University of Liège
Faculty of Applied Sciences
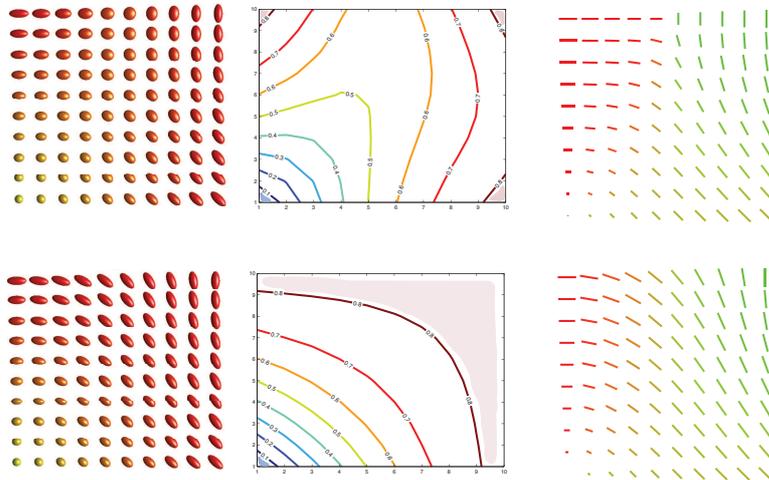Department of Electrical Engineering and Computer Science

Geometric statistical processing
of brain diffusion tensor images

PhD thesis of **Anne Collard**

Advisor: Prof. Rodolphe Sepulchre

Co-advisor: Dr. Christophe Phillips

July 2013

Jury members:

P. Maquet	(Chair)	Professor, University of Liège
R. Sepulchre	(Advisor)	Professor, University of Liège
C. Phillips	(Co-advisor)	Associate Professor, University of Liège
P-A. Absil		Associate Professor, Université Catholique de Louvain
D. C. Alexander		Professor, University College London (UK)
M. Van Droogenbroeck		Professor, University of Liège
K. Van Steen		Associate Professor, University of Liège

Abstract

Nowadays, the functioning of the human brain is still one of the biggest mysteries of humanity. The multiple holes in the understanding of the human brain explain why an intensification of brain-oriented research can be observed since a few years.

One of the most recent techniques to better understand the brain is Diffusion Tensor Imaging (DTI), a noninvasive imaging modality that provides information about orientation of nervous fibers, and their spatial density, with a high resolution. The particular nature of DTI images makes them multi-valued. Their processing therefore requires to adapt state-of-the-art techniques, which are fundamentally tailored to scalar-valued images.

The objective of this PhD thesis is to develop a novel framework for the processing of tensor diffusion images. The focus is threefold: first, we adopt a Riemannian geometric framework to generalize image processing from linear to nonlinear spaces. Second, we aim at developing a processing framework that retains the physical information of measurement data. Thirdly, the proposed algorithm must be computationally efficient in order to scale with the data size of clinical applications.

The main contribution of this thesis is the development of a novel processing method, which has the particularity to preserve the important features of diffusion tensors, while being computationally affordable. This technique is based on the decoupling between the two types of information conveyed by tensors: the diffusion intensity on one hand, and the orientation of diffusion on the other hand. Moreover, the computational cost is limited thanks to the use of unit quaternions to represent tensors orientation.

Another contribution of the thesis lies in the development of a statistical method for group comparison. This method uses the notion of similarity measure between the values, a notion that can be defined for multi-valued images, and which enables to reduce the computational cost. The use, for the statistical tests, of the similarity measure associated to our framework turns out to be efficient and informative.

The study of geometric methods for multi-valued images together with the study of potential applications of diffusion tensor images have enabled the introduction of a novel framework, which is particularly appropriate for those images. The basic operations developed in the thesis open the way to more sophisticated processing algorithms, while ensuring the preservation of the main information associated to the tensors.

Résumé

Le fonctionnement du cerveau humain reste de nos jours un des plus grands mystères de l'humanité. Les nombreuses zones d'ombre dans la compréhension du système nerveux expliquent pourquoi, depuis quelques années, on assiste à une intensification des recherches concernant le cerveau.

Une des plus récentes techniques pour mieux appréhender le cerveau est l'Imagerie par Tenseur de Diffusion (ITD), une technique non invasive qui fournit des informations sur l'orientation des fibres nerveuses et sur leur densité spatiale, avec une haute résolution. La nature particulière des images obtenues par ITD en fait des images à valeurs multiples. Leur traitement requiert donc l'adaptation des techniques de l'état de l'art, qui sont fondamentalement conçues pour les images à valeurs scalaires.

L'objectif de cette thèse de doctorat est de développer un nouvel ensemble de techniques pour le traitement des images par tenseur de diffusion. Cet objectif est triple : premièrement, nous adoptons un ensemble d'outils géométriques et Riemannien afin de généraliser le traitement d'images dans des espaces linéaires vers les espaces non-linéaires. Deuxièmement, nous cherchons à développer un système de traitement qui retienne les informations physiques des données mesurées. Troisièmement, l'algorithme proposé doit être efficace au niveau coût de calcul, de façon à être compatible avec la taille des données des applications cliniques.

La contribution principale de la thèse est le développement d'une technique de traitement des images qui a la particularité de préserver les caractéristiques principales des tenseurs de diffusion, tout en étant utilisable à des coûts de calculs raisonnables. Cette technique est basée sur le traitement découplé des deux types d'informations associés aux tenseurs: l'intensité de la diffusion d'une part, et l'orientation de la diffusion d'autre part. Le coût de calcul est par ailleurs limité grâce à l'utilisation des quaternions unitaires pour représenter l'orientation des tenseurs.

Une autre contribution de la thèse réside dans le développement d'une méthode statistique de comparaison de groupes. La méthode utilise la notion de distance entre les éléments, notion qui peut être définie pour les images non-linéaires, et qui permet de limiter les coûts de calcul. L'utilisation de la mesure de similarité fournie par notre méthode de traitement d'images pour les tests statistiques s'avère être efficace et informative.

La combinaison de l'étude des méthodes géométriques pour les images à valeurs multiples et de l'étude des applications potentielles des images par tenseur de diffusion a permis la création d'un nouvel ensemble d'outils particulièrement appropriés à ces images. Les opérations de base développées dans cette thèse ouvrent la voie aux méthodes d'analyse plus sophistiquées, tout en assurant la préservation des informations principales associées aux tenseurs.

Acknowledgements

Completing this thesis has been a wonderful and enriching experience, which has impacted my whole life. There are several people that I would like to thank for their help and support during the last four years. They have to know that it would not have been possible without them.

First and foremost, I would like to heartily thank my advisor, Rodolphe Sepulchre, who has offered me the wonderful opportunity to discover the research world. Rodolphe has played a multitude of roles for me during those four years, constantly juggling between a helpful guidance and a watchful eye. In many occasions, his invaluable advices together with his unwavering and catching enthusiasm have been a real source of motivation and optimism for me, helping me to go through various obstacles and always 'pushing' me further.
I am also warmly thankful to my co-advisor, Christophe Phillips, for his always prompt and pertinent answers to my many questions. In my opinion, Christophe has perfectly played the role of a co-advisor, offering me new paths for research and helping me with technical details when I needed it.
My two advisors together have offered me the opportunity to explore different sides of research, and I would like one more time to thank them for this.

I would like to mention with gratitude Silvère Bonnabel, who has directly contributed to this work. Moreover, I would like to thank him for his warm welcoming during my visits to Paris. I am also grateful to Pierre-Antoine Absil, whose questions and answers 'to the point' have always been helpful.

It is hard for people who have never experienced the world of research to clearly understand its different aspects. Luckily, I was not alone during those four years of research, and I would like to thank my colleagues at the Montefiore Institute for our constant sharing about our issues, accomplishments, or daily life. In particular, my thoughts go to Laura Trotta, my 'sister in research'. We have come through this together, and it would have been much less fun without her. I am heartily thankful to the two people with whom I have shared my office, Alexandre Mauroy and Raphaël Liégeois. Both of them are truly fascinated and fascinating guys, and I have enjoyed all our (sometimes completely crazy) conversations. Many thanks also to Pierre Sacré, for his promptness to offer help and his deadpan sense of humour; to Guillaume Drion, for the sharing of his funny videos; to Gilles Meyer, for his technical support and his auspicious enthusiasm; to Alain Sarlette, for showing us that it is possible to be hard working and full of humour at the same time; to Bamdev Mishra for his interesting and enjoyable company; to Julie Dethier, for his catching optimism and for our swimming lunch times; to Fulvio Forni, for the prompt

answer to 'distress calls' and his sparkling humour, and finally to our Italian guys, Alessio Franci, Mattia Zorzi and Enrico Amico, for showing us that the only one coffee is espresso. I am particularly proud to be a member of the 'biomed' group, and I would like to thank Sabine Paeme, Jessica Schrouff, Marie Schrynemackers, Laura Symul and Mélinda Choua, with whom I have always enjoyed to talk about work and many other 'girl stuffs'.
I also acknowledge Guy Lejeune, Etienne Michel, Matthieu Philippe, Gauthier Becker and Denis de Marneffe, for the enjoyable 'boulet' lunch times.

I would also like to express my genuine appreciation to all the members of the jury, Pierre-Antoine Absil, Daniel Alexander, Pierre Maquet, Christophe Phillips, Rodolphe Sepulchre, Marc Van Droogenbroeck and Kristel Van Steen, for devoting time and interest to the reading and evaluation of this manuscript.

Finally, some words in French for my family and friends.
Si les membres de ma famille n'ont pas toujours compris en quoi consistait exactement mon travail, ils n'ont jamais arrêté de me soutenir et de m'exprimer leurs encouragements. Un merci particulier à mes parents, qui m'ont permis de rester dans ce no man's land entre le statut d'étudiante et celui de 'travailleuse', à ma Mamy pour les excellents petits plats qu'elle m'a préparés chaque mardi, à Quentin pour m'avoir obligée à garder les pieds sur terre, à ma filleule chérie Emma, ainsi qu'à Jeanne et Gabin, pour leurs sourires qui ne manquent jamais de faire revenir le mien.

J'ai l'immense chance d'avoir une bande d'amies sur lesquelles je peux toujours compter. Merci donc à Anne-Sophie, pour ses précieux conseils et son amitié sans failles, à Auré, Béné, Elo et Nad pour les innombrables fous rires et merveilleux moments dont elles seules ont le secret (vive le PFC !), ainsi qu'à Marion, Solange et Laurence, pour ces magnifiques moments partagés.

Je ne saurais terminer sans remercier Stéphane, qui a été à de multiples reprises l'homme de la situation. Merci pour ça, merci pour tout, merci d'être toi.

Contents

Chapter 1

Introduction

> *Now, as humans, we can identify galaxies light years away.*
> *We can study particles smaller than an atom,*
> *but we still haven't unlocked the mystery*
> *of the three pounds of matter that sits between our ears.*

- Barack Obama, April 2013.

In April 2013, Barack Obama, President of the United States of America, announced his BRAIN (Brain Research through Advancing Innovative Neurotechnologies) initiative[1]. The goal of this initiative, often compared to the Human Genome Project, is to 'give to the scientists the tools they need to a get a dynamic picture of the brain and better understand how we think, learn, and remember'.

Some weeks before, the European FET (Future and Emerging Technologies) Flagship Programme had announced that it will support the *Human Brain Project* of the Ecole Polytechnique Fédérale de Lausanne[2]. The 86 institutions involved in this project will receive one billion euro in funding over ten years. The Human Brain Project aims to simulate the human brain with supercomputers. The goal is to better understand how it functions, and eventually to be able to mimic it. This should help to better diagnose different brain problems.

The attention that these two 'grand challenges' caught in the media provides a suggestive evidence that brain research is 'the place to be' for coming years.

Among the multitude of tools which will be developed or improved in the following years, there is no doubt that imaging methods will occupy a central place. All over the world, researchers develop novel methods to image the brain non invasively. It is in this particularly blossoming context that this dissertation takes place, as following sections will illustrate it.

---

[1]A summary infographic of this initiative can be found at www.whitehouse.gov/infographics/brain-initiative.

[2]http://www.humanbrainproject.eu

1.1 Research Context

The research presented in this thesis has the particularity of being rooted in two distinct fields. These two young fields have each attracted a growing interest over the last years, and the connection between them is recent. In the sequel, a brief presentation of these fields will enable to relate the present work to recent research development, and to motivate the path that our research has followed.

Diffusion Tensor Imaging

The method of Diffusion Tensor Imaging (DTI) is a relatively recent medical imaging method. The seminal paper which introduces the very concept of DTI was published in 1992 (Basser et al., 1992, 1994a). This method has attracted a lot of interest, because of both its potential applications and the very particular nature of the images themselves. It is often said that DTI enables to visualize, or reconstruct, the architecture of the brain, without having to open it. This wonderful ability easily explains why researchers have worked a lot to improve this imaging method. To understand how Diffusion Tensor Imaging works and what it shows, imagine a (terribly intricate) maze, in which many people walk freely. Suppose that you want to know where the walls of this maze are, and how the people go from one point to another. One solution is to put a GPS chip on every people, and record their displacements. If the record lasts long enough, you should be able to distinguish between the aisles of the maze (where many people walk through) and the walls of it (where no one goes through). The principle of DTI is the same: the brain is a maze, whose walls are the boundaries of axons. In this maze, water molecules move freely. The equivalent of the GPS chips here is the proton in water molecules. DTI uses different magnetic fields to 'label' those protons, and follow their displacements. The 'production' of those images has of course been the topic of much research, as it will be shown in Chapter 3. In this thesis, we focus on the images themselves.

In practice, the images obtained by Diffusion Tensor Imaging are matrix-valued: they associate a *diffusion tensor* to each of their pixels. This diffusion tensor, which can be visualized as an ellipsoid, represents the way the water move in each voxel of the brain. The resolution is high, with voxel size ranging from 5 to 1.5 mm. If the ellipsoid is spherical, it means that the water does not encounter any obstacle in this part of the brain. This is usually the case in gray matter, or in the cerebro-spinal fluid. To the contrary, if the ellipsoid is elongated, this indicates that water moves in a favorite direction. Water movement is restricted by the presence of 'walls', the boundaries of axons. This part of the brain is then viewed as white matter. The processing of diffusion tensor images thus implies to process all the information which is conveyed by the tensors: the direction of the movement and its intensity. In the early years after the development of the method, researchers usually processed those images by using only one kind of information, the shape of the tensors. The inherent 'loss of information' of this processing has however limited its use.

Mathematically, diffusion tensors are symmetric positive definite matrices, which among other things, do not live in a linear space, like ordinary matrices, but rather on a nonlinear space, nevertheless highly structured. This consideration is at the core of the connection between diffusion tensor imaging on one hand and processing on manifolds on the other hand. This connection has first been made in (Lenglet et al., 2006; Pennec et al., 2006; Fletcher and Joshi, 2007).

Optimization on manifolds

Optimization on manifolds relates to a multitude of algorithms which have the particularity to be defined over particular spaces, Riemannian manifolds. Those spaces can be viewed as sets of elements that satisfy some nonlinear constraints in a linear space (for instance, positive numbers are real numbers λ satisfying the inequality constraint $\lambda > 0$). The processing of those data implies to develop algorithms which preserve the constraints, in other words, data need to be geometrically processed in their own space. As an analogy, we can consider the field of musical acoustics. Musical sounds of course are sounds but they are more than that, and their particular properties enable to classify them as 'musical'. The writing and analysis of music should then be limited to those particular sounds. The signal processing of musical sounds should have musical sounds as outputs. These ideas can be translated to the field of Riemannian matrix manifolds: the processing of matrices belonging to a particular manifold should be constrained in this manifold.

Originally, the idea of treating optimization problems defined on Riemannian manifolds was presented in (Luenberger, 1972). Early work in the control community includes (Brockett, 1972, 1993), who focused on systems properties on a manifold state space. The monograph (Helmke and Moore, 1996) deals with optimization and dynamical systems from a differential geometric viewpoint. The monograph (Absil et al., 2008) presents recent advances in the field, focusing on algorithmic developments and addressing the problem of optimizing a smooth function on a manifold.

The concepts of optimization on manifolds apply to a broad variety of computational problems. For instance, (Ma et al., 2001) studies motion and structure recovery for computer vision, (Absil, 2003) deals with invariant subspace computation, (Journée, 2009) focuses on component analysis and analysis of gene expression data, (Sarlette, 2009) presents distributed consensus algorithms.

In this thesis, the manifold of interest is the manifold of symmetric positive definite matrices, which represent diffusion tensors. It seems indeed natural to demand that the processing of diffusion tensors results in diffusion tensors. As explained above, the 'manifold viewpoint' for the processing of diffusion tensor images has emerged in 2006. The principal tools which are needed to process those images are a similarity measure and a weighted mean of tensors. Those tools can be defined from the theory of Riemannian manifolds, as explained in Chapter 2 and Section 3.4.3.

Motivations of the thesis

The research presented in this dissertation comes at the intersection between Diffusion Tensor Imaging and optimization on manifolds. The first goal of the thesis was indeed to define and develop appropriate tools for the (nonlinear) processing and statistical analysis of those images.

Concerning the processing, the detailed study of the existing research in the field had highlighted three issues which can be viewed as the foundations of the research.

- Due to their nonlinear nature, image processing of diffusion tensors calls for a geometric framework.

- The two most basic processing operations are a similarity measure between tensors and an averaging (or interpolation) of tensors. These operations need to be computationally efficient, because they are widely and repetitively used.

- Existing Riemannian metrics suffer from an intrinsic limitation: they tend to degrade the anisotropy of tensors, *i.e.* their shape, through averaging. Since many analyses of diffusion tensor images are based on this information, it seems natural to favor algorithms which will preserve it.

The statistical analysis of Diffusion Tensor Images has attracted a lot of research in the recent years but is still largely under development. The study of the state-of-art has shown that most algorithms are based on single scalars, for instance the fractional anisotropy of tensors. One of our motivation was thus to propose analyses based on the whole information conveyed by the tensors, *i.e* intensity of diffusion as well as its orientation. Existing statistical analysis methods which use the whole tensor are often based on statistical models. However, the true distribution of tensors is not known. The research has thus focused on non-parametric methods, *i.e.* methods which do not need any assumptions about the distribution. Another motivation was to develop statistical analysis which can scale with the medical application needs.

1.2   Contributions

The answers to the various questions or objectives explained above have given rise to different contributions of our thesis to the topic of geometric processing of diffusion tensor images. These contributions are listed below.

- A first contribution of this thesis is the definition of a rank-preserving mean of N positive semi-definite matrices of fixed rank. This mean is related to an interpolating curve defined on the space of those matrices, which has the advantage to admit explicit formulas. This contribution highlighted the fact that interpolating curves should not necessarily be geodesics.

This contribution is developed in Section 2.3 and is the topic of a recent paper in *Linear Algebra and its Applications*, (Bonnabel et al., 2013).

Silvère Bonnabel, Anne Collard and Rodolphe Sepulchre. Rank-preserving geometric means of positive semi-definite matrices. *Linear Algebra and its Applications*, 438(8):3202-3216, 2013.

The following contributions concern the development of our novel geometric framework.

- The affine-invariant and Log-Euclidean frameworks are well-known in the field of processing of diffusion tensor images. However, as far as we know, this work is the first to highlight their important limitation regarding anisotropy. As explained in Section 3.4.3, both methods tend to degrade this information during the processing.

- To address this limitation, the method developed in this thesis has been specifically constructed to be *anisotropy preserving*.

- The thesis also introduces a novel anisotropy index, the *Hilbert Anisotropy* (see Section 4.3.4). The choice of this index is motivated by its intrinsic meaning on cones. Moreover, with the proposed framework, the Hilbert Anisotropy is linearly interpolated during interpolation between two tensors. The mean Hilbert anisotropy of N tensors is the arithmetic mean of the N Hilbert Anisotropies.

- In order to ensure the computational efficiency of the method, unit quaternions have been used to parametrize orientation of tensors. To the best of our knowledge, this parametrization had never been used for the processing of diffusion tensor images. Several illustrations of this thesis however suggest that the chordal mean of quaternions is an attractive substitute for the mean of rotation matrices.

- Besides the conservation of anisotropy and the attractive computational cost, a third important issue has been addressed in the research: the more a tensor is isotropic, the more its orientation is uncertain. This observation led us to propose a different interpolating curve between quaternions, depending upon the anisotropy of tensors. This notion has also been used to define a weighted term for the orientation component in the similarity measure.

- A particular attention has been devoted to the definition of a symmetric mean of N tensors. It is indeed crucial that the mean of a set of tensors is permutation invariant, that is, independent on the order in which the tensors are considered. A method has been proposed to meet this need, which is based on the appropriate selection of a reference tensor. The notion of information conveyed by the tensor has been used to perform this selection, see Section 4.4.1.

- Chapter 5 of this dissertation suggests that the developed framework can be used in all processing algorithms, and is competitive with the state-of-the-art framework.

The contributions above are presented in:

Anne Collard, Silvère Bonnabel, Christophe Phillips and Rodolphe Sepulchre. Anisotropy preserving DTI processing. Submitted in revised form to *International Journal of Computer Vision*, May 2013.

The 'statistical analysis' part of this thesis has led to the following contributions.

- Starting from the idea that the distribution of tensors is seldom known, this work focuses on permutation methods for the statistical comparison of groups. However, Section 6.1.1 explains how mean-based methods could be computationally expensive. A different method, based on dispersion of data, is thus introduced in Section 6.1.2. This method is novel for the analysis of diffusion tensor images.

- Chapter 6 also shows results of statistical analyses on synthetical data sets. Those results prove that the spectral-quaternion framework is efficient to compare diffusion tensor images.

- The weighting parameter introduced in the similarity measure associated to the spectral-quaternion framework (Equation (4.44)) provides a 'tunable' parameter which induces a tradeoff between the robustness and the sensitivity of the measure to a variation of orientation.

- A geometric parametrization is introduced in Equation (6.4), based on the spectral-quaternion framework. Section 6.2.3 shows that this framework is easily interpretable and that its use could represent a particularly useful input of information about the difference between diffusion tensor images.

The contributions above are presented in:

Anne Collard, Christophe Phillips and Rodolphe Sepulchre. Statistical tests for group comparison of manifold-valued data. Submitted to *52nd IEEE Conference on Decision and Control*, March 2013.

1.3   Outline

The dissertation is organized as follows.

Chapter 2 is devoted to the definition of interpolating curves, weighted means and similarity measures in two nonlinear spaces: the space of $n \times n$ symmetric positive definite matrices and the space of fixed-rank positive semi-definite matrices.

Chapter 3 provides a historical perspective on Diffusion Tensor imaging, emphasizing the physical and information properties of this particular modality.

Chapter 4 presents the core development of our novel framework for the processing of diffusion tensor images. Definitions and algorithms are provided for the interpolating curve between tensors, weighted mean of tensors and similarity measure.

Chapter 5 presents some simple image processing operations using our framework, and compare the results with state-of-the-art methods.

Finally, Chapter 6 introduces a novel method for the statistical analysis of diffusion tensor images, and evaluates the relevance of the proposed framework in that context.

"Water is the driving force of all nature."

"Realize that everything connects to everything else."

- Leonardo da Vinci.

Notations

The following notations for spaces and manifolds are used throughout the thesis.

$\mathbb{R}$	The set of real numbers.
$\mathbb{R}^+$	The set of positive real numbers.
$\mathbb{R}^n$	The set of n-dimensional real column vectors.
$\mathbb{R}^{n \times p}$	The set of $n \times p$ real matrices with n rows and p columns.
$\mathbb{R}_*^{n \times p}$	The set of $n \times p$ full rank real matrices with n rows and p columns.
$\mathcal{M}$	An arbitrary manifold.
S^1	The unit sphere in $\mathbb{R}^2$, the set of unit norm vectors in $\mathbb{R}^2$.
$\mathrm{Gl}(n)$	The general linear group of dimension n, the set of invertible $n \times n$ matrices.
$\mathrm{O}(n)$	The orthogonal group, the set of $n \times n$ orthogonal matrices.
$\mathrm{S}^+(n)$	The set of symmetric positive definite matrices of size n.
$\mathrm{S}^+(p, n)$	The set of symmetric positive semi-definite matrices of size n with rank p.
$\mathrm{V}_{n,p}$	The Stiefel manifold, the set of $n \times p$ orthonormal matrices.
$\mathrm{Gr}(p, n)$	The Grassmann manifold, the set of p-dimensional subspaces in $\mathbb{R}^n$.
$\mathrm{S}^+(3)$	The set of symmetric positive definite matrices of size 3, the group of diffusion tensors.
$\mathrm{S}_*^+(3)$	The set of symmetric positive definite matrices of size 3 with distinct eigenvalues.
$\mathrm{SO}(3)$	The special orthogonal group, the set of 3×3 rotation matrices.
$\mathrm{D}_+(3)$	The set of diagonal matrices of size 3×3, with positive elements.
$\mathbb{H}_1$	The set of unit quaternions.

Given an arbitrary matrix $\mathbf{A}$, we define the notations,

$\mathbf{A}_{i,j}$	The element at row i and column j in the matrix $\mathbf{A}$.
$\mathbf{A}^T$	The transpose of the matrix $\mathbf{A}$.
$\mathrm{Tr}(\mathbf{A})$	The trace of the square $n \times n$ matrix $\mathbf{A}$, $\mathrm{Tr}(\mathbf{A}) = \sum_{i=1}^n \mathbf{A}_{ii}$.
$\mathrm{Det}(\mathbf{A})$	The determinant of the square and invertible matrix $\mathbf{A}$.
$||\mathbf{A}||_2$	The Frobenius norm of the matrix $\mathbf{A}$, $||\mathbf{A}||_2 = \mathrm{Trace}(\mathbf{A}^T \mathbf{A})^{1/2}$.
$\mathbf{A} \succ 0$	The matrix $\mathbf{A}$ is (symmetric) positive definite, all its eigenvalues are strictly positive.
$\mathbf{A} \succeq 0$	The matrix $\mathbf{A}$ is (symmetric) positive semi-definite, all its eigenvalues are nonnegative.
$\exp(\mathbf{A})$	The matrix exponential of $\mathbf{A}$.
$\log(\mathbf{A})$	The matrix logarithm of (the positive definite matrix) $\mathbf{A}$.

Given a tensor $\mathbf{S} \in \mathrm{S}^+(3)$ whose spectral decomposition is $\mathbf{U}\Lambda\mathbf{U}^T$, we define the notations,

$\mathbf{U}$	The rotation matrix whose columns are the eigenvectors of the tensor.
Λ	The diagonal matrix of eigenvalues.
λ_i	The i-th eigenvalue of the tensor.
$\mathbf{q}$	The unit quaternion associated to the orientation of tensors.
$\mathcal{Q}$	The set of eight unit quaternions representing the different decompositions of the tensor.
$\mathbf{q}_r$	A realigned quaternion, *i.e.* a chosen quaternion in the set $\mathcal{Q}$.

The following notations are also used throughout the thesis.

t	The interpolating parameter.
w_i	The weight of element i in a weighted sum.
k	The weighting parameter of the spectral metrics.
κ	The weighting parameter of the spectral-quaternion similarity measure.
$\mathbf{M}$	A transformation matrix.
$\mathbf{t}$	A translation vector.
x	The spatial coordinates of a point, $x \in \mathbb{R}^d$, with $d = 2$ or 3 for 2D or 3D images.
$\mathcal{V}(x)$	The set of (spatial) neighbors of the point at x.
α	The level of uncertainty for statistical tests.

The following acronyms are also used in the thesis.

MRI	Magnetic Resonance Imaging.
DTI	Diffusion Tensor Imaging.
DWI	Diffusion Weighted Imaging.
FA	Fractional Anisotropy.
RA	Relative Anisotropy.
HA	Hilbert Anisotropy.
SVD	Singular Value Decomposition.
SPD	Semi-Positive Definite.

Interpolating curves and means in manifolds

The images considered in this thesis belong to the general field of matrix-valued images, a field which has received growing interest in the past few years. In the case of Diffusion Tensor Imaging, these (matrix-valued) images have an additional property: they belong to a nonlinear (but nevertheless highly structured) space, i.e. a Riemannian manifold. In this chapter, we show why those types of nonlinear images should be processed geometrically, i.e. with specific operations defined on manifolds. For that purpose, we exploit the concept of invariances. The basic operations of interpolating curves, means, and similarity measures are presented for two different manifolds: the space of symmetric positive definite matrices and the space of positive semi-definite matrices of fixed rank. The definition of a new mean for an arbitrary number of fixed-rank positive semi-definite matrices is a contribution of the present thesis. In particular, it highlights the fact that interpolating curves should not necessarily be curves of minimal length.

Multivalued variables appear in many fields of research, one of the most important being imaging. Multivariate images include color images, multi- and hyperspectral images. This thesis is particularly interested in Diffusion Tensor Imaging, a novel imaging method which produces matrix-valued images. These matrices are not 'simply' matrices, but belong to the particular space of symmetric positive definite matrices, also called the Riemannian manifold $S^+(3)$. The development of algorithms which specifically operate on manifolds is a current topic of research, which will be addressed in this chapter. Algorithms on manifolds are motivated by the fact that, for many applications, it appears that it is very important to geometrically process the data, *i.e.* to process them in their particular space, using the appropriate invariant metric. Invariance to a particular group action is indeed an important concept of the Riemannian manifold theory. The 'geometric processing' is in contrast with an arithmetic processing of data, which could be simpler but often has the disadvantage not to preserve some important features of the data.

Among the different operations which can be extended to manifold-valued data, this chapter will focus on interpolating curves and on the definitions of means. Indeed, these are two crucial operations in image processing, and in multivariate statistics, see (Thévenaz et al., 2000) and references therein. Moreover, these concepts have been shown to be important for Diffusion Tensor Images processing, for example in (Zhang et al., 2006; Kindlmann et al., 2007; Arsigny et al., 2007; Fletcher and Joshi, 2007; Fletcher et al., 2009).

The concepts of mean and interpolating curves are current topic of research in many other fields, see (Absil et al., 2004; Ando et al., 2004; Arsigny et al., 2007; Bonnabel et al., 2013; Bonnabel and Sepulchre, 2009; Fletcher and Joshi, 2007; Moakher, 2005; Petz and Temesi, 2005; Moakher and Zéraï, 2011).

In this chapter, we will focus on two spaces: the manifold of $n \times n$ symmetric positive definite matrices, $S^+(n)$ and the manifold of fixed-rank semi-definite positive matrices, $S^+(p, n)$. The first section will discuss interpolating curves in $S^+(n)$, introducing two distinct frameworks: the affine-invariant and Log-Euclidean frameworks. The second section will be devoted to the different definitions of means of two, or more than two, matrices. Finally, Section 2.3 will define those tools in $S^+(p, n)$.

2.1 Interpolating curves in $S^+(n)$

Symmetric positive definite matrices of size $n \times n$, which form the space $S^+(n)$, appear in many areas of engineering and applied mathematics, as explained in (Bonnabel et al., 2013). For example, they appear as covariance matrices in statistics, as variables in convex and semidefinite programming, as unknowns of important matrix (in)equalities in systems and control theory, and as kernels in machine learning, to cite a few. Diffusion tensor images, which are the topic of this dissertation, are fields of diffusion tensors, which belong to $S^+(3)$.

In this section, we will focus on the interpolating curves between two elements of $S^+(n)$. An interpolating curve of parameter t between two elements $\mathbf{S}_1$ and $\mathbf{S}_2$ of $S^+(n)$ is a one-parameter curve $\mathbf{S}(t)$ which starts from $\mathbf{S}(0) = \mathbf{S}_1$ and ends in $\mathbf{S}(1) = \mathbf{S}_2$. Each element $\mathbf{S}(t), 0 \leq t \leq 1$, belongs to $S^+(n)$.

2.1.1 The case of positive scalars

The space of positive scalars can be denoted by $S^+(1)$. Two different definitions of interpolation between scalars exist.

Definition 2.1.1. *Consider two scalars s_1 and s_2, with $s_1, s_2 > 0$. The* arithmetic *interpolation between those two scalars is given by*

$$s_a(t; s_1, s_2) = (1 - t)s_1 + t s_2 \,, \tag{2.1}$$

while the geometric *interpolation is given by*

$$s_g(t; s_1, s_2) = \exp((1 - t) \log s_1 + t \log s_2) = s_1 \exp(t \log(\frac{s_2}{s_1})) \,. \tag{2.2}$$

Those curves are illustrated in Figure 2.1, with $s_1 = 1$ and $s_2 = 3$.

The major difference between those two curves comes from their different interpretation of the space of positive scalars. Indeed, the arithmetic interpolation considers the set of positive scalars as a *subset* of the line of real numbers, $\mathbb{R}$. To the contrary, the geometric interpolation considers $\mathbb{R}^+$ as a *homogeneous space*. Those different conceptions lead to different definitions of interpolating curve, mean and distance.
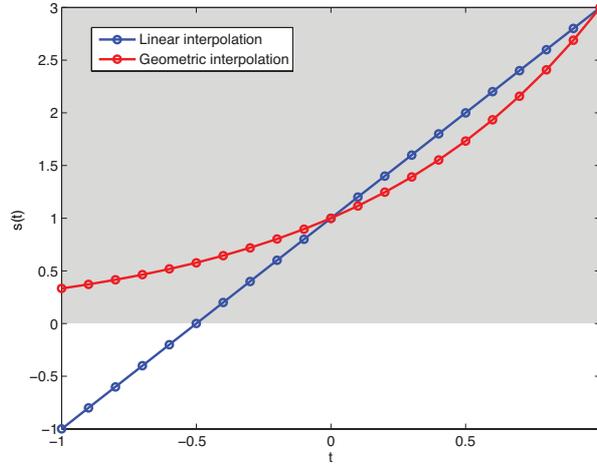
Figure 2.1: Difference between linear and geometric interpolation. The gray zone indicates the space of positive scalars, S$^+(1)$.

The two sets $\mathbb{R}$ and $\mathbb{R}^+$ are two different groups, which can be viewed as Lie groups, if equipped with an appropriate composition operation, or *group action*. If we denote by G a group, a group action associates to two elements g_1 and $g_2 \in G$ an element $g_1 g_2 \in G$. G is a Lie group if the group action satisfies the following properties:

Associativity If g_1, g_2 and $g_3 \in G$, then $(g_1 g_2) g_3 = g_1 (g_2 g_3) = g_1 g_2 g_3$.

Identity There exists an element $e \in G$, called the identity, such that for all $g \in G$, $eg = ge = g$.

Inverse For all $g \in G$, there exists an element $g^{-1} \in G$ such that $gg^{-1} = g^{-1}g = e$.

Smoothness The group operation should be smooth and differentiable.

The group of real numbers $\mathbb{R}$ is a Lie group if we consider the *addition* as group action, with zero as identity and $-g$ as the inverse of g. To the contrary, the addition can not be used as a group action for $\mathbb{R}^+$, and the appropriate group action for this set is the *multiplication*. In this case, the identity is 1 and the inverse of g is $1/g$.

An important concept when dealing with Lie groups is to require that the operations defined on these groups are *invariant by the group action*. This explains why the tools on $\mathbb{R}$ need to be invariant by translation while the ones on $\mathbb{R}^+$ should be invariant by scaling. In particular, if $\delta, \beta > 0$,

$$s_a(t; s_1 + \delta, s_2 + \delta) = s_a(t; s_1, s_2) + \delta, \tag{2.3}$$
$$s_g(t; \beta s_1, \beta s_2) = \beta s_g(t; s_1, s_2). \tag{2.4}$$

This concept of invariances also applies to the notion of *measure* in the spaces, which helps to explain how the length of a curve is computed, and depends on the space of interest. For example, the appropriate measure in $\mathbb{R}$ is the Euclidean norm. For the space $\mathbb{R}^+$, the norm is the Euclidean norm of the logarithm of the value. The concept of norm is used to define distances.

Definition 2.1.2. *The distance between two points is given by the length of the curve between them. Using the appropriate measure in each space, we have*

$$d_a(s_1, s_2) \quad = \quad |s_1 - s_2|, \tag{2.5}$$

$$d_g(s_1, s_2) \quad = \quad \left| \log\left(\frac{s_1}{s_2}\right) \right|. \tag{2.6}$$

It can be observed that the two distances are invariant by the action of their respective group actions, since

$$d_a(s_1 + \delta, s_2 + \delta) \quad = \quad d_a(s_1, s_2),$$

$$d_g(\beta s_1, \beta s_2) \quad = \quad d_g(s_1, s_2).$$

Finally, the interpolating curves can be used to define means. Indeed, means are defined as the midpoint of interpolating curves.

Definition 2.1.3. *The mean between two scalars is the midpoint of the corresponding interpolating curve.*
Using the arithmetic interpolation, the arithmetic mean is

$$\mathrm{mean}_a(s_1, s_2) = \frac{s_1 + s_2}{2}. \tag{2.7}$$

The geometric mean is the midpoint of the geometric interpolation, i.e.

$$\mathrm{mean}_g(s_1, s_2) = \exp(\frac{1}{2} \log s_1 s_2) = \sqrt{s_1 s_2}. \tag{2.8}$$

As expected, the means are invariant by their group action and we have

$$\mathrm{mean}_a(s_1 + \delta, s_2 + \delta) \quad = \quad \mathrm{mean}_a(s_1, s_2) + \delta, \tag{2.9}$$

$$\mathrm{mean}_g(\beta s_1, \beta s_2) \quad = \quad \beta \mathrm{mean}_g(s_1, s_2). \tag{2.10}$$

It is important to note that the geometric curve, distance and mean can be computed by using their arithmetics counterparts on the *logarithms* of the values, followed by an exponentiation. For example, we have

$$\sqrt{s_1 s_2} = \exp\left(\frac{\log(s_1) + \log(s_2)}{2}\right). \tag{2.11}$$

There is in fact a *one-to-one* mapping between the space $\mathbb{R}$ and the space $\mathbb{R}^+$, since for $s \in \mathbb{R}^+$ and $b \in \mathbb{R}$, we have

$$\log : \mathbb{R}^+ \to \mathbb{R} \quad : \quad b = \log(s), \tag{2.12}$$

$$\exp : \mathbb{R} \to \mathbb{R}^+ \quad : \quad s = \exp(b). \tag{2.13}$$

This mapping allows to handle positive values as simple scalars, using their logarithms. At the end of the processing, an exponentiation should be used to 'go back' to the space $\mathbb{R}^+$.

The example of positive scalars has enabled to define in a simple way some important concepts, which are the notions of group action and invariance to this action, the notions of interpolating curve, distance, and mean, as well as the importance of processing elements in their own space. The following sections will show how to extend those simple concepts to elements of $\mathrm{S}^+(n)$.

2.1.2 Geodesics are the natural extension of interpolating curves in $\mathrm{S}^+(n)$

Viewing positive definite matrices as matrix generalizations of positive numbers, we search for matrix analogues of the interpolating curve (2.2). Some necessary definitions to understand the developed frameworks will first be reminded. Most of those definitions come from (Absil et al., 2008).

Riemannian manifolds and metrics

Informally, a d-dimensional *manifold* can be defined as a set $\mathcal{M}$ covered with a 'suitable' collection of coordinate patches, or charts, that identify certain subsets of $\mathcal{M}$ with open subsets of $\mathbb{R}^d$ (Absil et al., 2008). Such a collection of coordinate charts can be thought of as the basic structure required to do differential calculus on $\mathcal{M}$.

Definition 2.1.4. *A* tangent vector ξ_x *to a manifold* $\mathcal{M}$ *at a point x is a mapping from* $\mathcal{F}_x(\mathcal{M})$ *(the set of smooth real-valued functions defined on a neighborhood of x) to* $\mathbb{R}$ *such that there exists a curve γ on $\mathcal{M}$ with $\gamma(0) = x$, satisifying*

$$\xi_x f = \dot{\gamma}(0) f := \left. \frac{\mathrm{d}(f(\gamma(t)))}{\mathrm{d}t} \right|_{t=0} \tag{2.14}$$

for all $f \in \mathcal{F}_x(\mathcal{M})$. Such a curve γ is said to realize the tangent vector ξ_x.

The *tangent space* to $\mathcal{M}$ at x, denoted by $T_x\mathcal{M}$, is the set of all tangent vectors to $\mathcal{M}$ at x. It is important to note that this set admits a structure of vector space. In order to define a notion of length that applies to tangent vectors, every tangent space $T_x\mathcal{M}$ is endowed by an inner product $\langle \cdot, \cdot \rangle_x$, which induces a norm

$$||\xi_x||_x := \sqrt{\langle \xi_x, \xi_x \rangle_x} \tag{2.15}$$

on $T_x\mathcal{M}$.

A manifold whose tangent spaces are endowed with a smoothly varying inner product is called a *Riemannian manifold*. The inner product is then called the *Riemannian metric*.

Definition 2.1.5. *The* length of a curve $\gamma : [a, b] \to \mathcal{M}$ *on a Riemannian manifold is defined by*

$$L(\gamma) = \int_a^b \sqrt{\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle} \mathrm{d}t \,. \tag{2.16}$$

The Riemannian distance *on a Riemannian manifold is given by*

$$\mathrm{dist} : \mathcal{M} \times \mathcal{M} \to \mathbb{R} : \mathrm{dist}(x, y) = \inf_\Gamma L(\gamma) \,, \tag{2.17}$$

where Γ is the set of all curves in $\mathcal{M}$ joining points x and y.

The curve for which the length is the minimal is called the *geodesic* between x and y. This geodesic is the generalization on manifolds of the concept of straight lines in $\mathbb{R}^n$. It can also be defined through the notion of curve with zero acceleration.

If we come back to the example of positive scalars, the curves (2.1) and (2.2) can be viewed as geodesics, for two different metrics. Indeed, the arithmetic interpolation is a geodesic if we consider the (translation) invariant metric of the Lie group $\mathbb{R}$. If the group $\mathbb{R}^+$ is considered, the group action is the multiplication and the metric has to be scale invariant, which leads to the geometric curve (2.2), whose length is given by (2.6).

In many works, geodesics have been used as interpolating curves between two points. It will however be shown in Section 2.3.1 and Chapter 4 that, for some applications, other curves are preferable, even if these are not of minimal length.

Affine-invariant geometry

The space $\mathrm{S}^+(n)$ is a Riemannian manifold, whose tangent space at identity is the space of symmetric (and not necessarily positive definite) matrices, denoted by $\mathrm{S}(n)$. As explained, the tangent space has the structure of a vector space. The simplest inner product on the tangent space at the identity matrix is given by $\langle W_1, W_2 \rangle = \mathrm{Tr}(W_1^T W_2)$, if W_1 and W_2 are tangent vectors, $W_1, W_2 \in \mathrm{S}(n)$. This product is the standard matrix scalar product. The requirement which is at the basis of the affine-invariant metric is that the scalar product at a point $\mathbf{S}$ is invariant by the action of any affine transformation, *i.e.*

$$\langle W_1, W_2 \rangle_{\mathbf{S}} = \langle A \star W_1, A \star W_2 \rangle_{A \star \mathbf{S}}, \tag{2.18}$$

where $A \star \mathbf{S} = A\mathbf{S}A^T$, A belongs to the space of general linear group $\mathrm{GL}(n)$, the space of $n \times n$ invertible matrices. This requirement is the matrix generalization of the scaling invariance for scalar curves, explained in Section 2.1.1. In this case, the group action is $A \star \mathbf{S}$. The invariance property should be true if $A = \mathbf{S}^{-1/2}$, which allows to define the scalar product at any $\mathbf{S}$ from the scalar product at the identity $\mathbf{I}$:

$$\begin{aligned} \langle W_1, W_2 \rangle_{\mathbf{S}} &= \langle \mathbf{S}^{-1/2} W_1 \mathbf{S}^{-1/2}, \mathbf{S}^{-1/2} W_2 \mathbf{S}^{-1/2} \rangle_{\mathbf{I}} \\ &= \mathrm{Tr}(\mathbf{S}^{-1/2} W_1 \mathbf{S}^{-1} W_2 \mathbf{S}^{-1/2}). \end{aligned}$$

This Riemannian metric is called the affine-invariant metric, and has been proposed in many works, as (Lenglet et al., 2006; Pennec et al., 2006; Lenglet, 2006; Moakher, 2005). Using this metric, the geodesic between two symmetric definite positives matrices can be defined.

Definition 2.1.6. *With the affine-invariant metric, the curve of minimal length (the geodesic) between two symmetric positive definite matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ is*

$$\mathbf{S}(t) = \mathbf{S}_1^{1/2} \exp(t \log(\mathbf{S}_1^{-1/2} \mathbf{S}_2 \mathbf{S}_1^{-1/2})) \mathbf{S}_1^{1/2}. \tag{2.19}$$

The length of this curve gives the Riemannian distance between those two matrices:

$$d(\mathbf{S}_1, \mathbf{S}_2) = ||\log(\mathbf{S}_1^{-1/2} \mathbf{S}_2 \mathbf{S}_1^{-1/2})||_2 \tag{2.20}$$

where $||A||_2$ is the Frobenius norm of matrix, $||A||_2 = \sqrt{\mathrm{Tr}(A^T A)}$.

It is important to note that this distance is globally defined and makes $\mathrm{S}_+(n)$ a complete metric space (Fletcher and Joshi, 2007). Indeed, symmetric matrices with null or negative eigenvalues are at an infinite distance from any symmetric positive definite matrices. This means that the original cone of positive definite symmetric matrices has been

changed into a regular and complete (but curved) manifold with an infinite development in each of its directions.

The computational cost of the affine-invariant distance (2.20) is high, due to the use of matrix inverses, square roots, logarithms and exponentials. This explains why other methods, more computationally efficient, have been proposed. The Log-Euclidean geometry, presented in the following, is one of these methods.

Log-Euclidean geometry

The motivation for the development of the Log-Euclidean geometry was to define a new Riemannian framework which overcomes the computational limitations of the affine-invariant framework while preserving some of its theoretical properties (Arsigny et al., 2006, 2007).
It should be noted that other 'computationally efficient' methods have been proposed as an alternative to the affine-invariant framework, as in (Sra, 2011). This paper introduced a novel metric for which the computational cost is reduced, because it avoids spectral decompositions or logarithms of matrices, and prefers the Cholesky decomposition.
In this thesis, we choose to focus on the Log-Euclidean method, because of the widespread use of this framework in our domain of application, the diffusion tensor imaging.

To introduce this framework, some important definitions are reminded.

- **Matrix exponential** For any matrix $\mathbf{A}$, its exponential is given by

$$\exp(\mathbf{A}) = \sum_{k=0}^{\infty} \frac{\mathbf{A}^k}{k!} \ .$$

  In the particular case of symmetric positive definite matrices, for which it always exist an orthonormal coordinate system in which these matrices are diagonal, the matrix exponential can be computed through diagonalization. If the spectral decomposition of a symmetric positive definite matrix $\mathbf{S}$ is equal to $\mathbf{R}\Lambda\mathbf{R}^T$ where $\mathbf{R}$ is a rotation matrix and Λ a diagonal matrix with the eigenvalues of $\mathbf{S}$ in its diagonal, then the exponential is equal to

$$\exp(\mathbf{S}) = \mathbf{R}\Lambda^*\mathbf{R}^T,$$

  where Λ^* is a diagonal matrix with the exponential of the eigenvalues in its diagonal.

- **Matrix logarithm** As in the scalar case, the matrix logarithm is defined as the inverse of the exponential. It is important to note that for general matrices, neither the uniqueness nor the existence of a logarithm is guaranteed for a given invertible matrix. However, in the particular case of symmetric positive definite matrices, the logarithm always exists and is unique. It can be defined in a similar way as the exponential, through

$$\log(\mathbf{S}) = \mathbf{R}\tilde{\Lambda}\mathbf{R}^T \ . \tag{2.21}$$

  $\tilde{\Lambda}$ is a diagonal matrix, whose diagonal elements are equal to the logarithms of the eigenvalues of $\mathbf{S}$.

The logarithm of a symmetric positive definite matrix is always a symmetric matrix. Conversely, the exponential of any symmetric matrix yields a symmetric positive definite matrix. There is a one-to-one correspondence between those two spaces.

The *one-to-one mapping* between the (non-linear) space of symmetric positive definite matrices and the (vector) space of symmetric matrices is at the core of the Log-Euclidean framework. Indeed, this means that standard algebraic operations can be transferred to symmetric positive definite matrices, thanks to the matrix exponential. In particular, the logarithmic multiplication $\odot$ and the logarithmic scalar multiplication $\circledast$ can be defined.

$$\mathbf{S}_1 \odot \mathbf{S}_2 = \exp(\log(\mathbf{S}_1) + \log(\mathbf{S}_2)), \tag{2.22}$$

$$\lambda \circledast \mathbf{S} = \exp(\lambda \log(\mathbf{S})) = \mathbf{S}^\lambda. \tag{2.23}$$

The logarithmic product (which is commutative) is based on the idea that, since the space of symmetric matrices $S(n)$ has an additive group structure, it 'suffices' to transport this additive structure to $S^+(n)$ with the exponential. This idea is similar to the one expressed in Section 2.1.1 for positive scalars. In this section, we said that geometric processing could be done by using the logarithms of positive scalars, and an exponentiation. This idea is exactly the same as the one used to construct Equation (2.22). Using $\odot$ and $\circledast$, the space $S^+(n)$ has by construction a *vector space structure.*

If only the multiplication $\odot$ is considered, a *Lie group* structure can be given to $S^+(n)$. A Lie group is a space that is both a smooth manifold and a group in which multiplication and inversion are smooth mappings. Since the logarithm and exponential mappings are smooth, this is also the case of $\odot$. Considering $S^+(n)$ as a Lie group, the goal is to associate to this group a metric. For Lie groups, *bi-invariant metrics*(*i.e.* metrics which are invariant by multiplication and inversion) are considered as the most convenient in practice. This kind of metrics exist for symmetric positive definite matrices, and are particularly simple. These metrics have been named *Log-Euclidean* metrics, since they correspond to Euclidean metrics in the domain of logarithms. All the theoretical developments of the Log-Euclidean framework are presented in (Arsigny et al., 2007).

Definition 2.1.7. *Using the Log-Euclidean metric, the distance between two symmetric positive matrices* $\mathbf{S}_1$ *and* $\mathbf{S}_2$ *is given by*

$$d(\mathbf{S}_1, \mathbf{S}_2) = ||\log(\mathbf{S}_1) - \log(\mathbf{S}_2)||_2. \tag{2.24}$$

This distance corresponds to the length of the geodesic between those two matrices, which is

$$\mathbf{S}(t) = \exp((1-t)\log(\mathbf{S}_1) + t\log(\mathbf{S}_2)). \tag{2.25}$$

In comparison to the affine-invariant metric, the Log-Euclidean metric enables a simpler processing of symmetric positive definite matrices. However, the price to pay is the limitation of the invariances: the Log-Euclidean framework is not invariant by any affine transformation, but is invariant by inversion, scaling and rotation, *i.e.*

$$d(\mathbf{S}_1, \mathbf{S}_2) = d(s\mathbf{R}\mathbf{S}_1\mathbf{R}^T, s\mathbf{R}\mathbf{S}_2\mathbf{R}^T), \tag{2.26}$$

with $\mathbf{R}$ an orthogonal matrix and s a positive scalar. These two invariances are often viewed as sufficient in applications such as medical imaging. The differences between

affine-invariant and Log-Euclidean metrics are rooted in the fact that the affine-invariant geometry is *intrinsic* while the Log-Euclidean one is *extrinsic*. This extrinsic behavior enables to perform computations in the embedding (Euclidean) space, but has the disadvantage of deforming the intrinsic notions. Extrinsic geometry produces a good approximation of the intrinsic one while close to the identity, but the quality of the approximation decreases while moving away from it.

2.2 The different definitions of means on S$^+(n)$

Just like interpolating curves, means appear in many processing algorithms. As in the scalar case, where the arithmetic and geometric definitions of means co-exist, different ways exist to define means on manifolds. In particular, means could be computed through a simple arithmetic average of matrices. However, for many applications where data represent some physical processes, it is important to define a *geometric* mean, in order to have a property of invariance of this mean with respect to inversion, as well as group action invariances.

In this section, the important case of means of two symmetric positive definite matrices (two ellipsoids) will first be introduced. The extension to means of more than two ellipsoids is the topic of Section 2.2.2.

2.2.1 Means of two ellipsoids

Two definitions of a geometric mean between two ellipsoids will be presented and compared in this section. If different in appearance, it can be shown that those means are in fact equivalent.

As the midpoint of the geodesic

An intuitive way to define a mean between two points is to take the midpoint of the curve of minimal length between those two points.

Definition 2.2.1. *The mean of two matrices* $\mathbf{S}_1$ *and* $\mathbf{S}_2$ *is given by the midpoint of the geodesic between them,* i.e

$$\mathrm{mean}(\mathbf{S}_1, \mathbf{S}_2) = \mathbf{S}(1/2; \mathbf{S}_1, \mathbf{S}_2),\qquad(2.27)$$

where $\mathbf{S}(t; \mathbf{S}_1, \mathbf{S}_2)$ *is the geodesic between* $\mathbf{S}_1$ *and* $\mathbf{S}_2$*, parametrized by* t*.*

This mean is of course unique and commutative, and corresponds to a *Riemannian mean* between two ellipsoids. Moreover, the geodesic can also be used to define a *weighted* mean between two ellipsoids, since

$$\mathrm{mean}(w_1\mathbf{S}_1, w_2\mathbf{S}_2) = \mathbf{S}(w_2; \mathbf{S}_1, \mathbf{S}_2)\qquad(2.28)$$

where w_1 and w_2 are the weights associated to the ellipsoids, and $w_1 + w_2 = 1$.

This definition is of course not limited to S$^+(n)$ and is general for any Riemannian manifolds for which a geodesic can be computed. It corresponds to the Kärcher mean (Kärcher, 1977) when only two elements are considered.

Through a notion of order

Another well-known definition of the mean is through a notion of order. If we denote by $\mathbf{S} \succ 0$ the fact that a matrix is positive definite, there exists an ordering of matrices, which is given by $\mathbf{A} \succ \mathbf{B} \iff \mathbf{A} - \mathbf{B} \succ 0$. The Ando geometric mean $\mathbf{A}\#\mathbf{B}$ is defined through an extremal problem.

Definition 2.2.2. *The Ando mean of two positive definite matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ is the solution of an extremal problem.*

$$\mathbf{A}\#\mathbf{B} = \max \left\{ \mathbf{X} \succ 0 : \begin{pmatrix} \mathbf{A} & \mathbf{X} \\ \mathbf{X} & \mathbf{B} \end{pmatrix} \succeq 0 \right\} \tag{2.29}$$

It is important to note that the solution of this extremal problem corresponds to the midpoint of the geodesic between the two matrices, *i.e.* that the two definitions (2.27) and (2.29) are equivalent. Those two definitions satisfy all the usual properties which are required to be geometric means (Ando et al., 2004; Bhatia, 2009).

2.2.2   Means of $N > 2$ ellipsoids

If the definition of a mean between two ellipsoids seems quite intuitive, this is not the case when more than two ellipsoids have to be averaged. Indeed, some problems arise in this case, such as the symmetry of the mean. This issue is well described in (Bhatia, 2009) and (Ando et al., 2004).

Considering the case of 3 ellipsoids can help to understand the case of N matrices. If we consider three positives matrices $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3 \in \mathrm{S}^+(n)$, their geometric mean $G(\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3)$ should be a positive matrix with the following properties.

- If the three matrices $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ commute with each other, their mean should be given by $(\mathbf{A}_1 \mathbf{A}_2 \mathbf{A}_3)^{1/3}$.

- **Symmetry** $G(\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3) = G(\mathbf{A}_{\pi(1)}, \mathbf{A}_{\pi(2)}, \mathbf{A}_{\pi(3)})$ for every permutation π of $\{1, 2, 3\}$.

- **Monotonicity** $G(\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3) \succeq G(\mathbf{A}_1', \mathbf{A}_2, \mathbf{A}_3)$ whenever $\mathbf{A}_1 \succeq \mathbf{A}_1'$.

None of the definitions given in the previous section for the mean of two matrices is easily extendable to the case of three matrices. This explains why another 'point of view' has to be adopted in this section. In a few words, the idea is to define the mean as the 'center' of the triangle that has the three matrices as its vertices. Transposing this idea to a manifold leads to the usual definition of the Kärcher mean.

Kärcher mean

The Kärcher mean (Kärcher, 1977) can be viewed as the extension to the case of N positive matrices of the mean between two matrices defined by the midpoint of the geodesic.

Definition 2.2.3. *For a group of matrices $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_N$, the Kärcher mean is given by*

$$\mathrm{mean}(\mathbf{S}_1, \mathbf{S}_2, \ldots \mathbf{S}_N) = \arg \min_{\mathbf{S} \in \mathrm{S}^+(n)} \sum_{i=1}^{N} d^2(\mathbf{S}, \mathbf{S}_i). \tag{2.30}$$

This definition corresponds to the 'center of mass' of the group of N matrices.

With the affine-invariant metric, the solution to (2.30) can be computed through an optimization algorithm, explained in (Pennec et al., 2006; Fletcher and Joshi, 2007). This algorithm uses a gradient descent scheme on the space $\mathrm{S}^+(n)$. The computational cost of this algorithm is prohibitive, as for the other tools of this framework.

With the Log-Euclidean metric, however, a *closed-form* formula for the solution of (2.30) is given by (Arsigny et al., 2006, 2007)

$$\mathrm{mean}_{LE}(\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_N) = \exp\left(\frac{1}{N} \sum_{i=1}^{N} \log(\mathbf{S}_i).\right) \qquad (2.31)$$

This formula does not require an optimization algorithm, while sharing most of the properties of the Riemannian mean. In particular, the determinant of the geometric mean is the geometric mean of the determinants. This property is not satisfied by the Euclidean mean

$$\mathrm{mean}_{E}(\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_N) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{S}_i, \qquad (2.32)$$

which induces a 'swelling effect'. This effect is due to the fact that the determinant of the Euclidean mean could be higher than all the determinants $\det(\mathbf{S}_i)$.

Ando mean

Another definition of a mean for many matrices is found in (Ando et al., 2004). In a few words, this mean is constructed by a recursive procedure, based on the mean between two matrices. Suppose first that the mean between k matrices $G(\mathbf{S}_1, \mathbf{S}_2, \ldots \mathbf{S}_k)$ has been defined, let $\mathbf{S} = (\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_{k+1})$, and let T denote the transformation

$$T(\mathbf{S}) \equiv (G((\mathbf{S}_i)_{i \neq 1}), G((\mathbf{S}_i)_{i \neq 2}), \ldots, G((\mathbf{S}_i)_{i \neq k+1}) \qquad (2.33)$$

Then, the sequence $\{T^r(\mathbf{S})\}_{r=1}^{\infty}$ has a limit, which has the form $(\tilde{\mathbf{S}}, \ldots, \tilde{\mathbf{S}})$. Ando *et al* define the geometric mean $G(\mathbf{S}_1, \mathbf{S}_2, \ldots \mathbf{S}_{k+1})$ to be $\tilde{\mathbf{S}}$.

However, this recursive procedure suffers from the same limitation as the affine-invariant mean: the computational cost of the method is prohibitive. Moreover, those two means, as well as the Log-Euclidean mean, are not appropriate for the processing of rank-deficient matrices (Bonnabel et al., 2013).

2.3 Interpolating curve and means in $\mathrm{S}^+(p, n)$

The preceding section presented the important tool of means for positive-definite matrices. However, those means are not appropriate for *semi*-definite positive matrices *of fixed rank*. Symmetric matrices of dimension n are said to be semi-definite positive if $x^T \mathbf{S} x \geq 0$ for all $x \in \mathbb{R}^n$. The rank of those matrices is denoted by $p < n$, and the space of symmetric positive semi-definite matrices of fixed rank p is called $\mathrm{S}^+(p, n)$. This section will explain how appropriate interpolating curve and means could be defined for this particular space and will be based on the work presented in (Bonnabel and Sepulchre, 2009; Bonnabel et al., 2013).

The main motivation for the development of a mean on $S^+(p,n)$ comes for the observation that typical matrix computations (like Singular Value Decomposition, QR factorization) require $O(n^3)$ operations for a matrix in $S^+(n)$, thus limiting their use in large-scale problems. A possible solution to decrease this cost is to work with low-rank approximations instead of the full matrices. A rank p approximation of a positive definite matrix can be factored as $\mathbf{A} = \mathbf{Z}\mathbf{Z}^T$, where $\mathbf{Z} \in \mathbb{R}^{n \times p}$ is of much reduced size if $p \ll n$, leading to a reduction of the numerical cost from $O(n^3)$ to $O(np^2)$, which grows only linearly with the size of the problem if p is kept to a moderate value.

The problem is that the Riemannian metric on $S^+(n)$ is only valid for full-rank matrices. It is thus important to define a novel metric, specific for the space $S^+(p,n)$. The principal idea of (Bonnabel and Sepulchre, 2009) is to use a decoupling between the subspace in which the matrices are living and the shapes of the ellipsoids of size p. Formally, for $\mathbf{A} \in S^+(p,n)$, we have

$$\mathbf{A} = \mathbf{Z}\mathbf{Z}^T = (\mathbf{U}\mathbf{R})(\mathbf{U}\mathbf{R})^T = \mathbf{U}\mathbf{R}^2\mathbf{U}^T \tag{2.34}$$

where $\mathbf{Z} \in \mathbb{R}_*^{n \times p}$, the space of full rank matrices of size $n \times p$, $\mathbf{R} \in S^+(p)$ and $\mathbf{U} \in V_{n,p}$, the Stiefel manifold, *i.e.* the space of $n \times p$ matrices with orthonormal columns, $\mathbf{U}^T\mathbf{U} = \mathbf{I}_p$. If we consider the following group action for $\mathbf{O} \in O(p)$, the space of orthogonal matrices of size p,

$$\begin{aligned} \mathbf{R} &\rightarrow \mathbf{O}^T\mathbf{R}\mathbf{O} &\in S^+(p) \\ \mathbf{U} &\rightarrow \mathbf{U}\mathbf{O} &\in V_{n,p} \end{aligned}$$

then the representation $\mathbf{A} = \mathbf{U}\mathbf{R}^2\mathbf{U}^T$ with $(\mathbf{U}, \mathbf{R}^2) \in V_{n,p} \times S^+(p)$ is univocal up to the equivalence relation

$$(\mathbf{U}, \mathbf{R}^2) \equiv (\mathbf{U}\mathbf{O}, \mathbf{O}^T\mathbf{R}\mathbf{O}) \tag{2.35}$$

for any $\mathbf{O} \in O(p)$. We indeed have that $\mathbf{A} = \mathbf{U}\mathbf{R}^2\mathbf{U}^T = \mathbf{U}\mathbf{O}(\mathbf{O}^T\mathbf{R}^2\mathbf{O})\mathbf{O}^T$.

Using (2.35), the set $S^+(p,n)$ can be defined through a *quotient manifold* representation

$$S^+(p,n) \cong (V_{n,p} \times S^+(p))/O(p). \tag{2.36}$$

Since the Grassmann manifold $Gr(p,n)$ is equivalent to $V_{n,p}/O(p)$, the metric associated to the product space (2.35) is naturally given by the sum of the metric in $Gr(p,n)$ and the metric in $S^+(p)$.

2.3.1 Interpolating curve and mean of two fixed-rank ellipsoids

In this section, an interpolating curve between two fixed-rank ellipsoids will be defined, based on the above definitions. This interpolating curve will be used to define the mean of those two ellipsoids.
Let $\mathbf{A}_1 = \mathbf{U}_1\mathbf{R}_1^2\mathbf{U}_1^T$ and $\mathbf{A}_2 = \mathbf{U}_2\mathbf{R}_2^2\mathbf{U}_2^T$ be two elements of $S^+(p,n)$. The representatives of the two matrices $(\mathbf{U}_i, \mathbf{R}_i^2)$, $i = 1, 2$ are defined up to an orthogonal transformation, as explained in (2.35). All the bases $\mathbf{U}_iO(p)$ correspond to the same p-dimensional subspace $\mathbf{U}_i\mathbf{U}_i^T$, as shown in Figure 2.2. This representation of a p-dimensional subspace as the set of bases $\mathbf{U}_iO(p)$ is at the core of the definition of the Grassmann manifold as a quotient manifold. The equivalence classes $\mathbf{U}_iO(p)$ are called the 'fibers'.

In the following, we will assume that the principal angles between $\text{span}(\mathbf{U}_1)$ and $\text{span}(\mathbf{U}_2)$ are less than $\pi/2$. This hypothesis is sufficient to ensure that their mean is unique.

Due to the equivalence relation explained above, there exists an ambiguity in the definition of a mean on $S^+(p, n)$. The solution proposed in (Bonnabel et al., 2013) to remove this ambiguity is to pick two particular representatives $\mathbf{Y}_1 = \mathbf{U}_1\mathbf{Q}_1$ and $\mathbf{Y}_2 = \mathbf{U}_2\mathbf{Q}_2$ in the fibers $\mathbf{U}_1\text{O}(p)$ and $\mathbf{U}_2\text{O}(p)$. Those particular bases will be chosen by imposing that their distance in the Stiefel manifold $\text{V}_{n,p}$ does not exceed the Grassmann distance between the fibers they generate. Mathematically,

$$d_{\text{V}_{n,p}}(\mathbf{Y}_1, \mathbf{Y}_2) = d_{\text{Gr}(p,n)}(\text{span}(\mathbf{U}_1), \text{span}(\mathbf{U}_2)). \tag{2.37}$$

Because the projection from $\text{V}_{n,p}$ to $\text{Gr}(p, n)$ is a Riemannian submersion, and because Riemannian submersions shorten the distances, the condition (2.37) admits the equivalent formulation:

$$(\mathbf{Q}_1, \mathbf{Q}_2) = \arg \min_{(\mathbf{O}_1, \mathbf{O}_2) \in \text{O}(p) \times \text{O}(p)} d_{\text{V}_{n,p}}(\mathbf{U}_1\mathbf{O}_1, \mathbf{U}_2\mathbf{O}_2). \tag{2.38}$$

This formulation is illustrated in Figure 2.2: a geodesic in the Grassmann manifold admits the representation of a *horizontal* geodesic in $\text{V}_{n,p}$, that is, in the present case, a geodesic whose tangent vector points everywhere to a direction normal to the fiber.
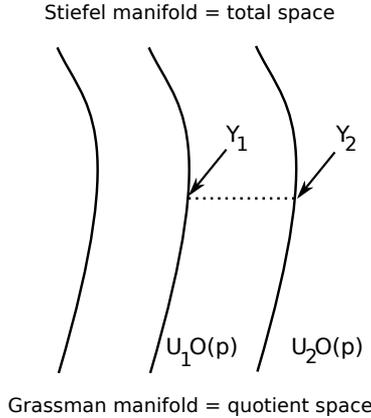

Figure 2.2: $(\mathbf{Y}_1, \mathbf{Y}_2)$ are two bases of the subspaces spanned by the columns of $\mathbf{U}_1$ and $\mathbf{U}_2$ that minimize the distance in $\text{V}_{n,p}$. The dashed line represents the shortest path between those two fibers, thus its horizontal lift in $\text{Gr}(p, n)$, viewed as a quotient manifold, is also a geodesic.

In the following, we will show how to compute the solution $(\mathbf{Q}_1, \mathbf{Q}_2)$ to the equation (2.38).

Proposition 2.3.1. *The compact singular value decomposition (SVD) of $\mathbf{U}_1^T\mathbf{U}_2$ writes*

$$\mathbf{U}_1^T\mathbf{U}_2 = \mathbf{O}_1(\cos \Sigma)\mathbf{O}_2^T \tag{2.39}$$

where Σ is a diagonal matrix whose entries are the principal angles between the p-dimensional subspaces spanned by $\mathbf{U}_1$ and $\mathbf{U}_2$. If the pair $(\mathbf{O}_1, \mathbf{O}_2)$ is defined via (2.39), it is a solution of (2.38).

Proof The proof will be based on a well-known result in the Grassmann manifold: the shortest path between two fibers in $V_{n,p}$ coincides with the geodesic path linking these two fibers in $\mathrm{Gr}(p, n)$, as the projection on the Grassmann manifold is a Riemannian submersion, and thus shortens the distances. Then, if two bases $\mathbf{Y}_1$ and $\mathbf{Y}_2$ of the fibers $\mathbf{U}_1 O(p)$ and $\mathbf{U}_2 O(p)$ are the endpoints of a geodesic in the Grassmann manifold, they must minimize (2.38). We thus have to prove that $\mathbf{Y}_1 = \mathbf{U}_1 \mathbf{O}_1$ and $\mathbf{Y}_2 = \mathbf{U}_2 \mathbf{O}_2$, where $\mathbf{O}_1$, $\mathbf{O}_2$ are defined via (2.39), are the endpoints of a minimizing Grassmann geodesic.

The general form of a geodesic in the Grassmann manifold with $\mathbf{Y}_1$ as starting point and Δ as tangent vector is

$$\gamma(t) = \mathbf{Y}_1 \mathbf{V} \cos(\Theta t) \mathbf{V}^T + \mathbf{U} \sin(\Theta t) \mathbf{V}^T \,, \tag{2.40}$$

where $\mathbf{U}\Theta\mathbf{V}^T = \Delta$ is the compact SVD of Δ. In (Bonnabel et al., 2013), the following curve is thus proposed

$$\mathbf{Y}(t) = \mathbf{Y}_1 \cos(\Sigma t) + \mathbf{X} \sin(\Sigma t) \,. \tag{2.41}$$

$\mathbf{X}$ is given by $\mathbf{X} = (\mathbf{Y}_2 - \mathbf{Y}_1 \cos(\Sigma))(\sin(\Sigma))^{-1}$, if we assume that all principal angles (all diagonal entries of Σ) are strictly positive.

The curve (2.41) is a geodesic, since it is of the form of (2.40), with $\Delta = \mathbf{X}\Sigma$. It can be proved that $\mathbf{X}\Sigma$ is a tangent vector, since

$$\mathbf{Y}_1^T \mathbf{X}\Sigma = (\mathbf{Y}_1^T \mathbf{Y}_2 - \mathbf{Y}_1^T \mathbf{Y}_1 \cos(\Sigma))(\sin(\Sigma))^{-1}\Sigma = 0 \,, \tag{2.42}$$

(using the fact that $\mathbf{Y}_1^T \mathbf{Y}_2 = \cos(\Sigma)$). Moreover, $\mathbf{X}\Sigma$ is a compact SVD, as

$$\mathbf{X}^T \mathbf{X} = (\mathbf{Y}_2^T - \cos(\Sigma)\mathbf{Y}_1^T)(\mathbf{Y}_2 - \mathbf{Y}_1 \cos(\Sigma))(\sin(\Sigma))^{-2} = (\mathbf{I} - (\cos(\Sigma))^2)(\sin(\Sigma))^{-2} = \mathbf{I} \,, \tag{2.43}$$

using the fact that $\mathbf{Y}_2^T \mathbf{Y}_1 = \mathbf{Y}_1^T \mathbf{Y}_2 = \cos(\Sigma)$. Finally, we can see that $\mathbf{Y}(0) = \mathbf{Y}_1$ and $\mathbf{Y}(1) = \mathbf{Y}_2$.

If there are null principal angles, it is clear that $\mathbf{Y}(t)$ is a geodesic. Denoting $\mathbf{C}^\dagger$ the pseudo-inversion of a matrix $\mathbf{C}$, $\mathbf{X}$ has to be chosen as $\mathbf{X} = (\mathbf{Y}_2 - \mathbf{Y}_1 \cos(\Sigma))(\sin(\Sigma))^\dagger$ along the directions corresponding to non-zero principal angles, and where $\mathbf{X}$ can be completed arbitrary with orthonormal vectors along the directions corresponding to null principal angles. Indeed, along those directions $\mathbf{Y}_1$ is equal to $\mathbf{Y}_2$ and thus $\mathbf{Y}(t)$ coincide, and the value of $\mathbf{X}$ does not play any role in the definition of $\mathbf{Y}(t)$.

Now that we have a method to choose the specific bases $\mathbf{Y}_1$ and $\mathbf{Y}_2$, it can be shown why this choice is relevant for defining a geometric mean. The following proposition proves that the rotation which maps $\mathbf{Y}_1$ to $\mathbf{Y}_2$ is equal to the rotation of *minimal energy* that maps $\mathrm{span}(\mathbf{A}_1)$ to $\mathrm{span}(\mathbf{A}_2)$.

Proposition 2.3.2. *Let $\mathbf{Y}_1 = \mathbf{U}_1 \mathbf{Q}_1$ and $\mathbf{Y}_2 = \mathbf{U}_2 \mathbf{Q}_2$ with $(\mathbf{Q}_1, \mathbf{Q}_2)$ a solution of (2.39). Then the rotation $\mathbf{R} \in SO(n)$ which maps the basis $\mathbf{Y}_1$ to the basis $\mathbf{Y}_2 = \mathbf{R}\mathbf{Y}_1$ is a rotation of minimal energy, that is, it minimizes $d_{SO(n)}(\mathbf{R}, \mathbf{I})$ among all rotation matrices that map $\mathbf{Y}_1$ to the subspace $\mathrm{span}(\mathbf{U}_2) = \mathrm{span}(\mathbf{Y}_2)$.*

Proof Let assume, without loss of generality, that $\mathbf{Y}_1 = [\mathbf{e}_1, \ldots, \mathbf{e}_r]$ where $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ is the canonical basis of $\mathbb{R}^n$. The search space can then be restricted to the rotations whose r first columns are of the form $\mathbf{Y}_2\mathbf{O}$, whereas the $n - r$ remaining columns coincide with the identity matrix (the rotation sought must indeed minimize the distance to identity). For any such rotation, we have $d_{V_{n,n}}(\mathbf{R}, \mathbf{I}) = d_{V_{n,p}}(\mathbf{Y}_2\mathbf{O}, \mathbf{I})$. Since $V_{n,n} = SO(n)$ and the fact that the metrics coincide, we have $d_{SO(n)}(\mathbf{R}, \mathbf{I}) = d_{V_{n,n}}(\mathbf{R}, \mathbf{I})$. The problem thus becomes equivalent to (2.38), and is solved taking $\mathbf{O} = \mathbf{I}$.

Once $\mathbf{Y}_1$ and $\mathbf{Y}_2$ have been computed, the full-rank $p \times p$ positive matrices $\mathbf{R}_1^2$ and $\mathbf{R}_2^2$ are given by the corresponding representatives

$$
\begin{aligned}
\mathbf{R}_1^2 &= \mathbf{Y}_1^T \mathbf{A}_1 \mathbf{Y}_1\,, \\
\mathbf{R}_2^2 &= \mathbf{Y}_2^T \mathbf{A}_2 \mathbf{Y}_2\,.
\end{aligned}
$$

The above developments enable to define a curve in $S^+(p, n)$, through the geodesic curves in the Grassmann manifold and in the space of symmetric positive definite matrices of size p.

Definition 2.3.1. *The interpolating curve between two matrices* $\mathbf{A}_1, \mathbf{A}_2 \in S^+(p, n)$ *is given by*

$$\gamma(t) = \mathbf{Y}(t)\mathbf{R}^2(t)\mathbf{Y}^T\,, \tag{2.44}$$

where

$$\mathbf{R}^2(t) = \mathbf{R}_1 \exp(t \log(\mathbf{R}_1^{-1}\mathbf{R}_2^2\mathbf{R}_1^{-1}))\mathbf{R}_1 \tag{2.45}$$

and $\mathbf{Y}(t)$ *is given by* (2.41).

It can be observed that (2.45) is the same as the affine-invariant geodesic (2.19), in the space $S^+(p)$.

The length of (2.44) is given by

$$l^2(\gamma) = ||\Theta||_2^2 + k||\log(\mathbf{R}_A^{-1}\mathbf{R}_B^2\mathbf{R}_A^{-1})||_2^2 \tag{2.46}$$

It is important to note that the length is not a Riemannian distance on $S^+(p, n)$, since the curve (2.44) is not a geodesic.

The mean between two fixed-rank ellipsoids can be defined in a similar way as the method proposed in (2.27): through the midpoint of the interpolating curve. This mean is thus equal to

$$\mathrm{mean}(\mathbf{A}, \mathbf{B}) = \gamma(1/2) = \mathbf{Y}(1/2)\mathbf{R}^2(1/2)\mathbf{Y}(1/2)^T\,. \tag{2.47}$$

A simple geometric intuition underlying the definition of the mean is the following: the mean of two flat ellipsoids $\mathbf{A}_1$ and $\mathbf{A}_2$ is defined in the mean subspace as the geometric mean of two full p-dimensional ellipsoids $\mathbf{R}_1^2$ and $\mathbf{R}_2^2$. Among the several ways to rotate the ellipsoid $\mathbf{A}_1$ into the subspace spanned by $\mathbf{A}_2$, the choice is made unique and sensible by selecting the *minimal rotation*.

As for the length of the curve, it is important to note that this mean does not necessarily agree with the Riemannian mean. However, this geometric mean is a good approximation of the Riemannian mean when k is small. Moreover, the geometric mean can be viewed as

a more natural extension of the Ando mean than the Riemannian mean. The geometric mean possesses most of the desirable properties of a matrix geometric mean, as shown in (Bonnabel and Sepulchre, 2009). Those properties will be studied in the following, for the extension of this geometric mean to N fixed rank matrices.

2.3.2 Rank-preserving mean of N PSD matrices of rank $p < n$

The extension of the mean between two semi-definite positive matrices of Equation (2.47), (Bonnabel and Sepulchre, 2009), is a first contribution of this thesis, (Bonnabel et al., 2013).

Let $(\mathbf{A}_1, \mathbf{A}_2, \ldots \mathbf{A}_N) \in \mathrm{S}^+(p, n)$ be the set of N semi-definite positive matrices for which a mean has to be computed. The main idea is to define a mean p-dimensional subspace and to bring all 'flat' ellipsoids to this mean subspace by *a minimal rotation.* When all the flat ellipsoids are in the common subspace, the problem becomes a 'usual' computation of the geometric mean of N matrices in $\mathrm{S}^+(p)$.

The construction is achieved through the following steps:

1. *Computation of a mean subspace:* Let $\mathbf{A}_i = \mathbf{U}_i \mathbf{R}_i^2 \mathbf{U}_i^T, \quad \forall i = 1, \ldots, N$. Suppose that the subspaces spanned by the columns of the $\mathbf{A}_i$ are enclosed in a geodesic ball of radius less than $\pi/(4\sqrt{2})$ in $\mathrm{Gr}(p, n)$, the Grassmann space. The mean $\mathbf{W} \in \mathrm{V}_{n,p}$ can then be computed as the orthonormal basis of the unique Kärcher mean of the $\mathbf{U}_i \mathbf{U}_i^T$. The computation can be made using a method described in (Sarlette and Sepulchre, 2009).

2. *Computation of the minimal rotations:* For each i, compute two bases $\mathbf{Y}_i$ and $\mathbf{W}_i$ of span$(\mathbf{U}_i)$ and span$(\mathbf{W})$ such that

$$d_{\mathrm{V}_{n,p}}(\mathbf{Y}_i, \mathbf{W}_i) = d_{\mathrm{Gr}(p,n)}(\mathrm{span}(\mathbf{U}_i), \mathrm{span}(\mathbf{W})) \tag{2.48}$$

   This can be viewed as searching for the minimal rotation that brings the flat ellipsoids to the mean subspace span$(\mathbf{W})$, a concept which is illustrated in Figure 2.3.

3. *Rotations of the ellipsoids to the mean subspace:* Compute $\mathbf{S}_i^2 = \mathbf{Y}_i^T \mathbf{A}_i \mathbf{Y}_i$. The ellipsoid $\mathbf{A}_i$ which has been rotated to the mean subspace writes $\mathbf{W}_i \mathbf{S}_i^2 \mathbf{W}_i^T$.

4. *Expression in a common basis and computation of the mean in $\mathrm{S}^+(p)$:* Let $\mathbf{M}$ denote the geometric mean on $\mathrm{S}^+(p)$ (computed using one of the method described above). For all i, compute $\mathbf{T}_i^2 = \mathbf{W}_0^T \mathbf{W}_i \mathbf{S}_i^2 \mathbf{W}_i^T \mathbf{W}_0 \in \mathrm{S}^+(p)$ where $\mathbf{W}_0 \in V_{n,p}$ is a fixed basis of the mean subspace. The geometric mean of the matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_N$ is defined the following way

$$\mathbf{A}_1 \circ \mathbf{A}_2 \circ \cdots \circ \mathbf{A}_N = \mathbf{W}_0[\mathbf{M}(\mathbf{T}_1^2, \mathbf{T}_2^2, \ldots, \mathbf{T}_N^2)]\mathbf{W}_0^T \tag{2.49}$$

It is important to note that, in the same way as the rank-preserving mean between two matrices was not a Riemannian mean, this geometric mean of N matrices is not a Kärcher mean. This is not viewed as a *disadvantage* of the mean, but rather as an interesting feature, since the proposed mean has the advantage of being rank-preserving, and based on explicit formulas.
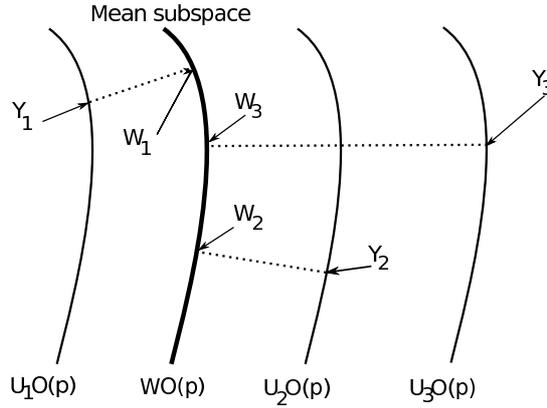
Figure 2.3: The (bold) fiber **WO** represent the mean subspace. Each pair $(\mathbf{Y}_i, \mathbf{W}_i)$ is chosen in order to have a minimal distance between the subspaces.

2.3.3 Properties of the mean

In this section, the fundamental properties of the mean will be studied.

Proposition 2.3.3. *On the set of rank p projectors, the mean* (2.49) *coincides with the Grassmann Riemannian mean. On the other hand, when the matrices in* S$^+(p, n)$ *are all supported by the same subspace, the mean* (2.49) *coincides with the geometric mean induced by* **M** *on the common range subspace of dimension p. More generally, the mean* (2.49) *coincides with* **M** *on the intersection of the ranges.*

Proof The first two properties are obvious, from the construction of the mean (2.49). The last one is due to the special choice of a minimal energy rotation. Indeed, on the intersection of the ranges, the rotation of minimal energy is the identity.

The next proposition will focus on the fact that the mean (2.49) deserves the designation 'geometric'. In fact, the fundamental and axiomatic approach of Ando *et al.* (Ando et al., 2004) reserves the adjective 'geometric' to a definition of mean that enjoys a list of seven properties. For a mean between N matrices, the extension of those properties can be based on the following properties (three of them were presented in Section 2.2.2).

(P1) If $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ commute, their mean is equal to $(\mathbf{A}_1\mathbf{A}_2\mathbf{A}_3)^{1/3}$.

(P2) If s_1, s_2 and s_3 are positive scalar, $s_1\mathbf{A}_1 \circ s_2\mathbf{A}_2 \circ s_3\mathbf{A}_3 = (s_1s_2s_3)^{1/3}(\mathbf{A}_1 \circ \mathbf{A}_2 \circ \mathbf{A}_3)$

(P3) $\mathbf{A}_1 \circ \mathbf{A}_2 \circ \mathbf{A}_3 = (\mathbf{A}_{\pi(1)} \circ \mathbf{A}_{\pi(2)} \circ \mathbf{A}_{\pi(3)})$ for every permutation π of $\{1, 2, 3\}$.

(P4) The map $(\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3) \rightarrow (\mathbf{A}_1 \circ \mathbf{A}_2 \circ \mathbf{A}_3)$ is monotone.

(P5) If $\{\mathbf{A}_{1,n}\}$, $\{\mathbf{A}_{2,n}\}$ and $\{\mathbf{A}_{3,n}\}$ are monotonic decreasing sequences converging to $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ then, $\lim(\mathbf{A}_{1,n} \circ \mathbf{A}_{2,n} \circ \mathbf{A}_{3,n}) = (\mathbf{A}_1 \circ \mathbf{A}_2 \circ \mathbf{A}_3)$.

(P6) For any $\mathbf{G} \in Gl(n)$, we have $(\mathbf{G}\mathbf{A}_1\mathbf{G}^T \circ \mathbf{G}\mathbf{A}_2\mathbf{G}^T \circ \mathbf{G}\mathbf{A}_3\mathbf{G}^T) = \mathbf{G}(\mathbf{A}_1 \circ \mathbf{A}_2 \circ \mathbf{A}_3)\mathbf{G}^T$.

(P7) $(\mathbf{A}_1 \circ \mathbf{A}_2 \circ \mathbf{A}_3)^{-1} = (\mathbf{A}_1^{-1} \circ \mathbf{A}_2^{-1} \circ \mathbf{A}_3^{-1})$.

When the goal is to define a *rank-preserving* mean, some adaptations to these properties must be defined. For example, **(P1)** is impossible to retain, unless the rank of $\mathbf{A}_1\mathbf{A}_2\mathbf{A}_3$ is equal to the rank of $\mathbf{A}_1$, $\mathbf{A}_2$ and $\mathbf{A}_3$. This property is then replaced by

(P1') If $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ commute and are supported by the same subspace, then $(\mathbf{A}_1 \circ \mathbf{A}_2 \circ \mathbf{A}_3) = (\mathbf{A}_1\mathbf{A}_2\mathbf{A}_3)^{1/3}$.

The property **(P6)** is also impossible to retain when dealing with rank-preserving mean. This is in fact the property which causes the limitation of the Ando mean for semi-definite positive matrices. The group of invariances thus have to be replaced by a smaller group. The property becomes

(P6') For $(s, \mathbf{P}) \in \mathbb{R}^* \times O(n)$, we have

$$(s\mathbf{P}^T\mathbf{A}_1s\mathbf{P} \circ s\mathbf{P}^T\mathbf{A}_2s\mathbf{P} \circ s\mathbf{P}^T\mathbf{A}_3s\mathbf{P}) = s\mathbf{P}^T(\mathbf{A}_1 \circ \mathbf{A}_2 \circ \mathbf{A}_3)s\mathbf{P} \tag{2.50}$$

Finally, the inversion of property **(P7)** must of course by replaced with pseudo-inversion.

(P7') $(\mathbf{A}_1 \circ \mathbf{A}_2 \circ \mathbf{A}_3)^\dagger = (\mathbf{A}_1^\dagger \circ \mathbf{A}_2^\dagger \circ \mathbf{A}_3^\dagger)$.

Proposition 2.3.4. *The mean* (2.49) *(with* $\mathbf{M}$ *a geometric mean on* $\mathrm{S}^+(p)$*) is well-defined and deserves the designation 'geometric' as it satisfies the properties* **(P1'),(P2-P5)** *and* **(P6'-P7')**.

Proof

(P1') Taking into account the fact that all the $\mathbf{A}_i$ have the same range, the Prop. 2.3.3 has proved that the mean of those matrices coincides with $\mathbf{M}$, and thus inherits of the **(P1)** property.

(P2) Since $\mathbf{M}$ satisfies this property, the construction of (2.49) ensures that the mean also satisfies the property.

(P3) Obvious since both the Grassmann mean and $\mathbf{M}$ are permutation invariant.

(P4) Suppose $\mathbf{A}_i \preceq \mathbf{A}_i^0$ for each i. The two matrices have the same range and can be written $\mathbf{A}_i = \mathbf{U}_i\mathbf{R}_i^2\mathbf{U}_i^T$, $\mathbf{A}_i^0 = \mathbf{U}_i\mathbf{R}_{i,0}^2\mathbf{U}_i^T$. The respective means have the same range, and **(P4)** is a mere consequence of the monotonicity of $\mathbf{M}$.

(P5) The same arguments as the proof of **(P4)** can be used to prove continuity from above of the mean, as a consequence of continuity of $\mathbf{M}$.

(P6) As for all $s > 0$ and i we have $s\mathbf{A}_i = \mathbf{U}_i(s\mathbf{R}_i^2)\mathbf{U}_i^T$, invariance with respect to scaling is a consequence of the invariance of $\mathbf{M}$. If we denote $\mathbf{O} \in O(n)$, the mean subspace in the Grassmann manifold of the rotated ranges of the $(\mathbf{O}\mathbf{A}_i\mathbf{O}^T)$ is the rotated mean subspace of the ranges of the $\mathbf{A}_i$. Since the matrices are transformed according to $\mathbf{W}_i \to \mathbf{O}\mathbf{W}_i$, the $\mathbf{T}_i$ are unchanged, and the mean of the rotated matrices is $\mathbf{O}\mathbf{W}_0[\mathbf{M}(\mathbf{T}_1^2, \mathbf{T}_2^2, \ldots, \mathbf{T}_N^2)]\mathbf{W}_0^T\mathbf{O}^T$

(P7) For each i, the pseudo-inverse writes $\mathbf{A}_i^\dagger = \mathbf{U}_i\mathbf{R}_i^{-2}\mathbf{U}_i^T$. The calculation of the mean of the pseudo inverse yields the inverse $\mathbf{T}_i^{-1}$ of $\mathbf{T}_i$ and **(P7)** is a consequence of the self-duality of $\mathbf{M}$.

2.3.4 Filtering application

The tools developed in this section can be applied to the processing of diffusion tensor images. Details about those images can be found in the following chapters. For now, we will simply explain that each point of a diffusion tensor image contains a matrix belonging to $S^+(3,3)$. Some of them are highly anisotropic, which means that they possess an elongated shape, with a favorite direction. In this condition, a good approximation of the tensors can be defined by considering on one hand the dominant direction of diffusion ($\in S^+(1,3)$) and on the other hand the non-dominant flat ellipsoid ($\in S^+(2,3)$). The smoothing of these images is performed by an extension of the Perona-Malik algorithm (Perona and Malik, 1990), which we decouple into a filtering problem on the Grassman manifold of one-dimensional subspaces and a filtering problem on the set $S^+(2,3)$.

In the Grassman manifold, using the chordal distance, the Perona-Malik algorithm becomes

$$\mathbf{V}_{i,j}^{t+1} = \mathbf{V}_{i,j}^t + \lambda \frac{\triangledown \mathbf{V}}{|\triangledown \mathbf{V}|}, \tag{2.51}$$

where $\mathbf{V}$ denotes the principal eigenvector of the tensor under consideration and,

$$\triangledown \mathbf{V} = c_N \triangledown_N \mathbf{V} + c_S \triangledown_S \mathbf{V} + c_E \triangledown_E \mathbf{V} + c_W \triangledown_W \mathbf{V}|_{i,j}^t. \tag{2.52}$$

$\triangledown$ denotes a difference with the north, south, east or west neighbor,

$$
\begin{aligned}
\triangledown_N \mathbf{V}_{i,j} &= \mathbf{V}_{i-1,j} - \mathbf{V}_{i,j}, \\
\triangledown_S \mathbf{V}_{i,j} &= \mathbf{V}_{i+1,j} - \mathbf{V}_{i,j}, \\
\triangledown_E \mathbf{V}_{i,j} &= \mathbf{V}_{i,j+1} - \mathbf{V}_{i,j}, \\
\triangledown_W \mathbf{V}_{i,j} &= \mathbf{V}_{i,j-1} - \mathbf{V}_{i,j}.
\end{aligned}
$$

The coefficients are defined by $c_{N_{i,j}}^t = g(|\triangledown_N \mathbf{V}_{i,j}^t|)$, e.g. for the north direction. g is a well-chosen function that allows to diffuse (and thus regularize) along the directions of low gradient but not along the directions of high gradient. This technique preserves the edges in the image, and thus prevents from blurring the structures.

The dominant eigenvalue is regularized with the usual algorithm for scalar images, see (Perona and Malik, 1990).

Once the dominant direction and its corresponding eigenvalue are regularized, we can proceed with the filtering of the non-dominant flat ellipsoid, which belongs to $S^+(2,3)$. For each pixel of the image, all of its four neighbors are mapped to its dominant subspace, and expressed in a common basis (using the ideas described in this section). The 2×2 positive definite matrices can then be filtered using a Perona-Malik algorithm on $S^+(2)$.

The results of this filtering algorithm are illustrated in Figure 2.4, where we can see that the fiber (in red) is well reconstructed by this method. The filtering is compared to a method using the affine-invariant metric (2.19) on $S^+(3)$. Some outliers are not well regularized with this method, an observation which will also be shared by the regularization algorithm using the Log-Euclidean method described in Section 5.2.
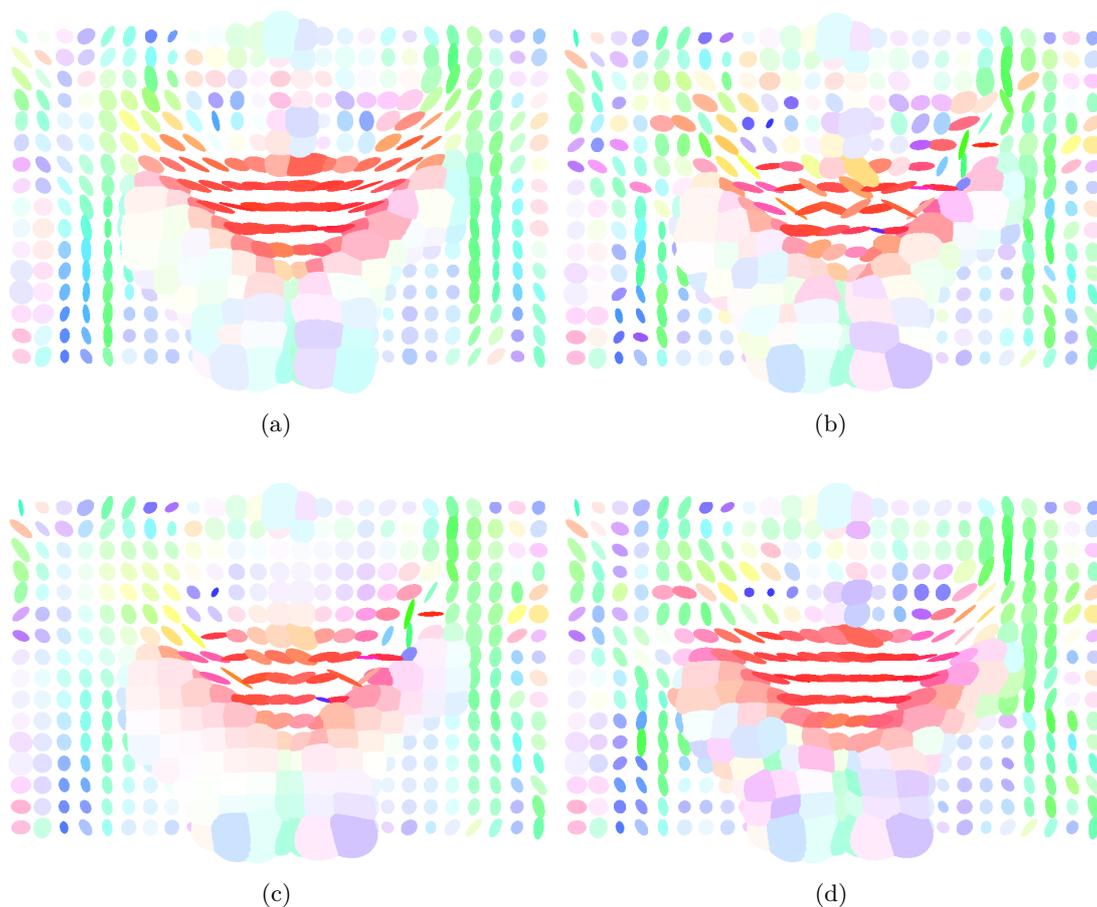
(a)

(b)

(c)

(d)

Figure 2.4: Implementation of a Perona-Malik filter on a Diffusion Tensor Image. (a): Slice of a Diffusion Tensor Image: zoom on a fiber of the corpus callosum (in red). Image courtesy of the Cyclotron Research Centre of the University of Liège. (b): Noise wad added on (a). (c): Result of the regularization with the natural metric on $S^+(3)$. (d): Result of the regularization with an approximation of the matrices, viewed as a dominant direction of diffusion coupled with a non dominant flat ellipsoid. The figure suggests that this regularization performs better than the first one in the reconstruction of the fibre.

2.4 Conclusion

This chapter was devoted to the introduction of the important notions of interpolating curves and of means in manifolds. The manifold of interest in this thesis being the space of symmetric positive definite matrices, the existing methods relative to this space have been presented. This chapter has also introduced the metric of Bonnabel and Sepulchre for the particular manifold of symmetric semi-definite positives matrices of fixed rank. The mean between two of those matrices has been presented. The extension of this mean to many tensors is a contribution of the thesis and the topic of a recent paper, (Bonnabel et al., 2013).

Silvère Bonnabel, Anne Collard and Rodolphe Sepulchre. Rank-preserving geometric means of positive semi-definite matrices. *Linear Algebra and its Applications*, 438(8):3202-3216, 2013.

The addressed notions of this chapter will be of crucial importance in the rest of this manuscript. In particular, the principal messages of this chapter are reminded below.

- The example of positive scalars suggests that invariant processing on nonlinear spaces calls for a Riemannian framework. The selection of relevant metrics is often driven by invariance considerations: the metric is chosen to be invariant with respect to a group action on the considered space.

- Affine-invariant and Log-Euclidean metrics are two different metrics for the same space, $S^+(n)$. If the affine-invariant metric possesses a larger invariance group, the Log-Euclidean distance is simpler to compute than the affine-invariant one. As explained, the affine-invariant geometry is intrinsic, while the Log-Euclidean one is extrinsic. Extrinsic geometry is an approximation of the intrinsic one in the embedding space, but this approximation degrades some important intrinsic notions.

- The Kärcher (or Riemannian) mean is a natural way of defining means on Riemannian manifolds. For positive definite matrices, this mean is well-defined. The complexity of its computation depends upon the chosen metric.

- In Riemannian manifolds, the 'shortest' curve between two points is given by the geodesic. However, interpolating curves should not necessary be geodesics. Computational considerations may suggest other choices. For example, if we consider the space $S^+(p, n)$ of positive semi-definite matrices of fixed-rank, we have proposed a sound notion of interpolating curves which are not the geodesics. Our interpolating curves admit explicit formulas whereas a closed-form expression is not known for the geodesics. The definition of a geometric mean of an arbitrary number of those matrices is a first contribution of this thesis.

Background of Diffusion Tensor Imaging

This chapter provides an historical perspective on Diffusion Tensor Imaging (DTI). We first introduce medical imaging in a general context and then emphasize the physical and information properties of the particular DTI modality. The state-of-the-art processing of those images is also presented. In particular, we show that the widespread Riemannian frameworks suffer from an intrinsic limitation regarding anisotropy of tensors: they tend to degrade this information through averaging.

One of the first motivations for this thesis was the development of a novel framework to process and analyze Diffusion Tensor Images. This imaging method being relatively recent, there is no clear consensus about the processing of the produced images. A diving in the literature has shown that this processing should be geometric, which justifies why we spent some time studying the concepts of curves and means in manifolds. This chapter will take us away from manifolds to explore the second ingredient of this dissertation, the Diffusion Tensor Imaging itself.

After a brief global history of medical imaging, we review the basic principles behind Diffusion Tensor Imaging, going from the Brownian motion of water molecules to the principal equations of this method. Section 3.4 presents early contributions about processing of DT images, and motivates the need for further developments.

3.1   Medical imaging

Medical imaging is a broad field, covering the techniques and processes used to create non-invasively an image of the human body (or parts and functions thereof) for clinical purposes (medical procedures seeking to reveal, diagnose, or examine disease) or medical science (including the study of normal anatomy and physiology). It includes a lot of different techniques, using different signals, aiming at showing different aspects of the human body.

The very first medical imaging technique, still used today, is the *radiography*. This modality uses X-rays for image acquisition. These were discovered in 1895 by Röntgen, whom first 'radiography' was of his wife's hand. The main idea behind diagnostic radiography is that, since the body is made up of various substances of different densities, the

differences in the absorption of the X-rays can be used to reveal the internal structure of the body. Many years later, the X-rays were used to develop the technique called *computed tomography*. This technique utilizes computer-processed X-rays to produce tomographic images (or 'slices') of specific areas of the body. This technique is based on the mathematical theory of the Radon Transform, developed in 1917, but has been used for the first time only in 1961, by William Oldendorf.

Another type of well-known modality of medical imaging is the *ultrasound*. This technique uses high frequency broadband sound waves that are reflected by tissue to varying degrees to produce images. Ultrasound produces less detailed images than computer tomography ones, but offers other advantages, such as the possibility of studying the function of moving structures in real-time and the fact that it does not use any ionizing radiation. Moreover, this technique is relatively inexpensive and quick to perform.

Among the medical imaging techniques using gamma cameras, *Positron Emission Tomography* (PET) is the most famous one. This technique uses short-lived emitting positron, for instance ^{18}F (incorporated in a substance such as glucose) to 'mark' metabolic utilization. The system detects pairs of gamma rays emitted indirectly by this marker, and a computer is used to reconstruct three-dimensional images of marker concentration in the body. Since the positron is associated to glucose, the concentrations of marker will indicate tissue metabolic activity, through the regional glucose uptake. One principal limitation to this technique is the high cost of cyclotrons needed to produce the radionuclides.

The fourth main type of medical images comes from *Magnetic Resonance Imaging* (MRI). Broadly speaking, this method uses powerful magnets to polarize and excite single protons in water molecules in human tissues. Magnetic field gradients cause nuclei at different locations to precess at different speeds, which allows spatial information to be recovered using Fourier analysis of the measured signal. Compared to CT or radiography, this technique has no known long-term effects of exposure to strong static fields, and there is no limit to the number of scans to which an individual can be subjected.

This thesis is about Diffusion Tensor Imaging, a method which uses the same devices and concepts than MRI. In the following, the main concepts of this modality will be defined, from the basic principle of Brownian motion to the final diffusion tensor images.

3.2 Basic principles of diffusion MRI

With respect to the the history of medical imaging, diffusion MRI is a newborn. If the Nuclear Magnetic Resonance (NMR) technique dates back to the 50's, the basis of diffusion MRI are due to (Le Bihan and Breton, 1985; Le Bihan et al., 1986), in 1985, and the first papers mentioning Diffusion Tensor Imaging (DTI) only appeared in the 90's (Basser et al., 1992, 1994a). However, this field has received broad attention, because it helps understanding the structural connections between areas of the brain. This type of information could be of great interest in the characterization of neuro-degenerative diseases, in surgical planning, in the study of the evolution of the structure of the brain and in other medical applications. Diffusion MRI has the capability of accurately describing the geometry of the underlying cerebral microstructures. This is made possible by the measure of the average diffusion of water molecules in the brain. This diffusion probes the

structure of the biological tissues at scales much smaller than the imaging resolution. The main physical concept related to diffusion MRI is the fact that under normal conditions, the diffusion of water molecules is Brownian. However, in fibrous structure such as the axons of the brain, water molecules tend to diffuse along fibers. Information about the neural structure in vivo can thus be obtained through the diffusion MRI modality. It is the only known modality to study the architecture of the brain in a noninvasive way. The basic physical principles of diffusion MRI will be briefly reviewed in the following.

3.2.1 Brownian motion

At a microscopic scale, the movement of water molecules in an *isotropic* medium is described by the Brownian motion. This motion is due to the fact that the molecules freely move and collide with each other. Figure 3.1 illustrates the Brownian motion, for a unique particle.

Figure 3.1: Illustration of a Brownian motion, for an unique particle. The red dot represents the origin of the movement.

At a macroscopic scale, this phenomenon yields a diffusion process. This is commonly illustrated as a drop of a water-soluble dye placed in a glass of water, which will spread out and whose color will become less intense until the glass is filled with a solution of uniform color. This phenomenon can be described by Fick's first law, derived by Adolf Fick in 1855 (Fick, 1855). This law explains how a population of particles is transported from regions of high concentration (the drop of dye) to regions of low concentration (the water) so as to decrease the concentration gradient. Mathematically, if $\mathbf{J}$ is the flux of particles, it is related to the concentration gradient $\nabla \mathbf{C}$ of the substance C by a proportional constant D, which is known as the diffusion coefficient. The governing equation of the diffusion process is thus

$$\mathbf{J} = -D\nabla \mathbf{C}\,. \tag{3.1}$$

The molecular displacements obey a three-dimensional Gaussian distribution, which means that the molecules travel randomly in space over a distance that is statistically well described by a diffusion coefficient D. This coefficient was related by Einstein (Einstein, 1956) to the root mean square of the diffusion distance as

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

where τ is the diffusion time, $\langle \dots \rangle$ denotes an ensemble average and $\mathbf{R}$ is the net displacement vector, *i.e* $\mathbf{R} = \mathbf{r} - \mathbf{r}_0$ with $\mathbf{r}_0$ the original position of a particle and $\mathbf{R}$ its position after the time τ.

In *anisotropic* tissues, such as most biological ones, things are a little bit different. The mobility of water molecules is indeed constrained by obstacles formed by surrounding structures, such as the axons in the brain. The actual diffusion distance of water molecules is reduced compared with that of free water, and the displacement distribution is no longer Gaussian. The general diffusion displacement probability density function (PDF), also called diffusion propagator, is extremely complex and still unknown today. This explains why simple *models* of diffusion have been proposed. Among those, the most popular one is the diffusion tensor (DT) model (Basser et al., 1994a), which is derived form the transport tensors theory previously studied in the mathematical literature. For this simplified model, Einstein's and Fick's laws of diffusion can be generalized, by replacing the scalar diffusion coefficient D which explains diffusion in isotropic medium with an operator $\mathbf{D}$. This operator is obtained by considering the covariance matrix of the net displacement vector, *i.e*

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

The probability density function of the diffusion displacement, denoted $P(\mathbf{R}, \tau)$, represents the probability that a water molecule originally located at $\mathbf{r}_0$ will have moved by amount $\mathbf{R} = \mathbf{r} - \mathbf{r}_0$ in time τ. Using the diffusion tensor model, the Taylor's expansion (ignoring the higher order terms) of P about $\mathbf{R}$ and τ is given by

$$\frac{\partial P(\mathbf{R}, \tau)}{\partial \tau} = \mathbf{D}\nabla^2 P(\mathbf{R}, \tau) \,, \tag{3.4}$$

using the Einstein's relation (3.3) and performing some mathematical derivations which are explained in (Campbell, 2004). The solution of this partial differential equation is P, the diffusion PDF of water molecules under the DT model assumption.

If we use the Fick's formulation, the diffusion tensor $\mathbf{D}$ can also be introduced to model the anisotropic diffusion of biological tissues. In this case, the first law becomes

$$\mathbf{J} = -\mathbf{D}\nabla\mathbf{C} \,. \tag{3.5}$$

Using this law together with the law of conservation of mass, *i.e.* $\partial C / \partial t = -\nabla \cdot \mathbf{J}$, we obtain

$$\frac{\partial C}{\partial t} = \nabla \cdot (\mathbf{D}\nabla\mathbf{C}) = \mathbf{D}\nabla^2 C \,, \tag{3.6}$$

which is the equivalent of equation (3.4) with the diffusion PDF P replaced by the concentration of the medium C. Both laws describe the classical diffusion equation. If $\mathbf{D}$ is an identity matrix, which means that the medium is isotropic, the law of diffusion is called the heat equation. For anisotropic medium, it is called the geometric heat equation.

3.3 From Nuclear Magnetic Resonance to Diffusion Tensor Imaging

In this section, we review the main steps that have been necessary for the development of Diffusion Tensor Imaging. Many of those concepts are shared with its 'older brother', the Magnetic Resonance Imaging, and have been defined years before the first Diffusion Tensor Image.

3.3.1 Historical notes

The key concepts of MRI (and of diffusion MRI) were used for a long time before the invention of these methods. Nuclear Magnetic Resonance (NMR) was indeed used to measure properties of substances in chemistry, and to perform NMR spectroscopy.

Felix Bloch and Edward Mills Purcell both described NMR in 1946, in separate works (Bloch et al., 1946; Purcell et al., 1946). They received a shared Nobel Prize in Physics in 1952 for their discovery. NMR's basic principle is the following: if a (very) strong external magnetic field is applied, magnetic nucleus as ^1H (proton) will align with this field. A perturbation of the alignment by an electromagnetic field will produce a characteristic response of the nucleus. In 1950, Herman Carr proposed to create the first one-dimensional MR images by introducing a *gradient* in the magnetic field (Carr and Purcell, 1954). In the same time, Hahn published his seminal paper (Hahn, 1950) on the NMR spin echo. In this paper, the authors 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 notion is fundamental to understand the principles of diffusion MRI.

In 1973, a method was proposed by Lauterbur (Lauterbur, 1973) which used gradients of magnetic fields to reconstruct two dimensional MR images. This method was further developed by Mansfield (Mansfield, 1977), who proposed a new ultrafast acquisition technique, known as the *echo-planar technique.* The 2003 Nobel prize in Physiology and Medicine was awarded to those two researchers for their discoveries in MRI.

The last step towards diffusion MRI was the development of the gradient spin echo sequence of Stejskal and Tanner (Stejskal and Tanner, 1965), explained below.

3.3.2 Pulse Gradient Spin Echo (PGSE)

The imaging sequence described by Stejskal and Tanner is used to measure the diffusion of water molecules in a given direction $\mathbf{g}_i$, $i = 1, \ldots, N$. The principal idea of this sequence is to use a pair of sharp magnetic field gradient pulses, the duration and the separation of them which can be adjusted. In an otherwise homogeneous field, the first pulse has for goal to magnetically 'labels' the protons that are carried by water molecules, according to their spatial location. This is possible since, for a short time, the magnetic field slowly varies along one direction. The second pulse is introduced slightly later to detect the changes in the location of nuclei, *i.e.* the displacement of nuclei that occurred during the

time interval between the two pulses. This time interval is also called 'diffusion time'. A change in location (along the gradient direction) of a proton carried by a diffusing water molecule results in a change of the magnetic field 'seen' by this proton. This change is proportional to the displacement. If we consider a population comprising a large number of water molecules, the overall effect is that the attached protons will experience various magnetic field changes, that closely reflect the statistical displacement distribution of this population, that is, the statistical diffusion process. This variation of the magnetic field seen by the population of nuclei results in an MRI radiowave signal that slightly differs from the one obtained from a population of nuclei in a perfectly homogeneous field.

The *attenuation of the signal* can be mathematically expressed, assuming the pulses to be infinitely narrow (this hypothesis is called the *narrow pulse approximation* and means that the duration of the gradient pulse is short enough to consider that the diffusion of water molecule during that time is negligible). Stejskal and Tanner showed that the signal attenuation $S(\mathbf{q}, \tau)$ can be expressed as the three-dimensional Fourier Transform of the ensemble average propagator P.

$$\frac{S(\mathbf{q}, \tau)}{S_0} = \int_{\mathbb{R}^3} P(\mathbf{r}|\mathbf{r}_0, \tau) \exp(-2\pi i \mathbf{q}^T \mathbf{R}) d\mathbf{r} = \mathcal{F}[P(\mathbf{r}|\mathbf{r}_0, \tau)], \qquad (3.7)$$

where $\mathbf{q} = \gamma \delta \mathbf{G}/2\pi$, with γ the nuclear gyromagnetic ratio for water protons, $\mathbf{G}$ the applied diffusion gradient vector, δ the gradient pulse duration, S_0 is the baseline image (acquired without any diffusion gradients, also called the $b = 0$ image) and $P(\mathbf{r}|\mathbf{r}_0, \tau)$ is the diffusion PDF introduced earlier, which is the function that we want to reconstruct. This reconstruction will only be possible by sampling the diffusion PDF along many $\mathbf{q}$ vectors. The space of all possible 3D $\mathbf{q}$ vectors is called the *q-space*.

If the diffusion PDF is assumed to be Gaussian, the Fourier integral can be computed analytically. The signal attenuation equation becomes

$$S(\mathbf{q}, \tau) = S_0 \exp(-\tau \mathbf{q}^T \mathbf{D} \mathbf{q}), \qquad (3.8)$$

where $D(\mathbf{q}) = \mathbf{q}^T \mathbf{D} \mathbf{q}$ is the Apparent Diffusion Coefficient (ADC). Indeed, it should be noted that the diffusion coefficient measure in diffusion MRI is not the *true* diffusion coefficient of water, unless the imaging object is pure water. In fact, the structure of the tissue restricts the diffusion. The measure of this restricted diffusion is called the Apparent Diffusion Coefficient.

The signal attenuation is often written with respect to a unit vector, $\mathbf{g} = \mathbf{q}/|\mathbf{q}|$. In this case, it is common to introduce the b-value, $b = \tau |\mathbf{q}|^2$ in the equation, which becomes

$$S(b, \mathbf{g}) = S_0 \exp(-b \mathbf{g}^T \mathbf{D} \mathbf{g}) = S_0 \exp(-b \, \mathrm{ADC}). \qquad (3.9)$$

Equation (3.9) is the most common formulation of the *Stejskal-Tanner equation under a Gaussian assumption*. It is at the heart of diffusion-weighted imaging (DWI).

3.3.3 Diffusion-Weighted Imaging (DWI)

The first diffusion MRI acquisitions were made in 1984-1985. At this time, diffusion NMR and available imaging techniques were put together to make those acquisitions. If Wesbey *et al* were the first to demonstrate DWI, their sequence was however not clinically feasible

(Wesbey et al., 1984a,b). The first DWI acquisition in vivo of the human brain using a whole-body scanner is due to Le Bihan *et al*, in 1985 (Le Bihan and Breton, 1985; Le Bihan et al., 1986). These papers also introduce the now famous *b*-value. At that time, a DWI was simply the unprocessed result of the application of the PGSE sequence in *one gradient direction.*

A few years later, Moseley *et al* (Moseley et al., 1990) had the idea to measure diffusion anisotropy in more than one direction. Their measurements were made on a cat brain. At this time, it was already presumed that diffusion is faster in the direction of fibers than in perpendicular directions. Using the log version of equation (3.9), *i.e.*

$$\log\left(\frac{S(b,\mathbf{g})}{S_0}\right) = -b\,\mathrm{ADC}\,, \qquad (3.10)$$

Moseley proposed to measure this ADC along the x and z axis and to use the anisotropy index $\mathrm{ADC}_z/\mathrm{ADC}_x$ to characterize the level of anisotropy in the tissue. This idea requires the acquisitions of at least two images.

The idea that diffusion MRI and measures of the ADC along two directions could be used to determine orientations of fiber bundles in the brain was first stated in 1991 by Douek (Douek et al., 1991). The problem at that time was that ADC measures and anisotropy index were very dependent on the direction of the gradient encoding used in the acquisition. Soon, the need for rotationally invariant measures became clear, along with a mathematical model with invariant properties to describe the 3D displacement distribution of water diffusion. The solution to this problem was provided by Peter Basser, in 1994, who introduced the Diffusion Tensor Imaging. This model overcomes the limitations of the preceding measures, and enables the introduction of rotationally invariant measures (Basser et al., 1994a).

3.3.4 Diffusion Tensor Imaging (DTI)

In his seminal paper of 1992 (Basser et al., 1992), Basser proposed to use a *second order symmetric and positive-definite tensor* to model the intrinsic diffusion properties of biological tissues, and particularly of the brain. This tensor is in fact the same as introduced earlier in Einstein's equations for anisotropic diffusion, Eqs. (3.3) and (3.5). As already explained, the solution of equation (3.4) is the Gaussian diffusion PDF. The probability P to find a molecule at position $\mathbf{r}$ which initially was at position $\mathbf{r}_0$ after a delay τ is given by

$$P(\mathbf{r}|\mathbf{r}_0,\tau) = \frac{1}{\sqrt{(4\pi\tau)^3|\mathbf{D}|}}\exp\left(-\frac{1}{4\tau}(\mathbf{r}-\mathbf{r}_0)^T\mathbf{D}^{-1}(\mathbf{r}-\mathbf{r}_0)\right)\,, \qquad (3.11)$$

where $|\mathbf{D}|$ is the determinant of the diffusion tensor $\mathbf{D}$. The idea of Basser is thus to approximates the diffusion PDF by a 3-variate normal distribution with zero mean. In this case, $\mathbf{D}$ can be viewed as the covariance matrix describing the Brownian motion of water molecules at each imaging voxel.

The symmetry of $\mathbf{D}$ implies that it has six unknown coefficients, which have to be estimated. This means that at least 6 DW images (and one unweighted diffusion image) have to be acquired to solve the system of equations. In practice, due to noise in the measures, more than 6 DW images are used. Typically, a *b*-value of 1000 s/mm^2 is used with 7 to 64 gradient directions.

Estimation of diffusion tensors through DWI acquisitions has been the subject of much research in the past fifteen years. First papers used classical linear or non-linear least-squares estimation methods (Basser et al., 1994b), while more recent papers are based on Riemannian frameworks (Lenglet, 2006; Arsigny, 2006; Fillard et al., 2007). Diffusion tensors are indeed elements of the Riemannian manifold of symmetric positive-definite matrices. The estimation methods based on Riemannian computations prevent degenerate tensors (non positive or non symmetric matrices).

Principal coordinate axes and principal diffusivities

The diffusion tensors have been shown to be a rich mathematical tool, with interesting properties that one can exploit for diffusion MRI visualization and analysis. As already explained in Basser (Basser et al., 1994a), for each estimated tensor, a local orthogonal coordinate system can be constructed. This system represents the principal coordinate axes, along which diffusive fluxes and concentration gradients are decoupled. The corresponding diffusion coefficients in these three principal directions (the principal diffusivities) can also be calculated. In fact, due to the symmetry and positive-definiteness of the tensor, its three eigenvectors are orthogonal, and they thus represent the principal coordinate directions. The three eigenvalues, which are positive, represent the principal diffusivities. If we denote by λ_i, $i = (1, 2, 3)$ the eigenvalues and by $\mathbf{e}_i$ the eigenvectors, we can write the *spectral decomposition* of a tensor $\mathbf{S}$ as

$$\mathbf{S} = \mathbf{U}\Lambda\mathbf{U}^T, \tag{3.12}$$

with $\mathbf{U} = (\mathbf{e}_1|\mathbf{e}_2|\mathbf{e}_3)$ and

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}. \tag{3.13}$$

The three eigenvalues can be ordered, to have $\lambda_1 \geq \lambda_2 \geq \lambda_3$. The main idea behind diffusion imaging was that in ordered structures such as brain white matter, the macroscopic anisotropy described by the tensor is due to microscopic heterogeneity - in particular to ordered, semipermeable membranes that retard diffusion. This idea is illustrated in Figure 3.2, left. The diffusion is easy along the fibers, but reduced in a direction perpendicular to them. This means that, in anisotropic tissues, the principal directions of the tensor must coincide with orthotropic directions of that tissue. In particular, the eigenvector associated with the largest eigenvalue λ_1 defines the 'fiber-tract axis' of the tissue. The two remaining eigenvectors, which are perpendicular to the first one, should define the two remaining orthotropic axes. This spectral decomposition is illustrated in Figure 3.2, (b). If $\lambda_1 = \lambda_2$, the tensor is said to be planar. The tensor is isotropic if all the eigenvalues are equal. Isotropic tensors are usually associated to isotropic medium, as gray matter or the cerebrospinal fluid.

Starting from the spectral decomposition of the tensor, its graphical representation can be visualized by a diffusion *ellipsoid*. The principal axes of the ellipsoid are given by the eigenvectors of the tensors, while the length of these axes is related to the eigenvalues. Ellipsoids can also be viewed as a three-dimensional representation of the diffusion distance covered in space by molecules in a given diffusion time. In the case of isotropic diffusion, the ellipsoid is a sphere, the size of which is proportional to the diffusion coefficient. In
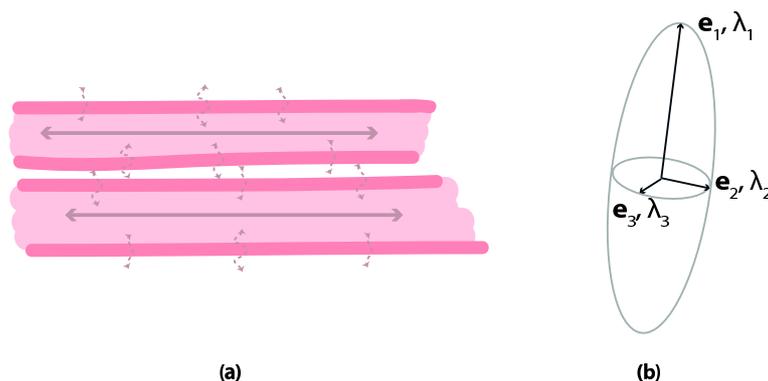
Figure 3.2: **(a):** Diffusion of water molecules is easier in the direction of fibers than in directions perpendicular to them. **(b):** Illustration of the spectral decomposition of tensors. The three principal directions of the ellipsoid correspond to the eigenvectors of the tensor. Their lengths are related to the corresponding eigenvalues.

the case of anisotropic diffusion, ellipsoids become elongated (cigar-shaped) if one diffusion direction predominates. If two directions are more important than a third one, ellipsoids are flat (pancake-shaped). These different shapes are illustrated in Figure 3.3.
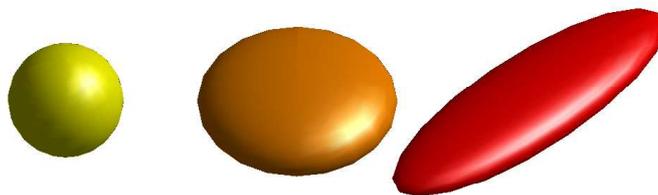
Figure 3.3: Diffusion tensor as ellipsoids. The color of ellipsoids is related to their anisotropy. From left to right: isotropic tensor, planar tensor (flat ellipsoid) ($\lambda_1 \simeq \lambda_2 > \lambda_3$), elongated ellipsoid ($\lambda_1 \gg \lambda_2 \geq \lambda_3$)

3.4 Early DTI processing

Parallel to the development of the imaging modality arose the question of how to process and analyze these new images. Many efficient methods had been developed for the processing of (scalar) Magnetic Resonance images. However, as it will be shown in the following, scalar processing of DT images is usually focused on a single type of information, forgetting all the other information. This has lead researchers to develop new, matrix-valued processing algorithms.

3.4.1 Projection on scalar maps

As explained above, one of the main reason to use diffusion tensor to model diffusion is to identify quantities that are independent of the fiber direction. This explains why many authors have defined different scalar invariants of tensors. In this context, 'invariants'

means that the scalar measures do not depend on the fiber direction. All these invariants are spectral quantities, that is, they are based on the eigenvalues of the tensor.

- **Trace** The trace of the tensor is given by $\text{Tr}(\mathbf{S}) = \text{Tr}(\Lambda) = \lambda_1 + \lambda_2 + \lambda_3$. The trace is related to the mean diffusivity in the voxel, which is equal to $\text{Tr}(\mathbf{S})/3$ (Basser et al., 1994a).

- **Determinant** The determinant is given by $\text{Det}(\mathbf{S}) = \text{Det}(\Lambda) = \lambda_1 \lambda_2 \lambda_3$. It measures the volume of the ellipsoid (Basser et al., 1994a).

- **Volume Ratio** The Volume Ratio (VR)(Pierpaoli et al., 1994) is given by

$$\text{VR} = \frac{\lambda_1 \lambda_2 \lambda_3}{\left(\frac{\lambda_1 + \lambda_2 + \lambda_3}{3}\right)^3} = 27 \frac{\text{Det}(\mathbf{S})}{\text{Tr}(\mathbf{S})^3} \, . \tag{3.14}$$

  This index has a geometrical interpretation: it represents the volume of an ellipsoid whose semi-major axes are the three eigenvalues of $\mathbf{S}$ divided by the volume of a sphere whose radius is the mean diffusivity ($\text{Tr}(\mathbf{S})/3$). A value of VR close to zero indicates an high anisotropy while a value of 1 represents complete isotropy.

- **Fractional Anisotropy** This index is very popular and one of the most used in the DTI literature. It is defined as (Basser and Pierpaoli, 1996)

$$\text{FA} = \sqrt{\frac{3}{2}} \frac{\sqrt{\sum_{i=1}^{3}(\lambda_i - \lambda_m)^2}}{\sqrt{\sum_{i=1}^{3} \lambda_i^2}} \tag{3.15}$$

  with $\lambda_m = (\lambda_1 + \lambda_2 + \lambda_3)/3$. FA measures the fraction of the 'magnitude' of the tensor that can be ascribed to anistropic diffusion. It goes from zero for isotropic tensor to one for anisotropic ones.

- **Relative Anisotropy** This index was also proposed in (Basser and Pierpaoli, 1996), and is given by

$$\text{RA} = \frac{1}{\sqrt{3}} \frac{\sqrt{\sum_{i=1}^{3}(\lambda_i - \lambda_m)^2}}{\lambda_m} \, . \tag{3.16}$$

  This index is in fact the ratio between the anisotropic part of the tensor and the magnitude of its isotropic part. It is equal to zero for isotropic tensors.

A comparison of different anisotropy indices is presented in (Pierpaoli and Basser, 1996), which shows the importance of rotationally invariant indices. This work also points out how the sorting of eigenvalues is affected by the background noise, therefore claiming that indices which are not based on this sorting are more robust to noise.

Many works have used scalar invariants to analyze DT images. The principle is to construct FA, or RA maps, and to (only) analyze those scalar images, instead of the tensor images. This is for example the case in (Jones et al., 2002), where the registration of images is based on FA maps, and in (Alexander et al., 2001) where RA maps are used. The Tract-Based Spatial Statistics (TBSS) software (Smith et al., 2006) is also based on FA data. As it will be shown in Chapter 6, the Fractional Anisotropy has also been widely used for statistical analyses of DT images.

However, in (Alexander and Gee, 2000), the authors used their registration algorithm with different similarity measures, based either on scalar differences or on more complex measures. They conclude from their tests that scalar measures are not appropriate for the processing of Diffusion Tensor Images. This work, among others, motivates multivariate methods to study DT images. Those methods are briefly reviewed in the next section.

3.4.2 Multivariate DTI processing

Different methods have been proposed for tensor analysis. Early multichannel algorithms used, for example, the 6 components of the tensor, or its three eigenvalues (Guimond et al., 2002), or a combination of those channels (Park et al., 2003). The preferred information depends upon the work and the evaluation methods. In (Leemans et al., 2005), another type of algorithm is used, which utilizes the DWI as the information channels. Another point of view of multivariate processing is the one of (Alexander and Gee, 2000), which perform multivariate processing without using multichannels. The principal information here is a *similarity measure* between tensors. This measure can be based on scalar derived from tensors, or be a simple Euclidean distance between tensors, *i.e.*

$$d(\mathbf{S}_1, \mathbf{S}_2) = \sqrt{\mathrm{Tr}(\mathbf{S}_2 - \mathbf{S}_1)^2}\,. \tag{3.17}$$

This distance is also used in (Cao et al., 2006; Goh and Vidal, 2006). However, the Euclidean processing of tensors has a well-known disadvantage: the interpolation of tensors suffers from the *swelling effect*, *i.e.* the volume of the tensors is not monotonically interpolated and goes through a maximum. This effect is physically unrealistic, since it amounts to the introduction of more diffusion.

The growing interest for multivariate processing slowly led researchers to focus on the *constraints* that the methods have to fulfill. One example of those constraints concerns the Principal Direction Diffusion (PDD), *i.e.* the eigenvector associated to the principal eigenvalue of the tensor. This vector indeed has to be of unit norm. Moreover, it should be noted that this eigenvector has *no sign*, *i.e.* if the PDD is denoted by $\mathbf{v}_1$, it is indistinguishable of $-\mathbf{v}_1$. Different regularization algorithms have been proposed, as for example in (Chen and Hsu, 2005; Coulon et al., 2004). The latter one also introduces the idea of decoupling the processing between the orientation of one hand and the diffusivities on the other hand. This idea is similar to the one of this thesis, which will further be explained in Chapter 4.

3.4.3 Riemannian processing

The idea of developing algorithms which work directly on the manifold of tensors, *i.e.* the manifold of symmetric positive definite matrices $\mathrm{S}^+(3)$ naturally emerged from the idea of processing eigenvectors as elements of the manifold $\mathbf{S}^2$ (the unit sphere). This idea is at the core of the affine-invariant framework.

Affine-invariant framework

The Riemannian metric and the geodesic in the space $\mathrm{S}^+(n)$ have been explained in Section 2.1.2. In this section, we focus on the space of diffusion tensors, $\mathrm{S}^+(3)$. As a reminder, the geodesic between $\mathbf{S}_1$ and $\mathbf{S}_2$ is given by

$$\mathbf{S}(t) = \mathbf{S}_1^{1/2} \exp(t \log(\mathbf{S}_1^{-1/2} \mathbf{S}_2 \mathbf{S}_1^{-1/2})) \mathbf{S}_1^{1/2}\,, \tag{3.18}$$

while the distance between tensors, corresponding to the length of geodesic, is

$$d(\mathbf{S}_1, \mathbf{S}_2) = || \log(\mathbf{S}_1^{-1/2} \mathbf{S}_2 \mathbf{S}_1^{-1/2})||_2 \,. \tag{3.19}$$

The use of the affine-invariant Riemannian framework was first proposed for DTI processing in 2006, see (Pennec et al., 2006; Lenglet et al., 2006), see also (Fletcher and Joshi, 2007; Fletcher et al., 2009; Castaño-Moraga et al., 2006; Gur and Sochen, 2007; Batchelor et al., 2005; Lenglet et al., 2009). The affine-invariant geometry is the natural geometry of $S^+(3)$ considered as the homogeneous space $Gl(3)/O(3)$, see (Smith, 2005).

Together with the development of geometric processing algorithms, a novel geometric anisotropy index has been proposed by Fletcher in (Fletcher and Joshi, 2007), which uses the logarithm of the eigenvalues. This index is called the geodesic anisotropy.

$$\mathrm{GA} = \sqrt{\sum_{i=1}^{3} (\log(\lambda_i) - \log(\lambda_{gm}))^2} \,, \tag{3.20}$$

with $\lambda_{gm} = \sqrt[3]{\lambda_1 \lambda_2 \lambda_3}$. It can be interpreted as the distance between the tensor and its closest isotropic tensor.

One example of (geodesic) interpolation between two tensors is shown in Figure 3.4 (top). The two tensors have the same shape, but different orientations. It can be observed that in this case, the framework performs both a modification of the orientation and the shape. In particular, the midpoint of the interpolation curve between two anisotropic tensors tends to be isotropic. Since the anisotropy is one of the most important information given by diffusion tensors, this decrease of anisotropy is a limitation of the affine-invariant framework for the processing of DT images. Moreover, as already explained in 2.1.2, this framework is computationally expensive.
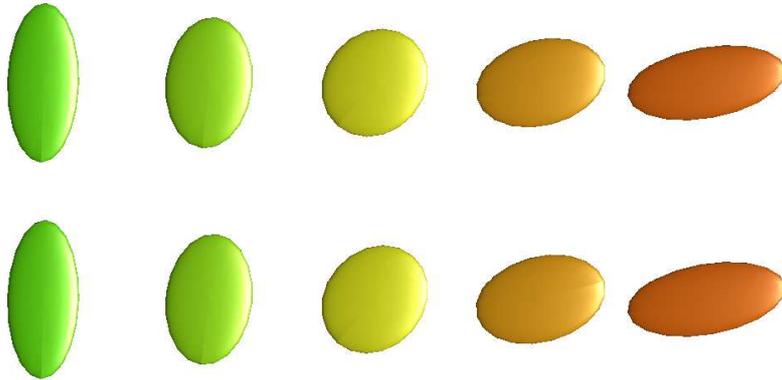

Figure 3.4: Geodesic interpolation between two tensors. *Top*: with the affine-invariant framework. *Bottom*: with the Log-Euclidean framework. It is clear from those figures that the Log-Euclidean framework is a good approximation of the affine-invariant one. However, both frameworks suffer from the same limitation regarding anisotropy.

Log-Euclidean framework

Section 2.1.2 has introduced the Log-Euclidean framework as an approximation of the affine-invariant framework, with the great advantage of being computationally cheaper. This advantage is the main reason why the Log-Euclidean method has been proposed to process DT images (Arsigny et al., 2006). The geodesic curve is in this case

$$\mathbf{S}(t) = \exp((1-t)\log(\mathbf{S}_1) + t\log(\mathbf{S}_2)) \tag{3.21}$$

and the distance between the tensors is

$$d(\mathbf{S}_1, \mathbf{S}_2) = ||\log(\mathbf{S}_1) - \log(\mathbf{S}_2)||_2. \tag{3.22}$$

The method became a popular framework and has been used by many authors (Goodlett et al., 2009; Chiang et al., 2008; Ingalhalikar et al., 2010; Castro et al., 2007; Weldeselassie and Hamarneh, 2007; Arsigny et al., 2006; Fillard et al., 2007; Yeo et al., 2009, 2008; Lepore et al., 2006; Awate et al., 2007). Figure 3.4 (bottom) suggests that it is a good substitute for the affine-invariant interpolation but that it suffers the same limitation regarding anisotropy.

This limitation was at the core of the development proposed in this thesis, which will be explained in details in the following chapter.

Figure 3.5 presents a timeline of the principal notions used to develop and study DTI, until the introduction of Riemannian methods for the processing of these images.
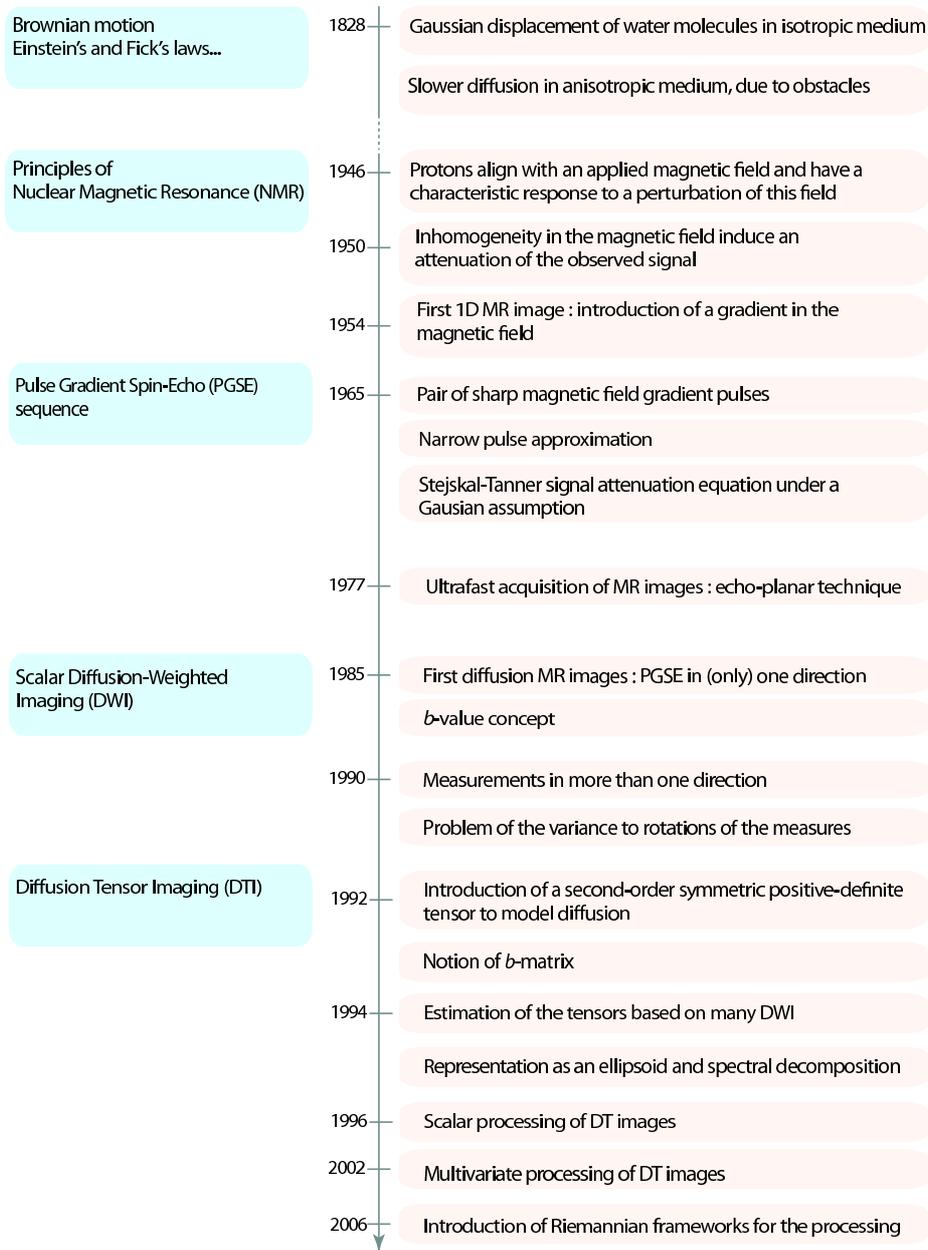
Summary

Figure 3.5: Illustrative summary of the chapter: timeline of the principal notions used to develop DTI.

Anisotropy preserving DTI processing

The objective of this thesis was to develop a suitable framework for the processing of Diffusion Tensor Images. This chapter introduces this novel framework, which is based upon two main motivations.

The first motivation is to avoid the limitation of the existing Riemannian frameworks regarding anisotropy of tensors. To address this limitation, we use a decoupling of the processing between intensities and orientation of diffusion, through the spectral decomposition of tensors.

Our second motivation concerns the computational cost of the processing operations, which has to be efficient in order to scale with data size of clinical applications. To this end, we use unit quaternions to represent the orientation of tensors. The combination of those two ideas explain the name 'spectral-quaternion' for our framework.

Besides those two concepts, we also deal with uncertainty of the orientation in isotropic tensors.

The chapter proposes novel algorithms for the basic processing operations: interpolation, averaging, and similarity measure. Those operations are invariant by scaling and rotation, and have the remarkable property of preserving all the characteristics of tensors.

The contributions of this chapter are presented in:

As emphasized in (Le Bihan, 2003), there are three 'meta-parameters' which can be derived from the diffusion tensor and are the three main sources of information about tissue microstructure and architecture of each voxel.

Those parameters are

- the *mean diffusivity*, which characterizes both the overall mean-squared displacement of molecules and the overall presence of obstacles to diffusion. This parameter is also related to the average ellipsoid size.

- the *main direction of diffusivities* (given by the eigenvectors) which is linked to the orientation in space of the structures.

- the *degree of anisotropy*, which describes the degree to which molecular displacements vary in space. The anisotropy is related to the presence and coherence of orientated structure. This information is related to the eccentricity of the ellipsoid and can be associated to many different indices, as seen in Chapter 3.

The major importance of those three parameters motivates a framework that preserves them during the processing, in order to preserve the information they convey. As explained in the preceding chapter, a main limitation of the Euclidean framework is its swelling effect on the mean diffusivity. Likewise, a main limitation of the affine-invariant and Log-Euclidean frameworks is to decrease the anisotropy of tensors along the interpolating curves. Those limitations motivated the development of an 'anisotropy preserving' framework in this thesis.

The idea of decoupling the processing between intensities of diffusion and orientation (see Section 4.1) , which seems the best way to preserve anisotropy, is not new, see for example (Tschumperlé and Deriche, 2001; Weldeselassie et al., 2009; Ingalhalikar et al., 2010). However, those methods did not become popular mostly due to their expensive computational cost, due to the difficult processing of orientations. Our second concern in this thesis has thus been to overcome this computational obstacle thanks to the use of unit quaternions, as it will be explained in Section 4.2.

The principal features of the developed framework are presented in Section 4.3, while Section 4.4 is devoted to the extension of interpolating curves to weighted means of more than two tensors. Finally, the important notion of similarity measure is the subject of Section 4.5.

4.1   Interpolating curves based on the spectral decomposition

The development of our anisotropy preserving framework is based on the spectral decomposition of tensors. As stated in (3.12), this decomposition writes $\mathbf{S} = \mathbf{U}\Lambda\mathbf{U}^T$ where $\mathbf{U}$ is the orthogonal matrix containing the eigenvectors and Λ is a diagonal matrix whose elements are the eigenvalues $\lambda_i, i = 1, 2, 3$. Since tensors are symmetric positive definite matrices, $\Lambda \in \mathrm{D}_+(3)$, the space of diagonal matrices with positive elements, and the matrix $\mathbf{U}$ belongs to the orthogonal group $\mathrm{O}(3)$. In this work, we will order the eigenvalues of the tensors such that $\lambda_1 \geq \lambda_2 \geq \lambda_3$ and we will impose $\det(\mathbf{U}) = 1$, i.e., $\mathbf{U}$ belongs to the space of special orthogonal matrices, $\mathrm{SO}(3)$.

The spectral decomposition of tensors enables to perform a decoupling between the intensities of diffusion (the eigenvalues) and the orientation (the eigenvectors). This decoupling is at the core of the spectral interpolating curve.

Definition 4.1.1. *The spectral interpolating curve between two diffusion tensors $\mathbf{S}_1$ and $\mathbf{S}_2$ is based on the decoupling between intensities and orientation of diffusion, and is given by the following equations.*

$$\begin{cases} \mathbf{S}(t) & = & \mathbf{U}(t)\Lambda(t)\mathbf{U}(t)^T, \\ \mathbf{U}(t) & = & \mathbf{U}_1 \exp(t \log(\mathbf{U}_1^T \mathbf{U}_2)), \\ \Lambda(t) & = & \exp((1 - t) \log(\Lambda_1) + t \log(\Lambda_2)), \end{cases} \tag{4.1}$$

where $\mathbf{U}_j$ is the matrix of eigenvectors of $\mathbf{S}_j$ and Λ_j the corresponding eigenvalues matrix.

This curve has the geometric interpretation of a geodesic in the product space $\mathcal{M} = \mathrm{SO}(3) \times \mathrm{D}_+(3)$, with the bi-invariant metric of $\mathcal{M}$ viewed as a Lie group. Indeed, both $\mathbf{U}(t)$ and $\Lambda(t)$ are computed through the geodesics of their respective spaces.

Figure 4.1 suggests that the decoupling of rotations and scalings in interpolation is favorable to the conservation of anisotropy along the interpolation curve. If this property is a great advantage for the processing of Diffusion Tensor Images, spectral interpolation is in fact not commonly used in applications. This is due to the computational cost of the framework, which is important because of the processing of orientation matrices, as explained in the following.
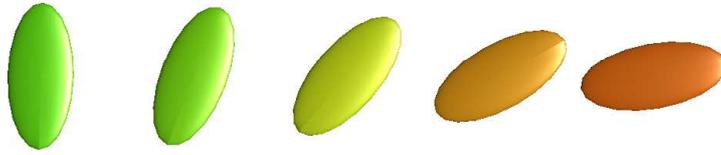

Figure 4.1: Geodesic interpolation between two tensors, in the product space $\mathrm{SO}(3) \times \mathrm{D}_+(3)$. Compared to the curves of Figure 3.4, it can be observed that the anisotropy of tensors is preserved along the interpolating curve. With this framework, the orientation is the only information which is interpolated.

4.1.1 The problem of the non-uniqueness of spectral decomposition

Consider $\mathrm{S}_*^+(3)$ the subset of $\mathrm{S}^+(3)$ with distinct eigenvalues $\lambda_1 > \lambda_2 > \lambda_3$. Every tensor of $\mathrm{S}_*^+(3)$ can be represented by four distinct elements of the Lie group $\mathcal{M}$, see Figure 4.2. Using the notation $\mathbf{e}_i$ for the i-th eigenvector, the four rotations corresponding to a single tensor $\mathbf{S}$ are

$$\mathbf{U} = \left(\ \mathbf{e}_1\ \middle|\ \mathbf{e}_2\ \middle|\ \mathbf{e}_3\ \right), \tag{4.2}$$

$$\mathbf{U}_x = \left(\ \mathbf{e}_1\ \middle|\ -\mathbf{e}_2\ \middle|\ -\mathbf{e}_3\ \right), \tag{4.3}$$

$$\mathbf{U}_y = \left(\ -\mathbf{e}_1\ \middle|\ \mathbf{e}_2\ \middle|\ -\mathbf{e}_3\ \right), \tag{4.4}$$

$$\mathbf{U}_z = \left(\ -\mathbf{e}_1\ \middle|\ -\mathbf{e}_2\ \middle|\ \mathbf{e}_3\ \right), \tag{4.5}$$

i.e. the three rotations of angle π around the principal axes of $\mathbf{U}$ plus the matrix $\mathbf{U}$ itself. The three matrices of rotation and the identity form the discrete group G.

The formal identification

$$\mathrm{S}_*^+(3) \simeq \mathrm{SO}(3)/\mathrm{G} \times \mathrm{D}_+(3),$$

where $\simeq$ is the standard notation to indicate a diffeomorphism between spaces, thus justifies the spectral interpolation (4.1) provided that $\mathbf{U}_1$ and $\mathbf{U}_2$ are univocally selected among the possible discrete spectral decompositions of $\mathbf{S}_1$ and $\mathbf{S}_2$.
This concept is explained (for 2-D ellipses) in the Figure 4.3. In this case, the orientation V_i and $-V_i$ are equivalent, in a same way as the four rotation matrices defined in (4.2)-(4.5) are equivalent for 3-D tensors. Let's consider that the spectral decompositions of
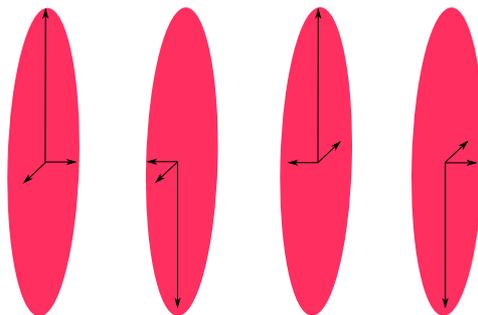
Figure 4.2: Non-uniqueness of the spectral decomposition. Four different rotation matrices parametrize the same tensor.

the two tensors give the orientation V_1 and V_2 in the Figure 4.3. If the interpolation between orientation is made without paying attention to the non-uniqueness of the spectral decomposition, the interpolating curve will follow the blue arc. However, it is clear from the figure that the pink arc, which connect V_1 to $-V_2$, is shorter than the blue one. This highlights the necessity to appropriately select the orientation used for the processing. This selection is often called the *realignment* procedure. In the 2D-case, if we consider V_1 as the reference, it simply consists in taking the direction of V_2 for which the scalar product between V_1 and V_2 is positive. In other words, we choose the orientation which is the *closest to the reference.*

The method of realignment for 3×3 rotation matrices follows the same principle. Denoting GU_2 the discrete group of four rotation matrices associated to the spectral decomposition of $\mathbf{S}_2$, a reasonable choice is to choose $\mathbf{U}_1$ arbitrarily and choose $\mathbf{U}_2$ according to

$$\mathbf{U}_2 = \arg \min_{\mathbf{U} \in GU_2} d_{SO(3)}(\mathbf{U}_1, \mathbf{U}) \tag{4.6}$$

$$= \arg \min_{\mathbf{U} \in GU_2} || \log(\mathbf{U}_1^T \mathbf{U})||_2 \tag{4.7}$$

The problem of the non-uniqueness of the spectral decomposition has been described in (Tschumperlé and Deriche, 2001; Chefd'hotel et al., 2002; Coulon et al., 2001, 2004), where different 'realignment procedures' have been proposed. The problem of those procedures is that they have to be performed at a local level, *i.e.* for each tensor of the images, leading to prohibitive computations for large-scale DTI processing. If the cost of this operation is decreased, the total cost of the processing could be drastically reduced. This is the basic idea behind the introduction of unit quaternions to represent orientations.

4.2   Spectral-quaternion interpolating curve

In this section, we present an important idea of the framework developed in the thesis. This idea consists in parametrizing the orientation of tensors as *unit quaternions* instead of rotation matrices. Using the embedding of SO(3) into the linear space of quaternion is very common in robotics in order to save on computational time, but has not been proposed for the processing of diffusion tensor images, to the best of our knowledge.
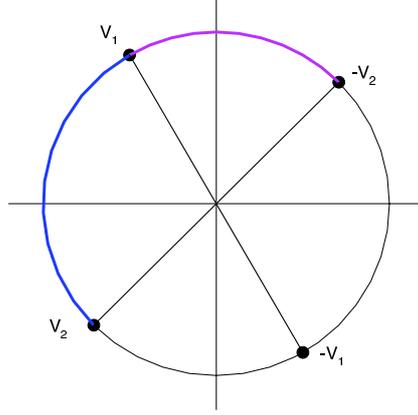
Figure 4.3: Difference between interpolation without realignment (blue curve) and with realignment (pink curve).

4.2.1 From rotation matrices to quaternions, and vice-versa

A unit quaternion is generally denoted by $\mathbf{q} = (a, \mathbf{V})$ where a is associated to the angle of rotation by $\theta = 2\arccos(a)$ and $\mathbf{V}$ is associated to the axis $\mathbf{w}$ of rotation through $\mathbf{w} = \mathbf{V}/\sin(\theta/2)$. The construction of $\mathbf{q}$ from $\mathbf{R}$ is given by

$$\theta = \arccos((\mathrm{Tr}(\mathbf{R}) - 1)/2),\tag{4.8}$$

$$\mathbf{w} = \frac{1}{2\sin\theta}\begin{pmatrix} R_{3,2} - R_{2,3} \\ R_{1,3} - R_{3,1} \\ R_{2,1} - R_{1,2} \end{pmatrix}.\tag{4.9}$$

Finally, we have $a = \cos(\theta/2)$, $\mathbf{V} = \sin(\theta/2)\mathbf{w}$. Note that the opposite quaternion given by $(-a, -\mathbf{V})$ represents the same rotation matrix.

From $\mathbf{q}$, the associated rotation matrix $\mathbf{R}$ is given by

$$\mathbf{R} = \exp\begin{pmatrix} 0 & -w_3\theta & w_2\theta \\ w_3\theta & 0 & -w_1\theta \\ -w_2\theta & w_1\theta & 0 \end{pmatrix}.\tag{4.10}$$

4.2.2 Interpolating curve between quaternions

Unit quaternions form the space $\mathbb{H}_1$. This space can be equipped with the Euclidean metric of its natural embedding in $\mathbb{R}^4$.

Definition 4.2.1. *Using the Euclidean metric of $\mathbb{R}^4$, the geodesic between two quaternions $\mathbf{q}_1$ and $\mathbf{q}_2$ is given by*

$$\mathbf{q}_g(t) = (1-t)\mathbf{q}_1 + t\mathbf{q}_2.\tag{4.11}$$

The *chordal* interpolation between two unit quaternions is then given by the projection of this geodesic on the space of unit quaternions. In this case, the projection corresponds to a normalization of the vector.

$$\mathbf{q}(t) = \frac{(1-t)\mathbf{q}_1 + t\mathbf{q}_2}{||(1-t)\mathbf{q}_1 + t\mathbf{q}_2||}.\tag{4.12}$$

The difference between a chordal and a geometric interpolation is illustrated in Figure 4.4, for the unit sphere S^1. The linear interpolation is in blue while the geometric one is in pink. The chordal interpolation is the projection on the sphere of the linear interpolation. In this case, the projection simply consists in normalizing the vector. It can be seen from the projection of the linear mean that it is very close to the geometric mean, for points that are not too far apart on the circle.

Using the same example of Figure 4.4, it is also interesting to compare the length of the curves. For two points θ_1 and θ_2, the length of the geometric curve is equal to

$$d_{S^1}(\theta_1, \theta_2) = |\theta_1 - \theta_2|, \tag{4.13}$$

while the chordal length is

$$d_{\mathbb{R}^2}(\theta_1, \theta_2) = 2 \left| \sin\left(\frac{\theta_1 - \theta_2}{2}\right) \right|. \tag{4.14}$$

It is clear from (4.14) that if θ_1 and θ_2 are not too far apart, the two distances are approximatively equal, since in this case $\sin(\theta_1 - \theta_2)/2$ can be approximated by $(\theta_1 - \theta_2)/2$. This approximation is often used to simplify algorithms, as in (Absil et al., 2004) for the computation of a mean on the Grassmann manifold.
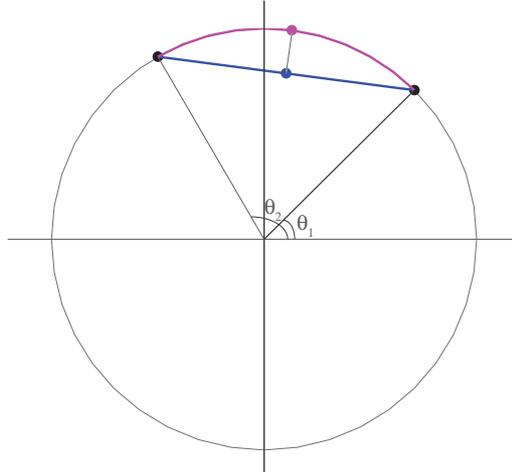

Figure 4.4: Difference between a linear and a geometric interpolation for the unit sphere S^1. The chordal interpolation corresponds to the projection on the sphere of the linear interpolation.

4.2.3 Realignment with quaternions

The realignment step in the quaternion space uses the Euclidean (chordal) distance

$$d(q_1, q_2) = ||q_1 - q_2||. \tag{4.15}$$

Because $\mathbb{H}_1$ provides a double covering of SO(3) (as already said, $\mathbf{q}$ and $-\mathbf{q}$ represent the same rotation matrix), eight different quaternions represent the same orientation (two for each of the four different rotation matrices of (4.2)-(4.5)). We denote by $\mathcal{Q}$ this set of quaternions, and we construct it as follows.

Definition 4.2.2. *If* $\mathbf{q}_i$ *is the quaternion computed from the spectral decomposition of the tensor and the transformation of the obtained rotation matrix, the set* $\mathcal{Q}_i$ *contains the four elements*

$$\left\{ \begin{pmatrix} \mathbf{q}_i(1) \\ \mathbf{q}_i(2) \\ \mathbf{q}_i(3) \\ \mathbf{q}_i(4) \end{pmatrix} \begin{pmatrix} \mathbf{q}_i(2) \\ -\mathbf{q}_i(1) \\ -\mathbf{q}_i(4) \\ \mathbf{q}_i(3) \end{pmatrix} \begin{pmatrix} \mathbf{q}_i(3) \\ \mathbf{q}_i(4) \\ -\mathbf{q}_i(1) \\ -\mathbf{q}_i(2) \end{pmatrix} \begin{pmatrix} \mathbf{q}_i(4) \\ -\mathbf{q}_i(3) \\ \mathbf{q}_i(2) \\ -\mathbf{q}_i(1) \end{pmatrix} \right\}$$

and their opposites.

Realignment in $\mathbb{H}_1$ thus involves solving

$$d(\mathcal{Q}_1, \mathcal{Q}_2) = \min_{\mathbf{q}_2 \in \mathcal{Q}_2} ||\mathbf{q}_1 - \mathbf{q}_2|| \tag{4.16}$$

with $\mathbf{q}_1$ arbitrarily chosen.

Because $\mathbf{q}_1$ and $\mathbf{q}_2$ have unit norm, we have the simplification

$$||\mathbf{q}_1 - \mathbf{q}_2||^2 = 2 - 2\,\mathbf{q}_1.\mathbf{q}_2 \tag{4.17}$$

which replaces (4.16) with

$$\mathbf{q}_2^a = \arg \max_{\mathbf{q}_2 \in \mathcal{Q}_2} \mathbf{q}_1^r.\mathbf{q}_2 \tag{4.18}$$

$$d(\mathcal{Q}_1, \mathcal{Q}_2) = ||\mathbf{q}_1^r - \mathbf{q}_2^a|| \tag{4.19}$$

where $\mathbf{q}_2^a$ is called the *realigned* quaternion.

The computation of (4.18) and (4.19) is in fact very fast compared to (4.7): the eight scalar products $\mathbf{q}_1^r.\mathbf{q}_2$ can be computed through a single matrix product between the 1×4 vector representing $(\mathbf{q}_1^r)^T$ and the 4×8 matrix formed by the eight quaternions $\mathbf{q}_2$. In contrast, computing the distance (4.7) requires four logarithms of product of 3×3 matrices, which is expensive. The selection of the parametrization of rotations as quaternions thus enables the framework to be computationally tractable.

Definition 4.2.3. *Using the realignment procedure of equations* (4.18) *and* (4.19)*, the* spectral-quaternion interpolating curve *is given by the equations*

$$\mathbf{S}(t) = \mathbf{U}_q(t)\Lambda(t)\mathbf{U}_q(t)^T \tag{4.20}$$

where

$$\Lambda(t) = \exp\left((1-t)\log(\Lambda_1) + t\log(\Lambda_2)\right),$$

$\mathbf{U}_q(t)$ *is the rotation matrix associated to the unit quaternion* $\mathbf{q}(t) \in \mathbb{H}_1$ *and* $\mathbf{q}(t)$ *satisfies*

$$\mathbf{q}(t) = \frac{(1-t)\mathbf{q}_1 + t\mathbf{q}_2^a}{||(1-t)\mathbf{q}_1 + t\mathbf{q}_2^a||}.$$

This curve is illustrated in Figure 4.5. The interpolating curve is very close to the curve shown in Figure 4.1. This observation is similar to the one of Section 3.4.3, where the Log-Euclidean framework was viewed as a good substitute of the affine-invariant method. In both cases, the embedding in a linear space has enabled the reduction of the computational cost while staying close of the intrinsic curves.

Figure 4.5: Interpolation between two tensors, in the product space $\mathbb{H}_1 \times D_+(3)$. Compared to the curve of Figure 4.1, it can be observed that the approximation of the 'spectral' geometric interpolation between matrices with the chordal interpolation of quaternions is very good. The interpolation of eigenvalues being the same for the two frameworks, the observation about the preservation of anisotropy is still true in this case.

4.3   Computational and invariance properties

This section will study and compare the merits of the spectral-quaternion interpolation method with respect to the ones of the other frameworks presented in this dissertation: the affine-invariant framework of Eq. (3.18), its extrinsic substitute, the Log-Euclidean method of Eq. (3.21), and the spectral framework of the beginning of this chapter, Eq. (4.1).

4.3.1   Computational cost

The two interpolation methods of equations (3.18) (affine-invariant) and (4.1) (spectral curve) are intrinsic, whereas the methods of equations (3.21) (Log-Euclidean) and (4.20) (spectral-quaternion) are extrinsic, that is, the interpolating curve is a projection on the manifold of an interpolating curve in the embedding space. Extrinsic geometries lead to significant computational savings because they reduce the use of matrix exponential and matrix logarithm. Table 4.1 illustrates how the difference translates into computational savings on a (non optimized) MATLAB code, where both extrinsic methods perform faster than their corresponding intrinsic curves. We expect that the qualitative difference between intrinsic curves and extrinsic interpolations will scale up dramatically in real DTI signal processing.

Affine-invariant	Log-Euclidean	Spectral	Spectral-quaternion
0.47 s	0.17 s	0.65 s	0.11 s

Table 4.1: Computational time of computing 1000 distances between a reference and random samples from a Wishart distribution. The computations are performed on a Intel Core 2 Duo 2,66 GHz with 4Go of RAM machine using a (non optimized) MATLAB code. The definitions of the similarity measure for the two spectral frameworks will be given in Section 4.5.

4.3.2 Invariances

Invariances of the processing are of great importance in the context of medical imaging. Indeed, it is important that the processing gives the same result, whatever the orientation of the head in the scanner, or the intensity of the applied gradient field. In other words, the processing has to be independent of the units, or the calibration. This is why two types of invariances are often required for a processing framework: invariance by scaling and invariance by rotation.

Definition 4.3.1. *A framework is said to be invariant by scaling and rotation if the interpolating curve from $\mathbf{S}_1$ to $\mathbf{S}_2$ which depends on t, $\mathbf{S}(t; \mathbf{S}_1, \mathbf{S}_2)$, satisfies*

$$\forall s \in \mathbb{R}^+, \forall \mathbf{R} \in \mathrm{SO}(3):$$

$$\mathbf{S}(t; s\mathbf{R}\mathbf{S}_1\mathbf{R}^T, s\mathbf{R}\mathbf{S}_2\mathbf{R}^T) = s\mathbf{R}\mathbf{S}(t; \mathbf{S}_1, \mathbf{S}_2)\mathbf{R}^T. \tag{4.21}$$

All the four frameworks presented in this thesis are invariant by rotation and scaling. The affine-invariant framework (3.18) possesses a larger invariance group:

$$\forall \mathbf{A} \in \mathrm{Gl}(3):$$

$$\mathbf{S}(t; \mathbf{A}\mathbf{S}_1\mathbf{A}^T, \mathbf{A}\mathbf{S}_2\mathbf{A}^T) = \mathbf{A}\mathbf{S}(t; \mathbf{S}_1, \mathbf{S}_2)\mathbf{A}^T \tag{4.22}$$

In fact, the interpolating curve (3.18) is the unique interpolation formula being invariant to an arbitrary congruence transformation, hence its name. This property can be desirable in other applications (see e.g. (Smith, 2005)), but it is the intrinsic cause of the anisotropy degradation observed in Figure 3.4.

The invariances of the Log-Euclidean framework have been studied in (Arsigny et al., 2006): it is invariant to scaling and rotation. Besides, the authors insist in this paper on the fact that these two invariances are sufficient for medical imaging.

Finally, both spectral frameworks satisfy (4.21), since

- The scaling only affects the eigenvalues, and it is easy to show that for each eigenvalue λ_i,

$$
\begin{aligned}
\lambda_i(t; s\lambda_{i,1}, s\lambda_{i,2}) &= \exp((1-t)\log(s\lambda_{i,1}) + t\log(s\lambda_{i,2})) \\
&= \exp((1-t)\log s + t\log s + (1-t)\log \lambda_{i,1} + t\log \lambda_{i,2}) \\
&= s\exp((1-t)\log \lambda_{i,1} + t\log \lambda_{i,2}) \\
&= s\lambda_i(t; \lambda_{i,1}, \lambda_{i,2})
\end{aligned}
$$

- The rotation only affects the interpolation between orientations. Using rotation matrices, it is easy to show that

$$
\begin{aligned}
\mathbf{U}(t; \mathbf{R}\mathbf{U}_1, \mathbf{R}\mathbf{U}_2) &= \mathbf{R}\mathbf{U}_1\exp(t\log(\mathbf{U}_1^T\mathbf{R}^T\mathbf{R}\mathbf{U}_2)) \\
&= \mathbf{R}\mathbf{U}(t; \mathbf{U}_1, \mathbf{U}_2),
\end{aligned}
$$

which leads to

$$
\begin{aligned}
\mathbf{S}(t, \mathbf{R}\mathbf{S}_1\mathbf{R}^T, \mathbf{R}\mathbf{S}_2\mathbf{R}^T)) &= \mathbf{R}\mathbf{U}(t; \mathbf{U}_1, \mathbf{U}_2)\Lambda(t)\mathbf{U}^T(t, \mathbf{U}_1, \mathbf{U}_2)\mathbf{R}^T \\
&= \mathbf{R}\mathbf{S}(t; \mathbf{S}_1; \mathbf{S}_2)\mathbf{R}^T.
\end{aligned}
$$

- Regarding the spectral quaternion framework, it can be proved that the chordal interpolation between quaternions is also invariant by rotations. First consider the multiplication of quaternions, which is also known as their Hamilton product. If we use the notation $\mathbf{q} = (a, \mathbf{V})$ introduced above, the multiplication of two quaternions $\mathbf{q}_1$ and $\mathbf{q}_2$ is given by

$$\mathbf{q}_1 \cdot \mathbf{q}_2 = (a_1 a_2 - \mathbf{V}_1.\mathbf{V}_2, a_1 \mathbf{V}_2 + a_2 \mathbf{V}_1 + \mathbf{V}_1 \times \mathbf{V}_2),$$

  where $\mathbf{V}_1.\mathbf{V}_2$ is the scalar product of vectors while $\mathbf{V}_1 \times \mathbf{V}_2$ is their cross product. The multiplication of quaternions is equivalent to the (matrix) multiplication of rotation matrices, *i.e* the rotation matrix corresponding to $\mathbf{q}_1 \cdot \mathbf{q}_2$ is equal to $\mathbf{U}_1 \mathbf{U}_2$. It should be noted that this multiplication is distributive and compatible with scalar multiplication. If the tensors are rotated by a matrix $\mathbf{R}$, their quaternions will be transformed as $\mathbf{q}_i^* = \mathbf{q}_R \cdot \mathbf{q}_i$. The interpolating curve between quaternions can thus be written as

$$
\begin{aligned}
\mathbf{q}(t; \mathbf{q}_R \cdot \mathbf{q}_1, \mathbf{q}_R \cdot \mathbf{q}_2) &= \frac{(1-t)\mathbf{q}_R \cdot \mathbf{q}_1 + t\mathbf{q}_R \cdot \mathbf{q}_2}{||(1-t)\mathbf{q}_R \cdot \mathbf{q}_1 + t\mathbf{q}_R \cdot \mathbf{q}_2||} \\
&= \frac{\mathbf{q}_R \cdot ((1-t)\mathbf{q}_1 + t\mathbf{q}_2)}{||\mathbf{q}_R \cdot ((1-t)\mathbf{q}_1 + t\mathbf{q}_2)||} \\
&= \frac{\mathbf{q}_R}{||\mathbf{q}_R||} \cdot \frac{(1-t)\mathbf{q}_1 + t\mathbf{q}_2}{||(1-t)\mathbf{q}_1 + t\mathbf{q}_2||} \\
&= \mathbf{q}_R \cdot \mathbf{q}(t; \mathbf{q}_1, \mathbf{q}_2)
\end{aligned}
$$

  where the equality $||\mathbf{q}_i \cdot \mathbf{q}_j|| = ||\mathbf{q}_i|| \, ||\mathbf{q}_j||$ has been used. The corresponding rotation matrix is then $\mathbf{R}\mathbf{U}(t)$, and the conclusion follows the one of the spectral interpolation.

- The combination of those results proves that spectral frameworks are invariant by rotation and scaling.

4.3.3 Volume and spectral interpolation

The goal of our framework is to retain the most of the information conveyed by tensors. It is thus important to study how the framework affects those properties.

Volume The volume of the ellipsoid represented by a positive definite matrix $\mathbf{S}$ is the product of its eigenvalues, *i.e.* its determinant. The four interpolation formulas discussed in this dissertation enjoy the remarkable property

$$\det(\mathbf{S}(t; \mathbf{S}_1, \mathbf{S}_2)) = \exp((1-t)\log(\det(\mathbf{S}_1)) + t\log(\det(\mathbf{S}_2))) \qquad (4.23)$$

In other words, the determinant of the tensor is geometrically interpolated by the curves. In particular, the volume is preserved along the interpolation curves between two tensors of same volume. It is in contrast with the 'swelling' effect observed with simpler interpolation methods and has been a main motivation to promote a Riemannian framework for DTI processing, as explained in (Pennec et al., 2006) for the affine-invariant metric, and in (Arsigny et al., 2006, 2007) for the Log-Euclidean one. Regarding the spectral frameworks, the interpolating curve between eigenvalues suffices to ensure the geometric interpolation of the determinant.

Eigenvalues Because the interpolation of rotations and eigenvalues is decoupled in the two spectral interpolation methods (4.1) and (4.20), they enjoy a further spectral interpolation property:

$$\lambda_i \left(\mathrm{S}(t; \mathbf{S}_1, \mathbf{S}_2) \right) = \exp((1 - t) \log(\lambda_i(\mathbf{S}_1)) + t \log(\mathbf{S}_2)) \qquad (4.24)$$

This spectral interpolation is a key property for the anisotropy interpolation. It can also be viewed as the 'missing' property of both affine-invariant and Log-Euclidean frameworks.

4.3.4 Anisotropy interpolation

As explained in the beginning of this chapter, the degradation of the anisotropy induced by the affine-invariant and Log-Euclidean interpolations was one of the main motivation to develop a novel, anisotropy preserving, framework. This important property will be explained in the sequel. First, we introduce a novel anisotropy index, the Hilbert Anisotropy (HA).

Intuitively, anisotropy is a measure of 'how far away' a diffusion tensor is from being isotropic, as explained in (Fletcher and Joshi, 2007). In this work, the authors introduced the Geodesic Anisotropy (GA) (3.20), by measuring the distance between a tensor and its closest isotropic tensor, in a geodesic sense *i.e.* $\mathrm{GA}(\mathbf{S}) = d(\mathbf{S}, \mathbf{S}_I)$ where d is the affine-invariant distance and $\mathbf{S}_I$ is the closest isotropic tensor. In a geometric framework, this tensor is given by $(\det(\mathbf{S}))^{1/3}\mathbf{I}$. It is important to note that this definition is scale-invariant, $\mathrm{GA}(\mathbf{S}) = \mathrm{GA}(\beta\mathbf{S})$.

Our novel definition of anisotropy is based on the fact that, geometrically, any anisotropy scalar measure should be a *scale invariant distance to identity*. It is known that projective distance are scale-invariant. We have thus decided to use the Hilbert metric (Birkhoff, 1957), which is precisely a projective distance that can be defined in arbitrary cones, to define our anisotropy index. It leads to the following definition that we refer to as Hilbert anisotropy (HA) in the sequel.

Definition 4.3.2. *The Hilbert Anisotropy of a tensor* $\mathbf{S}$ *is*

$$\mathrm{HA} = d_{\mathrm{H}}(\mathbf{S}, \mathbf{I}) = \log\left(\frac{\lambda_{\max}}{\lambda_{\min}} \right) \qquad (4.25)$$

where $\lambda_{\max}$ *and* $\lambda_{\min}$ *are respectively the maximum and minimum eigenvalues of* $\mathbf{S}$.

With a projective distance, the distance between any isotropic tensor and the identity is null, which explains why the index should be measured as a distance to identity.

The HA index possesses all the required properties for an anisotropy index, *i.e.*

- $\mathrm{HA} \geq 0$ and $\mathrm{HA} = 0$ only for isotropic tensors.

- HA is invariant to rotations: $\mathrm{HA}(\mathbf{S}) = \mathrm{HA}(\mathbf{U}\mathbf{S}\mathbf{U}^T)$ for all $\mathbf{U} \in \mathrm{O}(3)$.

- HA is invariant by scaling, $\mathrm{HA}(\mathbf{S}) = \mathrm{HA}(s\mathbf{S})$, $\forall s \in \mathbb{R}_+$ (it means that anisotropy only depends on the shape of the tensor and not on its size).

- HA is a dimensionless number. This property is desirable and natural, as the anisotropy of the tensor physically reflects the microscopic anisotropy of the tissues, which is independent from the diffusivity.

Figure 4.6 illustrates a comparison of HA with the three popular anisotropy indices: the fractional anisotropy (FA), the relative anisotropy (RA) and the geodesic anisotropy (GA). This figure is constructed with eigenvalues equal to $\lambda, (1-\lambda)/2, (1-\lambda)/2, \lambda \in]0,1]$, as explained in (Batchelor et al., 2005). The diffusion tensor varies from planar to spherical for $\lambda \in]0,1/3]$ and then tends to become unidirectional when λ increases. By definition, the Hilbert Anisotropy is more similar to the Geodesic Anisotropy, as they both are geometric.
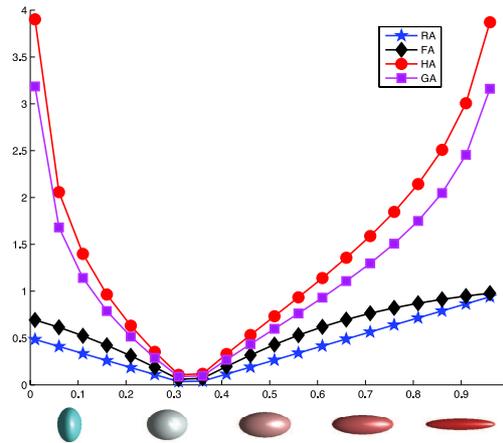

Figure 4.6: Evolution of four indices of anisotropy with respect to modifications of eigenvalues. The difference between 'linear' indices of anisotropy (FA, $\diamondsuit$ and RA, $\bigstar$) and geometric' ones (HA, $\bigcirc$ and GA,$\square$) is clearly visible. The eigenvalues are given by $\lambda, (1-\lambda)/2, (1-\lambda)/2, \lambda \in]0,1]$, (Batchelor et al., 2005).

The spectral interpolation formula (4.24) leads to a particularly attractive formula for Hilbert anisotropy interpolation

$$\text{HA}(\mathbf{S}(t;\mathbf{S}_1,\mathbf{S}_2)) = (1-t)\,\text{HA}(\mathbf{S}_1) + t\,\text{HA}(\mathbf{S}_2) \tag{4.26}$$

In other words, Hilbert anisotropy is *linearly interpolated* along the interpolation curves (4.1) and (4.20). The three other anisotropy indices do not lead to linear interpolation as the Hilbert anisotropy but they are qualitatively equivalent, since they evolve monotonically along the interpolating curves. This is a major difference between the affine-invariant and Log-Euclidean frameworks on one hand and the spectral frameworks on the other hand.

Figure 4.7 illustrates how the different features of the tensors evolve along the Log-Euclidean and spectral-quaternion interpolating curves. In particular, it can be observed how the anisotropy may degrade along the curve (3.21). This 'anisotropy swelling effect' is undesirable for a processing framework in which anisotropy carries important information. Some difference also appears in the orientation interpolation. The angle between the principal eigenvectors is quasi linearly interpolated when using quaternions, which is not the case with the Log-Euclidean framework.
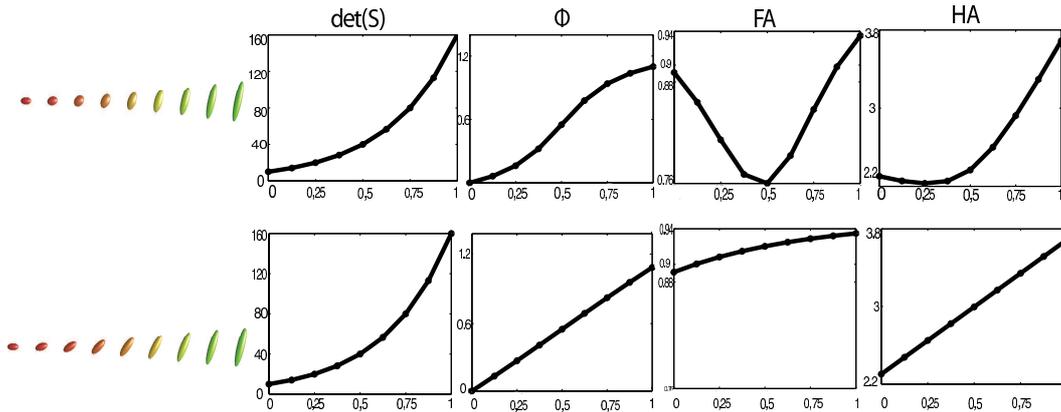
Figure 4.7: Interpolating curves between two tensors. *Top*: Log-Euclidean interpolation. *Bottom*: Spectral-quaternion interpolation. The represented indices are (from left to right) the determinant, the angle between the principal eigenvectors of the first tensor and the considered one, the fractional anisotropy and the Hilbert anisotropy. Anisotropy evolves monotonically along spectral-quaternion interpolating curves, in contrast to Log-Euclidean interpolation.

4.3.5 Singularities and uncertainty

A potential shortcoming of the proposed spectral-interpolation framework is that it is only defined on the subset $S_*^+(3)$ where eigenvalues are distinct. In situations of two or three similar eigenvalues, small measurements errors may lead to significant differences in the interpolating curves.

One should bear in mind that an isotropic measurement is also much less certain about orientation than an anisotropic measurement (Parker et al., 2003). In other words, eigenvalue separation of positive definite matrix correlates with the orientation information contained in the physical measurement. A similar idea was stated in (Coulon et al., 2004), which explains that the principal direction of diffusion map does not have any organization or meaning in isotropic regions. This idea led the authors of (Coulon et al., 2004) to the introduction of a weighting factor in their regularization algorithm.

In order to account for this uncertainty, we introduce the real function

$$f(x) = \frac{(\beta x)^4}{(1 + (\beta x)^4)} \tag{4.27}$$

where β is a parameter to be defined in the sequel. The function f defines a smooth sigmoidal transition between the limits $f(0) = 0$ and $f(\infty) = 1$. The function of Eq. (4.27) will be used to rescale the interpolation between orientations of tensors in a way that mitigates the importance of the orientation information in case of low anisotropy.

Consider two tensors $\mathbf{S}_1, \mathbf{S}_2$, with $\mathbf{S}_1$ isotropic. The orientation of $\mathbf{S}_1$ (encoded by the associated quaternion $\mathbf{q}_1$) is totally arbitrary. As a result, it seems logical *not* to take into account this orientation in the interpolation between $\mathbf{S}_1$ and $\mathbf{S}_2$. This suggest to fix $\mathbf{q}_2$ along the entire interpolation curve, which means weighting $\mathbf{q}_1$ by a vanishing weight in the formula (4.20).
On the other hand, if we consider two very anisotropic tensors (sharing the same eigenvalues), then their orientations $\mathbf{q}_1, \mathbf{q}_2$ should be fully taken into account in the interpolation

process by opting for the linear interpolation (4.20).

The function f can be used as a smooth transition between those two limit cases. Noticing that the case where one tensor is isotropic corresponds to

$$\min(\mathrm{HA}_1, \mathrm{HA}_2) = 0\,,$$

while the case where both are very anisotropic corresponds to $\min(\mathrm{HA}_1, \mathrm{HA}_2)$ being large, we thus let

$$\kappa(\mathrm{HA}_1, \mathrm{HA}_2) = f(\min(\mathrm{HA}_1, \mathrm{HA}_2))\,. \tag{4.28}$$

The linear quaternion interpolation in (4.20) can be replaced with

$$\mathbf{q}(t) = w_1^*(t)\mathbf{q}_1 + w_2^*(t)\mathbf{q}_2 \tag{4.29}$$

where

$$w_1^*(t) = (1-t)\frac{\kappa(\mathrm{HA}_1, \mathrm{HA}_t)}{\bar{\kappa}}\,,$$

$$w_2^*(t) = t\frac{\kappa(\mathrm{HA}_t, \mathrm{HA}_2)}{\bar{\kappa}}$$

and where $\bar{\kappa} = (1-t)\kappa(\mathrm{HA}_1, \mathrm{HA}_t) + t\kappa(\mathrm{HA}_t, \mathrm{HA}_2)$ ensures that $w_1^* + w_2^* = 1$. HA_t is defined in Eq.(4.26).

Defined this way, the interpolation is consistent with the orientation uncertainty. Indeed when one tensor, say $\mathbf{S}_1$, is isotropic we have $w_1^*(t) \equiv 0$. More generally as soon as $\mathrm{HA}_1 < \mathrm{HA}_2$ we have $w_1^* < 1-t$ as $f(\mathrm{HA}_1) < f(\mathrm{HA}_t)$, meaning that the basis $\mathbf{q}_1$ is given less weight than in the linear interpolation case (4.20). Finally, when $\min(\mathrm{HA}_1, \mathrm{HA}_2) \to \infty$, we recover the linear interpolation (4.20). Figure 4.8 illustrates the effect of this rescaling in the case of unequal anisotropies. For the simulations of this paper, we choose $\beta = 0.6$ so that $\kappa(\mathrm{HA}_1, \mathrm{HA}_2)$ becomes very close to 1 (more than 0.9) when both anisotropies are at least equal to 3, which means the Hilbert ratio between the largest and smallest eigenvalue is at least 20 (for example, $\lambda_1 = 10$ and $\lambda_3 = 0.5$). In this case both tensors can be considered as very anisotropic and the linear interpolation (4.20) is recovered. The parameter β thus appears as a tuning parameter such that $f^{-1}(0.9)/\beta$ (where f^{-1} is the inverse of f) is considered as a high Hilbert anisotropy. The parameter β can be considered as a user parameter.

It should be noted that the proposed rescaling does not handle the singularities of partial isotropy ($\lambda_1 \simeq \lambda_2 \neq \lambda_3$). This theoretical limitation of the quaternion framework appears as a price to pay for the computational savings in the quaternions space.

4.4 Weighted means and multidimensional interpolation

As explained in the first chapter of this thesis, weighted means are crucial in many steps of the processing. They appear in interpolation, transformation of images, statistics, averages of images, construction of atlas,... If the mean between two tensors is naturally given by the interpolating curve, the extension to many tensors is less straightforward, as it will be explained in the following.

4.4.1 Weighted means

An interpolating curve $\mathbf{S}(t; \mathbf{S}_1, \mathbf{S}_2)$ provides a straightforward definition of weighted mean from

$$\mathrm{mean}((1-t)\mathbf{S}_1, t\mathbf{S}_2) = \mathbf{S}(t; \mathbf{S}_1, \mathbf{S}_2)\,, 0 \leq t \leq 1\,. \tag{4.30}$$
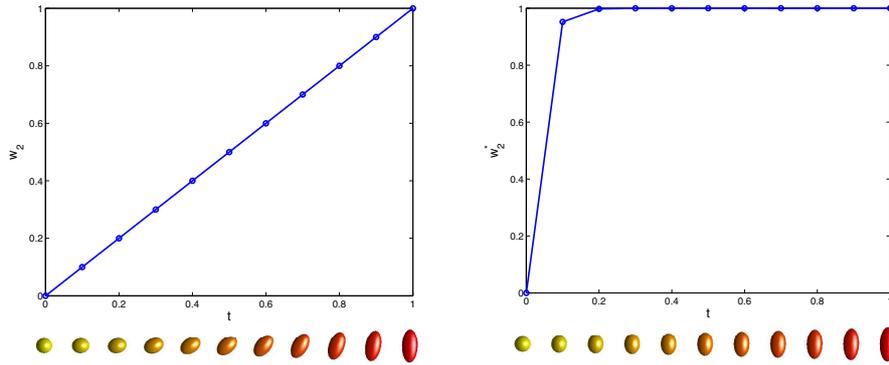
Figure 4.8: Impact of the rescaling on interpolating curve. For each interpolation, the quaternion interpolation follows $\mathbf{q}(t) = w_1^*(t)\mathbf{q}_1 + w_2^*(t)\mathbf{q}_2$. *Left*: $w_1^* = 1 - t, \quad w_2^* = t$, which corresponds to the linear interpolation of Eq. (4.20). *right*: anisotropy is used to rescale the interpolation, and $w_2^*(t) = t\frac{\kappa(\mathrm{HA}_t, \mathrm{HA}_2)}{\bar{\kappa}}$. Since the anisotropy of the first tensor is very low, the interpolated orientation is mostly based on the one of the second tensor (which is equivalent to $w_2^* \approx 1$).

In particular, the unweighted mean obtained for $t = 1/2$ corresponds to the midpoint of the interpolation curve, a definition which agrees with the usual notion of Riemannian mean when the interpolating curve is a geodesic (see 2.2.1).

Extending the concept of mean to more than two positive definite matrices is less straightforward and the topic of current research (Pennec et al., 2006; Ando et al., 2004; Petz and Temesi, 2005; Moakher and Zéraï, 2011; Moakher, 2005). As explained in the first chapter, mean on manifolds is usually defined by the Kärcher mean (Kärcher, 1977), which minimizes the sum of squared distances. Usually, this definition does not enable to derive a closed-form formula for the computation of the mean of N elements, and this mean is computed through an optimization algorithm, as for example in (Pennec et al., 2006; Fletcher and Joshi, 2007). This is another attractive feature of the Log-Euclidean framework, which provides a closed-form formula for the mean of several matrices (Arsigny et al., 2006).

One of the main problem which arises for the extension of means to many tensors is the issue of the uniqueness of the mean, and of its symmetry. For example, the mean of three tensors should satisfy

$$\mathrm{mean}(\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3) = \mathrm{mean}(\mathbf{S}_2, \mathbf{S}_1, \mathbf{S}_3) = \cdots = \mathrm{mean}(\mathbf{S}_3, \mathbf{S}_2, \mathbf{S}_1). \qquad (4.31)$$

At first sight, this problem seems simple. However, due to non-uniqueness of the spectral decomposition, the solution to this problem necessitates an appropriate procedure.

This issue is explained in Figure 4.9, for the orientation of 2D tensors. In this figure, the orientations of three tensors are represented (dotted lines). To the left, the pink orientation is chosen as the reference for the realignment, with the direction indicated by the pink dot. Using the procedure described above, the chosen directions for the other orientations are indicated by blue and khaki dots. Using those directions, the orientation of the mean is given by the plain red line. To the right, however, the khaki orientation is chosen as the reference. This changes the selection of 'realigned' directions, and

the orientation of the mean is different, which means that three identical tensors could give different means. This is in contradiction with our postulate of uniqueness of the mean.

It could be decided to always choose the first (or last) tensor as a reference. However, in this case, the mean would be non-commutative. Another procedure of selection of the reference has to be defined. In this work, we decided to select the *most informative* tensor as the reference. The notion of information is related to the weight of the tensor in the sum and to its anisotropy.
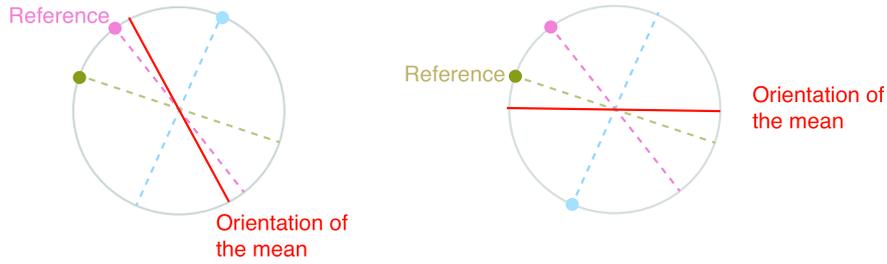

Figure 4.9: Depending upon the chosen reference for the 'realignment' of orientations, the orientation of the mean is very different.

Given N weights $w_1, \ldots, w_N$ that satisfy $\sum_i w_i = 1$, the weighted w-mean of N tensors $\mathbf{S}_1, \ldots, \mathbf{S}_N$ is defined as follows:

- The eigenvalues of the mean tensor are defined from the scalar formula (2.2):

$$\lambda_{\mu,k} = \exp(\sum_{i=1}^{N} w_i \log(\lambda_{i,k})), \quad k = 1, 2, 3. \tag{4.32}$$

- We select the (weighted) most informative tensor as the reference quaternion $\mathbf{q}^r$. Namely, we choose r as

$$r = \arg\max_i w_i \mathrm{HA}_i \tag{4.33}$$

The realigned quaternions are denoted by $\mathbf{q}_{i,r}$.

- We define the mean quaternion as the (chordal) mean in the quaternion space

$$\mathbf{q}_m = \sum_i w_i \mathbf{q}_{i,r} \tag{4.34}$$

$$\overline{\mathbf{q}} = \frac{\mathbf{q}_m}{||\mathbf{q}_m||} \tag{4.35}$$

The chordal mean of quaternions is the Riemannian mean over the set of rotations using the chordal distance on the set of quaternions (Dai et al., 2010; Sarlette et al., 2007). We emphasize that the only singularity of the proposed mean occurs in the situation $\mathbf{q}_m = 0$, which reflects a situation in which there is so much uncertainty about the average orientation that any choice for $\overline{\mathbf{q}}$ is equally good (or bad).

The properties of this mean are similar to those of the interpolating curve, since this mean is 'information preserving': the determinant of the mean of N tensors is the (geometric) mean of the determinant, the length of each principal axis is the geometric mean of the principal axis lengths, and the Hilbert anisotropy of the mean tensor is the arithmetic mean of the anisotropies

$$\overline{\mathrm{HA}} = \sum_{i=1}^{N} w_i \mathrm{HA}_i . \tag{4.36}$$

Finally, as explained in Section 4.3.5, it is possible to weight the orientations differently than the eigenvalues, in order to take into account the uncertainty of orientation for isotropic tensors. Following the developments of Section 4.3.5, a novel weighting of the quaternions, similar to the curve (4.29), is given by replacing the weights in Eq. (4.34) by

$$w_i^* = w_i \frac{\kappa(\mathrm{HA}_i, \overline{\mathrm{HA}})}{\bar{\kappa}} \tag{4.37}$$

where $\kappa(\mathrm{HA}_i, \mathrm{HA}_j)$ is given by Eq. (4.28) and $\bar{\kappa}$ is the normalizing factor given by $\bar{\kappa} = \sum_i w_i \kappa(\mathrm{HA}_i, \overline{\mathrm{HA}})$.

4.4.2 Multidimensional interpolation

As previously advocated in (Zhang et al., 2006; Kindlmann et al., 2007; Arsigny et al., 2007), an adequate interpolation method is important for the processing of diffusion tensor images and particularly for the extension of usual registration techniques (for scalar images) to the case of tensor images. This interpolation scheme is necessary to resample images. Here, we provide a direct generalization of classical interpolation method, where the interpolated value is computed as the weighted mean of the original tensors.

A common definition for multidimensional interpolation is through a weighted average of diffusion tensors (Pennec et al., 2006; Fletcher and Joshi, 2007; Arsigny et al., 2006). The weight associated to each tensor is a function of the grid distance between this tensor and the location of the interpolated tensor. In this work, if $(x_1, x_2, x_3) \in [0,1] \times [0,1] \times [0,1]$ are the coordinates of the interpolated tensor and $(\alpha_1, \alpha_2, \alpha_3) \in \{0,1\} \times \{0,1\} \times \{0,1\}$ the coordinates of the point α of the grid, the following function will be used

$$w_\alpha(x_1, x_2, x_3) = \prod_{i=1}^{3} (1 - \alpha_i + (-1)^{1-\alpha_i} x_i). \tag{4.38}$$

The interpolation between two tensors has already been illustrated in Figure 4.7, where the evolution of the different information conveyed by the tensors is also shown, as in (Zhou, 2010).

Using the method described above for computing the weighted means of many tensors, the interpolation of four tensors at the corners of a grid can be computed, as illustrated in Figures 4.10 and 4.11. For those figures, the alternative weights proposed in (4.37) are used. In the left columns of those figures, colors of the tensors are determined by HA. In their middle columns, the evolution of the anisotropy in the grid is represented as a contour map. Each curve in these subfigures represent a constant level of anisotropy. In Figure 4.10, which performs the weighted mean of 3 very anisotropic (prolate) tensors and

an isotropic one, the difference is clearly visible between the two interpolation schemes. The monotone evolution of anisotropy achieved with the spectral interpolation is strongly distorted with the Log-Euclidean interpolation. Figure 4.11 uses an isotropic tensor (bottom left), a prolate tensor (bottom right) and two oblate tensors (for which $\lambda_1 \approx \lambda_2 > \lambda_3$). In this case, anisotropy varies in a same way using any of the two methods. Finally, the direction of the principal eigenvector of tensors is also shown (right columns). Some differences can be observed, for example on the first row of the subfigures in Figure 4.10, which corresponds to an interpolation between two orthogonal tensors with the same shape.With the Log-Euclidean method, the principal eigenvector does not vary until the midpoint of the interpolation, where it is rotated by $\pi/2$ rad. On the contrary, the spectral quaternion interpolation results in a monotone rotation of the tensor. In Figure 4.11, it is interesting to note the difference for the orientation interpolation between the prolate tensor and the oblate one (last columns of the subfigures). Since the principal direction of diffusion (first eigenvector) of the prolate tensor is in the diffusion plan of the oblate tensor, it is not necessary to vary the orientation of the prolate tensor. This is indeed the case when using the Log-Euclidean method, but not with the Spectral-Quaternion framework, which performs a smooth variation of this orientation. This disadvantage of our method could be avoided by using the whole rotation matrix to represent eigenvectors. This could help to weight differently the orientation of a prolate tensor if its principal direction is included in the diffusion plane of an oblate tensor. However, this would imply to weight the orientation interpolation by a factor depending upon the full orientation, which is not permitted by the quaternion representation.
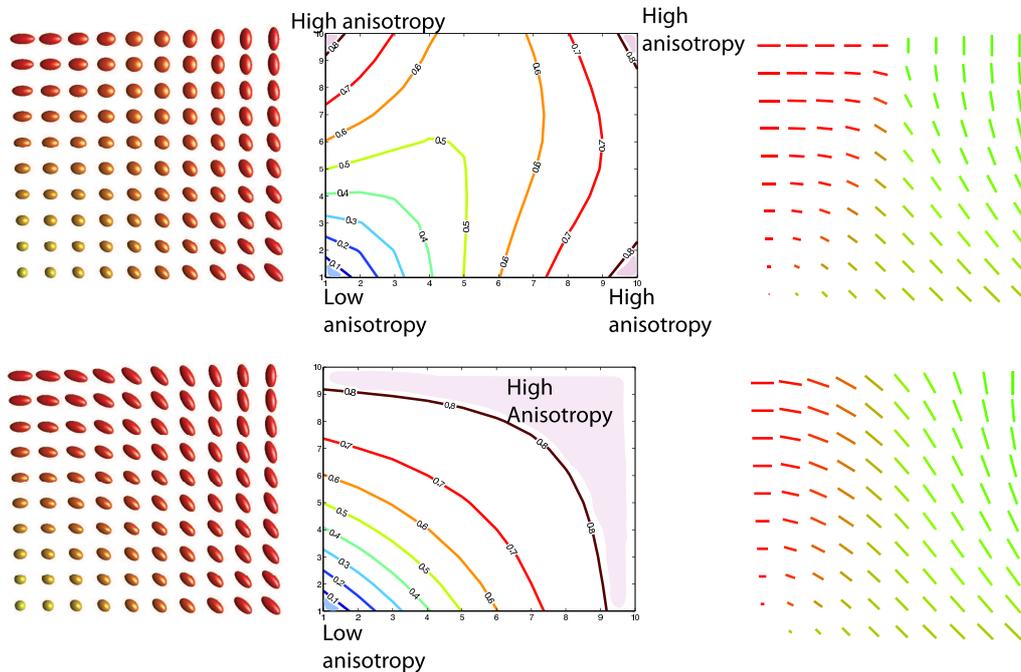

Figure 4.10: Multi-dimensional interpolation between four tensors at the corners of the grids. *Top*: Log-Euclidean interpolation. *Bottom*: Spectral interpolation. Colors of the ellipsoids indicates their anisotropy from yellow (low anisotropy) to red for higher anisotropies. The subfigures in the middle represent a contour map of the Fractional Anisotropy. Areas of constant anisotropy are represented by color lines. In the right, the orientation of the first eigenvector of tensors is represented. The lengths of the lines are proportional to the fractional anisotropy and the colors depend on the direction.
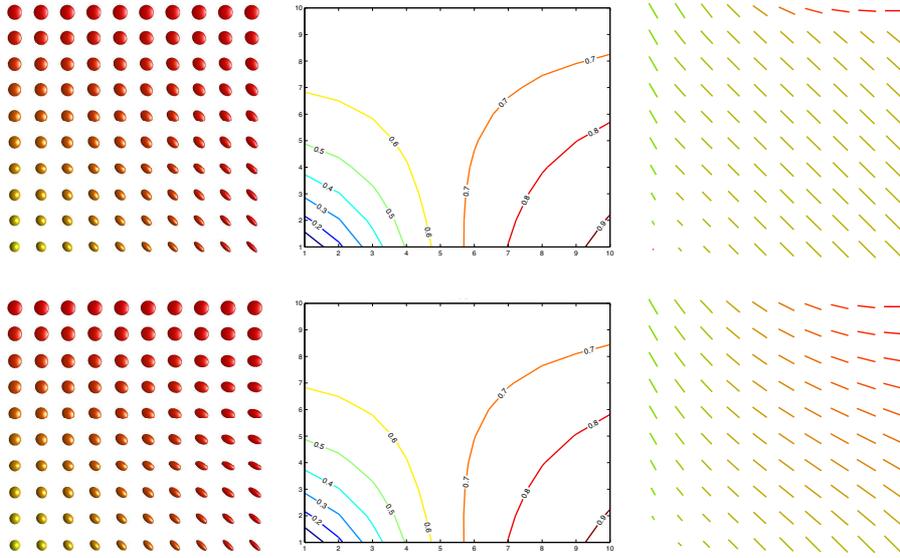
Figure 4.11: Multi-dimensional interpolation between four tensors at the corners of the grids. *Top*: Log-Euclidean interpolation. *Bottom*: Spectral interpolation. Colors of the ellipsoids indicates their anisotropy from yellow (low anisotropy) to red for higher anisotropies. The subfigures in the middle represent a contour map of the Fractional Anisotropy. Areas of constant anisotropy are represented by color lines. In the right, the orientation of the first eigenvector of tensors is represented. The lengths of the lines are proportional to the fractional anisotropy and the colors depend on the direction.

In order to further compare the two frameworks, we implemented some synthetic processing examples which illustrate how approximation errors could propagate in the processing pipeline. The simulations are chosen as in (Thévenaz et al., 2000). The protocol is the following: a succession of $r = 15$ rotations of $2\pi/15 = 24°$ each was applied to some image, such that the output of any given step r_i is the input for the next step r_{i+1}. The final output is then compared to the initial image, see Figure 4.12.

To enable a better interpretation of the results, the maps of squared differences between the properties of the original image and the ones of the output are also shown. The errors in FA are localized, and are bigger in the case of the Log-Euclidean interpolation. There are more errors for the direction of the principal eigenvector, but it should be noted that the largest errors are localized in low anisotropy areas. The uncertainty of orientation is high in these areas, which accounts for large orientation errors.

4.5 Distance, Riemannian metric and similarity measure

As explained in Section 2.1.2, interpolation curves provide a natural measure of similarity (or closeness) as the length of the interpolating curve in a suitable metric space.
For the scalar interpolation formula (2.2), the definition

$$d(s_1, s_2) = \int_0^1 \sqrt{\langle \dot{s}(t), \dot{s}(t) \rangle_{s(t)}} dt = |\log\left(\frac{s_1}{s_2}\right)| \tag{4.39}$$

where $g_{s(t)}$ is the chosen scalar product at $s(t)$, coincides with the (scale-invariant) geometric distance between two positive numbers.
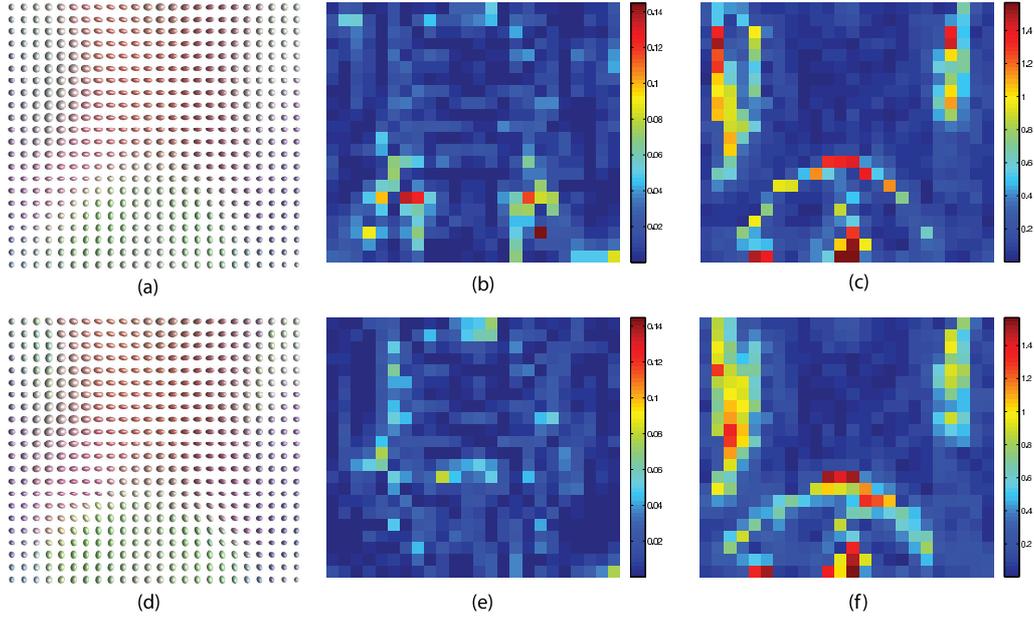
Figure 4.12: Results of the successive rotations of $2\pi/15$ applied to a real image. *Top:* Means were computed with the Log-Euclidean framework. (a): Output of the protocol, (b) Map of the squared differences in FA between the original image and the output. (c) Map of the squared differences in the direction of principal diffusion between the original image and the output. *Bottom:* Idem to top with the spectral-quaternion framework. It can be observed that the (localized) errors in FA are higher with the Log-Euclidean framework (color scales are identical). In both cases, the errors in orientation are concentrated in low anisotropy areas.

Likewise, for the four interpolating formulas considered in this paper, the definition

$$d(\mathbf{S}_1, \mathbf{S}_2) = \int_0^1 \sqrt{\langle \dot{\mathbf{S}}(t), \dot{\mathbf{S}}(t) \rangle_{\mathbf{S}(t)}}\, dt \tag{4.40}$$

qualifies as valid distance on $\mathrm{S}_+(3)$.

The Riemannian distances associated with the affine-invariant metric and the Log-Euclidean metric have been defined in Equations (3.19) and (3.22).

For the spectral geometry, the distance on the Lie group $\mathrm{SO}(3) \times \mathrm{D}_+(3)$ could be taken as given by formula (4.40) with the following metric.

Definition 4.5.1. *The metric associated to the Lie group* $\mathrm{SO}(3) \times \mathrm{D}_+(3)$ *is given by*

$$g_{\mathbf{S}(t)}(\dot{\mathbf{S}}(t), \dot{\mathbf{S}}(t)) = k^2(\mathbf{S}(t))\, g_{\mathbf{U}(t)}^{\mathrm{SO}(3)}(\dot{\mathbf{U}}(t), \dot{\mathbf{U}}(t)) + \sum \left(\frac{\dot{\Lambda}_i(\mathbf{S}(t))}{\Lambda_i(\mathbf{S}(t))} \right)^2 \tag{4.41}$$

which is a weighted sum of the (infinitesimal) Riemannian distances on $\mathrm{SO}(3)$ *and* $\mathrm{D}_+(3)$, *and k is a weighting factor to be defined in the sequel.*

The Riemannian distance on $\mathrm{SO}(3)$ is not global because of cut-locus singularities.

The spectral-quaternion geometry replaces the geodesic distance on $\mathrm{SO}(3)$ by a chordal distance in the space of quaternion, leading to another metric.

Definition 4.5.2. *Using the chordal distance in the space of quaternion, a novel metric can be defined as*

$$g_{\mathbf{S}(t)}(\dot{\mathbf{S}}(t), \dot{\mathbf{S}}(t)) = k^2(\mathbf{S}(t)) \, ||\dot{\mathbf{q}}(t)||^2 + \sum \left(\frac{\dot{\Lambda}_i(\mathbf{S}(t))}{\Lambda_i(\mathbf{S}(t))} \right)^2 . \tag{4.42}$$

Building upon the interpolating curves (4.29), we propose to let the weighting factor in (4.42) be

$$k(\mathbf{S}(t)) = f(\mathrm{HA}(t)), \tag{4.43}$$

where f is defined in Eq. (4.27). Thus defined, it appears as a sigmoid that discards the (infinitesimal) distance with the orientation term when the considered tensor is isotropic (since this term contains no information), and fully accounts for orientation in the case of large anisotropy (that is, k is close to 1 when the ratio between the largest and smallest eigenvalue becomes larger than 20).

It is hopeless to compute in explicit form the geodesic distance d associated to the metric (4.42). However, the geodesic distance satisfies

$$\sum |\log \frac{\lambda_i(\mathbf{S}_1)}{\lambda_i(\mathbf{S}_2)}| \leq d(\mathbf{S}_1, \mathbf{S}_2) \leq$$

$$\kappa(\mathbf{S}_1, \mathbf{S}_2)||\mathbf{q}_1 - \mathbf{q}_2^r||_2 + \sum |\log \frac{\lambda_i(\mathbf{S}_1)}{\lambda_i(\mathbf{S}_2)}| := \Delta(\mathbf{S}_1, \mathbf{S}_2) \tag{4.44}$$

with

$$\kappa(\mathbf{S}_1, \mathbf{S}_2) = f(\min(\mathrm{HA}_1, \mathrm{HA}_2)).$$

The lower bound is obvious since $k(\mathbf{S}(t)) \geq 0$ along the path; the upper bound is the (Riemannian) length in the sense of metric (4.42) of a curve originating at the tensor of minimal isotropy and made of two arcs: the first one linking the orientations at constant (minimal) anisotropy and the second one linking the eigenvalues while the orientation remains fixed. The geodesic distance being by definition smaller than the length of any path, the upper bound is obtained.

In the limit case of an isotropic tensor, the inequality above becomes an equality, and when it is close to identity, $\Delta(\mathbf{S}_1, \mathbf{S}_2)$ is thus a good approximation of the geodesic distance. It is meaningful, as the weighting factor goes to zero when the orientation information becomes uncertain. This measure captures the fact that the cost of rotation is weighted with the orientation information. We propose formula (4.44) as a convenient substitute for the geodesic distance associated to the Riemannian metric (4.41-4.43). This similarity measure is not a distance because it does not satisfy the triangular inequality. However, the bounds in (4.44) show that it approximates the distance when the anisotropy is small and weights the distance between quaternion only when the anisotropy is high, that is, when the orientation of tensors carries physical information.

It should be noted that the proposed interpolating curve (4.29), metric (4.42), and similarity measure (4.44) are not very sensitive to the form of the sigmoid f, nor to the parameter β. The rationale for the term $k(\mathbf{S}(t))$ is simply to let the orientation quaternion $\mathbf{q}$ play a role only when the anisotropy exceeds a certain threshold.

We close this section with a conceptual remark for the interested reader. The interpolating curve (4.29), metric (4.42), and similarity measure (4.44) approximate a Riemmanian framework in $S^+(3)$ by properly combining a Riemmanian framework in $D^+(3)$ and in $SO(3)$. This means that distances between ellipsoids are approximated by the weighted sum of distances between rotations and distances between dilations. This idea is in close analogy with the framework recently developed by the authors (Bonnabel and Sepulchre, 2009; Bonnabel et al., 2013) to define distances between flat ellipsoids, that is, positive semidefinite tensors of fixed rank: the proposal being to approximate the distance between the flat ellipsoids by the weighted sum of distances between the subspaces that support them and distances between full ellipsoids in a common subspace. The Riemmanian framework proposed for flat ellipsoids is probably not directly relevant for the DTI application but it is of interest to notice that different applications motivate a separate weighting of rotations and dilations, leading to similar theoretical and computational issues about reconciling the geometry of the product space to the geometry of the original space.

4.6 Computation of the spectral-quaternion tools

The basic steps for the implementation of the tools presented in this chapter are summarized in the following pages.

Algorithm 1: Computation of interpolation

`Inputs`: two tensors $\mathbf{S}_1$ and $\mathbf{S}_2$, a parameter t
`Output`: the interpolated tensor for the value t, $\mathbf{S}_t$

1. $\mathbf{S}_1 = \mathbf{U}_1 \Lambda_1 \mathbf{U}_1^T$, $\mathbf{S}_2 = \mathbf{U}_2 \Lambda_2 \mathbf{U}_2^T$, with the eigenvalues in decreasing order.

2. Compute the quaternions representing the orientations, $\mathbf{q}_1$ and $\mathbf{q}_2$, and the associated set $\mathcal{Q}_2$. This set contains the four elements

$$\left\{ \begin{pmatrix} \mathbf{q}_2(1) \\ \mathbf{q}_2(2) \\ \mathbf{q}_2(3) \\ \mathbf{q}_2(4) \end{pmatrix} \begin{pmatrix} \mathbf{q}_2(2) \\ -\mathbf{q}_2(1) \\ -\mathbf{q}_2(4) \\ \mathbf{q}_2(3) \end{pmatrix} \begin{pmatrix} \mathbf{q}_2(3) \\ \mathbf{q}_2(4) \\ -\mathbf{q}_2(1) \\ -\mathbf{q}_2(2) \end{pmatrix} \begin{pmatrix} \mathbf{q}_2(4) \\ -\mathbf{q}_2(3) \\ \mathbf{q}_2(2) \\ -\mathbf{q}_2(1) \end{pmatrix} \right\}$$

   and their opposites.

3. Compute the interpolated eigenvalues matrix, using $\lambda_i(t) = \exp((1-t)\log(\lambda_{i,1}) + t\log(\lambda_{i,2})$.

4. Select the 'realigned' quaternion $\mathbf{q}_2^a$ with respect to $\mathbf{q}_1$ using $\mathbf{q}_2^a = \arg\max_{\mathbf{q}_2 \in \mathcal{Q}_2} \mathbf{q}_1^r . \mathbf{q}_2$

5. Compute the interpolated quaternion as

$$\mathbf{q}_m(t) = (1-t)\mathbf{q}_1 + t\mathbf{q}_2^a$$

$$\mathbf{q}(t) = \mathbf{q}_m(t)/||\mathbf{q}_m(t)||$$

6. Compute $\mathbf{U}(t)$ as the rotation matrix corresponding to $\mathbf{q}(t)$.

7. The interpolated tensor is given by

$$\mathbf{S}(t) = \mathbf{U}(t)\Lambda(t)\mathbf{U}(t)^T .$$

Algorithm 2: Computation of weighted mean

`Inputs`: A set of tensors $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_N$ and their associate weights $w_i, i = 1, \ldots, N$.
`Output`: the mean tensor $\mathbf{S}_\mu$

1. Perform the spectral decomposition of each of the N tensors.

2. Compute the weighted mean of eigenvalues, through

$$\lambda_{\mu,k} = \exp(\sum_{i=1}^{N} w_i \log(\lambda_{i,k})), \quad k = 1, 2, 3.$$

3. Select as the reference tensor the one that maximizes the product $w_i \mathrm{HA}_i$. The reference quaternion is $\mathbf{q}^r$.

4. For each tensor, select the 'realigned' quaternion $\mathbf{q}_i^a$ with respect to $\mathbf{q}^r$ using
$\mathbf{q}_i^a = \arg\max_{\mathbf{q}_i \in \mathcal{Q}_i} \mathbf{q}^r . \mathbf{q}_i$

5. The weighted mean of quaternions is

$$\mathbf{q}_m = \sum_i w_i \mathbf{q}_i^a$$

$$\mathbf{q}_\mu = \frac{\mathbf{q}_m}{||\mathbf{q}_m||}$$

6. Compute $\mathbf{U}_\mu$ from $\mathbf{q}_\mu$ according to (4.10).

7. The mean tensor is $\mathbf{S}_\mu = \mathbf{U}_\mu \Lambda_\mu \mathbf{U}_\mu^T$.

Algorithm 3: Computation of similarity measure

`Inputs`: Two tensors $\mathbf{S}_1$ and $\mathbf{S}_2$.
`Output`: the measure of their similarity $\Delta(\mathbf{S}_1, \mathbf{S}_2)$

1. Perform the spectral decomposition of each tensor, and compute the quaternions corresponding to the orientation matrices.

2. The distance between eigenvalues is given by

$$d_\Lambda = \sum |\log \frac{\lambda_i(\mathbf{S}_1)}{\lambda_i(\mathbf{S}_2)}|$$

3. The weighting factor is given by $\kappa(\mathbf{S}_1, \mathbf{S}_2) = \frac{(\beta \min(\mathrm{HA}_1, \mathrm{HA}_2))^4}{1 + (\beta \min(\mathrm{HA}_1, \mathrm{HA}_2))^4}$, with $\beta = 0.6$.

4. Compute the chordal distance between quaternions, using one of the quaternion as the reference and realigning the other one to this reference, *i.e.* if $\mathbf{q}_1$ is the reference, select the 'realigned' quaternion $\mathbf{q}_2^a$ with respect to $\mathbf{q}_1$ using $\mathbf{q}_2^a = \arg\max_{\mathbf{q}_2 \in \mathcal{Q}_2} \mathbf{q}_1 . \mathbf{q}_2$.

   The distance between quaternions is

$$d_\mathbf{q} = ||\mathbf{q}_1 - \mathbf{q}_2^a||_2 .$$

5. Finally, $\Delta(\mathbf{S}_1, \mathbf{S}_2) = \kappa d_\mathbf{q} + d_\Lambda$.

4.7   Conclusion

This chapter had introduced the novel framework for DTI processing which is called the *spectral-quaternion* framework. The development of this method was motivated by the observed limitation of the affine-invariant and Log-Euclidean methods. Indeed, those two methods have the drawback of degrading the anisotropy during the processing, although this information is one of the most important conveyed by the tensor. The decoupling between the processing of intensities and the processing of orientations had enabled to avoid this limitation.

Together with the goal of anisotropy preservation, we focused on the computational cost as another important issue. The processing of rotations matrices can indeed be expensive, and the idea was to use the alternative parametrization of unit quaternions. The proposed framework is both information preserving and computationally efficient, as illustrated in Figure 4.13.
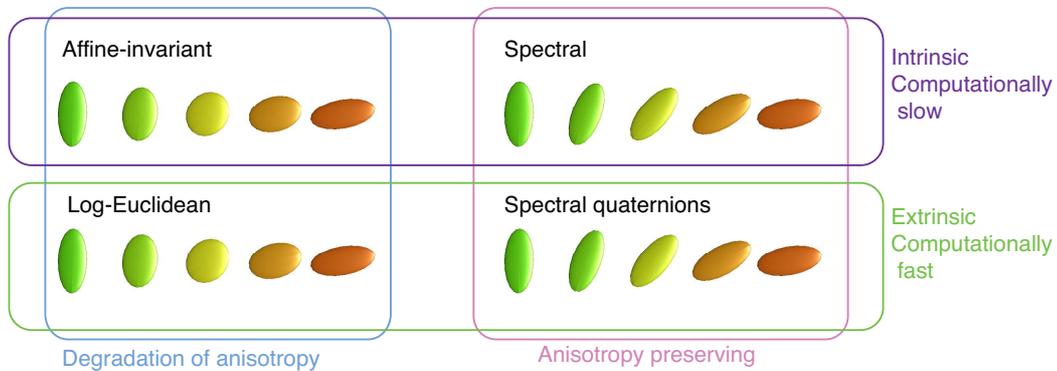

Figure 4.13: Comparison of the four frameworks presented in this dissertation: the affine-invariant framework (3.18), the Log-Euclidean framework (3.21), the spectral framework (4.1) and the spectral quaternion framework (4.20).

Another contribution of this chapter concerns the uncertainty of orientation for isotropic (or nearly isotropic) tensors. Indeed, the orientation of those tensors is highly uncertain, even irrelevant for isotropic tensors. Based upon this observation, a weighted interpolating curve between orientations has been proposed. This uncertainty is also taken into account in the development of our similarity measure.

Finally, the important concept of weighted mean has been defined, leading to a general definition of mean that is univoque and commutative. This mean satisfies the properties of invariance by scaling and rotation, and preserves all the important features of tensors.

The work presented in this chapter is the subject of a submitted paper to *International Journal of Computer Vision*:

Anne Collard, Silvère Bonnabel, Christophe Phillips and Rodolphe Sepulchre. Anisotropy preserving DTI processing. Submitted in revised form to *International Journal of Computer Vision*, May 2013.

Transformations and filtering of Diffusion Tensor Images

> *The goal of this chapter is to apply the basic operations defined in the preceding chapter to some simple processing algorithms.*
> *Using existing reorientation strategies, we show how to transform Diffusion Tensor Images with our spectral-quaternion framework. A simple rigid-body registration algorithm is also presented.*
> *Weighted means of tensors are used to filter DT images, with a Gaussian kernel and through a Perona-Malik anisotropic filter. The results of our framework are compared with algorithms using the Log-Euclidean method.*
> *The results of this chapter suggest that our framework can be used in all processing algorithms, and is competitive with state-of-the-art methods.*

In the preceding chapters, four different Riemannian frameworks for the processing of Diffusion Tensor Images have been studied. The main focus was on three basic operations: interpolation, averaging, and similarity measure. In this chapter, we use these basic operations to illustrate the generalization of some usual processing algorithms to the case of matrix-valued images.

To analyse Diffusion Tensor Images, multiple pre- and post-processing steps are required. In this chapter, we will focus on the transformation of Diffusion Tensor Images, which can occur in many of those steps, for example in registration, or in the construction of an atlas. The second application of interest here is the filtering of DT images. Filtering of images is often necessary to reduce the measurement noise. The simple Gaussian filtering will be illustrated, followed by one version of an anisotropic filtering based on the well-known developments of Perona and Malik (Perona and Malik, 1990).

Since both the Log-Euclidean and spectral-quaternion frameworks are computationally efficient, a main objective will be to compare the merits of the two methods in DTI processing.

5.1 Transformations of DT images

Transformations of images appear in many processing algorithms, and the way they are performed might have a huge impact on the final outputs of those algorithms, as explained

in (Muñoz-Moreno et al., 2009). Two major concepts have to be considered when deal-
ing with transformations of Diffusion Tensor Images: the interpolation method and the
reorientation strategy. Those two concepts will be explained in the following. We first
introduce the notations which will be used throughout this section.

If the transformed image f' is computed from an original image f through the trans-
formation M, this image f' satisfies

$$f'(x) = f(M(x)) \tag{5.1}$$

where x denotes a position 'on the grid', $x \in \mathbb{N}^3$. If the transformation is affine *i.e.* it
does not depend on the position on the grid, it has the form $M(x) = \mathbf{M}x + \mathbf{t}$, where $\mathbf{M}$ is
a $n \times n$ linear transformation matrix, and $\mathbf{t}$ is a vector which expresses a rigid translation
of the image. Equation (5.1) thus becomes

$$f'(x) = f(\mathbf{M}x + \mathbf{t}) \tag{5.2}$$

5.1.1   Interpolation method

During a transformation, it generally occurs that the value $M(x)$ is not a natural, *i.e.* it
does not correspond to a new point in the grid. This means that the value $f(M(x))$ is not
defined. It is here that interpolation methods come to rescue: they enable to compute an
'approximation' of this tensor, which will be used to assign its value to $f'(x)$ (Thévenaz
et al., 2000).

As explained in the preceding chapter, interpolation between tensors amounts to com-
puting weighted means. However, it should be noted that other non mean-based interpo-
lation methods are used. This is the case for example in (Alexander et al., 2001; Whitcher
et al., 2007), which rely on a *nearest neighbor resampling* in order to avoid averaging of
diffusion tensors. This method consists in taking the tensor at the nearest location as the
output of the interpolation.

In this work, we have decided to compute interpolated tensors through weighted means.
There are different ways of assigning the interpolation weights, which would lead to dif-
ferent 'interpolation methods'. A simple trilinear method was proposed in Chapter 4, but
other more complex methods could be used, as polynomial interpolation, methods based
on B-splines, or on a Gaussian kernel.

5.1.2   Reorientation strategy

The important concept of *reorientation of Diffusion Tensors* has been introduced in
(Alexander et al., 2001) and widely used since then.

The paper addresses the issue of applying spatial transformations to Diffusion Tensor
Images. Indeed, compared to scalar images, the transformation of Diffusion Tensor Images
is complicated by the fact that the tensors contain orientation information, which is itself
affected by the transformation. This problem is illustrated in Figure 5.1. A rigid rotation
of $\pi/6$ rad is applied to a Diffusion Tensor Image ((a) in the Figure). If the rotation
is naively applied to the image, by simply constructing the value at each pixel in the
transformed image f' from an interpolation between the tensors in the neighborhood of

the corresponding position in the original image f, we obtain the result shown in Figure 5.1, (b). It can clearly be seen that the orientation of tensors is problematic. The red box highlights some positions where this problem is particularly visible. The tensors which were horizontal in the original image are still horizontal after the transformation. This observation explains why a reorientation strategy has to be implemented together with the transformation of the image. In a few words, if the image f has been rotated by $\mathbf{R}$, each tensor in the transformed image f' should be replaced by

$$\mathbf{S}' = \mathbf{R}\mathbf{S}\mathbf{R}^{T}. \tag{5.3}$$

The Figure 5.1,(c) illustrates how this reorientation enables to obtain tensors which are adequately orientated. In particular, the orientation of tensors in the red box is now coherent with respect to the true image. It is important to note that if the transformation is rigid, each tensor will be rotated by the same rotation matrix. However in the case of an elastic transformation, it could be beneficial to compute local estimates of the 'reorientation matrix', *i.e.* for each voxel.
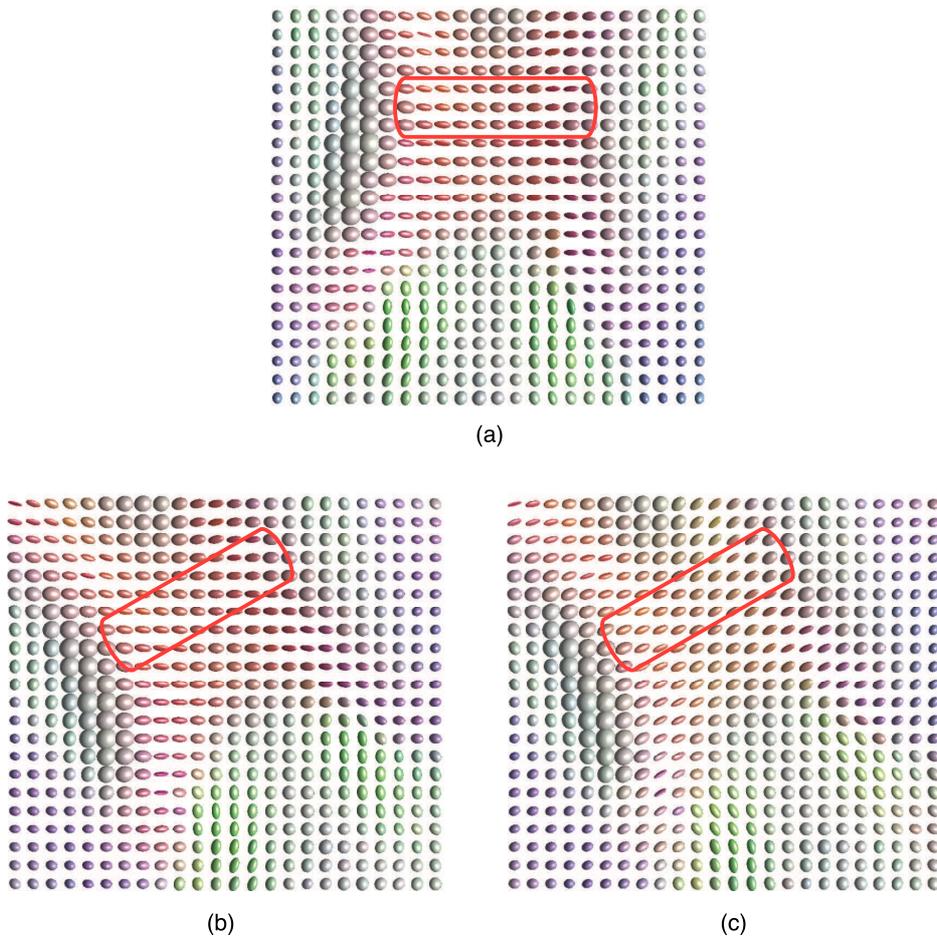
(a)

(b)

(c)

Figure 5.1: Illustration of the problem which could arise if no reorientation procedure is implemented during an image transformation.

Different reorientation strategies have been proposed in (Alexander et al., 2001), depending upon the type of transformation applied. The simplest but less accurate strategy consists in ignoring the problem of reorientation, and to simply interpolate the tensors.

The second strategy, called the *Finite Strain* (FS) strategy, is particularly adapted for affine transformations, and widely used in the literature thanks to its simplicity. If we consider the (nonsingular) transformation matrix $\mathbf{M}$, it can always be decomposed to a *rigid* rotation component $\mathbf{R}$ and a deformation component $\mathbf{D}$, through $\mathbf{M} = \mathbf{DR}$. One strategy is thus to 'extract' the rotation component and to apply it to each tensor of the image. This rotation component can be computed through

$$\mathbf{R} = (\mathbf{MM}^T)^{-1/2}\mathbf{M}. \tag{5.4}$$

This method can easily be used with the spectral-quaternion framework. In this case, the obtained rotation matrix $\mathbf{R}$ is converted to a unit quaternion, which will be used throughout the computations.

The third strategy proposed in (Alexander et al., 2001) is called the *preservation of principal direction* (PPD). This strategy was developed in order to overcome a possible limitation of the finite strain one. Indeed, with the FS strategy, the deformation component of the transformation is discarded. However, transformations as shearing or stretching also affect the orientation of the image. This impact, which usually depends on the original orientation of the structure, should be accounted for in the transformation.

Briefly, the chosen rotation should map the first eigenvector of $\mathbf{S}$, $\mathbf{e}_1$ to $\mathbf{n}_1 = \mathbf{Me}_1/|\mathbf{Me}_1|$ and the second eigenvector $\mathbf{e}_2$ to a unit vector perpendicular to $\mathbf{n}_1$ in the plane spanned by $\mathbf{n}_1$ and $\mathbf{n}_2 = \mathbf{Me}_2/|\mathbf{Me}_2|$. In other words, the PPD method preserves the principal direction of the diffusion tensor through the transformation, as well as the plane of the first two eigenvectors. The detailed method to compute the rotation matrix $\mathbf{R}$ can be found in the appendix of (Alexander et al., 2001).

In practice, this method is less often used than the Finite Strain one, because it is more complicated. Moreover, it can not be directly used within our spectral-quaternion framework. To apply it with this framework would imply to 'go back' to the rotation matrix space, in order to obtain the eigenvectors. The increase in accuracy would then be obtained at the expense of a significant increase in computational cost.

5.1.3   Rigid-body registration

To compare a pair of images, it is necessary to make sure that each pixel of those images is related to the same physical position in the brain. The procedure which finds the geometrical transformation that maps one image into another, so that objects or structures in both images are aligned, is called *image registration*. A lot of procedures exist for the registration of scalar images, but their extension to the diffusion tensor case is complicated by the many features of tensors. Multiple issues arise when considering tensors, as for example the problem of their reorientation. As explained in the preceding section, tensors must be reoriented after a transformation. However, it is also important to take this step into account when searching for the best transformation.

A detailed review of the principal issues of Diffusion Tensor Images registration is found in (Muñoz-Moreno et al., 2009). Many papers are concerned with this topic, using varying similarity measures, or different degrees of freedom for the desired transformation. For example, in (Goh and Vidal, 2006), a closed-form formula is given for the rigid registration of images. Affine registration is the topic of (Pollari et al., 2006), while (Alexander and Gee, 2000) focus on elastic matching. More complex frameworks, such as diffeomorphic ones, can be found in (Cao et al., 2006) and in (Yeo et al., 2009, 2008) which derive an exact differential using the Finite Strain strategy.

Registration methods are of course based on a *similarity measure* between images. The spectral-quaternion framework and its associated measure could be used to drive registration procedures, since this method has been shown to be particularly efficient for the processing of the interesting features of the tensors. As a proof of concept, a very simple rigid-body algorithm is considered in this thesis, using the spectral-quaternion similarity measure, the corresponding definition of weighted means and the Finite Strain strategy for the reorientation of tensors. If f is the *moving* image and g is the *target* image, the goal of the algorithm is to search the minimum of a cost function χ^2. In the scalar case, this cost function is the sum over all pixels of the squared difference between values. A simple extension of this function to the case of diffusion tensors is

$$\chi^2 = \sum_i d^2(\mathbf{R}f(\mathbf{Mx}_i + \mathbf{t})\mathbf{R}^T, g(\mathbf{x}_i)) \tag{5.5}$$

where $\mathbf{R}$ is the rotation component of $\mathbf{M}$, computed using (5.4), and d is the spectral-quaternion similarity measure. For comparison purposes, the method was also implemented with the Log-Euclidean distance. The cost function can be optimized by using a Gauss-Newton optimization, since it can be written as

$$\chi^2 = \sum_i b_i^2$$
$$b_i = d(\mathbf{R}f(\mathbf{Mx}_i + \mathbf{t})\mathbf{R}^T, g(\mathbf{x}_i))$$

For a rigid body registration, the transformation $\mathbf{Mx} + \mathbf{t}$ is dependent upon 6 parameters : three translation parameters and three angles of rotation. The translation $\mathbf{t}$ writes

$$\mathbf{t} = \begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix} \tag{5.6}$$

while the matrix $\mathbf{M}$ is equal to a composition of three rotation matrices, $\mathbf{M} = \mathbf{R}_x\mathbf{R}_y\mathbf{R}_z$, *i.e.*

$$\mathbf{M} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(m_4) & -\sin(m_4) \\ 0 & \sin(m_4) & \cos(m_4) \end{pmatrix} \begin{pmatrix} \cos(m_5) & 0 & -\sin(m_5) \\ 0 & 1 & 0 \\ \sin(m_5) & 0 & \cos(m_5) \end{pmatrix} \begin{pmatrix} \cos(m_6) & -\sin(m_6) & 0 \\ \sin(m_6) & \cos(m_6) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{5.7}$$

In this particular case, the matrix $\mathbf{R}$ is equal to $\mathbf{M}$, since the translation does not affect the orientation of tensors. The gradient of the functions b_i with respect to the

parameters of the transformation is needed to compute the optimization. An adaption of a finite-difference scheme has been defined, and the gradient will be approximated by

$$\frac{\partial b_i(m)}{\partial m_j} = \frac{d(\mathbf{R}^+ f(\mathbf{M}^+ x_i + \mathbf{t}^+)(\mathbf{R}^+)^T, g(x_i) - d(\mathbf{R}^- f(\mathbf{M}^- x_i + \mathbf{t}^-)(\mathbf{R}^-)^T, g(x_i))}{\Delta m_j} , \quad (5.8)$$

where $\mathbf{M}^+$ and $\mathbf{t}^+$ are the transformation matrix and translation vector where m_j is replaced by $m_j^+ = m_j + \frac{1}{2}\Delta m_j$, and $\mathbf{M}^-,\mathbf{t}^-$ are the same matrix and vector with $m_j^- = m_j - \frac{1}{2}\Delta m_j$. $\mathbf{R}^+$ and $\mathbf{R}^-$ are the corresponding rotation components.

This simple algorithm has been applied to a pair of images from a single subject which has been scanned twice, in two different positions (this justifies the fact that only rotation and translation are considered in the transformation). The same image slice for each of the acquisitions is shown in Figure 5.2. For the algorithm, the figure (a) is used as the fixed image, while the figure is (b) the moving image.

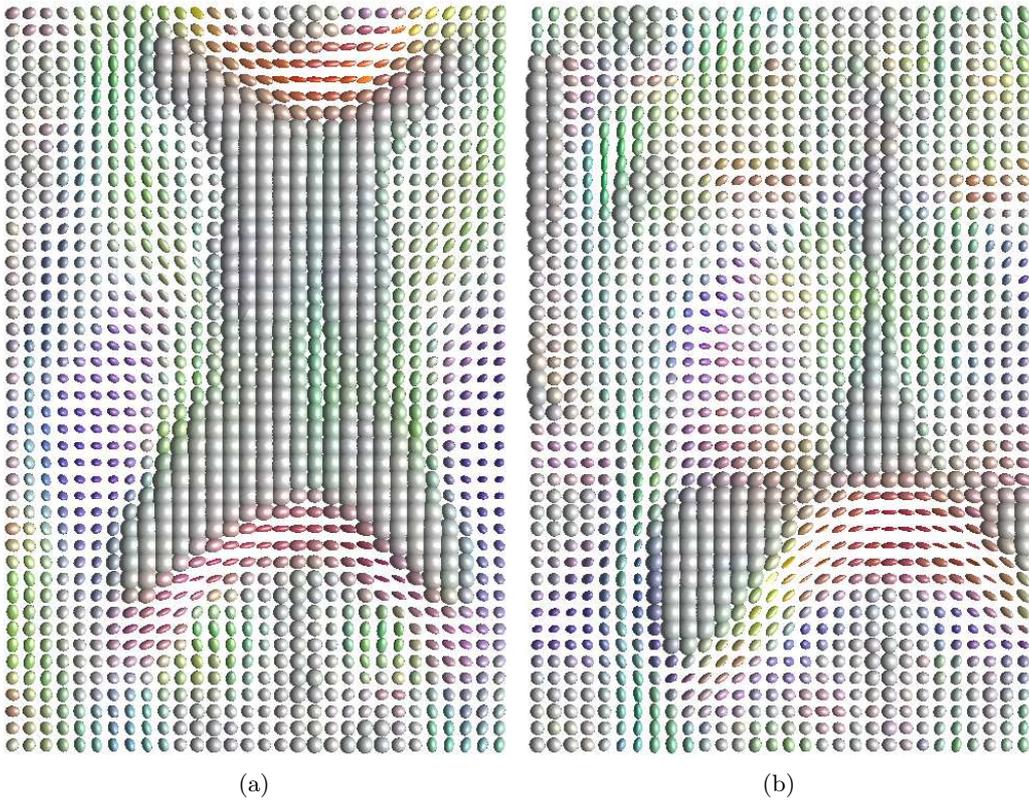(a)   (b)

Figure 5.2: Slice of the pair of images to be registered. The same slice was selected in the two images. They are very different, which is due to the different positions of the head of the subject in the scanner. For the algorithm, image (a) is the fixed image, while (b) is the moving one.

Figure 5.3 shows the results of the rigid-body registration: on the left, the Log-Euclidean similarity measure was used to compute the cost function, and the image was transformed using a Log-Euclidean interpolation. On the right, the spectral-quaternion method was used. In both cases, the images are reoriented using the Finite Strain strategy. At first sight, both figures are very close to the fixed image. This indicates that the algorithms are efficient at finding the 'true' displacement of the subject. The two results are very similar, which suggests that the spectral-quaternion method is competitive with respect to the Log-Euclidean framework, and that its application in practice should be considered. The comparison of the obtained parameters for the transformation indicates a similar observation, as we have (where the units are the size of voxels).

$$\mathbf{t}_{LE} = \begin{pmatrix} 2.94 \\ 3.76 \\ 7.98 \end{pmatrix}, \mathbf{t}_{SQ} = \begin{pmatrix} 2.53 \\ 3.53 \\ 7.93 \end{pmatrix} \tag{5.9}$$

and

$$\mathbf{M}_{LE} = \begin{pmatrix} 0.9998 & 0.0213 & 0.0038 \\ -0.0213 & 0.9998 & 0.0055 \\ -0.0037 & -0.0056 & 1 \end{pmatrix}, \mathbf{M}_{SQ} = \begin{pmatrix} 0.999 & 0.0208 & 0.0114 \\ -0.0209 & 0.9998 & 0.0072 \\ -0.0112 & -0.0074 & 0.9999 \end{pmatrix}. \tag{5.10}$$

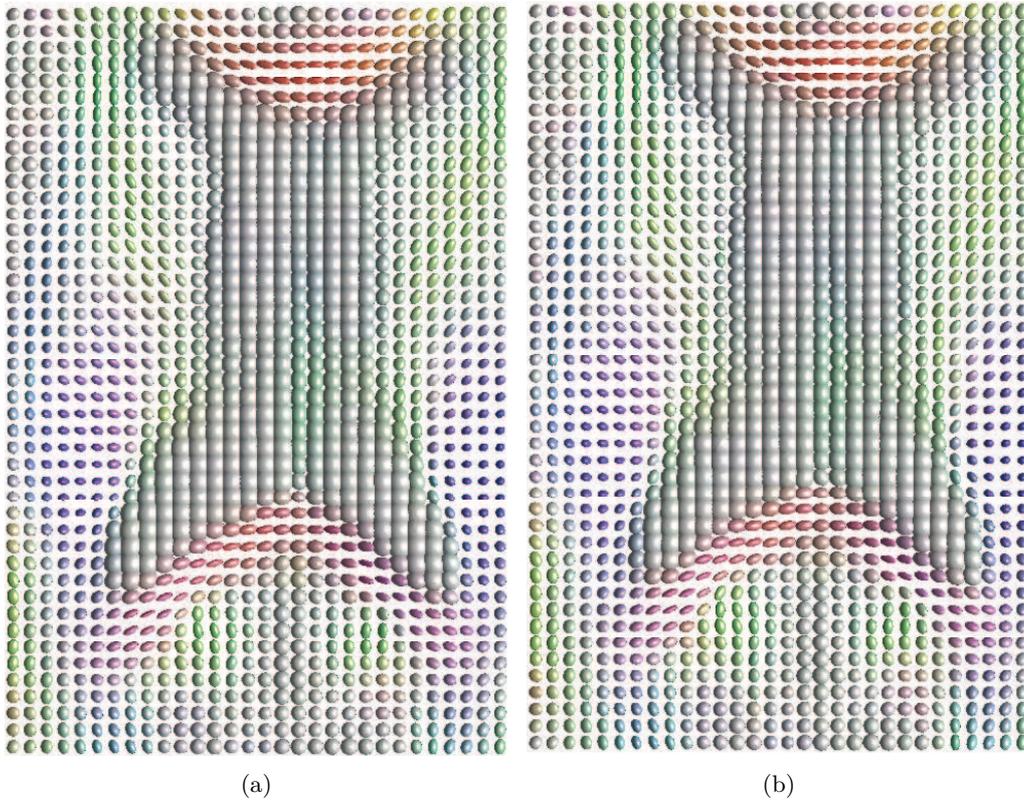<div align="center">(a)   (b)</div>

Figure 5.3: Results of the simple registration algorithm. (a): With the Log-Euclidean similarity measure and interpolation. (b): With the spectral-quaternion similarity measure and interpolation. Both results are very similar, and similar to the fixed image shown in Figure 5.2, (a). This is a good indication that our framework could be used to develop more complex algorithms.

5.2 Filtering of DT images

In this section, we will illustrate how the framework introduced in Chapter 4 affects the results of some important processing methods. All these methods are based on the main notion of weighted means. Many image processing tasks imply the convolution of the image with some kernels. This is the case of denoising, smoothing, edge detection, image sharpening and filtering. The convolution of diffusion tensor images is straightforward from the definition of means.

Gaussian filtering

In a continuous setting, the convolution of a vector field $F_0(x)$ by a Gaussian G_σ is given by

$$F(x) = \int_y G_\sigma(y - x)F_0(y)dy \,.$$

In the case of images (*i.e* pixels or voxels on a discrete grid), the coefficients are renormalized since the neighborhood $\mathcal{V}$ is usually limited to points within one to three times the standard deviation σ. The discretization is given by

$$F(x) = \frac{\sum_{u\in\mathcal{V}(x)} G_\sigma(u)F_0(x + u)}{\sum_{u\in\mathcal{V}(x)} G_\sigma(u)} \,, \tag{5.11}$$

which is nothing else than a weighted averaging, where the weights are given by the coefficients G_σ. We illustrate tensor image smoothing on a 2D example: one slice through a true image (Figure 5.4 (a)) is corrupted by additive noise (Figure 5.4 (b)) then Gaussian filtered. The size of the Gaussian filter was 7×7, and $\sigma = 3$ pixels. The Log-Euclidean filtering (Figure 5.4 (c)) is compared to a convolution based on the spectral-quaternion framework (Figure 5.4 (d)). It can be observed that both methods produce smoothened images.

FA maps were derived from the noisy tensor image (Figure 5.5 (a)), and the smoothed images obtained with the Log-Euclidean and spectral-quaternion approaches (Figure 5.5 (c) and Figure 5.5 (d) respectively). The FA map of the noisy tensor was also smoothed (Figure 5.5 (b)) as would be classically done. It is clear from those figures that tensors which are highly anisotropic in the true image become much more isotropic with the Log-Euclidean filtering (Figure (c)). This (expected) degradation of anisotropy is fortunately less important with the spectral-quaternion framework (Figure (d)). For comparison purpose, a Gaussian filtering was applied directly on the FA map of the noisy image. The resulting map is shown in Figure 5.5,(b). It is interesting to note that the FA map of the spectral-quaternion filtered image (Figure (d)) is more similar to the filtered FA map (Figure (b)) than the Log-Euclidean one. This means that the spectral-quaternion smoothing affects FA as if working directly on the scalar image.

Anisotropic filtering

Anisotropic regularization of images corrupted by noise is very important in medical image processing, since it allows for a reduction of the noise level while preserving boundaries and structures. The main idea of anisotropic filtering is to 'penalize' the smoothing in the directions where the derivative is important (Perona and Malik, 1990). As proposed in (Pennec et al., 2006), if $c(\cdot)$ is a weighting function decreasing from $c(0) = 1$ to $c(\infty) = 0$,
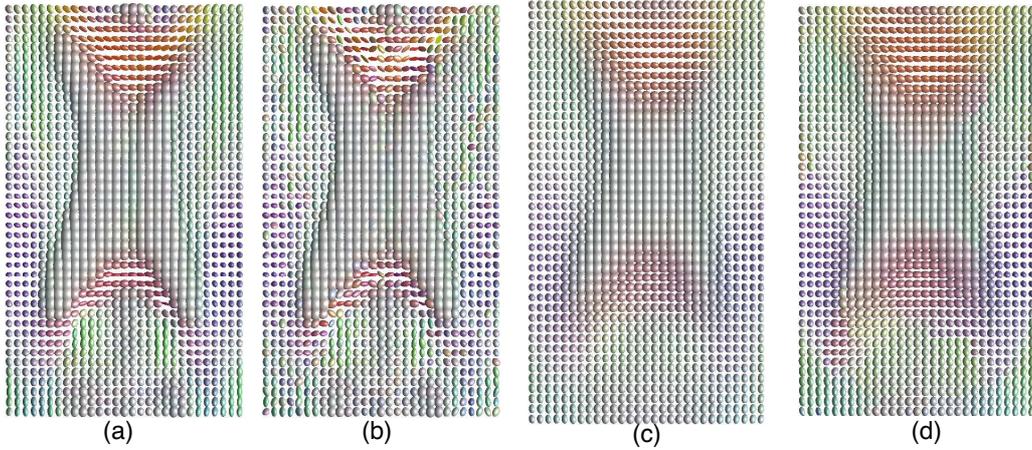
Figure 5.4: Gaussian filtering of a true image. *From left to right:* Slice of an image constructed from real data, noisy image, result of the Gaussian filtering with the Log-Euclidean method, Gaussian filtering with the spectral-quaternion framework. For both methods, boundaries are less visible after the filtering.

anisotropic filtering can be realized using a discrete implementation of the Laplacian operator. The contribution $\Delta_u\mathbf{S}$ of the spatial direction u to the Laplace-Beltrami operator is weighted by the decreasing function, according to the norm of the gradient in that direction. The norm of the gradient should of course be evaluated according to the corresponding metric or measure.

Finite difference approximations of the derivatives are explained in (Pennec et al., 2006) for the affine-invariant metric. The update of a tensor $\mathbf{S}(x)$ at position x is computed through

$$\mathbf{S}_{n+1}(x) = \mathbf{S}(x)^{1/2} \exp\left(\mathbf{S}(x)^{-1/2} 2\varepsilon\Delta_a\mathbf{S}(x)\mathbf{S}(x)^{-1/2}\right)\mathbf{S}(x)^{1/2}, \tag{5.12}$$

where $\Delta_a\mathbf{S}(x)$ is the Laplace-Beltrami operator for the anisotropic filtering. Using a finite set $\mathcal{V}$ of directions u, this operator can be approximated as

$$\Delta_a\mathbf{S}(x) \simeq \frac{2d}{\text{card}(\mathcal{V})} \sum_{u\in\mathcal{V}} \left[c\left(\|\log(\mathbf{S}(x)^{-1/2}\mathbf{S}(x+u)\mathbf{S}(x)^{-1/2}\|_2\right) \right.$$

$$\left. \times \mathbf{S}(x)^{1/2}\log\left(\mathbf{S}(x)^{-1/2}\mathbf{S}(x+u)\mathbf{S}(x)^{-1/2}\right)\mathbf{S}(x)^{1/2}\right] \tag{5.13}$$

where d is the dimension of the image and $\text{card}(\mathcal{V})$ is the cardinality of the set $\mathcal{V}$. It is clear from the equations that this algorithm is expensive to compute, since it requires many logarithms and square roots of matrices.

In the case of the Log-Euclidean metric, the schemes are simpler since the dependence of the gradient on the current tensor disappears on the logarithms of tensors. The norm of gradient is then easily computed, as explained in (Arsigny et al., 2006; Fillard et al., 2007).
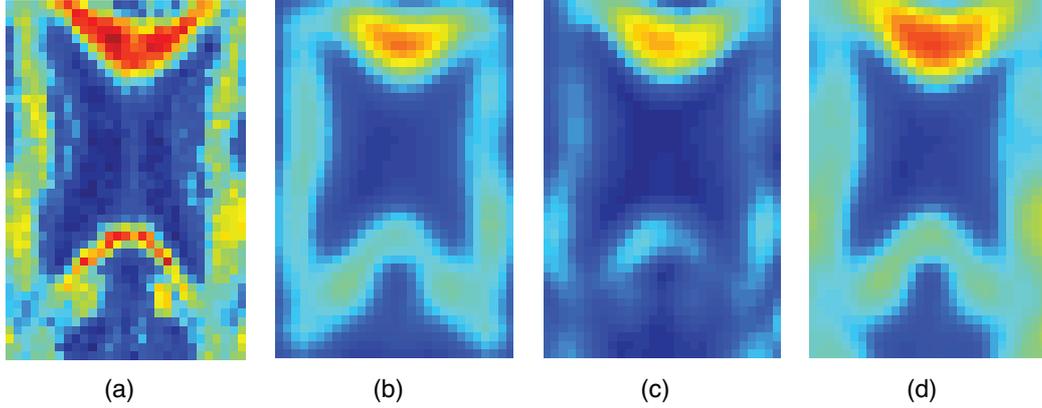
(a)   (b)   (c)   (d)

Figure 5.5: FA maps of the Gaussian filtering of a true image. *From left to right:* FA map of the noisy image (Figure 5.4 (b)), Gaussian filtering directly applied to the noisy FA map, FA map of the Gaussian filtering with the Log-Euclidean method (Figure 5.4 (c)), FA map of Gaussian filtering with the spectral-quaternion framework (Figure 5.4 (d)). It can be observed that tensors which are highly anisotropic in the true image are more isotropic with the Log-Euclidean framework. This degradation is less important with the spectral-quaternion method. Moreover, the FA map of the spectral-quaternion filtering is closer to the Gaussian filtering of the noisy FA map than the one of the Log-Euclidean filtering.

In this case, the update is given by

$$\mathbf{S}_{n+1}(x) = \exp(2\varepsilon\Delta_{LE}\mathbf{S}(x)) \tag{5.14}$$

where the Laplace-Beltrami operator can be computed through

$$\Delta_{LE}\mathbf{S}(x) = \frac{2d}{\text{card}(\mathcal{V})} \sum_{u\in\mathcal{V}} \left[c\left(||\log\mathbf{S}(x) - \log\mathbf{S}(x+u)||_2\right)\right.$$

$$\times \left.(\log\mathbf{S}(x+u) - \log\mathbf{S}(x))\right]. \tag{5.15}$$

It is important to note that all the computations can be made on the logarithms of the tensors, and that the exponential can be applied only at the end of the algorithm, to come back to the tensor space. This reduces the cost of the filtering.

With the spectral-quaternion framework, the situation is similar to the one of the Log-Euclidean method. Indeed, with this framework, all the computations can be done in an Euclidean way by using the quaternions and the logarithms of eigenvalues. For each pixel at position x and at each time step, both the orientation and the eigenvalues have to be computed.

For the quaternions, the update equation is a simple chordal sum, followed by a normalization of the quaternion.

$$\mathbf{q}_{\text{temp}}(x) = \mathbf{q}_n(x) + 2\varepsilon_q\Delta\mathbf{q}(x), \tag{5.16}$$

$$\mathbf{q}_{n+1}(x) = \frac{\mathbf{q}_{\text{temp}}(x)}{||\mathbf{q}_{\text{temp}}(x)||}, \tag{5.17}$$

where

$$\Delta\mathbf{q}(x) = \frac{2d}{\text{card}(\mathcal{V})} \sum_{u \in \mathcal{V}} c(\sqrt{\kappa}||\mathbf{q}_n^*(x + u) - \mathbf{q}_n(x)||)\Delta_u\mathbf{q}(x) \tag{5.18}$$

with $\mathbf{q}_n^*(x + u)$ the realigned quaternion with respect to $\mathbf{q}_n(x)$ and

$$\Delta_u\mathbf{q}(x) = \mathbf{q}_n^*(x + u) - \mathbf{q}_n(x). \tag{5.19}$$

The parameter κ in the argument of function $c(\cdot)$ is the one used in the similarity measure (4.44).

The logarithms of eigenvalues are used to compute their filtering, which explains why the update equation has a geometric expression.

$$\Lambda_{n+1}(x) = \exp(\log \Lambda_n(x) + 2\varepsilon_\Lambda \Delta\Lambda(x)) \tag{5.20}$$

where

$$\Delta\Lambda(x) = \frac{2d}{\text{card}(\mathcal{V})} \sum_{u \in \mathcal{V}} c(||\Delta_u\Lambda(x)||)\Delta_u\Lambda(x) \tag{5.21}$$

with

$$\Delta_u\Lambda(x) = \log(\Lambda_n(x + u)) - \log(\Lambda_n(x)). \tag{5.22}$$

If $\mathbf{U}_q(x)$ is the rotation matrix corresponding to $\mathbf{q}_{n+1}(x)$, the tensor is updated through

$$\mathbf{S}_{n+1}(x) = \mathbf{U}_q(x)\Lambda_{n+1}(x)\mathbf{U}_q(x)^T. \tag{5.23}$$

The same remark as in the Log-Euclidean case can be made: the reconstruction of the tensor does not need to be performed at each step of the processing, but only when the optimization is done. This enables substantial computational savings.

The Log-Euclidean and spectral-quaternion algorithms have been applied to the same noisy image that had been used for the Gaussian filtering, Figure 5.4, and the results are shown in Figure 5.6. The function $c(\cdot)$ used in both cases is $c(t) = \exp(-t^2/\sigma^2)$. For the spectral-quaternion method, different σ are used for the orientation and the eigenvalues ($\sigma_q = 0.1$ and $\sigma_\Lambda = 1$). The used time steps are also different, with $\varepsilon_q = 0.1$ and $\varepsilon_\Lambda = 0.001$ in our example. For a 2D-image, $d = 2$ and $\text{card}(\mathcal{V}) = 4$. In the figure one can see that both methods perform relatively well. Compared to the results of the Gaussian filtering shown in Figure 5.4, the structures are better preserved. Some 'outliers' (very different tensors) are not regularized with either method. Moreover, the results suggest that the spectral-quaternion method produces better orientation results, for highly anisotropic tensors. In the subfigures showing the angular difference between the principal eigenvectors, low errors areas are observable for the spectral-quaternion framework. Those structures correspond to high anisotropy areas. This property is important, since the orientation of highly anisotropic tensors is crucial in many processing algorithms.
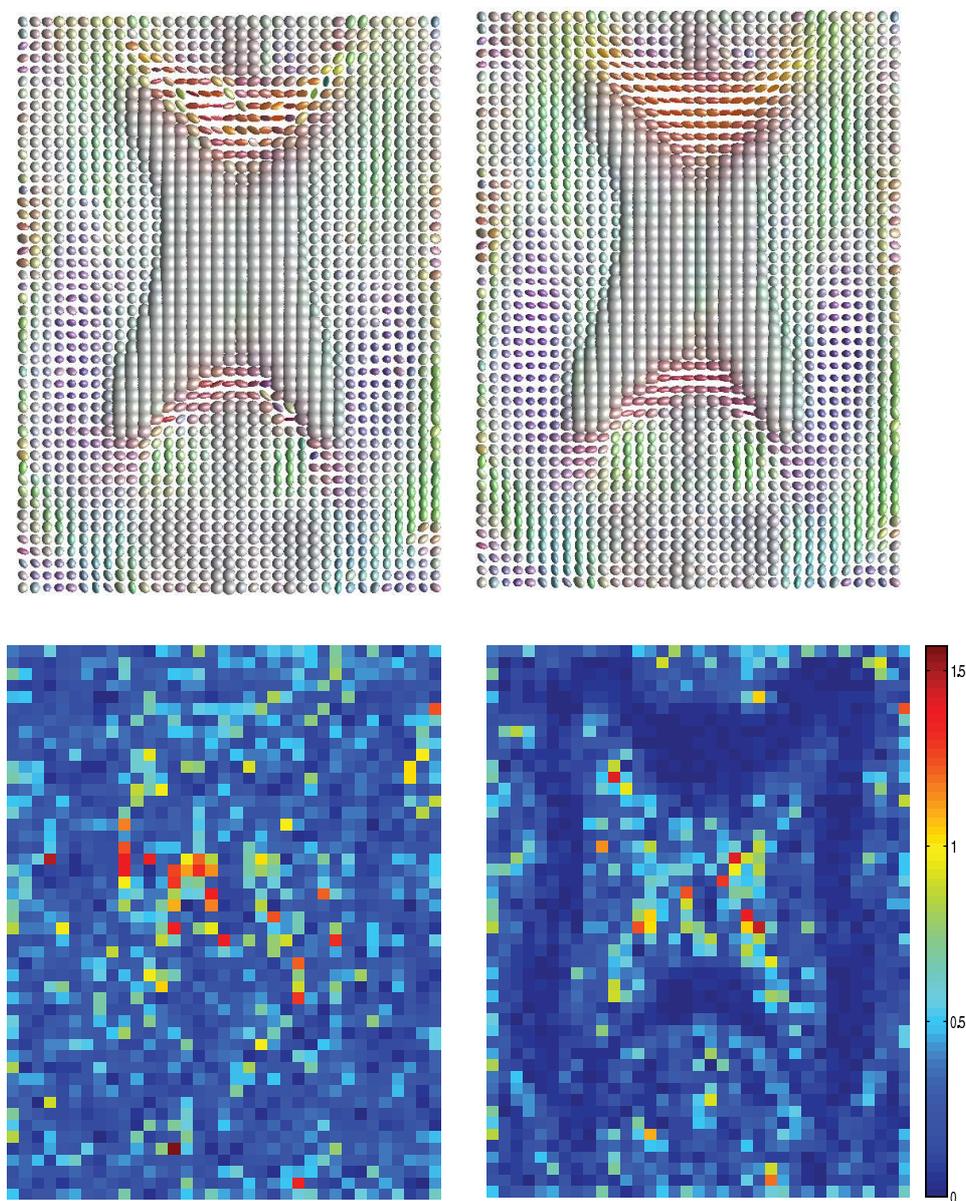
Figure 5.6: Anisotropic filtering of a true image. *Top left*: Log-Euclidean filtering. *Top right*: Filtering based on the spectral-quaternion framework. *Bottom*: angular difference between the first eigenvector of the filtered tensor and the one of the 'true' tensor. It can be observed that some 'very noisy' tensors are not well regularized. Compared to the Gaussian filtering, the structures are well preserved. Moreover, the orientation of highly anisotropic tensors is better regularized with the spectral-quaternion framework, as can be observed from the visible 'low errors areas'.

5.3 Conclusions and perspectives

This chapter has presented some processing algorithms for Diffusion Tensor Images, using both the Log-Euclidean and spectral-quaternion framework. The algorithms studied here have been kept simple, but they proved to be efficient. The ideas developed in this chapter could be applied to other more complex algorithms, since for example transformations of images are crucial in many post-processing procedures.

The framework proposed in this thesis has been compared with the Log-Euclidean, which has been widely used in many works, and for different procedures (see for example (Castro et al., 2007; Fillard et al., 2007; Yeo et al., 2009, 2008; Castro et al., 2007; Van Hecke et al., 2008)).

Regarding the spectral-quaternion framework, this chapter can be seen as a proof-of-concept: the developed framework is easily adapted to many processing algorithms and has been shown to be efficient in applications. Its principal properties should enable a better processing of tensor images while preserving their core characteristics, such as anisotropy. Compared to the Log-Euclidean framework, the computational cost is not prohibitive. This observation opens the way to further developments using our framework, where both the theoretical and implementation aspects will have to be considered. All the spectral-quaternion tools have been implemented in Matlab® and gathered in a single object, which should enable the easy implementation of more complex algorithms.

Statistics on DTI

Statistical analysis of Diffusion Tensor Images is an important topic of current research.
In this chapter, we introduce a novel method for the statistical analysis of matrix-valued images. This method is based upon two motivations: (1) non-parametric methods are more appropriate than parametric ones for the analysis of diffusion tensor images, since the distribution of tensors is seldom known, and (2) those methods should be computationally tractable.
We propose a statistical test based upon the dispersion of data. This notion is related to the similarity measure between elements.
In order to evaluate the relevance of the spectral-quaternion framework of Chapter 4 in that context, the proposed test is used with the spectral-quaternion similarity measure on synthetical data sets. The spectral-quaternion method is shown to be promising and interpretable. In addition, it allows for a tunable tradeoff between sensitivity and robustness to a variation of orientation.

Medical imaging is about the (non-invasive) observation of the human body. The goal is to *see something*, or to *look after something* in the human body. However, how can we assess the importance of what we see ? How can we quantify that an observation is normal, or not ?

Answering this question is the role of statistics. Statistical studies enable to find consistent, significant differences between groups of subjects. They enable to follow, for a single person, the modifications or alterations which appear in the brain. They enable to answer many types of questions, with a certain level of confidence.

This chapter is devoted to the development of adequate methods for the statistical analysis of Diffusion Tensor Imaging. In particular, it focuses on the problem of 'group comparison'. Indeed, if statistical analysis of scalar-valued images is a well established central component of contemporary science, the evolution of sensor technology and data

storage had increasingly produced images that are multimodal and nonlinear in nature. This has motivated significant work in the recent years to extend signal processing techniques from scalar-valued data to manifold-valued data, see e.g. (Jensen et al., 1996). The challenge is methodological as well as computational because clinical studies usually involve large populations and many voxels, that is, large-scale statistical analysis.

Regarding Diffusion Tensor Imaging, the usual approach consists in converting the tensor information into a scalar information (usually fractional anisotropy, see below). In this case though, both the intensity information and the orientation information potentially contain valuable statistical information, calling for new methodological developments.

State of the art for statistical analysis of DTI

There is a vast recent literature on DTI statistical analysis. Among these papers, we can distinguish the one using univariate tests to compare DT images and the ones trying to exploit further information. For example, (Chappell et al., 2008), (Chung et al., 2008), and (Jones et al., 2005) focused on the fractional anisotropy or the mean diffusivity of tensors, without paying any attention to the orientation of these tensors. These scalar values are indeed geometric quantities linked to the shape of tensors, but they can not detect any difference in orientation between tensors. In (Goodlett et al., 2009), the evolution of the fractional anisotropy along a fiber tract is studied. Some other papers have tried to use the whole information contained in the tensor, namely through the use of the Log-Euclidean metric (*i.e.* they use the logarithms of tensors instead of the tensors themselves). This is the case in (Lee et al., 2007), where an Hotelling's T^2 test is developped for the Log-Euclidean metric, a framework similar to the ones in (Fouque et al., 2011) and in (Grigis et al., 2012). Some other papers presented a rigorous conceptual framework based on the Riemannian manifold $S^+(3)$, as in (Pennec, 2006), (Fletcher and Joshi, 2007) and (Lenglet et al., 2006). However, those papers have not addressed the statistical significance of a test for group comparison. This is also the case of (Dryden et al., 2009), which uses another parametrization of tensors. Statistical tests were proposed in (Schwartzman et al., 2010), through a decoupled analysis of the eigenvalues and the eigenvectors of the tensors. The tests are based on distributional assumptions, which is a potential limitation for diffusion tensors. The closest published work to the present paper is (Whitcher et al., 2007), where the authors propose a multivariate dispersion-based permutation test, see Section 6.1.2 for more details.

In Section 6.1, usual mean-based permutation tests will be introduced, before the introduction of dispersion-based permutation tests (Section 6.1.2). Those tests are introduced as a computationally cheap alternative to mean-based tests. The case of multivariate tests will be explained in Section 6.1.3. Subsequently, Section 6.2 focuses on the results of those tests in the case of synthetical data sets.

6.1 Statistical tests for group comparison

Statistical analyses of scalar images are often performed through the use of parametric methods (Friston et al., 2007), such as the Student t^2 test for comparison of Gaussian variables. However, the distribution of multivariate data is rarely known and, if known, is often not Gaussian. This explains why many authors opt for non-parametric methods to

study multivariate images. Among these non parametric methods, permutation tests are often used because of their relative simplicity. Permutation methods provide statistical significance testing of difference between groups without having to assume a distribution of the data (Nichols and Holmes, 2002). These methods have the ability to directly estimate the null distribution of the statistics describing the difference. Moreover, these methods are easily applicable to any statistical test, which is interesting to compare results obtained with different parametrizations of the data.

6.1.1 Mean-based permutation tests

Permutation tests are based on a simple idea. Consider that two populations C and D have to be compared with respect to the variable x. Suppose that Δ_0 is the difference between the means of the two groups. Permutation tests enable to quantify, without any assumptions about the distribution of the variable, if this difference is significant or not. Indeed, if the difference is not significant, it should not be altered by random permutations between C and D. A permutation here consists in assigning a random labeling to an element, no matter in which group it truly belongs. The only rule is to keep the number of elements by group constant. Permutation tests are based on the null hypothesis that the two groups are not different with respect to the variable x. This means that an element of the group C could belong to the group D, and that this permutation should imply no significant difference. In other words, the null hypothesis is that the labeling of elements is arbitrary. Given this hypothesis and the fact that data are exchangeable (*i.e.* they provide from different subjects), the significance of x can be assessed by comparison with the distribution of values under all possible permutations. This is illustrated by the histogram in Figure 6.1, B. In practice, this distribution is approximated: only a certain number of permutations are computed. For each permutation, the novel mean difference is computed and denoted by Δ_k^*, where k is the index of the permutation. The histogram of Figure 6.1 is built from all computed values Δ_k^*, and represents the distribution of the mean-difference under the null hypothesis. If the observed difference Δ_0 is in the tail of the distribution, it means that very few permutations of the data have a larger value by chance, and the true labeling is unlikely arbitrary. The p-value of the test is given by the ratio between the number of times that a permuted statistics Δ_k^* is higher or equal than the observed value Δ_0 and the total number of permutations. The test is statistically significant at a level α if the p-value is smaller than α. This means that this value has less than $100\alpha\%$ of chance to have be found randomly.

The generalization of a permutation test to data that takes values on a manifold is conceptually straightforward because it only requires a proper notion of mean. In the case of DT images, this notion of mean has been widely studied, as shown in preceding chapters.

However, it should be emphasized that a mean-based permutation test may represent a formidable computational task. For instance, computing the Riemannian mean of several positive definite matrices with the affine-invariant metric (3.18) has been shown to be computationally expensive. For a population of size N classified in G groups of size n_g, this computation must be repeated $M = N! / \prod_{g=1}^{G} n_g!$ which is prohibitive for large populations, even if the full distribution is not computed. In the context of DTI, the statistics are typically computed for a large number of voxels, which adds to the computational burden.
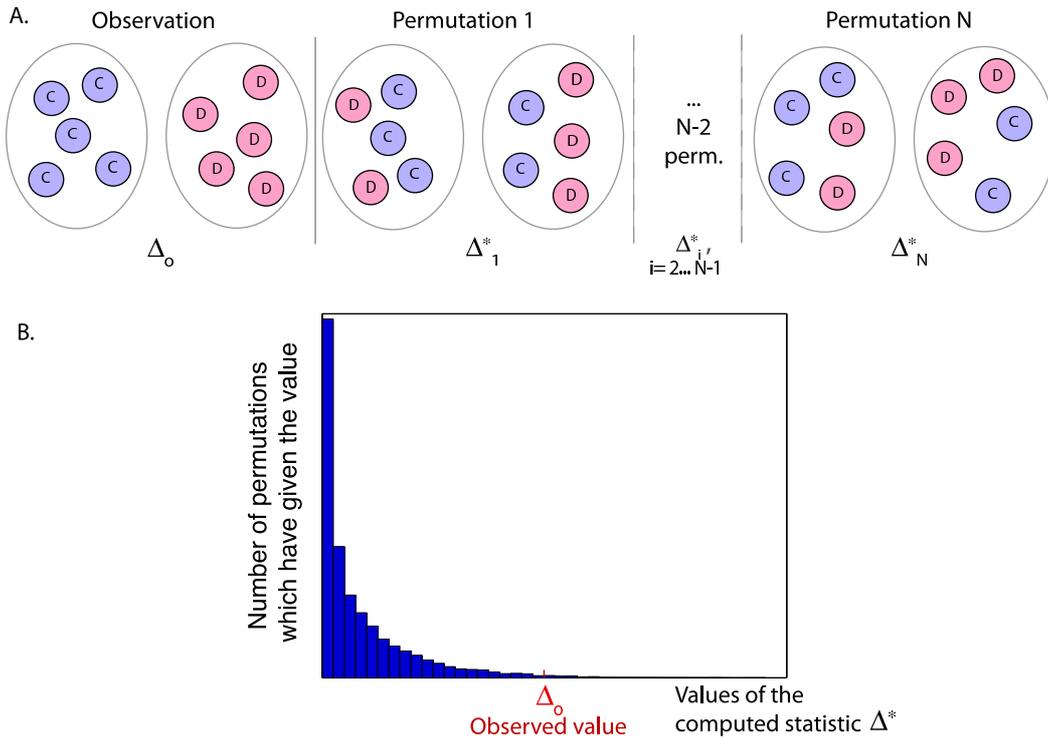
Figure 6.1: General procedure of a permutation test. Given the observed value of the statistic, values corresponding to N permutations of the labels are computed. This gives an approximation of the (unknown) distribution of the statistic. The comparison between the observed value and this distribution enables to compute the p-value of the test.

One computational remedy for positive definite tensors is to compute the arithmetic mean of tensors (that is, to work with the Euclidean metric) or, better, to compute a matrix geometric mean which is known as less computationally demanding, as the Log-Euclidean mean (3.21) or the spectral-quaternion mean (4.20).
Even with such computational simplifications, the computational burden of a mean-based approach remains prohibitive for a real application of group studies using Diffusion Tensor Images because the matrix log operation will scale in a factorial way with the population size.

6.1.2 Dispersion-based permutation tests

The MultiResponse Permutation Procedure (MRPP) proposed in (Mielke and Berry, 2001) is not based on repeated computations of means but only requires to build a similarity matrix between all data points. Given a symmetric similarity measure $s(i, j)$ between data points x_i and x_j, the similarity matrix is defined as the symmetric matrix S with element $S_{ij} = s(i, j)$.

The statistical test proposed in (Mielke and Berry, 2001) is based on a measure of dispersion of the data points within each group rather than on a measure of mean.

The dispersion δ_g of a group g of n_g elements is defined as

$$\delta_g = \frac{2(n_g - 2)!}{n_g!} \sum_{\substack{i<j \\ i,j \in g}} s(i,j)\,, \tag{6.1}$$

where the sum is computed over all data points of the group.

The overall dispersion δ of the variable x in the population is defined as a weighted sum of the dispersions in each group:

$$\delta = \sum_{g=1}^{G} C_g \delta_g\,,$$

where $C_g > 0, g = 1 \ldots G$ are the weights of each group (their sum must be 1).

The rest of the procedure follows the permutation test described in Section 6.1.1, with the mean-difference Δ replaced by the dispersion δ. The observed dispersion is judged statistically significant only if it occurs in the lower tail of the histogram among all possible permutations of the population.

The distance-based permutation test has the same advantages than the mean-based permutation test: it does not require any assumption about the statistical distribution and only requires a similarity measure between data points, for instance a distance on Riemannian manifolds. But a significant computational advantage is that the similarity matrix must be computed only once, requiring $O(N^2)$ computations of pairwise similarity for a total population of size N.

Every distance on the manifold of positive tensors qualifies for a similarity measure. Four distances have been studied in this thesis, and two of them will be used for statistical studies: the computationally fast Log-Euclidean and spectral-quaternion measures.

Another type of dispersion test (but not based on permutations) is the one presented in (Whitcher et al., 2007). In this case, the test statistic is the difference of the sum of all Euclidean interpoint distances between the random variables from the two different groups and one-half of the two corresponding sums of distances of the variables within the same group. If we denote $\mathbf{s}_{1,j}$ the vector representing the j-th tensor of the group 1, and by n_1, n_2 the number of elements in each group, the test statistic is

$$T_{n_1,n_2} = \frac{n_1 n_2}{n_1 + n_2} \left[\frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} ||\mathbf{s}_{1,i} - \mathbf{s}_{2,j}|| - \frac{1}{2n_1^2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} ||\mathbf{s}_{1,i} - \mathbf{s}_{1,j}|| \right.$$
$$\left. - \frac{1}{2n_2^2} \sum_{i=1}^{n_2} \sum_{j=1}^{n_2} ||\mathbf{s}_{2,i} - \mathbf{s}_{2,j}|| \right]\,. \tag{6.2}$$

This test assume that $\mathbf{s}_j$ is identically distributed with a distribution function $\mathcal{G}_j$, and the null hypothesis is that $\mathcal{G}_1 = \mathcal{G}_2$. As explained in (Whitcher et al., 2007), the null hypothesis is rejected for large values of T_{n_1, n_2}. Different methods can be used to obtain critical values for this statistic. Different parametrizations of the tensors can be used, since it is sufficient to replace $\mathbf{s}_j$ by the appropriate vector.

6.1.3   Multivariate testing

Our description of permutation tests has so far assumed an univariate statistical testing but is easily extended to multivariate testing. A widely used method for the statistical comparison of multivariate data consists in the computation of 'marginal' or 'partial' tests (one for each of the n_v considered variable) and to combine the p-values obtained for each partial test by a combining function, which can be of different forms (Pesarin, 2001). This method involves two stages of computation.

- First, the marginal p-values (denoted $\xi_i, i = 1, \ldots, n_v$) are computed through a permutation test (here, we will use a MRPP test for univariate data).

- Then, the combining function is used to compute the combined observed value, $T_o = \mathcal{C}(\xi_1, \ldots, \xi_{n_v})$. The distribution of this T is computed through the combination of the p-values computed for each permutation of the first step, *i.e.* $T_k^* = \mathcal{C}(\xi_{1,k}^*, \ldots, \xi_{n_v,k}^*)$. The combined p-value for the test is then estimated via the ratio between the number of occurrences where T_k^* was larger than T_o and the total number of performed permutations.

Regarding multivariate testing, we propose to compare an 'Euclidean' statistical test based on the six independent quantities of a 3×3 positive definite tensor (as proposed in (Whitcher et al., 2007)) to a 'geometric' test. For a tensor $\mathbf{S}$, the Euclidean parametrization writes

$$\mathbf{v}_E = \begin{pmatrix} S_{xx} \\ S_{yy} \\ S_{zz} \\ \sqrt{2}S_{xy} \\ \sqrt{2}S_{xz} \\ \sqrt{2}S_{yz} \end{pmatrix}, \tag{6.3}$$

while the 'geometric' statistical test is based on the six geometric quantities that define scaling and orientation of the tensor: the three eigenvalues and the first three components of the quaternion (the fourth element of the quaternion is not useful, as it can be computed from the others three and the fact that the quaternion is of unit norm). We use a geometric similarity measure for the (positive) eigenvalues $s(\lambda, \mu) = \sqrt{\log^2\left(\frac{\lambda}{\mu}\right)}$ and an Euclidean (chordal) measure for the quaternions. This parametrization of the tensor is of course related to the spectral-quaternion framework of Chapter 4.

$$\mathbf{v}_g = \begin{pmatrix} \log(\lambda_1) \\ \log(\lambda_2) \\ \log(\lambda_3) \\ \mathbf{q}_1 \\ \mathbf{q}_2 \\ \mathbf{q}_3 \end{pmatrix}. \tag{6.4}$$

A further alternative would be a Log-Euclidean statistical test based on the six elements of the logarithm of the tensor, as investigated in (Whitcher et al., 2007). We do not include this comparison here since the work of (Whitcher et al., 2007) suggests that it does not offer significant advantages compared to the Euclidean test.

For comparison purposes, we will also compute an univariate test using the fractional anisotropy of tensors (Chung et al., 2008).

6.2 Results on synthetic data sets

In this section, dispersion tests with different measures will be compared through their results for some synthetic data sets. The construction of those sets is based upon plausible differences, which may appear in the brain, as explained in Section 6.2.1.

The performance of the different tests in many situations will be assessed through the computation of the *power* of these tests. This quantity is computed by performing the same test a large amount of times and by counting the occurrences of significant results. The power is the ratio between this number of occurrences and the total amount of performed tests. If the test is not efficient, the power will be close to the significance level α. A value of the power equal to one means that the test is particularly efficient for the analyzed situation. The different values used (number of tests, level of significance, ...) are summarized in Table 6.1.

Parameter	Value
Level of significance α	0.05
Number of group G	2
Number of samples by groups n_g	10
Weight of each group C_g	0.5
Number of permutations N_p	20 000
Number of tests by situation N_t	500

Table 6.1: Parameters of the statistical tests.

6.2.1   Construction of the data sets

The power of the proposed statistical tests will be compared on the following synthetic data sets.

We will generate two groups of tensors starting from a reference tensor and a transformed tensor. The parameter γ quantifies the amount of difference in a specific parameter between the two groups. Denoting by $\lambda_i, i = \{1, 2, 3\}$, the eigenvalues of the reference tensor and by λ'_j the ones of the transformed tensor, the following five different geometric transformations of the reference diffusion tensor are studied (Boisgontier et al., 2012):

- *Decrease of longitudinal diffusion (DL)*: λ'_1 will be given by $\lambda'_1 = \lambda_1 - \gamma(\lambda_1 - \lambda_2)$, $\gamma \in [0, 1]$. The two others eigenvalues will be left unchanged.

- *Increase of radial diffusion (IR)*: λ_2 and λ_3 will be replaced by $\lambda'_j = \lambda_j + \gamma \frac{\lambda_1 - \lambda_j}{2}$, $j = \{2, 3\}$, $\gamma \in [0, 1]$, while λ_1 will be left unchanged.

- *Increase of mean diffusion (IM)*: all the eigenvalues of the deformed tensor will be given by $\lambda'_j = (1 + \gamma)\lambda_j$, $j = \{1, 2, 3\}$, $\gamma \in [0, 1]$.

- *Change of diffusion orientation (CO)*: the angle θ between principal directions of the reference tensors will change following $\theta' = \theta + \gamma \frac{\pi}{2}$, $\gamma \in [0, 1]$.

- *Change in both eigenvalues and orientation*: in this case, we will combine a difference in orientation (CO) with one of the first three differences. Both modifications will evolve following the same γ.

The statistical comparison will be tested in a situation of high anisotropy, where the eigenvalues of the reference are $\Lambda = (5, 1, 0.5)$, a situation of small anisotropy where they are given by $\Lambda = (3, 1, 1)$, and a situation of near isotropy, $\Lambda = (1.3, 1, 1)$.

Starting from the reference tensor and its reference deformation, we will generate a population of $N = 20$ tensors by sampling from a Wishart distribution 10 tensors around the reference tensor and 10 tensors around the deformed tensor. The Wishart distribution is a generalization to multiple dimensions of the χ^2 distribution. It is a distribution for symmetric positive semidefinite matrices, typically covariance matrices, the diagonal elements of which are each χ^2 random variables. The Wishart distribution is parameterized with a symmetric, positive semidefinite matrix, Σ, and a positive scalar degrees of freedom parameter, v. v is analogous to the degrees of freedom parameter of a univariate χ^2 distribution, and Σv is the (Euclidean) mean of the distribution. It should be noted that a large number of degree of freedom corresponds to a low level of noise. In the following, $v = 32$ (except when otherwise indicated), which corresponds to a large level of noise (Whitcher et al., 2007).

6.2.2 Univariate tests

Figure 6.2 illustrates an uncommon situation, where the desired result is not a maximal power of the test, but a minimal one. The simulated situation is the one of near isotropic tensors, with a progressive change of orientation. As the tensors are all nearly spherical, the uncertainty in their orientation is high, as explained in (Parker et al., 2003) and in Section 4.3.5. There should be no noticed difference between them, and the tests should not be sensitive to the simulated deformations. However, it can be seen that this is not the case using Euclidean or Log-Euclidean measures. In this case, the spectral-quaternion measure and the test based on Fractional Anisotropy perform better, as they do not detect any difference between the reference and the deformed tensors. It is interesting to note that these curves of power can also be interpreted as curves of sensibility and robustness of the measures. The power of a test is a measure of its sensitivity to the considered deformation. Conversely, flat curves indicate a robustness of the test to a given deformation.
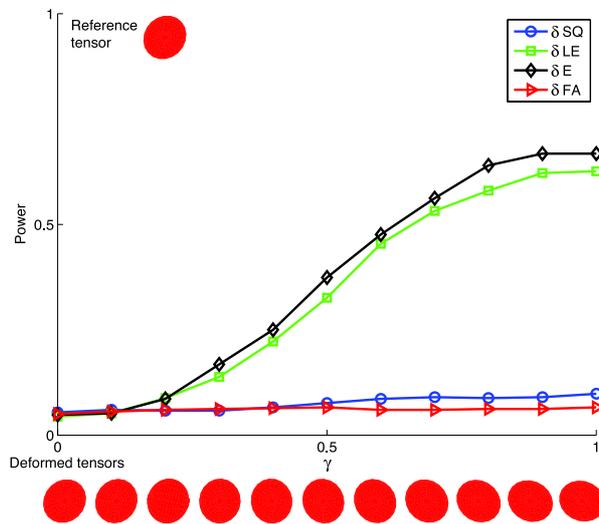

Figure 6.2: Power of the statistical tests in a situation of very low anisotropy (the tensors are nearly isotropic). The tests use the spectral-quaternion measure (δ SQ, blue circles), the Log-Euclidean measure (δ LE, green squares), the Euclidean measure (δ E, black diamonds) and the Fractional Anisotropy of tensors (δ FA, red triangles). A difference in the diffusion orientation (CO) is simulated. In this case, as illustrated on the figure, the tensors are very similar (all almost spheric) and we argue here that no difference should be noted. However, this is not the case for the Log-Euclidean and Euclidean tests.

This analysis of robustness is also applicable to the example of Figure 6.3, showing a difference in orientation in the case of low anisotropy. In this figure, a clear difference can be observed between the Fractional Anisotropy test and the Euclidean and Log-Euclidean tests. As expected, the Fractional Anisotropy test is never significant for this type of difference, showing one more time the limitation of using a unique scalar to represent multivariate data, as the orientation information is totally lost. To the contrary, the Euclidean and Log-Euclidean tests are very sensitive to this difference and thus exhibit strong performance. The spectral-quaternion measure offers an intermediate situation: it is sensitive to the orientation information, but less than the two Euclidean tests. In fact, as explained in Section 4.5, the orientation term of this measure is weighted by a parameter κ, which depends on the anisotropy of tensors. The role of this parameter is to decrease the importance of the orientation term in case of low anisotropy, due to high

uncertainty of this information. This explains why the spectral-quaternion measure is not sensitive to the deformations shown in Figure 6.2. It is important to understand the impact of this parameter on the results. If it is increased, the orientation term will become more important, which will produce an increase of the sensitivity of the measure (*i.e.* an increase of the performance of the test). Depending upon the tradeoff between sensitivity and robustness, the curve of the spectral-quaternion test can be closer to the Fractional Anisotropy one, or to the contrary, closer to the Log-Euclidean test. The fact that this tradeoff can be tuned is of relevance for clinical applications.
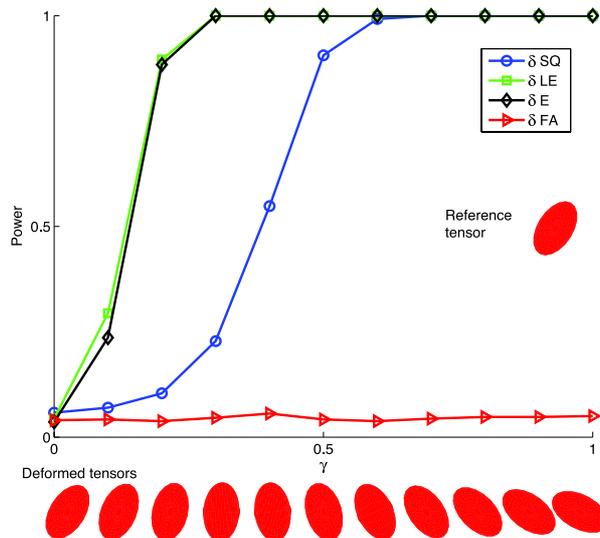

Figure 6.3: Power of the statistical tests in a situation of low anisotropy. A difference in the orientation of the tensors is simulated. The tests use the spectral-quaternion measure (δ SQ, blue circles), the Log-Euclidean measure (δ LE, green squares), the Euclidean measure (δ E, black diamonds) and the Fractional Anisotropy of tensors (δ FA, red triangles). For $\gamma = 0$, which means no difference between the two references of the groups, the power of the tests is about α. As expected, the test based on the Fractional Anisotropy fails to detect the difference in orientation. The Euclidean and Log-Euclidean tests are very sensitive to this kind of deformation, while the spectral-quaternion measure is between those two situations.

Figure 6.4 presents another type of tests, with variation of the eigenvalues. The anisotropy is low in Figure (a) and tensors are highly anisotropic in Figure (b). The simulated transformation is an increase of the mean diffusion (IM). It can be observed that the Fractional Anisotropy is not efficient in those situations, since the shapes of the tensors are not affected. To the contrary, the spectral-quaternion measure is very efficient. Both Euclidean tests have good performances, but are a little less efficient than the spectral framework.

Figure 6.5 shows the results for a transformation combining variation of orientation and decrease of the first eigenvalue. The Euclidean and Log-Euclidean tests are very sensitive in this case. The FA test is better than in Figure 6.2 and Figure 6.3, since the first eigenvalue also varies. Finally, it can be observed that the spectral-quaternion test is better for the high anisotropy case than for the low anisotropy case. This can be explained by the weighting factor of the similarity measure (4.44), which depends upon anisotropies.

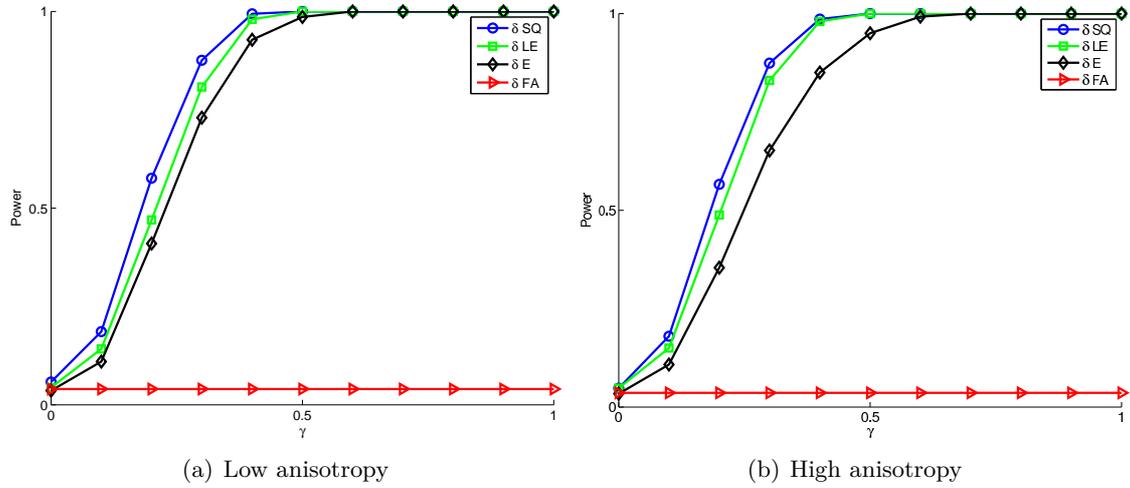(a) Low anisotropy   (b) High anisotropy

Figure 6.4: Power of the statistical tests in a situation of low anisotropy (Figure (a)) and of high anisotropy (Figure (b)). An increase of the mean diffusion is simulated. The tests use the spectral-quaternion measure (δ SQ, blue circles), the Log-Euclidean measure (δ LE, green squares), the Euclidean measure (δ E, black diamonds) and the Fractional Anisotropy of tensors (δ FA, red triangles). The tests based on the Fractional Anisotropy fail to detect this modification, since only the size is varying, and not the shape of the tensors. The spectral-quaternion test is very efficient in both cases, being in all situations slightly better than both Euclidean tests.
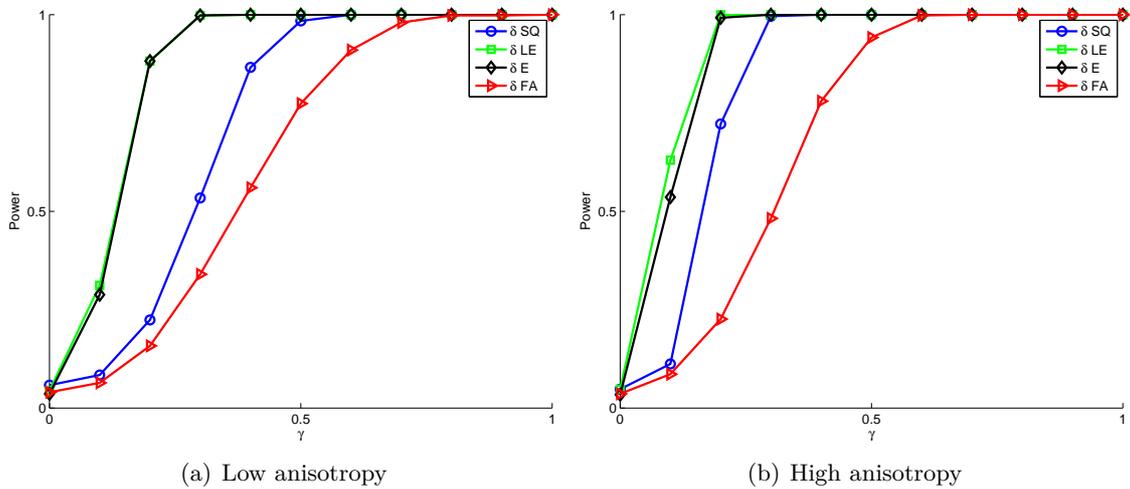

(a) Low anisotropy   (b) High anisotropy

Figure 6.5: Power of the statistical tests in a situation of low anisotropy (Figure (a)) and of high anisotropy (Figure (b)). The simulated transformation is a variation of both the first eigenvalue (DL) and the orientation (CO). The tests use the spectral-quaternion measure (δ SQ, blue circles), the Log-Euclidean measure (δ LE, green squares), the Euclidean measure (δ E, black diamonds) and the Fractional Anisotropy of tensors (δ FA, red triangles). It can be observed that the spectral-quaternion framework is more sensitive when the tensors are highly anisotropic, an observation which is also explained by the weighting factor κ of the spectral-quaternion similarity measure (4.44). Due to the combination of transformations, the FA test is sensitive in this case. The Euclidean and Log-Euclidean tests are very efficient, and seem better in the high anisotropy case.

The simulated transformation in Figure 6.6 is an increase of the radial diffusion. The two tests using a geometric distance (spectral-quaternion and Log-Euclidean tests) are efficient, and better than the two other tests. The Fractional Anisotropy measure seems more sensitive when the anisotropy is high.

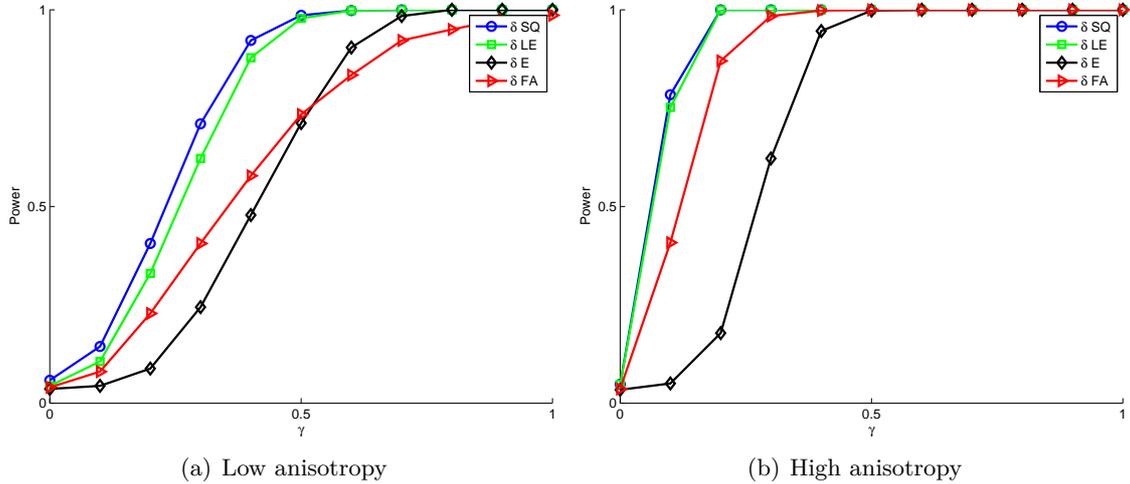(a) Low anisotropy   (b) High anisotropy

Figure 6.6: Power of the statistical tests in a situation of low anisotropy (Figure (a)) and of high anisotropy (Figure (b)). The simulated transformation is an increase of the radial diffusion (IR). The tests use the spectral-quaternion measure (δ SQ, blue circles), the Log-Euclidean measure (δ LE, green squares), the Euclidean measure (δ E, black diamonds) and the Fractional Anisotropy of tensors (δ FA, red triangles). In both cases, the results of the spectral-quaternion and Log-Euclidean tests are very similar, and those two tests are more performant than the FA and Euclidean ones. In particular, the FA test seems more sensitive to this transformation when the anisotropy is high.

The impact of the level of noise on the results is illustrated in Figure 6.7 . The simulated transformation is a decrease of the longitudinal diffusion (DL). The two tests associated to a high level of noise are very similar. In both cases, the four methods are equivalent, with a slight predominance of the Euclidean test. These observations are shared with the results of the same tests with a lower level of noise. However, it can be seen that in this case, the tests are more efficient for a smaller transformation. This seems logical, since for a lower noise level, there should be less 'covering' between the two distributions of samples.

(a) Low anisotropy, high noise

(b) High anisotropy, high noise

(c) Low anisotropy, low noise
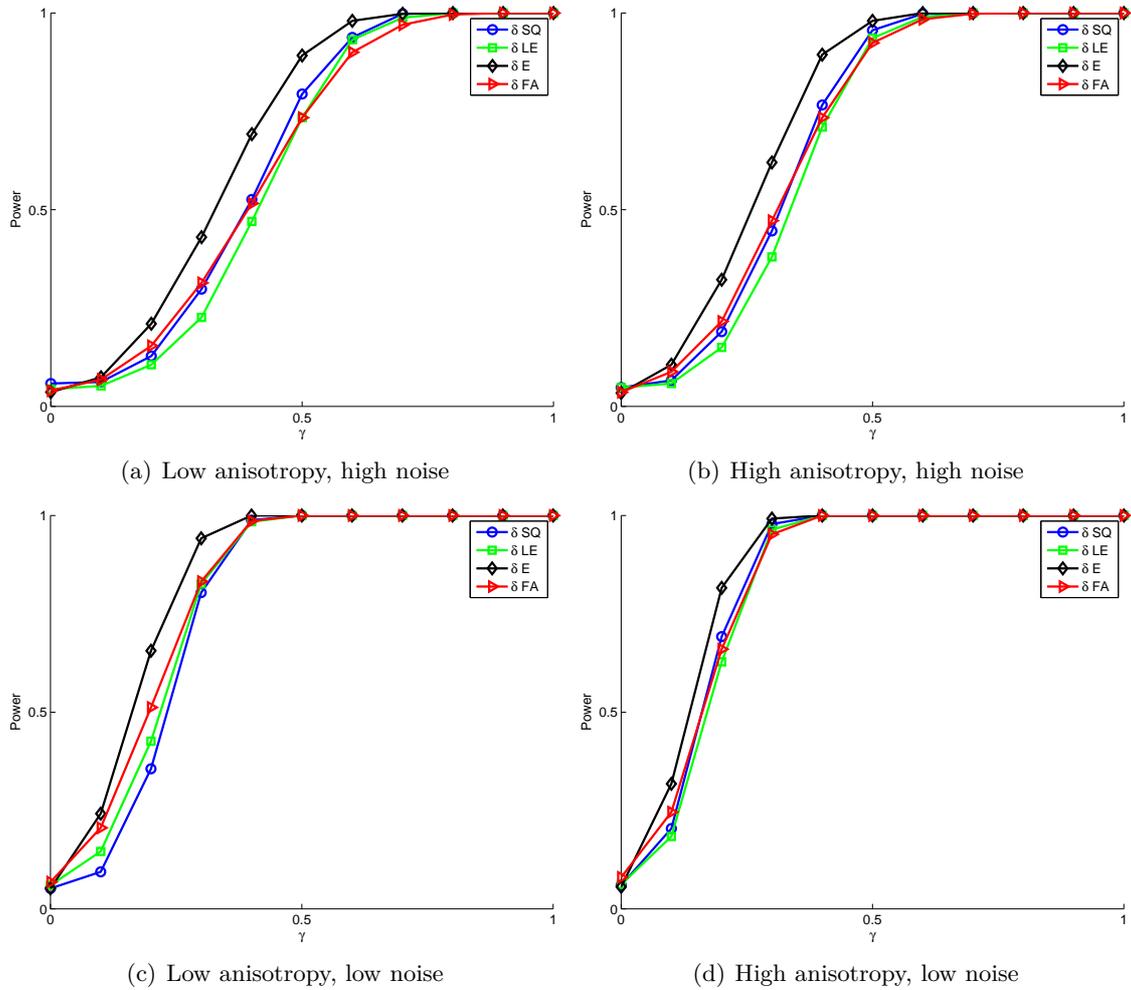
(d) High anisotropy, low noise

Figure 6.7: *Top*: Power of the statistical tests in a situation of low anisotropy (Figure (a)) and of high anisotropy (Figure (b)), with a high level of noise (32 degrees of freedom for the Wishart distribution). *Bottom*: Power of the statistical tests in a situation of low anisotropy (Figure (c)) and of high anisotropy (Figure (d)), with a low level of noise (128 degrees of freedom for the Wishart distribution). The simulated transformation is a decrease of the longitudinal diffusion (DL). The tests use the spectral-quaternion measure (δ SQ, blue circles), the Log-Euclidean measure (δ LE, green squares), the Euclidean measure (δ E, black diamonds) and the Fractional Anisotropy of tensors (δ FA, red triangles). It can be observed that for a low noise level, small transformations are better detected than in a high level of noise.

6.2.3 Multivariate tests

In the following, we will focus on the interpretation of the results of multivariate tests.

Figure 6.8 illustrates the results of each partial tests for multivariate parametrizations of the tensors, for a simulated change of orientation in a case of high anisotropy. The comparison between geometric parametrization (*left*) and algebraic one (*right*) is straightforward. As the geometric parametrization clearly shows that the orientation only has been changed (since only the partial tests related to $\mathbf{q}_2$ and $\mathbf{q}_3$ are efficient), this interpretation can not be drawn from the results of the Euclidean tests. This information could however be of great importance in clinical studies.

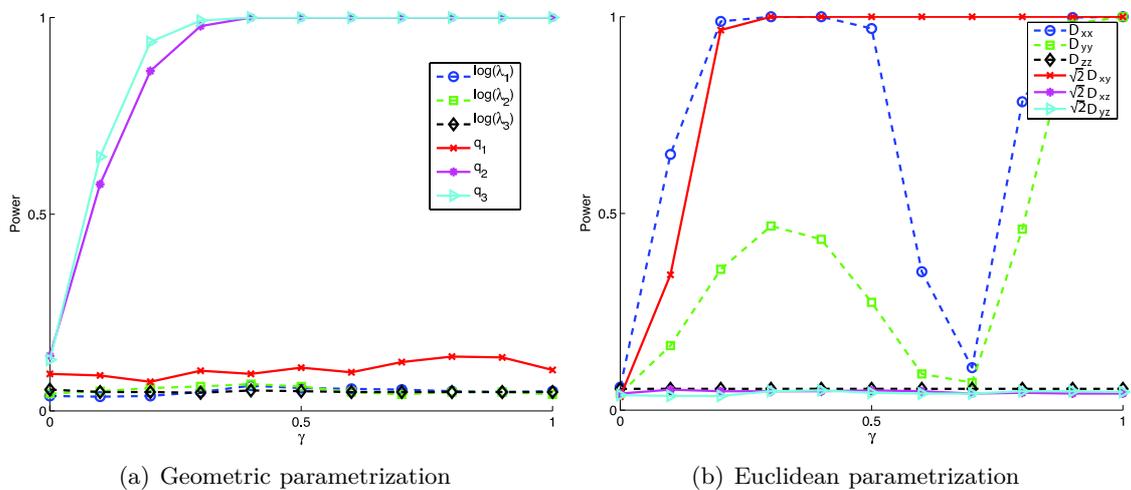(a) Geometric parametrization   (b) Euclidean parametrization

Figure 6.8: Power of each of the partial statistical tests for multivariate parametrizations of the tensors. (a) A geometric parametrization is used. (b) An Euclidean parametrization is used. The simulated situation is a deformation of orientation in a case of high anisotropy. The results of the geometric tests are easily interpretable, as only the partial tests associated to the orientation are performant. To the contrary, the Euclidean tests are poorly understandable.

In a similar way, the decrease of longitudinal diffusion simulated in Figure 6.9 is clearly seen with the geometric parametrization, while this is not the case for the Euclidean one. Indeed, for a geometric parametrization, the partial test of the first eigenvalue is the only one to detect a difference. It should be noted that, from $\gamma = 0.9$, the first eigenvalue is very close to the second one, which increases the uncertainty in orientation (this explains why other partial tests become significant). The easy interpretation of the statistical tests using the geometric parametrization is a desirable feature, which opens the way to many applications.

Combining multivariate tests

Figure 6.10 shows the results of the multivariate parametrizations in a situation similar to the one of Figure 6.4: an increase of the mean diffusion for highly anisotropic tensors. As expected, the tests based on three eigenvalues are particularly efficient. It should be noted that the Euclidean parametrization is also efficient, as the three diagonal elements clearly detect the simulated transformation.

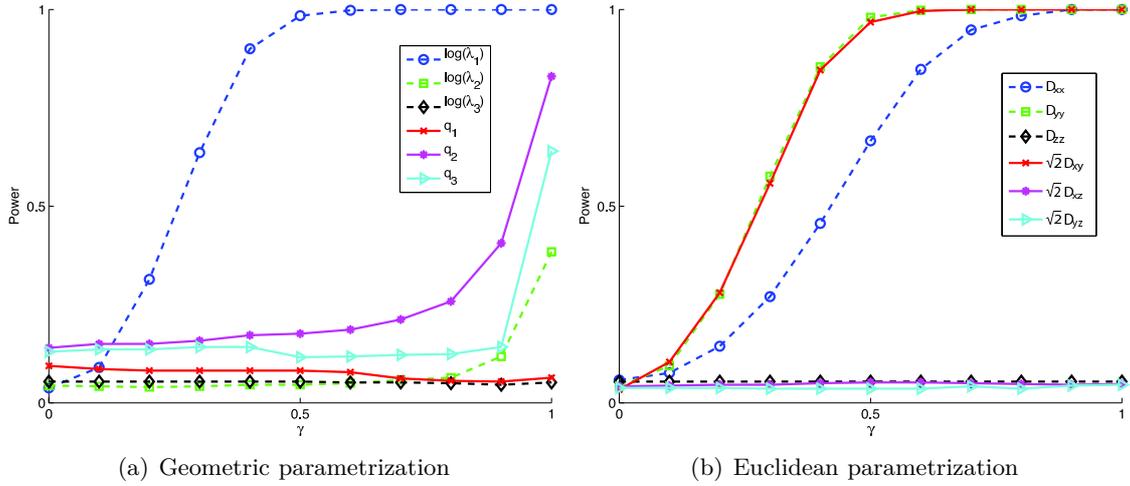(a) Geometric parametrization   (b) Euclidean parametrization

Figure 6.9: Power of each of the partial statistical tests for multivariate parametrizations of the tensors. (a) A geometric parametrization is used. (b) An Euclidean parametrization is used. The simulated situation is a decrease of the longitudinal diffusion in a case of high anisotropy. The results of the geometric tests are easily interpretable, as only the partial test associated to the first eigenvalue is performant. To the contrary, the Euclidean tests are difficult to interpret.

As explained above, the partial tests can be combined, in order to obtain a unique value for each test. For example, Figure 6.11 illustrates the results of the combination of the geometric partial tests (in blue) and of the combination of the algebraic partial tests (in green). The combining function used is the Fisher function

$$T_k^* = \mathcal{C}(\xi_{1,k}^*, \xi_{2,k}^*, \dots \xi_{n_v,k}^*) = -2 \sum_{i=1}^{n_v} \log(\xi_{i,k}^*), \tag{6.5}$$

where ξ_j is the p-value of the test using the j-th variable, and n_v is the number of variables, $n_v = 6$ for both parametrizations.

Both tests are efficient to detect the simulated transformation, the geometric one being slightly better than its Euclidean counterpart. It is interesting to note that, even if some partial tests do not detect any difference, the combination of all the tests give performant results. One more time, this observation is in agreement with the fact that studies based on a single scalar as Fractional Anisotropy could be misleading.

The results presented in Figure 6.11 can be compared with the one of Figure 6.4. This comparison is shown in Figure 6.12, where only the spectral-quaternion and Log-Euclidean univariate tests are shown, for the sake of clarity. The tests based on the spectral-quaternion framework are slightly better. From those results, it can be observed that the multivariate tests do not lead to better results, and that univariate tests based on appropriate measures are sufficient. However, those tests are less interpretable than their multivariate counterparts. The selection of a type of tests should depend upon the application: if the goal is to detect differences, the univariate tests should be used. If the nature of the difference is needed, the geometric multivariate tests could be of great interest.

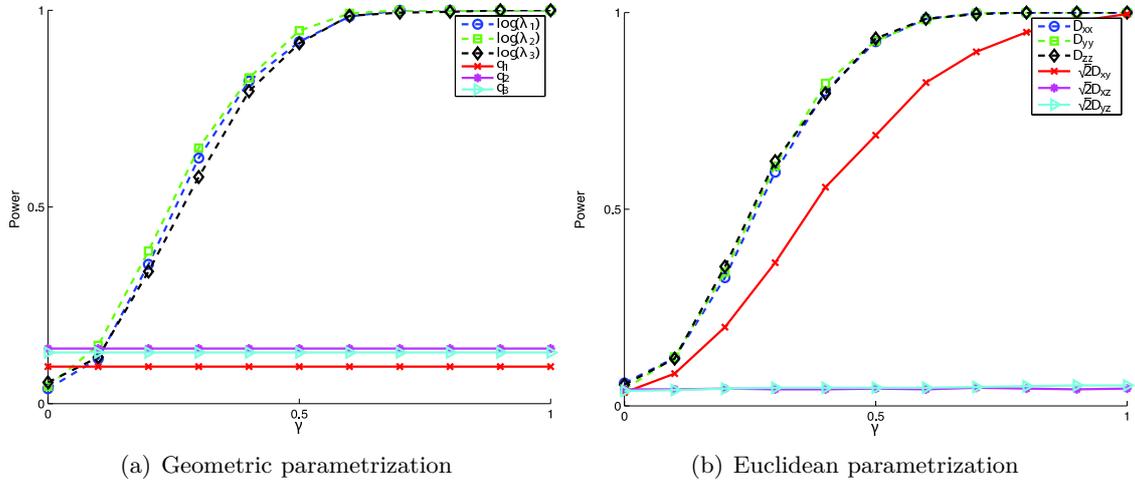(a) Geometric parametrization   (b) Euclidean parametrization

Figure 6.10: Power of each of the partial statistical tests for multivariate parametrizations of the tensors. (a) A geometric parametrization is used. (b) An Euclidean parametrization is used. The simulated situation is an increase of the mean diffusion in a case of high anisotropy. For the geometric parametrization, only the three eigenvalues show a consistent difference. The three diagonal elements of the tensor are significantly affected by the transformation, as well as D_{xy}, the first off-diagonal element.
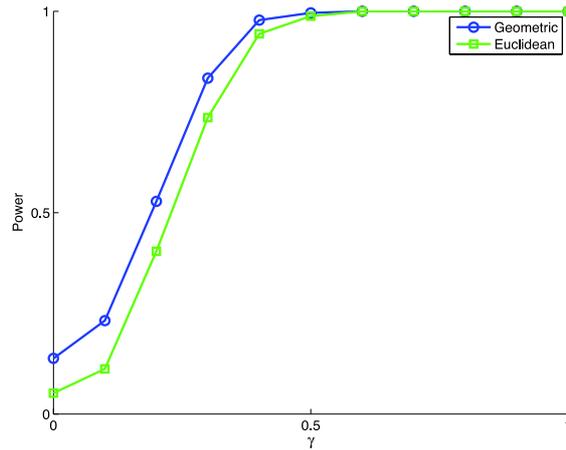

Figure 6.11: Combination of the partial tests presented in Figure 6.10, using the Fisher function (6.5). Both tests are efficient for a sufficiently large simulated difference. The combination of geometric results is slightly better in this case.
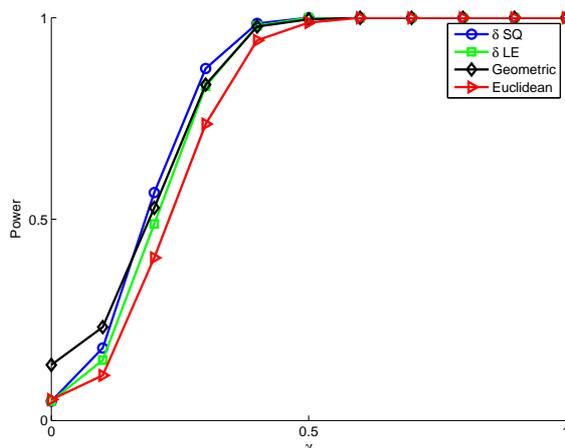
Figure 6.12: Comparison between the combination results of Figure 6.11 and the univariate results of Figure 6.4. It can be observed that all methods are efficient. The tests using a combination of multivariate partial results are very similar to their univariate counterparts.

6.3 Conclusions and perspectives

In this chapter, we have shown that existing methods in the field of group comparison could be advantageously used for the statistical analyses of data lying on Riemannian manifolds. These methods have several advantages as they do not use any assumptions about the distribution of the data (which is seldom known). Moreover, it has been shown that the only specific tool which is needed for this group comparison is an appropriate similarity measure between the data (or a parametrization of them). We have illustrated the computational advantage of basing the permutation test on dispersion rather than means.

Regarding the statistical analysis of Diffusion Tensor Images, we have shown how different measures or different parametrizations of the data can affect the results of the tests. Moreover, two interesting features of the spectral-quaternion measure (and the geometric parametrization associated to this measure) have been highlighted, which both could be relevant for clinical applications:

- The sensitivity of the framework to variation in orientation is tunable, via the parameter κ in the similarity measure (4.44).

- The geometric parametrization (6.4) enables an easier interpretation of the results in terms of what could have physically cause the difference between tensors.

Both for computational and conceptual reasons, dispersion based permutation tests offer an appealing framework for group comparison of manifold-valued data.

Future works should focus on the extension of those 'single voxel' tests to whole brain tests, using real data. Statistical tests can be used for many different applications, as change detection for a single subject (Chung et al., 2008), investigation of different brain pathology as cerebral ischemia, schizophrenia, cognitive impairment, epilepsy, brain tumors,... A review of the potential clinical applications of DTI is found in (Sundgren et al., 2004).

<div align="right">

Chapter 7

</div>

Conclusions and perspectives

The research described in this thesis was motivated by the need to provide a suitable set of algorithmic tools to process and analyze Diffusion Tensor Images. To this end, we have adopted a geometric framework with the threefold aim of addressing the nonlinear nature of the data, preserving their physical information, and allowing for their statistical processing at a computational cost scalable to medical applications.

The first step was to understand the non-linear nature of the images to motivate a geometric framework. The general need of processing data in nonlinear spaces is addressed in an optimization framework on manifolds. Since the majority of processing algorithms are based on the three concepts of interpolating curve, weighted mean and similarity measure, the study of these three mathematical objects in the context of algorithms on manifolds has constituted an important part of this thesis, see Chapter 2. The developments of the thesis have been limited to two cases of interest: the manifold of symmetric positive definite matrices and the manifold of *semi*-definite positive matrices of fixed rank. This latter case has highlighted an important point: the geodesic curve between two elements of a Riemannian manifold is not necessarily the 'best' curve for applications, and the same goes for Kärcher means. If geodesic curve will always enable to follow the shortest path between two elements, it is possible that this path possesses some detrimental properties, or that its computational cost is prohibitive.

Chapter 3 briefly reviews the state of the art for DTI processing in order to understand the physical nature of the measurements. We have seen that scalar indices such as anisotropy are informative, but not sufficient to process diffusion tensor images, and that two important frameworks have been developed for the processing of Diffusion Tensor Images, namely the affine-invariant and Log-Euclidean frameworks. An important message of the chapter is to underline an intrinsic limitation of those frameworks: they do not preserve the anisotropy of tensors during the processing. Specifically, the mean of two anisotropic tensors tends to be isotropic, so that the anisotropy information is degraded through averaging.

The limitation of the existing Riemannian frameworks is a fundamental motivation for the development of a new geometry. The second motivation is the necessity to keep the computational cost as low as possible. Starting from those two pillars, we have developed what we call the *spectral-quaternion* framework.

The term 'spectral' refers to the fact that we have used a decoupling between the intensities and orientations of diffusion. This decoupling enables to preserve the anisotropy of tensors. Moreover, a novel anisotropy index has been proposed, which has an intrinsic meaning and the particularly attractive property to be linearly interpolated through the proposed interpolation method.

The 'quaternion' term refers to a second important feature of the novel method, which is to parametrize orientation of tensors by unit quaternions instead of rotation matrices. This different parametrization is responsible for great computational saving, since it enables Euclidean processing of the orientation.

These developments are the core of Chapter 4. An additional contribution is to deal with the uncertainty of orientations in isotropic tensors. Building upon those theoretical foundations, the chapter proposes novel algorithms for the basic processing operations: interpolation, averaging, and similarity measure.

The application of the basic operations developed in Chapter 4 to some simple processing algorithms is the topic of Chapter 5. We have explained transformations of images, with the important notion of reorientation, and have implemented a simple rigid body algorithm for the registration of two images. The topic of filtering of images has also been covered, in particular the simple Gaussian algorithm and a 'transformed' version of the Perona-Malik anisotropic filtering are explained. These illustrations are viewed as a proof of concept that our framework could be used in more complex algorithms. This could be facilitated by the development of a DTI toolbox, which has been implemented in Matlab$^{\circledR}$.

In the field of Diffusion Tensor Imaging, as well as in the general field of medical imaging, the final goal is not to process images, but to analyze them. Statistical analysis of DTI images is an important topic of current research. We want to compare images, to infer where significant differences are located. In Chapter 6, we have introduced a novel method for the statistical analyses of diffusion tensor images. The introduction of this framework is based on two motivations: non-parametric methods are more appropriate than parametric ones for those images, since the distribution of tensors is seldom known, and those methods should be computationally tractable. The proposed method is based on the dispersion of data, and only uses a similarity measure. The use of the spectral-quaternion framework was shown to be promising, interpretable and allowing for a tunable tradeoff between sensitivity and robustness to a variation of orientation.

To summarize, the thesis offers a novel framework for DTI processing, which combines geometric foundations, physical relevance, and computational efficiency.

A general lesson from the PhD experience is that the modeling of 'special' nonlinear data should always be developed prior to algorithmic developments. Those questions should concern the physical and information properties of the data, and in particular those that motivate the particular measurement technique. At the other end of the spectrum, computational efficiency is a fundamental requirement for the end user.

Future work and perspectives for future research

The thesis has introduced a novel set of tools and preliminary investigations suggest that those tools could be competitive in medical applications. The list of perspectives that this work opens is listed below, together with the planned future work.

- The toolbox containing all the tools presented in this dissertation should be made available for anyone who would want to use them.

- An atlas of DT images should be created, based on the average between a large sample of subjects. The spectral-quaternion mean could be advantageously used.

- The construction of this atlas implies to use appropriate and efficient registration methods. Those complex methods should use the spectral-quaternion framework.

- Once registration methods and atlases will be developed, statistical tests will have to be tested on real images, and a procedure will have to be developed for the analysis of whole brain images.

- Although not addressed in this thesis, tractography algorithms could benefit from the work presented here, since anisotropy is often used as a stopping criterion for those algorithms.

- The framework proposed in Chapter 4 could be extended to $S^+(n)$. An alternative to unit quaternions should be found, as unit quaternions are only representative of orientations in 3D. Data on $S^+(n)$ include zero-mean Gaussian models (represented by their covariance matrix), and anisotropy in this context refers to highly informative directions.

Bibliography

P-A. Absil. *Invariant Subspace Computation: A Geometric Approach.* PhD thesis, University of Liège, 2003.

P-A. Absil, R. Mahony, and R. Sepulchre. Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. *Acta Applicandae Mathematicae*, 80(2):199–220, 2004.

P-A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds.* Princeton University Press, 2008.

D.C. Alexander and J.C. Gee. Elastic matching of Diffusion Tensor Images. *Computer Vision and Image Understanding*, 77(2):233 – 250, 2000.

D.C. Alexander, C. Pierpaoli, P.J. Basser, and J.C. Gee. Spatial transformations of diffusion tensor magnetic resonance images. *IEEE Trans. Med Imaging*, 20(11):1131–1139, 2001.

T. Ando, C-K. Li, and R. Mathias. Geometric means. *Linear Algebra and its Applications*, 385(0):305 – 334, 2004.

V. Arsigny. *Processing data in Lie groups : an algebraic approach. Application to nonlinear registration and Diffusion Tensor MRI.* PhD thesis, Ecole polytechnique, 2006.

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.

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*, 29(1):328–347, 2007.

S. Awate, H. Zhang, and J.C. Gee. Fuzzy Nonparametric DTI Segmentation for Robust Cingulum-Tract Extraction. In *MICCAI*, pages 294–301, 2007.

P.J. Basser and C. Pierpaoli. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *Journal of Magnetic Resonance-Series B*, 111(3):209–219, 1996.

P.J. Basser, J. Mattiello, and D. LeBihan. Diagonal and off-diagonal components of the self-diffusion tensor: their relation to and estimation from the NMR spin-echo signal. In *11th annual meeting of the SMRM, Berlin*, page 1222, 1992.

P.J. Basser, J. Mattiello, and D. Le Bihan. MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66(1):259 – 267, 1994a.

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-Series B*, 103(3):247–254, 1994b.

P.G. Batchelor, M. Moakher, D. Atkinson, F. Calamante, and A. Connelly. A rigorous framework for diffusion tensor calculus. *Magnetic Resonance in Medicine*, 53(1):221–225, 2005.

R. Bhatia. *Positive Definite Matrices*. Princeton Series in Applied Mathematics. Princeton University Press, 2009.

G. Birkhoff. Extensions of Jentzsch's theorem. *Transactions of the American Mathematical Society*, 85(1):219–227, 1957.

F. Bloch, W.W. Hansen, and M. Packard. Nuclear induction. *Physical review*, 70(7-8): 460–474, 1946.

H. Boisgontier, V. Noblet, F. Heitz, L. Rumbach, and J-P. Armspach. Generalized likelihood ratio tests for change detection in diffusion tensor images: Application to multiple sclerosis. *Medical Image Analysis*, 16(1):325–338, 2012.

S. Bonnabel and R. Sepulchre. Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank. *SIAM Journal on Matrix Analysis and Applications*, 31(3):1055–1070, 2009.

S. Bonnabel, A. Collard, and R. Sepulchre. Rank-preserving geometric means of positive semi-definite matrices. *Linear Algebra and its Applications*, 438(8):3202–3216, 2013.

R.W. Brockett. System theory on group manifolds and coset spaces. *SIAM J. Control*, 10 (2):265–284, 1972.

R.W. Brockett. Differential geometry and the design of gradient algorithms. In Amer. Math. Soc., editor, *Proceedings of Symposia in Pure Mathematics*, volume 54, pages 69–92, 1993.

J.S. Campbell. *Diffusion Imaging of White Matter Fiber Tracts*. PhD thesis, McGill University, 2004.

Y. Cao, M.I. Miller, S. Mori, R.L. Winslow, and L. Younes. Diffeomorphic matching of diffusion tensor images. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*, pages 67–67. IEEE, 2006.

H.Y. Carr and E.M. Purcell. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical Review*, 94(3):630, 1954.

C.A. Castaño-Moraga, C. Lenglet, R. Deriche, and J. Ruiz-Alzola. A fast and rigorous anisotropic smoothing method for DT-MRI. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 93–96, 2006.

F.J.S. Castro, O. Clatz, J. Dauguet, N. Archip, J-P. Thiran, and S. Warfield. Evaluation of brain image nonrigid registration algorithms based on Log-Euclidean MR-DTI consistency measures. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 45–48, 2007.

M.H. Chappell, J.A. Brown, J.C. Dalrymple-Alford, A.M. Uluğ, and R. Watts. Multivariate analysis of diffusion tensor imaging data improves the detection of microstructural damage in young professional boxers. *Magn Reson Imaging*, 26(10):1398–405, 2008.

C. Chefd'hotel, D. Tschumperlé, R. Deriche, and O.D. Faugeras. Constrained flows of matrix-valued functions: Application to Diffusion Tensor regularization. In *ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part I*, pages 251–265, London, UK, 2002. Springer-Verlag.

B. Chen and E.W. Hsu. Noise Removal in Magnetic Resonance Diffusion Tensor Imaging. *Magnetic Resonance in Medicine*, 54:393–407, 2005.

M-C. Chiang, A.D. Leow, A.D. Klunder, R.A. Dutton, M. Barysheva, S. Rose, K.L. McMahon, G.I. de Zubicaray, A.W. Toga, and P.M. Thompson. Fluid registration of Diffusion Tensor Images using information theory. *IEEE Transactions on Medical Imaging*, 27, 2008.

S. Chung, D. Pelletier, M. Sdika, Y. Lu, J.I. Berman, and R.G. Henry. Whole brain voxel-wise analysis of single-subject serial DTI by permutation testing. *Neuroimage*, 39 (4):1693–705, 2008.

O. Coulon, D.C. Alexander, and S. Arridge. A regularization scheme for Diffusion Tensor Magnetic Resonance Images. In Michael Insana and Richard Leahy, editors, *Information Processing in Medical Imaging*, volume 2082 of *Lecture Notes in Computer Science*, pages 92–105. Springer Berlin / Heidelberg, 2001.

O. Coulon, D.C. Alexander, and S. Arridge. Diffusion Tensor Magnetic Resonance Image Regularization. *Medical Image Analysis*, 8:47–67, 2004.

Y. Dai, J. Trumpf, H. Li, N. Barnes, and R. Hartley. Rotation Averaging with Application to Camera-Rig Calibration. In *Computer Vision – ACCV 2009*, pages 335–346, 2010.

P. Douek, R. Turner, J. Pekar, N. Patronas, and D. Le Bihan. MR color mapping of myelin fiber orientation. *Journal of computer assisted tomography*, 15(6):923–929, 1991.

I.L. Dryden, A. Koloydenko, and D. Zhou. Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *The Annals of Applied Statistics*, 3 (3):1102–1123, 2009.

A. Einstein. *Investigations on the Theory of the Brownian Movement.* Courier Dover Publications, 1956.

A. Fick. V. On liquid diffusion. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 10(63):30–39, 1855.

P. Fillard, X. Pennec, V. Arsigny, and N. Ayache. Clinical DT-MRI estimation, smoothing and fiber tracking with Log-Euclidean metrics. *IEEE Transactions on Medical Imaging.*, 26, 2007.

P.T. Fletcher and S. Joshi. Riemannian Geometry for the Statistical Analysis of Diffusion Tensor Data. *Signal Processing*, 87:250–262, 2007.

P.T. Fletcher, S. Venkatasubramanian, and S. Joshi. The geometric median on Riemannian manifolds with application to robust atlas estimation. *NeuroImage*, 45(1, Supplement 1):S143 – S152, 2009.

A-L. Fouque, P. Fillard, A. Bargiacchi, A. Cachia, M. Zilbovicius, B. Thyreau, E. Le Floch, P. Ciuciu, and E. Duchesnay. Voxelwise multivariate statistics and brain-wide machine learning using the full diffusion tensor. *Med Image Comput Comput Assist Interv*, 14 (Pt 2):9–16, 2011.

K.J. Friston, J.T. Ashburner, S.J. Kiebel, T.E. Nichols, and W.D. Penny. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press, 2007.

A. Goh and R. Vidal. Algebraic methods for direct and feature based registration of diffusion tensor images. In *Computer Vision–ECCV 2006*, pages 514–525. Springer, 2006.

C.B. Goodlett, P.T. Fletcher, J.H. Gilmore, and G. Gerig. Group analysis of DTI fiber tract statistics with application to neurodevelopment. *NeuroImage*, 45(1, Supplement 1):S133 – S142, 2009.

A. Grigis, V. Noblet, F. Heitz, F. Blanc, J. de Sèze, S. Kremer, L. Rumbach, and J-P. Armspach. Longitudinal change detection in diffusion MRI using multivariate statistical testing on tensors. *Neuroimage*, 60(4):2206–21, 2012.

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*, pages 761–764. IEEE, 2002.

Y. Gur and N.A. Sochen. Fast invariant Riemannian DT-MRI regularization. In *International Conference on Computer Vision*, pages 1–7, 2007.

E.L. Hahn. Spin echoes. *Physical Review*, 80(4):580, 1950.

U. Helmke and J. Moore. *Optimization and Dynamical Systems*. Springer, 1996.

M. Ingalhalikar, J. Yang, C. Davatzikos, and R. Verma. DTI-DROID: Diffusion tensor imaging-deformable registration using orientation and intensity descriptors. *International Journal of Imaging Systems and Technology*, 20(2):99–107, 2010.

J.R. Jensen et al. *Introductory digital image processing: a remote sensing perspective*. Prentice-Hall Inc., 1996.

D.K. Jones, L.D. Griffin, D.C. Alexander, M. Catani, M.A. Horsfield, R. Howard, and S.CR. Williams. Spatial normalization and averaging of diffusion tensor MRI data sets. *Neuroimage*, 17(2):592–617, 2002.

D.K. Jones, M.R. Symms, M. Cercignani, and R.J. Howard. The effect of filter size on VBM analyses of DT-MRI data. *Neuroimage*, 26(2):546–54, 2005.

M. Journée. *Geometric Algorithms for Component Analysis with a view to gene-expression analysis*. PhD thesis, University of Liège, 2009.

H. Kärcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30(5):509–541, 1977.

G. Kindlmann, R. San Jose Estepar, M. Niethammer, S. Haker, and C-F. Westin. Geodesic-loxodromes for diffusion tensor interpolation and difference measurement. In *MICCAI*, pages 1–9, 2007.

P.C. Lauterbur. Image formation by induced local interactions: examples employing nuclear magnetic resonance. *Nature*, 242(5394):190–191, 1973.

D. Le Bihan. Looking into the functional architecture of the brain with diffusion MRI. *Nature Reviews Neuroscience*, 4(6):469–480, 2003.

D. Le Bihan and E. Breton. Imagerie de diffusion in-vivo par résonance magnétique nucléaire. *Comptes-Rendus de l'Académie des Sciences*, 93(5):27–34, 1985.

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.

A.D. Lee, N. Lepore, F. Lepore, F. Alary, P. Voss, Y. Chou, C. Brun, M. Barysheva, A.W. Toga, and P.M. Thompson. Brain differences visualized in the blind using tensor manifold statistics and diffusion tensor imaging. In *Frontiers in the Convergence of Bioscience and Information Technologies, 2007. FBIT 2007*, pages 470–476. IEEE, 2007.

A. Leemans, J. Sijbers, S. De Backer, E. Vandervliet, and P.M. Parizel. Affine coregistration of diffusion tensor magnetic resonance images using mutual information. In *Advanced Concepts for Intelligent Vision Systems*, pages 523–530. Springer, 2005.

C. Lenglet. *Geometric and variational methods for diffusion tensor MRI processing*. PhD thesis, Université de Nice-Sophia Antipolis, 2006.

C. Lenglet, M. Rousson, R. Deriche, and O. Faugeras. Statistics on the manifold of multivariate normal distributions: theory and application to diffusion tensor MRI processing. *Journal of Mathematical Imaging and Vision*, 25:423–444, 2006.

C. Lenglet, J.S.W. Campbell, M. Descoteaux, G. Haro, P. Savadjiev, D. Wassermann, A. Anwander, R. Deriche, G.B. Pike, G. Sapiro, K. Siddiqi, and P.M. Thompson. Mathematical methods for diffusion MRI processing. *NeuroImage*, 45(1, Supplement 1):S111 – S122, 2009.

N. Lepore, C. Brun, M-C. Chiang, Y-Y. Chou, R. Dutton, K. Hayashi, O. Lopez, H. Aizenstein, A. Toga, J. Becker, and P. Thompson. Multivariate Statistics of the Jacobian Matrices in Tensor Based Morphometry and Their Application to HIV/AIDS. In *MICCAI*, pages 191–198, 2006.

D.G. Luenberger. The gradient projection method along geodesics. *Management Science*, 18(11):620–631, 1972.

Y. Ma, J. Kosecka, and S. Sastry. Optimization criteria and geometric algorithms for motion and structure estimation. *International Journal of Computer Vision*, 44(3): 219–249, 2001.

P. Mansfield. Multi-planar image formation using NMR spin echoes. *Journal of Physics C: Solid State Physics*, 10(3):L55, 1977.

P.W. Mielke and K.J. Berry. *Permutation methods: a distance function approach.* Springer Verlag, 2001.

M. Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 26(3): 735–747, 2005.

M. Moakher and M. Zéraï. The Riemannian Geometry of the Space of Positive-Definite Matrices and Its Application to the Regularization of Positive-Definite Matrix-Valued Data. *Journal of Mathematical Imaging and Vision*, 40:171–187, 2011.

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.

E. Muñoz-Moreno, R. Cárdenes-Almeida, and M. Martin-Fernandez. Review of techniques for registration of diffusion tensor imaging. In S. Aja-Fernández, R. Luis García, D. Tao, and X. Li, editors, *Tensors in Image Processing and Computer Vision*, Advances in Pattern Recognition, pages 273–297. Springer London, 2009.

T.E. Nichols and A.P. Holmes. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping*, 15(1):1–25, 2002.

H-J. Park, M. Kubicki, M.E. Shenton, A. Guimond, R.W. McCarley, S. Maier, R. Kikinis, F.A. Jolesz, and C-F. Westin. Spatial normalization of diffusion tensor MRI using multiple channels. *Neuroimage*, 20(4):1995, 2003.

G.J.M Parker, H.A. Haroon, and C.A.M. Wheeler-Kingshott. A framework for a streamline-based probabilistic index of connectivity (PICo) using a structural interpretation of MRI diffusion measurments. *Journal of Magnetic Resonance Imaging*, 18: 245–254, 2003.

X. Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006.

X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, 2006.

P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(7):629–639, 1990.

F. Pesarin. *Multivariate Permutation Tests : With Applications in Biostatistics.* Wiley, 2001.

D. Petz and R. Temesi. Means of positive numbers and matrices. *SIAM J. Matrix Anal. Appl.*, 27(3):712–720, 2005.

C. Pierpaoli and P.J. Basser. Toward a quantitative assessment of diffusion anisotropy. *Magnetic Resonance in Medicine*, 36:893–906, 1996.

C. Pierpaoli, J. Mattiello, D. Le Bihan, G. Di Chiro, and P.J. Basser. Diffusion tensor imaging of brain white matter anisotropy. In *Proc. SMR, 2nd Meeting, San Francisco*, 1994.

M. Pollari, T. Neuvonen, and J. Lötjönen. Affine registration of Diffusion Tensor MR images. In *MICCAI*, pages 629–636, 2006.

E.M. Purcell, H.C. Torrey, and R.V. Pound. Resonance absorption by nuclear magnetic moments in a solid. *Physical review*, 69(1-2):37, 1946.

A. Sarlette. *Geometry and Symmetries in Coordination Control*. PhD thesis, University of Liège, Belgium, 2009.

A. Sarlette and R. Sepulchre. Consensus optimization on manifolds. *SIAM Journal on Control and Optimization*, 48(1):56–76, 2009.

A. Sarlette, R. Sepulchre, and N. Leonard. Cooperative attitude synchronization in satellite swarms: a consensus approach. *Proceedings of the 17th IFAC Symposium on Automatic Control in Aerospace*, 2007.

A. Schwartzman, R.F. Dougherty, and J.E. Taylor. Group comparison of eigenvalues and eigenvectors of diffusion tensors. *Journal of the American Statistical Association*, 105 (490):588–599, 2010.

S.M. Smith, M. Jenkinson, H. Johansen-Berg, D. Rueckert, T. E. Nichols, C.E. Mackay, K.E. Watkins, O. Ciccarelli, M.Z. Cader, P.M. Matthews, et al. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage*, 31(4):1487–1505, 2006.

S.T. Smith. Covariance, subspace, and intrinsic Cramér-Rao bounds. *IEEE Transactions on Signal Processing*, 53:1610–1630, 2005.

S. Sra. Positive definite matrices and the symmetric Stein divergence. *arXiv preprint arXiv:1110.1773*, 2011.

E.O. Stejskal and J.E. Tanner. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *The journal of chemical physics*, 42(1):288, 1965.

P.C. Sundgren, Q. Dong, D. Gómez-Hassan, S.K. Mukherji, P. Maly, and R. Welsh. Diffusion tensor imaging of the brain: review of clinical applications. *Neuroradiology*, 46 (5):339–50, 2004.

P. Thévenaz, T. Blu, and M. Unser. Interpolation revisited [medical images application]. *Medical Imaging, IEEE Transactions on*, 19(7):739–758, 2000.

D. Tschumperlé and R. Deriche. Diffusion tensor regularization with constraints preservation. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:948, 2001.

W. Van Hecke, J. Sijbers, E. D'Agostino, F. Maes, S. De Backer, E. Vandervliet, P.M. Parizel, and A. Leemans. On the construction of an inter-subject diffusion tensor magnetic resonance atlas of the healthy human brain. *Neuroimage*, 43(1):69–80, 2008.

Y.T. Weldeselassie and G. Hamarneh. DT-MRI segmentation using graph cuts. In *SPIE*, volume 6512, 2007.

Y.T. Weldeselassie, G. Hamarneh, F. Beg, and S. Atkins. Novel decomposition of tensor distance into shape and orientation distances. In *MICCAI*, pages 173–180, 2009.

G.E. Wesbey, M.E. Moseley, and R.L. Ehman. Translational molecular self-diffusion in magnetic resonance imaging: I. Effects on observed spin-spin relaxation. *Investigative radiology*, 19(6):484–490, 1984a.

G.E. Wesbey, M.E. Moseley, and R.L. Ehman. Translational molecular self-diffusion in magnetic resonance imaging: II. Measurement of the self-diffusion coefficient. *Investigative radiology*, 19(6):491–498, 1984b.

B. Whitcher, J.J. Wisco, N. Hadjikhani, and D.S. Tuch. Statistical group comparison of diffusion tensors via multivariate hypothesis testing. *Magn Reson Med*, 57(6):1065–74, 2007.

B.T.T. Yeo, T. Vercautern, P. Fillard, X. Pennec, P. Gotland, N. Ayache, and O. Clatz. DTI registration with exact finite-strain differential. In *IEEE Internation Symposium on Biomedical Imaging: From Nano to Macro*, 2008.

B.T.T. Yeo, T. Vercauteren, P. Fillard, J-M. Peyrat, X. Pennec, P. Golland, N. Ayache, and O. Clatz. DT-REFinD: Diffusion Tensor registration with exact finite-strain differential. *IEEE Transactions on Medical Imaging*, 28, 2009.

H. Zhang, P.A. Yushkevich, D.C. Alexander, and J.C. Gee. Deformable registration of diffusion tensor MR images with explicit orientation optimization. *Medical Image Analysis*, 10(5):764 – 785, 2006.

D. Zhou. *Statistical Analysis of Diffusion Tensor Imaging.* PhD thesis, University of Nottingham, January 2010.