Double diffeomorphism: combining morphometry and structural connectivity analysis
Pietro Gori, Olivier Colliot, Linda Kacem, Yulia Worbe, Alexandre Routier, Cyril Poupon, Andreas Hartmann, Nicholas Ayache, Stanley Durrleman

To cite this version:
Pietro Gori, Olivier Colliot, Linda Kacem, Yulia Worbe, Alexandre Routier, et al.. Double diffeomorphism: combining morphometry and structural connectivity analysis. IEEE Transactions on Medical Imaging, Institute of Electrical and Electronics Engineers, 2018, 37 (9), pp.2033-2043. 10.1109/TMI.2018.2813062 . hal-01709847v2

HAL Id: hal-01709847
https://hal.archives-ouvertes.fr/hal-01709847v2
Submitted on 23 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Double diffeomorphism: combining morphometry and structural connectivity analysis

Pietro Gori*, Olivier Colliot, Linda Marrakchi Kacem, Yulia Worbe, Alexandre Routier, Cyril Poupon, Andreas Hartmann, Nicholas Ayache and Stanley Durrleman

Abstract—The brain is composed of several neural circuits which may be seen as anatomical complexes composed of grey matter structures interconnected by white matter tracts. Grey and white matter components may be modelled as 3D surfaces and curves respectively. Neurodevelopmental disorders involve morphological and organizational alterations which can not be jointly captured by usual shape analysis techniques based on single diffeomorphisms. We propose a new deformation scheme, called double diffeomorphism, which is a combination of two diffeomorphisms. The first one captures changes in structural connectivity, whereas the second one recovers the global morphological variations of both grey and white matter structures. This deformation model is integrated into a Bayesian framework for atlas construction. We evaluate it on a data-set of 3D structures segmented from Magnetic Resonance (MR) T1-w images. Neural projections of the white matter are instead modelled as bundles of 3D streamlines, called fiber bundles, which result from tractography algorithms applied on diffusion MR images. Every streamline is an estimate of the trajectories of large groups of neural axons. The geometrical representation of the neural circuits combines thus both surface and curve meshes into a single multi-object complex, called shape complex.

A. Related Work

The statistical shape analysis of 3D meshes has been the subject of several works. One of the most popular strategy relies on the selection of consistent correspondences between the structures of the subjects [9]. Correspondences can be, for instance, manually chosen by an expert (i.e. landmarks) [10], estimated using shape descriptors [11]–[13] or found with the iterative closest point (ICP) algorithm (or variants of it [14]). All structures are then aligned to a common reference frame where both mean and covariance matrix can be estimated. Principal Component Analysis (PCA) or Principal Geodesic Analysis (PGA) [15] can be employed to analyse the main morphological variations [16]. This strategy has been successfully employed with several brain structures. However, most of the shape descriptors are conceived for only a particular kind of mesh, i.e. genus-zero surfaces or streamlines. Thus, they can not be used to handle both grey and white matter structures into a single framework. Moreover, this approach may not preserve the anatomical organization of the neural circuits, which means that separated structures may intersect when computing the average or the main morphological variations.

Another class of statistical shape models, which naturally allows the combination of different mesh types, is based on the Grenander’s pattern theory [17]. Every shape complex is modelled as a deformation of a reference shape complex called template complex. Deformations put into correspondence the components of the template complex with the homologous ones of the subjects. The “amount” of deformation needed to warp the template complex to the subject complex quantifies their morphological differences. The joint estimate of template

Index Terms—shape, morphometry, complex, multi-object, atlas, structural connectivity, Tourette, neural circuits

I. INTRODUCTION

The pathophysiology of neurodegenerative and neurodevelopmental disorders, such as Parkinson’s disease and Gilles de la Tourette syndrome (GTS), often involves morphological alterations of the cortico-basal ganglia and corticothalamus circuits [1], [2]. These networks are composed of neural projections connecting specific areas of the cortical surface and sub-cortical nuclei. Abnormalities can affect: i) the shape of every component of the circuits from both grey and white matter, ii) the relative position between grey matter structures and iii) the structural connectivity, namely the areas where white matter tracts integrate grey matter structures. Most of the studies present in the literature focus either on the first or on the last point [2], [3]. Few of them analyse the first two points together [4], [5]. In this paper, we propose a new method to tackle all points at the same time.

Every component of the neural circuits may be segmented as a 3D object. Grey matter structures, such as cortical surface and basal ganglia, are represented as surface meshes...
complex and deformations is called atlas construction \cite{7}–\cite{20}. They capture the common anatomical characteristics and the morphological variability of the population respectively.

Deformations are usually defined as single diffeomorphisms of the entire ambient space which are smooth invertible transformations with smooth inverse. This kind of deformation preserves the anatomical organization of the components of the template complex, namely they can not intersect, fold or shear. Moreover, deformations are defined locally and they can vary across different areas of the ambient space. This makes it possible to capture the variations in relative position between separate structures. However, using a single diffeomorphism, one implicitly assumes that the relative position between structures in contact with each other or, in practice, close to each other, does not change across subjects. This implies that a particular fiber bundle of the neural circuits should integrate the same areas of the cortical surface and basal ganglia across the whole population. This assumption precludes the study of changes in structural connectivity which could be caused by an abnormal brain development. In Fig\[1\] we present a toy example composed of a template complex and a subject shape complex characterized by a different structural connectivity. A single diffeomorphism could not put into correspondence all structures and capture the differences in structural connectivity.

Structural connectivity analysis is usually based on the partition of the cortical surface and sub-cortical nuclei in consistent parcels across subjects \cite{21}. Every parcel is considered as a node of a graph and the number of streamlines connecting two nodes (or other quantities such as the projected Fractional Anisotropy) represents the weighted edge. Variability in structural connectivity across subjects can be analysed in each parcel independently or with indexes and methods from the complex network theory \cite{2}, \cite{22}. In both cases, the analysis highly depends on the chosen parcellation scheme and it does not take into consideration the morphological variability of grey and white matter structures.

![Template complex and Subject complex](image)

**Fig. 1.** Two complexes composed of a pseudo cortex, divided into black and green gyri, a blue sub-cortical nucleus and a red fiber bundle. A single diffeomorphism could not put into correspondence all 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 complex or to the red 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 structures, producing an accurate matching and capturing also the dissimilarities in structural connectivity.

### B. Our contribution

In this paper, extending \cite{23}, we propose to join together shape and structural connectivity analysis into a unified framework based on a double diffeomorphic atlas construction. The template complex is warped towards every shape complex of the population using a composition of two diffeomorphisms. The first diffeomorphism acts only on the white matter of the template complex, keeping fixed the grey matter. During this transformation, the fiber bundles are repositioned with respect to the grey matter structures, capturing the variations in structural connectivity. The second diffeomorphism acts on the whole template complex, namely on both the resulting white matter and grey matter, bringing all structures of the template complex into the subject’s space. White matter tracts are re-arranged by the first diffeomorphism so that the second one can correctly put into correspondence all the components of the template complex. The two diffeomorphisms are optimised together minimising a single cost function. The data-term only depends on the deformed template complex resulting from the second diffeomorphism. Using again the example in Fig\[1\] the first diffeomorphism would move the fiber bundle from the left gyrus to the right one. The second diffeomorphism would then modify the shape of all structures producing an accurate matching. The first diffeomorphism would capture the changes in structural connectivity, whereas the second one would recover the global morphological differences. Both diffeomorphisms are parametrized using control points as proposed in \cite{24}. The number of control points is fixed by the user and their position is automatically adjusted during the atlas construction. To note that, we estimate two distinct deformation fields (no composition is performed) and that smoothness across white and grey matter is guaranteed by the fact that they are jointly deformed only by the second diffeomorphism.

Our approach is different from other multi-diffeomorphic methods with sliding conditions such as \cite{25}–\cite{27}. These methods aim to correctly register longitudinal scans or anatomical complexes characterized by sliding regions. Every region is smoothly and independently deformed. Contrary to that, we are interested in studying the relative variation of one region, white matter, with respect to another one, grey matter. The aforementioned sliding registrations, if applied to the example shown in Fig\[1\] would result into two independent deformations, one for the white matter and one for the grey matter. It would be thus impossible to understand whether the deformation of the white matter is due to a difference in grey matter or to a variation in structural connectivity. Furthermore, the proposed method differs from multi-scale diffeomorphisms, such as \cite{28}, \cite{29}, which combine multiple kernels at different scales to create one single diffeomorphic deformation. In this case, the goal is mainly to improve the registration accuracy and remove the scale tuning.

In order to deal with the considerable amount of streamlines resulting from tractography algorithms, we rely on the parsimonious representation, based on weighted prototypes, introduced in \cite{30}. Both prototypes and streamlines are modelled as weighted currents \cite{7}. This model is well suited for any kind of fiber bundle, both sheet-like \cite{31} and tubular \cite{12}. Furthermore, we propose to model grey matter structures as varifolds \cite{32}, \cite{33}, the non-oriented extension of the framework of currents \cite{7}, \cite{34}, or landmarks, if correspondences across subjects are available. The atlas is estimated within a Bayesian
framework based on a generative model similar to the one proposed in [35] and adapted to double diffeomorphisms.

The paper is organized as follows. We first describe the double diffeomorphic generative model and the Bayesian framework for atlas construction. We initially model all shapes with landmarks and then we describe how to integrate the computational models of varifolds and weighted currents. We conclude Sec III showing how to compute the diffeomorphic deformations and with a description of the optimization procedure. Eventually, we evaluate the discriminative power of the proposed method to distinguish between a population of controls and one of GTS patients. We also compare the resulting classification scores with the ones obtained using a single diffeomorphism.

II. METHODS

A. 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 \( M = M^G + M^W \) 3D discrete geometric representations (points, polylines or polygon meshes) from both grey (\( M^G \)) and white matter (\( M^W \)). We define the representation of structure \( j \) belonging to subject \( j \) as \( S_{ij} \). Every subject shape complex \( S_i \), defined for the moment in a generic way as a double deformation of a common template complex \( T \) plus a residual noise \( \epsilon_i \). Both \( T \) and \( \epsilon_i \) are also defined as the ensembles of the templates \( T_j \) and residuals \( \epsilon_{ij} \). The first deformation \( \phi^W \) acts only on the white matter structures of the template complex: \( T^W \). The grey matter of the template complex \( T^G \) is not modified. The second deformation \( \phi^All \) deforms both the resulting white matter \( \phi^W(T^W) \) and \( T^G \). This formulation derives from the forward model [35], [38], [39] where we assume that all elements belong to an algebraic structure where addition is defined. It results:

\[
S_i = \phi^All_i(\phi^W_i(T^W) \cup T^G) + \epsilon_i
\tag{1}
\]

The two deformations, \( \phi^W_i \) and \( \phi^All_i \), are two diffeomorphisms of the entire ambient space. They follow one another creating a cascade of diffeomorphisms. White matter streamlines of \( T^W \) are re-positioned by \( \phi^W_i \) within the grey matter \( T^G \), which is kept fixed. This can be seen as a relative change of coordinates with respect to \( T^G \), which is considered as a fixed reference frame common to all subjects \( i \). The entire template complex, both \( T^G \) and \( \phi^W_i(T^W) \), are then registered to the subject shape complex \( S_i \) by \( \phi^All_i \). This is instead a global change of coordinates which brings the template complex to the subject space. The two deformations, \( \phi^W_i \) and \( \phi^All_i \), capture the differences in structural connectivity and the global morphological changes, common to both white and grey matter, respectively. A diagram based on the toy example of Fig. I can be found in the Appendix.

B. Bayesian Atlas Construction

The goal of the atlas construction is to estimate the template complex \( T = T^W \cup T^G \), the variations in structural connectivity within the population described by the ensemble of first diffeomorphisms \( \{\phi^W_i\} \) and the global morphological variations captured by the second diffeomorphisms \( \{\phi^All_i\} \). Both diffeomorphisms are parametrized by a set of parameters, \( \alpha^W \) and \( \alpha^All \) respectively, specific to every subject \( i \). We assume that these parameters follow a Gaussian distribution with zero mean and covariance matrix equal to \( \Gamma^W \) and \( \Gamma^All \) respectively: \( \alpha^W_i \sim N(0, \Gamma^W) \) and \( \alpha^All_i \sim N(0, \Gamma^All) \). Moreover, as usual in statistical learning, we assume that the residuals follow a Gaussian distribution centered at 0 and with a scalar matrix as covariance matrix (\( \epsilon_{ij} \sim N(0, \alpha^2_i \Lambda^S) \)). For now, we model all structures with landmarks, which are 3D points reproducible among subjects that establish a point correspondence. For every subject, structure \( j \) is modelled using \( \Lambda_j \) landmarks. Thus, \( \Lambda_j \) is the identity matrix of size \( \Lambda_j \). The norm of the difference between two meshes is the \( L^2 \)-norm (|| \cdot ||_2, i.e. the sum of squared differences between corresponding landmark pairs). The likelihoods of the residuals of white (\( W \)) and grey (\( G \)) matter structures modelled as landmarks are:

\[
p(\epsilon_{ij}^W | \sigma^W_{ij}) \propto \frac{1}{|\sigma^W_{ij}|^{1/2}} \exp\left[-\frac{1}{\sigma^W_{ij}^2}\|S_{ij} - \phi^W_i(\phi^W_i(T^W))\|_2^2\right]
\]

\[
p(\epsilon_{ij}^G | \sigma^G_{ij}) \propto \frac{1}{|\sigma^G_{ij}|^{1/2}} \exp\left[-\frac{1}{\sigma^G_{ij}^2}\|S_{ij} - \phi^All_i(\phi^W_i(T^G))\|_2^2\right]
\tag{2}
\]

where \( \sigma^G_{ij} \) and \( \sigma^W_{ij} \) refer to grey and white matter structures respectively. In the following, we refer to \( \sigma_j \) when we make no distinction between grey and white matter structures. In Sec II-C, we will make clear how to adapt these equations when modeling a structure as weighted current or varifold.

Whatever the model employed, the variance only depends on the structure-dependent parameter \( \sigma^2_j \). Moreover, from Eq. 1 and Eq. 2, it follows that all shapes \( S_{ij} \) follow a Gaussian distribution: \( S_{ij}^W \sim N(\phi^W_i(\phi^W_i(T^W)), \sigma^W_j \Lambda^S_j) \), \( S_{ij}^G \sim N(\phi^All_i(\phi^W_i(T^G)), \sigma^G_j \Lambda^G_j) \). The two covariance matrices of the deformation parameters, \( \Gamma^W \) and \( \Gamma^All \), are also considered as parameters of the model. We can thus reformulate the goal of the atlas construction as estimating \( T, \Gamma^W, \Gamma^All \), knowing the shape complexes \( \{S_i\} \) and assuming they follow a Gaussian distribution. This can be achieved by maximizing the joint posterior distribution of \( T, \sigma^2_j, \Gamma^W, \Gamma^All \), assuming independence between all random variables and considering \( \{\alpha^W_i\} \) and \( \{\alpha^All_i\} \) as hidden variables, it results:

\[
\{T^*, \Gamma^W, \Gamma^All, \sigma^2\} = \arg\max_{T, \Gamma^W, \Gamma^All, \sigma^2} \prod_{i,j} p(T_j, \Gamma^W, \Gamma^All, \alpha^W_j, \alpha^All_j, S_{ij})d\alpha^Alld\alpha^W
\tag{3}
\]

Not using priors for \( \Gamma^W, \Gamma^All \) and \( \sigma^2 \) can produce degenerate estimates with small training data-sets, as demonstrated in [35]. A possible solution is to regularize the estimates using adapted versions of the inverse Wishart distributions as priors \( \sigma^2_j \sim W^{-1}(P_j, w_j), \Gamma^W \sim W^{-1}(P^W, w^W), \Gamma^All \sim W^{-1}(P^All, w^All) \):
shapes \( S \), modelled as varifolds or weighted currents, to finite dimensional spaces. For each structure \( j \), we define a regular grid composed of \( \Lambda_j \) points which covers the ambient space and where pdf can be computed. The reader is referred to [41] Chapter 4.2.3 for more details.

**Varifolds** - In the space of varifolds \( W^* \), the inner-product between two surfaces \( X \) and \( Y \) is: \( \langle X, Y \rangle_{W^*} = \sum_{l=1}^L \sum_{h=1}^H \exp \left( \frac{-||p_l-h||^2}{\lambda^2_h} \right) \left( \frac{n_l^{bh}}{|n_h||n_l|} \right)^2 |n_h| |u_h| |u_l| \) where \( n_l \) (resp. \( u_h \)) is the normal of \( X \) (resp. \( Y \)) at point \( p_l \) (resp. \( q_h \)). The only user-defined parameter is the kernel bandwidth \( \lambda_h \). The distance between \( X \) and \( Y \) is: \( ||X-Y||_{W^*} = \langle X, Y \rangle_{W^*} \). Two important characteristics of this metric are: the absence of correspondences and the invariance to a change of orientation of some normals of the surfaces. For more information, the user is referred to [32].

**Weighted currents** - The inner product between two oriented 3D polygonal curves, \( A \) and \( B \), modelled as weighted currents and composed of \( G \) and \( F \) segments respectively, is: \( \langle C_A, C_B \rangle_{Q^*} = \sum_{g=1}^G \sum_{f=1}^F \exp \left( -\frac{||x_g-y_f||^2}{2\sigma^2_y} \right) \alpha_g^\beta_f \exp \left( -\frac{(||f^\alpha-e^\alpha||^2)}{2\sigma^2_y} \right) \exp \left( -\frac{(||f^\beta-e^\beta||^2)}{2\sigma^2_y} \right) \). As usual, currents, curves need to have a consistent orientation. This can be achieved by tracing all streamlines of a bundle from one ROI (Region Of Interest) to another one, as it is the case in this paper. For more details, please see [30].

**Weighted prototypes** - We approximate white matter fiber bundles with a parsimonious representation of weighted streamlines prototypes. Prototypes are chosen among the streamlines by minimizing an approximation error based on the metric of weighted currents. Every prototype represents an ensemble of streamlines which share similar endpoints and pathway. The weight of the prototype is related to the number of streamlines approximated. An outlier detection and removal step is also performed during the algorithm. This approximation preserves the global shape of the bundle and its structural connectivity, which is fundamental for the scope of this paper. For more information the user is referred to [30].

\[ p(\sigma^2_j, \varphi_j^W) \propto \exp \left( \frac{-1}{2\sigma^2_j} \right) \exp \left( \frac{-1}{2\sigma^2_j} \right) \]

\[ p(\Gamma^W_\alpha) \propto |\Gamma^W_\alpha|^{-\frac{\alpha}{2}} \exp \left( -\frac{1}{2|\Gamma^W_\alpha|} \right) \]

\[ p(\Gamma^{All}_\alpha) \propto |\Gamma^{All}_\alpha|^{-\frac{\alpha}{2}} \exp \left( -\frac{1}{2|\Gamma^{All}_\alpha|} \right) \]

The scalars \( w_{ij}, p_j, w^W_{ij} \) and \( w^{All}_{ij} \) are strictly positive and \( P^W_\alpha \) and \( P^{All}_\alpha \) are positive symmetric matrices. They are hyper-parameters fixed by the user (see Sec. II-E to get more insight). Since the maximization of Eq. (4) 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. See [35], [37] for more information, the user is referred to [35].

\[ \text{C. Similarity metrics for shape complexes} \]

When landmarks are not available, we propose to use two correspondence-free shape metrics: varifolds [32], for grey matter surfaces, and weighted currents [30], for white matter streamlines. In both frameworks, meshes are embedded into a infinite-dimensional Hilbert space where the union of surfaces or streamlines is equal to a sum of varifolds (\( V \)) or weighted currents (\( C \)) respectively. We define Gaussian variables in these two spaces similarly to [40] for the framework of currents. Since they do not have probability density functions (pdf) in infinite dimension, we project both templates \( T \) and
\[ \frac{\partial \phi_i(x,t)}{\partial t} = v_i \phi_i(x,t) \] over \( t \in [0,1] \) where \( v_i(x_i(t), t) \) is a time-varying vector field representing the instantaneous velocity of a point located at \( x_i(t) \) at time \( t \). Both vector fields \( v^{All} \) and \( v^W \) belong to the same RKHS \( D \) with Gaussian kernel \( K_D \). They are defined by two different sets of 3D control points, \( c^{All} \) and \( c^W \), shared among all subjects, and by two distinct sets of 3D vectors, called momenta, \( \alpha^{All} \) and \( \alpha^W \), linked to the control points and specific to each subject: \( v^{All}(x_i(t), t) = K_D(x_i(t), c^{All}(t)) \alpha^{All}(t) \) and \( v^W(x_i(t)) = K_D(x_i(t), c^W(t)) \alpha^W(t) \), where \( x_i(0) = x \) and \( K_D(x_i(t), c(t)) \) represents a block matrix of Gaussian kernels with an equal fixed bandwidth for both \( v^{All} \) and \( v^W \). The deformation of every point \( x \) in the ambient space depends on its initial position at \( t = 0 \) and on the evolution of the system \( L^{All}(t) = \{c^{All}(t), \alpha^{All}(t)\} \) if the point belongs to the grey matter, and on both systems \( L^W(t) = \{c^W(t), \alpha^W(t)\} \) and \( L^{All}(t) \) if the point belongs to the white matter. At \( t = 0 \) the deformations \( \phi^{All}_0 \) and \( \phi^W_0 \) are equal to the identity transformations. For both systems, the path from \( \phi^{W}_0 \) to \( \phi^{All}_i \) (resp. \( \phi^W_0 \)), 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 \( \int_0^1 ||v^{All}_i||^2 \) (resp. \( \int_0^1 ||v^W_i||^2 \)). It has been shown in [20] that the extremal paths are such that both systems \( L^W(t) \) and \( L^{All}(t) \) satisfy:

\[ \begin{aligned} \dot{c}_i(t) &= K_D(c_i(t), c_i(t)) \alpha_i(t) = F^*(c_i(t), \alpha_i(t)) \\ \dot{\alpha}_i(t) &= -\alpha_i(t)^T \nabla_i K_D(c_i(t), c_i(t)) = F^*(c_i(t), \alpha_i(t)) \end{aligned} \]

s.t. \( c_i(0) = c(0) = c_0 \), \( \alpha_i(0) = \alpha_0 \) \( \alpha^{All} \) and \( \alpha^W \) are minimised using a gradient descent scheme. Instead, \( \Gamma^{All}_\alpha, \Gamma^W_\alpha \) and \( \{\sigma_j^2\} \) have closed form solutions due to the use of conjugate priors:

\[ \begin{aligned} \tilde{\Gamma}^W_\alpha &= \sum_{i=1}^N \left( [(\alpha^{All}_i(\alpha^W_0)^T] + w^W_0 (P^W_\alpha)^T \right) \\ \tilde{\Gamma}^{All}_\alpha &= \sum_{i=1}^N \left( [\alpha^{All}_i(\alpha^{All}_0)^T] + w^{All}_0 (P^{All}_\alpha)^T \right) \\ (\sigma^j)^2 &= \sum_{i=1}^N ||P_i(S_{ij} - \phi^{All}_i(T^W_i))||^2_{W_{s_j}} + w_j P_j \\ (\tilde{\sigma}^j)^2 &= \sum_{i=1}^N ||P_i(S_{ij} - \phi^{W}_i(T^W_i))||^2_{W_{s_j}} + w_j P_j \end{aligned} \]

where \( W_{s_j} \) and \( Q_{s_j} \) are the finite dimensional spaces where warp fields and weighted currents are projected to. For more details about the projection \( \Pi \), the user is referred to [41] Chapter 4.2.3.3.

The two covariance matrices \( \tilde{\Gamma}^W_\alpha \) and \( \tilde{\Gamma}^{All}_\alpha \) are equal to a weighted sum between the sample covariance matrix of the initial momenta and a prior. We choose \( P^W_\alpha = K_D^{-1}(c^W_0, c^W_0) \) and \( P^{All}_\alpha = K_D^{-1}(c^{All}_0, c^{All}_0) \) which are block matrices of Gaussian kernels between the initial control points. Note that \( K_D \) is the kernel of the RKHS to which belong both vector fields \( v^{All} \) and \( v^W \). This choice is motivated by the fact that, when \( N \ll w^W_0 \), it results \( \tilde{\Gamma}^W_\alpha \approx K_D^{-1}(c^W_0, c^W_0) \) and, consequently, the regularity term in \( \text{Eq. (7)} \) becomes \( \sum_{i=1}^N (\alpha^{All}_0)^T K_D(c^W_i, c^W_i) \alpha^{All}_0 = \sum_{i=1}^N ||v^{All}_i||^2_D \) which is the sum of the lengths of the geodesic paths over all subjects. This kind of regularity term has been often employed in previous

A diagram of the double diffeomorphism is shown in Fig. 2.
The two other parameters, \( \hat{\sigma}^A_t \) and \( \hat{\sigma}^W_t \), are equal to a weighted sum between the data-term of the \( j \)-th structure and the prior \( P_j \). Each parameter balances the importance of structure \( j \) with respect to the other structures and with respect to the regularity terms of both diffeomorphisms. The prior \( P_j \) imposes a minimum value to \( \hat{\sigma}_j \) 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 \( \hat{\sigma}^2_t \rightarrow 0 \) and therefore also to \( \log(\hat{\sigma}^2_t) \rightarrow -\infty \).

The gradients of the cost function in Eq.9 with respect to \( \alpha^A_{s,t} \) and \( \alpha^W_{s,t} \) are equal to:

\[
\nabla_{\alpha^A_{s,t}} E = \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{1}{2\sigma^2} \nabla_{\alpha^A_{s,t}} D_{ij} \\
\nabla_{\alpha^W_{s,t}} E = \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{1}{2\sigma^2} \nabla_{\alpha^W_{s,t}} D_{ij}
\]

where \( D^G_{ij} = \| \Pi(S_{ij} - \phi^A_{\alpha^A_t}(T^G_{ij})) \|^2_{\Sigma_{ij}} \) and \( D^W_{ij} = \| \Pi(S_{ij} - \phi^W_{\alpha^W_t}(T^W_{ij})) \|^2_{\Sigma_{ij}} \) and refer to the data-terms of grey and white matter structures respectively whereas \( D_{ij} \) refers to the data-term of any structure.

To calculate the gradients of the data terms \( \{ D_{ij} \} \), we need the deformed template complex at the end of the second diffeomorphism for every subject \( i \) and for all structures, namely \( \phi^A_{\alpha^A_t}(T^G) \) and \( \phi^W_{\alpha^W_t}(T^W) \). First, we integrate forward in time \( L^W_{t} \) and \( L^G_{t} \) (Eq.7). Then, we integrate \( L^A_{t} \) and \( L^W_{t} \) with respect to the vertices of \( T^A_{\alpha^A_t} \) and \( T^W_{\alpha^W_t} \). Using the calculus of variations, this information is brought back from \( t = 1 \) to \( t = 0 \) to update first \( T^A_{\alpha^A_t}(0) \) and \( T^G \) and then \( L^W_{t}(0) \) and \( \{ \alpha^A_{i,0}, \alpha^W_{i,0} \} \) and \( T^W \). The optimisation is based on a set of adjoint equations describing the evolution of four auxiliary variables \( \theta_{i,t}^A, \xi_{i,t}^W, \hat{\theta}_{i,t}^W, \hat{\xi}_{i,t}^W \):

\[
\begin{align*}
\theta_{i,t}^A &= (d_{L^A_t} F^W_{t}(t))^T \xi_{i,t}^W \\
\hat{\theta}_{i,t}^W &= -\partial_{T^W} Z^W_{i,t}(t) \beta_{i,t}^A(t) \\
\xi_{i,t}^W &= -\partial_{L^A_t} Z^A_{i,t}(t) \beta_{i,t}^W(t) + \partial_{L^W_t} F^W_{t}(t) \hat{\theta}_{i,t}^W \\
\hat{\xi}_{i,t}^W &= -\partial_{T^W} Z^W_{i,t}(t) \beta_{i,t}^W(t) + \partial_{L^A_t} F^W_{t}(t) \hat{\theta}_{i,t}^A(t)
\end{align*}
\]

where \( Z^A_{i,t}(t) = Z[T^A_{i,t}(t), L^A_{i,t}(t)] \) and \( Z^W_{i,t}(t) = Z[T^W_{i,t}(t), L^W_{i,t}(t)] \). The size of \( \theta_{i,t}^A \) (resp. \( \theta_{i,t}^W \)) and \( \xi_{i,t}^A \) (resp. \( \xi_{i,t}^W \)) are the same as the ones of \( T^A \) (resp. \( T^W \)) and \( L^A_{i,t} \) (resp. \( L^W_{i,t} \)) respectively. We first integrate backward in time Eq.11 obtaining \( \theta_{i,t}^A(0) \) and \( \xi_{i,t}^A(0) = \{ \xi_{i,t}^W(0), \xi_{i,t}^A(0) \} \). Then, we use \( \theta_{i,t}^A(0) \), which are the initial values of \( \theta_{i,t}^A \) relative to the white matter, as final values for \( \theta_{i,t}^W \) and we integrate backward in time Eq.13 and Eq.14 obtaining \( \theta_{i,t}^W(0) \) and \( \xi_{i,t}^W(0) = \{ \xi_{i,t}^W(0), \xi_{i,t}^A(0) \} \). From this set of equations, we can notice that the optimisation of the two diffeomorphisms is linked by the constraint \( \theta_{i,t}^W(1) = \theta_{i,t}^A(0) \).

Fig. 3. Diagram of the optimisation procedure. We omit the subject index \( i \) for clarity purpose.

F. Atlas Parameters Initialisation

Since we use a gradient descent scheme, we need to initialise the atlas parameters. Control points of both diffeomorphisms are initialised as a regular lattice covering the entire ambient space with an inter-points distance equal to the bandwidth of the diffeomorphic kernel \( K_D \). Momenta are initialised to zero. The template of surface meshes is initialised as the average of the population when a vertex-correspondence is available. Otherwise, we use a centred and scaled ellipsoid as in [20]. For the template of fiber bundles, all subjects'
bundles are first gathered together into a single bundle which is then approximated as a set of weighted prototypes. The weights of the prototypes are scaled so that the norm of the template is equal to the average norm of the population.

### III. Experiments and Results

In this section, we first describe the dataset used in the following experiments and some numerical aspects of the proposed algorithm. Then, we use both a toy example and real data to compare the registrations based on a single and double diffeomorphism. After that, we present an explanatory toy example where we show how one could use the proposed double diffeomorphism to compare two groups of subjects. Eventually, we assess the effectiveness of our algorithm by showing that it better discriminates between controls and GTS patients than a single diffeomorphism.

#### A. Materials

The dataset used in this paper contains 76 subjects: 27 controls and 49 GTS patients divided in three sub-groups based on their symptoms: ST=simple-tics (17 patients), CT=complex-tics (15), OCD=complex-tics with Obsessive Compulsive Disorders (12). Anatomical images are acquired using 3D MR T1-w sequences with a voxel size of 1x1x1 mm^3. MR diffusion weighted images are obtained with 50 gradient-directions, a B-factor of 1000 and a voxel size of 2x2x2 mm^3. Artefacts related to spike, motion, susceptibility and eddy currents are corrected using Connectomist-2.0. Diffusion and T1-w data are matched using a mutual information based registration technique. In the experiments, we use the left hemisphere of the cortical surface, left putamen and the fiber bundles connecting them. For the selection of the tracts, we use a specific technique conceived for the cortico-striatal circuits explained in [2]. Cortical surfaces are segmented using FreeSurfer v5.3 followed by a pipeline of BrainVisa v4.3.0 which produces a vertex-correspondence between subjects. Putamens are segmented with FSL. Fiber bundles result from a deterministic tractography algorithm (1 seed per voxel, SDT model) available in Connectomist-2.0. See [2] for more details about the data-set, pre-processing and tractography.

#### B. Numerical aspects

In the following experiments, cortical surfaces are modelled with landmarks, putamens as varifolds with λ_W=3mm and fiber bundles, approximated with weighted prototypes, as weighted currents with λ_β=7mm, λ_a=10mm (cortex) and λ_β=5mm (putamen). The bandwidths of both diffeomorphic kernels are equal to 11mm, which produce 804 control points. The maximum number of iterations for the atlas construction is 120 and the computations are performed on an 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 suite deformetrica (www.deformetrica.org). The computational time for an atlas with 10 subjects is about 37 hours. All shape complexes are previously rigidly registered to a reference shape complex.

#### C. Toy example - registration

In Fig 4 we compare the registrations of a toy template complex (blue) towards a toy subject complex (red) based on a single (first row) and double diffeomorphism (second row). Both complexes are composed of a pseudo cortical surface, sub-cortical structure and fiber bundle linking them, all modelled as varifolds. We use the same parameters for the two deformation schemes. Grey matter components have a similar shape but they do not share the same structural connectivity. As it is possible to notice, a single diffeomorphism can not correctly put into correspondence all structures. On the contrary, a double diffeomorphism makes first the fiber bundle move, keeping fixed the grey matter structures, and then it accurately registers all structures with the second diffeomorphism. In this way, it is possible to disentangle the differences in structural connectivity, captured by the first diffeomorphism, from the global morphological changes, captured by the second diffeomorphism.

#### D. Real data - registration

In Fig 5 we compare the results of a single and double diffeomorphism using real data. We match a control subject to a Gilles de la Tourette patient. A double diffeomorphism better aligns both white and grey matter (see Fig. 4 in the Appendix) than a single diffeomorphism, capturing at the same time the variations in structural connectivity (i.e. φ^W).

#### E. Toy Example - group differences

We present here an explanatory toy example of the proposed atlas construction procedure based on a toy data-set constituted of 6 pseudo shape complexes representing two different populations (3 controls and 3 GTS patients). They are shown in Fig 6 where it is possible to notice that the complexes of population A have a different organization and shape with respect to the ones of population B. The Bayesian atlas construction results in a final template complex and in the covariance matrices of the momenta of both diffeomorphisms. The template shows the characteristics common to both populations. The two covariance matrices describe the
organisational and global morphological variability within the 6 subjects respectively. We compute a Principal Component Analysis (PCA) for each covariance matrix and we deform the final template complex at ±σ (standard deviation) along the first modes of both PCAs. The main variations captured by the first 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. The first mode of the second PCA describes instead the main global morphological variations. We can notice that the 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 method and it is based on a simple toy data-set where the intra-group variations are definitely smaller than the inter-group ones. This is probably exaggerated 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 distinct clusters. In the next section, we will exploit this hypothesis by looking for the discriminant hyperplane that separates the two groups.

**F. Real data - Classification**

Here, we use the estimated initial momenta of the two diffeomorphisms, \( \alpha_{\phi_i^W} \) and \( \alpha_{\phi_i^{All}} \), as features to discriminate between controls and patients. Then, we compare the resulting 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). Since we use subjects from both groups, the final template should be positioned in between them in the shape space. The estimated template is successively warped to all the remaining \( J = 66 \) subjects by minimizing a cost function similar to Eq.5 where we do not sum over all subjects \( i \) and where we fix the controls points, \( \Gamma_{\alpha}^W \), \( \Gamma_{\alpha}^{All} \) and \( \sigma^2 \) to the values estimated during the initial atlas. The resulting initial momenta, \( \alpha_{\phi_i^{All}}^{W} \) and \( \alpha_{\phi_i^{All}}^{W} \), represent the input features of the classifier. We employ a Linear Discriminant Analysis (LDA) with a leave-one-out cross validation strategy. We assume that the class-conditional densities of the initial momenta are Gaussian with a covariance matrix equal to the one estimated during the initial atlas. This can be seen as a regularized LDA since the covariance matrix is estimated as in Eq.9. We separately test the discriminative power of the two diffeomorphisms by using either only \( \alpha_{\phi_i^{All}}^{W} \) or \( \alpha_{\phi_i^W} \). Moreover, we compare these results with the ones obtained using the initial momenta of a single diffeomorphism where we employ either only the fiber bundles or all structures from both grey and white matter. Resulting sensibility, sensitivity and balanced accuracy are shown in Table I where we separately use either all patients or each sub-group alone. We assess the statistical significance of the classification scores with a randomization test (1000 permutations). It is possible to notice that the classification scores based on the first (white) diffeomorphism, especially for the most severe patients (CT and OCD), are definitely better than using a single diffeomorphism.

Due to the variability of the results, we also investigate the sampling distributions of sensitivity, sensibility and balanced accuracy 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, we pick a random sample (with replacement) of the 44 patients which is classified, together with the 22 controls, using LDA. We repeat this process 1000 times. The histograms of balanced accuracy for the double and single diffeomorphic approach are shown in Fig.7. The average sensitivity and specificity is respectively: 74% and 51% for the global diffeomorphism, 73% and 64% for the white diffeomorphism, 64% and 48% for the single diffeomorphism, considering both white and grey matter, and 70% and 52% for the single diffeomorphism, using only the fiber bundles.
TABLE I
CLASSIFICATIONS SCORES

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity %</th>
<th>Specificity %</th>
<th>Balanced Accuracy %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Diffeomorphism - White and Grey Matter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST</td>
<td>12</td>
<td>36</td>
<td>24</td>
</tr>
<tr>
<td>CT</td>
<td>33</td>
<td>64</td>
<td>48</td>
</tr>
<tr>
<td>OCD</td>
<td>58</td>
<td>59</td>
<td>59</td>
</tr>
<tr>
<td>CT+OCD</td>
<td>52</td>
<td>64</td>
<td>58</td>
</tr>
<tr>
<td>ST+CT+OCD</td>
<td>54</td>
<td>41</td>
<td>48</td>
</tr>
<tr>
<td>Single Diffeomorphism - Only White Matter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST</td>
<td>53</td>
<td>54</td>
<td>54</td>
</tr>
<tr>
<td>CT</td>
<td>33</td>
<td>45</td>
<td>39</td>
</tr>
<tr>
<td>OCD</td>
<td>50</td>
<td>54</td>
<td>52</td>
</tr>
<tr>
<td>CT+OCD</td>
<td>59</td>
<td>59</td>
<td>59</td>
</tr>
<tr>
<td>ST+CT+OCD</td>
<td>66</td>
<td>45</td>
<td>56</td>
</tr>
<tr>
<td>Double Diffeomorphism - First (white) diffeomorphism</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST</td>
<td>47</td>
<td>59</td>
<td>53</td>
</tr>
<tr>
<td>CT</td>
<td>67</td>
<td>77</td>
<td>72*</td>
</tr>
<tr>
<td>OCD</td>
<td>50</td>
<td>82</td>
<td>66*</td>
</tr>
<tr>
<td>CT+OCD</td>
<td>74</td>
<td>64</td>
<td>69*</td>
</tr>
<tr>
<td>ST+CT+OCD</td>
<td>73</td>
<td>41</td>
<td>57</td>
</tr>
<tr>
<td>Double Diffeomorphism - Second (global) diffeomorphism</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST</td>
<td>29</td>
<td>50</td>
<td>40</td>
</tr>
<tr>
<td>CT</td>
<td>40</td>
<td>45</td>
<td>43</td>
</tr>
<tr>
<td>OCD</td>
<td>50</td>
<td>68</td>
<td>59</td>
</tr>
<tr>
<td>CT+OCD</td>
<td>52</td>
<td>68</td>
<td>60</td>
</tr>
<tr>
<td>ST+CT+OCD</td>
<td>70</td>
<td>50</td>
<td>60</td>
</tr>
</tbody>
</table>

*: p-value < 0.05

44 patients) using either $\alpha^{Alh}$ or $\alpha^{W}$. The typical configurations of patients and controls are found by deforming the template complex at $\mu - w^*$ and $\mu + w^*$ respectively, where $\mu = \frac{1}{2}(\mu_c + \mu_p)$ and $||w^*|| = ||\mu_c - \mu_p||$ with $\mu_c$ and $\mu_p$ equal to the averages of initial momenta of controls and patients respectively. In Fig. 8 we compare the typical structural connectivity of the two groups. The main differences are in the supplementary motor, premotor, superior frontal areas, insula and in the dorsal and ventro-lateral part of the putamen. These results are in line with those reported in the literature [2]. In Fig. 9 we compare the typical grey matter configurations of controls and patients. In this case, there is mainly a compression in the premotor and frontal area of the cortex, insula and occipital lobe. About the putamen, the main variations are in the fronto-dorsal and postero-ventral areas. In Fig. 5 of the Appendix, we show for comparison the main variations in structural connectivity and morphology only within the population of controls.

G. Most discriminative deformation axis

Eventually, we also compute the organizational and morphological characteristics proper to each group by deforming the template complex along the most discriminative deformation axis. We estimate the best linear decision boundary (i.e. $\alpha^T w^* - b^*$) with all the $J$ test subjects (22 controls and

IV. DISCUSSION AND CONCLUSIONS

We presented a double-diffeomorphic mesh-based atlas construction method. In contrast to standard single-diffeomorphic registrations, the cascade of two diffeomorphisms can put into correspondence anatomical complexes characterised by a different structural connectivity. We showed that this approach makes it possible to characterise, localise and quantify both organisational and morphological pathological anomalies altering grey and white matter structures.
It is important to notice that it is fundamental to first deform the white matter of the template complex and then the grey matter in order to retrieve the main variations in structural connectivity. In fact, 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 grey matter of the template complex. If one changed the order, deforming first the grey matter and then the white matter, it would not be possible to compare the variations in structural connectivity since the reference frame, given by the grey matter, would be different across subjects. A diagram describing these two approaches can be found in the Appendix.

White matter fiber bundles are not constrained to always stay in contact with the grey matter during the deformation. We only enforce, by modelling streamlines as weighted currents, that they will be close to the grey matter at the end of the second diffeomorphism (See Eq.\( \text{III-F} \)). To note that, the two diffeomorphisms are not explicitly weighted during the optimization procedure in Eq.\( \text{III-B} \) - \( \text{III-D} \). However, they both depend on the gradients of the data-terms, and therefore on the parameters of their corresponding computational models. Furthermore, the precision and flexibility of each diffeomorphism depend on its kernel bandwidth. In this work, since we aim on all fiber bundles. A possible solution would be to substitute the first diffeomorphism with \( N \) diffeomorphisms, where \( N \) would be equal to the number of fiber bundles. Every bundle would be then independently deformed by a diffeomorphism. In this way, we could capture the variations in structural connectivity proper to each bundle and the global morphological changes associated to the entire neural circuit.

In the proposed method, we assumed that the initial momenta of the two diffeomorphisms are independent, that is to say that \( p(\alpha_i^{All}, \alpha_i^W) = p(\alpha_i^{All})p(\alpha_i^W) \), even if the update rule for \( \alpha_i^{All} \) and \( \alpha_i^W \) are related as explained in Sec.\( \text{II-E} \). It would seem more reasonable to take that into account by modelling directly \( p(\alpha_i^{All}, \alpha_i^W) \) without the assumption of independence. We could model, for instance, their joint distribution as a single Gaussian distribution. However, the statistical relationship between \( \alpha_i^{All} \) and \( \alpha_i^W \) is highly complex since they are related by the linearised ODEs shown in Sec.\( \text{II-E} \) 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 captures useful and relevant information since it better discriminates between controls and patients than a single diffeomorphism. In particular, we observed that
the importance of evaluating the role of pathophysiological mechanisms underlying GTS.

ACKNOWLEDGMENT

The research leading to these results received funding from ANR-10-IAIHU-06 and from l’Association Française du Syndrome Gilles de la Tourette (AFSGT).

REFERENCES


