Statistical models to learn the structural organisation of neural circuits from multimodal brain images - Application to Gilles de la Tourette syndrome

Pietro Gori

University Pierre et Marie Curie - UPMC

Doctoral School ED3C
The Brain-Cognition-Behaviour Doctoral School

P H D T H E S I S

Specialty : Computer Science

Presented by: Pietro Gori

Statistical models to learn the structural organisation of neural circuits from multimodal brain images

Application to Gilles de la Tourette syndrome

Thesis Advisors:
Stanley Durrleman, Olivier Colliot, Nicholas Ayache

prepared at ICM, Aramis LAB, Paris

defended on the 8^{th} of January 2016

Jury

Advisors:	Stanley Durrleman	-	Inria (Aramis LAB)
	Olivier Colliot	-	CNRS (Aramis LAB)
	Nicholas Ayache	-	Inria (Asclepios)
Reviewers:	Alain Trouvé	-	ENS Cachan (CMLA)
	Daniel Rueckert	-	Imperial College London (BioMedIA)
Examiners:	Benoit Perthame	-	UPMC (Lab Jacques-Louis Lions)
	Jean-François Mangin	-	CEA (NeuroSpin)
	Rachid Deriche	-	Inria (Athena)
Invited:	Andreas Hartmann	-	AP-HP (ICM)
	Yulia Worbe	-	AP-HP (ICM)

Abstract

In this Thesis, we propose a statistical framework to analyse *morphological* and *organisational* anomalies altering the anatomy of the neural circuits of the brain in neurodevelopmental disorders. We evaluate it on a disease model about Gilles de la Tourette syndrome (GTS). Every circuit is composed of white matter projections and grey matter structures which are all virtually represented as 3D meshes. All components of the neural circuits are then integrated into a single complex. This makes possible to study their organisation, namely their relative position, and in particular the structural connectivity (i.e. the areas of the grey matter structures integrated by white matter fibers). Moreover, the use of meshes facilitates the visualisation of the circuits and the interpretation of their pathological alterations.

The proposed methodology is based on a *generative* model. Given a population, the neural circuits of each subject are modelled as a deformed template complex plus a residuals noise. The template complex captures the common morphological characteristics within the population and it can be thought as an average. The deformations, usually defined as diffeomorphisms of the entire ambient space, model instead the morphological variability of the population. The joint estimate of the template complex and deformations is called *atlas construction*.

In the first part of this Thesis, we propose a Bayesian framework to embed the aforementioned generative model. It is general and it can be applied to any parametric deformation framework and to all shape models with which is possible to define probability density functions. Using this framework, we can automatically estimate important balancing parameters which were fixed by the user in previous methods not based on a statistical setting, namely the trade-off between data-terms and deformation regularity. Moreover, it is also possible to estimate from the data a well-conditioned covariance matrix of the deformation parameters which can be directly employed in statistical analysis such as Principal Component Analysis. Furthermore, we propose to model both curve and surface meshes as Gaussian random varifolds for which we define finite-dimensional approximation spaces where it is possible to define probability density functions. This computational model does not need point-correspondences and it has a closed-form metric easily derivable.

In the second part, we define a computational model for white matter fiber bundles called *weighted currents*. Similarly to currents, it does not need point correspondences or fiber correspondences and it augments its definition taking into consideration not only the pathway of the fibers but also the locations of their extremities. This makes thus possible to correctly register also the extremities of the bundles in the template-to-subjects deformations and not only the most dense parts of the bundles as in currents. This is fundamental to retrieve the variations in structural connectivity. Moreover, we also propose an approximation scheme for fiber bundles based on the framework of weighted currents. It results in a parsimonious representation which preserves both the shape and the structural connectivity of the original bundles. It facilitates the visualisation and interpretation and it makes it computationally possible to consider at the same time multiple fiber

bundles and grey matter structures in an atlas construction.

In the last part, we describe a new deformation scheme based on a cascade of two diffeomorphisms. It allows us to locate variations in structural connectivity and at the same time to capture global anatomical changes. This was not possible with previous single-diffeomorphic deformations. The proposed deformation setting is integrated into the Bayesian atlas construction previously presented. We show its effectiveness by comparing the cortico-putamen circuits of a group of GTS patients with the ones of a group of controls. Preliminary results highlight differences about both the shape of the grey matter structures and the structural connectivity. Moreover, we also show that the proposed approach leads to better classification scores than a single diffeomorphic method. This suggests that it might better characterise the anatomical alterations associated to GTS and therefore also improve our understanding of the pathophysiological mechanisms underlying this syndrome.

Acknowledgments

First of all, I would like to warmly thank my three supervisors for guiding me during the last three years.

I am extremely grateful to Stanley Durrleman for his countless and precious advice and for his endless availability. He taught me how research should be conducted and I hope to become a researcher as scrupulous, creative and curious as he is. I am also truly grateful to Olivier Colliot for accepting me in his wonderful team, for helping me every time I was stuck with new ideas and points of view and for his sincerity and generosity. I also had the honour to be supervised by Nicholas Ayache. I wish to express my deepest gratitude to him for regularly coming to Paris, for always instilling positivity and good mood in our meetings, and for showing me how to communicate research results in an elegant, clear and nice way.

I also would like to sincerely thank Alain Trouvé and Daniel Rueckert for accepting to review this manuscript and all the other jury members Benoit Perthame, Jean-François Mangin, Rachid Derich, Andreas Hartmann and Yulia Worbe. Their feedbacks and suggestions definitely improved the quality of the manuscript and they helped me to see beyond the results of this Thesis, paving the way for new and exciting research perspectives.

I will never forget all the extraordinary people of the Aramis team (ex Cogimage) that I met during these three years. I would like to thank Mario, Fabrizio and Marie for the pleasant conversations about research and life. I am also grateful to all students, postdocs and engineers who tolerated (or helped me with) my jokes: Alexandre (un vrai Gaulois!), Claire (Calamity Chouquette!), Thomas S (docteur à la puissance 2!), Thomas E (rappelle-toi que je te vois...), Ludovic (je te vois aussi...), François (personne n'a compris notre art participatif!), JB (je t'attends à la cantine...), Barbara (Roar!), Ana (la plus gentille de tous!), Ali, Linda, Alexandre P, Hugo, Jérémy, Jorge, Catalina, Junhao, Mathieu, Fanny, Chabha, Sonia, Johanne, Corinne, Laurent and... obviously my favourite prey... Takoua! (souviens-toi de verrouiller l'écran ...)

I also would like to express my gratitude to all the Asclepios team members I got to know at MICCAI and during the wonderful scientific retreats at Auron. In particular, I would like to thank Xavier Pennec, Hervé Delingette and Maxime Sermesant for their precious suggestions and comments.

Eventually, I would like to thank both my Italian and French family. They took care of me and gave me anything a person might wish. In particular, I would like to dedicate this Thesis to my mother.

Contents

Chapter 1

Introduction

Contents

1.1 Atypical brain configuration in neurodevelopmental disorders

Neurodevelopmental disorders such as autism (Belger et al., 2011; Tye and Bolton, 2013), attention deficit-hyperactivity disorder (ADHD) (Konrad et al., 2010; Xia et al., 2012) or Gilles de la Tourette syndrome (GTS) (Liu et al., 2013; Worbe et al., 2015) are characterised by an atypical brain development which is thought to be associated to a different "wiring" architecture of the white matter with respect to healthy subjects. This would involve dysfunctions of the cortico-striato-thalamo-cortical (CSTC) circuits of the brain which could underlie the pathological symptoms.

The CSTC neural loops, schematically represented in Fig.1.1, are topographically and functionally well separated into sensorimotor, associative and limbic circuits which are implicated in motor, cognitive and motivational aspect of behaviour respectively (Alexander et al., 1986). Every circuit is composed of neural projections from several functionally related cortical areas which are sent to restricted portion

of the striatum (caudate nucleus and putamen). These striatal regions send further projections to the globus pallidus and substantia nigra, which in turn project to a specific region of the thalamus. Each thalamic region projects back to one (or more) cortical areas. The role of the basal ganglia and thalamus is to integrate the information coming from different inputs of the entire cerebral cortex and to "funnel" back these influences again to particular areas of the cortex.

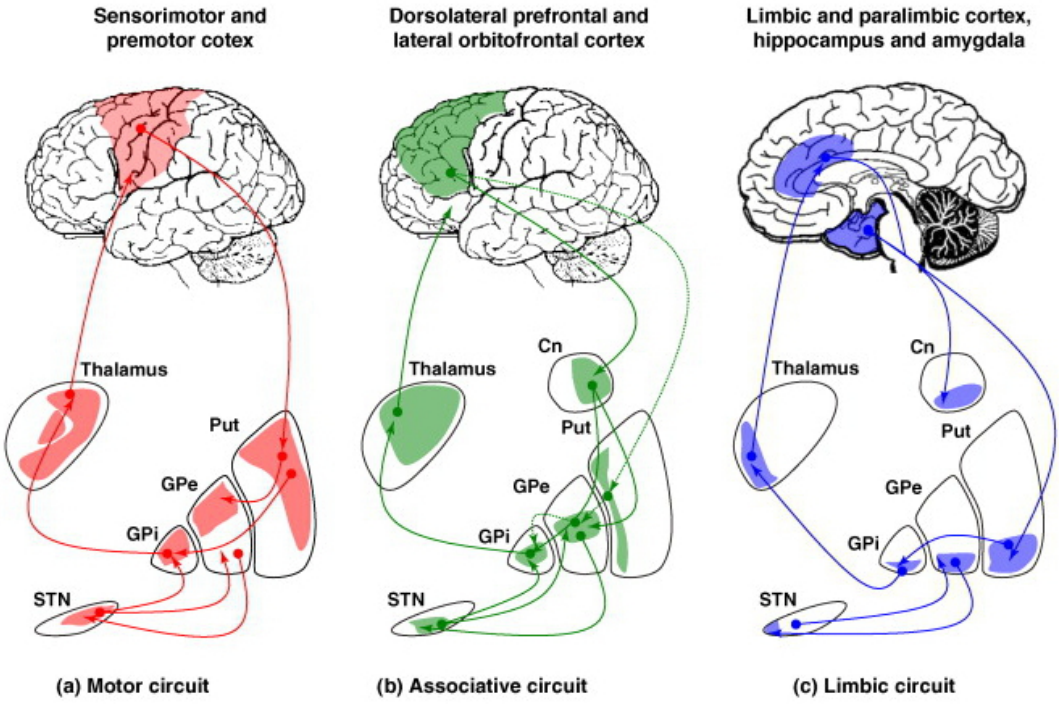

Figure 1.1: Schematic diagram illustrating the three main cortico-striato-thalamo-cortical circuits. Figure taken from (Krack et al., 2010).

Dysfunctions in these circuits are thought to be associated with *organisational* and *morphological* alterations of their anatomical components. The analysis of the configuration of these structures, namely their relative position, implies a systemic approach. All the components of the circuits, from both white and grey matter, need to be studied as a whole and not as a set of independent components.

Moreover, neurodevelopmental disorders are characterised by an important clinical heterogeneity. Each symptom is thought to be related to a specific pattern of anatomical alterations (Tuite and Dagher, 2013). Thus, it seems more likely that a syndrome would present a range of abnormalities rather than a single specific one. This pathological variability would be added to the normal variability presents in the healthy population.

The analysis of neurodevelopmental disorders need therefore *holistic* and *statistical* approaches to estimate the typical brain configuration associated to a disorder and its variability. This should then be compared with the characteristic organisation of a population of healthy controls in order to distinguish the normal variations

from the pathological ones. Alterations should be clearly *localised*, *quantified* and *characterised* in order to be easily *interpreted* by clinicians and permit to better characterise the pathophysiological mechanisms underlying the disorder.

Furthermore, the estimated typical organisations of both patients and healthy controls, together with their variability, could be employed to classify the configuration of a new patient in order to improve the diagnostic accuracy.

To this end, we propose in this thesis a computational and statistical framework to analyse morphological and organisational anomalies altering the anatomy of the neural circuits in neurodevelopmental disorders and we evaluate it on a disease model about Gilles de la Tourette syndrome.

1.2 Gilles de la Tourette syndrome

Gilles de la Tourette syndrome (GTS) is a heritable, childhood onset, neuropsychiatric disorder characterized by the presence of sudden, rapid, recurrent and non-rhythmic movements (motor tics) or vocalisations (phonic tics) (Hallett, 2015). In more than 80-90% of patients, tics are accompanied by psychiatric comorbidities such as obsessive-compulsive disorder (OCD) (Kimber, 2010). Typically the disorders begin in early childhood (6-7 years) with transient bouts of simple motor tics which are usually followed several years later by phonic tics. Tics may initially wax and wane but eventually they become persistent reaching the maximum severity at an age around 10-12 years. After that there is usually a steady decline in symptoms (Leckman, 2002; Jankovic and Kurlan, 2011). Once thought to be rare, current estimates change this view. GTS occurs in all countries, ethnic groups and social classes and epidemiological studies suggest that its overall prevalence could be as high as 1% of the general paediatric population in most countries (Leckman, 2002; Stern et al., 2005; Jankovic and Kurlan, 2011).

The exact aetiology of GTS is still unclear as well as its underlying pathophysiological mechanisms (Worbe and Hartmann, 2013). Different living environment and lifestyle factors are thought to be implicated in the development of GTS (McNaught and Mink, 2011). Genetic factors have also a major role and the clinical heterogeneity of this syndrome suggests that it might be related to different genes and to several genetic variants. Unfortunately, at present the genes that predispose individuals to develop the disorder are unknown (Leckman, 2002).

1.3 Driving hypothesis about GTS pathophysiology

In the last years, there has been an increasing interest for the neuroanatomical causes of this syndrome due also to the arrival of new high-resolution imaging tools. Different hypotheses have been proposed. Among them, inference from various approaches support that GTS is associated with dysfunctions of the cortico-striato-thalamo-cortical (CSTC) circuits (Worbe et al., 2015; Singer, 2013; Ganos et al., 2013; McNaught and Mink, 2011). According to that, GTS symptoms would arise

due to an atypical integration of information which would be associated to anatomical and functional abnormalities in the CSTC circuits. Alterations could be the result of a different brain maturation in GTS patients. In a typical brain development there is an initial increase in the number of neural connections which is followed by a selective pruning during brain maturation which optimizes the efficiency of the neural connections. An erroneous pruning in GTS, together with defects in the axonal myelinization, could lead to less efficient CSTC circuits and consequently also to the development of the symptoms. The heterogeneous clinical expression of GTS could be explained by dysfunctions impairing different CSTC circuits. For instance, simple tics are thought to be associated with abnormalities in the motor circuit (Worbe et al., 2010; Fahim et al., 2010; Sowell et al., 2008) whereas complex tics could be linked to dysfunctions of the associative circuit (Worbe et al., 2010; Neuner et al., 2010) and the presence of psychiatric co-morbidities, such as OCD and ADHD, could instead be associated with abnormalities in the limbic circuit (Peterson et al., 2007; Worbe et al., 2010).

These hypotheses were corroborated by experiments based on primate models (Worbe et al., 2009, 2013). Micro-injections of bicuculline, a γ-aminobutyric acid (GABA)ergic antagonist, were performed in different locations of the CSTC circuits to a group of 8 monkeys. These injections induced neuronal perturbations similar to three GTS symptoms, namely tic-like movements, hyperactivity and stereotyped behaviours. To localize the neurons perturbed by the bicuculline, the authors performed injections of axonal tracers followed by an histological analysis. It was shown that animals with abnormal movements resembling to simple motor tics had dysfunctions in the premotor and sensorimotor circuits. Instead, monkeys presenting behavioural disorders resembling to complex tics and compulsions had dysfunctions in the associative and limbic circuits.

These results were consistent with the aforementioned hypothesis but, as pointed out by the authors, they can not be directly translated to humans. Post-mortem studies in humans, like in (Kataoka et al., 2010), have also corroborated this hypothesis. Unfortunately, the small sample size of brain specimens, the large clinical heterogeneity of GTS, the presence of comorbidities and its evolution with age can confound the clinical conclusions. Nowadays, the precise mechanisms of the dysfunctions acting on the CSTC circuits are still poorly understood.

For this reason, the role of neuroimaging has already become fundamental since it is a non-invasive and painless technique which could help answering fundamental questions such as "Which are the anatomical abnormalities associated to the dysfunction of the CSTC circuits?" and more importantly "Where are they located and which is their extent?"

1.4 The role of structural and diffusion imaging

Structural and diffusion imaging studies seek to characterise the anatomy of the brain and the abnormalities associated to a pathology.

1.4. The role of structural and diffusion imaging

Magnetic Resonance Imaging (MRI) is a radiologist technique used to visualize internal structures of the body in detail. It makes use of the property of nuclear magnetic resonance (NMR) to produce high quality two- or three-dimensional images. Different modalities exist and two of the most employed are the T1-weighted (T1-w) and the diffusion MRI (dMRI). The first one provides an excellent soft tissue contrast which is especially useful for characterising the anatomy of both grey and white matter in the brain. The second one provides instead images whose contrast is influenced by the differences in water molecule diffusivity and it is used to study the micro-structure and the organisation of the white matter axons.

Different studies based on these two modalities have already produced a reasonable body of evidence consistent with the aforementioned hypothesis about GTS. For instance, morphometric studies on T1-w scans have shown a volume reduction in the caudate and in the sensorimotor, parietal, orbitofrontal and anterior cingulate cortical areas. Moreover, it was also reported an enlarged hippocampus, amygdala, putamen and corpus callosum in GTS patients (Felling and Singer, 2011). These results were obtained by comparing scans of healthy subjects (controls) with GTS patients using mainly two techniques. The first one is based on the comparison of morphometric measurements, such as the volume, of single and clearly defined grey matter structures of the circuits (i.e. caudate and putamen) which are manually or automatically segmented (Felling and Singer, 2011; Peterson BS et al., 2003). This technique produces results which are rather easy to interpret but they give just an overall insight on the pathological anatomical effects of GTS without being able to reveal the exact location of the abnormalities and whether inter-linked neighbour components are also affected.

The second technique is called Voxel-Based Morphometry (VBM) (Ashburner and Friston, 2000) and it is based on non-rigid registrations of the structural scans of both controls and patients to a reference scan called template. This permits to bring all the scans in the same reference frame allowing a voxel-wise comparison between the deformed images. After segmenting every subject scan into different tissue types and smoothing the resulting segmentations, mass-univariate statistics are used to detect statistically significant differences in the local concentration of grey matter between the two groups. A similar approach, called Tensor-Based Morphometry (TBM) (Davatzikos et al., 1996), uses instead the deformations to quantify and model the anatomical differences between subjects. It is based on the comparison of the Jacobian matrix fields of the deformations (or their determinants) using multi-variate (or mass univariate) statistics. These two techniques can also be combined in an augmented version of VBM where every voxel of the deformed image is multiplied by its volume change which is encoded in the Jacobian determinant of the deformation field (Ashburner and Friston, 2000). In this way it is possible to compare the absolute amount of gray matter instead than the relative concentration of grey matter with respect to the other tissues like white matter. These methods do not focus on single structures and they result in a comprehensive evaluation of anatomical alterations across the whole brain. However, they can investigate only focal differences at each voxel without taking into consideration

the organisation of the different components. Moreover, these techniques are highly sensitive to misalignement and misclassification errors which can bring to erroneous interpretations. Their target is usually the grey matter but they could be used to examine also the volume variations in the white matter. However, white matter in T1-w scans is usually characterized by large homogeneous regions with only subtle changes in intensity.

For this reason, white matter is predominantly analysed with dMRI scans where the diffusion at each voxel is modelled as a tensor (Diffusion Tensor Imaging - DTI) or using more accurate model-free approaches such as Q-Ball Imaging (QBI) (Tuch, 2004) which can resolve intra-voxel fiber crossing, contrarily to DTI. White matter microstructure is then characterised by scalar quantities such as fractional anisotropy (FA) (Pierpaoli and Basser, 1996) or generalized fractional anisotropy (GFA) (Tuch, 2004; Cohen-Adad et al., 2008) which are based on DTI and QBI respectively. They define the anisotropy level of diffusion at each voxel and they can be interpreted as measures of the structural integrity of the fiber bundles. These quantities can be compared voxel-wise across groups of subjects using VBM or TBSS (Tract-based spatial statistics) (Smith et al., 2006). The latter method, similarly to VBM, starts by aligning all FA (or GFA) images to a template using a cascade of affine and non-linear registration. These images are then projected to a "skeletonised" mean FA (or GFA) image, which represents the centres of all tracts common to the group of subjects, where mass-univariate statistics are carried out. Differently from VBM, this technique permits to avoid the image smoothing step, where it is still unclear how to fix the optimal smoothing parameter, and it is supposed to reduce the registration errors thanks to the projection to the mean FA (or GFA) image. Both techniques have been employed for studying GTS and with VBM it was shown a decrease in FA in the corticospinal tract, internal capsule and corpus callosum as well as an increase in the white matter underlying the ventral and posteriolateral part of the thalamus in patients subject to GTS (Thomalla et al., 2009; Neuner et al., 2010) whereas it was shown no differences in FA using TBSS (Govindan et al., 2010). However, in another study based on TBSS (Liu et al., 2013), the authors found a significant axial diffusivity (AD) and mean diffusivity (MD) increases in the right anterior thalamic radiation.

1.5 A holistic approach

Grey and white matter are thus usually studied separately and independently using either T1-w or dMRI scans respectively. However, due to the interwoven nature of the brain, the syndrome of Gilles de la Tourette is supposed to affect all the components of the circuits, namely both grey matter structures such as sub-cortical nuclei and cortical surface as well as their inter-linking white matter tracts. Associated abnormalities could alter the *shape* of each one of these components as well as their *relative position*. Following Dryden and Mardia (1998), we define the shape of a structure as the geometrical information invariant to the Euclidean similarity trans-

formations. This implies that two structures have the same shape if the first one can be transformed into the second one using only translations, scaling and rotations. When analysing an anatomical complex, such as the CSTC circuits, abnormalities can affect also the position of a structure with respect to the other ones, namely their relative position. For instance, a dysfunction of the flux of information in the CSTC circuits can modify the location where white matter tracts integrate grey matter structures, namely the structural connectivity, as well as the spatial organisation of the sub-cortical nuclei. These anatomical alterations could be captured only analysing the CSTC circuits as a whole, that is studying both grey and white matter structures simultaneously.

As previously seen, this would imply that the information about the grey matter structures captured from T1-w scans should be combined with the information about white matter from dMRI scans. However, it is still unclear how to consistently integrate these two different imaging modalities into a single computational framework. Few integrative multi-modal approaches have been proposed. Two examples are (Avants et al., 2010) and (Savadjiev et al., 2014) where the authors proposed to investigate the relationship between geometrical characteristics of grey and white matter using correlation analysis and mutual information respectively. These approaches can show statistically related abnormalities in grey and white matter which can help to clarify how a certain syndrome acts on the whole brain. However, these methods *independently* extract only specific morphological characteristics from each scan which are then statistically related. This does not permit to take into consideration the organisation of the circuits and the conclusions that can be drawn are limited to the morphological characteristics chosen (i.e. volume, grey matter concentration, FA).

In this thesis, we follow a different strategy which allows us to integrate the information of different modalities into a single virtual representation of the brain anatomy. We propose to shift the morphological analysis from images to complexes of 3D meshes segmented from the different structural scans. In the case of CSTC circuits, cortical surface and sub-cortical nuclei are segmented from T1-weighted scans and they are modelled as 3D surfaces. Estimates of the trajectories of large groups of white matter neural axons are instead traced from dMRI scans using tractography algorithms and they are modelled as bundles of 3D curves called streamlines. An example of anatomical complex of 3D meshes is shown in Fig.1.2. Tractography is the only non invasive way to model the white matter architecture and therefore it is also the only technique capable to capture the information about structural connectivity. Studying how and where neural fibers integrate the grey matter structures is essential in order to understand which are the functional areas implicated in the abnormal brain development of GTS and how this syndrome modify the organisation of the CSTC circuits.

The proposed approach permits to visualise and analyse all the components of the CSTC circuits as a single complex. Our goal is to verify and, if confirmed, even localize and describe the presence of anatomical abnormalities associated to GTS within the CSTC circuits. This can be achieved by comparing two groups

of anatomical complexes representing the CSTC circuits of controls and patients respectively. The first step is to define a computational model and metric for both 3D curve and surface meshes to compare the representations of the circuits across individuals. This model should then be integrated into a statistical framework to characterise the invariants and the organisational and morphological variability of each clinical group. Eventually, we should be able to compare these statistics in order to localize and quantify the potential abnormalities affecting the GTS group. Furthermore, the use of a statistical method also increases the robustness of the results with respect to possible imperfections due to segmentation or tractography errors. Since these errors should not be systematic, they should not influence (or at least to a smaller extent) the estimate of the common morphological and organisational characteristics as well as their variability.

It is important to say that another possible strategy would be to use an iconic-geometric approach using both images and meshes like in Siless et al. (2012); Graciano Fouquier et al. (2014). However, these strategies have been used so far just to improve the accuracy of image-based registrations. Moreover, results based on images and meshes would be more difficult to visualise, and therefore also to interpret, than using only meshes as proposed in this thesis.

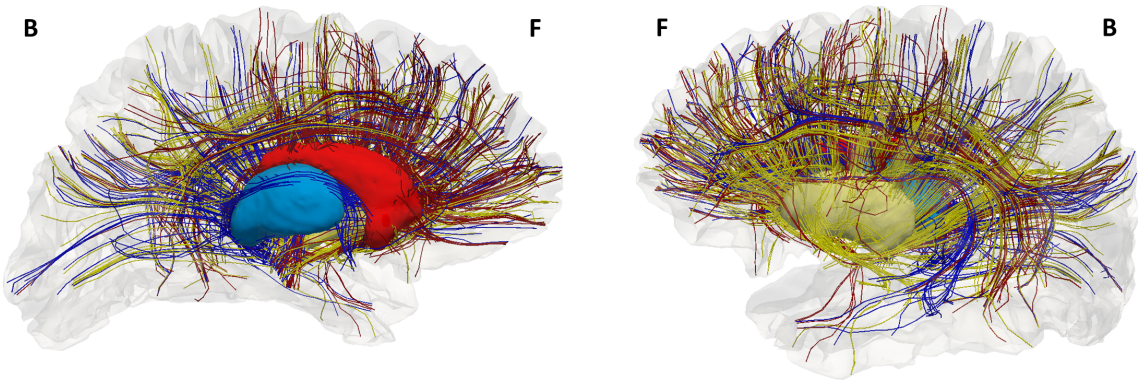

Figure 1.2: Two views of the same anatomical complex of 3D meshes representing the caudate, putamen and thalamus of the left hemisphere of the cortex in red, yellow and blue respectively. Streamlines connecting these three sub-cortical nuclei to the left hemisphere of the cortex are also shown with darker colours. Fibers stem from a deterministic tractography and they have been uniformly downsampled for visualisation purpose. B=back of the head, F= front of the head.

1.6 Mesh-based morphometry

A 3-dimensional surface mesh is composed of connected polygonal faces (usually triangles) which recreate the 3D shape of the grey matter objects. Meshes are built

starting from the outlines of the structures on each MRI scan, namely the segmentation results, using algorithms such as marching cubes (Lorensen and Cline, 1987). Curve meshes resulting from tractography algorithms are called streamlines, they link two areas of the dMRI scans and they are constituted of segments connecting neighbouring voxels. In this thesis we will consider only streamlines connecting two grey-matter structures of the CSTC circuits.

1.6.1 Computational models for complexes of 3D meshes

Different computational models for both surface and curve meshes have been proposed and probably the most well known is the point distribution model (Cootes et al., 1995). It relies on the definition of consistent point correspondences (landmarks) between meshes of different subjects. They can be either manually identified by an expert (Bookstein, 1997) or automatically estimated using, for instance, dynamic particles (Cates et al., 2008) or by propagating the landmarks of an atlas through an affine (Brett and Taylor, 1998) or non-rigid transformation (Frangi et al., 2002). Landmarks have been used mainly for 3D surface meshes since finding point-correspondences between white matter fiber bundles composed of hundreds of thousands of streamlines is a difficult, if not impossible, task especially when the streamlines do not share a similar length and starting/ending anatomical locations.

Several extensions of this method have been proposed and they are based on correspondences defined using other shape parametrisations such as spherical harmonics (Brechbühler et al., 1995), minimum-description length (Davies et al., 2002), skeleton-based representations (Golland et al., 1999), m-reps (Yushkevich et al., 2001), cubic B-splines (Corouge et al., 2006) and Fourier descriptors (Székely et al., 1996; Batchelor et al., 2006). These shape descriptors have been broadly and successfully employed to analyse several brain structures in both single- and multi-object studies. However, most of them are conceived for only a particular kind of mesh, i.e. genus-zero surfaces (Lu et al., 2007; Gorczowski et al., 2010) or streamlines (O'Donnell et al., 2012), which prevents the integration of all the components of the CSTC circuits in a unified framework.

In contrast to previous shape parametrisations, the frameworks of currents (Glaunès, 2005; Vaillant and Glaunès, 2005; Durrleman, 2010) and varifolds (Charon and Trouvé, 2013b) are not limited to a specific kind of mesh and they do not need point-correspondences or streamline-correspondences when dealing with white matter fiber bundles. The framework of varifolds can be seen as the non-oriented version of the one of currents, in the sense that meshes do not need a consistent orientation among subjects. These models can thus provide a unifying framework to process all the components of the CSTC circuits. Moreover, they embed all structures in a Hilbert space which means that, for instance, the union of different streamlines forming a fiber bundle is modelled as a sum of vectors. This simplifies standard computations such as norms, distances and calculus making these frameworks two valid candidates on which ground the subsequent statistical shape analysis.

Eventually, an important characteristic of shape descriptors is their computa-

tional efficiency, namely the computational load required to process the modelled meshes. In the case of white matter fiber bundles this is fundamental since the huge number of streamlines resulting from tractography algorithms can make any analysis infeasible. To this end, different parsimonious representations have been proposed (see Chapter 3 for a small review) such as the one in Durrleman et al. (2011a) based on a matching-pursuit algorithm for currents. This approximation is very concise but it has the drawback to accurately approximate only the areas of the bundle characterised by a high density of streamlines, like the central mass, almost ignoring the external smaller fascicles. Moreover, this approximation does not conserve the topology of streamlines since it can be seen as a set of disconnected points. This prevents the study of structural connectivity and therefore it can not be used to test the aforementioned hypothesis about GTS.

1.6.2 Statistical shape analysis of complexes of 3D meshes

Once described how to mathematically represent the structures of an anatomical complex and how to measure a distance between them, one needs to define how to statistically analyse their morphological and organisational variability within each clinical group. Differences between the two groups will characterise the abnormalities related to GTS.

The first step of a statistical shape analysis is to globally align all the complexes using rigid or affine transformations in order to filter out the uninformative shape variability[1]. Transformations are applied to all the structures of the complex at the same time and the remaining differences regard the local relative position between components and their individual shape. The alignment can be obtained using, for instance, Procrustes analysis (Dryden and Mardia, 1998) which is based on the minimisation of the sum-of-squared distances between the structures with respect to the similarity transforms chosen. Distances are defined as the standard Euclidean metrics when dealing with landmarks and as geodesic distances based on appropriate Riemannian metrics when dealing with other shape descriptors such as m-reps (Gorczowski et al., 2010).

Once all complexes are in the same reference frame it is possible to compute the mean configuration and the main morphological variations of the population based on a Principal Component Analysis (PCA) when working with Euclidean vector spaces (i.e. landmarks) or Principal Geodesic Analysis (PGA) when complexes lie on nonlinear Riemannian spaces (Fletcher et al., 2004). Average configurations and main modes of variations of the two clinical groups can be compared in order to localize and quantify the abnormalities. Otherwise, differences can also be visualised by deforming the average configuration of one clinical group towards the orthogonal direction to the most discriminative hyperplane in the feature space, namely the direction that best separates the two clinical groups. The feature space can be composed by the parameters describing the shape of each individual component

[1] Depending on the hypothesis of the study, scaling can be an important feature.

and/or their relative position, depending on the hypothesis to test. This strategy is not computationally intensive and the results are easy to interpret but it has the drawback that it does not take into consideration the fact that the anatomical organisation of the circuits should not change, namely that separated structures should not intersect, when computing the average, the main morphological variations or when deforming the average configuration towards the most discriminative direction.

From a different perspective, rooted in the seminal book of D'Arcy Wentworth Thompson (D'Arcy Wentworth, 1917), Grenander proposed to quantify the morphological differences between two anatomical complexes as the "amount" of deformation needed to warp the first one to the second one. Moreover, he proposed also a *generative* statistical model where every complex of a population is modelled as a deformation of a common anatomical complex called template complex (Grenandner, 1993; Miller et al., 1993; Christensen et al., 1996; Grenander and Miller, 1998; Allassonnière et al., 2007). Deformations act on the entire ambient space and they modify the template complex such that it resembles to the subject complexes putting into correspondence the components of the template complex with the homologous ones of the subjects. Similarity between correspondent structures is defined by the metric of the computational models previously described. First examples used a fixed template chosen a priori like the Talairach atlas (Davatzikos et al., 1996) or a randomly-selected subject (Csernansky et al., 1998). However, this kind of templates might have imperfections due to segmentation errors and they might not be "centred" with respect to the population under study, biasing consequently the whole analysis. Later works have thus proposed to automatically estimate the template without imposing it at the beginning (Joshi and Miller, 2000; Qiu et al., 2010; Durrleman et al., 2014). The joint estimate of the template complex and the deformations has been called: *atlas construction*.

In this setting, the template complex captures the common morphological characteristics of the population, and it can be thought as an average. The deformations model instead how each subject varies with respect to this average and they represent therefore the morphological variability of the population. It is important to highlight that this framework is generic and it can be applied with any kind of mesh and shape parametrisation, on condition that they are compatible with the type of deformation employed, and even with T1-w images like in Joshi and Miller (2000); Avants and Gee (2004); Bhatia et al. (2004); Zhang et al. (2013). This avoids the use of segmentation and tractography algorithms but it prevents the analysis of structural connectivity since it is impossible to understand where white matter fibers integrate grey matter structures from T1-w scans. Furthermore, in the case of 3D surfaces and 2D/3D curves (open and closed) other authors have also proposed to define the deformations directly on the meshes rather than assuming an underlying deformation of the whole space (Joshi et al., 2007; Michor and Mumford, 2007; Younes et al., 2008; Kurtek et al., 2011; Bauer and Bruveris, 2011; Bauer et al., 2014; Gutman et al., 2015). However, these methods have been usually designed for single-object studies and only for one specific kind of mesh. Thus, they can not be

employed for the analysis of the entire CSTC network.

In order to preserve the anatomical organisation of the template complex, deformations are usually defined as diffeomorphisms, namely smooth bijective deformations with smooth inverse[2]. This kind of deformation prevents the structures to intersect, fold, torn or collapse to single points. This is very important in order to obtain an anatomically relevant analysis.

The group of diffeomorphisms constitutes a non-linear curved manifold which is not a vector space. Statistics such as average or principal modes need to be defined in the tangent space of this manifold (Vaillant et al., 2004; Pennec, 2006). In the framework known as large deformation diffeomorphic metric mapping (LDDMM) (Dupuis et al., 1998; Trouvé, 1998; Beg et al., 2005), diffeomorphic deformations are defined as the last ones of a flow produced by integrating a time-varying vector field which is constrained to be smooth. Based on the metric for such a group of diffeomorphisms defined in Trouvé (1995), the flow that warps the template complex to every subject complex is chosen as the geodesic path between the identity transform and the last deformation of the flow (i.e. the path with the minimal cost required to make the template similar to the subject complex). In Miller et al. (2006) it is shown that the (final) deformation of the template complex depends only on the dual of the initial time-varying vector field, called initial momentum, which belongs to a vector space. In this space, it is therefore possible to use linear statistical analysis such as PCA (Principal Component Analysis) to study the non-linear shape space defined as the orbit of the template complex under the action of the group of diffeomorphisms (Vaillant et al., 2004). Moreover, it offers also a computationally efficient way to compute the deformations, called geodesic shooting (Miller et al., 2006). Different algorithms have been proposed based on this principle such as in Cotter and Holm (2006); Marsland and McLachlan (2007); Avants et al. (2008); Ashburner and Friston (2011); Vialard et al. (2011).

In this thesis, we will use the one proposed in Durrleman et al. (2011b) where the time-varying vector fields are defined with a small set of control points. The number of control points, chosen by the user, determines the number of initial momenta and it does not increase with the number of analysed structures in the complex. This makes such a model well suited for the statistical analysis of anatomical complexes like the CSTC circuits. Moreover, it has already been used with both the framework of currents (Durrleman et al., 2011a) and varifolds (Durrleman et al., 2014) on which we will ground our analysis in the following chapters. Nevertheless, it is important to notice that for the scope of this thesis any parametric diffeomorphic deformation model could have been used (provided that it allows to compute statistics).

For instance, another powerful setting which allows to compute statistics on groups of diffeomorphisms is the one proposed in Arsigny et al. (2006) where the authors parametrised diffeomorphic deformations with smooth stationary velocity fields (SVF). They presented a Log-Euclidean framework which permits to com-

[2]Topology preservation is assured even by a homeomorphism (that is a continuous deformation with a continuous inverse). Every diffeomorphism is a homeomorphism but with an addition constraint on the regularity (smoothness) of both the transformation and its inverse.

pute the exponential map and the logarithm map on the manifold of the group of diffeomorphisms using the scaling-and-squaring method. In this setting the statistical analysis is performed in the vector space given by the logarithms of the diffeomorphisms. Several extension and applications have been proposed like in Ashburner (2007); Vercauteren et al. (2008); Modat et al. (2012); Lombaert et al. (2012); Lorenzi et al. (2013).

Other diffeomorphic registration algorithms exist such as the one in Christensen et al. (1996) which models the template as a viscous fluid whose deformations are based on the Navier-Stokes equations (non-linear partial differential equations) or the one in Rueckert et al. (2006) based on Free-Form deformations. For a more exhaustive description about diffeomorphic registration algorithms the user is referred to Sotiras et al. (2013).

All the previous diffeomorphic models provide dense deformations which vary locally across the ambient space permitting to capture the variations in the relative position between *separated* structures such as the sub-cortical nuclei (Wang et al., 2007). However, using a single diffeomorphism we implicitly assume that the relative position between structures *in contact with* each other does not change across subjects. This implies that every white matter fiber tract of the CSTC circuits should link the same areas of the cortical surface and basal ganglia across the whole population. This assumption is in contradiction with the aforementioned hypothesis about GTS and it precludes the study of abnormalities in structural connectivity which could be caused by an abnormal brain development.

GTS should be studied with an holistic approach where morphometry and structural connectivity analysis are *unified* into a single computational framework. However, structural connectivity is usually studied independently of the integrated grey matter structures without considering neither their morphology nor their organisation.

1.6.3 Structural connectivity analysis

Structural connectivity refers to the density of connections resulting from a tractography method and integrating grey matter structures.

The most popular strategy to study differences in structural connectivity between two clinical groups is based on the subdivision of the cortical surface and sometimes even sub-cortical nuclei in parcels which are reproducible across the subjects of both groups. This step is usually based on the registration of all subject T1-w scans to a common atlas (i.e. the MNI-space, Montreal Neurological Institute) which is already parcelled. After that, the spatial location of the parcels in the T1-w scan is transferred to the DWI scan of the same subject through registration. Then, it is used a tractography algorithm on the parcelled DWI scans to trace streamlines from specific ROI (Region of Interest) or throughout the whole brain. Eventually, one counts the number of streamlines connecting two ROIs, two parcels or a determined ensemble of parcels of the cortical surface (i.e. motor area) to a sub-cortical nucleus. The density of streamlines quantifies the structural connectivity. Multi-

variable or uni-variate statistics can then be used to find the most discriminating parcels between the two groups. Discordant results exist about GTS. For instance in Makki et al. (2009); Cheng et al. (2014) the authors have found that in patients both caudate and thalamus have a *reduced* connectivity with the anterior dorsolateral frontal cortex and that the frontal cortical areas have a *lower* connectivity with the basal ganglia. On the other hand, in Worbe et al. (2015) the authors pointed out that both the motor circuit (primary motor and sensory cortical areas) and the associative circuit (lateral orbitofrontal cortex) presented an abnormally *enhanced* structural connectivity with the striatum and thalamus in patients subject to GTS. The discrepancy between these studies could be due to methodological differences or to the small size of the data samples in the first two studies (15 and 18 subjects respectively with respect to 77 in Worbe et al. (2015)). Nevertheless, this also indicates that there is a need of new, rigorous and robust statistical methods to brighten the GTS mechanism affecting structural connectivity.

The field of complex networks (Bullmore and Sporns, 2009; Rubinov and Sporns, 2010), based on graph theory, proposes innovative solutions to study structural connectivity as a network where the nodes are the parcels and the edges are constituted by the number of streamlines (or other indices) linking two nodes. Different measures characterising a network exist such as the degree distribution, small-world topology, highly connected hubs and modularity. They usually have a physical meaning which makes them easily interpretable. There are already some works about GTS which make use of these graph measures (Worbe et al., 2012) but most of them are about functional networks and not about structural connectivity networks. Moreover, to the best of our knowledge, there are no methods based on the complex network theory which integrate morphology and structural connectivity analysis.

1.7 Contributions and Manuscript Overview

The purpose of this thesis is to provide a methodology which is able to discover and characterise morphological and organisational abnormalities in anatomical complexes composed of both grey and white matter structures. To this end, the complexes must be studied as a whole and not as a set of independent components. The proposed method should be able to discover both local and distributed shape alterations across the whole anatomical complex. Moreover, it should unveil also pathological changes altering the relative position of both separated components (i.e. sub-cortical nuclei) as well as connected structures (i.e. structural connectivity). All these abnormalities should be quantified, accurately localised and easily visualised in order to be useful in the clinical domain. The proposed method has been driven by the will to test the aforementioned hypothesis about GTS. Nevertheless, it could also be applied to the study of other neurodevelopmental syndromes which are thought to be associated with alterations of the CSTC circuits or other neural networks.

We propose to ground our method on the generative model of the previously described Grenander's atlas construction procedure. This allows us to estimate the common features of a population of anatomical complexes, represented by the template complex, and at the same time the morphological variability described by the set of diffeomorphisms, one for each subject. These characteristics can then be compared between a group of healthy subjects and one of GTS patients in order to localize the possible abnormalities due to the syndrome. This strategy has been mainly employed for studying single structures or complexes composed of *separated* structures. Here we propose to extend this method to anatomical complexes composed of grey matter structures *interconnected* by white matter tracts. This kind of anatomical complexes can not be studied with previous *deterministic*, *single-diffeomorphic* atlas procedures based on *computationally expensive* or *topology-neglecting* white matter shape models.

To this end, we propose to embed the atlas construction in a *Bayesian framework* which permits to automatically estimate important balancing parameters, one for each structure of the complex, which were fixed by the user in previous methods and which highly influenced the results. Moreover, this statistical setting also permits to directly estimate a consistent and well-conditioned covariance matrix, which is fundamental for the statistical analysis of the morphological and organisational variability. In order to capture the variations in the relative position of both separated and attached structures, making thus possible the analysis of structural connectivity, we propose also an innovative deformation setting called double diffeomorphism. It is a cascade of two diffeomorphisms which can model the variation in relative position between white and grey matter structures and at the same time correctly warp the template complex to every subject complex. Eventually, to reduce the computational resources required to work with white matter fiber bundles, we also propose to approximate them with a parsimonius representation based on weighted prototypes. This makes feasible the computation of a Bayesian double diffeomorphic atlas.

These contributions are thoroughly explained in the following chapters. It follows a brief description of each one of them.

1. **Chapter 2**: The estimate of the atlas is usually based on a *deterministic* approach where a gradient descent scheme is used to minimize a cost function constituted of M data-terms, one for each structure of the complex, and a regularity term on the diffeomorphisms. The influence of the data-terms is balanced by M trade-off parameters which are *fixed* at the beginning. The choice of these values is crucial since they influence the results. A value too small might weight too much the data term leading to a situation of overfitting. On the contrary, a value too large might penalize the deformations reducing the accuracy of the matching. Moreover, the number of deformation parameters (initial momenta) is usually much higher than the number of subjects which makes the atlas construction a high-dimensional low samples size (HDLSS) problem. The study of the morphological variability of a population

is often based on a Principal Component Analysis (PCA) which needs the computation of the covariance matrix of the deformation parameters. In a HDLSS setting one needs to regularize this estimate *at the end* of the atlas construction making therefore the computed covariance matrix *inconsistent* with the optimal deformations resulting from the atlas procedure. To this end, we propose to embed the atlas construction in a *Bayesian framework* taking advantage of the prior distributions introduced in (Allassonnière et al., 2007). This allows us to estimate both the trade-off parameters and a well-conditioned covariance matrix of the deformation parameters which can be directly employed in statistical analysis such as PCA. Furthermore, we propose to model all meshes as Gaussian random varifolds. Since the space of varifolds is of infinite dimension, we define a grid for every structure, similarly to (Durrleman, 2010), on which both shapes and template are projected to. This determines a finite-dimensional approximation space on which it is possible to define probability density functions. Eventually, we extend also the proposed Bayesian framework to a multi-population atlas construction. We take advantage of the clinical diagnosis of the subjects to divide them into different groups (i.e. controls and patients). The groups share the same template and their deformation parameters are modelled with different distributions which are defined on the same space. This allows us to compare them and to test hypothesis of equality between the moments of their distributions such as their covariance matrices.

2. **Chapter 3**: Fiber bundles stemming from tractography algorithms contain many streamlines, from few thousands up to millions. They require therefore a great amount of computer memory and computational resources to be stored, visualised and processed. This can limit their use in computationally-intensive processes such as the atlas construction. To this end, we propose to approximate a fiber bundle with a parsimonious representation of *weighted prototypes*. Prototypes are chosen among the streamlines and they represent groups of similar streamlines. Their weight is related to the number of streamlines approximated. Both streamlines and prototypes are modelled with a new extension of the framework of currents, which we have called *weighted currents*. This computational model inherits all the qualities of usual currents, such that it does not need point-to-point correspondences, and it augments its geometry-based metric with information about structural connectivity in the spirit of functional currents (Charon and Trouvé, 2013a). Two streamlines are considered similar if their endpoints are close to each other and if their pathways are similar. Moreover, the space of weighted currents is also a vector space with a closed-form metric. This permits to easily compute the approximation error and to select the prototypes based on the minimisation of this error. We propose an iterative algorithm which approximates independently and simultaneously all the fascicles of the bundle. This allows us to approximate the whole bundle in a fast and accurate way. We evaluate our algorithm

on bundles resulting from both deterministic and probabilistic tractography algorithms. We show that this approximations use on average only 2% of the original streamlines as prototypes. This drastically reduces the computational burden required for any processing such as atlas construction while preserving both the geometry and the structural connectivity of the original bundles. The latter quality is fundamental for the scope of this thesis.

3. **Chapter 4**: Single-diffeomorphic deformations can correctly put into correspondence anatomical complexes composed of *separated* structures. Since they preserve topology, they do not alter the organisation of the anatomical complexes preventing structures to fold, intersect or shear. However, this also implies that the relative position between structures *in contact* with each other should not change across subjects. This means, for instance, that white matter fiber bundles should link the same functional areas of the cortical surface and basal ganglia across the whole population. This assumption precludes the study of changes in structural connectivity and therefore it also impedes to test the aforementioned hypothesis about GTS. We propose to tackle this problem with a new deformation setting called *double diffeomorphism*. The template complex is warped towards every shape complex of the population using a cascade of two diffeomorphisms. The first one acts only on the white matter of the template complex, keeping fixed the grey matter. It can be seen as a relative change of coordinates within the gray matter considered as a fixed reference frame. During this transformation the fiber bundles slide on the fixed gray matter structures, capturing the variation in structural connectivity. These variations can be compared among subjects since they are all computed with respect to the same reference frame, namely the fixed gray matter of the template complex. The second diffeomorphism acts on the whole template complex, namely on both the resulting white matter and on the gray matter. This is a global change of coordinates which has the role to put into correspondence both gray and white matter structures of the template complex with the homologous ones of the subject's shape complex. The first diffeomorphism re-positions the white matter tracts within the gray matter such that all the components of the template complex can be correctly put into correspondence by the second diffeomorphism. The two diffeomorphisms are optimised together minimising a single cost function and using a gradient descent scheme. This new deformation setting is combined with the Bayesian framework presented in Chapter 2 into a Bayesian double diffeomorphic atlas construction. We use this framework to study the morphological and organisational variations of the cortico-putamen circuit in both controls and GTS patients. By comparing these results we are able to visualise and quantify the differences between the two clinical groups both in terms of shape and structural connectivity. The feasibility of this study is due to the parsimonious representation for white matter fiber bundles proposed in Chapter 3. Eventually, we also show that a double diffeomorphic approach leads to better

classification scores than using the single diffeomorphic deformation presented in Chapter 2. We use 49 GTS patients and 27 healthy controls with a Linear Discriminant Analysis (LDA) and a leave-one-out cross-validation strategy.

1.8 List of Publications

Journal Articles

- **Pietro Gori**, Olivier Colliot, Linda Marrakchi-Kacem, Yulia Worbe, Cyril Poupon, Andreas Hartmann, Nicholas Ayache and Stanley Durrleman. (2015) A Bayesian Framework for Joint Morphometry of Surface and Curve meshes in Multi-Object Complexes. *submitted to Medical Image Analysis.* Chapter 2

- **Pietro Gori**, Olivier Colliot, Linda Marrakchi-Kacem, Yulia Worbe, Fabrizio De Vico Fallani, Mario Chavez, Cyril Poupon, Andreas Hartmann, Nicholas Ayache and Stanley Durrleman. (2015) Parsimonious Approximation for White Matter Bundles of Streamlines. *submitted to IEEE Transactions on Medical Imaging.* Chapter 3

- **Pietro Gori**, Olivier Colliot, Linda Marrakchi-Kacem, Yulia Worbe, Alexandre Routier, Cyril Poupon, Andreas Hartmann, Nicholas Ayache and Stanley Durrleman. (2015) Unified morphometry and structural connectivity analysis of shape complexes using double diffeomorphisms. *to submit to NeuroImage.* Chapter 4

Peer-Reviewed Conference Articles

- (Gori et al., 2013a) **Pietro Gori**, Olivier Colliot, Yulia Worbe, Linda Marrakchi-Kacem, Cyril Poupon, Andreas Hartmann, Nicholas Ayache and Stanley Durrleman. (2013) Bayesian Atlas Estimation for the Variability Analysis of Shape Complexes. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (Eds.). *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013.* No. 8149 in Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 267–274. Poster

- (Gori et al., 2014) **Pietro Gori**, Olivier Colliot, Linda Marrakchi-Kacem, Yulia Worbe, Fabrizio De Vico Fallani, Mario Chavez, Cyril Poupon, Andreas Hartmann, Nicholas Ayache and Stanley Durrleman (2014) A Prototype Representation to Approximate White Matter Bundles with Weighted Currents. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (Eds.). *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014.* No. 8675 in Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 289–296. Poster

- (Gori et al., 2015a) **Pietro Gori**, Olivier Colliot, Linda Marrakchi-Kacem, Yulia Worbe, Alexandre Routier, Cyril Poupon, Andreas Hartmann, Nicholas Ayache and Stanley Durrleman. (2015). Joint Morphometry of Fiber Tracts and Gray Matter Structures Using Double Diffeomorphisms. In: Ourselin, S., Alexander, D. C., Westin, C.-F., Cardoso, M. J. (Eds.), *Information Processing in Medical Imaging. - IPMI 2015.* No. 9123 in Lecture Notes in Computer Science. Springer International Publishing, pp. 275–287. Oral - Podium Presentation

Abstracts

- (Gori et al., 2013b) **Pietro Gori**, Olivier Colliot, Yulia Worbe, Linda Marrakchi-Kacem, Sophie Lecomte, Cyril Poupon, Andreas Hartmann, Nicholas Ayache and Stanley Durrleman. (2013) Towards joint morphometry of white matter tracts and gray matter surfaces. *Organization for Human Brain Mapping - OHBM 2013.* Oral - Podium Presentation + Trainee Abstract Travel Award

- (Gori et al., 2015b) **Pietro Gori**, Olivier Colliot, Linda Marrakchi-Kacem, Yulia Worbe, Alexandre Routier, Cyril Poupon, Andreas Hartmann, Nicholas Ayache and Stanley Durrleman. (2015). Unified analysis of shape and structural connectivity of neural pathways. *Organization for Human Brain Mapping - OHBM 2015.* Poster

Workshops

- (Routier et al., 2014) Alexandre Routier, **Pietro Gori**, Ana Beatriz G. Fouquier, Sophie Lecomte, Olivier Colliot and Stanley Durrleman. (2014) Evaluation of morphometric descriptors of deep brain structures for the automatic classification of patients with Alzheimer's disease, mild cognitive impairment and elderly controls. *MICCAI challenge on Computer-Aided Diagnosis of Dementia based on structural MRI data (CADDementia 2014).* Oral - Podium Presentation

Chapter 2

A Bayesian framework for multi-object atlas construction

This chapter has been submitted to Medical Image Analysis and it has been partly published in Gori et al. (2013a). *The C++/CUDA code of the described Bayesian framework will be soon available at* `www.deformetrica.org`

Contents

2.1 Introduction

Morphological analysis of human organs based on Magnetic Resonance Imaging (MRI) or Computed Tomography (CT) scans is an important field in medical imaging. An example of clinical application is the identification and quantification of the

effect of a pathology on the anatomy of the brain (i.e. hippocampal atrophy in the Alzheimer's disease (Gerardin et al., 2009)) or of the heart (i.e. Tetralogy of Fallot (Mansi et al., 2011), stroke or myocardial ischemia). This provides precious insights into the pathological development giving the possibility to identify biomarkers of disease progression or quantitative rules for disease evaluation, therapy planning and monitoring.

The analysis can be carried out either directly on images or on selected structures which are previously segmented as 3D meshes. The second strategy has the drawback that it depends on an accurate segmentation, which can be challenging in some cases, but it has also some advantages. First, it permits to focus only on certain structures belonging to a particular area of an organ (i.e. basal ganglia in patients subject to Gilles de la Tourette syndrome). Second, it allows to visualise and directly analyse the entire 3D geometry of a structure without scrolling through a stack of images. And third, it gives the possibility to combine different imaging modalities since one can analyse together meshes obtained from structural MRI images, CT scans and diffusion MRI images. Furthermore, the two strategies can be combined together in an iconic-geometric setting (Siless et al., 2012; Graciano Fouquier et al., 2014).

The structures of an organ can be modelled as 3D surfaces or 3D curves. In the brain, surfaces are used as models of sub-cortical nuclei or cortical surface (Durrleman et al., 2009; Auzias et al., 2011). In the heart they are used as models of the left, right ventricle or of the entire myocardium (Mansi et al., 2011; Palit et al., 2015; Lombaert and Peyrat, 2013). Lungs and liver can also be modelled as 3D surfaces (Gorbunova et al., 2010). 3D curves are employed in the brain to model trajectories of ensembles of neural tracts (O'Donnell et al., 2009; Durrleman et al., 2011a) which are commonly called fiber bundles. In the heart they can be used as models of cardiac myofibers (Palit et al., 2015) whereas in the lungs or liver they may be employed to model blood vessels (Gorbunova et al., 2010). The geometrical representation of an entire organ may thus combine both surface and curve meshes into a single multi-object complex, which we call *shape complex*.

There are several examples of mesh based morphometry in the literature. Most of them use a single-object approach since they select and analyse only one particular structure of an organ (Golland et al., 2005; Niethammer et al., 2007; Davies et al., 2010; Hufnagel et al., 2009; Kurtek et al., 2011; Savadjiev et al., 2012; Cury et al., 2015). This strategy limits the extent of the clinical conclusion to the chosen object, thus neglecting the information given by the surrounding structures. This approach does not seem appropriate to study intricate and complex systems of interconnected objects such as the human organs. There is a growing awareness that multi-object studies could be more befitting. Some examples are Gorczowski et al. (2010); Cates et al. (2008); Bossa et al. (2011); Qiu et al. (2010); Durrleman et al. (2014) where the authors have shown the importance of the information about the relative position between separated structures to discriminate controls from patients using sub-cortical nuclei of the brain.

In this chapter, we focus on a general morphometric approach called atlas con-

struction. It consists of estimating an average shape complex of a population called template complex and the deformations of the embedding space which warp the *template complex* to the shape complexes of every subject. Deformations put into correspondence the template with every shape complex. The template captures the common morphological characteristics of the population and the deformations capture the variations in shape across the population. Every deformation is a single 3D diffeomorphism (a smooth deformation with smooth inverse) which transforms the entire 3D space. All meshes are therefore deformed together whatever their number or kind. This kind of deformation preserves the spatial organization of the structures since neither intersection, folding or shearing may occur. This makes thus possible to study a human organ as a whole and not as a set of independent components, analysing not only the shape of each structure, but also their relative position.

The deformations transform the template complex so that each one of its structures superimposes with the homologous one of the subject shape complex. To this end, one needs to define a similarity measure for every structure. Different metrics have been proposed for surfaces and curves. Some of them assume that it is possible to find correspondences between points of homologous structures, called landmarks. This is usually hard (or even impossible) especially when working with groups of curves. This is why we have opted for a metric based on the framework of *varifold* (Charon and Trouvé, 2013b) which can be applied to both surfaces and curves. It does not require point-correspondence between structures or curve-correspondence between ensemble of curves. Moreover, surfaces and curves are treated as instances of the same mathematical object. This simplifies the atlas construction and the following statistical analysis.

A standard procedure to estimate a multi-object atlas (Durrleman et al., 2014; Avants and Gee, 2004; Ma et al., 2010) involves an optimization scheme where all template structures and deformations are optimised together by minimizing a single cost function. Every structure is weighted by a scalar value which balances the importance of the structure with respect to the others and to a regularity term on the deformations. These weights are fixed by the user and the results are rather sensitive to them. Optimal values should be determined by cross-validation which is time consuming and computationally intensive. Moreover, the computational load increases with the number of structures and subjects under study, making sometimes unfeasible the automatic estimate of the weights via cross-validation.

Another strategy adopted by different authors is to embed the atlas procedure into a statistical setting. Using a generative model, every shape can be interpreted as the sum between a deformation of the template and a noise (Allassonnière et al., 2007). Moreover, one considers the deformation parameters, template and noise as random variables. This makes possible to use Maximum Likelihood (ML) or Maximum A Posteriori (MAP) estimations to infer the atlas. If the noise is modelled as a Gaussian distribution, it turns out that the weight of a structure in the standard optimisation procedure can be interpreted as the variance of the noise (Allassonnière et al., 2007; Zhang et al., 2013; Simpson et al., 2012; Folgoc et al., 2014). This can be

automatically estimated with ML or MAP techniques. Other authors also proposed to infer from the data also the level of regularisation by modelling it as the variance of the deformation parameters or deformation field (Simpson et al., 2012; Folgoc et al., 2014; Risholm et al., 2013; Wassermann et al., 2014). However, most of these works are about registration and atlas construction of *images*. Few of them focus on *shapes* and they are usually based on single-object complexes and the variance of the data noise as well as the regularity level are usually fixed by the user. Three examples are Ma et al. (2010) where the authors introduced a surface single-object diffeomorphic atlas procedure using a hyper-template and Durrleman et al. (2014, 2011a) where the authors employed a generative model adapted for complexes of only surfaces or only curves respectively.

Here we propose to embed a diffeomorphic multi-object atlas construction into a Bayesian framework. Shape complexes can be composed of curves and surfaces and they are both modelled as varifolds. The noise of each structure is assumed to be a random Gaussian varifold which belongs to an infinite dimensional space. We define a finite-dimensional approximation space where the noise is projected to and where it is possible to define probability distributions. Furthermore, we use a prior distribution on the variance of the noise as in Allassonnière et al. (2007) adapted for shape complexes modelled as varifolds. This allows us to automatically estimate the balancing weights in one single atlas construction. Moreover, we assume that also the deformation parameters follow a Gaussian distribution. We use a prior distribution on their covariance matrix similarly to Allassonnière et al. (2007) which allows us to estimate it automatically even with a small sample size. Furthermore, the estimated covariance matrix is always well-conditioned whatever the number of subjects and it can be directly employed to compute statistics about the population under study.

In addition to that, we also take advantage of the proposed Bayesian framework to define a multi-population atlas construction. We use the information given by the clinical diagnosis of the subjects to subdivide the data-set in different groups. Their deformation parameters are then modelled as different Gaussian distributions. Since all sets of deformation parameters are defined on the same space, we can quantitatively compare them and the moments of their distributions such as the covariance matrices. The proposed Bayesian framework permits us to perform this multi-population analysis at no additional cost, namely without tuning any new balancing weight.

The chapter is organised as follows. In Section 2.2, we first formulate the proposed Bayesian framework in a rather general way, showing that it could be employed with different shape models and deformation settings. We then present how to model multi-object shape complexes with varifolds and how to define a varifold random variable. Afterwards, we show the diffeomorphic framework used in the experiments and how to integrate it in our statistical setting. Then, we present the gradient-descent scheme used to optimise the cost function and how to initialise the template of both surfaces and curves. In the last part of Section 2.2, we describe how to compute a multi-population atlas construction. Eventually, we test

the robustness of our algorithm with respect to the hyperparameters of the prior distributions and we show also how the proposed method can be employed to assess morphological differences between two groups of subjects.

2.2 Methods

2.2.1 Bayesian atlas construction

We define the mesh of structure j belonging to subject i as S_{ij}. The total number of subjects is N and the number of analysed structures is M. Every subject shape complex $\boldsymbol{S}_i$, defined for the moment in a generic way as the ensemble of all the meshes S_{ij} of subject i, is modelled as a deformed template complex $\phi_i(\boldsymbol{T})$ plus a residual $\boldsymbol{\varepsilon}_i$. Both $\boldsymbol{T}$ and $\boldsymbol{\varepsilon}_i$ are also defined as the ensembles of the templates T_j and residuals ε_{ij}. This formulation is also known as forward model where we assume that all elements belong to an algebraic structure where addition is defined (Durrleman et al., 2011b; Allassonnière et al., 2007; Ma et al., 2008). It writes:

$$\boldsymbol{S}_i = \phi_i(\boldsymbol{T}) + \boldsymbol{\varepsilon}_i \tag{2.1}$$

The deformation ϕ_i, proper to subject i, can belong to any diffeomorphic framework present in the literature. The only requirement, for the scope of this chapter, is that it deforms the entire ambient space, namely all the structures of the template simultaneously. The goal of the atlas construction is to estimate the template and its morphological variations within the population of shape complexes. The variations are described by the ensemble of deformations $\{\phi_i\}$ and each one of them is parametrised by a set of parameters $\boldsymbol{\alpha}_i \in \mathbb{R}^q$. We assume that these parameters follow a Gaussian distribution with zero mean and covariance matrix Γ_α:

$$\boldsymbol{\alpha}_i \sim N(0, \Gamma_\alpha) \qquad p(\boldsymbol{\alpha}_i|\Gamma_\alpha) \propto \frac{1}{|\Gamma_\alpha|^{1/2}} \exp\left[-\frac{1}{2}\boldsymbol{\alpha}_i^T \Gamma_\alpha^{-1} \boldsymbol{\alpha}_i\right] \tag{2.2}$$

From Eq.2.2 we can notice that the distribution of each set of deformation parameters $\boldsymbol{\alpha}_i$ is completely described by Γ_α and therefore we can rephrase our goal as estimating the template $\boldsymbol{T}$ and the covariance matrix Γ_α knowing the shape complexes of the population $\{\boldsymbol{S}_i\}$, or more formally maximizing their joint posterior distribution:

$$\{\boldsymbol{T}^*, \Gamma_\alpha^*\} = \arg\max_{\boldsymbol{T}, \Gamma_\alpha} p(\boldsymbol{T}, \Gamma_\alpha | \{\boldsymbol{S}_i\}) \tag{2.3}$$

This maximization is constrained by the fact that the deformed template $\phi_i(\boldsymbol{T})$ should resemble to the shape complex $\boldsymbol{S}_i$ or, in other words, that the residual $\boldsymbol{\varepsilon}_i$ should be small. This is a common problem in statistical learning and it is usually tackled by assuming that the residual follows a Gaussian distribution centred at 0. The maximization of its likelihood is equivalent to minimize its squared norm. We will start by modelling the structures of the shape complexes $\boldsymbol{S}_i$ and of the template

complex $\boldsymbol{T}$ with landmarks. Every structure j has a number of landmarks equal to Λ_j and the norm of the difference between two meshes of the same structure is defined as the square root of the sum of squared differences between pair of landmarks (L^2-norm, $||\cdot||_2$). The likelihood of the residuals modelled as landmarks is thus defined as:

$$\varepsilon_{ij} \sim N(0, \sigma_j^2 Id_{\Lambda_j}) \qquad p(\varepsilon_{ij}|\sigma_j^2) \propto \frac{1}{|\sigma_j^2|^{\Lambda_j/2}} \exp\left[-\frac{1}{2\sigma_j^2}||S_{ij} - \phi_i(T_j)||_2^2\right] \tag{2.4}$$

In section 2.2.2 we will make clear how to adapt this equation to varifolds by changing only the norm $||\cdot||$ and the definition of the parameter Λ_j. The variance of the noise of structure j depends only on the parameter σ_j^2 which is structure-dependent and it can also be modelled as a random variable.

It is important to notice that in Eq.2.1 a value of $\boldsymbol{\alpha}_i$ equal to zero means no deformation. The mean of $\boldsymbol{\alpha}_i$ is set to zero because the template is supposed to represent the average of the shape complexes (i.e. $S_{ij} \sim N(\phi_i(T_j), \sigma_j^2 Id_{\Lambda_j})$). In other words, using the terminology of mixed models, the deformation parameters $\boldsymbol{\alpha}_i$ are the random effects associated to the fixed effect $\boldsymbol{T}$ which is therefore the expected value of the shape complexes.

Assuming independence between all random variables and considering $\boldsymbol{\alpha}_i$ as an unobserved nuisance variable we can rewrite Eq.2.3 as:

$$\{\boldsymbol{T}^*, \Gamma_\alpha^*, \sigma_j^{2*}\} = \underset{\boldsymbol{T},\Gamma_\alpha,\sigma_j^2}{\arg\max} \left[\prod_i^N \prod_j^M \int p(T_j, \Gamma_\alpha, \sigma_j^2, \boldsymbol{\alpha}_i, S_{ij}) d\boldsymbol{\alpha}_i\right] \tag{2.5}$$

The maximization of Eq.2.5 is not tractable analytically. A possible solution is to employ an iterative algorithm like the EM (Expectation Maximization) and to approximate the conditional distribution of the E step with a Dirac distribution at its mode. Let $\boldsymbol{\Theta} = \{\Gamma_\alpha, \boldsymbol{T}, \{\sigma_j^2\}\}$ be the parameters of interest and $\boldsymbol{Z} = \{\{\boldsymbol{\alpha}_i\}\}$ the unobserved nuisance variables, it results:

$$\{\boldsymbol{\Theta}^*, \boldsymbol{Z}^*\} = \underset{\boldsymbol{\Theta},\boldsymbol{Z}}{\arg\min} - \log\left[p(\{\boldsymbol{S}_i\}/\boldsymbol{Z}, \boldsymbol{\Theta})\right] - \log\left[p(\boldsymbol{Z}/\boldsymbol{\Theta})\right] - \log\left[p(\boldsymbol{\Theta})\right] \tag{2.6}$$

This algorithm is an approximation of the MAP estimator which does not converge when applied to a dataset with low signal-to-noise ratio (SNR), as demonstrated in Allassonnière et al. (2007). In order to ensure the convergence of the iterative estimation scheme one may use sampling algorithms like MCMC (Markov Chain Monte Carlo) (Allassonnière et al., 2010). Unfortunately, these methods require a great computational load and execution time.

Not using priors for σ_j^2 and Γ_α is equivalent to a ML estimation and it can produce degenerate estimates. Instead, as demonstrated in Allassonnière et al. (2007), the introduction of inverse Wishart distributions makes possible to obtain good es-

timates even when the number of subjects is small. They regularize the estimates of σ_j^2 and Γ_α. As in Allassonnière et al. (2007), the probability density functions employed here are adapted versions of the standard inverse Wishart distributions:

$$\sigma_j^2 \sim \mathcal{W}^{-1}(P_j, w_j) \qquad p(\sigma_j^2; P_j, w_j) \propto (\sigma_j^2)^{-\frac{w_j}{2}} \exp\left[-\frac{1}{2}\frac{w_j P_j}{\sigma_j^2}\right] \tag{2.7}$$

$$\Gamma_\alpha \sim \mathcal{W}^{-1}(P_\alpha, w_\alpha) \qquad p(\Gamma_\alpha; P_\alpha, w_\alpha) \propto |\Gamma_\alpha|^{-\frac{w_\alpha}{2}} \exp\left[-\frac{1}{2} w_\alpha Tr(P_\alpha^T \Gamma_\alpha^{-1})\right] \tag{2.8}$$

The scalars w_j, P_j and w_α are strictly positive and P_α is a positive symmetric matrix. Using these priors for σ_j^2 and Γ_α and assuming that the template $\boldsymbol{T}$ has a non-informative prior distribution, Eq.2.6 becomes:

$$\boxed{\sum_{j=1}^{M}\sum_{i=1}^{N}\frac{1}{2\sigma_j^2}\left(||S_{ij}-\phi_i(T_j)||_2^2+\frac{P_j w_j}{N}\right)}+\sum_{j=1}^{M}\frac{1}{2}(w_j+\Lambda_j N)\log(\sigma_j^2)+$$
$$\boxed{\frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{\alpha}_i)^T(\Gamma_\alpha)^{-1}\boldsymbol{\alpha}_i}+\frac{1}{2}(w_\alpha+N)\log(|\Gamma_\alpha|)+\frac{w_\alpha}{2}tr((\Gamma_\alpha)^{-1}P_\alpha) \tag{2.9}$$

The framed terms refer respectively to the data-terms and to the regularity of the deformations. If we fix both σ_j^2 and Γ_α and we set $w_j = 0$, their sum is the cost function minimised in standard atlas construction procedures (Avants and Gee, 2004; Durrleman et al., 2014) where it is not employed a statistical setting. This minimisation is often referred to as the computation of the Fréchet mean. It can be noticed that σ_j^2 weights the contribution of structure j as mentioned in the Introduction.

The proposed statistical framework is general and it could be applied to any shape model and diffeomorphic setting provided that one defines a probability density function for the noise and a parametric deformation model. Modelling shapes with landmarks simplifies the computation of the metric and of the gradient and it eases also the definition of random variables. Unfortunately, the choice of corresponding points between homologous structures of different subjects is not always an easy task and in some cases, like for fiber bundles in the brain, it is almost impossible. It is better to opt for a correspondence-free metric like the one based on the framework of varifolds (Charon and Trouvé, 2013b).

2.2.2 Varifolds

The framework of varifolds is an extension of the one of currents (Vaillant and Glaunès, 2005) and it can be used for both surfaces and curves. It does not require point-correspondence between structures or curve-correspondence between ensemble

of tracts. This means that the same structure may have a different number of vertices among subjects. It is robust to mesh imperfections such as holes, spikes or tract interruptions. Moreover, it does not require a consistent orientation of the normal or tangent vectors among the population since the distance is invariant to a change of orientation of some normals or tangent vectors. Another important characteristic is that varifolds prevent what is known as the "currents cancelling effect". This happens when two surface/curve elements with opposite orientation cancel reciprocally their effect on the estimation of the template. For a more detailed discussion about varifolds the user is referred to Durrleman et al. (2014); Charon and Trouvé (2013b).

Let X and Y be two curves (oriented or unoriented rectifiable curves), they can be modelled as varifolds based on:

$$V_X(\omega) = \int_X \omega(x, \overleftrightarrow{\beta(x)})|\beta(x)|_2 dx \qquad V_Y(\omega) = \int_Y \omega(y, \overleftrightarrow{\gamma(y)})|\gamma(y)|_2 dy \tag{2.10}$$

where $V_X(\omega)$ and $V_Y(\omega)$ are scalars, $\beta(p)$ and $\gamma(y)$ are the tangents of X and Y at the point x and y respectively and $\overleftrightarrow{\beta_p}$ is the unoriented unit vector with the same direction of β_p (respectively $\overleftrightarrow{\gamma_q}$ for γ_q). The set of unoriented unit vectors $\overleftrightarrow{\beta} \in \overleftrightarrow{\mathbb{S}}$ can be formally defined as the quotient of the unit sphere in $\mathbb{R}^3$ by the two elements group $\{\pm \mathrm{Id}_3\}$. Instead, the test field $\omega \in W$ is a function of both the points in the ambient space $\mathbb{R}^3$ and the unoriented unit tangent vectors in $\overleftrightarrow{\mathbb{S}}$.

The curves X and Y can be seen as polygonal lines of P and Q segments respectively. Every segment of X is completely described by its center point $x_p \in \mathbb{R}^3$ and tangent vector $\beta_p \in \mathbb{R}^3$ centred at x_p (respectively $y_q \in \mathbb{R}^3$ and $\gamma_q \in \mathbb{R}^3$ for Y). Let c_p be the length of the tangent vector β_p (respectively d_q for γ_q), one can approximate Eq.2.10 with: $V_X(\omega) \approx \sum_{p=1}^{P} \omega(x_p, \overleftrightarrow{\beta_p})c_p$ and $V_Y(\omega) \approx \sum_{q=1}^{Q} \omega(y_q, \overleftrightarrow{\gamma_q})d_q$. It is important to notice that, even if this approximation seems reasonable, it is still an open question how to ensure its convergence when the number of segments tends to infinite, i.e. with a more accurate sampling of the curve (Charlier et al., 2014).

Furthermore, Eq.2.10 shows that the space of varifolds W^* is a linear functional from W to $\mathbb{R}$. This makes W^* a vector space and therefore the union of meshes is equal to a sum in the space of varifolds. But, contrary to currents, inverting the orientation of a mesh does not change the varifold representation. One chooses W to be a separable RKHS on $\mathbb{R}^3 \mathrm{x} \overleftrightarrow{\mathbb{S}}$ in order to have an explicit definition of distance between two meshes modelled as varifolds. More precisely, since W is a product space, one defines its kernel K_W as the tensor product between two kernels k_x, Gaussian, and k_β, Cauchy-Binet, defined on $\mathbb{R}^3$ and on $\overleftrightarrow{\mathbb{S}}$ respectively. Thus, the inner product between X and Y modelled as varidolds is defined as:

$$< V_X, V_Y >_{W^*} = \sum_{p=1}^{P}\sum_{q=1}^{Q} k_x(x_p, y_q) k_\beta(\overleftrightarrow{\beta_p}, \overleftrightarrow{\gamma_q}) c_p d_q$$
$$= \sum_{p=1}^{P}\sum_{q=1}^{Q} \exp\left(\frac{-||x_p - y_q||^2}{\lambda_W^2}\right)\left(\frac{\beta_p^T \gamma_q}{c_p d_q}\right)^2 c_p d_q \qquad (2.11)$$

The distance between V_X and V_Y is therefore: $||V_X - V_Y||^2_{W^*} = < V_X, V_X >_{W^*} + < V_Y, V_Y >_{W^*}$ -2 $< V_X, V_Y >_{W^*}$. This metric is completely parametrized by the standard deviation λ_W of the Gaussian kernel k_X which is fixed by the user. This framework can be easily extended also to surfaces by replacing x_p (y_q) and β_p (γ_q) with the centres and normals of the faces of the surface, as shown in Chapter 4 and in Charon and Trouvé (2013b). The computation of the gradient with respect to the points of the mesh can be found in Charon and Trouvé (2013b); Durrleman et al. (2014).

Varifolds random variable In Eq.2.4 we defined the residual $\boldsymbol{\varepsilon}$ as a multivariate Gaussian random variable on the finite-dimensional Euclidean space of landmarks. Here we extend this definition to the framework of varifolds following the same line of reasoning presented in Durrleman (2010) for the framework of currents. We define a random Gaussian varifold as a linear map between every vector field $\omega \in W$ to a real random Gaussian variable $G(\omega)$ such that, given two vector fields ω_1 and ω_2, $E[G(\omega_1)] = 0$ and $E[G(\omega_1), G(\omega_2)] =< \omega_1, \omega_2 >_W$. This shows that the kernel K_W of the space W completely defines the covariance matrix of the Gaussian varifold. However, since it is infinite-dimensional, it has no probability density function and therefore we can not simulate instances of it. To tackle this problem, we define for each structure j a finite-dimensional space $W^*_{\Lambda_j}$ on which we project the template T_j and all the shapes $\{S_{ij}\}_{i=1\ldots N}$ modelled as varifolds. This finite dimensional space $W^*_{\Lambda_j}$ is defined as the span of a set of delta Dirac varifolds: $\text{Span}\{\delta_{(x_u, \overleftrightarrow{\beta_k})}\}$ where both the points $\{x_u\}$ and the unoriented unit vectors $\{\overleftrightarrow{\beta_k}\}$ are constrained to belong to two predefined grids, respectively Υ_x and Υ_β. The first one is a linearly spaced grid in the ambient space and Υ_β is a regular sampling of the half unit sphere in $\mathbb{R}^3$. The number of points of the two grids is respectively Λ_x^j and Λ_β and Λ_j is their product where the index j refers to the j-th structure.

For every finite-dimensional space $W^*_{\Lambda_j}$, it is possible to define a block matrix $\boldsymbol{K}_{W,\Lambda_j}$ whose blocks are the RKHS kernel $K_{W,\Lambda_j}\left((x_u, \overleftrightarrow{\beta_k}), (\cdot,\cdot)\right)_{\substack{u=1,\ldots,\Lambda_x^j \\ k=1,\ldots,\Lambda_\beta}}$ between every possible combination of the couples $\{x_u, \overleftrightarrow{\beta_k}\}$. The matrix $\boldsymbol{K}_{W,\Lambda_j}$ has a dimension of $[\Lambda_j \text{x} \Lambda_j]$. To project a delta Dirac varifold $\delta_{(y,\overleftrightarrow{\alpha})}|\alpha|$ onto $W^*_{\Lambda_j}$, using for instance the closest neighbour projection, we look for the closest point to y of the grid Υ_x and for the closest direction to $\overleftrightarrow{\alpha}$ among the ones given by Υ_β and we assign the scalar $|\alpha|$ to that particular couple of grid points. The projection is

therefore completely parametrised by the vector $\{|\alpha|\}$ of size Λ_j. Using this scheme, the squared norm of a projected varifold $||\sum_{l=1}^{L}\delta_{(y_l,\overleftrightarrow{\alpha_l}}|\alpha_l|)||^2_{W^*_{\Lambda j}}$ is equal to $\sum_{u\in\Upsilon_x}\sum_{p\in\Upsilon_x}\sum_{k\in\Upsilon_\beta}\sum_{q\in\Upsilon_\beta} c_{uk}K_{W,\Lambda_j}\left((x_u,\overleftrightarrow{\beta_k}),(x_p,\overleftrightarrow{\beta_q})\right)c_{pq}$ where the scalars c_{uk} and c_{pq} refer to the values obtained at the end of the projection of the varifold in $\{x_u,\overleftrightarrow{\beta_k}\}$ and $\{x_p,\overleftrightarrow{\beta_q}\}$ respectively.

Scaling the kernel $\boldsymbol{K}_{W,\Lambda_j}$ with a scalar σ_j^2, we can define the likelihood of the residuals modelled as varifolds in the following way:

$$p(\varepsilon_{ij}|\sigma_j^2) \propto \frac{1}{|\sigma_j^2|^{\Lambda_j/2}}\exp\left[-\frac{1}{2\sigma_j^2}||S_{ij}-\phi_i(T_j)||^2_{W^*_{\Lambda j}}\right] \tag{2.12}$$

where it is important to notice that $\boldsymbol{K}_{W,\Lambda_j}$ is a constant matrix that is computed only once at the beginning of the atlas construction. This equation has the same formulation of the likelihood of the residuals modelled as landmarks in Eq.2.4. The only two differences are the definitions of the norm and of the parameter Λ_j. These are also the only changes to apply to the cost function in Eq.2.9 when using the framework of varifolds instead than landmarks.

In practice, due to the important computational load and time required to calculate $||\cdot||^2_{W^*_{\Lambda j}}$, we choose to directly compute the exact expression of the norm $||(\cdot)||^2_{W^*}$ using a fast GPU (CUDA) implementation. Moreover, we define Υ_x^j as a regular linearly-spaced grid containing both shapes and template of structure j and where the distance between every couple of points is given by λ_W. The resulting number of points Λ_x^j is used to compute $\Lambda_j{=}\Lambda_x^j$. In this way Λ_j depends only on λ_W which is fixed by the user and it is related to the distance at which two shapes can be compared and at the same time to the degree of detail of the anatomical features one wants to detect. In Sec.2.3.3 we show that the number of grid points does not affect the results above a certain threshold and that the proposed choice seems a good one.

2.2.3 Diffeomorphic Transformations

We define here how to compute the diffeomorphic deformations of the template complex. Our approach relies on the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework based on the control point formulation presented in Durrleman et al. (2011b). Template transformations are built by integrating a time-varying vector field $v_t(x)$ over $t\in[0,1]$ where $v_t(x)$ represents the instantaneous velocity of every point x belonging to the ambient space at time t. Calling $\phi_t(x)$ the position of a point at time t which was located in x at time $t=0$, its evolution is given by: $\frac{\partial\phi_t(x)}{\partial t}=v_t(\phi_t(x))$ with $\phi_0(x)=x$, which produces a flow of deformations $\{\phi_t\}_{t\in[0,1]}$. Furthermore, v_t belongs to a Reproducible Kernel Hilbert Space (RKHS) with Gaussian kernel K_V. In Glaunès (2005) it is shown that if such a vector field is square integrable in the interval [0,1] then every deformation of the flow $\{\phi_t\}_{t\in[0,1]}$ is a diffeomorphism. Here we define v_t using a dynamical system of C_p control points

$\boldsymbol{c}=\{c_k \in \mathbb{R}^3\}$ scattered in the ambient space and a set of time-varying vectors called momenta $\boldsymbol{\alpha}=\{\alpha_k \in \mathbb{R}^3\}$ linked to each control point. This means that the velocity of every point of the space is given by:

$$\dot{x}(t) = v_t(\phi_t(x)) = \sum_{p=1}^{C_p} K_V(x(t), c_p(t))\alpha_p(t) \qquad \phi_0(x) = x(0) = x \tag{2.13}$$

where $\phi_t(x)=x(t)$. With this definition every ϕ_t is a diffeomorphism provided that all $\boldsymbol{\alpha}(t)$ are square integrable for every t. The evolution of a point x depends only on its initial position $x(0) = x_0$ and on the evolution of the system $\boldsymbol{L}(t) = \{\boldsymbol{c}(t), \boldsymbol{\alpha}(t)\}$. In an atlas construction, the deformations of the template complex are the last ones (ϕ_1) of a flow of diffeomorphisms. Among all the possible paths connecting ϕ_0 to ϕ_1 we use the geodesic one, which means the one that minimizes the total kinetic energy along the path:

$$\int_0^1 ||v_t||_V^2 dt = \int_0^1 \sum_{k=1}^{C_p}\sum_{p=1}^{C_p} \alpha_k(t)^T K_V(c_k(t), c_p(t))\alpha_p(t)dt \tag{2.14}$$

It has been demonstrated in Durrleman et al. (2014) that the extremal paths are such that the system $\boldsymbol{L}(t)$ satisfies:

$$\begin{cases} \dot{c}_k(t) = \sum_{p=1}^{C_p} K(c_k(t), c_p(t))\alpha_p(t) \\ \dot{\alpha}_k(t) = -\sum_{p=1}^{C_p} \alpha_k(t)^T \alpha_p(t) \nabla_1 K(c_k(t), c_p(t)) \end{cases} \tag{2.15}$$

Denoting $\boldsymbol{L}(0)=\boldsymbol{L}_0=\{\boldsymbol{c}(0), \boldsymbol{\alpha}(0)\}$ the initial condition of the system, the previous set of ODEs can be rewritten as: $\dot{\boldsymbol{L}}(t) = F[\boldsymbol{L}(t)]$ with $\boldsymbol{L}(0) = \boldsymbol{L}_0$. This equation shows that the entire flow of diffeomorphisms, and more precisely the last one, is completely determined by the initial state of the system $\boldsymbol{L}_0$. Thus, given $\boldsymbol{c}(0)$ and $\boldsymbol{\alpha}(0)$, one first integrates Eq.2.15 obtaining the values of $\boldsymbol{c}(t)$ and $\boldsymbol{\alpha}(t)$ at any time t. Then, every point x in the ambient space is deformed by integrating Eq. 2.13.

In geometrical term, the value $\boldsymbol{L}_0$ defines the initial velocity of the geodesic path in the tangent space of our group of diffeomorphisms at identity. The parameters $\{\boldsymbol{c}(0), \boldsymbol{\alpha}(0)\}$ can be used to perform the so-called tangent-space statistics (Vaillant et al., 2004).

2.2.4 Optimization procedure

This diffeomorphic setting can be easily included into the proposed Bayesian framework. We choose to employ a unique set of control points $\boldsymbol{c}$ for the whole population and N subject-specific sets of momenta $\{\boldsymbol{\alpha}_i\}_{i=1,\dots,N}$. The set of initial control points and momenta $\{\boldsymbol{c}(0), \{\boldsymbol{\alpha}_i(0)\}_{i=1,\dots,N}\} = \{\boldsymbol{c}_0, \{\boldsymbol{\alpha}_{i0}\}_{i=1,\dots,N}\}$ represents the deformation parameters to warp the template complex towards the shape complex of subject

i. Thus, the hidden variables in Eq.2.6 become: $\boldsymbol{Z} = \{\{\boldsymbol{\alpha}_{i0}\}_{i=1,\dots,N}, \boldsymbol{c}_0\}$. Assuming that all random variables are independent and that the control points $\boldsymbol{c}_0$ have a non-informative prior distribution, the expression of the cost function in Eq.2.9 does not change except for the second framed term which becomes $\frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{\alpha}_{i0})^T(\Gamma_\alpha)^{-1}\boldsymbol{\alpha}_{i0}$.

The parameters $\boldsymbol{T}, \{\boldsymbol{\alpha}_0^i\}$ and $\boldsymbol{c}_0$ are minimised using an accelerated version of the line search gradient descent method based on the Nesterov's scheme (Nesterov, 1983). Instead, the use of conjugate priors makes possible to compute the optimal values for $\{\sigma_j^2\}$ and Γ_α in a closed form:

$$\hat{\Gamma}_\alpha = \frac{\sum_{i=1}^{N}\left[(\boldsymbol{\alpha}_{i0})(\boldsymbol{\alpha}_{i0})^T\right] + w_\alpha P_\alpha^T}{(w_\alpha + N)} \quad \hat{\sigma}_j^2 = \frac{\sum_{i=1}^{N}||S_{ij} - \phi_i(T_j)||^2_{W^*_{\Lambda j}} + w_j P_j}{(w_j + N\Lambda_j)} \tag{2.16}$$

The first parameter $\hat{\Gamma}_\alpha$ is equal to a weighted sum between the sample covariance matrix of the initial momenta $\boldsymbol{\alpha}_{i0}$ and the prior P_α. A good choice for the prior seems to be: $P_\alpha = K_V^{-1}$, where K_V is a block matrix whose blocks are 3D Gaussian kernels between two different control points. It is exactly the kernel of the RKHS to which belongs the velocity field v_t building the diffeomorphism ϕ_1. If the number of subjects N is small with respect to w_α then $\hat{\Gamma}_\alpha \propto K_V^{-1}$ and this means that the deformation regularity part in Eq.2.9 becomes $\sum_{i=1}^{N}(\boldsymbol{\alpha}_{i0})^T K_V \boldsymbol{\alpha}_{i0}$, which is the sum of the geodesic distances between the template complex and all its transformations. This is exactly the deformation regularity term used in previous atlas constructions not based on a statistical setting (Durrleman et al., 2014). Moreover, the use of this prior makes possible the inversion of Γ_α even when the number of subjects N is smaller than the number of deformation parameters.

The second parameter $\hat{\sigma}_j^2$ is equal to a weighted sum between the data-term of the j-th structure and the prior. This term balances the contribution of the different structures to the cost function (Eq.2.9) and at the same time it is a trade-off between the data-terms and the the other parts of the cost function. The automatic estimate of this parameter is rather important as shown in Sec.2.3.2. The use of an inverse Wishart prior helps to avoid over-fitting. In fact, using a maximum likelihood estimator, the minimization process might focus only on a structure k reducing its residuals almost to zero and consequently $\log(\sigma_k^2) \rightarrow$-$\infty$. This could also result in almost ignoring all the other structures. The role played by the prior is to impose a minimum value to σ_k^2 in order to avoid such a situation.

It is important to notice that the closed-form solutions for both $\hat{\Gamma}_\alpha$ and $\hat{\sigma}_j^2$ in Eq.2.16 would be different if using the standard Inverse Wishart distributions instead than the ones proposed in Eq.2.7 and Eq.2.8. They would not be equal to a weighted average between prior and empirical estimates, making therefore more difficult their interpretation and use.

The gradients of the cost function E (Eq.2.9) with respect to $\boldsymbol{T}$,$\{\boldsymbol{\alpha}_0^i\}$ and $\boldsymbol{c}_0$ are respectively equal to:

$$\begin{aligned}
\nabla_{T_k} E &= \sum_{i=1}^{N} \frac{1}{2\sigma_k^2} \nabla_{T_k}[D_{ik}] \\
\nabla_{\boldsymbol{\alpha}_{s0}} E &= \sum_{j=1}^{M} \frac{1}{2\sigma_j^2} \nabla_{\boldsymbol{\alpha}_{s0}}[D_{sj}] + \Gamma_\alpha^{-1} \boldsymbol{\alpha}_{s0} \\
\nabla_{\boldsymbol{c}_0} E &= \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{1}{2\sigma_j^2} \nabla_{\boldsymbol{c}_0}[D_{ij}] + \frac{w_\alpha}{2} \nabla_{\boldsymbol{c}_0} \left[tr((\Gamma_\alpha)^{-1} P_\alpha) \right]
\end{aligned} \tag{2.17}$$

where $D_{ij}{=}||S_{ij} - \phi_i(T_j))||^2_{W^*_{\Lambda j}}$. The differentiation of the prior P_α with respect to $\boldsymbol{c}_0$ is not taken into account since its norm is negligible with respect to the one of the data term.

In order to compute the three derivatives of the data-terms D_{ij}, we need first to compute the deformed template complex $\phi_i(\boldsymbol{T})$ where the deformation ϕ_i is the final diffeomorphism of the flow generated by integrating forward in time Eq.2.15 and then Eq.2.13 which can be rewritten in matrix notation as: $\dot{\boldsymbol{T}}_i(t){=}G[\boldsymbol{T}_i(t), \boldsymbol{L}_i(t)]$ with $\boldsymbol{T}_i(0){=}\boldsymbol{T}(0){=}\boldsymbol{T}$. After that, we can compute the gradient of D_{ij} with respect to the deformed template complex $\boldsymbol{T}_i(1) = \phi_i(\boldsymbol{T})$ and then bring back this information to $t = 0$ in order to update the *initial* template points $\boldsymbol{T}$, control points $\boldsymbol{c}_0$ and momenta $\boldsymbol{\alpha}_{i0}$. This is done by integrating backward in time from $t{=}1$ to $t{=}0$ a set of linearised ODEs called adjoint equations like in Durrleman et al. (2014). It results:

$$\begin{aligned}
\nabla_{\boldsymbol{T}} E &= \sum_{i=1}^{N} \boldsymbol{\theta}_i(0) \\
\nabla_{\boldsymbol{\alpha}_{i0}} E &= \boldsymbol{\xi}_i^{\boldsymbol{\alpha}}(0) + \Gamma_\alpha^{-1} \boldsymbol{\alpha}_{i0} \\
\nabla_{\boldsymbol{c}_0} E &= \sum_{i=1}^{N} \boldsymbol{\xi}_i^{\boldsymbol{c}}(0)
\end{aligned} \tag{2.18}$$

where the auxiliary variables $\boldsymbol{\xi}_i(t) = \{\boldsymbol{\xi}_i^{\boldsymbol{\alpha}}(t), \boldsymbol{\xi}_i^{\boldsymbol{c}}(t)\}$ and $\boldsymbol{\theta}_i(t)$ satisfy the linearised ODEs:

$$\dot{\boldsymbol{\theta}}_i(t) = -\left(\partial_{\boldsymbol{T}_i} G[\boldsymbol{T}_i(t), \boldsymbol{L}_i(t)]\right)^T \boldsymbol{\theta}_i(t) \qquad \boldsymbol{\theta}_i(1) = \frac{1}{2\boldsymbol{\sigma}^2} \nabla_{\boldsymbol{T}_i(1)}[\boldsymbol{D}_i] \tag{2.19}$$

$$\dot{\boldsymbol{\xi}}_i(t) = -\left(\partial_{\boldsymbol{L}_i} G[\boldsymbol{T}_i(t), \boldsymbol{L}_i(t)]\right)^T \boldsymbol{\theta}_i(t) + d_{\boldsymbol{L}_i} F[\boldsymbol{L}_i(t)]^T \boldsymbol{\xi}_i(t) \qquad \boldsymbol{\xi}_i(1) = 0 \tag{2.20}$$

ODEs are integrated with the Euler's method using 10 steps. More details can be found in the Appendix of Durrleman et al. (2014).

2.2.5 Template initialisation

Since we use a gradient descent scheme to minimize the cost function, we need an initial template. This should not be the mesh of a subject otherwise it would inherit all the imperfections due to segmentation errors and data noise, biasing consequently the atlas construction.

For surfaces, we use a regular sphere which is first centred in the middle of the population and then it is scaled to an ellipsoid by using the three main modes of variability of the surfaces of the population considered as a points cloud.

For bundles of curves, we propose instead to initialise the template as the "most relevant" curves in the population.

We first gather the curves F of every subject bundle into a raw initial template $T = \cup_{h=1}^{P} F_h$, where P is the total number of curves. Then we select the set of H curves which resembles the most to the raw template T. Calling $\{F\}_k$ one of the possible $\binom{P}{H}$ combinations of H curves out of P, it results:

$$\arg\min_k ||T - \{F\}_k||^2_{W^*} \qquad k = 1, ..., \binom{P}{H} \tag{2.21}$$

which means that we choose the best combination of H curves from T. In many applications the number of curves of the raw template can be huge, thus we have developed a greedy algorithm which does not test all the possible combinations of curves but it is based on an iterative scheme. At each step, we divide T in subsets of Z curves, we look for the best one and we remove it from the raw template T saving it in a new template. This process is repeated until the new template has H curves, where H is the average number of curves in the population and Z is a submultiple of H. At each iteration, the best subset of Z curves is selected minimising:

$$\arg\min_k ||T - \sum_{h=(kZ)+1}^{(k+1)Z} F_h||^2_{W^*} \qquad k = 0, ..., (H/Z) - 1 \tag{2.22}$$

after having randomized the order of the curves in T. The new template is the initial template used in the atlas construction.

Since the number of points per curve and the number of curves in the bundles can be different among the population, we have implemented this algorithm with the metric of varifolds, but any metric could be used.

An example of template initialisation applied to three brain fiber bundles is shown in Fig.2.1. The last row represents the initial templates used in the following experiments.

2.2.6 Multi-population atlas construction

In the previous subsections we assumed that all subjects belonged to the same group. In a clinical application, a subject can be characterised by a label which can be for instance his/her clinical status. Depending on this label, subjects can be

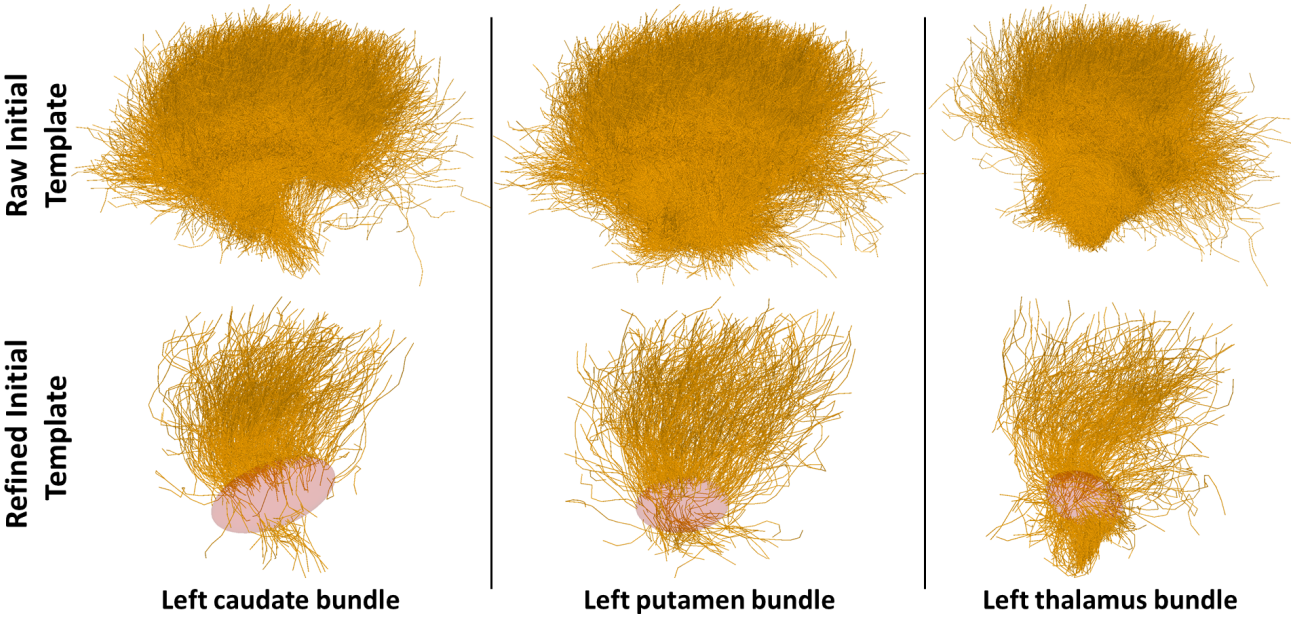

Figure 2.1: Template initialisation for three brain fiber bundles: left caudate, left putamen and left thalamus bundle. The top row contains the three raw initial templates where the fibers of every subject present in the population have been gathered together (see 2.3.1). The second row presents the initial templates obtained at the end of the proposed initialisation process cut at the intersection with their respective sub-cortical initial templates, transparent and coloured in red.

divided into different groups. Here we propose to take advantage of this information computing a multi-population atlas construction.

Let $\boldsymbol{L}$ be the set of labels that can be given to a subject shape complex $\boldsymbol{S}_i$. We assume that $\boldsymbol{L}$ has only two elements for simplicity purpose: $\boldsymbol{L} = \{c, p\}$, they can be interpreted as "control" and "patient". The extension to more than two labels is straightforward but we will not consider it here. With this notation, we can rewrite Eq.2.1 for both groups as:

$$\boldsymbol{S}_i^c = \phi_i^c(\boldsymbol{T}) + \boldsymbol{\varepsilon}_i^c \qquad \boldsymbol{S}_i^p = \phi_i^p(\boldsymbol{T}) + \boldsymbol{\varepsilon}_i^p \tag{2.23}$$

where we assume that the control points $\boldsymbol{c}$ and the template $\boldsymbol{T}$ are shared between the two populations. Instead, the initial momenta $\boldsymbol{\alpha}_{0i}$ are divided into two classes which can be modelled as distinct Gaussian distributions:

$$\boldsymbol{\alpha}_{i0}^c \sim N(\boldsymbol{\mu}^c, \Gamma_\alpha^c) \qquad \boldsymbol{\alpha}_{i0}^p \sim N(\boldsymbol{\mu}^p, \Gamma_\alpha^p) \tag{2.24}$$

We can quantitatively compare $\boldsymbol{\alpha}_{i0}^c$ with $\boldsymbol{\alpha}_{i0}^p$ since they are defined on the same set of control points and the groups share the same template. Moreover, in order to avoid identifiability issues about the template, we impose that the sum of their means $\boldsymbol{\mu}^c$

and $\boldsymbol{\mu}^p$ is equal to zero obtaining: $\boldsymbol{\mu}^c$=$\boldsymbol{\mu}$ and $\boldsymbol{\mu}^p$=-$\boldsymbol{\mu}$. This constraint forces the estimated template complex to be in between the two groups at the end of the atlas construction. Or, in other words, it pushes the estimated template complex to be at equal distance from the averages of the shapes of the two groups.

We can then assume either that the two groups share the same variability (Γ_α^c=Γ_α^p=Γ_α) or that their within-group variations are different ($\Gamma_\alpha^c \neq \Gamma_\alpha^p$). The computations for both cases can be found in 2.A.1. Using the second assumption, we could test the hypothesis of equality between the two covariance matrices or quantitatively compare their respective modes. On the contrary, if we ignored the information of the label computing a single-population atlas, we would implicitly assume that the two different groups belong to the same population and that they share the same variability. Furthermore, we could also adopt another strategy by constructing an atlas for each group and obtaining therefore two different template complexes and sets of control points. However, in this case, we could neither quantitatively compare the related populations of initial momenta nor directly compare the resulting template complexes.

The multi-population extension is computationally feasible only when using the proposed Bayesian framework. As previously shown, we use the log-likelihood of the initial momenta as regularity term of the deformations. Dividing the initial momenta into two different groups entails the addition of another regularity term into the cost function (see 2.A.1). This new term should be balanced by a trade-off parameter tuned using cross-validation. Increasing the number of groups would also augment the number of regularity terms and balancing values to tune. With the proposed statistical setting, there is no need to tune these weights since they are automatically taken into account in the estimates of the covariance matrices.

2.3 Experiments

In this section, we first present the dataset used in the following experiments. Then, we evaluate the robustness of our Bayesian framework with respect to the hyper-parameters, and we compare it with the robustness of a previous atlas procedure without automatic estimates with respect to σ_j^2. Furthermore, we also analyse how stable are the results of the Bayesian atlas construction when varying the number of points of the varifold grid Λ_j. Eventually, we present how the proposed multi-population atlas construction approach can be used to highlight morphological differences between two groups.

2.3.1 Materials

The dataset used throughout this chapter contains 40 subjects: 20 controls and 20 patients subject to Gilles de la Tourette syndrome. This neuropsychiatric disorder is thought to be associated with dysfunctions of the cortico-striato-pallido-thalamic circuits which are composed of sub-cortical structures linked to the cortical surface

by fiber bundles. We will investigate the morphological characteristics of these structures which distinguish the group of patients from the one of controls and which might be due to the Gilles de la Tourette syndrome.

For each subject, we consider three sub-cortical structures (left caudate, left putamen and left thalamus) and the fiber bundles connecting them to the left hemisphere of the cortical surface. The three sub-cortical structures are segmented with FSL (Patenaude et al., 2011) from T1-weighted images (3T). Since we want to analyse as a single structure nucleus accumbens and caudate, we merge these two segmentations together. We use the marching cubes algorithm to create the 3D meshes for all structures (AimsMesh function of Brainvisa 4.3.0). Fiber bundles come from a probabilistic tractography method using 8 seeds (Perrin et al., 2005) applied to diffusion-weighted images (50 directions, B-factor=1000) using an analytical Q-Ball model to estimate the local underlying orientation distribution function (ODF) (Descoteaux et al., 2007). Every subject bundle is downsampled by randomly selecting 10% of its fibers and it is then approximated as proposed in Durrleman et al. (2009) using a Matching Pursuit Algorithm. For more details about the acquisition or the tractography, the reader is referred to Worbe et al. (2015).

2.3.2 Robustness with respect to the hyperparameters

We evaluate here the robustness of the proposed algorithm with respect to the hyperparameter values $\{w_j, P_j, w_\alpha\}$. We do not consider P_α since, as explained in Sec.2.2.4, we fix it to $P_\alpha = K_V^{-1}$. We compute 18 different atlases changing every time only one of the hyperparameters and keeping fixed the others at a certain value. Both w_j and w_α are previously normalised in order to use the same range of values, this means: $w_\varepsilon = w'_\varepsilon N\Lambda$ and $w_\alpha = w'_\alpha N$ giving as result:

$$\hat{\Gamma}_\alpha = \frac{\frac{1}{N}\sum_{i=1}^{N}\left[(\boldsymbol{\alpha}_{i0})(\boldsymbol{\alpha}_{i0})^T\right] + w'_\alpha P_\alpha^T}{(1+w'_\alpha)} \quad \hat{\sigma}_j^2 = \frac{\frac{1}{N\Lambda_j}\sum_{i=1}^{N}||S_{ij}-\phi_i(T_j)||^2_{W^*} + w'_j P_j}{(1+w'_j)} \tag{2.25}$$

In order to understand a plausible range of testing values, we compute the residuals $\frac{1}{N\Lambda_j}\sum_{i=1}^{N}||S_{ij}-\phi_i(T_j)||^2_{W^*}$ using the aforementioned dataset and initial templates and we notice that the maximum value considering either only the sub-cortical structures or only the fiber bundles is never above 10 and 100 respectively. This means that the product $w'_j P_j$ should not surpass this limit, otherwise a reduction of the data-term would impact very little the estimate of $\hat{\sigma}_j^2$ and consequently the minimization of the cost function. Noting that, we choose as fixed value 10^{-2} and as testing range for all hyperparameters the interval $[10^{-2}$ - $10^3]$ for the sub-cortical structures and $[10^{-2}$ - $10^4]$ for the fiber bundles.

In order to test the robustness of the results we compute the norm of the difference between the resulting template complex of all 18 atlases and a reference template complex. Moreover, we analyse also the norm of the residuals obtained at the end of the atlas constructions.

To show the importance of an automatic estimate of σ_j^2 and Γ_α, we compare our results with the ones obtained using a previous atlas procedure (Fixed) where both σ_j^2 and Γ_α are fixed. In order to be coherent with the proposed Bayesian algorithm we set $\Gamma_\alpha = K_V^{-1}$, which is the value of our prior P_α and we let σ_j^2 change along the same range of values of the hyperparameters. This means building 6 other atlases changing only the value of σ_j^2. In all 24 atlas procedures, the other parameters, like λ_W and λ_V, are kept equal. Results using only left caudates or left caudate bundles of the 20 controls are shown in Fig.2.2 where we have used as reference template the one obtained using the Fixed method with σ_j^2=10^{-1}.

In Fig.2.3 we compare the robustness of the results optimizing simultaneously two structures in each atlas construction: left caudate and left caudate bundle. This means that we need to fix two σ_j^2 values with the Fixed method and 5 hyperparameters in the Bayesian framework (w_α, two w_j and two P_j). Using as range of values [10^{-2} - 10^3], the number of possible combinations of the values of the parameters for the Fixed technique is 36 whereas for the hyperparameters of the Bayesian framework is 7776. This makes infeasible an extensive and complete analysis of the robustness of the results. Nevertheless, we have decided to compute 30 atlases (15 Fixed + 15 Bayesian) picking randomly the values of the hyperparameters in the range [10^{-2} - 10^2]. The used hyperparameter and σ_j^2 values are listed in Table 2.2. The reference template is the one obtained at the end of the 9-th case using the Fixed method.

It is possible to conclude that the Fixed method leads to much more variable results than using Bayesian priors. The choice of the hyperparameters in our Bayesian framework is therefore easier than the one of σ_j^2. We decide to set them at: ω_j=$0.01\Lambda_j N$ and P_j=$\frac{0.05R_j^0}{\omega_j}$ where R_j^0 is the initial data-term of structure j. This brings to σ_j^2=$\frac{R_j+0.05R_j^0}{N\Lambda_j(1+0.01)}$ which means that the minimum value of σ_j^2 is equal to about 5% of $\frac{R_j^0}{N\Lambda_j}$. This choice reduces the risk of over-fitting since a value of R_j smaller than $\frac{0.05R_j^0}{N\Lambda_j}$ would almost not affect the estimate of σ_j^2. We also fix ω_α to 10^{-2}. This makes negligible the contribution of P_α to the estimate of $\hat{\Gamma}_\alpha$, especially when N is big, but it still permits to invert $\hat{\Gamma}_\alpha$.

2.3.3 Robustness with respect to the number of points Λ of the varifold grid

Another parameter of our algorithm is the number of points Λ of the varifold grid where template and shape complexes are projected to. We decided to make it dependent on λ_W which is the bandwidth of the varifold kernel, as explained in Sec.2.2.2. In Fig.2.4 we show that the results are not so sensitive to a change of the number of grid points and that our choice seems a good one. We compute 9 different atlases using the left-caudate surfaces of the 20 controls. All atlas constructions share the same set of parameters except for Λ which is tested in the range: [1-10^5].

As before we analyse the norm of the difference between a reference template (Λ=1) and the estimated templates at the end of the atlas procedures. We also show the norm of the residuals and the logarithm of the determinant of the covariance matrix of the initial momenta ($\log(|\Gamma_\alpha|)$). As it is possible to notice there is a minimum value of the number of grid points ($\simeq$1000) above which the results are very similar and also more satisfactory.

In order to explain these results, it can be noticed that replacing σ_j^2 with $\hat{\sigma}_j^2$ into Eq.2.9 one obtains a new data-term which is equal to $(w + \Lambda N)\log(\hat{\sigma}^2)$. When Λ is small, the reduction of the residual, and consequently of $\log(\hat{\sigma}^2)$, is not enough amplified by the multiplication with Λ to compensate a great increase in the regularity term. Thus, the estimated initial momenta keep a norm close to zero throughout the atlas construction. The deformations of the template are therefore insignificant and the covariance matrix Γ_α depends only on the prior which explains why $\log(|\Gamma_\alpha|)$ does not vary. The number of grid points obtained with the proposed method is equal to: 3456.

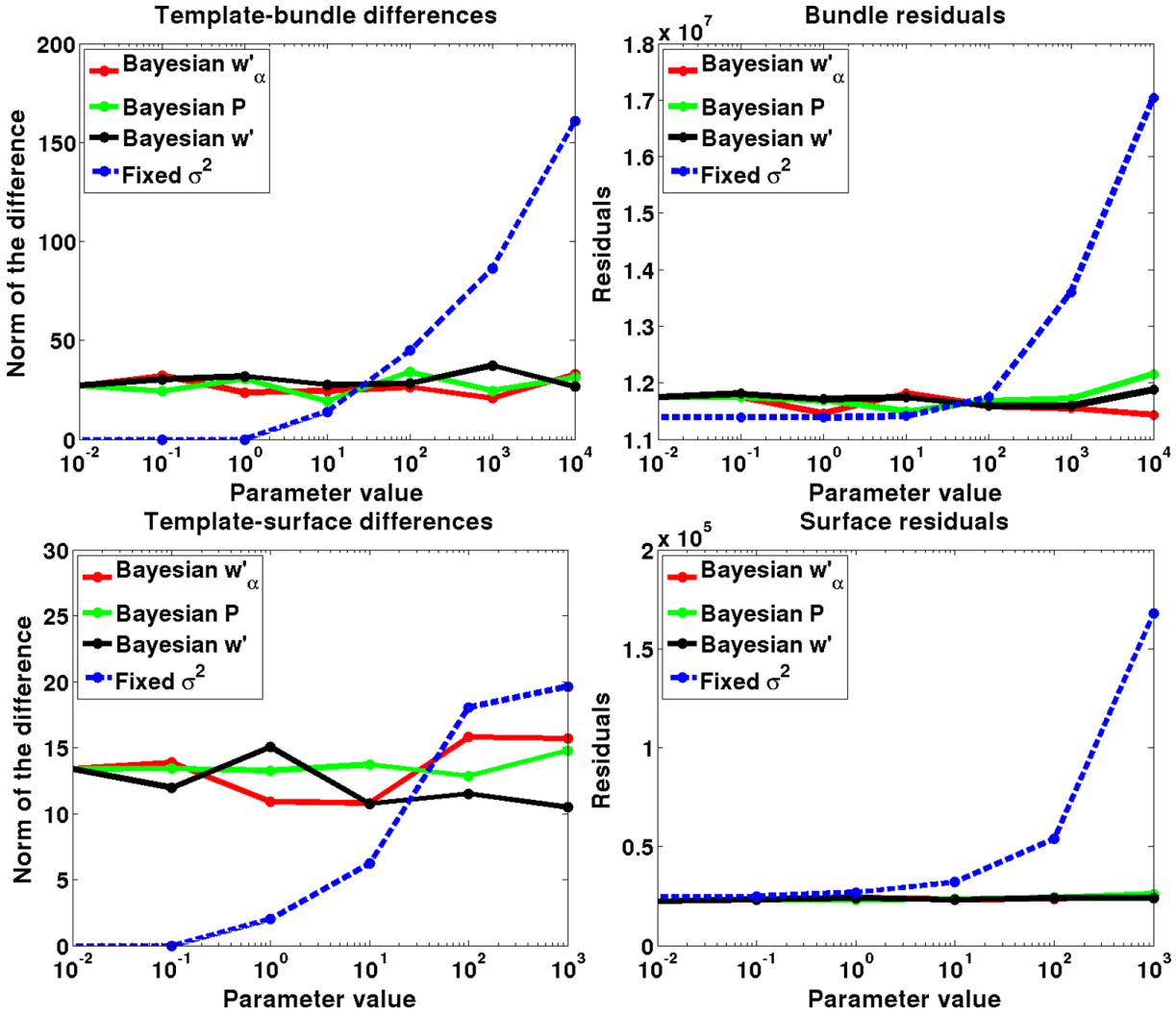

Figure 2.2: Comparison between the robustness of the proposed algorithm (Bayesian) when varying the hyper-parameters and the robustness of a previous algorithm without automatic estimates (Fixed) with respect to σ_j^2. In all atlas estimations based on the Bayesian method, it has been changed only one of the hyperparameters fixing the others to 0.01. "Surfaces" are the left caudates and "Bundles" are the fibers connecting them to the left hemisphere of the cortical surface. Figures on the left represent the norm of the difference between the template obtained at the end of the atlas constructions and a reference template, which is the one obtained using the Fixed method with σ_j^2=0.01. Figures on the right refer to the norm of the residuals obtained at the end of the atlas procedures.

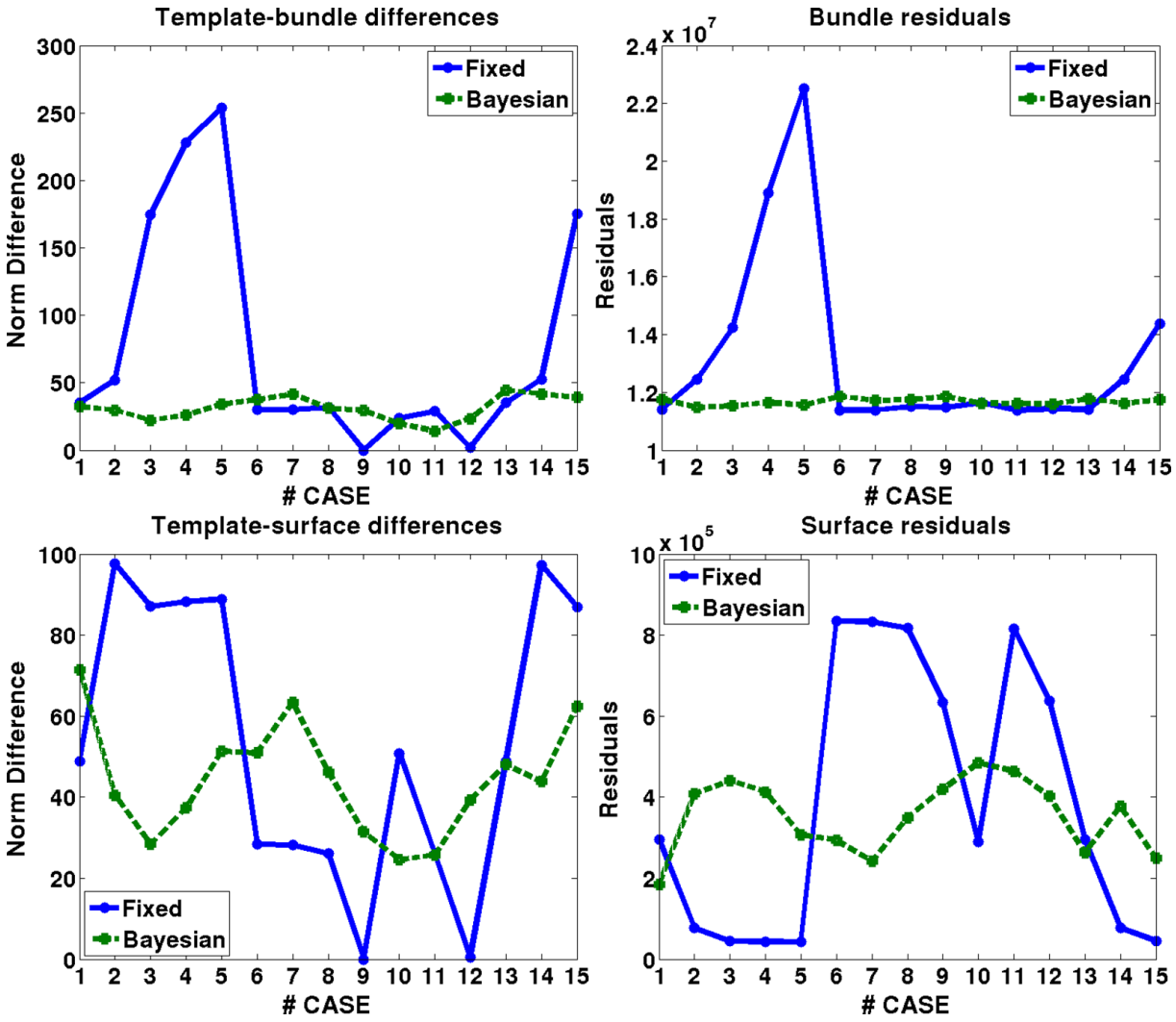

Figure 2.3: Comparison between the robustness of the proposed algorithm (Bayesian) when varying the hyper-parameters and the robustness of a previous algorithm without automatic estimates (Fixed) with respect to σ_j^2. The template complex in every atlas construction is composed of both the left caudate and the left caudate bundle. The values of the hyperparameters for the Bayesian estimation and the values of the two σ_j^2 with the Fixed method are listed in Table 2.2. Figures on the left represent the norm of the difference between the template obtained at the end of the atlas constructions and a reference template, which is the one obtained in the 9-th case using the Fixed method. Figures on the right refer to the norm of the residuals obtained at the end of the atlas procedures.

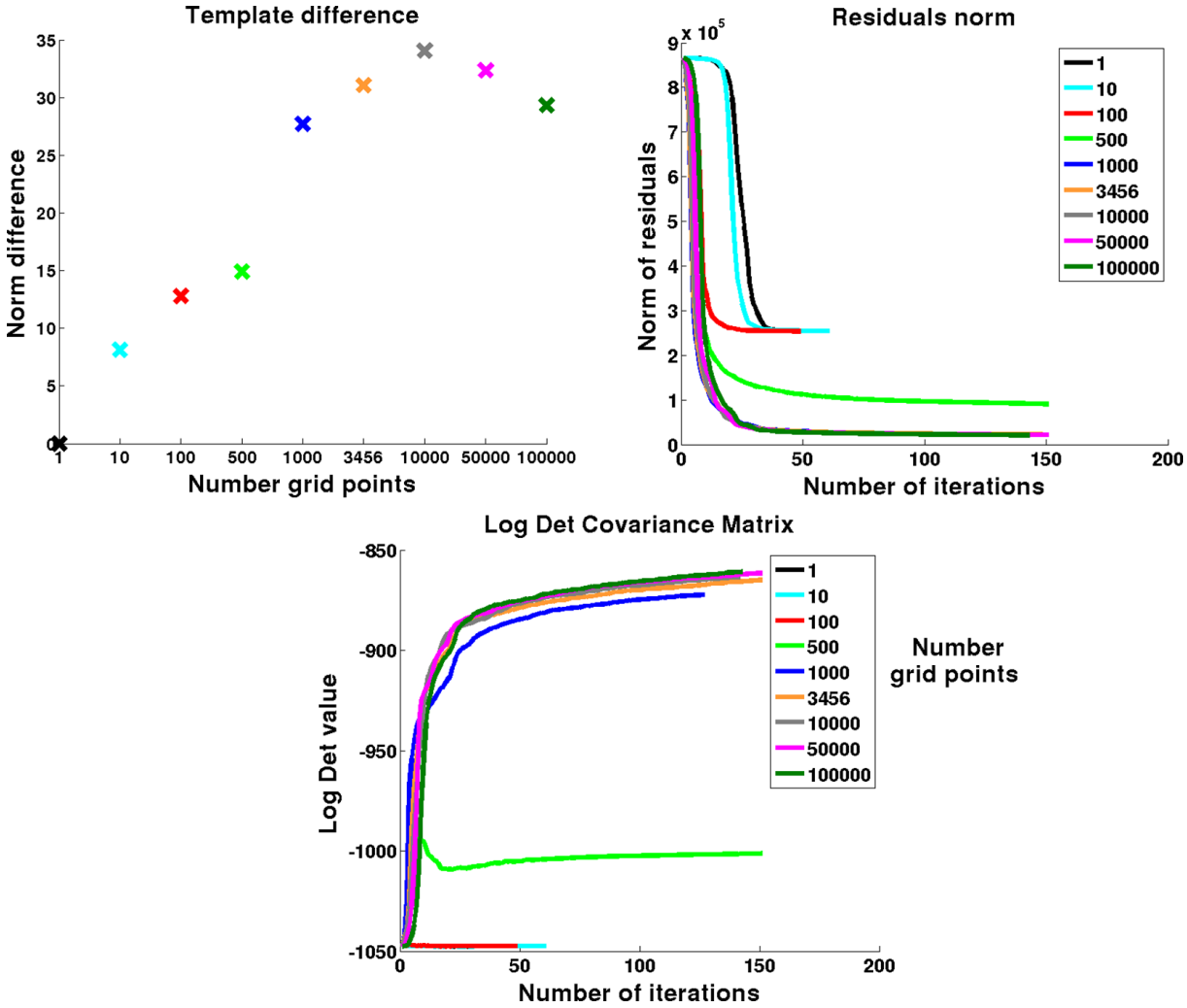

Figure 2.4: Robustness with respect to the number of points Λ of the varifold grid. We compute 9 atlases using 20 left-caudate surfaces and changing only Λ within the range [1-10^5]. Top-left figure represents the norm of the difference between a reference template (Λ=1) and the templates obtained at the end of the atlas constructions. Top-right figure shows the norm of the residuals during the atlas constructions. Bottom figure is about the evolution of $\log(|\Gamma_\alpha|)$ during the atlas procedures. The maximum number of iterations in every atlas procedure is 150. Colors refer to the different number of grid points.

2.3.4 Population differences

In all the following atlas constructions we use the same parameters. All structures, except for the caudate, have a varifold parameter λ_W equal to 2mm. For the caudate we choose a value of 3mm. About the bandwidth of the diffeomorphic kernel K_V, we use a value of 7mm with a consequent number of control points equal to 1080. The maximum number of iterations is 120 and all the computations are performed on a Intel Xeon, 32 cores, CPU E5-2650, 2.60GHz with a graphic card NVIDIA Quadro 5000. All shape complexes are previously rigidly registered to a reference shape complex.

Multi-population Atlas (Γ_α^c=Γ_α^p=Γ_α) We compute a multi-population atlas using the two groups of controls and patients. The template complex and control points are shared between the two groups. Initial momenta are assumed to follow two Gaussian distributions with opposite mean and equal covariance matrix. The overall processing time is about 100 hours based on a single-CUDA implementation.

In Fig.2.5 we show a scheme of the process. On the left there are the initial template complex and the initial set of control points which is initialised as an equally-distanced grid covering the entire ambient space. The distance is the bandwidth of the diffeomorphic kernel: λ_V. Momenta $\boldsymbol{\alpha}_{i0}$ are initially all set to zero. At the end of the atlas construction we obtain the final template, the updated set of control points, the subject-specific sets of initial momenta, the average $\boldsymbol{\mu}$ and the covariance matrix Γ_α. The template shows the common shape features of both groups. At the top and at the bottom we present the average initial momenta for each group: $\boldsymbol{\mu}$ and $-\boldsymbol{\mu}$. Then, on the right, we show the final common template deformed using $\boldsymbol{\mu}$ and $-\boldsymbol{\mu}$. The two complexes represent the anatomical configurations typical of each group and they can be directly compared since they both stem from the final template complex. We take advantage of that by computing the absolute value of the difference between the displacements from the final template along the two average directions. This is shown in Fig.2.6 for sub-cortical structures and in Fig.2.7 for fiber bundles. From the first figure it is clear that the main differences are in the dorso-lateral part of the three sub-cortical structures, especially for the caudate. About the bundles, the differences are mainly in the central part of the caudate-cortico and putamen-cortico bundles.

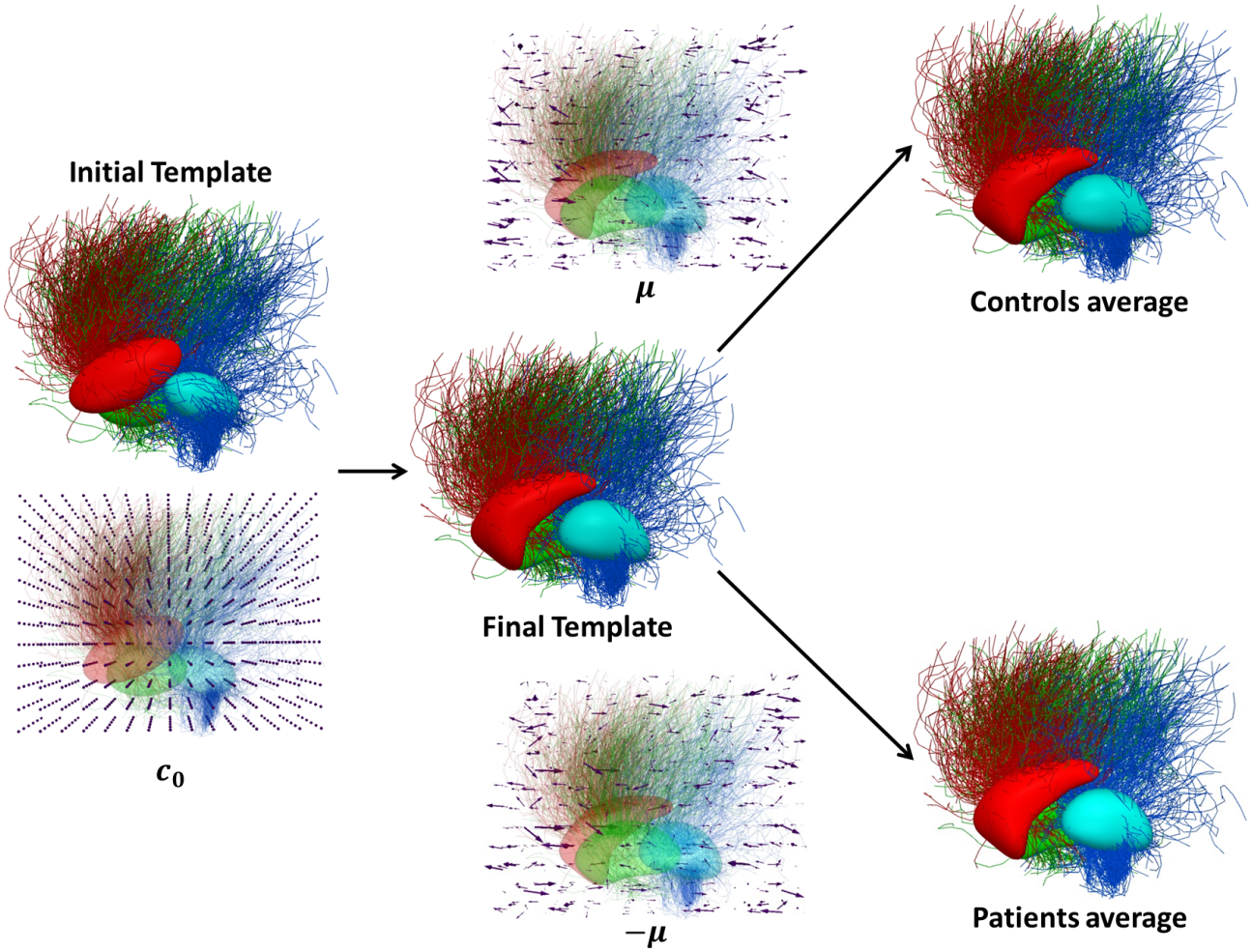

Figure 2.5: Atlas construction process. From left to right, we present first the initial template complex and the initial set of control points. Then, we show the final template obtained at the end of the atlas construction. The top and bottom arrows point to the final template deformed accordingly to the averages of initial momenta of controls μ (top) and patients -μ (bottom). The averages of initial momenta (μ and -μ) are shown respectively above and below the two arrows.

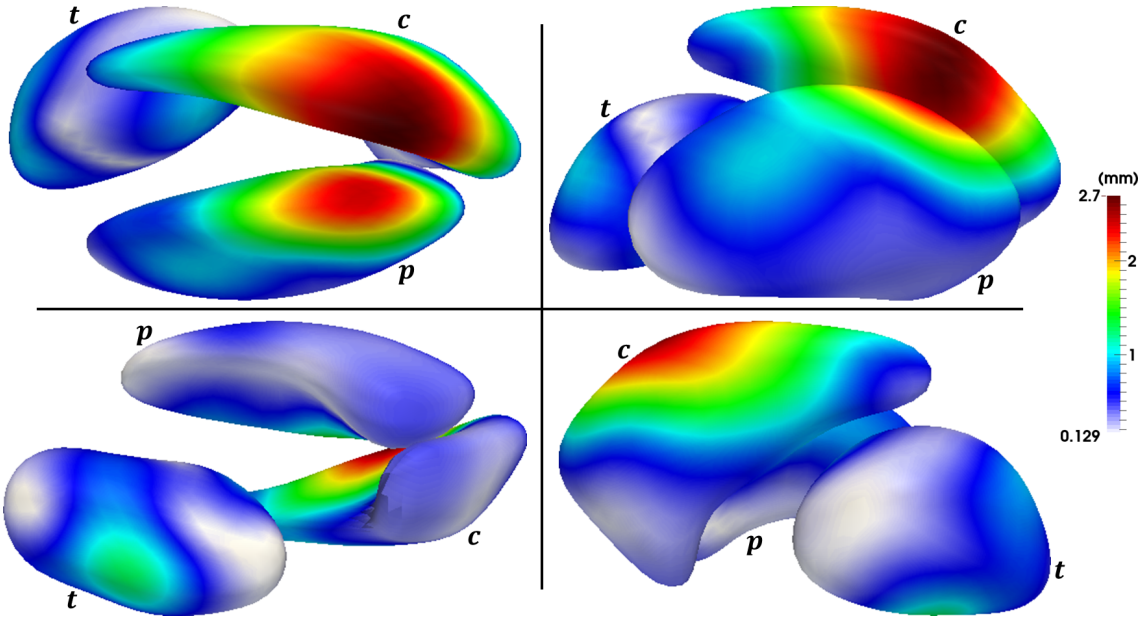

Figure 2.6: Shape dissimilarities about the sub-cortical structures between the two groups. The colors refer to the absolute value of the difference between the displacements from the final template to the average configurations of patients and controls. The two average configurations are obtained by deforming the final template along the directions given by μ and -μ. The four frames represent the same three structures from different points of view. Letters $\{c, p, t\}$ refer to caudate, putamen and thalamus respectively.

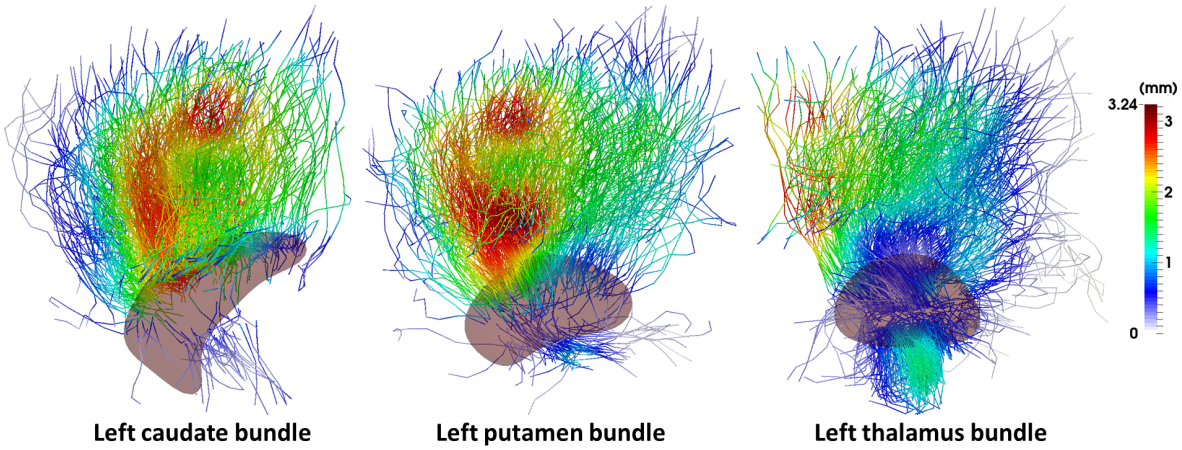

Figure 2.7: Shape dissimilarities about the fiber bundles between the two groups. The colors refer to the absolute value of the difference between the displacements from the final template to the average configurations of patients and controls. The two average configurations are obtained by deforming the final template with μ and -μ.

Multi-population Atlas ($\Gamma_\alpha^c \neq \Gamma_\alpha^p$) Here we assume that the two groups of patients and controls do not share the same covariance matrix. To test this assumption, we use the test proposed in Srivastava and Yanagihara (2010). The p-value (0.3461) is too big to reject the null hypothesis: $\Gamma_\alpha^c = \Gamma_\alpha^p$. This might be due to the fact that the sample size (20 subjects per group) is too small with respect to the number of deformation parameters (3240). Another possible explication could be that the two groups share most of their morphological variability which means that the first modes of variation of both covariance matrices are similar. To test this hypothesis, we compute a Principal Component Analysis (PCA) for both covariance matrices and we calculate the angles between the modes of variation of the group of controls with the ones of the group of patients. The results for the first 5 modes are shown in Table 2.1. As it is possible to notice, the first mode of the group of controls is almost parallel to the first one of the group of patients. This means that they produce similar morphological changes of the template complex. Moreover, they explain almost 35% and 45% respectively of the total variability of their groups, see Fig.2.8. This might explain why the previous test fails in rejecting the hypothesis of equality between the two covariance matrices.

The first modes of both groups indicate the shared morphological variability between controls and patients. The second modes should instead produce shape variations which are more characteristic of each group (angle $\sim 60°$) and which contribute also substantially ($\sim 10\%$) to the total variability. In Fig.2.9 we show the second mode of variation of both groups. Colors refer to the displacement from the final common template shown in gray in the middle of the figure. It is possible to notice that the dorso-lateral part of the putamen varies almost in the same way between the two groups. Instead, the posterior-lateral part of the caudate and the ventral part of the thalamus present morphological variations which are different between the two groups.

		Patients				
		1	2	3	4	5
Controls	1	25.71	82.49	87.10	87.77	83.32
	2	87.70	60.39	72.57	83.62	80.30
	3	80.60	62.59	67.17	84.39	79.76
	4	86.74	75.15	82.60	88.98	77.35
	5	83.70	81.24	89.51	85.39	83.80

Table 2.1: Angles in degrees between the first 5 modes of variations of the group of controls and the first five of the group of patients. The average angle between all modes is 85.39°.

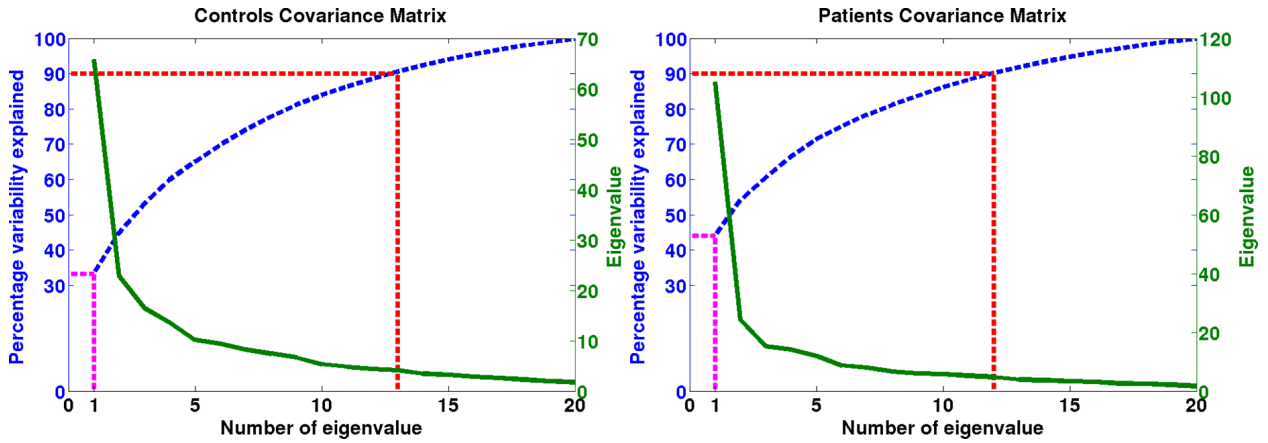

Figure 2.8: PCA of the two covariance matrices of the initial momenta based on the assumption $\Gamma_\alpha^c \neq \Gamma_\alpha^p$. The magenta line represents the percentage of variability explained by the first mode. The red line shows the number of components needed to explain 90% of the total variability. The green line indicates the magnitude of the eigenvalues and the blue line shows the explained variability using all the previous components.

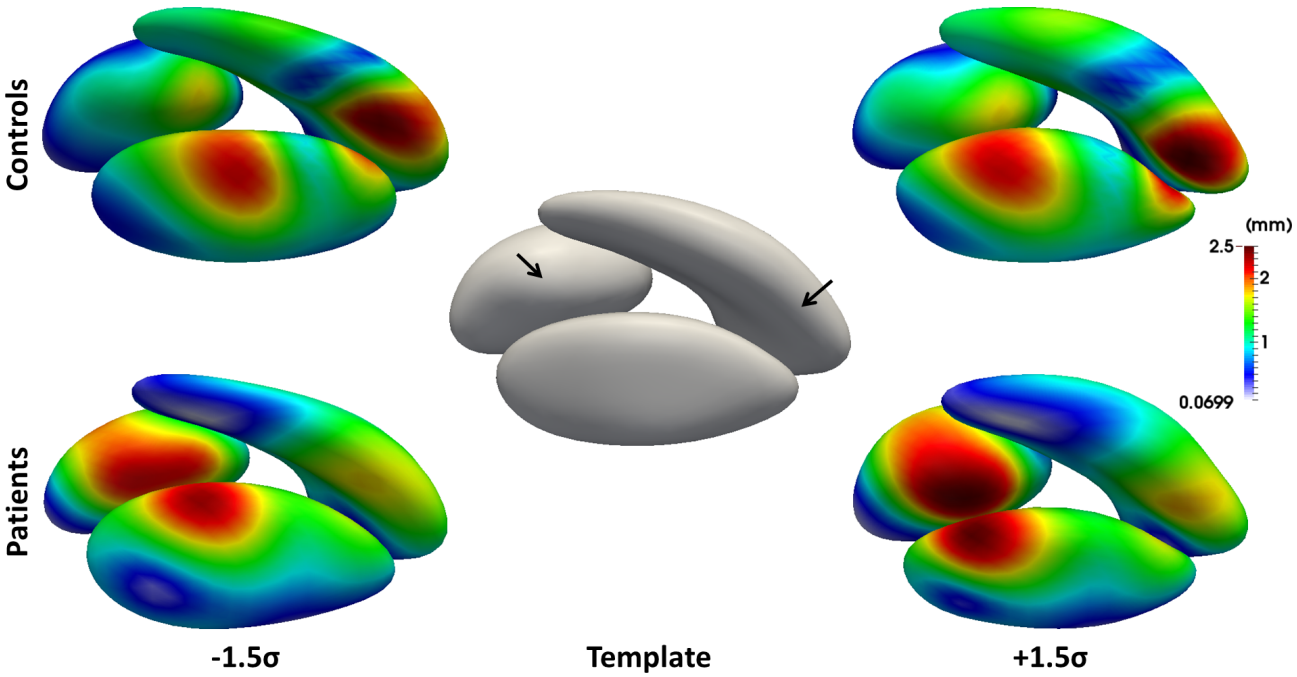

Figure 2.9: Second mode of variation at $\pm$ 1.5 standard deviation (σ) of two PCAs based on the two estimated covariance matrices of the initial momenta Γ_α^c and Γ_α^p. Colors refer to the displacement from the final common template shown in the middle of the figure in gray. Arrows indicate the areas of the template complex which vary in a different way between the two groups.

2.4 Discussion and Conclusion

This work provides a Bayesian framework to embed a multi-object atlas construction into a statistical setting. It is general and it can be applied to any parametric deformation framework and to all shape models with which is possible to define probability density functions. It allows to automatically estimate important balancing parameters which were fixed by the user in previous methods, namely the noise variances of every structure. We have demonstrated that these parameters influence much more the results than the hyper-parameters introduced with the proposed Bayesian priors. This statistical setting makes therefore multi-object atlas constructions more feasible and reproducible since the user is not obliged to tune or fix balancing weights which grow with the number of analysed structures.

The proposed method also allows us to estimate the covariance matrix of the deformation parameters. It is used for the statistical analysis of the morphological variability of the template complex within the population under study. In previous works (Durrleman et al., 2014; Avants and Gee, 2004; Ma et al., 2010), the authors proposed to use the sample covariance matrix computed at the end of the atlas procedure. Since the number of subjects is usually smaller than the number of deformation parameters, the sample covariance matrix might be ill-conditioned. A standard solution is to regularize it a posteriori using an identity matrix multiplied by a small scalar. This strategy is not satisfactory since it means that the estimate of the deformation parameters is based on a degenerate sample covariance matrix or on a wrong covariance matrix. Moreover, the regularization seems also too simplistic. In our approach, we directly estimate a well-conditioned covariance matrix using a more natural and coherent regularisation term given by the kernel of the RKHS to which belongs the vector field used to compute the diffeomorphisms. Furthermore, the estimate of the covariance matrix does not change if the regularity term with its related prior terms are multiplied by a scalar. This means that we automatically take into account also the estimate of the so-called regularity level.

We have also extended the proposed statistical setting to a multi-population strategy where we take into consideration the affiliation of a subject to a certain group defined, for instance, by its clinical diagnosis. This allows us to employ more complicated but also more pertinent models at no additional cost, meaning without fixing or tuning new parameters. One of the main advantages is to estimate different populations of initial momenta and not only one. Moreover, we can also quantitatively compare them since they are defined on the same set of control points and the groups share the same template. Furthermore, this extension also allows us to estimate a different covariance matrix for every group. We can therefore quantitatively compare them and their modes. It is interesting to notice that one could also compute the sample covariance matrices of different groups at the end of a single-population atlas and then compare their modes. But this would not be consistent with the fact that in a single-population atlas the initial momenta of different groups are estimated as belonging to the same Gaussian distribution.

We propose to model both curve and surface meshes as Gaussian random var-

ifolds. They can be seen therefore as instances of the same mathematical object. Since the space of varifolds is of infinite dimension, we create a grid for every structure j on which we project both shapes and template. We decide to make the number of grid points Λ_j dependent on the varifold bandwidth λ_W^j. This choice does not influence the results in a single-object atlas as demonstrated in Sec.2.3.3. In a multi-object analysis, it makes the contribution to the cost function (Eq.2.9) of structure j even more dependent on λ_W^j. In fact, the contribution of structure j depends mainly on the residual and on the value of Λ_j which are both defined by λ_W^j. On the one hand, this choice simplifies the atlas construction since the user has to fix only one parameter, on the other hand it makes the choice of λ_W^j even more crucial. In future works, we will investigate how to automatically estimate λ_W^j.

Another important parameter fixed by the user is the diffeomorphic bandwidth λ_V. It is chosen based on the desired registration accuracy and it defines the number of control points. Its estimate, together with the one of λ_W^j, would make the atlas construction completely automatic whatever the number or kind of meshes. A possible strategy could be to select a range of suitable values of λ_V and then use sparse multi-scale methods (Sommer et al., 2012). Otherwise, one could estimate the best deformation modules from a dictionary using sparse techniques in order to disentangle the single complicated diffeomorphism into interpretable transformations (Gris et al., 2015). This would augment the computational load and execution time but it would also make the analysis more objective.

Another improvement for our method would be the employment of sampling algorithms like MCMC to ensure the convergence of the EM scheme whatever the quality of the data. As already said, these methods require a great computational burden. A possible solution would be to use a GPU implementation based on Multi-Graphics Processing Units (Multi-GPU) as in Ha et al. (2009). The authors showed that a Multi-GPU implementation gives a computational gain up to sixty times faster than a single CPU implementation.

We tested the proposed algorithm by comparing two populations of shape complexes, one of controls and one of Gilles de la Tourette patients, obtained from MR images of the brain and consisting of three sub-cortical structures and the fiber bundles connecting them to the cortical surface. Results proved the effectiveness of our method in detecting morphological differences between the two populations. They could indicate atypical connections resulting from abnormal brain development due to Gilles de la Tourette syndrome. Future works will aim to confirm this study by adding more structures as the cortical surface and possibly more subjects. Moreover, we plan to apply our algorithm to other human organs such as the heart or the liver.

2.A Appendix

2.A.1 Multi-population atlas construction

Here we derive the changes of the cost function (Eq.2.9) and relative gradients for the multi-population extension of the atlas construction.

2.A.1.1 Γ^c_α=Γ^p_α=Γ_α

Let N_c and N_p be the number of subjects of the two groups and N=N_c+N_p. Moreover, we define $\boldsymbol{\alpha}^c_{i0} \sim N(\boldsymbol{\mu}, \Gamma_\alpha)$ and $\boldsymbol{\alpha}^p_{i0} \sim N(-\boldsymbol{\mu}, \Gamma_\alpha)$. Thus, Eq.2.9 becomes:

$$\boxed{\sum_{j=1}^{M}\sum_{i=1}^{N_c}\frac{1}{2\sigma_j^2}\left(||S^c_{ij}-\phi^c_i(T_j)||^2+\frac{P_j w_j}{N}\right)}+\boxed{\frac{1}{2}\sum_{i=1}^{N_c}(\boldsymbol{\alpha}^c_{i0}-\boldsymbol{\mu})^T(\Gamma_\alpha)^{-1}(\boldsymbol{\alpha}^c_{i0}-\boldsymbol{\mu})}+$$
$$\boxed{\sum_{j=1}^{M}\sum_{i=1}^{N_p}\frac{1}{2\sigma_j^2}\left(||S^p_{ij}-\phi^p_i(T_j)||^2+\frac{P_j w_j}{N}\right)}+\boxed{\frac{1}{2}\sum_{i=1}^{N_p}(\boldsymbol{\alpha}^p_{i0}+\boldsymbol{\mu})^T(\Gamma_\alpha)^{-1}(\boldsymbol{\alpha}^p_{i0}-\boldsymbol{\mu})}+$$
$$\sum_{j=1}^{M}\frac{1}{2}(w_j+\Lambda_j N)\log(\sigma_j^2)+\frac{1}{2}(w_\alpha+N)\log(|\Gamma_\alpha|)+\frac{w_\alpha}{2}tr((\Gamma_\alpha)^{-1}P_\alpha) \tag{2.26}$$

where the norm $||\cdot||$ can be both the L^2-norm and the varifold one. The gradients with respect to $\boldsymbol{T}$, $\boldsymbol{c}_0$ and $\{\sigma_j^2\}$ are exactly the same as in Eq.2.17 and Eq.2.16. The one with respect to $\boldsymbol{\alpha}_{i0}$ depends on the group and it is equal to:

$$\nabla_{\boldsymbol{\alpha}^c_{s0}}E=\sum_{j=1}^{M}\frac{1}{2\sigma_j^2}\nabla_{\boldsymbol{\alpha}^c_{s0}}D^c_{sj}+\Gamma_\alpha^{-1}(\boldsymbol{\alpha}^c_{s0}-\boldsymbol{\mu}) \qquad \nabla_{\boldsymbol{\alpha}^p_{s0}}E=\sum_{j=1}^{M}\frac{1}{2\sigma_j^2}\nabla_{\boldsymbol{\alpha}^p_{s0}}D^p_{sj}+\Gamma_\alpha^{-1}(\boldsymbol{\alpha}^p_{s0}+\boldsymbol{\mu}) \tag{2.27}$$

There is a closed form solution for both μ and Γ_α which is equal to:

$$\hat{\Gamma}_\alpha=\frac{\sum_{i=1}^{N_c}\left[(\boldsymbol{\alpha}^c_{i0}-\boldsymbol{\mu})(\boldsymbol{\alpha}^c_{i0}-\boldsymbol{\mu})^T\right]+\sum_{i=1}^{N_p}\left[(\boldsymbol{\alpha}^p_{i0}+\boldsymbol{\mu})(\boldsymbol{\alpha}^p_{i0}+\boldsymbol{\mu})^T\right]+w_\alpha P_\alpha^T}{(N+w_\alpha)}$$
$$\hat{\boldsymbol{\mu}}=\frac{\sum_{i=1}^{N_c}\boldsymbol{\alpha}^c_{i0}-\sum_{i=1}^{N_p}\boldsymbol{\alpha}^p_{i0}}{N} \tag{2.28}$$

2.A.1.2 $\Gamma^c_\alpha \neq \Gamma^p_\alpha$

As before, let N_c and N_p be the number of subjects of the two groups and N=N_c+N_p. Moreover, we define $\boldsymbol{\alpha}^c_{i0} \sim N(\boldsymbol{\mu}, \Gamma^c_\alpha)$ and $\boldsymbol{\alpha}^p_{i0} \sim N(-\boldsymbol{\mu}, \Gamma^p_\alpha)$. The priors on Γ^p_α and

Γ^p_α are both equal to $\mathcal{W}^{-1}(P_\alpha, w_\alpha)$. Thus, Eq.2.9 becomes:

$$\boxed{\sum_{j=1}^{M}\sum_{i=1}^{N_c}\frac{1}{2\sigma_j^2}\left(||S^c_{ij}-\phi^c_i(T_j)||^2+\frac{P_j w_j}{N}\right)}+\boxed{\frac{1}{2}\sum_{i=1}^{N_c}(\boldsymbol{\alpha}^c_{i0}-\boldsymbol{\mu})^T(\Gamma^c_\alpha)^{-1}(\boldsymbol{\alpha}^c_{i0}-\boldsymbol{\mu})}+$$
$$\boxed{\sum_{j=1}^{M}\sum_{i=1}^{N_p}\frac{1}{2\sigma_j^2}\left(||S^p_{ij}-\phi^p_i(T_j)||^2+\frac{P_j w_j}{N}\right)}+\boxed{\frac{1}{2}\sum_{i=1}^{N_p}(\boldsymbol{\alpha}^p_{i0}+\boldsymbol{\mu})^T(\Gamma^p_\alpha)^{-1}(\boldsymbol{\alpha}^p_{i0}-\boldsymbol{\mu})}+$$
$$\sum_{j=1}^{M}\frac{1}{2}(w_j+\Lambda_j N)\log(\sigma_j^2)+\frac{1}{2}(w_\alpha+N_c)\log(|\Gamma^c_\alpha|)+\frac{w_\alpha}{2}tr((\Gamma^c_\alpha)^{-1}P_\alpha)+$$
$$+\frac{1}{2}(w_\alpha+N_p)\log(|\Gamma^p_\alpha|)+\frac{w_\alpha}{2}tr((\Gamma^p_\alpha)^{-1}P_\alpha) \tag{2.29}$$

where the norm $||\cdot||$ can be both the L^2-norm and the varifold one. The gradients with respect to $\boldsymbol{T}$, $\boldsymbol{c}_0$ and $\{\sigma_j^2\}$ are exactly the same as in Eq.2.17 and Eq.2.16. The one with respect $\boldsymbol{\alpha}_{i0}$ depends on the covariance matrix of the group and its formulation is equal to the one in Eq.2.27. Furthermore, there is a closed form solution for both μ, Γ^c_α and Γ^p_α. The last two have the same expression of Eq.2.16 with a sum over N_c and N_p. Instead, the optimal value for the average momenta is :

$$\hat{\boldsymbol{\mu}}=\frac{\sum_{i=1}^{N_c}(\Gamma^c_\alpha)^{-1}\boldsymbol{\alpha}^c_{i0}-\sum_{i=1}^{N_p}(\Gamma^p_\alpha)^{-1}\boldsymbol{\alpha}^p_{i0}}{N_c(\Gamma^c_\alpha)^{-1}+N_p(\Gamma^p_\alpha)^{-1}} \tag{2.30}$$

2.A.2 Robustness Analysis

We present in Table 2.2 the values of the fixed parameters and hyperparameters used in the experiments shown in Fig.2.3.

Case	Fixed		Bayesian				
	$\sigma^2_{surface}$	σ^2_{bundle}	w_α	$w'_{surface}$	$P_{surface}$	w'_{bundle}	P_{bundle}
1	0.01	0.01	1	10	0.1	10	0.1
2	0.01	0.1	0.01	0.01	0.01	0.01	0.01
3	0.01	1	0.1	0.1	0.1	0.1	0.1
4	0.01	10	1	1	1	1	1
5	0.01	100	10	10	10	10	10
6	100	0.01	100	100	100	100	100
7	100	0.1	10	1	0.1	0.01	100
8	100	1	0.1	10	1	10	0.1
9	100	10	0.01	100	10	1	0.01
10	100	100	100	0.1	0.1	10	1
11	1	0.01	0.1	0.1	100	1	10
12	1	0.1	0.1	10	10	1	1
13	1	1	0.01	1	10	0.1	100
14	1	10	10	10	0.1	100	1
15	1	100	10	1	0.01	0.1	1

Table 2.2: Values of the two σ^2 for the Fixed method and values of the hyper-parameters for the Bayesian estimation used in the robustness analysis shown in Fig.2.3.

Chapter 3

Parsimonious approximation for white matter fiber bundles based on weighted currents

This chapter has been submitted to IEEE Transactions on Medical Imaging and it has been partly published in Gori et al. (2014). *The C++ code about the computational model of weighted currents will be soon integrated to the software Deformetrica.*

Contents

3.1 Introduction

Tractography (Conturo et al., 1999; Mori and van Zijl, 2002) from diffusion-weighted magnetic resonance imaging (DW-MRI) (Basser et al., 1994) is the only non-invasive technique capable to trace in vivo the wiring architecture of the human brain white matter. It is widely employed for both clinical (i.e. stroke (Mukherjee, 2005), surgical procedures (Ciccarelli et al., 2008)) and research purposes (i.e. Alzheimer's disease, schizophrenia (Kubicki et al., 2007)). The 3D polylines stemming from tractography algorithms, called streamlines or fibers, are only estimates of the trajectories of large groups of neural axons. Streamlines are traced from points inside a starting voxel, called seeds, and they are constituted of segments connecting neighbouring voxels. The direction of these segments is defined by a local diffusion model (i.e. tensor, Q-ball (Fillard et al., 2011)) computed at each voxel and by a tractography method: deterministic or probabilistic (Fillard et al., 2011). Deterministic algorithms produce segments which follow the principal direction of the local diffusion model whereas probabilistic ones use randomly perturbed versions of the main direction. Seeds are usually placed in every voxel of the white matter (whole-brain tractography) and the resulting streamlines can then be divided into different fiber bundles based on clustering algorithms or starting/ending Regions of Interest (ROI). Fiber bundles may be then decomposed into *fascicles* which are groups of fibers with a similar pathway and whose extremities are close to each other, connecting therefore the same functional territories.

Fiber bundles are difficult to analyse both qualitatively and quantitatively due to their considerable number of streamlines. The size of a bundle can make intractable processes such as clustering (O'Donnell and Westin, 2007), registration (Siless et al., 2012), atlas construction (Durrleman et al., 2011a) or shape analysis (Corouge et al., 2004) due to the great computational load and execution time. It can also complicate the rendering and the visualisation, thus limiting possible clinical applications. Moreover, the great quantity of streamlines might impede understanding the pattern of a bundle making therefore difficult its interpretation.

In this chapter, we propose to approximate a fiber bundle with a parsimonious representation of weighted streamline prototypes. We exploit the fact that many streamlines starting from seeds in the same voxel or in neighbour voxels share the same pathway and ending area. We approximate these streamlines with one of them, called prototype. We use a computational model for both streamlines and prototypes characterised by an explicit and easily computable metric. This allows us to control the approximation error and to select the streamlines which minimize it as prototypes.

3.2 Related Work

In the last years, there has been a great effort to compactly represent a fiber bundle. A pragmatic strategy is to randomly choose a subset of the streamlines. The

sampling is not driven by the minimisation of an approximation error and it is not possible to control the selection of the streamlines. This can cause the loss of the smallest fascicles of the bundle which might be important for the purpose of the study. More sophisticated solutions have been proposed and they can be separated into two categories. The first group gathers the computational methods which compactly parametrise the single streamlines. The second category pools instead the strategies focused on concisely representing the entire fiber bundle.

3.2.1 Compact representation of streamlines

Streamlines are composed of contiguous variable-length segments whose number might also vary among fibers. Different computational models have been proposed whose goal is to concisely parametrise a streamline. A widely employed method consists of defining point-to-point correspondence among streamlines parametrising them as sets of points (O'Donnell et al., 2012) or with cubic B-splines (Corouge et al., 2006) for instance. This technique eases the computations but it can be applied only if streamlines have a similar length and the definition of corresponding points can be very challenging. Other authors proposed to characterise a streamline using only its extremities (Brun et al., 2003) or its connectivity signature (Tunç et al., 2013), namely the probabilities to be linked to a defined set of ROIs. These methods have been used for clustering, visualisation and interpretation purposes but they do not take into consideration the shape of the streamlines which is important for registration and morphometry. Conversely, different authors proposed to evaluate only the local pathway of the fibers without taking into consideration their extremities. A first example is given by the methods based on Fourier descriptors (Batchelor et al., 2006; Chung et al., 2010), which result in a concise parametrisation useful for clustering and shape analysis. In these models the number of descriptors needs to be fixed though and the optimal number depends on the length and shape of the streamlines which might vary even within a single bundle. Lately, other authors proposed to represent a streamline as a blurred indicator function modelled as a Gaussian process (Wassermann et al., 2010; Liu et al., 2012). This representation permits to easily compare and average streamlines but it is not a geometric primitive and it is therefore difficult to employ in multi-object registrations together with other geometric primitives such as 3D surface meshes.

3.2.2 Compact representation of the whole bundle

The second category is composed of methods which approximate the entire fiber bundle. The most common strategy is to divide the fiber bundle into subsets, usually called clusters, which are then characterised by representative fibers (Zvitia et al., 2010) or centroids (O'Donnell et al., 2009; Guevara et al., 2011; Garyfallidis et al., 2012). Other authors have also employed isosurfaces to represent the spatial variation of the centroid within the bundle (Maddah et al., 2007). This representation can be used only for tubular-shaped bundles that can be modelled as convex

envelopes. Other bundles, such as the corpus callosum and the rostral part of the corticospinal, have a different topology and they are defined as sheet-like bundles. In Maddah et al. (2011) the authors proposed to represent those bundles as 3D surface meshes whereas in Yushkevich et al. (2008) it was suggested to use deformable medial models (cm-reps). In both cases, the medial surface representations are employed only for visualisation and clustering and to provide statistics about diffusion coefficients. A different representation, which can be employed for any kind of bundle, is the tract probability map (Hua et al., 2008; Bürgel et al., 2006; Wassermann et al., 2010). It indicates the probability of a voxel to belong to a given bundle. This method is very concise but it is not based on a geometrical primitive and it has been used for visualisation, interpretation and clustering. A last example is the sparse representation based on the matching pursuit algorithm for currents presented in Durrleman et al. (2011a). In the framework of currents (Vaillant and Glaunès, 2005) a bundle is considered as a single mathematical object composed of disconnected oriented points which model the local orientation of the streamlines. The approximation presented in Durrleman et al. (2011a) represents a bundle with a sparse set of oriented points. This representation is very concise but it has the drawback to approximate accurately only the areas of the bundle characterised by a high density of streamlines, like the central mass of the bundle. Thus, the small fascicles may not be well approximated. Moreover, the framework of currents does not take into account the extremities of the streamlines. This prevents the analysis of the structural connectivity, namely the areas of the gray matter connected by the bundle.

3.3 Our contribution

In this chapter, we propose to approximate *any* fiber bundle with a set of weighted prototypes. Prototypes are chosen among the streamlines and they represent ensembles of similar fibers. Their weights are related to the number of streamlines approximated. Both prototypes and streamlines are modelled as weighted currents, an extension of the framework of currents. This computational model takes into consideration both the *pathway* of the fiber/prototype and the *anatomical location* of its extremities. Two fibers/prototypes are therefore considered similar if their endpoints are close to each other and if their trajectories are similar. The space of weighted currents is a *vector space* with an explicit and easily computable metric. This permits to approximate a bundle of streamlines controlling the approximation error. The resulting parsimonious representation, up to a reasonable approximation level, preserves both the shape and the structural connectivity (computed as streamline density) of the original bundle. Moreover, the framework of weighted currents inherits from the one of usual currents (Vaillant and Glaunès, 2005) the fact that it does *not* need either point-to-point or streamline-to-streamline correspondences. Nevertheless, it requires to define the starting and ending point of each streamline. The uncertainty of tractography algorithms near the grey/white matter

boundaries is taken into account by a Gaussian smoothing kernel. Prototypes are visualised as tubes whose constant radii are proportional to their weights. This concise representation can be easily combined in multi-object studies with other geometric primitives such as surface meshes modelled as landmarks, currents or varifolds (Gori et al., 2013a; Durrleman et al., 2014).

The chapter is organised as follows. In Section 3.4 we first present the framework of weighted currents and then the different steps of the proposed approximation scheme. After that, we demonstrate the effectiveness of our algorithm on deterministic and probabilistic fiber bundles from both a qualitative and quantitative point of view. The last two sections are about the discussion of the results and conclusions.

3.4 Method

3.4.1 Terminology

The proposed approximation scheme is conceived for fiber bundles resulting from both deterministic and probabilistic streamline tractography algorithms. The definitions of streamline, fascicle and bundle, as employed throughout this chapter, are as follows. A streamline is a curve composed of a finite, ordered and connected sequence of 3D points. The distance between connected points is not assumed to be constant. The number of points may vary between two different streamlines. A fascicle is an ensemble of streamlines with similar pathway and whose extremities are close to each other. A bundle is a group of streamlines with a consistent orientation and connecting two specific ROIs defined by the user. Every bundle is composed of one or more fascicles.

3.4.2 Weighted currents

The framework of weighted currents is an extension of the one of currents (Vaillant and Glaunès, 2005). A streamline is considered as a set of disconnected oriented points which are *weighted* by the spatial coordinates of the streamline extremities. In this way, every oriented point encodes not only the local orientation of a streamline, as in usual currents, but also its connectivity. It is an adaptation of the framework of functional currents (Charon and Trouvé, 2013a).

A streamline X is a polygonal line of N segments which is assumed to be an oriented and rectifiable curve in $\boldsymbol{R}^3$. The coordinates of the two extremities f^a and f^b are two 3D vectors defined in the space $\mathrm{Q}=\boldsymbol{R}^3\mathrm{x}\boldsymbol{R}^3$. The fiber X is modelled as

a 1-weighted current C_X via a line integral of a vector field ω:

$$\begin{aligned} C_X(\omega) = \int_X w(x, f^a, f^b)^T \alpha(x) dx &\approx \sum_{i=1}^{N} \omega^T_{(x_i, f^a, f^b)}(\alpha_i) \\ &\approx \sum_{i=1}^{N} \delta^{\alpha_i}_{(x_i, f^a, f^b)}(\omega) \end{aligned} \tag{3.1}$$

where x_i and α_i are respectively the center and the tangent vector of segment i which is approximated by a Dirac delta weighted current δ. We take advantage of the fact that the tangent vectors $\alpha(x)$ are constant over every segment and it can be shown that the approximation error tends to zero as the sampling becomes more accurate i.e. the length of the segments decreases (Glaunès, 2005). The vector field ω belongs to a reproducing kernel Hilbert space (RKHS) W defined on the product space $\boldsymbol{R}^3$xQ. The space of weighted currents is a continuous linear form on W and every weighted current C_X belongs to its dual space W^*. A natural way to build a kernel K associated to the product space W is as tensor product of two kernels defined separately in $\boldsymbol{R}^3$ (K_g) and in Q (K_f): $K{=}K_g \otimes K_f$. Since even Q is a product space, K_f is also defined as a tensor product between two kernels K_a and K_b. Thus, the kernel K results: $K((x, f^a, f^b), (y, t^a, t^b)){=}K_g(x, y)K_a(f^a, t^a)K_b(f^b, t^b)$. All kernels K_g, K_a and K_b are defined as Gaussian and they are parametrised by their bandwidths λ_g, λ_a and λ_b. Using these kernels, the inner product in the framework of weighted currents between two Diracs is defined as: $\langle \delta^{\alpha}_{(x,f^a,f^b)}, \delta^{\beta}_{(y,t^a,t^b)} \rangle_{W*}$ $= K_a(f^a, t^a)\ K_b(f^b, t^b)\ (\alpha^T K_g(x, y)\beta)$. By linearity, the inner product between two streamlines X and Y ($C_Y(\omega) \approx \sum_{j=1}^{M} \omega_{(y_j, t^a, t^b)}(\beta_j)$) is:

$$\begin{aligned} \langle C_X, C_Y \rangle_{W*} &= K_a(f^a, t^a) K_b(f^b, t^b) \boxed{\sum_{i=1}^{N} \sum_{j=1}^{M} \alpha_i^T K_g(x_i, y_j) \beta_j} \\ &= \exp\left(\frac{-||f^a - t^a||^2}{\lambda_a^2}\right) \exp\left(\frac{-||f^b - t^b||^2}{\lambda_b^2}\right) \\ &\quad \boxed{\sum_{i=1}^{N} \sum_{j=1}^{M} \exp\left(\frac{-||x_i - y_j||^2}{\lambda_g^2}\right) \alpha_i^T \beta_j} \end{aligned} \tag{3.2}$$

The framed part would be the inner product between X and Y if modelled as usual currents. It measures overall differences between the geometry of their trajectories. The two other terms take into account the differences in the location of the streamline extremities. As shown in Fig.3.1, the framework of currents is almost "blind" to a change of the positions of the end-points. Even if the extremities of the two streamlines are far from each other, with respect to the kernel bandwidth

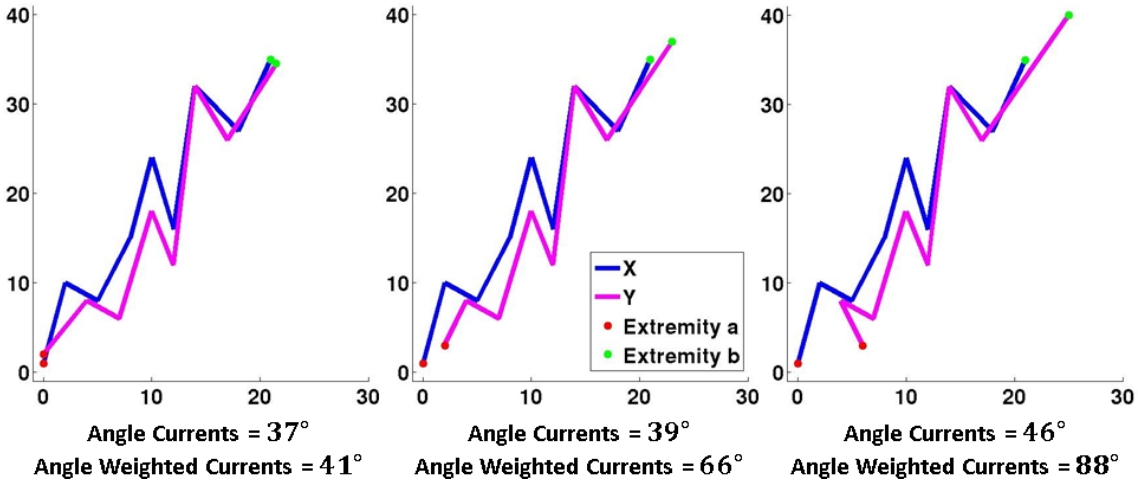

Figure 3.1: Two 2D streamlines X and Y are compared using both the framework of currents and the one of weighted currents. In the three figures only the position of the extremities change, the overall pathway remains almost unchanged. The framework of weighted currents is more sensitive to the distance between the extremities which explains why the two streamlines X and Y are almost orthogonal in the last figure on the right. The bandwidths of all kernels employed in both frameworks are equal to 5.

λ_g, the angle $(cos^{-1}\frac{\langle C_X, C_Y\rangle_{W*}}{|C_X|_{W*}|C_Y|_{W*}})$ between the two streamlines varies by only few degrees. On the contrary, the streamlines become almost orthogonal in the framework of weighted currents when the extremities are apart. This makes the definition of similarity twofold in the framework of weighted currents. Two streamlines are considered similar if their pathways are alike, as in usual currents, but also if their endpoints are close to each other. In Fig.3.2 we show the most similar fibers to the red streamline in the framework of currents (green) and in the one of weighted currents (blue). The green fibers share a similar pathway with the red one but some of them connect different anatomical areas. On the contrary, the blue fibers are similar to the red streamline both in terms of geometry and connectivity.

As usual currents, the framework of weighted currents does not need point-to-point correspondence, except for the extremities. This can be obtained, for instance, by tracing all the streamlines of a bundle from one ROI to another one, as it is done for the bundles considered in this chapter and throughout the Thesis. Moreover, every streamline S_i is considered as a vector in a Hilbert space. Thus, a fiber bundle, which is the union of many fibers $B = \cup_i S_i$, is represented as a sum in this framework: $C_B = \sum_i^N C_{S_i}$. The difference between two streamlines is defined as their sum with the orientation of the second fiber inverted. If two fibers are equal, their difference cancels out. Furthermore, it is also possible to compute the average weighted current $\bar{S}$ of a fiber bundle as: $C_{\bar{S}} = \frac{1}{N}\sum_i^N C_{S_i}$. Given the inner product defined in Eq.3.2, the squared norm of the difference between

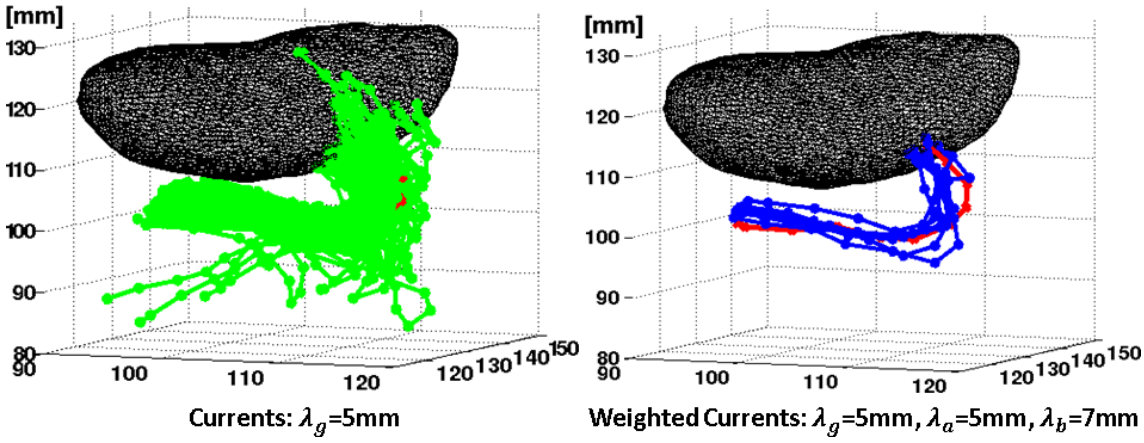

Figure 3.2: Streamlines that have an angle smaller than 45 degrees with the red one using currents (green, #118) and weighted currents (blue, #8). Green streamlines are more spread than the blue ones, connecting anatomical locations far from the ones of the red fiber. The concept of similarity in the framework of weighted currents is more stringent than using usual currents.

two bundles $C_B = \sum_{i=1}^{N} C_{S_i}$ and $C_{B'} = \sum_{p=1}^{M} C_{S'_p}$ is equal to: $||C_B - C_{B'}||^2_{W^*} = \sum_{i=1}^{N}\sum_{j=1}^{N}\langle C_{S_i}, C_{S_j}\rangle_{W^*} + \sum_{p=1}^{M}\sum_{q=1}^{M}\langle C_{S'_p}, C_{S'_q}\rangle_{W^*}$ - 2 $\sum_{i=1}^{N}\sum_{p=1}^{M}\langle C_{S_i}, C_{S'_p}\rangle_{W^*}$.

A bundle B composed of two streamlines X and Y is modelled as $C_B = C_X + C_Y$ in the framework of weighted currents. If the two streamlines X and Y are similar in this framework, their sum can be well approximated by $C_B = 2C_X$ or $C_B = 2C_Y$. This is crucial for the scope of this chapter since an ensemble of streamlines can thus be represented with a single *weighted prototype* where the weight is related to the number of streamlines approximated. In the previous example both X and Y could be chosen as prototype and the weight would be 2. A weighted prototype can be visualised as a tube where the streamline chosen as prototype is the central axis and the constant radius is proportional to the weight (see Fig.3.3). In the following, we will describe how to use this idea to approximate a complex bundle stemming from a tractography algorithm. We will also assume that both streamlines and bundles are modelled as weighted currents writing simply S (resp. B) instead than C_S (resp. C_B).

3.4.3 Approximation scheme

The goal of the proposed approximation scheme is to represent a fiber bundle B with a set of weighted prototypes $\{\tau_k P_k\}$. The resulting parsimonious representation should preserve both the shape and the structural connectivity of the original bundle. It is based on a *greedy approach* where we first subdivide the bundle into fascicles and then select the prototypes in each fascicle independently.

Fascicles detection A fascicle is a group of streamlines which are considered similar in the framework of weighted currents, namely they have a similar pathway and end-points close to each other. The subdivision of a bundle into fascicles is based on the maximization of a quality function called *modularity*, a concept borrowed from network theory (Blondel et al., 2008):

$$Q = \sum_{F=1}^{N_F} \left\{ \| \sum_{i \in F} S_i \|_{W^*}^2 \| \sum_{j \notin F} S_j \|_{W^*}^2 - (\sum_{i \in F} \sum_{j \notin F} \langle S_i, S_j \rangle_{W^*})^2 \right\} \tag{3.3}$$

where F is a fascicle, N_F is the number of fascicles and it is constrained by $\sum_{F=1}^{N_F} \sum_{i \in F} S_i = B$. In the simple case of N_F=2, Eq.3.3 can be rewritten as: $Q = \|\bar{S}_1\|_{W^*}^2 \|\bar{S}_2\|_{W^*}^2 - \langle \bar{S}_1, \bar{S}_2 \rangle_{W^*}^2$ where $\bar{S}_1$ and $\bar{S}_2$ are the averages of the two fascicles. Maximizing Q means therefore dividing the bundle into two fascicles which tend to be orthogonal to each other and such that their averages have a similar norm. In the general form of Eq.3.3, one looks for N_F fascicles with balanced norms and which tend to have streamlines orthogonal to the streamlines of the other fascicles and parallel to the streamlines of their own fascicle.

Modularity is often employed in the field of complex networks to detect densely connected communities of nodes within a network (Blondel et al., 2008). It has been demonstrated that exact modularity optimization is strongly NP-complete (Brandes et al., 2006). Several approximation schemes exist in the literature and one of the state-of-the-art methods is the "Louvain" algorithm (Blondel et al., 2008). It is a greedy solution where every fascicle is considered as a vertex of a graph. Two vertices F_1 and F_2 have a weighted edge equal to the sum of the inner products between the streamlines of the fascicles $\sum_{i \in F_1} \sum_{j \in F_2} \langle S_i, S_j \rangle_{W^*}$. At the beginning, every streamline is considered as an independent fascicle. The algorithm is divided into two parts which are repeated iteratively. The first part consists of moving all the streamlines of a vertex to its neighbour vertices finding the relocation that leads to the greatest increase in modularity. If none movement produces a positive gain in modularity, the streamlines remain in their initial vertex. This part is repeated until no change would produce an increase in the modularity. In the second part, one redefines the graph by discarding the empty vertices and recomputing the weighted edges between the new vertices. The two steps are repeated until no change would produce an increase in modularity. At the end of this algorithm the fiber bundle is separated into different fascicles without fixing in advance neither the number of fascicles nor the number of streamlines in each fascicle.

Prototypes Streamline Selection (PSS) Once defined the fascicles, a PSS is performed in each fascicle independently. We propose an iterative algorithm in the spirit of orthogonal matching pursuit (Tropp and Gilbert, 2007). Let F be a fascicle with L streamlines modelled as weighted currents, the first prototype P_1 is chosen as the streamline minimising the residual squared error, namely: $P_1 = \arg\min_{S_i} ||F - \tau_1 S_i||_{W^*}^2$. Since the space of weighted currents is a vector space, we

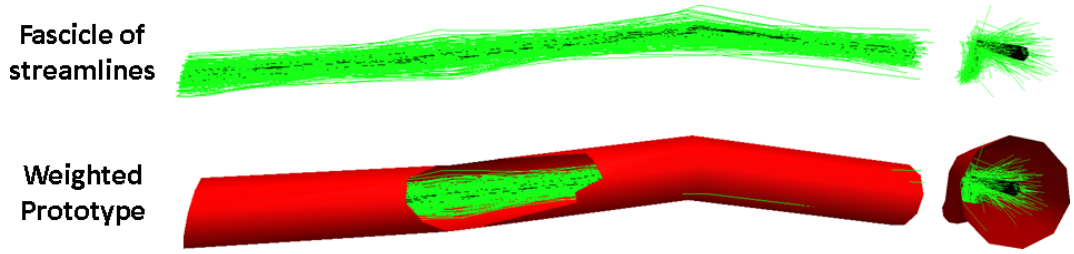

Figure 3.3: Example of weighted prototype visualised as a red tube. It approximates the fascicle of streamlines coloured in green. The streamline chosen as prototype is the central axis of the tube and it is coloured in black. The radius of the tube is proportional to the weight of the prototype and it does not represent the spatial coverage of the prototype. The spatial coverage is the same for all prototypes and it depends on the three bandwidths λ_g, λ_a and λ_b.

can easily minimize it. The optimal weight is: $\tau_1 = \frac{\langle F,P_1\rangle_{W^*}}{||P_1||^2_{W^*}}$ and the prototype is: $P_1 = \arg\max_{S_i}\langle F, \frac{S_i}{||S_i||_{W_*}}\rangle^2_{W^*} = \arg\max_{S_i} L^2\langle \bar{S}, \frac{S_i}{||S_i||_{W^*}}\rangle^2_{W^*}$ with i=1,...,L. The prototype is therefore the most parallel streamline to the average $\bar{S}$ of the fascicle.

Once the first prototype is selected, we remove from each streamline S_i its orthogonal projection onto the prototype, resulting in the residual: $r(S_i) = S_i - \pi(Si) = S_i - \frac{\langle S_i,P_1\rangle_{W_*}P_1}{||P_1||^2_{W_*}}$. We keep therefore only the components of the streamlines orthogonal to the prototype P_1. In this new space, we select the second prototype as: $P_2 = \arg\max_{r(S_i)}\langle r(F), \frac{r(S_i)}{\|r(S_i)\|_{W_*}}\rangle^2_{W_*}$. We iterate this process until: $||F - \sum_{k=1}^{K}\tau_k P_k||_{W_*} \leq \gamma||F||_{W_*}$ where $||F||_{W_*}$ is the norm of the fascicle, K is the number of prototypes and γ indicates the required approximation level. At each iteration t, the set of weights $\{\tau_k\}_{k=1,...,t}$ is computed as the orthogonal projection of all the streamlines of F to the space spanned by the selected set of prototypes $\{P_k\}_{k=1,...,t}$.

It is important to notice that all these computations are based on the Gram matrix Γ of the fascicle F which has size $[L\text{x}L]$. Thus, instead of computing directly $r(S_i)$, we simply update Γ as: $\langle r(S_i), r(S_j)\rangle_{W_*} = \langle S_i, S_j\rangle_{W_*} - \frac{\langle S_i,P\rangle_{W_*}\langle S_j,P\rangle_{W_*}}{||P||^2_{W_*}} = \Gamma_{(i,j)} - \frac{\Gamma_{(i,P)}\Gamma_{(j,P)}}{||\Gamma_{(P,P)}||^2_2}$. A sketch of the algorithm can be found in *Algorithm* 1 where $\Gamma_{(i,j)}$ indicates the value of the matrix Γ at row i and column j, $\Gamma_{(K,L)}$ refers to the submatrix of Γ containing the K rows of the prototypes and all the L columns, $\Gamma_{(K,K)}$ is the square submatrix with the rows and columns of the K prototypes and $\mathbf{1}_{(L,1)}$ is a L-dimensional column vector of ones.

After selecting the prototypes of each fascicle independently, they are all gathered into a single bundle of prototypes B_P. The weights are then recomputed as the orthogonal projection of the whole bundle B to the entire set of prototypes B_P in order to retrieve the correct values also for the weights of the prototypes close to the boundary between two different fascicles. Moreover, before the PSS algorithm, we also perform an outlier detection step in every fascicle. The streamlines char-

Algorithm 1 Prototype Fiber Selection

Input: Fascicle F with L streamlines S_i: $F = \sum_{i=1}^{L} S_i$;
threshold γ ; pre-computed matrix Γ ; K=1

$P_1 \leftarrow \arg\max_{S_i} \langle F, \frac{S_i}{||S_i||_{W^*}} \rangle^2_{W^*}$
$\tau_1 \leftarrow \frac{\langle F, P_1 \rangle_{W^*}}{||P_1||^2_{W^*}}$
for $i = 1$ **to** L **do**
  for $j = 1$ **to** L **do**
    $\langle S_i, S_j \rangle_{W^*} \leftarrow \langle S_i, S_j \rangle_{W^*} - \frac{\langle S_i, P_1 \rangle_{W^*} \langle S_j, P_1 \rangle_{W^*}}{||P_1||^2_{W^*}}$
  end for
end for
while $||F - \sum_{k=1}^{K} \tau_k P_k||_{W_*} \leq \gamma ||F||_{W_*}$ **do**
  $K \leftarrow K + 1$
  $P_K \leftarrow \arg\max_i \frac{(\Gamma_{(i,L)} \mathbf{1}_{(L,1)})^2}{\Gamma_{(i,i)}}$
  $\{\tau_k\}_{k=1,...,K} \leftarrow \frac{\Gamma_{(K,L)} \mathbf{1}_{(L,1)}}{\Gamma_{(K,K)}}$
  for $i = 1$ **to** L **do**
    for $j = 1$ **to** L **do**
      $\Gamma_{(i,j)} \leftarrow \Gamma_{(i,j)} - \frac{\Gamma_{(i,P_K)} \Gamma_{(j,P_K)}}{||\Gamma_{(P_K,P_K)}||_2^2}$
    end for
  end for
end while

Output: $\{\tau_k\}$, $\{P_k\}$ k=1,...,K

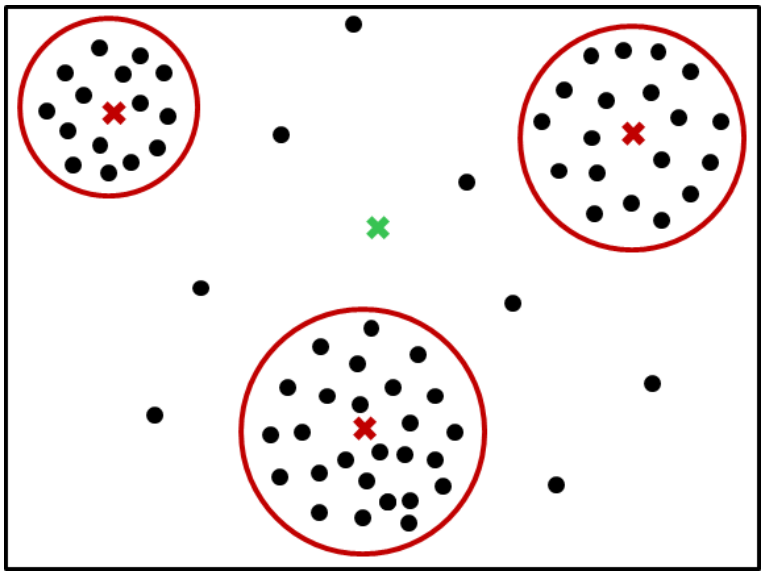

Figure 3.4: Visual explanation of the bundle subdivision into fascicles (modes). Dots and crosses represent the streamlines of a fiber bundle in the space of weighted currents. The green cross is the most similar streamline to the average of the bundle. It is far from almost all the other streamlines and it could be considered as an outlier. The three circles represent the fascicles (modes) composed by similar fibers. The red crosses are the prototypes of the fascicles considered independently. These fibers are more representative than the green cross and they better approximate the fiber bundle.

acterised by an average angle with the other streamlines between 88° and 90° are considered as outliers and discarded from the analysis.

Remark Performing the PSS in each fascicle independently allows us to distribute the computations to different processors, reducing therefore the computational time. Moreover, it also decreases the chance to select as prototype a streamline which might be considered an outlier. This is explained in Fig.3.4 where every dot represents a streamline of a bundle modelled as weighted current. We oversimplify this space assuming it is simply $\mathbb{R}^2$. In this space, it is likely that a tractography bundle has a multi-modal distribution, where every mode is a fascicle. If we wanted to approximate the whole bundle with a single prototype, it would be the most parallel streamline to the average of the bundle. In Fig.3.4 we would choose the streamline represented by the green cross. This fiber is far from almost all the other streamlines and it could be considered as an outlier. Instead, if we apply the same selection process in each fascicle independently we would obtain the three prototypes highlighted in red. These streamlines are more representative than the green fiber and they better approximate the bundle. Obviously, an actual fascicle does not lie in a 2D space and therefore we may need more than one prototype to explain its variability.

3.5 Experiments and Results

In this section, we first describe the dataset used in the following experiments and some technical details about the implementation of the proposed algorithm. Then, we present the approximation of two probabilistic bundles and we show that their structural connectivity is similar to the one of the original bundles. Furthermore, we qualitatively evaluate the effect of the parameters on the approximation of a deterministic bundle. Eventually, we assess quantitatively the performance of our algorithm showing that the registration between two approximated fiber bundles is definitely faster than using the original bundles for the same registration accuracy.

3.5.1 Materials

We test the proposed approximation scheme on 25 subjects. Diffusion weighted scans are acquired with sequences of 50 directions with a B-factor of 1000 and a voxel size of 2x2x2 mm^3. We use the Spherical Deconvolution Transform (SDT) model (Descoteaux et al., 2009) to estimate the local underlying orientation distribution function (ODF). Whole brain connectivity is then inferred within an anatomy-based tractography mask (Guevara et al., 2011) using both a deterministic (1 seed per voxel) and a probabilistic (8 seeds per voxel) tractography algorithm available in BrainVISA/Connectomist-2.0 (Perrin et al., 2005). In this chapter, we consider three distinct fiber bundles connecting the left hemisphere of the cortical surface to the left thalamus, putamen and caudate respectively. We extract them from both the deterministic and probabilistic whole brain tractography as explained in Worbe et al. (2015). All bundles also include the commisural fibers which are truncated at the inter-hemispheric plane. All the other fibers are cut at the intersection with their respective sub-cortical structure and at the border between white and gray matter of the cortex. Sub-cortical structures are segmented with FSL (Patenaude et al., 2011) from 3D T_1-weighted images (voxel size: 1x1x1 mm^3) and we merge the segmentations of nucleus accumbens and caudate in order to consider them as a single structure. The 3D meshes are created using the marching cubes algorithm of BrainVISA v4.4.0. The cortical surface is segmented using FreeSurfer v5.3 (Fischl et al., 2004). More information about the acquisition and the preprocessing of both T1-weighted and diffusion-weighted images can be found in Worbe et al. (2015).

3.5.2 Numerical aspects

The parameters needed to be fixed by the user are the bandwidths of the three kernels of weighted currents λ_g, λ_a and λ_b and the approximation level γ. In the following, λ_g refers to the kernel of usual currents, λ_a to the end-point on the nuclei and λ_b to the end-point on the cortical surface.
All experiments shown in this chapter are computed on a Intel Xeon, 8 cores, CPU E5-1620, 3.60 GHz with a memory of 16Gb and a graphic card NVIDIA Quadro K4000.

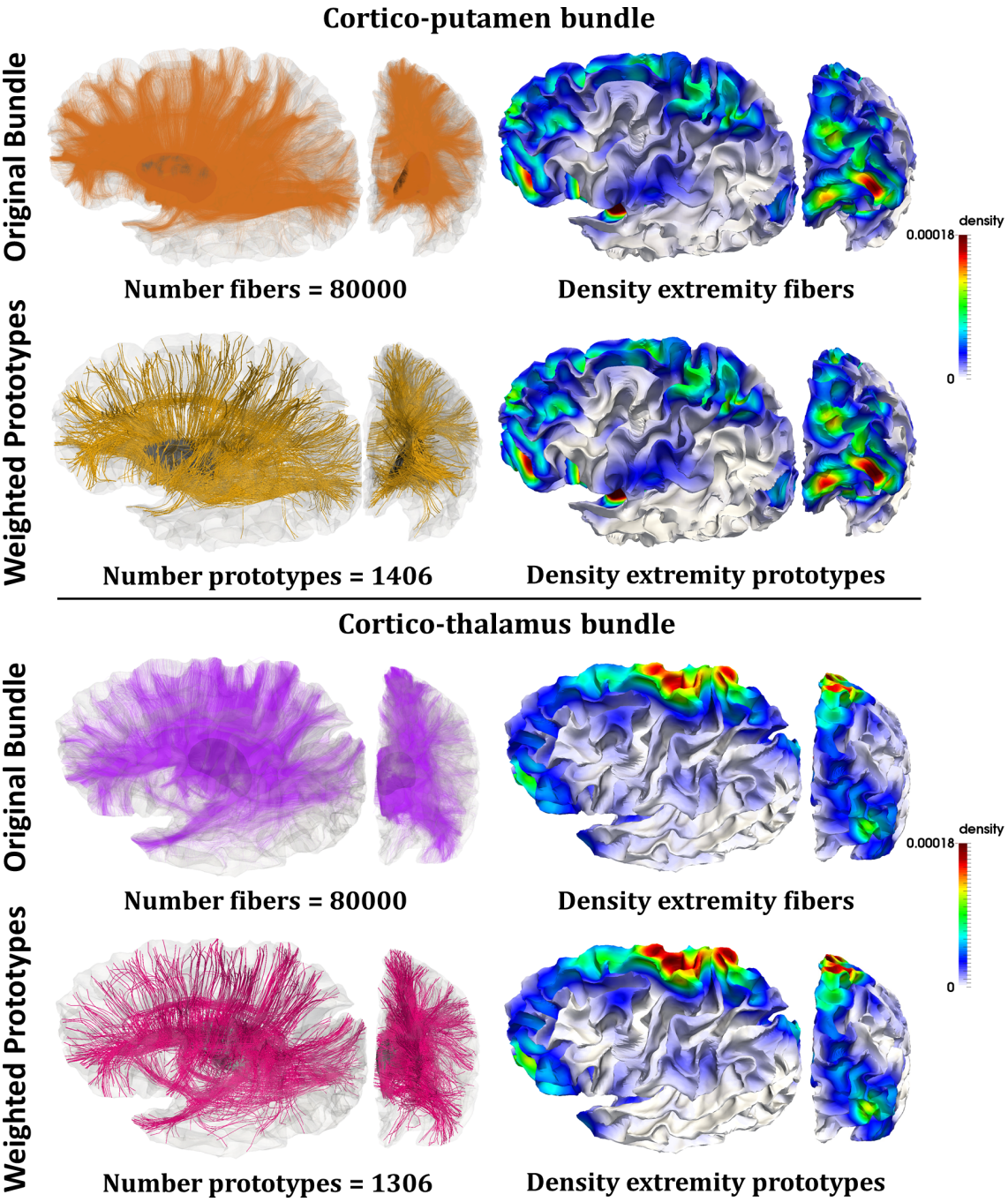

Figure 3.5: Weighted prototype approximations of two probabilistic bundles: a cortico-putamen and a cortico-thalamus. As it is possible to notice, our approximation alters neither the global shape of the bundle nor the densities of the endpoints onto the cortical surface. We use: γ=0.13, λ_g=7mm, λ_a=5mm and λ_b=10mm.

The computational times for the approximations of the fiber bundles shown in

Deterministic			
Bundle	Cortex-caudate	Cortex-putamen	Cortex-thalamus
N Streamlines	17079±4881	28056±5247	28371±6806
N Prototypes	344±58	409±55	341±87
Compression	97.85 %	98.49 %	98,77 %

Table 3.1: Average compression ratio (%) of all the bundles obtained from the deterministic tractography algorithm

Probabilistic			
Bundle	Cortex-caudate	Cortex-putamen	Cortex-thalamus
N Streamlines	75389±4646	78125±2223	68640±8568
N Prototypes	1182±358	1411±393	1000±319
Compression	98,41 %	98,19 %	98,54 %

Table 3.2: Average compression ratio (%) of all the bundles obtained from the probabilistic tractography algorithm

Fig.3.5 and in Fig.3.6 composed of 80.000 and 35.674 were of 150 and 19 minutes respectively.

3.5.3 Weighted prototypes representation

We present in Fig.3.5 the weighted prototype approximations of two probabilistic bundles using γ= 0.13, λ_g=7mm, λ_a=5mm and λ_b=10mm. The proposed representation preserves the global shape of the bundle and it approximates thoroughly all the fascicles. We also show the probability densities of the endpoints (i.e. structural connectivity) on the cortical surface computed using either the streamlines of the original bundles or the weighted prototypes. The two densities are very similar from a qualitative point of view and the Kolmogorov-Smirnov test fails to show statistically significant differences between them at the 5% level. Probability densities are computed using Gaussian kernels, taking into account the weights of the prototypes for the proposed approximation. Furthermore, we use the same parameters to approximate all the bundles of our data-set (75 deterministic and 75 probabilistic) and in no case the density of the endpoints is statistically significantly different from the one of the original bundle. The average compression ratios, 100(1- K/N) where N is the number of streamlines of the original bundles and K is the number of prototypes, are shown in Table 3.5.3 and 3.5.3 for the deterministic and probabilistic bundles respectively. These results show that our algorithm leads to a much more compact representation of the bundle while preserving the overall structural connectivity, which may be used for further analysis.

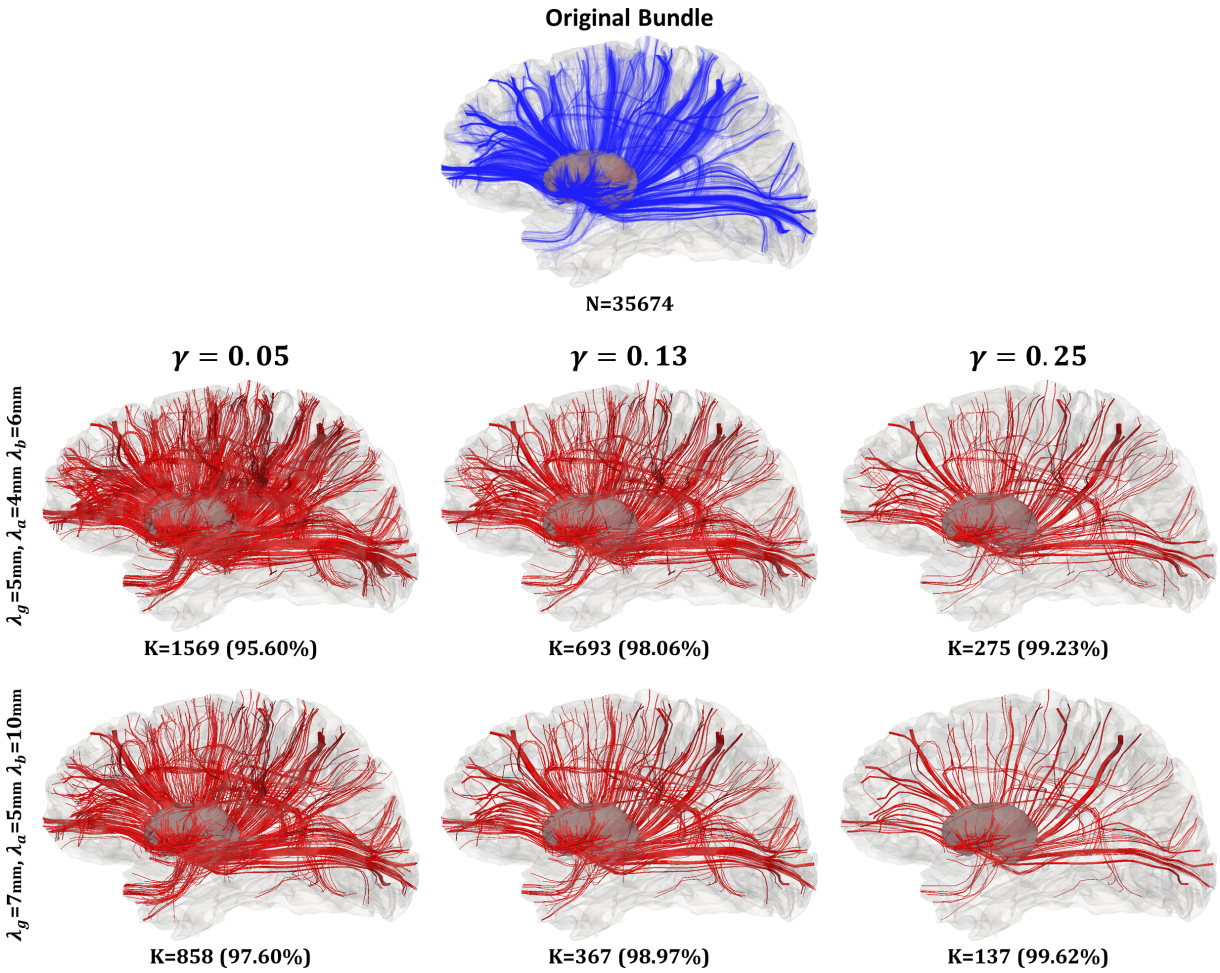

Figure 3.6: Weighted prototype representations (in red) at different approximation levels of a deterministic cortico-putamen bundle (in blue). Each row is based on a different set of kernel bandwidths. The letters N and K refer respectively to the number of streamlines of the bundle and to the number of prototypes. The compression ratios are indicated in brackets.

3.5.4 Qualitative evaluation of the parameters influence

In Fig.3.6 we evaluate the influence of the parameters of our algorithm on a deterministic cortico-putamen bundle. In the first row we employ λ_g=5mm, λ_a=4mm, λ_b=6mm and in the second row λ_g=7mm, λ_a=5mm, λ_b=10mm. Every column corresponds to a different approximation level. It can be noticed that at γ=0.05 and γ=0.13 all fascicles are well approximated, whatever the set of parameters. At γ=0.25 and higher values of γ (not shown here) only the denser parts of the fascicles are kept. Moreover, the results based on the first row use almost twice the prototypes than in the second row. This is expected since the values of the kernels are smaller and therefore the definition of similarity between two streamlines is more

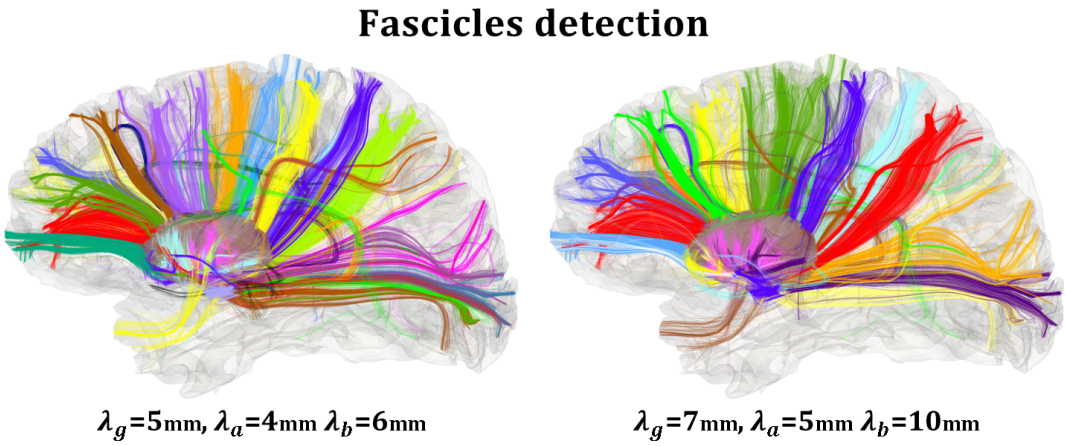

Figure 3.7: Results of the fascicles detection step applied to the bundle shown in Fig.3.6 using the two sets of kernels' bandwidths. The number of fascicles are respectively: 65 (left) and 35 (right). Colours are chosen randomly.

stringent. Thus, for a given γ, one needs more prototypes to approximate the same number of streamlines. Furthermore, this also influences the number/size of the fascicles, as shown in Fig.3.7. The fascicles obtained with the first set of parameters are smaller in size and greater in number than the ones obtained using the second set of parameters.

3.5.5 Registration-based evaluation of the algorithm

Here we evaluate the impact of the proposed approximation scheme on the quality of a registration between two deterministic cortico-putamen bundles of different subjects. We use the diffeomorphic transformation implemented in the software *Deformetrica* (`www.deformetrica.org`) with a kernel bandwidth equal to 10mm and 1309 control points. The source bundle B_S is the one shown in Fig.3.6. First, we approximate both B_S and the target bundle B_T, composed of 25916 streamlines, at different approximation levels using λ_g=7mm, λ_a=5mm, λ_b=10mm. Then, for each level, we register the approximation of B_S onto the one of B_T. We apply then the obtained deformation to the original bundle B_S and we measure the residual error between the transformed original source bundle $\phi(B_S)$ and the original target bundle B_T: $||\phi(B_S) - B_T||^2_{W_*}$ in the framework of weighted currents. Ideally, we would compare the obtained residual errors with the one of the registration between the original fiber bundles. Unfortunately, the computational time would be too long (see Table 3.5.5) and therefore we decide to use the registrations between smaller sub-samples of the original fiber bundles for comparison. In Fig.3.8 we show the deformations of the original fiber bundles using the transformations obtained with the approximation at γ=0.13 and with the sub-sample of 5000 streamlines. It is possible to notice that the results look very similar. This is confirmed in Table 3.5.5 where we show that the difference between their residual errors is very small.

	Registration Error	Computational Time
γ=**0.40**	1.75e+09	2h 39min
γ=**0.25**	1.42e+09	4h 34min
γ=**0.13**	9.88e+08	5h 51min
γ=**0.05**	9.79e+08	18h 35min
1000 streamlines	9.96e+08	23h 12min
5000 streamlines	9.94e+08	547h 32min
8000 streamlines	-	$\sim$ 1120h
15000 streamlines	-	$\sim$ 4484h
30000 streamlines	-	∞

Table 3.3: Registration error and computational time using different approximation levels and samples of the original bundles.

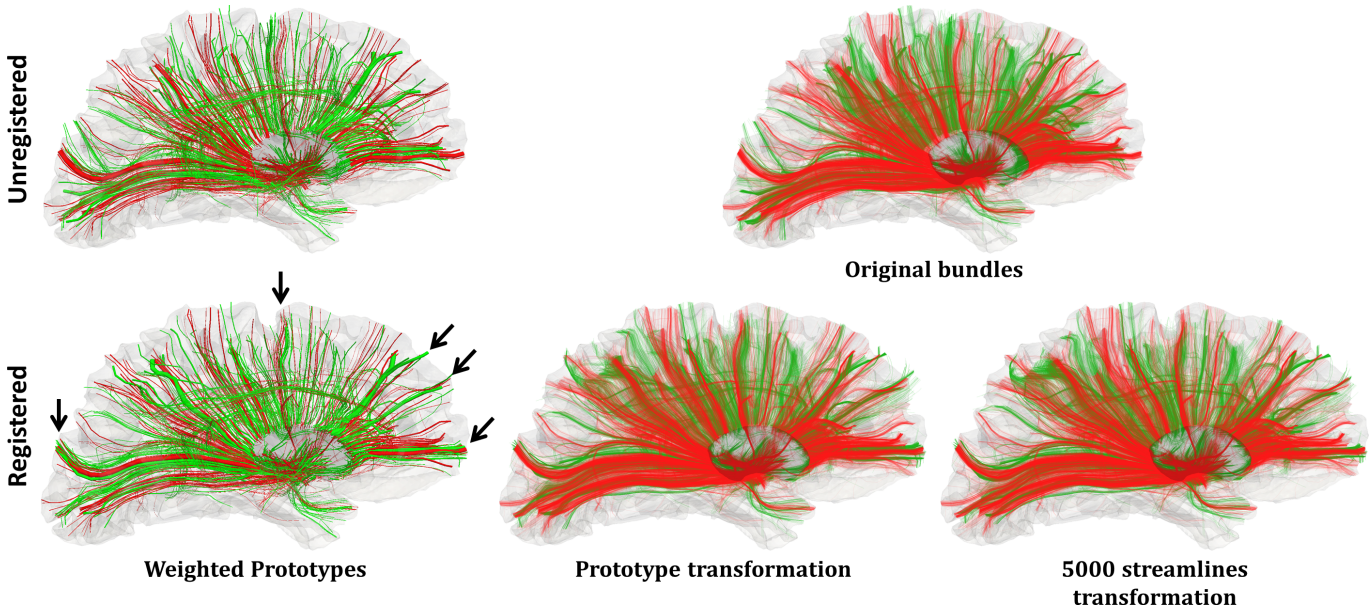

Figure 3.8: On the left: registration between the green source bundle and the red target bundle approximated with weighted prototypes. On the right: original fiber bundles and the transformations based on the approximation at γ=0.13 and with the sub-sample of 5000 streamlines. Black arrows highlight the areas where the alignment is more noticeable.

We can conclude that the registration based on the approximation at γ=0.13 is as accurate as using the sub-sample of 5000 streamlines but 93 times faster! We also present the results for other approximation levels. Compared to γ=0.13, the other registrations are either less accurate or slower and with a similar accuracy.

3.6 Discussion and Conclusions

We presented an algorithm to approximate a fiber bundle with a small and scattered set of weighted streamlines prototypes. We tested it on 150 bundles resulting from both deterministic and probabilistic tractography algorithms. The number of prototypes was on average 2% of the number of streamlines of the original bundles. We showed that such a parsimonious representation preserves both the shape and the structural connectivity of the bundles.

The streamlines considered in this chapter have been truncated at the intersection between gray and white matter which is an area usually characterised by a low Signal to Noise Ratio. This means that the estimates of the end-points of the streamlines are not very robust. To account for this uncertainty, we use Gaussian kernels for measuring the dissimilarity between two streamlines. Theoretically, the greater the uncertainty, the greater the bandwidths of the kernels. In the proposed algorithm, these bandwidths (i.e. λ_a and λ_b) are considered as parameters fixed by the user. Their values are chosen by looking at how much the streamlines fan out when approaching to the boundary between white and grey matter. Streamlines deviate more when they are close to the cortex than to the sub-cortical nuclei, which explains why λ_b is always greater than λ_a in our experiments. It would be of interest to automatically estimate these parameters taking into consideration the type of bundle, the SNR of the diffusion image, the tractography algorithm and the diffusion model.

Another parameter fixed at the beginning of the algorithm is the approximation level γ. It defines the stopping criteria and the value $(1-\gamma)$ is the minimal percentage of the norm of the fascicle explained by the prototypes. In fact, thanks to the triangle inequality, we can rewrite the stopping criteria obtaining $(1-\gamma)||F||_{W_*} \leq ||\sum_{k=1}^{K} \tau_k P_k||_{W_*} \leq (1+\gamma)||F||_{W_*}$. This means that, using γ=0.13, the norm of the prototypes will be at least 87% of the norm of the fascicle at the end of the algorithm. Furthermore, we noticed that between γ=0.5 and γ=0.01 the number of prototypes grows exponentially. For instance, using the set of parameters of Fig.3.5, the number of prototypes is 70 at γ=0.5, 367 at γ=0.13 and 2307 at γ=0.01. This means that we need less prototypes to reduce γ from 0.5 to 0.13 than from 0.13 to 0.01. This is because the first prototypes approximate the parts of the fascicle with an higher density of streamlines (i.e. greater redundancy). Thus, their weights have a great value and few prototypes can explain a considerable percentage of the norm of the fascicle. Instead, between γ=0.1 and γ=0.01, most of the streamlines have already been approximated and every new prototype can explain only a few of the remaining fibers. We found that a value of γ=0.13 results in a parsimonious representation which exhaustively approximates all the fascicles. An interesting improvement would be to automatically estimate the value of γ taking into consideration both the approximation error and the number of prototypes.

We showed also that our representation can approximate not only the central and more dense mass of a bundle, as usual currents, but also its smaller fascicles and extremities. We demonstrated its usefulness in the registration between two

approximated bundles where we correctly match the whole pathway of the weighted prototypes, including their extremities (see Fig.3.8). Furthermore, bundle registrations based on our approximation scheme present a lower computational time and memory footprint than using the original fiber bundles. This makes thus possible population studies, like the atlas construction, based on multiple template-to-subject non-linear registrations, which would be very time-consuming - or even infeasible - with the original fiber bundles.

3.A Appendix

3.A.1 Interpretation of τ

We show here that the weight τ of a prototype P is related to the number of fibers approximated by P. Given a bundle B composed of 3 fibers: $B = \sum_{i=1}^{3} S_i$ we want to approximate it with one prototype P. Let assume that S_1 is the prototype P (the reasoning does not change modifying the prototype), the value of its weight τ is $1 + \frac{||S_2||^2_{W^*}}{||P||^2_{W^*}} \cos(P, S_2)_{W^*} + \frac{||S_3||^2_{W^*}}{||P||^2_{W^*}} \cos(P, S_3)_{W^*}$. This means that if S_1 is parallel to the other fibers and their norms are similar, the value of τ will be about 3. Instead, if either S_2 or S_3 is orthogonal to S_1, the prototype will not approximate that fiber and the weight will be smaller than 3. This shows that τ is related to the number of fibers approximated by the prototype or, more precisely, to how much similar is the prototype with respect to the other fibers of the bundle. When dealing with more prototypes, every τ also depends on the inner product between the prototypes. Since we project at each iteration all the streamlines onto the orthogonal space of the last estimated prototype, the inner product between prototypes should be small.

3.A.2 Modularity based on weighted currents

The definition of modularity in Blondel et al. (2008) is:

$$Q = \sum_{c=1}^{N_C} \left[\frac{W_c}{m} - \left(\frac{S_c}{2m} \right)^2 \right] \tag{3.4}$$

where N_C is the number of modules (fascicles), $W_c = 1/2 \| \sum_{i \in c} S_i \|^2_{W^*}$ is the sum of the weights of all the edges joining only the vertices of module c, $S_c = \sum_{i \in c} \sum_{j=1}^{N} \langle S_i, S_j \rangle_{W^*}$ is the sum of the weights of the edges between the vertices in c and all the N vertices in the graph and $m = 1/2 \| \sum_{i=1}^{N} S_i \|^2_{W^*}$ is the sum of the weights of all edges in the graph. Substituting these equations in Eq.3.4 and noting that m is a constant term and that it can be rewritten as $2m = \| \sum_{i \in c} S_i \|^2_{W^*} + \| \sum_{j \notin c} S_j \|^2_{W^*} + 2 \sum_{i \in c} \sum_{j \notin c} \langle S_i, S_j \rangle_{W^*}$, one obtains Eq.3.3.

Chapter 4

A deformation framework to unify morphometry and structural connectivity analysis

This chapter will be submitted to NeuroImage and it has been partly published in Gori et al. (2015a). *The C++/CUDA code of the described double diffeomorphic deformation will be soon integrated to the software Deformetrica.*

Contents

4.1 Introduction

The pathophysiology of neurodevelopmental disorders such as autism, attention deficit-hyperactivity disorder (ADHD) and Gilles de la Tourette syndrome (GTS) involves morphological alterations of the cortico-basal ganglia and cortico-thalamus neural circuits (Tye and Bolton, 2013; Konrad et al., 2010; Worbe et al., 2015). These networks are composed of neural projections linking particular areas of the cortical surface to specific parts of the sub-cortical nuclei. Abnormalities can affect: i) the shape of the single components of the circuits, including both gray matter structures and white matter tracts, ii) the relative position among gray matter structures and iii) the structural connectivity, namely the areas where gray matter integrate white matter tracts. Most of the studies present in the literature focus only on the first or last point. Few of them propose to analyse the first two points together and, to the best of our knowledge, no method has been proposed to tackle the three points in a unified framework.

In order to analyse the relative position between different structures, one needs to study the neural circuits as a whole. This requires a single framework where both grey and white matter components are studied together and at the same time. Grey matter structures are usually analysed with T1-w scans whereas white matter tracts using dMRI scans. As already explained in the Introduction, we integrate the information from these two modalities by shifting the morphological analysis from images to complexes of 3D meshes, called *shape complexes*, representing the different components of the neural circuits. Structures of the gray matter such as cortical surface and basal ganglia are represented as closed surfaces segmented from T1-w scans whereas the neural projections of the white matter are modelled as bundles of 3D streamlines, called fiber bundles, resulting from tractography algorithms applied on diffusion scans.

Different strategies exist for the statistical shape analysis of a population of shape complexes, representing each one the structures of a subject. One which naturally allows the combination of different mesh types is based on the Grenander's pattern theory (Grenandner, 1993). Every shape complex is modelled as a deformation of a reference complex called *template complex*. The deformations are defined on the whole ambient space where the template complex lies and they put into correspondence the template complex with the homologous structures of the subjects. First examples used a fixed template chosen a priori like the Talairach atlas (Davatzikos et al., 1996) or a randomly-selected subject (Csernansky et al., 1998). However, this kind of templates might have imperfections due to segmentation errors and they might not be "centred" with respect to the population under study, biasing consequently the whole analysis. Later works have thus proposed to automatically estimate the template (Joshi and Miller, 2000; Durrleman et al., 2014; Qiu et al., 2010). The joint estimate of the template complex and the deformations has been called *atlas construction*. The template complex captures the common morphological characteristics of the population. The set of deformations explain the variations in shape across the population.

4.1. Introduction

Deformations are defined as diffeomorphisms of the whole ambient space which are smooth invertible transformations with smooth inverse. This kind of deformation preserves the anatomical organisation of the components of the template complex, namely they can not intersect, fold or shear. Moreover, since the transformations are defined locally and they can vary across different areas in the ambient space, it is also possible to capture the variations in relative position between *separated* structures. However, using a single diffeomorphism we implicitly assume that the relative position between structures *in contact* with each other or close to each other does not change across subjects. This implies that a particular fiber bundle should link the same areas of the cortical surface and basal ganglia in each subject within the population. This assumption precludes the study of changes in structural connectivity which could be caused by an abnormal brain development. In Fig.4.1 we present a toy example composed of a template complex and a subject shape complex characterised by a different structural connectivity. A single diffeomorphism could not put into correspondence all the structures and capture the differences in structural connectivity.

Structural connectivity analysis is usually tackled by dividing the cortical surface and the sub-cortical nuclei in parcels based on anatomical, histological or functional schemes reproducible across subjects (Craddock et al., 2013). Every parcel is then considered as a node of a graph and the number of streamlines connecting two nodes represents their weighted edge. Variability in structural connectivity across subjects can be defined by simply analysing independently the variations in each parcel or by using indexes and methods from the complex network theory (Bullmore and Sporns, 2009; Worbe et al., 2015). In both cases, the analysis depends on the chosen parcelling scheme and it does not take into consideration the morphological variability of the gray matter structures and the trajectories of the streamlines.

Here, we propose to join *shape* and *structural connectivity* analysis in a unified framework based on a *double diffeomorphic* atlas construction. The template complex is warped towards every shape complex of the population using a cascade of two diffeomorphisms. The first diffeomorphism acts only on the white matter of the template complex keeping fixed the gray matter. It can be seen as a relative change of coordinates within the gray matter which is considered as a fixed reference frame. During this deformation the fiber bundles are "detached" from the fixed gray matter structures and "re-attached" in another location, capturing the variation in structural connectivity. These variations can be compared among subjects since they are all computed with respect to the same reference frame, namely the fixed gray matter of the template complex. The second diffeomorphism acts on the whole template complex, namely on both the resulting white matter and on the gray matter. This is a global change of coordinates which has the role to put into correspondence both gray and white matter structures of the template complex with the homologous ones of the subject's shape complex. The first diffeomorphism re-positions the white matter tracts within the gray matter such that all the components of the template complex can be correctly put into correspondence by the second diffeomorphism. The two diffeomorphisms are optimised together minimising a single

cost function and using a gradient descent scheme. The data-term depends only on the deformed template complex obtained at the end of the second diffeomorphism. Looking again at the example in Fig.4.1, the first diffeomorphism would make the fiber bundle slide from the left gyrus to the right gyrus. The second diffeomorphism would then modify the shape of all structures producing an accurate matching. The first diffeomorphism would capture therefore the changes in structural connectivity whereas the second one would recover the global morphological differences.

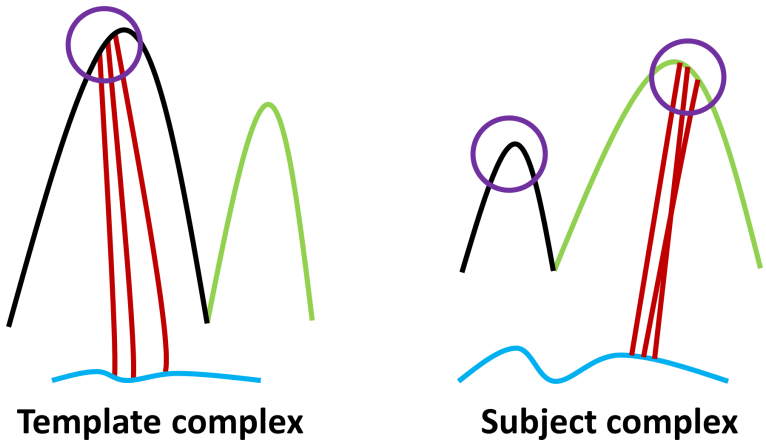

Figure 4.1: A template and a subject shape complex composed by a pseudo cortex, divided into a black and green gyrus, a blue pseudo sub-cortical nucleus and a red pseudo fiber bundle. A single diffeomorphism could not put into correspondence all the structures and capture the differences in structural connectivity. The points within the violet circle in the template complex would be matched either to the black gyrus of the subject shape complex or to its fiber bundle. A double diffeomorphism would first move the fiber bundle from the left to the right gyrus and then it would change the shape of all the structures, producing an accurate matching and capturing also the variation in structural connectivity.

It is important to remark that our approach is different from other multi-diffeomorphic methods with sliding conditions such as Risser et al. (2013); Pace et al. (2013); Arguillère et al. (2015). These methods aim to correctly register longitudinal scans or anatomical complexes characterised by sliding regions/components. This is achieved by ensuring that the components of the displacement field orthogonal to the motion boundaries vanish at the boundaries. In this way, every region deforms smoothly and *independently* from the others and discontinuities are allowed only at the boundaries. In our case, we are interested in studying the *relative variation* of one region, white matter, with respect to another one, gray matter. The aforementioned sliding registrations, if applied to the template complex shown in Fig.4.1, would result in two independent deformations, one for the white matter and one for the grey matter, and it would be impossible to understand whether the changes of the white matter depend on a difference in gray matter or in structural connectivity. Moreover, due to the sliding conditions, fiber bundles should always

stay in contact with the grey matter surface during deformation. This would require a lot of "deformation energy" to move the fiber bundle from the left gyrus to the right one and it would produce a very complicated deformation. Instead, with the proposed method the fiber bundle is not constrained to slide on the cortex and therefore the resulting deformation would be "simpler". Note that, we enforce that the template fiber bundle will be near the grey matter *at the end* of the double diffeomorphic deformation by modelling streamlines as weighted currents. As explained in Chapter 3, this is an extension of the framework of currents and two streamlines are considered similar not only if their pathways are alike but also if their endpoints are close to each other. This makes possible to match correctly also the extremities of two fiber bundles and not only their central part as in usual currents. The template fiber bundle is thus deformed in such a way that at the end of the second diffeomorphism its extremities match the ones of the subject bundle and therefore they are also close to the grey matter structures.

In the proposed method, both diffeomorphisms are parametrized using control points as presented in Durrleman et al. (2011b). The number of control points is fixed by the user and their positions are automatically adjusted close to the most variable parts of the template complex during the atlas construction.

In order to deal with the considerable amount of streamlines resulting from tractography algorithms, we rely on the approximation scheme introduced in Chapter 3. Fiber bundles are approximated with a parsimonious representation of weighted streamlines prototypes. Prototypes are chosen among the streamlines of the fiber bundle and they represent groups of similar streamlines. This approximation is controlled by the metric of weighted currents.

We propose to model gray matter structures as varifolds (Charon and Trouvé, 2013b) or landmarks if correspondences across subjects are available. The framework of varifolds is the non-oriented extension of the one of currents. It does not require neither point-correspondence nor a consistent orientation of the normal vectors among the population. Moreover, it prevents the "currents cancelling effect" which happens when two surface cells modelled as usual currents and with opposite orientation (i.e. a bump) cancel reciprocally their effect on the estimation of the template. A structure which is typically affected by this effect is the thin tail of the caudate nucleus.

The atlas is estimated within a Bayesian framework based on a generative statistical model similar to the one presented in Chapter 2 (Gori et al., 2013a; Allassonnière et al., 2007) and adapted to double diffeomorphisms. Every shape is modelled as a sum between a double diffeomorphism of the template complex and a noise. The noise and the deformation parameters of both diffeomorphisms are modelled as Gaussian random variables centred at zero. We use priors similar to the ones in Allassonnière et al. (2007) which enable us to automatically estimate the noise variance of each structure and a well-conditioned covariance matrix of the deformation parameters for both diffeomorphisms. The noise variances balance the importance of the structures in the cost function with respect to the other structures and with respect to the regularity terms of the deformations. As demonstrated in Chapter 2

the results are rather sensitive to their values and their automatic estimate becomes even more important increasing the number of structures.

This chapter is organised as follows. In Section 4.2, we first present a double diffeomorphic generative model which is integrated into a Bayesia framework for atlas construction. Afterwards, we recapitulate how to model surface meshes as varifolds (Charon and Trouvé, 2013b) and streamlines as weighted currents (Chapter 3) and how to define probability densities in these infinite-dimensional spaces. Then, we present how to compute double diffeomorphic deformations based on a control-points formulation of the LDDMM framework and how to integrate them in the proposed statistical setting. Furthermore, we use this method to compare the organisation and morphology of the cortico-putamen neural circuit of a population of healthy controls with the ones of GTS patients. Eventually, we show that the proposed double diffeomorphic scheme discriminates better between controls and GTS patients than using a single diffeomorphism.

4.2 Methods

4.2.1 Double Diffeomorphic Generative Model

The proposed atlas construction is based on a generative statistical model. We assume that the population under study is composed of N subjects. Every subject i is characterised of M structures segmented from both structural and diffusion images. We define the mesh of structure j belonging to subject i as S_{ij}. Every subject shape complex $\boldsymbol{S}_i$, defined for the moment in a generic way as the ensemble of all the meshes S_{ij} of subject i, is modelled as a *double deformation* of a common template complex $\boldsymbol{T}$ plus a residual noise $\boldsymbol{\varepsilon}_i$. Both $\boldsymbol{T}$ and $\boldsymbol{\varepsilon}_i$ are also defined as the ensembles of the templates T_j and residuals ε_{ij}. The first deformation ϕ^W acts only on the white matter of the template complex: $\boldsymbol{T}^W$. The second deformation ϕ^{All} deforms both the resulting white matter $\phi^W(\boldsymbol{T}^W)$ and the gray matter of the template complex $\boldsymbol{T}^G$. This formulation derives from the forward model where we assume that all elements belong to an algebraic structure where addition is defined(Allassonnière et al., 2007; Ma et al., 2008; Durrleman et al., 2008, 2009) and it results:

$$\boldsymbol{S}_i = \phi_i^{All}\left(\phi_i^W(\boldsymbol{T}^W) \cup \boldsymbol{T}^G\right) + \boldsymbol{\varepsilon}_i \tag{4.1}$$

The two deformations ϕ_i^W and ϕ_i^{All}, proper to subject i, are two diffeomorphisms of the entire ambient space. They follow one another creating a cascade of diffeomorphisms. The first one ϕ_i^W, called also white diffeomorphism, acts only on the white matter of the template complex. It represents a relative change of coordinates with respect to the gray matter of the template complex which is kept fixed. The second (or global) diffeomorphism ϕ_i^{All} transforms the whole template complex, both white and gray matter, and it is a global change of coordinates which brings the template complex to the subject space.

The first deformation ϕ_i^W makes the fiber bundles slide onto the gray matter structures. The streamlines are detached and re-attached in another area such that the whole template complex, both gray and white matter, can be well matched to the subject shape complex by the second diffeomorphism ϕ_i^{All}. The first diffeomorphism ϕ_i^W captures the difference in structural connectivity, namely the variations in the connected areas of the gray matter. The second diffeomorphism ϕ_i^{All} describes the global morphological changes of both white and gray matter.

It is important to notice that it is fundamental to first deform the white matter of the template complex and then the gray matter in order to retrieve the main variations in structural connectivity within the population. In fact, the transformations due to the first diffeomorphisms $\{\phi_i^W\}$ are comparable across subjects since they are all computed with respect to the same reference frame, namely the *fixed* gray matter of the template complex. If we changed the order, deforming first the gray matter and then the white one, we would not be able to compare the variations in structural connectivity since the reference frame given by the gray matter would be specific to each subject.

Furthermore, a question that naturally arises using the proposed method is about the uniqueness of the decomposition into two diffeomorphisms in the regions containing only white matter structures. In these areas, the white matter could be deformed in two different but equivalent ways. A possible solution to obtain a unique decomposition is to choose a scale of deformation so that white matter objects are deformed by the second diffeomorphism in a correlated way with respect to at least one gray matter structure.

4.2.2 Bayesian Atlas Construction

The goal of the atlas construction is to estimate the template complex $\boldsymbol{T} = \boldsymbol{T}^W \cup \boldsymbol{T}^G$, the variations in structural connectivity within the population described by the ensemble of white diffeomorphisms $\{\phi_i^W\}$ and the global morphological variations captured by the set of global diffeomorphisms $\{\phi_i^{All}\}$. Both diffeomorphisms are parametrised by a set of parameters called respectively $\boldsymbol{\alpha}_i^W$ and $\boldsymbol{\alpha}_i^{All}$ specific to each subject i. We assume that these parameters follow a Gaussian distribution with zero mean and covariance matrix equal to Γ_α^W and Γ_α^{All} respectively:

$$\begin{aligned} &\boldsymbol{\alpha}_i^W \sim N(0, \Gamma_\alpha^W) && p(\boldsymbol{\alpha}_i^W | \Gamma_\alpha^W) \propto \frac{1}{|\Gamma_\alpha^W|^{1/2}} \exp\left[-\frac{1}{2}(\alpha_i^W)^T (\Gamma_\alpha^W)^{-1} \boldsymbol{\alpha}_i^W\right] \\ &\boldsymbol{\alpha}_i^{All} \sim N(0, \Gamma_\alpha^{All}) && p(\boldsymbol{\alpha}_i^{All} | \Gamma_\alpha^{All}) \propto \frac{1}{|\Gamma_\alpha^{All}|^{1/2}} \exp\left[-\frac{1}{2}(\alpha_i^{All})^T (\Gamma_\alpha^{All})^{-1} \boldsymbol{\alpha}_i^{All}\right] \end{aligned} \tag{4.2}$$

Both distributions are completely described by their covariance matrix. Moreover, as usual in statistical learning, we assume that the residuals $\boldsymbol{\varepsilon}_i$ follow a Gaussian distribution centred at 0. For now, we model all structures of the shape complexes $\boldsymbol{S}_i$ and of the template complex $\boldsymbol{T}$ with landmarks. Every structure j has a

number of landmarks equal to Λ_j and the norm of the difference between two meshes of the same structure is defined as the square root of the sum of squared differences between pair of landmarks (L^2-norm, $||\cdot||_2$). Thus, the likelihoods of the residuals of white and gray matter structures modelled as landmarks are respectively:

$$\begin{aligned} &\varepsilon_{ij}^W \sim N(0, \sigma_j^2 Id_{\Lambda_j}) \qquad p(\varepsilon_{ij}|\sigma_j^2) \propto \frac{1}{|\sigma_j^2|^{\Lambda_j/2}} \exp\left[-\frac{1}{2\sigma_j^2}||S_{ij} - \phi_i^{All}\left(\phi_i^W(T_j^W)\right)||_2^2\right] \\ &\varepsilon_{ij}^G \sim N(0, \sigma_j^2 Id_{\Lambda_j}) \qquad p(\varepsilon_{ij}|\sigma_j^2) \propto \frac{1}{|\sigma_j^2|^{\Lambda_j/2}} \exp\left[-\frac{1}{2\sigma_j^2}||S_{ij} - \phi_i^{All}\left(T_j^G\right)||_2^2\right] \end{aligned} \tag{4.3}$$

In Sec.4.2.3 we will make clear how to adapt these equations when a mesh is modelled as varifold or weighted current. Whatever the model employed, the variance depends exclusively on the structure-dependent parameter σ_j^2 which is also assumed to be a random variable. Moreover, from Eq.4.1 and Eq.4.3 it follows that all shapes $\{S_{ij}\}$ follow a Gaussian distribution:

$$S_{ij}^W \sim N(\phi_i^{All}\left(\phi_i^W(T_j^W)\right), \sigma_j^2 Id_{\Lambda_j}) \qquad S_{ij}^G \sim N(\phi_i^{All}\left(T_j^G\right), \sigma_j^2 Id_{\Lambda_j}) \tag{4.4}$$

We can thus reformulate the goal of the atlas construction as estimating the template complex $\boldsymbol{T}$, the two covariance matrices of the deformation parameters Γ_α^W and Γ_α^{All} and σ_j^2, knowing the shape complexes of the population $\{S_{ij}\}$ and assuming they follow a Gaussian distribution as in Eq.4.4. This can be achieved by maximizing the joint posterior distribution of $\boldsymbol{T}$, σ_j^2, Γ_α^W and Γ_α^{All}, which are also modelled as random variables. Assuming independence between all random variables and considering $\{\boldsymbol{\alpha}_i^W\}_{i=1,\dots,N}$ and $\{\boldsymbol{\alpha}_i^{All}\}_{i=1,\dots,N}$ as hidden variables, it results:

$$\{\boldsymbol{T}^*, \Gamma_\alpha^{W*}, \Gamma_\alpha^{All*}, \sigma_j^{2*}\} = \underset{\boldsymbol{T},\Gamma_\alpha^W,\Gamma_\alpha^{All},\sigma_j^2}{\arg\max} \left[\prod_i^N \prod_j^M \int\int p(T_j, \Gamma_\alpha^W, \Gamma_\alpha^{All}, \sigma_j^2, \boldsymbol{\alpha}_i^W, \boldsymbol{\alpha}_i^{All}, S_{ij}) d\boldsymbol{\alpha}_i^{All} d\boldsymbol{\alpha}_i^W\right] \tag{4.5}$$

Not using priors for σ_j^2, Γ_α^W and Γ_α^{All} can produce degenerate estimates. Instead, as demonstrated in (Allassonnière et al., 2007), the introduction of inverse Wishart distributions makes possible to obtain good estimates even when the number of subjects is small. Their probability density functions are:

$$\sigma_j^2 \sim \mathcal{W}^{-1}(P_j, w_j) \qquad p(\sigma_j^2; P_j, w_j) \propto (\sigma_j^2)^{-\frac{w_j}{2}} \exp\left[-\frac{1}{2}\frac{w_j P_j}{\sigma_j^2}\right] \tag{4.6}$$

$$\Gamma_\alpha^W \sim \mathcal{W}^{-1}(P_\alpha^W, w_\alpha^W) \quad p(\Gamma_\alpha^W; P_\alpha^W, w_\alpha^W) \propto |\Gamma_\alpha^W|^{-\frac{w_\alpha^W}{2}} \exp\left[-\frac{1}{2} w_\alpha^W Tr((P_\alpha^W)^T(\Gamma_\alpha^W)^{-1})\right] \tag{4.7}$$

$$\Gamma_\alpha^{All} \sim \mathcal{W}^{-1}(P_\alpha^{All}, w_\alpha^{All}) \quad p(\Gamma_\alpha^{All}; P_\alpha^{All}, w_\alpha^{All}) \propto |\Gamma_\alpha^{All}|^{-\frac{w_\alpha^{All}}{2}} \exp\left[-\frac{1}{2} w_\alpha^{All} Tr((P_\alpha^{All})^T(\Gamma_\alpha^{All})^{-1})\right] \tag{4.8}$$

The scalars w_j, P_j, w_α^W and w_α^{All} are strictly positive and P_α^W and P_α^{All} are positive symmetric matrices. Since the maximization of Eq.4.5 is not tractable analytically, we use the EM (Expectation Maximization) algorithm where we approximate the conditional distribution of the E step with a Dirac distribution at its mode. Assuming that the template $\boldsymbol{T}$ has a non-informative prior distribution, it results:

$$\begin{aligned}
&\boxed{\sum_{j=1}^{M}\sum_{i=1}^{N}\frac{1}{2\sigma_j^2}\left(||S_{ij} - \phi_i^{All}\left(\phi_i^W(T_j^W)\right)||_2^2 + \frac{P_j w_j}{N}\right)} + \\
&\boxed{\sum_{j=1}^{M}\sum_{i=1}^{N}\frac{1}{2\sigma_j^2}\left(||S_{ij} - \phi_i^{All}\left(T_j^G\right)||_2^2 + \frac{P_j w_j}{N}\right)} + \sum_{j=1}^{M}\frac{1}{2}(w_j + \Lambda_j N)\log(\sigma_j^2) + \\
&\boxed{\frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{\alpha}_i^{All})^T(\Gamma_\alpha^{All})^{-1}\boldsymbol{\alpha}_i^{All}} + \frac{(w_\alpha^{All}+N)}{2}\log(|\Gamma_\alpha^{All}|) + \frac{w_\alpha^{All}}{2}tr((\Gamma_\alpha^{All})^{-1}P_\alpha^{All}) + \\
&\boxed{\frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{\alpha}_i^{W})^T(\Gamma_\alpha^{W})^{-1}\boldsymbol{\alpha}_i^{W}} + \frac{(w_\alpha^{W}+N)}{2}\log(|\Gamma_\alpha^{W}|) + \frac{w_\alpha^{W}}{2}tr((\Gamma_\alpha^{W})^{-1}P_\alpha^{W})
\end{aligned} \tag{4.9}$$

This equation represents the cost function of our algorithm. We use a gradient descent scheme to minimize it. The framed terms refer respectively to the data-terms and to the regularity terms of both diffeomorphisms. The other terms are due to the use of the inverse Wishart prior distributions. The proposed statistical framework is general since it can be employed with any shape model, provided it is possible to define probability density functions, and any parametric diffeomorphic model.

4.2.3 Shape models

The framework of landmarks has been extensively used to model shapes since the metric and its gradients are easy to compute and it is also simple to define random variables. However, the definition of correspondences between 3D meshes is a tedious and difficult task which might also be impossible when dealing with fiber bundles. To this end, it is better to opt for correspondence-free shape models such as the framework of varifolds or currents. We propose to model gray matter structures as varifolds and white matter streamlines as weighted currents. Moreover, since fiber bundles resulting from tractography algorithms might have a number of streamlines so big to make infeasible the atlas construction, we propose to approximate them with the parsimonious representation of weighted streamlines prototypes presented in Chapter 3.

In the following, we will first recapitulate the framework of varifolds and weighted currents, then we will explain how to define random variables based on these two shape models.

4.2.3.1 Varifolds

The framework of varifolds is the non-oriented extension of the one of currents (Vaillant and Glaunès, 2005). It can be used to model both surfaces and curves as in Chapter 2 but here we will use it only for surfaces. The definition of varifold for a 3D rectifiable surface X is :

$$V_X(\omega) = \int_X \omega(p, \overleftrightarrow{n(p)})|n(p)|_2 dp \approx \sum_{l=1}^{L} \omega(p_l, \overleftrightarrow{n_l})|n_l|_2 \tag{4.10}$$

where $n(p) \in \mathbb{R}^3$ is the normal of X at the point $p \in \mathbb{R}^3$. Since the surfaces used throughout this chapter are meshes composed of a finite number of faces, it seems reasonable to approximate the integral in Eq.4.10 as a sum over the centres $\{p_l\}$ of the L faces of the surface as in Chapter 2. However, it is still an open problem the demonstration of the convergence of this approximation when the area of the faces decreases, i.e. the sampling becomes more accurate (Charlier et al., 2014). The test field $\omega \in W$ is function of both the position p and the unoriented unit normal vector $\overleftrightarrow{n(p)}$. The set of unoriented unit vectors $\overleftrightarrow{n} \in \overleftrightarrow{\mathbb{S}}$ can be formally defined as the quotient of the unit sphere in $\mathbb{R}^3$ by the two elements group $\{\pm \mathrm{Id}_3\}$. This means that two unoriented unit vectors which are superimposed (and thus parallel) but with a different orientation are considered equivalent. Eq.4.10 shows that the space of varifolds W^* is a linear functional from W to $\mathbb{R}$. This makes W^* a vector space and therefore the union of meshes is equal to a sum in the space of varifolds. But, contrary to currents, inverting the orientation of a mesh does not change the varifold representation.

The test space W is defined as a product space between the ambient space $\mathbb{R}^3$ and the space of unoriented unit normal vectors $\overleftrightarrow{\mathbb{S}}$, $W : (\mathbb{R}^3 \mathrm{x} \overleftrightarrow{\mathbb{S}})$. More

precisely, as proposed in (Charon and Trouvé, 2013b), one chooses W to be a Reproducible Kernel Hilbert Space (RKHS) whose kernel K_W is defined as the tensor product between two kernels k_g, Gaussian (continuous red line), and k_n, Cauchy-Binet (dashed blue line), defined on $\mathbb{R}^3$ and on $\overleftrightarrow{\mathbb{S}}$ respectively: $K_W((p,\overleftrightarrow{n(p)}),(q,\overleftrightarrow{u(q)})) = k_g(p,q)k_n(\overleftrightarrow{n(p)},\overleftrightarrow{u(q)})$. Thanks to the reproducing property of the RKHS, the inner product between X and another surface Y modelled as varifold ($V_Y \approx \sum_{h=1}^{H} \omega(q_h, \overleftrightarrow{u_h})|u_h|_2$) results:

$$\langle V_X, V_Y \rangle_{W^*} = \sum_{l=1}^{L}\sum_{h=1}^{H} \exp\left(\frac{-||p_l - q_h||_2^2}{\lambda_W^2}\right)\left(\frac{n_l^T u_h}{|n_l|_2|u_h|_2}\right)^2 |n_l|_2|u_h|_2 \tag{4.11}$$

The only parameter is the standard deviation λ_W of the Gaussian kernel. The distance between V_X and V_Y is: $||V_X - V_Y||^2_{W^*} = \langle V_X, V_X\rangle_{W^*} + \langle V_Y, V_Y\rangle_{W^*} - 2\langle V_X, V_Y\rangle_{W^*}$. An important characteristic of this metric is the absence of correspondences: every normal of the first surface is compared with all the normals of the other surface (Cauchy-Binet kernel) and their contribution is weighted by the distance between the centres (Gaussian kernel). Moreover, it is interesting to notice that the Cauchy-Binet kernel is equal to the squared cosine of the angle between the normals (i.e. $\langle n_l, u_h\rangle_2 = n_l^T u_h = |n_l|_2|u_h|_2\cos(\cdot)$). This makes the metric invariant to a change of orientation of some normals of the surfaces. Moreover, this property is particularly important in the estimation of the template since it permits to prevent the so called "currents cancelling effect". This happens in surfaces modelled as currents when two faces with opposite orientation, like in a thin protrusion, cancel reciprocally their effect on the estimate of the template. For a more detailed discussion about varifolds the user is referred to Charon and Trouvé (2013b).

4.2.3.2 Weighted currents

The framework of weighted currents, presented in Chapter 3, is also an extension of the one of currents. It has been conceived for curves and particularly to model the streamlines resulting from tractography algorithms. An oriented 3D polygonal curve A composed of G segments is modelled as a 1-weighted current C_A via a line integral of a square-integrable vector field ω:

$$C_A(\omega) = \int_A w(x, f_a, f_b)^T \alpha(x) dx \approx \sum_{g=1}^{G} w(x_g, f_a, f_b)^T \alpha_g \tag{4.12}$$

where $x_g \in \mathbb{R}^3$ and $\alpha_g \in \mathbb{R}^3$ are respectively the centre and the tangent vector of segment g and, as previously for varifolds, we approximate the integral with a sum over the segments of A. The two 3D vectors f_a and f_b are the coordinates of the end-points of the curve and they are defined in the space $\Psi = \mathbb{R}^3\text{x}\mathbb{R}^3$. Similarly to varifolds, the vector field ω belongs to a RKHS Q defined on the product space

$\mathbb{R}^3$xΨ and the space of weighted currents Q^* is a continuous linear form on Q. Thus, even in this case, every curve is considered as a vector in a Hilbert space and the union of several curves, such as a white matter fiber bundle, is represented as a sum. The kernel K_Q associated to the RKHS Q is defined as: $K_Q((x, f^a, f^b), (y, t^a, t^b)) = k_g(x, y)k_a(f^a, t^a)k_b(f^b, t^b)$ where the three kernels k_g, k_a, k_b are Gaussian and defined in $\mathbb{R}^3$. As explained in Chapter 3, the inner product between A and another curve B modelled as weighted current ($C_B(\omega) \approx \sum_{f=1}^{F} w(y_f, t_a, t_b)^T \beta_f$) results in:

$$\langle C_A, C_B \rangle_{Q^*} = \sum_{g=1}^{G} \sum_{f=1}^{F} \exp\left(\frac{-||x_g - y_f||_2^2}{\lambda_g^2}\right) \alpha_g^T \beta_f \exp\left(\frac{-||f^a - t^a||_2^2}{\lambda_a^2}\right) \exp\left(\frac{-||f^b - t^b||_2^2}{\lambda_b^2}\right) \tag{4.13}$$

The continuous red line highlights the inner product between A and B if modelled as usual currents. It measures overall geometrical differences between their pathways. The two other terms, underlined with dashed lines, constrain the definition of similarity. Two curves are considered similar if their pathways are alike, as usual currents, but also if their endpoints are close to each other. The inner product is parametrised by the three bandwidths: λ_g, λ_a, λ_b. The distance between two curves is defined as: $||C_A - C_B||^2_{Q^*} = \langle C_A - C_B, C_A - C_B \rangle_{Q^*}$. As usual currents, this framework does not need point-to-point correspondences, except for the endpoints. This means that curves need to have a consistent orientation. This can be complicated when comparing whole-brain tractograms but here we need to compare only fiber bundles extracted using two specific ROIs (Region Of Interest) defining the initial and last point respectively. This permits to easily ensure a consistent orientation between the streamlines of the same bundle and also among bundles of different subjects. Furthermore, the use of Gaussian kernels on the extremities of the streamlines permits to be more robust towards tractography errors at the boundary between grey and white matter which is an area usually characterised by a low SNR (Signal to Noise Ratio).

4.2.3.3 Varifolds and Currents random variables

In order to integrate the models of varifolds and weighted currents to the statistical framework previously presented, we need to define random Gaussian variables in these two spaces. In both cases, they are defined as a linear map between every vector field $\omega \in W$ (or $\omega \in Q$) to a real random Gaussian variable $G(\omega)$ such that $E[G(\omega)] = 0$ and, given two vector fields ω_1 and ω_2, $E[G(\omega_1), G(\omega_2)] = \langle \omega_1, \omega_2 \rangle_{W(orQ)}$. This means that the kernels K_W and K_Q define the covariance structures of the Gaussian varifolds and weighted currents respectively. Since both of them are infinite-dimensional, we define for each structure j a finite dimensional space on which we project both the template T_j and the shapes $\{S_{ij}\}_{i=1...N}$ mod-

elled as varifolds or weighted currents. We follow the line of reasoning presented in Durrleman (2010) for the framework of currents.

Varifolds The finite dimensional space $W^*_{\Lambda_j}$ for varifolds is defined as the span of a set of delta Dirac varifolds: $\text{Span}\{\delta_{(x_u,\overleftrightarrow{\beta_k})}\}$ where the points $\{x_u\}$ and the unoriented unit vectors $\{\overleftrightarrow{\beta_k}\}$ are constrained to belong to two predefined grids, respectively Υ^j_x and Υ_β. The first one is a linearly spaced grid in the ambient space and Υ_β is a regular sampling of the half unit sphere in $\mathbb{R}^3$. The number of points of the two grids is respectively Λ^j_x and Λ_β and Λ_j is their product. For every finite-dimensional space $W^*_{\Lambda_j}$, it is possible to define a block matrix $\boldsymbol{K}_{W,\Lambda_j}$ whose blocks are the RKHS kernel $K_{W,\Lambda_j}\left((x_u,\overleftrightarrow{\beta_k}),(\cdot,\cdot)\right)_{\substack{u=1,\dots,\Lambda^j_x\\k=1,\dots,\Lambda_\beta}}$ between every possible combination of the couples $\{x_u,\overleftrightarrow{\beta_k}\}$. The matrix $\boldsymbol{K}_{W,\Lambda_j}$ has a dimension of $[\Lambda_j \text{x} \Lambda_j]$. We can then use the closest neighbour projection as explained in Sec.2.2.2 to project a varifold on this grid. The squared norm of the projection of the varifold defined in Eq.4.10 results:

$$||\sum_l \delta_{(p_l,\overleftrightarrow{n_l})}|n_l|_2||^2_{W^*_{\Lambda j}} = \sum_{u\in\Upsilon_x}\sum_{p\in\Upsilon_x}\sum_{k\in\Upsilon_\beta}\sum_{q\in\Upsilon_\beta} c_{uk}K_{W,\Lambda_j}\left((x_u,\overleftrightarrow{\beta_k}),(x_p,\overleftrightarrow{\beta_q})\right)c_{pq} \quad (4.14)$$

where the scalars c_{uk} and c_{pq} refer to the values resulting from the projection in $(x_u,\overleftrightarrow{\beta_k})$ and $(x_p,\overleftrightarrow{\beta_q})$ respectively.

Weighted currents The finite dimensional space $Q^*_{\Lambda_j}$ for weighted currents is also defined as the span of a set of delta Dirac weighted currents: $\text{Span}\{\delta^\alpha_{(y_l,f^a_m,f^b_n)}\}$ where $\alpha\in\mathbb{R}^3$ and both the points $\{y_l\}$ and the endpoints $\{f^a_m\}$, $\{f^b_n\}$ are constrained to belong to predefined grids: $y_l\in\Upsilon^j_y$, $f^a_m\in\Upsilon^j_a$ and $f^b_n\in\Upsilon^j_b$. The three grids are composed of points linearly spaced in the ambient space and their sizes are equal to Λ^j_y, Λ^j_a, Λ^j_b respectively. The product of their sizes is equal to Λ_j. As for varifolds, we can define a block matrix $\boldsymbol{K}_{Q,\Lambda_j}$ for every finite-dimensional space $Q^*_{\Lambda_j}$, whose blocks are the RKHS kernel $K_{Q,\Lambda_j}\left((y_l,f^a_m,f^b_n),(\cdot,\cdot,\cdot)\right)_{\substack{l=1,\dots,\Lambda^j_y\\m=1,\dots,\Lambda^j_a\\n=1,\dots,\Lambda^j_b}}$ between every possible combination of the triples $\{y_l,f^a_m,f^b_n\}$. The matrix $\boldsymbol{K}_{Q,\Lambda_j}$ has a dimension of $[\Lambda_j \text{x} \Lambda_j]$. The squared norm of the projection of the weighted current defined in Eq.4.12 is:

$$||\sum_{g=1}^{G} w(x_g,f_a,f_b)^T\alpha_g||^2_{Q^*_{\Lambda j}} = \sum_{l\in\Upsilon^j_y}\sum_{m\in\Upsilon^j_a}\sum_{n\in\Upsilon^j_b}\sum_{p\in\Upsilon^j_y}\sum_{q\in\Upsilon^j_a}\sum_{r\in\Upsilon^j_b}\alpha_{lmn}K_{Q,\Lambda_j}\left((y_l,f^a_m,f^b_n),(y_p,f^a_q,f^b_r)\right)\alpha_{pqr} \quad (4.15)$$

where the 3D vectors α_{lmn} and α_{pqr} refer to the values resulting from the projection in (y_l, f^a_m, f^b_n) and (y_p, f^a_q, f^b_r) respectively. The projected norm in Eq.4.15 can be computed using, for instance, the numerical scheme $p3m$ proposed in Durrleman (2010) and based on FFT (Fast Fourier Transform).

Likelihood Once all shapes $\{S_{ij}\}_{i=1...N}$ and the template T_j are projected onto a finite dimensional space ($W^*_{\Lambda_j}$ or $Q^*_{\Lambda_j}$) it is possible to compute the norm of the projected residuals and the likelihood of the noise. The latter is defined for white and gray matter structures respectively as:

$$p(\varepsilon^W_{ij}|\sigma^2_j) \propto \frac{1}{|\sigma^2_j|^{\Lambda_j/2}} \exp\left[-\frac{1}{2\sigma^2_j}||S_{ij} - \phi^{All}_i\left(\phi^W_i(T^W_j)\right)||^2_{Q^*_{\Lambda_j}}\right] \tag{4.16}$$
$$p(\varepsilon^G_{ij}|\sigma^2_j) \propto \frac{1}{|\sigma^2_j|^{\Lambda_j/2}} \exp\left[-\frac{1}{2\sigma^2_j}||S_{ij} - \phi^{All}_i(T^G_j)||^2_{W^*_{\Lambda_j}}\right]$$

where we have scaled the kernels $\boldsymbol{K}_{W,\Lambda_j}$ and $\boldsymbol{K}_{Q,\Lambda_j}$ with the scalar σ^2_j. As it is possible to notice, these equations are very similar to the ones in Eq.4.3 where we modelled the structures as landmarks. The only differences concern the metric and the definition of the parameter Λ_j. These are also the only changes to take into consideration in Eq.4.9 when modelling the structures as varifolds and weighted currents.

In practice, due to the important computational load and time required to calculate both $||\cdot||^2_{Q^*_{\Lambda j}}$ and $||\cdot||^2_{W^*_{\Lambda j}}$, we choose to compute directly $||(\cdot)||^2_{Q^*}$ and $||(\cdot)||^2_{W^*}$ with a fast GPU (CUDA) implementation. Moreover, choosing the number of points for each grid is a tricky task. This becomes even harder when mixing different computational models such as varifolds and weighted currents. Here, we propose to select Λ_j as the number of points required to build a regular linearly-spaced grid covering the ambient space where the shapes and the template of structure j lie. The distance between every couple of points is equal to λ_W or λ_g for varifolds and weighted currents respectively. In Sec.4.3.2 we will show that results are stable varying the size of the grids and that we obtain reasonable results with the proposed heuristic rule.

4.2.4 Diffeomorphic deformations

We define here how to compute the diffeomorphic deformations of the template complex. Our approach relies on the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework based on the control point formulation presented in Durrleman et al. (2011b). For every subject i, both ϕ^W_i and ϕ^{All}_i are defined as the last deformations of two flows of diffeomorphisms $\{\phi^W_{it}\}_{t\in[0,1]}$ and $\{\phi^{All}_{it}\}_{t\in[0,1]}$. Calling $\phi_i(\boldsymbol{x},t) = \phi_{it}(\boldsymbol{x}) = \boldsymbol{x}_i(t)$ the position of a point at time t which was located in $\boldsymbol{x}$ at time $t = 0$, each flow is built by integrating: $\frac{\partial\phi_i(\boldsymbol{x},t)}{\partial t} = v_i(\phi_i(\boldsymbol{x},t),t) = v_i(\boldsymbol{x}_i(t),t)$

over $t \in [0,1]$ where $v_i(\boldsymbol{x}_i(t),t)$ is a time-varying vector field representing the instantaneous velocity of a point located in $\boldsymbol{x}_i(t)$ at time t. Both vector fields v_i^{All} and v_i^W belong to the same RKHS D with Gaussian kernel $\boldsymbol{K}_D$. They are defined by two different sets of 3D control points, $\boldsymbol{c}^{All}$ and $\boldsymbol{c}^W$, shared among all subjects, and by two distinct sets of 3D vectors, called momenta, $\boldsymbol{\alpha}_i^{All}$ and $\boldsymbol{\alpha}_i^W$ linked to the control points and specific to each subject i: $v_i^{All}(\boldsymbol{x}_i(t),t) = \boldsymbol{K}_D(\boldsymbol{x}_i(t),\boldsymbol{c}^{All}(t))\boldsymbol{\alpha}_i^{All}(t)$ and $v_{it}^W(\boldsymbol{x}_i(t)) = \boldsymbol{K}_D(\boldsymbol{x}_i(t),\boldsymbol{c}^W(t))\boldsymbol{\alpha}_i^W(t)$, where $\boldsymbol{x}_i(0) = \boldsymbol{x}$ and $\boldsymbol{K}_D(\boldsymbol{x}_i(t),\boldsymbol{c}(t))$ represents a block matrix of Gaussian kernels with an equal fixed bandwidth for both v_i^{All} and v_i^W. The deformation of every point $\boldsymbol{x}$ in the ambient space depends on its initial position at $t=0$ and on the evolution of the system $\boldsymbol{L}_i^{All}(t) = \{\boldsymbol{c}^{All}(t), \boldsymbol{\alpha}_i^{All}(t)\}$ if the point belongs to the grey matter and on both systems $\boldsymbol{L}_i^W(t) = \{\boldsymbol{c}^W(t), \boldsymbol{\alpha}_i^W(t)\}$ and $\boldsymbol{L}_i^{All}(t)$ if the point belongs to the white matter. At $t=0$ the deformations ϕ_{i0}^W and ϕ_{i0}^{All} are equal to the identity transformations. For both systems, the path from ϕ_{i0}^{All} (resp. ϕ_{i0}^W) to ϕ_{i1}^{All} (resp. ϕ_{i1}^W), the latter being the deformation of interest, is chosen as the geodesic one, which means the one that minimizes the total kinetic energy along the path: $\int_0^1 ||v_{it}^{All}||_D^2$ (resp. $\int_0^1 ||v_{it}^W||_D^2$). It has been shown in Durrleman et al. (2014) that the extremal paths are such that both systems $\boldsymbol{L}_i^W(t)$ and $\boldsymbol{L}_i^{All}(t)$ satisfy:

$$\begin{aligned}
&\dot{\boldsymbol{c}}_i(t) = \boldsymbol{K}(\boldsymbol{c}_i(t),\boldsymbol{c}_i(t))\boldsymbol{\alpha}_i(t) = F^c(\boldsymbol{c}_i(t),\boldsymbol{\alpha}_i(t)) && \boldsymbol{c}_i(0) = \boldsymbol{c}(0) = \boldsymbol{c}_0 \\
&\dot{\boldsymbol{\alpha}}_i(t) = -\boldsymbol{\alpha}_i(t)^T\boldsymbol{\alpha}_i(t)\nabla_1\boldsymbol{K}(\boldsymbol{c}_i(t),\boldsymbol{c}_i(t)) = F^\alpha(\boldsymbol{c}_i(t),\boldsymbol{\alpha}_i(t)) && \boldsymbol{\alpha}_i(0) = \boldsymbol{\alpha}_{i0}
\end{aligned} \tag{4.17}$$

which can be summarized as $\dot{\boldsymbol{L}}_i^{All}(t) = F(\boldsymbol{L}_i^{All}(t))$ (resp. $\dot{\boldsymbol{L}}_i^W(t) = F(\boldsymbol{L}_i^W(t))$). The last diffeomorphisms ϕ_{i1}^{All} and ϕ_{i1}^W are completely parametrized by the initial conditions of the systems: $\boldsymbol{L}_i^{All}(0) = \boldsymbol{L}_{i0}^{All} = \{\boldsymbol{c}_0^{All}, \boldsymbol{\alpha}_{i0}^{All}\}$ (resp. $\boldsymbol{L}_i^W(0) = \boldsymbol{L}_{i0}^W = \{\boldsymbol{c}_0^W, \boldsymbol{\alpha}_{i0}^W\}$). Thus, in order to put into correspondence the template complex $\boldsymbol{T}$ with the subject shape complex $\boldsymbol{S}_i$, we first integrate forward in time $\dot{\boldsymbol{L}}_i^W(t)$ starting from $\boldsymbol{L}_{i0}^W$ and we use these values to deform only the white matter structures of the template complex $\boldsymbol{T}^W$ integrating forward in time:

$$\dot{\boldsymbol{T}}_i^W(t) = \boldsymbol{K}(\boldsymbol{T}_i^W(t),\boldsymbol{c}_i^W(t))\boldsymbol{\alpha}_i^W(t) = G[\boldsymbol{T}_i^W(t),\boldsymbol{L}_i^W(t)] \qquad \boldsymbol{T}_i^W(0) = \boldsymbol{T}_{i0}^W = \boldsymbol{T}^W \tag{4.18}$$

The deformed white matter template $\phi_{i1}^W(\boldsymbol{T}^W) = \boldsymbol{T}_{i1}^W$, together with the undeformed gray matter template $\boldsymbol{T}^G = \boldsymbol{T}_{i0}^G$ are then transformed by the second diffeomorphism ϕ_{i1}^{All} integrating first forward in time $\dot{\boldsymbol{L}}_i^{All}(t)$ and then:

$$\dot{\boldsymbol{T}}_i^{All}(t) = \boldsymbol{K}(\boldsymbol{T}_i^{All}(t),\boldsymbol{c}_i^{All}(t))\boldsymbol{\alpha}_i^{All}(t) = G[\boldsymbol{T}_i^{All}(t),\boldsymbol{L}_i^{All}(t)] \qquad \boldsymbol{T}_i^{All}(0) = \boldsymbol{T}_{i0}^{All} = \boldsymbol{T}_{i1}^W \cup \boldsymbol{T}_{i0}^G \tag{4.19}$$

A sketch of the cascade of diffeomorphisms is shown in Fig.4.2.

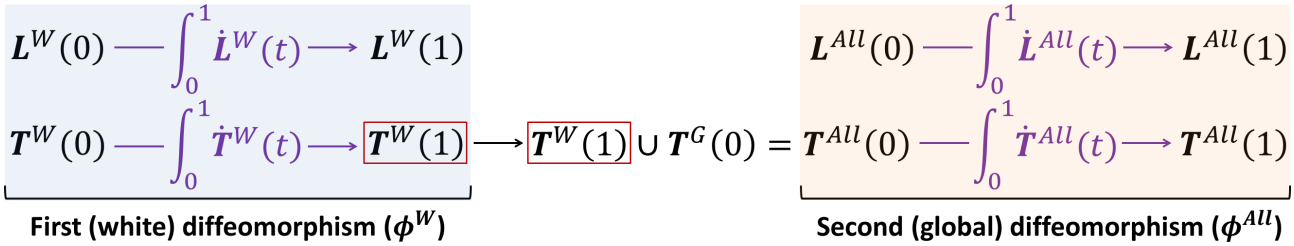

Figure 4.2: Sketch of the deformation of the template complex based on the proposed cascade of two diffeomorphisms ϕ^W and ϕ^{All}. We omit the subject index i for clarity purpose.

4.2.5 Optimization procedure

The double diffeomorphic framework presented in the previous sub-section can be easily integrated in the Bayesian setting for atlas construction explained in Sec.4.2.2. The two sets of initial control points and momenta $\{\boldsymbol{c}_0^{All}, \{\boldsymbol{\alpha}_{i0}^{All}\}_{i=1,\dots,N}\}$ and $\{\boldsymbol{c}_0^{W}, \{\boldsymbol{\alpha}_{i0}^{W}\}_{i=1,\dots,N}\}$ represent the deformation parameters to warp the template complex $\boldsymbol{T}$ towards the subject shape complex $\boldsymbol{S}_i$. The initial control points $\boldsymbol{c}_0^{All}$ and $\boldsymbol{c}_0^{W}$ are shared among the whole population and we assume they follow a non-informative prior distribution. The initial momenta $\boldsymbol{\alpha}_i^{All}$ and $\boldsymbol{\alpha}_i^{W}$ are specific to each subject i and they follow a Gaussian distribution exactly as in Eq.4.2. Modelling the grey matter structures as varifolds and the white matter bundles as weighted currents, assuming that all random variables are independent and considering $\{\boldsymbol{\alpha}_{i0}^{W}\}_{i=1,\dots,N}$, $\{\boldsymbol{\alpha}_{i0}^{All}\}_{i=1,\dots,N}$, $\boldsymbol{c}_0^{All}$ and $\boldsymbol{c}_0^{W}$ as hidden variables, we can rewrite the cost function in Eq.4.9 as:

$$\boxed{\sum_{j=1}^{M^W}\sum_{i=1}^{N}\frac{1}{2\sigma_j^2}\left(||\Pi(S_{ij}-\phi_{i1}^{All}\left(\phi_{i1}^{W}(T_j^W)\right))||^2_{Q^*_{\Lambda_j}}+\frac{P_j w_j}{N}\right)}+\sum_{j=1}^{M^W}\frac{1}{2}(w_j+\Lambda_j N)\log(\sigma_j^2)+$$
$$\boxed{\sum_{j=1}^{M^G}\sum_{i=1}^{N}\frac{1}{2\sigma_j^2}\left(||\Pi(S_{ij}-\phi_{i1}^{All}(T_j^G))||^2_{W^*_{\Lambda_j}}+\frac{P_j w_j}{N}\right)}+\sum_{j=1}^{M^G}\frac{1}{2}(w_j+\Lambda_j N)\log(\sigma_j^2)+$$
$$\boxed{\frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{\alpha}_{i0}^{All})^T(\Gamma_\alpha^{All})^{-1}\boldsymbol{\alpha}_{i0}^{All}}+\frac{(w_\alpha^{All}+N)}{2}\log(|\Gamma_\alpha^{All}|)+\frac{w_\alpha^{All}}{2}tr((\Gamma_\alpha^{All})^{-1}P_\alpha^{All})+$$
$$\boxed{\frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{\alpha}_{i0}^{W})^T(\Gamma_\alpha^{W})^{-1}\boldsymbol{\alpha}_{i0}^{W}}+\frac{(w_\alpha^{W}+N)}{2}\log(|\Gamma_\alpha^{W}|)+\frac{w_\alpha^{W}}{2}tr((\Gamma_\alpha^{W})^{-1}P_\alpha^{W}) \tag{4.20}$$

where M^W and M^G are the numbers of white matter and grey matter structures respectively. The EM algorithm employed here is an approximation of the Maximum

a Posterior (MAP) estimator where the hidden variables are actually considered as parameters to be optimised. The whole set of parameters is therefore: $\{\boldsymbol{T}, \Gamma_\alpha^{All}, \Gamma_\alpha^{W}, \{\sigma_j^2\}_{j=1,...,M}, \{\boldsymbol{\alpha}_{i0}^{All}\}_{i=1,...,N}, \{\boldsymbol{\alpha}_{i0}^{W}\}_{i=1,...,N}, \boldsymbol{c}_0^{All}, \boldsymbol{c}_0^{W}\}$. We use a gradient descent scheme for $\{\boldsymbol{T}, \{\boldsymbol{\alpha}_{i0}^{All}\}_{i=1,...,N}, \{\boldsymbol{\alpha}_{i0}^{W}\}_{i=1,...,N}, \boldsymbol{c}_0^{All}, \boldsymbol{c}_0^{W}\}$. Instead the other parameters, namely Γ_α^{All}, Γ_α^{W}, $\{\sigma_j^2\}_{j=1,...,M}$, have closed form solutions thanks to the use of conjugate priors. They result:

$$\hat{\Gamma}_\alpha^{All} = \frac{\sum_{i=1}^{N}\left[(\boldsymbol{\alpha}_{i0}^{All})(\boldsymbol{\alpha}_{i0}^{All})^T\right] + w_\alpha^{All}(P_\alpha^{All})^T}{(w_\alpha^{All}+N)} \qquad (\hat{\sigma}_j^2)^G = \frac{\sum_{i=1}^{N}||\Pi(S_{ij} - \phi_{i1}^{All}(T_j^G))||^2_{W^*_{\Lambda j}} + w_j P_j}{(w_j + N\Lambda_j)}$$

$$\hat{\Gamma}_\alpha^{W} = \frac{\sum_{i=1}^{N}\left[(\boldsymbol{\alpha}_{i0}^{W})(\boldsymbol{\alpha}_{i0}^{W})^T\right] + w_\alpha^{W}(P_\alpha^{W})^T}{(w_\alpha^{W}+N)} \qquad (\hat{\sigma}_j^2)^W = \frac{\sum_{i=1}^{N}||\Pi(S_{ij} - \phi_{i1}^{All}\left(\phi_{i1}^{W}(T_j^W)\right))||^2_{Q^*_{\Lambda j}} + w_j P_j}{(w_j + N\Lambda_j)}$$

where $(\hat{\sigma}_j^2)^G$ and $(\hat{\sigma}_j^2)^W$ refer to the grey and white matter structures respectively.

The two covariance matrices $\hat{\Gamma}_\alpha^{All}$ and $\hat{\Gamma}_\alpha^{W}$ are equal to a weighted sum between the sample covariance matrix of the initial momenta and the prior. We choose $P_\alpha^{All}=\boldsymbol{K}_D^{-1}(\boldsymbol{c}_0^{All}, \boldsymbol{c}_0^{All})$ and $P_\alpha^{W}=\boldsymbol{K}_D^{-1}(\boldsymbol{c}_0^{W}, \boldsymbol{c}_0^{W})$ which are block matrices of Gaussian kernels between the initial control points. The matrix $\boldsymbol{K}_D$ is the kernel of the RKHS to which belong both vector fields v_i^{All} and v_i^{W}. This choice is motivated by the fact that if the number of subjects N is definitely smaller than w_α^{All} and w_α^{W} then $\hat{\Gamma}_\alpha^{All} \propto \boldsymbol{K}_D^{-1}(\boldsymbol{c}_0^{All}, \boldsymbol{c}_0^{All})$ and $\hat{\Gamma}_\alpha^{W} \propto \boldsymbol{K}_D^{-1}(\boldsymbol{c}_0^{W}, \boldsymbol{c}_0^{W})$. This means that the framed regularity terms in Eq.4.20 become $\sum_{i=1}^{N}(\boldsymbol{\alpha}_{i0}^{All})^T\boldsymbol{K}_D(\boldsymbol{c}_0^{All}, \boldsymbol{c}_0^{All})\boldsymbol{\alpha}_{i0}^{All} = \sum_{i=1}^{N}||v_{i0}^{All}||^2_D$ and $\sum_{i=1}^{N}(\boldsymbol{\alpha}_{i0}^{W})^T\boldsymbol{K}_D(\boldsymbol{c}_0^{W}, \boldsymbol{c}_0^{W})\boldsymbol{\alpha}_{i0}^{W} = \sum_{i=1}^{N}||v_{i0}^{W}||^2_D$ which are the sums of the energies of the geodesic paths of all subjects for both diffeomorphisms. This kind of regularity term is often employed in deterministic atlas constructions, namely not based on a statistical setting (Joshi and Miller, 2000; Beg et al., 2005; Durrleman et al., 2012). Furthermore, the use of these priors regularises the estimate of the covariance matrices since it makes possible their inversion even when the number of subjects N is smaller than three times the number of controls points (i.e. size of α_{i0}^{All} and α_{i0}^{W}). This also implies that they can be *directly* employed in linear statistical analyses such as PCA at the end of the atlas construction without being regularised a posteriori.

The two other parameters, $(\hat{\sigma}_j^2)^G$ and $(\hat{\sigma}_j^2)^W$, are equal to a weighted sum between the data-term of the j-th structure and the prior P_j. Every parameter balances the importance of structure j with respect to the other structures and with respect to the regularity terms of both diffeomorphisms. The chosen prior imposes a minimum value to $\hat{\sigma}_j^2$ which is useful to avoid overfitting. In fact, without a prior, the minimisation process might focus only on a structure k reducing its residuals almost to zero and ignoring the other structures. This would result in $\hat{\sigma}_k^2 \to 0$ and therefore also in $\log(\hat{\sigma}_k^2) \to$-$\infty$.

The gradients of the cost function E in Eq.4.20 with respect to the other parameters are equal to:

$$\nabla_{T_k^G} E = \sum_{i=1}^{N} \frac{1}{2\sigma_k^2} \nabla_{T_k^G} D_{ik}^G \qquad \nabla_{T_k^W} E = \sum_{i=1}^{N} \frac{1}{2\sigma_k^2} \nabla_{T_k^W} D_{ik}^W$$
$$\nabla_{\boldsymbol{\alpha}_{s0}^{All}} E = \sum_{j=1}^{M} \frac{1}{2\sigma_j^2} \nabla_{\boldsymbol{\alpha}_{s0}^{All}} D_{sj} + (\Gamma_\alpha^{All})^{-1} \boldsymbol{\alpha}_{s0}^{All} \qquad \nabla_{\boldsymbol{\alpha}_{s0}^{W}} E = \sum_{j=1}^{M} \frac{1}{2\sigma_j^2} \nabla_{\boldsymbol{\alpha}_{s0}^{W}} D_{sj} + (\Gamma_\alpha^{W})^{-1} \boldsymbol{\alpha}_{s0}^{W}$$
$$\nabla_{\boldsymbol{c}_0^{All}} E = \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{1}{2\sigma_j^2} \nabla_{\boldsymbol{c}_0^{All}} D_{ij} \qquad \nabla_{\boldsymbol{c}_0^{W}} E = \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{1}{2\sigma_j^2} \nabla_{\boldsymbol{c}_0^{W}} D_{ij} \tag{4.21}$$

where $D_{ij}^G = ||\Pi(S_{ij} - \phi_{i1}^{All}(T_j^G))||^2_{W^*_{\Lambda j}}$ and $D_{ij}^W = ||\Pi(S_{ij} - \phi_{i1}^{All}\left(\phi_{i1}^W(T_j^W)\right))||^2_{Q^*_{\Lambda j}}$ refer to the data terms of the grey and white matter structures respectively whereas D_{ij} refers to the data-term of any structure. Moreover, we have not taken into account the differentiation of the prior P_α with respect to the control points since its norm is negligible with respect to the one of the data term.

To calculate the gradients of the data terms $\{D_{ij}\}$ we need to compute the deformed template complex obtained at the end of the second diffeomorphism for every subject i and for all structures, namely $\phi_{i1}^{All}(\boldsymbol{T}^G)$ and $\phi_{i1}^{All}\left(\phi_{i1}^W(\boldsymbol{T}^W)\right)$. First, we integrate forward in time $\dot{\boldsymbol{L}}_i^W(t) = F(\boldsymbol{L}_i^W(t))$ (Eq.4.17) and $\dot{\boldsymbol{T}}_i^W(t) = G[\boldsymbol{T}_i^W(t), \boldsymbol{L}_i^W(t)]$ (Eq.4.18). Then, we integrate $\dot{\boldsymbol{L}}_i^{All}(t) = F(\boldsymbol{L}_i^{All}(t))$ (Eq.4.17) and, using as initial value $\boldsymbol{T}_i^{All}(0) = \boldsymbol{T}_{i1}^W \cup \boldsymbol{T}_{i0}^G$, we integrate also $\dot{\boldsymbol{T}}_i^{All}(t) = G[\boldsymbol{T}_i^{All}(t), \boldsymbol{L}_i^{All}(t)]$ (Eq.4.19). After that, we can compute the data term D_{ij} and its gradient with respect to the vertices of $\boldsymbol{T}_i^{All}(1) = \boldsymbol{T}_{i1}^{All}$. The computations of the gradient for surfaces modelled as varifolds are shown in the Appendix of Durrleman et al. (2014) whereas the calculations about the gradient for fiber bundles modelled as weighted prototypes are presented in Sec.4.A.1. Using the calculus of variations (see Sec.4.A.2), the information about the gradient is brought back from $t = 1$ to $t = 0$ to update first $\boldsymbol{L}_i^{All}(0) = \{\boldsymbol{c}_0^{All}, \boldsymbol{\alpha}_{i0}^{All}\}$ and $\boldsymbol{T}^G$ and then $\boldsymbol{L}_i^W(0) = \{\boldsymbol{c}_0^W, \boldsymbol{\alpha}_{i0}^W\}$ and $\boldsymbol{T}^W$. The optimisation is based on a set of linearised ODEs describing the evolution of four auxiliary variables $\theta_i^{All}, \xi_i^{All} = \{\xi_{\alpha i}^{All}, \xi_{ci}^{All}\}, \theta_i^W, \xi_i^W = \{\xi_{\alpha i}^W, \xi_{ci}^W\}$:

$$\dot{\theta}_i^{All}(t) = -(\partial_{\boldsymbol{T}_i^{All}} G_i^{All}(t))^T \theta_i^{All}(t) \qquad \theta_i^{All}(1) = \nabla_{\boldsymbol{T}_i^{All}(1)} \boldsymbol{D}_i \tag{4.22}$$
$$\dot{\xi}_i^{All}(t) = -(\partial_{\boldsymbol{L}_i^{All}} G_i^{All}(t))^T \theta_i^{All}(t) + (d_{\boldsymbol{L}_i^{All}} F_i^{All}(t))^T \xi_i^{All}(t) \qquad \xi_i^{All}(1) = 0 \tag{4.23}$$
$$\dot{\theta}_i^W(t) = -(\partial_{\boldsymbol{T}_i^W} G_i^W(t))^T \theta_i^W(t) \qquad \boxed{\theta_i^W(1) = \theta_i^{All,W}(0)} \tag{4.24}$$
$$\dot{\xi}_i^W(t) = -(\partial_{\boldsymbol{L}_i^W} G_i^W(t))^T \theta_i^W(t) + (d_{\boldsymbol{L}_i^W} F_i^W(t))^T \xi_i^W(t) \qquad \xi_i^W(1) = 0 \tag{4.25}$$

The size of θ_i^{All} and ξ_i^{All} are the same as the ones of $\boldsymbol{T}$ and $\boldsymbol{L}_i^{All}$ respectively. Instead, the size of θ_i^W and ξ_i^W are equal to the ones of $\boldsymbol{T}^W$ and $\boldsymbol{L}_i^W$ respectively. We first integrate backward Eq.4.22 and Eq.4.23 obtaining $\theta_i^{All}(0)$ and $\xi_i^{All}(0) = \{\xi_{\alpha i}^{All}(0), \xi_{ci}^{All}(0)\}$. Then, we use $\theta_i^{All,W}(0)$, which are the initial values of θ_i^{All} relative to the white matter structures, as final values for θ_i^W and we integrate backward Eq.4.24 and Eq.4.25 obtaining $\theta_i^W(0)$ and $\xi_i^W(0) = \{\xi_{\alpha i}^W(0), \xi_{ci}^W(0)\}$. From this set of equations, we can notice that the optimisation of the two diffeomorphisms is linked by the constraint $\theta_i^W(1) = \theta_i^{All,W}(0)$ and that the information given by $\nabla_{\boldsymbol{T}_i^{All}(1)}\boldsymbol{D}_i$ where $\boldsymbol{D}_i = \{D_{ij}\}_{j=1,\dots,M}$ flows first through the second diffeomorphism All and then through the first one W and eventually it is used to update all the parameters in the following way:

$$
\begin{aligned}
\nabla_{\boldsymbol{T}^G} E &= \sum_{i=1}^{N} \theta_i^{All,G}(0) & \nabla_{\boldsymbol{T}^W} E &= \sum_{i=1}^{N} \theta_i^W(0) \\
\nabla_{\boldsymbol{\alpha}_{s0}^{All}} E &= \xi_{\alpha s}^{All}(0) + (\Gamma_\alpha^{All})^{-1}\boldsymbol{\alpha}_{s0}^{All} & \nabla_{\boldsymbol{\alpha}_{s0}^W} E &= \xi_{\alpha s}^W(0) + (\Gamma_\alpha^W)^{-1}\boldsymbol{\alpha}_{s0}^W \\
\nabla_{\boldsymbol{c}_0^{All}} E &= \sum_{i=1}^{N} \xi_{ci}^{All}(0) & \nabla_{\boldsymbol{c}_0^W} E &= \sum_{i=1}^{N} \xi_{ci}^W(0)
\end{aligned}
\tag{4.26}
$$

where $\theta_i^{All,G}$ refers to the values of θ_i^{All} relative to the grey matter structures. A sketch of the optimisation procedure to update all parameters is shown in Fig.4.3. It should be noted that the flow of information goes in the opposite sense, namely from the second global diffeomorphism All to the first white one W, with respect to the one shown in Fig.4.2 to deform the template complex $\boldsymbol{T}$. More details about the computations can be found in Sec.4.A.2.

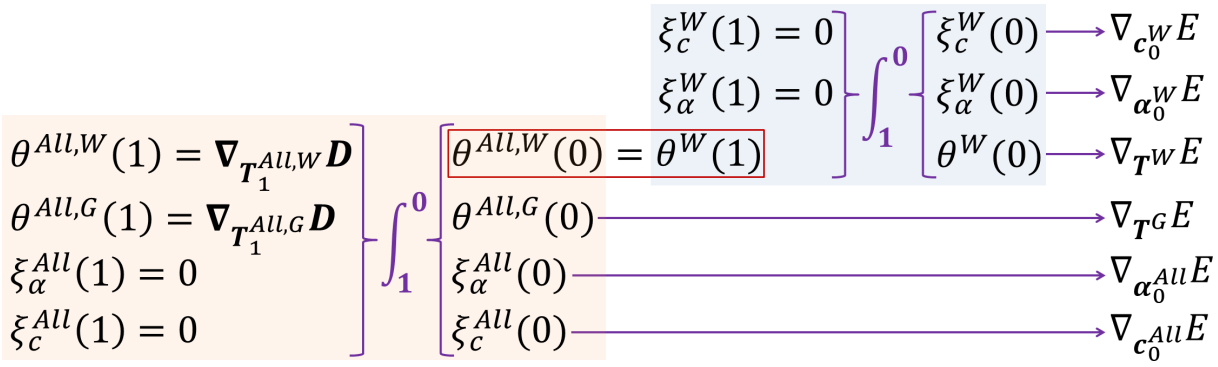

Figure 4.3: Sketch of the optimisation procedure for the template complex and the deformation parameters. We omit the subject index i for clarity purpose.

4.2.6 Initialisation of the parameters

The use of a gradient descent scheme to optimise the atlas parameters requires an initialisation. Control points of both diffeomorphisms are initialised as a regular lattice covering the entire ambient space with a inter-points distance equal to the bandwidth of the diffeomorphic kernel $\boldsymbol{K}_D$. Momenta are initialised to zero, namely no deformations. The third parameter is the template complex which is composed of two kinds of meshes: 3D surfaces and bundles of streamlines approximated as weighted prototypes. The initial template should not be chosen among the meshes of the subjects since it would inherit all the imperfections due to segmentation errors or data noise, biasing consequently the atlas estimation.

For surfaces, we use the average of the population when it is available a vertex-correspondence, otherwise we use a regular sphere which is first centred with respect to the ensemble of subject meshes and then it is scaled to an ellipsoid using the three main modes of variations of the meshes considered as points cloud.

For bundles of weighted prototypes, we propose to gather all the original streamlines of the subjects into a single bundle which is then approximated as a set of weighted prototypes. The weights of the prototypes are then scaled so that the norm of the template is equal to the average norm of the population.

4.3 Experiments and Results

The dataset used in this chapter contains 76 subjects: 27 controls and 49 patients subject to Gilles de la Tourette syndrome (GTS) and divided in three sub-groups based on their symptoms: $G1$=simple-tics, $G2$=complex-tics, $G3$=complex tics with obsessive compulsive disorders (OCD).

In this section, we will first evaluate the robustness of the proposed framework with respect to the size of both the varifold and weighted current grid using left caudate and the bundle of prototypes connecting it to the left hemisphere of the cortical surface. Then, we will take advantage of the proposed atlas construction procedure to investigate morphological and organisational abnormalities of the left cortico-putamen circuit in the population of GTS patients. These alterations are likely to be associated to a pathological development of the brain. The employed neural circuit is composed of left putamen, left hemisphere of the cortical surface and the streamlines (prototypes) connecting them.

In the following experiments, we first compute an atlas of 50 iterations using only the sub-cortical structures employed. This permits to transform the ellipsoid in a structure more similar to the sub-cortical nucleus. After that, we cut the prototypes of the template bundle at the intersection with the inter-linked grey matter structures.

4.3.1 Materials

Anatomical scans of all subjects were acquired using 3D T1-weighted sequences with a voxel size of 1x1x1 mm^3. Diffusion weighted scans were acquired with sequences of 50 directions with a B-factor of 1000 and a voxel size of 2x2x2 mm^3. For each subject, we consider the left hemisphere of the cortical surface, left putamen, left caudate and the fiber bundles connecting these sub-cortical nuclei to the left hemisphere of the cortex. The cortical surface is segmented using FreeSurfer v5.3 (Fischl et al., 2004) from the T1-weighted images followed by a pipeline of BrainVisa 4.3.0 which allows to have vertex-correspondence between subjects. The putamen and caudate are segmented with FSL (Patenaude et al., 2011) from the T1-weighted images. Since we want to analyse as a single structure nucleus accumbens and caudate, we merge these two segmentations together. We use the marching cubes algorithm to create the 3D mesh (AimsMesh function of Brainvisa 4.3.0). The fiber bundles result from a deterministic tractography algorithm (1 seed per voxel) available in BrainVISA/Connectomist-2.0 (Perrin et al., 2005). We use the Spherical Deconvolution Transform (SDT) model (Descoteaux et al., 2009) to estimate the local underlying orientation distribution function (ODF). Whole brain connectivity is then inferred within an anatomy-based tractography mask (Guevara et al., 2011) from which the fiber bundles are extracted as explained in (Worbe et al., 2015). The bundles include also the commisural fibers which are truncated at the interhemispheric fissure. All the other fibers are cut at the intersection with their respective sub-cortical structure and at the border between white and gray matter of the cortex. For more details about the acquisition, preprocessing of both T1-weighted and diffusion-weighted images or the tractography, the reader is referred to (Worbe et al., 2015).

4.3.2 Robustness with respect to the grid size Λ_j

We evaluate here the robustness of the proposed method with respect to the number of points of the grids (Λ_j) where both the structures and the template modelled as varifolds or weighted currents are projected to. We estimate 16 atlases using a group of 8 anatomical complexes composed of left caudate and left caudate bundle. The left caudate is modelled as varifold using λ_W=3mm and the left caudate bundle as weighted currents with λ_g=7mm, λ_a=10mm (cortex) and λ_b=5mm (caudate). Atlases share the same parameters except for the grid size of both structures. In Fig.4.4 we show the results where V and WC indicate the number of points of the Varifold and Weighted Currents grid respectively. The first row represents the evolution of the residuals of both structures during the atlas construction. Similarly, the second row represents the evolution of the logarithm of the determinant of the two covariance matrices, namely $\log(|\Gamma_\alpha^W|)$ and $\log(|\Gamma_\alpha^{All}|)$. The last row describes instead the point-wise Euclidean difference $||\cdot||_2$ between the estimated template and the initial one.

It is possible to notice that when the grid size of the bundle is greater than

the one of the nucleus ($WC > V$), results are very stable and satisfactory. The only exception is when the difference is too great, namely with $V = 1000$ and $WC = 500.000$. In such a case, the residuals of the caudate are higher than in the other examples and the template does not evolve since its data-term is too small with respect to the one of the bundle (see Fig.4.5 top-left). On the contrary, when $V > WC$ like with $V = 100.000$ and $WC = 1.000$, we have the opposite effect since the residuals of the bundle are higher and its template does not converge (see Fig.4.5 top-right).

In the following experiments, we propose to use an heuristic to define V and WC. They are selected as the number of points required to build a regular linearly-spaced grid covering *only* the subject shapes and the template of each structure and the distance between every couple of points is given by the smallest value between λ_W and λ_g among all structures. In this example, the smallest value is the one of the caudate, namely 3mm, and the grid sizes result $V \approx 5.000$ and $WC \approx 110.000$. Results based on this choice are satisfactory and as it is possible to see from Fig.4.5 at the bottom, both templates are updated during the atlas construction.

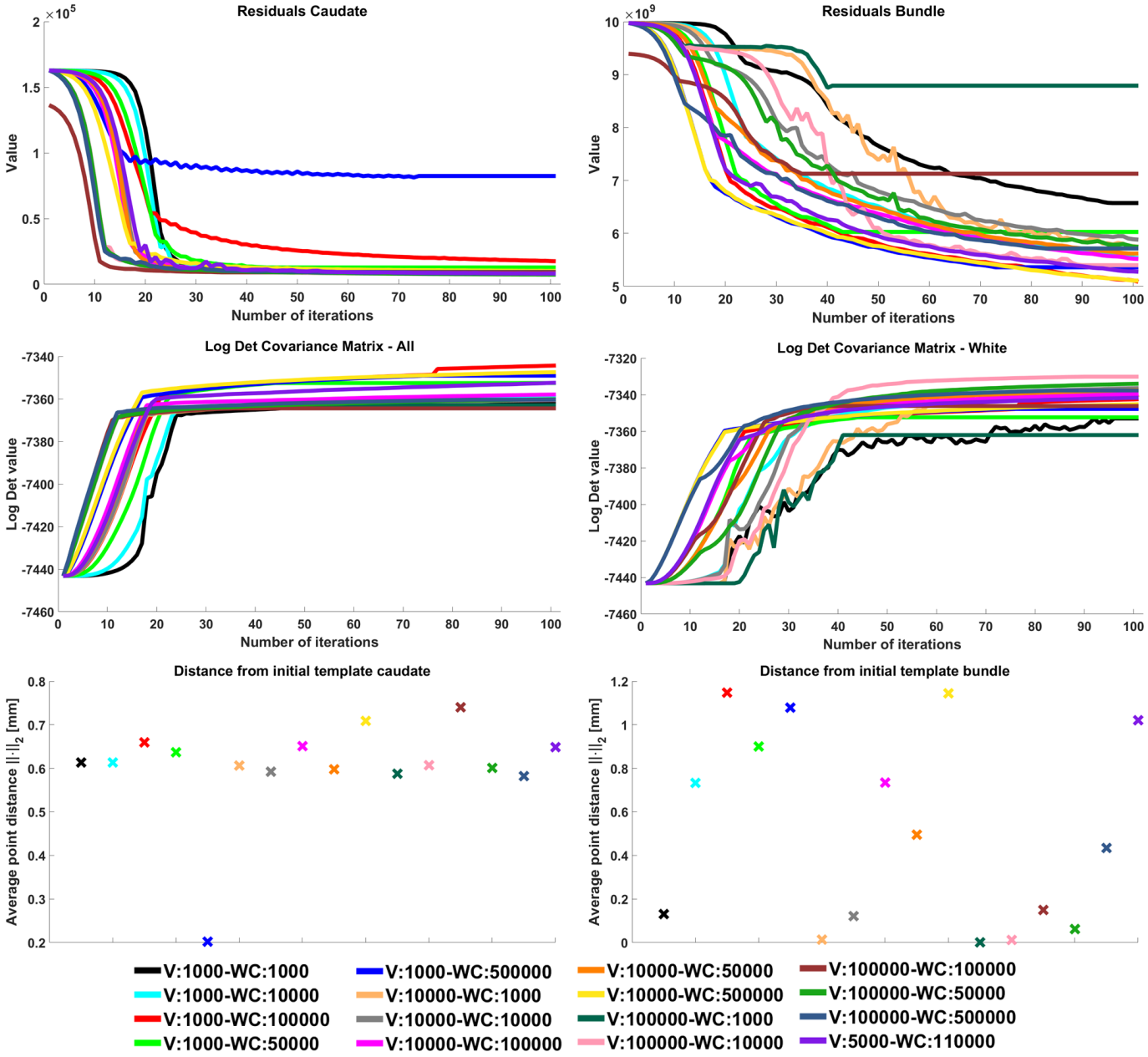

Figure 4.4: Evaluation of the robustness of the results with respect to the number of points of the varifold (V) and weighted current (WC) grids. Colours refer to 16 atlases computed using different values of grid sizes. Each atlas is based on 8 anatomical complexes composed of left caudate, modelled as varifold, and left caudate bundle, modelled as weighted current. The maximum number of iterations for all atlases is 100.

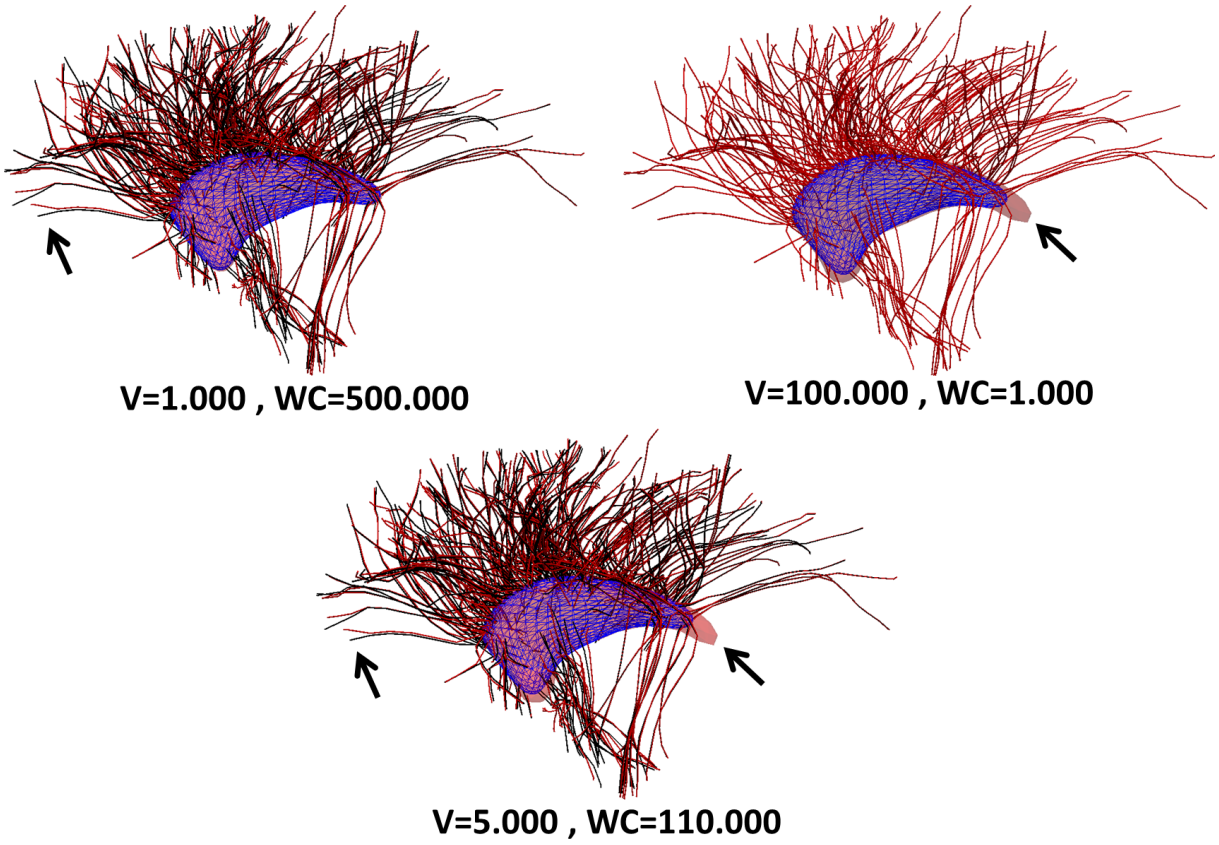

Figure 4.5: Comparison of three estimated template complexes (in red) with respect to the initial template complex (in blue and black). The left caudate is modelled as varifold and the bundle as weighted current. Atlases share the same parameters except for the number of points of the grid, V for varifolds and WC for weighted currents. Black arrows indicate the areas where the initial template complex has evolved during the atlas construction.

4.3.3 Group differences

In all the following experiments we use the cortico-putamen neural circuit. The cortical surface is modelled with landmarks, the putamen as varifold with λ_W=3mm and the fiber bundle connecting them as weighted currents with λ_g=7mm, λ_a=10mm (cortex) and λ_b=5mm (putamen). All bundles are approximated as weighted prototypes. The bandwidths of both diffeomorphic kernels is equal to 11mm and the number of control points is 816 for both diffeomorphisms. The maximum number of iterations is 120 and all the computations are performed on a Intel Xeon, 32 cores, CPU E5-2650, 2.60GHz with a graphic card NVIDIA Quadro 5000. The code is written in C++ and CUDA and it is an extension of the freely available software *deformetrica* (www.deformetrica.org). The computational time for an atlas of 30 subjects is equal to 144 hours for 120 iterations. All shape complexes are previously rigidly registered to a reference shape complex.

It is important to notice that a bandwidth of 11mm permits to deform all white matter fiber bundles in a correlated way with at least one grey matter mesh. This permits to have an identifiable model with a unique decomposition of the two diffeomorphisms all over the ambient space.

4.3.3.1 Atlas with pooled subjects

Toy example We start this section with an explanatory example based on a toy data-set constituted of 6 pseudo shape complexes belonging to two different populations (i.e. controls and GTS patients). Each complex is composed of a pseudo cortical surface, a pseudo sub-cortical structure and a pseudo fiber bundle linking them. They are shown in Fig.4.6 where it is possible to notice that the complexes of population A have a differ organisation and shape with respect to the ones of population B. Indeed, subjects from A have a bigger cortex and their sub-cortical structure have a bump on the right whereas in population B it is on the left and the fiber bundles of the two populations integrate completely different grey matter areas. In population A fibers go from the right area of the sub-cortical structure to the left area of the cortex whereas in population B it is exactly the contrary.

We estimate an atlas considering both populations together as if they were from a single group. We start with a simple and unbiased initial template complex and we obtain a final template complex and the covariance matrices of the momenta of both diffeomorphisms. The template shows the features common to both populations. The two covariance matrices describe the *morphological* and *organisational* variability within the 6 subjects. We compute a Principal Component Analysis (PCA) for each covariance matrix and we deform the final template complex at $\pm\sigma$ (standard deviation) along the first modes of both PCAs.

Since the intra-group variations are definitely smaller than the inter-group ones, the main variations among the 6 subjects captured by the white diffeomorphisms $\{\phi_i^W\}$, which affect only the fiber bundles, explain the principal differences in structural connectivity between the two populations. The positions of the fiber bundle

at $-\sigma$ and $+\sigma$ are the ones of population A and B respectively. Moreover, even the first mode of the global PCA describes the main global morphological variations between the two populations. As previously, we can notice that both grey matter structures at $-\sigma$ and $+\sigma$ reproduce the morphological characteristics of population A and B respectively.

This example shows the exploratory potential of the proposed algorithm even if it is based on a simple toy data-set where the inter-group differences are probably exaggerate compared to a real-data example. Nevertheless, given the important structural changes that are likely to occur in syndromes such as GTS, we may assume that controls and patients create separate clusters. The first modes of variations of the two diffeomorphisms should therefore describe the main differences between these two clusters highlighting the effects of the pathology on both the anatomy and structural connectivity.

This could not be achieved using a single diffeomorphic atlas since the template-to-subject registrations would not be able to put into correspondence all the structures of the complex. An example of registration is shown in Fig.4.7, where it is possible to notice that the matching using a single diffeomorphism is not accurate. Instead, a double diffeomorphic registration makes first the fiber bundle sliding onto the grey matter structures and then it accurately aligns with the second (global) diffeomorphism all structures of the template complex to the homologous ones of the subject complex. This permits to disentangle the differences in structural connectivity, captured by the first diffeomorphism, from the global morphological changes, captured by the second diffeomorphism.

Cortico-putamen circuit Following the same strategy we also build an atlas with pooled subjects based on real data. We use the cortico-putamen circuits of 10 controls and 10 patients subject to GTS. In Fig.4.8 we show the estimated template complex and its deformations along the first mode at ± 3 std of a PCA based on the initial momenta of the white diffeomorphisms. The first and last column present the densities of the endpoints of the template bundle onto the fixed grey matter structures. Colours in the middle column highlight the displacement of the deformed template bundle along the first mode. The two arrows point to the most variable areas. The first one, at the top, is in the superior-frontal area of the cortical surface. As it is possible to notice from the densities on the cortex, fibers tend to concentrate in a specific area in the first row which is then left almost empty in the bottom row. The second arrow points to the lateral part of the putamen where fibers move from the bottom to the central part in the first and last row respectively. This movement is clearly visible in the right column.

In Fig.4.9 we present the main global morphological variations of both the cortical surface and the putamen. We compute the first two modes of a PCA based on the initial momenta of the global diffeomorphisms. Then, we deform the estimated template (shown in the middle in gray) along the two modes at ± 3 std. Colours highlight the displacement of the deformed template cortex. The first mode captures changes mainly in the parietal area of the cortex whereas the second mode

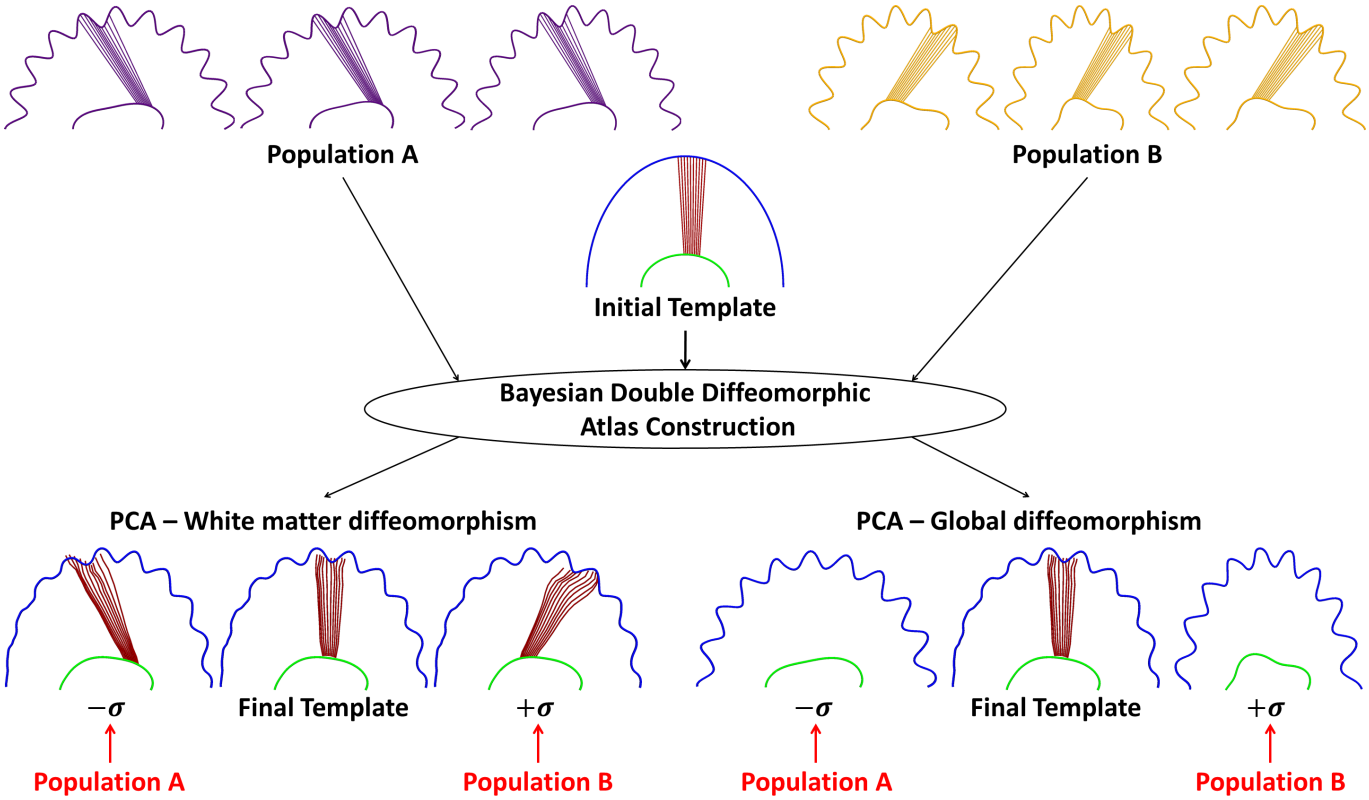

Figure 4.6: Toy example of double diffeomorphic atlas construction. At the top we present the shape complexes of two different populations. The intra-variability of each group is smaller than the inter-variability. Subjects of A have a bigger cortex than subjects in B and both the shape of the sub-cortical nucleus and the structural connectivity are different. In the middle we show the initial template which is updated during the atlas construction. The final template is presented at the bottom and it is deformed at $\pm\sigma$ along the first modes of two Principal Component Analysis (PCA) computed with the covariance matrices of the momenta of the two diffeomorphisms. Due to the great inter-variability between the two groups, the first mode of the white PCA (first diffeomorphism) presents the main differences in structural connectivity between the two populations. The first mode of the global PCA (second diffeomorphism) shows instead the global morphological variations between the two groups.

in the middle and superior frontal areas. The variations in the putamen concern principally its relative position with respect the other structures. The first and second modes present almost a rigid movement along the horizontal and vertical axis respectively. The second mode presents also a global expansion/shrinking.

The changes presented in these two figures could highlight the main variations between the two groups or they could show their common main variability. In the next sub-section we will compare these results with the ones obtained by analysing

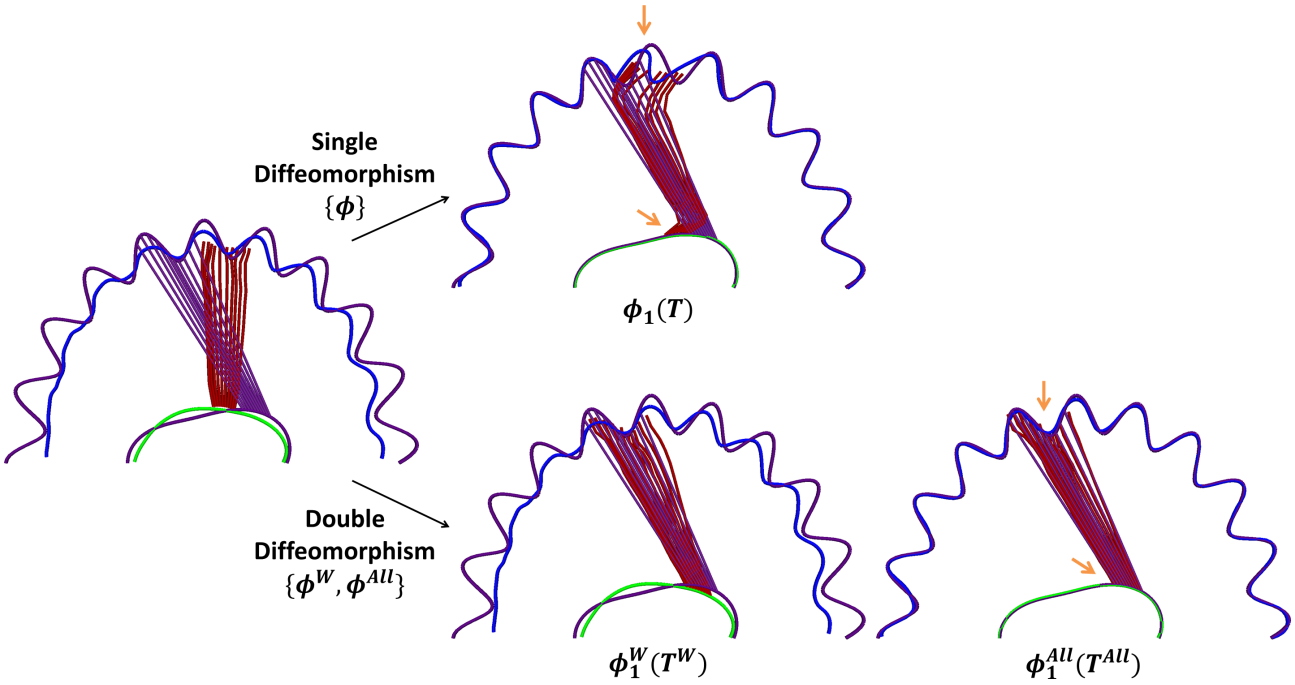

Figure 4.7: Example of template-to-subject registration using both a single and a double diffeomorphism. Due to the variation in structural connectivity between the template complex (blue, green and red) and the subject complex (violet), a single diffeomorphism can not put into correspondence all structures. Instead, a double diffeomorphism can first make sliding the fibers onto the grey matter structures, showing the variation in structural connectivity, and then it can correctly register all structures. Orange arrows indicate the changes in structural connectivity.

each group independently.

We also computed the average of the initial momenta of each population and we deformed the template complex along these two directions. This gave us a quantitative way to compare the average configurations of both groups. Results (not shown) were quite similar and no differences stood out.

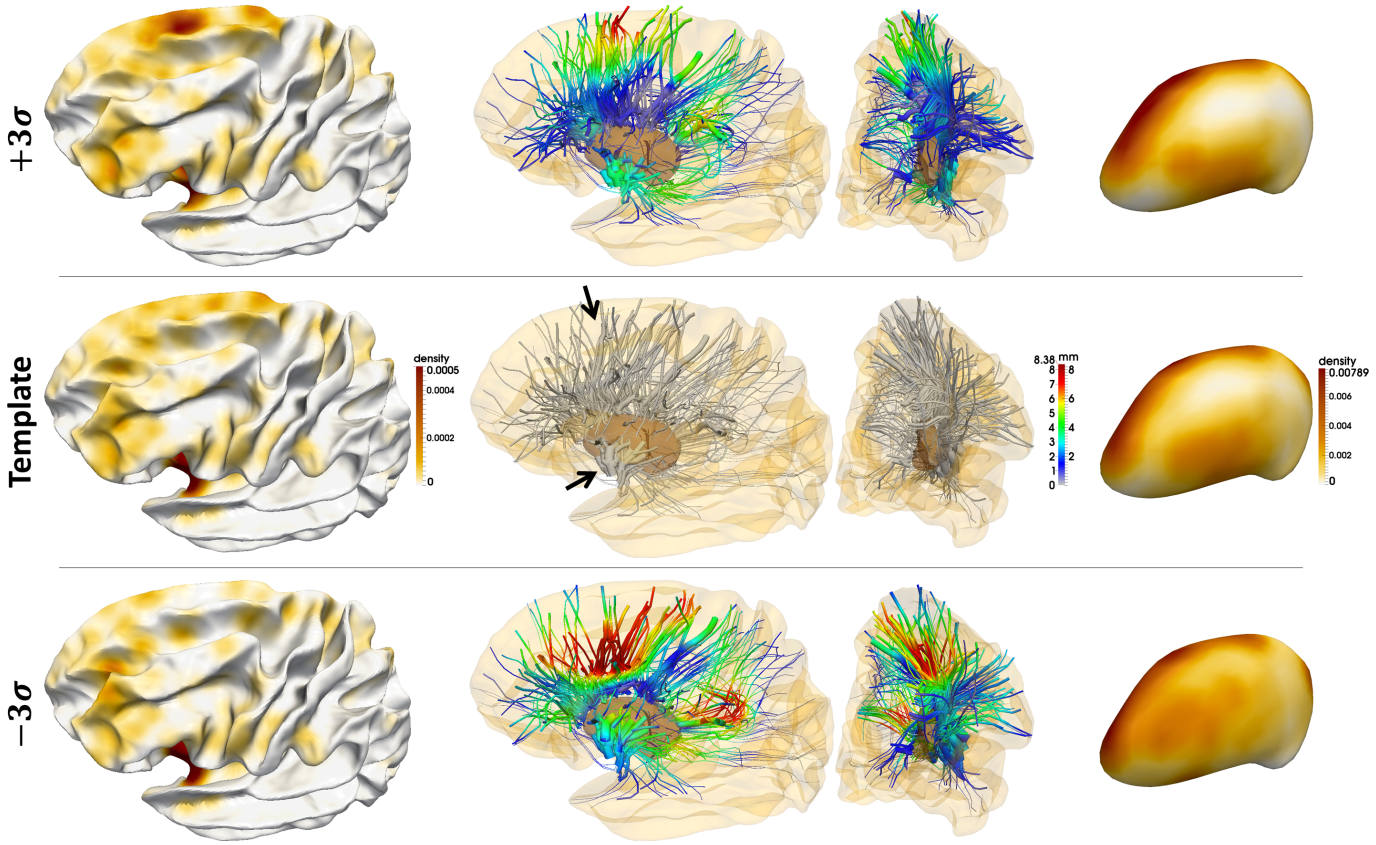

Figure 4.8: Main variations in structural connectivity of an atlas composed of 10 controls and 10 patients. The final template bundle, shown in the middle, is deformed at ± 3 std along the first mode of a PCA based on the initial momenta of the white diffeomorphisms. Colours of the left and right columns indicate the densities of the endpoints of the template bundle onto the fixed grey matter structures. Colours in the middle column highlight instead the displacement of the template bundle at the end of the deformations. Arrows point to the most variable areas.

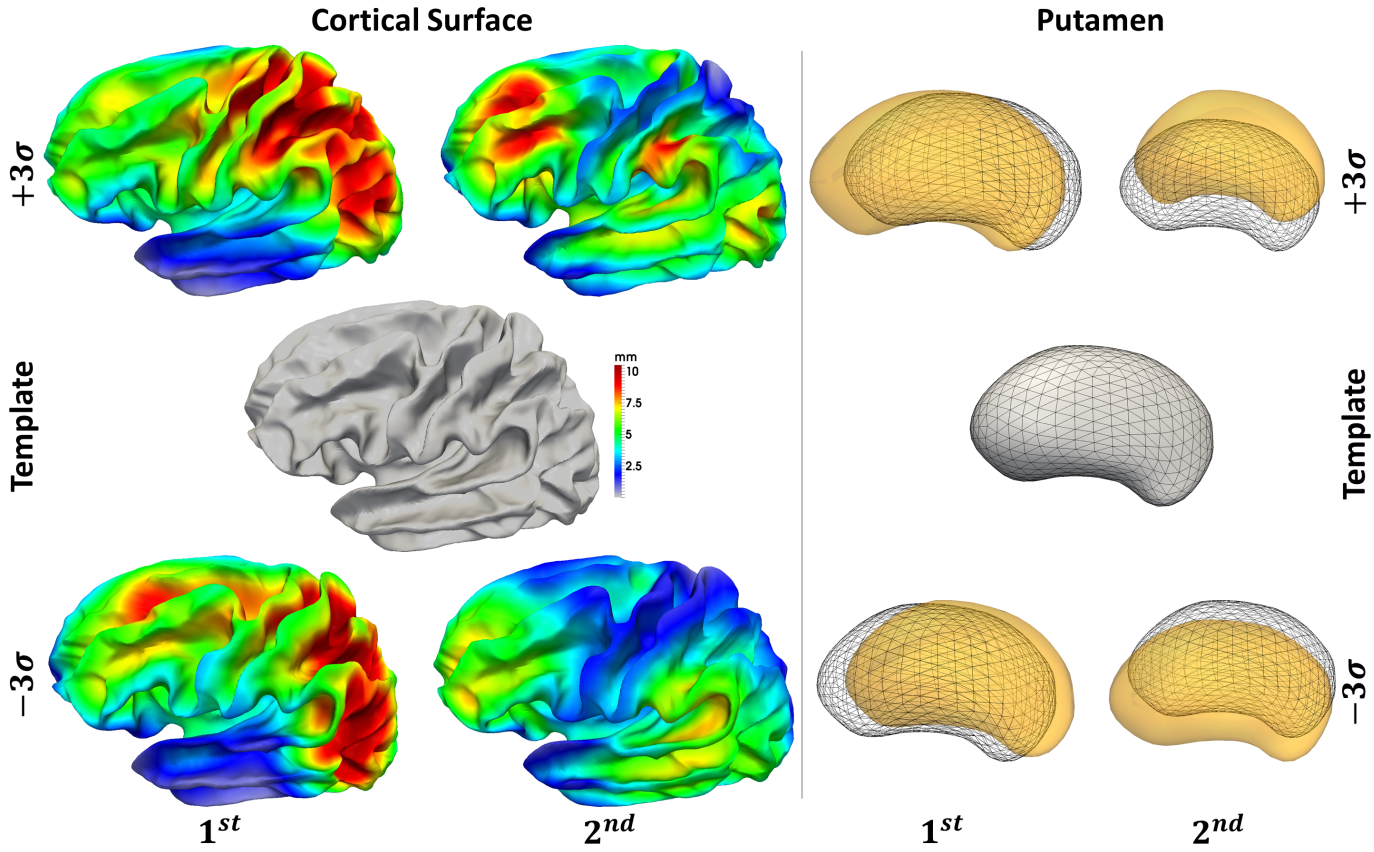

Figure 4.9: Main global morphological variations of an atlas composed of 10 controls and 10 patients. The final templates of the grey matter, shown in grey in the middle row, are deformed at ±3 std along the first and second mode of a PCA based on the initial momenta of the global diffeomorphisms. Colours on the left highlight the displacement of the template cortex at the end of the deformations.

4.3.4 Separated Atlases

We compute here two different atlases using 30 patients ($10G1$ + $10G2$ +$10G3$) and 27 controls respectively. We use the same initial template complex as in the previous sub-section. We compute the first three modes of a PCA based on the initial momenta of the white diffeomorphisms for both groups. Then, we deform the final template complex along every mode at ± 3 std. Deformations are shown in Fig.4.10 for controls and in Fig.4.11 for patients. These two figures present the main variability in structural connectivity within each group.

By comparing them, we can notice that the first mode of each group captures variations relative to the commisural fibers passing through the corpus callosum. This means that variations in this area are probably common to both groups. The other two modes show instead changes affecting different areas. In particular, in controls the most variable area is the superior-frontal instead in patients the variations are more heterogeneous. Interestingly, the changes of the second mode in patients regard almost the same areas as in Fig.4.8. This might signify that the group of patients, due to the vast clinical heterogeneity of GTS, might create a large cluster in the shape space which might even intersect the one of controls creating a kind of continuum from the group of controls to the most severe patients. If we suppose then that the intra-variability of the group of patients is greater than the one of controls we could explain why the main variations found considering controls and patients together are very similar to the greatest changes estimated among only patients.

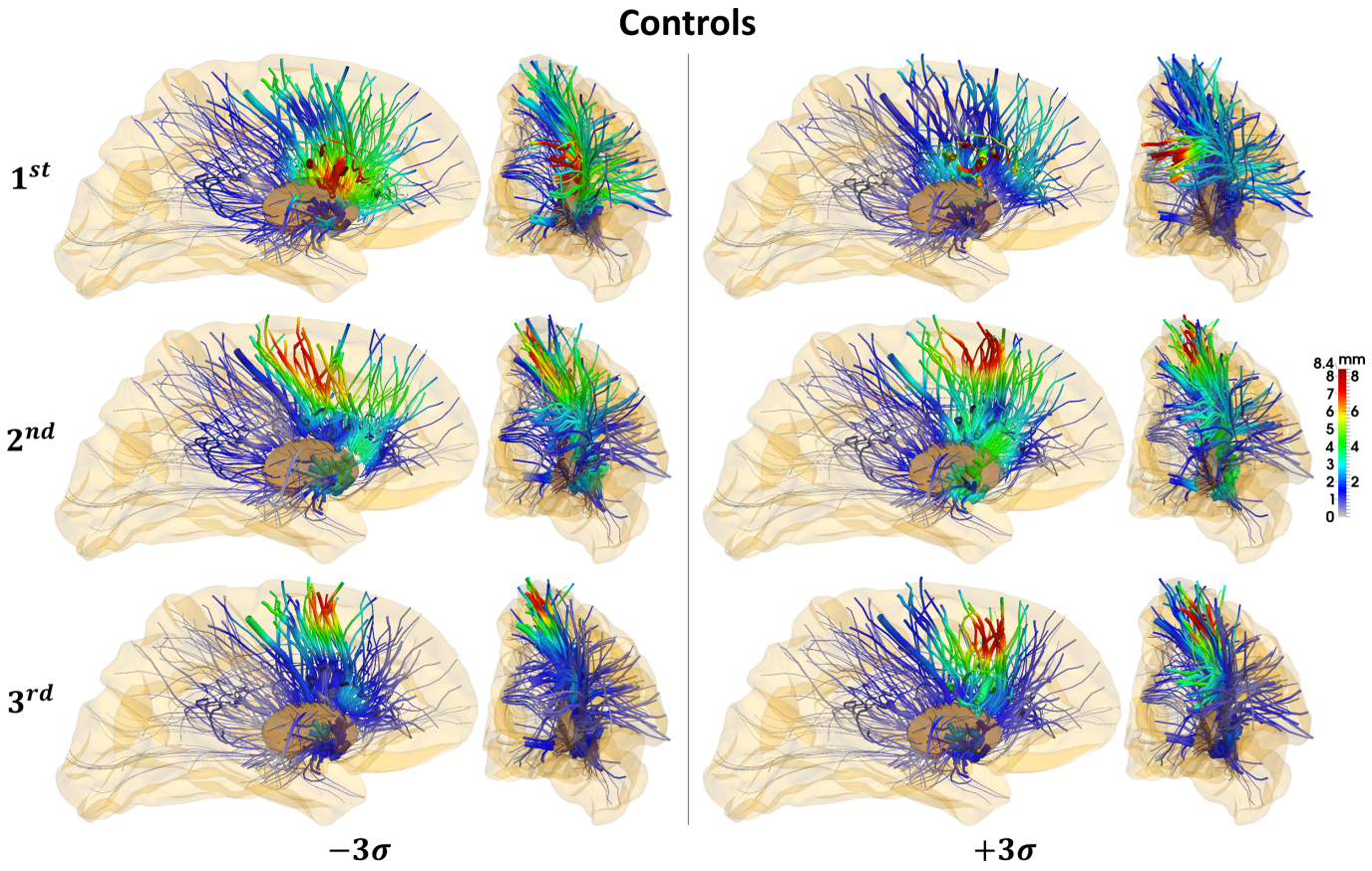

Figure 4.10: Main variations in structural connectivity of an atlas estimated using 27 healthy controls. Each row shows the final template bundle deformed at ±3 std along the first three modes of a PCA based on the initial momenta of the white diffeomorphisms. Colours refer to the displacement (in mm) of the template bundle at the end of the deformations.

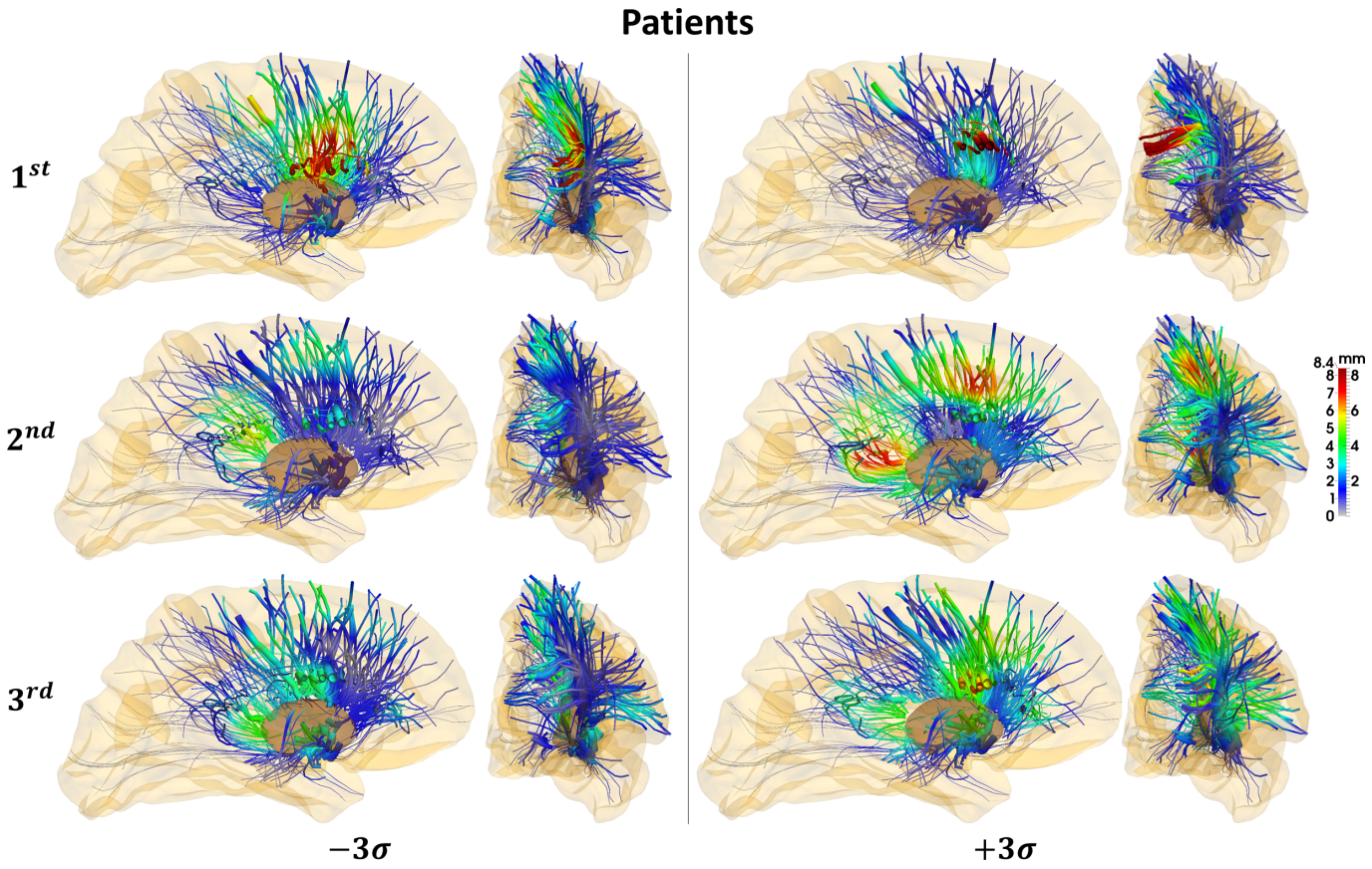

Figure 4.11: Main variations in structural connectivity of an atlas estimated using 30 GTS patients. Each row shows the final template bundle deformed at ±3 std along the first three modes of a PCA based on the initial momenta of the white diffeomorphisms. Colours refer to the displacement (in mm) of the template bundle at the end of the deformations.

4.3.5 Classification

Eventually, we use the initial momenta of the proposed double diffeomorphic deformation to discriminate between controls and patients and we compare the classification scores with the ones obtained using the initial momenta of a single diffeomorphism.

First of all, we build an atlas with 10 subjects (5 controls and 5 patients) which results in a template, two sets of control points and two well-conditioned covariance matrices of deformation parameters of the white (Γ_α^W) and global (Γ_α^{All}) diffeomorphism respectively. Since we use subjects from both clinical groups, the template should be positioned in between them in the shape space. The template is successively warped to all the *remaining* J subjects fixing the control points equal to the estimated ones and using the two covariance matrices as regularity terms, namely minimising for each subject:

$$\begin{aligned}&\sum_{j=1}^{M^W} \frac{1}{2\sigma_j^2}||\Pi(S_j - \phi_1^{All}\left(\phi_1^W(T_j^W)\right))||^2_{Q^*_{\Lambda_j}} + (\boldsymbol{\alpha}_0^{All})^T(\Gamma_\alpha^{All})^{-1}\boldsymbol{\alpha}_0^{All} \\ &\sum_{j=1}^{M^G} \frac{1}{2\sigma_j^2}||\Pi(S_j - \phi_1^{All}(T_j^G))||^2_{W^*_{\Lambda_j}} + (\boldsymbol{\alpha}_0^{W})^T(\Gamma_\alpha^{W})^{-1}\boldsymbol{\alpha}_0^{W}\end{aligned} \tag{4.27}$$

which is equal to Eq.4.20 but with only one subject (index i disappears) and with a constant σ_j^2 which is fixed to the value estimated in the atlas. The resulting initial momenta $\{\boldsymbol{\alpha}_{i0}^{All}\}_{1...J}$ and $\{\boldsymbol{\alpha}_{i0}^{W}\}_{1...J}$ represent the features used as input in the classifier. We employ Linear Discriminant Analysis (LDA) with a leave-one-out cross validation strategy. We test the discriminative power of the two diffeomorphisms separately by using either only $\{\boldsymbol{\alpha}_{i0}^{All}\}_{1...J}$ or only $\{\boldsymbol{\alpha}_{i0}^{W}\}_{1...J}$. In each experiment, we suppose that the class-conditional densities of the initial momenta for both clinical groups are Gaussian, namely $f_c(\boldsymbol{\alpha}_{i0}) = N(\boldsymbol{\mu}_c, \Gamma_c)$ and $f_p(\boldsymbol{\alpha}_{i0}) = N(\boldsymbol{\mu}_p, \Gamma_p)$, and that they have priors equal to their relative number, namely $\pi_c = J_c/J$ and $\pi_p = J_p/J$ where the indices c and p refer to controls and patients respectively, π_c and π_p are the prior distributions, $J_c = 22$ is the number of controls and $J_p = 44$ is the number of patients.

In LDA the feature space is divided into two parts by a decision boundary which is based on the Bayes factor $\mathscr{B}$ defined as:

$$\begin{aligned}\mathscr{B}(\boldsymbol{\alpha}_t) =& \log\left(\frac{f_c(\boldsymbol{\alpha}_t)}{f_p(\boldsymbol{\alpha}_t)}\right) + \log\left(\frac{\pi_c}{\pi_p}\right) \\ =& -\frac{1}{2}(\boldsymbol{\alpha}_t - \boldsymbol{\mu}_c)^T\Gamma_c^{-1}(\boldsymbol{\alpha}_t - \boldsymbol{\mu}_c) - \frac{1}{2}\log|\Gamma_c| \\ &+ \frac{1}{2}(\boldsymbol{\alpha}_t - \boldsymbol{\mu}_p)^T\Gamma_p^{-1}(\boldsymbol{\alpha}_t - \boldsymbol{\mu}_p) + \frac{1}{2}\log|\Gamma_p| + \log(\frac{J_c}{J}) - \log(\frac{J_p}{J})\end{aligned} \tag{4.28}$$

where $\boldsymbol{\alpha}_t$ is the initial momenta of the test subject t and we omitted the index

0 for clarity purpose. If $\mathscr{B}$ is smaller than a threshold τ the subject is classified as patient otherwise as control. Moreover, in LDA the covariance matrices of the two groups are assumed to be equal and with full rank (i.e. $\Gamma_c = \Gamma_p = \Gamma$) which means that Eq.4.28 becomes:

$$\mathscr{B}(\boldsymbol{\alpha}_t) = \boldsymbol{\alpha}_t^T \underbrace{\Gamma^{-1}(\boldsymbol{\mu}_c - \boldsymbol{\mu}_p)}_{\boldsymbol{w}} - \underbrace{\left(\frac{1}{2}\boldsymbol{\mu}_c^T\Gamma^{-1}\boldsymbol{\mu}_c - \frac{1}{2}\boldsymbol{\mu}_p^T\Gamma^{-1}\boldsymbol{\mu}_p - \log(\frac{J_c}{J}) + \log(\frac{J_p}{J})\right)}_{b} \tag{4.29}$$

The decision boundary is therefore linear, namely it is the hyperplane in the feature space: $\boldsymbol{\alpha}_t^T\boldsymbol{w} - b - \tau = 0$. Depending on the side where the $\boldsymbol{\alpha}_t$ lies, the correspondent test subject t is classified as patient < 0 or control > 0.

To estimate the accuracy of our model without incurring in over-fitting, we use a leave-one-out cross validations technique. At each iteration, we estimate the best hyperplane using J-1 subjects and we test it on the left out subject t. First, we compute the averages of both controls and patients within the J-1 subjects. Assuming that the test subject t is a control, it results:

$$\boldsymbol{\mu}_c = \frac{1}{J_c - 1}\sum_{i=1}^{J_c-1}\boldsymbol{\alpha}_{ic} \qquad \boldsymbol{\mu}_p = \frac{1}{J_p}\sum_{i=1}^{J_p}\boldsymbol{\alpha}_{ip} \tag{4.30}$$

After that, using the covariance matrix estimated in the initial atlas as Γ in Eq.4.29, we compute the best threshold τ as the one which gives the highest balanced accuracy BA, namely the average between specificity and sensitivity. When there is a continuous range of values at which BA is maximal, we choose the middle one. Once estimated the best hyperplane in the training set, we use it to test the left-out subject t. We iterate this process for all the J subjects of the data-set. Resulting sensibility, sensitivity and balanced accuracy are shown in Table 4.1.

Since the group of patients is composed of three sub-groups, namely $G1$=simple-tics (17 patients), $G2$=complex tics (15) and $G3$=OCD (12), we test our classification algorithm using these sub-groups separately or together. Moreover, we compare these results with the ones obtained using the initial momenta of a single diffeomorphism. Following the same strategy, we first estimate a single-diffeomorphic atlas (like in Chapter 2) using the same 10 subjects as before. Then, the estimated template is warped with a single-diffemorphic deformation towards the remaining J subjects. Even in this case we fix the control points and we use the estimated covariance matrix as regularity term. Eventually, we employ LDA with a leave-one-out cross validation technique to classify the initial momenta exactly as before.

As it is possible to notice from Table 4.1, the classification scores based on a single diffeomorphism are worse than using the proposed double diffeomorphism especially when using the most severe groups of patients ($G2$ and $G3$).

Due to the variability of the results in Table 4.1, we also investigate their distributions within the group of patients with a bootstrap analysis. More precisely, we

perform it on the top of the previous leave-one-out cross validation classification. At each iteration of the bootstrap, we pick a random sample (with replacement) of the 44 patients which is classified, together with the (un-modified) 22 controls, using LDA. We repeat this process 1000 times. Each bootstrap sample gives as result an estimate of sensitivity, sensibility and balanced accuracy. The histograms of balanced accuracy for both the double diffeomorphism and the single diffeomorphism are shown in Fig.4.12. The averages for sensitivity and specificity are respectively: 74% and 51% for the global diffeomorphism and 73% and 64% for the white diffeomorphism. Instead, the averages sensitivity and specificity for the single diffeomorphism are 64% and 48% respectively.

The results obtained with the double diffeomorphism are, even in this case, better than using a single diffeomorphism. This demonstrates that the proposed double diffeomorphic approach captures important and relevant information which better characterise the pathological anomalies of GTS. Moreover, looking at Table 4.1, we can see that there is a general trend in the classification scores. Whatever the diffeomorphic scheme employed, we always classify the most severe patients, namely groups $G2$ and $G3$, better than the ones with only simple tics ($G1$). This confirms the hypothesis that there might be a continuum in the shape space between healthy controls and GTS patients where the subgroup $G1$ would be "closer" to the healthy group than the subgroup $G3$. Furthermore, it also confirms that the atlas construction technique might be a good model to investigate GTS and more generally neurodevelopmental syndromes.

Single Diffeo			
	Sensitivity [%]	**Specificity [%]**	**Balanced Accuracy [%]**
G1	**12**	**36**	**24**
G2	**33**	**64**	**48**
G3	**58**	**59**	**59**
G2+G3	**52**	**64**	**58**
G1+G2+G3	**54**	**41**	**48**
Double Diffeo - White			
	Sensitivity [%]	**Specificity [%]**	**Balanced Accuracy [%]**
G1	**47**	**59**	**53**
G2	**67**	**77**	**72**
G3	**50**	**82**	**66**
G2+G3	**74**	**64**	**69**
G1+G2+G3	**73**	**41**	**57**
Double Diffeo - Global			
	Sensitivity [%]	**Specificity [%]**	**Balanced Accuracy [%]**
G1	**29**	**50**	**40**
G2	**40**	**45**	**43**
G3	**50**	**68**	**59**
G2+G3	**52**	**68**	**60**
G1+G2+G3	**70**	**50**	**60**

Table 4.1: Classifications scores using LDA and based on the deformation parameters of a single (top) and double diffeomorphism (middle and bottom) respectively. We show sensitivity, specificity and balanced accuracy using always 22 controls and either one of the patient sub-groups ($G1$,$G2$,$G3$) or combinations of them.

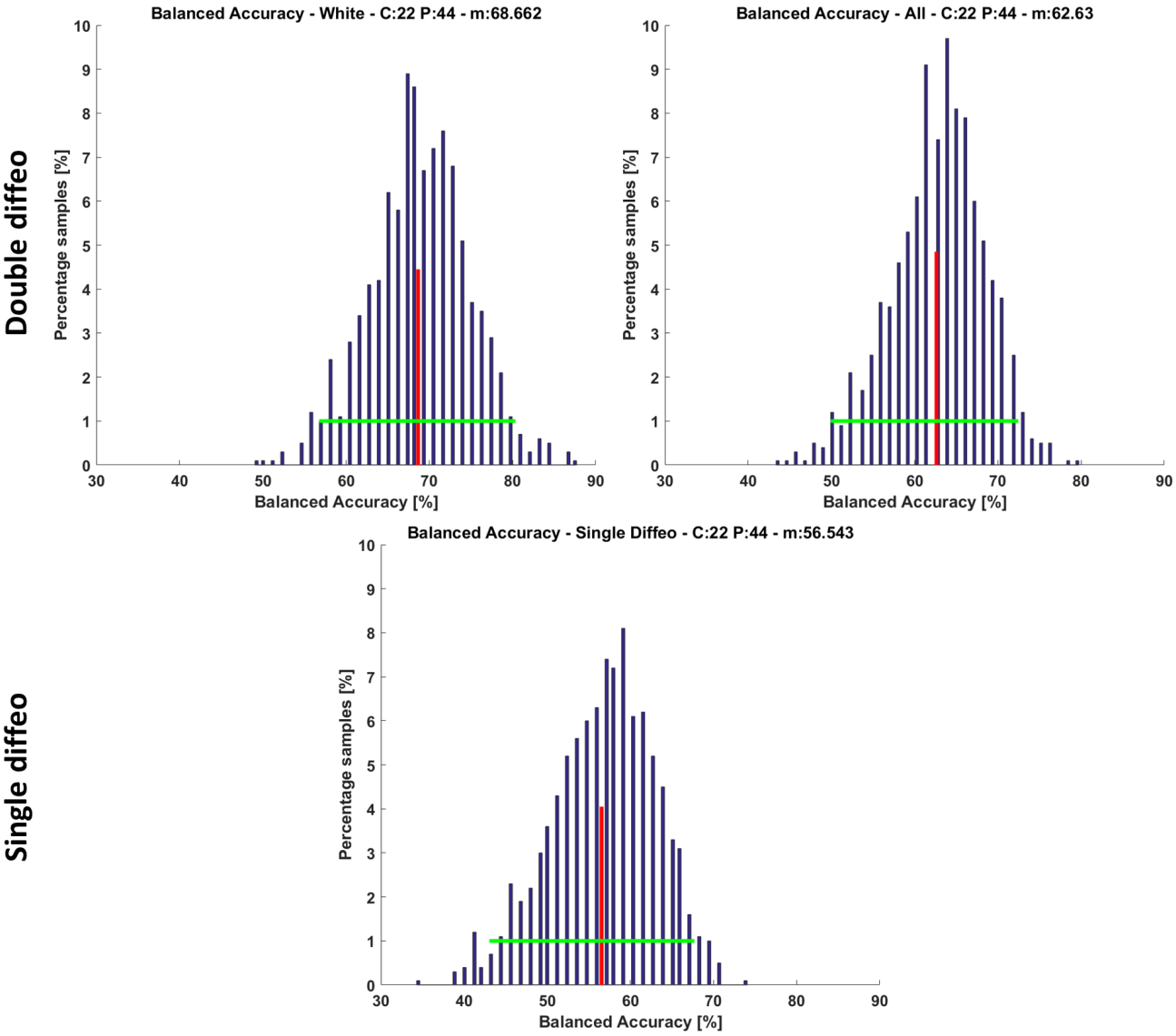

Figure 4.12: Bootstrap analysis of 1000 iterations performed on the top of a LDA with a leave-one-out cross validation. Each sample of the histogram represents the classification score obtained using 44 patients chosen randomly (with replacement) among all the sub-groups and 22 fixed controls. Red and green lines show the average and the 95% confidence interval respectively.

4.4 Discussion and Conclusion

We presented a method to analyse the organisational and morphological variability of a population of anatomical complexes composed of grey matter surfaces *interconnected* by white matter fiber bundles. It is built on a double-diffeomorphic mesh-based atlas construction embedded in a Bayesian framework. The resulting atlas consists of a template complex, showing the common characteristics of the population, and two well-conditioned covariance matrices describing the distributions of the deformation parameters, namely the initial momenta, of both diffeomorphisms.

The cascade of two diffeomorphisms permits to put into correspondence anatomical complexes characterised by fiber bundles connecting *different* areas of the grey matter, namely with a different structural connectivity. The first (white) diffeomorphism modifies the relative position of the fiber bundles with respect to the fixed grey matter structures in such a way that it is possible to correctly align the entire brain, both grey and white matter structures, with the second (global) diffeomorphism. In this way, we can disentangle the morphological variations into two categories: the ones related to the structural connectivity are captured by the first diffeomorphism and the ones about global morphological differences, affecting both white and grey matter structures, are instead described by the second diffeomorphism. The deformations obtained with the white diffeomorphisms can be compared across subjects since they are computed with respect to the same reference frame, that it the fixed grey matter. On the contrary, it should be noted that the deformations of the fiber bundles obtained with the global diffeomorphisms can not be compared across subjects. Global diffeomorphisms can be used to study the morphological variability of the grey matter structures. Both white and global changes can be visualised as a continuous deformation of the template complex. This permits to easily *characterise*, *localise* and *quantify* both organisational and morphological pathological anomalies altering grey and white matter structures.

Furthermore, our algorithm is limited for now to cases in which we can assume that there is a point-correspondence among the cortical surfaces in the population. In future works, we will try to adapt our technique for considering also subjects showing different cortical gyrifications. Another possible extension would be the use of a probabilistic tractography instead than the deterministic one employed here. This would make the analysis computationally more demanding but thanks to the approximation scheme presented in Chapter 3 it should still be feasible.

Both diffeomorphisms are parametrised by a different set of control points. The number of control points defines the dimension of the initial momenta. These are the features used in statistical analysis such as PCA or classification. It would be of interest investigating how a reduction of the number of control points would influence both the atlas construction and the statistical analysis. In Durrleman et al. (2014), the authors used a single-diffeomorphic atlas construction method similar to the one we proposed. They demonstrated that the statical performance in a classification augments by decreasing the number of control points until a certain threshold. It seems therefore reasonable to expect the same behaviour for the proposed method.

This brings to another question which is how to choose the position and the number of control points. A possible solution was presented in Allassonnière et al. (2015) by integrating the selection of the best control points in the minimisation process. The authors proposed to start from a regular grid which is trimmed by keeping only the control points that participate to the template-to-subject deformations of the whole population. The idea is brought from the Group LASSO Literature and it seems to work well in practice. It would be of interest to integrate it to the proposed model.

In this chapter, like in Chapter 2, we approximate the conditional distribution of the E step in the optimisation procedure with a Dirac distribution at its mode. It was demonstrated in Allassonnière et al. (2007) that this algorithm might not converge when the SNR (Signal to Noise Ratio) is low. A possible solution would be to employ sampling techniques such as Markov Chain Monte Carlo (MCMC) methods (i.e. SAEM-MCMC) which would converge whatever the quality of the data. Moreover, this would also permit to estimate the entire posterior distribution of the deformations, and not only its mode, which contains important information such as the registration uncertainty (Risholm et al., 2013; Simpson et al., 2012).

Another perspective regards the initialisation of the template of fiber bundles. In Chapter 2 and in this chapter we proposed two different methods. In both cases, we first combined the original streamlines of the bundles of the subjects in a single raw template. Then, we approximated it either by looking for the most representative streamlines (Sec.2.2.5) or by representing it with few scattered weighted prototypes (Sec.4.2.6). These methods can be considered as data-driven and they seem to work well in practice. However, due to the importance of the initialisation in gradient-descent algorithms, we think that it would be relevant to investigate the robustness of the results with respect to a change in the initial template. For instance, we could randomly divide a group of subjects into different sub-groups. Then, we could create the initial template and estimate the atlas in each sub-group separately. Results could then be compared across sub-groups giving an estimate of the robustness. However, this would require a considerable computational time.

All experiments shown in this chapter were about one neural circuit. To fulfil the goal of this thesis, namely an holistic approach integrating the whole CSTC network in a single framework, we should use the proposed method with all the neural circuits at the same time. However, this might not work well since every circuit might be affected by different pathological alterations. This means that every bundle should be deformed in a different and independent way with respect to the others. Thus, we should not use a single diffeomorphism to deform the entire white matter when composed by multiple bundles. A possible solution would be to use $N+1$ diffeomorphisms. Every bundle would be independently deformed by a white diffeomorphism. Then, all bundles would be deformed, together with the grey matter structures, by another global diffeomorphism. In this way, we could capture the different variations in structural connectivity proper to each bundle and also the global morphological changes associated to all structures of the CSTC circuits.

In this chapter, we used three different computational models: landmarks, varifolds and weighted currents. When using the last two infinite-dimensional spaces,

we need to project both the subject shapes and the template to a finite-dimensional space where probability density functions can be computed. We assumed that the noise of each structure j follows a Gaussian distribution whose covariance matrix has the size of the finite-dimensional approximation space for weighted currents and varifolds and it is equal to the number of point-correspondences when working with landmarks. In both cases we define it as Λ_j.

Using ω_j=$0.01\Lambda_j N$ and P_j=$\frac{0.05R_j^0}{\omega_j}$ with R_j^0 equal to the initial residual of structure j, as proposed in Sec.2.3.2, we can notice that the data-terms of the cost function of the atlas construction (Eq.4.20) can be approximated by:

$$\approx \sum_{j=1}^{M} \frac{N\Lambda_j}{2} \log\left(\frac{R_j + 0.05R_j^0}{N\Lambda_j}\right) \tag{4.31}$$

after removing the constant terms, with R_j equal to the residual of structure j and where N and M are the number of subjects and structures respectively. This shows that the data-term of each structure j depends on the metric of the computational model chosen and on the value of Λ_j. If all structures are modelled with the same computational model, like in Chapter 2, the data-terms are commensurable since we use the same rule to compute both R_j and Λ_j but if the structures are modelled with different computational models, like in this chapter, nothing guarantees that the data-terms are commensurable. Depending on the number of landmarks chosen or on the size of the grid where a varifold or weighted current is projected to, a structure could be over-weighted with respect to the others (or the contrary). In this chapter, we used some heuristics to find a good balance between the different data-terms which gave reasonable results but future works will have to find a way to "harmonize" the choice of Λ_j among the different computational models.

In the proposed method, we assumed that the initial momenta of the two diffeomorphisms are independent, that is to say that $p(\boldsymbol{\alpha}_i^{All}, \boldsymbol{\alpha}_i^W) = p(\boldsymbol{\alpha}_i^{All})p(\boldsymbol{\alpha}_i^W)$, even if the update rule for $\boldsymbol{\alpha}_i^{All}$ and $\boldsymbol{\alpha}_i^W$ are related as explained in Sec.4.2.5. It would seem more reasonable to take that into account by modelling directly $p(\boldsymbol{\alpha}_i^{All}, \boldsymbol{\alpha}_i^W)$ without the assumption of independence. We could model, for instance, their joint distribution as a single Gaussian distribution. However, the statistical relationship between $\boldsymbol{\alpha}_i^{All}$ and $\boldsymbol{\alpha}_i^W$ is highly complex since they are related by the linearised ODEs shown in Sec.4.2.5 and we have not found yet a satisfactory solution to model their joint distribution. This is left as future work.

Nevertheless, we demonstrated that the proposed double diffeomorphic approach can detect some useful and relevant information since it leads to better classification scores than a single diffeomorphism. This means that the information about structural connectivity might play an important role in the characterisation of GTS and that the proposed method can bring an important contribution in the description of the pathophysiological mechanisms underlying GTS.

4.A Appendix

4.A.1 Data-term and gradient of weighted prototypes

Given two sets of weighted prototypes: $\sum_{i=1}^{N} \tau_i F_i$ and $\sum_{p=1}^{M} w_p T_p$ where $F_i = \sum_{n=1}^{L_i} \delta^{\alpha_{in}}_{(x_{in},f_i,t_i)}$ and $T_p = \sum_{l=1}^{Z_p} \delta^{\beta_{pl}}_{(y_{pl},v_p,u_p)}$ we first present how to compute the squared norm of their difference $||\sum_{i=1}^{N} \tau_i F_i - \sum_{p=1}^{M} w_p T_p||^2_{Q^*}$ and then its gradient with respect to every point of the fiber $F_k = \sum_{n=1}^{L_k} \delta^{\alpha_{kn}}_{(x_{kn},f_k,t_k)}$ where x_{kn} is the center of segment n of fiber k and α_{kn} is its tangent vector. They are defined as $x_{kn} = (p_{kn} + m_{kn})/2$ and $\alpha_{kn} = p_{kn} - m_{kn}$, where p_{kn} and m_{kn} are the first and the last point of segment n of fiber k respectively. This means that every couple of neighbour segments n and $n+1$ has one point in common: $p_n = m_{n+1}$ (except for the extremities). We assume also that all fibers are oriented from the extremity f_i (v_p) towards t_i (u_p) and we call k_a the kernel related to the first extremity (f_i) and k_b the kernel for the second extremity (t_i). Since weighted currents are defined with respect to the center and tangent vectors, we need to use the chain rule. Moreover, every point (except for the extremities) depends on two segments since it can be seen as p_n but also as m_{n+1} which means that its update rule will depend on the center and tangent vectors of both segments n and $n+1$.

The data-term is defined as:

$$||\sum_{i=1}^{N} \tau_i F_i - \sum_{p=1}^{M} w_p T_p||^2_{Q^*} = \underbrace{||\sum_{i=1}^{N} \tau_i F_i||^2_{Q^*}}_{\mathcal{A}} + \underbrace{||\sum_{p=1}^{M} w_p T_p||^2_{Q^*}}_{\mathcal{B}} \underbrace{-2 \left\langle \sum_{i=1}^{N} \tau_i F_i, \sum_{p=1}^{M} w_p T_p \right\rangle_{Q^*}}_{\mathcal{C}}$$

$$\mathcal{A} : \sum_{i=1}^{N}\sum_{j=1}^{N} \tau_i \tau_j k_a(f_i, f_j) k_b(t_i, t_j) \langle F_i, F_j \rangle_{U^*}$$

$$\mathcal{B} : \sum_{p=1}^{M}\sum_{d=1}^{M} w_p w_d k_a(v_p, v_d) k_b(u_p, u_d) \langle T_p, T_d \rangle_{U^*}$$

$$\mathcal{C} : \sum_{i=1}^{N}\sum_{p=1}^{M} \tau_i w_p k_a(f_i, v_p) k_b(t_i, u_p) \langle F_i, T_p \rangle_{U^*} \tag{4.32}$$

where U^* is the space of usual currents, namely $\langle F_i, F_j \rangle_{U^*} = \left(\sum_{n=1}^{L_i}\sum_{m=1}^{L_j} \alpha_{in}^T k_g(x_{in}, x_{jm}) \alpha_{jm}\right)$, $\langle T_p, T_d \rangle_{U^*} = \left(\sum_{l=1}^{Z_p}\sum_{o=1}^{Z_d} \beta_{pl}^T k_g(y_{pl}, y_{do}) \beta_{do}\right)$, $\langle F_i, T_p \rangle_{U^*} = \left(\sum_{n=1}^{L_i}\sum_{l=1}^{Z_p} \alpha_{in}^T k_g(x_{in}, y_{pl}) \beta_{pl}\right)$.

The derivative with respect to the point p_{ks} (resp. m_{ks}), that is $\frac{\partial||\sum_{i=1}^{N} \tau_i F_i - \sum_{p=1}^{M} w_p T_p||^2_{Q^*}}{\partial p_{ks}}$ (resp. $\frac{\partial||\cdot||^2_{Q^*}}{\partial m_{ks}}$), if it is not the extremity ($p_{ks} \neq t_k$, resp.

$m_{ks} \neq f_k$), results:

$$2\tau_k \sum_{i=1}^{N} \tau_i k_a(f_i, f_k) k_b(t_i, t_k) \sum_{n=1}^{L_i} \left[\pm \alpha_{in} \exp(-\frac{||x_{ks} - x_{in}||^2}{\lambda_g^2}) - \alpha_{ks}^T \alpha_{in} (\frac{x_{ks} - x_{in}}{\lambda_g^2}) \exp(-\frac{||x_{ks} - x_{in}||^2}{\lambda_g^2}) \right] -$$
$$2\tau_k \sum_{p=1}^{M} w_p k_a(v_p, f_k) k_b(u_p, t_k) \sum_{l=1}^{Z_p} \left[\pm \beta_{pl} \exp(-\frac{||x_{ks} - y_{pl}||^2}{\lambda_g^2}) - \alpha_{ks}^T \beta_{pl} (\frac{x_{ks} - y_{pl}}{\lambda_g^2}) \exp(-\frac{||x_{ks} - y_{pl}||^2}{\lambda_g^2}) \right] \tag{4.33}$$

If $p_{ks} = t_k$ one needs to add the following to the previous result:

$$2\tau_k \sum_{\substack{i=1 \\ i \neq k}}^{N} \left[\tau_i k_a(f_i, f_k) (\frac{-2(t_k - t_i)}{\lambda_b^2}) k_b(t_i, t_k) \langle F_k, F_i \rangle_{U^*} \right] -$$
$$2\tau_k \sum_{p=1}^{M} \left[w_p k_a(v_p, f_k) (\frac{-2(t_k - u_p)}{\lambda_b^2}) k_b(t_k, u_p) \langle F_k, T_p \rangle_{U^*} \right]$$

Instead when computing the derivative with respect to m_{ks} and $m_{ks} = f_k$ one needs to add:

$$2\tau_k \sum_{\substack{i=1 \\ i \neq k}}^{N} \left[\tau_i k_a(f_i, f_k) (\frac{-2(f_k - f_i)}{\lambda_a^2}) k_b(t_i, t_k) \langle F_k, F_i \rangle_{U^*} \right] -$$
$$2\tau_k \sum_{p=1}^{M} \left[w_p k_a(v_p, f_k) (\frac{-2(f_k - v_p)}{\lambda_a^2}) k_b(t_k, u_p) \langle F_k, T_p \rangle_{U^*} \right]$$

If $p_{ks} \neq t_k$, the update rule of the point p_{ks} is thus equal to $\frac{\partial ||\sum_{i=1}^{N} \tau_i F_i - \sum_{p=1}^{M} w_p T_p||_{Q^*}^2}{\partial p_{ks}} + \frac{\partial ||\sum_{i=1}^{N} \tau_i F_i - \sum_{p=1}^{M} w_p T_p||_{Q^*}^2}{\partial m_{ks+1}}$ otherwise if $p_{ks} = t_k$ it is simply $\frac{\partial ||\sum_{i=1}^{N} \tau_i F_i - \sum_{p=1}^{M} w_p T_p||_{Q^*}^2}{\partial p_{ks}}$. The update rule of the first point $m_{ks} = f_k$ is instead equal to $\frac{\partial ||\sum_{i=1}^{N} \tau_i F_i - \sum_{p=1}^{M} w_p T_p||_{Q^*}^2}{\partial m_{ks}}$.

4.A.2 Gradient of the Atlas construction procedure

We compute here the gradients of the atlas construction as explained in Sec.4.2.5. A variation of the initial control points and momenta of the first diffeomorphism $\delta \boldsymbol{L}_{i0}^W$ produces a variation in their evolution $\delta \boldsymbol{L}_i^W(t)$ and consequently a variation of the path of the flow of the first diffeomorphism $\delta \boldsymbol{T}_i^W(t)$. In parallel, $\delta \boldsymbol{L}_{i0}^{All}$ produces a variation in $\delta \boldsymbol{L}_i^{All}(t)$ which , together with $\delta \boldsymbol{T}_{i1}^W$, induces a variation in the path of the flow of the second diffeomorphism $\delta \boldsymbol{T}_i^{All}(t)$ and consequently also in the criterion δE. Here we consider only the variations related to the data term $D_i[\boldsymbol{T}_{i1}^{All}]$. The

variations of the regularity terms can be found in Sec.4.2.5. It results:

$$\delta E = \sum_{i=1}^{N} \nabla_{\boldsymbol{T}_{i1}^{All,G}}(D_i[\boldsymbol{T}_{i1}^{All,G}])^T \delta \boldsymbol{T}_{i1}^{All,G} + \nabla_{\boldsymbol{T}_{i1}^{All,W}}(D_i[\boldsymbol{T}_{i1}^{All,W}])^T \delta \boldsymbol{T}_{i1}^{All,W}$$

$$\begin{aligned}
&\delta \dot{\boldsymbol{L}}_i^{All}(t) = (d_{\boldsymbol{L}_{it}^{All}} F_i^{All}(t))^T \delta \boldsymbol{L}_{it}^{All} && \delta \boldsymbol{L}_i^{All}(0) = \delta \boldsymbol{L}_{i0}^{All} \\
&\delta \dot{\boldsymbol{L}}_i^{W}(t) = (d_{\boldsymbol{L}_{it}^{W}} F_i^{W}(t))^T \delta \boldsymbol{L}_{it}^{W} && \delta \boldsymbol{L}_i^{W}(0) = \delta \boldsymbol{L}_{i0}^{W} \\
&\delta \dot{\boldsymbol{T}}_i^{W}(t) = (\partial_{\boldsymbol{T}_{it}^{W}} G_i^{W}(t))^T \delta \boldsymbol{T}_{it}^{W} + (\partial_{\boldsymbol{L}_{it}^{W}} G_i^{W}(t))^T \delta \boldsymbol{L}_{it}^{W} && \delta \boldsymbol{T}^{W}(0) = \delta \boldsymbol{T}_0^{W} \\
&\delta \dot{\boldsymbol{T}}_i^{All}(t) = (\partial_{\boldsymbol{T}_{it}^{All}} G_i^{All}(t))^T \delta \boldsymbol{T}_{it}^{All} + (\partial_{\boldsymbol{L}_{it}^{All}} G_i^{All}(t)])^T \delta \boldsymbol{L}_{it}^{All} && \delta \boldsymbol{T}_i^{All}(0) = \delta \boldsymbol{T}_{i1}^{W} \cup \delta \boldsymbol{T}_0^{G}
\end{aligned}$$

The two first ODEs are linear and their solution is given by:

$$\delta \boldsymbol{L}(t) = \exp\left(\int_0^t d_{\boldsymbol{L}(u)} F(u) du\right) \delta \boldsymbol{L}_0 = R_{0t} \delta \boldsymbol{L}_0 \tag{4.34}$$

The last two ODEs are instead linear with source term and their solution is given by:

$$\begin{aligned}
\delta \boldsymbol{T}(t) &= \int_0^t \exp\left(\int_u^t \partial_{\boldsymbol{T}(u)} G(u) du\right) \partial_{\boldsymbol{L}(u)} G(u) \delta \boldsymbol{L}(u) du + \exp\left(\int_0^t \partial_{\boldsymbol{T}(s)} G(s) ds\right) \delta \boldsymbol{T}_0 \\
&= \int_0^t V_{ut} \partial_{\boldsymbol{L}(u)} G(u) \delta \boldsymbol{L}(u) du + V_{0t} \delta \boldsymbol{T}_0
\end{aligned} \tag{4.35}$$

where, as in (Durrleman et al., 2014), we denote: $R_{st} = \exp(\int_s^t d_{\boldsymbol{L}(u)} F(u) du)$ and $V_{st} = \exp(\int_s^t \partial_{\boldsymbol{T}(u)} G(u) du)$. We omit the upper indices W and All and the index i for clarity purpose. Substituting Eq.4.34 into Eq.4.35 and noting that $\delta \boldsymbol{T}^{All,W}(0) = \delta \boldsymbol{T}^{W}(1)$, we obtain:

$$\delta \boldsymbol{T}^{All,W}(t) = \left(\int_0^t V_{ut}^{All} \partial_{\boldsymbol{L}^{All}(u)} G^{All}(u) R_{0u}^{All} du\right) \delta \boldsymbol{L}_0^{All} + V_{0t}^{All} \delta \boldsymbol{T}_0^{W} + \tag{4.36}$$

$$V_{0t}^{All} \int_0^1 \partial_{\boldsymbol{T}^{W}(s)} G^{W}(s) \delta \boldsymbol{T}^{W}(s) ds + V_{0t}^{All} \int_0^1 \partial_{\boldsymbol{L}^{W}(s)} G^{W}(s) \delta \boldsymbol{L}^{W}(s) ds \tag{4.37}$$

Using the Fubini's theorem we can rewrite the 3^{rd} term as:

$$V_{0t}^{All}\int_0^1 \partial_{\boldsymbol{T}^W(s)} G^W(s)\delta\boldsymbol{T}^W(s)ds = \left(V_{0t}^{All}\int_0^1 V_{u1}^W \partial_{\boldsymbol{L}^W(u)} G^W(u) R_{0u}^W du\right)\delta\boldsymbol{L}_0^W - V_{0t}^{All}\delta\boldsymbol{T}_0^W - \left(V_{0t}^{All}\int_0^1 \partial_{\boldsymbol{L}^W(u)} G^W(u) R_{0u}^W du\right)\delta\boldsymbol{L}_0^W + (V_{0t}^{All}V_{01}^W)\delta\boldsymbol{T}_0^W \tag{4.38}$$

Instead the 4^{th} term becomes:

$$V_{0t}^{All}\int_0^1 \partial_{\boldsymbol{L}^W(s)} G^W(s)\delta\boldsymbol{L}^W(s)ds = \left(V_{0t}^{All}\int_0^1 \partial_{\boldsymbol{L}^W(s)} G^W(s) R_{0s}^W ds\right)\delta\boldsymbol{L}_0^W \tag{4.39}$$

Plugging them with Eq.4.35 for $\delta\boldsymbol{T}^{All,G}(t)$ into δE we obtain:

$$\nabla_{\boldsymbol{L}_0^{All}} E = \left(\int_0^1 (R_{0u}^{All})^T (\partial_{\boldsymbol{L}^{All}} G^{All}(u))^T (V_{u1}^{All})^T du\right)\nabla_{\boldsymbol{T}^{All}(1)} D[\boldsymbol{T}^{All}(1)]$$
$$\nabla_{\boldsymbol{L}_0^W} E = \left(\int_0^1 (R_{0u}^W)^T (\partial_{\boldsymbol{L}^W} G^W(u))^T (V_{u1}^W)^T du\right)(V_{01}^{All})^T \nabla_{\boldsymbol{T}^{All,W}(1)} D[\boldsymbol{T}^{All,W}(1)]$$
$$\nabla_{\boldsymbol{T}_0^W} E = (V_{01}^W)^T (V_{01}^{All})^T \nabla_{\boldsymbol{T}^{All,W}(1)} D[\boldsymbol{T}^{All,W}(1)]$$
$$\nabla_{\boldsymbol{T}_0^G} E = (V_{01}^{All})^T \nabla_{\boldsymbol{T}^{All,G}(1)} D[\boldsymbol{T}^{All,G}(1)]$$

Calling $\theta^{All,G}(t) = (V_{t1}^{All})^T \nabla_{\boldsymbol{T}_1^{All,G}} D$ and $\theta_t^{All,W} = (V_{t1}^{All})^T \nabla_{\boldsymbol{T}_1^{All,W}} D$ which together form $\theta^{All}(t) = \{\theta^{All,G}(t),\ \theta^{All,W}(t)\}$ and $\xi^{All}(t) = \int_t^1 (R_{tu}^{All})^T (\partial_{\boldsymbol{L}^{All}} G^{All}(u))^T \theta^{All}(u)du$ for the second diffeomorphism and $\theta^W(t) = (V_{t1}^W)^T \theta_0^{All,W}$ with $\xi^W(t) = \int_t^1 (R_{tu}^W)^T\ (\partial_{\boldsymbol{L}^W} G^W(u))^T \theta^W(u)du$ for the first diffeomorphism, we obtain the results in Eq.4.26.

Chapter 5

Conclusions and Perspectives

Contents

5.1 Conclusive summary

In this thesis, we proposed a new approach to *localise*, *characterise* and *visualise* the anatomical abnormalities associated to a neurodevelopental disorder and affecting both the structural organisation and shape of neural circuits. This methodology is based on the Grenander's atlas construction procedure which permits to analyse the neural circuits as a single interwoven complex where every component is virtually represented as a 3D mesh. This makes possible the integration of white matter fiber bundles and grey matter structures in a single framework. Moreover, the use of meshes facilitates the visualisation of the complex and the interpretation of its pathological alterations.

In Chapter 2 we proposed a Bayesian framework in which embed the generative model of the atlas construction procedure. It allows to automatically estimate the noise variance of each component of the neural circuits which should otherwise be fixed by the user. We showed that these values can highly influence the results and their automatic estimate becomes even more crucial if the number of analysed structures increases. This is therefore essential to perform an holistic analysis of the entire neural circuits. Moreover, the proposed statistical setting also permits to estimate a well-conditioned covariance matrix of the deformation parameters. This makes thus possible to directly use it for statistical analysis, such as PCA, without regularising it a posteriori.

Furthermore, we proposed to model every surface of the grey matter as a Gaussian random varifold for which we define finite-dimensional approximation spaces

similarly to Durrleman (2010) where it is possible to define probability density functions. This computational model does not need point-correspondences and it has a closed-form metric which is easily derivable. Moreover, it can be used for any kind of shape, including 3D curves such as the white matter fibers. However, this framework takes into consideration only the geometry (pathway) of the fibers. This makes the template-to-subject registrations accurate only in the most dense parts of the bundles and it almost ignores the extremities, namely where the streamlines integrate grey matter structures. Thus, it hampers the study of variations in structural connectivity. In addition to that, fiber bundles resulting from tractography algorithms would require an enormous computational load and time to be directly used in the atlas construction procedure due to their huge number of streamlines.

To this end, we proposed in Chapter 3 an approximation scheme for fiber bundles. The resulting parsimonious and computationally-efficient representation makes feasible the integration of fiber bundles in the atlas procedure together with the other components of the neural circuits. Moreover, it is visualised as a scattered set of tubes whose radii are proportional to the number of streamlines approximated. The smaller number of tubes, with respect to the original amount of streamlines, eases the visualisation and interpretation of the bundles since it allows to clearly identify and analyse all its fascicles. Moreover, it drastically reduces the computational load for visualisation software thus accelerating any kind of manipulation. The proposed approximation is based on the computational model of weighted currents, an adaptation of the one of functional currents (Charon and Trouvé, 2013a). Similarly to currents, it does not need point correspondences or streamlines correspondences and it augments its definition taking into consideration not only the pathway of the streamlines but also the locations of their extremities. This makes thus possible to correctly register also the extremities of the bundles in the template-to-subjects deformations and, theoretically, to capture the variations in structural connectivity.

However, the deformations of the template complex in standard atlas construction procedures are defined as single diffeomorphisms of the entire ambient space. This precludes the study of variations in topology such as the changes in structural connectivity. To this end, we proposed in Chapter 4 a new diffeomorphic scheme based on a cascade of two diffeomorphisms. This permits to model variations in structural connectivity with the first diffeomorphism and global anatomical changes with the second one. The parameters of each diffeomorphism are assumed to follow a different and independent Gaussian distribution centred at zero. A double diffeomorphic atlas construction embedded in the Bayesian framework presented in Chapter 2 results in a template complex, showing the common anatomical characteristics of the population under analysis, and two well-conditioned covariance matrices, describing the organisational and morphological variability captured by the double diffeomorphisms.

The proposed method describes the organisation and morphology of a group of anatomical complexes composed of both grey matter surfaces and white matter tracts. We showed its effectiveness by comparing the cortico-putamen circuits of a group of patients subject to Gilles de la Tourette syndrome and a control group.

Preliminary results highlight abnormal differences about both the shape of the grey matter structures and the structural connectivity. These alterations are in line with the hypothesis put forth in Worbe et al. (2010, 2015). Moreover, we also showed that the proposed approach leads to better classification scores than the single diffeomorphic method described in Chapter 2. This suggests that it might better characterise the anatomical alterations associated to GTS and therefore that it could bring an important contribution in the description of the pathophysiological mechanisms underlying GTS.

The algorithms presented in this thesis have been coded mostly in C++ and some parts also in CUDA. The code of the proposed Bayesian framework (Chapter 2) and of the double diffeomorphic deformation (Chapter 4) will be integrated to the software deformetrica `www.deformetrica.org`.

5.2 Limitations and Perspectives

The proposed methodology offers a mathematical tool which can help understanding the pathophysiology of neurodevelopmental syndromes. This opens a wide spectrum of possible future clinical applications such as autism and attention deficit-hyperactivity disorders.

Moreover, to fully characterise the vast clinical heterogeneity and complexity of neurodevelopmental disorders we still need to extend the proposed method. This offers a challenging opportunity for further methodological improvements. In the following, we present perspectives and possible solutions to some of the limitations highlighted in the previous chapters.

5.2.1 Mixture of templates

A possible extension would be to use a mixture of templates instead than a single one as proposed in Allassonnière et al. (2007). Statistical learning techniques could be used to automatically estimate the number of templates. Each template would characterise a sub-group of the population. This could then be used to cluster the patients in smaller groups which would be (hopefully) clinically more homogeneous. The analysis of the templates and their variability could help understanding whether these sub-groups share anatomical alterations, distinctive therefore of the disorder, or if they are characterised by specific abnormalities. Eventually, this could help clarifying the pathophysiological mechanisms related to the different symptoms.

5.2.2 Cortical surface

In this thesis, we modelled the cortical surface with landmarks which is a computationally convenient solution but it is based on a strong hypothesis of point-correspondence which might not be correct. A different strategy would be modelling it as a varifold with explicit sulcal constraints given, for instance, by the 120

BrainVisa sulci (Perrot et al., 2011; Auzias et al., 2011), the 56 major sulci of MNI Display (Holmes et al., 1996; Joshi et al., 2012) or the so called "sulcal pits", namely the deepest points of the folding pattern (Auzias et al., 2015). In this way, we would account for the cortical shape variation and at the same time we would be able to align the cortical folding architecture.

However, we should also take into consideration the fact that patients with Gilles de la Tourette syndrome might have an abnormal structural pattern of cortical sulci, as it was demonstrated in (Muellner et al., 2015).

5.2.3 Integrating functional information

Another possible extension would be to perform an even more holistic approach by adding a functional signal to the anatomical complexes. This could be easily taken into account by augmenting the framework of varifolds and weighted currents as in functional currents (Charon and Trouvé, 2013a). The functional signal could be represented, for instance, by functional MRI or EEG. The functional information would be directly taken into account by augmenting the dimensionality of the space where shapes lie. It would change from a shape space to a functional-shape space where the deformation of the template complex towards a subject complex would be driven by a similarity measure which considers both shape and functional signal. This approach would be in contrast with most of the works present in the literature which either study functional and structural imaging separately or use the latter to restrain the functional analysis to specific areas (Schumann et al., 2011; Vaidya, 2012; Singh et al., 2015).

The framework of functional currents could also be used to augment the approximation scheme proposed in Chapter 3 with information about the white matter microstructure. Every streamline could be characterised by a vector of scalar quantities, such as FA or GFA, describing the white matter structural integrity of the voxels through which the fiber has been traced. A similarity measure between two streamlines would then take into account their pathways, connectivity and also the local white matter microstructure. This could be combined with a new deformation setting always in the spirit of the Grenander's approach like in Charlier et al. (2014). The differences between two streamlines would always be described by warping the first one onto the other. However, in this case the deformation would change not only the pathway of the source streamline but also the scalar values attached to it such as FA or GFA. This would permit to discover differences related to both the geometry of the streamlines and to their microstructure.

5.2.4 Topological changes in fiber bundles

The double diffeomorphic scheme presented in Chapter 4 can not take into account all the topological variations of the white matter. For instance, it can not put into correspondence circuits with *intersecting* or *shrinking* fibers. In Fig.5.1 we show on the left a template with parallel streamlines connecting homogeneously

two toy grey matter structures and on the right a subject complex with *intersecting streamlines* connecting only specific grey matter areas. The only way to attain a correct matching would be to use a diffeomorphism for each streamline. Instead, in Fig.5.2 we show an example of *shrinking fibers*. This happens when the fibers of the template, shown on the left, do not have correspondent fibers on the subject bundle, shown on the right, within an area equal to the diffeomorphic kernel bandwidth. This produces a data-term which depends entirely on the norm of the template and therefore the only way to decrease it is shrinking the template fibers.

The first problem would happen mainly when considering several fiber bundles at the same time, like all the bundles of the CSTC circuits, instead the second one when there are important differences in structural connectivity among the population under study or between the chosen initial template and the subjects bundles.

We envisage here two possible solutions to solve these problems. The first one would be to use a different deformation framework: the one of metamorphosis (Trouvé and Younes, 2005; Holm et al., 2009; Younes and Richardson, 2013). It combines a diffeomorphic mapping with a topological change of the template. This would permit to make disappear the shrinking fibers from the template or to change the arrangement of its streamlines in order to take into account the *intersecting streamlines*. Moreover, in this framework it is possible to compute population statistics based on the initial momenta parametrising the deformations, as shown in Richardson and Younes (2015).

A second (and more radical) solution would be to discard the pathway of the streamlines keeping only the information about the structural connectivity. Given two ROIs (Region Of Interest), which could be the entire putamen and the entire left hemisphere of the cortex as in Chapter 4, we would compute a probability map at each vertex of the first ROI indicating the probability to be linked with any vertex of the second ROI. This probability map would describe the structural connectivity. Then we could use functional currents (or functional varifolds) (Charon and Trouvé, 2013a) to model the grey matter surfaces where the functional signal would be the probability map. Eventually, one could use a kind of double deformation to modify first the probability maps and then both the shape of the structures and their functional signal. This would avoid all topological problems about fiber bundles but it would discard the information about the pathway of the streamlines, which might be important, and it would depend on the choice of the ROIs.

5.2.5 Clinical variables

A last perspective is about the integration of clinical variables in the proposed framework. A clinical variable is a scalar value which quantifies a specific characteristic of a pathology affecting a patient like the severity of the symptoms. An example is the scalar YTGTSS/50 which quantifies tic severity.

Clinical variables could be used in different ways. We could estimate the PCA modes that best explain the clinical variables with linear regression analysis and model reduction methods such as Akaike information criterion (AIC). Furthermore,

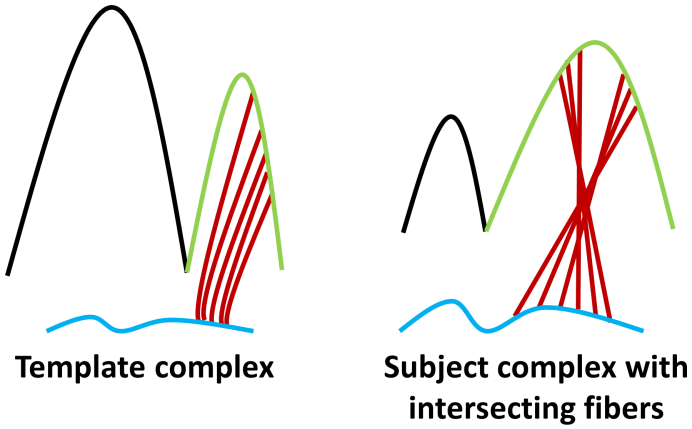

Figure 5.1: Example of subject complex with intersecting fibers.

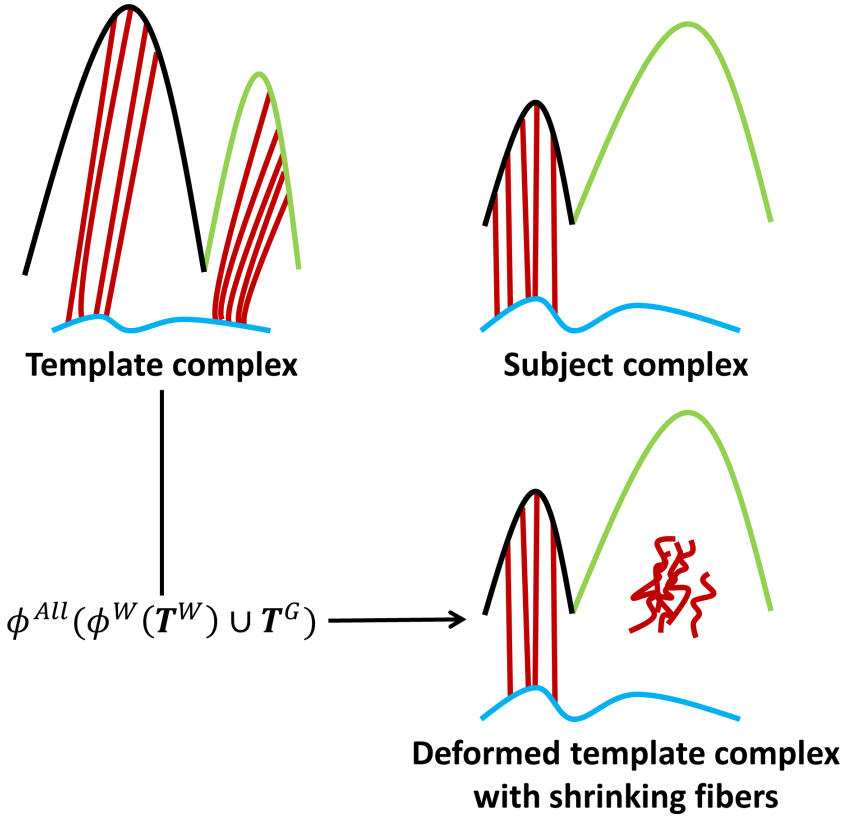

Figure 5.2: Example of deformed template complex with shrinking fibers.

we could also use Partial Least Squares (PLS) regression to automatically extract the modes that are relevant to both the shape variability and the clinical variables. Such an approach may be useful to investigate whether an increase in the severity of tics is correlated with more pronounced atypical connectivity patterns or shape alterations in grey matter structures.

Another interesting extension would be to estimate the relationship between the

template and the clinical variables, which means understanding how the template changes as the clinical variable varies. A similar approach was proposed for the quantification and prediction of cardiac remodelling in Mansi et al. (2011).

Eventually, we could also integrate shape and clinical information into a single framework by using manifold learning techniques (Aljabar et al., 2011). Pairwise distances between couples of subjects could be defined by the amount of deformation needed to warp the first one onto the second one, i.e. the norm of the initial velocity field parametrising the deformation. In our case, we could use the two diffeomorphisms together or separately. These distances could then be used as input of a manifold learning algorithm *together* with the clinical information like in Wolz et al. (2011) or Fiot et al. (2012). In the resulting low-dimensional representation of the data we could then use classification algorithms such as Support Vector Machine (SVM).

Bibliography

Alexander, G. E., DeLong, M. R., and Strick, P. L. (1986). Parallel Organization of Functionally Segregated Circuits Linking Basal Ganglia and Cortex. *Annual Review of Neuroscience*, 9(1):357–381. 1

Aljabar, P., Wolz, R., Srinivasan, L., Counsell, S., Rutherford, M., Edwards, A., Hajnal, J., and Rueckert, D. (2011). A Combined Manifold Learning Analysis of Shape and Appearance to Characterize Neonatal Brain Development. *IEEE Transactions on Medical Imaging*, 30(12):2072–2086. 125

Allassonnière, S., Amit, Y., and Trouvé, A. (2007). Towards a Coherent Statistical Framework for Dense Deformable Template Estimation. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 69(1):3–29. 11, 16, 23, 24, 25, 26, 27, 77, 78, 80, 112, 121

Allassonnière, S., Durrleman, S., and Kuhn, E. (2015). Bayesian Mixed Effect Atlas Estimation with a Diffeomorphic Deformation Model. *SIAM Journal on Imaging Sciences*, pages 1367–1395. 112

Allassonnière, S., Kuhn, E., and Trouvé, A. (2010). Construction of Bayesian deformable models via stochastic approximation algorithm: a convergence study. *Bernoulli*, 16(3):641–678. 26

Arguillère, S., Trélat, E., Trouvé, A., and Younes, L. (2015). Multiple Shape Registration using Constrained Optimal Control. *arXiv:1503.00758 [math]*. arXiv: 1503.00758. 76

Arsigny, V., Commowick, O., Pennec, X., and Ayache, N. (2006). A Log-Euclidean Polyaffine Framework for Locally Rigid or Affine Registration. In Pluim, J. P. W., Likar, B., and Gerritsen, F. A., editors, *Biomedical Image Registration*, number 4057 in Lecture Notes in Computer Science, pages 120–127. Springer Berlin Heidelberg. DOI: 10.1007/11784012_15. 12

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1):95–113. 13

Ashburner, J. and Friston, K. J. (2000). Voxel-Based Morphometry—The Methods. *NeuroImage*, 11(6):805–821. 5

Ashburner, J. and Friston, K. J. (2011). Diffeomorphic registration using geodesic shooting and Gauss–Newton optimisation. *NeuroImage*, 55(3):954–967. 12

Auzias, G., Brun, L., Deruelle, C., and Coulon, O. (2015). Deep sulcal landmarks: Algorithmic and conceptual improvements in the definition and extraction of sulcal pits. *NeuroImage*, 111:12–25. 122

Auzias, G., Colliot, O., Glaunes, J., Perrot, M., Mangin, J.-F., Trouvé, A., and Baillet, S. (2011). Diffeomorphic Brain Registration Under Exhaustive Sulcal Constraints. *IEEE Transactions on Medical Imaging*, 30(6):1214–1227. 22, 122

Avants, B. and Gee, J. C. (2004). Geodesic estimation for large deformation anatomical shape averaging and interpolation. *NeuroImage*, 23, Supplement 1:S139–S150. 11, 23, 27, 48

Avants, B. B., Cook, P. A., Ungar, L., Gee, J. C., and Grossman, M. (2010). Dementia induces correlated reductions in white matter integrity and cortical thickness: A multivariate neuroimaging study with sparse canonical correlation analysis. *NeuroImage*, 50(3):1004–1016. 7

Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C. (2008). Symmetric Diffeomorphic Image Registration with Cross-Correlation: Evaluating Automated Labeling of Elderly and Neurodegenerative Brain. *Medical image analysis*, 12(1):26–41. 12

Basser, P. J., Mattiello, J., and LeBihan, D. (1994). Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance. Series B*, 103(3):247–254. 54

Batchelor, P. G., Calamante, F., Tournier, J.-D., Atkinson, D., Hill, D. L. G., and Connelly, A. (2006). Quantification of the shape of fiber tracts. *Magnetic Resonance in Medicine*, 55(4):894–903. 9, 55

Bauer, M. and Bruveris, M. (2011). A New Riemannian Setting for Surface Registration. In *Proceedings of the Third International MICCAI Workshop on Mathematical Foundations of Computational Anatomy*, pages 182–193. 11

Bauer, M., Bruveris, M., Marsland, S., and Michor, P. W. (2014). Constructing reparameterization invariant metrics on spaces of plane curves. *Differential Geometry and its Applications*, 34:139–165. 11

Beg, M. F., Miller, M. I., Trouvé, A., and Younes, L. (2005). Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms. *International Journal of Computer Vision*, 61(2):139–157. 12, 89

Belger, A., Carpenter, K. L. H., Yucel, G. H., Cleary, K. M., and Donkers, F. C. L. (2011). The neural circuitry of autism. *Neurotoxicity Research*, 20(3):201–214. 1

Bhatia, K., Hajnal, J., Puri, B., Edwards, A., and Rueckert, D. (2004). Consistent groupwise non-rigid registration for atlas construction. In *IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2004*, volume 1, pages 908–911. 11

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008. 61, 72

Bookstein, F. L. (1997). Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Medical Image Analysis*, 1(3):225–243. 9

Bossa, M., Zacur, E., and Olmos, S. (2011). Statistical analysis of relative pose information of subcortical nuclei: Application on ADNI data. *NeuroImage*, 55(3):999–1008. 22

Brandes, U., Delling, D., Gaertler, M., Goerke, R., Hoefer, M., Nikoloski, Z., and Wagner, D. (2006). Maximizing Modularity is hard. *arXiv:physics/0608255*. arXiv: physics/0608255. 61

Brechbühler, C., Gerig, G., and Kübler, O. (1995). Parametrization of Closed Surfaces for 3-D Shape Description. *Computer Vision and Image Understanding*, 61(2):154–170. 9

Brett, A. D. and Taylor, C. J. (1998). A Method of Automated Landmark Generation for Automated 3d PDM Construction. pages 91.1–91.10. British Machine Vision Association. 9

Bürgel, U., Amunts, K., Hoemke, L., Mohlberg, H., Gilsbach, J. M., and Zilles, K. (2006). White matter fiber tracts of the human brain: Three-dimensional mapping at microscopic resolution, topography and intersubject variability. *NeuroImage*, 29(4):1092–1105. 56

Brun, A., Park, H.-J., Knutsson, H., and Westin, C.-F. (2003). Coloring of DT-MRI Fiber Traces Using Laplacian Eigenmaps. In Moreno-Díaz, R. and Pichler, F., editors, *Computer Aided Systems Theory - EUROCAST 2003*, number 2809 in Lecture Notes in Computer Science, pages 518–529. Springer Berlin Heidelberg. 55

Bullmore, E. and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198. 14, 75

Cates, J., Fletcher, P. T., Styner, M., Hazlett, H. C., and Whitaker, R. (2008). Particle-Based Shape Analysis of Multi-object Complexes. In Metaxas, D., Axel, L., Fichtinger, G., and Székely, G., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008*, number 5241 in Lecture Notes in Computer Science, pages 477–485. Springer Berlin Heidelberg. 9, 22

Charlier, B., Charon, N., and Trouvé, A. (2014). The fshape framework for the variability analysis of functional shapes. *arXiv:1404.6039 [cs, math]*. arXiv: 1404.6039. 28, 82, 122

Charon, N. and Trouvé, A. (2013a). Functional Currents: A New Mathematical Tool to Model and Analyse Functional Shapes. *Journal of Mathematical Imaging and Vision*, 48(3):413–431. 16, 57, 120, 122, 123

Charon, N. and Trouvé, A. (2013b). The varifold representation of non-oriented shapes for diffeomorphic registration. *SIAM Journal on Imaging Sciences*, 6(4):2547–2580. 9, 23, 27, 28, 29, 77, 78, 83

Cheng, B., Braass, H., Ganos, C., Treszl, A., Biermann-Ruben, K., Hummel, F. C., Müller-Vahl, K., Schnitzler, A., Gerloff, C., Münchau, A., and Thomalla, G. (2014). Altered intrahemispheric structural connectivity in Gilles de la Tourette syndrome. *NeuroImage: Clinical*, 4:174–181. 14

Christensen, G., Rabbitt, R., and Miller, M. (1996). Deformable templates using large deformation kinematics. *IEEE Transactions on Image Processing*, 5(10):1435–1447. 11, 13

Chung, M. K., Adluru, N., Lee, J. E., Lazar, M., Lainhart, J. E., and Alexander, A. L. (2010). Cosine series representation of 3d curves and its application to white matter fiber bundles in diffusion tensor imaging. *Statistics and its interface*, 3(1):69–80. 55

Ciccarelli, O., Catani, M., Johansen-Berg, H., Clark, C., and Thompson, A. (2008). Diffusion-based tractography in neurological disorders: concepts, applications, and future developments. *The Lancet Neurology*, 7(8):715–727. 54

Cohen-Adad, J., Descoteaux, M., Rossignol, S., Hoge, R. D., Deriche, R., and Benali, H. (2008). Detection of multiple pathways in the spinal cord using q-ball imaging. *NeuroImage*, 42(2):739–749. 6

Conturo, T. E., Lori, N. F., Cull, T. S., Akbudak, E., Snyder, A. Z., Shimony, J. S., McKinstry, R. C., Burton, H., and Raichle, M. E. (1999). Tracking neuronal fiber pathways in the living human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 96(18):10422–10427. 54

Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (1995). Active Shape Models-Their Training and Application. *Computer Vision and Image Understanding*, 61(1):38–59. 9

Corouge, I., Fletcher, P. T., Joshi, S., Gouttard, S., and Gerig, G. (2006). Fiber tract-oriented statistics for quantitative diffusion tensor MRI analysis. *Medical Image Analysis*, 10(5):786–798. 9, 55

Corouge, I., Gouttard, S., and Gerig, G. (2004). Towards a shape model of white matter fiber bundles using diffusion tensor MRI. In *IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2004*, pages 344–347 Vol. 1. 54

Cotter, C. J. and Holm, D. D. (2006). Singular solutions, momentum maps and computational anatomy. *arXiv:nlin/0605020*. arXiv: nlin/0605020. 12

Craddock, R. C., Jbabdi, S., Yan, C.-G., Vogelstein, J. T., Castellanos, F. X., Di Martino, A., Kelly, C., Heberlein, K., Colcombe, S., and Milham, M. P. (2013). Imaging human connectomes at the macroscale. *Nature Methods*, 10(6):524–539. 75

Csernansky, J. G., Joshi, S., Wang, L., Haller, J. W., Gado, M., Miller, J. P., Grenander, U., and Miller, M. I. (1998). Hippocampal morphometry in schizophrenia by high dimensional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, 95(19):11406–11411. 11, 74

Cury, C., Glaunès, J., Chupin, M., and Colliot, O. (2015). Analysis of anatomical variability using diffeomorphic iterative centroid in patients with Alzheimer's disease. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pages 1–9. 22

D'Arcy Wentworth, T. (1917). *On growth and form*. Cambridge University Press. 11

Davatzikos, C., Vaillant, M., Resnick, S. M., Prince, J. L., Letovsky, S., and Bryan, R. N. (1996). A computerized approach for morphological analysis of the corpus callosum. *Journal of Computer Assisted Tomography*, 20(1):88–97. 5, 11, 74

Davies, R., Twining, C., Cootes, T., and Taylor, C. (2010). Building 3-D Statistical Shape Models by Direct Optimization. *IEEE Transactions on Medical Imaging*, 29(4):961–981. 22

Davies, R. H., Twining, C. J., Cootes, T. F., Waterton, J. C., and Taylor, C. J. (2002). A minimum description length approach to statistical shape modeling. *Medical Imaging, IEEE Transactions on*, 21(5):525–537. 9

Descoteaux, M., Angelino, E., Fitzgibbons, S., and Deriche, R. (2007). Regularized, fast, and robust analytical Q-ball imaging. *Magnetic Resonance in Medicine*, 58(3):497–510. 37

Descoteaux, M., Deriche, R., Knosche, T., and Anwander, A. (2009). Deterministic and Probabilistic Tractography Based on Complex Fibre Orientation Distributions. *IEEE Transactions on Medical Imaging*, 28(2):269–286. 65, 93

Dryden, I. L. and Mardia, K. V. (1998). *Statistical shape analysis*. John Wiley & Sons, Chichester. 6, 10

Dupuis, P., Grenander, U., and Miller, M. I. (1998). Variational problems on flows of diffeomorphisms for image matching. *Quarterly of applied mathematics*, 56(3):587. 12

Durrleman, S. (2010). *Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution*. PhD thesis, University of Nice-Sophia Antipolis. 9, 16, 29, 85, 86, 120

Durrleman, S., Fillard, P., Pennec, X., Trouvé, A., and Ayache, N. (2009). A statistical model of white matter fiber bundles based on currents. In *Information Processing in Medical Imaging*, pages 114–125. Springer Berlin Heidelberg. 22, 37, 78

Durrleman, S., Fillard, P., Pennec, X., Trouvé, A., and Ayache, N. (2011a). Registration, atlas estimation and variability analysis of white matter fiber bundles modeled as currents. *NeuroImage*, 55(3):1073–1090. 10, 12, 22, 24, 54, 56

Durrleman, S., Pennec, X., Trouvé, A., Ayache, N., and others (2008). A forward model to build unbiased atlases from curves and surfaces. In *2nd MICCAI Workshop on Mathematical Foundations of Computational Anatomy*, pages 68–79. 78

Durrleman, S., Prastawa, M., Charon, N., Korenberg, J. R., Joshi, S., Gerig, G., and Trouvé, A. (2014). Morphometry of anatomical shape complexes with dense deformations and sparse parameters. *NeuroImage*, 101:35–49. 11, 12, 22, 23, 24, 27, 28, 29, 31, 32, 33, 48, 57, 74, 87, 90, 111, 116

Durrleman, S., Prastawa, M., Gerig, G., and Joshi, S. (2011b). Optimal Data-Driven Sparse Parameterization of Diffeomorphisms for Population Analysis. In Székely, G. and Hahn, H. K., editors, *Information Processing in Medical Imaging*, number 6801 in Lecture Notes in Computer Science, pages 123–134. Springer Berlin Heidelberg. 12, 25, 30, 77, 86

Durrleman, S., Prastawa, M., Korenberg, J. R., Joshi, S., Trouvé, A., and Gerig, G. (2012). Topology Preserving Atlas Construction from Shape Data without Correspondence Using Sparse Parameters. In Ayache, N., Delingette, H., Golland, P., and Mori, K., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*, number 7512 in Lecture Notes in Computer Science, pages 223–230. Springer Berlin Heidelberg. 89

Fahim, C., Yoon, U., Das, S., Lyttelton, O., Chen, J., Arnaoutelis, R., Rouleau, G., Sandor, P., Frey, K., Brandner, C., and Evans, A. C. (2010). Somatosensory–motor bodily representation cortical thinning in Tourette: Effects of tic severity, age and gender. *Cortex*, 46(6):750–760. 4

Felling, R. J. and Singer, H. S. (2011). Neurobiology of Tourette Syndrome: Current Status and Need for Further Investigation. *Journal of Neuroscience*, 31(35):12387–12395. 5

Fillard, P., Descoteaux, M., Goh, A., Gouttard, S., Jeurissen, B., Malcolm, J., Ramirez-Manzanares, A., Reisert, M., Sakaie, K., Tensaouti, F., Yo, T., Mangin, J.-F., and Poupon, C. (2011). Quantitative evaluation of 10 tractography algorithms on a realistic diffusion MR phantom. *NeuroImage*, 56(1):220–234. 54

Fiot, J.-B., Fripp, J., and Cohen, L. D. (2012). Combining imaging and clinical data in manifold learning: distance-based and graph-based extensions of Laplacian eigenmaps. In *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, pages 570–573. IEEE. 125

Fischl, B., Kouwe, A. v. d., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., Busa, E., Seidman, L. J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., and Dale, A. M. (2004). Automatically Parcellating the Human. *Cerebral Cortex*, 14(1):11–22. 65, 93

Fletcher, P., Lu, C., Pizer, S., and Joshi, S. (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, 23(8):995–1005. 10

Folgoc, L. L., Delingette, H., Criminisi, A., and Ayache, N. (2014). Sparse Bayesian Registration. In Golland, P., Hata, N., Barillot, C., Hornegger, J., and Howe, R., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*, number 8673 in Lecture Notes in Computer Science, pages 235–242. Springer International Publishing. 23, 24

Frangi, A., Rueckert, D., Schnabel, J., and Niessen, W. (2002). Automatic construction of multiple-object three-dimensional statistical shape models: application to cardiac modeling. *IEEE Transactions on Medical Imaging*, 21(9):1151–1166. 9

Ganos, C., Roessner, V., and Münchau, A. (2013). The functional anatomy of Gilles de la Tourette syndrome. *Neuroscience & Biobehavioral Reviews*, 37(6):1050–1062. 3

Garyfallidis, E., Brett, M., Correia, M. M., Williams, G. B., and Nimmo-Smith, I. (2012). QuickBundles, a Method for Tractography Simplification. *Frontiers in Neuroscience*, 6. 55

Gerardin, E., Chételat, G., Chupin, M., Cuingnet, R., Desgranges, B., Kim, H.-S., Niethammer, M., Dubois, B., Lehéricy, S., Garnero, L., Eustache, F., and Colliot, O. (2009). Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. *NeuroImage*, 47(4):1476–1486. 22

Glaunès, J. (2005). *Transport par difféomorphismes de points, de mesures et de courants pour la comparaison de formes et l'anatomie numérique*. PhD thesis, Université Paris 13. 9, 30, 58

Golland, P., Grimson, W. E. L., and Kikinis, R. (1999). Statistical Shape Analysis Using Fixed Topology Skeletons: Corpus Callosum Study. In Kuba, A., Šáamal, M., and Todd-Pokropek, A., editors, *Information Processing in Medical Imaging*, number 1613 in Lecture Notes in Computer Science, pages 382–387. Springer Berlin Heidelberg. 9

Golland, P., Grimson, W. E. L., Shenton, M. E., and Kikinis, R. (2005). Detection and analysis of statistical differences in anatomical shape. *Medical Image Analysis*, 9(1):69–86. 22

Gorbunova, V., Durrleman, S., Lo, P., Pennec, X., and de Bruijne, M. (2010). Lung CT registration combining intensity, curves and surfaces. In *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 340–343. 22

Gorczowski, K., Styner, M., Jeong, J. Y., Marron, J., Piven, J., Hazlett, H., Pizer, S., and Gerig, G. (2010). Multi-Object Analysis of Volume, Pose, and Shape Using Statistical Discrimination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(4):652–661. 9, 10, 22

Gori, P., Colliot, O., Marrakchi-Kacem, L., Worbe, Y., Fallani, F. D. V., Chavez, M., Lecomte, S., Poupon, C., Hartmann, A., Ayache, N., and Durrleman, S. (2014). A Prototype Representation to Approximate White Matter Bundles with Weighted Currents. In Golland, P., Hata, N., Barillot, C., Hornegger, J., and Howe, R., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*, number 8675 in Lecture Notes in Computer Science, pages 289–296. Springer International Publishing. 18, 53

Gori, P., Colliot, O., Marrakchi-Kacem, L., Worbe, Y., Routier, A., Poupon, C., Hartmann, A., Ayache, N., and Durrleman, S. (2015a). Joint Morphometry of Fiber Tracts and Gray Matter Structures Using Double Diffeomorphisms. In Ourselin, S., Alexander, D. C., Westin, C.-F., and Cardoso, M. J., editors, *Information Processing in Medical Imaging*, number 9123 in Lecture Notes in Computer Science, pages 275–287. Springer International Publishing. 19, 73

Gori, P., Colliot, O., Marrakchi-Kacem, L., Worbe, Y., Routier, A., Poupon, C., Hartmann, A., Ayache, N., and Durrleman, S. (2015b). Unified analysis of shape and structural connectivity of neural pathways. In *Organisation for Human Brain Mapping*, Honolulu, Hawaii, United States. 19

Gori, P., Colliot, O., Worbe, Y., Marrakchi-Kacem, L., Lecomte, S., Poupon, C., Hartmann, A., Ayache, N., and Durrleman, S. (2013a). Bayesian Atlas Estimation for the Variability Analysis of Shape Complexes. In Mori, K., Sakuma, I., Sato, Y., Barillot, C., and Navab, N., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, number 8149 in Lecture Notes in Computer Science, pages 267–274. Springer Berlin Heidelberg. 18, 21, 57, 77

Gori, P., Colliot, O., Worbe, Y., Marrakchi-Kacem, L., Lecomte, S., Poupon, C., Hartmann, A., Ayache, N., and Durrleman, S. (2013b). Towards joint morphometry of white matter tracts and gray matter surfaces. In *Organisation for Human Brain Mapping*. 19

Govindan, R. M., Makki, M. I., Wilson, B. J., Behen, M. E., and Chugani, H. T. (2010). Abnormal water diffusivity in corticostriatal projections in children with Tourette syndrome. *Human Brain Mapping*, 31(11):1665–1674. 6

Graciano Fouquier, A. B., Durrleman, S., Yelnik, J., Fernández-Vidal, S., and Bardinet, E. (2014). Iconic-Geometric Nonlinear Registration of a Basal Ganglia Atlas for Deep Brain Stimulation Planning. In *2nd International MICCAI Workshop on Deep Brain Stimulation Methodological Challenges (DBSMC)*. 8, 22

Grenander, U. and Miller, M. I. (1998). Computational Anatomy: An Emerging Discipline. *Q. Appl. Math.*, LVI(4):617–694. 11

Grenandner, U. (1993). *General pattern theory - A mathematical study of regular structures*. Oxford, Clarendon Press,. 11, 74

Gris, B., Durrleman, S., and Trouvé, A. (2015). A Sub-Riemannian Modular Approach for Diffeomorphic Deformations. In Nielsen, F. and Barbaresco, F., editors, *Geometric Science of Information*, number 9389 in Lecture Notes in Computer Science, pages 39–47. Springer International Publishing. 49

Guevara, P., Poupon, C., Rivière, D., Cointepas, Y., Descoteaux, M., Thirion, B., and Mangin, J. F. (2011). Robust clustering of massive tractography datasets. *NeuroImage*, 54(3):1975–1993. 55, 65, 93

Gutman, B. A., Fletcher, P. T., Cardoso, M. J., Fleishman, G. M., Lorenzi, M., Thompson, P. M., and Ourselin, S. (2015). A Riemannian Framework for Intrinsic Comparison of Closed Genus-Zero Shapes. In Ourselin, S., Alexander, D. C., Westin, C.-F., and Cardoso, M. J., editors, *Information Processing in Medical Imaging*, number 9123 in Lecture Notes in Computer Science, pages 205–218. Springer International Publishing. 11

Ha, L. K., Krüger, J., Fletcher, P. T., Joshi, S., and Silva, C. T. (2009). Fast Parallel Unbiased Diffeomorphic Atlas Construction on Multi-graphics Processing Units. In *Proceedings of the 9th Eurographics Conference on Parallel Graphics and Visualization*, EG PGV'09, pages 41–48, Aire-la-Ville, Switzerland, Switzerland. Eurographics Association. 49

Hallett, M. (2015). Tourette Syndrome: Update. *Brain and Development*, 37(7):651–655. 3

Holm, D., Trouvé, A., and Younes, L. (2009). The Euler-Poincaré theory of metamorphosis. *Quarterly of Applied Mathematics*, 67(4):661–685. 123

Holmes, C. J., MacDonald, D., Sled, J. G., Toga, A. W., and Evans, A. C. (1996). Cortical peeling: CSF/grey/white matter boundaries visualized by nesting isosurfaces. In Höhne, K. H. and Kikinis, R., editors, *Visualization in Biomedical Computing*, number 1131 in Lecture Notes in Computer Science, pages 99–104. Springer Berlin Heidelberg. DOI: 10.1007/BFb0046942. 122

Hua, K., Zhang, J., Wakana, S., Jiang, H., Li, X., Reich, D. S., Calabresi, P. A., Pekar, J. J., van Zijl, P. C. M., and Mori, S. (2008). Tract probability maps in stereotaxic spaces: Analyses of white matter anatomy and tract-specific quantification. *NeuroImage*, 39(1):336–347. 56

Hufnagel, H., Pennec, X., Ehrhardt, J., Ayache, N., and Handels, H. (2009). Computation of a Probabilistic Statistical Shape Model in a Maximum-a-posteriori Framework:. *Methods of Information in Medicine*, 48(4):314–319. 22

Jankovic, J. and Kurlan, R. (2011). Tourette syndrome: Evolving concepts. *Movement Disorders*, 26(6):1149–1156. 3

Joshi, S., Cabeen, R., Joshi, A., Sun, B., Dinov, I., Narr, K., Toga, A., and Woods, R. (2012). Diffeomorphic Sulcal Shape Analysis on the Cortex. *IEEE Transactions on Medical Imaging*, 31(6):1195–1212. 122

Joshi, S., Klassen, E., Srivastava, A., and Jermyn, I. (2007). A Novel Representation for Riemannian Analysis of Elastic Curves in Rn. In *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR '07*, pages 1–7. 11

Joshi, S. and Miller, M. (2000). Landmark matching via large deformation diffeomorphisms. *IEEE Transactions on Image Processing*, 9(8):1357–1370. 11, 74, 89

Kataoka, Y., Kalanithi, P. S., Grantz, H., Schwartz, M. L., Saper, C., Leckman, J. F., and Vaccarino, F. M. (2010). Decreased number of parvalbumin and cholinergic interneurons in the striatum of individuals with Tourette syndrome. *The Journal of Comparative Neurology*, 518(3):277–291. 4

Kimber, T. E. (2010). An Update on Tourette Syndrome. *Current Neurology and Neuroscience Reports*, 10(4):286–291. 3

Konrad, A., Dielentheis, T. F., El Masri, D., Bayerl, M., Fehr, C., Gesierich, T., Vucurevic, G., Stoeter, P., and Winterer, G. (2010). Disturbed structural connectivity is related to inattention and impulsivity in adult attention deficit hyperactivity disorder. *The European Journal Of Neuroscience*, 31(5):912–919. 1, 74

Krack, P., Hariz, M. I., Baunez, C., Guridi, J., and Obeso, J. A. (2010). Deep brain stimulation: from neurology to psychiatry? *Trends in Neurosciences*, 33(10):474–484. 2

Kubicki, M., McCarley, R., Westin, C.-F., Park, H.-J., Maier, S., Kikinis, R., Jolesz, F. A., and Shenton, M. E. (2007). A review of diffusion tensor imaging studies in schizophrenia. *Journal of Psychiatric Research*, 41(1–2):15–30. 54

Kurtek, S., Klassen, E., Ding, Z., Jacobson, S. W., Jacobson, J. B., Avison, M., and Srivastava, A. (2011). Parameterization-Invariant Shape Comparisons of Anatomical Surfaces. *IEEE Transactions on Medical Imaging*, 30(3):849–858. 11, 22

Leckman, J. F. (2002). Tourette's syndrome. *The Lancet*, 360(9345):1577–1586. 3

Liu, M., Vemuri, B. C., and Deriche, R. (2012). Unsupervised automatic white matter fiber clustering using a Gaussian mixture model. *Proceedings / IEEE International Symposium on Biomedical Imaging: from nano to macro. IEEE International Symposium on Biomedical Imaging*, 2012(9):522–525. 55

Liu, Y., Miao, W., Wang, J., Gao, P., Yin, G., Zhang, L., Lv, C., Ji, Z., Yu, T., Sabel, B. A., He, H., and Peng, Y. (2013). Structural Abnormalities in Early Tourette Syndrome Children: A Combined Voxel-Based Morphometry and Tract-Based Spatial Statistics Study. *PLoS ONE*, 8(9):e76105. 1, 6

Lombaert, H., Grady, L., Pennec, X., Ayache, N., and Cheriet, F. (2012). Spectral Demons–Image Registration via Global Spectral Correspondence. In *Computer Vision–ECCV 2012*, pages 30–44. Springer. 13

Lombaert, H. and Peyrat, J.-M. (2013). Joint Statistics on Cardiac Shape and Fiber Architecture. In Mori, K., Sakuma, I., Sato, Y., Barillot, C., and Navab, N., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, number 8150 in Lecture Notes in Computer Science, pages 492–500. Springer Berlin Heidelberg. 22

Lorensen, W. E. and Cline, H. E. (1987). Marching Cubes: A High Resolution 3d Surface Construction Algorithm. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '87, pages 163–169, New York, NY, USA. ACM. 9

Lorenzi, M., Ayache, N., Frisoni, G. B., and Pennec, X. (2013). LCC-Demons: A robust and accurate symmetric diffeomorphic registration algorithm. *NeuroImage*, 81:470–483. 13

Lu, C., Pizer, S. M., Joshi, S., and Jeong, J.-Y. (2007). Statistical Multi-Object Shape Models. *International Journal of Computer Vision*, 75(3):387–404. 9

Ma, J., Miller, M. I., Trouvé, A., and Younes, L. (2008). Bayesian template estimation in computational anatomy. *NeuroImage*, 42(1):252–261. 25, 78

Ma, J., Miller, M. I., and Younes, L. (2010). A Bayesian Generative Model for Surface Template Estimation. *International Journal of Biomedical Imaging*, 2010. 23, 24, 48

Maddah, M., Miller, J. V., Sullivan, E. V., Pfefferbaum, A., and Rohlfing, T. (2011). Sheet-Like White Matter Fiber Tracts: Representation, Clustering, and Quantitative Analysis. In Fichtinger, G., Martel, A., and Peters, T., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*, number 6892 in Lecture Notes in Computer Science, pages 191–199. Springer Berlin Heidelberg. 56

Maddah, M., Wells, W. M., Warfield, S. K., Westin, C.-F., and Grimson, W. E. L. (2007). Probabilistic Clustering and Quantitative Analysis of White Matter Fiber Tracts. *Information Processing in Medical Imaging*, 20:372–383. 55

Makki, M. I., Govindan, R. M., Wilson, B. J., Behen, M. E., and Chugani, H. T. (2009). Altered Fronto-Striato-Thalamic Connectivity in Children with Tourette Syndrome Assessed with Diffusion Tensor MRI and Probabilistic Fiber Tracking. *Journal of Child Neurology*, 24(6):669–678. 14

Mansi, T., Voigt, I., Leonardi, B., Pennec, X., Durrleman, S., Sermesant, M., Delingette, H., Taylor, A., Boudjemline, Y., Pongiglione, G., and Ayache, N. (2011). A Statistical Model for Quantification and Prediction of Cardiac Remodelling: Application to Tetralogy of Fallot. *IEEE Transactions on Medical Imaging*, 30(9):1605–1616. 22, 125

Marsland, S. and McLachlan, R. (2007). A Hamiltonian particle method for diffeomorphic image registration. In *Information Processing in Medical Imaging*, pages 396–407. Springer. 12

McNaught, K. S. P. and Mink, J. W. (2011). Advances in understanding and treatment of Tourette syndrome. *Nature Reviews Neurology*, 7(12):667–676. 3

Michor, P. W. and Mumford, D. (2007). An overview of the Riemannian metrics on spaces of curves using the Hamiltonian approach. *Applied and Computational Harmonic Analysis*, 23(1):74–113. 11

Miller, M. I., Christensen, G. E., Amit, Y., and Grenander, U. (1993). Mathematical textbook of deformable neuroanatomies. *Proceedings of the National Academy of Sciences*, 90(24):11944–11948. 11

Miller, M. I., Trouvé, A., and Younes, L. (2006). Geodesic Shooting for Computational Anatomy. *Journal of Mathematical Imaging and Vision*, 24(2):209–228. 12

Modat, M., Daga, P., Cardoso, M., Ourselin, S., Ridgway, G., and Ashburner, J. (2012). Parametric non-rigid registration using a stationary velocity field. In

2012 IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA), pages 145–150. 13

Mori, S. and van Zijl, P. C. M. (2002). Fiber tracking: principles and strategies – a technical review. *NMR in Biomedicine*, 15(7-8):468–480. 54

Muellner, J., Delmaire, C., Valabrégue, R., Schüpbach, M., Mangin, J.-F., Vidailhet, M., Lehéricy, S., Hartmann, A., and Worbe, Y. (2015). Altered structure of cortical sulci in gilles de la Tourette syndrome: Further support for abnormal brain development. *Movement Disorders*, 30(5):655–661. 122

Mukherjee, P. (2005). Diffusion Tensor Imaging and Fiber Tractography in Acute Stroke. *Neuroimaging Clinics of North America*, 15(3):655–665. 54

Nesterov, Y. E. (1983). A method for solving the convex programming problem with convergence rate $O(1/k^2$). *Dokl. Akad. Nauk SSSR*, 269:543–547. 32

Neuner, I., Kupriyanova, Y., Stöcker, T., Huang, R., Posnansky, O., Schneider, F., Tittgemeyer, M., and Shah, N. J. (2010). White-matter abnormalities in Tourette syndrome extend beyond motor pathways. *NeuroImage*, 51(3):1184–1193. 4, 6

Niethammer, M., Reuter, M., Wolter, F.-E., Bouix, S., Peinecke, N., Koo, M.-S., and Shenton, M. E. (2007). Global Medical Shape Analysis Using the Laplace-Beltrami Spectrum. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 10(Pt 1):850–857. 22

O'Donnell, L. and Westin, C.-F. (2007). Automatic Tractography Segmentation Using a High-Dimensional White Matter Atlas. *IEEE Transactions on Medical Imaging*, 26(11):1562–1575. 54

O'Donnell, L. J., Iii, W. M. W., Golby, A. J., and Westin, C.-F. (2012). Unbiased Groupwise Registration of White Matter Tractography. In Ayache, N., Delingette, H., Golland, P., and Mori, K., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*, number 7512 in Lecture Notes in Computer Science, pages 123–130. Springer Berlin Heidelberg. 9, 55

O'Donnell, L. J., Westin, C.-F., and Golby, A. J. (2009). Tract-based morphometry for white matter group analysis. *NeuroImage*, 45(3):832–844. 22, 55

Pace, D., Aylward, S., and Niethammer, M. (2013). A Locally Adaptive Regularization Based on Anisotropic Diffusion for Deformable Image Registration of Sliding Organs. *IEEE Transactions on Medical Imaging*, 32(11):2114–2126. 76

Palit, A., Bhudia, S. K., Arvanitis, T. N., Turley, G. A., and Williams, M. A. (2015). Computational modelling of left-ventricular diastolic mechanics: Effect of fibre

orientation and right-ventricle topology. *Journal of Biomechanics*, 48(4):604–612. 22

Patenaude, B., Smith, S. M., Kennedy, D. N., and Jenkinson, M. (2011). A Bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage*, 56(3):907–922. 37, 65, 93

Pennec, X. (2006). Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154. 12

Perrin, M., Poupon, C., Cointepas, Y., Rieul, B., Golestani, N., Pallier, C., Rivière, D., Constantinesco, A., Bihan, D. L., and Mangin, J.-F. (2005). Fiber Tracking in q-Ball Fields Using Regularized Particle Trajectories. In Christensen, G. E. and Sonka, M., editors, *Information Processing in Medical Imaging*, number 3565 in Lecture Notes in Computer Science, pages 52–63. Springer Berlin Heidelberg. 37, 65, 93

Perrot, M., Rivière, D., and Mangin, J.-F. (2011). Cortical sulci recognition and spatial normalization. *Medical Image Analysis*, 15(4):529–550. 122

Peterson, B. S., Choi, H. A., Hao, X., Amat, J. A., Zhu, H., Whiteman, R., Liu, J., Xu, D., and Bansal, R. (2007). Morphologic Features of the Amygdala and Hippocampus in Children and Adults With Tourette Syndrome. *Archives of general psychiatry*, 64(11):1281–1291. 4

Peterson BS, Thomas P, Kane MJ, and et al (2003). Basal ganglia volumes in patients with gilles de la tourette syndrome. *Archives of General Psychiatry*, 60(4):415–424. 5

Pierpaoli, C. and Basser, P. J. (1996). Toward a quantitative assessment of diffusion anisotropy. *Magnetic Resonance in Medicine*, 36(6):893–906. 6

Qiu, A., Brown, T., Fischl, B., Ma, J., and Miller, M. (2010). Atlas Generation for Subcortical and Ventricular Structures With Its Applications in Shape Analysis. *IEEE Transactions on Image Processing*, 19(6):1539–1547. 11, 22, 74

Richardson, C. L. and Younes, L. (2015). Metamorphosis of images in reproducing kernel Hilbert spaces. *Advances in Computational Mathematics*. 123

Risholm, P., Janoos, F., Norton, I., Golby, A. J., and Wells III, W. M. (2013). Bayesian characterization of uncertainty in intra-subject non-rigid registration. *Medical Image Analysis*, 17(5):538–555. 24, 112

Risser, L., Vialard, F.-X., Baluwala, H. Y., and Schnabel, J. A. (2013). Piecewise-diffeomorphic image registration: Application to the motion estimation between 3d CT lung images with sliding conditions. *Medical Image Analysis*, 17(2):182–193. 76

Routier, A., Gori, P., Fouquier, A. B. G., Lecomte, S., Colliot, O., and Durrleman, S. (2014). Evaluation of morphometric descriptors of deep brain structures for the automatic classification of patients with Alzheimer's disease, mild cognitive impairment and elderly controls. In *MICCAI Workshop*, page 8. 19

Rubinov, M. and Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3):1059–1069. 14

Rueckert, D., Aljabar, P., Heckemann, R. A., Hajnal, J. V., and Hammers, A. (2006). Diffeomorphic Registration Using B-Splines. In Larsen, R., Nielsen, M., and Sporring, J., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006*, number 4191 in Lecture Notes in Computer Science, pages 702–709. Springer Berlin Heidelberg. DOI: 10.1007/11866763_86. 13

Savadjiev, P., Rathi, Y., Bouix, S., Smith, A. R., Schultz, R. T., Verma, R., and Westin, C.-F. (2014). Fusion of white and gray matter geometry: A framework for investigating brain development. *Medical Image Analysis*, 18(8):1349–1360. 7

Savadjiev, P., Strijkers, G. J., Bakermans, A. J., Piuze, E., Zucker, S. W., and Siddiqi, K. (2012). Heart wall myofibers are arranged in minimal surfaces to optimize organ function. *Proceedings of the National Academy of Sciences*, 109(24):9248–9253. 22

Schumann, C. M., Bauman, M. D., and Amaral, D. G. (2011). Abnormal structure or function of the amygdala is a common component of neurodevelopmental disorders. *Neuropsychologia*, 49(4):745–759. 122

Siless, V., Glaunès, J., Guevara, P., Mangin, J.-F., Poupon, C., Bihan, D. L., Thirion, B., and Fillard, P. (2012). Joint T1 and Brain Fiber Log-Demons Registration Using Currents to Model Geometry. In Ayache, N., Delingette, H., Golland, P., and Mori, K., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*, number 7511 in Lecture Notes in Computer Science, pages 57–65. Springer Berlin Heidelberg. 8, 22, 54

Simpson, I. J. A., Schnabel, J. A., Groves, A. R., Andersson, J. L. R., and Woolrich, M. W. (2012). Probabilistic inference of regularisation in non-rigid registration. *NeuroImage*, 59(3):2438–2451. 23, 24, 112

Singer, H. S. (2013). Motor control, habits, complex motor stereotypies, and Tourette syndrome. *Annals of the New York Academy of Sciences*, 1304(1):22–31. 3

Singh, S., Modi, S., Goyal, S., Kaur, P., Singh, N., Bhatia, T., Deshpande, S. N., and Khushu, S. (2015). Functional and structural abnormalities associated with

empathy in patients with schizophrenia: An fMRI and VBM study. *Journal of Biosciences*, 40(2):355–364. 122

Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., Watkins, K. E., Ciccarelli, O., Cader, M. Z., Matthews, P. M., and Behrens, T. E. J. (2006). Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *NeuroImage*, 31(4):1487–1505. 6

Sommer, S., Lauze, F., Nielsen, M., and Pennec, X. (2012). Sparse Multi-Scale Diffeomorphic Registration: The Kernel Bundle Framework. *Journal of Mathematical Imaging and Vision*, 46(3):292–308. 49

Sotiras, A., Davatzikos, C., and Paragios, N. (2013). Deformable Medical Image Registration: A Survey. *IEEE Transactions on Medical Imaging*, 32(7):1153–1190. 13

Sowell, E. R., Kan, E., Yoshii, J., Thompson, P. M., Bansal, R., Xu, D., Toga, A. W., and Peterson, B. S. (2008). Thinning of sensorimotor cortices in children with Tourette syndrome. *Nature Neuroscience*, 11(6):637–639. 4

Srivastava, M. S. and Yanagihara, H. (2010). Testing the equality of several covariance matrices with fewer observations than the dimension. *Journal of Multivariate Analysis*, 101(6):1319–1329. 46

Stern, J. S., Burza, S., and Robertson, M. M. (2005). Gilles de la Tourette's syndrome and its impact in the UK. *Postgraduate Medical Journal*, 81(951):12–19. 3

Székely, G., Kelemen, A., Brechbühler, C., and Gerig, G. (1996). Segmentation of 2-D and 3-D objects from MRI volume data using constrained elastic deformations of flexible Fourier contour and surface models. *Medical Image Analysis*, 1(1):19–34. 9

Thomalla, G., Siebner, H. R., Jonas, M., Baumer, T., Biermann-Ruben, K., Hummel, F., Gerloff, C., Muller-Vahl, K., Schnitzler, A., Orth, M., and Münchau, A. (2009). Structural changes in the somatosensory system correlate with tic severity in Gilles de la Tourette syndrome. *Brain*, 132(3):765–777. 6

Tropp, J. A. and Gilbert, A. C. (2007). Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit. *IEEE Transactions on Information Theory*, 53(12):4655–4666. 61

Trouvé, A. (1995). An Approach of Pattern Recognition through Infinite Dimensional Group Action. Technical report. 12

Trouvé, A. (1998). Diffeomorphisms Groups and Pattern Matching in Image Analysis. *International Journal of Computer Vision*, 28(3):213–221. 12

Trouvé, A. and Younes, L. (2005). Metamorphoses Through Lie Group Action. *Foundations of Computational Mathematics*, 5(2):173–198. 123

Tuch, D. S. (2004). Q-ball imaging. *Magnetic Resonance in Medicine*, 52(6):1358–1372. 6

Tuite, P. and Dagher, A. (2013). *Magnetic Resonance Imaging in Movement Disorders: A Guide for Clinicians and Scientists*. Cambridge University Press. 2

Tunç, B., Smith, A. R., Wasserman, D., Pennec, X., Wells, W. M., Verma, R., and Pohl, K. M. (2013). Multinomial Probabilistic Fiber Representation for Connectivity Driven Clustering. *Information processing in medical imaging : proceedings of the ... conference*, 23:730–741. 55

Tye, C. and Bolton, P. (2013). Neural connectivity abnormalities in autism: Insights from the Tuberous Sclerosis model. *BMC Medicine*, 11(1):1–4. 1, 74

Vaidya, C. J. (2012). Neurodevelopmental Abnormalities in ADHD. *Current Topics in Behavioral Neurosciences*, 9:49–66. 122

Vaillant, M. and Glaunès, J. (2005). Surface Matching via Currents. In Christensen, G. E. and Sonka, M., editors, *Information Processing in Medical Imaging*, number 3565 in Lecture Notes in Computer Science, pages 381–392. Springer Berlin Heidelberg. 9, 27, 56, 57, 82

Vaillant, M., Miller, M. I., Younes, L., and Trouvé, A. (2004). Statistics on diffeomorphisms via tangent space representations. *NeuroImage*, 23, Supplement 1:S161–S169. 12, 31

Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2008). Symmetric Log-Domain Diffeomorphic Registration: A Demons-Based Approach. In Metaxas, D., Axel, L., Fichtinger, G., and Székely, G., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008*, number 5241 in Lecture Notes in Computer Science, pages 754–761. Springer Berlin Heidelberg. DOI: 10.1007/978-3-540-85988-8_90. 13

Vialard, F.-X., Risser, L., Rueckert, D., and Cotter, C. J. (2011). Diffeomorphic 3d Image Registration via Geodesic Shooting Using an Efficient Adjoint Calculation. *International Journal of Computer Vision*, 97(2):229–241. 12

Wang, L., Lee, D. Y., Bailey, E., Hartlein, J. M., Gado, M. H., Miller, M. I., and Black, K. J. (2007). Validity of large-deformation high dimensional brain mapping of the basal ganglia in adults with Tourette syndrome. *Psychiatry Research: Neuroimaging*, 154(2):181–190. 13

Wassermann, D., Bloy, L., Kanterakis, E., Verma, R., and Deriche, R. (2010). Unsupervised white matter fiber clustering and tract probability map generation: Applications of a Gaussian process framework for white matter fibers. *NeuroImage*, 51(1):228–241. 55, 56

Wassermann, D., Toews, M., Niethammer, M., and Iii, W. W. (2014). Probabilistic Diffeomorphic Registration: Representing Uncertainty. In Ourselin, S. and Modat, M., editors, *Biomedical Image Registration*, number 8545 in Lecture Notes in Computer Science, pages 72–82. Springer International Publishing. 24

Wolz, R., Aljabar, P., Hajnal, J., Lotjonen, J., and Rueckert, D. (2011). Manifold learning combining imaging with non-imaging information. In *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1637–1640. 125

Worbe, Y., Baup, N., Grabli, D., Chaigneau, M., Mounayar, S., McCairn, K., Féger, J., and Tremblay, L. (2009). Behavioral and Movement Disorders Induced by Local Inhibitory Dysfunction in Primate Striatum. *Cerebral Cortex*, 19(8):1844–1856. 4

Worbe, Y., Gerardin, E., Hartmann, A., Valabregue, R., Chupin, M., Tremblay, L., Vidailhet, M., Colliot, O., and Lehericy, S. (2010). Distinct structural changes underpin clinical phenotypes in patients with Gilles de la Tourette syndrome. *Brain*, 133(12):3649–3660. 4, 121

Worbe, Y. and Hartmann, A. (2013). Neuroimaging of Gilles de la Tourette syndrome. In *Magnetic Resonance Imaging in Movement Disorders: A Guide for Clinicians and Scientists*. Cambridge University Press. 3

Worbe, Y., Malherbe, C., Hartmann, A., Pelegrini-Issac, M., Messe, A., Vidailhet, M., Lehericy, S., and Benali, H. (2012). Functional immaturity of cortico-basal ganglia networks in Gilles de la Tourette syndrome. *Brain*, 135(6):1937–1946. 14

Worbe, Y., Marrakchi-Kacem, L., Lecomte, S., Valabregue, R., Poupon, F., Guevara, P., Tucholka, A., Mangin, J.-F., Vidailhet, M., Lehericy, S., Hartmann, A., and Poupon, C. (2015). Altered structural connectivity of cortico-striato-pallido-thalamic networks in Gilles de la Tourette syndrome. *Brain*, 138(2):472–482. 1, 3, 14, 37, 65, 74, 75, 93, 121

Worbe, Y., Sgambato-Faure, V., Epinat, J., Chaigneau, M., Tandé, D., François, C., Féger, J., and Tremblay, L. (2013). Towards a primate model of Gilles de la Tourette syndrome: Anatomo-behavioural correlation of disorders induced by striatal dysfunction. *Cortex*, 49(4):1126–1140. 4

Xia, S., Li, X., Kimball, A. E., Kelly, M. S., Lesser, I., and Branch, C. (2012). Thalamic shape and connectivity abnormalities in children with

attention- deficit/hyperactivity disorder. *Psychiatry Research: Neuroimaging*, 204(2–3):161–167. 1

Younes, L., Michor, P., Shah, J., and Mumford, D. (2008). A metric on shape space with explicit geodesics. *Rendiconti Lincei - Matematica e Applicazioni*, pages 25–57. 11

Younes, L. and Richardson, C. L. (2013). Computing metamorphoses between discrete measures. *Journal of Geometric Mechanics*, 5(1):131–150. 123

Yushkevich, P., Pizer, S. M., Joshi, S., and Marron, J. S. (2001). Intuitive, localized analysis of shape variability. In *Information Processing in Medical Imaging*, pages 402–408. Springer. 9

Yushkevich, P. A., Zhang, H., Simon, T. J., and Gee, J. C. (2008). Structure-specific statistical mapping of white matter tracts. *NeuroImage*, 41(2):448–461. 56

Zhang, M., Singh, N., and Fletcher, P. T. (2013). Bayesian Estimation of Regularization and Atlas Building in Diffeomorphic Image Registration. In Gee, J. C., Joshi, S., Pohl, K. M., Wells, W. M., and Zöllei, L., editors, *Information Processing in Medical Imaging*, number 7917 in Lecture Notes in Computer Science, pages 37–48. Springer Berlin Heidelberg. 11, 23

Zvitia, O., Mayer, A., Shadmi, R., Miron, S., and Greenspan, H. (2010). Co-registration of White Matter Tractographies by Adaptive-Mean-Shift and Gaussian Mixture Modeling. *IEEE Transactions on Medical Imaging*, 29(1):132–145. 55