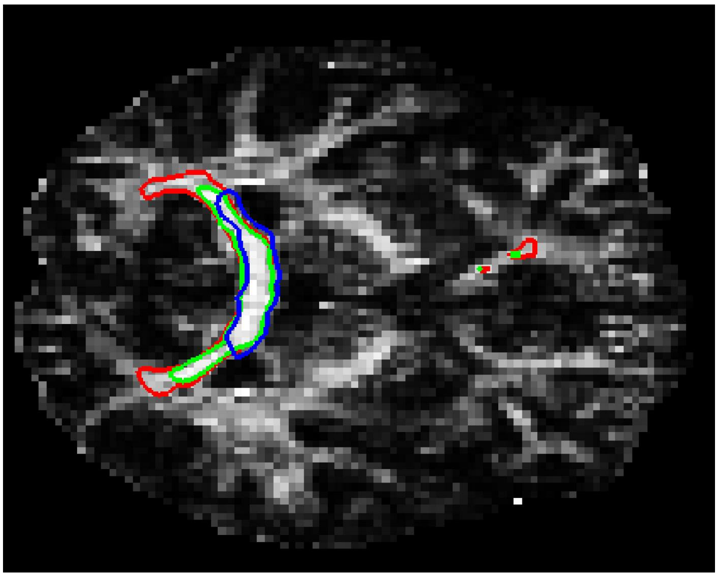


Figure 8.17: Segmentation results in the region of the splenium (blue: Euclidean distance, green: J-divergence, red: geodesic distance)

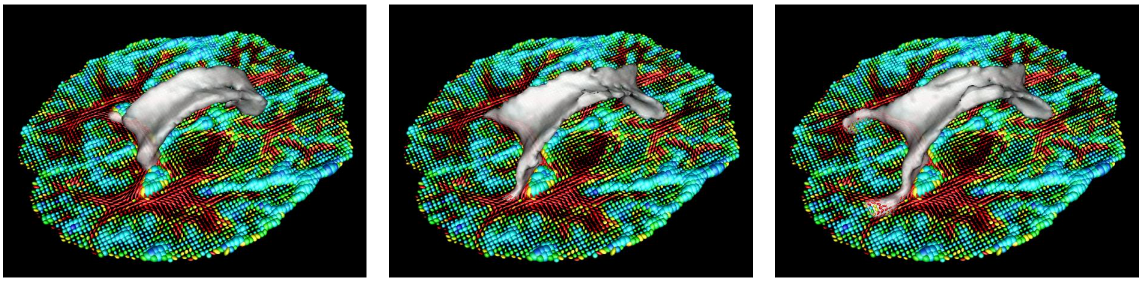


Figure 8.18: Segmentation of the corpus callosum using the Euclidean distance (left), J-divergence (center), and geodesic distance (right)

8.5 CONCLUSION

We have presented a unified statistical surface evolution framework for the segmentation of diffusion tensor images. Since a diffusion tensor can be understood as the covariance matrix of a three-dimensional normal distribution with zero mean, we have introduced various probability metrics (Euclidean distance, J-divergence and geodesic distances), i.e. dissimilarity measures between probability density functions, to derive statistics on DT images. These statistical parameters (mean and covariance matrix) allowed us to define a notion of Gaussian density for diffusion tensors, depending on the probability metric, which was used to model the distribution of a set of tensors. Finally, we have shown how to estimate the norm of the spatial gradient of a DT image, using the three different dissimilarity measures, in order to detect boundaries between structures of the white matter. By fusing these statistical and geometrical information within a variational framework, we derived a powerful level-set based DTI segmentation technique. At this point, our claim was

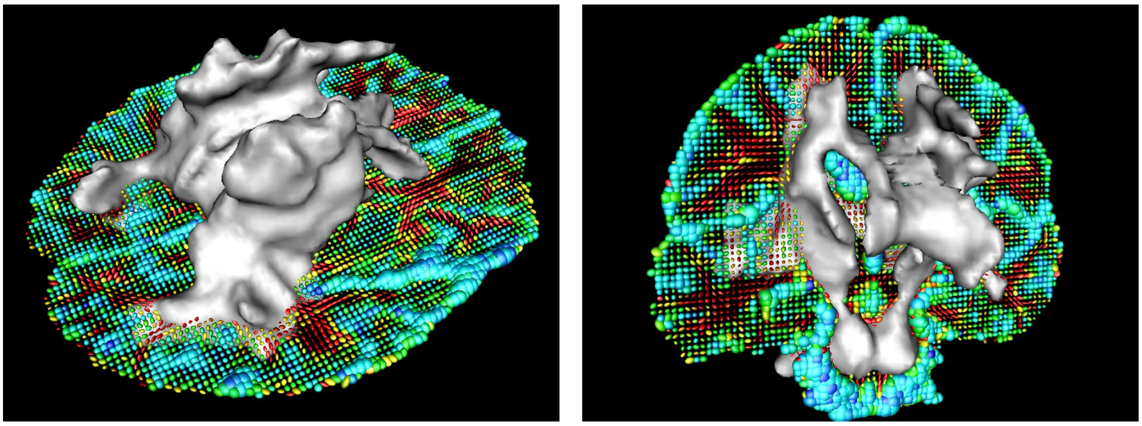


Figure 8.19: Segmentation of the corpus callosum and intermingling fiber tracts

that the special properties of the space of 3×3 diffusion tensors (symmetry and positivity) were naturally handled by working in the Riemannian framework. It must consequently yield more adequate tools to deal with tensors than the Euclidean or J-divergence approaches. The former, by seeing $S^+(3)$ as a linear space is completely blind to its curvature. The latter was shown to be equivalent to the geodesic distance only for infinitesimally close tensors. The Riemannian framework was proposed to derive the proper tools to work within the space of 3×3 diffusion tensors while taking into account its special properties.We proved that the choice of the probability metric, i.e. the dissimilarity measure, has a deep impact on the tensor statistics and, hence, on the segmentation results.

Through numerical experiments on synthetic data-sets, a biological rat spinal cord phantom, as well as on human brain DT images, we could demonstrate the superiority of the geodesic probability metric over the J-divergence which, in turn, performed better than the Euclidean distance. This order was found on synthetic data-sets with increasing complexity and for which, ground truth being known, it was very easy to undoubtedly evaluate the quality of the segmentations. The biological phantom data-set, because of its known and relatively simple geometry, allowed to test the robustness to the initialization of our algorithm and, again, to demonstrate on a single realistic data-set that the best results were obtained with the geodesic distance. Finally, on human brain DTI data-sets, the Riemannian approach was the only one capable of correctly segmenting highly variable tensor fields. It achieved better results than the other metrics (Euclidean distance and J-divergence), by comparison with neuroanatomical knowledge, for the segmentation of the corpus callosum or the corticospinal tract.

We now address one last processing technique for DTI before presenting two applications of the tools developed in this thesis: the non-rigid registration of diffusion tensor images.

CHAPTER 9

NON RIGID REGISTRATION OF DIFFUSION TENSOR IMAGES

Contents

OVERVIEW

We propose a novel variational framework for the dense non-rigid registration of diffusion tensor images. Our approach relies on the differential geometrical properties of the Riemannian manifold of multivariate normal distributions endowed with the metric derived from the Fisher information matrix, as presented in chapter 5. The availability of closed form expressions for the geodesics and the Christoffel symbols allows us to define statistical quantities and to perform the parallel transport of tangent vectors in this space. We propose a matching energy that aims to minimize the difference in the local statistical content (means and covariance matrices) of two DT images I_1 and I_2 through a gradient descent procedure. The result of the algorithm is a dense vector field h that can be used to wrap I_2 onto I_1. This chapter is essentially a mathematical study of the registration problem. Some numerical experiments are provided in section 9.6 as a proof of concept.

Keywords: registration, non rigid registration, partial differential equations, differential geometry, parallel transport, statistics, linear elasticity

9.1 INTRODUCTION

We deal with the problem of estimating the geometric deformations between two diffusion tensor images. This is reminiscent of the problem of estimating the deformation of two images where the values at each voxel are real numbers [110]. This is solved by minimizing with respect to the deformation field h an error criterion that takes into account two sources of a priori knowledge:

1. The properties of the intensities, to define the similarity of the images.
2. The constraints on the possible geometric deformations.

In our case, the "intensities" are diffusion tensors. The problem of measuring their similarity is much more complicated and the corresponding gradient descent scheme becomes significantly more involved.

Previous works on the subject was initiated by Alexander *et al.* [4] by extending multiresolution registration techniques to DTI after having introduced various possible dissimilarity measures for such images [5]. In [263] and [262] the authors proposed to register three-dimensional scalar, vector and tensor data by matching areas with a high degree of structure and then interpolating the sparse estimated displacement field to the complete dataset. Other approaches like [155], [130], [229] and [255] rely on one or several transformation invariant tensor characteristics like the eigenvalues, the anisotropy measures, the apparent diffusion coefficient or even the tensor components to perform the registration. When several characteristics are used, which is often the case, multiple input channel registration methods like the demons algorithm [131] are used. In [321, 319] and then [320], the authors proposed a piecewise affine registration technique based on the L^2 inner product of diffusion profiles. They also investigate the tensors reorientation issue raised by Alexander *et al.* in [6]. Recently, Cao *et al.* [52] proposed to apply the framework of the Large Deformations Diffeomorphic Metric Mapping to DTI. Finally, Leemans [177] introduced an affine multi-channel registration technique based on the mutual information as well as an original feature based registration method based on the curvature and torsion of fibers pathways. Although different from the DTI registration problem, we also want to point out a few recent works [11, 236] that have used the Riemannian or Log-Euclidean metrics to characterize the properties of deformation fields, obtained through scalar images registration algorithms.

Contributions of this chapter:

In this chapter, we extend the approach presented in [110] to matrix-valued images $I : \Omega \rightarrow S^+(3)$. To our knowledge, this is the very first work to make use of the Riemannian structure of $S^+(3)$, proved to be relevant for DTI processing for instance in [238, 116, 14, 186], in a non-rigid DTI registration algorithm. The numerical implementation of the method is very tedious. We will illustrate the feasibility of the approach on two-dimensional synthetic datasets.

Organization of this chapter:

We first set up the registration problem in section 9.2 and recall some important notions on $S^+(3)$. We then detail the regularization (section 9.3) and the data (section 9.4) terms of the initial value problem 9.3. We detail the computation of the gradient of the data term in section 9.5. Finally, we present numerical experiments in section 9.6.

9.2 THE REGISTRATION PROBLEM

We consider the problem of estimating the geometric deformations between two Diffusion Tensor Images (DTI). At a conceptual level, DT images are integrable bounded functions defined in $\mathbb{R}^n$, $n = 2, 3$ with values in $S^+(3)$ (noted S^+ in the sequel). As briefly recalled below, this space has a natural Riemannian structure. Bounded means that all observed diffusion tensors are within a ball of center $\mathbb{I}$, the 3×3 identity matrix, for the distance defined by equation 9.4. The same equation shows that the eigenvalues must lie between two strictly positive constants and therefore that the set of bounded diffusion tensors (for the Riemannian metric) is also bounded for the 2-norm and therefore for all the usual p-norms and the Frobenius norm.

These abstract images are not directly observable because of the physics of acquisition. What we call an image is an element of $C^\infty(\mathbb{R}^n, S^+)$, the space of infinitely differentiable functions. They are bounded and Lipschitz continuous as well as all their derivatives.

9.2.1 Statement of the problem

Let I_1 and I_2 be two images and $h : \Omega \to \mathbb{R}^n$ a vector field defined on a bounded and regular region of interest $\Omega \subset \mathbb{R}^n$. The registration or matching problem may be defined as that of finding a vector field h^* minimizing an error criterion between I_1 and $I_2 \circ h$. The search for this function is done within a set $\mathcal{F}$ of admissible functions so as to minimize an energy functional $\mathcal{I} : \mathcal{F} \to \mathbb{R}^+$ of the form

$$\mathcal{I}(h) = \mathcal{J}(h) + \mathcal{R}(h).$$

The term $\mathcal{J}$ is designed to measure the "dissimilarity" between the reference image I_1 and the h-warped image, noted $\mathcal{T}_h(I_2)$. We have the following proposition

Proposition 9.2.1.1. *If the relation between the two images I_1 and I_2 is a change of coordinates $x' = h(x)$ then the value $I_1(x)$ should be equal to the value $\mathcal{T}_h(I_2)(x)$, where*

$$\mathcal{T}_h(I_2) = Dh^{-1} I_2(h) Dh^{-T}. \tag{9.1}$$

Proof. $I_1(x)$ is a twice contravariant tensor. In the new coordinate system defined by $x' = h(x)$ it is equal to

$$I_1'(x') = Dh(x) I_1(x) Dh^T(x),$$

because of the way tensor components change with changes of coordinates. This new tensor should be equal to $I_2(x')$ and this yields the expression for $\mathcal{T}_h(I_2)(x)$. $\square$

Note that other possibilities for $\mathcal{T}_h(I_2)$ have been considered in the literature (see [6] for instance). R. Sierra [270], has considered the case where one wants to preserve the determinant of I_2; this leads to

$$\mathcal{T}_h(I_2) = (\det(Dh))^{2/3} Dh^{-1} I_2(h) Dh^{-T}$$

In the following we consider that $\mathcal{T}_h(I_2)$ is defined by equation 9.1. The term $\mathcal{R}(h)$ is designed to penalize fast variations of the function h. It is a regularization term introducing an a priori preference for smoothly varying functions. Our error criterion is classically the sum of a data term $\mathcal{J}$ and a regularization term $\mathcal{R}$.

The set $\mathcal{F}$ is a dense linear subspace of a Hilbert space H whose scalar product is denoted by $(\cdot, \cdot)_H$. If $\mathcal{I}$ is sufficiently regular, its first variation (also called the Gâteaux derivative) at $h \in \mathcal{F}$ is defined (see, e.g., [81]) as

$$\delta_k \mathcal{I}(h) = \lim_{\varepsilon \to 0} \frac{\mathcal{I}(h + \varepsilon k) - \mathcal{I}(h)}{\varepsilon} \tag{9.2}$$

If the mapping $k \to \delta_k \mathcal{I}(h)$ is linear and continuous, the Riesz representation theorem [108] guarantees the existence of a unique vector, denoted by $\nabla_H \mathcal{I}(h)$, and called the gradient of $\mathcal{I}$, which satisfies the equality

$$\delta_k \mathcal{I}(h) = (\nabla_H \mathcal{I}(h), k)_H,$$

for every $k \in H$. The gradient depends on the choice of the scalar product $(\cdot,\cdot)_H$ though, a fact which explains our notation. If a minimizer h^* of $\mathcal{I}$ exists, then the set of equations $\delta_k \mathcal{I}(h^*) = 0$ must hold for every $k \in H$, which is equivalent to $\nabla_H \mathcal{I}(h^*) = 0$.

These equations are called the Euler-Lagrange equations associated with the energy functional $\mathcal{I}$. They give necessary conditions for the existence of a minimizer but they are not sufficient since they only guarantee the existence of a critical point of the functional $\mathcal{I}$. These critical points can be found in many ways, including methods for nonlinear equations. Rather than solving them directly the search for a minimizer of $\mathcal{I}$ is done using a gradient descent strategy. Given an initial estimate $h_0 \in \mathcal{F}$, a time-dependent differentiable function (also denoted by h) from the interval $[0, +\infty[$ into H is computed as the solution of the following initial value problem:

$$\begin{cases} \dfrac{dh}{dt} = -\Big(\nabla_H \mathcal{J}(h) + \nabla_H \mathcal{R}(h)\Big), \\ h(0)(\cdot) = h_0(\cdot). \end{cases} \tag{9.3}$$

The asymptotic state (i.e. when $t \to \infty$) of $h(t)$ is then chosen as the solution of the matching problem, provided that $h(t) \in \mathcal{F} \quad \forall t > 0$.

9.2.2 Precisions on the Riemannian structure of $S^+(n)$

In this section, we remind some basic concepts that will be useful for the following. We recall that $S^+(n)$ denotes the set of $n \times n$ real symmetric positive definite matrices, Σ. It is a subset of $M_n(\mathbb{R})$, the set of $n \times n$ real matrices. It is also a m_n-dimensional C^∞ submanifold of $\mathbb{R}^{m_n}$ ($m_n = n(n+1)/2$) whose local coordinates can be chosen as the m_n algebraically independent components of the elements of Σ. We note $\varphi_n : S^+(n) \to \mathbb{R}^{m_n}$ the natural coordinates mapping of this manifold. We recall that $T_\Sigma S^+(n)$, the tangent space at Σ of $S^+(n)$, coincides with the set $S(n)$ of $n \times n$ real symmetric matrices. This is a vector space which can be identified with $\mathbb{R}^{m_n}$ through the mapping φ_n. Elements in that space are contravariant vectors. We finally denote by $T^*_\Sigma S^+(n)$ the cotangent space at Σ of $S^+(n)$, the dual space of $T_\Sigma S^+(n)$. Elements in that space are covariant vectors. The basis of $T_\Sigma S^+(n)$ and $T^*_\Sigma S^+(n)$ were given in section 5.2 of chapter 5.

We recall the following theorem, see, e.g. [209]:

Theorem 9.2.2.1. *Let $\mathcal{E}$ be the set of real $n \times n$ matrices such that all the eigenvalues λ_i are such that $|Im(\lambda_i)| < \pi$. The restriction to $\mathcal{E}$ of the exponential is a diffeomorphism between $\mathcal{E}$ and $\exp \mathcal{E}$.*

There are two consequences of this theorem that are used in the sequel. The first one is the

Corollary 9.2.2.1. *The exponential is a diffeomorphism between $S(n)$ and $S^+(n)$.*

In other words, the exponential of any symmetric matrix is a positive definite symmetric matrix and the inverse of the exponential (i.e. the principal logarithm) of any positive definite symmetric matrix is a symmetric matrix. Moreover, both the exponential and the logarithm are continuously differentiable in $S(n)$ and $S^+(n)$, respectively.
The second one is the

Corollary 9.2.2.2. *The logarithm of a matrix with positive eigenvalues exists, and is unique and differentiable.*

Proof. Any such matrix belongs to $\exp \mathcal{E}$ defined in theorem 9.2.2.1. Therefore its logarithm exists, is unique and differentiable. $\square$

We introduce two notations

Definition 9.2.2.1. *We note* $\exp$ *and* $\log$ *the exponential and its inverse, the logarithm. Given $M \in M_n(\mathbb{R})$, we note* $\operatorname{dexp}(M, X)$ *the derivative of* $\exp$ *at M, applied to the element $X \in M_n(\mathbb{R})$. This is also sometimes called the derivative of the function* $\exp$ *at M in the direction X. In a similar manner, given $M \in \exp \mathcal{E}$ we note* $\operatorname{dlog}(M, X)$ *the derivative of the function* $\log$ *at M in the direction X.*

Details on the directional derivative of the matrix exponential and its computation can be found in [215]. However, to our knowledge, there is no previous work on the computation of the directional derivative of the matrix logarithm. As we will see in section 9.5, this will be a key component of our method. In appendix A, we propose a novel formulation for the directional derivative of the matrix logarithm $\operatorname{dlog}(M, X)$ based on the spectral decomposition of M. We will show that it is in fact a linear function of its second argument X.

The geodesic distance between two elements Σ_1 and Σ_2 of $S^+(n)$ was introduced in chapter 5 and is defined by

$$\mathcal{D}(\Sigma_1, \Sigma_2) = \sqrt{\frac{1}{2}\mathrm{tr}\left(\log^2\left(\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2}\right)\right)}, \tag{9.4}$$

It is justified by corollary 9.2.2.1. At each point Σ of S^+, the metric tensor G acts on pairs of tangent vectors of $T_\Sigma S^+$ and defines an inner product. Its inverse G^{-1} is twice contravariant. For any real differentiable function f defined on S^+, one defines its differential, noted $Df = [\frac{\partial f}{\partial \mathbf{x}^1}, \cdots, \frac{\partial f}{\partial \mathbf{x}^{m_n}}]$, with respect to the coordinates defined by the chart φ_n, a covariant vector, and its gradient, noted ∇f, which is a vector of $T_\Sigma S^+$. The relation between Df and ∇f is through the metric tensor:

$$\nabla f = G^{-1} Df \tag{9.5}$$

Equation 9.4 defines a real function on $S^+(n) \times S^+(n)$ which is differentiable. The gradient of $\mathcal{D}^2$ with respect to Σ_1, noted ∇_{Σ_1}, at Σ_1 and for some fixed Σ_2, is equal to [211]:

$$\nabla_{\Sigma_1} \mathcal{D}^2(\Sigma_1, \Sigma_2) = \Sigma_1 \log\left(\Sigma_2^{-1}\Sigma_1\right). \tag{9.6}$$

It is a vector of $T_{\Sigma_1} S^+(n)$, hence a symmetric matrix. This can also be seen from the general relation

$$\log(A^{-1}BA) = A^{-1}(\log B)A, \tag{9.7}$$

by writing

$$\Sigma_1 \log\left(\Sigma_2^{-1}\Sigma_1\right) = \Sigma_1 \log\left(\Sigma_1^{-1}\Sigma_1\Sigma_2^{-1}\Sigma_1\right) = \log\left(\Sigma_1\Sigma_2^{-1}\right)\Sigma_1 = \left(\Sigma_1 \log\left(\Sigma_2^{-1}\Sigma_1\right)\right)^T$$

It is tangent at Σ_1 to the (unique) geodesic between Σ_1 and Σ_2. In the following, we use the cases $n = 3$ and $n = 6$. To facilitate the reading of the formulas, indexes running from 1 to 3 are lower case Latin characters, e.g., $i = 1, 2, 3$, indexes running from 1 to 6 are upper case Latin characters, e.g., $I = 1, \ldots, 6$, and indexes running from 1 to 9 are lower case Greek characters, e.g., $\kappa = 1, \ldots, 9$.

9.3 REGULARIZATION TERM

This section studies the regularization part of the initial value problem 9.3, i.e. the term $\nabla_H \mathcal{R}(h)$. We choose concrete functional spaces $\mathcal{F}$ and H and specify the domain of the regularization operators.

9.3.1 Function spaces and boundary conditions

We begin by a brief description of the function spaces that will be appropriate for our purposes. In doing this, we will make reference to Sobolev spaces, denoted by $W^{k,p}(\Omega)$. We refer to the books of Evans [108] and Brezis [40] for formal definitions and in-depth studies of the properties of these functional spaces.

For the definition of $\nabla_H \mathcal{I}$, we use the Hilbert space

$$H = \mathbf{L}^2(\Omega) = \underbrace{L^2(\Omega) \times \cdots \times L^2(\Omega)}_{n \text{ terms}} = \left(W^{0,2}(\Omega)\right)^n.$$

The regularization functionals that we consider are of the form

$$\mathcal{R}(h) = \kappa \int_\Omega \varphi(Dh(x))\, dx, \tag{9.8}$$

where $Dh(x)$ is the Jacobian of h at x, φ is a quadratic form of the elements of the matrix $Dh(x)$ and $\kappa > 0$. Therefore the set of admissible functions $\mathcal{F}$ will be contained in the space

$$\mathbf{H}^1(\Omega) = \left(W^{1,2}(\Omega)\right)^n.$$

Additionally, the boundary conditions for h will be specified in $\mathcal{F}$. We consider Dirichlet conditions of the form $h = 0$ almost everywhere on $\partial\Omega$ (in fact, because of the regularity of h, this condition holds everywhere on $\partial\Omega$), and set

$$\mathcal{F} = \mathbf{H}^1_0(\Omega) = \left(W^{1,2}_0(\Omega)\right)^n.$$

Because of the special form of $\mathcal{R}(h)$, the corresponding regularization operator is a second order differential one, and we therefore will need the space

$$\mathbf{H}^2(\Omega) = \left(W^{2,2}(\Omega)\right)^n$$

for the definition of its domain.

9.3.2 Linearized elasticity

The family that we consider is inspired from the equilibrium equations of linearized elasticity (we refer to [69] for a formal study of three-dimensional elasticity) obtained by defining φ in equation 9.8 by

$$\varphi(Dh) = \frac{1}{2}\left(\xi \mathrm{Tr}(Dh^T Dh) + (1-\xi)\mathrm{Tr}(Dh)^2\right), \tag{9.9}$$

where $1/2 < \xi \leq 1$. It is straightforward to verify that the Euler-Lagrange equation corresponding to equation 9.8 in this case is:

$$\nabla_H \mathcal{R}(h) = \mathrm{div}\left(D\varphi(Dh)\right) = \xi\Delta h + (1-\xi)\nabla(\nabla \cdot h)$$

We thus define the corresponding regularization operator as follows.

Definition 9.3.2.1. *The linear operator $A : \mathcal{D}(A) \to H$ is defined as*

$$\begin{cases} \mathcal{D}(A) = \mathbf{H}^1_0(\Omega) \cap \mathbf{H}^2(\Omega), \\ \\ Ah = \nabla_H \mathcal{R}(h) = \xi\Delta h + (1-\xi)\nabla(\nabla \cdot h) \end{cases}$$

for $1/2 < \xi \leq 1$

9.4 DEFINITION OF THE DATA TERM $\mathcal{J}$

We want to compare the values of the image I_1 in a neighborhood of a voxel x to those of I_2 in the corresponding neighborhood transformed by h. We propose a statistical framework for doing so, in the spirit of block-matching techniques. Local statistics (mean, covariance matrix) has been found to be very useful for warping scalar images, i.e. real-valued images. This idea can be generalized to tensor-valued images as follows.

9.4.1 Local mean and covariance matrix

Given a voxel x in the volume Ω, the local mean $\hat{\mu}_1(x)$ is defined as one of the minima with respect to its first argument of the following function defined on $S^+ \times \Omega$

$$\mathcal{C}_1(\mu_1, x) = \frac{1}{\mid \Omega \mid} \int_\Omega \mathcal{D}^2(\mu_1, I_1(y)) G_\gamma(x-y)\, dy,$$

where G_γ is a three-dimensional Gaussian with 0-mean and variance γ^2

$$G_\gamma(x) = \frac{1}{(2\pi\gamma^2)^{3/2}} \exp\left(-\frac{|x|^2}{2\gamma^2}\right).$$

$|x|$ is the Euclidean norm of the vector x and $\mathcal{D}$ is the geodesic distance defined in equation 9.4 between the two elements μ_1 and $I_1(y)$ of S^+ [186]. $\mathcal{C}$ is a weighted average of the squared geodesic distances between the element μ_1 of S^+ and the elements of the image I_1. The amount of locality is controlled by the parameter γ^2, the variance of the Gaussian. We call $\hat{\mu}_1(x)$ the element of S^+ that minimizes $\mathcal{C}_1$. It is the weighted Riemannian mean of the family $I_1(y)$ of elements of S^+, where y varies in Ω, a notion introduced by Grove, Karcher and Ruh [129]. We have already explained in section 5 why this mean is unique.

Hence we have

$$\hat{\mu}_1(x) = \operatorname*{argmin}_{\mu_1} \frac{1}{\mid \Omega \mid} \int_\Omega \mathcal{D}^2(\mu_1, I_1(y)) G_\gamma(x-y)\, dy$$

Because of equation 9.6 we can write an expression for the gradient of the function $\mathcal{C}$ with respect to μ_1, at μ_1:

$$\nabla_{\mu_1} \mathcal{C}_1(\mu_1, x) = \frac{\mu_1}{\mid \Omega \mid} \int_\Omega \log\left(I_1(y)^{-1}\mu_1\right)\, G_\gamma(x-y)\, dy$$

At the minimum $\hat{\mu}_1(x)$, this gradient is equal to 0:

$$\frac{\hat{\mu}_1(x)}{\mid \Omega \mid} \int_\Omega \log\left(I_1(y)^{-1}\hat{\mu}_1(x)\right)\, G_\gamma(x-y)\, dy = 0 \tag{9.10}$$

An interpretation of this relation is the following. Each of the matrices

$$\beta_1(x,y) \overset{def}{=} -\hat{\mu}_1(x) \log\left(I_1(y)^{-1}\hat{\mu}_1(x)\right)\, G_\gamma(x-y)$$

belongs to the tangent space $T_{\hat{\mu}_1(x)}S^+$, a copy of S, the space of symmetric matrices. Since $T_{\hat{\mu}_1(x)}S^+$ is identified to $\mathbb{R}^6$ through the chart φ_3, one can define the covariance matrix of the vectors $\varphi_3(\beta_1(x,y))$, noted $\beta_1(x,y)$ for simplicity, which have zero-mean according to equation 9.10:

$$\Lambda_1(x) = \frac{1}{\mid \Omega \mid} \int_\Omega \beta_1(x,y) \beta_1^T(x,y)\, dy.$$

This is a twice contravariant tensor defined on $T_{\hat{\mu}_1(x)}S^+$.

Applying the transformation h to the second image I_2, we can define the corresponding quantities. The local mean at the voxel $h(x)$ becomes:

$$\hat{\mu}_2(x,h) = \underset{\mu_2}{\operatorname{argmin}}\, \mathcal{C}_2(\mu_2, x) = \underset{\mu_2}{\operatorname{argmin}} \frac{1}{|\,\Omega\,|} \int_\Omega \mathcal{D}^2(\mu_2, \mathcal{T}_h(I_2)(y)) G_\gamma(x-y)\, dy,$$

and satisfies

$$\nabla_{\mu_2} \mathcal{C}_2(\hat{\mu}_2(x,h), x) = \frac{\hat{\mu}_2(x,h)}{|\,\Omega\,|} \int_\Omega \log\left(\mathcal{T}_h(I_2)^{-1}(y)\hat{\mu}_2(x,h)\right)\, G_\gamma(x-y)\, dy = 0. \tag{9.11}$$

The tangent vectors to S^+ at $\hat{\mu}_2(x,h)$ are

$$\beta_2(x,y,h) \overset{def}{=} -\hat{\mu}_2(x,h) \log\left(\mathcal{T}_h(I_2)^{-1}(y)\hat{\mu}_2(x,h)\right)\, G_\gamma(x-y) \tag{9.12}$$

and their covariance matrix is

$$\Lambda_2(x,h) = \frac{1}{|\,\Omega\,|} \int_\Omega \beta_2(x,y,h)\beta_2^T(x,y,h)\, dy. \tag{9.13}$$

We now face a difficulty. We would like to compare the tangent vectors $\beta_1(x,y)$ and $\beta_2(x,y,h)$ but this cannot be done in a straightforward manner since they live in two different vector spaces, $T_{\hat{\mu}_1(x)}S^+$ and $T_{\hat{\mu}_2(x,h)}S^+$. In order to compare them, we must either parallel transport the vectors $\beta_1(x,y)$ to $T_{\hat{\mu}_2(x,h)}S^+$ (obtaining the vectors $\tilde{\beta}_1(x,y,h)$) or the vectors $\beta_2(x,y,h)$ to $T_{\hat{\mu}_1(x)}S^+$ (obtaining the vectors $\tilde{\beta}_2(x,y,h)$).

We can then define the covariance matrices $\tilde{\Lambda}_{12}(x,h)$ and $\tilde{\Lambda}_{21}(x,h)$ of the parallel transported vectors $\tilde{\beta}_1(x,y,h)$ and $\tilde{\beta}_2(x,y,h)$, respectively:

$$\tilde{\Lambda}_{12}(x,h) = \frac{1}{|\,\Omega\,|} \int_\Omega \tilde{\beta}_1(x,y,h)\tilde{\beta}_1^T(x,y,h)\, dy \tag{9.14}$$

and

$$\tilde{\Lambda}_{21}(x,h) = \frac{1}{|\,\Omega\,|} \int_\Omega \tilde{\beta}_2(x,y,h)\tilde{\beta}_2^T(x,y,h)\, dy \tag{9.15}$$

9.4.2 Parallel transport

We now detail the process of parallel transport as illustrated on figure 9.1.

The equations

To parallel transport a vector $\beta_1(x,y)$ from $T_{\hat{\mu}_1(x)}S^+$ along the curve $\mathcal{G}(t)$ such that $\mathcal{G}(0) = \hat{\mu}_1(x)$ and $\mathcal{G}(1) = \hat{\mu}_2(x,h)$ one has to solve the first-order linear differential equation

$$\nabla_{\dot{\mathcal{G}}(t)}\beta(t) = 0 \tag{9.16}$$

with initial condition $\beta(0) = \beta_1(x,y)$. $\nabla_{\dot{\mathcal{G}}(t)}$ stands for the covariant derivative in the direction $\dot{\mathcal{G}}(t)$ of $T_{\mathcal{G}(t)}S^+$ and equation 9.16 can be rewritten in local coordinates as:

$$\left(\nabla_{\dot{\mathcal{G}}(t)}\beta(t)\right)^I = \frac{d\beta^I}{dt} + \Gamma^I_{JK}(\mathcal{G}(t))\dot{\mathcal{G}}(t)^J\beta^K(t) = 0 \tag{9.17}$$

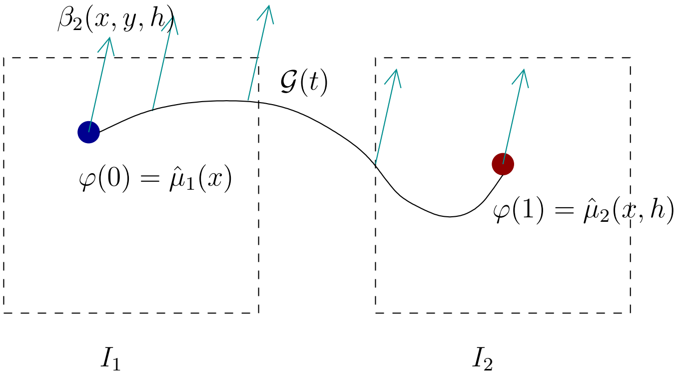


Figure 9.1: Parallel transport of vector $\beta_2(x,y,h)$ along $\mathcal{G}(t)$

where the Γ^I_{JK}'s are the Christoffel symbols of the second kind associated to the metric of S^+. This linear differential equation can be written in the form

$$\frac{d\,\beta(t)}{dt} = -\mathcal{A}(t)\beta(t) \tag{9.18}$$

with the same initial condition $\beta(0) = \beta_1(x,y)$. The 6×6 matrix $\mathcal{A}(t)$ is equal to

$$(\mathcal{A})^I_K\,(t) = \Gamma^I_{JK}(\mathcal{G}(t))\dot{\mathcal{G}}(t)^J. \tag{9.19}$$

We recall that $\mathcal{G}(t)$ is the geodesic between $\hat{\mu}_1(x)$ and $\hat{\mu}_2(x,h)$ whose equation is [186]

$$\mathcal{G}(t) = \hat{\mu}_1(x)^{1/2}\exp\left(tX\right)\hat{\mu}_1(x)^{1/2}, \tag{9.20}$$

where

$$X = \log\left(\hat{\mu}_1(x)^{-1/2}\hat{\mu}_2(x,h)\hat{\mu}_1(x)^{-1/2}\right) \tag{9.21}$$

Similar considerations apply to the problem of the parallel transport of the vector $\beta_2(x,y,h)$ along the geodesic $\mathcal{G}(t)$ from $T_{\hat{\mu}_2(x,h)}S^+$to $T_{\hat{\mu}_1(x)}S^+$ by introducing the matrix $\mathcal{B}(t)$.

The matrices $\mathcal{A}$ and $\mathcal{B}$

In the following, we prove that the matrices $\mathcal{A}(t)$ and $\mathcal{B}(t)$ do not depend on the curve parameter t. The solution of equation 9.18 is therefore

$$\tilde{\beta}_1(x,y,h) = \beta(1) = \exp\left(-\mathcal{A}\right)\beta(0) = \exp\left(-\mathcal{A}\right)\beta_1(x,y) \tag{9.22}$$

Similarly

$$\tilde{\beta}_2(x,y,h) = \exp\left(-\mathcal{B}\right)\beta_2(x,y,h) \tag{9.23}$$

We have the following

Proposition 9.4.2.1. *The matrix* $\mathcal{A}(t)$ *is independent of* t. *It is given by the following expression*

$$\mathcal{A}(x,h) = \tag{9.24}$$

$$-\begin{bmatrix} \psi_1^1 & \psi_1^2/2 & 0 & \psi_1^3/2 & 0 & 0 \\ \psi_2^1 & (\psi_1^1+\psi_2^2)/2 & \psi_1^2 & \psi_2^3/2 & \psi_1^3/2 & 0 \\ 0 & \psi_2^1/2 & \psi_2^2 & 0 & \psi_2^3/2 & 0 \\ \psi_3^1 & \psi_3^2/2 & 0 & (\psi_1^1+\psi_3^3)/2 & \psi_1^2/2 & \psi_1^3 \\ 0 & \psi_3^1/2 & \psi_3^2 & \psi_2^1/2 & (\psi_2^2+\psi_3^3)/2 & \psi_2^3 \\ 0 & 0 & 0 & \psi_3^1/2 & \psi_3^2/2 & \psi_3^3 \end{bmatrix} \overset{Def}{=} \mathcal{M}(\psi),$$

where the matrix ψ *is equal to* $\log\left(\hat{\mu}_2(x,h)\hat{\mu}_1(x)^{-1}\right)$.

The matrix $\mathcal{B}(t)$ *is also independent of* t *and its expression is similar to that of* $\mathcal{A}$ *by replacing the matrix* ψ *with the matrix* $\theta = \log\left(\hat{\mu}_1(x)^{-1}\hat{\mu}_2(x,h)\right)$.

Proof. It can be shown [271] that the Christoffel symbols, at $\mathcal{G}(t) \in S^+$, are given by:

$$\Gamma\left(E_{pq}, E_{rs}; E_{uv}^*\right) = -\frac{1}{2}\mathbf{tr}\left(E_{pq}\mathcal{G}(t)^{-1}E_{rs}E_{uv}^*\right) - \frac{1}{2}\mathbf{tr}\left(E_{rs}\mathcal{G}(t)^{-1}E_{pq}E_{uv}^*\right)$$
$$\forall p,q,r,s,u,v = 1,2,3$$

The indices p,q,r,s,u,v are used to access the components of the basis elements in matrix form and therefore run from 1 to 3. We introduce the indices I,J,K which run from 1 to 6 since they correspond to the coordinates of a given matrix expressed in the associated local coordinate system (see, for example, equations 9.17 or 9.19). Hence we use the following convention:

$$\begin{array}{lll} K = 3(r-1)+s & \text{if} & r \le s \\ J = 3(p-1)+q & \text{if} & p \le q \\ I = 3(u-1)+v & \text{if} & u \le v \end{array} \tag{9.25}$$

We can now express the quantity $\Gamma_{JK}^I\left(\mathcal{G}(t)\right)\dot{\mathcal{G}}^J(t)$ as:

$$\Gamma\left(E_J, E_K; E_I^*\right)\dot{\mathcal{G}}(t)^J = -\frac{1}{2}\mathbf{tr}\left(\dot{\mathcal{G}}(t)^J E_J\mathcal{G}(t)^{-1}E_K E_I^*\right) - \frac{1}{2}\mathbf{tr}\left(E_K\mathcal{G}(t)^{-1}\dot{\mathcal{G}}(t)^J E_J E_I^*\right)$$

Noting that $\dot{\mathcal{G}}(t)^J E_J = \dot{\mathcal{G}}(t)$, this reduces to:

$$\Gamma\left(E_J, E_K; E_I^*\right)\dot{\mathcal{G}}(t)^J = -\frac{1}{2}\mathbf{tr}\left(\dot{\mathcal{G}}(t)\mathcal{G}(t)^{-1}E_K E_I^*\right) - \frac{1}{2}\mathbf{tr}\left(E_K\mathcal{G}(t)^{-1}\dot{\mathcal{G}}(t)E_I^*\right)$$

In our case $\mathcal{G}(t) \in S^+$, and since

$$\dot{\mathcal{G}}(t) = \hat{\mu}_1(x)^{1/2}X\exp\left(tX\right)\hat{\mu}_1(x)^{1/2} = \hat{\mu}_1(x)^{1/2}\exp\left(tX\right)X\hat{\mu}_1(x)^{1/2}$$

and

$$\mathcal{G}^{-1}(t) = \hat{\mu}_1(x)^{-1/2} \exp(-tX) \hat{\mu}_1(x)^{-1/2}$$

we have

$$\dot{\mathcal{G}}(t)\mathcal{G}^{-1}(t) \overset{def}{=} \psi(x,h) = \hat{\mu}_1(x)^{1/2} X \hat{\mu}_1(x)^{-1/2} = \log\left(\hat{\mu}_2(x,h)\hat{\mu}_1(x)^{-1}\right)$$

$$\mathcal{G}^{-1}(t)\dot{\mathcal{G}}(t) \overset{def}{=} \theta(x,h)^T = \hat{\mu}_1(x)^{-1/2} X \hat{\mu}_1(x)^{1/2} = \log\left(\hat{\mu}_1(x)^{-1}\hat{\mu}_2(x,h)\right)$$

which do not depend on t. Note that ψ and θ are once contravariant and once covariant tensors.

The last equality in the previous two equations holds because of equation 9.7.
The matrix $\mathcal{A}(t)$ is therefore independent of t but depends on x and h, thus we note $\mathcal{A}(x,h)$. The contraction of the Christoffel symbols then yields the once covariant, once contravariant tensor:

$$\begin{aligned} \Gamma\left(E_J, E_K; E_I^*\right) \dot{\mathcal{G}}(t)^J &= \mathcal{A}_K^I(x,h) \\ &= -\frac{1}{2}\mathbf{tr}\left(\psi(x,h)E_K E_I^*\right) - \frac{1}{2}\mathbf{tr}\left(E_K \psi(x,h)^T E_I^*\right) \\ &= -\frac{1}{2}\mathbf{tr}\left(\left(\psi(x,h)E_K + \left(\psi(x,h)E_K\right)^T\right) E_I^*\right) \end{aligned}$$

This provides an expression for $\mathcal{A}(x,h)$, as a function of $\psi(x,h)$, in terms of $\hat{\mu}_1(x)$ and $\hat{\mu}_2(x,h)$ (We use the notation ψ_q^p to denote the pq^{th} element of $\psi(x,h)$). ☐

9.4.3 The data term $\mathcal{J}$

We are now ready to define the data term $\mathcal{J}$ in the error criterion for the registration of two DT images I_1 and I_2. This data term is the combination of two terms. The first enforces that the means $\hat{\mu}_1(x)$ and $\hat{\mu}_2(x,h)$ at two corresponding voxels x and $h(x)$ are sufficiently similar. We define the *local* energy:

$$\mathcal{J}_{\text{Mean}}(x,h) = \mathcal{D}^2(\hat{\mu}_2(x,h), \hat{\mu}_1(x)), \tag{9.26}$$

where $\mathcal{D}^2$ is the geodesic distance 9.4 in S^+.

We also want the covariance matrices $\Lambda_1(x)$ and $\tilde{\Lambda}_{21}(x,h)$ (respectively $\Lambda_2(x,h)$ and $\tilde{\Lambda}_{12}(x,h)$) to be as close as possible. We therefore define the *local* energy

$$\mathcal{J}_{\text{AC}}(x,h) = \frac{1}{2}\left(|\Lambda_1(x) - \tilde{\Lambda}_{21}(x,h)|_F^2 + |\Lambda_2(x,h) - \tilde{\Lambda}_{12}(x,h)|_F^2\right) \tag{9.27}$$

where $|\cdot|_F$ denotes the Frobenius norm. A more consistent definition of $\mathcal{J}_{\text{AC}}(x,h)$ can be obtained by using the geodesic distance instead of the Frobenius norm

$$\mathcal{J}_{\text{AC}}(x,h) = \frac{1}{2}\left(\mathcal{D}^2\left(\Lambda_1(x), \tilde{\Lambda}_{21}(x,h)\right) + \mathcal{D}^2\left(\Lambda_2(x,h), \tilde{\Lambda}_{12}(x,h)\right)\right) \tag{9.28}$$

This has very little incidence on the final form of the gradient of $\mathcal{J}_{\mathrm{AC}}$ but may be the source of numerical problems. Indeed, if the region of interest is homogeneous, the covariance matrices tend to be degenerate and the geodesic distance is not defined anymore. In practice, we check if that case happens and only use the Euclidean distance if this is the case. Otherwise, we use the geodesic distance.

We combine these two local criteria into a local data term

$$\mathcal{J}(x,h) = \mathcal{J}_{\mathrm{Mean}}(x,h) + \alpha_1 \mathcal{J}_{\mathrm{AC}}(x,h), \tag{9.29}$$

where α_1 is a positive weight. The *global* criterion is obtained by integrating the local one over Ω:

$$\mathcal{J}(h,Dh) = \int_\Omega \mathcal{J}(x,h)\,dx = \mathcal{J}_{\mathrm{Mean}}(h,Dh) + \alpha_1 \mathcal{J}_{\mathrm{AC}}(h,Dh). \tag{9.30}$$

For the sake of clarity, we usually do not express the dependence in Dh of $\mathcal{J}(x,h)$. However, we have to keep in mind that $\mathcal{J}_{\mathrm{Mean}}(x,h)$ and $\mathcal{J}_{\mathrm{AC}}(x,h)$ do depend on the Jacobian of the vector field h because of equation 9.1. Hence the equation 9.30.

9.5 THE GRADIENT OF THE DATA TERM

We show that the gradient of the data term exists in H and can be effectively computed and implemented for numerical experiments. The main ingredient in the proof is to show that $\delta_k \mathcal{J}(h,Dh)$ can be written as $(\mathcal{J}_h, k)_H + (\mathcal{J}_{Dh}, Dk)_H$, where $\mathcal{J}_h$ and $\mathcal{J}_{Dh}$ are complicated but computable functions of H. We spend the next sections to prove the following

Theorem 9.5.0.1. *For any $k \in \mathcal{F}$ the quantity*

$$\delta_k \mathcal{J}(h,Dh) = \lim_{\varepsilon \to 0} \frac{\mathcal{J}(h+\varepsilon k, Dh + \varepsilon Dk) - \mathcal{J}(h,Dh)}{\varepsilon}$$

exists and is equal to

$$(\mathcal{J}_h, k)_H - (\mathcal{J}_{Dh}, Dk)_H,$$

where the functions $\mathcal{J}_h$ and $\mathcal{J}_{Dh}$ are defined in the proof.

9.5.1 The first variation of $\mathcal{J}_{\mathrm{Mean}}(h,Dh)$

Because of equation 9.30 we have

$$\delta_k \mathcal{J}_{\mathrm{Mean}}(h,Dh) = \int_\Omega \delta_k \mathcal{J}_{\mathrm{Mean}}(x,h)\,dx.$$

Because of equations 9.5, 9.6 and 9.26 we have

$$\begin{aligned}\delta_k \mathcal{J}_{\mathrm{Mean}}(x,h) = D\mathcal{J}_{\mathrm{Mean}}(x,h)\varphi_3(\delta_k \hat{\mu}_2(x,h)) = \\ G\left(\varphi_3(\hat{\mu}_2(x,h)\log\left((\hat{\mu}_1(x))^{-1}\hat{\mu}_2(x,h)\right))\right)\varphi_3(\delta_k \hat{\mu}_2(x,h)),\end{aligned} \tag{9.31}$$

where $D\mathcal{J}_{\text{Mean}}(x,h)$ is the differential with respect to its first argument of the function $\mathcal{D}^2(\hat{\mu}_2(x,h),\hat{\mu}_1(x))$. Note that

$$G\left(\varphi_3(\hat{\mu}_2(x,h)\log\left((\hat{\mu}_1(x))^{-1}\,\hat{\mu}_2(x,h)\right))\right)$$

is a covariant vector of $T_{\hat{\mu}_2(x,h)}S^+$ while $\varphi_3(\delta_k\hat{\mu}_2(x,h))$ is a contravariant vector. We thus need to compute $\delta_k\hat{\mu}_2(x,h)$.

Computation of $\delta_k\,\hat{\mu}_2(x,h)$

We remember that the minimum of the functional

$$\mathcal{C}_2(\mu_2,x)=\frac{1}{\mid\Omega\mid}\int_\Omega \mathcal{D}^2(\mu_2,\mathcal{T}_h(I_2)(y))G_\gamma(x-y)\,dy,$$

is achieved at $\hat{\mu}_2(x,h)$. It therefore satisfies

$$\nabla_{\mu_2}\mathcal{C}_2(\hat{\mu}_2,x)=0.$$

To simplify notation, we note $\mathcal{L}$ the vector $\nabla_{\mu_2}\mathcal{C}_2$ of $T_{\hat{\mu}_2}S^+$. This vector is a 3×3 symmetric matrix which we identify with its image by φ_3, a vector of $\mathbb{R}^6$. The previous equation becomes

$$\mathcal{L}(\hat{\mu}_2(x,h),x,h)=0,$$

where the notation indicates that $\mathcal{L}$ depends upon h indirectly through $\hat{\mu}_2$ and directly through its definition. We compute δ_k of the lefthand side and equal it to zero.

$$\frac{\partial\mathcal{L}}{\partial\mu_2}(\hat{\mu}_2,x,h)\varphi_3(\delta_k\hat{\mu}_2(x,h))+\delta_k\mathcal{L}(\hat{\mu}_2,x,h)=0$$

We next compute $\frac{\partial\mathcal{L}}{\partial\mu_2}$, a once contravariant and once covariant tensor, a 6×6 matrix, as well as $\delta_k\mathcal{L}$. For the sake of clarity we drop in the sequel the index 2 in μ_2.

Computation of $\frac{\partial\mathcal{L}}{\partial\mu}$: According to equation 9.11 we have

$$\begin{aligned}\varphi_3^{-1}\left(\frac{\partial\mathcal{L}}{\partial\mu^I}\right)=&\frac{E_I}{\mid\Omega\mid}\int_\Omega\log\left(\mathcal{T}_h(I_2)^{-1}(y)\mu\right)\,G_\gamma(x-y)\,dy+\\&\frac{\mu}{\mid\Omega\mid}\int_\Omega\text{dlog}\left(\mathcal{T}_h(I_2)^{-1}(y)\mu,\frac{\partial\left(\mathcal{T}_h(I_2)^{-1}(y)\mu\right)}{\partial\mu^I}\right)\,G_\gamma(x-y)\,dy,\ I=1,\cdots,6,\end{aligned}$$

where the notation E_I has been defined in section 9.4.2. We also have

$$\frac{\partial\mathcal{T}_h(I_2)^{-1}(y)\mu}{\partial\mu^I}=\mathcal{T}_h(I_2)^{-1}(y)E_I.$$

Computation of $\delta_k \mathcal{L}$: $\delta_k \mathcal{L}$ is the sum of two terms corresponding to the variation with respect to h and Dh, respectively. We note them $(\delta_k \mathcal{L})^1$ and $(\delta_k \mathcal{L})^2$. We have

$$\varphi_3^{-1}((\delta_k \mathcal{L})^1) = \frac{\mu}{\mid \Omega \mid} \int_\Omega \text{dlog}\left(\mathcal{T}_h(I_2)^{-1}(y)\mu, \frac{\partial \left(\mathcal{T}_h(I_2)^{-1}(y)\mu\right)}{\partial h^i(y)} k^i(y) \right) G_\gamma(x-y)\, dy.$$

Because of equation 9.1,

$$\frac{\partial \mathcal{T}_h(I_2)^{-1}(y)\mu}{\partial h^i(y)} = -Dh(y)^T I_2(h(y))^{-1} \left(DI_2(h(y))\right)_i I_2(h(y))^{-1} Dh(y)\mu,$$

where DI_2 is the twice contravariant once covariant tensor obtained by taking the derivative of I_2 with respect to the space coordinates.

We introduce the thrice covariant tensor $\frac{\partial \mathcal{T}_h(I_2)^{-1}}{\partial h}$ such that

$$\frac{\partial \mathcal{T}_h(I_2)^{-1}}{\partial h}(y) = -Dh^T(y) I_2(h(y))^{-1} DI_2(h(y)) I_2(h(y))^{-1} Dh(y) \tag{9.32}$$

Because of the linearity of the function $\text{dlog}\,(,)$ with respect to its second argument we obtain

$$\varphi_3^{-1}((\delta_k \mathcal{L})^1) = -\frac{\mu}{\mid \Omega \mid} \int_\Omega \text{dlog}\left(\mathcal{T}_h(I_2)^{-1}(y)\mu, \frac{\partial \mathcal{T}_h(I_2)^{-1}}{\partial h}(y)\mu \right) k(y)\, G_\gamma(x-y)\, dy. \tag{9.33}$$

Note that for each value of the covariant index i, the matrix

$$\mu \text{dlog}\left(\mathcal{T}_h(I_2)^{-1}(y)\mu, \left(\frac{\partial \mathcal{T}_h(I_2)^{-1}}{\partial h} \right)_i (y)\mu \right)$$

is symmetric. At this point we introduce the corresponding once contravariant and once covariant tensor, noted $\frac{\partial \mathcal{L}}{\partial h}$:

$$\begin{aligned} &\left(\frac{\partial \mathcal{L}}{\partial h} \right)_i^{\cdot} (x, y, h) = \\ &-\frac{G_\gamma(x-y)}{\mid \Omega \mid} \varphi_3 \left(\hat{\mu}_2(x,h) \text{dlog}\left(\mathcal{T}_h(I_2)^{-1}(y) \hat{\mu}_2(x,h), \frac{\partial \mathcal{T}_h(I_2)^{-1}}{\partial h}(y) \hat{\mu}_2(x,h) \right) \right). \end{aligned} \tag{9.34}$$

We write

$$(\delta_k \mathcal{L})^1(x,h) = \int_\Omega \frac{\partial \mathcal{L}}{\partial h}(x,y,h) k(y)\, dy$$

$(\delta_k \mathcal{L})^2$ is obtained in a similar manner:

$$\varphi_3^{-1}((\delta_k \mathcal{L})^2) = \frac{\mu}{\mid \Omega \mid} \int_\Omega \text{dlog}\left(\mathcal{T}_h(I_2)^{-1}(y)\mu, \frac{\partial \left(\mathcal{T}_h(I_2)^{-1}(y)\mu\right)}{\partial Dh_m^l(y)} Dk_m^l(y) \right) G_\gamma(x-y)\, dy$$

where, because of equation 9.1,

$$\frac{\partial \mathcal{T}_h(I_2)^{-1}(y)\mu}{\partial Dh_m^l(y)} = 1_{ml} I_2(h(y))^{-1} Dh(y)\mu + Dh(y)^T I_2(h(y))^{-1} 1_{lm}\mu,$$

where $1_{lm}$and 1_{ml} are matrices whose only non zero element is located respectively at line l, row m or line m, row l. We introduce a once contravariant thrice covariant tensor $\frac{\partial \mathcal{T}_h(I_2)^{-1}}{\partial Dh}$ such that

$$\left(\frac{\partial \mathcal{T}_h(I_2)^{-1}}{\partial Dh}\right)_l^m (y) = 1_{ml} I_2(h(y))^{-1} Dh(y) + Dh(y)^T I_2(h(y))^{-1} 1_{lm} \tag{9.35}$$

and therefore

$$\frac{\partial \mathcal{T}_h(I_2)^{-1}}{\partial Dh}(y) Dk(y) = \left(\frac{\partial \mathcal{T}_h(I_2)^{-1}}{\partial Dh}\right)_l^m (y) Dk_m^l(y),$$

where l and m vary from 1 to 3. Using again the linearity of the function $\mathrm{dlog}\,(\,,\,)$ with respect to its second argument we obtain

$$\varphi_3^{-1}((\delta_k \mathcal{L})^2) = \frac{\mu}{\mid \Omega \mid} \int_\Omega \mathrm{dlog}\left(\mathcal{T}_h(I_2)^{-1}(y)\mu, \frac{\partial \mathcal{T}_h(I_2)^{-1}}{\partial Dh}(y)\mu\right) Dk(y)\, G_\gamma(x-y)\, dy. \tag{9.36}$$

Note that for each value of the covariant index l and contravariant index m, the matrix

$$\mu \mathrm{dlog}\left(\mathcal{T}_h(I_2)^{-1}(y)\mu, \left(\frac{\partial \mathcal{T}_h(I_2)^{-1}}{\partial Dh}\right)_l^m (y)\mu\right)$$

is symmetric. At this point we introduce the twice contravariant and once covariant tensor, noted $\frac{\partial \mathcal{L}}{\partial Dh}$, such that

$$\left(\frac{\partial \mathcal{L}}{\partial Dh}\right)_l^{\cdot m} (x, y, h) =$$
$$\frac{G_\gamma(x-y)}{\mid \Omega \mid} \varphi_3\left(\hat{\mu}_2(x,h) \mathrm{dlog}\left(\mathcal{T}_h(I_2)^{-1}(y)\hat{\mu}_2(x,h), \left(\frac{\partial \mathcal{T}_h(I_2)^{-1}}{\partial Dh}\right)_l^m (y)\hat{\mu}_2(x,h)\right)\right). \tag{9.37}$$

We write

$$(\delta_k \mathcal{L})^2(x,h) = \int_\Omega \frac{\partial \mathcal{L}}{\partial Dh}(x,y,h) Dk(y)\, dy$$

This allows us to compute $\delta_k \hat{\mu}_2(x,h)$ if the matrix $\frac{\partial \mathcal{L}}{\partial \mu_2}$ is invertible:

$$\varphi_3(\delta_k \hat{\mu}_2(x,h)) = -\left(\frac{\partial \mathcal{L}}{\partial \mu_2}(x,h)\right)^{-1} \int_\Omega \frac{\partial \mathcal{L}}{\partial h}(x,y,h) k(y)\, dy -$$
$$\left(\frac{\partial \mathcal{L}}{\partial \mu_2}(x,h)\right)^{-1} \int_\Omega \frac{\partial \mathcal{L}}{\partial Dh}(x,y,h) Dk(y)\, dy \tag{9.38}$$

all expressions being evaluated at $(\hat{\mu}_2, x, h)$.

We define the once contravariant and once covariant tensor

$$\mathbf{T}(x,y,h) = -\left(\frac{\partial \mathcal{L}}{\partial \mu_2}(x,h)\right)^{-1} \frac{\partial \mathcal{L}}{\partial h}(x,y,h), \tag{9.39}$$

and the twice contravariant and once covariant tensor

$$\mathbf{U}(x,y,h) = \left(\frac{\partial \mathcal{L}}{\partial \mu_2}(x,h)\right)^{-1} \frac{\partial \mathcal{L}}{\partial Dh}(x,y,h), \tag{9.40}$$

and rewrite equation 9.38 in a more compact manner, namely:

$$\varphi_3(\delta_k \hat{\mu}_2(x,h)) = \int_\Omega \mathbf{T}(x,y,h)k(y)\,dy - \int_\Omega \mathbf{U}(x,y,h)Dk(y)\,dy. \tag{9.41}$$

Using indexes, this is equivalent to

$$(\varphi_3(\delta_k \hat{\mu}_2(x,h)))^I = \int_\Omega \mathbf{T}_l^I(x,y,h)k^l(y)\,dy - \int_\Omega \mathbf{U}_l^{Im}(x,y,h)Dk_m^l(y)\,dy,$$
$$I = 1,\cdots,6,\ l,m = 1,\cdots,3$$

In the next sections we will also need the twice contravariant and once covariant tensor $\varphi_3^{-1}(\mathbf{T})$ and the thrice contravariant and once covariant tensor $\varphi_3^{-1}(\mathbf{U})$ which we note $\mathbf{t}$ and $\mathbf{u}$. The previous formula can be rewritten as

$$(\delta_k \hat{\mu}_2(x,h))^{ij} = -\int_\Omega \mathbf{t}_l^{ij}(x,y,h)k^l(y)\,dy - \int_\Omega \mathbf{u}_l^{ijm}(x,y,h)Dk_m^l(y)\,dy,$$
$$i,j = 1,\cdots,3,\ l,m = 1,\cdots,3 \tag{9.42}$$

An expression for $\delta_k \mathcal{J}_{\text{MEAN}}(h, Dh)$

We are in a position to prove the following

Proposition 9.5.1.1. *$\delta_k \mathcal{J}_{\text{MEAN}}(h, Dh)$ exists and is of the form of theorem 9.5.0.1.*

Proof. We introduce the once covariant tensor

$$\mathbf{t}_{\text{Mean}}(x,y,h) = D\mathcal{J}_{\text{Mean}}(x,h)\mathbf{T}(x,y,h),$$

the once covariant and once contravariant tensor

$$\mathbf{u}_{\text{Mean}}(x,y,h) = D\mathcal{J}_{\text{Mean}}(x,h)\mathbf{U}(x,y,h),$$

and write

$$\delta_k \mathcal{J}_{\text{Mean}}(x,h) = \int_\Omega \mathbf{t}_{\text{Mean}}(x,y,h)k(y)\,dy - \int_\Omega \mathbf{u}_{\text{Mean}}(x)Dk(y)\,dy,$$

or, using indexes

$$\delta_k \mathcal{J}_{\text{Mean}}(x,h) = \int_\Omega \mathbf{t}_{\text{Mean l}}(x,y,h)k^l(y)\,dy - \int_\Omega \mathbf{u}^m_{\text{Mean l}}(x,y,h)Dk_m^l(y)\,dy,$$

where, for example:

$$\mathbf{t}_{\text{Mean l}}(x,y,h) = D\mathcal{J}_{\text{Mean I}}(x,h)\mathbf{T}_l^I(x,y,h)$$

We finally introduce the once covariant tensor

$$\mathbf{T}_{\text{Mean}}(x,h) = \int_\Omega \mathbf{t}_{\text{Mean}}(z,x,h)\,dz,$$

and the once covariant and once contravariant tensor

$$\mathbf{U}_{\text{Mean}}(x,h) = \int_\Omega \mathbf{u}_{\text{Mean}}(z,x,h)\,dz,$$

From this follows the fact that $\delta_k \mathcal{J}_{\text{Mean}}(h, Dh)$ can be written in the form of theorem 9.5.0.1

$$\delta_k \mathcal{J}_{\text{Mean}}(h, Dh) = \int_\Omega \mathbf{T}_{\text{Mean}}(x,h)k(x)\,dx - \int_\Omega \mathbf{U}_{\text{Mean}}(x,h)Dk(x)\,dx$$

□

It is then possible to rewrite $\delta_k \mathcal{J}_{\text{Mean}}(h, Dh)$ in the form of a scalar product with $k(x)$. Indeed, integrating by part, and using the fact that k is zero on $\partial\Omega$, we obtain

$$(\mathbf{U}_{\text{Mean}}, Dk)_H = -\int_\Omega \text{div}\mathbf{U}_{\text{Mean}}(x,h)k(x)\,dx = -\int_\Omega \frac{\partial}{\partial x^m}\mathbf{U}^m_{\text{Mean}\,l}(x,h)k^l(x)\,dx$$

with x^m the $m^{th} \in [1,2,3]$ coordinate of the position $x \in \mathbb{R}^3$. Hence,

$$\delta_k \mathcal{J}_{\text{Mean}}(h, Dh) = \int_\Omega \left(\mathbf{T}_{\text{Mean}}(x,h) + \text{div}\mathbf{U}_{\text{Mean}}(x,h)\right)k(x)\,dx$$

and

$$\nabla_H \mathcal{J}_{\text{Mean}}(x,h) = \mathbf{T}_{\text{Mean}}(x,h) + \text{div}\mathbf{U}_{\text{Mean}}(x,h)$$

This is the first contribution to $\nabla_H \mathcal{J}(x,h)$ (see equation 9.3) to be used in our gradient descent.

9.5.2 The First Variation of $\mathcal{J}_{\text{AC}}(h)$

We would like to compute the second contribution to $\nabla_H \mathcal{J}(x,h)$, associated to the matching term for covariance matrices. The calculations are similar to the previous ones but very much involved so we detail them in appendix B.

9.5.3 Conclusion

We are now ready to give the proof of theorem 9.5.0.1.

Proof. It suffices to combine propositions 9.5.1.1, B.2.0.1, B.3.0.2 and B.4.0.3 to obtain

$$\delta_k \mathcal{J}(h, Dh) = (\mathbf{T}_{\text{Mean}} + \mathbf{T}^1_{\text{AC}} + \mathbf{T}^2_{\text{AC}} + \mathbf{T}^3_{\text{AC}}, k)_H - (\mathbf{U}_{\text{Mean}} + \mathbf{U}^1_{\text{AC}} + \mathbf{U}^2_{\text{AC}} + \mathbf{U}^3_{\text{AC}}, Dk)_H \quad (9.43)$$

□

This yields the existence of the gradient of the data term:

Proposition 9.5.3.1. *The gradient $\nabla_H \mathcal{J}(h)$ of data term $\mathcal{J}(h)$ exists and is given by*

$$\nabla_H \mathcal{J}(h) \;=\; \mathbf{T}_{\text{Mean}} + \mathbf{T}^1_{\text{AC}} + \mathbf{T}^2_{\text{AC}} + \mathbf{T}^3_{\text{AC}} + \text{div}\left(\mathbf{U}_{\text{Mean}} + \mathbf{U}^1_{\text{AC}} + \mathbf{U}^2_{\text{AC}} + \mathbf{U}^3_{\text{AC}}\right)$$

Proof. This is a direct consequence of the proof of theorem 9.5.0.1 and of equation 9.43. □

We thus know how to compute the gradient of the data and regularization terms. In the next section, we address some of the many numerical difficulties arising in the implementation of this registration technique. We also provide examples on two dimensional synthetic datasets, as a proof of concept.

9.6 NUMERICAL EXPERIMENTS

In this section, we will illustrate our method on three different examples. An in-depth study of many numerical aspects of this problem is still needed but we believe that current results, though simple, demonstrate the feasibility of the approach. The code was written in C++.

Up to now, we concentrated on the implementation of the gradient of the matching term $\mathcal{J}_{\text{Mean}}(x,h)$. The gradient of the other term $\mathcal{J}_{\text{AC}}(x,h)$ is much more tricky to compute and, most importantly, extremely time consuming because of the numerous numerical integrations to perform. The examples below were thus generated by only using $\nabla_H \mathcal{J}_{\text{Mean}}(x,h)$, which makes sense since we definitely want the local means to match before the local covariance matrices do. Regarding the computation of the gradient of the linear elasticity regularization term, $\nabla_H \mathcal{R}(h)$ (equation 9.3.2.1), we refer to [142] where adequate numerical schemes were given.

In order to recover large displacements, we used a multiresolution approach. The original images I_1 and I_2 were subsampled such that every level of the multiscale pyramid had a resolution equal to half of the resolution at the previous level. Details on that point can also be found in [142]. In the following experiments, we used 2 levels in the pyramid in addition to the original images. Subsampling was performed by computing local Riemannian averages of I_1 and I_2. Whenever the algorithm converged at a given level, it is easy to propagate the resulting diffeomorphism to the next level by bilinear interpolation. This serves as the initial value of the evolution for the next level. At the lowest resolution, the vector field h is initialized with the identity.

At each resolution, the evolution 9.3 requires the definition of a few parameters. First of all, we check for convergence by simply looking at the evolution of the energy 9.30. Whenever it stops decreasing for many iterations, we stop the gradient descent at the current level and propagate the estimated diffeomorphism to the next scale. The scale parameter γ, used to smooth the images at each resolution was fixed, in all our experiments to a small value, typically between 0.5 and 1 pixel.
Finally, the coefficient κ of the regularization term (see equation 9.8) and the time

step dt can be chosen, as proposed in [110], as follows. In order to control the range of the regularization term, κ is normalized by the maximum value $\kappa_0 = |\nabla_H \mathcal{J}(h)|_\infty$ such that $C = \kappa\kappa_0$. Using C instead of κ makes the algorithm much more stable and we used $\kappa = 10$. The time step dt is adapted at each level of the pyramid so that Cdt is less than a specified value. In our experiments, we set $Cdt = 0.2$.

It also makes sense to favor rigid transformations like translations and rotations at coarse resolutions. Consequently, at the first and second levels of the pyramid, we fit the estimated non-rigid deformation field with the best rigid transformation, expressed as the combination of a rotation and a translation, and this rigid transformation is used in place of the estimated non-rigid one. This is easily achieved by solving, for the 2D case, the following linear system

$$\underbrace{\begin{bmatrix} h(x_1)^1 & \cdots & h(x_N)^1 \\ h(x_1)^2 & \cdots & h(x_N)^2 \\ 1 & \cdots & 1 \end{bmatrix}}_{\tilde{X}} = \underbrace{\begin{bmatrix} R_{11} & R_{12} & T^1 \\ R_{12} & R_{22} & T^2 \\ 0 & 0 & 1 \end{bmatrix}}_{M} \underbrace{\begin{bmatrix} x_1^1 & \cdots & x_N^1 \\ x_1^2 & \cdots & x_N^2 \\ 1 & \cdots & 1 \end{bmatrix}}_{X}$$

where, for an image containing N pixels, $\tilde{X}$ is the $3 \times N$ matrix whose columns contain the components of the estimated diffeomorphism, X is the $3 \times N$ matrix whose columns contain the coordinates of each pixel and M is the 3×3 matrix containing the rotation matrix R and the translation T. $\tilde{X}$ and X being given, we simply have $M = \tilde{X}\left(X^T X\right)^{-1} X^T$. A QR decomposition of the submatrix R of M should be used to replace R by only its rotational component.

In figure 9.2, we present the very simple example of a translation. It is perfectly recovered. In figure 9.3, the more complicated example of a rotation is shown. We can see that the rotation is well recovered and, most importantly, that the tensor reorientation is correctly performed. Finally, we considered the case of a non-rigid transformation taking an ellipse onto a circle (figure 9.4). During the evolution, the algorithm first estimated a translation that maximized the overlap of the ellipse on the circle. Afterwards, non-rigid effects took the advantage in order to fully deform the ellipse into the circle.

9.7 CONCLUSION

We have presented a mathematical study of the non-rigid registration problem for diffusion tensor images. We setup a variational formulation taking into account the Riemannian structure of the space of diffusion tensors to derive the matching energy. To our knowledge, this is the first time that the properties of the manifold S^+ are exploited for DTI registration. As shown in this chapter, the computations are a bit tedious and the numerical implementation must be done carefully. We demonstrated the feasibility of the approach by successfully applying the algorithm to three different transformations. We now move into the third and last part of this manuscript. It is dedicated to the application of the connectivity mapping techniques, developed in this part, for the investigation of the architecture of the human brain visual and motor systems.

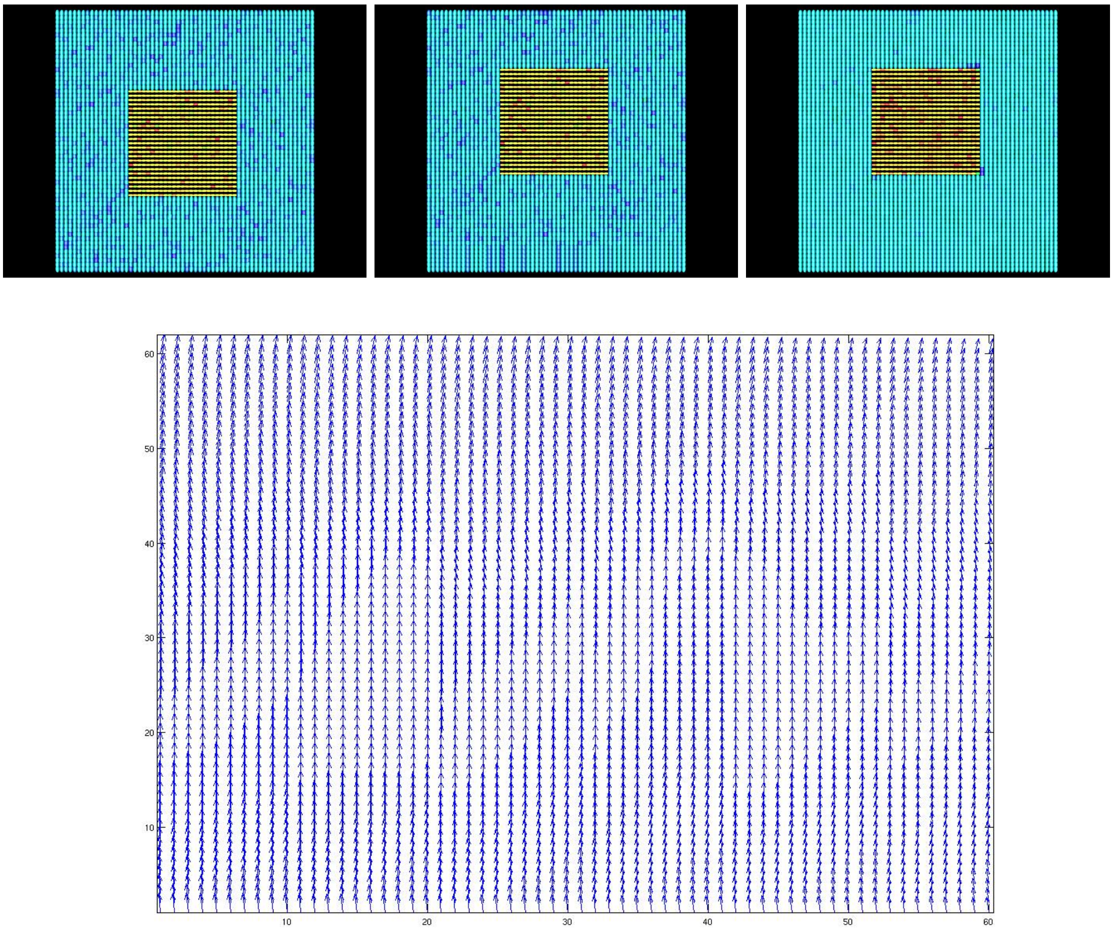


Figure 9.2: Estimation of the translation (6 pixels) of a square. I_2, I_1 and $\mathcal{T}_h(I_2)$ on top, h at bottom.

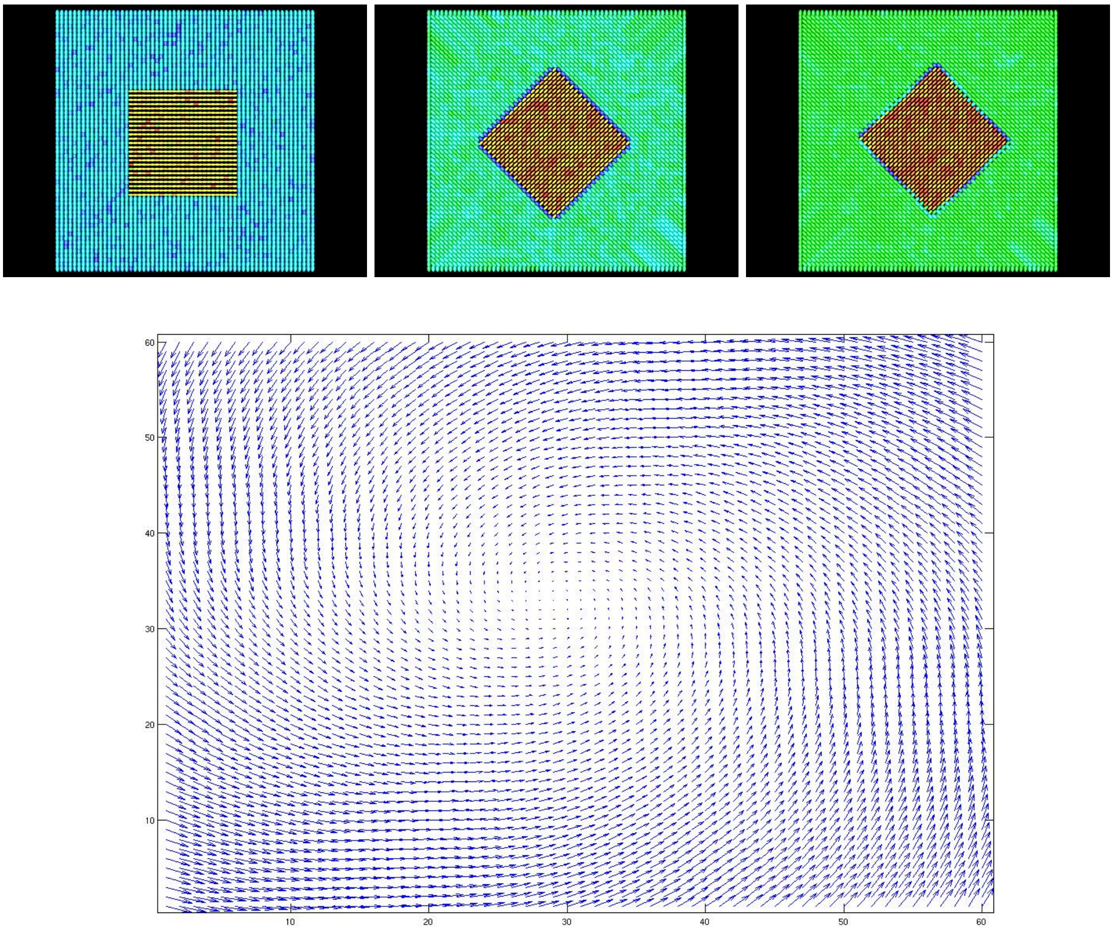


Figure 9.3: Estimation of the rotation ($\pi/4$) of a square. I_2, I_1 and $\mathcal{T}_h(I_2)$ on top, h at bottom.

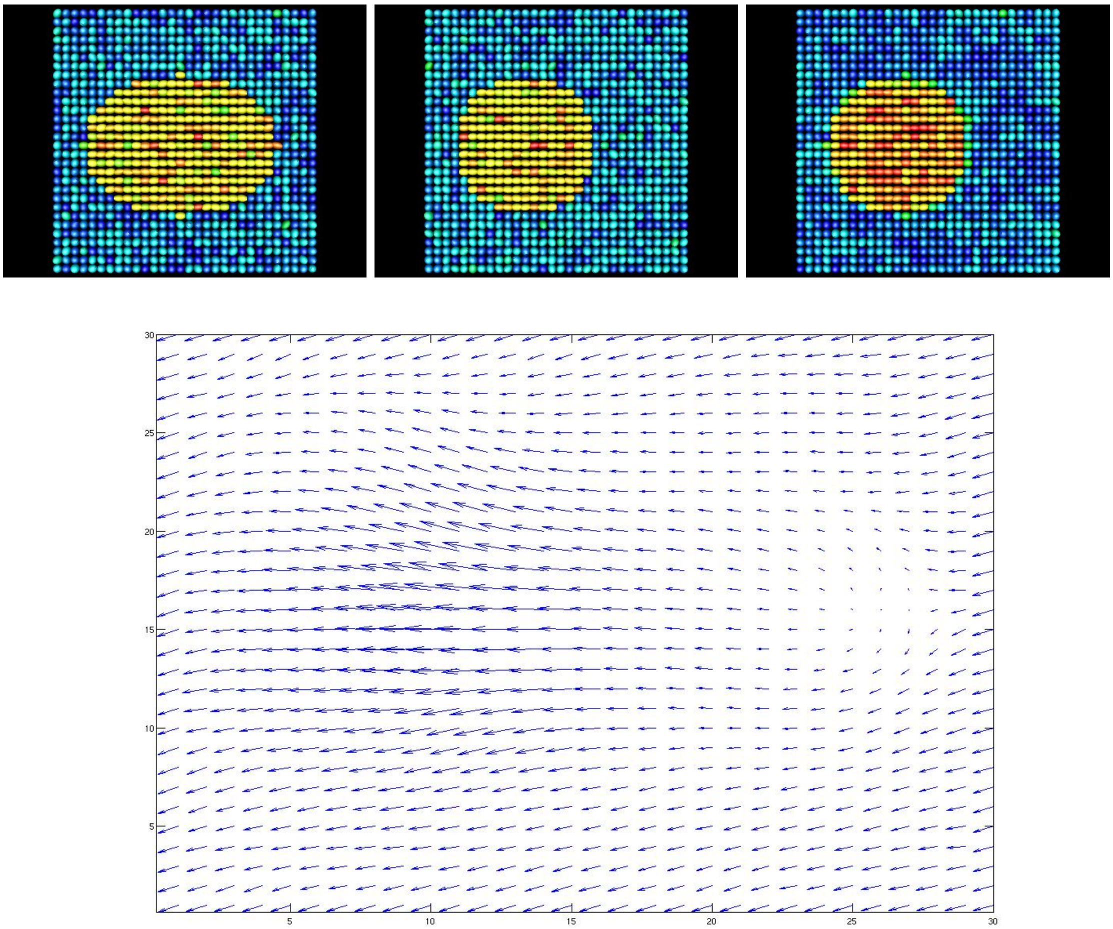


Figure 9.4: Estimation of the diffeomorphism taking an ellipse onto a circle. I_2, I_1 and $\mathcal{T}_h(I_2)$ on top, h at bottom.

Part III

Applications

CHAPTER 10

MAPPING CORTICO-STRIATAL CONNECTIONS INVOLVED IN MOTOR LEARNING

Contents

10.1 INTRODUCTION

In this short chapter, we report an application of the connectivity mapping technique detailed in section 7.3 to investigate the cortico-striatal circuitry in humans. We recall that this method relies on the use of the diffusion tensor as a Riemannian metric to measure the geodesic distance from a point of interest to the rest of the brain white matter. From this distance function, optimal paths (i.e geodesics) can be computed by back-propagation in the direction of the gradient of the distance function. Finally, it is possible to evaluate the likelihood of each estimated curve (i.e. geodesic) to represent a true fiber by looking at the statistics of some local connectivity index along the curve.

As reported by Lehéricy *et al.* in [178], it is possible to parcel the anatomical connections of the human striatum by the means of a deterministic tractography algorithm based on the method presented in [26]. Invasive animal studies showed indeed that the white matter fibers projections existing between the cortex and the striatum are organized in a set of discrete circuits [8]. Each circuit is assumed to perform distinct behavioral functions, such as movement execution and preparation, decision making, planning and learning, and thus to convey distinct information (sensorimotor, associative or limbic). In [178], the authors presented the first human study of the corticobasal ganglia connectivity and could successfully demonstrate that the posterior (sensorimotor), anterior (associative) and ventral (limbic) compartments of the human striatum had distinct connectivity pattern with the cortical areas. In this chapter, we try to precise these connectivity patterns by looking at connectivity maps associated to two different regions of interest of the striatum. But first, we would like to give a brief overview of the functions related to the sensorimotor, associative and limbic compartments of the striatum.

The sensorimotor (posterior) compartment of the striatum is involved in movement execution. Most recent PET and fMRI studies have shown that the striatum is constantly activated for all type of movements, including complex or simple finger movements [151], [180]. The striatum is probably not a key structure in coding basic movement parameters, such as frequency [254], amplitude [301] and force [93], as these parameters were not correlated until now to the amount of activation in this structure. Preparatory activation was also reported in the striatum, either in the putamen or the caudate nucleus. Activation in the putamen during preparation was located anteriorly than during movement execution (see [7] for instance), similarly to preparatory activity in cortical structures (SMA, motor cortex) suggesting that corticostriatal projections for preparatory and movement circuits are represented along distinct circuits, in line with animal studies [152], [256]. This is precisely what will be shown in the following.

The associative circuit is expected to be activated for more complex type of motor acts. Complex movements require additional motor or cognitive demands [86]. Numerous recent studies found that striatal activation depended upon the nature of the task and have shown activation in the associative (anterior) compartment during complex movement, including movement selection / decision, mental representation including mental simulation of grasping and hand movement simulation, working memory learning, and planning tasks. The selective involvement of the anterior striatum when subjects had to prepare a sequential action based on stored information and not during simple maintenance of information is in line with the hypothesis that striatal neurons convey information that are useful for behavioral acts. Contrary to older studies comparing skilled and newly learned movements [126], more recent studies have found that the caudate nucleus was activated during new explicit learning [279] by trial and error of a sequence of finger movements or arbitrary visuomotor associations. On the opposite, implicit learning (e.g. in the absence of any explicit knowledge) was associated with predominant activation in the sensorimotor part of the putamen [126]. Whereas putamen activity increased with learning in some studies [100] but not in others, the caudate nucleus was activated early in learning and decreased to baseline levels as learning progressed [279].

The limbic circuit is implicated in the motivational aspects of behavior. Several fMRI studies have reported activation in the ventral striatum and orbitomedial frontal regions during monetary reward or punishment, and expectation of monetary reward. Signal in the ventral striatum increased in response to novelty even in the absence of awareness, and when reward was unpredictable. The ventral striatum may thus function as a structure able to use prediction error signal to update successive predictions of future reward-associated events whereas the dorsal striatum encodes stimulus-response-reward associations so that actions associated with greater reward are chosen more frequently.

In summary, these results suggest that mental processes associated with motor, cognitive, and motivational aspects of behavior are represented within distinct cortico–basal ganglia circuits. They suggest that the basal ganglia have not only motor but also behavioral functions, and that they are particularly involved in the selection and inhibition of action commands, and reward-based learning. These results motivated our investigation of the 3D functional organization of the sensorimotor circuits. In particular, Although it is now pretty clear that the basal ganglia play a predominant role in the acquisition of motor skills [127], this study was motivated by the fact that the way such skills are built up and stored in the basal ganglia is not fully understood yet.

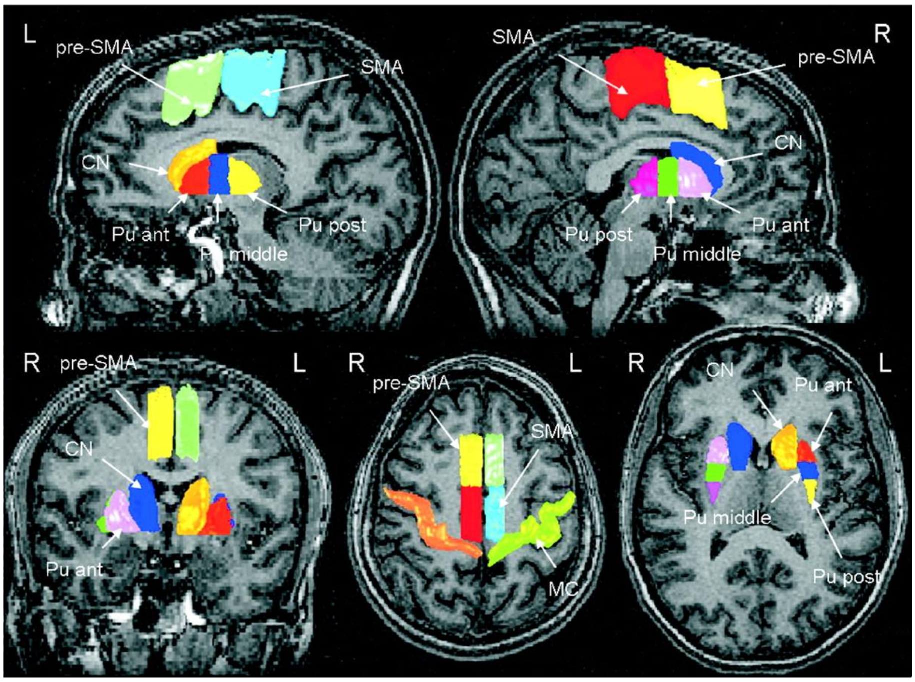


Figure 10.1: Regions of interest in the basal ganglia and the motor cortex. Abbreviations: ant, anterior; CN, caudate nucleus; L, left; MC, motor cortex; post, posterior; Pu, putamen; R, right; SMA, supplementary motor area

Keywords: basal ganglia, striatum, putamen, cortico-striatal connections, motor cortex, motor skills, learning, storage

10.2 FMRI/DTI STUDY OF MOTOR SKILLS LEARNING

As presented in [181] and [179], we tested the hypothesis that activation is transferred from the premotor / associative territories to the sensorimotor territory of the basal ganglia during the explicit learning of a motor sequence and that those territories were respectively connected to the premotor cortex (preSMA on figure 10.1) and the sensorimotor and posterior promotor cortex (sensory cortex and posterior SMA on figure 10.1).

Since it is now known that cortical areas project to the striatum in separate associative, premotor, and sensorimotor circuits [8], Lehéricy *et al.* postulated [181] that different cortico-basal ganglia circuits may be involved in the acquisition process of new motor skills. They showed indeed that anterior (associative) striatal regions are implicated during the acquisition of new motor skills, whereas posterior (sensorimotor) regions may be critical for the long-term storage of those skilled behaviors. The associative cortico–basal ganglia circuit is thus believed to be engaged at the beginning of learning and to contribute to the acquisition of an accurate representation of the sequence, whereas the sensorimotor circuit is thought to maintain a speedy representation of that skill when it is well learned and has become automatic.

In [181], 14 right-handed subjects were followed over a period of training of 4 weeks. At 4 weeks of training, automaticity was evaluated to determine whether subjects could perform the tasks with minimal interference.

10.2.1 Experimental Design

Subjects were asked to practice a trained sequence of eight moves by using fingers 2 to 5 of the non-dominant left hand over a period of 4 weeks. Subjects were asked to practice this sequence during 10 to 20 minutes daily, during which they were instructed to repeatedly tap a sequence in a rapid self-paced and accurate manner. The subjects' performance was assessed inside and outside the scanner by using a four key keyboard. Outside the scanner, control of task performance was performed 5 days of each of the 4 weeks as well as before and after every scanning session. These test consisted in performing the sequence as fast as possible during 30 seconds. fMRI acquisitions occurred on day 1, 14 and 28. During each session, subjects performed one run with the known sequence on which they trained and one run on an unknown sequence. On day 1, subjects were given two more runs of the known sequence: one before and one immediately after 30 min of additional practice on the known sequence while they were still lying on the scanner's bed, but without scanning.

10.2.2 Data Acquisition

The MR protocol was carried out at the Center for Magnetic Resonance Research, University of Minnesota, with a 3 Tesla whole-body system (Siemens, Erlangen, Germany). EPI BOLD (TR/TE/angle: 4.5s/40ms/90, voxel size: 1.5x1.5x2.5 mm^3, 123 acquisitions) and diffusion-weighted images (TR/TE/angle: 9.2s/92ms/90, voxel size: 2x2x2 mm^3, b-value: 1000s/mm2) images were obtained with 12 gradient directions. The anatomical image was acquired with a 3D MP-RAGE sequence (voxel size: 1x1x1 mm^3). Movements were audio-paced at 2Hz and alternated with rest. Sequences of identical length but different pattern were used for the unpracticed control state. Random effects group analysis was performed using SPM99.

10.2.3 Results

After 4 weeks of training, subjects made 58% less errors and were 97% faster (ANOVA, all $p < 0.05$). There was no significant change for the untrained sequences. With learning, activation decreased in bilateral anterior premotor, associative parietal areas, anterolateral thalamus, subthalamic nuclei, red nuclei, pons and cerebellum (lobules V,VI and crus I) (see figure 10.2).

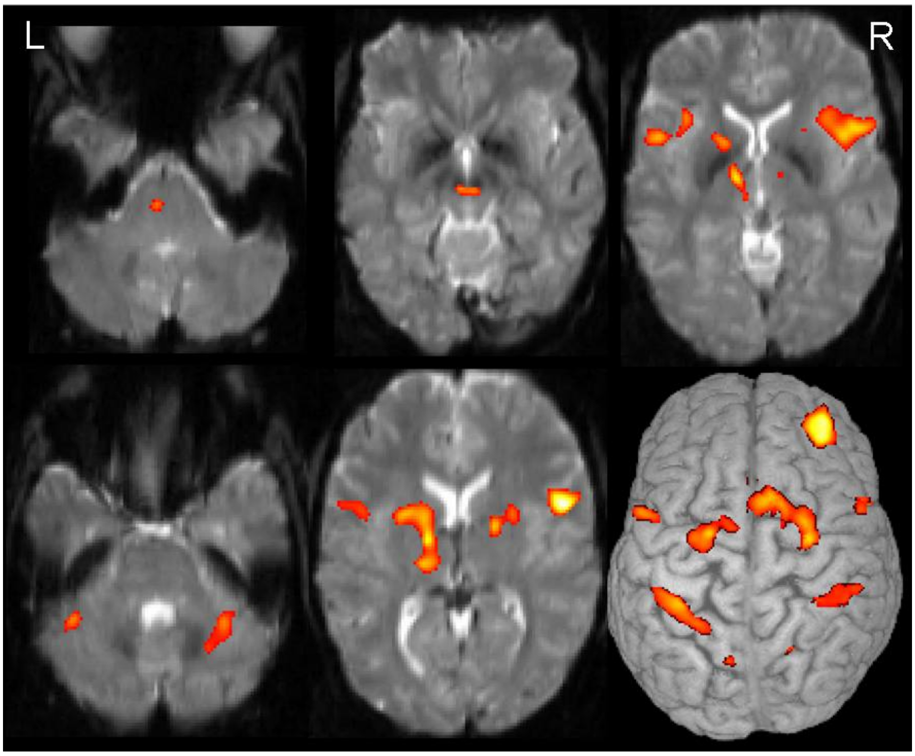


Figure 10.2: Areas with activation decrease during session 1, (before) versus (after 50 min of practice)

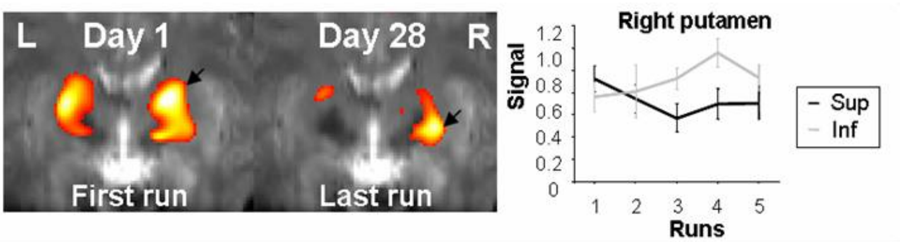


Figure 10.3: Observed foci of activation in the putamen

Two main foci of activation were observed in the anterodorsal (Figure 10.3, arrow in Day 1, coronal view) and more posteroventral parts of the putamen (arrow in Day 28, clusters corrected at $P < 0.05$). Regression analysis on percentage signal increase showed that activation decreased with practice in the dorsal part of both putamen (right putamen: $R2 = 0.64$, $P = 0.030$, Figure 10.3 right, dark grey curve), whereas activation increased bilaterally in the posteroventral areas ($R2 = 0.67$, $P = 0.046$, light grey curve). ANOVA showed a significant effect of the regions (ventral > dorsal, $P = 0.011$) and an interaction between runs and regions ($P = 0.045$).

By applying the DTI connectivity mapping technique presented in section 7.3, we could show that the anterodorsal putamen (Figure 10.4, left) was connected

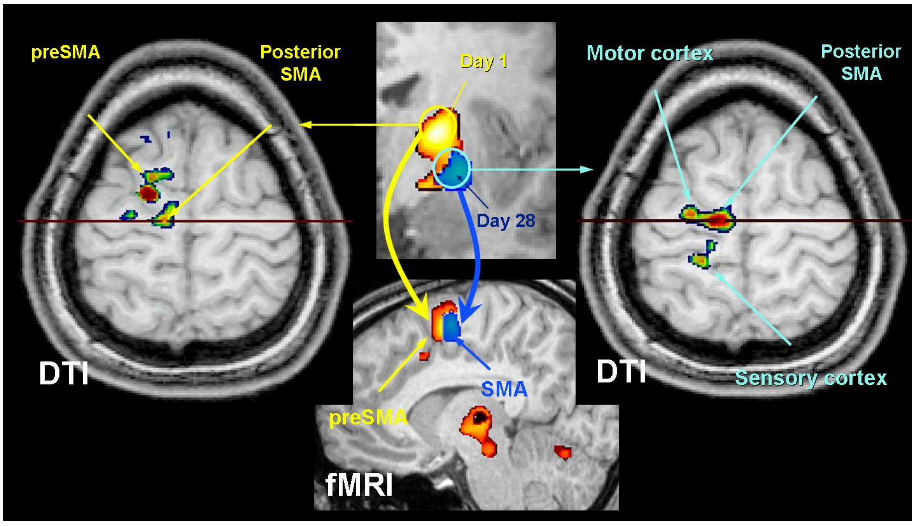


Figure 10.4: DTI connectivity of striatal activation maps before and after learning

with anterior premotor areas while the posteroventral putamen (Figure 10.4, right) was connected with the sensorimotor and posterior premotor areas. For each of the 14 subjects, diffusion weighted images were corrected for the distortions induced by eddy currents because of the large diffusion gradients. We used the algorithm proposed in [202]. Diffusion tensor images were then computed from the 12 diffusion-weighted images with the method proposed in section 6.1.2. A mask of the white matter was created by thresholding and applying morphological transformations (dilation, erosion) to the fractional anisotropy (FA) maps. We restricted those masks to the right side since we were only interested in corticostriatal connections of this hemisphere. Moreover, this permitted to reduce the computational time. We then chose the voxel with the highest activation in the activated anterodorsal and in the posteroventral putamen as points of interest. In other words, they were the origin of the computed distance functions (two per subject). Fiber tracking by back-propagation (equation 7.9) toward the origin of each distance function was performed and the average value of the connectivity index 7.10 (values in the range $[0, 1]$) was computed along each curve starting from each voxel of the mask and ending at the origin of the distance function. The highest this value at each voxel, the more confident in the connection originating for the voxel (and linked to the anterodorsal or posteroventral putamen) we are. Connectivity maps were thresholded to only display values at voxels with value greater than 0.8.

Figure 10.5 shows, for 4 subjects, the obtained connectivity maps for the anterodorsal putamen (top) and the posteroventral putamen (bottom). There were computed in each subject's native space, ie. no spatial normalization was performed. We can see that the anterodorsal putamen (top) is connected to anterior premotor areas and the posteroventral putamen (bottom) is connected to the sensorimotor and

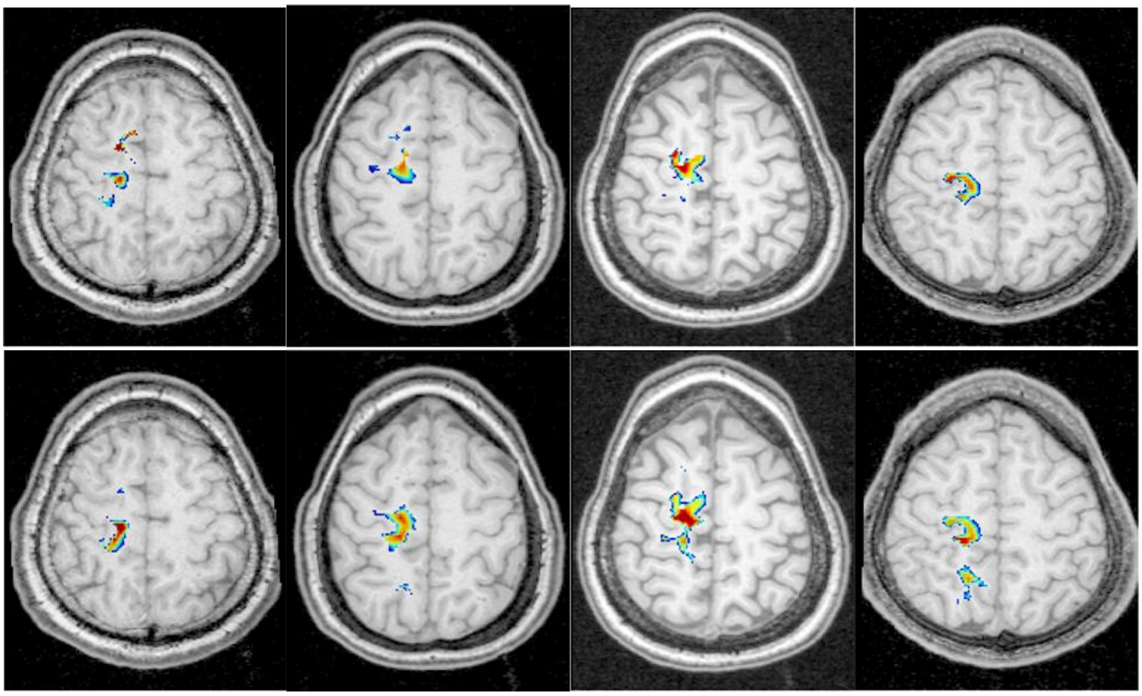


Figure 10.5: DTI connectivity of striatal activation before maps (top) and after (bottom) learning for 4 subjects in native space.

posterior premotor areas. In figure 10.6, we show the evolution of the connectivity index, for one subject, along the inferior-superior direction, before and after learning. We also present in figure 10.7 the corresponding estimated fibers. They were computed by the back-propagation method for all the voxels exhibiting a connectivity index greater than 0.8. They clearly show the shift of the corticostriatal projections.

Finally, we proceeded to the spatial normalization in Talairach space, with SPM99, of the 28 obtained connectivity maps (2 per subject). This allowed us to average the 14 normalized maps for the anterodorsal putamen and the other 14 for the posteroventral putamen, thus yielding group connectivity maps for the two regions of interest. These maps are presented in figure 10.8, after thresholding, and also clearly highlight the shift from the associative/premotor cortex to the sensorimotor cortex.

10.3 CONCLUSION

These results demonstrate a shift of motor representations from the associative/premotor to the sensorimotor territories of the basal ganglia during the course of explicit learning of motor sequences. The premotor territory of the striato-pallidal complex was engaged at the beginning of learning. It probably contributes to the acquisition of an accurate representation of the sequence. The sensorimotor ter-

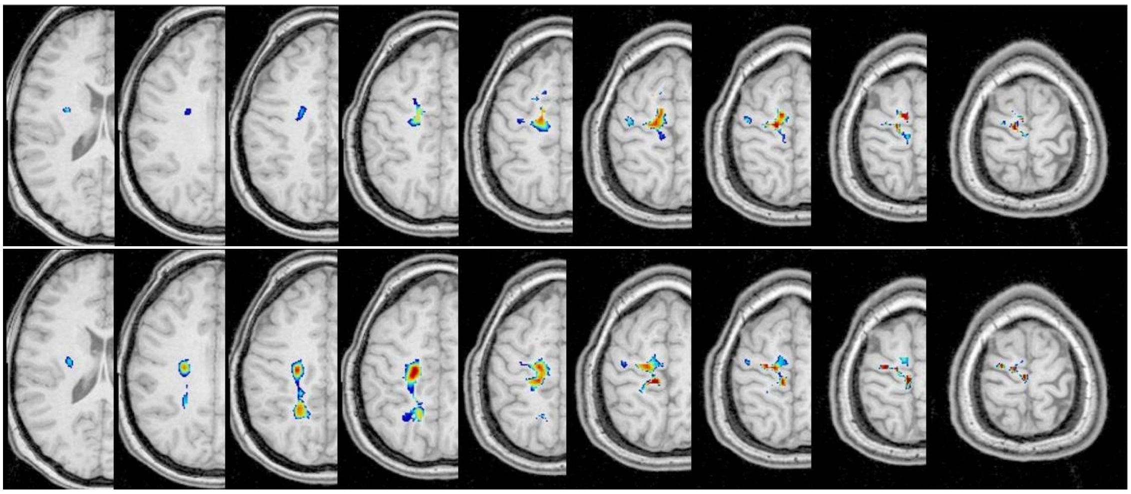


Figure 10.6: Evolution of the connectivity index, for one subject, along the inferior-superior direction, before (top) and after (bottom) learning

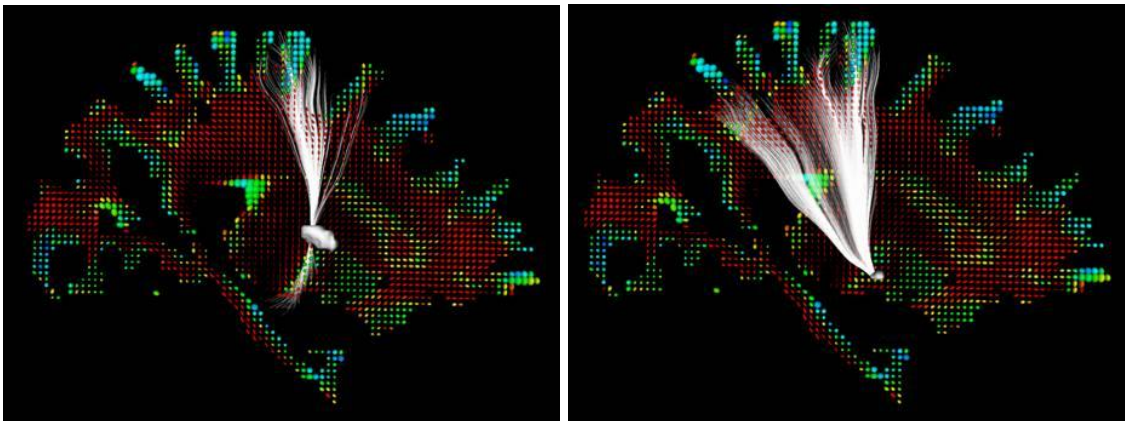


Figure 10.7: Estimated corticostriatal connections for subject of figure 10.6, before (left) and after (right) learning

ritory of the striato-pallidal complex was recruited during the late learning stage. It would maintain a speedy representation of the motor skill when it is well learned and has become automatic. By applying the connectivity mapping technique of section 7.3 to the anterodorsal and posteroventral putamen, we could successfully demonstrate their respective connections with the anterior premotor areas and the sensorimotor and posterior premotor areas, hence validating the hypothesis of the existence of distinct cortico–basal ganglia circuits involved in the acquisition and storage of motor skills. In the next chapter, we investigate the architecture of the human visual cortex by using the Geodesic Connectivity Mapping algorithm presented in chapter 7.

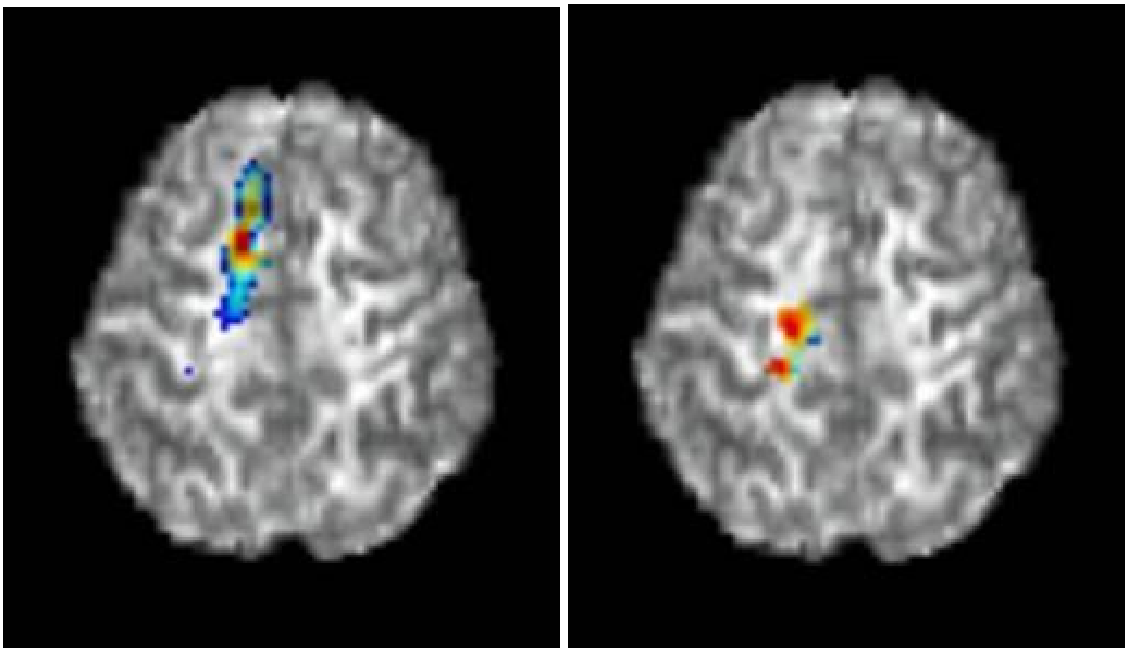


Figure 10.8: Average DTI connectivity of striatal activation maps before (left) and after (right) learning for 14 spatially normalized subjects

CHAPTER 11

MAPPING THE ANATOMICAL CONNECTIONS IN THE HUMAN VISUAL CORTEX

Contents

OVERVIEW

In this chapter, we validate the geometric tractography technique, Geodesic Connectivity Mapping (GCM), introduced in chapter 7 and presented as able to overcome the main limitations of geometrical approaches. Using the GCM technique, we could successfully characterize anatomical connections in the human low-level visual cortex. We reproduce previous findings regarding the topology of optic radiations linking the LGN to V1 and the regular organization of splenium fibers with respect to their origin in the visual cortex. Moreover, our study brings further insights regarding the connectivity of the human MT complex (hMT+) and the retinotopic areas.

Keywords: diffusion tensor MRI, Riemannian geometry, anatomical connectivity, retinotopic areas, hMT/V5+

11.1 INTRODUCTION

Understanding the relationship between anatomical structure and function is a fundamental issue in neuroscience. In the last decades, neuroimaging advances have been providing ever more promising means to non-invasively address this fundamental issue, thus opening the possibility to investigate in vivo normal and patients cerebral architecture and activity. This is especially true for Magnetic Resonance Imaging (MRI), which allows to combine anatomical (structural MRI), functional (functional MRI) and white matter connectivity (diffusion MRI) information at a spatial resolution of a few millimeters.
High resolution anatomical images are routinely used to segment cerebral tissues and extract 3D models of different tissues including the cerebral cortex. Various approaches have been proposed to achieve these computations [82, 77, 295]. Since its discovery in the early 90's [220], Blood Oxygen Level Dependent (BOLD) fMRI has been increasingly used to identify and functionally characterize various cortical areas. This is especially the case in the visual cortex were retinotopic mapping and functional localization can successfully be used [128, 302, 291].

Previous work on human visual cortex connectivity mapping: DTI based connectivity mapping of the human visual cortex has been addressed by various groups with different protocols and methods. Using a classical streamline tractography with smooth interpolation of the tensor field [212], [78] could reconstruct various bundles including visual pathway fibers. They showed fibers passing through the splenium, with a specific topology: anterior splenium fibers head to the parietal lobe while dorsal splenium fibers progress toward the occipital cortex. However, although the distinction between these two bundles is clear, the fibers they show fail to reach any precisely defined target on the grey matter and callosal-occipital fibers seem to rapidly converge to a single path (see figure 1 in [78]). They could, more convincingly, show a topology within the geniculo-occipital fibers, where medial (respectively lateral) LGN fibers terminate in a more superior (resp. inferior) part of the occipital cortex. Using a similar streamline tractography algorithm, [55] have identified in a single subject different visual fiber bundles including occipito-frontal and occipito-temporal fibers. In a subsequent study, [56] have identified different visual fiber bundles: (i) the optic tract from the chiasm to the LGN; (ii) the optic radiations from the LGN to the occipital cortex, which can be further divided into a ventro-temporal bundle ending in the lower lip of the calcarine sulcus and a dorsal bundle terminating in the upper lip of the calcarine sulcus; (iii) a splenium bundle connecting both occipital poles; (iv) U-shaped occipito-temporal fibers; (v) the inferior longitudinal fasciculus (ilf) directly connecting the extrastriate occipital cortex and temporal lobes. Despite these interesting findings, no functionally defined areas were used and data were mainly analyzed on a single mean DT image obtained through the averaging of different subjects, hence diminishing the possible interpretations of their findings when one considers the strong anatomico-functional variance across subjects.

The group of Ciccarelli, Toosy *et al.* used the Fast Marching Tractography (FMT) technique [231] to investigate 3 fiber bundles: the pyramidal tract, anterior callosal fibers and, more interestingly for our study, the optic radiations. They first concentrated on methodological issues, demonstrating the inter-subject reproducibility [72] as well as the inter-observer reproducibility [70] of the reconstructed tracts. More recently, they applied this technique to study changes in the optic radiation of patients affected by a specific optic nerve pathology (optic neuritis) [71]. Note that the FMT method has been partly validated in a combined study on macaques and humans, showing a part of the cortico-spinal tract and the optic radiations in both species [233]. Note however that the authors acknowledge the currently limited spatial resolution in DTI which prevents reliable tractography in macaques and therefore a true validation.

As of today, only a few studies, like [164], have combined fMRI and DTI to study the visual cortex. In the above mentioned study, [78] used fMRI activation maps to roughly identify the LGN and the occipital visual cortex. [309] demonstrated that the fractional anisotropy was lower in the activated occipital cortex than in the optic radiations. This is consistent with the known relative isotropy of grey-matter voxels as compared with white-matter voxels [243]. Using a probabilistic tractography method, [280] completed this work, showing a correlation between the estimated optic tracts mean FA values and the degree of fMRI activity within the visual cortex but, like the former, they did not functionally identify the occipital visual areas they were considering. To our knowledge, the most complete and precise study in the literature was done by [99] at Stanford. They combined a classical streamline tractography method with a functional identification of occipital retinotopic areas to recover occipital fiber tracts passing through the splenium. They demonstrated a good spatial matching in the splenium of independently estimated fibers starting from left and right occipital poles. More specifically, they found a topology in these tracts in which (i) dorsal (respectively ventral) cortical regions project dorsally (resp. ventrally) into the splenium, in agreement with a macaque autoradiography study, (ii) a foveal-periphery gradient can be found in the anterior-posterior direction of the splenium.

In the present study, we used the Geodesic Connectivity Mapping (GCM) technique, introduced in chapter 7 to assess the anatomical connectivity across individually identified areas in the low level visual cortex. Besides its robustness and efficiency, this approach naturally restricts the estimation within the white matter voxels and further provides statistics of a local connectivity measure along each estimated fiber. We first validate this methodology by recovering typically known fiber tracts, reproducing previous findings [99, 78, 56, 72, 70], before addressing new issues regarding intra-hemisphere connectivity in the occipital visual cortex.

11.2 METHODS

11.2.1 MR Data Acquisition

3 healthy subjects with normal or corrected to normal vision participated in two separate scanning sessions, to acquire the functional and diffusion weighted images, respectively. In each session, a high resolution anatomical scan was acquired and later used as references to coregister the two sessions. All scans were acquired on a 3 Tesla MEDSPEC 30/80 AVANCE (Bruker) at the Centre IRMf de Marseille, France, using a quadrature bird-cage head coil.

In the first session, the functional scans, later used to identify the retinotopic areas and the hMT+ complex, and diffusion weighted images were acquired. However, due to an acquisition problem, the phase map could not be reconstructed which is particularly problematic considering the important geometric distorsions of the echo-planar diffusion weighted images (see below). As soon as this problem was solved, we acquired in a second session new diffusion weighted images and the corresponding phase map for the same subjects. As we were not acquiring functional images, we took advantage of the saved time to increase the number of repetitions for each direction, thus increasing the diffusion-weighted images signal to noise ratio.

Structural MRI

High resolution anatomical scans were acquired through a 15 min. 3D-gradient echo sequence with inversion-recovery (TE=5ms, TR=25ms, IT=800ms, acquisition matrix size: 256x192x104 and reconstructed to 256x256x128, voxel size 1x0.75x1.22mm^3).

fMRI

During each functional scan, 151 Echo Planar Images were acquired over 5 min. 19 sec. using a coronal sequence (TE=35ms, TR=2111ms, Angle=70°). Each functional image spans 20 coronal slices 3mm thick and 2x2mm^2 in plane resolution, approximately orthogonal to the calcarine sulcus covering the occipital retinotopic areas and extending anteriorly to confidently include hMT+ region [103]. The first five images were systematically discarded to avoid magnetic saturation effects. The 144 following images correspond to the visual stimulus per se. The last two images were taken to allow slice-timing correction preprocessing.

Diffusion Tensor Imaging

In a pilot study, we first tried different acquisition parameters to obtain the best diffusion weighted images, having fixed TR=10000ms, TE=86ms, voxel size 2x2x2mm^3. We finally used 12 diffusion directions for a single b-value of 1000 s.mm^{-2}, which is consistent with other studies [140] and allows to increase the number of repetitions for each gradient direction to achieve a better SNR. Each gradient directions was hence repeated 10 times. 8 standard T2 images (i.e. without diffusion sensitization or b =0 s.mm^{-2}) were also acquired.

11.2.2 Visual Stimuli

Stimuli were generated under Matlab 6.1 using the Image Processing Toolbox (Matlab, The Mathworks), providing an AVI file with eighteen 300x300 pixels frames per second and lasting 5 min. 04 sec. Visual stimulation was synchronized with the acquisition through a trigger sent by the scanner at the beginning of each scan. Stimuli are displayed at 72Hz by a SONY video-projector, placed in a custom-designed Faraday cage inside the scanner room, onto a large adjustable mirror then onto a translucent screen inside the bore at the back of subjects head and finally reflected by a custom designed mirror placed at 5cm above subjects eyes. This setup leads to a display subtending a visual angle of 20.9°x20.9°. The stimuli are all presented within a circular aperture of 19.5° in diameter. During the first 5 and last 2 scans, a mid grey-level image with the 0.5° red fixation cross was shown to the subjects.

Phase-encoded stimulus for the retinotopic mapping consisted of a 9Hz flickering black-and-white checkerboard into a 80° rotating wedge; 8 complete rotations, either clockwise or counter-clockwise, were performed in each scan.

A classical block design paradigm was used to reveal the human mid-temporal complex, hMT+, supposed to comprise human homologues of at least macaque MT and MST [102, 145]; the stimulus was a black and white random dot pattern (maximum contrast) on a mid-grey background with a 10.28 dots.deg^{-2} density presented in 2 different conditions, each presented during blocks of 8 RT (16.888sec): static (STA), i.e. the same image for all the block duration and coherent (COH) when dots were moving at 7.53 deg.sec^{-1} radially in a 2Hz inward and outward alternation. Dots leaving the mask were replaced through a radial wrap-around constrained to keep the dots density constant.

11.2.3 Image Processing Pipeline

Each dataset was analyzed on a subject basis to avoid undesirable effects such as the strong smoothing implied by standard normalization procedures. [309] showed it was possible to obtain geometrically matched fMRI and DWI with appropriate acquisition sequences, therefore avoiding various distortions correction steps. However, these images are not coregistered with the anatomical image and this procedure supposes to acquire data with the same volume prescriptions. In this study, since (i) we considered complementary information from 3 different MRI modalities (anatomical, functional and diffusion-weighted images), (ii) acquired two distinct sessions for each subject and (iii) used different slice prescriptions for the different modalities, a common or reference analysis space was required. We used the mean T2-weighted image, further corrected for EPI geometric distortions (see below), as reference space. We note *umean T2* this reference image. This choice minimized the deformations and interpolations of the diffusion-weighted images acquired within the same run using a similar sequence. Each type of image received specific processing detailed in the following paragraphs and the extracted useful information was finally coregistered to the *umean T2* reference image. Figure 11.1 summarizes the overall processing pipeline used in this study.

Anatomical image

High resolution anatomical images acquired in both sessions allowed precise intersession coregistration using SPM2 algorithm. We note $M1$ the estimated transformation mapping anatomical image from session1 to anatomical image from session2. The latter was further coregistered with the *umean T2* reference image by transformation $M2$. Structural information was processed from both anatomical scans to segment grey and white matter tissues and build a 3D model of the cortical surface using a combination of the BRAINVISA[1] package and local methods [109].

---

[1] `http://brainvisa.info/index.html`

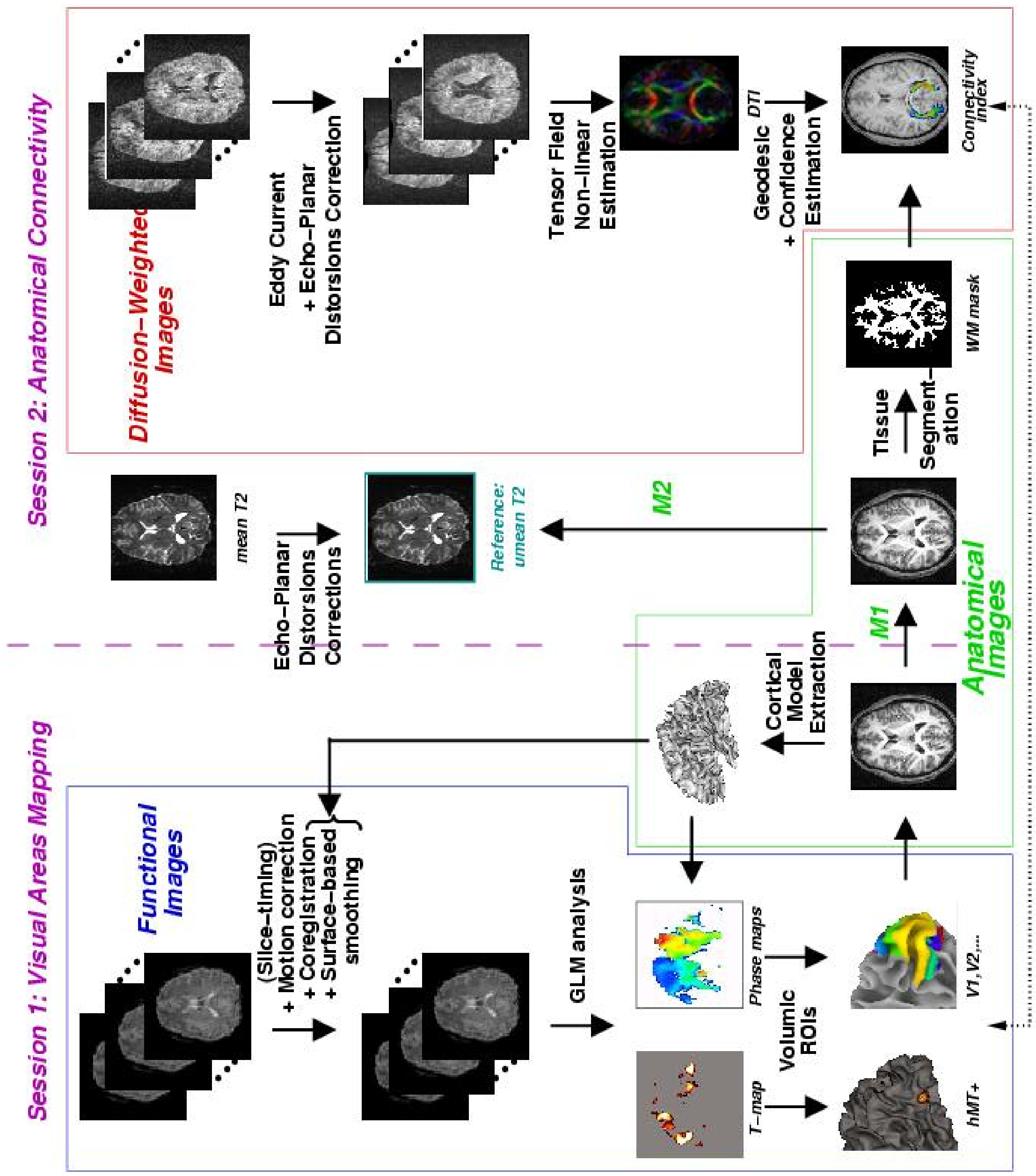


Figure 11.1: The image processing pipeline. Functional, diffusion-weighted and anatomical images receive specific processing and are realigned on a reference image, *umean T2*, for the connectivity maps computation (see text for details).

Visual areas functional definition

Statistical Parametric Mapping (SPM2[2]) was used to analyze functional data for the retinotopic and hMT+ mapping. Images were first realigned with the INRIalign software [120] and coregistered with the anatomical image with SPM2, then smoothed through an appropriate cortical surface based smoothing method described elsewhere [313], with a 3mm equivalent Gaussian filter FWHM. Temporal high-pass and low-

---

[2] `www.fil.ion.ucl.ac.uk/spm/`

pass filtering were also performed on the time-courses to respectively remove low-frequency signal drifts and high frequency noise. Retinotopic mapping and hMT+ localizer data were then analyzed separately with respective general linear models [123].

Retinotopic stimuli were modeled by cosine and sine functions at the stimulus frequency and a voxelwise F-test thresholded at $p < 0.001$ was used to identify responding voxels. The local signal phase, related to the stimulus position, was then given by the arctangent of the ratio of the parameter estimates for both regressors [314].

Each condition of the hMT+ localizer stimulus was modeled by a classical boxcar function convoluted with a canonical hemodynamic response function model. A T-contrast was computed to compare COH and STA conditions and thresholded at $p < 0.001$. We considered as hMT+ the activated cluster within or close to the inferior temporal sulcus [306, 103].

Subsets of connected voxels were extracted for each area and further used as ROIs in the connectivity analysis. Since diffusion anisotropy is relatively low in grey matter voxels [243], such as in the visual cortex [309], we defined white matter ROIs, considering the white matter voxels closest to the cortical ROIs. Specifically, volumic ROIs were automatically computed from their identification on the GM/WM interface by projecting the respective surface-based labels along the surface normal inside the white matter voxels. Each ROI voxels subset was then coregistered to the *umean T2* reference image by the transformation $M1 \circ M2$ and further masked to solely lie within the white-matter mask extracted from the high resolution anatomical image. Possible intersections between each pair of white matter ROIs were automatically removed from the analysis.

Diffusion weighted images (DWI)

T2 image: the 8 T2-weighted images were motion corrected using the INRIAlign software before being averaged. The resulting mean T2 image was then processed to correct geometric EPI distortions caused by magnetic susceptibility inhomogeneities, i.e. magnetic field inhomogeneities particularly found at the interfaces between different tissues [153]. Based on the phase map acquired during the DWI data recording session, which maps the spatial distribution of field inhomogeneities, we used the SPM interfaced toolbox “Fieldmap”[3] to compute and apply a voxel displacement map accounting for these susceptibility artifacts. As mentioned above, the resulting *umean T2* image served as reference image for connectivity maps computation.

DTI: DWI data were first preprocessed to minimize the distortions induced by eddy-currents and related to the large diffusion-sensitizing gradients [201]. We used

---

[3] `www.fil.ion.ucl.ac.uk/spm/toolbox/fieldmap`

the algorithm proposed by [201] and implemented within the BrainVisa package. Briefly, this method uses a 2D image registration technique to realign each DWI slice with its corresponding standard T2-weighted slice. A scale factor, a translation and a shearing are the parameters for the slice and image dependent affine transformation searched. The mutual information is used as a similarity measure to estimate the transformation parameters. We then applied to the resulting images the EPI geometric distortions correction algorithm used for the mean T2 image. The diffusion tensor image (DTI), a field of 3x3 real symmetric positive-definite tensors along the image domain, was finally computed with the method presented in section 6.1.2.

11.2.4 Seed Voxels Placement

A crucial aspect for any fiber tracking method is the location of the initial seed. The seeds for the GCM algorithm were selected depending on the considered tracts.

LGN seed voxels identification

Lacking a precise functional localization of the LGN, we first identified LGN seeds voxels with a classical streamline technique. To do so, we manually selected in each hemisphere a rough thalamus sub-region which obviously included the expected LGN location. More specifically, based on both anatomical and diffusion tensor image prior information, the initial region was identified anterior to the lateral ventricles and only voxels with a relatively high anisotropy (FA ≥ 0.15) were kept. Diffusion tracts starting from each selected voxel were estimated with an implementation of a classical streamline tractography technique [174] and further automatically filtered to keep the fibers heading to the ipsilateral retinotopically identified area V1. Only fibers reaching a 3 voxels wide band around the functionally defined V1 region were kept. This approach is very similar to that of Conturo *et al.* [78], although we did not oversample the DTI data, thus getting less fibers than in the latter work. The starting voxels of the remaining fibers were finally labeled as the LGN voxels. We typically found a region of 5 connected voxels in each hemisphere, consistent with the reported LGN size both in previous anatomical [143] and fMRI studies [65]. Beyond yielding an anatomical connectivity based delineation of the LGN, the reproduction of the well-known visual pathway as well as the likely extent and location of the LGN ROIs validates our diffusion-weighted images quality as well as our image processing pipeline.

Splenium seed voxels

Seed voxels in the splenium were manually traced on a mid-sagittal slice of the anatomical image as the most posterior and ventral portion of the corpus callosum. The voxels subset was further masked by the white matter tissue mask.

Other tracts

For connectivity mapping starting from the functionally identified visual areas, we simply used the white-matter ROIs defined with the procedure detailed above.

11.2.5 Connectivity Maps and Fiber Tracts Computation

Streamline technique

To date, streamline techniques are very often used since their implementation is quite straightforward and the basic idea underlying the algorithm is intuitive: at each voxel, the eigenvector associated to the largest eigenvalue is supposed to reflect the microstructure of the underlying tissue and parallels the mean fiber orientation in that voxel. Hence, the major eigenvector field is treated as a flow field and integrated in order to estimate streamlines, expected to coincide with white matter fibers. In the implementation we used, the integration is typically performed numerically with subvoxel precision through a 4^{th} order Runge-Kutta scheme [182].

The Riemannian geometry framework applied to DTI

We used the approach proposed in chapter 7 to compute (i) a distance function to a given point of interest (or *seed point*) x_0, (ii) the putative fiber path linking any voxel of a given brain region V to x_0 and (iii) a connectivity map, i.e. a confidence measure associated with each fiber. We refer to this technique as Geodesic Connectivity Mapping (GCM). Considering the low anisotropy in the grey matter tissue, we will consider for V the set of white matter voxels obtained from a segmentation of the anatomical image (see methods). Moreover, we used for the confidence measure, the statistics along the fibers of the following local quantity:

$$\mathcal{C}(x) = |f^*(x)|_E$$

where $|.|_E$ is the Euclidean norm. $f^*(x)$ corresponds to the so-called optimal dynamics at voxel x (see section 7.5). Since each voxel $x \in V$ can be assigned a geodesic reaching x_0, we can compute a pair $(\mu(x), \sigma(x))$ at each voxel. $\mu(x)$ is the mean of $\mathcal{C}(x)$ along the geodesic and $\sigma(x)$, the standard deviation. In the following, we call μ-map and σ-map the respective images of μ and σ values.

The results presented below were systematically obtained with the GCM algorithm. Notice however that the level set method of section 7.3 led to qualitatively similar results, although the computation time was by far higher and the numerical issues mentioned in chapter 7 could lead to anatomically impossible front propagation and connections, requiring iterative manual modifications of the white matter mask.

Connectivity indices computation

We used the following methods to compute connectivity within the splenium region and to estimate hMT+ connectivity with retinotopic areas.

To study splenium tracts, we analyzed connectivity maps in the splenium voxels, starting from our retinotopically (or functionally for hMT+) defined ROIs. Each ROI was considered separately. We note $X = (x_i)_{i=1,...,n}$ a specific n voxels seed ROI (e.g. left hemisphere V1) and $Y = (y_j)_{j=1,...,m}$ the m splenium voxels identified on a mid-sagittal slice. For each seed voxel x_i, the corresponding μ-map and σ-map were computed with the GCM method. We therefore have the mean and sigma values for each optimal path $\gamma_{i,j}$ linking x_i to y_j. We then filter these maps to remove the highest variance paths and compute a single mean μ-map in the splenium. Specifically, for each splenium voxel y_{j_0}, we have n putative paths γ_{i,j_0}. We discard a given proportion p of these n connectivity paths, removing paths with highest variance σ. The mean connectivity indices of the remaining putative fibers are then averaged, leading to a single mean value μ at voxel y_{j_0}. The procedure is repeated for each $y_j, j = 1, ..., m$. The resulting μ map is interpreted as the mean connectivity between area X and the splenium. p was arbitrarily set to 10%, but the qualitative results did not differ for $5\% \leq p \leq 20\%$.

Regarding hMT+ connectivity with retinotopic areas, the connectivity indices were computed as follows. Taking as seeds each hMT+ voxel $(x_i)_{i=1,...,n}$, we compute the μ and σ maps with the GCM method. For each x_i, we then discard a given proportion p of paths with highest σ values among the m paths linking target ROI voxels $(y_j)_{j=1,...,m}$ to x_i. The mean μ value of the remaining paths is then computed and assigned to voxel x_i. At each hMT+ voxel, we end up with a mean connectivity value for each retinotopically defined target ROI. Similarly to splenium fibers, p was arbitrarily set to 10%, but the qualitative results did not differ for $5\% \leq p \leq 20\%$.

11.3 RESULTS

We first validated our protocol and connectivity mapping technique on the previously characterized optic radiation tracts before investigating callosal connectivity and intra-cortical connectivity across the functionally identified visual areas.

11.3.1 Optic Radiations

Since the optic radiations were often reconstructed in diffusion tractography studies [78, 56, 70, 72], we decided to start the validation of our fiber tracking approach by considering this well characterized fiber bundle, which links the Lateral Geniculate Nucleus (LGN) to area V1 in the occipital cortex.

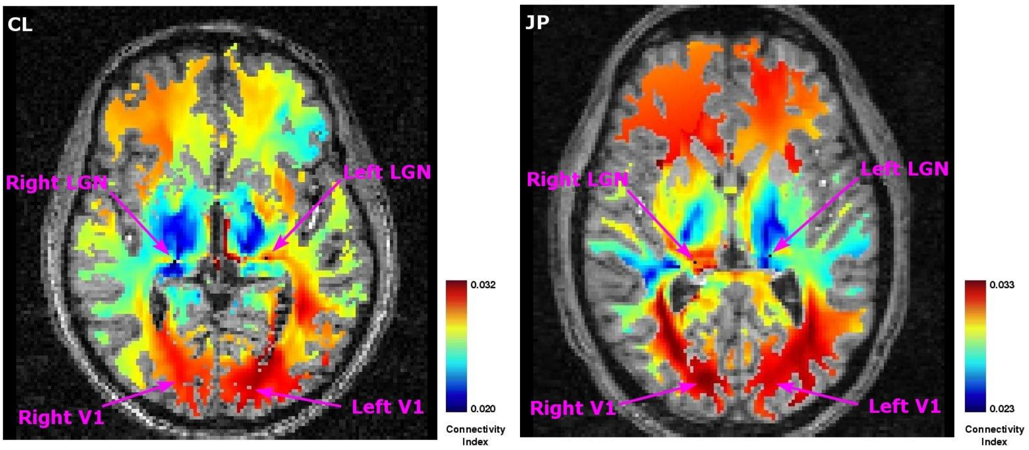


Figure 11.2: Connectivity μ-maps obtained in two subjects estimated separately from one LGN seed voxel (in black) for each hemisphere. The highest connectivity index values are found along the putative optic radiations paths, with the maximum value within area V1.

Starting from each previously identified LGN voxels (see methods), we computed the connectivity index maps with the GCM method. As we were not concerned here with inter-hemispheric connections, the GCM computation was restricted to the ipsilateral hemisphere of the seed voxel. To compare our method with a standard approach, we also reconstructed the fibers leaving the LGN with a classical streamline technique. Figure 11.2 shows one μ-map per hemisphere in two subjects. As each map is restricted to its respective hemisphere, we merged them in a single image and overlaid the result on an axial slice of subjects' anatomical image. The seed voxel of each μ-map is shown in black. The highest connectivity index values (dark red) are systematically found within the typical path of the optic radiations. Besides, highest values were found in the retinotopically identified V1 region. These results were found for each LGN seed voxel in the 6 hemispheres we analyzed. Consequently, for each connectivity maps, the voxel with maximum connectivity mapping index (which lay in area V1) was identified and the geodesic linking that voxel and the seed point was traced. Figure 11.3 shows the reconstructed fiber bundles obtained with the streamline technique (blue) and the GCM technique (red). Although the thalamo-occipital fibers estimated with streamline and geodesic methods qualitatively match and are consistent with known anatomy, we noted some differences between reconstructed tracts. Most fibers estimated by streamline propagation fail to reach the V1 white matter ROI, unexpectedly heading in a ventral direction a few millimeters before reaching the V1 region. We attribute this unexpected trajectory ending to an improbable connection with another fiber bundle crossing the thalamo-occipital track. This observation led us to use a relaxed constraint to filter the fibers passing closely to V1, as mentioned in the above description of the LGN seed voxel identification procedure (see methods). Note that [78] also used a 1cm band within the white matter, lat-

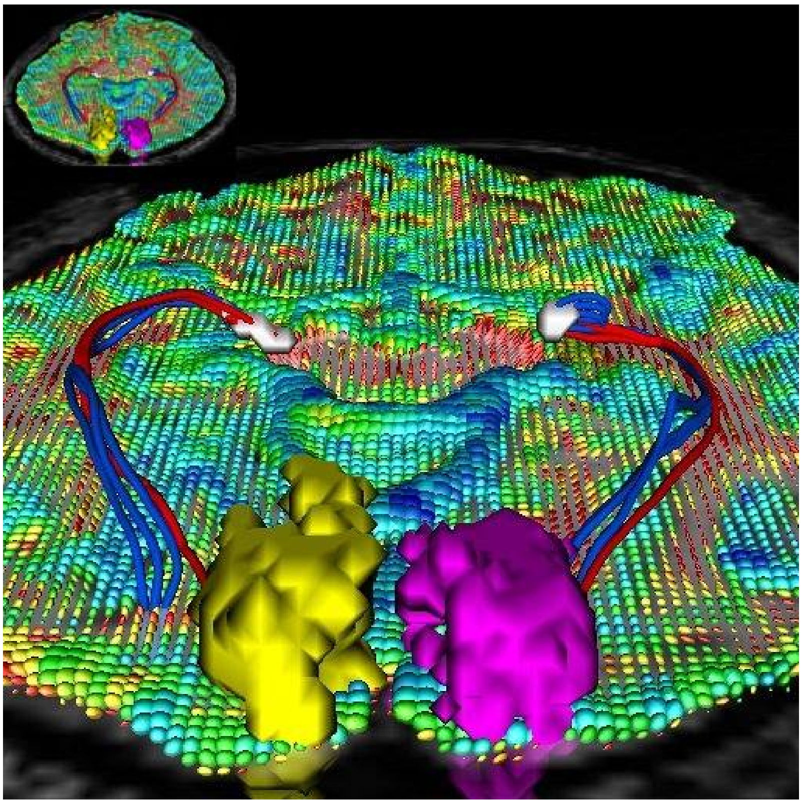


Figure 11.3: Optic radiation tracts estimated with a classical streamline method (blue) and with our GCM technique (red). The LGN seeds voxels (grey), left V1 (yellow), right V1 (purple) and an axial slice of the DTI are also represented. Tracts obtained with the streamline method can fail to reach area V1, as in the left hemisphere here. On the other hand, tracts estimated with our GCM technique systematically reach area V1, but they rapidly converge into a single bundle, illustrating the more global behavior of the method.

erally located to the activated occipital cortex to filter their thalamo-occipital fibers (see [78], figure 3). Besides, other DTI tractography works showing this bundle do not exhibit an actual connection with a accurately defined V1 ROI, letting open the question of the fibers termination location.

On the other hand, GCM estimated tracts systematically reach our white matter V1 region, which illustrates an important advantage of the geometric front propagation method over local approaches. However, GCM fibers tend to converge rapidly after leaving the seed voxels, which denotes the less local characteristic of the method (see discussion).

11.3.2 Callosal Connections

For each hemisphere, the low-level visual areas represent and analyze only one half of the visual field, i.e. their respective contralateral hemifield. Nonetheless, homologue areas of both sides, such as left and right V1, have been shown to be connected, at least for the vertical meridian representations, through the splenium, a portion of the corpus callosum [74]. Following [99], we studied the GCM estimated connectivity maps of our functionally defined areas. We were interested in testing the capability of our GCM method to replicate the broad connection topology that Dougherty *et al.* reported in the region of the splenium. Figure 11.4 shows the resulting mean μ-maps for visual areas hMT+, V1, V3A and V4 respectively taken as starting ROIs in a mid-sagittal section of the brain for each subject. We do not represent here the mean μ-maps for areas V2v, V2d, V3v and V3d, as they do not significantly differ from their clothest neighboring areas on the cortical surface, i.e. V1v, V1d, V4 and V3A respectively (see discussion below).

Connectivity values are ordered similarly for each areas, with a smooth gradient from lowest values in the posterior/dorsal portion of the splenium to highest values in its anterior/ventral portion. Comparing the different origin areas, lowest connectivity values in the splenium were systematically found for hMT+. Regarding areas V1, V3A and V4, values are not consistent enough across subjects to infer a systematic topology in the occipito-callosal connections. However, V3A connectivity is higher than for V4 in 4 out of 6 hemispheres, suggesting a stronger callosal connectivity for dorsal with respect to ventral areas. Finally, we observed a systematic asymmetry between the maps associated to each hemisphere. The highest values were found for putative connections originating from the left hemisphere. Figure 11.5 represents the most probable fibers linking each hMT+ voxel from both hemispheres to the splenium. We employed a similar method to that used to obtain the optic radiations fibers. More specifically, for each hMT+ voxel considered as a seed, we computed the related connectivity index maps. We then identified the splenium voxel with highest connectivity index and constructed the related geodesic. The estimated fiber tracts from the two hemispheres show a great spatial agreement.

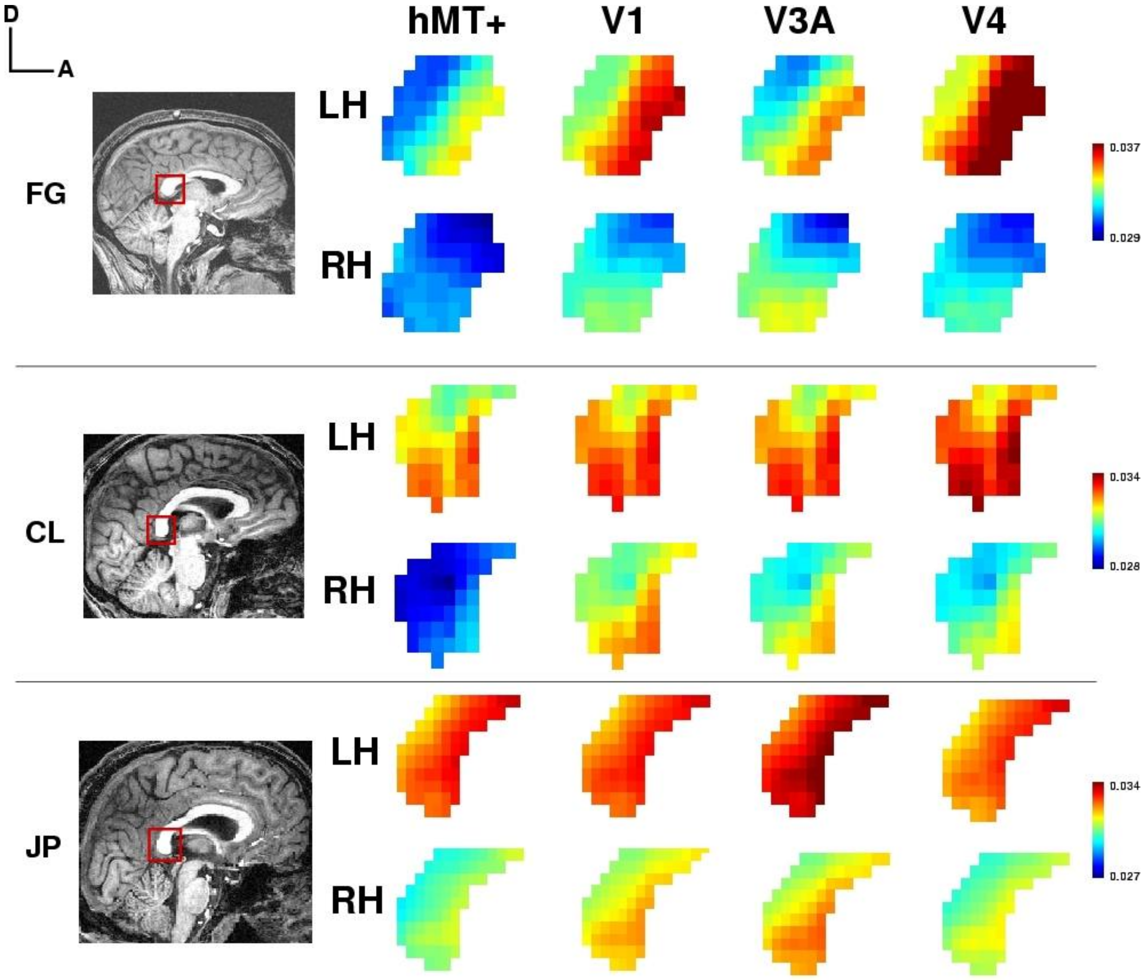


Figure 11.4: Mean connectivity indices from distinct visual areas to the splenium voxels. The mean connectivity values show a smooth gradient from posterior/dorsal to anterior/ventral splenium portions. Lowest connectivity values in the splenium are systematically found for hMT+.

11.3.3 hMT+ Intra-hemispheric Connectivity

Using a similar approach, we finally studied the GCM of the human MT complex with the ipsilateral occipital retinotopic areas. Figure 11.6 shows a box plot of the mean connectivity values distribution for the different seed voxels of hMT+ across retinotopic areas. The boxes edges depict the values of the first quartile, the median and the third quartile. Values outside this box are also shown, to completely represent the distribution dispersion.

V1 and V2 systematically showed the highest connectivity values, suggesting highly probable connections with hMT+. V1 and V2 can hardly be distinguished, which can be attributed to their very close anatomical locations given our voxel size (see discussion). On the other hand, V4 systematically showed the lowest connectivity values, suggesting a weak direct anatomical connection with hMT+. It is more difficult to

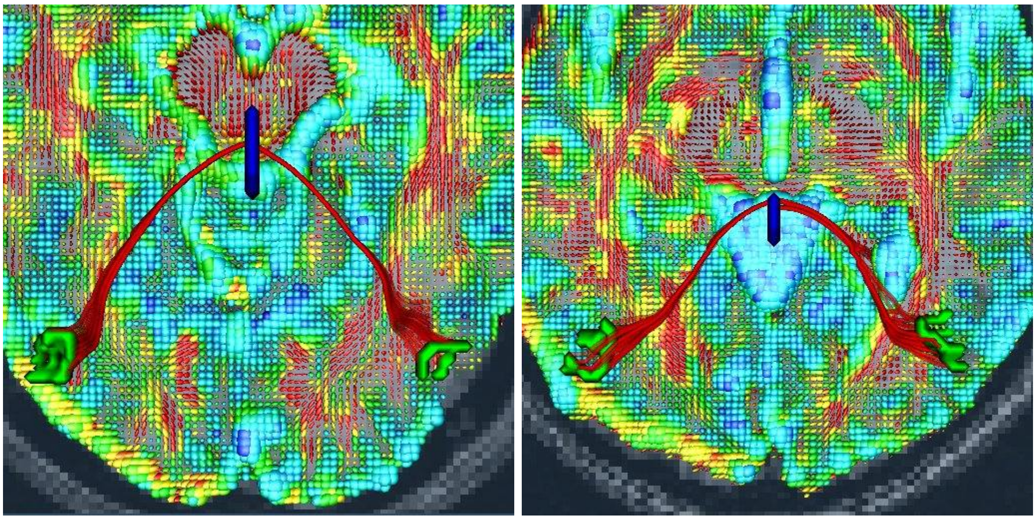


Figure 11.5: Independently estimated most probable fibers linking left and right hMT+ (green) to the splenium (blue) from two subjects (left: CL and right: JP).

clearly distinguish the remaining areas V3v, V3d and V3A.
Similarly to the splenium data analysis, we clearly found higher connectivity values for the left hemisphere as compared to the right, regardless of the area considered (see the values range on the vertical axes).

11.4 DISCUSSION

We could successfully characterize connectivity along projection, callosal and association tracts in the human visual system with our GCM approach. Although our analysis is restricted to 3 subjects, we could successfully reproduce known results about the white matter tissue organization in this region of the brain. This validates our method. We shall first discuss methodological issues regarding DTI based tractography, with a particular emphasis on our GCM technique, before we address the results regarding current knowledge on the human visual brain connectivity.

11.4.1 Methodological Issues

DTI Geodesic Connectivity Mapping: validity and limitations

The current study provides a validation of the GCM approach to estimate DTI based connectivity mapping in the visual system. With its other application to the human motor system in the previous chapter and [179], this Riemannian geometrical approach, using the full tensor information, appears very useful to study anatomical connectivity in various cognitive systems. Geometrical tractography methods, such as the current GCM used in our study, have three main advantages over other tractography approaches. First they provide a connectivity measure between any pair of points within the white matter. This information can be used to build

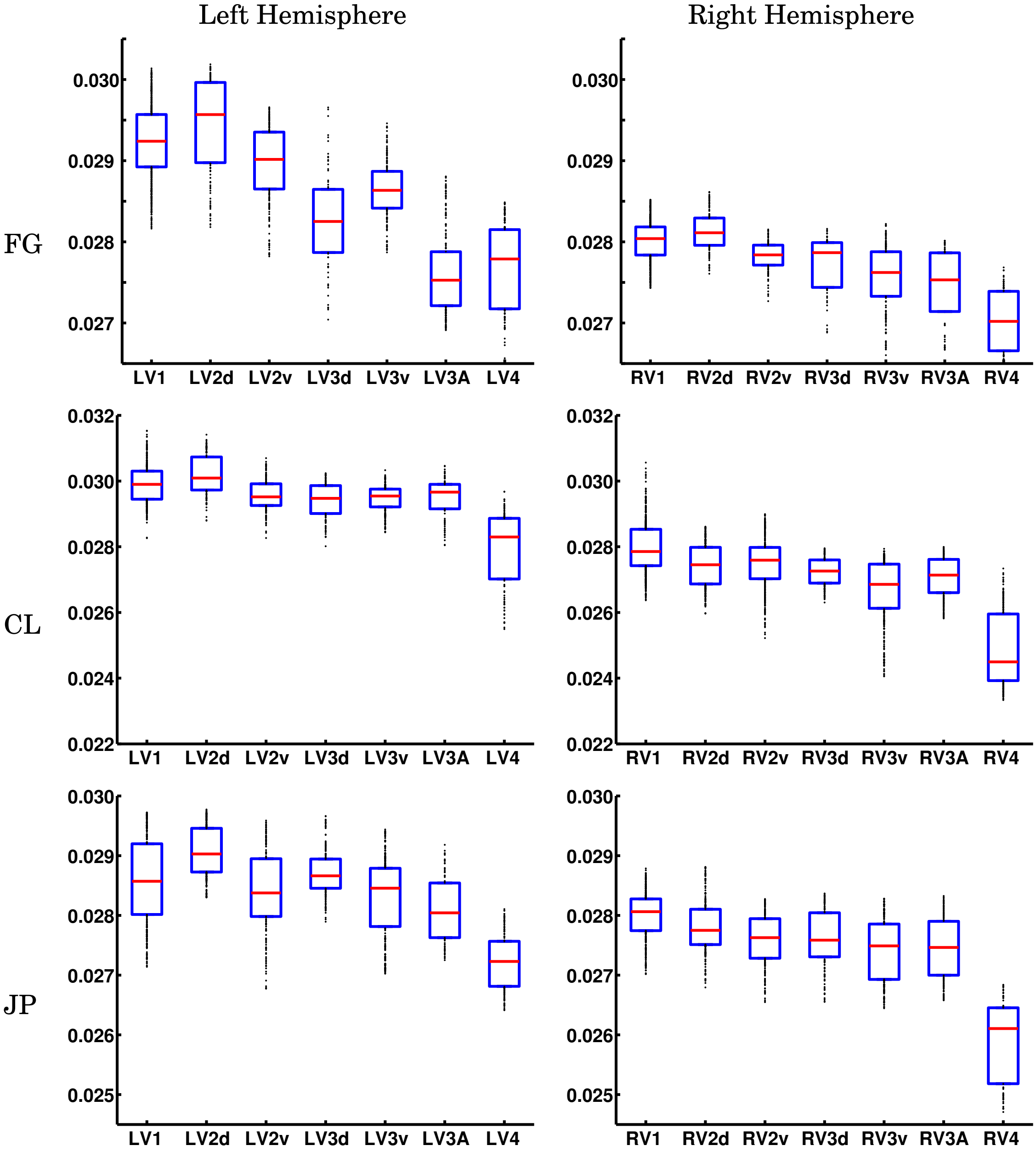


Figure 11.6: Mean connectivity between hMT+ and the retinotopic areas. V1 and V2 systematically shows the highest values while V4 exhibits the lowest values. Note also the lowest values found in the right hemisphere compared to the left in the 3 subjects.

connectivity matrices over the whole brain or to rank putative connections pathways in the white matter. Then, geometrical approaches can deal with locally isotropic tensors occurring at fibers kissing or crossing. This is not the case with deterministic or probabilistic approaches where a FA threshold condition is often necessary to avoid unreliable fibers. Finally, these methods are less sensitive to acquisition noise, since they take advantage of the complete tensor information and of the less local behavior of the algorithm by comparison with streamline or stochastic approaches.

There are however limitations both due to DTI by itself and to the geometrical connectivity mapping framework. First and foremost, the relatively poor spatial resolution of DTI (typically a few mm^3) when compared to actual white matter fibers diameter (between 0.2 and 20 μm) has important implications:

- Only white matter "highways" may be properly recovered, which hardly represent every cortico-cortical connections; false negative connections are thus unavoidable and a precise discrimination between spatially close regions is still difficult to obtain. Our results on the visual system illustrate these spatial limitations. As mentioned above, we could hardly distinguish mean connectivity maps for areas V1 and V2. Although surprising at first sight, this result can actually find a simple explanation when considering together the anatomical layout of these areas and the current spatial resolution of DTI. Areas V1d and V2d (and similarly V1v and V2v) respectively lie on the opposite banks of the same gyrus[4]. The white matter tissue separating the latter is therefore relatively thin, especially with 2mm isotropic voxels. Thus we cannot expect to easily distinguish the connectivity maps obtained with two opposite voxels in this gyrus. Improvement of the spatial resolution appears as the only way to solve this problem. Although still to be considered for the areas couples V3d/V3A dorsally and V3v/V4 ventrally, this gyral proximity is less pronounced since these areas borders appear less constrained by the sulco-gyral pattern than for V1 and V2 borders. Improvements in image acquisition protocols, such as parallel imaging, may overcome this spatial limitation, but a precise physical lower bound is still to be estimated.

- The tensor model cannot handle properly fibers crossings or kissings that may occur within a voxel. Emerging approaches using higher order models based on High Angular Resolution Diffusion Imaging (HARDI) [118, 287, 225, 50, 90, 92] may provide an answer to this issue.

An intrinsic problem of the geometrical connectivity mapping approach used here comes from the absence of absolute threshold to confidently estimate fiber tracts from the connectivity maps [233]. Depending on the threshold choice (the p proportion),

---

[4] Note that Van Essen proposed an interesting mechanical tension-based theory to explain this particular folding pattern [294].

false positive or false negative connections may arise. Combination of complementary connectivity indices associated with each geodesics may prove to minimize this limitation. Furthermore, most tractography methods to date, including ours, are not symmetrical in the sense that putative paths reaching a position y while starting from x may not necessarily coincide with those linking x when starting from y. Besides, tracking within GM, although theoretically possible with geometrical approaches like the one we employed, still leads to difficult interpretations of the reconstructed connectivity maps and related tracts as the diffusion signal is poor in the cortical tissue. Last but not least, a direct validation of DTI based methods is still missing. Although reconstructed tracts such as the optic radiations in the current study or the motor pathway found in [179] are consistent with known anatomy, a quantitative validation could indicate the advantages and weaknesses of DTI based tractography methods. Ultimately, an animal study comparing the different tractography approaches with invasively identified connections would be of great interest to demonstrate their respective advantages and current DTI based tractography limits.

Combined fMRI and DTI

The fMRI areas identification confidently restrained our analysis into known brain regions. We thus avoided possible operator-dependent bias in seed placement or rough anatomically based inference. No obvious false-positive connections were found in our study.

11.4.2 Visual Cortex Connectivity

Thalamo-occipital fibers

We first reproduced tracking of the thalamo-occipital fibers bundle connecting the LGN and V1. This fiber bundle was identified in various DTI tractography works, either with a deterministic streamline [78, 56] or a Fast Marching Tractography [72, 70] method. Although our methodology to identify the LGN seed voxels might appear biased as it is already based on DTI information, we stress that the estimated LGN location and extent consistently fits known anatomy [143] and previous imaging reports [65]. Furthermore, this method is not prone to operator dependent seed selection. The comparison between streamline methods and our GCM technique illustrates the lower local sensitivity of the latter, since spatially close seeds lead to relatively similar connectivity maps, hence to close fibers tracts (figure 11.3). This can be an advantage over classical streamline approaches as it is less prone to noise, but might also obscure local topology across spatially close fibers, such as those shown by Conturo *et al.* in the thalamo-occipital fiber bundle [78].

Splenium fibers

We investigated the topology of callosal fibers with respect to their origin in the low level visual cortex. We could reproduce with our GCM method the antero-ventral localization of fibers linking occipital retinotopic areas to the splenium (figure 11.4), as found by Dougherty *et al.* [99] using a classical streamline approach. Our results also suggest higher connectivity values for V3A when compared to V4, which is consistent with [99]. We could not however identify the precise topological organization of connections within the splenium they observed, neither with our GCM approach nor with a streamline technique similar to the one they used. A lower quality in our diffusion-weighted images may be responsible for this discrepancy.

We found the lowest connectivity values in the splenium for hMT+ as compared to occipital retinotopic areas (figure 11.4). This result should be related to a clinical study demonstrating that visual motion perception, strongly correlated with hMT+ activity, is not affected by posterior callosal destruction [73]. On the other hand, a weaker activation during bilateral visual field stimulation was found in the patient left hemisphere calcarine region compared to 20 normal subjects, correlated with severely impaired reading and color naming performances. These findings suggested other, probably parallel, pathways conveying interhemispheric visual motion information. Possible candidates for the alternative routes include anterior part of the corpus callosum, anterior commissure and subcortical (via the superior colliculus, the intercollicular commissure and the pulvinar) connections. Future work will shortly assess these alternative interhemispheric connections for hMT+.

hMT+ and occipital areas connectivity

We studied the connectivity between hMT+ and various low-level retinotopic areas. To the best of our knowledge, this is the first DTI connectivity study considering this cornerstone of the visual motion pathway. V1 systematically showed the highest connectivity index values with hMT+ (figure 11.6), consistent with the known highly myelinated white matter fiber bundles linking both areas [296]. Besides, lowest connectivity values between the retinotopic areas and hMT+ were systematically found for area V4, further supporting the famous distinction between ventral and dorsal streams [293, 207].

We also found similar hMT+ connectivity values for V3d and V3v, despite their relatively important distance along the cortical sheet. This observation could be an other evidence to consider V3d and V3v as the two quarterfields representation of a single area V3, as also demonstrated with anatomical connectivity studies in various species of monkeys [196, 197].

Hemisphere asymmetry

Our results in the visual system suggested a significant asymmetry in our connectivity maps between the two hemispheres. The left hemisphere exhibits higher connectivity values than its right counterpart. A similar result was also reported in [99], where more occipito-callosal fibers could be reconstructed in the left than in the right hemisphere by the employed streamline algorithm. The authors suggested a bias in the hemispheres respective size may account for this difference. Note that such an asymmetry between both hemispheres was also reported in the motor system [133] and could possibly be attributed to handedness. Similarly, [125] reported a greater asymmetry in the left than in the right hemisphere in most parts of the cingulum, but no significant correlation with handedness could be demonstrated [124]. We suggest an alternative hypothesis, based on a perhaps more straightforward brain observation: hemispheric functional specialization. Undoubtedly, the hemispheres are functionally asymmetric and this should imply a different, asymmetrical wiring within each hemisphere. This may in particular be the case in the occipital cortex, the right lobe possibly presenting more fiber crossings than its left counterpart. As a consequence, the local diffusion tensors would not be equivalently anisotropic in both sides, leading to more difficult fiber tracking for streamline methods or lower connectivity values for our GCM algorithm.

11.5 CONCLUSION

We evaluated the ability of our new geometrical DTI analysis framework, GCM, to infer anatomical connectivity in the visual cortex. We could successfully reconstruct the well-known optic radiations connecting the LGN and V1 with our fast connectivity mapping method. We also showed a plausible topology of occipito-callosal connections in the splenium, consistent with previous works. Finally, we assessed the anatomical connectivity between hMT+ and occipital retinotopic areas, supporting the view of parallel ventral and dorsal processing streams. With both image acquisition and methodological improvements, diffusion MRI should provide a new means to uncover the architecture of the visual system and further relate it to its functional characterization.

CHAPTER 12

CONCLUSION

In this thesis, we have proposed new methods for diffusion tensor MR images processing (Part II). This required some preliminary theoretical work in order to define and understand the proper mathematical concepts that would constitute the core tools of our work (Part I & Appendix A). Finally, we could successfully apply these results to investigate the motor and visual anatomo-functional networks of the human brain (Part III). Our contributions are thus threefold: theoretical, methodological and applied. All along this thesis, we tried to make the right mathematical choices to model the problems of interest. We believe this enabled us to propose adequate and efficient algorithms such that we could finally tackle challenging neuroscience questions. To summarize,

- We proposed an original information geometric point of view to work with diffusion tensors. The introduction of a Riemannian structure on the “space of diffusion tensors” yielded intrinsic numerical schemes that naturally preserve the properties of these mathematical objects. We showed, for instance, how to draw random tensors following an imposed Gaussian distribution or how to interpolate tensor-valued images.

- We presented new algorithms for the estimation and regularization of diffusion tensor images which naturally enforce the tensors properties.

- We introduced front propagation techniques that use the full diffusion tensor to estimate the degree of connectivity of a region of interest with the rest of the brain and to approximate neural fibers as geodesics.

- We demonstrated that it was possible to use our intrinsic statistics on diffusion tensors within a surface evolution framework to perform the segmentation of DTI. The choice of the dissimilarity measure was shown to be crucial since only the Riemannian metric derived from the Fisher information matrix and used in chapter 5 recovered the expected results on particularly difficult datasets.

- We proposed a novel approach to the non-rigid registration of DTI. Although essentially theoretical, our work is the first to resort to intrinsic statistics on the manifold S^+ and simple numerical examples proved the feasibility of the approach.

- Finally, we applied our connectivity mapping techniques to the investigation of the human motor and visual systems. We reproduced previous findings but, most importantly, we proposed new insights into the anatomo-functional organization of the human brain.

Although our contributions can undoubtedly still be improved, we believe they can benefit to a wide range of clinical and neuroscience applications. The proper estimation and regularization of DTI datasets can have an important impact on the statistical analysis of large DTI databases. Our connectivity mapping techniques, since they quantify the degree of connectivity between region of interest, can not only be of great interest to better understand the wiring of the cerebral white matter but also to study neurological pathologies like multiple sclerosis, Alzheimer's or Parkinson's disease, schizophrenia ...etc. Christine Delmaire and Stéphane Lehéricy (La Pitié-Salpêtrière Hospital, Paris, France) are currently studying whether our tools can be used in the diagnosis of dystonia. By combining our registration and segmentation methods, it could also be interesting to investigate the variability of the shape of fiber bundles. Building atlases of those structures would be of great help to simply incorporate prior knowledge in our segmentation algorithm or to help detecting abnormal three-dimensional configurations that could indicate ischemia, MS plaques, tumors ...etc

We are also currently investigating the possibility to extend the various methods presented in this manuscript to higher order models. It is now indeed well-known that the diffusion tensor model is only good when there exists one major fiber orientation within a voxel, which is rarely the case. High Angular Resolution Diffusion Imaging (HARDI) [118, 287, 225, 50, 90] is more and more frequently used to compute the so-called Orientation Distribution Functions (ODFs). Those spherical functions may provide enough information to resolve fibers crossings, branching, kissing ...etc. It has also been shown that, by working with spherical harmonics, they can be mapped onto higher order symmetric and positive definite diffusion tensors [91]. It should thus be possible to generalize at least some of the algorithms proposed in this thesis to tensor representations of ODFs.

CHAPTER 13

CONCLUSION (FRANÇAIS)

Dans cette thèse, nous avons proposé de nouvelles méthodes pour le traitement des IRM du tenseur de diffusion, ou Imagerie du Tenseur de Diffusion (ITD), (Partie II). Une étude théorique préliminaire a été nécessaire pour mieux comprendre et définir les notions mathématiques adéquates qui constituent les outils centraux de notre travail (Partie I & Appendice A). Enfin, nous avons appliqué ces résultats avec succès à l'investigation des réseaux anatomo-fonctionnels moteurs et visuels du cerveau humain (Partie III). Nos contributions sont donc à la fois théoriques, méthodologiques et appliquées. Tout au long de cette thèse, nous nous sommes efforcés de faire des choix mathématiques judicieux afin de modéliser le mieux possible les problèmes d'intérêt. Nous sommes convaincus que cela nous a permis de proposer des algorithmes pertinents et efficaces pour tenter de répondre à des questions fondamentale en neurosciences. En résumé,

- Nous avons adopté un point de vue géométrique original pour travailler avec les entités complexes que sont les tenseurs de diffusion. L'introduction d'une structure Riemannienne sur "l'espace des tenseurs de diffusion" nous a permis de proposer des schémas numériques préservant les propriétés de ces objets mathématiques. Nous avons notamment montré comment générer des tenseurs aléatoires, distribués selon une loi Gaussienne imposée, ou comment interpoler des images à valeur matricielle.

- Nous avons présenté de nouveaux algorithmes pour l'estimation et la régularisation d'ITD. Ceux-ci préservent naturellement les propriétés des tenseurs.

- Nous avons introduit des techniques de propagation de fronts, faisant usage de toute l'information contenue par le tenseur de diffusion, afin d'estimer le degré de connectivité d'une région d'intérêt avec le reste du cerveau et d'approximer les fibres nerveuses par des géodésiques.

- Nous avons montré comment utiliser les statistiques intrinsèques sur les tenseurs de diffusion, dans un algorithme d'évolution de surface, pour réaliser la segmentation d'ITD. Le choix de la mesure de dissimilarité entre tenseurs s'est révélé être crucial. Seule la métrique Riemannienne, dérivée de la matrice d'information de Fisher et utilisée dans le chapitre 5 a été en mesure de trouver les résultats attendus sur des données particulièrement difficiles.

- Nous avons proposé une approche originale pour le recalage non-rigide d'ITD. Bien qu'essentiellement théorique, notre travail est le premier à exploiter les statistiques intrinsèques sur la variété S^+ et des exemples numériques simples ont prouvé la faisabilité de l'approche.

- Finalement, nous avons appliqué nos techniques d'analyse de la connectivité anatomique à l'investigation des systèmes moteurs et visuels de l'homme. Nous avons pu reproduire certains résultats de la littérature et, surtout, avons présenté de nouveaux éléments de réponse sur l'organisation anatomo-fonctionnelle du cerveau humain.

Bien que nos contributions puissent encore sans aucun doute être améliorées, nous sommes convaincus qu'elles peuvent s'avérer très utiles en neuroscience et pour un grand nombre d'applications cliniques. L'estimation et la régularisation des IRM du tenseur de diffusion, avec des méthodes appropriées, ont très certainement un impact important sur l'analyse statistique d'un grand nombre de ces images. Nos techniques d'analyse de la connectivité, puisqu'elles permettent de quantifier cette connectivité entre régions d'intérêt, peuvent non seulement être utiles pour mieux comprendre l'architecture des réseaux de la matière blanche cérébrale mais aussi pour étudier les pathologies neurologiques telles que la sclérose en plaques, la maladie d'Alzheimer ou la maladie de Parkinson, la schizophrénie ...etc. Christine Delmaire et Stéphane Lehéricy (Hôpital La Pitié-Salpêtrière, Paris, France) utilisent actuellement nos outils afin de mieux comprendre les origines d'une autre pathologie : la dystonie. En combinant nos méthodes de recalage et de segmentation, il pourrait également être intéressant d'étudier la variabilité de la forme des faisceaux de fibres. La construction d'atlas pour ces structures serait une aide précieuse pour incorporer des connaissances a priori dans notre algorithme de segmentation ou pour aider à la détection de configurations tridimensionnelles anormales, indiquant la présence possible d'un accident vasculaire cérébral, d'une tumeur ...etc.

Nous travaillons actuellement sur l'extension des techniques exposées dans ce manuscrit à des modèles d'ordres supérieurs. Le modèle du tenseur de diffusion (ordre 2) n'est effet valide que lorsque les fibres présentes dans un voxel sont principalement alignées selon une unique direction, ce qui est rarement le cas. L'IRM de diffusion à haute résolution angulaire [118, 287, 225, 50, 90] est une technique récente permettant d'estimer des fonctions de distribution angulaire (FDAs). Ces fonctions sphériques fournissent beaucoup plus d'information et constituent la piste à suivre pour distinguer les divers types de croisement, convergence, divergence, contact ...etc. réalisés par les fibres nerveuses. Il a aussi été montré, en utilisant les propriétés des harmoniques sphériques, que ces fonctions peuvent être transformées en tenseurs de diffusion (symétriques et définis positifs) d'ordre supérieur [91]. Il devrait donc être possible de généraliser les algorithmes (ou au moins certains d'entre eux) proposés dans cette thèse à ces représentations tensorielles des fonctions de distribution angulaire.

Appendices

APPENDIX **A**

DIRECTIONAL DERIVATIVES OF MATRIX FUNCTIONS

A.1 DERIVATIVES & SPECTRAL REPRESENTATION

As developed in [84], a matrix function f can be defined in terms of Cauchy integral formula or by employing a similarity transformation of the form:

Proposition A.1.0.1. *Let* $t \to \mathbf{X}(t)$ *be a smooth function from a neighborhood of* t_0 *to* $M_n(\mathbb{R})$*. Let* f *be a function which domain is* $\mathcal{D}$*. If* $\mathbf{X}(t)$ *is diagonalizable in a neighborhood* $\mathcal{V}$ *of* t_0 *with eigenvalues in the domain* $\mathcal{D}$*, then in this neighborhood* $f(\mathbf{X}(t))$ *is defined as:*

$$f(\mathbf{X}(t)) = \mathbf{Z}(t) f(\mathbf{D}(t)) \mathbf{Z}^{-1}(t) \ ,$$

where $\mathbf{Z}(t)$ *and* $\mathbf{D}(t)$ *are defined in* $\mathcal{V}$ *as one solution of the diagonalization equation:*

$$\mathbf{X}(t) = \mathbf{Z}(t)\mathbf{D}(t)\mathbf{Z}^{-1}(t) \ , \tag{A.1}$$

where $\mathbf{D}(t)$ *is a diagonal matrix and* $f(\mathbf{D}(t))$ *is the diagonal matrix obtained by applying the function* f *to the diagonal coefficients of* $\mathbf{D}(t)$*.*

When involved in optimization procedure, it is necessary to obtain the derivatives of these functions of matrices. This is the purpose of the present appendix. We will prove the following proposition and evaluate its numerical properties.

Proposition A.1.0.2. *If furthermore all eigenvalues of* $\mathbf{X}(t)$ *are distinct in a neighborhood* $\mathcal{V}' \subset \mathcal{V}$ *(which makes the decomposition A.1 unique up to a permutation) then its derivative* $\frac{df(\mathbf{X}(t))}{dt}$ *exists in* $\mathcal{V}'$ *and is a function of* $\mathbf{X}(t)$ *and* $\frac{d\mathbf{X}}{dt}$*, noted* $\mathrm{df}\,(,)$

$$\frac{df(\mathbf{X}(t))}{dt} = \mathrm{df}\left(\mathbf{X}(t), \frac{d\mathbf{X}}{dt}\right) \tag{A.2}$$

An expression for $\mathrm{df}\,(,)$ *can be obtained from the diagonalization equation A.1 as:*

$$\frac{df(\mathbf{X})}{dt} = \frac{d\mathbf{Z}}{dt}\mathbf{Z}^{-1} f(\mathbf{X}) - f(\mathbf{X})\frac{d\mathbf{Z}}{dt}\mathbf{Z}^{-1} + \mathbf{Z}\frac{d\mathbf{D}}{dt}\mathbf{Z}^{-1}\mathbf{X}^{-1} \ ,$$

where time dependence of functions has been dropped for readability.

The derivative $\frac{d\mathbf{D}}{dt}$ *can be obtained as:*

$$\frac{d\mathbf{D}}{dt} = diag\left(\mathbf{Z}^{-1}\frac{dX}{dt}\mathbf{Z}\right) \ ,$$

where $diag(\mathbf{M})$ *means the matrix obtained by keeping only the diagonal elements of the matrix* $\mathbf{M}$.

The term $\frac{d\mathbf{Z}}{dt}\mathbf{Z}^{-1}$ *is equal to* $\mathbf{ZQZ}^{-1}$ *where*

$$\mathbf{Q}_j^i = \begin{cases} \left(\mathbf{Z}^{-1}\frac{d\mathbf{X}}{dt}\mathbf{Z} - \frac{d\mathbf{D}}{dt}\right)_j^i / (\lambda_j - \lambda_i) & \text{if } i \neq j \\ 0 & \textit{otherwise} \end{cases}$$

and where we denote by $\lambda_i, \ i = 1, ..., n$ *the elements of the diagonal matrix* $\mathbf{D}$.

The first part of the proposition concerning the existence of the derivative is a consequence of corollary 9.2.2.2. The proof of the other assertions occupies the next three sections.

A.1.1 Spectral representation of the derivative

We recall that λ_i are the eigenvalues of the matrix $\mathbf{X}$ and note $\mathbf{Z}_i$ the associated eigenvectors, e.g. the columns of $\mathbf{Z}$. From proposition A.1.0.1, we have $f(\mathbf{X}) = \mathbf{Z}f(\mathbf{D})\mathbf{Z}^{-1}$. Hence,

$$\frac{d}{dt}f(\mathbf{X}) = \frac{d\mathbf{Z}}{dt}f(\mathbf{D})\mathbf{Z}^{-1} + \mathbf{Z}\frac{df(\mathbf{D})}{dt}\mathbf{Z}^{-1} + \mathbf{Z}f(\mathbf{D})\frac{d\mathbf{Z}^{-1}}{dt} \ .$$

$\frac{d\mathbf{Z}^{-1}}{dt}$ is easily computed as follows:

$$\mathbf{ZZ}^{-1} = \mathbf{I} \Rightarrow \frac{d\mathbf{Z}}{dt}\mathbf{Z}^{-1} + \mathbf{Z}\frac{d\mathbf{Z}^{-1}}{dt} = 0 \Rightarrow \frac{d\mathbf{Z}^{-1}}{dt} = -\mathbf{Z}^{-1}\frac{d\mathbf{Z}}{dt}\mathbf{Z}^{-1} \ .$$

We can now go back to the expression of $\frac{df(\mathbf{X})}{dt}$ and obtain:

$$\begin{aligned} \frac{df(\mathbf{X})}{dt} &= \frac{d\mathbf{Z}}{dt}\mathbf{Z}^{-1}\mathbf{Z}f(\mathbf{D})\mathbf{Z}^{-1} + \mathbf{Z}f(\mathbf{D})\mathbf{Z}^{-1}\mathbf{Z}\frac{d\mathbf{Z}^{-1}}{dt} + \mathbf{Z}\frac{df(\mathbf{D})}{dt}\mathbf{Z}^{-1} \\ &= \frac{d\mathbf{Z}}{dt}\mathbf{Z}^{-1}f(\mathbf{X}) - f(\mathbf{X})\frac{d\mathbf{Z}}{dt}\mathbf{Z}^{-1} + \mathbf{Z}\frac{df}{dt}(\mathbf{D})\frac{d\mathbf{D}}{dt}\mathbf{Z}^{-1} \\ &= \left[\frac{d\mathbf{Z}}{dt}\mathbf{Z}^{-1}, f(\mathbf{X})\right] + \mathbf{Z}\frac{df}{dt}(\mathbf{D})\frac{d\mathbf{D}}{dt}\mathbf{Z}^{-1} \ , \end{aligned} \tag{A.3}$$

where $[.,.]$ denotes the commutator and $\frac{df}{dt}(\mathbf{D})$ can be easily computed using proposition A.1.0.1 for the functional $\frac{df}{dt}$.

Thus, in the previous expression, we only have to evaluate $\frac{d\mathbf{D}}{dt}$ and $\frac{d\mathbf{Z}}{dt}\mathbf{Z}^{-1}$. As we will see in the next two sections, we actually only need to compute $\frac{d\mathbf{X}}{dt}$.

A.1.2 Computation of $\frac{d\mathbf{D}}{dt}$

Computing $\frac{d\mathbf{D}}{dt}$ boils down to computing the derivatives of the eigenvalues λ_i, $i = 1, ..., n$. As shown in the next paragraph (for the case $n = 3$, although the calculations could be generalized to any values of n), this can be obtained rather easily from the characteristic polynomial of $\mathbf{X}$. However, the method proposed just below is simpler, faster and much more general:

From equation A.3 applied to the identity function, we have:

$$\begin{aligned}
\mathbf{Z}^{-1}\frac{d\mathbf{X}}{dt}\mathbf{Z} &= \mathbf{Z}^{-1}\frac{d\mathbf{Z}}{dt}\mathbf{Z}^{-1}\mathbf{X}\mathbf{Z} - \mathbf{Z}^{-1}\mathbf{X}\frac{d\mathbf{Z}}{dt} + \frac{d\mathbf{D}}{dt} \\
&= \mathbf{Z}^{-1}\frac{d\mathbf{Z}}{dt}\mathbf{D} - \mathbf{D}\mathbf{Z}^{-1}\frac{d\mathbf{Z}}{dt} + \frac{d\mathbf{D}}{dt} \\
&= \left[\mathbf{Z}^{-1}\frac{d\mathbf{Z}}{dt}, \mathbf{D}\right] + \frac{d\mathbf{D}}{dt} = [\mathbf{Q}, \mathbf{D}] + \frac{d\mathbf{D}}{dt} \ ,
\end{aligned}$$

with $\mathbf{Q} = \mathbf{Z}^{-1}\frac{d\mathbf{Z}}{dt}$. Because $\mathbf{D}$ is a diagonal matrix, the $(i, j)^{th}$ element of $[\mathbf{Q}, \mathbf{D}]$ is:

$$Q^i_j \left(\lambda_j - \lambda_i\right) \ ,$$

which vanishes when $i = j$. Thus, the diagonal elements of the commutator $[\mathbf{Q}, \mathbf{D}]$ are zero and we have

$$\frac{d\mathbf{D}}{dt} = diag\left(\mathbf{Z}^{-1}\frac{d\mathbf{X}}{dt}\mathbf{Z}\right).$$

Alternative computation of $\frac{d\mathbf{D}}{dt}$ for $\mathrm{n} = 3$:

The derivative $\frac{d\mathbf{D}}{dt}$ can easily be obtained from the following expression of the characteristic polynomial of $\mathbf{X}$, when $n = 3$:

$$\mathcal{P}(\lambda, t) = -\lambda^3 + \mathbf{tr}(\mathbf{X}(t))\lambda^2 - \left(\sum_{i=1}^{3} \mathcal{M}^i_i(\mathbf{X}(t))\right)\lambda + |\mathbf{X}(t)| = 0$$

where $\mathcal{M}^i_i\left(\mathbf{X}(t)\right)$ are the second-order principal minors of $\mathbf{X}(t)$ for $i = 1, 2, 3$ and $|\mathbf{X}(t)|$ denotes its determinant.

$\mathcal{P}$ being a function of λ and t, its total derivative with respect to t is:

$$\begin{aligned}
\frac{d\mathcal{P}(\lambda, t)}{dt} &= \frac{\partial\mathcal{P}(\lambda, t)}{\partial\lambda}\frac{d\lambda(t)}{dt} + \frac{\partial\mathcal{P}(\lambda, t)}{\partial t} = 0 \\
\Rightarrow \frac{d\lambda(t)}{dt} &= -\frac{\partial\mathcal{P}(\lambda, t)}{\partial t} / \frac{\partial\mathcal{P}(\lambda, t)}{\partial\lambda}
\end{aligned} \tag{A.4}$$

with

$$\begin{aligned}
\frac{\partial\mathcal{P}(\lambda, t)}{\partial t} &= \frac{d\mathbf{tr}(\mathbf{X})}{dt}\lambda^2 - \left(\sum_{i=1}^{3}\frac{d\mathcal{M}^i_i(\mathbf{X})}{dt}\right)\lambda + \frac{d|\mathbf{X}|}{dt} \\
&= \mathbf{tr}\left(\frac{d\mathbf{X}}{dt}\right)\lambda^2 + \left(\mathbf{tr}\left(\mathbf{X}\frac{d\mathbf{X}}{dt}\right) - \mathbf{tr}(\mathbf{X})\mathbf{tr}\left(\frac{d\mathbf{X}}{dt}\right)\right)\lambda + \mathbf{tr}\left(\mathbf{X}^\star\frac{d\mathbf{X}}{dt}\right)
\end{aligned} \tag{A.5}$$

and

$$\frac{d\mathcal{P}(\lambda,t)}{d\lambda} \quad = \quad -3\lambda^2 + 2\mathrm{tr}(\mathbf{X})\lambda - \sum_{i=1}^{3}\mathcal{M}_i^i(\mathbf{X}) \tag{A.6}$$

Let us detail the expressions for the derivative of the determinant (Jacobi's formula) and the derivative of the minors.

Denoting by $(\mathbf{X}^\star)_j^i = C_j^i = (-1)^{i+j}\mathcal{M}_j^i$ the cofactors of $\mathbf{X}$, the determinant is expressed as $|\mathbf{X}| = \sum_{k=1}^n C_k^i \mathbf{X}_i^k$. Hence,

$$\begin{aligned}
\frac{\partial|\mathbf{X}|}{\partial \mathbf{X}_j^i} &= C_i^j \\
\Rightarrow \frac{d|\mathbf{X}|}{dt} &= \sum_{i,j=1}^{n}\frac{\partial|\mathbf{X}|}{\partial \mathbf{X}_j^i}dX_j^i = \sum_{i,j=1}^{n} C_i^j dX_j^i = \langle \mathbf{X}^\star, \frac{d\mathbf{X}}{dt}\rangle = \mathrm{tr}\left(\mathbf{X}^\star\frac{d\mathbf{X}}{dt}\right)
\end{aligned}$$

where $\langle .,.\rangle$ denotes the canonical inner product.

The derivative of $\sum_{i=1}^3 \mathcal{M}_i^i(\mathbf{X})$ can be computed at once from the trace of the adjugate matrix derivative. The proof is given for non-singular $\mathbf{X}$ but the result is true even without this condition (it can be easily deduced by a continuity argument). Writing $\mathbf{X}^\star = |\mathbf{X}|\mathbf{X}^{-1}$, the characteristic polynomial associated to $\mathbf{X}$ can be written as (remember that $\mathbf{X}$ is a 3×3 matrix):

$$\lambda^3 - \mathrm{tr}(\mathbf{X})\lambda^2 + \mathrm{tr}(\mathbf{X}^\star)\lambda - |\mathbf{X}| = 0\ .$$

since $\mathrm{tr}(\mathbf{X}^\star) = \sum_{i=1}^3 \mathcal{M}_i^i(\mathbf{X})$.

Using the Cayley-Hamilton theorem, we have:

$$\mathbf{X}^3 - \mathrm{tr}(\mathbf{X})\mathbf{X}^2 + \mathrm{tr}(\mathbf{X}^\star)\mathbf{X} - |\mathbf{X}|I = 0$$

Multiplying this expression by $\mathbf{X}^\star$ and using the identity $\mathbf{X}\mathbf{X}^\star = |\mathbf{X}|I$ gives:

$$|\mathbf{X}|\left(\mathbf{X}^2 - \mathrm{tr}(\mathbf{X})\mathbf{X} + \mathrm{tr}(\mathbf{X}^\star)I - \mathbf{X}^\star\right) = 0\ .$$

Simplifying by the determinant $|\mathbf{X}|$ (which is assumed to be non-zero), and taking the trace of the expression yields:

$$\mathrm{tr}(\mathbf{X}^2) - \mathrm{tr}(\mathbf{X})^2 + 2\mathrm{tr}(\mathbf{X}^\star) = 0\ .$$

Differentiating, this expression with respect to time t and using the fact that $\mathrm{tr}(\mathbf{X}\frac{d|\mathbf{X}|}{dt}) = \mathrm{tr}(\frac{d|\mathbf{X}|}{dt}\mathbf{X})$ gives:

$$2\mathrm{tr}\left(\mathbf{X}\frac{d\mathbf{X}}{dt}\right) - 2\mathrm{tr}(\mathbf{X})\mathrm{tr}\left(\frac{d\mathbf{X}}{dt}\right) + 2\mathrm{tr}\left(\frac{d\mathbf{X}^\star}{dt}\right) = 0\ .$$

Finally:

$$\frac{d\sum_{i=1}^3\mathcal{M}_i^i(\mathbf{X})}{dt} = \mathrm{tr}\left(\frac{d\mathbf{X}^\star}{dt}\right) = \mathrm{tr}\left(\mathbf{X}\right)\mathrm{tr}\left(\frac{d\mathbf{X}}{dt}\right) - \mathrm{tr}\left(\mathbf{X}\frac{d\mathbf{X}}{dt}\right).$$

A.1.3 Computation of $\frac{d\mathbf{Z}}{dt}\mathbf{Z}^{-1}$

Applying again equation A.3 to the identity function, it follows that:

$$\frac{d\mathbf{X}}{dt} - \mathbf{Z}\frac{d\mathbf{D}}{dt}\mathbf{Z}^{-1} = \left[\frac{d\mathbf{Z}}{dt}\mathbf{Z}^{-1}, \mathbf{X}\right] . \tag{A.7}$$

From the previous section, we know how to compute $\frac{d\mathbf{D}}{dt}$. Then, setting $\mathbf{P} = \frac{d\mathbf{Z}}{dt}\mathbf{Z}^{-1}$, equation A.7 should easily be solved for $\mathbf{P}$ by seeking the solution of the linear system:

$$\mathbf{WY} = \mathbf{U} ,$$

with $\mathbf{U} = \varphi_n\left(\frac{d\mathbf{X}}{dt} - \mathbf{Z}\frac{d\mathbf{D}}{dt}\mathbf{Z}^{-1}\right)$ and $\mathbf{Y} = \varphi_n(\mathbf{P})$ where φ_n denotes the map associating to each $n \times n$ matrix $\mathbf{M} = [\mathbf{M}_1, ..., \mathbf{M}_n]$ the vector $\varphi_n(\mathbf{M}) = [\mathbf{M}_1^T, ..., \mathbf{M}_n^T]^T \in \mathbb{R}^{n^2}$.

The components of the $n^2 \times n^2$ matrix $\mathbf{W}$ can be computed as follows. Considering the commutator $\mathbf{C} = [\mathbf{P}, \mathbf{X}]$, each of its elements writes:

$$C^i_j = \sum_{k=1}^{n} P^i_k X^k_j - X^i_k P^k_j \tag{A.8}$$

This yields:

$$W^\kappa_\nu = \frac{\partial C^i_j}{\partial P^k_l} = X^l_j \delta(i,k) - X^i_k \delta(l,j) ,$$

where $i, j, k, l, = 1, ..., n$, $\kappa = n(i-1) + j$, $\nu = n(k-1) + l$ and $\delta(i_1, i_2)$ denotes the Kronecker symbol $\delta(i_1, i_2) = 1$ if $i_1 = i_2$ and $\delta(i_1, i_2) = 0$ otherwise.

However, it turns out that the matrix $\mathbf{W}$ is of rank $n^2 - n = n(n-1)$ and thus singular. This situation actually arises from the structure of the problem and it is possible to take advantage of that to derive a better method for the computation of $\mathbf{P}$.

Setting $\mathbf{F} = f(\mathbf{X})$ and rewriting the commutator in the right hand side of equation A.3 as:

$$\begin{aligned}\frac{d\mathbf{Z}}{dt}\mathbf{Z}^{-1}f(\mathbf{X}) - f(\mathbf{X})\frac{d\mathbf{Z}}{dt}\mathbf{Z}^{-1} &= \mathbf{ZQZ}^{-1}\mathbf{F} - \mathbf{FZQZ}^{-1} \\ &= \mathbf{Z}\left(\mathbf{QZ}^{-1}\mathbf{FZ} - \mathbf{Z}^{-1}\mathbf{FZQ}\right)\mathbf{Z}^{-1} \\ &= \mathbf{Z}[\mathbf{Q}, \mathbf{Z}^{-1}\mathbf{FZ}]\mathbf{Z}^{-1}\end{aligned}$$

we can see that the logarithm derivative does not depend on the diagonal elements of $\mathbf{Q} = \mathbf{Z}^{-1}\frac{d\mathbf{Z}}{dt}$. Indeed, we have:

$$[\mathbf{Q}, \mathbf{Z}^{-1}\mathbf{FZ}] = [\mathbf{Q}, \mathbf{Z}^{-1}f(\mathbf{X})\mathbf{Z}] = [\mathbf{Q}, f(\mathbf{Z}^{-1}\mathbf{XZ})] = [\mathbf{Q}, f(\mathbf{D})]$$

But the $(i,j)^{th}$ element of $[\mathbf{Q}, f(\mathbf{D})]$ is:

$$Q^i_j\left(f(\lambda_j) - f(\lambda_i)\right) ,$$

which again vanishes when $i = j$ for all Q^i_j.

Equation A.7 is thus rewritten as follows to only solve for the off-diagonal components of $\mathbf{Q}$ instead of directly seeking $\mathbf{P}$:

$$\mathbf{Z}^{-1}\frac{d\mathbf{X}}{dt}\mathbf{Z} - \frac{d\mathbf{D}}{dt} = \mathbf{Z}^{-1}\frac{d\mathbf{Z}}{dt}\mathbf{D} - \mathbf{D}\mathbf{Z}^{-1}\frac{d\mathbf{Z}}{dt} = [\mathbf{Q}, \mathbf{D}]$$

Q^i_j is then straightforward to obtain as:

$$Q^i_j = \begin{cases} \left(\mathbf{Z}^{-1}\frac{d\mathbf{X}}{dt}\mathbf{Z} - \frac{d\mathbf{D}}{dt}\right)^i_j / \left(\lambda_j - \lambda_i\right) & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases} \tag{A.9}$$

$\frac{d\mathbf{Z}}{dt}\mathbf{Z}^{-1}$ is in turn obtained as $\mathbf{Z}\mathbf{Q}\mathbf{Z}^{-1}$ and this completes the proof of proposition A.1.0.2.

A.2 NUMERICAL EXPERIMENTS: THE log FUNCTION

We recall that the principal logarithm $\log\left(\mathbf{X}(t)\right)$ can be defined by the series:

$$\log\left(\mathbf{X}(t)\right) = -\sum_{k=1}^{\infty} \frac{\left(\mathbf{I} - \mathbf{X}(t)\right)^k}{k}$$

whenever $\|\mathbf{I} - \mathbf{X}(t)\| < 1$ for any norm $\|.\|$. We could use a truncated version of this series expansion to approximate the derivative of the logarithm, however the hypothesis $\|\mathbf{I} - \mathbf{X}(t)\| < 1$ does not hold in general and, in particular, for the cases of interest in chapter 9. Our procedure is much more general.

As pointed out in [211] and detailed in [95], the derivative of the matrix logarithm can be defined through the following integral:

$$\frac{d}{dt}\log\left(\mathbf{X}(t)\right) = \int_0^1 \left(\left(\mathbf{X}(t) - \mathbf{I}\right)s + \mathbf{I}\right)^{-1} \frac{d}{dt}\mathbf{X}(t) \left(\left(\mathbf{X}(t) - \mathbf{I}\right)s + \mathbf{I}\right)^{-1} ds \tag{A.10}$$

The numerical evaluation of this integral can be a difficult and computationally time-consuming task. We hereafter show that our method is a fast and accurate procedure to evaluate the derivative of the matrix logarithm.

We have implemented in C++ equation A.10 by resorting to a Gauss-Kronrod approximation (with 21 points) of the integral, which is interesting since it yields an estimate of the error. We have also implemented two versions of our proposed method in C++. The first version (called version 1) relies on the computation of $\frac{d\mathbf{D}}{dt}$ from the characteristic polynomial. The second version (called version 2) uses the result of section A.1.2.

We first applied theses three methods to the following symmetric case:

$$\mathbf{X} = \begin{pmatrix} 9 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 9 \end{pmatrix} \qquad \frac{d\mathbf{X}}{dt} = \begin{pmatrix} 0.211325 & 0.543185 & 0.424983 \\ 0.543185 & 0.665381 & 0.657061 \\ 0.424983 & 0.657061 & 0.878216 \end{pmatrix}$$

The quantity $\frac{d \log \mathbf{X}(t)}{dt}$ was estimated by the Gauss-Kronrod approximation and versions 1 and 2 of our algorithm. Results were identical with a precision of 10^{-6} and we found

$$\frac{d \log \mathbf{X}(t)}{dt} = \begin{pmatrix} 0.0105527 & 0.0694973 & 0.0320969 \\ 0.0694973 & 0.134841 & 0.0825508 \\ 0.0320969 & 0.0825508 & 0.083588 \end{pmatrix}$$

moreover, the error estimate of the Gaussian quadrature was

$$\pm \begin{pmatrix} 2.41888 \times 10^{-6} & 1.50485 \times 10^{-6} & 3.28319 \times 10^{-6} \\ 1.50485 \times 10^{-6} & 7.18358 \times 10^{-7} & 1.82484 \times 10^{-6} \\ 3.28319 \times 10^{-6} & 1.82484 \times 10^{-6} & 4.42678 \times 10^{-6} \end{pmatrix}$$

Regarding the timing of the three methods, Gauss-Kronrod approximation took 2.32 seconds for $10,000$ evaluations of $\frac{d \log \mathbf{X}(t)}{dt}$ while versions 1 and 2 of our algorithm respectively took 0.45 seconds and 0.37 seconds.

We next applied the three methods to the general (non symmetric) case:

$$\mathbf{X} = \begin{pmatrix} 9 & 1 & 1 \\ 2 & 4 & 1 \\ 3 & 7 & 9 \end{pmatrix} \qquad \frac{d\mathbf{X}}{dt} = \begin{pmatrix} 0.211325 & 0.543185 & 0.424983 \\ 0.543185 & 0.665381 & 0.657061 \\ 0.424983 & 0.657061 & 0.878216 \end{pmatrix}$$

Hence again, results were identical with a precision of 10^{-6} with

$$\frac{d \log \mathbf{X}(t)}{dt} = \begin{pmatrix} 0.00668372 & 0.0574211 & 0.0362147 \\ 0.0608406 & 0.0805234 & 0.0896055 \\ 0.00321555 & 0.00143868 & 0.0507173 \end{pmatrix}$$

and the error estimate of the Gaussian quadrature was

$$\pm \begin{pmatrix} 5.86227 \times 10^{-6} & 5.20053 \times 10^{-6} & 4.75582 \times 10^{-6} \\ 4.78581 \times 10^{-6} & 3.85294 \times 10^{-6} & 3.51304 \times 10^{-6} \\ 1.40296 \times 10^{-5} & 1.23989 \times 10^{-5} & 1.13835 \times 10^{-5} \end{pmatrix}$$

Gauss-Kronrod approximation took 2.3 seconds for $10,000$ evaluations of $\frac{d \log \mathbf{X}(t)}{dt}$ while versions 1 and 2 of our algorithm respectively took 0.61 seconds and 0.56 seconds.

We can thus conclude that our method is always about 4 to 6 times faster that Gaussian quadrature even though we should notice that it may depend on the properties of the matrix $\mathbf{X}$. Versions 1 and 2 are indeed slightly slower in the general case than in the symmetric one because of the higher computational cost of the diagonalization.

APPENDIX B

DETAILS ON THE FIRST VARIATION OF $\mathcal{J}_{\mathrm{AC}}(h, Dh)$

B.1 INTRODUCTION

In this appendix, we compute the first variation of the term $\mathcal{J}_{\mathrm{AC}}(h, Dh)$ introduced in section 9.4 and corresponding to the matching term, for the DTI registration problem, between local covariance matrices of the images I_1 and $\mathcal{T}_h(I_2)$.

We identify the covariance matrices, elements of $S(6)$ with their images by the canonical map φ_6. Because of (9.30) we have

$$\delta_k \mathcal{J}_{\mathrm{AC}}(h, Dh) = \int_\Omega \delta_k \mathcal{J}_{\mathrm{AC}}(x, h)\, dx.$$

Because of (9.27) we have

$$\begin{aligned}\delta_k \mathcal{J}_{\mathrm{AC}}(x, h) = \frac{1}{2}\Big(\quad & \tfrac{\partial}{\partial \Lambda_2} \|\Lambda_2(x, h) - \tilde{\Lambda}_{12}(x, h)\|_F^2\, \delta_k \Lambda_2(x, h) + \\ & \tfrac{\partial}{\partial \tilde{\Lambda}_{12}} \|\Lambda_2(x, h) - \tilde{\Lambda}_{12}(x, h)\|_F^2\, \delta_k \tilde{\Lambda}_{12}(x, h) + \\ & \tfrac{\partial}{\partial \tilde{\Lambda}_{21}} \|\Lambda_1(x) - \tilde{\Lambda}_{21}(x, h)\|_F^2\, \delta_k \tilde{\Lambda}_{21}(x, h) \Big)\end{aligned}$$

In this equation, the partial derivatives are covariant vectors and the variations $\delta_k \cdot$ are contravariant vectors. The expression of the partial derivatives follows from the fact that $\|A - B\|_F^2 = \mathrm{tr}\left((A - B)(A - B)^T\right)$ and $\frac{\partial}{\partial A}\|A - B\|_F^2 = A - B = -\frac{\partial}{\partial B}\|A - B\|_F^2$

$$\begin{aligned}\frac{1}{2}\frac{\partial}{\partial \Lambda_2}\|\Lambda_2(x, h) - \tilde{\Lambda}_{12}(x, h)\|_F^2 &= \Lambda_2(x, h) - \tilde{\Lambda}_{12}(x, h) \overset{\mathrm{def}}{=} \Theta(x) \\ \frac{1}{2}\frac{\partial}{\partial \tilde{\Lambda}_{12}}\|\Lambda_2(x, h) - \tilde{\Lambda}_{12}(x, h)\|_F^2 &= \tilde{\Lambda}_{12}(x, h) - \Lambda_2(x, h) = -\Theta(x) \\ \frac{1}{2}\frac{\partial}{\partial \tilde{\Lambda}_{21}}\|\Lambda_1(x) - \tilde{\Lambda}_{21}(x, h)\|_F^2 &= \tilde{\Lambda}_{21}(x, h) - \Lambda_1(x) \overset{\mathrm{def}}{=} \Phi(x)\end{aligned}$$

Note that in these formulas, Θ and Φ are 21-dimensional covariant vectors that we identify for convenience with their images by φ_9^{-1}. Θ and Φ are therefore twice covariant tensors. Note that, since we know how to compute the gradient of the geodesic distance function $\mathcal{D}$, it is straightforward to define these quantities when they involve $\mathcal{D}$ instead of the Frobenius norm.

We have obtained an expression for $\delta_k \mathcal{J}_{\rm AC}(x,h)$ and thus, by integration, $\delta_k \mathcal{J}_{\rm AC}(h, Dh)$ as the sum of three terms

$$\begin{aligned}
\delta_k \mathcal{J}^1_{\rm AC}(x,h) &= \Theta(x)\delta_k \Lambda_2(x,h)\\
\delta_k \mathcal{J}^2_{\rm AC}(x,h) &= -\Theta(x)\delta_k \tilde{\Lambda}_{12}(x,h)\\
\delta_k \mathcal{J}^3_{\rm AC}(x,h) &= \Phi(x)\delta_k \tilde{\Lambda}_{21}(x,h)
\end{aligned}$$

Let us write the first equation using indexes:

$$\delta_k \mathcal{J}^1_{\rm AC}(x,h) = \Theta_{IJ}\delta_k \Lambda_2^{IJ}(x,h)$$

We thus need to compute the three quantities $\delta_k\Lambda_2(x,h)$, $\delta_k\tilde{\Lambda}_{12}(x,h)$, $\delta_k\tilde{\Lambda}_{21}(x,h)$. It involves taking derivatives of logarithms and exponentials of matrices which require some numerical care. Appendix A detailed how to evaluate the directional derivative of matrix logarithms and we already referred the reader to [215] for details on the exponential case. The following computations are not difficult but tend to be a little bit involved.

B.2 COMPUTATION OF $\delta_k\Lambda_2(x,h)$

Because of (9.13) we have

$$\delta_k\Lambda_2(x,h) = \frac{1}{\mid \Omega \mid}\int_\Omega \Big((\delta_k\beta_2(x,y,h))\beta_2^T(x,y,h) + \beta_2(x,y,h)(\delta_k\beta_2(x,y,h))^T \Big)\, dy \qquad \text{(B.1)}$$

Because of (9.12) we have

$$\begin{aligned}
\delta_k\beta_2(x,y,h) = -G_\gamma(x-y)\varphi_3\Big(& (\delta_k\hat{\mu}_2(x,h))\log\left(\mathcal{T}_h(I_2)^{-1}(y)\hat{\mu}_2(x,h)\right) + \\
& \hat{\mu}_2(x,h)\left(\delta_k \log\left(\mathcal{T}_h(I_2)^{-1}(y)\hat{\mu}_2(x,h)\right)\right)\Big) \qquad \text{(B.2)}
\end{aligned}$$

We have computed $\delta_k\hat{\mu}_2(x,h)$ in section 9.5.1, we now compute $\delta_k \log\left(\mathcal{T}_h(I_2)^{-1}(y)\hat{\mu}_2(x,h)\right)$.

Computation of $\delta_k \log\left(\mathcal{T}_h(I_2)^{-1}(y)\hat{\mu}_2(x,h)\right)$: We note that since the matrix $\mathcal{T}_h(I_2)^{-1}(y)\hat{\mu}_2(x,h)$ is similar to $\hat{\mu}_2(x,h)^{1/2}\mathcal{T}_h(I_2)^{-1}(y)\hat{\mu}_2(x,h)^{1/2}$ which belongs to S^+, it satisfies the hypotheses of corollary 9.2.2.2 and we can write

$$\delta_k\left(\log\left(\mathcal{T}_h(I_2)^{-1}(y)\hat{\mu}_2(x,h)\right)\right) \;=\; \operatorname{dlog}\left(\mathcal{T}_h(I_2)^{-1}(y)\hat{\mu}_2(x,h), \delta_k\left(\mathcal{T}_h(I_2)^{-1}(y)\hat{\mu}_2(x,h)\right)\right)$$

We need to compute $\delta_k\left(\mathcal{T}_h(I_2)^{-1}(y)\hat{\mu}_2(x,h)\right)$. Using the formula for the derivative of a product

$$\delta_k\left(\mathcal{T}_h(I_2)^{-1}(y)\hat{\mu}_2(x,h)\right) = \delta_k\left(\mathcal{T}_h(I_2)^{-1}(y)\right)\hat{\mu}_2(x,h) + \mathcal{T}_h(I_2)^{-1}(y)\delta_k\hat{\mu}_2(x,h)$$

We have already computed the second term in the righthand side (equations (9.32) and (9.35)), hence

$$\delta_k \left(\mathcal{T}_h(I_2)^{-1}(y)\right) \hat{\mu}_2(x,h) = \frac{\partial \mathcal{T}_h(I_2)^{-1}}{\partial h}(y)\hat{\mu}_2(x,h)k(y) + \frac{\partial \mathcal{T}_h(I_2)^{-1}}{\partial Dh}(y)\hat{\mu}_2(x,h)Dk(y)$$

Hence we get,

$$\begin{aligned}\delta_k \left(\mathcal{T}_h(I_2)^{-1}(y)\hat{\mu}_2(x,h)\right) &= \mathcal{T}_h(I_2)^{-1}(y)\delta_k \hat{\mu}_2(x,h) \\ &\quad + \frac{\partial \mathcal{T}_h(I_2)^{-1}}{\partial h}(y)\hat{\mu}_2(x,h)k(y) + \frac{\partial \mathcal{T}_h(I_2)^{-1}}{\partial Dh}(y)\hat{\mu}_2(x,h)Dk(y),\end{aligned}$$

which we write in tensor form

$$\begin{aligned}&\delta_k \left(\mathcal{T}_h(I_2)^{-1}(y)\hat{\mu}_2(x,h)\right) = \\ &\quad \int_\Omega \mathbf{t}_1(x,y,z)k(z)\,dz + \mathbf{t}_2(x,y)k(y) - \int_\Omega \mathbf{u}_1(x,y,z)Dk(z)\,dz - \mathbf{u}_2(x,y)Dk(y), \quad \textbf{(B.3)}\end{aligned}$$

where $\mathbf{t}_1$ is the once contravariant and twice covariant tensor obtained by contracting the second covariant index of $\mathcal{T}_h(I_2)^{-1}(y)$ with the first contravariant index of $\mathbf{t}$:

$${\mathbf{t}_1}^{i}_{\ jl}(x,y,z) = -(\mathcal{T}_h(I_2)^{-1})_{jm}(y)\mathbf{t}_l^{mi}(x,z),$$

and $\mathbf{u}_1$ is the twice covariant and twice contravariant tensor obtained from $\mathbf{U}$ in a similar fashion:

$${\mathbf{u}_1}^{im}_{\ jl}(x,y,z) = -(\mathcal{T}_h(I_2)^{-1})_{jn}(y)\mathbf{u}_l^{nim}(x,z).$$

The tensor $\mathbf{t}_2 = \frac{\partial \mathcal{T}_h(I_2)^{-1}}{\partial h}(y)\hat{\mu}_2$ is once contravariant and twice covariant; its coordinates are given by

$${\mathbf{t}_2}^{i}_{\ jl} = \left(\frac{\partial \mathcal{T}_h(I_2)^{-1}}{\partial h}\right)_{jlm} \hat{\mu}_2^{mi}.$$

The tensor $\mathbf{u}_2 = \frac{\partial \mathcal{T}_h(I_2)^{-1}}{\partial Dh}(y)\hat{\mu}_2$ is twice contravariant and twice covariant; its coordinates are given by

$${\mathbf{u}_2}^{im}_{\ jl} = \left(\frac{\partial \mathcal{T}_h(I_2)^{-1}}{\partial Dh}\right)^{m}_{jln} \hat{\mu}_2^{ni}.$$

Computation of $\delta_k\beta_2$ and $\delta_k\Lambda_2$: We now do a bit of rewriting in order to get an expression for $\delta_k\beta_2$ and $\delta_k\Lambda_2$. This is tedious but not difficult. We first prove the following

Lemma B.2.0.1. *$\delta_k\beta_2$ can be written as*

$$\begin{aligned}\delta_k\beta_2(x,y) = \int_\Omega \mathbf{T}_{1\beta_2}(x,y,z)k(z)\,dz + \mathbf{T}_{2\beta_2}(x,y)k(y) - \\ \int_\Omega \mathbf{U}_{1\beta_2}(x,y,z)Dk(z)\,dz - \mathbf{U}_{2\beta_2}(x,y)Dk(y),\end{aligned}$$

where the expressions of the tensors $\mathbf{T}_{1\beta_2}$, $\mathbf{T}_{2\beta_2}$, $\mathbf{U}_{1\beta_2}$ and $\mathbf{U}_{2\beta_2}$ are given in the proof.

Proof. We combine equations (9.41) and (B.3). Using equation (B.2) we can then write

$$
\begin{aligned}
&\delta_k\beta_2(x,y) = \\
&\quad -\varphi_3\Big(\Big(\Big(\int_\Omega \mathbf{t}(x,z)k(z)\,dz - \int_\Omega \mathbf{u}(x,z)Dk(z)\,dz\Big)\log\big(\mathcal{T}_h(I_2)^{-1}(y)\hat{\mu}_2(x,h)\big) + \\
&\qquad\qquad \hat{\mu}_2(x,h)\Big(\int_\Omega \mathbf{t}_1(x,y,z)k(z)\,dz + \mathbf{t}_2(x,y)k(y) - \\
&\qquad\qquad\qquad \int_\Omega \mathbf{u}_1(x,y,z)Dk(z) - \mathbf{u}_2(x,y)Dk(y)\Big)\Big)G(x-y)\Big),
\end{aligned}
$$

and obtain

$$
\begin{aligned}
\mathbf{T}_{1\beta_2}(x,y,z) &= -\varphi_3((\mathbf{t}(x,z)\log\big(\mathcal{T}_h(I_2)^{-1}(y)\hat{\mu}_2(x)\big) + \hat{\mu}_2(x,h)\mathbf{t}_1(x,y))G(x-y)) \\
\mathbf{T}_{2\beta_2}(x,y) &= -\varphi_3(\hat{\mu}_2(x,h)\mathbf{t}_2(x,y)G(x-y)) \\
\mathbf{U}_{1\beta_2}(x,y,z) &= \varphi_3((\mathbf{u}(x,z)\log\big(\mathcal{T}_h(I_2)^{-1}(y)\hat{\mu}_2(x,h)\big) + \hat{\mu}_2(x,h)\mathbf{u}_1(x,y))G(x-y)) \\
\mathbf{U}_{2\beta_2}(x,y) &= \varphi_3(\hat{\mu}_2(x,h)\mathbf{u}_2(x,y)G(x-y))
\end{aligned}
$$

□

This lemma allows us to prove the following proposition concerning the form of the first term in the expression of $\delta_k\mathcal{J}_{AC}$.

Proposition B.2.0.1. *The first term $\delta_k\mathcal{J}^1_{\mathrm{AC}}$ in the expression of $\delta_k\mathcal{J}_{\mathrm{AC}}$ is of the form described in theorem 9.5.0.1.*

Proof. The previous manipulations and equation (B.1) yield

$$
\begin{aligned}
&(\delta_k\Lambda_2(x))^{IJ} = \\
&\qquad \frac{1}{\mid\Omega\mid}\int_\Omega\Big(\Big(\int_\Omega \mathbf{T}^I_{1\beta_2\ l}(x,y,z)k^l(z)\,dz + \mathbf{T}^I_{2\beta_2\ l}(x,y)k^l(y) \\
&\qquad - \int_\Omega \mathbf{U}^{Im}_{2\beta_2\ l}(x,y,z)Dk^l_m(z)\,dz - \mathbf{U}^{Im}_{1\beta_2\ l}(x,y)Dk^l_m(y)\Big)\beta^J_2(x,y) \\
&\qquad + \beta^I_2(x,y)\Big(\int_\Omega \mathbf{T}^J_{1\beta_2\ l}(x,y,z)k^l(z)\,dz + \mathbf{T}^J_{2\beta_2\ l}(x,y)k^l(y) \\
&\qquad\qquad - \int_\Omega \mathbf{U}^{Jm}_{1\beta_2\ l}(x,y,z)Dk^l_m(z)\,dz - \mathbf{U}^{Jm}_{2\beta_2\ l}(x,y)Dk^l_m(y)\Big)\Big)\,dy
\end{aligned}
$$

The corresponding term $\delta_k\mathcal{J}^1_{\mathrm{AC}}(x,h) = \Theta_{IJ}(x)(\delta_k\Lambda_2(x))^{IJ}$ in $\delta_k\mathcal{J}_{\mathrm{AC}}(x,h)$ is

$$
\begin{aligned}
&\frac{\Theta_{IJ}(x)}{\mid\Omega\mid}\int_\Omega\Big(\Big(\int_\Omega \mathbf{T}^I_{1\beta_2\ l}(x,y,z)k^l(z)\,dz + \mathbf{T}^I_{2\beta_2\ l}(x,y)k^l(y) \\
&\qquad - \int_\Omega \mathbf{U}^{Im}_{1\beta_2\ l}(x,y,z)Dk^l_m(z)\,dz - \mathbf{U}^{Im}_l(x,y)Dk^{2\beta_2\ l}_m(y)\Big)\beta^J_2(x,y) \\
&\qquad + \beta^I_2(x,y)\Big(\int_\Omega \mathbf{T}^J_{1\beta_2\ l}(x,y,z)k^l(z)\,dz + \mathbf{T}^J_{2\beta_2\ l}(x,y)k^l(y) \\
&\qquad\qquad - \int_\Omega \mathbf{U}^{Jm}_{1\beta_2\ l}(x,y,z)Dk^l_m(z)\,dz - \mathbf{U}^{Jm}_{2\beta_2\ l}(x,y)Dk^l_m(y)\Big)\Big)\,dy \qquad \text{(B.4)}
\end{aligned}
$$

We define

$$\mathbf{H}_l(x,y) = \frac{\Theta_{IJ}(x)}{\mid \Omega \mid} \int_\Omega \Big(\mathbf{T}^I_{1\beta_2\ l}(x,z,y)\beta_2^J(x,z) + \beta_2^I(x,z)\mathbf{T}^J_{1\beta_2\ l}(x,z,y) \Big)\, dz + \\ \frac{\Theta_{IJ}(x)}{2 \mid \Omega \mid} \Big(\mathbf{T}^I_{2\beta_2\ l}(x,y)\beta_2^J(x,y) + \beta_2^I(x,y)\mathbf{T}^J_{2\beta_2\ l}(x,y) \Big),$$

and

$$\mathbf{K}_l^m(x,y) = \frac{\Theta_{IJ}(x)}{\mid \Omega \mid} \int_\Omega \Big(\mathbf{U}^{Im}_{1\beta_2\ l}(x,z,y)\beta_2^J(x,z) + \beta_2^I(x,z)\mathbf{U}^{Jm}_{1\beta_2\ l}(x,z,y) \Big)\, dz + \\ \frac{\Theta_{IJ}(x)}{2 \mid \Omega \mid} \Big(\mathbf{U}^{Im}_{2\beta_2\ l}(x,y)\beta_2^J(x,y) + \beta_2^I(x,y)\mathbf{U}^{Jm}_{2\beta_2\ l}(x,y) \Big),$$

and rewrite (B.4) as

$$\delta_k \mathcal{J}^1_{\mathrm{AC}}(x,h) = \int_\Omega \mathbf{H}_l(x,y,h) k^l(y)\, dy - \int_\Omega \mathbf{K}_l^m(x,y,h) D k^l_m(y)\, dy$$

The corresponding term in $\delta_k \mathcal{J}_{\mathrm{AC}}(h)$ is

$$\int_\Omega \left(\int_\Omega \mathbf{H}_l(x,y,h) k^l(y)\, dy \right) dx - \int_\Omega \left(\int_\Omega \mathbf{K}_l^m(x,y,h) D k^l_m(y)\, dy \right) dx. \tag{B.5}$$

Now define

$$\overline{\mathbf{H}}_l(x,h) = \int_\Omega \mathbf{H}_l(z,x,h)\, dz,$$

and

$$\overline{\mathbf{K}}_l^m(x,h) = \int_\Omega \mathbf{K}_l^m(z,x,h)\, dz.$$

Exchanging the order of summation in the second term of (B.5) and renaming the variables, we obtain a new form of (B.5):

$$\int_\Omega \overline{\mathbf{H}}_l(x,h) k^l(x)\, dx - \int_\Omega \overline{\mathbf{K}}_l^m(x,h) D k^l_m(x)\, dx \overset{Def}{=} \\ \int_\Omega \mathbf{T}^1_{\mathrm{AC}}(x,h) k(x)\, dx - \int_\Omega \mathbf{U}^1_{\mathrm{AC}}(x,h) D k(x)\, dx$$

□

B.3 COMPUTATION OF $\delta_k \tilde{\Lambda}_{12}(x,h)$

According to equation (9.14), we have

$$\delta_k \tilde{\Lambda}_{12}(x,h) \quad = \quad \frac{1}{\mid \Omega \mid} \int_\Omega \left(\delta_k \tilde{\beta}_1(x,y,h) \tilde{\beta}_1^T(x,y,h) + \tilde{\beta}_1(x,y,h) \left(\delta_k \tilde{\beta}_1(x,y,h) \right)^T \right) dy$$

Because of (9.22), we have

$$\delta_k \tilde{\beta}_1(x,y,h) = \delta_k \Big(\exp\left(-\mathcal{A}(x,h)\right) \beta_1(x,y) \Big) = \Big(\delta_k \exp\left(-\mathcal{A}(x,h)\right) \Big) \beta_1(x,y), \tag{B.6}$$

since the vectors $\beta_1(x,y)$ are not functions of h. $\mathcal{A}(x,h)$ is defined in section 9.4.2. In order to compute $\delta_k \exp\left(-\mathcal{A}(x,h)\right)$, we use the following result from [215]. Let X be a diagonalizable matrix of $M_n(\mathbb{R})$, and V a matrix. We are interested in computing the directional derivative of the exponential of X in the direction V

$$\mathrm{dexp}\,(X, V) = \lim_{t \to 0} \frac{1}{t}(\exp(X + tV) - \exp(X))$$

The following theorem (page 41 of [215]) provides an answer and shows that $\mathrm{dexp}\,(X, V)$ is, like the matrix logarithm, linear in its second argument V.

Theorem B.3.0.1. *If $X = ZDZ^{-1}$ is the spectral decomposition of the semi-simple matrix X, its directional derivative in the direction V is given by*

$$\mathrm{dexp}\,(X, V) = Z\left(\overline{V} \bullet \Xi\right) Z^{-1}$$

where $\overline{V} = Z^{-1}VZ$ and $\overline{V} \bullet \Xi$ denote the Hadamard (entry-by-entry) product of $\overline{V}$ with the matrix Ξ whose entries are given by:

$$\Xi^i_j = \Xi^j_i = \begin{cases} \frac{e^{\lambda_i} - e^{\lambda_j}}{\lambda_i - \lambda_j} & if \lambda_i \neq \lambda_j \\ e^{\lambda_i} & if \lambda_i = \lambda_j \end{cases}$$

Computation of $\delta_k \exp\left(-\mathcal{A}(x, h)\right)$: According to definition 9.2.2.1 and the chain rule

$$\delta_k \exp\left(-\mathcal{A}(x, h)\right) = -\mathrm{dexp}\left(-\mathcal{A}(x, h), \delta_k \mathcal{A}(x, h)\right).$$

According to equation (9.24), in order to compute $\delta_k \mathcal{A}(x, h)$, we need to compute $\delta_k \psi(x, h)$ where

$$\psi(x, h) = \log\left(\hat{\mu}_2(x, h)\hat{\mu}_1(x)^{-1}\right).$$

According to corollary 9.2.2.2 and definition 9.2.2.1 we can write

$$\delta_k \psi(x, h) = \mathrm{dlog}\left(\hat{\mu}_2(x, h)\hat{\mu}_1(x)^{-1}, \left(\delta_k \hat{\mu}_2(x, h)\right)\hat{\mu}_1(x)^{-1}\right).$$

Because $\mathrm{dlog}\,(,)$ is a linear function of its second argument, using equation (9.41), the previous equation can be rewritten as

$$\delta_k \psi(x, h) = \int_\Omega \mathbf{t}_\psi(x, z)k(z)\,dz - \int_\Omega \mathbf{u}_\psi(x, z)Dk(z)\,dz,$$

where

$$\mathbf{t}_\psi(x, z) = \mathrm{dlog}\left(\hat{\mu}_2(x, h)\hat{\mu}_1(x)^{-1}, \mathbf{t}(x, z)\hat{\mu}_1(x)^{-1}\right)$$
$$\mathbf{u}_\psi(x, z) = \mathrm{dlog}\left(\hat{\mu}_2(x, h)\hat{\mu}_1(x)^{-1}, \mathbf{u}(x, z)\hat{\mu}_1(x)^{-1}\right).$$

Using indexes,

$$\mathbf{t}_\psi{}^{\cdot\cdot}{}_l = \sum_{m=1}^{3} \mathrm{dlog}\left(\hat{\mu}_2(x, h)\hat{\mu}_1(x)^{-1}, \mathbf{T}_l^{\cdot\,m}\left(\hat{\mu}_1(x)^{-1}\right)_{m\cdot}\right)$$

Since the relation between ψ and $\mathcal{A}$ is linear ($\mathcal{A} = \mathcal{M}(\psi)$, see equation (9.24)), we have

$$\delta_k \mathcal{A}(x,h) = \int_\Omega \mathcal{M}(\mathbf{t}_\psi(x,z))k(z)\,dz - \int_\Omega \mathcal{M}(\mathbf{u}_\psi(x,z))Dk(z)\,dz,$$

and therefore, using the linearity in the second argument of $\operatorname{dexp}(\cdot,\cdot)$,

$$\begin{aligned}&\delta_k \exp\left(-\mathcal{A}(x,h)\right) = \\ &-\int_\Omega \operatorname{dexp}\left(-\mathcal{A}(x,h), \mathcal{M}(\mathbf{t}_\psi(x,z))\right) k(z)\,dz + \int_\Omega \operatorname{dexp}\left(-\mathcal{A}(x,h), \mathcal{M}(\mathbf{u}_\psi(x,z))\right) Dk(z)\,dz.\end{aligned}$$

We obtain an expression for $\delta_k \tilde{\beta}_1$

$$\delta_k \tilde{\beta}_1(x,y) = \int_\Omega \mathbf{T}_{\tilde{\beta}_1}(x,y,z)k(z)\,dz - \int_\Omega \mathbf{U}_{\tilde{\beta}_1}(x,y,z)Dk(z)\,dz,$$

where

$$\begin{aligned}\mathbf{T}_{\tilde{\beta}_1}(x,y,z) &= -\operatorname{dexp}\left(-\mathcal{A}(x), \mathcal{M}(\mathbf{t}_\psi(x,z))\right)\beta_1(x,y)\\ \mathbf{U}_{\tilde{\beta}_1}(x,y,z) &= -\operatorname{dexp}\left(-\mathcal{A}(x), \mathcal{M}(\mathbf{u}_\psi(x,z))\right)\beta_1(x,y)\end{aligned}$$

This allows us to prove the following

Proposition B.3.0.2. *The second term, $\delta_k \mathcal{J}^2_{AC}$, in the expression of $\delta_k \mathcal{J}_{AC}$ is of the form described in theorem 9.5.0.1.*

Proof. The previous manipulations yield

$$\begin{aligned}\left(\delta_k \tilde{\Lambda}_{12}(x)\right)^{IJ} = \frac{1}{\mid\Omega\mid}\int_\Omega \Big(\Big(\int_\Omega \Big(\mathbf{T}^{I}_{\tilde{\beta}_1\ l}(x,y,z)k^l(z) - \mathbf{U}^{Im}_{\tilde{\beta}_1\ l}(x,y,z)Dk^m_l(z)\Big)\,dz\Big)\tilde{\beta}^J_1(x,y) + \\ \tilde{\beta}^I_1(x,y)\int_\Omega \Big(\mathbf{T}^{J}_{\tilde{\beta}_1\ l}(x,y,z)k^l(z) - \mathbf{U}^{Jm}_{\tilde{\beta}_1\ l}(x,y,z)Dk^m_l(z)\Big)\,dz\Big)\,dy\end{aligned}$$

The corresponding term, $\delta_k \mathcal{J}^2_{\mathrm{AC}}(x,h)$, in $\delta_k \mathcal{J}_{\mathrm{AC}}(x,h)$ is

$$\begin{aligned}-\frac{\Theta_{IJ}(x)}{\mid\Omega\mid}\int_\Omega \Big(\Big(\int_\Omega \Big(\mathbf{T}^{I}_{\tilde{\beta}_1\ l}(x,y,z)k^l(z) - \mathbf{U}^{Im}_{\tilde{\beta}_1\ l}(x,y,z)Dk^m_l(z)\Big)\,dz\Big)\tilde{\beta}^J_1(x,y) + \\ \tilde{\beta}^I_1(x,y)\int_\Omega \Big(\mathbf{T}^{J}_{\tilde{\beta}_1\ l}(x,y,z)k^l(z) - \mathbf{U}^{Jm}_{\tilde{\beta}_1\ l}(x,y,z)Dk^m_l(z)\Big)\,dz\Big)\,dy\end{aligned}$$

This results in the following expression for $\delta_k \mathcal{J}^2_{\mathrm{AC}}(h)$

$$\delta_k \mathcal{J}^2_{\mathrm{AC}}(h) = \int_\Omega \mathbf{T}^2_{\mathrm{AC}}(x,h)k(x)\,dx - \int_\Omega \mathbf{U}^2_{\mathrm{AC}}(x,h)Dk(x)\,dx,$$

where

$$\mathbf{T}^2_{\mathrm{AC}\ l}(x,h) = -\frac{1}{\mid\Omega\mid}\int_\Omega\int_\Omega \Theta_{IJ}(y)\Big(\mathbf{T}^{I}_{\tilde{\beta}_1\ l}(y,z,x)\tilde{\beta}^J_1(y,z) + \tilde{\beta}^I_1(y,z)\mathbf{T}^{J}_{\tilde{\beta}_1\ l}(y,z,x)\Big)\,dy\,dz,$$

and

$$\mathbf{U}^{2\ m}_{\mathrm{AC}\ l}(x,h) = -\frac{1}{\mid\Omega\mid}\int_\Omega\int_\Omega \Theta_{IJ}(y)\Big(\mathbf{U}^{Im}_{\tilde{\beta}_1\ l}(y,z,x)\tilde{\beta}^J_1(y,z) + \tilde{\beta}^I_1(y,z)\mathbf{U}^{Jm}_{\tilde{\beta}_1\ l}(y,z,x)\Big)\,dy\,dz$$

$\square$

B.4 COMPUTATION OF $\delta_k \tilde{\Lambda}_{21}(x,h)$

According to equation (9.15), we have

$$\delta_k \tilde{\Lambda}_{21}(x,h) = \frac{1}{|\,\Omega\,|} \int_\Omega \left(\delta_k \tilde{\beta}_2(x,y,h) \tilde{\beta}_2^T(x,y,h) + \tilde{\beta}_2(x,y,h) \left(\delta_k \tilde{\beta}_2(x,y,h) \right)^T \right) dy \quad \textbf{(B.7)}$$

Because of (9.23), we have

$$\begin{aligned} \delta_k \tilde{\beta}_2(x,y,h) &= \delta_k \Big(\exp\left(-\mathcal{B}(x,h)\right) \beta_2(x,y,h) \Big) \\ &= \Big(\delta_k \exp\left(-\mathcal{B}(x,h)\right) \Big) \beta_2(x,y,h) + \exp\left(-\mathcal{B}(x,h)\right) \Big(\delta_k \beta_2(x,y,h) \Big) \end{aligned}$$

We have already derived, in the previous sections, all we need to evaluate this derivative.

Computation of the first term of $\delta_k \tilde{\beta}_2(x,y,h)$: The first term of $\delta_k \tilde{\beta}_2(x,y,h)$, namely

$$\Big(\delta_k \exp\left(-\mathcal{B}(x,h)\right) \Big) \beta_2(x,y,h),$$

is readily obtained from the derivations carried out in section B.3. Since

$$\delta_k \exp\left(-\mathcal{B}(x,h)\right) = -\mathrm{dexp}\left(-\mathcal{B}(x,h), \delta_k \mathcal{B}(x,h)\right),$$

all the arguments used previously to derive an expression for $\delta_k \exp\left(-\mathcal{A}(x,h)\right)$ are still valid. Replacing ψ by

$$\theta(x,h) = \log\left(\hat{\mu}_1(x)^{-1} \hat{\mu}_2(x,h)\right),$$

results in the expressions

$$\delta_k \theta(x,h) = \int_\Omega \mathbf{t}_\theta(x,z) k(z)\, dz - \int_\Omega \mathbf{u}_\theta(x,z) Dk(z),$$

where

$$\begin{aligned} \mathbf{t}_\theta(x,z) &= \mathrm{dlog}\left(\hat{\mu}_1(x)^{-1} \hat{\mu}_2(x,h), \hat{\mu}_1(x)^{-1} \mathbf{t}(x,z)\right) \\ \mathbf{u}_\theta(x,z) &= \mathrm{dlog}\left(\hat{\mu}_1(x)^{-1} \hat{\mu}_2(x,h), \hat{\mu}_1(x)^{-1} \mathbf{u}(x,z)\right). \end{aligned}$$

Using indexes,

$$\mathbf{t}_{\theta}{}^{\cdot\cdot}_{\,l} = \sum_{m=1}^{3} \mathrm{dlog}\left(\hat{\mu}_1(x)^{-1} \hat{\mu}_2(x,h), \left(\hat{\mu}_1(x)^{-1}\right)_{m\,\cdot} \mathbf{t}_l^{\cdot\, m}\right)$$

Since the relation between θ and $\mathcal{B}$ is linear ($\mathcal{B} = \mathcal{M}(\theta)$, see equation (9.24)), we have

$$\delta_k \mathcal{B}(x,h) = \int_\Omega \mathcal{M}(\mathbf{t}_\theta(x,z)) k(z)\, dz - \int_\Omega \mathcal{M}(\mathbf{u}_\theta(x,z)) Dk(z)\, dz,$$

and therefore, using the linearity of $\mathrm{dexp}\left(\cdot,\cdot\right)$ with respect to its second argument:

$$\begin{aligned} &\delta_k \exp\left(-\mathcal{B}(x,h)\right) \beta_2(x,y,h) = \\ &\qquad - \int_\Omega \mathrm{dexp}\left(-\mathcal{B}(x,h), \mathcal{M}(\mathbf{t}_\theta(x,z))\right) \beta_2(x,y,h) k(z)\, dz + \\ &\qquad\qquad \int_\Omega \mathrm{dexp}\left(-\mathcal{B}(x,h), \mathcal{M}(\mathbf{u}_\theta(x,z))\right) \beta_2(x,y,h) Dk(z)\, dz \end{aligned}$$

Computation of the second term of $\delta_k\tilde{\beta}_2(x, y, h)$: The second term of $\delta_k\tilde{\beta}_2(x, y, h)$, namely

$$\exp\left(-\mathcal{B}(x, h)\right)\Big(\delta_k\beta_2(x, y, h)\Big)$$

is readily obtained from the derivations lead in section B.2. Indeed the derivative $\delta_k\beta_2(x, y, h)$ was proved in lemma B.2.0.1 to be equal to

$$\delta_k\beta_2 = \int_\Omega \mathbf{T}_{1\beta_2}(x, y, z)k(z)\,dz + \mathbf{T}_{2\beta_2}(x, y)k(y) - \\ \int_\Omega \mathbf{U}_{1\beta_2}(x, y, z)Dk(z)\,dz - \mathbf{U}_{2\beta_2}(x, y)Dk(y)$$

Combining these two results allows us to prove the following lemma, analog to lemma B.2.0.1

Lemma B.4.0.2. *$\delta_k\tilde{\beta}_2$ can be written as*

$$\delta_k\beta_2(x, y) = \int_\Omega \mathbf{T}_{1\tilde{\beta}_2}(x, y, z)k(z)\,dz + \mathbf{T}_{2\tilde{\beta}_2}(x, y)k(y) - \\ \int_\Omega \mathbf{U}_{1\tilde{\beta}_2}(x, y, z)Dk(z)\,dz - \mathbf{U}_{2\tilde{\beta}_2}(x, y)Dk(y),$$

where the expressions of the tensors $\mathbf{T}_{1\tilde{\beta}_2}$, $\mathbf{T}_{2\tilde{\beta}_2}$, $\mathbf{U}_{1\tilde{\beta}_2}$ and $\mathbf{U}_{2\tilde{\beta}_2}$ are given in the proof.

Proof. We can write immediately

$$\delta_k\tilde{\beta}_2 = \int_\Omega \mathbf{T}_{1\tilde{\beta}_2}(x, y, z, h)k(z)\,dz + \mathbf{T}_{2\tilde{\beta}_2}(x, y, h)k(y) - \\ \int_\Omega \mathbf{U}_{1\tilde{\beta}_2}(x, y, z, h)Dk(z)\,dz - \mathbf{U}_{2\tilde{\beta}_2}(x, y, h)Dk(y),$$

where

$$\begin{aligned}
\mathbf{T}_{1\tilde{\beta}_2}(x, y, z) &= -\mathrm{dexp}\left(-\mathcal{B}(x, h), \mathcal{M}(\mathbf{t}_\theta(x, z))\right)\beta_2(x, y, h) + \exp\left(-\mathcal{B}(x, h)\right)\mathbf{T}_{1\beta_2}(x, y, z, h) \\
\mathbf{T}_{2\tilde{\beta}_2}(x, y) &= \exp\left(-\mathcal{B}(x, h)\right)\mathbf{T}_{2\beta_2}(x, y, h) \\
\mathbf{U}_{1\tilde{\beta}_2}(x, y, z) &= -\mathrm{dexp}\left(-\mathcal{B}(x, h), \mathcal{M}(\mathbf{u}_\theta(x, z))\right)\beta_2(x, y, h) + \exp\left(-\mathcal{B}(x, h)\right)\mathbf{U}_{1\beta_2}(x, y, z, h) \\
\mathbf{U}_{2\tilde{\beta}_2}(x, y) &= \exp\left(-\mathcal{B}(x, h)\right)\mathbf{U}_{2\beta_2}(x, y, h)
\end{aligned}$$

□

This allows us to prove the following

Proposition B.4.0.3. *The third term, $\delta_k\mathcal{J}^3_{\mathrm{AC}}$, in the expression of $\delta_k\mathcal{J}_{\mathrm{AC}}$ is of the form described in theorem 9.5.0.1.*

Proof. The proof is completely analog to that of proposition B.2.0.1. We end up with

$$\delta_k\mathcal{J}^3_{\mathrm{AC}}(h) = \int_\Omega \mathbf{T}^3_{\mathrm{AC}}(x, h)k(x)\,dx - \int_\Omega \mathbf{U}^3_{\mathrm{AC}}(x, h)Dk(x)\,dx,$$

where

$$\begin{aligned}\mathbf{T}^{3}_{\mathrm{AC}\ l}(x,h) = \overline{\widetilde{\mathbf{H}}}(x,h) & \qquad\qquad = \int_{\Omega} \overline{\widetilde{\mathbf{H}}}(z,x,h)\,dz \\ \mathbf{U}^{3\ m}_{\mathrm{AC}\ l}(x,h) = \overline{\widetilde{\mathbf{K}}}(x,h) & \qquad\qquad = \int_{\Omega} \overline{\widetilde{\mathbf{K}}}(z,x,h)\,dz,\end{aligned}$$

where the once covariant tensor $\widetilde{\mathbf{H}}(z,x,h)$ is given by the following expression

$$\begin{aligned}\widetilde{\mathbf{H}}_{l}(x,y) = \frac{\Phi_{IJ}(x)}{\mid \Omega \mid} \int_{\Omega} \Big(\mathbf{T}^{I}_{1\tilde{\beta}_2\ l}(x,z,y)\tilde{\beta}^{J}_{2}(x,z) + \tilde{\beta}^{I}_{2}(x,z)\mathbf{T}^{J}_{1\tilde{\beta}_2\ l}(x,z,y)\Big)\,dz + \\ \frac{\Phi_{IJ}(x)}{2 \mid \Omega \mid}\Big(\mathbf{T}^{I}_{2\tilde{\beta}_2\ l}(x,y)\tilde{\beta}^{J}_{2}(x,y) + \tilde{\beta}^{I}_{2}(x,y)\mathbf{T}^{J}_{2\tilde{\beta}_2\ l}(x,y)\Big),\end{aligned}$$

and the once covariant once contravariant tensor $\widetilde{\mathbf{K}}(z,x,h)$ is given by the following expression

$$\begin{aligned}\widetilde{\mathbf{K}}^{m}_{l}(x,y) = \frac{\Phi_{IJ}(x)}{\mid \Omega \mid} \int_{\Omega} \Big(\mathbf{U}^{Im}_{1\tilde{\beta}_2\ l}(x,z,y)\tilde{\beta}^{J}_{2}(x,z) + \tilde{\beta}^{I}_{2}(x,z)\mathbf{U}^{Jm}_{1\tilde{\beta}_2\ l}(x,z,y)\Big)\,dz + \\ \frac{\Phi_{IJ}(x)}{2 \mid \Omega \mid}\Big(\mathbf{U}^{Im}_{2\tilde{\beta}_2\ l}(x,y)\tilde{\beta}^{J}_{2}(x,y) + \tilde{\beta}^{I}_{2}(x,y)\mathbf{U}^{Jm}_{2\tilde{\beta}_2\ l}(x,y)\Big).\end{aligned}$$

□

APPENDIX **C**

PUBLICATIONS OF THE AUTHOR

Book chapters

- R. Deriche, D. Tschumperlé, C. Lenglet and M. Rousson
  Variational Approaches to the Estimation, Regularization and Segmentation of Diffusion Tensor Images, Mathematical Models in Computer Vision: The Handbook, Springer 2005, Editors: N. Paragios, Y. Chen and O. Faugeras

Journal papers

- C.A. Castaño Moraga, C. Lenglet, R. Deriche, J. Ruiz-Alzola
  A Riemannian Approach to Anisotropic Filtering of Tensor Fields, Signal Processing, Special issue on Tensor Signal Processing, 2006 (In press)
- C. Lenglet, M. Rousson, R. Deriche
  DTI Segmentation by Statistical Surface Evolution, IEEE Transactions on Medical Imaging, 25(6), 685-700, June 2006
- C. Lenglet, M. Rousson, R. Deriche, O. Faugeras
  Statistics on the Manifold of Multivariate Normal Distributions: Theory and Application to Diffusion Tensor MRI Processing, Journal of Mathematical Imaging and Vision, 25(3), 423-444, October 2006
- O. Faugeras, G. Adde, G. Charpiat, C. Chefd'Hotel, M. Clerc, T. Deneux, R. Deriche, G. Hermosillo, R. Keriven, P. Kornprobst, J. Kybic, C. Lenglet, L. Lopez-Perez, T. Papadopoulo, J.P. Pons, F. Segonne, B. Thirion, D. Tschumperlé, T. Viéville, N. Wotawa
  Variational, Geometric and Statistical Methods for Modeling Brain Anatomy and Function, Neuroimage, 23(1), S46-S55, 2004

Conference papers and abstracts

- M. Descoteaux, C. Lenglet, R. Deriche
  Diffusion Tensor Sharpening Improves White-matter Tractography, Proc. SPIE

Medical Imaging, San Diego, CA, February 17-22, 2007 (To appear)

- E. Prados, C. Lenglet, J.P. Pons, N. Wotawa, R. Deriche, O. Faugeras, S. Soatto
  Control Theory and Fast Marching Methods for Brain Connectivity Mapping, Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, June 17-22, 2006
- C. Lenglet, M. Rousson, R. Deriche
  A Statistical Framework for DTI Segmentation, Proc. IEEE Intl. Symposium on Biomedical Imaging: From Nano to Macro, Arlington, VA, April 6-9, 2006
- C.A. Castaño Moraga, C. Lenglet, R. Deriche, J. Ruiz-Alzola
  A Fast and Rigorous Anisotropic Smoothing Method for DT-MRI, Proc. IEEE Intl. Symposium on Biomedical Imaging: From Nano to Macro, Arlington, VA, April 6-9, 2006
- C. Lenglet, M. Rousson, R. Deriche, O. Faugeras, S. Lehéricy, K. Ugurbil
  A Riemannian Approach to Diffusion Tensor Images Segmentation, Proc. Information Processing in Medical Imaging, Glenwood Springs, CO, July 11-15, 2005
- S. Lehéricy, C. Lenglet, J. Doyon, H. Benali, P.F. Van de Moortele, G. Sapiro, O. Faugeras, R. Deriche, K. Ugurbil
  Activation Shifts from the Premotor to the Sensorimotor Territory of the Striatum during the Course of Motor Sequence Learning, Proc. 11th Annual Meeting of the Organization for Human Brain Mapping, Toronto, June 12-16, 2005
- N. Wotawa, C. Lenglet, M. Roth, B. Nazarian, J.L. Anton, R. Majhoub, R. Deriche, O. Faugeras
  Combined fMRI and DTI of the Human Low Level Visual Cortex, Proc. 11th Annual Meeting of the Organization for Human Brain Mapping, Toronto, June 12-16, 2005
- C. Lenglet, R. Deriche, O. Faugeras, S. Lehéricy, K. Ugurbil
  A Riemannian Approach to Diffusion Tensors Estimation and Streamline-based Fiber Tracking, Proc. 11th Annual Meeting of the Organization for Human Brain Mapping, Toronto, June 12-16, 2005
- C. Lenglet, O. Faugeras, T. Papadopoulo, R. Deriche
  Diffeomorphic Matching of Symmetric Positive Definite Matrix Fields, Proc. 2nd Ntl. Congress of Industrial and Applied Mathematics, Evian, May 23-27, 2005
- C. Lenglet, M. Rousson, R. Deriche
  Segmentation of 3D Probability Density Fields by Surface Evolution: Application to diffusion MRI, Proc. 7th Intl. Conf. on Medical Image Computing and Computer Assisted Intervention, Saint-Malo, France, Septembre 26-30, 2004

- C. Lenglet, R. Deriche, O. Faugeras
  Anatomical Connectivity Mapping Inferred from DTI-based White Matter Geometry, Proc. 10th Annual Meeting of the Organization for Human Brain Mapping, Budapest, June 13-17, 2004
- M. Rousson, C. Lenglet, R. Deriche
  Level Set and Region Based Surface Propagation for Diffusion Tensor MRI Segmentation, Proc. Computer Vision Approaches to Medical Image Analysis, Prague, May 11-14, 2004
- C. Lenglet, R. Deriche, O. Faugeras
  Inferring White Matter Geometry from Diffusion Tensor MRI: Application to Connectivity Mapping, Proc. 8th European Conference on Computer Vision, Prague, May 11-14, 2004
- R. Deriche, D. Tschumperlé, C. Lenglet
  DT-MRI Regularization & Fiber Tractography, Proc. IEEE Intl. Symposium on Biomedical Imaging: From Nano to Macro, Arlington, VA, April 15-18, 2004

Technical reports

- E. Prados, C. Lenglet, J.P. Pons, N. Wotawa, R. Deriche, O. Faugeras, S. Soatto
  Control Theory and Fast Marching Methods for Brain Connectivity Mapping, INRIA Research Report 5845, February 2006 and UCLA Computer Science Department Technical Report, February 2006
- C. Lenglet, M. Rousson, R. Deriche
  DTI Segmentation by Statistical Surface Evolution, INRIA Research Report 5843, February 2006
- C. Lenglet, M. Rousson, R. Deriche, O. Faugeras
  Toward Segmentation of 3D Probability Density Fields by Surface Evolution: Application to Diffusion MRI, INRIA Research Report 5243, June 2004
- C. Lenglet, M. Rousson, R. Deriche, O. Faugeras
  Statistics on Multivariate Normal Distributions: A Geometric Approach and its Application to Diffusion Tensor MRI, INRIA Research Report 5242, June 2004
- C. Lenglet, R. Deriche, O. Faugeras
  Diffusion Tensor Magnetic Resonance Imaging: Brain Connectivity Mapping, INRIA Research Report 4983, October 2003

Bibliography

[1] D. Adalsteinsson and J.A. Sethian. The fast construction of extension velocities in level set methods. *Journal of Computational Physics*, 148(1):2–22, 1999.

[2] G.W. Albers, M.G. Lansberg, A.M. Norbash, D.C. Tong, M.W. O'Brien, A.R. Woolfenden, M.P. Marks, and M.E. Moseley. Yield of diffusion-weighted MRI for detection of potentially relevant findings in stroke patients. *Neurology*, 54:1562–1567, 2000.

[3] A.L. Alexander, Hasan K.M., M. Lazar, J.S. Tsuruda, and D.L. Parker. Analysis of partial volume effects in diffusion-tensor MRI. *Magnetic Resonance in Medicine*, 45(5):770–780, 2001.

[4] D.C. Alexander and J.C. Gee. Elastic matching of diffusion tensor images. *Computer Vision and Image Understanding*, 77:233–250, 2000.

[5] D.C. Alexander, J.C. Gee, and R. Bajcsy. Similarity measures for matching diffusion tensor images. In *British Machine Vision Conference*, page 251, 1999.

[6] D.C. Alexander, J.C. Gee, and R. Bajcsy. Strategies for data reorientation during non-rigid warps of diffusion tensor images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 463–472, 1999.

[7] G.E. Alexander and M.D. Crutcher. Preparation for movement: neural representations of intended direction in three motor areas of the monkey. *J. Neurophysiol*, 64:133–150, 1990.

[8] G.E. Alexander, M.R. DeLong, and P.L. Strick. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu. Rev. Neurosci.*, 9:357–381, 1986.

[9] S.I. Amari. *Differential-Geometrical Methods in Statistics*. Lectures Notes in Statistics. Springer-Verlag, 1990.

[10] B.A. Ardekani, J. Nierenberg, M.J. Hoptman, D.C. Javitt, and K.O. Lim. MRI study of white matter diffusion anisotropy in schizophrenia. *NeuroReport*, 14(16):2025–2029, November 2003.

[11] V. Arsigny, O. Commowick, X. Pennec, and N. Ayache. A Log-Euclidean framework for statistics on diffeomorphisms. In *Proceedings of the 9th International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 924–931, 2006.

[12] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Fast and simple calculus on tensors in the log-euclidean framework. In *Proceedings of the 8th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, pages 115–122, 2005.

[13] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 2006. In press.

[14] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Log-euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine*, 56:411–421, 2006.

[15] C. Atkinson and A.F.S. Mitchell. Rao's distance measure. *Sankhya: The Indian Journal of Stats.*, 43(A):345–365, 1981.

[16] V. Balasubramanian. A geometric formulation of Occam's razor for inference of parametric distributions. PUPT 1588, Princeton University, December 2001.

[17] C. Ballester, V. Caselles, and M. González. Affine invariant segmentation by variational method. *SIAM Journal of Applied Mathematics*, 56:294–325, 1996.

[18] F. Barbaresco, C. Germond, and N. Rivereau. Advanced radar diffusive CFAR based on geometric flow and information theories and its extension for Doppler and polarimetric data. In *Proc. International Conference on Radar Systems*, 2004.

[19] M. Bardi and I. Capuzzo-Dolcetta. *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*. Birkhauser, 1997.

[20] A. Barmpoutis, B.C. Vemuri, and J.R. Forder. Robust tensor splines for approximation of diffusion tensor MRI data. In *IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, page 86, 2006.

[21] P. Basser and S. Pajevic. A normal distribution for tensor-valued random variables: Applications to diffusion tensor MRI. *IEEE Transactions on Medical Imaging*, 22(7):785–794, 2003.

[22] P.J. Basser. Inferring microstructural features and the physiological state of tissues from diffusion-weighted images. *Nuclear Magnetic Resonance in Biomedicine*, 8:333–344, 1995.

[23] P.J. Basser, J. Mattiello, and D. Le Bihan. Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance*, B(103):247–254, 1994.

[24] P.J. Basser, J. Mattiello, and D. Le Bihan. MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66(1):259–267, 1994.

[25] P.J. Basser, J. Mattiello, R. Turner, and D. Le Bihan. Diffusion tensor echo-planar imaging of human brain. In *Proceedings of the SMRM*, page 584, 1993.

[26] P.J. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi. In vivo fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine*, 44:625–632, 2000.

[27] P.J. Basser and C. Pierpaoli. Microstructural and physiological features of tissues elucidated by quantitative diffusion tensor MRI. *Journal of Magnetic Resonance*, 111(3):209–219, 1996.

[28] P.G. Batchelor, D.L.G. Hill, F. Calamante, and D. Atkinson. Study of connectivity in the brain using the full diffusion tensor from MRI. In *Image Processing in Medical Imaging*, pages 121–133, 2001.

[29] C. Beaulieu. The basis of anisotropic water diffusion in the nervous system - a technical review. *NMR in Biomedicine*, 15:435–455, 2002.

[30] T.E.J. Behrens, H. Johansen-Berg, M.W. Woolrich, S.M. Smith, 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*, 6(7):750–757, 2003.

[31] 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*, 50:1077–1088, 2003.

[32] Marcel Berger. *Riemannian Geometry During the Second Half of the Twentieth Century*, volume 17 of *University Lecture Series*. American Mathematical Society, 2000.

[33] Marcel Berger and Bernard Gostiaux. *Differential Geometry: Manifolds, Curves, and Surfaces*. Number 115 in Graduate Texts in Mathematics. Springer–Verlag, 1988.

[34] D. Le Bihan and E. Breton. Imagerie de diffusion *in vivo* par résonnance magnétique nucléaire. *CR Académie des Sciences*, 301:1109–1112, 1985.

[35] D. Le Bihan, E. Breton, D. Lallemand, P. Grenier, E. Cabanis, and M. Laval-Jeantet. Mr imaging of intravoxel incoherent motions: Application to diffusion and perfusion in neurologic disorders. *Radiology*, pages 401–407, 1986.

[36] D. Le Bihan, E. Breton, D. Lallemand, P. Grenier, E. Cabanis, and M. Laval-Jeantet. MR imaging of intravoxel incoherent motions: Application to diffusion and perfusion in neurologic disorders. *Radiology*, 161(2):401–407, 1986.

[37] D. Le Bihan, J.F. Mangin, C. Poupon, C.A. Clark, S. Pappata, N. Molko, and H. Chabriat. Diffusion tensor imaging: Concepts and applications. *Journal of Magnetic Resonance Imaging*, 13:534–546, 2001.

[38] M. Bjornemo, A. Brun, R. Kikinis, and C.F. Westin. Regularized stochastic white matter tractography using diffusion tensor MRI. In *MICCAI*, pages 435–442, 2002.

[39] F. Bloch. Nuclear induction. *Physical Review*, 70:460–474, 1946.

[40] H. Brezis. *Analyse fonctionnelle. Théorie et applications*. Masson, 1983.

[41] K. Brodmann. *Vergleichende Lokalisationslehre der Grobhirnrinde*. J.A.Barth, Leipzig, 1909.

[42] R. Brown. A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Phil. Mag.*, 4:161–173, 1828.

[43] T. Brox, J. Weickert, B. Burgeth, and P. Mrázek. Nonlinear structure tensors. Technical Report 113, Department of Mathematics, Saarland University, Saarbrücken, Germany., October 2004.

[44] J. Burbea. Informative geometry of probability spaces. *Expositiones Mathematica*, 4:347–378, 1986.

[45] J. Burbea and C.R. Rao. Entropy differential metric, distance and divergence measures in probability spaces: A unified approach. *Journal of Multivariate Analysis*, 12:575–596, 1982.

[46] P.T. Callaghan. *Principles of nuclear magnetic resonance microscopy*. Clarendon Press, Oxford, 2nd edition, 1995.

[47] M. Calvo and J.M. Oller. A distance between multivariate normal distributions based in an embedding into the Siegel group. *Journal of Multivariate Analysis*, 35(2):223–242, 1990.

[48] M. Calvo and J.M. Oller. An explicit solution of information geodesic equations for the multivariate normal model. *Statistics and Decisions*, 9:119–138, 1991.

[49] J. Campbell. *Diffusion Imaging of White Matter Fiber Tracts*. PhD thesis, McGill University, 2004.

[50] J.S.W. Campbell, K. Siddiqi, V.V. Rymar, A. Sadikot, and B.G. Pike. Flow-based fiber tracking with diffusion tensor q-ball data: Validation and comparison to principal diffusion direction techniques. *NeuroImage*, 27(4):725–736, October 2005.

[51] J.S.W. Campbell, K. Siddiqi, B.C. Vemuri, and G.B Pike. A geometric flow for white matter fibre tract reconstruction. In *IEEE International Symposium on Biomedical Imaging Conference Proceedings*, pages 505–508, July 2002.

[52] Y. Cao, M.I. Miller, S. Mori, R.L. Winslow, and L. Younes. Diffeomorphic matching of diffusion tensor images. In *Conference on Computer Vision and Pattern Recognition Workshop*, page 67, 2006.

[53] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *The International Journal of Computer Vision*, 22(1):61–79, 1997.

[54] C.A. Castaño-Moraga, M.A. Rodriguez-Florido, L. Alvarez, C.F. Westin, and J. Ruiz-Alzola. Anisotropic interpolation of DT-MRI. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 343–350, 2004.

[55] 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*, 17:77–94, 2002.

[56] M. Catani, D.K. Jones, R. Donato, and D. H. ffytche. Occipito-temporal connections in the human brain. *Brain*, 126(9):2093–2107, September 2003.

[57] A. Caticha. Change, time and information geometry. In *Proc. Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, volume 568, pages 72–82, 2000.

[58] M. Cercignani and M.A. Horsfield. The physical basis of diffusion-weighted MRI. *J. Neurol.*, 186(1):S11–S14, 2001.

[59] T. Chan and L. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, February 2001.

[60] L.C. Chang, D.K. Jones, and C. Pierpaoli. Restore: Robust estimation of tensors by outlier rejection. *Magnetic Resonance in Medicine*, 53:1088–1095, 2005.

[61] P. Charbonnier, G. Aubert, M. Blanc-Féraud, and M. Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. In *Proceedings of the International Conference on Image Processing*, volume II, pages 168–172, 1994.

[62] G. Charpiat, O. Faugeras, and R. Keriven. Approximations of shape metrics and application to shape warping and empirical shape statistics. *Foundations of Computational Mathematics*, 5(1):1–58, February 2005.

[63] C. Chefd'hotel, D. Tschumperlé, R. Deriche, and O. Faugeras. Constrained flows on matrix-valued functions : application to diffusion tensor regularization. In *Proceedings of ECCV'02*, page 251, June 2002.

[64] C. Chefd'hotel, D. Tschumperlé, R. Deriche, and O. Faugeras. Regularizing flows for constrained matrix-valued images. *Journal of Mathematical Imaging and Vision*, 20(1-2):147–162, 2004.

[65] W. Chen, X. Zhu, et al. Retinotopic mapping of lateral geniculate nucleus in humans using functional magnetic resonance imaging. *Proc. Natl. Acad. Sci. USA*, 96:2430–2434, March 1999.

[66] Y.G. Chen, Y. Giga, and S. Goto. Uniqueness and existence of viscosity solutions of generalized mean curvature flow equations. *Journal on Differential Geometry*, 33:749–786, 1991.

[67] David L. Chopp. Computing minimal surfaces via level set curvature flow. *Journal of Computational Physics*, 106:77–91, 1993.

[68] M.K. Chung, M. Lazar, A.L. Alexander, Y. Lu, and R. Davidson. Probabilistic connectivity measure in diffusion tensor imaging via anisotropic kernel smoothing. Technical Report TR-1081, Department of Statistics, University of Wisconsin-Madison, 2003.

[69] P.G. Ciarlet. *Mathematical Elasticity*, volume 1. North Holland, 1988.

[70] O. Ciccarelli, G.J.M. Parker, A.T. Toosy, C.A.M. Wheeler-Kingshott, G.J. Barker, P.A. Boulby, D.H. Miller, and A.J. Thompson. From diffusion tractography to quantitative white matter tract measures: a reproducibility study. *NeuroImage*, 18(2):348–359, February 2003.

[71] O. Ciccarelli, A.T. Toosy, S.J. Hickman, G.J.M. Parker, C.A.M. Wheeler-Kingshott, D.H. Miller, and A.J. Thompson. Optic radiation changes after optic neuritis detected by tractography-based group mapping. *Human Brain Mapping*, 25(3):308–316, 2005.

[72] O. Ciccarelli, A.T. Toosy, G.J.M. Parker, C.A.M. Wheeler-Kingshott, G.J. Barker, D.H. Miller, and A.J. Thompson. Diffusion tractography based group mapping of major white-matter pathways in the human brain. *NeuroImage*, 19(4):1545–1555, August 2003.

[73] S. Clarke, P. Maeder, R. Meuli, F. Staub, A. Bellmann, L. Regli, N. de Tribolet, and Assal G. Interhemispheric transfer of visual motion information after a posterior callosal lesion: a neuropsychological and fMRI study. *Experimental Brain Research*, 132(1):127–133, 2000.

[74] S. Clarke and J. Miklossy. Occipital cortex in man: Organization of callosal connections, related myelo- and cytoarchitecture, and putative boundaries of functional visual areas. *The Journal of Comparative Neurology*, 298(2):188–214, 1990.

[75] O. Clatz, M. Sermesant, P.Y· Bondiau, H. Delingette, S.K. Warfield, G. Malandain, and N. Ayache. Realistic simulation of the 3D growth of brain tumors in MR images coupling diffusion with mass effect. *IEEE Transactions on Medical Imaging*, 24(10):1334–1346, 2005.

[76] L. Cohen. Minimal paths and fast marching methods for image analysis. In *Mathematical Models in Computer Vision: The Handbook*, chapter 7. Springer, 2005.

[77] Y. Cointepas, J.-F. Mangin, Line Garnero, J.-B. Poline, and H. Benali. BrainVISA: Software platform for visualization and analysis of multi-modality brain data. In *Proc. 7th HBM*, page S98, Brighton, United Kingdom, 2001.

[78] 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. *Proceedings of the National Academy of Sciences*, 96:10422–10427, August 1999.

[79] I. Corouge, S. Gouttard, and G. Gerig. A statistical shape model of individual fiber tracts extracted from diffusion tensor MRI. In *MICCAI*, volume 3217 of *LNCS*, pages 671–679, 2004.

[80] O. Coulon, D.C. Alexander, and S.R. Arridge. Diffusion tensor magnetic resonance image regularisation. *Medical Image Analysis*, 8(1):47–67, 2004.

[81] R. Courant. *Calculus of Variations*. New York, 1946.

[82] A.M. Dale, B. Fischl, and M.I. Sereno. Cortical surface-based analysis I: Segmentation and surface reconstruction. *NeuroImage*, 9:179–194, 1999.

[83] R.W.R. Darling. Intrinsic location parameter of a diffusion process. Technical Report Tech Report 493, UC Berkeley, Department of Statististics, 1998.

[84] P.I Davies and N.J. Higham. A Schur-Parlett algorithm for computing matrix functions. *SIAM J. Matrix Anal. Appl.*, 25(2):464–485, 2003.

[85] M. de Lara. Geometric and symmetry properties of a nondegenerate diffusion process. *An. of Probability*, 23(4):1557–1604, 1995.

[86] M.P. Deiber, R.E. Passingham, J.G. Colebatch, K.J Friston, P.D. Nixon, and R.S. Frackowiak. Cortical areas and the selection of movement: a study with positron emission tomography. *Exp. Brain. Res.*, 84:393–402, 1991.

[87] J.J. Dejerine. *Anatomie des Centres Nerveux*. Paris, Rueff & Cie., 1901.

[88] A. Dervieux and F. Thomasset. A finite element method for the simulation of Rayleigh-Taylor instability. *Lecture Notes in Mathematics*, 771:145–159, 1979.

[89] A. Dervieux and F. Thomasset. Multifluid incompressible flows by a finite element method. In W. Reynolds and R.W. MacCormack, editors, *Seventh International Conference on Numerical Methods in Fluid Dynamics*, volume 141 of *Lecture Notes in Physics*, pages 158–163, June 1980.

[90] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche. A linear and regularized ODF estimation algorithm to recover multiple fibers in Q-Ball imaging. Technical Report 5768, INRIA, November 2005.

[91] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R Deriche. Apparent diffusion coefficients from high angular resolution diffusion imaging: Estimation and applications. *Magnetic Resonance in Medicine*, 56:395–410, 2006.

[92] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R Deriche. A fast and robust ODF estimation algorithm in Q-ball imaging. In *Third IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 81–84, Arlington, Virginia, USA., April 2006.

[93] C. Dettmers, G.R. Fink, R.N. Lemon, K.N. Stephan, R.E. Passingham, D. Silbersweig, A. Holmes, M.C. Ridding, D.J. Brooks, and R.S. Frackowiak. Relation between cerebral activity and force in the motor areas of the human brain. *J. Neurophysiol.*, 74:802–815, 1995.

[94] L. Devroye. *Non-Uniform Random Variate Generation*. Springer–Verlag, 1986.

[95] L. Dieci, B. Morini, and A. Papini. Computational techniques for real logarithms of matrices. *SIAM J. Matrix Anal. Appl.*, 17:570–593, 1996.

[96] M. P. Do Carmo. *Differential Geometry of Curves and Surfaces*. Prentice-Hall, 1976.

[97] M. P. Do Carmo. *Riemannian Geometry*. Birkhäuser, 1992.

[98] Q. Dong, R.C. Welsh, T.L. Chenevert, R.C. Carlos, P. Maly-Sundgren, D.M. Gomez-Hassan, and S.K. Mukherji. Clinical applications of diffusion tensor imaging. *Journal of Magnetic Resonance Imaging*, 19:6–18, 2004.

[99] 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 Science*, 102(20):7350–7355, May 2005.

[100] J. Doyon, V. Penhune, and L.G. Ungerleider. Distinct contribution of the cortico-striatal and cortico-cerebellar systems to motor skill learning. *Neuropsychologia*, 41:252–262, 2003.

[101] J. Dubois, L. Hertz-Pannier, G. Dehaene-Lambertz, Y. Cointepas, and D. Le Bihan. Assessment of the early organization and maturation of infants' cerebral white matter fiber bundles: a feasibility study using quantitative diffusion tensor imaging and tractography. *Neuroimage*, 30(4):1121–1132, 2006.

[102] S.P. Dukelow, J.F.X. DeSouza, J.C. Culham, A.V. Van Den Berg, R. Menon, and T. Vilis. Distinguishing subregions of the human MT+ complex using visual fields and pursuit eye movements. *Journal of Neurophysiology*, 86(4):1991–2000, October 2001.

[103] S. Dumoulin, R. Bittar, N.J. Kabani, C.L. Baker, G. Le Goualher, G.B. Pike, and A.C. Evans. A new anatomical landmark for reliable identification of human area V5/MT: a quantitative analysis of sulcal patterning. *Cerebral Cortex*, 10:454–463, May 2000.

[104] H.M. Duvernoy, P. Bourgouin, E.A. Cabanis, E.A. Cattin, J. Guyot, M.T. Iba-Zizen, P. Maeder, B. Parratte, L. Tatu, and F. Vuillier. *The Human brain: surface, three-dimensional sectional anatomy with MRI, and blood supply*. Springer, 1999.

[105] A. Einstein. *Investigations on the Theory of the Brownian Movement*. New York: Dover, 1956.

[106] I. Ekeland and R. Temam. *Analyse Convexe et Problèmes Variationnels*. Etudes mathématiques. Dunod; Gauthier-Villars, Paris, Bruxelles, Montreal, 1974.

[107] P.S. Eriksen. Geodesics connected with the fisher metric on the multivariate manifold. Technical Report 86-13, Institute of Electronic Systems, Aalborg University, 1986.

[108] L.C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. Proceedings of the American Mathematical Society, 1998.

[109] O. Faugeras, G. Adde, G. Charpiat, C. Chefd'Hotel, M. Clerc, T. Deneux, R. Deriche, G. Hermosillo, R. Keriven, P. Kornprobst, J. Kybic, C. Lenglet, L. Lopez-Perez, T. Papadopoulo, J.-P. Pons, F. Ségonne, B. Thirion, D. Tschumperlé, T. Viéville, and N. Wotawa. Variational, geometric, and statistical methods for modeling brain anatomy and function. *NeuroImage*, 23S1:S46–S55, 2004.

Special issue: Mathematics in Brain Imaging - Edited by P.M. Thompson, M.I. Miller, T. Ratnanather, R.A. Poldrack and T.E. Nichols.

[110] O. Faugeras and G. Hermosillo. Well-posedness of two non-rigid multimodal image registration methods. *Siam Journal of Applied Mathematics*, 64(5):1550–1587, 2004.

[111] C. Feddern, J. Weickert, and B. Burgeth. Level-set methods for tensor-valued images. In *Proc. Second IEEE Workshop on Variational, Geometric and Level Set Methods in Computer Vision*, pages 65–72, Nice, France, 2003.

[112] C. Feddern, J. Weickert, B. Burgeth, and M. Welk. Curvature-driven PDE methods for matrix-valued images. Technical Report 104, Department of Mathematics, Saarland University, Saarbrücken, Germany, April 2004.

[113] P. Fillard, V. Arsigny, X. Pennec, and N. Ayache. Clinical DT-MRI estimation, smoothing and fiber tracking with log-euclidean metrics. In *Proceedings of the Third IEEE International Symposium on Biomedical Imaging*, pages 786–789, 2006.

[114] P. Fillard, V. Arsigny, X. Pennec, K.M. Hayashi, P.M. Thompson, and N. Ayache. Measuring brain variability by extrapolating sparse tensor fields measured on sulcal lines. *Neuroimage*, 2006. In press.

[115] P. T. Fletcher, C. Lu, and S. Joshi. Statistics of shape via principal geodesic analysis on lie groups. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 95–101, 2003.

[116] P.T. Fletcher and S. Joshi. Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In *Proc. Computer Vision Approaches to Medical Image Analysis*, pages 87–98, Prague, May 2004.

[117] W. Förstner and B. Moonen. A metric for covariance matrices. Technical report, Stuttgart University, Dept. of Geodesy and Geoinformatics, 1999.

[118] L.R. Frank. Anisotropy in high angular resolution diffusion-weighted MRI. *Magnetic Resonance in Medicine*, 45:935–939, 2001.

[119] M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré*, X(IV):215–310, 1948.

[120] L. Freire, A. Roche, and J.F. Mangin. What is the best similarity measure for motion correction in fMRI time series? *IEEE Transactions on Medical Imaging*, 21(5):470–484, May 2002.

[121] O. Friman, G. Farneback, and C.F. Westin. A Bayesian approach for stochastic white matter tractography. *IEEE Transactions on Medical Imaging*, 25:965–978, 2006.

[122] O. Friman and C.F. Westin. Uncertainty in white matter fiber tractography. In *Proceedings of MICCAI*, pages 107–114. Springer-verlag, 2005.

[123] K.J. Friston, A.P. Holmes, J.B. Poline, P.J. Grasby, S.C. Williams, R.S. Frackowiak, and R. Turner. Analysis of fMRI time series revisited. *Neuroimage*, 2:45–53, 1995.

[124] G. Gong, T. Jiang, C. Zhu, Y. Zang, Y. He, S. Xie, and J. Xiao. Side and handedness effects on cingulum from diffusion tensor imaging. *NeuroReport*, 16(15):1701–1705, 2005.

[125] G. Gong, T. Jiang, C. Zhu, Y. Zang, F. Wang, S. Xie, J. Xiao, and X. Guo. Asymmetry analysis of cingulum based on scale-invariant parameterization by diffusion tensor imaging. *Human Brain Mapping*, 24(2):92–98, 2005.

[126] S.T. Grafton, J.C Mazziotta, S. Presty, K.J. Friston, R.S. Frackowiak, and M.E. Phelps. Functional anatomy of human procedural learning determined with regional cerebral blood flow and pet. *J. Neurosci.*, 12:2542–2548, 1992.

[127] A.M. Graybiel. Building action repertoires: memory and learning functions of the basal ganglia. *Curr. Opin. Neurobiol.*, 5(6):733–741, 1995.

[128] K. Grill-Spector and R. Malach. The human visual cortex. *Annual Reviews Neuroscience*, 27:649–677, 2004.

[129] K. Grove, H. Karcher, and E.A. Ruh. Group actions and curvature. *Inventiones Mathematicae*, 23:31–48, 1974.

[130] A. Guimond, C.R.G. Guttmann, S.K. Warfield, and C.F. Westin. Deformable registration of DT-MRI data based on transformation invariant tensor characteristics. In *IEEE International Symposium on Biomedical Imaging*, Washington (DC), USA, July 7–10 2002.

[131] A. Guimond, A. Roche, N. Ayache, and J. Meunier. Three-dimensional multimodal brain warping using the demons algorithm and adaptive intensity corrections. *IEEE Transactions in Medical Imaging*, 20(1):58–69, 2001.

[132] A.C. Guo, J.R. MacFall, and J.M. Provenzale. Multiple sclerosis: diffusion tensor MR imaging for evaluation of normal-appearing white matter. *Radiology*, 222(3):729–736, 2002.

[133] M. Guye, G. J.M. Parker, M. Symms, P. Boulby, C.A.M. Wheeler-Kingshott, A. Salek-Haddadi, G.J. Barker, and J.S. Duncana. Combined functional MRI and tractography to demonstrate the connectivity of the human primary motor cortex in vivo. *NeuroImage*, 19:1349–1360, 2003.

[134] M. Guye, G.J.M. Parker, M. Symms, P. Boulby, C.A.M. Wheeler-Kingshott, A. Salek-Haddadi, G.J. Barker, and J.S. Duncana. Combined functional MRI and tractography to demonstrate the connectivity of the human primary motor cortex in vivo. *Neuroimage*, 19:1349–1360, 2003.

[135] N. Hageman, D. Shattuck, K. Narr, and A.W. Toga. A diffusion tensor imaging tractography method based on Navier-Stokes fluid mechanics. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 798–801, 2006.

[136] P. Hagmann, J.P. Thiran, L. Jonasson, P. Vandergheynst, S. Clarke, P. Maeder, and R. Meuli. DTI mapping of human brain connectivity: Statistical fiber tracking and virtual dissection. *NeuroImage*, 19:545–554, 2003.

[137] E.L. Hahn. Spin echoes. *Physical Review*, 80:580–594, 1950.

[138] A. Harten, B. Engquist, S. Osher, and S. Chakravarthy. Uniformly high-order accurate essentially non oscillatory schemes iii. *Journal of Computational Physics*, 71:231–303, 1987.

[139] K.M. Hasan, D.L. Parker, and A.L. Alexander. Comparison of optimization procedures for diffusion-tensor encoding directions. *Journal of Magnetic Resonance Imaging*, 13:769–780, 2001.

[140] K.M. Hasan, D.L. Parker, and A.L. Alexander. Magnetic resonance water self-diffusion tensor encoding optimization methods for full brain acquisition. *Image Analysis and Stereology*, 21(2):87–96, 2002.

[141] R.H. Hashemi and W.O. Bradley. *MRI–The Basics*. Lippincott Williams and Wilkins, Philadelphia, 2nd edition, 2004.

[142] Gerardo Hermosillo. *Variational Methods for Multimodal Image Matching*. PhD thesis, Université de Nice-Sophia Antipolis, May 2002.

[143] J.C. Horton, K. Landau, P. Maeder, and W.F. Hoyt. Magnetic resonance imaging of the human lateral geniculate body. *Archives of Neurology*, 47(11):1201–1206, November 1990.

[144] P.J. Huber. *Robust Statistics*. New York: Wiley, 1981.

[145] A. Huk, R. Dougherty, and D. Heeger. Retinotopy and functional subdivision of human areas MT and MST. *The Journal of Neuroscience*, 22(16):7195–7205, August 2002.

[146] R. Iyer, R. Holsapple, and D. Doman. Optimal control problems on parallelizable Riemannian manifolds: theory and applications. *ESAIM: Control, Optimisation and Calculus of Variations*, 12:1–11, January 2006.

[147] M. Jackowski, C.Y. Kao, M. Qiu, R.T. Constable, and L.H. Staib. White matter tractography by anisotropic wavefront evolution and diffusion tensors imaging. *Medical Image Analysis*, 9:427–440, 2005.

[148] S. Jbabdi, E. Mandonnet, H. Duffau, L. Capelle, K.R. Swanson, M. Pélégrini-Issac, R. Guillevin, and H. Benali. Simulation of anisotropic growth of low-grade gliomas using diffusion tensor imaging. *Magnetic Resonance in Medicine*, 54:616–624, 2005.

[149] H. Jeffreys. *Theory of probability*. Wiley, New York, 1959.

[150] B.J. Jellison, A.S. Field, J. Medow, M. Lazar, M.S. Salamat, and A.L. Alexander. Diffusion tensor imaging of cerebral white matter: A pictorial review of physics, fiber tract anatomy, and tumor imaging patterns. *American Journal of Neuroradiology*, 25:356–360, 2004.

[151] I.H. Jenkins, D.J. Brooks, P.D. Nixon, R.S. Frackowiak, and R.E. Passingham. Motor sequence learning: a study with positron emission tomography. *Journal of Neuroscience*, 14:3775–3790, 1994.

[152] I.H. Jenkins, M. Jahanshahi, M. Jueptner, R.E. Passingham, and D.J. Brooks. Self-initiated versus externally triggered movements. ii. the effect of movement predictability on regional cerebral blood flow. *Brain*, 123:1216–1228, 2000.

[153] P. Jezzard and R. S. Balaban. Correction for geometric distortions in echoplanar images from b0 field variations. *Magnetic Resonance in Medicine*, 34:65–73, 1995.

[154] L. Jonasson, X. Bresson, P. Hagmann, O. Cuisenaire, R. Meuli, and J.P. Thiran. White matter fiber tract segmentation in DT-MRI using geometric flows. *Medical Image Analysis*, 9(3):223–236, 2005.

[155] D. K. Jones, L.D. Griffin, D.C. Alexander, M. Catani, M.A. Horsfield, R. Howard, and S.C. Williams. Spatial normalization and averaging of diffusion tensor MRI data sets. *Neuroimage*, 17(2):592–617, 2002.

[156] D.K. Jones. The effect of gradient sampling schemes on measures derived from diffusion tensor MRI: a Monte Carlo study. *Magnetic Resonance in Medicine*, 51:807–815, 2004.

[157] D.K. Jones, M.A. Horsfield, and A. Simmons. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magnetic Resonance in Medicine*, 42:515–525, 1999.

[158] D.K. Jones, A. Simmons, S.C.R. Williams, and M.A. Horsfield. Noninvasive assessment of axonal fiber connectivity in the human brain via diffusion tensor MRI. *Magnetic Resonance in Medicine*, 42:37–41, 1999.

[159] N. Kang, J. Zhang, E.S. Carlson, and D. Gembris. White matter fiber tractography via anisotropic diffusion simulation in the human brain. *IEEE Transactions on Medical Imaging*, 24:1127–1137, 2005.

[160] H. Karcher. Riemannian centre of mass and mollifier smoothing. *Comm. Pure Appl. Math*, 30:509–541, 1977.

[161] J. Karger and W. Heink. The propagator representation of molecular transport in microporous crystallites. *Journal of Magnetic Resonance Imaging*, 51:1–7, 1983.

[162] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. In *First International Conference on Computer Vision*, pages 259–268, London, June 1987.

[163] W. Kendall. Probability, convexity, and harmonic maps with small image i: uniqueness and fine existence. *Proc. London Math. Soc.*, 61(2):371–406, 1990.

[164] D.S. Kim and M. Kim. Combining functional and diffusion tensor mri. *Ann. N.Y. Acad. Sci.*, 1064:1–15, 2005.

[165] R. Kimmel, A. Amir, and A.F. Bruckstein. Finding shortest paths on surfaces using level set propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(6):635–640, June 1995.

[166] R. Kimmel and J. A. Sethian. Computing Geodesic Paths on Manifolds. *Proc. of the Nat. Academy of Science*, 95, 1998.

[167] C.G. Koay, J.D. Carew, A.L. Alexander, P.J. Basser, and M.E. Meyerand. Investigation of anomalous estimates of tensor-derived quantities in diffusion tensor imaging. *Magnetic Resonance in Medicine*, 55:930–936, 2006.

[168] C.G. Koay, L.C. Change, J.D. Carew, C. Pierpaoli, and P.J. Basser. A unifying theoretical and algorithmic framework for least squares methods of estimation in diffusion tensor imaging. *Journal of Magnetic Resonance*, 182:115–125, 2006.

[169] M. Koch, V. Glauche, J. Finsterbusch, U. Nolte, J. Frahm, and C. Büchel. Estimation of anatomical connectivity from diffusion tensor data. In *7th Annual Meeting of the Organization for Human Brain Mapping*, page 1974, 2001.

[170] M.A. Koch, D.G. Norris, and M. Hund-Georgiadis. An investigation of functional and anatomical connectivity using magnetic resonance imaging. *NeuroImage*, 16:241–250, 2002.

[171] P.C. Lauterbur. Image formation by induced local interactions: examples employing nuclear magnetic resonance. *Nature*, 242:190–191, 1973.

[172] M. Lazar and A. Alexander. Bootstrap white matter tractography (BOOTTRAC). *Neuroimage*, 24:524–532, 2005.

[173] M. Lazar and A.L. Alexander. White matter tractography using random vector (rave) perturbation. In *10th ISMRM Annual Meeting*, page 539, 2002.

[174] 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. In *Human Brain Mapping*, volume 18, pages 306–321, 2003.

[175] Y.G. Leclerc. Constructing simple stable description for image partitioning. *The International Journal of Computer Vision*, 3(1):73–102, 1989.

[176] John M. Lee. *Introduction to Smooth Manifolds*. Springer–Verlag, 2003.

[177] Alexander Leemans. *Modeling and Processing of Diffusion Tensor Mangetic Resonance Images for Improved Analysis of Brain Connectivity*. PhD thesis, University of Antwerp, Belgium, 2006.

[178] S. Lehericy, M. Ducros, P.F. Van de Moortele, C. Francois, L. Thivard, C. Poupon, N. Swindale, K. Ugurbil, and D.S. Kim. Diffusion tensor fiber tracking shows distinct corticostriatal circuits in humans. *Ann. Neurol.*, 55(4):522–529, 2004.

[179] S. Lehericy, C. Lenglet, J. Doyon, H. Benali, P.F. Van de Moortele, G. Sapiro, O. Faugeras, R. Deriche, and K. Ugurbil. Activation shifts from the premotor to the sensorimotor territory of the striatum during the course of motor sequence learning. In *11th Annual Meeting of the Organization for Human Brain Mapping*, Toronto, June 2005.

[180] S. Lehericy, P.F. van de Moortele, E. Lobel, A.L. Paradis, M. Vidailhet, V. Frouin, P. Neveu, Y. Agid, C. Marsault, and D. Le Bihan. Somatotopical organization of striatal activation during finger and toe movement: a 3-t functional magnetic resonance imaging study. *Ann. Neurol.*, 44:398–404, 1998.

[181] S. Lehéricy, H. Benali, P.F. Van de Moortele, M. Pélégrini-Issac, T. Waechter, K. Ugurbil, and J. Doyon. Distinct basal ganglia territories are engaged in early and advanced motor sequence learning. *PNAS*, 102(35):12566–12571, 2003.

[182] C. Lenglet, R. Deriche, and O. Faugeras. Diffusion tensor magnetic resonance imaging: Brain connectivity mapping. Technical Report 4983, INRIA, November 2003.

[183] C. Lenglet, R. Deriche, and O. Faugeras. Inferring white matter geometry from diffusion tensor MRI: Application to connectivity mapping. In *ECCV'2004*, pages 127–140, May 2004.

[184] C. Lenglet, R. Deriche, O. Faugeras, S. Lehericy, and K. Ugurbil. A Riemannian approach to diffusion tensors estimation and streamline-based fiber tracking. In *11th Annual Meeting of the Organization for Human Brain Mapping*, Toronto, June 2005.

[185] C. Lenglet, M. Rousson, and R. Deriche. Segmentation of 3D probability density fields by surface evolution: Application to diffusion MRI. In *Proc. 7th Intl. Conf. on Medical Image Computing and Computer Assisted Intervention*, pages 18–25, Saint-Malo, France, Sep 2004.

[186] C. Lenglet, M. Rousson, R. Deriche, and 0. Faugeras. Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor MRI processing. *Journal of Mathematical Imaging and Vision*, August 2006.

[187] C. Lenglet, M. Rousson, R. Deriche, and O. Faugeras. Statistics on multivariate normal distributions: A geometric approach and its application to diffusion tensor MRI. Technical Report 5242, INRIA, June 2004.

[188] C. Lenglet, M. Rousson, R. Deriche, O. Faugeras, S. Lehericy, and K. Ugurbil. A Riemannian approach to diffusion tensor images segmentation. In *Information Processing in Medical Imaging*, pages 591–602, 2005.

[189] R.J. LeVeque. *Numerical Methods for Conservation Laws*. Birkhäuser, Basel, 1992.

[190] T. Levi-Civita. Nozione di parallelismo in una varietà qualunque. *Rend. Circ. Mat. Palermo*, 42:173–205, 1917.

[191] T. Levi-Civita and G. Ricci. Méthodes de calcul différentiel absolu et leurs applications. *Math. Ann. B*, 54:125–201, 1900.

[192] C.P. Lin, V.J. Wedeen, J.H. Chen, C. Yao, and W.Y. I. Tseng. Validation of diffusion spectrum magnetic resonance imaging with manganese-enhanced rat optic tracts and ex vivo phantoms. *NeuroImage*, 19:482–495, 2003.

[193] P.-L. Lions. *Generalized Solutions of Hamilton–Jacobi Equations*. Number 69 in Research Notes in Mathematics. Pitman Advanced Publishing Program, 1982.

[194] X.D. Liu, S. Osher, and T. Chan. Weighted essentially non oscillatory schemes. *J. Comput. Phys.*, 115:200–212, 1994.

[195] M. Lovric, M. Min-Oo, and E.A. Ruh. Multivariate normal distributions parametrized as a Riemannian symmetric space. *Journal of Multivariate Analysis*, 74(1):36–48, 2000.

[196] D.C. Lyon and J.H. Kaas. Evidence for a modified V3 with dorsal and ventral halves in macaque monkeys. *Neuron*, 33:453–461, January 2002.

[197] D.C. Lyon and J.H. Kaas. Evidence from V1 connections for both dorsal and ventral subdivisions of V3 in three species of new world monkeys. *Journal of Comparative Neurology*, 449(3):281–297, 2002.

[198] X. Ma, S.M. LaConte, Y.M. Kadah, D.H. Frakes, A.P. Yoganathan, and X. Hu. Adaptive control grid interpolation of DTI data. In *International Society for Magnetic Resonance in Medicine*, page 1214, 2004.

[199] L. Madsen. *The geometry of statistical models*. PhD thesis, University of Copenhagen, 1978.

[200] P. Mahalanobis. On the generalized distance in statistics. In *Proceedings of the Indian National Institute Science*, volume 2, pages 49–55, 1936.

[201] J.-F. Mangin, C. Poupon, Y. Cointepas, D. Rivière, D. Papadopoulos-Orfanos, C. A. Clark, J. Régis, and D. Le Bihan. A framework based on spin glass models for the inference of anatomical connectivity from Diffusion-Weighted MR data. *NMR in Biomedicine*, 15:481–492, 2002.

[202] J.F. Mangin, C. Poupon, C. Clark, and I. Le Bihan, D.and Bloch. Distortion correction and robust tensor estimation for MR diffusion imaging. *Med Image Anal*, 6(3):191–8, September 2002.

[203] P. Mansfield. Multi-planar image formation using NMR spin echoes. *Journal of Physics C*, 10:55–58, 1977.

[204] C. Mantegazza and A.C. Mennucci. Hamilton-Jacobi equations and distance functions on Riemannian manifolds. *App. Math. and Optim.*, 47(1):1–25, 2002.

[205] K.D. Merboldt, W. Hanicke, and J. Frahm. Self-diffusion NMR imaging using stimulated echoes. *J. Magn. Reson.*, 64:479–486, 1985.

[206] C.A. Micchelli and L. Noakes. Rao distances. *Journal of Multivariate Analysis*, 92(1):97–115, 2005.

[207] A.D. Milner and M.A. Goodale. *The visual brain in action*. Oxford University Press, 1995.

[208] A. Mishra, Y. Lu, J. Meng, A.W. Anderson, and Z. Ding. Unified framework for anisotropic interpolation and smoothing of diffusion tensor images. *Neuroimage*, 31:1525–1535, 2006.

[209] R. Mneimné and F. Testard. *Introduction à la théorie des groupes de Lie classiques*. Hermann, 1986.

[210] M. Moakher. Means and averaging in the group of rotations. *SIAM Journal on Matrix Analysis and Applications*, 24(1):1–16, 2002.

[211] M. Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.*, 26(3):735–747, April 2005.

[212] S. Mori, B.J. Crain, V.P. Chacko, and P.C.M. Van Zijl. Three-dimensional tracking of axonal projections in the brain by Magnetic Resonance Imaging. *Annals of Neurology*, 45(2):265–269, February 1999.

[213] S. Mori and P. van Zijl. Fiber tracking: Principles and strategies - a technical review. *Nuclear Magnetic Resonance in Biomedicine*, 15(7–8):468–480, 2003.

[214] M.E. Moseley, Y. Cohen, J. Mintorovitch, J. Kucharczyk, J. Tsuruda, P. Weinstein, and D. Norman. Evidence of anisotropic self-diffusion. *Radiology*, 176:439–445, 1990.

[215] I. Najfeld and T. Havel. Derivatives of the matrix exponential and their computation. Technical Report TR-33-94, Harvard Medical School, Department of Biological Chemistry and Molecular Pharmacology, November 1994.

[216] K.L. Narr, P.M. Thompson, T. Sharma, J. Moussai, A.F. Cannestra, and A.W. Toga. Mapping morphology of the corpus callosum in schizophrenia. *Cerebral Cortex*, 10(1):40–49, 2000.

[217] M. Niethammer, R. San-Jose Estepar, S. Bouix, M. Shenton, and C.F. Westin. On diffusion tensor estimation. In *28th IEEE EMBS, New York City, NY, USA*, pages 343–350, 2006.

[218] L. O'Donnell, S. Haker, and C.F. Westin. New approaches to estimation of white matter connectivity in Diffusion Tensor MRI: Elliptic PDEs and geodesics in a tensor-warped space. In *MICCAI*, 2002. 459–466.

[219] L. O'Donnell, M. Kubicki, M.E. Shenton, M. Dreusicke, W.E.L. Grimson, and C.F. Westin. A method for clustering white matter fiber tracts. *American Journal of Neuroradiology*, 27:1032–1036, 2006.

[220] S. Ogawa, T.M. Lee, A.R. Kay, and D.W. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, 87(24):9868–9872, December 1990.

[221] J.M. Oller and C.M. Cuadras. Rao's distance for negative multinomial distributions. *Sankhya: The Indian Journal of Stats.*, 47:75–83, 1985.

[222] S. Osher. A level set formulation for the solution of the Dirichlet problem for a Hamilton-Jacobi equations. *SIAM Journal on Mathematical Analysis*, 24(5):1145–1152, 1993.

[223] S. Osher and J.A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton–Jacobi formulations. *Journal of Computational Physics*, 79(1):12–49, 1988.

[224] P.A. Osment, K.J. Packer, M.J. Taylor, J. J. Attard, T. A. Carpenter, L. D. Hall, S. J. Doran, and N. J. Herrod. NMR imaging of fluids in porous solids. *Phil. Trans. Roy. Soc.*, 333:441–452, 1990.

[225] E. Ozarslan, B. C. Vemuri, and T. Mareci. Fiber orientation mapping using generalized diffusion tensor imaging. In *ISBI*, pages 1036–1039, 2004.

[226] S. Pajevic, A. Aldroubi, and P.J. Basser. A continuous tensor field approximation of discrete DT-MRI data for extracting microstructural and architectural features of tissue. *Journal of Magnetic Resonance*, 154:85–100, 2002.

[227] N.G. Papadakis, D. Xing, C.L. Huang, L.D. Hall, and T.A. Carpenter. A comparative study of acquisition schemes for diffusion tensor imaging using MRI. *Journal of Magnetic Resonance*, 137:67–82, 1999.

[228] N. Paragios and R. Deriche. Geodesic active regions: a new paradigm to deal with frame partition problems in computer vision. *Journal of Visual Communication and Image Representation, Special Issue on Partial Differential Equations in Image Processing, Computer Vision and Computer Graphics*, 13(1/2):249–268, march/june 2002.

[229] H.J. Park, M. Kubicki, M.E. Shenton, A. Guimond, R.W. McCarley, S.E. Maier, R. Kikinis, F.A. Jolesz, and C.F. Westin. Spatial normalization of diffusion tensor MRI using multiple channels. *Neuroimage*, 20(4):1995–2009, 2003.

[230] 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. *Journal of Magnetic Resonance Imaging*, 18(2):242–254, August 2003.

[231] G.J.M. Parker. Tracing fibers tracts using Fast Marching. In *Proceedings of the International Society of Magnetic Resonance*, volume 85, 2000.

[232] G.J.M. Parker and D.C Alexander. Probabilistic Monte Carlo based mapping of cerebral connections utilising whole-brain crossing fibre information. In *IPMI*, pages 684–695, 2003.

[233] G.J.M. Parker, K.E. Stephan, G.J. Barker, J.B. Rowe, D.G. MacManus, C.A.M. Wheeler-Kingshott, O. Ciccarelli, R.E. Passingham, R.L. Spinks, R.N. Lemon, and R. Turner. Initial demonstration of in vivo tracing of axonal projections in the macaque brain and comparison with the human brain using Diffusion Tensor Imaging and Fast Marching Tractography. *NeuroImage*, 15(4):797–809, April 2002.

[234] G.J.M. Parker, C.A.M. Wheeler-Kingshott, and G.J. Barker. Estimating distributed anatomical connectivity using Fast Marching Methods and Diffusion Tensor Imaging. *Transactions in Medical Imaging*, 21(5):505–512, 2002.

[235] X. Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, July 2006.

[236] X. Pennec. Left-invariant Riemannian elasticity: A distance on shape diffeomorphisms? In *Proceedings of International Workshop on the Mathematical Foundations of Computational Anatomy*, pages 1–13, 2006.

[237] X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. Technical Report 5255, INRIA, Sophia Antipolis, July 2004.

[238] X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, 2006.

[239] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, July 1990.

[240] E. Pichon, G. Sapiro, and A. Tannenbaum. *Directions in Mathematical Systems Theory and Optimization*, volume 286 of *Lecture Notes in Control and Information Sciences*, chapter Segmentation of Diffusion Tensor Imagery, pages 239–247. Springer-Verlag Heidelberg, 2003.

[241] E. Pichon, C.F. Westin, and A. A. Tannenbaum. A hamilton-jacobi-bellman approach to high angular resolution diffusion tractography. In *Proceedings of MICCAI'05*, pages 180–187, 2005.

[242] C. Pierpaoli and P.J. Basser. Toward a quantitative assessment of diffusion anisotropy. *Magnetic Resonance in Medicine*, 36:893–906, 1996.

[243] C. Pierpaoli, P. Jezzard, P.J. Basser, A. Barnett, and G. Di Chiro. Diffusion Tensor MR imaging of human brain. *Radiology*, 201:637–648, 1996.

[244] E. Prados. *Application of the theory of the viscosity solutions to the Shape From Shading problem*. PhD thesis, University of Nice-Sophia Antipolis, 2004.

[245] E. Prados, O. Faugeras, and E. Rouy. Shape from shading and viscosity solutions. Technical Report 4638, INRIA, November 2002.

[246] E. Prados, C. Lenglet, J.P. Pons, N. Wotawa, R. Deriche, O. Faugeras, and S. Soatto. Control theory and fast marching methods for brain connectivity mapping. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1076–1083, New York, NY, June 2006.

[247] E. Prados, C. Lenglet, J.P. Pons, N. Wotawa, R. Deriche, O. Faugeras, and S. Soatto. Control theory and fast marching techniques for brain connectivity mapping. Technical Report RR-5845, INRIA Research Report, Sophia-Antipolis, 2006.

[248] E. Prados and S. Soatto. Fast marching method for generic shape from shading. In *VLSM'05*, pages 320–331, 2005.

[249] Emmanuel Prados, Fabio Camilli, and Olivier Faugeras. A viscosity solution method for shape-from-shading without image boundary data. *Mathematical Modelling and Numerical Analysis (M2AN)*, 40(2):393–412, 2006.

[250] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, 1988.

[251] K. Pribam and P. MacLean. Neuronographic analysis of medial and basal cerebral cortex. *J. of Neurophysiology*, 16:324–340, 1953.

[252] E.M. Purcell, H.C. Torrey, and R.V. Pound. Resonance absorption by nuclear magnetic moments in a solid. *Physical Review*, 69:37–38, 1946.

[253] C.R. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*, 37:81–91, 1945.

[254] S.M. Rao, P.A. Bandettini, J.R. Binder, J.A. Bobholz, T.A. Hammeke, E.A. Stein, and J.S. Hyde. Relationship between finger movement rate and functional magnetic resonance signal change in human primary motor cortex. *J. Cereb. Blood. Flow. Metab.*, 16:1250–1254, 1996.

[255] G.K. Rohde, S. Pajevic, and C. Pierpaoli. Multi-channel registration of diffusion tensor images using directional information. In *IEEE International Symposium on Biomedical Imaging*, pages 712–715, Washington (DC), USA, April 15–18 2004.

[256] R. Romo and W. Schultz. Role of primate basal ganglia and frontal cortex in the internal generation of movements. iii. neuronal activity in the supplementary motor area. *Exp. Brain. Res.*, 91:396–407, 1992.

[257] SE. Rose, F. Chen, J.B. Chalk, F.O. Zelaya, W.E. Strugnell, M. Benson, J. Semple, and D.M. Doddrell. Loss of connectivity in Alzheimer's disease: an evaluation of white matter tract integrity with colour coded MR diffusion tensor imaging. *J. Neurol. Neurosurg. Psychiatry.*, 69(4):528–530, 2002.

[258] M. Rousson. *Cues Integrations and Front Evolutions in Image Segmentation*. PhD thesis, Université de Nice-Sophia Antipolis, December 2004.

[259] M. Rousson and R. Deriche. A variational framework for active and adaptative segmentation of vector valued images. In *Proc. IEEE Workshop on Motion and Video Computing*, pages 56–62, Orlando, Florida, December 2002.

[260] M. Rousson, C. Lenglet, and R. Deriche. Level set and region based surface propagation for diffusion tensor MRI segmentation. In *Computer Vision Approaches to Medical Image Analysis (CVAMIA) and Mathematical Methods in Biomedical Image Analysis (MMBIA) Workshop*, pages 123–134, Prague, May 2004.

[261] E. Rouy and A. Tourin. A Viscosity Solutions Approach to Shape-from-Shading. *SIAM Journal of Numerical Analysis*, 29(3):867–884, June 1992.

[262] J. Ruiz-Alzola, C.F. Westin, S.K. Warfield, C. Alberola, S.E. Maier, and R. Kikinis. Nonrigid registration of 3d tensor medical data. *Medical Image Analysis*, 6(2):143–161, 2002.

[263] J. Ruiz-Alzola, C.F. Westin, S.K. Warfield, A. Nabavi, and R. Kikinis. Nonrigid registration of 3d scalar, vector and tensor medical data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 541–550, 2000.

[264] C. Scherfler, M.F. Schocke, K. Seppi, R. Esterhammer, C. Brenneis, W. Jaschke, G.K. Wenning, and W. Poewe. Voxel-wise analysis of diffusion weighted imaging reveals disruption of the olfactory tract in parkinson's disease. *Brain*, 129(2):538–542, 2006.

[265] A. Schwartzman. *Random ellipsoids and false discovery rates; statistics for diffusion tensor imaging data*. PhD thesis, Stanford University, 2006.

[266] N.R. Selden, D.R. Gitelman, N. Salamon-Murayama, T.B. Parrish, and M.M. Mesulam. Trajectories of cholinergic pathways within the cerebral hemispheres of the human brain. *Brain*, 121:2249–2257, 1998.

[267] J.A. Sethian. A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences*, 93(4):1591–1694, 1996.

[268] J.A. Sethian. *Level Set Methods*. Cambridge University Press, 1996.

[269] J.A. Sethian and A. Vladimirsky. Ordered upwind methods for static Hamilton–Jacobi equations: Theory and algorithms. *SIAM Journal on Numerical Analysis*, 41(1):325–363, 2003.

[270] Raimundo Sierra. Nonrigid registration of diffusion tensor images. Master's thesis, Swiss Federal Institute of Technology (ETHZ), March 2001.

[271] L.T. Skovgaard. A Riemannian geometry of the multivariate normal model. Technical Report 81/3, Statistical Research Unit, Danish Medical Research Council, Danish Social Science Research Council, 1981.

[272] L.T. Skovgaard. A Riemannian geometry of the multivariate normal model. *Scandinavian Journal of Statistics*, 11:211–233, 1984.

[273] H. M. Soner. Optimal control with state space constraints. *SIAM J. Contr. Optim*, 24:Part I: 552–562, Part II: 1110–1122, 1986.

[274] C. Sotak. The role of diffusion tensor imaging (DTI) in the evaluation of ischemic brain injury. *NMR Biomed.*, 15:561–569, 2002.

[275] E.O. Stejskal and J.E. Tanner. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *Journal of Chemical Physics*, 42:288–292, 1965.

[276] E.V. Sullivan and A. Pfefferbaum. Diffusion tensor imaging in normal aging and neuropsychiatric disorders. *Eur. J. Radiol.*, 45(3):244–255, 2003.

[277] I-F Talos, L. O'Donnell, C-F Westin, S. Warfield, W. Wells III, S-S Yoo, L. Panych, A. Golby, H. Mamata, S. Maier, P. Ratiu, C. Guttmann, P. Black, F. Jolesz, and R. Kikinis. Diffusion tensor and functional MRI fusion with anatomical MRI for image-guided neurosurgery. In *Sixth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'03)*, 2003. 407-415.

[278] D.G. Taylor and M.C. Bushell. The spatial mapping of translational diffusion coefficients by the NMR imaging technique. *Phys. Med. Bio.*, 30:345–349, 1985.

[279] I. Toni, M. Krams, R. Turner, and R.E. Passingham. The time course of changes during motor sequence learning: a whole-brain fMRI study. *Neuroimage*, 8:50–61, 1998.

[280] A. Toosy, O. Ciccarelli, et al. Characterising function-structure relationships in the human visual system with functional MRI and Diffusion Tensor Imaging. *NeuroImage*, 21:1452–1463, 2004.

[281] J.-D. Tournier, F. Calamante, D.G. Gadian, and A. Connelly. Diffusion-weighted magnetic resonance imaging fibre tracking using a front evolution algorithm. *NeuroImage*, 20:276–288, 2003.

[282] Y.-H. Tsai, L.-T. Cheng, S. Osher, and H.-K. Zhao. Fast sweeping algorithms for a class of Hamilton-Jacobi equations. *SIAM J. Numer. Anal.*, 41(2):673–694, 2003.

[283] Y-H. Tsai, Y. Giga, and S. Osher. A level set approach for computing discontinuous solutions of hamilton-jacobi equations. *Math. Comput.*, 72(241):159–181, 2003.

[284] D. Tschumperlé and R. Deriche. DT-MRI images : Estimation, regularization and application. In J.C. Rodriguez R. Morena-Diaz Jr, A. Quesada-Arencibia, editor, *EUROCAST'2003*, 9th International Workshop on Computer Aided Systems Theory, pages 530–541, Las Palmas de Gran Canaria, 2003.

[285] D. Tschumperlé and R. Deriche. Variational frameworks for DT-MRI estimation, regularization and visualization. In *Proceedings of the 9th International Conference on Computer Vision*, pages 116–121, Nice, France, 2003. IEEE Computer Society, IEEE Computer Society Press.

[286] J.N. Tsitsiklis. Efficient algorithms for globally optimal trajectories. *IEEE Trans. Aut. Control*, 40:1528–1538, 1995.

[287] D. Tuch. Q-ball imaging. *Magnetic Resonance in Medicine*, 52(6):1358–1372, 2004.

[288] D.S. Tuch. Mapping cortical connectivity with diffusion MRI. In *ISBI*, pages 392–394, 2002.

[289] D.S. Tuch, M.R. Wiegell, T.G. Reese, J.W. Belliveau, and V.J. Weeden. Measuring cortico-cortical connectivity matrices with diffusion spectrum imaging. In *Int. Soc. Magn. Reson. Med.*, volume 9, page 502, 2001.

[290] D.S. Tuch, J.J. Wisco, M.H. Khachaturian, L.B. Ekstrom, R. Kotter, and W. Vanduffel. Q-ball imaging of macaque white matter architecture. *Philosophical Transactions of the Royal Society B*, 360:869–879, May 2005.

[291] C.W. Tyler, L.T. Likova, L.L. Kontsevich, M.M. Schira, and A.R. Wade. Enhanced concepts of occipital retinotopy. *Current Medical Imaging Reviews*, 1:319–329, 2005.

[292] A.M. Ulug. Monitoring brain development with quantitative diffusion tensor imaging. *Developmental Science*, 5(3):286–292, 2002.

[293] L.G. Ungerleider and M. Mishkin. *Two cortical visual systems.*, pages 549–586. MIT Press, 1982.

[294] D.C. Van Essen. A tension-based theory of morphogenesis and compact wiring in the central nervous system. *Nature*, 385:313–318, January 1997.

[295] D.C. Van Essen, H.A. Drury, J. Dickson, J. Harwell, D. Hanlon, and C.H. Anderson. An integrated software suite for surface-based analyses of cerebral cortex. *Journal of American Medical Informatics Association*, 8:443–459, 2001.

[296] D.C. Van Essen, J.H.R. Maunsell, and J.L. Bixby. The Middle Temporal visual area in the macaque: myeloarchitecture, connections, functional properties and topographic organization. *Journal of Comparative Neurology*, 199:293–326, 1981.

[297] B. Vemuri, Y. Chen, M. Rao, T. McGraw, T. Mareci, and Z. Wang. Fiber tract mapping from diffusion tensor MRI. In *1st IEEE Workshop on Variational and Level Set Methods in Computer Vision (VLSM'01)*, page 81, July 2001.

[298] R. Venkataraman, R. Holsapple, and D. Doman. Optimal control problems on Riemannian manifolds: Theory and applications. Technical Report A433804, Air force research lab Wright-Patterson AFB OH air vehicles directorate, 2002.

[299] A. Virta, A. Barnett, and C. Pierpaoli. Visualizing and characterizing white matter fiber structure and architecture in the human pyramidal tract using diffusion tensor MRI. *Magnetic Resonance Imaging*, 17:1121–1133, 1999.

[300] S. Wakana, H. Jiang, L.M. Nagae-Poetscher, P.C. van Zijl, and S. Mori. Fiber tract-based atlas of human white matter anatomy. *Radiology*, 230(1):77–87, 2004.

[301] D. Waldvogel, P. van Gelderen, K. Ishii, and M. Hallett. The effect of movement amplitude on activation in functional magnetic resonance imaging studies. *J. Cereb. Blood. Flow. Metab.*, 19:1209–1212, 1999.

[302] B.A. Wandell, A.A. Brewer, and R.F. Dougherty. Visual field map clusters in human cortex. *Phil. Trans. of the Royal Society London*, 360:693–707, 2005.

[303] Z. Wang and B.C. Vemuri. An affine invariant tensor dissimilarity measure and its application to tensor-valued image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 228–233, Washington, DC., June 2004.

[304] Z. Wang and B.C. Vemuri. Tensor field segmentation using region based active contour model. In *ECCV'2004*, pages 304–315, Prague, Czech Republic, May 2004.

[305] Z. Wang, B.C. Vemuri, Y. Chen, and T.H. Mareci. A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex DWI. *IEEE Transactions on Medical Imaging*, 23(8):930–939, 2004.

[306] J. Watson, R. Myers, R. Frackowiak, J. Hajnal, R. Woods, J. Mazziotta, S. Shipp, and S. Zeki. Area V5 of the human brain: evidence from a combined study using Positron Emission Tomography and Magnetic Resonance Imaging. *Cerebral Cortex*, 3:79–94, 1993.

[307] J. Weickert. *Anisotropic Diffusion in Image Processing*. Teubner-Verlag, Stuttgart, 1998.

[308] J. Weickert and T. Brox. Diffusion and regularization of vector and matrix-valued images. Technical report, Universitat des Saarlandes, 2002.

[309] D.J. Werring, C.A. Clark, G.J.M. Parker, D.H. Miller, A.J. Thompson, and G.J. Barker. A direct demonstration of both structure and function in the visual system: combining Diffusion Tensor Imaging with functional Magnetic Resonance Imaging. *NeuroImage*, 9:352–361, 1999.

[310] C.F Westin, S.E Maier, H. Mamata, A. Nabavi, F.A. Jolesz, and R. Kikinis. Processing and visualization of diffusion tensor MRI. In *In proceedings of Medical Image Analysis*, volume 6(2), pages 93–108, 2002.

[311] C.F. Westin, M. Martin-Fernandez, C. Alberola-Lopez, J. Ruiz-Alzola, and H. Knutsson. Tensor field regularization using normalized convolution and Markov random fields in a Bayesian framework. In J. Weickert and H. Hagen, editors, *Visualization and Image Processing of Tensor Fields. Series: Mathematics and Visualization*, pages 381–398,464–467. Springer, 2006. ISBN:3-540-25032-8.

[312] M.R. Wiegell, D.S. Tuch, H.W.B. Larson, and V.J. Wedeen. Automatic segmentation of thalamic nuclei from diffusion tensor magnetic resonance imaging. *NeuroImage*, 19:391–402, 2003.

[313] N. Wotawa, J.-P. Pons, L. Lopez, R. Deriche, and O. Faugeras. fMRI data smoothing constrained to the cortical surface: a comparison of the level-set and mesh-based approaches. In *NeuroImage (HBM'04)*, 2004.

[314] N. Wotawa, B. Thirion, E. Castet, J.L. Anton, and O. Faugeras. Human retinotopic mapping using fMRI. Technical Report 5472, INRIA, January 2005.

[315] L. Yatziv, A. Bartesaghi, and G. Sapiro. O(N) implementation of the fast marching algorithm. *Journal of computational physics*, 212:393–399, 2006.

[316] S. Yoshizawa and K. Tanabe. Dual differential geometry associated with the Kullback-Leibler information on the Gaussian distributions and its 2-parameter deformations. *Science University of Tokyo Journal of Mathematics*, 35:113–137, 1999.

[317] M.P. Young, G. Burns, and J.W. Scannell. *The Analysis of Cortical Connectivity*. Landes Bioscience, October 1995.

[318] E. Yörük, B. Acar, and R. Bammer. A physical model for DT-MRI based connectivity map computation. In *Proceedings of MICCAI'05*, pages 213–220, 2005.

[319] H. Zhang, P.A. Yushkevich, and J.C. Gee. Towards diffusion profile image registration. In *IEEE International Symposium on Biomedical Imaging*, pages 324–327, Washington (DC), USA, April 15–18 2004.

[320] H. Zhang, P.A. Yushkevich, and J.C. Gee. Deformable registration of diffusion tensor MR images with explicit orientation optimization. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 172–179, 2005.

[321] Hui Zhang, Paul A. Yushkevich, and James C. Gee. Registration of diffusion tensor images. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 842–847, 2004.

[322] Y. Zhang, M. Brady, and S. Smith. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1), 2001.

[323] Zhengyou Zhang. Parameter estimation techniques: A tutorial with application to conic fitting. *Image and Vision Computing Journal*, 15(1):59–76, 1997.

[324] H.-K. Zhao, T. Chan, B. Merriman, and S. Osher. A variational level set approach to multiphase motion. *Journal of Computational Physics*, 127(1):179–195, 1996.

[325] S. Zhu and A. Yuille. Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):884–900, September 1996.

[326] L. Zhukov, K. Museth, D. Breen, R. Whitaker, and A.H. Barr. Level set segmentation and modeling of DT-MRI human brain data. *Journal of Electronic Imaging*, 12:125–133, 2003.