Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution

Stanley Durrleman

UNIVERSITE DE NICE - SOPHIA ANTIPOLIS

ECOLE DOCTORALE STIC
SCIENCES ET TECHNOLGIES DE L'INFORMATION ET DE LA COMMUNICATION

T H E S E

pour obtenir le grade de

Docteur ès Sciences

de l'Université Nice - Sophia Antipolis

Mention : AUTOMATIQUE, TRAITEMENT DU SIGNAL ET DES IMAGES

presentée et soutenue par

Stanley DURRLEMAN

Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution

Thèse dirigée par
Nicholas AYACHE, Xavier PENNEC et Alain TROUVÉ
soutenue le 26 mars 2010

Jury :

Rapporteurs:	Polina GOLLAND	Professeur	MIT
	Jean-François MANGIN	Docteur	CEA - Neurospin
	David MUMFORD	Professeur	Brown University
Examinateurs:	Guido GERIG	Profeseur	University of Utah
	Stéphane MALLAT	Professeur	Ecole Polytechnique
Membre invité:	Laurent YOUNES	Professeur	Johns Hopkins University
Directeurs thèse:	Nicholas AYACHE	Docteur	INRIA - Sophia Antipolis
	Xavier PENNEC	Docteur	INRIA - Sophia Antipolis
	Alain TROUVÉ	Professeur	ENS Cachan

Abstract

This thesis is about the definition, the implementation and the evaluation of statistical models of variability of curves and surfaces based on currents in the context of Computational Anatomy. Currents were recently introduced in medical imaging in order to define a metric between curves and surfaces which does not assume point correspondence between structures. This metric was used to drive the registration of anatomical data. In this thesis, we propose to extend this tool to analyze the variability of anatomical structures via the inference of generative statistical models. Besides the definition and discussion of these models, we provide a numerical framework to deal efficiently with their estimation. Several applications on real anatomical database in brain and cardiac imaging tend to show the generality and relevance of the approach.

In the first part of the manuscript, we extend the computational framework of currents by introducing new numerical tools for approximation and compression purposes. First, a rigorous discretization framework based on linearly spaced grids is provided: it enables to give finite-dimensional projection of currents which converges to the initial continuous representation as the grids become finer. This leads to a generic way to derive robust and efficient algorithms on currents, while controlling the numerical precision. This gives for instance a more stable numerical implementation of the registration algorithm of currents. Then, we define an approximation algorithm which gives a sparse representation of any currents at any desired accuracy via the search of an adapted basis for currents decomposition. This sparse representation is of great interest to compress large sets of anatomical data and to give interpretable representation of statistics on such data sets.

In the second part, we define an original statistical model which considers a set of curves or surfaces as the result of random deformations of an unknown template plus random residual perturbations in the space of currents. The inference of such models on anatomical data enables to decompose the variability into a geometrical part (captured by diffeomorphisms) and a "texture" part (captured by the residual currents). This approach allows us to address three anatomical problems: first, the analysis of variability of a set of sulcal lines is used to describe the variability of the cortex surface, second, the inference of the model on set of white matter fiber bundles shows that both the geometrical part and the texture part may contain relevant anatomical information and, third, the variability analysis is used in a clinical context for the prediction of the remodeling of the right ventricle of the heart in patients suffering from Tetralogy of Fallot.

In the third part, we define statistical models for shape evolution. First, we define a spatiotemporal registration scheme which maps the sets of longitudinal data of two subjects. This registration does not only account for the morphological differences between subjects but also for the difference in terms of speed of evolution. Then, we propose a statistical model which jointly estimates a mean scenario of evolution from a set of longitudinal data along with its spatiotemporal variability in the population. This four-dimensional analysis opens up new possibilities for characterizing pathologies in terms of variations of the growth process of anatomical structures.

Résumé

Le but de cette thèse est de définir, implémenter et évaluer des modèles statistiques de variabilité de courbes et de surfaces basés sur des courants en anatomie numérique. Les courants ont été récemment introduits en imagerie médicale dans le but de définir une métrique entre courbes et surfaces qui ne dépend pas de correspondance de points entre les structures. Cette métrique a été utilisée pour guider le recalage de données anatomiques. Dans cette thèse, nous proposons d'étendre cet outil pour analyser la variabilité de structures anatomiques grâce à l'inférence de modèles statistiques génératifs. Outre la définition et la discussion de tels modèles, nous proposons un cadre numérique pour les estimer efficacement. Plusieurs applications en imagerie cérébrale et cardiaque tendent à montrer la généralité et la pertinence de cette approche.

Dans la première partie, nous introduisons de nouveaux outils numériques d'approximation et de compression des courants. Tout d'abord, un cadre rigoureux de discrétisation basé sur des grilles régulières est proposé: il définit des projections en dimension finie des courants qui convergent vers le courant initial quand la grille devient plus fine. Cela permet de définir de manière générique des algorithmes robustes et efficaces pour traiter les courants, avec un contrôle de la précision numérique. En particulier, cela donne une implémentation plus stable de l'algorithme de recalage de courants. Enfin, nous définissons une méthode d'approximation qui calcule une représentation éparse d'un courant à n'importe quelle précision grâce à la recherche d'une base adaptée au signal. Cette représentation éparse est d'un grand intérêt pour compresser de grands ensembles de données anatomiques et pour interpréter les statistiques sur de tels ensembles.

Dans la deuxième partie, nous définissons un modèle statistique original qui considère un ensemble de courbes ou de surfaces comme le résultat de déformations aléatoires d'une forme prototype inconnue plus des perturbations résiduelles aléatoires dans l'espace des courants. L'inférence de tels modèles sur des données anatomiques décompose la variabilité en une partie géométrique (capturée par des difféomorphismes) et une partie de "texture" (capturée par les courants résiduels). Cette approche nous permet de traiter trois problèmes anatomiques: d'abord l'analyse de la variabilité d'un ensemble de lignes sulcales est utilisée pour décrire la variabilité de la surface corticale, ensuite l'inférence du modèle sur un ensemble de faisceaux de fibres de la matière blanche montre qu'à la fois la partie géométrique et la texture peuvent contenir de l'information anatomiquement pertinente et enfin l'analyse de la variabilité est utilisée dans un contexte clinique pour la prédiction de la croissance du ventricule droit du cœur chez des patients atteints de la Tétralogie de Fallot.

Dans la troisième partie, nous définissons des modèles statistiques pour l'évolution de formes. Nous proposons d'abord une méthode de recalage spatio-temporel qui met en correspondance les ensembles de données longitudinales de deux sujets. Ce recalage prend en compte à la fois les différences morphologiques entre les sujets et la différence en terme de vitesse d'évolution. Nous proposons ensuite un modèle statistique qui estime conjointement un scénario moyen d'évolution à partir d'un ensemble de données longitudinales et sa variabilité spatio-temporelle dans la population. Cette analyse ouvre de nouvelles perspectives pour caractériser des pathologies par une différence de vitesse de dévelopement des organes.

Acknowledgments

I am extremely grateful to my three advisors who had guided me for the last four years. On the one hand, Nicholas Ayache and Xavier Pennec introduced me to the medical imaging community and showed me how mathematical tools can be used to address cutting-edge problems in this exciting field. On the other hand, Alain Trouvé shared with me its deep scientific vision and showed me how one can transform the most innocuous problems or results into an endless source of questions and of new methodological developments. I thank them for making me a young scientist.

I would like also to thank INRIA for offering me outstanding working conditions for these last years and ENS Cachan for hosting me during my regular visits at Cachan. I thank their respective leaders, Nicholas Ayache and Alain Trouvé for having made possible this fruitful collaboration between both teams.

I warmly thank Polina Golland, David Mumford and Jean-François Mangin for the precious time they dedicated to review the manuscript. Their positive feedbacks are invaluable. I am thankful to Stéphane Mallat who kindly accepted to chair the defense committee, Guido Gerig who accepted to come in Nice, before bringing me back to Salt Lake City and Laurent Younes who kindly accepted to take advantage of his stay in Paris to attend my defense.

I owe a great amount of gratitude to all the people I had the chance to closely work with: José Braga, Pierre Fillard, Joan Glaunès, Vladlena Gorbunova, Tommaso Mansi, and Paul Thompson. Without their crucial input, this manuscript would have been very different.

At this precise moment when my time as PhD candidate comes to an end, my thoughts go to Polina Golland, Guido Gerig and Michael Miller. The kind interest they showed for my work during the years 2008 and 2009 played an important role in my decision to start an academic career. Their encouragements also are a strong source of motivation for the future.

I am very thankful to all the people who went along with me during these years. In particular, I would like to thank Pierre Fillard for his friendship, Olivier Clatz for our endless discussions, Alexandre Gramfort for all the time spent in my car between Nice and Sophia, Demian Wasserman for introducing me to tango, Ender Konukoglu, Jean-Marc Peyrat, Nicolas Toussaint, Tom Vercauteren, Sylvain Vallaghé, François-Xavier Vialard, Thomas Yeo, Nicolas Charon and all the lab-mates at Cachan and at INRIA for the good time we spent together.

Above all, I thank Stéphanie for her constant support, patience and love.

"And while I have sought to show the naturalist how a few mathematical concepts and dynamical principles may help and guide him, I have tried to show the mathematician a field for his labour - a field which few have entered and no man has explored. Here may be found homely problems, such as often tax the highest skill of the mathematician, and reward his ingenuity all the more for their trivial associations and outward semblance of simplicity."

D'Arcy W. Thompson, On Growth and Form (1917)

"Et tandis que je me suis efforcé de montrer au naturaliste combien quelques concepts mathématiques et quelques principes dynamiques pourraient lui venir en aide et le guider, j'ai tenté aussi d'indiquer au mathématicien une nouvelle direction de travail - une voie que peu ont empruntée, et que nul n'a encore explorée. Il se trouvera certainement dans cette voie quelques problèmes modestes, qui mettent souvent à rude épreuve l'habileté du mathématicien, et qui le recompensent d'autant mieux que leur résolution nécessite des associations évidentes et qu'elles revêtent un semblant de simplicité."

D'Arcy W. Thompson, On Growth and Form (1917)
traduction Dominique Teyssié

Contents

Databases

The methodology developed in this thesis has been driven by the analysis of anatomical data and has been used to provide solutions to the scientific problems which motivated the constitution of these data. Six different databases were used: a set of sulcal lines, a set of meshes of subcortical structures of the brain, white matter fiber bundles extracted from Diffusion Tensor Images, a set of fossil endocasts of bonobos, meshes of the right ventricle of the heart from patients suffering from Tetralogy of Fallot and CT images of the lungs.

We give here the technical description of the databases which will be used throughout the thesis:

- As part of a collaborative project involving the Asclepios-LONI associated team `Brain-Atlas`, we used a dataset of cortical sulcal landmarks (72 per brain) delineated manually by neuroanatomists experts in 34 subjects scanned with 3D MRI (age: 51.8 +/- 6.2 years). Detailed protocol of image acquisition and sulci delineation can be found at `http://www.loni.ucla.edu/ khayashi/Public/medial_surface/`.

- We access a database of deep brain structures (hippocampus, amygdala, putamen, pallidum and caudate for each hemisphere) segmented in 50 autistics, 4 developmental delays and 7 control children scanned at about 2 and 4 years old. Semi automatic, semi manually segmentation of the structures have been performed. See [Hazlett 2005] for a detailed description of the protocol.

- As part of a collaboration with P. Fillard at CEA/Neurospin, we access to a set of fiber bundles segmented from Diffusion Tensors Images. Six brain DTI data sets acquired on a 1.5T GE scanner on healthy volunteers were used. Image dimensions are $128 \times 128 \times 30$, and resolution is $1.8 \times 1.8 \times 4$mm. 25 non-collinear diffusion gradients and a b-value of $1000 s.mm^{-2}$ were used. Fiber tractography was performed using MedINRIA [1], which includes a robust tensor estimation and a streamline tractography algorithm using log-Euclidean tensor interpolation [Fillard 2007b]. Manual segmentation of five fiber bundles was done: the entire corpus callosum, the corticospinal and the corticobulbar tracts, and the left and right arcuate fasciculi.

- As part of the collaborative project "INRIA-ARC 3D-Morphine", we access a database of 60 fossil endocasts of *Pan paniscus* (bonobos) and 59 endocasts of *Pan troglodytes* (chimpanzees). The original endocasts are from the collection of "Musée de l'Afrique centrale" in Tervuren, Belgium (conservator: E. Gilissen). They have been scanned by J.Braga using a Siemens Somatom Esprit Spiral CT, with slide thickness between 0.33 and 0.50 mm.

- Joint work with T. Mansi in Chapter 8 involved images of the heart of patients suffering from Tetralogy of Fallot acquired in the framework of the European IP-project Health-e-child. Steady-State Free Precession cine-MRI of the heart were acquired with

[1] http://www-sop.inria.fr/asclepios/software/MedINRIA/

a 1.5T MR scanner (Avanto, Siemens). Images were acquired in the short-axis view covering entirely both ventricles (10-15 slices; isotropic in-plane resolution: 1.1x1.1mm to 1.7x1.7mm; slice thickness: 6-10mm; 25-40 phases). Images were made isotropic through tri-cubic resampling. Segmentation of the ventricle has been performed as explained in 8.3.1.

- Joint work with V. Gorbunova in Chapter 4 involved publicly available CT images of the lungs [Castillo 2009]. Segmentation of blood vessels and surface of the lungs have been performed as explained in Section 4.5.

We would like to warmly thank P. Thompson (sulcal lines), G. Gerig and M. Styner (deep brain structures), P. Fillard (white matter fiber bundles) and J. Braga (fossil endocasts of bonobos) for providing us with the data. This work would not have been possible without their precious help and support.

Notations

Mathematical symbols

- $\Lambda^m\mathbb{R}^d$: m-th exterior power of $\mathbb{R}^d$, i.e. set of m-vectors in $\mathbb{R}^d$ (Def. A.2)
- $(\Lambda^m\mathbb{R}^d)^*$: space of m-forms on $\mathbb{R}^d$ (Def. A.4)
- $\mathcal{C}^0(\mathbb{R}^d, (\Lambda^m\mathbb{R}^d)^*)$: space of differential m-forms on $\mathbb{R}^d$ which are continuous and tend to 0 at infinity (Def. A.6).
- $d_x\phi$: Jacobian matrix of the diffeomorphism ϕ at point x.
- $|d_x\phi|$: (signed) determinant of the Jacobian matrix of the diffeomorphism ϕ
- ϕ^*: pull-back action of the diffeomorphism ϕ on the scalar or vector field (Def. A.11)
- ϕ_*: push-forward action of the diffeomorphism ϕ on the space of currents (Def. 1.15)
- $\phi_*^\dagger$: adjoint push-forward action on the space of currents (Eq. (5.3.21))
- $|v|$: if v is a scalar: absolute value of v, if $v \in \mathbb{R}^d$: Euclidean norm of v
- $\langle ., . \rangle_H$: inner-product on the Hilbert space H
- $\|.\|_H$: norm of the Hilbert space H
- $u \times v$: cross-product between 3D vectors u, v
- $\mathcal{L}_W$: canonical isometric map between the Hilbert space W and its dual space W^* (Def. B.8)
- $\|.\|_\infty$: for a scalar or vector field, its supremum norm. For a current, the supremum norm of its representation in terms of vector fields (Eq. (1.5.20)).
- I_n: the n-by-n identity matrix.
- $A \star B$: discrete nD-convolution between two n-by-n matrices A and B.
- $M(T)$: the mass of a current T (i.e. the mass-norm) (Def. 1.3)
- $\lambda(T)$: the Lebesgue measure of the sub-manifold T.

Vocabulary

δ_x^α is called a Dirac delta current. The couple (x, α) is called a momentum. x is the location of the momentum. α is the coefficient of the momentum. A linear combination of Dirac delta current is denoted: $\sum_{x_i} \delta_{x_i}^{\alpha_i}$. The set of points x_i is called the support of this linear combination.

The word "shape" in this thesis has the very general sense of any kind of geometrical data (set of points, curves, surfaces, volumes, etc.).

Introduction

The not so random variability of living organisms

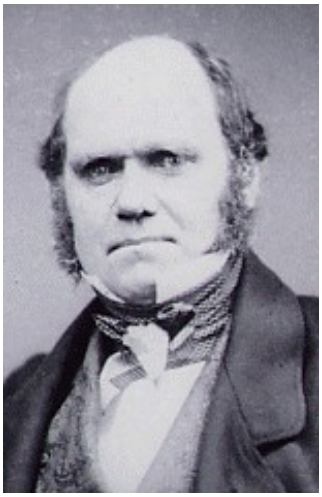


Figure 1: Charles Darwin (1809-1882) is among the first ones to imagine that the observed anatomies result from the selection of random variations of a common anatomy.

Charles Darwin was born two hundreds years ago. He dedicated most of his scientific activity to set up a theory which can explain the anatomical variability of the living organisms. Since Antiquity, one has observed that, beneath the incredible variety of the forms of life, large groups of animals share the same inner framework. For instance, most of the vertebrates share a similar skeleton with two posterior and two anterior limbs. As shown in Figure 2, the anterior limb has an homologous structure in many different species. In the $18th$ century, Linné (1707-1778) and Buffon (1707-1788) define species as a group of individuals which share common anatomical features and can reproduce within the group. They are the first ones to introduce the idea that the observed anatomies may be considered as variations ("degenerations" according to Buffon) of a common prototype anatomy.

Darwin went one step further by considering these variations in a temporal perspective. The common prototype anatomy is the one of a common ancestor and the variability results from random innovations over ages and a selection of these innovations according to competitive advantages. The discovery of genetics strengthens his idea of random innovations introduced during reproduction. This builds a theory, namely a set of simple hypotheses, which can explain both the large variability of the anatomy of living organisms and the presence of a structure in this randomness.

D'Arcy Thompson (1860-1948) introduced the idea that the possible innovations which can be introduced during evolution are constrained by the inherent structure of the biological tissues and the physical forces exerted to them [Thompson 1917]. This idea is strengthen by recent experiments which manage to drive the growth of a plant by controlling the external pressure or by modifying the physical forces exerted to the cells [Mulder 2008, Hamant 2008]. As a consequence, the number of degrees of freedom for introducing new features during evolution might be much smaller than what is expected by the current stage of the theory of evolution. For instance, the neck could only elongate or shrink, any other deformations being unlikely to appear from a biological point of view. In any case, the conclusion is the same: the huge variability observed in nature is not pure randomness but is structured by underlying rules. This holds whether the random process is intrinsically constrained or whether a selection principle makes eventually the actual variability structured.

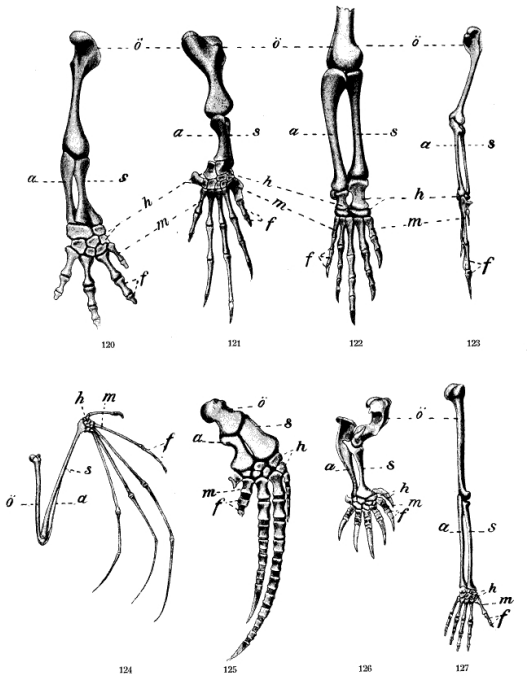


Figure 2: Comparative study of the skeleton of the arm of several vertebrates, by Wilhelm Leche (1909). Beneath the apparent huge variability of the skeletons, one can find homologous anatomical structures across species.

Historically, the classification of species was supported mostly by the analysis of skeletons. However, the presence of anatomical invariants is not limited to the skeleton. We can find homologous structures in most of the organs and anatomical structures of the living organisms. For instance, Fig. 3 shows four different brains of healthy humans. The number and the location of the cortical sulci (the folding patterns of the cortex surface) seem to vary a lot from one subject to another. However, some main sulci such as the Sylvian and Calgarine fissure, the central sulcus appear in almost every human.

Analyzing and understanding the variability of anatomical structures is of great interest, especially for humans. Knowing the normal variability of an organ would help to detect abnormal developments of a given subject via large deviations from the normal model. This would help also to find discriminative anatomical features, to divide a population into consistent clusters according to some anatomical characteristics, to discover subtypes of pathologies, etc. Understanding the causes of this variability should offer a way to treat abnormal developments.

Deterministic integrative models to explain the variability

To understand the variability of structures, we can try to find the physical or biological causes of this variability. This requires to build generic models which integrate the biological phenomena which we think are involved in the process. Such models try to explain how every

particular anatomy derives from a common framework. For instance, the "physiome project" is an attempt to build integrative models which can derive the anatomy of macroscopic structures from genetic and proteomic information [Hunter 2003]. A recent morphogenesis hypothesis supposes that the variability of the cortical folding patterns in the adult brain derives from a common architecture of folding locations at the fetal stage [Regis 2005]. In oncology, several models have been proposed to explain the growth of a tumor by the integration of physical and biological interactions between cancerous and healthy cells, like in [Mayneord 1932, Levine 2000] for instance.

However, it not always possible to design such integrative models, either because the number and the complexity of the interactions make the derivation of the model impossible or because we simply lack the physical or biological knowledge to build such models. This is illustrated by the example of the brain. On the one hand, one has a pretty good understanding of how an individual neuron transports the action potential via the integrate-and-fire model of [Hodgkin 1952] for instance. On the other hand, we have almost no idea on how to model the interactions between every neuron of the brain in order to explain brain functions such as memory for instance. We are also far from having a global model of the physical interactions between neurons, which can explain the anatomy of the brain.

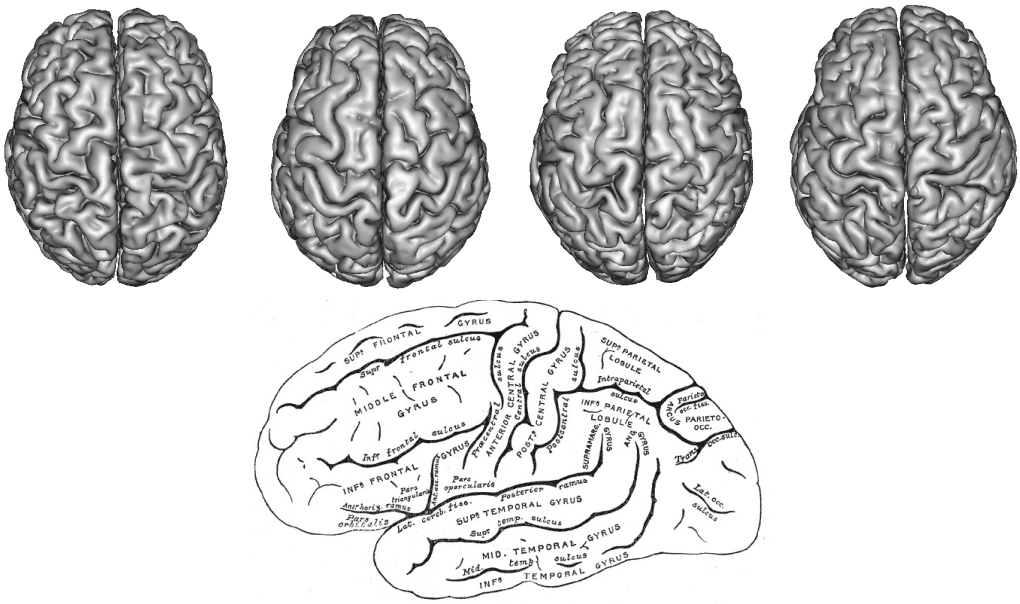


Figure 3: Top: four different brains of healthy humans. The folding patterns of the cortex surface vary a lot across individuals, although some major sulci seems to build a common architecture of the cortex as illustrated in the bottom figure (top images courtesy of J.-F. Mangin, bottom image from Gray's Anatomy of the Human Body)

Stochastic generative models

To validate the biological assumptions on which a model is based, we need to measure how well this model can explain the experimental observations. For this purpose, one personalizes the parameters of the model (such as size, weight or genes for instance) to a particular patient and compare the prediction of the model with the anatomy of this patient. Such models establish therefore a deterministic link between the observed anatomies and a set of physiological parameters. The variability of observations is due to various initial conditions.

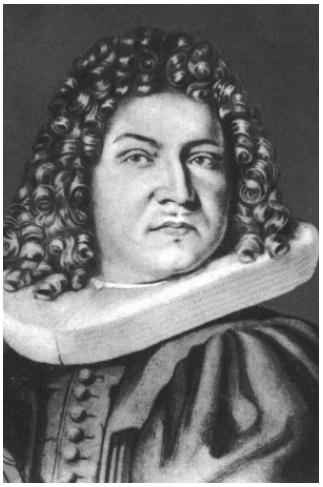


Figure 4: Jacques Bernoulli (1654-1705), one of the father of the theory of probabilities, which gives mathematical tools to find the structures of randomness.

When integrative models are not available, one can replace such deterministic approaches by stochastic processes. Such an approach is common across various fields of science, see [Grenander 1994] for a general stochastic pattern theory or [Mumford 2006] for a focus on visual perception, for instance. Here, the idea is to see the observations as some instances of a random process: we consider that they derive randomly from a common prototype anatomy. Such an approach probes the data to answer the questions: what is invariant across individuals? what is variable? how is it variable? The first question is answered by the estimation of the prototype anatomy, the next two questions by the estimation of the law of the random process. Such "bottom-up" approaches can be set up even in the absence of biological knowledge about the possible causes of the variability. As such, they are complementary with the "top-down" integrative approaches.

The inference of such stochastic models extracts automatically invariant geometrical structures from anatomical data. These anatomical findings can be used, in turn, to drive the search for relevant biological hypotheses which can explain the structure of the stochastic model. Indeed, having a clear representation of the similarities and differences of highly variable data should help to find the causes of this variability. Moreover, we can think of integrating biological variables into the statistical framework and therefore give a way to evaluate the relevance of some biological assumptions. The two approaches, deterministic biological modeling and stochastic geometrical modeling, should be investigated in parallel. Coming back and forth between the data and the biological hypotheses should enable firstly to accurately describe and measure the variability of the anatomy and, secondly, to understand the biological causes of this variability.

The purpose of this thesis is to design a generic statistical framework which can describe and measure the variability of a set of anatomical data. In this perspective, *generative* statistical models have to play a tremendous role over other statistical approaches. Descriptive approaches test the significance of some hypotheses: this is adapted when one knows beforehand what we are looking for. By contrast, generative models do not make strong assumptions on the variability we are looking for. They infer a random process which can

generate new data which can be compared to the experimental observations. This can be used to predict the anatomy of new subjects, thus giving a relevant way to evaluate the model. More importantly, this offers a way to include biological assumptions in the variability model, a key feature toward the search of the biological causes of the observed phenomenon.

To define such generative statistical models, we need first to look carefully at the data we want to analyze. This will drive the definition of a relevant mathematical and statistical framework to analyze shape[2] variations.

The starting point: the data

The evaluation of the theory of evolution relies on very small amount of data: for example, few fossil skeletons, which are often incomplete and which have been deformed during fossilization (taphonomic deformations). This makes the quantitative evaluation of the theory particularly difficult. Working with organs or anatomical structures of still living organisms makes the evaluation of the models in Computational Anatomy much easier. Indeed, there is an increasing use of in-vivo imaging devices and modalities (Magnetic Resonance Imaging (MRI), Diffusion Tensor Imaging (DTI), Ultrasound imaging, Positron Emission Tomography (PET), etc.) both in the clinical context and for neuroscience studies. This makes available larger and larger database of images which can be used to understand the normal anatomy as well as the pathological cases.

The raw experimental data are thus 3D grey-level images of patients or control subjects. Several methods in Computational Anatomy focus on the 3D images. This has the advantage to base the statistical estimation directly on the experimental data. However, the images offer a view of the anatomy through a predefined window, whereas one is interested on some particular anatomical structures *contained* into this field of view. This is the reason why we believe that we should *select* the relevant information in the images before applying the statistical analysis on these extracted data.

However, this approach raises an important methodological question: how can we reliably extract this information if we have no prior about the normal variability of this information in the population? Using segmentation from anatomical experts is not a completely satisfactory answer: first the increasing number of available images prevents us from a systematic manual segmentation of structures and second this extraction is based on an implicit and uncontrollable prior, which is what the expert thinks the structure should look like in images. The automatic segmentation of structures is necessarily based on some priors such as the regularity of the contour of the structure for instance. Of course, these priors bias the following statistical analysis. But, this statistical analysis helps also to better understand how the structure varies in the population. It can be used then to adapt the segmentation priors to take into account the new anatomical knowledge. Defining such a feedback loop in the segmentation process should increase both the reliability of the segmentation and the power of the statistical estimation. However, this approach is not addressed

[2] In this thesis, the word "shape" denotes any kind of geometrical data like curves, surfaces, volumes or point sets for instance.

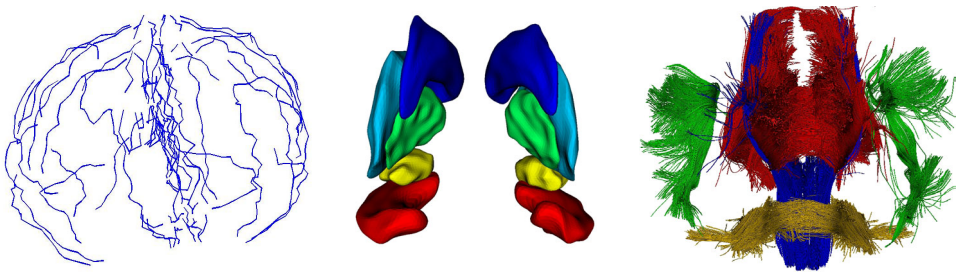


Figure 5: Three examples of anatomical data. Left: trace of sulcal lines drawn on the convex hull of the cortex. Middle: meshes of deep brain structures (hippocampus, amygdala, putamen, pallidum and caudate for each hemisphere). Right: Five white matter fiber bundles (corpus callosum, corticospinal and corticobulbar tracts, left and right arcuate fasciculi). Each bundle consists of thousands of curves.

in this thesis. Our starting point is anatomical structures, which have been either manually or automatically extracted from images, and we focus on the direct way: the statistical estimation of the variability from these data.

The anatomical data which we want to deal with may have very various forms. Figure 5 shows three typical examples of data for which we need to design statistical models. The data may be the contours of the anatomical structures whose representation may be as structured as a surface mesh or as unstructured as a set of points supposed to belong to the surface. The data may be also volumetric: a volume mesh or a set of points inside the structure. However, the anatomical structures of interest are not limited to volumes or contours of volumes. For instance the sulci of the brain may be represented as a surface sheet delimited by the gyri as in [Mangin 2004b]: the surface is not necessarily closed. According to the theory of sulcal roots [Regis 2005], it is interesting to focus also on the curve drawn on the cortex surface at the bottom of the sulci. In this case, the anatomical data is a set of curves whose topologies are arbitrary: a sulcal curve may have branching or be cut in several parts. In this thesis, we will also deal with white matter fiber bundles (a representation of the collection of neurons which connects two functional areas of the brain). The extraction and selection of fiber bundles from diffusion tensor images lead to a large set of individual curves gathered into consistent clusters. The number of curves as well as the connectivity of the curves within each bundle is highly variable according to the subject and the extraction method.

Defining generic statistical models for such data requires to answer two fundamental questions: (1) how can we define a generic mathematical framework which can deal with so different kind of data? (2) how can we measure the geometrical variations of these data which would be robust enough to the change of representation of the structure (change of topology, non-homologous points across subjects, different number of curves within a bundle, etc.)? This includes, for instance, the comparison between a sulcal curve which has a branching with the same sulcal curve in another subject which has no branching, or the comparison between two fiber bundles (a reproducible anatomical structure across subject)

which have a very different number of curves.

Which mathematical model for shapes?

Several mathematical frameworks have been proposed to deal with geometrical data, mostly in the field of Computer Vision. Popular methods used level-set representations for closed contours: curves in 2D and surfaces in 3D. Such approaches use geometrical constraints (length, area, volume, curvature, etc.) as priors, like in [Leventon 2000, Leventon 2003, Pardo 2004, Cremers 2006] for instance. However, it is difficult to figure out how to automatically learn such priors from typical data sets. Indeed, curves in this setting are embedded into a space which is not a vector space and on which no natural metric is defined. This makes the definition of statistics particularly difficult, for example, there is no explicit and simple way to compute a mean.

Other models are based on point distribution and therefore assume point-correspondences between structures like in [Cootes 1995, Cootes 2008]. Such approaches can lead to statistical models as in [Twining 2005] for instance. In medical imaging, however, samples of curves or surfaces segmented from MR images, do not generally correspond from one structure to another. For instance, it is not possible to assume that points of two fiber bundles (such as the ones in Fig. 5) are homologous across subjects. Assuming such correspondence introduces therefore a non-realistic geometrical prior, which eventually may lead to an important bias in the statistical estimations. Improvements of point distribution models were proposed to relax the constraint of point-correspondences, for instance Minimum Description Length approach [Marsland 2008], statistical versions of the Iterative Closest Points algorithm [Hufnagel 2008] or fuzzy correspondences approaches [Chui 2003]. These techniques, though, are still built on the idea of point correspondences. Moreover, they deal only with collection of points and do not take into account normals or tangents of the structures.

Alternatively, Medial Axis representations have been proposed to build statistical models on surfaces [Pizer 2003, Fletcher 2004]. As a consequence, these methods are only available for surfaces, they are not robust to changes of connectivity of the structures and are quite sensitive to variations of the segmentation algorithm. In [Charpiat 2005], dissimilarities between compact subset of $\mathbb{R}^2$ or $\mathbb{R}^3$ (seen as characteristic functions) are measured by the Hausdorff distance. In [Mio 2007, Joshi 2007], a metric between closed contours is provided which is invariant under a certain group of transformations (rigid transformations and/or scaling) on the shapes. The definition of general Riemannian metrics on the space of closed planar curves, which are invariant under re-parameterization of the curves, are given in [Michor 2006, Michor 2007].

In contrast to these methods, the framework based on currents, as introduced in [Vaillant 2005, Glaunès 2005], is not limited to a particular kind of data. Indeed, it provides a unifying framework to process any sets of points, curves and surfaces or mix of them. No hypothesis on the topology of the shapes is assumed. In particular, it is robust to the change of connectivity of the structures. Moreover, it is weakly sensitive to the sampling of shapes. The main advantage of this stetting is that shapes are embedded into

a vector space provided with an inner-product. This gives a mathematically well-grounded framework to compute statistics like mean and principal modes of a population of shapes, without selecting a set of arbitrary features like volume, length or curvature for instance. Eventually, the metric on currents does not assume any kind of correspondences between structures.

For these reasons, we choose to root our statistical models into the framework of currents, since it seems particularly adapted to process the kind of anatomical data such as the ones in Figure 5.

Which statistical model for shape variations?

Once one has defined a model for shapes, one must find a way to compare them. There are mostly two different perspectives for measuring shape variations. In the perspective of Kendall's work [Kendall 1989], a metric is directly defined in the space of shapes. In this original work, shapes are first aligned using similarity transformations and then a metric is derived in the manifold of the set of N-points identified by similarity (quotient space). In this setting, a shape is defined by invariance properties: the shape is what remained after registration, once a class of deformations has been defined (rigid transformation, similarity, diffeomorphism, etc.). Shape comparison is performed directly in the shape space, once the deformations between different instances have been discounted.

By contrast, following the ideas of D'Arcy Thompson [Thompson 1917], later rooted in a mathematical framework by Grenander [Grenander 1994], the distance between shapes is measured by the "quantity of" deformation which is needed to warp one shape onto another. In this setting the variable of interest is precisely the (elastic) deformation between shapes, whereas the remaining part is considered as uninformative noise. This framework has the advantage to be compatible to many shape representations, provided that the deformation of a shape can still be represented in the same setting.

As we shall discuss in Chapter 5, none of the approaches is completely satisfactory, since they assume that the interesting part of the variability is either on the deformation or on the residual which remain after registration. The statistical model proposed in this thesis is an attempt to bridge the gap between the two points of view. We propose to consider a collection of shapes as a random deformation of a template shape plus a random perturbation in the shape space (here the space of currents). Then, the consistent estimation of both the deformations and the residual perturbations leads to the decomposition of the variability into two terms. The variability which is captured by the deformations is mostly geometric and describe torque, elongation, shrinking effects, etc. The variability which is captured by the residual describes topology changes, matter creation or deletion, etc. This approach has the advantage to take into account the whole variability without making strong assumptions on the kind of variability one is looking for.

This statistical model focuses on sets of 3D shapes. In some sense, it can be seen as a systematic version of the approach followed by Buffon in the 18th century to classify living organisms. After Darwin, we need to put this analysis of variability into a temporal perspective. Indeed, to understand the differences between species, it is better to compare

the evolution of the anatomy of the species over ages rather than their anatomical difference at a given time point. Similarly, the pathology of an organ might not be characterized by the anatomical differences observed at a given age, since it may affect the development of the organ more than its shape. Therefore, we propose to extend the previous statistical models of shape to statistical models of shape evolution. The prototype anatomy is replaced by a prototype scenario of evolution. The geometrical variability is replaced by a spatiotemporal variability of the growth scenario. This variability describes both the morphological changes and variations of the speed of evolution across individuals.

Manuscript overview

The purpose of this thesis is therefore to use the framework of currents to define generative statistical models of shapes. The manuscript is divided into three parts. In the first part, we introduce the currents as a unifying tool to model shapes. Although currents have been successfully applied for the registration of pair of anatomical data [Vaillant 2005, Glaunès 2008], a clear bottleneck appears when dealing with group-wise statistics. Therefore, we need to develop a new numerical framework to define robust and numerically stable algorithms to process collection of currents.

In the second part, we focus on the statistical modeling to measure the variability of a set of anatomical shapes. This model decomposes the variability into a geometrical part captured by the deformations and an "texture" part captured by the residual, as an attempt to conciliate Kendall's and Grenander's approach. Then, this model is used first to estimate the variability of the cortex surface knowing the positions of several sulcal lines, second to describe the anatomical variability of white matter fiber bundles both in terms of geometrical deformations and in terms of variations of the density of fibers and third to predict the remodeling of the heart of patients suffering from repaired Tetralogy of Fallot.

In the third part, we extend this statistical model of shapes to a statistical model of shape evolution. The variability of longitudinal data is described in terms of morphological changes and of change of speed of evolution. This offers a way to detect possible developmental delays in a population and to characterize pathologies by its impact on the development of anatomical structures.

In details, each part is divided as follows:

Part I: Currents for modeling curves and surfaces

- **Chapter 1:** We introduce the theory of currents in the perspective of [Glaunès 2005]. We propose a unifying framework to process various kind of geometrical data like point sets, curves, surfaces and volumes for instance, which may be associated with some attributes. We show also how the metric on currents addresses the problem of point correspondence in Computational Anatomy.

- **Chapter 2:** We propose finite-dimensional approximation spaces for currents via linearly spaced grids. This provides a numerical framework to compute the usual operations on currents via sampling, partial volume interpolation and Fast Fourier Transforms. We prove that the approximated currents converge to the initial current

as the grid becomes finer and provide speed of convergence. This gives a generic way to derive fast and numerically stable algorithms dealing with currents, in particular the matching pursuit algorithm of Chapter 3, the registration of currents of Chapter 4 and atlas estimation of Chapters 5 and 9.

- **Chapter 3:** We introduce a matching pursuit algorithm for currents, which defines a sparse approximation of currents at any desired accuracy via the search of an adapted basis for currents decomposition. This allows us, in particular, to have a numerically stable representation of statistics on currents. Indeed, the addition in the space of currents represents the union of curves or surfaces. Therefore, the mean or modes of a collection of shapes has an increasing complexity as the size of the database is increasing, whereas the statistics converge to their asymptotic values. The sparse approximation of currents enables to approximate these statistics with a rather constant complexity. It offers a tractable way to use such statistics in more general algorithms, like for the estimation of the statistical model of Chapter 5 for instance.

- **Chapter 4:** This chapter deals with the diffeomorphic deformation of currents. First, we recall how currents may be used to register geometrical data in the framework of [Vaillant 2005]. The numerical framework set up in Chapter 2 is used to give a new numerical implementation of the registration algorithm. Then, we extend this registration framework to the temporal regression of a set of time-indexed shapes. Eventually we show two original applications: (1) the joint registration of both curves and surfaces extracted from CT-scans of the lung, which is shown to compare with a registration based on image intensity and (2) the temporal regression of fossil endocasts of bonobos, which provides a continuous growth scenario from data sparsely distributed in time.

Part II: Statistical models of sets of curves and surfaces

- **Chapter 5:** We introduce and discuss statistical models on geometrical data modeled as currents. The data are seen as the random deformation of an unknown current plus an additive Gaussian noise in the space of currents. The inference of this model leads to the estimation of a template shape (a current) and the variability of this template shape in the population. The variability is decomposed into two terms: a geometrical part captured by the deformations and a "texture" part captured by the residual currents. A simulated example shows how this model can be used to find discriminative features between two sets of shapes.

- **Chapter 6:** In this chapter, we address the problem of estimating the variability of the cortex surface from the positions of sulcal landmarks. We take advantage here of the registration framework based on currents to integrate consistently a set of geometrical constraints. In this work, the template is given and statistics are performed on template-to-subjects deformations. The estimated measure of variability is compared with a more pragmatic method applied to the same database. *This chapter can be read independently of the rest of the thesis.*

- **Chapter 7:** In this chapter, we describe the anatomical variability of a set of white matter fiber bundles extracted from diffusion tensor images. Following Chapter 5, a set of prototype fiber bundles is estimated along with its variability in the population. Statistics on the diffeomorphisms and the residual currents show that both the geometrical part and the "texture" part of the variability may contain relevant anatomical information. *This chapter can be read independently of the rest of the thesis.*

- **Chapter 8:** This chapter is an attempt to bridge the gap between the estimated geometrical variability and clinical variables. We use our atlas construction method on the right ventricle of the heart segmented in images of patients suffering from repaired Tetralogy of Fallot. Then, we statistically analyze the correlations between the deformations and clinical parameters to provide a personalized scenario of the remodeling of the heart. This scenario is used to predict the future evolution of new patients, thus showing the relevance of the approach. *This chapter can be read independently of the rest of the thesis.*

Part III: Statistical models of shape evolution

- **Chapter 9:** We extend the 3D-statistical models of Chapter 5 to take into account shape evolution along time. First, we define a spatiotemporal registration scheme which aligns two sets of time-indexed anatomical shapes. This registration does not only account for the morphological differences between subjects but also for the difference in terms of the speed of evolution (i.e. it synchronizes the two evolutions). Then, we introduce a statistical model which enables to estimate a mean scenario of evolution and its spatiotemporal variability in a set of longitudinal data. We use this model on a set of deep brain structures segmented in autistics, developmental delayed and control children scanned at 2 and 4 years old. The results tend to show that such pathologies may be characterized by variations of the growth of the anatomical structures rather than by morphological differences at a given age.

Guidelines for reading the thesis

In this thesis, we try to show the links between the mathematical modeling based on currents, its inclusion into statistical models, the derivation of the theory as algorithms and the results of these algorithms on real data. Each processing is one piece of a jigsaw: the assumptions of the one influence the result of the others. One of the main contribution of the thesis is precisely to embed all the processing units into a single consistent framework for which all the underlying hypotheses are explicit. As a result, the global framework is controlled by a small set of parameters and the impact of these parameters on the final result is relatively clear. This also allows us to guarantee the stability of the numerical implementation of the theory and therefore its ability to deal with various kinds of data.

Having said that, the different chapters have been written so that they are as independent as possible. As a consequence, the thesis can be read in several ways, depending if the reader is interested in:

- *The mathematical modeling:* how the mathematical modeling of shapes addresses the targeted applications is mostly discussed in the presentation of currents in Chapter 1 (along with the Appendix A on differential forms and B on RKHS), the presentation of the sparse representation of currents in Chapter 3 (especially Section 3.2 about the orthogonal matching pursuit) and Chapter 4 about registration (in particular Section 4.3 which couples the diffeomorphic deformations and the modeling of currents). Discussions on the model parameters can be found also in Chapter 6 in case of joint registration of a set of sulcal lines.

- *The statistical modeling:* the definition and discussion of the statistical models can be found in the presentation of the sparse representation of currents in Chapter 3 (especially Section 3.4 about the sparse representation of statistics on currents and 3.5 about the deconvolution of noise of currents), in the presentation of the joint statistical model "deformation+texture" in Chapter 5 and its 4D extension in Chapter 9.

- *The computational framework:* to know how the theory can be translated as algorithms, one should focus mainly on the presentation of the discretization framework in Chapter 2 and the approximation of currents in Section 3.3 which are used throughout the thesis to derive numerically stable algorithms. This is the case for the matching pursuit algorithm (Algorithm 2 in Section 3.2.2), the atlas construction (Algorithm 3 and 4 in Section 5.3.3) and its spatiotemporal extension (Algorithm 5 in Section 9.3.3.3). The new optimization scheme for the registration of currents can be found in Section 4.4 (the details of the implementation of the original algorithm can be found in [Glaunès 2005]).

- *The applications in Computational Anatomy:* to know how this theoretical and computational framework can be used to address concrete anatomical problems, the reader should focus on the joint registration of vessels and lung surfaces in Section 4.5, on the temporal regression of fossil endocasts of bonobos in Section 4.6), on the whole three chapters about the estimation of the cortex variability from sulcal lines (Chapter 6), about the description of the anatomical variability of white matter fiber bundles (Chapter 7) and about the prediction of the remodeling of the heart (Chapter 8), which can be read independently of the rest of the thesis, and the section about the use of the spatiotemporal statistical model to detect developmental delay in longitudinal data set (Section 9.4). In addition, a non mathematical introduction of the theory of currents is provided in Section 1.2.

Part I

Currents for modeling curves and surfaces

Chapter 1

Curves and surfaces embedded in a metric space

Contents

In this chapter, we present mathematical objects called "currents", which are used as a model for general geometrical objects. In particular, we show that any sets of curves or surfaces may be embedded in the space of currents. This gives a non-parametric representation of such geometrical objects. These objects inherit from the *metric* defined in the space of currents: a geometric similarity measure between objects is provided, which does not assume any kind of point-correspondence between structures. The space of currents is a *vector space*: any set of geometrical structures may be decomposed into the union of several parts and each part may be weighted separately. This allows us to compare two sets of anatomical structures while adjusting the level of correspondence: correspondence between clusters of curves or between individual curves, for instance. This allows us to use any anatomical knowledge as prior without introducing arbitrary correspondences. The space of currents is also topologically *complete* and therefore allows us to process in the

same framework both discrete geometrical structures and continuous objects seen as the limit of sampled structures. This guarantees the robustness of the framework with respect to different sampling of the data.

Note: The first two sections may be read independently of the rest of the chapter. If you are not interested in the mathematical details of the modeling, you can switch to Chapter 2 after reading Section 1.2.

1.1 The giants on whose shoulders we stand

The idea of currents is deeply rooted into the theory of distributions, as set up by L. Schwartz in the 1940's [Schwartz 1966]. The distributions generalize the concept of function and measure on an open sub-space of $\mathbb{R}^d$. A distribution is characterized by its action on any infinitely differentiable (i.e. smooth) functions with compact support. A function f for instance, is completely determined by the collection of the integrals $\int f\phi$ for any smooth functions ϕ with compact support. Similarly, a measure μ is determined by the collection of $\int \phi d\mu$. More generally, a distribution is a continuous linear operator on the "test space" of the smooth functions with compact support. This idea of seeing an object (a distribution) via its action on a test space allowed to extend the concept of differentiability to non differentiable functions. In particular, this enables to state rigorously that the derivative of the Heaviside function is the Dirac delta distribution defined by $\delta_x(\phi) = \phi(x)$. The theory of distributions plays therefore a crucial role in the theory of partial differential equations and in Fourier analysis.

Distributions extend the concept of scalar functions. Currents is a similar construction but which extends the concept of differential forms on an open subset of $\mathbb{R}^d$ or on a smooth manifold. The theory of homological currents was developed by de Rham, as outlined in [Cartan 1970] and reported by Raoul Bott:

> *"When I met de Rham in 1949 at the Institute in Princeton he was lecturing on the Hodge theory in the context of his "currents". These are the natural extensions to manifolds of the distributions which had been introduced a few years earlier by Laurent Schwartz and of course it is only in this extended setting that both the de Rham theorem and the Hodge theory become especially complete."*

He named these objects "currents" by analogy with electromagnetism. For instance the Faraday's law of induction states that the intensity within a wire loop $\mathcal{C}$ induced by variations of a magnetic field $\mathbf{B}$ is proportional to the variations of the flux of this magnetic field through the surface S delimited by the wire (i.e. the boundary of S is $\mathcal{C}$): $\Phi(\mathbf{B}) = \int_S \mathbf{B}^t \mathbf{n} d\lambda$ (where $\mathbf{n}$ is the normal of the surface). This means that if we measure the intensity of the current within the wire (via the flux $\Phi(\mathbf{B})$) for every possible variations of the magnetic field (created by a moving magnet for instance), then we can retrieve the geometry of the wire. On the contrary, we can fix the magnetic field and move the wire $\mathcal{C}$ in the space. We can retrieve the value of the magnetic field by measuring the intensity of the Eddy currents in the wire for every possible motions of the wire. In this case, the wire is used to probe the magnetic field. In the first case, the magnetic field was used to probe

the geometry of the wire. These two examples show that the current $\mathcal{C}$ and the vector field $\mathbf{B}$ are two "dual" objects.

Initially, de Rham developed currents in order to find algebraic characterization of topological invariants on manifolds, which leads to his famous cohomology groups. However, currents quickly spread far beyond the field of algebraic and differential topology. In particular, they played a key role in the emergence of the "Geometric Measure Theory", pioneered by H. Federer [Federer 1969, Morgan 1987]. This theory tries to extend the measure theory (which leads to the integration theory of Lebesgue) to sub-manifolds, which are usually described by some parameterization (parametric curve or surface for instance). As explained by H. Federer, this should lead to a parameterization-free characterization of sub-manifolds:

> *"(...) one must abandon the idea of describing all the competing surfaces by continuous maps from a single predetermined parameter space. One should rather think of surfaces as m-dimensional mass distributions, with tangent m-vectors attached."*

This seminal vision leads to many theoretical developments and applications in various fields such as image processing and computational geometry. For instance, Wintgen and Zähle [Wintgen 1982] introduced particular currents called "Normal cycles" which generalize for singular objects the unit normal bundle of a smooth manifold. This tool was used to define curvature measures on a large class of geometric objects. Recent results of J.-M. Morvan and D. Cohen-Steiner [Cohen-Steiner 2003a, Cohen-Steiner 2003b] give an upper-bound of the error between the curvature measures of a polyhedron "close to" a sub-manifold and the curvature measures of the sub-manifold itself.

In 2005, J. Glaunès and M. Vaillant introduced the concept of currents in the field of Computational Anatomy [Glaunès 2005, Vaillant 2005]. Their purpose was to give a dissimilarity measure between meshes or polygonal curves which does not assume point-correspondence between structures, a key feature for comparing anatomical structures segmented automatically from Magnetic Resonance Images. They used this dissimilarity metric to drive the deformation of a source object (a set of curves or surfaces) to a target object. They proposed also to use the framework of reproducible kernel Hilbert space (RKHS) to give tractable formula of the metric as well as its derivatives.

In this chapter, we present the currents in the perspective of J. Glaunès' work with an emphasis on these two seminal ideas of currents: (1) currents model geometrical objects via their action on a test space of vector fields and (2) the modeling based on currents consider objects as a mass distribution without any kind of parameterization which would give a particular label to each point. Moreover, the topological properties of the space of currents enable to embed in a single framework smooth geometrical objects (on which the usual metric properties are naturally defined) and their discrete representation (the only objects to be accessible from a computational point of view).

1.2 An overview of currents in Computational Anatomy

The purpose of this section is to give a concise introductions of the currents without going into too much mathematical details. We introduce the concept of currents and the

main properties which are useful in the context of Computational Anatomy. The rigorous definitions and the proofs of the claimed properties will be given in the next section (see Section 1.3).

1.2.1 Currents: an object which integrates vector fields

Curves and surfaces tested on vector fields

As emphasized in the previous section, the main idea of currents is to probe shapes by vector fields. The word "shape" here is a generic word, which denotes a set of piecewise smooth curves or piecewise smooth surfaces (which will be modeled as rectifiable subsets of $\mathbb{R}^2$ and $\mathbb{R}^3$ in Section 1.3). As a consequence, a shape can be represented by an infinite set of oriented points: the set of all normals of the surfaces (resp. tangents of the curves). Such oriented points are called "momenta" in the sequel. In the discrete setting, shapes are given as meshes (resp. polygonal lines): the direction of the normals (resp. tangents) is constant over each mesh cell (reps. each segment).

Given ω a square integrable 3D vector field (a mapping from $\mathbb{R}^3$ to $\mathbb{R}^3$), any set of piecewise smooth surfaces S integrates ω thanks to the flux equation:

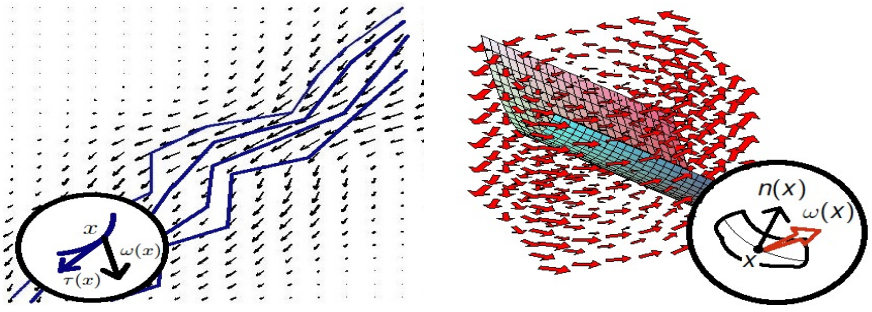
$$S(\omega) = \int_S \omega(x)^t n(x) d\lambda(x), \tag{1.2.1}$$

where $n(x)$ is the unit normal of the surface at point x and $d\lambda$ the Lebesgue measure on the surface. This equation computes the flux of the vector field ω through the surface S.

Similarly, any set of piecewise smooth curves L integrates a vector field ω thanks to the path-integral:

$$L(\omega) = \int_L \omega(x)^t \tau(x) d\lambda(x), \tag{1.2.2}$$

where τ is the tangent of the curve at point x and $d\lambda$ the Lebesgue measure on the curve. This equation computes the flux of the field of tangents through the equipotential surfaces of ω.



$$L(\omega) = \sum_{i=1\ldots4} \int_{L_i} \omega(x)^t \tau(x) d\lambda(x) \qquad S(\omega) = \int_S \omega(x)^t n(x) d\lambda(x)$$

Figure 1.1: In the framework of currents, curves and surfaces are tested on vector fields via the path-integral of the vector field along the curves (left) or via the flux-integral of the vector field through the surface (right). When we know the result of this operation for every vector fields possible, we get a characterization of the geometrical object.

The idea of currents is to characterize a shape by the collection of the real numbers $S(\omega)$ or $L(\omega)$ (given in Eq. (1.2.1) and (1.2.2) respectively) for all possible vector fields ω. For this purpose, we need to make precise the idea of 'all possible vector fields' by defining a proper test space of vector field: W.

Remark 1.1. We distinguish here between the case of curves and the case of surfaces. In Section 1.3, we will define a unifying framework for modeling curves, surfaces, unconnected point sets and volumes. In this framework, vector fields are replaced by differential forms. As shown in Appendix A, it appears that differential forms can be identified to vector fields when modeling curves in 2D or 3D and surfaces in 3D, thus leading to Eq. (1.2.1) and (1.2.2). These two particular cases are the one of most interest for the applications in Computational Anatomy. □

Test space of vector fields

We choose for the test space W the set of the convolutions between any square integrable vector fields and a smoothing kernel. Formally, W is defined as a Reproducing Kernel Hilbert Space (see Appendix B). The kernel plays the role of the transfer function of a low-pass filter. It enables to map every square integrable vector field to a smooth one. As a consequence, W cannot contain vector fields with too high spatial frequencies. In our applications, we will use a Gaussian kernel: $K^W(x,y) = \exp(-|x-y|^2/\lambda_W^2)\mathrm{Id}$ for any points (x,y), where Id stands for the identity map. In this case, the standard deviation λ_W is the typical scale at which the vector fields ω in W may vary spatially. As we shall see below, this kernel will allow us to control the metric on the space of currents and hence the measure of the distance between shapes.

In contrast to the usual space of square integrable vector fields (L^2), the RKHS of vector fields W have two important properties:

- W is the closed span of the vector fields of the form $\omega(x) = K^W(x,y)\beta$ for any fixed points y and vectors β (i.e. momentum (y,β)), meaning that any vector field ω can be written as an infinite linear combination of the basis elements $K^W(x,y)\beta$
- W is provided with an inner product which is defined on these basis vectors by

$$\left\langle K^W(.,x)\alpha, K^W(.,y)\beta \right\rangle_W = \alpha^t K^W(x,y)\beta \tag{1.2.3}$$

If we denote ω the vector field $K^W(.,y)\beta$ in Eq. (1.2.3), this equation can be written as:

$$\left\langle K^W(.,x)\alpha, \omega \right\rangle_W = \alpha^t \omega(x)\,. \tag{1.2.4}$$

Since the set of vector field of the form $K^W(.,y)\beta$ is dense in the RKHS W, this equation still holds for any vector fields ω in W (see Appendix B for details). It is called the "reproducing property".

The space of currents

The space of currents, denoted W^*, is the space of the continuous linear mappings from W to $\mathbb{R}$. Equations (1.2.1) and (1.2.2) show two examples of such mappings, thus making

any set of curves or surfaces a particular cases of currents. This means that shapes can be embedded into the space of currents. Each test space W defines a different embedding space W^*. As we shall see in the sequel, the parameter of the kernel λ_W will allow us to tune the metric properties of the embedding space of currents.

As a space of mappings, the space of currents is a *vector space*. Let T and T' be two surfaces (or two curves). The flux through the sum of the two currents $T+T'$ is equal to the sum of the flux through each surface: $(T+T')(\omega) = T(\omega) + T'(\omega)$. The sum in the space of currents is equivalent to the union of geometrical data. The opposite surface $-T$ in the space of current is the same surface but with opposite orientation (since the flux through the surface has then the opposite sign). A surface T may be weighted by a coefficient λ: $((\lambda T)(\omega) = \lambda(T(\omega))$. This allows us to give a relative weight to different pieces of surfaces, or to different surfaces within a set of surfaces.

Representation of currents in terms of vector fields

As a consequence of this definition, the Riesz representation theorem ensures that there is a linear mapping between the space of vector fields W and its dual space W^*, the space of currents (see Appendix B and Section 1.5 for more details). We denote this mapping $\mathcal{L}_W : W \longrightarrow W^*$. It is defined by:

$$\mathcal{L}_W(\omega)(\omega') = \langle \omega, \omega' \rangle_W \tag{1.2.5}$$

for all vector fields $(\omega, \omega') \in W$ (for $\omega \in W$, $\mathcal{L}_W(\omega)$ is a current, i.e. a mapping from W to $\mathbb{R}$). We call $\mathcal{L}_W(\omega)$ the dual representation of the vector field ω.

The dual representation of the basis vectors $K^W(x,.)\alpha$ are called the Dirac delta currents: $\delta_x^\alpha = \mathcal{L}_W(K^W(x,.)\alpha)$ (where the couple (x,α) is called a momentum). This shows that K^W is the Green function of the differential operator[1] $\mathcal{L}_W$. Combining Eq. (1.2.4) and Eq. (1.2.5), we get:

$$\delta_x^\alpha(\omega) = \left\langle K^W(x,.)\alpha, \omega \right\rangle_W = \alpha^t \omega(x)\,. \tag{1.2.6}$$

Decomposition of sub-manifolds as discrete currents

The previous equation shows that $\delta_x^\alpha(\omega) = \alpha^t\omega(x)$, which is the term within integrals in Eq. (1.2.1) and Eq. (1.2.2). A Dirac delta current may be interpreted therefore as an infinitesimal segment (or normal) α entirely concentrated at point x. Since W is a closed span of the vector fields $K^W(x,.)\alpha$, the space of currents is a closed span of the Dirac delta currents δ_x^α. This means that any currents may be decomposed into an *infinite* set of Dirac currents, like any piecewise smooth curves or surfaces is decomposed into the set of its tangents or normals.

In the discrete setting, the tangents of polygonal lines or normals of meshes are constant over the segments or the mesh cells. Therefore, one can approximate the set of tangents (resp. normals of a given segment (resp. mesh cell) by a single Dirac delta current δ_x^α

[1] An equivalent construction would consist of fixing a differential operator $\mathcal{L}_W$ and to denote K^W its Green function. However, we prefer here to have a closed form for the kernel instead of the differential operator. See Appendix B for more details.

where the momenta (x, α) is located at the center of mass the segment (resp. mesh cell) and the magnitude of the coefficient α encodes the length of the segment (resp. the are of the mesh cell). As a consequence the whole set of polygonal lines (resp. the meshes) may be approximated by a *finite* sum:

$$T \sim \sum_k \delta_{x_k}^{\alpha_k} . \tag{1.2.7}$$

Here we see again that the addition in the space of currents plays the role of the union of shapes: a surface mesh is seen as the union of its cells, each cells being approximated by a single Dirac delta current. We will prove in Section 1.3 that this approximation converges in the space of currents when the sampling of the discrete shapes becomes finer and finer. This shows that this modeling of curves and surfaces is weakly sensitive to the sampling of the geometrical objects. Moreover, the description in terms of the collection of momenta (i.e. oriented points) accounts only for local properties of the shapes. It makes the framework based on currents fully robust to topology changes or the change of connectivity between structures (like curves interruption or reconnection for instance). See illustrative example in Fig. 1.2.

The dual representation of this approximation in terms of vector field is given by (applying the linear map $\mathcal{L}_W^{-1}$ to Eq. (1.2.7) and combining with Eq. (1.2.5)):

$$\mathcal{L}_W^{-1}(T)(x) \sim \sum_k K^W(x, x_k)\alpha_k,$$

for any point $x \in \mathbb{R}^3$. This representation in terms of vector field is simply given by the convolution of every momentum (x, α) by the smoothing kernel K^W.

Thanks to this approximation, the integrals in Eq. (1.2.1) and Eq. (1.2.2) are replaced by their Riemann sums:

$$S(\omega) = \int_S \omega(x)^t n(x) d\lambda(x) \sim \sum_k \omega(x_k)^t n_k$$
$$L(\omega) = \int_L \omega(x)^t \tau(x) d\lambda(x) \sim \sum_k \omega(x_k)^t \tau_k$$

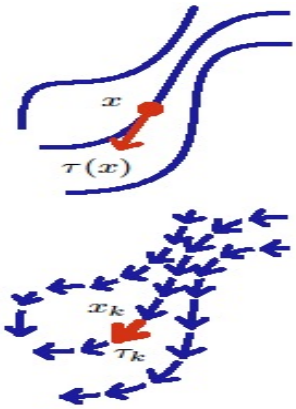


Figure 1.2: Both continuous and discrete shapes are handled in the same space of currents. In the continuous form, smooth curves are decomposed into the infinite set of their tangents. If curves are sampled, they can be approximated by a finite set of oriented points (called momenta) encoding each segment. The integral of a vector field ω on the smooth curve (i.e. a continuous current) is given as $L(\omega) = \int_L \omega(x)^t \tau(x) dx$. For the discrete approximation, this integral becomes a finite sum $L(\omega) = \sum_k \omega(x_k)^t \tau_k$. The discrete current converges to the continuous one as the sampling of the curves becomes finer and finer.

1.2.2 Correspondence-less distance between curves or surfaces

In this section, we introduce the metric on the space of currents. It will allow us to define a measure of the dissimilarity between two shapes without assuming point correspondences between structures.

The test space of vector field W is provided with an inner-product which satisfies Eq. (1.2.4). We can carry this inner-product to the space of current W^* via the linear map $\mathcal{L}_W$. The inner-product between two currents T and T' is then given as: $\langle T, T' \rangle_{W^*} = \left\langle \mathcal{L}_W^{-1}(T), \mathcal{L}_W^{-1}(T') \right\rangle_W$ where $\mathcal{L}_W^{-1}(T)$ denotes the vector field associated to T (if T is a discrete curve, $\mathcal{L}_W^{-1}(T)$ is the convolution of its tangents by the kernel). This makes $\mathcal{L}_W$ an isometric map between the space of vector field W and the space of currents W^*.

In W, the basis elements are the vector fields of the form $K^W(.,x)\alpha$. In W^*, the corresponding basis elements are the Dirac delta currents: $\delta_x^\alpha = \mathcal{L}_W^{-1}(K^W(.,x)\alpha)$. Therefore, the this inner-product between Dirac delta currents is given (thanks to Eq. (1.2.3)):

$$\begin{aligned} \left\langle \delta_x^\alpha, \delta_y^\beta \right\rangle_{W^*} &= \langle K(.,x)\alpha, K(.,y)\beta \rangle_W \\ &= \alpha^t K^W(x,y)\beta \,. \end{aligned} \tag{1.2.8}$$

By linearity, the inner product between two finite sets of Dirac currents $T = \sum_i \delta_{x_i}^{\alpha_i}$ and $T' = \sum_j \delta_{y_j}^{\beta_j}$ (which may model two discrete surfaces or two discrete curves) is given by:

$$\langle T, T' \rangle_{W^*} = \sum_i \sum_j \alpha_i^t K^W(x_i, y_j)\beta_j. \tag{1.2.9}$$

This equation gives *explicit* and *easily tractable* formula to compute the inner product between two discrete shapes. For continuous curves or surfaces, the sums in Eq. (1.2.9) are replaced by integrals.

We define now the distance between two shapes modeled as currents as the norm of their difference:

$$d(T, T') = \|T - T'\|_{W^*} = \sqrt{\langle T - T', T - T' \rangle_{W^*}}. \tag{1.2.10}$$

Combining with the definition of the map $\mathcal{L}_W$ in Eq. (1.2.5), we get:

$$\|T - T'\|_{W^*}^2 = (T - T')(\mathcal{L}_W^{-1}(T - T')). \tag{1.2.11}$$

In this equation, $T - T'$ is a current, namely an object which integrates vector fields, which is applied here to the vector field $\mathcal{L}_W^{-1}(T - T')$. Let us denote this vector field $\Delta(x) = \mathcal{L}_W^{-1}(T - T')(x)$. If T and T' are two curves whose tangents are denoted $\tau(x)$ and $\tau'(x)$ respectively, then:

$$\|T - T'\|_{W^*}^2 = \int_T \Delta(x)^t \tau(x) dx - \int_{T'} \Delta(x)^t \tau'(x) dx, \tag{1.2.12}$$

as illustrated in Fig. 1.3. If T and T' are discretized as finite sets of momenta ((x_p, α_p) and (y_q, β_q) respectively), then this squared norm becomes:

$$\|T - T'\|_{W^*}^2 = \sum_p \Delta(x_p)^t \alpha_p - \sum_q \Delta(y_q)^t \beta_q. \tag{1.2.13}$$

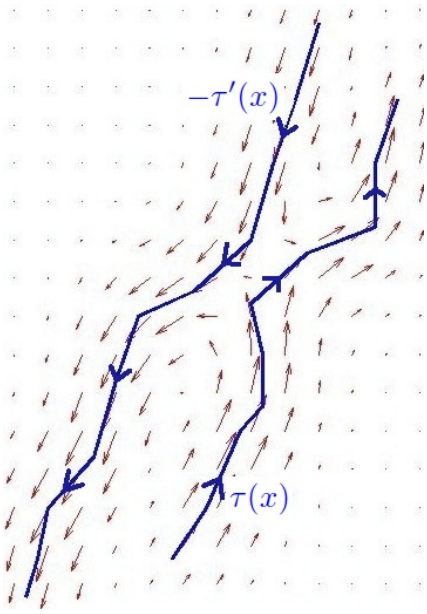


Figure 1.3: Distance between two curves L and L'. One builds the current $L - L'$ by inverting the orientation of L' and building the union of the tangents of both curves. The associated vector field $\Delta(x) = \mathcal{L}_W^{-1}(L - L')(x)$ is shown in red (i.e. the convolution of all the momenta of the L and $-L'$). The distance between both curves is given by the integration of this vector field along the curve L and the curve $-L'$: $\|L - L'\|^2 = \int_L \Delta(x)^t \tau(x) dx - \int_{L'} \Delta(x)^t \tau'(x) dx$. If the polygonal lines are approximated with a finite number of momenta, these integrals become finite sums over the segments of the lines.

The dense vector field $\Delta(x)$ is the vector field associated to the current $T - T'$. In case T and T' are given by the set of momenta $(x_p, \alpha_p)_{p=1\dots N}$ and $(y_q, \beta_q)_{q=1\dots N'}$, then the current $-T'$ is given by the momenta $(y_q, -\beta_q)$ (i.e. the orientation of the curve or the surface is changed) and eventually the current $T - T'$ is given by the union of all the momenta: $\{(x_p, \alpha_p)_{p=1\dots N}, (y_q, -\beta_q)_{q=1\dots N'}\}$. Its associated vector field is given by the convolution of these momenta by the kernel K^W, namely:

$$\Delta(x) = \mathcal{L}_W^{-1}(T - T')(x) = \sum_{p=1}^{N} K^W(x, x_p)\alpha_p - \sum_{q=1}^{N'} K^W(x, y_q)\beta_q \tag{1.2.14}$$

Combining this last equation with Eq. (1.2.13) leads to:

$$\|T - T'\|_{W^*}^2 = \sum_{p=1}^{N}\sum_{q=1}^{N} \alpha_p^t K^W(x_p, x_q)\alpha_q - 2\sum_{p=1}^{N}\sum_{q=1}^{N'} \alpha_p^t K^W(x_p, y_q)\beta_q + \sum_{p=1}^{N'}\sum_{q=1}^{N'} \beta_p^t K^W(y_p, y_q)\beta_q \tag{1.2.15}$$

This gives a closed form of the distance between two discrete curves or two discrete surfaces, which implies the kernel and every momenta representing the segments or the mesh cells of the shapes. Using Eq. (1.2.9), this last equation can be written as:

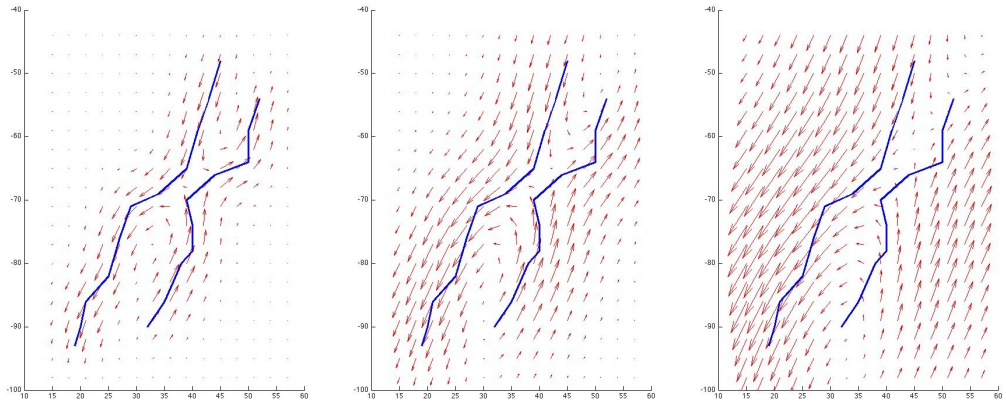
$$\|T - T'\|_{W^*}^2 = \|T\|_{W^*}^2 - 2\langle T, T'\rangle_{W^*} + \|T'\|_{W^*}^2 \tag{1.2.16}$$

which is the usual formula for computing the norm form inner-products.

We could have derived the closed form for the norm in Eq. (1.2.15) using only the usual identity in Eq. (1.2.16) and Eq. (1.2.9). However, we prefer to compute the distance $\|T - T'\|_{W^*}$ by introducing the vector field $\Delta(x)$, since this vector field has a geometrical interpretation. Indeed, we will show in Section 1.3 that this vector field is the one which achieves the supremum:

$$\sup_{\|\omega\|_W \neq 0} |T(\omega) - T'(\omega)| \,/\, \|\omega\|_W \,. \tag{1.2.17}$$

This means that if T and T' are two curves, the vector field $\Delta(x) = \mathcal{L}_W^{-1}(T - T')(x)$ is the one which maximizes the difference between the integral $T(\omega) = \int_T \omega(x)^t \tau(x) dx$ and the integral $T'(\omega) = \int_{T'} \omega(x)^t \tau'(x) dx$ over all the possible vector fields ω in the test space W. In some sense, this vector field is the one in W which best separates the two curves. Of course, if one changes the test space W, one changes this maximizing vector field and hence the measured distance between the curves. As illustrated in Fig 1.4, the highest the spatial frequencies of the vector fields in W, the more differences between both curves the maximizing vector field captures, the further the curves in the space of currents. The bandwidth of the vector fields in W is determined by the kernel (by the standard deviation λ_W for Gaussian kernel). This parameter can be tuned to set of "scale of noise" of shapes under which the geometrical details of shapes will be neglected, as illustrated in Fig. 1.5. Fig. 3.15 and 3.16 in Chapter 3 will also illustrate of the impact of the kernel on the distance between shapes.



$\lambda_W = 5$ $\lambda_W = 10$ $\lambda_W = 20$

Figure 1.4: Impact of the kernel on the distance between two curves L and L' in blue. As in Fig. 1.3, the vector field associated to the current $L - L'$ is shown in red. This is the vector in the test space W which best separates the two curves. The result is shown for 3 different W: RKHS with Gaussian kernel and standard deviation: $\lambda_W = 5, 10$ and 20. For small λ_W, the vector field can vary fast enough, so that it can follows almost every small details of the curves and therefore almost perfectly interpolates between the directions of the curves: the two curves are almost orthogonal in the space of currents ($\arccos\left(\frac{|\langle L_1, L_2\rangle_{W^*}|}{\|L_1\|_{W^*}\|L_2\|_{W^*}}\right) = 85°$ for $\lambda_W = 5$). For large λ_W, the highest spatial frequencies are excluded from W: the vector field cannot adapt to the small-scale variations of the curves: the two curves become more and more aligned in the space of currents ($\arccos\left(\frac{|\langle L_1, L_2\rangle_{W^*}|}{\|L_1\|_{W^*}\|L_2\|_{W^*}}\right) = 65°$ for $\lambda_W = 10$ and $38°$ for $\lambda_W = 20$).

Figure 1.5: The choice of the kernel enables to adjust the metric on the space of currents. In particular, the rate of decay of the kernel (λ_W) determines the scale under which shape variations are considered as noise. On the left hand side, the bump is smoothed by the kernel and both lines are considered similar as currents. On the right-hand side the metric detects the bump as a shape dissimilarity. More precise discussion about the ability of the metric on currents to capture shape dissimilarity can be found in Section 3.5.2

1.2.3 Diffeomorphic deformations of currents

To include registration into the analysis of variability of anatomical structures, we need to define the deformation of currents in a way which is compatible with the usual geometric deformation of shapes.

Let ϕ be a diffeomorphism (a smooth deformation of the underlying 3D space, with smooth inverse) and S a surface. As a surface, S may be deformed by ϕ into $\phi(S)$ (the geometrical transportation of the points of S which still draw a surface). If we model S as a current, we define the push-forward current $\phi_* S$ such that the flux of any vector field ω through $\phi_* S$ is equal to the flux of ω through the transported surface $\phi(S)$. A change of variable within integrals of Eq. (1.2.1) and Eq. (1.2.2) leads to the definition: $\phi_* S(\omega) = S(\phi^* \omega)$ where the pull-back vector field $\phi^* \omega$ is equal to $|d_x \phi| \, (d_x \phi)^{-1} \omega(\phi(x))$ for surfaces and $d_x \phi^t \omega(\phi(x))$ for curves ($d_x \phi$ denotes the Jacobian matrix of ϕ and $|d_x \phi|$ its determinant). This action replaces for curves and surfaces the usual action on images: $(\phi_* I)(x) = I(\phi^{-1}(x))$. This is here slightly more complex since we do not transport points but tangents or normals (differential 1 and 2-forms, as will be explained in Section 1.3).

In practice, the push-forward action on the basis vectors is simply given by:

$$\phi_* \delta_x^\alpha = \delta_{\phi(x)}^{d_x \phi(\alpha)}, \tag{1.2.18}$$

in case α is a tangent of a curve. And

$$\phi_* \delta_x^{u \times v} = \delta_{\phi(x)}^{d_x \phi(u) \times d_x \phi(v)}, \tag{1.2.19}$$

in case $u \times v$ is the normal of a surface. One notices that by definition of the cross product, for any vector w, we have:

$$\begin{aligned} (d_x \phi(u) \times d_x \phi(v))^t w &= \det(d_x \phi(u), d_x \phi(v), w) \\ &= |d_x \phi| \det(u, v, d_x \phi^{-1}(w)) \\ &= |d_x \phi| \, (u \times v)^t d_x \phi^{-1}(w) = \left(|d_x \phi| \, d_x \phi^{-t}(u \times v) \right)^t w, \end{aligned} \tag{1.2.20}$$

where A^{-t} stands for $(A^{-1})^t$. Therefore, we have: $d_x \phi(u) \times d_x \phi(v) = |d_x \phi| \, d_x \phi^{-t}(u \times v)$.

And the deformation of an infinitesimal normal α is given by:

$$\phi_* \delta_x^\alpha = \delta_{\phi(x)}^{|d_x\phi| d_x\phi^{-t}\alpha}. \tag{1.2.21}$$

1.2.4 Currents: a solution to the (point)-correspondence issue

As emphasized in the introduction, this framework of currents has been chosen to measure dissimilarities between anatomical data. The anatomical structures extracted from MRI consist mostly of set of points which draw polygonal lines or surface meshes, as shown in Figure 1.6. These data may be seen as a hierarchical structure: points build curves or surfaces, individual curves or surfaces are gathered into clusters, a set of cluster builds complex multi-objects structures. At a certain level, these structures are labeled as anatomical structures, which have been proved to be stable features across the population. In the context of brain imaging for instance, some sulcal lines such as the Sylvian fissure can be found in almost every subject, whereas the individual points of the delineated Sylvian fissure are determined by the segmentation process and are not a stable anatomical feature.

It is crucial to account for this anatomical knowledge when comparing two sets of anatomical structure: the distance between these two sets should put into correspondence the clusters of points only at the anatomically relevant level. Comparing the data at a higher level leads to a less constrained distance, which will be less able to capture fine geometrical differences. It will also compare parts of the data which have different anatomical roles. Comparing the data at a lower level introduces correspondence without any anatomical reason. Such arbitrary constraints introduce bias in the analysis of the variability.

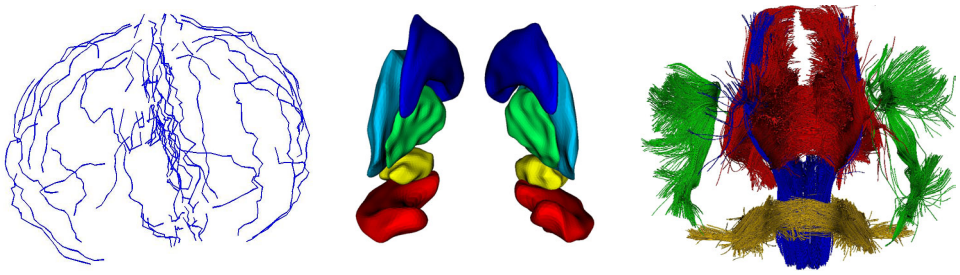
In the example of the fiber tracts of Fig. 1.6 (which will be explained with more details in Chapter 7), *individual* fibers extracted from diffusion images have never been shown to be a representation of some biological structures and have never been shown to be a stable feature across subjects. By contrast, *clusters* of these fibers draw fiber tracts which are considered as a representation, up to a certain precision, of the underlying white matter fiber bundles which connect two different functional areas of the cortex. In this example, we must compare pairs of fiber tracts and not pair of individual fibers.

The structure of vector space of the currents enables precisely to *adjust the level of correspondence* according to the anatomical knowledge. To compare the anatomical data of two subjects, one decomposes the data of each subject into the set of every tangent or normal. Let $T = \sum_i \delta_{x_i}^{\alpha_i}$ and $U = \sum_j \delta_{y_j}^{\beta_j}$ be the set of such oriented points for each subject. Then, we divide this set into clusters $\mathcal{C}_k$ according to the anatomical labels. Note that the number of points in a cluster may be very different for both subjects. By contrast, every subject is supposed to share the same anatomical description and, in particular, to have the same number of clusters. A cluster may be reduced to one single point if this point is considered as anatomically relevant, such as the anterior or the posterior commissure of the brain for instance. Then, a metric between the two data sets which account for the anatomical prior without introducing arbitrary correspondences can be written as:

$$\|T - U\|_{W^*}^2 = \sum_{\mathcal{C}_k} \lambda_k \left\| \left(\sum_{i \in \mathcal{C}_k} \delta_{x_i}^{\alpha_i} \right) - \left(\sum_{j \in \mathcal{C}_k} \delta_{y_j}^{\beta_j} \right) \right\|_{W^*}^2. \tag{1.2.22}$$

The two extreme cases are: (1) in absence of any anatomical knowledge, we compare the whole data-set as a single cluster $\left(\|T-U\|_{W^*} = \left\|\sum_i \delta_{x_i}^{\alpha_i} - \sum_{y_j}^{\beta_j}\right\|_{W^*}\right)$ and (2) each point has an anatomical label and we can assume correspondence between every pair of points (which requires that all the subjects have the same number of points) $\left(\|T-U\|_{W^*}^2 = \sum_k \lambda_k \left\|\delta_{x_i}^{\alpha_i} - \delta_{y_i}^{\beta_i}\right\|_{W^*}^2\right)$. The parameters λ_k enable to weight one anatomical structure with respect to the others. They can be used to normalize the total length or area of each anatomical structure for instance.

We remark that each cluster is considered as a *global* feature: a collection of infinitesimal tangents or normals. The topology of the shape is not taken into account. This makes the framework robust to curve interruption or reconnection for instance, a key feature for the comparison of fiber bundles as emphasized in Chapter 7. The distance on currents is also blind to the number of connected components of a mesh for instance.



points	points	points
sulcal lines	meshes	individual fibers
set of sulci	set of meshes	fiber bundles
–	–	all fibers

Figure 1.6: Three examples of anatomical data-sets. Left: a set of sulcal lines. Each line is labeled (Sylvian fissure, central sulcus, etc.) and is supposed to be present in every subject in a normal population. Middle: set of 5 internal structures of the brain for each hemisphere. These structures are labeled (hippocampus, amygdala, etc..), whereas no point on this surface has been proved to play a particular anatomical role. Right: 5 white matter fiber bundles. Each bundle (set of thousands of curves) is labeled (corpus callosum, arcuate fasciculi, etc.) but not individual curves, whose number within a bundle may vary a lot across subjects. Only labeled structures are stable features across the population. Each of these structures must be compared as a global feature, without introducing arbitrary correspondence at a lower level (such as the individual curve level or point level, for instance). The framework based on currents allows us precisely to adjust the right level of correspondence according to the anatomical knowledge.

1.3 The mathematical construction of currents

In this section, we give the rigorous definitions of the concepts introduced in the previous one, prove the claimed properties and discuss the relevance of this model for anatomical data.

1.3.1 A unified model of geometrical data

The framework of currents has been presented in [Vaillant 2005] for modeling surfaces and in [Glaunès 2008] for modeling curves. A global framework for curves and surfaces has been introduced in [Glaunès 2005], along with objects called "measures" which model unstructured point-sets. In this thesis, we adopt a slightly different point of view and present a unified framework for modeling unstructured point sets, curves, surfaces, volumes and more generally any sub-manifold of dimension m in $\mathbb{R}^d$. This framework allows us also to account for possible scalar attributes on the geometrical structures. This feature could be used for image matching purposes for instance.

The key tool of the theory is the differential m-forms (see Appendix A). These objects generalize the concept of vector field (to a field of normals for instance, since the normal is the cross-product between two vectors). Therefore, they enable to embed in the same framework all kind of geometrical data.

The general definition of currents is given as a continuous linear map from a space of differential m-forms to $\mathbb{R}$, namely a linear form on differentiable m-forms. The parameter m determines the dimension of the sub-manifold which can be seen as a current. If $m = 0$ (differential 0-forms are scalar fields), a current is simply a distribution of Schwartz. For $m = 1$ and $m = 2$ we retrieve the case of curves and surfaces introduced in the previous section.

Definition 1.2 (currents)**.** *The space of m-currents is the dual space of the space of differential m-forms $\mathcal{C}^0(\mathbb{R}^d, (\Lambda^m\mathbb{R}^d)^*)$ (as in Definition A.6). The topological dual is meant in the same sense as for the Schwartz distributions [Schwartz 1966].*

Therefore, a m-current T maps every m-differential form ω to a real $T(\omega)$ such that:

$$T(\omega) \leq C_T \left\|\omega\right\|_\infty , \tag{1.3.1}$$

for a fixed constant C_T ($\|.\|_\infty$ denotes the supremum norm on $\mathcal{C}^0(\mathbb{R}^d, (\Lambda^m\mathbb{R}^d)^*)$ as in Definition A.6).

As a space of linear mappings, the space of m-currents is a vector space. For all currents T and T' and real λ, the map $(T+\lambda T')(\omega) = T(\omega)+\lambda T'(\omega)$ defines a m-current. Moreover, the space of current is provided with the following operator norm:

Definition 1.3 (mass-norm on currents)**.** *Let T be a m-current. The mass-norm of T is defined as the operator norm:*

$$M(T) = \sup_{\|\omega\|_\infty \leq 1} |T(\omega)| . \tag{1.3.2}$$

The main interest of currents is that sub-manifolds in $\mathbb{R}^d$ can be seen as currents. The following proposition shows that this is achieved via the integration of differential m-forms on sub-manifolds as introduced in Appendix A. In this proposition, we account for a possible image I which is drawn on the sub-manifold.

Proposition 1.4. *Let T be an oriented rectifiable sub-manifold of dimension m in $\mathbb{R}^d$ and I a scalar function on T, such that $\int_T |I(x)|\, d\lambda(x) < \infty$.*

Then, for any m-differential form ω in $\mathcal{C}^0(\mathbb{R}^d, (\Lambda^m \mathbb{R}^d)^*)$ the mapping:

$$T_I(\omega) = \int_T I\omega, \tag{1.3.3}$$

defines a m-current (the integral having the sense given in Definition A.7).

Proof. The mapping T_I defined in Eq. (1.3.3) is obviously linear with respect to ω. To make T_I a current, we must verify that this mapping is continuous. Using the notations of Definition A.7, we have:

$$\begin{aligned} |T_I(\omega)| &\leq \int_T |I(x)| \left| \omega(x) \left(\frac{u_1(x) \wedge \ldots \wedge u_m(x)}{|u_1(x) \wedge \ldots \wedge u_m(x)|} \right) \right| d\lambda(x) \\ &\leq \sup_{x \in \mathbb{R}^d} \sup_{|v_1 \wedge \ldots \wedge v_m| \leq 1} |\omega(x)(v_1 \wedge \ldots \wedge v_m)| \int_T |I(x)|\, d\lambda(x) \\ &\leq \|\omega\|_\infty \, \lambda(T_I), \end{aligned} \tag{1.3.4}$$

where $\lambda(T_I) = \int_T |I(x)|\, d\lambda(x)$ is the measure of the colored manifold T (if $I = 1$, $\lambda(T)$ is the length, area or volume of T according to the dimension m). This proves the continuity of the mapping. ■

Remark 1.5. In Proposition 1.4, the definition of the current T depends on the orientation of the sub-manifold. If we change the orientation of the sub-manifold T, then the integral has the opposite sign. Therefore the sub-manifold with the opposite orientation corresponds to the current $-T$. In other words, the sign of a current encodes an orientation. A sub-manifold vanishes in the space of currents if it is added to itself with opposite orientation. □

Now, we can show that the mass-norm introduced in Definition 1.3 is the generalization of the volume of a m-dimensional sub-manifold.

Proposition 1.6. *Let T be a bounded oriented rectifiable sub-manifold of dimension m in $\mathbb{R}^d$ and I a bounded non-negative scalar function on T. Proposition 1.4 makes the couple (T, I) a current denoted T_I. Then,*

$$M(T_I) = \lambda(T_I), \tag{1.3.5}$$

where $M(T_I)$ is the mass-norm of the current T_I (see Definition 1.3) and $\lambda(T_I) = \int_T I(x) d\lambda(x)$ the Lebesgue measure of the colored sub-manifold.

Proof. Equation (1.3.4) shows precisely that for all $\omega \in \mathcal{C}^0(\mathbb{R}^d, (\Lambda^m \mathbb{R}^d)^*)$:

$$|T_I(\omega)| \leq \|\omega\|_\infty \, \lambda(T_I) \tag{1.3.6}$$

This shows that $M(T_I) \leq \lambda(T_I)$. To show the equality, we construct a differential m-form ω which achieves the supremum in Equation (1.3.4).

Let $x \in T$ and $(u_1(x), \ldots, u_m(x))$ a positively oriented *orthonormal* basis of the tangent space of T at point x, defined almost everywhere. Then, we define the m-form $\omega(x)$ such that $\omega(x)(u_1(x) \wedge \ldots \wedge u_m(x)) = 1$ and $\omega(x)(\eta) = 0$ in every direction η orthogonal to $u_1(x) \wedge \ldots \wedge u_m(x)$ in $\Lambda^m \mathbb{R}^d$. Such a ω is build such that $T_I(\omega) = \int_T I(x) d\lambda(x)$. This collection of m-forms $\omega(x)$ (for $x \in T$) is therefore a good candidate to achieve the supremum. However, we still need to prove that it can be extended to continuous differential m-form defined on the whole space $\mathbb{R}^d$.

Since the sub-manifold is rectifiable, we can choose the orthonormal basis vectors $u_i(x)$ on each tangent space such that the map $x \to u_i(x)$ is continuous almost everywhere (we recall that the m-mutivector $\frac{u_1(x) \wedge \ldots \wedge u_m(x)}{|u_1(x) \wedge \ldots \wedge u_m(x)|}$ is invariant under a change of positively oriented basis in the tangent space of T at point x). This makes the mapping $x \to \omega(x)$ continuous almost everywhere on T. If T is continuously differentiable, then $x \to \omega(x)$ is continuous on T and it can be extended to a continuous differential m-form from $\mathbb{R}^d$ to $\Lambda^m \mathbb{R}^d$ which tends to zero at infinity. This constructed ω achieves the supremum. If T is only piecewise smooth, $x \to \omega$ is piecewise continuous. It can be approximated by a continuous differential m-form at any precision. This leads to the same supremum. ■

Remark 1.7. In this proof, we supposed that the bounded scalar function I is non-negative. This is not a strong limitation, since we can always shift this bounded function so that it is non-negative. In any case we have $M(T_I) \leq \lambda(T_I)$. If I is the linear, surface or volume mass density of the sub-manifold T, then $M(T_I)$ is the total mass of the manifold. This function I can also be used to give a weight to different parts of the shape T, for comparison purpose for instance. I can be also a gray-level image drawn on the manifold.

If $I = 1$ and if T is bounded (i.e. $\lambda(T) < \infty$), then

$$M(T) = \lambda(T). \tag{1.3.7}$$

In this case the mass of T is equal to the length, the area or the volume of the sub-manifold. If the sub-manifold is of dimension 0, the mass of T equals the number of points of T. □

1.3.2 Discretization in the space of currents

In this section, we show how polygonal lines, surface or volume meshes can be discretized in the space of currents: each face of the mesh can be approximated by a single Dirac delta current which represents an infinitesimal tangent or normal. Then, the approximation converges when the sampling of the meshes tend to zero, namely when the discrete mesh converges to a continuous curve or surface.

Before introducing the Dirac delta currents, we define general m-meshes in 3D. A 0-mesh is a finite set of points, a 1-mesh is set of segments, a 2-mesh is a set of triangles and a 3-mesh is a set of tetrahedrons.

Definition 1.8 (m-mesh)**.** *We define an oriented m-mesh in $\mathbb{R}^3$ as a finite collection of oriented m-dimensional simplexes ($0 \leq m \leq 3$). Each simplex is called the face or the cell of the m-mesh. It is denoted f_i for $i = 1, \ldots, N$ and N the total number of mesh cells. For*

each cell i, we denote $v_1^i, \dots, v_{m+1}^i$ its vertices and $c_i = \frac{1}{m+1}\sum_{k=1}^{m+1} v_k^i$ its center of mass. We define also $u_k^i = (v_{k+1}^i - v_0^i)/(m!)^{1/m}$, m vectors parallel to the edges of the simplex i (for $k = 1 \dots m$). We suppose that the order of the vertices have been chosen so that the basis $(u_k^i)_{k=1\dots m}$ has the same orientation as the mesh cell. Therefore, the m-multivector (see Appendix A) $u_1^i \wedge \dots \wedge u_m^i$ has the same sign as the orientation of the mesh cell and its norm equals the m-volume of the mesh cell (here m-volume denotes length, area or volume according to the dimension m).

In this definition, we limit the mesh cells to be simplexes, so that surface mesh cells must be triangles and volume mesh cells tetrahedrons. However, it is possible to extend this definition of arbitrary polyhedrons. In this case, we must define carefully a basis of vector space spanned by each polyhedron so that the m-multivector has the same orientation as the mesh cell and whose norm is equal to the m-volume of the cell. We also limit the definition of m-mesh to the 3D case. The definition could also be extended for arbitrary dimension, like the 4D dimension for modeling moving surfaces for instance.

Definition 1.9 (Dirac delta current)**.** *We denote by $\delta_x^{u_1\wedge\dots\wedge u_m}$ the linear form defined by:*

$$\forall \omega \in \mathcal{C}^0(\mathbb{R}^d, (\Lambda^m \mathbb{R}^d)^*), \quad \delta_x^{u_1\wedge\dots\wedge u_m}(\omega) = \omega(x)(u_1 \wedge \dots \wedge u_m)\,. \tag{1.3.8}$$

This mapping is obviously continuous and defines therefore a m-current.

Remark 1.10 (computing with Dirac delta currents)**.** We notice that this definition leads to the following rule:

$$a\delta_x^\alpha + \delta_x^\beta = \delta_x^{a\alpha} + \delta_x^\beta = \delta_x^{a\alpha+\beta}, \tag{1.3.9}$$

for any scalar a, m-multivectors α, β and point $x \in \mathbb{R}^d$. As a consequence, we will write linear combination of Dirac delta currents as $\sum_i \delta_{x_i}^{\alpha_i}$ where the upper-scripts include the weighting coefficients. In such sums, each point x_i are supposed to be distinct. □

According to Proposition 1.4, any bounded m-mesh with scalar attributes I is a current. In this setting, the mesh is considered as a piecewise $\mathcal{C}^1$ sub-manifold. However, from a computational point of view, a mesh is a discrete structure with a finite number of faces. In the following proposition, we show how a m-mesh can be approximated by a finite set of Dirac delta currents. The idea is simply to replace each cell of the mesh by a Dirac delta current: the entire m-volume of the cell is concentrated at its center of mass. This proposition shows also that currents can handle continuous and discrete structures in the same framework.

Proposition 1.11. *Let T be a m-mesh in the sense of Definition 1.8 and I a bounded scalar function defined on T. Thanks to Proposition 1.4, T_I is a current.*

We assume that the differential m-forms $\omega \in \mathcal{C}^0(\mathbb{R}^d, (\Lambda^m \mathbb{R}^d)^*)$ are C^1 and verify $\|\nabla_x \omega\|_\infty \leq C_\infty \|\omega\|_\infty$ for a fixed constant C_∞.

With the notations of Definition 1.8, let $I_i = \frac{1}{V_i}\int_{f_i} I(x)d\lambda(x)$ be the mean intensity of I over each face f_i of T, where $V_i = \left|u_1^i \wedge \dots \wedge u_m^i\right|$ is the m-volume of the face f_i. We define the current

$$\tilde{T}_I = \sum_i I_i \delta_{c_i}^{u_1^i\wedge\dots\wedge u_m^i}. \tag{1.3.10}$$

Then, the dissimilarity between T and $\tilde{T}$ in the space of currents is such that:

$$M(T_I - \tilde{T}_{\tilde{I}}) \leq C_\infty \|I\|_\infty M(T) \max_i \operatorname{diam}(f_i), \tag{1.3.11}$$

where $\operatorname{diam}(f_i)$ *denotes the diameter of the face* f_i.

Therefore, when the sampling of the mesh becomes finer ($\max_i \operatorname{diam}(f_i) \longrightarrow 0$*), the current* $\tilde{T}_{\tilde{I}}$ *converges to* T_I.

Proof. Let $\omega \in \mathcal{C}^0(\mathbb{R}^d, (\Lambda^m \mathbb{R}^d)^*)$. By definition of the Dirac delta currents, we have:

$$\tilde{T}_{\tilde{I}}(\omega) = \sum_i I_i \omega(c_i)(u_1^i \wedge \ldots \wedge u_m^i) = \sum_i \int_{f_i} I(x)\omega(c_i) \left(\frac{u_1^i \wedge \ldots \wedge u_m^i}{|u_1^i \wedge \ldots \wedge u_m^i|} \right) d\lambda(x)\,.$$

Thanks to Eq. (1.3.3), we have for T_I:

$$T_I(\omega) = \sum_{f_i} \int_{f_i} \omega(x) \left(\frac{u_1(x) \wedge \ldots \wedge u_m(x)}{|u_1(x) \wedge \ldots \wedge u_m(x)|} \right) I(x) d\lambda(x),$$

where $u_1(x), \ldots u_m(x)$ denotes a positively oriented basis of the tangent plane of T at point x, this integral being independent of the choice of the positively oriented basis. One notices now that every point on the flat face f_i shares the same tangent space, so that one can choose the same basis $u_1^i, \ldots, u_m^i$ at every point on the face. We have therefore:

$$\begin{aligned}
\left| T_I(\omega) - \tilde{T}_{\tilde{I}}(\omega) \right| &\leq \sum_i \int_{f_i} |I(x)| \left| \omega(x) \left(\frac{u_1^i \wedge \ldots \wedge u_m^i}{|u_1^i \wedge \ldots \wedge u_m^i|} \right) - \omega(c_i) \left(\frac{u_1^i \wedge \ldots \wedge u_m^i}{|u_1^i \wedge \ldots \wedge u_m^i|} \right) \right| \\
&\leq \|I\|_\infty \sum_i \int_{f_i} \underbrace{\sup_{|v_1 \wedge \ldots \wedge v_m| = 1} |(\omega(x) - \omega(c_i))(v_1 \wedge \ldots \wedge v_m)|}_{|\omega(x) - \omega(c_i)|_{(\Lambda^m \mathbb{R}^d)^*}} d\lambda(x) \\
&\leq \|I\|_\infty \|\nabla_x \omega\|_\infty \sum_i \int_{f_i} |x - c_i|\, d\lambda(x) \\
&\leq \|I\|_\infty C_\infty \|\omega\|_\infty (\max_i \operatorname{diam}(f_i)) \int_T d\lambda(x)\,.
\end{aligned} \tag{1.3.12}$$

According to Proposition 1.6, $\lambda(T) = M(T)$. This leads to:

$$\sup_{\|\omega\|_\infty \leq 1} \left| T_I(\omega) - \tilde{T}_{\tilde{I}}(\omega) \right| \leq C_\infty M(T) \|I\|_\infty \max_i \operatorname{diam}(f_i).$$

■

The following proposition shows that the discretization of a mesh in the space of currents preserves the mass-norm of the mesh.

Proposition 1.12. *Let T be a current of the form:*

$$T = \sum_{i=1}^N \delta_{x_i}^{\eta_i}, \tag{1.3.13}$$

for N distinct *points x_i and N m-multivectors η_i.*

Then the mass-norm of T is given as:

$$M(T) = \sum_{i=1}^{N} |\eta_i| \ . \tag{1.3.14}$$

Proof. For any m-differential form ω, we have:

$$T(\omega) = \sum_{i=1}^{N} \omega(x_i)(\eta_i) \, . \tag{1.3.15}$$

Therefore, we have that $T(\omega) \leq \sum\limits_{i=1}^{N} \sup\limits_{|u|\leq 1} |\omega(x_i)(u)| \, |\eta_i| \leq \|\omega\|_\infty \sum\limits_{i=1}^{N} |\eta_i|$. This implies that $M(T) \leq \sum_{i=1}^{N} |\eta_i|$.

To show the equality, we build a particular differential m-form ω which achieves the supremum in the previous equation (similarly as in the proof of Proposition 1.6). We choose a m-form $\omega(x_i)$ for every point x_i so that $\omega(x_i)(\eta_i / |\eta_i|) = 1$ in the direction of η_i and $\omega(x_i)(\eta_i^\perp) = 0$ in the directions $\eta_i^\perp$ orthogonal to η_i in $\Lambda^m \mathbb{R}^d$. Since, every points are distinct, we can interpolate between the points x_i so that the interpolated m-form is continuous and tends to 0 at infinity. For such a ω, $T(\omega) = \sum_{i=1}^{N} |\eta_i|$. ■

Corollary 1.13. *Let T be a m-mesh and I a non-negative bounded map on T. Let $\tilde{T}_I = \sum_{f_i} I_i \delta_{c_i}^{u_1^i \wedge \ldots \wedge u_m^i}$ be the discrete current which approximate T_I in the sense of Proposition 1.11. Then T_I and $\tilde{T}_I$ have the same mass-norm:*

$$M(T_I) = M(\tilde{T}_I) \, . \tag{1.3.16}$$

Proof. By application of Proposition 1.6, the mass-norm of T_I is equal to:

$$M(T_I) = \int_T I(x) d\lambda(x) = \sum_{f_i} \int_{f_i} I(x) d\lambda(x) = \sum_{f_i} V_i I_i = \sum_{f_i} I_i \left| u_1^i \wedge \ldots \wedge u_m^i \right| \, . \tag{1.3.17}$$

This last expression is precisely the mass-norm of $\tilde{T}_I$ by application of Proposition 1.12. ■

Remark 1.14 (Geometry and attributes: ambiguities?)**.** This discretization of colored currents in terms of Dirac delta currents may lead to some ambiguities. Indeed, let $I_i \delta_{x_i}^{u_m^i \wedge \ldots \wedge u_m^i}$ be an element of the discretization of a m-mesh. Due to the properties of the Dirac delta currents (see Eq. (1.3.9)), for any scalar a, we have:

$$I_i \delta_{x_i}^{u_m^i \wedge \ldots \wedge u_m^i} = \frac{I_i}{a} \delta_{x_i}^{a u_m^i \wedge \ldots \wedge u_m^i} . \tag{1.3.18}$$

This means that a change of attribute (I_i becomes I_i/a) may balance a scaling of the mesh cell. In other words, the discretization of two meshes with different size of mesh cells could be undistinguishable if the change of size is overcome by a change of attributes. However, such meshes, seen as continuous currents, are distinct. This is a bad effect of the discretization.

In this thesis, we will not use scalar attributes. As a consequence, the magnitude of the momenta of the discrete approximation of a m-mesh encodes the area of the mesh cells.

The possible ambiguity between geometry and scalar attributes may be a problem for using currents to match grey-level images. During registration of discretized images with

currents, the algorithm may tend to change the size of the voxel to accommodate to the change of intensity. Although we believe that image registration can be performed in this framework, one must design specific solution to this problem. A re-meshing of the image during registration could be investigated, for instance. □

1.3.3 Action of the group of diffeomorphism on the space of currents

In this section, we define the deformation of a general current T so that, when T is the current associated to a sub-manifold, the deformation of the current (denoted $\phi_* T$) corresponds to the deformed sub-manifold $\phi(T)$.

If T is an oriented rectifiable sub-manifold of dimension m in $\mathbb{R}^d$ and ϕ is diffeomorphism of $\mathbb{R}^d$, then $\phi(T)$ remains an oriented rectifiable sub-manifold (with the same regularity as T). If I is a scalar map on the sub-manifold T, then $I \circ \phi^{-1}$ is a map on $\phi(T)$: the attribute of $\phi(T)$ at point $\phi(x)$ is the same as the attribute of T at point x: attributes are not affected by the deformation. Therefore, Proposition 1.4 makes the couple $(\phi(T), I \circ \phi^{-1})$ a colored current. We denote this current $\phi_* T_{I \circ \phi^{-1}}$. This current maps every differential m-forms $\omega \in \mathcal{C}^0(\mathbb{R}^d, (\Lambda^m \mathbb{R}^d))$ to $\phi_* T_{I \circ \phi^{-1}}(\omega) = \int_{\phi(T)} (I \circ \phi^{-1}) \omega$. We can now apply the change of variable formula as in Eq. (A.3.8) to this integral. This gives:

$$\phi_* T_{I \circ \phi^{-1}}(\omega) = \int_{\phi(T)} I \circ \phi^{-1} \omega = \int_T I \phi^* \omega = T_I(\phi^* \omega)\,, \tag{1.3.19}$$

where T_I in the last term is the colored current associated to the sub-manifold T and the map I. ϕ^* denotes the pullback action on differential m-forms as in Definition A.11. The equality: $\phi_* T_{I \circ \phi^{-1}}(\omega) = T_I(\phi^* \omega)$ still makes sense even if T_I is not the current associated to a sub-manifold but a more general current. This allows us to define the action a diffeomorphism ϕ on any current T as follows:

Definition 1.15 (push-forward action on currents)**.** *Let T be a m-current in $\mathbb{R}^d$ and ϕ a diffeomorphism of $\mathbb{R}^d$ such that $\sup_{x \in \mathbb{R}^d} |d_x \phi| < \infty$. The push-forward action of ϕ on T is defined by:*

$$\phi_* T(\omega) = T(\phi^* \omega) \tag{1.3.20}$$

for all differential m-forms $\omega \in \mathcal{C}^0(\mathbb{R}^d, (\Lambda^m \mathbb{R}^d)^*)$.

We can easily check that this defines an action of the group of diffeomorphism on the space of currents: $(\phi \circ \psi)_* T = \phi_*(\psi_* T)$, since the pullback is also an action as mentioned in Appendix A. Moreover, the action is linear:

$$\phi_*(T + \lambda T') = \phi_* T + \lambda \phi_* T', \tag{1.3.21}$$

for any currents T and T' and any real numbers λ.

Proposition 1.16. *If T is a sub-manifold of $\mathbb{R}^d$ and I a map on T such that $\int_T |I(x)|\, d\lambda(x) < \infty$, then the sub-manifold $\phi(T)$ associated to the map $I \circ \phi^{-1}$ is a current. This current is equal to $\phi_*(T_{I \circ \phi^{-1}})$.*

Proof. This is exactly what we proved in Eq. (1.3.19). ■

Now, we can apply Definition 1.15 on Dirac delta currents:

Proposition 1.17. *Let ϕ be a diffeomorphism of $\mathbb{R}^d$. The deformation of a m-Dirac delta current is given by:*

$$\phi_* \left(\delta_x^{u_1 \wedge \ldots \wedge u_m}\right) = \delta_{\phi(x)}^{(d_x\phi(u_1)) \wedge \ldots \wedge (d_x\phi(u_m))} . \tag{1.3.22}$$

Proof. For any differential m-form ω, we have by definition of the push-forward action on currents:

$$\begin{aligned} \phi_* \left(\delta_x^{u_1 \wedge \ldots \wedge u_m}\right)(\omega) &= \delta_x^{u_1 \wedge \ldots \wedge u_m}(\phi^*\omega) \\ &= \phi^*\omega(x)\left(u_1 \wedge \ldots \wedge u_m\right) \\ &= \omega(\phi(x))\left(d_x\phi(u_1) \wedge \ldots \wedge d_x\phi(u_m)\right) \\ &= \delta_{\phi(x)}^{d_x\phi(u_1) \wedge \ldots \wedge d_x\phi(u_m)}(\omega) \end{aligned} \tag{1.3.23}$$

which gives the expected result. ■

1.4 Particular cases of practical interest

In the applications, we are mainly interested in modeling unstructured point-sets in 2D or 3D ($m = 0, d = 2, 3$), curves in 2D or 3D ($m = 1$, $d = 2, 3$), surfaces in 3D ($m = 2$, $d = 3$) and volumes in 3D ($m = 3$, $d = 3$). These cases, all of great practical interest, fall into one of these 4 categories: they are of dimension 0 ($m = 0$), co-dimension 0 ($d - m = 0$), dimension 1 ($m = 1$) or co-dimension 1 ($d - m = 1$). In these cases, the m-forms can be represented by scalar fields (dimension or co-dimension 0) or vector fields (dimension or co-dimension 1), as shown in Appendix A.

In this section, we apply the construction of the previous section to these particular cases. We retrieve then the properties claimed in Section 1.2.

1.4.1 Unstructured point sets

The 0-forms are constant mappings and differential 0-forms map every point $x \in \mathbb{R}^d$ to a scalar $\omega(x)$. A differential 0-form is therefore a scalar field.

Let A be a discrete set of points $\{x_i\}$ associated to some scalar I_i such that $\int_A I(x)d\lambda(x) = \sum_{i \in A} I_i < \infty$. We recall that for sub-manifold of dimension 0, the measure $d\lambda = \sum_{x \in A} \delta_x$ counts the number of elements in A. Therefore, A may be modeled as the 0-current via :

$$A_I(\omega) = \int_A I(x)\omega(x)d\lambda(x) = \sum_i I_i\omega(x_i). \tag{1.4.1}$$

This shows that the sub-manifolds of dimension 0 are directly given as a sum of Dirac delta currents:

$$A = \sum_i \delta_{x_i}^{I_i} \tag{1.4.2}$$

Let ϕ be a diffeomorphism of $\mathbb{R}^d$. The push-forward action of ϕ on the 0-current A is given as:

$$\phi_* A = \delta_{\phi(x_i)}^{I_i}. \tag{1.4.3}$$

In these equations, we notice that the diffeomorphism ϕ just moves the point of A while keeping unchanged the attributes I_i. These coefficients are seen as intrinsic weights of each points, or probabilities if $\sum_i I_i = 1$.

1.4.2 Curves in any dimension

A 1-form is a linear form on $\mathbb{R}^d$ (i.e. linear mapping from $\mathbb{R}^d$ to $\mathbb{R}$). Thanks to the Riesz theorem, any linear form may be represented by the inner product: $\omega(u) = \overline{\omega}^t u$ for a constant vector $\overline{\omega} \in \mathbb{R}^d$. A differential m-form may be represented therefore by a vector field $\overline{\omega}(x)$ such that for all points $x \in \mathbb{R}^d$ and all vectors $u \in \mathbb{R}^d$, $\omega(x)(u) = \overline{\omega}(x)^t u$. In the sequel, we denote ω both the 1-form $\omega(x)$ and the vector field $\overline{\omega}(x)$.

Let L be a set of piecewise continuous curves and $I(x)$ an integrable scalar map on these curves. L may be seen as a 1-current via:

$$L_I(\omega) = \int_L I(x)\omega(x)^t \tau(x) d\lambda(x) \tag{1.4.4}$$

for every vector field ω, where $\tau(x)$ is the unit tangent vector of the curves L at point x. We recall that $d\lambda$ denotes the Lebesgue measure on the curves, so that $\int_L d\lambda(x)$ equals the total length of the curves.

If the curves L are polygonal lines (whose segments are denoted s_i), they can be approximated as

$$\tilde{L}_I = \sum_{s_i} \delta_{c_i}^{I_i \tau_i}, \tag{1.4.5}$$

where τ_i is the oriented segment s_i and $I_i = \frac{1}{|\tau_i|} \int_{s_i} I(x) d\lambda(x)$ the mean attribute over the segment s_i. Its action on a vector field ω is given by: $\tilde{L}_I(\omega) = \sum_i I_i \omega(c_i)^t \tau_i$.

If ϕ is a diffeomorphism of $\mathbb{R}^d$, then the deformed current $\phi_* L_{I \circ \phi^{-1}}$ is given by

$$\begin{aligned} \phi_* L_{I\circ\phi^{-1}}(\omega) = L_I(\phi^*\omega) &= \int_L I(x)\omega(\phi(x))^t (d_x\phi)\tau(x) d\lambda(x) \\ &= \int_L I(x) \left(d_x\phi^t \omega(\phi(x))\right)^t \tau(x) d\lambda(x)\,. \end{aligned} \tag{1.4.6}$$

In the discrete case, we have:

$$\phi_* \left(\sum_i I_i \delta_{c_i}^{\alpha_i} \right) = \sum_i I_i \delta_{c_i}^{d_{c_i}\phi(\alpha_i)}\,. \tag{1.4.7}$$

In this equation, we remark that the density $I(x)$ remains unchanged during the deformation and that the tangents of the curves are deformed according to the Jacobian of the deformation ϕ.

1.4.3 Surfaces in 3D

As shown in Appendix A, the space of 2-forms in dimension 3 is of dimension 3. Each 2-form ω is associated isometrically to a 3D-vector $\overline{\omega}$, such that $\omega(u,v) = \det(u,v,\overline{\omega}) = \overline{\omega}^t(u \times v)$ for all vectors (u,v), where $\times$ denotes the cross-product in $\mathbb{R}^3$. Therefore, a differential 2-form may be represented by a vector field $\overline{\omega}(x)$: $\omega(x)(u,v) = \overline{\omega}(x)^t (u \times v)$. In the sequel, $\omega(x)$ denotes both the differential 2-form and its associated vector field.

Let S be a set of piecewise continuous surfaces and I an integrable scalar map on S. S may be seen as a 2-current via:

$$S_I(\omega) = \int_S I(x)\omega(x)^t \left(u(x) \times v(x)\right) d\lambda(x) = \int_S I(x)\omega(x)^t n(x) d\lambda(x) \tag{1.4.8}$$

for every vector field ω, where $n(x) = u(x) \times v(x)$ is the unit normal vector of the surfaces S at point x ($(u(x), v(x))$ being an orthonormal basis of the tangent plane of the surface S at point x).

If the surfaces S are surface meshes (whose mesh cells are denoted f_i), they can be approximated as

$$\tilde{S}_I = \sum_{f_i} \delta_{c_i}^{I_i n_i}, \tag{1.4.9}$$

where $n_i = u_1^i \times u_2^i$ (in the sense of Definition 1.8) is the oriented normal of face f_i whose norm is equal to the surface of f_i, $I_i = \frac{1}{|n_i|}\int_{f_i} I(x)d\lambda(x)$ the mean attribute over the face f_i and c_i the center of mass of the face f_i. Its action on a vector field ω is given by: $\tilde{S}_I(\omega) = \sum_i I_i \omega(c_i)^t n_i$. Note that in this equation we write $\delta_x^{u\wedge v}$ for the 2-multivector $u \wedge v$ in 3D as $\delta_x^{u\times v}$ where $u \times v$ is the 3D-vector which characterizes $u \wedge v$ (see Appendix A). This notation, however, may be misleading since the dimension of the current ($m = 2$) is no more visible in this expression (see Remark 1.18).

If ϕ is a diffeomorphism of $\mathbb{R}^d$, then the deformed current $\phi_* S_{I\circ\phi^{-1}}$ is given by

$$\begin{aligned} \phi_* S_{I\circ\phi^{-1}}(\omega) = S_I(\phi^*\omega) &= \int_S I(x)\omega(\phi(x))^t \left(d_x\phi u(x) \times d_x\phi v(x)\right) d\lambda(x) \\ &= \int_S I(x) \left(|d_x\phi|\, d_x\phi^{-1}\omega(\phi(x))\right)^t n(x) d\lambda(x). \end{aligned} \tag{1.4.10}$$

In the discrete case, we have:

$$\phi_* \left(\sum_i I_i \delta_{c_i}^{n_i} \right) = \sum_i I_i \delta_{c_i}^{\left|d_{c_i}\phi\right| d_{c_i}\phi^{-t} n_i}. \tag{1.4.11}$$

1.4.4 Volumes in any dimension

As shown in Appendix A, all d-forms in dimension d are proportional to the determinant. This means that every d-form in $\mathbb{R}^d$ has the form: $\omega(u_1 \wedge \ldots \wedge u_d) = \overline{\omega} \det(u_1, \ldots, u_d)$, where $\overline{\omega}$ is a scalar which characterizes the d-form ω. As a consequence, a differential d-form in $\mathbb{R}^d$ is characterized by a scalar field $\overline{\omega}(x)$ such that $\omega(x)(u, v, w) = \overline{\omega}(x) \det(u, v, w)$. In the sequel, $\omega(x)$ denotes both the differential d-form and its associated scalar field.

Let V be a set of continuous volumes and I an integrable scalar map on V. V may be seen as a 3-current via:

$$V_I(\omega) = \int_V I(x)\omega(x) \det(u_1(x), \ldots, u_d(x)) d\lambda(x) = \int_V \omega(x) d\lambda(x) \tag{1.4.12}$$

for every scalar field ω, where $u_1(x), \ldots, u_d(x)$ is a positively oriented orthonormal basis of $\mathbb{R}^d$ and as such $\det(u_1(x), \ldots, u_d(x)) = 1$.

If the volumes V are volume meshes in 3D (whose mesh cells are denoted f_i), they can be approximated as

$$\tilde{V}_I = \sum_{f_i} \delta_{c_i}^{I_i v_i} \tag{1.4.13}$$

where c_i denotes the center of the polyhedron f_i, v_i the volume of the polyhedron f_i and $I_i = \frac{1}{v_i}\int_{f_i} I(x)d\lambda(x)$ the mean intensity in the volume f_i. The action of this 3-current on a vector field ω is given by: $\tilde{V}_I(\omega) = \sum_i I_i v_i \omega(c_i)$.

If ϕ is a diffeomorphism of $\mathbb{R}^d$, then the deformed current $\phi_* V_{I\circ\phi^{-1}}$ is given by

$$\begin{aligned}\phi_* V_{I\circ\phi^{-1}}(\omega) = V_I(\phi^*\omega) &= \int_V I(x)\omega(\phi(x))\det(d_x\phi u_1(x),\ldots,d_x\phi u_d(x))d\lambda(x)\\ &= \int_V I(x)\omega(\phi(x))\,|d_x\phi|\,d\lambda(x)\,.\end{aligned} \tag{1.4.14}$$

We notice in particular that this action is different from the one on the scalar field associated to a 0-current due to the Jacobian $|d_x\phi|$.

In the discrete case, we have:

$$\phi_*\left(\sum_i I_i\delta_{x_i}^{v_i}\right) = \sum_i I_i\delta_{x_i}^{|d_{c_i}\phi|v_i}\,. \tag{1.4.15}$$

Note that, like the 0-currents, 3-currents are build on scalar fields. But for 3-currents, the action of a diffeomorphism takes into account the deformation of the geometry of the volumes. The volume v_i is changed according to the determinant of the Jacobian of the deformation.

Remark 1.18 (On the notations)**.** The notation of Dirac Delta current δ_x^α in this context is misleading since it may denote 0-,1-,2- and 3-currents. Indeed α may be a scalar, a vector, the vector associated to a 2-multivector or the scalar associated to a 3-multivector. In particular, the reader has to keep in mind that the upper-script in this notation **has a unit**! For the 0-current δ_x^α, α is a pure scalar (without any unit) and it is not affected by the deformation. On the contrary, if δ_x^α denotes a 3-current, then the scalar α is the measure of a volume which is mutliplied by the Jacobian of the deformation when the current moves in space. Similarly, a vector α in the notation δ_x^α has the unit of a length for a 1-current and the unit of a surface for a 2-current. It is deformed by a diffeomorphism according to its dimension. □

Remark 1.19 (Take care of the dimension!)**.** In this section, we introduced m-currents for modeling sub-manifolds of dimension m in $\mathbb{R}^3$. However, it is possible to consider a manifold of dimension m as a collection of manifolds of lower dimension. For instance, curves, surfaces or volumes can be all considered as continuous point sets. Similarly, a surface may be described as dense collection of curves (a moving curve which sweeps the surface). In this example, the surface may be modeled as a single 2-current or as a sum of 1-currents. One must be aware that these two currents are *not* equivalent. They do not have the same geometrical properties. This is particularly visible when considering the deformation of the currents. The action of a diffeomorphism on the collection of curves deforms the surface in the direction of the curves only, whereas the action on the 2-current affects also the surface in the direction orthogonal to the curves.

Similarly, let $\mathcal{C}$ be a polygonal line $\{x_1,\ldots,x_n\}$. Modeled as a 1-current this curve is represented by $C_1 = \sum_{i=1}^{n-1}\delta_{x_i}^{x_{i+1}-x_i}$. Modeled as a collection of points, the curve is represented by the 0-current: $C_0 = \sum_{i=0}^{n}\delta_{x_i}$. Let ϕ be a diffeomorphism of the space.

The deformed 1-current is given by: $\phi_* C_1 = \sum_{i=1}^{n-1} \delta_{\phi(x_i)}^{d_{x_i}\phi(x_{i+1}-x_i)}$, whereas the deformed 0 current is written as: $\phi_* C_0 = \sum_{i=1}^{n} \delta_{\phi(x_i)}$. Only the first current $\phi_* C_1$ is the discretization of the deformed curve $\phi(\mathcal{C})$ (in particular the mass-norm of $\phi_* C_1$ is equal to the length of $\phi(\mathcal{C})$). By contrast, $\phi_* C_0$ does not take the tangential information into account. $\mathcal{C}$ and *the set of points on* $\mathcal{C}$ are two different objects: the one is of dimension 1, the other of dimension 0. As a conclusion, any current is associated to a particular dimension m. And there is no simple way to describe a m-current as a collection of $(m-1)$-currents while preserving the same geometrical properties.

In practice, the segmentation of an anatomical surfaces or volumes may be given as an unstructured point-set. In absence of surface or volume mesh, one has no other choice than modeling this set of points as a 0-current. One must be aware, however, that building a surface or a volume mesh from this point set would reconstruct the geometry of the anatomical structure and lead eventually to different measures of similarity, which would account for this richer geometrical information. □

Remark 1.20 (Why using m-forms instead of scalar/vector fields?)**.** There are some cases of practical interest which are not modeled with differential m-forms of dimension or co-dimension 0 or 1. For instance, one may be interested in the temporal evolution of a curve in 3D. This can be seen as a tubular surface ($m = 2$) in a 4D space ($d = 4$). In this case, the representation in terms of vector field is no more possible, whereas the framework based on currents still applies. However, from a computational point of view, the algorithms presented in this thesis take advantage of the representation in terms of vector/scalar fields. Dealing with other cases would require to develop new algorithms. □

1.5 The space of currents as a RKHS

1.5.1 Why the mass-norm is not adapted to measure shape dissimilarity

In the previous section, we introduced the mass-norm of a current. However, the following proposition shows that this norm cannot be used in practice for measuring shape dissimilarities. Indeed, this measure is insensitive to the relative distance between two sub-manifolds, as long as they do not intersect.

Proposition 1.21. *Let T and T' be two smooth compact sub-manifolds with disjoint supports. Then*

$$M(T - T') = M(T) + M(T')\,. \tag{1.5.1}$$

Let T and T' be two finite sums of Dirac delta currents located at different points. Then,

$$M(T - T') = M(T) + M(T')\,. \tag{1.5.2}$$

Proof. The proof is a generalization of the proofs of Proposition 1.6 and Proposition 1.12 and relies on the same idea. The triangle inequality states that $M(T - T') = \sup_{\|\omega\|_\infty \leq 1} |T(\omega) - T'(\omega)| \leq M(T) + M(T')$ in both cases. To show the equality, one finds a particular ω which enables to achieve the supremum. We focus on the case of sub-manifolds

since the case of Dirac currents is direct consequence of Proposition 1.12. For any x in T and T', we choose the m-form $\omega(x)$ such that $\omega(x)(u_1(x) \wedge \ldots \wedge u_m(x)) = |u_1(x) \wedge \ldots \wedge u_m(x)|$ and 0 in the directions orthogonal to $u_1(x) \wedge \ldots \wedge u_m(x)$ in $\Lambda^m \mathbb{R}^d$, where $u_i(x)$ denotes a basis of the tangent plane of T at point x. Since the sub-manifolds are smooth, we can choose basis vectors $u_i(x)$ which varies continuously on each sub-manifold, thus making $\omega(x)$ continuous on T and T'. Now, since T and T' are disjoint, we can interpolate continuously the values of $\omega(x)$ in-between T and T'. Since the sub-manifolds are compact, we can also make $\omega(x)$ tend to 0 at infinity. Therefore, this constructed ω belongs to $\mathcal{C}^0(\mathbb{R}^d, (\Lambda^m \mathbb{R}^d)^*)$ and achieves the supremum in the definition of the norm. ■

Remark 1.22. There is no contradiction between this proposition and Proposition 1.11 since the support of the discretized current $\tilde{T}$ is included into the support of the original current T. □

Remark 1.23 (support of a current)**.** This Proposition can be expressed in a more general form for any currents T and T' with disjoint supports. However, for the sake of simplicity, we do not define the support of a current and focus on two particular cases: sub-manifolds and Dirac delta currents. However, the definition of the support of a distribution [Schwartz 1966] extends straightforwardly to currents. □

Let us take two Dirac currents δ_x^α and δ_y^α which model two small mesh cells (infinitesimal surface for instance) with the same orientation but located at two distinct points x and y. The dissimilarity between these two currents is equal to $M(\delta_x^\alpha - \delta_y^\alpha) = 2\,|\alpha|$, thus meaning that the distance between both structures is constant as long as $x \neq y$. But once $x = y$, then $M(\delta_x^\alpha - \delta_y^\alpha) = 0$. This shows that the metric M is blind to shape dissimilarities until the two shapes are perfectly aligned! As a consequence, it is not possible to use this metric to drive the registration of one shape onto another. Moreover, the discontinuous behavior of the metric with respect to the positions x and y prevents us from using the mass-norm from a numerical point of view.

The main problem is that the mass-norm is the dual norm for the supremum norm for continuous differential m-form. The key argument in the proofs of Propositions 1.6, 1.12 and 1.21 is that we can always find a differential m-form ω which continuously interpolates between $\omega(x) = \alpha$ and $\omega(y) = -\alpha$ for two distinct points x and y. If we impose some constraints on the variations of ω, then it may not be possible to interpolate with bounded variations between arbitrary points x and y. Then the norm $M(T - T')$ would start to decrease smoothly to zero as T is approaching T'. This point is illustrated in Fig. 1.4.

This analysis justifies to change the test space of differential m-forms for a test space in which the differential m-forms are more regular (i.e. whose variations are bounded). Until now we have considered the space of continuous differential forms which tend to zero at infinity (i.e. such that $\|\omega\|_\infty < \infty$). We would like to consider now differential forms which are differentiable and such that the derivatives are controlled: $\|\omega\|_\infty + \|\nabla\omega\|_\infty < \infty$. This leads naturally to define our test space as Sobolev spaces. Most of these spaces are Reproducible Kernel Hilbert Space (RKHS) as shown in Appendix B. Let W denote this new test space of differential forms. The associated space of currents is denoted W^*: the space of continuous mapping from W to $\mathbb{R}$. Then the mass-norm ($M(T) = \sup_{\|\omega\|_\infty \leq 1} |T(\omega)|$) is

replaced by the dual norm: $\|T\|_{W^*} = \sup_{\|\omega\|_W \leq 1} |T(\omega)|$. The kernel of the RKHS acts as a low-pass filter on the space of square integrable differential forms: this excludes from W the differential forms with too high frequencies (i.e. whose variations are not bounded).

Using RKHS as test space of differential m-forms have compelling advantages. First, the choice of the kernel determines how smooth the differential forms are. In particular, we can set a scale parameter which determines the rate of decay of the RKHS norm ($\|T - T'\|_{W^*}$) to zero when T is "converging" to T'. See Figure 3.16 for instance. If this scale tends to zero, then the RKHS norm tends to the mass-norm. For a larger parameter, shape dissimilarities are captured up to this scale. Second, in the framework of RKHS, the space of currents is provided with an inner-product and the norm, which is defined by a supremum, has a closed form. This makes the overall framework particularly well suited from a computational point of view. Third, the RKHS norm in the space of currents allows us to define random Gaussian variables. This will allow us to define statistical models of currents to measure the variability of shapes.

Remark 1.24 (Flat norm)**.** To workaround the bad behavior of the mass-norm, one introduces often the flat-norm defined as:

$$F(T) = \sup_{\|\omega\|_\infty \leq 1, \|d\omega\|_\infty \leq 1} |T(\omega)|\,, \tag{1.5.3}$$

where $d\omega$ denotes the exterior derivative of the differential m-form ω (see [Federer 1969, Cohen-Steiner 2003a], for instance). This norm introduces explicitly a control on the variations of the differential forms. However, by contrast to RKHS norm, this norm does not derive from an inner-product and has no closed form. This norm is therefore difficult to use from a computational point of view. □

1.5.2 The space of currents as the dual space of a RKHS

In this section, we adapt the construction of currents in Section 1.3 to a test space of differential forms which is a RKHS. We will show also how this framework enables to have a norm which derives from an inner-product and for which we have a closed-form.

For the sake of simplicity, we suppose, from now on, that we deal only with the practical cases of Section 1.4: sub-manifolds of dimension $0, 1, 2$ and 3 in $\mathbb{R}^3$ and sub-manifold of dimension $0, 1$ and 2 in $\mathbb{R}^2$. In all these cases, the differential m-forms can be seen as a scalar or vector field: a continuous mapping from $\mathbb{R}^d$ ($d = 2$ or $d = 3$) to $\mathbb{R}^p$ ($p = 0$ or $p = d$). The case $p = 0$ is for the scalar fields, the case $p = d$ is for the vector fields.

Note that the following construction could be made also for general differential m-forms (see Remark 1.28). However, this would involve more sophisticated notations without any benefits for the targeted applications.

1.5.2.1 A new space of currents

Let W be a Hilbert space of vector fields in which the vector fields are continuous and verifies:

$$\|\omega\|_\infty \leq C_W \|\omega\|_W \tag{1.5.4}$$

As shown in Appendix B, this condition makes W a RKHS (see Proposition B.4). From a numerical point of view, this condition implies that numerical error measured in the space W are numerically small.

We denote W^* the dual space of W, now our space of currents. The following proposition shows that the previous space of currents (i.e. the dual space of $\mathcal{C}^0(\mathbb{R}^d, (\Lambda^m\mathbb{R}^d)^*)$) is continuously embedded into W^*.

Proposition 1.25. *If T is a continuous linear form on $\mathcal{C}^0(\mathbb{R}^d, (\Lambda^m\mathbb{R}^d)^*)$ then it is a continuous linear form on W and*

$$\|T\|_{W^*} \leq C_W M(T). \tag{1.5.5}$$

Proof. By definition of a continuous linear map, we have that there is a constant C_T such that:

$$|T(\omega)| \leq C_T \|\omega\|_\infty , \tag{1.5.6}$$

which leads according to Eq. (1.5.4)

$$|T(\omega)| \leq C_T C_W \|\omega\|_W . \tag{1.5.7}$$

This shows that the T can be considered as current in W^*.

Moreover, we have:

$$\begin{aligned} \|T\|_{W^*} &= \sup_{\|\omega\|_W \leq 1} |T(\omega)| = \sup_{\|\omega\|_W \neq 0} \left| T\left(\frac{\omega}{\|\omega\|_W}\right)\right| \\ &\leq C_W \sup_{\|\omega\|_W \neq 0} \left| T\left(\frac{\omega}{\|\omega\|_\infty}\right)\right| \\ &\leq C_W \sup_{\|\omega\|_\infty \neq 0} \left| T\left(\frac{\omega}{\|\omega\|_\infty}\right)\right| = C_W M(T) \end{aligned} \tag{1.5.8}$$

■

This proposition shows that the RKHS norm on the space of currents is more precise than the mass-norm. It shows also that the currents introduced in the previous sections can be considered as currents in W^*. In particular, the Dirac delta currents as introduced in Definition 1.9 are currents in W^*. Proposition 1.4 which shows that a sub-manifold of $\mathbb{R}^d$ defines a current is still valid in this framework: under the same conditions sub-manifolds define a current in W^*. Indeed the inequality in Eq. (1.5.5) shows that a linear form on W which is continuous with respect to the $\|.\|_\infty$ norm is also continuous with respect to the W-norm. Proposition 1.11, which enables to approximate m-meshes T with finite sum of Dirac delta currents $\tilde{T}$, is also valid in W^*. Indeed, the proof relies on the inequality: $\left|T(\omega) - \tilde{T}(\omega)\right| \leq \varepsilon \|\omega\|_\infty$ for all $\omega \in \mathcal{C}^0(\mathbb{R}^d, (\Lambda^m\mathbb{R}^d)^*)$. This implies that for all $\omega \in W$, $\left|T(\omega) - \tilde{T}(\omega)\right| \leq \varepsilon C_W \|\omega\|_W$ and therefore that $\left\|T - \tilde{T}\right\|_{W^*} \leq \varepsilon C_W$ which tends to zero under the same conditions as in Proposition 1.11.

Eventually, the pull-back action on continuous differential m-forms (Definition A.11) extends straightforwardly to differential m-form in W under the same condition as in Definition A.11. Indeed, if ϕ is a diffeomorphism such that $\sup_{x\in\mathbb{R}^d} |d_x\phi| < \infty$, then

$|\phi^*\omega(x)| \leq \|\omega\|_\infty \|d_x\phi\|_\infty \leq C_{x,\phi} \|\omega\|_W$ which proves that the pullback vector field $\phi^*\omega$ belong to W. Therefore, we can define the push-forward action on currents as in Definition 1.15, which is now an action on the space of currents W^*.

1.5.2.2 Norm and inner-product in the RKHS

It is shown in Appendix B that there is an isometric mapping between the test space of vector/scalar field W and the space of current W^*. This mapping is denoted $\mathcal{L}_W$ and the kernel of the RKHS W is denoted K^W. This provides the space of currents W^* with a Hilbert structure. In this section, we show how this inner-product allows us to give a closed form to the norm $\|.\|_{W^*}$ in practical cases.

Proposition B.9 shows that the map $\mathcal{L}_W$ is isometric. Therefore the norm of a current $T \in W^*$ satisfies:

$$\|T\|_{W^*} = \sup_{\|\omega\|_W \leq 1} |T(\omega)| = \left\|\mathcal{L}_W^{-1}(T)\right\|_W , \tag{1.5.9}$$

and that the supremum is achieved for the vector field $\omega = \mathcal{L}_W^{-1}(T)$, which implies that $\|T\|_{W^*}^2 = T(\mathcal{L}_W^{-1}(T))$.

Now, we will show how to compute $\mathcal{L}_W^{-1}(T)$ and the norm of T when T is a sub-manifold or a finite set of Dirac delta currents. Assume first that T is a sub-manifold of $\mathbb{R}^3$ which can be seen as a current under the assumptions of Proposition 1.4 (assuming that the attribute map $I = 1$). We focus on the vectorial case (T is dimension 1 or 2) but the computations can be very easily adapted in the scalar case (T of dimension 0 or 3). Let $\omega \in W$ be a vector field, then by definition of $T(\omega)$ in Prop. 1.4 and thanks to the reproducing property in W (see Eq. (B.2.5)), we have:

$$\begin{aligned} T(\omega) &= \int_T \omega(x)^t \alpha(x) d\lambda(x) \\ &= \int_T \left\langle \omega, K^W(x,.)\alpha(x) \right\rangle_W d\lambda(x) \\ &= \left\langle \int_T K^W(x,.)\alpha(x) d\lambda(x), \omega \right\rangle_W \end{aligned} \tag{1.5.10}$$

where $\alpha(x)$ denotes the unit tangent of T at x if T is of dimension 1 or the unit normal of T at x if T is of dimension 2.

In section B.3 (Eq. (B.3.5)), it is shown that:

$$T(\omega) = \langle T, \mathcal{L}_W(\omega) \rangle_{W^*} \tag{1.5.11}$$

The combination of this equation with Eq. (1.5.10) shows that the mapping $\mathcal{L}_W^{-1}$ can be computed explicitly in this case as:

$$\mathcal{L}_W^{-1}(T)(x) = \int_T K^W(x,y)\alpha(y) d\lambda(y) \tag{1.5.12}$$

The representation of the sub-manifold in the space of vector fields W is given as the convolution between the kernel K^W and the dense field of its tangents or normals.

The application of the isometric mapping $\mathcal{L}_W$ (see Eq. (1.5.9)) leads to the norm of T as:

$$\begin{aligned}\|T\|^2_{W^*} &= \left\langle \mathcal{L}_W^{-1}(T), \mathcal{L}_W^{-1}(T) \right\rangle_W = \left\langle \int_T K^W(x,.)\alpha(x)d\lambda(x), \int_T K^W(x,.)\alpha(x)d\lambda(x) \right\rangle_W \\ &= \int_T \int_T \left\langle K^W(x,.)\alpha(x), K^W(y,.)\alpha(y) \right\rangle_W d\lambda(x)d\lambda(y) \\ &= \int_T \int_T \alpha(x)^t K^W(x,y)\alpha(y)d\lambda(x)d\lambda(y),\end{aligned} \tag{1.5.13}$$

by linearity of the integration and thanks to the reproducing property of the kernel (see Eq (B.2.5)).

Therefore, the Hilbert norm of the sub-manifold seen as a current is given by the double integration of the kernel on the manifold. Similar computations show that the inner-product between two sub-manifolds of the same dimension, T and T', is given by:

$$\langle T, T' \rangle_{W^*} = \int_T \int_{T'} \alpha'(y)^t K^W(y,x)\alpha(x)d\lambda(y)d\lambda(x), \tag{1.5.14}$$

where $\alpha(x)$ (resp. $\alpha'(x)$) denotes the tangent or normal of T (resp. T') at point x.

If the sub-manifold T is given as a m-mesh, it can be approximated, in the sense of Proposition 1.11 by a finite set of Dirac Delta currents: $T = \sum_{i=1}^n \delta_{x_i}^{\alpha_i}$. In this case, using Eq. (B.3.9), we have also explicit formulation for the representation of the current as a vector field:

$$\mathcal{L}_W^{-1}(T)(x) = \sum_{i=1}^n \mathcal{L}_W^{-1}(\delta_{x_i}^{\alpha_i})(x) = \sum_{i=1}^n K^W(x,x_i)\alpha_i, \tag{1.5.15}$$

also for the norm of this current:

$$\|T\|^2_{W^*} = T(\mathcal{L}_W^{-1}(T)) = \sum_{j=1}^n \delta_{x_j}^{\alpha_j}(\mathcal{L}_W^{-1}(T)) = \sum_{i=1}^n \sum_{j=1}^n \alpha_j^t K^W(x_j,x_i)\alpha_i \tag{1.5.16}$$

and the inner-product between T and $T' = \sum_{j=1}^m \delta_{y_j}^{\beta_j}$

$$\langle T, T' \rangle_{W^*} = \sum_{i=1}^n \sum_{j=1}^m \beta_j^t K^W(y_j,x_i)\alpha_i. \tag{1.5.17}$$

The comparison of these last two equations with Equations (1.5.13) and (1.5.14) shows that the approximation of Proposition 1.11 consists in replacing the continuous integrals by their Riemann sums. In other words, the continuous current T is decomposed into its *infinite* set of tangents or normals, whereas its discretization in the space of currents consists in sampling this field of tangents/normals. When the sampling becomes finer, the approximation converges in the space of currents. This shows that using RKHS enables to compare continuous objects with their discrete representation. This is a direct consequence of the construction of the RKHS as the completion of a pre-Hilbert space in the proof of Theorem B.6. This guarantees that the metric on currents is weakly sensitive to the sampling of the geometrical data, since each sampling is an approximation of the same continuous quantity. This ensures the robustness and numerical stability of this metric used in practical algorithms.

Table 1.2 summarizes these operations on currents for any m-currents in $3D$.

Remark 1.26. In this section we omit the scalar function I which can model attributes attached on the sub-manifold T. Taking into account such map induces only very slight changes in Eqs. (1.5.10) to (1.5.17). □

Remark 1.27 (Computational cost of the distance between currents)**.** In practice, we need mainly to compute the distance $\|T - T'\|_{W^*}$ between two set of Dirac Delta currents: $T = \sum_{i=1}^{N} \delta_{x_i}^{\alpha_i}$ (with N terms) and $T' = \sum_{j=1}^{N'} \delta_{x'_j}^{\alpha'_j}$ (with N' terms). For this purpose, one can use the equality: $\|T - T'\|_{W^*}^2 = \|T\|_{W^*}^2 + \|T'\|_{W^*}^2 - 2\langle T, T'\rangle_{W^*}$. Or one can form the current $S = \sum_{i=1}^{N+N'} \delta_{y_i}^{\beta_i}$ where $(y_i = x_i, \beta_i = \alpha_i)$ if $i = 1 \dots N$ and $(y_i = x'_i, \beta_i = -\alpha'_i)$ if $i = N{+}1, \dots, N{+}N'$ (i.e. the concatenation of the list of points (x,x') and the list of vectors $(\alpha,{-}\alpha')$) and compute directly $\|S\|_{W^*}$. The first solution requires to sum $N^2{+}(N')^2{+}NN'$ terms, whereas the second solution requires to sum $(N + N')^2 = N^2 + (N')^2 + 2NN'$ terms. The first solution is computationally less expensive than the second one. However, in chapter 2 we will present a computational framework which enables to compute these double sums at almost a constant cost, independently of the number of terms. This will make eventually the second solution about 3 times faster than the first one (3 double sums to be computed versus 1). □

1.5.2.3 Three norms on the space of currents

We can define 3 norms in the space of currents W^*. We introduced previously the mass-norm:

$$M(T) = \sup_{\|\omega\|_\infty \leq 1} |T(\omega)| \tag{1.5.18}$$

and the W^*-norm:

$$\|T\|_{W^*} = \sup_{\|\omega\|_W \leq 1} |T(\omega)|\,, \tag{1.5.19}$$

where T denotes a generic current in W^* and ω a generic scalar/vector field in W.

We saw that the mass norm generalizes the notion of volume of a m-dimensional sub-manifold and that this norm is not adapted to measure shape dissimilarity. The W^*-norm is a regularized version of the mass-norm which will allow us to drive the registration of shapes and define statistical models on shapes. Moreover, this norm is easily computable in practical cases.

In addition to these two norms, we can define the L^∞-norm of currents as:

$$\|T\|_\infty = \left\|\mathcal{L}_W^{-1}(T)\right\|_\infty, \tag{1.5.20}$$

where the norm on the right-hand side is the supremum norm of the scalar/vector field associated to the current T (this scalar/vector field belongs to W and therefore is continuous and tend to zero at infinity).

This last norm raises naturally when one considers the current T via the inner-products $\langle T, \delta_x^{\epsilon_k}\rangle_{W^*}$ for any points $x \in \mathbb{R}^d$ (ϵ_k denotes the canonical basis of $\mathbb{R}^3$ for vector fields and $\mathbb{R}$ for scalar field). Since $\langle T, \delta_x^{\alpha}\rangle_{W^*} = \mathcal{L}_W^{-1}(T)(x)^t\alpha$, we have the following property:

$$\|T\|_\infty = \sup_{|\alpha|=1, x\in\mathbb{R}^d} |\langle T, \delta_x^{\alpha}\rangle_{W^*}| \tag{1.5.21}$$

Norm on vector fields	Norm on currents	Comments
$\|\omega\|_\infty = \sup_{x\in\mathbb{R}^3} \|\omega(x)\|$	$M(T) = \sup_{\|\omega\|_\infty \leq 1} \|T(\omega)\|$	If T is a submanifold, $M(T)$ is the length, area, volume of T
$\|\omega\|_W$ (regularized L^2-metric)	$\|T\|_{W^*} = \sup_{\|\omega\|_W \leq 1} \|T(\omega)\|$	derives from an inner-product
	$\begin{aligned}\|T\|_\infty &= \left\|\mathcal{L}_W^{-1}(T)\right\|_\infty \\ &= \sup_{\substack{\|\alpha\|=1 \\ x\in\mathbb{R}^3}} \|\langle T, \delta_x^\alpha\rangle_{W^*}\|\end{aligned}$	corresponds to the weak-topology on W^*

Relations between norms:
$$\begin{aligned}&\|\omega\|_\infty \leq C_W \|\omega\|_W \\ &\|T\|_{W^*} \leq C_W M(T) \\ &\|T\|_\infty \leq \left(\sup_{x\in\mathbb{R}^3} \left|K^W(x,x)\right|\right) \|T\|_{W^*}\end{aligned}$$

Table 1.1: The norms on currents and vector fields

Since the span of the Dirac delta currents is dense in W^*, this equation shows that a sequence of currents T_n converges to 0 with respect to the L^∞-norm ($\|T_n\|_\infty \to 0$) if and only if $\langle T_n, T'\rangle_{W^*} \to 0$ for every currents $T' \in W^*$. This shows that the L^∞-norm is associated to the weak topology of the RKHS.

As a consequence, we can control the L^∞- by the W^*-norm. Indeed, we have:

$$|\langle T, \delta_x^\alpha\rangle_{W^*}| \leq \|T\|_{W^*} \|\delta_x^\alpha\|_{W^*} . \tag{1.5.22}$$

Moreover, $\|\delta_x^\alpha\|_{W^*}^2 = \alpha^t K^W(x,x)\alpha \leq \sup_{x\in\mathbb{R}^3} \left|K^W(x,x)\right| |\alpha|^2$ where $\left|K^W(x,x)\right|$ denotes the spectral norm of the matrix $K^W(x,x)$. Therefore, $\sup_{|\alpha|=1} |\langle T, \delta_x^\alpha\rangle_{W^*}| \leq \sup_{x\in\mathbb{R}^3} \left|K^W(x,x)\right|^{1/2} \|T\|_{W^*}$, which leads to:

$$\|T\|_\infty \leq \left(\sup_{x\in\mathbb{R}^3} \left|K^W(x,x)\right|^{1/2}\right) \|T\|_{W^*} \tag{1.5.23}$$

Both L^∞- and W^*-norm will be used in the next chapters, for instance to control the convergence of the matching pursuit in Chapter 3.

Remark 1.28 (RKHS of differential forms)**.** In this section, we provide the space of currents with a norm in the case of currents which can be represented by scalar or vector fields. This has been done for the sake of simplicity. We could have built also a RKHS of differential m-forms. The reproducing property would have been: $\omega(x)(u_1, \ldots, u_m) = \langle \omega, K(x,.)(u_1, \ldots, u_m)\rangle_W$ and the kernel K would have been a m-covariant, m-contravariant tensor. The construction would have been very similar to the one presented here. □

Remark 1.29 (Units)**.** The physical objects defined in this chapter (differential form, vector fields, currents, etc.) have dimension and therefore their measure depends on the choice of the unit of the ambient Euclidean space $\mathbb{R}^2$ or $\mathbb{R}^3$. The action of a m-form on a set

of m vectors leads to a real number which has no dimension and which is invariant under a change of coordinate. Since the vectors have the dimension of a length, denoted here L, a m-form is of dimension L^{-m}. A current integrates a m forms to give an adimensional real number: it is of dimension L^m. Therefore, it has the same dimension that the geometric object, which it models. The norm of the current $\|T\|_{W^*} = \sup_{\omega \neq 0} |T(\omega)| \, \|\omega\|$ has the dimension L^m, namely the dimension of a length for a 1-current and the dimension of a area for a 2-current. The map $\mathcal{L}_W$ from W to W^* depends also on the choice of units: it is of dimension $L^m/L^{-m} = L^{2m}$. The inverse map is implemented by the matrix $K^W(x, y)$ whose elements is of dimension L^{-2m}.

In Section 1.2 and 1.5 as well as in Appendix A, we identify differential 1-forms and vector fields, whereas the former is of dimension L^{-1} and the later of dimension L. Therefore, this identification depends on the choice of the units in the ambient Euclidean space: if the units change, the vector field associated to the 1-form also changes. This is well-known in differential geometry: the form is a covariant tensor, the vector field a contravariant tensor. One transforms the former into the later by applying the metric tensor, which is of dimension L^2. Therefore, these kind of identification can be made only once the units are fixed. In our applications, units are given by the imaging device. □

1.5.3 Random Gaussian Currents

Defining the test space W as a RKHS has another advantage: it allows us to define random Gaussian currents. Indeed, there is a standard way to define random Gaussian variables in infinite-dimensional Hibert spaces, so that their projection on any finite-dimensional subspace is a usual multi-variate Gaussian variable. In this setting, the kernel of a RKHS gives the covariance of the random variables. In this section, we show how such variables can be defined. In Section 2.3.3, we will show that the projection of the Gaussian currents on finite-dimensional spaces are usual Gaussian vectors.

First, we will show how the finite-dimensional case can be generalized to define infinite-dimensional random variables. Let Γ be a zero-mean Gaussian vector in $\mathbb{R}^n$ with covariance K. By definition, for any n-dimensional vector ω, $\omega^t \Gamma$ is a zero-mean Gaussian real variable with variance $\omega^t K \omega$. This shows that the covariance matrix K can be seen as a metric on $\mathbb{R}^n$. Actually, one can consider that the Gaussian variable Γ maps every eigenvectors $(e_i)_{i=1,\ldots,n}$ of K to an independent Gaussian real variables γ_i with variance given by the eigenvalues λ_i^2. Therefore, any vector ω is mapped into $\omega^t \Gamma = \sum_{i=1}^{n} \omega^t e_i \gamma_i$. This shows that a given metric K determines a Gaussian vector Γ. This idea of mapping each eigenvector of the metric to an independent Gaussian variable can be generalized in infinite-dimension as follows.

A current T is a linear mapping from a test space of vector field W (with metric K^W) to the space of real numbers $\mathbb{R}$. A random Gaussian current is a linear mapping from the test space W to a Gauss space $\mathcal{G}$[2]. This means that a random Gaussian current maps

[2] a Gauss space is a set of random Gaussian variables. The typical example of a Gauss space is the linear span of N independent Gaussian variables.

every vector field ω to a real random Gaussian variable $X^*(\omega)$, whereas a deterministic current maps the test vector field ω to a real number $T(\omega)$. We define this mapping such that the random current X^* tested on two *orthogonal* vector fields ω and ω' leads to two *independent* Gaussian variables. X^* can be seen as the Gaussian variable associated to the RKHS W with kernel K^W. However, since it is infinite-dimensional, it has no probability density function.

To give a precise definition, we assume that the test space W is separable. As such, W can be provided with an orthogonal basis w_n. We define the linear mapping X^* from W to $\mathcal{G}$ via its value on the orthogonal basis $(w_n)_{n=0,\dots,\infty}$: we set $X^*(w_n) = \gamma_n$, where γ_n is an infinite sequence of independent normal variables (zero mean and unit variance). Since every vector field ω can be decomposed into $\omega = \sum_{n=0}^{\infty} \langle \omega, w_n \rangle_W w_n$ such that $\|\omega\|_W^2 = \sum_{n=0}^{\infty} |\langle \omega, w_n \rangle_W|^2 < \infty$, we have by linearity of the mapping X^*: $X^*(\omega) = \sum_{n=0}^{\infty} \langle \omega, w_n \rangle_W \gamma_n$. Therefore for all $\omega \in W$:

$$\begin{aligned} &\mathbb{E}\left(X^*(\omega)\right) = 0 \\ &\mathbb{E}\left(X^*(\omega)^2\right) = \sum_{n=0}^{\infty} |\langle \omega, w_n \rangle_W|^2 = \|\omega\|_W^2 \end{aligned} \tag{1.5.24}$$

Moreover, given two vector fields ω and ω', we have:

$$\begin{aligned} \mathbb{E}\left(X^*(\omega)X^*(\omega')\right) &= \mathbb{E}\left(\left(\sum_{n=0}^{\infty} \langle \omega, w_n \rangle_W \gamma_n\right)\left(\sum_{m=0}^{\infty} \langle \omega, w_m \rangle_W \gamma_m\right)\right) \\ &= \sum_{n=0}^{\infty} \langle \omega, w_n \rangle_W \langle \omega', w_n \rangle_W = \langle \omega, \omega' \rangle_W . \end{aligned} \tag{1.5.25}$$

These intrinsic definitions of the variance and covariance of real variables $X^*(\omega)$ in Equations (1.5.24) and (1.5.25) shows that the definition of the random Gaussian current does not depend on the choice of the basis on W. This leads to the following definition:

Definition 1.30. *Let W be a separable Hilbert space of vector fields and $\mathcal{G}$ a Gaussian space. Let X^* be the isometric mapping between W and $\mathcal{G}$:*

$$\begin{array}{rccc} X^*: & W & \longrightarrow & \mathcal{G} \\ & \omega & & X^*(\omega) \end{array} \tag{1.5.26}$$

such that for all $\omega, \omega' \in W$:

$$\begin{aligned} &\mathbb{E}\left(X^*(\omega)\right) = 0 \\ &\mathbb{E}\left(X^*(\omega)X^*(\omega')\right) = \langle \omega, \omega' \rangle_W \end{aligned} \tag{1.5.27}$$

If the test space W is a RKHS, then we can test the Gaussian current on the basis vector $K^W(x,.)\alpha$. This leads to:

$$\mathbb{E}\left(X^*\left(K^W(x,.)\right)X^*\left(K^W(y,.)\beta\right)\right) = \alpha^t K^W(x,y)\beta. \tag{1.5.28}$$

This shows that the kernel determines the covariance of the random Gaussian currents. If the kernel is diagonal, in the sense $K(x,y) = 0$ if $x \neq y$, then the vectors $K(x,.)\alpha$ build an

orthogonal basis of the RKHS W. This is not the case in general but we can always build an orthogonal basis from these vectors via the Gram-Schmidt process.

This definition constructs theoretically a random Gaussian current with zero mean and a covariance structure determined by the kernel. However, in absence of probability function, there is no simple way to simulate some instances of this random current. In the next chapter, we will introduce finite-dimensional spaces to approximate the space of currents. The projections of this random current on these subspaces have a probability density function (pdf) as we will show in Section 2.3.3 and can be simulated numerically as we shall show and discuss in Section 3.5.

1.6 Conclusion

In this chapter, we gave a rather general presentation of currents. We discussed the ability of this framework to model a large range of geometrical objects, some with local orientation such as curves or surfaces, other without local orientation like point sets or volumes, some defined in the continuous domain, other defined as discrete structures, possibly provided with scalar attributes. Eventually, we provided the space of currents with a RKHS norm and inner-product. This offers a way to adapt the metric on currents to every particular anatomical data.

However, dealing with discrete currents, such as the ones resulting from the approximation in Proposition 1.11, may be computationally expensive, especially when the number of mesh cells increases, as mentioned in Remark 1.27. The following chapters are precisely dedicated to the design of an efficient numerical framework for currents.

From now on, we focus on the cases of curves and surfaces modeled as 1- and 2-currents, as the most interesting cases for our applications. For the sake of simplicity, we will use also 0-currents in some synthetic examples. We will use the scalar attributes to weight different parts of the currents as in Chapter 7, but we will not use such attributes for modeling non-geometrical properties of the shapes, according to Remark 1.14.

Dimension m	0-current	1-current	2-current	3-current
Differential m-form $x \to \omega(x)$	$a \in \mathbb{R} \to a\omega(x)$ scalar field	$\alpha \to \omega(x)^t \alpha$ vector field	$(\alpha, \beta) \to \omega(x)^t(\alpha \times \beta)$ vector field	$(\alpha, \beta, \gamma) \to$ $\det(\alpha, \beta, \gamma)\omega(x)$ scalar field
Sub-manifolds as continuous currents	set of points $\{x_1, \ldots, x_N\}$	param. curve $t \to L(t)$	param. surface $(u, v) \to S(u, v)$	Volume
Example of discrete currents	$\sum_i \delta_{x_i}^{a_i}$ a_i: scalars	$\sum_i \delta_{x_i}^{\alpha_i}$ α_i: vectors	$\sum_i \delta_{x_i}^{\alpha_i \times \beta_i}$ $\alpha_i \times \beta_i$: vectors	$\sum_i \delta_{x_i}^{v_i}$ $v_i = \det(\alpha_i, \beta_i, \gamma_i)$: scalars
Action on diff. forms (sub-manifolds)	$T(\omega) =$ $\sum_i a_i \omega(x_i)$	$T(\omega) =$ $\int_L \omega\left(L(t)\right)^t \frac{dL(t)}{dt} dt$	$T(\omega) =$ $\int_S \omega\left(S(u,v)\right)^t \left(\frac{\partial S}{\partial u} \times \frac{\partial S}{\partial v}\right) dudv$	$T(\omega) =$ $\int_V \omega(x) dx$
Action on diff. forms (discrete currents)	$\sum_i a_i \omega(x_i)$	$\sum_i \omega(x_i)^t \alpha_i$	$\sum_i \omega(x_i)^t (\alpha_i \times \beta_i)$	$\sum_i v_i \omega(x_i)$
Push-forward action	$\phi_* T(\omega) =$ $T(\omega \circ \phi)$	$\phi_* T(\omega) =$ $T(d_x\phi^t \omega \circ \phi)$	$\phi_* T(\omega) =$ $T\left(\lvert d_x\phi \rvert\, d_x\phi^{-1} \omega \circ \phi\right)$	$\phi_* T(\omega) =$ $T\left(\lvert d_x\phi \rvert\, \omega \circ \phi\right)$
Push-forward (discrete currents)	$\phi_*\left(\delta_x^a\right) = \delta_{\phi(x)}^a$	$\phi_*\left(\delta_x^\alpha\right) = \delta_{\phi(x)}^{d_x\phi(\alpha)}$	$\phi_*\left(\delta_x^{\alpha\times\beta}\right) = \delta_{\phi(x)}^{d_x\phi(\alpha)\times d_x\phi(\beta)}$ $= \delta_{\phi(x)}^{\lvert d_x\phi \rvert d_x\phi^{-t}(\alpha\times\beta)}$	$\phi_*\left(\delta_x^v\right)$ $= \delta_{\phi(x)}^{\lvert d_x\phi \rvert v}$
Inner-product btw sub-manifolds	$\sum_{i,j} a_i K(x_i, x_j) a_j$	$\iint_{L\times L'} \tau(t)^t K(L(t), L'(u)) \tau'(u) dtdu$ $\tau(t) = \frac{dL(t)}{dt} \quad \tau'(u) = \frac{dL'(u)}{du}$	$\iint_{S\times S'} n(x)^t K(S(x), S'(y)) n'(y) dxdy$ $n(x) = \frac{\partial S(x)}{\partial x_1} \times \frac{\partial S(x)}{\partial x_2}$ $n'(y) = \frac{\partial S'(y)}{dy_1} \times \frac{\partial S'(y)}{\partial y_2}$	$\iint_{V\times V'} K(x, y) dxdy$
Inner-product btw discrete currents	$\sum_{i,j} a_i K(x_i, x_j) a_j$	$\sum_{i,j} \tau_i^t K(x_i, y_j) \tau_j'$	$\sum_{i,j} n_i^t K(x_i, y_j) n_j'$	$\sum_{i,j} v_i K(x_i, y_j) v_j'$

Table 1.2: Summary of the operations on m-currents in $3D$, for $m = 0, \ldots, 3$

Chapter 2

Numerical tools for computing with currents and vector fields

Contents

The purpose of this chapter is to define an efficient computational framework to deal with currents and vector fields. The main idea is to introduce linearly spaced grids to define finite-dimensional sub-spaces of the RKHS of vector fields W and of the space of currents W^*. The usual operations on vector fields and currents in these discrete spaces can be performed efficiently via FFT and standard linear algebra operations. Then, we introduce a stable numerical scheme to project arbitrary currents on these approximation spaces. The resulting approximation error is shown to converge to zero as the step of the grid tends to zero. We provide an estimation of the speed of convergence. This offers a way to determine the step of the grid which guarantees a fixed approximation error, independently of the data to be processed.

This framework is the core tool for the implementation of the algorithms presented in this thesis. It serves in particular to define the discrete version of the matching pursuit algorithm in Chapter 3, to optimize the registration of currents in Chapter 4, to implement the atlas estimation in Chapter 5 and its 4D-extension in Chapter 9.

For instance, registration of anatomical data would not possible with real data without such an optimization framework. The original implementation in [Glaunès 2005] uses Fast Gauss Transform (FGT). We will show that the framework we proposed here is more robust and more generic, especially because the approximation error does not depend on the position of the points to be matched. More importantly, this framework is not application-dependent and can be used in different contexts to derive fast and robust algorithms based on currents and vector fields.

2.1 Introduction of lattices in the framework of currents

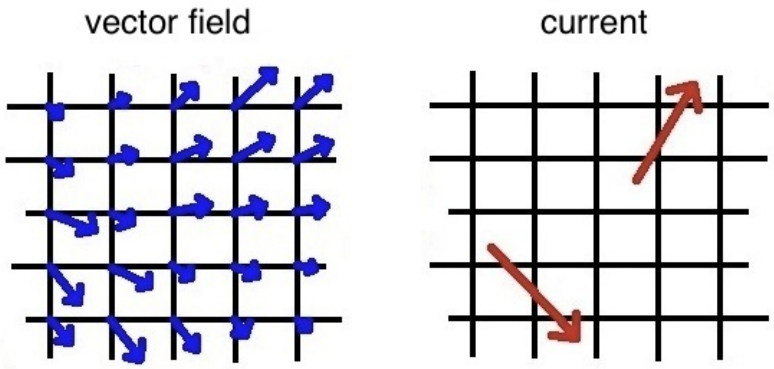


Figure 2.1: Principle of grid approximation: on the right hand side, the momenta of the currents are defined in continuous coordinates, whereas its associated vector field on the left hand side is computed only at the nodes of a fixed grid. In the framework of RKHS, this can be seen as the projection of the dense vector field on finite-dimensional spaces. On these spaces, usual operations on currents can be efficiently computed using FFT. We show in particular that these approximated operations converge to their true value as the grid step tends to zero.

In the previous chapter, we showed that a current, which models polygonal lines or surface meshes, is given by a finite combination of Dirac delta currents: $T = \sum_{i=1}^{N} \delta_{x_i}^{\alpha_i}$. Computing the norm of this current (and hence the inner-product with another current of the same form) can be done very easily via a double sum: $\|T\|_{W^*}^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i^t K(x_i, x_j)\alpha_j$ (see Eq. (1.5.16) and Eq. (1.5.17)) which can be written in matrix form as $\boldsymbol{\alpha}^t \mathbf{K} \boldsymbol{\alpha}$. This offers a way to efficiently compute the metric on currents via matrix multiplication when the number of momenta in the decomposition of currents is finite. However, the main problem is that the matrix $\mathbf{K}$ depends on the positions of the points x_i and therefore on the current T. This means that we need to compute and store a different matrix $\mathbf{K}$ for each current. By contrast, the purpose of this chapter is to define a discretization framework of currents, in which the momenta can be located only at some fixed positions. Therefore, the matrix $\mathbf{K}$ can be pre-computed and stored once for all, thus allowing for a fast parallelization of the

computation on currents. Moreover, if the discretization points build a regular lattice, it is possible to use FFT to compute this matrix multiplication, which reduces the complexity of the computations from $\mathcal{O}(N^2)$ to $\mathcal{O}(N\log(N))$, where N is the number of momenta in the input current.

To define such a discretization framework, we get inspired by the Hamiltonian particle-mesh method [Frank 2002, Frank 2003] which were further developed to give the variational particle-mesh methods proposed in [Cotter 2008, Cotter 2009]. These methods propose a semi-Lagrangian and semi-Eulerian numerical scheme for integrating partial differential equations. The discretization is given by as set of moving particles in continuous coordinates, whereas the velocity field which drives the evolution of the particle is computed on a fixed lattice. An interpolation scheme enables to extend the velocity at the grid nodes to the particle positions.

In this chapter, we adapt this idea to give efficient numerical scheme to compute with currents. The positions of the momenta of an arbitrary current correspond to the moving particles of the variational-particle mesh method. They are given in continuous coordinates. Normally, the vector field associated to this current is dense. Here, we restrict this vector field to be in a discrete space, so that it is entirely characterized by its samples at the grid nodes. When needed, an interpolation scheme enables to reconstruct a dense vector field from these samples.

In our framework based on RKHS, the introduction of grids can be considered as an approximation of the *infinite*-dimensional RKHS by *finite*-dimensional sub-spaces. As the grid step tends to zero, the finite-dimensional sub-spaces tend to the original RKHS, thus making the result of the operations performed in the discrete framework to converge to their true value. A complete computational framework should include not only the operations in the discrete spaces but also a way to project the data given in the continuous spaces to the discrete spaces.

The chapter is organized as follows. In Section 2.2, we introduce the discrete approximation spaces and show how operations on currents can be performed efficiently in this setting. In Section 2.3, we use orthogonal projections to map continuous currents to discrete spaces. This operation, unfortunately, is ill-posed from a numerical point of view. Therefore, in Section 2.4, we propose to use closest neighbors strategies to compute this projection, which converges at a polynomial speed with respect to the grid step. Eventually, Section 2.5 shows how this framework can be used in practice.

2.2 Finite-dimensional approximations of RKHS

2.2.1 Construction of sub-RKHS

The key idea for approximating currents is to define sub-spaces of the general space of currents W in which momenta are constrained to be located in a particular set of points Λ. We show then that when the set Λ becomes dense in $\mathbb{R}^d$, the approximation space converges to the original space.

Definition 2.1. *Let W be a RKHS of mappings from $\mathbb{R}^d$ to $\mathbb{R}^p$ with kernel K. Let Λ be a*

subset of $\mathbb{R}^d$. *We define the vector space* W_Λ *as:*

$$W_\Lambda = \overline{\text{Span}}\,\{K(x,.)\alpha; \quad x \in \Lambda, \alpha \in \mathbb{R}^p\} \tag{2.2.1}$$

($\overline{\text{Span}}$ *means the closure of the set spanned by the vectors* $K(x,.)\alpha$*)*

The following properties are deduced directly from the definition:

- W_Λ is included in (possibly equal to) W.
- If Λ has $N < \infty$ points, then W_Λ is of dimension N. Otherwise, W_Λ is of infinite dimension.

Proposition 2.2. W_Λ *as defined in Definition 2.1 is a RKHS. The norm of this RKHS coincides with the norm of* W *on* W_Λ[1].

Proof. Since W_Λ is a closed subspace of the Hilbert space W, it is a Hilbert space when provided with the restriction of the norm on W to W_Λ. The evaluation functionals $\delta_x^\alpha : W \to \mathbb{R}$ are continuous on W, so they are also continuous on the subspace W_Λ. According to definition B.2, W_Λ is therefore a RKHS.

For all $x \in \Lambda$, $K(x,.)\alpha \in W_\Lambda$. Therefore, for all $\omega \in W_\Lambda$:

$$\omega(x)^t\alpha = \langle \omega, K(x,.)\alpha\rangle_W = \langle \omega, K(x,.)\alpha\rangle_{W_\Lambda}$$

This shows that the restriction of the inner-product on W to W_Λ satisfies the reproducing property on W_Λ. It is therefore the norm on the RKHS W_Λ. ■

Definition 2.3. *We define* $\mathcal{L}_{W_\Lambda} : W_\Lambda \to W^*$*, the restriction of* $\mathcal{L}_W : W \to W^*$ *to* W_Λ.
We denote W_Λ^* *the image of* W_Λ *via the mapping* $\mathcal{L}_W$*:*

$$W_\Lambda^* = \mathcal{L}_W(W_\Lambda)$$

We deduce from these definitions the following properties:

- W_Λ^* is included in W^*: we provide W_Λ^* with the restriction of the norm on W^* to W_Λ^*.
- This makes $\mathcal{L}_{W_\Lambda}$ an isometric mapping between W_Λ and W_Λ^*

The currents in W_Λ^* are like general current in W^* except that we restrict the momenta (x_i, α_i) to be located in Λ (i.e. $x_i \in \Lambda$). If Λ is a discrete set, W_Λ^* is a discretization of the general space of currents W^*. If Λ is finite, all the usual operations on currents may be computed with vectors and matrices, as we shall see in the next section.

Remark 2.4. We want to make clear that W_Λ^* is *not* the dual space of W_Λ (i.e. the space of continuous linear forms on W_Λ), contrary to what the notation suggests. Let W_Λ^d be the dual space of W_Λ. The restriction of the current $\delta_x^\alpha \in W^*$ (seen as a map from W to $\mathbb{R}$) to W_Λ belongs to W_Λ^d for any x, whereas it belongs to W_Λ^* only if $x \in \Lambda$ (provided that Λ is a discrete set). However, one can show that there is an isomorphic mapping between W_Λ^d and W_Λ^*. In the following, we will use only the space W_Λ^*. □

[1] This is the reason why we will not use the notations $\langle .,. \rangle_{W_\Lambda}$ and $\|.\|_{W_\Lambda}$, but $\langle .,. \rangle_W$ and $\|.\|_W$ instead.

2.2.2 Linear algebra in finite-dimensional RKHS

From now on, we suppose that Λ is a finite set of points of $\mathbb{R}^d$: $\Lambda = \{x_i^\Lambda\}_{i\in\Lambda}$ (note that Λ denotes the set of points as well at the set of indices).

In this case, W_Λ is finite-dimensional and a basis of this vector space is given by the vectors $K(x,.)\varepsilon_k$ for $x \in \Lambda$ and ε_k and orthonormal basis of $\mathbb{R}^p$. Indeed, by definition, these vectors span W_Λ and they are linearly independent since the kernel K is supposed to be positive definite. Therefore, the dimension of W_Λ is Np and a vector field γ in W_Λ has a unique decomposition of the form:

$$\gamma(x) = \sum_{i\in\Lambda} K(x, x_i)\alpha_i \tag{2.2.2}$$

This last equation suggests to use matrix notations. This was not possible in the general space of currents: if the position of the momenta are arbitrary, the vectors should be infinite dimensional. But once momenta are constrained to be in a finite set Λ, the operations on W^* can be performed using linear algebra in $\mathbb{R}^{Np}$.

Definition 2.5. *Let $T = \sum_{i\in\Lambda} \delta_{x_i}^{\alpha_i}$ a current in W_Λ^*. We define $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, two vectors of dimension Np such that:*

- *$\boldsymbol{\alpha}$ is the concatenation of the N vectors of dimension p: $(\alpha_i)_{1\leq i\leq N}$*
- *$\boldsymbol{\gamma}$ is the concatenation of the N vectors of dimension p: $(\mathcal{L}_W^{-1}(T)(x_i))_{1\leq i\leq N}$.*

We define the Np-by-Np block matrix $\mathbf{K}_\Lambda$ such that the block (i,j) is given by the p-by-p matrix $K(x_i,x_j)$ for $1 \leq i,j \leq N$. Since the kernel K is symmetric and positive definite, so is the matrix $\mathbf{K}_\Lambda$ (this is a direct deduction from Definition B.5).

Let $\gamma = \mathcal{L}_W^{-1}(T)$ be the vector field associated to a generic current T in W_Λ^*. $\gamma(x)$ has the general form given in Eq. (2.2.2). $\boldsymbol{\gamma}$ is the sampling of γ on the grid nodes. Applying Eq. (2.2.2) to every $x_i \in \Lambda$ leads to the fundamental equality:

$$\boldsymbol{\gamma} = \mathbf{K}_\Lambda \boldsymbol{\alpha}. \tag{2.2.3}$$

On the contrary, let γ be a generic vector field in W_Λ and $T = \mathcal{L}_W(\gamma)$. T is therefore of the form $\sum_{i\in\Lambda} \delta_{x_i}^{\alpha_i}$. Then, applying the equality $\left\langle T, \delta_{x_i}^{\varepsilon_k}\right\rangle_{W^*}$ for a given grid nodes x_i and ε_k the canonical basis of $\mathbb{R}^p$ leads to: $\sum_{j\in\Lambda} K(x_i,x_j)\alpha_j = \gamma(x_i)$. Therefore, the vectors (α_i) are solution of a linear set of Np equations whose solution is given in matrix form as:

$$\boldsymbol{\alpha} = \mathbf{K}_\Lambda^{-1}\boldsymbol{\gamma}. \tag{2.2.4}$$

The Equations (2.2.3) and (2.2.4) show first that the mapping $\mathcal{L}_W^{-1}$ on the space W_Λ^* is implemented by the multiplication by the matrix $\mathbf{K}_\Lambda^{-1}$ and, second, that the mapping $\mathcal{L}_W$ on the space W_Λ is implemented by the multiplication by the matrix $\mathbf{K}_\Lambda$. Since all the usual operations with currents are computed via the mapping $\mathcal{L}_W$, all these operations have a translation in matrix form using the matrix $\mathbf{K}_\Lambda$, when dealing with the discrete space W_Λ.

	Continuous form	Matrix form
current in W^*	$T = \sum_{i\in\Lambda} \delta_{x_i}^{\alpha_i}$	$\boldsymbol{\alpha}$
vector field in W	$\gamma(x), x \in \mathbb{R}^d$	$\boldsymbol{\gamma}$ (sampling of γ on Λ)
mapping $\mathcal{L}_W$	$T = \mathcal{L}_W(\gamma)$	$\boldsymbol{\alpha} = \mathbf{K}_\Lambda^{-1}\boldsymbol{\gamma}$
mapping $\mathcal{L}_W^{-1}$	$\gamma = \mathcal{L}_W^{-1}(T)$	$\boldsymbol{\gamma} = \mathbf{K}_\Lambda\boldsymbol{\alpha}$
action on vector fields	$T(\gamma) \in \mathbb{R}$	$\boldsymbol{\alpha}^t\boldsymbol{\gamma}$
inner-product in W_Λ	$\langle \gamma, \gamma' \rangle_W$	$\boldsymbol{\gamma}^t\mathbf{K}_\Lambda^{-1}\boldsymbol{\gamma}$
inner-product in W_Λ^*	$\langle T, T' \rangle_{W^*}$	$\boldsymbol{\alpha}^t\mathbf{K}_\Lambda\boldsymbol{\alpha}$
identities in W_Λ	$\langle \gamma, \gamma' \rangle_W = \mathcal{L}_W(\gamma)(\gamma')$ $= \langle \mathcal{L}_W(\gamma), \mathcal{L}_W(\gamma') \rangle_{W^*}$	$\boldsymbol{\gamma}^t\mathbf{K}_\Lambda^{-1}\boldsymbol{\gamma} = \left(\mathbf{K}_\Lambda^{-1}\boldsymbol{\gamma}\right)^t\boldsymbol{\gamma}'$ $= \left(\mathbf{K}_\Lambda^{-1}\boldsymbol{\gamma}\right)^t\mathbf{K}_\Lambda\left(\mathbf{K}_\Lambda^{-1}\boldsymbol{\gamma}'\right)$
identities in W_Λ^*	$\langle T, T' \rangle_{W^*} = T(\mathcal{L}_W^{-1}(T'))$ $= \left\langle \mathcal{L}_W^{-1}(T), \mathcal{L}_W^{-1}(T') \right\rangle_W$	$\boldsymbol{\alpha}^t\mathbf{K}_\Lambda\boldsymbol{\alpha} = \boldsymbol{\alpha}^t\left(\mathbf{K}_\Lambda\boldsymbol{\alpha}'\right)$ $= \left(\mathbf{K}_\Lambda\boldsymbol{\alpha}\right)^t\mathbf{K}_\Lambda^{-1}\left(\mathbf{K}_\Lambda\boldsymbol{\alpha}'\right)$

Table 2.1: The central column shows the usual operations on currents and vector fields written as for a general space of currents. When the space of current is a finite-dimensional RKHS W_Λ (the momenta are constrained to be located at a finite set of points Λ), then the quantities of the central column can be written using matrices and vectors as shown in the right column.

Table 2.1 summarizes the translation in matrix form of the usual operations on currents. Let's take the action (i.e. dual bracket) of a current on a vector field as an example. If γ is a vector field in W_Λ and $T = \sum_{i\in\Lambda} \delta_{x_i}^{\alpha_i}$, then the action of T on γ is given by:

$$T(\gamma) = \sum_{i\in\Lambda} \delta_{x_i}^{\alpha_i}(\gamma) = \sum_{i\in\Lambda} \gamma(x_i)^t\alpha_i = \boldsymbol{\gamma}^t\boldsymbol{\alpha} \tag{2.2.5}$$

Similarly, the inner-product in W^* between $T = \sum_{i\in\Lambda} \delta_{x_i}^{\alpha_i}$ and $T' = \sum_{i\in\Lambda} \delta_{x_i}^{\alpha_i'}$ (two currents in W_Λ^*) is given by:

$$\langle T, T' \rangle_{W^*} = \sum_{i\in\Lambda}\sum_{j\in\Lambda} \alpha_i^t K(x_i, x_j)\alpha_j = \boldsymbol{\alpha}^t\mathbf{K}_\Lambda\boldsymbol{\alpha}'. \tag{2.2.6}$$

The other equalities in the table 2.1 can be proven in the same way.

The expression of the inner-product in Eq. (2.2.6) shows in particular that the metric on the space W_Λ^* of dimension $\mathbb{R}^{Np}$ is given by the matrix $\mathbf{K}_\Lambda$. Similarly, the metric on the finite-dimensional space W_Λ is given by the matrix $\mathbf{K}_\Lambda^{-1}$.

2.2.3 Discretization of RKHS with linearly spaced grids

The matrix form of the usual operations on currents introduced in the previous section has the advantage to be simple and easily computable. However, from a computational point of view, these operations have the complexity of p^2N^2 (the cost of a multiplication by a Np-by-Np matrix). This may be a clear bottleneck when the number of nodes of Λ increases (and, as we shall see below, the number of points has to increase for W_Λ to be a good approximation of W^*). In this section, we show how we can compute these operations

at a cost of $Np^2 \log(N)$ provided that the grid is linearly spaced with periodic boundary conditions.

Definition 2.6 (Identification vectors in $\mathbb{R}^{pN}$ and image of vectors)**.** *Let Λ be a linearly spaced grid of size $n_x \times n_y$ (in 2D $d = 2$ and $N = n_x n_y$) or of size $n_x \times n_y \times n_z$ (in 3D $d = 3$ and $N = n_x n_y n_z$). Then, a Np dimensional vector $\boldsymbol{\gamma}$ can be seen as an "image of vectors". We call here "image of vectors" an n_x-by-n_y (or n_x-by-n_y-by-n_z) array whose elements $\boldsymbol{\gamma}(i,j)$ (or $\boldsymbol{\gamma}(i,j,k)$) is a p-dimensional vector. Such images of vectors may be stored as p matrices of size $n_x \times n_y$ ($n_x \times y \times n_z$), one for each coordinate. We denote these p matrices $\boldsymbol{\gamma}_k$ for $1 \leq k \leq p$. And the elements of these matrices are denoted $\gamma_k(i)$ for $i = 1, \ldots, N$.*

In the sequel, we identify the Np dimensional vectors and the n_x-by-n_y (n_x-by-n_y-by-n_z) images of vectors.

Proposition 2.7. *Let Λ be linearly spaced grid with periodic boundary conditions[2] and K a translation-invariant isotropic kernel ($K(x,y) = K(|x-y|)$). In this case the matrix $\mathbf{K}_\Lambda$ is block-circulant. Let $(\mathbf{k}[i])_{1 \leq i \leq p}$ be the first p rows of $\mathbf{K}_\Lambda$ ($\mathbf{k}[i]$ is a vector of dimension Np). If we identify Np-dimensional vectors with image of vectors, the image of vectors $\mathbf{k}[i]$ may decomposed into p vectors of size N: $\mathbf{k}[i]_j$. Then, the following equality holds for all $\boldsymbol{\alpha} \in \mathbb{R}^{Np}$:*

$$(\mathbf{K}_\Lambda \boldsymbol{\alpha})_i = \sum_{j=1}^{p} \mathbf{k}[i]_j \star \boldsymbol{\alpha}_j, \tag{2.2.7}$$

for $i = 1, \ldots, p$, where $\star$ denotes the usual 2D (or 3D) discrete convolutions between matrices.

Proof. First, we show that the matrix $\mathbf{K}_\Lambda$ is block-circulant. Indeed the block (i,j) of the matrix $\mathbf{K}_\Lambda$ (denoted here $\mathbf{K}_\Lambda(i,j)$) verifies:

$$\mathbf{K}_\Lambda(i,j) = K(\left|x_i^\Lambda - x_j^\Lambda\right|) = K(\left|x_0^\Lambda - x_{j-i}^\Lambda\right|) = \mathbf{K}_\Lambda(0, j-i), \tag{2.2.8}$$

since Λ is linearly spaced with boundary conditions.

This means that the matrix $\mathbf{K}_\Lambda$ is entirely determined by its first p rows $\mathbf{k}[i]$, each row being a vector of dimension Np. $\mathbf{k}[i]$ is seen as an image of vectors: p matrices $\mathbf{k}[i]_j$ whose element are denoted $\mathbf{k}[i]_j(n)$ for $n = 1, \ldots, N$.

By definition, we have:

$$(\mathbf{K}_\Lambda \boldsymbol{\alpha})(j) = \sum_{q=1}^{N} K(x_j, x_q)\alpha_q = \sum_{q=1}^{N} \mathbf{K}_\Lambda(j-q)\boldsymbol{\alpha}(q) \tag{2.2.9}$$

The i*th* coordinate of the p-by-p matrix multiplication $\mathbf{K}_\Lambda(j-q)\boldsymbol{\alpha}(q)$ is given by: $\sum_{s=1}^{p} \mathbf{k}[i]_s(j-q)\boldsymbol{\alpha}_s(q)$. Therefore,

$$(\mathbf{K}_\Lambda \boldsymbol{\alpha})_i(j) = \sum_{s=1}^{p}\sum_{q=1}^{N} \mathbf{k}[i]_s(j-q)\boldsymbol{\alpha}_s(q) = \sum_{s=1}^{p} (\mathbf{k}[i]_s \star \boldsymbol{\alpha}_s)(j). \tag{2.2.10}$$

which is Eq. (2.2.7) written with coordinates. ■

[2] This means that the grid has the geometry of a torus: points at two opposite borders are at a distance 0 one from each other.

This proves that $\mathbf{K}_\Lambda \boldsymbol{\alpha}$ may be computed via p^2 discrete convolution of matrices with N elements. Each discrete convolution between matrices can be computed via FFTs at a cost $\mathcal{O}(N\log(N))$: $\gamma \star \gamma' = \mathcal{FFT}^{-1}(\mathcal{FFT}(\gamma).\mathcal{FFT}(\gamma))$ (where '.' denotes the element-wise multiplication of matrices). Therefore, the total cost of the convolution between images of vectors is $\mathcal{O}(p^2 N\log(N))$.

If the kernel is scalar and translation-invariant (i.e. $K(x,y) = K(|x-y|)\mathrm{Id}$), then the computations are even simpler. Indeed, in this case the block-circulant matrix $\mathbf{K}_\Lambda$ (N^2p^2 elements) can be reduced to a circulant matrix with N^2 elements. This matrix is completely determined by its first row $\mathbf{k}$ of size N which can be seen as a matrix (image of vector with a single coordinate). Therefore,

$$(\mathbf{K}_\Lambda \boldsymbol{\alpha})_i = \mathbf{k} \star \boldsymbol{\alpha}_i \tag{2.2.11}$$

In this case, the matrix multiplication requires only p discrete convolutions. Its complexity is of $\mathcal{O}(pN\log(N))$, to be compared to the complexity of the direct computation $\mathcal{O}(N^2p^2)$. Typical grids have $N = (10^2)^3 = 10^6$ nodes.

Besides the improvement in terms of time complexity, linearly spaced grids allow us also to save memory. Indeed, there is no need to store the matrix $\mathbf{K}_\Lambda$ which has $p^2N^2 = p^2 n_x^2 n_y^2 n_z^2$ elements. The computational framework depends only on the discrete Fourier transforms of the p^2 matrices $\mathbf{k}$ of size $n_x \times n_y \times n_z$. Due to the periodic boundary conditions, each matrix $\mathbf{k}$ is symmetric. Its Fourier transform is therefore real and symmetric. It can be stored in a matrix of size $(n_x/2+1)\times(n_y/2+1)\times(n_z/2)$ for even dimensions. Therefore the total memory space needed to store the metric is of order $p^2N/8$, to be compared to the initial memory space of p^2N^2. For scalar kernel, only one matrix $\mathbf{k}$ needs to be stored, thus leading to a memory space of order $N/8$ instead of N^2 for the storage of the whole matrix $\mathbf{K}_\Lambda$.

Remark 2.8. Note that due to the boundary conditions the result of the matrix computation ($\boldsymbol{\gamma} = \mathbf{K}_\Lambda \boldsymbol{\alpha}$) is not strictly equal to the result of the discrete convolutions, since momenta close to a border of the grid may impact the result of $\boldsymbol{\gamma}$ at a point close to the opposite border. However, this has little impact on the computation as soon as the grid borders are located at a distance much further than λ_W (or the rate of decay of the kernel) from the momenta (x_i, α_i). In the following, we suppose that we deal with current in W_Λ which have non-zero momenta only at a distance much further than λ_W from the border of the grid. This is always possible, provided that we choose a large enough grid. □

2.2.4 The ill-posed nature of the metric on discrete currents

The purpose of introducing the spaces W_Λ is twofold: (1) computations in these spaces are particularly easy and efficient, as we saw in the previous section and (2) these spaces can be used to approximate the original continuous space W as the grid step tends to zero, as we shall see in the next sections. However, we must be aware that the matrix $\mathbf{K}_\Lambda$ used as the metric in the space W_Λ may have a bad conditioning. Here, we prove precisely that the conditioning of this matrix becomes worse as the grid step tend to zero, thus meaning that the inversion of the matrix $\mathbf{K}_\Lambda$ (or, equivalently, that the deconvolution via the inversion of the FFT of matrix $\mathbf{k}$) is not possible in practice for fine grids.

This is not surprising since the low-pass filter $\mathbf{K}_\Lambda$ implements the regularizing operator $\mathcal{L}_W^{-1}$ (which is intrinsically a convolution operator), whereas the high-pass filter $\mathbf{K}_\Lambda^{-1}$ implements the irregular differential operator $\mathcal{L}_W$.

This ill-posed nature of the metric has two consequences. First, it explains why the orthogonal projection of currents on the approximation space W_Λ can not be computed in practice (see next section). Second, it justifies the need for a robust deconvolution scheme to implement the inversion of the matrix $\mathbf{K}_\Lambda$, like the one we will introduce in Chapter 3 via matching pursuit.

Proposition 2.9. *Let Λ be a linearly spaced grid of size n_x, n_y, n_z (with $N = n_x n_y n_z$), with step Δ and with periodic boundary conditions. Let K be a scalar translation-invariant kernel: $K(x,y) = G(x-y)\mathrm{I}_p$ for a scalar function G and $\mathbf{K}_\Lambda$ the Np-by-Np matrix as in Definition 2.5. Let $\mathbf{k}$ be the n_x-by-n_y-byn_z matrix defined by:*

$$\mathbf{k} = \{G(i\Delta, j\Delta, k\Delta)\}, \tag{2.2.12}$$

for $-(n_x/2) \leq i < n_x/2$, $-(n_y/2) \leq j < n_y/2$ and $-(n_z/2) \leq k < n_z/2$. $\hat{\mathbf{k}}$ denotes its discrete Fourier transform.

Then, the spectrum of the matrix $\mathbf{K}_\Lambda$ is given by the N eigenvalues $\hat{\mathbf{k}}(i,j,k)$, each with multiplicity p. In particular, we have:

$$|\mathbf{K}_\Lambda| = \left(\prod_{i,j,k} \hat{\mathbf{k}}(i,j,k) \right)^p \tag{2.2.13}$$

Moreover, at the limit $\Delta \to 0$ and $V = N\Delta^3 \to \infty$ (the volume covered by the grid tends to infinity), we have:

$$\hat{\mathbf{k}}(i,j,k) \sim \frac{\sqrt{N}}{V} \hat{G}\left(\frac{2\pi i}{n_x \Delta}, \frac{2\pi j}{n_y \Delta}, \frac{2\pi k}{n_z \Delta} \right), \tag{2.2.14}$$

where $\hat{G}$ denotes the Fourier transform of the function G.

Proof. The assumptions on the grid makes the matrix $\mathbf{K}_\Lambda$ a block-circulant matrix and as such, is diagonal in a Fourier basis.

First, we notice that, since the kernel is scalar, each block (i,j) of $\mathbf{K}_\Lambda$ may be reduced to a single scalar $G(x_i - x_j)$. This transforms the Np-by-Np matrix $\mathbf{K}_\Lambda$ to a N-by-N matrix $\widetilde{\mathbf{K}}_\Lambda$. If $(v_1, \dots, v_N)$ is an eigenvector of $\widetilde{\mathbf{K}}_\Lambda$, then the p vectors $\big(\underbrace{v_1, 0, \dots, 0}_{\text{p elements}}, \dots, \underbrace{v_N, 0, \dots, 0}_{\text{p elements}}\big)$, ..., $\big(\underbrace{0, \dots, v_1}_{\text{p elements}}, \dots, \underbrace{0, \dots, v_N}_{\text{p elements}}\big)$ are linearly independent eigenvectors of $\mathbf{K}_\Lambda$ associated with the same eigenvalue. This proves, in particular, that $|\mathbf{K}_\Lambda| = \left|\widetilde{\mathbf{K}}_\Lambda\right|^p$.

Then, we need to solve the equation $\widetilde{\mathbf{K}}_\Lambda \mathbf{v} = \mu \mathbf{v}$ for $v \in \mathbb{R}^N$. This equation might be written as $\mathbf{k} \star \mathbf{v} = \mu \mathbf{v}$ provided that we identify the $N = n_x n_y n_z$-dimensional vectors with the n_x-by-n_y-by-n_z matrices. As in Section 2.2.3, $\mathbf{k}$ denotes the first row of the $\widetilde{\mathbf{K}}_\Lambda$ and is seen as a n_x-by-n_y-by-n_z matrix. This convolution is equivalent in the Fourier domain to $\hat{\mathbf{k}}.\hat{\mathbf{v}} = \mu \hat{\mathbf{v}}$ where '.' denotes the element-wise multiplication of matrices. The canonical basis of the n_x-by-n_y-by-n_z matrices $(\delta_{i,j,k})$ is solution of this equation with $\mu = \hat{\mathbf{k}}_{i,j,k}$ $(0 \leq i < n_x, 0 \leq l < n_y, 0 \leq k < n_z)$. Since the Fourier transform is an isometric mapping,

the inverse Fourier transform of the canonical basis (i.e. the Fourier basis itself) is an orthonormal basis of eigenvectors of $\widetilde{\mathbf{K}}_\Lambda$. Eventually, the spectrum of $\mathbf{K}_\Lambda$ is the N values $\hat{\mathbf{k}}(i,j,k)$ each one with multiplicity p. In particular, $|\mathbf{K}_\Lambda| = \left(\prod_{i,j,k} \hat{\mathbf{k}}(i,j,k)\right)^p$.

Now, we want to give an approximation of these eigenvalues as the grid step tends to zero and the domain covered by the grid tends to infinity. First, we notice that the first row of the matrix $\widetilde{\mathbf{K}}_\Lambda$ is the sampling of the function G as written in Eq. (2.2.12). In one dimension, one would have:

$$\hat{\mathbf{k}}(n) = \frac{1}{\sqrt{N}} \sum_{-N/2 \leq p < N/2} G(p\Delta) \exp(-\frac{2i\pi}{N} pn).$$

This last expression may be seen, in the limit when Δ tends to zero, as the Riemann sum of $\frac{1}{\Delta\sqrt{N}} \int_{-N\Delta/2}^{N\Delta/2} G(t) \exp(-\frac{2i\pi}{N} nt/\Delta) dt$. If we suppose that the size of the grid $L = N\Delta$ is much larger than λ_W then this integral is approximately: $\frac{1}{\Delta\sqrt{N}} \int_{-\infty}^{\infty} G(t) \exp(-\frac{2i\pi}{N} nt/\Delta) dt = \frac{1}{\sqrt{L\Delta}} \hat{G}\left(\frac{2\pi n}{N\Delta}\right)$ for $N\Delta = L$. This results can be extended directly in 3D, which leads to:

$$\hat{\mathbf{k}}(i,j,k) \sim \frac{\sqrt{N}}{V} \hat{G}\left(\frac{2\pi i}{n_x\Delta}, \frac{2\pi j}{n_y\Delta}, \frac{2\pi k}{n_z\Delta}\right). \tag{2.2.15}$$

■

If we apply this result with G the zero-mean Gaussian function with variance $\lambda_W \mathrm{I}_3$, we get:

$$\begin{aligned} \hat{\mathbf{k}}(i,j,k) &\sim \frac{\sqrt{N}}{V} \hat{G}\left(\frac{2\pi i}{n_x\Delta}, \frac{2\pi j}{n_y\Delta}, \frac{2\pi k}{n_z\Delta}\right) \\ &\sim C \frac{\lambda_W^3}{V} \frac{\Delta^{3/2}}{V^{1/2}} \exp\left(-C' \frac{\lambda_W^2}{\Delta^2} \left(\frac{i^2}{n_x^2} + \frac{j^2}{n_y^2} + \frac{k^2}{n_z^2}\right)\right), \end{aligned} \tag{2.2.16}$$

with $V = N\Delta^3$ the volume covered by the grid and C, C' two positive numerical constants.

This shows that the conditioning number of the matrix $\mathbf{K}_\Lambda$ (i.e. the ratio between its largest and smallest eigenvalue) is given by:

$$\frac{\mathbf{k}(0,0,0)}{\mathbf{k}(n_x - 1, n_y - 1, n_z - 1)} \sim \exp\left(C' \frac{\lambda_W^2}{\Delta^2}\right). \tag{2.2.17}$$

This number growths exponentially as the ratio between the grid step and the standard deviation $(\Delta/\lambda_W)^2$ tends to zero, thus showing the bad conditioning of the matrix $\mathbf{K}_\Lambda$ for small grid steps. This prevents us from computing the inverse of $\mathbf{K}_\Lambda$ or equivalently to compute $1/\hat{\mathbf{k}}(i,j,k)$ (required to solve the inversion problem via deconvolution). See for instance the experiments in Section. 3.5.1.

Of course, this behavior still holds for non-Gaussian kernel. In this case, $1/\lambda_W^2$ must be replaced by the typical rate of decay of the Fourier transform of the function G.

2.3 Orthogonal projections on discrete spaces

2.3.1 What we need

In the previous section, we show how the map $\mathcal{L}_W^{-1}$ can be computed efficiently on discrete spaces of currents, using the multiplication by the matrix $\mathbf{K}_\Lambda$ or the convolutions

by FFT for linearly-spaced grid with periodic boundary conditions. This operation is represented in Fig. 2.2 by an arrow between the space of discrete currents W_Λ^* and the space of discrete vector fields W_Λ.

Now, the question is to know how to use these discrete spaces of currents to approximate the usual operations on currents. By usual operations on currents, we mean for instance the action of a current on a vector field $T(\omega)$, the norm of a current $\|T\|^2_{W^*}$, the inner-product between two currents $\langle T, T' \rangle_{W^*}$. Let us take the norm of the current as the prototype example of such operations. Let $T = \sum_{i=1}^{N} \delta_{x_i}^{\alpha_i}$ be a generic current in W^* with a finite number of momenta given in continuous coordinates (this current is in the upper-right corner in Fig. 2.2).

The exact computation of the norm requires to compute the vector field associated to T: $\gamma(x) = \mathcal{L}_W^{-1}(T)(x) = \sum_{i=1}^{N} K(x, x_i)\alpha_i$ (which is on the upper-left corner in Fig. 2.2) at the arbitrary positions x_i. Then, the norm is given as: $\|T\|^2_{W^*} = \sum_{i=1}^{N} \gamma(x_i)^t \alpha_i$.

To take advantage of the fast computations in the discrete spaces, we project the current $T \in W^*$ into a finite-dimensional space W_Λ^* (from the upper-right to the bottom-right corner in Fig. 2.2). This leads to a Np-dimensional vector $\boldsymbol{\alpha}_\Lambda$. Then we compute the associated vector field in W_Λ via FFT (from the bottom-right to the bottom-left corner of Fig. 2.2): $\boldsymbol{\gamma}_\Lambda = \mathbf{K}\boldsymbol{\alpha}_\Lambda$. A this point, two options are possible: (1) we compute the norm directly in the discrete spaces as $\boldsymbol{\alpha}_\Lambda^t \boldsymbol{\gamma}_\Lambda$ or (2) we interpolate the samples on grid nodes $\boldsymbol{\gamma}_\Lambda$ at the initial positions x_i (from the bottom-left to the upper-left corner in Fig. 2.2). This leads to $\tilde{\gamma}(x_i)$ and we compute the norm as: $\sum_{i=1}^{N} \tilde{\gamma}(x_i)^t \alpha_i$.

At the first glance, the first solution seems much simpler. However, a precise discussion of these two options in Section 2.5.1 will reveal that this is not always the case. Anyway, this shows that for defining a complete computational framework, we need to define first a projection from the continuous space of currents to the discrete space of currents and, second, an interpolation scheme which can reconstruct a dense vector field from samples on grid nodes. One remarks that the vector field in the RKHS W are intrinsically band-limited. Therefore, Shannon theorem shows that such a reconstruction scheme can be defined for fine enough grids.

In this section, we investigate the orthogonal projections from continuous to discrete spaces. We will show that, theoretically, the projected currents converge to their initial value as the grid step tends to zero, thus showing the potential of discrete spaces to be used as approximation spaces. However, this operation is ill-posed from a numerical point of view. This will require to investigate alternative strategies in the forthcoming section. We will also show the orthogonal projection of random Gaussian currents on discrete spaces have a probability density function, a key feature for defining statistics on currents in the next chapters.

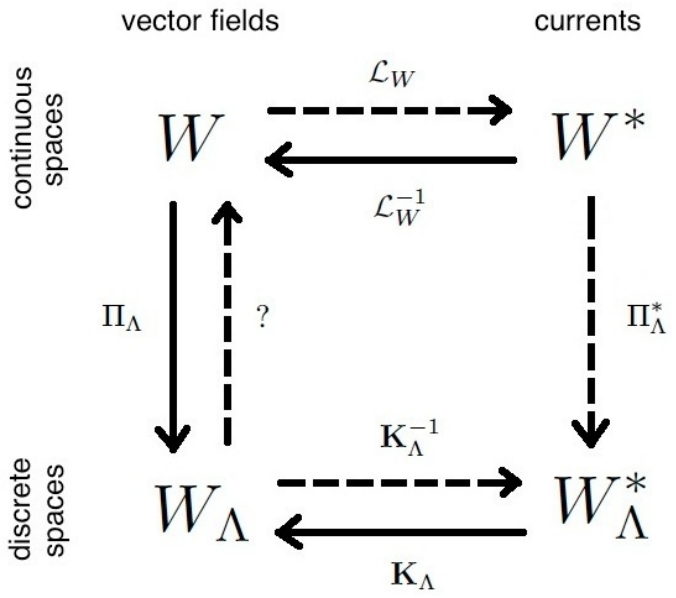


Figure 2.2: Global picture of the computational framework for approximating operations on currents. This framework is based on projections from continuous to discrete spaces: Π_Λ for vector field and Π_Λ^* for currents, as well as reconstruction formula to define a dense vector field from samples on grid nodes ('?' arrow). Using the orthogonal projections is possible theoretically but suffer from the ill-posed nature of the matrix $\mathbf{K}_\Lambda$ from a numerical point of view. Indeed, whereas the orthogonal projection for vector fields (Π_Λ) is simply the sampling of the dense vector field on grid nodes, the orthogonal projection for currents Π_Λ^* is equal to $\mathbf{K}_\Lambda^{-1} \circ \Pi_\Lambda \circ \mathcal{L}_W^{-1}$ which is ill-posed because of $\mathbf{K}^{-1}$. Other projection methods need to be investigated. *(dashed lines correspond to ill-posed or non-defined operations)*

2.3.2 Projection of currents and vector fields

Let T be a generic current in W^*. Our purpose is to give an approximation of T by a current T^Λ in W_Λ^* whose momenta are located at the grid nodes (i.e. which is characterized by a Np-dimensional vector $\boldsymbol{\alpha}$). By isometry, this give an approximation of a generic vector field γ in W by a vector field γ^Λ in W_Λ which is completely determined by its samples at the grid nodes: $\boldsymbol{\gamma} = \{\gamma(x_i)\}_{i\in\Lambda}$.

Since W_Λ is a closed subspace of W, the best approximation (in the sense of the norm in W) is given by the orthogonal projection of γ on W_Λ (and the orthogonal projection of T on W_Λ^*). First, we give a characterization of the orthogonal space $W_\Lambda^\perp$.

Proposition 2.10. *Let $W_\Lambda^\perp$ be the orthogonal space of W_Λ in W: $W = W_\Lambda \overset{\perp}{\oplus} W_\Lambda^\perp$. Then,*

$$W_\Lambda^\perp = \{\gamma \in W; \quad \forall x \in \Lambda, \gamma(x) = 0\} \tag{2.3.1}$$

Proof. If $\gamma \in W_\Lambda^\perp$, then it is orthogonal to every vector field of the form $K(x,.)\alpha$. Therefore, for all $x \in \Lambda$ and all $\alpha \in \mathbb{R}^p$,

$$\langle \gamma, K(x,.)\alpha \rangle_W = \gamma(x)^t \alpha = 0, \tag{2.3.2}$$

and then $\gamma(x) = 0$.

Conversely, if $\gamma(x) = 0$ for all $x \in \Lambda$, then the same equation shows that γ is orthogonal to every $K(x, .)\alpha$ ($x \in \Lambda$). Since W_Λ is the closed span of such vectors, we get the result by density. ■

Corollary 2.11. *Let γ be dense vector field in W, γ^Λ its orthogonal projection on W_Λ and $\boldsymbol{\gamma}$ its sampling at the points of Λ (i.e. $\boldsymbol{\gamma} = \{\gamma(x_i)\}_{i\in\Lambda}$ a vector of dimension Np). Then, γ^Λ is equal to:*

$$\gamma^\Lambda(x) = \sum_{i\in\Lambda} K(x, x_i)\alpha_i, \tag{2.3.3}$$

where the concatenation of the vectors α_i form a Np-dimensional vector $\boldsymbol{\alpha}$ which satisfies:

$$\boldsymbol{\alpha} = \mathbf{K}_\Lambda^{-1}\boldsymbol{\gamma} \tag{2.3.4}$$

In particular, for all $x_i \in \Lambda$, $\gamma^\Lambda(x_i) = \gamma(x_i)$.

Proof. By construction, the dense vector field γ^Λ verifies $\gamma^\Lambda(x) = \gamma(x)$ for every point x in Λ. Therefore $\gamma - \gamma^\Lambda$ vanishes on Λ and then belongs to $W_\Lambda^\perp$. We get the result by uniqueness of the decomposition on two orthogonal spaces. ■

This corollary shows that computing $\gamma^\Lambda(x)$ at the grid nodes is very easy: one needs to sample γ at the grid nodes (this is the implementation of the projection Π_Λ in Fig. 2.2) and that computing $\gamma^\Lambda(x)$ at an arbitrary point requires to compute the momenta $\boldsymbol{\alpha}$ by inverting the matrix $\mathbf{K}_\Lambda$: this computation is not possible numerically (this would be the implementation of the '?' arrow in Fig. 2.2).

Corollary 2.12. *Let T be a current in W^* and T^Λ its orthogonal projection on W_Λ^*. Let $\gamma = \mathcal{L}_W^{-1}(T)$ be a dense vector field in W and $\boldsymbol{\gamma}$ its sampling on the points in Λ. Then,*

$$T^\Lambda = \sum_{i\in\Lambda} \delta_{x_i}^{\alpha_i}, \tag{2.3.5}$$

where the concatenation of the vectors α_i form a Np-dimensional vector $\boldsymbol{\alpha}$ which satisfies:

$$\boldsymbol{\alpha} = \mathbf{K}_\Lambda^{-1}\boldsymbol{\gamma} \tag{2.3.6}$$

Proof. Since $\mathcal{L}_W$ is an isometric mapping, $\mathcal{L}_W^{-1}(T^\Lambda) \in W_\Lambda$ is the orthogonal projection of $\mathcal{L}_W^{-1}(T) \in W$ on W_Λ. Then, we apply the previous result. ■

This corollary shows that the orthogonal projection in the space of currents (denotes Π_Λ^* in Fig. 2.2) is equal to $K_\Lambda^{-1} \circ \Pi_\Lambda \circ \mathcal{L}_W^{-1}$ and therefore is ill-posed numerically due to the bad-conditioning of the matrix $\mathbf{K}_\Lambda$ (although the computation of the vector $\boldsymbol{\gamma}$ can be done via discrete convolution).

These results show that we need to find alternative numerical scheme both to estimate the momenta of T^Λ and to reconstruct a dense vector field from its samples at the points of Λ. Beforehand, we will show that the orthogonal projections converge to the original object in W or W^* when the sampling of the grid tends to zero. This will prove that the discrete spaces W_Λ can be used to approximate the currents and the vector fields at any desired accuracy. We will use two norms: the norm W and the L^∞-norm.

Definition 2.13. *Let $(\Lambda_n)_{n\geq 0}$ be a sequence of finite sets of points of $\mathbb{R}^d$. We say that such a sequence is an acceptable sequence if for every point $x \in \mathbb{R}^d$, there is a sequence $x_n \in \Lambda_n$, such that $|x - x_n| \xrightarrow[n\to\infty]{} 0$*

This definition establishes the conditions under which we can say that the points of the sets Λ_n become dense in $\mathbb{R}^d$. It is possible to build such an acceptable sequence: let Λ_n be linearly spaced grid whose step tends to zero ($\Delta_n = \Delta/2^n$) and whose borders tend to infinity in every direction.

Proposition 2.14. *Let Λ_n be an acceptable sequence of finite sets of points of $\mathbb{R}^d$ in the sense of Definition 2.13.*

Let W be a RKHS of continuous mappings from $\mathbb{R}^d$ to $\mathbb{R}^p$ which tends to zero at infinity and such that $\|\gamma\|_\infty \leq C_W \|\gamma\|_W$. We assume also that the kernel K is continuous on $\mathbb{R}^d \times \mathbb{R}^d$. Let $\gamma \in W$ and γ_n the orthogonal projection of γ on Λ_n. Then:

$$\begin{aligned} \|\gamma - \gamma_n\|_W &\xrightarrow[n\to\infty]{} 0 \\ \|\gamma - \gamma_n\|_\infty &\xrightarrow[n\to\infty]{} 0 \end{aligned} \tag{2.3.7}$$

Proof. First, we show that $\gamma_n(x)$ converge to $\gamma(x)$ for all x. Let $x_n \in \Lambda_n$ be a sequence converging to x. By definition of the orthogonal projection on W_{Λ_n}, $\gamma_n(x_n) = \gamma(x_n)$. This implies:

$$|\gamma(x) - \gamma_n(x)| \leq |\gamma(x) - \gamma(x_n)| + |\gamma_n(x_n) - \gamma_n(x)|$$

Since γ is continuous, the first term tends to zeros as n tends to infinity. The second term satisfies:

$$\begin{aligned} |\gamma_n(x) - \gamma_n(x_n)| &\leq \sup_{|\alpha|=1} |\langle \gamma_n, K(.,x)\alpha - K(.,x_n)\alpha\rangle_W| \\ &\leq \|\gamma\|_W \sup_{|\alpha|=1} \|K(.,x)\alpha - K(.,x_n)\alpha\|_W \\ &\leq \|\gamma\|_W \left(|K(x,x_n) - K(x,x)|_2 + |K(x_n,x_n) - K(x_n,x)|_2\right)^{1/2}, \end{aligned} \tag{2.3.8}$$

where we used the fact that $\|\gamma_n\|_W \leq \|\gamma\|_W$ (orthogonal projection) and we denoted $|.|_2$ the spectral norm of matrices. Since K is continuous on $\mathbb{R}^d \times \mathbb{R}^d$ this last expression tends also to 0 when $n \to \infty$. This shows the point-wise convergence.

Now, if γ is of the form $\gamma = \sum_{i=1}^k K(.,x_i)\alpha_i$, then

$$\|\gamma - \gamma_n\|_W^2 = \langle \gamma, \gamma - \gamma_n\rangle_W = \sum_i \alpha_i^t(\gamma(x_i) - \gamma_n(x_i))$$

which tends to zero thanks to the point-wise convergence. Let γ be arbitrary in W, by density there is a sequence of $\gamma^p \in W$ which are finite linear combinations of vector fields of the form $K(.,x_i^p)\alpha_i^p$ such that $\|\gamma - \gamma^p\|_W \xrightarrow[p\to\infty]{} 0$. Let P_n denotes the orthogonal projection on W_{Λ_n}:

$$\begin{aligned} \|\gamma - \gamma_n\|_W &\leq \|\gamma - \gamma^p\|_W + \|\gamma^p - P_n(\gamma^p)\|_W + \|P_n(\gamma^p) - P_n(\gamma)\|_W \\ &\leq 2\|\gamma - \gamma^p\|_W + \|\gamma^p - P_n(\gamma^p)\|_W \end{aligned} \tag{2.3.9}$$

Let's choose p such that $\|\gamma - \gamma^p\|_W \leq \varepsilon/4$. Then, we choose n such that $\|\gamma^p - P_n(\gamma^p)\|_W \leq \varepsilon/2$ (thanks to the previous result, $P_n(\gamma^p)$ tend to γ^p for all p). This proves that $\|\gamma - \gamma_n\|_W \leq \varepsilon$ and therefore the convergence for the norm W.

Eventually, since $\|\gamma - \gamma_n\|_\infty \leq C_W \|\gamma - \gamma_n\|_W$, the convergence still holds for the L^∞ norm. ■

This proof is rather general since it does not make any assumptions on the form of the vector field $\gamma \in W$. This makes also difficult the estimation of a speed of convergence. By contrast, if we assume that γ is parameterized by a finite number of momenta (as the most common case in practice), then we will show in the next section that the speed of convergence is at least polynomial. Indeed, we will show a polynomial speed of convergence for a suboptimal projection on discrete spaces in Prop. 2.19 and 2.20 (the orthogonal projection being the projection which minimizes the projection error).

2.3.3 Trace of random Gaussian currents on discrete spaces

In Section 1.5.3, we defined random Gaussian currents via a mapping X^* between the space of vector field W to a Gauss space $\mathcal{G}$. This function maps a vector field to a real random Gaussian variable instead of a real number like for a deterministic current. These random Gaussian currents have no probability density functions in infinite dimension. It is therefore particularly difficult to simulate such currents and to use it in a statistical framework, which usually requires the use of the likelihood of a random variable.

In this section, we use the orthogonal projections to show that the restriction of the map X^* to our discrete spaces W_Λ converge to the initial random variable as the grid Λ becomes finer and finer. Moreover, we show that this projected random variables are of finite-dimension and, as such, admit a probability density function.

First, we define X^*_Λ by the restriction of X^* to W_Λ.

Definition 2.15. *Let X^* be the mapping as in Definition 1.30. Let X^*_Λ be the mapping from W to the Gauss space $\mathcal{G}$ defined by:*

$$\forall \omega \in W, \quad X^*_\Lambda(\omega) = X^*\left(\Pi_{W_\Lambda}(\omega)\right), \tag{2.3.10}$$

where Π_{W_Λ} denotes the orthogonal projection on W_Λ.

Thanks to this definition, X^*_Λ coincides with X^* on W_Λ and is equal to 0 almost surely on $W_\Lambda^\perp$. We can now prove the following convergence result:

Proposition 2.16. *Let Λ_n be an acceptable sequence of finite sets of points of $\mathbb{R}^d$ as in Definition 2.13. Let W^* a Gaussian random current as in Definition 1.30 and W^*_Λ as in Definition 2.15. Then, for all $\omega \in W$:*

$$\mathbb{E}\left((X^*(\omega) - X^*_{\Lambda_n}(\omega))^2\right) \xrightarrow[n\longrightarrow\infty]{} 0. \tag{2.3.11}$$

Proof. By definition of $X^*_{\Lambda_n}$, we have:

$$\mathbb{E}\left((X^*(\omega) - X^*_{\Lambda_n}(\omega))^2\right) = \mathbb{E}\left(X^*\left(\Pi_{W^\perp_{\Lambda_n}(\omega)}\right)^2\right) = \left\|\Pi_{W^\perp_{\Lambda_n}}(\omega)\right\|^2_{W^*} \tag{2.3.12}$$

This last expression tends to zero as n tends to infinity by application of Proposition 2.14. ■

Moreover, we state now that the random Gaussian current X^*_Λ is a finite dimensional random Gaussian process which admits a probability density function.

Proposition 2.17. *Let Λ be a finite set of points of $\mathbb{R}^d$ and X^*_Λ the random Gaussian current as in Definition 2.15. Then,*

$$X^*_\Lambda = \sum_{i\in\Lambda} \delta^{\alpha_i}_{x_i}, \tag{2.3.13}$$

where the concatenation of the vectors α_i (denoted $\boldsymbol{\alpha}$) is a Np-dimensional Gaussian vector:

$$\boldsymbol{\alpha} \sim \mathcal{N}\left(0, \mathbf{K}_\Lambda^{-1}\right). \tag{2.3.14}$$

Proof. We denote $(\varepsilon_k)_{k=1,\dots,p}$ the canonical basis of $\mathbb{R}^p$. Let $\omega \in W_\Lambda$ be parametrized by the momenta $\boldsymbol{\alpha}$: $\omega = \sum_{i\in\Lambda} K(x_i,.)\alpha_i$. Then, by linearity of X^*_Λ, we verify that:

$$\begin{aligned} X^*_\Lambda(\omega) &= X^*_\Lambda\left(\sum_{i\in\Lambda} K(x_i,.)\alpha_i\right) \\ &= \sum_{i\in\Lambda}\sum_{k=1}^{p} X^*_\Lambda(K(x_i,.)\varepsilon_k)\alpha_i^k \\ &= \boldsymbol{\chi}^t\boldsymbol{\alpha}, \end{aligned} \tag{2.3.15}$$

where $\boldsymbol{\chi}$ is a Np-dimensional random vectors which results of the concatenation of the Gaussian variables: $X^*_\Lambda(K(x_i,.)\varepsilon_k)$. To make explicit the linear dependency of this last expression with respect to ω, we introduce $\boldsymbol{\gamma} = \mathbf{K}_\Lambda\boldsymbol{\alpha}$ which is such that $\gamma_i = \omega(x_i)$ for all $i \in \Lambda$. We have therefore:

$$\begin{aligned} X^*_\Lambda(\omega) = \boldsymbol{\chi}^t\boldsymbol{\alpha} &= \left(\mathbf{K}_\Lambda^{-1}\boldsymbol{\chi}\right)^t\boldsymbol{\gamma} \\ &= \sum_{i\in\Lambda}\left(\mathbf{K}_\Lambda^{-1}\boldsymbol{\chi}\right)^t_i\omega(x_i) \\ &= \sum_{i\in\Lambda}\delta_{x_i}^{\left(\mathbf{K}_\Lambda^{-1}\boldsymbol{\chi}\right)_i}(\omega) \end{aligned} \tag{2.3.16}$$

This last equality still holds for $\omega \in W_\Lambda^\perp$ (both terms equal 0). We have:

$$X^*_\Lambda = \sum_{i\in\Lambda}\delta_{x_i}^{\left(\mathbf{K}_\Lambda^{-1}\boldsymbol{\chi}\right)_i} \tag{2.3.17}$$

Moreover, thanks to the definition of X^*, the random vector $\boldsymbol{\chi}$ is zero-mean and has $\mathbf{K}_\Lambda$ as covariance matrix. Indeed,

$$\mathbb{E}(\chi_i^n\chi_j^m) = \mathbb{E}\left(X^*(K(x_i,.)\varepsilon_n)X^*(K(x_j,.)\varepsilon_m)\right) = K(x_i,x_j)_{n,m}. \tag{2.3.18}$$

Therefore, $\mathbf{K}_\Lambda^{-1}\boldsymbol{\chi}$ has the covariance matrix $\mathbf{K}_\Lambda^{-1}\mathbf{K}_\Lambda\mathbf{K}_\Lambda^{-1} = \mathbf{K}_\Lambda^{-1}$. However, since the components of the vector $\mathbf{K}_\Lambda^{-1}\boldsymbol{\chi}$ are not independent, this does not prove that it is a Gaussian vector. To show this fact, we shall prove that $\tilde{\boldsymbol{\chi}} = \mathbf{K}_\Lambda^{-1/2}\boldsymbol{\chi}$ is vector of independent

normal variables. This will imply that $\mathbf{K}_\Lambda^{-1}\boldsymbol{\chi} = \mathbf{K}_\Lambda^{-1/2}\tilde{\boldsymbol{\chi}}$ follows a Gaussian distribution as a linear combination of independent Gaussian variables. Indeed, by linearity of X^*, we have:

$$\tilde{\boldsymbol{\chi}}_i^k = \sum_{j=1}^{N} (\mathbf{K}_\Lambda^{-1/2})_{i,j}\chi_j^k = X^*\left(\sum_{j=1}^{N} (\mathbf{K}_\Lambda^{-1/2})_{i,j} K(x_j,.)\varepsilon_k\right). \tag{2.3.19}$$

This proves that every component of $\tilde{\boldsymbol{\chi}}$ is a Gaussian variable. Moreover, a simple computation shows that:

$$\begin{aligned}\mathbb{E}\left((\mathbf{K}_\Lambda^{-1/2}\boldsymbol{\chi})_i^n (\mathbf{K}_\Lambda^{-1/2}\boldsymbol{\chi})_j^m\right) &= \sum_{i',j',n',m'} (\mathbf{K}_\Lambda^{-1/2})_{ii'}^{nn'} (\mathbf{K}_\Lambda^{-1/2})_{jj'}^{mm'} \mathbb{E}(\boldsymbol{\chi}_{i'}^{n'}\boldsymbol{\chi}_{j'}^{m'}) \\ &= \left(\mathbf{K}_\Lambda^{-1/2}\mathbf{K}_\Lambda\mathbf{K}_\Lambda^{-1/2}\right)_{ij}^{nm} = \delta(i-j)\delta(n-m).\end{aligned} \tag{2.3.20}$$

This shows that the Gaussian variables $\tilde{\boldsymbol{\chi}}_i^n$ are independent. As a consequence, $\mathbf{K}_\Lambda^{-1}\boldsymbol{\chi}$ follows the multivariate Gaussian distribution with zero mean and covariance $\mathbf{K}_\Lambda^{-1}$. ■

This proposition shows that the random Gaussian current X_Λ^* can be identified with a random Gaussian vector of momenta $\boldsymbol{\alpha}$. We can therefore define the likelihood of the current X_Λ^*, $p(X_\Lambda^*)$ as the likelihood of the momenta $\boldsymbol{\alpha}$:

$$p(X_\Lambda^*) = p(\boldsymbol{\alpha}) \propto \exp\left(-\frac{\boldsymbol{\alpha}^t\mathbf{K}_\Lambda\boldsymbol{\alpha}}{2}\right) = \exp\left(-\frac{\|X_\Lambda^*\|_{W^*}^2}{2}\right), \tag{2.3.21}$$

where '$\propto$' means "proportional to". As expected, the log-likelihood of a Gaussian random current is proportional to the squared norm of the current (up to a fixed constant). This likelihood is of great interest to estimate statistics on currents as we will show in Chapter 5.

We notice that the direct simulation of such random currents is not possible due to the bad conditioning of the matrix $\mathbf{K}_\Lambda$ for fine grids. However, the robust implementation of the map $\mathcal{L}_W$ introduced in the next chapter (Chapter 3) will enable the simulation of Gaussian currents, as we will show in Section 3.5).

Remark 2.18. As mentioned already in Section 1.5.3, the variance of the Gaussian currents are completely determined by the kernel K^W. In particular, scaling the kernel K^W into $\sigma_W^2 K^W$ enables to adjust the magnitude to the variance of the Gaussian currents. In the discrete space, this allows us to write the likelihood of the discrete Gaussian currents as: $\exp\left(-\frac{\|X_\Lambda^*\|_{W^*}^2}{2\sigma_W^2}\right)$. This general form of the likelihood will be used in Chapter 5 to estimate statistical models on currents. □

2.4 Discrete approximation of currents and vector fields

In this section, we propose first an alternative way to compute the projection from the continuous space of currents W^* to the discrete space W_Λ^* for a grid Λ and, second, an interpolation scheme which reconstruct a dense vector field from its sample on grid nodes.

2.4.1 Closest neighbors projections of currents

In this section, we propose to project a current T into W_Λ, in a way which does not require to inverse the matrix $\mathbf{K}_\Lambda$ contrary to the orthogonal projection. This approach is sub-optimal since the orthogonal projection is the best approximation of T by a current in W_Λ in the sense of the W-norm. However, we will show in the sequel that the approximation error still converges to zero as the grid becomes finer and finer. We also provide the speed of convergence.

In this section, we suppose that T has a finite number of momenta and therefore of the form $T = \sum_{k=1}^{P} \delta_{y_k}^{\beta_k}$. This is the case if T is the approximation of a mesh or a polygonal lines in the sense of proposition 1.11. We denote $M(T) = \sum_k |\beta_k|$ the mass norm of T as shown in Proposition 1.12. We denote Λ a linearly spaced grid, so that all points y_k are at a distance much further than λ_W from the grid borders.

We propose two numerical schemes: a closest neighbor approach which project every momentum at the closest node of the grid and a partial volume projection (PVP) which project the momentum at every surrounding nodes.

2.4.1.1 Closest neighbor projection

We denote $\pi_\Lambda(y_k)$ the grid node which is the closest to the point y_k. Then we define $\tilde{T} = \sum_k \delta_{\pi_\Lambda(y_k)}^{\beta_k}$. The following proposition suggests that the convergence of $\tilde{T}$ to T as the grid step tends to zero has the speed at least of order Δ/λ_W, where λ_W is the "rate of decay" of the kernel (the standard deviation of a Gaussian kernel, for instance). The more "flat" the kernel at 0, the faster the convergence.

Proposition 2.19. *Let K be a translation-invariant scalar kernel ($K(x,y) = g\left(\frac{x-y}{\lambda_W}\right)\mathrm{Id}$) and p the smallest integer such that the pth derivative of g at zero (a p-covariant tensor denoted $d_0^{(p)}g$) is non null. Then in the limit $\Delta \to 0$,*

$$\left\|T - \tilde{T}\right\|_{W^*} \leq CM(T) \left|d_0^{(p)} g\right|^{1/2} \left(\frac{\Delta}{\lambda_W}\right)^{p/2} \Delta^{p/2}, \tag{2.4.1}$$

and this upper-bound is optimal.

Proof. We have in the limit $\Delta \to 0$:

$$\begin{aligned}
\left\|T - \tilde{T}\right\|_{W^*}^2 &\leq \sum_k \left\|\delta_{y_k}^{\beta_k} - \delta_{\pi_\Lambda(y_k)}^{\beta_k}\right\|_{W^*}^2 \\
&\leq \sum_k |\beta_k|^2 \, 2 \left| g(0) - g\left(\frac{\pi_\Lambda(y_k) - y_k}{\lambda_W}\right)\right| \\
&\leq \sum_k |\beta_k|^2 \, 2C \underbrace{\left|d_0^{(p)} g\left(\frac{\pi_\Lambda(y_k) - y_k}{\lambda_W}, \ldots, \frac{\pi_\Lambda(y_k) - y_k}{\lambda_W}\right)\right|}_{\leq \left|d_0^{(p)} g\right| \left|\frac{\pi_\Lambda(y_k) - y_k}{\lambda_W}\right|^p} \\
&\leq CM(T)^2 \left|d_0^{(p)} g\right| \left(\frac{\Delta}{\lambda_W}\right)^p,
\end{aligned} \tag{2.4.2}$$

where the norm of the tensor is defined as $\left|d_0^{(p)}g\right| = \sup_{|u_i|\leq 1}\left|d_0^{(p)}g(u_1,\ldots,u_p)\right|$.

If T has only one momenta located at point y. We can choose the grid such that y is at the center of a mesh of the grid, so that $|y-x| = \Delta/2$. Let $T = \delta_y^\beta$. Then, $\tilde{T} = \delta_x^\beta$ and $\left\|T-\tilde{T}\right\|_{W^*}^2 = 2\left|\beta\right|^2\left|g(0)-g(x-y)\right|$ which is asymptotically equivalent to $2/p!M(T)^2\left|k^{(p)}(0)\right|\Delta^p$. ■

For a Gaussian kernel, we have $g(x) = \exp(-\left|x\right|^2)$ and $p = 2$ (i.e. only the first derivative at zero is null), therefore,

$$\left\|T-\tilde{T}\right\|_{W^*} \leq C'M(T)\frac{\Delta}{\lambda_W}$$

2.4.1.2 Partial Volume Projection (PVP)

Instead of a closest neighbor approach, we can project the momenta (y_k,β_k) at the surrounding nodes (8 nodes in dimension 3) so that the partial volume interpolation (PVI) of these momenta retrieves the value β_k at point y_k. We call this inverse operation of the PPI, the partial volume projection (PVP). We recall that if f_i is the sampling of a function f on the grid nodes, then the partial volume interpolation inside this voxel has the form[3]:

$$f(x) = \sum_{i\in V(x)}\rho_i(x)f_i,$$

where $V(x)$ denotes the set of surrounding nodes of point x. The interpolating functions $\rho_i(x)$ satisfies two important properties:

- $\sum_{i\in V(x)}\rho_i(x) = 1$
- $\sum_{i\in V(x)}\rho_i(x)(x-x_i) = 0$

The partial volume projection of the momentum (y_k,β_k) consists in adding the momenta $(\pi_\Lambda^i(y_k),\rho_i(y_k)\beta_k)$ at each nodes $\pi_\Lambda^i(y_k)\in V(y_k)$ around the point y_k. Therefore, the new approximation of T is written as:

$$\bar{T} = \sum_k\sum_{\pi_\Lambda^i(y_k)\in V(y_k)}\delta_{\pi_\Lambda^i(y_k)}^{\rho_i(y_k)\beta_k}.$$

The following proposition shows that the speed of convergence is the same than for the closest neighbor procedure. However, we will show below than the error measured with the L^∞-norm (i.e. the numerical error) is smaller for the PVP.

Proposition 2.20. *Let K be a translation-invariant scalar kernel ($K(x,y) = g\left(\frac{x-y}{\lambda_W}\right)\mathrm{Id}$) and p the smallest integer such that the pth derivative of g at zero (a p-covariant tensor denoted $d_0^{(p)}g$) is non null. Then in the limit $\Delta\to 0$,*

$$\left\|T-\bar{T}\right\|_{W^*} \leq CM(T)\left|d_0^{(p)}g\right|^{1/2}\left(\frac{\Delta}{\lambda_W}\right)^{p/2} \tag{2.4.3}$$

and this upper-bound is optimal.

[3] In 1D, a point x have two neighbors: x_+ and x_- and we have the interpolating formula: $f(x) = \frac{x-x_-}{x_+-x_-}f_+ + \frac{x_+-x}{x_+-x_-}f_-$ so that $\rho_+(x) = (x-x_-)/\Delta$ and $\rho_-(x) = (x_+-x)/\Delta$. This formula can be extended easily in 3D: the point x divides the cube delimited by the 8 neighboring nodes into 8 sub-volumes and $\rho_i(x)$ is equal to the sub-volume opposite to the node i divided by Δ^3

Proof. Since $\sum_i \rho_i(x) = 1$, we have:

$$\begin{aligned}
\left\|T - \bar{T}\right\|_{W^*} &\leq \sum_k \left| \sum_{\pi^i_\Lambda(y_k) \in V(y_k)} \delta_{y_k}^{\rho_i(y_k)\beta_k} - \delta_{\pi^i_\Lambda(y_k)}^{\rho_i(y_k)\beta_k} \right| \\
&\leq \sum_k \sum_{\pi^i_\Lambda(y_k) \in V(y_k)} \rho_i(y_k) \left|\beta_k\right| \sqrt{2\left(g(0) - g\left(\frac{y_k - \pi^i_\Lambda(y_k)}{\lambda_W}\right)\right)} \qquad (2.4.4) \\
&\leq C \sum_k \left|\beta_k\right| \left| d_0^{(p)} g \right|^{1/2} \frac{\Delta^{p/2}}{\lambda_W^{p/2}}
\end{aligned}$$

Similar arguments than for the closest neighbor approach show that we cannot expect a speed of convergence better than $\left|k^{(p)}(0)\right|^{1/2} \Delta^{p/2}$ in a general setting. ■

For a Gaussian kernel, we have:

$$\left\|T - \bar{T}\right\|_{W^*} \leq C' M(T) \frac{\Delta}{\lambda_W} \qquad (2.4.5)$$

2.4.1.3 Convergence in L^∞_Λ-norm

Once the momenta have been projected on the grid nodes, the resulting currents $\tilde{T}$ and $\bar{T}$ belong to W^*_Λ. We can use then discrete convolution to compute the images of vectors $\tilde{\boldsymbol{\gamma}}$ and $\bar{\boldsymbol{\gamma}}$ which correspond to the sampling of the dense vector fields $\mathcal{L}_W^{-1}(\tilde{T})$ and $\mathcal{L}_W^{-1}(\bar{T})$ at the grid nodes.

In this section, we want measure the numerical precision of the vectors $\tilde{\boldsymbol{\gamma}}$ and $\bar{\boldsymbol{\gamma}}$ compared to the true value $\boldsymbol{\gamma} = \{\mathcal{L}_W^{-1}(T)(x_i)\}_{i \in \Lambda}$. For this purpose, we introduce the L^∞_Λ-norm defined by:

$$\left\|T - \tilde{T}\right\|_{\infty,\Lambda} = \sup_{x \in \Lambda} \left| \mathcal{L}_W^{-1}(T - \tilde{T})(x) \right|, \qquad (2.4.6)$$

which correspond to the infinity norm between the Np-dimensional vectors $\boldsymbol{\gamma}$.

In particular, we show that both approximations converges for the L^∞_Λ-norm. However, the speed of convergence is one order better for the PVP approach than for the closest neighbor approach.

Proposition 2.21. *Let K be a translation-invariant scalar kernel $K(x,y) = g((x-y)/\lambda_W)\mathrm{Id}$ with g a scalar function such that the pth derivative $d^p_x g$ (seen as a p-covariant tensor) is uniformly bounded*[4] *($\sup_{x \in \mathbb{R}^d} \sup_{|u_i| \leq 1} \left| d_x^{(p)} g(u_1, \ldots, u_p) \right| \leq Cte$).*

We have then,

- $\left\|T - \tilde{T}\right\|_{\infty,\Lambda} \leq C M(T) \frac{\Delta}{\lambda_W}$
- $\left\|T - \bar{T}\right\|_{\infty,\Lambda} \leq C M(T) \frac{\Delta^2}{\lambda_W^2}$

where C is a numerical constant independent of T, λ_W and the grid. This upper-bounds are optimal.

[4] Notice that the Gaussian kernel for which $g(x) = \exp(-|x|^2)$ satisfies these conditions.

Lemma 2.22. *Let $f_{x,y,y'}(t)$ be the function:*

$$f_{x,y,y'}(t) = g\left(\frac{x-y}{\lambda_W} + t\frac{y-y'}{\lambda_W}\right),$$

for $t \in [0,1]$. Then, in the limit $|y-y'| \to 0$:

- $\sup_{x\in\mathbb{R}^d} \left|f^{(p)}_{x,y,y'}(0)\right| = C\frac{|y-y'|^p}{\lambda_W^p}$,
- $\sup_{x\in\mathbb{R}^d, t\in[0,1]} \left|f^{(p)}_{x,y,y'}(t)\right| \leq C\frac{|y-y'|^p}{\lambda_W^p}$,

where C denotes a generic constant which does not depend either on x, y, y', or on λ_W.

Proof of Lemma 2.22. The p*th* derivative of the function f is given by:

$$\begin{aligned} f^{(p)}_{x,y,y'}(t) &= d^{(p)}_{\xi_t} g\left(\frac{y-y'}{\lambda_W}, \ldots, \frac{y-y'}{\lambda_W}\right) \\ &= \left|\frac{y-y'}{\lambda_W}\right|^p d^{(p)}_{\xi_t} g\left(\frac{y-y'}{|y-y'|}, \ldots, \frac{y-y'}{|y-y'|}\right), \end{aligned} \tag{2.4.7}$$

where $\xi_t = \frac{x-y}{\lambda_W} + t\frac{y-y'}{\lambda_W}$. Since the differential of g is supposed to be uniformly bounded we have:

$$\begin{aligned} \left|f^{(p)}_{x,y,y'}(0)\right| &\leq \left|\frac{y-y'}{\lambda_W}\right|^p \sup_{x\in\mathbb{R}^d} \sup_{|u|=1} \left|d^{(p)}_x g(u, \ldots, u)\right| \\ &\leq \left\|d^{(p)} g\right\|_\infty \left|\frac{y-y'}{\lambda_W}\right|^p \end{aligned} \tag{2.4.8}$$

Similarly, since only ξ_t depends on t in the upper-bound in Eq. (2.4.7), we also have:

$$\sup_{t\in[0,1]} \left|f^{(p)}_{x,y,y'}(t)\right| \leq \left\|d^{(p)} g\right\|_\infty \left|\frac{y-y'}{\lambda_W}\right|^p \tag{2.4.9}$$

■

Proof of Proposition 2.21. Now, we can prove the proposition 2.21. We recall that $T = \sum_p \delta^{\beta_p}_{y_p}$ and $\tilde{T} = \sum_p \delta^{\beta_p}_{\pi_\Lambda(y_p)}$ where $\pi_\Lambda(y_p)$ is the closest node to y_p. We have then:

$$\left\|T - \tilde{T}\right\|_{\infty,\Lambda} \leq \sum_p |\beta_p| \sup_{x\in\Lambda} |K(x, y_p) - K(x, \pi_\Lambda(y_p))| \tag{2.4.10}$$

With the notations of the lemma 2.22: $|K(x,y_p) - K(x,\pi_\Lambda(y_p))| = \left|f_{x,\pi_\Lambda(y_p),y_p}(1) - f_{x,\pi_\Lambda(y_p),y_p}(0)\right| = \left|f'(0) + \int_0^1 f''(t)(1-t)dt\right|$. And the uniform bounds of the lemma 2.22 imply that:

$$\sup_{x\in\Lambda} |K(x,y_p) - K(x,\pi_\Lambda(y_p))| = C\frac{|y_p - \pi_\Lambda(y_p)|}{\lambda_W} + \mathcal{O}\left(\frac{|y_p - \pi_\Lambda(y_p)|^2}{\lambda_W^2}\right) \leq C\frac{\Delta}{\lambda_W} + \mathcal{O}\left(\frac{\Delta^2}{\lambda_W^2}\right) \tag{2.4.11}$$

This leads to:

$$\left\|T - \tilde{T}\right\|_{\infty,\Lambda} \leq CM(T)\frac{\Delta}{\lambda_W} + \mathcal{O}\left(\frac{\Delta^2}{\lambda_W^2}\right)$$

and $\left\|T-\tilde{T}\right\|_{\infty,\Lambda}$ is asymptotically equivalent to this upper-bound when T has a single momentum.

Using tri-linear interpolation instead of closest neighbor projection leads to:

$$\bar{T} = \sum_p \sum_{\pi^i_\Lambda(y_p)\in V(y_p)} \delta^{\rho_i(y_p)\beta_p}_{\pi^i_\Lambda(y_p)}.$$

Since $\sum_p \rho_i(y_p) = 1$, we have:

$$\begin{aligned}\left\|T-\bar{T}\right\|_{\infty,\Lambda} &\leq \sum_p |\beta_p| \sup_{x\in\Lambda} \left| \sum_{i\in V(y_p)} \rho_i(y_p)(K(x,y_p) - K(x,\pi^i_\Lambda(y_p))) \right| \\ &\leq \sum_p |\beta_p| \sup_{x\in\Lambda} \left| \sum_i \rho_i(y_p) f'_{x,y_p,\pi^i_\Lambda(y_p)}(0) + \right. \\ &\left. \frac{1}{2}\sum_i \rho_i(y_p) f''_{x,y_p,\pi^i_\Lambda(y_p)}(0) + \frac{1}{2}\sum_i \rho_i(y_p) \int_0^1 f'''_{x,y_p,\pi^i_\Lambda(y_p)}(t)(1-t)^2 dt \right|\end{aligned} \tag{2.4.12}$$

However, by linearity of the differential,

$$\sum_{i\in V(y_p)} \rho_i(y_p) f'_{x,y_p,\pi^i_\Lambda(y_p)}(0) = d_{\frac{x-y_p}{\lambda_W}} g \underbrace{\left(\sum_{i\in V(y_p)} \rho_i(y_p)(y_p - \pi^i_\Lambda(y_p)) \right)}_{=0} = 0,$$

so that the first term of the upper-bound of Eq. (2.4.12) is null. This allows us to have a better precision than the closest neighbor projection. More precisely, thanks to the uniform upper-bounds proved in lemma 2.22, we get that the next two terms are bounded by:

$$\frac{C}{\lambda_W^2} \sum_i \rho_i(y_p) \left| y_p - \pi^i_\Lambda(y_p) \right|^2 + \mathcal{O}\left(\frac{\Delta^3}{\lambda_W^3}\right) \leq C\frac{\Delta^2}{\lambda_W^2} + \mathcal{O}\left(\frac{\Delta^3}{\lambda_W^3}\right)$$

and eventually:

$$\left\|T-\bar{T}\right\|_{\infty,\Lambda} \leq CM(T)\frac{\Delta^2}{\lambda_W^2} + \mathcal{O}\left(\frac{\Delta^3}{\lambda_W^3}\right)$$

■

2.4.2 Reconstruction of dense vector fields

In the previous section, we showed how grids can be used to project a current T on a finite-dimensional approximation space W^*_Λ. On such a space, the usual operations on currents can be performed efficiently via discrete convolutions and FFTs.

In particular, we saw how to approximate the samples of the vector field $\mathcal{L}_W^{-1}(T)$ at the grid nodes (i.e. approximating the Np dimensional vectors $\boldsymbol{\gamma}$ by $\tilde{\boldsymbol{\gamma}}$ or $\bar{\boldsymbol{\gamma}}$). We still need to propose an interpolation scheme to reconstruct a *dense* vector field from the samples $\tilde{\boldsymbol{\gamma}}$ or $\bar{\boldsymbol{\gamma}}$ and to show that the reconstructed dense vector field is a good approximation of the true vector field $\mathcal{L}_W^{-1}(T)$.

Here, we propose two interpolation schemes: $\tilde{\gamma}(x)$ is a piecewise constant interpolation of the samples $\tilde{\boldsymbol{\gamma}}$ and $\bar{\gamma}(x)$ a partial volume interpolation of the samples $\bar{\boldsymbol{\gamma}}$. The following property determines the speed of convergence of such reconstructed vector fields to the true $\gamma = \mathcal{L}_W^{-1}(T)$ for the L^∞-norm.

Definition 2.23. *For any point x in the convex hull of the grid Λ, we denote $\pi_\Lambda(x)$ the grid node closest to x and $V(x) = \{\pi_\Lambda^i(x)\}_{i=1,\dots,2^d}$ the set of nodes surrounding point x.*

Let $T = \sum_p \delta_{y_p}^{\beta_p}$ be a generic current in T, we denote $\tilde{T} = \sum_p \delta_{\pi_\Lambda(y_p)}^{\beta_p}$ and $\bar{T} = \sum_p \sum_{i \in V(y_p)} \delta_{\pi_\Lambda^i(y_p)}^{\rho_i(y_p)\beta_p}$ the approximations of T by the closest neighbor approach and the PVP approach respectively (as defined in the previous section).

Let y be a point in the convex hull of the grid, then we define $\tilde{\gamma}$ and $\bar{\gamma}$ by:

$$\begin{aligned} \tilde{\gamma}(y) &= \tilde{\gamma}(\pi_\Lambda(y)), \\ \bar{\gamma}(y) &= \sum_{k \in V(y)} \rho_k(y) \bar{\gamma}(\pi_\Lambda^k(y)), \end{aligned} \tag{2.4.13}$$

where the values of $\tilde{\gamma}$ at grid nodes has been computed as $\{\tilde{\gamma}(x_i)\}_{i\in\Lambda} = \tilde{\boldsymbol{\gamma}} = \mathbf{K}_\Lambda \tilde{\boldsymbol{\beta}}$ where $\tilde{\boldsymbol{\beta}}$ is Np-dimensional vectors corresponding the momenta of $\tilde{T}$ and the values of $\bar{\gamma}$ as: $\{\bar{\gamma}(x_i)\}_{i\in\Lambda} = \bar{\boldsymbol{\gamma}} = \mathbf{K}_\Lambda \bar{\boldsymbol{\beta}}$ where $\bar{\boldsymbol{\beta}}$ is Np-dimensional vectors corresponding the momenta of $\bar{T}$ (involving the coefficients ρ_i).

Remark 2.24. These formula define $\tilde{\gamma}$ and $\bar{\gamma}$ only inside the domain delimited by the grid. At this point, we can consider either that the grid is an infinite 3D lattice, or that these vectors fields are equal to 0 outside the domain delimited by the grid. In this last case, we assume that the grid borders are far enough from the momenta of T so that $\gamma = \mathcal{L}_W^{-1}(T)$ is smaller than any positive constant outside the domain delimited by the grid. Therefore the values of $\gamma(x)$ outside the grid do not impact the speed of convergence proved in the following proposition. □

Remark 2.25. The construction of the dense vector field can be written also as:

$$\begin{aligned} \tilde{\gamma}(x) &= \sum_p \tilde{K}(x, y_p)\beta_p \\ \bar{\gamma}(x) &= \sum_p \bar{K}(x, y_p)\beta_p \end{aligned} \tag{2.4.14}$$

where $\tilde{K}$ and $\bar{K}$ are two approximations of the kernel K defined by:

$$\begin{aligned} \tilde{K}(x, y) &= K(\pi_\Lambda(x), \pi_\Lambda(y)) \\ \bar{K}(x, y) &= \sum_{i \in V(x)} \sum_{j \in V(y)} \rho_i(x) K(\pi_\Lambda^i(x), \pi_\Lambda^j(y)) \rho_j(y) \end{aligned} \tag{2.4.15}$$

for any point x, y. $\tilde{K}$ is a piecewise constant approximation of K and $\bar{K}$ a piecewise linear approximation of the kernel.

Compared to the exact formula $\gamma(x) = \sum_p K(x, y_p)\beta_p$, these expressions show that the approximated kernels implement an approximation of the mapping $\mathcal{L}_W^{-1}$. This can be

written:

$$\begin{aligned}\tilde{\gamma} &= \tilde{\mathcal{L}_W^{-1}}(T)\\ \bar{\gamma} &= \bar{\mathcal{L}_W^{-1}}(T)\end{aligned} \tag{2.4.16}$$

□

We can now prove that the partial volume approach converges at a faster rate than the closest neighbor strategy:

Proposition 2.26. *With the notations of the definition 2.23 and under the same assumptions as in Proposition 2.21,*

$$\begin{aligned}\sup_{x\in \mathrm{Conv}(\Lambda)} |\gamma(x) - \tilde{\gamma}(x)| &\leq CM(T)\frac{\Delta}{\lambda_W}\\ \sup_{x\in \mathrm{Conv}(\Lambda)} |\gamma(x) - \bar{\gamma}(x)| &\leq CM(T)\frac{\Delta^2}{\lambda_W^2},\end{aligned} \tag{2.4.17}$$

where $\mathrm{Conv}(\Lambda)$ *denotes the convex hull of the grid* Λ *(see Remark 2.24).*

Proof. The proof uses the bounds of lemma 2.22. More precisely, one has:

$$|\gamma(y) - \tilde{\gamma}(y)| \leq \sum_p |\beta_p|\,|K(y, y_p) - K(\pi_\Lambda(y), \pi_\Lambda(y_p))| \tag{2.4.18}$$

We bound the absolute value by: $|K(y, y_p) - K(\pi_\Lambda(y), y_p)| + |K(\pi_\Lambda(y), y_p) - K(\pi_\Lambda(y), \pi_\Lambda(y_p))|$. Thanks to lemma 2.22, $|K(y, y_p) - K(\pi_\Lambda(y), y_p)| = C\frac{|y-\pi_\Lambda(y)|}{\lambda_W} + \mathcal{O}\left(\frac{|y-\pi_\Lambda(y)|^2}{\lambda_W^2}\right)$, and since $\sup_{y\in\mathbb{R}^d} |y - \pi_\Lambda(y)| \sim \Delta$ inside the convex hull of the grid, this shows that the first term is asymptotically equivalent to Δ/λ_W. The second term is bounded by $C\frac{|y_p-\pi_\Lambda(y_p)|}{\lambda_W} + \mathcal{O}\left(\frac{|y_p-\pi_\Lambda(y_p)|^2}{\lambda_W^2}\right)$. This upper bound is independent of y and is equivalent to Δ/λ_W. Then, similar computations as in the proof of Proposition 2.21 achieve the proof.

For the partial volume interpolation, one has:

$$|\gamma(y) - \bar{\gamma}(y)| \leq \sum_p |\beta_p| \left| \sum_{i\in V(y)} \sum_{j\in V(y_p)} \rho_i(y) \left(K(y, y_p) - K(\pi_\Lambda^i(y), \pi_\Lambda^j(y_p))\right) \rho_j(y_p) \right| \tag{2.4.19}$$

However,

$$\begin{aligned}\left|\sum_{i,j} \rho_i(y) \left(K(y, y_p) - K(\pi_\Lambda^i(y), \pi_\Lambda^j(y_p)\right) \rho_j(y_p)\right| &\leq \left|\sum_{i,j} \rho_i(y) \left(K(y, y_p) - K(\pi_\Lambda^i(y), y_p)\right) \rho_j(y_p)\right|\\ &+ \left|\sum_{i,j} \rho_i(y) \left(K(\pi_\Lambda^i(y), y_p) - K(\pi_\Lambda^i(y), \pi_\Lambda^j(y_p))\right) \rho_j(y_p)\right|\end{aligned} \tag{2.4.20}$$

Since $\sum_{i\in V(y)} \rho_i(y)(y - \pi_\Lambda^i(y)) = 0$ (resp. $\sum_{j\in V(y_p)} \rho_j(y_p)(y_p - \pi_\Lambda^j(y_p)) = 0$) the first term of the Taylor expansion of the first (resp. second) absolute value is null. Therefore, the first absolute value is of order $\left|y - \pi_\Lambda^i(y)\right|^2/\lambda_W^2$. The supremum with respect to y is

then of order Δ^2/λ_W^2. The second absolute value does not depend on y and is of order $\left|y_p - \pi_\Lambda^j(y_p)\right|^2/\lambda_W^2 \sim \Delta^2/\lambda_W^2$. We conclude with the same arguments as in the proof of Proposition 2.21. ■

Remark 2.27. Note that the reconstructed dense vector fields $\tilde{\gamma}$ and $\bar{\gamma}$ do not belong to W. Therefore, it does not make sense to compare these approximations with the norm W. □

Remark 2.28 (Higher-order interpolation)**.** In this section, we used piecewise constant and piecewise linear interpolation method, as well as an inverse method to distribute the momentum at the grid nodes. The first one uses B-splines of order 0 and lead to an approximation error of Δ/λ_W. The second one uses B-splines of order 1 and lead to an approximation error of Δ^2/λ_W^2. Using other interpolating functions of greater order must lead to smaller approximation errors, namely of order Δ^n/λ^n for n greater than 2. See [Thévenaz 2000, Meijering 2002] for surveys of interpolation methods and approximation theory. □

2.5 Approximation spaces in practice

Figure 2.3 summarizes what we have done so far. In Section 2.2, we introduced approximation spaces of vector fields W_Λ (bottom-left corner in Fig. 2.3) and their corresponding approximation spaces of currents W_Λ^* (bottom-right corner in Fig. 2.3). We showed in particular that the mapping $\mathcal{L}_W^{-1}$ can be computed in such spaces thanks to discrete convolutions. In Section 2.3, we showed that the orthogonal projection from the continuous space of vector fields (W) to the discrete space of vector fields (W_Λ) consists simply in the sampling of the vector field. The orthogonal projection in the spaces of currents, however, is not stable numerically. Then, in Section 2.4.1, we proposed instead to use a partial volume projection (PVP) to project any general currents in W^* to the approximation space W_Λ^* and prove the convergence to this projection when the grid step tends to zero. Eventually, in Section 2.4.2 we introduced an interpolation scheme which allows us to reconstruct a dense vector field from their samples on the grid nodes (inverse operation of the sampling). We showed that the succession of the projection, the convolution and the interpolation is an approximation of the map $\mathcal{L}_W^{-1}$ which implies a numerical error of order Δ^2/λ_W^2, where Δ is the step of the grid and λ_W the "rate of decay" of the kernel.

In this section, we show how this numerical framework can used to implement efficiently usual operations on currents.

2.5.1 How to compute usual operations on currents

We illustrate the use of the discrete approximations for the computation of the inner-product between two currents. This operation is the core of many algorithms, such as the computation of a fidelity-to-data term in the context of registration or the computation of the correlations between principal modes of a set of currents and each instance, for instance.

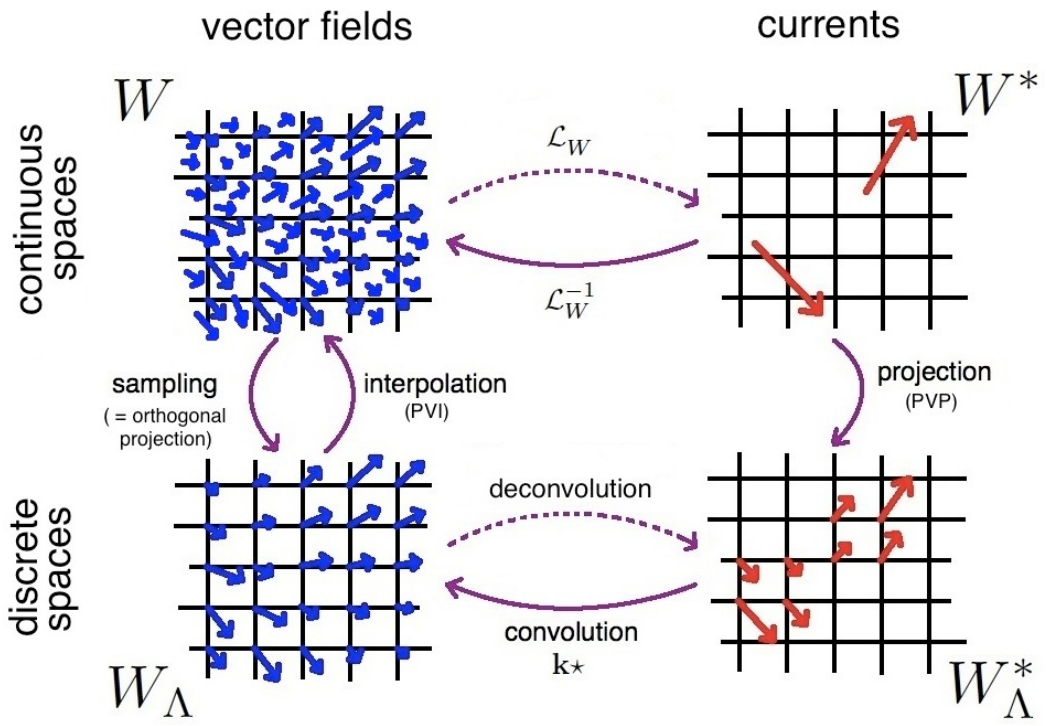


Figure 2.3: Global picture of the numerical framework proposed in this chapter. On the top row, continuous space of currents and vectors fields correspond via the map $\mathcal{L}_W$ and $\mathcal{L}_W^{-1}$. On the bottom row, linearly spaced grids define discrete approximation spaces of currents and vector field. The implementation of the map $\mathcal{L}_W^{-1}$ in this discrete setting can be reduced to a convolution problem. The inversion of this map in the discrete like in the continuous domain is ill-posed numerically. To map the continuous space to the discrete space, we use the orthogonal projection from W to W_Λ (which is equivalent to the sampling of the vector field) and a partial volume procedure to map the continuous space of currents to its discrete counterpart (Partial Volume Projection) and to map the discrete space of vector field to its continuous counterpart (Partial Volume Interpolation).

Given two currents $T = \sum_{i=1}^{N_\alpha} \delta_{x_i}^{\alpha_i}$ and $U = \sum_{j=1}^{N_\beta} \delta_{y_j}^{\beta_j}$, there are least three ways to compute the inner-product $\langle T, U \rangle_{W^*}$:

1. the exact computation with the double summation,
2. approximating T and U by $\bar{T}$ and $\bar{U}$ respectively and approximating $\langle T, U \rangle_{W^*}$ by $\langle \bar{T}, \bar{U} \rangle_{W^*}$,
3. approximating the vector field associated to T by $\bar{\gamma}(x)$ and approximating $\langle U, T \rangle_{W^*} = U(\mathcal{L}_W^{-1}(T))$ by $U(\bar{\gamma}) = \sum_{j=1}^{N_\beta} \bar{\gamma}(y_j)^t \beta_j$.

In the following, we compare the advantages and drawbacks of each method. In particular, we compare the approximation error, the time complexity and the space complexity (i.e. the memory needed to store the data) of each approach. Note that we use always the partial

volume projection instead of the closest neighbor approach since it enables to achieve the highest speed of convergence.

2.5.1.1 Solution ♯1: exact computation

We can use the analytical formula:

$$\langle T, U \rangle_{W^*} = \sum_{i=1}^{N_\alpha} \sum_{j=1}^{N_\beta} \beta_j^t K(y_j, x_i) \alpha_i \tag{2.5.1}$$

The time complexity of this computation is equal to $N_\alpha N_\beta$. It requires to store only the list of momenta of both currents. The space complexity is of $N_\alpha + N_\beta$.

2.5.1.2 Solution ♯2: approximating both currents

We set a grid Λ such that the grid borders are much further than λ_W from the data points (x_i) and (y_j) and such that Δ/λ_W is small (typically 0.2). Then, we follow the steps:

1. project the momenta (x_i, α_i) on the grid to give $\bar{T}$, stored as the image of vectors $\bar{\boldsymbol{\alpha}}$,
2. project the momenta (y_j, β_j) on the grid to give $\bar{U}$, stored as the image of vectors $\bar{\boldsymbol{\beta}}$,
3. compute the convolution $\mathbf{k} \star \bar{\boldsymbol{\alpha}}$ to give the image of vectors $\bar{\boldsymbol{\gamma}}$,
4. compute $\bar{\boldsymbol{\gamma}}^t \bar{\boldsymbol{\beta}}$ (as vectors of dimension Np).

The result is $\bar{\boldsymbol{\beta}}^t \mathbf{K}_\Lambda \bar{\boldsymbol{\alpha}} = \langle \bar{U}, \bar{T} \rangle_{W^*}$ which is an approximation of $\langle U, T \rangle_{W^*}$.
The speed of convergence of this approximation is given by:

$$\left| \langle \bar{U}, \bar{T} \rangle_{W^*} - \langle U, T \rangle_{W^*} \right| \leq \left| \langle \bar{U} - U, \bar{T} \rangle_{W^*} \right| + \left| \langle U, \bar{T} - T \rangle_{W^*} \right| \tag{2.5.2}$$

The first term is bounded by:

$$\begin{aligned} \left| \langle \bar{T}, U - \bar{U} \rangle_{W^*} \right| &= \left| \sum_{i=1}^{N_\alpha} \sum_{k \in V(x_i)} \mathcal{L}_W^{-1}(U - \bar{U})(\pi_\Lambda^k(x_i))^t \rho_k(x_i) \alpha_i \right| \\ &\leq \sum_{i=1}^{N_\alpha} \sum_{k \in V(x_i)} \rho_k(x_i) \left| \alpha_i \right| \sup_{x \in \Lambda} \left| \mathcal{L}_W^{-1}(U - \bar{U})(x) \right| \\ &= M(\tilde{T}) \left\| U - \bar{U} \right\|_{\infty, \Lambda} \end{aligned} \tag{2.5.3}$$

Moreover, the mass-norm satisfies $M(\bar{T}) \leq M(T)$. Indeed,

$$\begin{aligned} M(\bar{T}) &= \sum_{x_n^\Lambda \in \Lambda} \left| \sum_{\left\{ i; \quad \exists k, \pi_\Lambda^k(x_i) = x_n^\Lambda \right\}} \rho_k(x_i) \alpha_i \right| \\ &\leq \sum_{x_n^\Lambda \in \Lambda} \sum_{\left\{ i; \quad \exists k, \pi_\Lambda^k(x_i) = x_n^\Lambda \right\}} \rho_k(x_i) \left| \alpha_i \right| \\ &= \sum_{x_i} \sum_{k \in V(x_i)} \rho_k(x_i) \left| \alpha_i \right| = M(T) \end{aligned} \tag{2.5.4}$$

Therefore the first term of the upper-bound in Eq. (2.5.2) is bounded by: $M(T)\left\|U-\bar{U}\right\|_{\infty,\Lambda}$. Thanks to Proposition 2.21, this term is of order $\mathcal{O}(\Delta^2/\lambda_W^2)$.

Similarly, the second term of the upper-bound in Eq. (2.5.2) is bounded by:

$$\begin{aligned}\left|\left\langle U, T-\bar{T}\right\rangle_{W^*}\right| &= \left|\sum_{j=1}^{N_\beta} \mathcal{L}_W^{-1}(T-\bar{T})(y_j)^t\beta_j\right| \\ &\leq \sum_{j=1}^{N_\beta}|\beta_j| \sup_{x\in\mathbb{R}^d}\left|\mathcal{L}_W^{-1}(T-\bar{T})(x)\right| = M(U)\left\|T-\bar{T}\right\|_\infty\end{aligned} \tag{2.5.5}$$

For the eight-neighbor interpolation scheme, Proposition 2.26 shows that $\left\|T-\bar{T}\right\|_\infty$ is of order $\mathcal{O}\left(\frac{\Delta^2}{\lambda_W^2}\right)$.

Eventually, the speed of convergence is given by:

$$\begin{aligned}\left|\left\langle \bar{U},\bar{T}\right\rangle_{W^*} - \left\langle U,T\right\rangle_{W^*}\right| &\leq \left(M(T)\left\|U-\bar{U}\right\|_{\infty,\Lambda} + M(U)\left\|T-\bar{T}\right\|_\infty\right) \\ &\leq C\left(M(T)+M(U)\right)\frac{\Delta^2}{\lambda_W^2}.\end{aligned} \tag{2.5.6}$$

Remark 2.29. We bound the inner-product between two currents T and T' by:

$$\left|\left\langle T,T'\right\rangle_{W^*}\right| \leq M(T)\left\|T'\right\|_\infty, \tag{2.5.7}$$

instead of

$$\left|\left\langle T,T'\right\rangle_{W^*}\right| \leq \left\|T\right\|_W\left\|T'\right\|_W. \tag{2.5.8}$$

The first inequality (which is true only if the current T has a *finite* number of momenta) is more precise than the second inequality. Indeed, we have for generic currents $\left\|T\right\|_\infty \leq C\left\|T\right\|_W$.

Actually, measuring the approximation error with the W-norm instead of the L^∞-norm leads to a speed of convergence of Δ/λ_W only (instead of Δ^2/λ_W^2). □

2.5.1.3 Solution ♯3: approximating the dual representation of one current

As the computations of the speed of convergence of the previous case suggest, projecting both currents on grids is not the most natural idea. A much simpler idea in the framework of currents is to write:

$$\left\langle T,U\right\rangle_{W^*} = U(\mathcal{L}_W^{-1}(T)) = \sum_{j=1}^{N_\beta}\beta_j^t\gamma(y_j) \tag{2.5.9}$$

where $\gamma(y_j) = \mathcal{L}_W^{-1}(T)(y_j) = \sum_{i=1}^{N_\alpha}K(y_j,x_i)\alpha_i$.

In this expression, the vector field γ could be approximated thanks to the scheme introduced in Section 2.4.2. This leads to the following computations:

1. project the momenta of T: (x_i,α_i) on the grid nodes to give the image of vector $\bar{\boldsymbol{\alpha}}$,
2. compute $\bar{\boldsymbol{\gamma}} = \mathbf{k}\star\bar{\boldsymbol{\alpha}}$,
3. interpolate the values of $\bar{\boldsymbol{\gamma}}$ at points y_j to give $\bar{\gamma}(y_j)$,

solution	numerical error	space complexity	time complexity
♯1: Exact computation	0	$N_\alpha + N_\beta$	$N_\alpha N_\beta$
♯2: Approx. two currents	$(M(T) + M(U))\frac{\Delta^2}{\lambda_W^2}$	$2N_{\text{grid}}$	$N_{\text{grid}} \log(N_{\text{grid}})$ $+N_{\text{grid}} + N_\alpha + N_\beta$
♯3: Using rep. vector field	$M(U)\frac{\Delta^2}{\lambda_W^2}$	N_{grid}	$N_{\text{grid}} \log(N_{\text{grid}}) + N_\alpha + N_\beta$
♯4: FGT	depends on point distrib.	$N_\alpha + N_\beta$	$N_\alpha + N_\beta$

Table 2.2: Comparison of numerical solution to compute the inner-product between a current with N_α momenta and another with N_β momenta. In solution ♯2, N_{grid} is the number of point of a grid which covers the momenta of both currents. In solution ♯3, the grid needs only to cover the points of one current. Note that these number is an order of magnitude of the space and time complexity. In particular, it does not account for multiplicative constants. These constants do not depend much on the number of momenta for the implementation with grids, by contrast to the FGT implementation. Grid-based optimizations, however, are more sensitive to the spreading of the points in space.

4. compute $\sum_{j=1}^{N_\alpha} \beta_j^t \bar{\gamma}(y_j)$.

The speed of convergence of this approximation is given by (denoting the reconstructed vector field $\bar{\gamma}$ as $\bar{\mathcal{L}_W^{-1}}(T)$):

$$\begin{aligned}\left|U\left(\mathcal{L}_W^{-1}(T)\right) - U\left(\bar{\mathcal{L}_W^{-1}}(T)\right)\right| &= \left|\sum_{j=1}^{N_\beta}\left(\mathcal{L}_W^{-1}(T) - \bar{\mathcal{L}_W^{-1}}(T)\right)(y_j)^t\beta_j\right| \\ &\leq M(U)\left\|T - \bar{T}\right\|_\infty \leq CM(U)\frac{\Delta^2}{\lambda_W^2}\end{aligned} \tag{2.5.10}$$

thanks to Proposition 2.26.

Remark 2.30. This approach is not symmetric with respect of U and T, although switching the two currents leads to exactly the same result with the same approximation error. From a numerical point of view, however, only the current T needs to be projected and convoluted with the kernel. If one has a fixed current T and one needs to perform a bench of inner-product with several currents U_i, then it is of course much faster to perform only one convolution. If one needs to perform only one inner-product, then projecting T or U is equivalent. □

We can now compare the solution ♯2 and ♯3. In terms of speed of convergence, the two approaches are of the same order, whereas the solution ♯3 has a smaller constant $M(U)$ versus $M(U) + M(T)$.

In terms of time complexity, the solution ♯2 requires (1) to project both momenta on the grid (complexity $N_\alpha + N_\beta$), (2) compute one convolution (complexity $N_{\text{grid}} \log(N_{\text{grid}})$)

and (3) to compute an inner-product between images of vectors (complexity N_{grid}). The solution ♯3 requires (1) to project one momenta on the grid (complexity N_α), (2) to compute one convolution (complexity $N_{\text{grid}} \log(N_{\text{grid}})$) and (3) to interpolate the image of vectors at the locations of the momenta of U (complexity N_β). In both cases, the limiting factor is the same, namely the convolution of cost $N \log(N)$. However, once this cost has been discounted, the solution ♯2 has an extra cost of order N_{grid} compared with solution ♯3. This extra cost may be not negligible in case of repetition of such computations.

In terms of space complexity, the solution ♯2 requires to store 2 grids (one for the projection of each currents), whereas the solution ♯1 needs to store only 1 grid. Moreover, in case of solution ♯2, the grid must cover the points of both currents, whereas in solution ♯3 the grid must cover only the points of a single current (and we can choose the current whose momenta are contained in the smaller domain in space). Therefore, if the two currents are far one from the other, or if one current is spread over a much larger domain than the other, then using solution ♯3 has a significant advantage toward solution ♯2 in terms of memory storage.

Moreover, suppose that we would like to compute a bench of inner-product of the form $\langle T, U_i \rangle_{W^*}$ for a fixed current T and several currents U_i. Then, using the solution ♯2 would require to compute as many convolutions as the number of currents U_i. By contrast, in case of solution ♯3, only one convolution has to be computed as a preprocessing (the computation of the image of vectors associated to T) and only interpolation at different location has to be computed for each U_i. In this case, the solution ♯3 definitely outperforms the solution ♯2.

We can compare now these approximation with the direct approach which implements the analytical sum. The approximations with grids are faster if $N_{\text{grid}} \log(N_{\text{grid}}) \leq N_\alpha N_\beta$. The typical values are $N_{\text{grid}} = 10^6$ and $N_\alpha, N_\beta \geq 10^3$. In practical cases, the grid approximation method is almost always faster. Note that the number of grid nodes is determined by the spreading of the point set in space and the approximation parameter Δ/λ_W. The most favorable case is when points are concentrated locally in a small domain of space. The least favorable case is when few points are spread in a large domain (typically a long straight line). In this case, a multi-grid approach could be investigated: several small grids cover the point set and the combination of the local convolutions leads to an approximation of the convolution with a unique grid.

2.5.2 Comparison with fast multipole approximations

There are other numerical schemes to optimize the implementation of the map $\mathcal{L}_W^{-1}$, namely the computation of sums of the form: $\sum_{i=1}^{N_\alpha} K(y_j, x_i)\alpha_i$ at N_β points y_j for a given set of momenta $(x_i, \alpha_i)_{i=1,\ldots,N_\alpha}$.

In [Glaunès 2005, Glaunès 2008], J. Glaunès proposed to use the Improved Fast Gauss Transform [Greengard 1991, Yang 2003] as a fast multipole approximation of these sums. The principle is to gather the point set (x_i) into clusters and to use the Taylor expansion of the kernel at every center of the clusters. The sum over all the points becomes a sum over all centers of cluster weighted by some coefficients which are computed for each cluster

independently of the point y (i.e. which can be computed once for all for a given distribution of points x_i).

This method depends on two parameters: the number of clusters (or equivalently the maximum radius of the clusters) and the number of terms used in the truncation of the Taylor expansion. The resulting approximation error does not depend only on the values of these parameters but also on the distribution of the points x_i. This means that the values of the parameters does not determine a fixed approximation error. For a desired numerical precision, the user has to adapt the number of clusters to the particular distribution of the points on which he wants to apply the FGT.

There is simple empirical rule to determine automatically the parameters for a given approximation error. To set them, we usually compare the result of the approximation with the exact computation and determine the parameters for this particular distribution of points. Setting the parameters cannot be done automatically and is time-consuming. More importantly, this prevents us from using this method routinely to process arbitrary current. By contrast, the framework based on grid guarantees a fixed numerical precision, once the step of the grid and the kernel is fixed, independently of the current to be approximated.

It has been shown in [Glaunès 2005], that this approximation decreases the time complexity from $\mathcal{O}(N_\alpha N_\beta)$ to $\mathcal{O}(N_\alpha + N_\beta)$. This time-complexity may overcome the one of the grid-based approximations. This is the case when there are few momenta spread in a large domain of space (like a long curve compared to a volume or a concentrated set of points). An empirical comparison between both methods will be performed in Chapter 4 in the case of registration (see Section 4.4.4.2). Moreover, the FGT approximation does not require more memory space than the one needed to store the original currents ($N_\alpha + N_\beta$). By contrast, grid-based approximations require to store image of vectors which usually requires much more memory space.

The Fast Gauss Transform is less generic than the grid-based approximations, since it is dedicated to the Gaussian kernel. For general kernels, other multipole approximations could be derived, though, provided that we have an explicit Taylor expansion of the kernel.

2.5.3 Toward a complete computational framework

In this chapter, we provided a computational framework to compute any operations on currents which involves the mapping $\mathcal{L}_W^{-1}$, such as the computation of norm and inner-product as we have shown above. Given a current T on the continuous domain, we performed the map $\mathcal{L}_W^{-1}$ by following the 3 steps: first projection of the current on the grid, second discrete convolution and third interpolation. One of the main advantage of this framework is that the approximation error depends only on the ratio between the grid step Δ and the rate of decay of the kernel λ_W. In particular, it does not depend on the input data. This makes the overall framework very stable, generic and easy to set up. Once the grid and the kernel are fixed, we can perform any operations on currents while guaranteeing a fixed numerical precision.

This numerical framework will be used in Chapter 4 to give a new optimization scheme for the registration of currents. The purpose is to compute efficiently the dense vector field (the vector field driving the deformation, for instance) associated to current given as

a finite set of momenta in the continuous coordinates (the parameterization of the vector field). Using the discrete spaces based on linearly spaced grids allows us to approximate this operation. We proved in this chapter that the numerical error of this approximation is of order Δ^2/λ_W^2 where Δ is the grid step and λ_W^2 the typical rate of decay of the kernel. This reconstruction error is constant even if the momenta are moving in space. This makes a great difference compared to the optimization scheme based on Fast Gauss Transform proposed in the original implementation of the registration algorithm in [Glaunès 2005] (See Section 4.4.4).

However, this computational framework is not yet complete. Indeed, we do not provide any robust implementation of the map $\mathcal{L}_W$, or equivalently in the discrete case, the implementation of the multiplication with the matrix $\mathbf{K}_\Lambda^{-1}$. As a consequence, there is no way to map vector fields (on the left-hand side of Fig. 2.3) to currents (on the right-hand side of Fig. 2.3). Such an operation is not needed to compute the most usual operations in the space of currents (as we have shown in the example above). However, more complex operations on currents are only defined on their representation in terms of vector fields, like the adjoint push-forward action which will be defined in Chapter 5 (See Section 5.3.3). In this case, a robust deconvolution scheme is required. This problem will be addressed by the matching pursuit introduced in the next chapter (See Section 3.3 and Fig. 3.1). In addition to its ability to give a sparse representation of any currents, this algorithm enables to complete the computational framework presented in this chapter and to compute any operations from currents to vector spaces, from continuous space to discrete spaces and vice versa.

Chapter 3

Sparse representation of currents

This chapter has been partly published in [Durrleman 2008b, Durrleman 2009c]

Contents

3.1 The need for optimal decomposition of currents

In Chapter 1, we showed that discrete geometrical data could be approximated in the space of currents as a finite sum of Dirac delta currents. In the next chapters, we will use this representation to compute group-wise statistics. Such statistics result from a combination of all input currents. As such, they are still given as a sum of Dirac delta currents which has as many terms as the total number of Dirac delta currents in the database. For instance, linear statistics like mean and principal modes results from a linear combination of the input currents. The estimation of an atlas in Chapter 5 results also from a combination of the input currents, but in a non-linear way. The number of terms in the representation of such statistics increases linearly as the sampling of the input shape becomes finer (i.e. as the number of Dirac delta currents increases). This number of terms increases also as the number of subjects increases. In both cases, however, the estimated statistics converge to their true values in the space of currents. This representation of the statistics, which results from the standard operations on currents, is not stable from a computational point of view.

This instability is due to the fact that a representation of a current in terms of Dirac delta currents may be very redundant at the scale of analysis. This scale of analysis is

determined by the rate of decay of the kernel (λ_W), which is given for a Gaussian kernel as its standard deviation. If two Dirac delta currents in the decomposition of a current point in the same direction α and are located at two points x and y much closer than λ_W ($|x-y| \ll \lambda_W$), then they contribute to the same mode of the corresponding vector field in W: $K(.,x)\alpha + K(.,y)\alpha$. This mode may be approximated by a single Dirac delta current located at $(x+y)/2$ with a doubled magnitude: $K(.,x)\alpha + K(.,y)\alpha \sim 2K(.,(x+y)/2)\alpha$. This approximation is controlled by the ratio $|x-y|/\lambda_W$. Using such approximations, it would be possible to compress the representation of currents, which is highly redundant at the scale λ_W, with much fewer terms.

For this purpose, we take advantage of the fact that the set of Dirac delta currents is not a Hilbert basis of the space of currents. Admittedly, the Dirac delta currents are linearly independent and span the space of currents, but they are not orthogonal[1]: $\left\langle \delta_x^\alpha, \delta_y^\beta \right\rangle_{W^*} = \alpha^t K(x,y)\beta \neq 0$. This means that the decomposition of a current T as a (possibly infinite) linear combination of Dirac delta currents is not unique. Note that this is not surprising: we proved that the approximation of a surface mesh as a finite sum converges to the continuous limit when the sampling of the mesh becomes finer. Two different sampling converge to the same continuous limit. Now, if the decomposition of a current $T = \sum_{i=1}^{\infty} \delta_{x_i}^{\alpha_i}$ is redundant, we would like to re-write this current on a new series of Dirac delta currents: $T = \sum_{i=1}^{\infty} \delta_{x'_i}^{\alpha'_i}$. Among all the possible decompositions, a "good" decomposition satisfies two properties: (1) Dirac currents should be almost orthogonal to avoid redundancy: $\langle \delta_x, \delta_y \rangle_{W^*} \sim 0$ which implies in particular that $|x-y| \gg \lambda_W$ and (2) the Dirac currents should be chosen so that they are adapted to the signal, namely so that they have a strong correlation with the original current T (i.e. so that they maximize $\langle T, \delta_x^\alpha \rangle_{W^*}$). Consider the N first terms of this series $\tilde{T} = \sum_{i=1}^{N} \delta_{x_i}^{\alpha_i}$. The first condition implies that for a fixed approximation error ($\left\| T - \tilde{T} \right\|_{W^*}$), the number of terms N is minimized. The second condition implies that for a fixed number of terms N, the momenta are chosen to minimize the approximation error. Eventually, the first terms of this new decomposition allows us to give a sparse representation of the current at any desired accuracy.

The purpose of this chapter is precisely to propose a matching pursuit algorithm, which is a greedy approach to estimate such an optimal decomposition of currents. We show that the optimal Dirac delta currents are given as the main "modes" of the dense vector field associated to the current T: $\gamma = \mathcal{L}_W^{-1}(T)$. Indeed, this vector field which results from a convolution of every momentum, already integrates the redundancy of the decomposition at the scale λ_W. We show that this iterative procedure leads to an approximation of the initial current which converges to the initial current. This offers a way to control the precision of the approximation.

The possible applications of the matching pursuit are threefold:

[1] By contrast to finite dimensional spaces, a set of linearly independent vectors which spans an infinite dimensional vector space is not necessarily a basis of this vector space, in the sense that every vector has a unique decomposition on this set of vectors. In infinite dimension, an extra condition is required like the orthonormality condition in Hilbert spaces for instance (which leads, up to a change of the metric, to the general Riesz bases).

- **Compression:** The raw anatomical data may be given with a huge number of points, like fiber bundles (see Fig. 3.10) which may contain up to 10^5 segments concentrated in a small domain of space. This geometrical information is often highly redundant at the relevant scale of analysis. The matching pursuit could be used to give a sparse representation of these data with much fewer points without altering the geometrical information contained in the data at the determined scale. This is of uttermost importance to store and handle the data efficiently. For instance, this may divide drastically the time needed to register the data.

- **Interpretation:** The representation of a current which results from standard computations in the space of currents may be particularly difficult to interpret. The representation of a mean surface as a collection of weighted normals does not give a precise idea of the common anatomical features that this mean captures. By contrast, the dense vector field associated to this mean surface averages the contribution of each surface and therefore highlights the areas where the information is "concentrated". The sparse representation of the mean shows precisely the main "modes" of this distribution which leads to much more interpretable results.

- **Deconvolution:** Some operations on currents are defined via computations on the space of vector fields W. This is the case for the adjoint push-forward action on currents, which will be introduced in Chapter 5. The simulation of a Gaussian noise is only possible numerically in the space W. In such cases, the parameterization of the vector field in terms of momenta is not given beforehand. The direct deconvolution of the vector field (i.e. the computation of the mapping $\mathcal{L}_W$) is ill-posed numerically. The matching pursuit precisely gives an approximation of the solution of the equation $\gamma = \mathcal{L}_W(T)$ (when one knows γ and looks for T) at any desired accuracy. In this context, the matching pursuit is seen as a robust deconvolution scheme.

3.2 Orthogonal matching pursuit for currents

The matching pursuit algorithm is quite a natural idea which has emerged independently in different communities. In [Friedman 1981, Huber 1985, Jones 1987], projection pursuit regression was introduced for statistical purposes: the goal was to find an estimation of the conditional expectation of a random variable Y with respect to a very large number of variables $X_1, \ldots, X_n$. One finds the best linear combination of the variables X_i that makes the conditional expectation the highest possible, and one iterates on the residual expectation. Eventually, one ends up with a much smaller set of variables which gives a good approximation of the conditional expectation. In some sense, the estimated variables best explain the variable Y. In [Mallat 1993, Pati 1993], matching pursuit was proposed in the signal processing community to find the best decomposition of a signal on a redundant dictionary of time-frequency atoms and later to find adapted wavelets bases for image decomposition. Further improvements of the algorithm have been proposed, especially in the field of compressed sensing [Donoho 2006, Needell 2008]. The underlying idea remains the same: one find iteratively the atom which best correlates with the residual signal. In

this section, we introduce an adaptation of this algorithm for currents. To the best of our knowledge, this is the first time that this idea is introduced in the framework of currents for geometrical data compression and as a robust deconvolution scheme.

3.2.1 The continuous case

Let $\gamma \in W$ be a vector field and T its associated current ($\gamma = \mathcal{L}_W^{-1}(T)$). Our goal is to write T as an infinite series with fast decreasing terms, so that the first n terms of this series will provide the best approximation of the current. So, if we fix n, we want to find a set of n points (x_i) and n vectors (α_i) such that the current $\Pi_n(T) = \sum_{i=1}^n \delta_{x_i}^{\alpha_i}$ is as close as possible to T.

If one knows the set of optimal point positions $\mathcal{X}_n = \{x_i\}_{1 \leq i \leq n}$, we can define the n-dimensional sub-space of currents as in Section 2.2:

$$W_{\mathcal{X}_n}^* = \text{Span}\left(\delta_{x_i}^{\epsilon_k}; \quad k = 1, \ldots, p, i = 1, \ldots, n\right),$$

where $(\epsilon_k)_{k=1,\ldots,p}$ is the canonical basis of $\mathbb{R}^p$ (in the following the index k always takes values $k = 1, \ldots, p$, where $p = 0$ for point sets and volumes, $p = 2$ for planar curves, $p = 3$ for curves and surfaces in 3D). Therefore, the current $\Pi_n(T)$ in $W_{\mathcal{X}_n}^*$ which is the closest to T is given as the orthogonal projection of T onto $W_{\mathcal{X}_n}^*$. This orthogonal projection is given in Corollary 2.12. We can write also directly the orthogonality conditions: $\langle \Pi_n(T), \delta_{x_i}^{\epsilon_k} \rangle_{W^*} = \langle T, \delta_{x_i}^{\epsilon_k} \rangle_{W^*}$, which leads to, applying the isometric mapping $\mathcal{L}_W^{-1}$:

$$\sum_{j=1}^n \alpha_j {}^t K^W(x_i, x_j)\epsilon_k = \langle \mathcal{L}_W^{-1}(T), K^W(., x_i)\epsilon_k \rangle_W$$

Thanks to the reproducing property in Eq. (1.2.4), this last expression is equal to $\mathcal{L}_W^{-1}(T)(x_i)^t \epsilon_k = \gamma(x_i)_k$. Finally, this gives the set of pn linear equations:

$$\sum_{j=1}^n \left(K^W(x_i, x_j)\alpha_j\right)_k = \gamma(x_i)_k \tag{3.2.1}$$

Solving this linear system leads to the optimal vectors $(\alpha_i)_{1 \leq i \leq n}$, once one knows the optimal positions $(x_i)_{1 \leq i \leq n}$. Finding the optimal points (x_i) has been proved to be NP-hard in general [Davis 1997]. The orthogonal matching pursuit algorithm is a suboptimal greedy approach to this problem: the first point x_1 is the one for which the projection of T on $W_{\mathcal{X}_n}^* = \text{Span}(\delta_x^{\epsilon_k})$ is maximal. δ_{x_1} is the basis vector which enables to explain the largest part of the signal. Since

$$\langle T, \delta_x^{\epsilon_k} \rangle_{W^*} = \langle \gamma, K^W(., x)\epsilon_k \rangle_W = \gamma(x)_k$$

x_1 is a point where $|\gamma(x)|$ is maximal. Solving Eq. (3.2.1) gives $\alpha_1^1 = K^W(x_1, x_1)^{-1}\gamma(x_1)$. Then we remove from γ its orthogonal projection on $\mathcal{L}_W^{-1}(\delta_{x_1}^{\epsilon_k})$ (i.e. the first basis vector in the space W): this gives the first residual vector field: $\gamma_1(x) = \gamma(x) - K^W(x, x_1)\alpha_1^1$. And we iterate the procedure on this residual.

Eventually, at step n of the algorithm, γ_{n-1} stores the residual vector field, x_n is determined as the point which achieves the maximum of γ_{n-1} and the vectors $\alpha_1^n, \ldots, \alpha_n^n$ are

such that $\Pi_n(T) = \sum_{i=1}^{n} \delta_{x_i}^{\alpha_i^n}$ is the orthogonal projection of T on $W^*_{\mathcal{X}_n}$. This leads to the equations:

$$\begin{cases} x_n = \arg\max_{x\in\mathbb{R}^3} |\gamma_{n-1}(x)| & \text{(find the next position)} \\ \boldsymbol{\alpha}^n = \mathbf{K}_{\mathcal{X}_n}^{-1}\boldsymbol{\gamma}^n & \text{(compute associated momenta)} \\ \gamma_n(x) = \sum_{i=1}^{n} K(x,x_i)\alpha_i^n, \quad \text{for all } x\in\mathbb{R}^d & \text{(compute residual vector field)} \end{cases},$$

where we denote as in Chapter 2: $\boldsymbol{\alpha}^n$ (resp. $\boldsymbol{\gamma}^n$) the concatenation of vectors $(\alpha_i^n)_{1\leq i\leq n}$ (resp. $(\gamma(x_i))_{1\leq i\leq n}$) and $\mathbf{K}_{\mathcal{X}_n}$ the np-by-np block matrix whose block (i,j) is given as $K^W(x_i,x_j)$. About the possible bad conditioning of this matrix, see Remark 3.9.

Eventually, the algorithm is summarized in Algorithm 1.

Algorithm 1 Orthogonal Matching Pursuit for Currents (continuous version)

Input: a vector field γ, a threshold $\eta > 0$
$\gamma_0 = \gamma$, $n = 0$
while $\|\gamma_n\|_\infty \geq \eta$ **do**
  $x_{n+1} = \text{argmax}_{x\in\mathbb{R}^3} |\gamma_n(x)|$
  Find $(\alpha_i^{n+1})_{1\leq i\leq n+1}$ by solving $\sum_{q=1}^{n+1} \left(K^W(x_i,x_q)\alpha_q^{n+1}\right)_k = \gamma(x_i)_k$
  $\gamma_{n+1} = \gamma - \sum_{i=1}^{n+1} K^W(.,x_i)\alpha_i^{n+1}$
  $n \leftarrow n+1$
end while
Output: list of $(x_i,\alpha_i^n)_{i=1\ldots n}$

After n steps, the algorithm gives an approximation of $T = \mathcal{L}_W(\gamma)$ with n delta Dirac currents: $\Pi_n(T) = \sum_{i=1}^{n} \delta_{x_i}^{(\alpha_i^n)}$. The theorem 3.1 shows that $\Pi_n(T)$ converges to T as n tends to infinity (e.g. $\|\Pi_n(T) - T\|_{W^*} \to 0$). The auxiliary variable $\gamma_n = \gamma - \mathcal{L}_W^{-1}(\Pi_n)$ stores the residual vector field that remains to be explained. We prove also in theorem 3.1 that the L^∞-norm of this residue tends towards zeros as n tends towards infinity (e.g. $\|\gamma_n\|_\infty = \sup_{x\in\mathbb{R}^3} |\gamma_n(x)| \to 0$). This means, in particular, that the norm of the residue is below any positive threshold in finite time, thus proving that the algorithm actually finishes.

The threshold η has to be specified for every application. If T is a linear combination of p currents $T_1,\ldots,T_p$ (such as mean current, principal mode, difference between two currents, etc.), we can choose η as a fixed ratio of the standard deviation of the set of currents: $\eta = \tau\sigma$ where $\sigma^2 = \frac{1}{p-1}\sum_{i=1}^{p} \left\|T_p - \overline{T}\right\|_\infty^2$ and $\overline{T} = \frac{1}{p}\sum_{i=1}^{p} T_i$ (for sake of simplicity, $\|T\|_\infty$ denotes $\left\|\mathcal{L}_W^{-1}(T)\right\|_\infty$). This means that the algorithm finishes when the maximum approximation's error is smaller than $\tau\%$ of the standard deviation. In our applications, we usually fix this sparsity parameter at $\tau = 5\%$.

Theorem 3.1. *Let γ be a vector field in W and $T = \mathcal{L}_W(\gamma)$. Let $\Pi_n = \sum_{i=1}^{n} \delta_{x_i}^{\alpha_i^n}$ be the current constructed after n steps of the algorithm 1. Then,*

- $\|\Pi_n(T) - T\|_{W^*} \xrightarrow[n\to\infty]{} 0$

- $\|\Pi_n(T) - T\|_\infty = \left\|\mathcal{L}_W^{-1}(\Pi_n(T)) - \mathcal{L}_W^{-1}(T)\right\|_\infty \xrightarrow[n\to\infty]{} 0$

Proof. For the sake of simplicity, we denote by $E_n = W^*_{\mathcal{X}_n}$ the finite-dimensional space of currents spanned by the iteratively estimated points (x_i). The estimated current Π_n is defined as the W-orthogonal projection onto E_n. We denote R_n the residue so that $\mathcal{L}_W(\gamma) = \Pi_n + R_n$. The corresponding residual error in W is denoted by $\gamma_n = \mathcal{L}_W^{-1}(R_n)$. K denotes here a generic kernel, such that $u^t K(x,x)u \geq c|u|^2$ for a given constant $c > 0$ and for any $x \in \mathbb{R}^3$ and $u \in \mathbb{R}^p$. This includes the translation-invariant definite positive kernels, since then $K(x,x) = K(0) > 0$.

Convergence for the L^∞ norm

At every iteration, the point x_{n+1} is chosen so that: $|\gamma_n(x_{n+1})| = \|\gamma_n\|_\infty$. For any $\alpha \in \mathbb{R}^p$, since $\Pi_n + \delta^\alpha_{x_{n+1}} \in E_{n+1}$, we have:

$$\|R_{n+1}\|^2_{W^*} \leq \left\|\Pi_n + \delta^\alpha_{x_{n+1}} - T\right\|^2_{W^*} \tag{3.2.2}$$

The right-hand side of this equation is equal to: $\|R_n\|^2_{W^*} + \alpha^t K(x_{n+1}, x_{n+1})\alpha - 2\gamma_n(x_{n+1})^t\alpha$. Minimizing the right-hand side of Eq.(3.2.2) with respect to α leads to $\alpha_* = K(x_{n+1}, x_{n+1})^{-1}\gamma_n(x_{n+1})$ for which

$$\begin{aligned}\|R_{n+1}\|^2_{W^*} &\leq \|R_n\|^2_{W^*} - \gamma_n(x_{n+1})^t K(x_{n+1}, x_{n+1})\gamma_n(x_{n+1}) \\ &\leq \|R_n\|^2_{W^*} - c^2\|\gamma_n\|^2_\infty\end{aligned} \tag{3.2.3}$$

Therefore the sequence $\|R_n\|_{W^*}$ is monotonically decreasing and hence converges. We deduce also from this equation that $c\sum_{k=0}^n \|\gamma_k\|^2_\infty \leq \|T\|^2_{W^*} - \|R_{n+1}\|^2_{W^*} \leq \|T\|^2_{W^*}$. This series is therefore convergent and $\|\gamma_n\|_\infty$ converges to 0 as n tends to infinity.

Convergence for the W-norm

We introduce F_n such that $E_{n+1} = E_n \overset{\perp}{\oplus} F_n$ and p_n the W-orthogonal projection of T on F_n. As a consequence, each pair of distinct currents p_k are orthogonal and $\Pi_n = \sum_{k=0}^{n-1} p_k$. Therefore $\|\Pi_n\|^2_{W^*} = \sum_{k=0}^{n-1}\|p_k\|^2_{W^*} \leq \|T\|^2_{W^*}$. This shows that the series converges. Now, we can show that the sequence R_n satisfies the Cauchy condition. Indeed, for $n \geq m$, we have:

$$\|R_n - R_m\|^2_{W^*} = \|\Pi_n - \Pi_m\|^2_{W^*} = \sum_{k=m}^{n-1}\|p_k\|^2_{W^*} \leq \sum_{k=m}^{\infty}\|p_k\|^2_{W^*}$$

This Cauchy sequence in the Hilbert space W^* converges therefore to a current R for the W^*-norm.

For all points x and vectors α, we have:

$$\begin{aligned}|\langle R, \delta^\alpha_x\rangle_{W^*}| &\leq |\langle R - R_n, \delta^\alpha_x\rangle_{W^*}| + |\langle R_n, \delta^\alpha_x\rangle_{W^*}| \\ &\leq Cte\,\|R - R_n\|_{W^*} + \left|\gamma_n(x)^t\alpha\right|\end{aligned} \tag{3.2.4}$$

Since we proved that $\|\gamma_n\|_\infty$ converge to zero as n tends to infinity, so does the right-hand side of this equation. This shows that R is orthogonal to every δ^α_x, a dense family of vectors in W^*. This proves that $R = 0$. ■

Remark 3.2. Equation (3.2.3) states that: $\|R_{n+1}\|^2_{W^*} \leq \|R_n\|^2_{W^*} - c^2\|\gamma_n\|^2_\infty$ and therefore tells us how much the residue decays at each step. This must be useful to characterize

the speed of convergence of the algorithm. Assume that there is a constant β such that $\|\gamma_n\|_\infty \geq \beta \|\gamma_n\|_W$. Then, Eq. (3.2.3) would give $\|R_{n+1}\|^2_{W^*} \leq (1 - c^2\beta^2) \|R_n\|^2_{W^*}$ and:

$$\|R_n\|_{W^*} \leq (1 - c^2\beta^2)^{n/2} \|T\|_{W^*}, \tag{3.2.5}$$

which would prove the convergence of the algorithm at least at an exponential speed.

Unfortunately, the inequality $\|\gamma\|_\infty \geq \beta \|\gamma\|_W$ is not satisfied for all $\gamma \in W$. Since W is a RKHS, one already has: $\|\gamma\|_\infty \leq C_W \|\gamma\|_W$ (see Eq. (1.5.4)). The equivalence between both norm is false in general. However, it becomes true for finite-dimensional spaces. In the next section, we will give precisely a discrete version of the algorithm using grids to define finite-dimensional approximations of the space of currents. In this case, we will prove the exponential speed of convergence of the algorithm (see Proposition 3.6) and discuss the behavior of this convergence speed when the approximation space tend to the infinite-dimensional space of currents. □

Remark 3.3. For computational reasons, one would like to find the point x_{n+1} such that $|\gamma_n x_{n+1}| \geq C_\infty \|\gamma_n\|_\infty$ for some constant $C_\infty \leq 1$, instead of the point that reaches exactly the maximum. In this case, Eq. (3.2.3) becomes $\|R_{n+1}\|^2_{W^*} \leq \|R_n\|^2_{W^*} - c^2 C^2_\infty \|\gamma_n\|^2_\infty$ and the same proof shows the convergence of the algorithm. This may help, for instance, to speed up the search for the maximum by limiting this search on a finite set of finely distributed points. □

Remark 3.4. In this algorithm, we control the approximation error with the infinity norm. However, by definition of the RKHS, there is a constant C_W such that for all $\gamma \in W$, $\|\gamma\|_\infty \leq C_W \|\gamma\|_W$. This shows that the speed of convergence is faster with the infinity norm than with the W-norm. Therefore, controlling the error with the W-norm would result in approximation with a larger number of estimated momenta. However, as mentioned in Section 1.5.2.3, a current T tends to zero with respect to the L^∞ norm if $\langle T, U\rangle_{W^*}$ tends to zero for any test current U. More precisely, given a test current U with a finite number of momenta, the numerical error between $\langle T, U\rangle_{W^*}$ and $\left\langle \tilde{T}, U \right\rangle_{W^*}$ is given by:

$$\left| \left\langle T - \tilde{T}, U \right\rangle_{W^*} \right| \leq M(U) \left\| T - \tilde{T} \right\|_\infty, \tag{3.2.6}$$

where $M(U)$ is the mass-norm of U.

As it appears from Section 2.5.1 and especially Remark 2.29, it seems that L^∞-norm is the good norm to measure the approximation of currents. □

3.2.2 Fast computations in a discrete setting

From a computational point of view, the two most expensive steps of the algorithm are the search of the maximum of the residual vector field γ_n over the *entire* space (on line 4 of Algorithm 1) and the computation of this residual vector field at *every* point of the space (on line 6 of Algorithm 1). To give a more efficient version of this algorithm, we use the approximation spaces introduced in Chapter 2 via linearly spaced grids. First, we impose the estimated momenta to be located at the grid nodes. Second, we sample the dense vector fields on the grid. This allows us to use discrete convolution and FFT to compute efficiently the sampling of the residual vector field on the grid nodes.

3.2.2.1 Discrete algorithm

In this section, we suppose that the kernel K is translation-invariant. The input vector field γ belongs to W and, as such, tends to zero at infinity and is band-limited (this band being determined by the typical decay rate of the Fourier transform of the kernel, λ_W for the Gaussian kernels). Then, we set a grid Λ with periodic boundaries conditions, which is large enough so that the norm of $\gamma(x)$ outside the grid is negligible compared to the required precision of the approximation and whose step Δ is small compared to the spatial period of the kernel: $\Delta \ll \lambda_W$ (in practice, we choose $\Delta/\lambda_W = 1/5$). Then, we sample the input vector field γ on the grid nodes and store it as an image of vectors $\boldsymbol{\gamma}$. As seen in Chapter 2, $\boldsymbol{\gamma}$ parametrizes γ^Λ: the orthogonal projection of γ on the approximation space W_Λ.

In the discrete version of the matching pursuit, we will impose that the estimated momenta are located at the grid nodes (i.e. $x_i \in \Lambda$). This means that the estimated current is forced to belong to discrete set of currents W_Λ^* (defined as $\mathcal{L}_W(W_\Lambda)$ as in Chapter 2). The residual vector field is the difference between γ^Λ and the vector field associated to the estimated current: this residual still belongs to W_Λ and therefore can be stored as an image of vectors. Therefore the residual remains in the space W_Λ along the iterations and can be stored and handled as an image of vectors.

Eventually, we will show that this algorithm build a current in W_Λ which tends to $\mathcal{L}_W^{-1}(\gamma^\Lambda)$: the orthogonal projection of $\mathcal{L}_W(\gamma)$ on W_Λ^*.

Computations in this framework are particularly easy and efficient. The estimated point x_{n+1} is the grid node where the the residual γ_n has the largest magnitude: $x_{n+1} = \arg\max_{x\in\Lambda} |\gamma_n(x)|$ (we denote $\|\gamma_n\|_{\infty,\Lambda}$ this supremum). This is done by searching the maximum magnitude in the image of vectors $\boldsymbol{\gamma}_n$. The coefficients α_i^n are solution of the same set of $3n$ linear equations as in Algorithm 1. Once these coefficients have been computed, we build the image of vectors $\boldsymbol{\alpha}$ which contains the momenta $(x_i, \alpha_i^n)_{1\leq i\leq n}$. Then, the residual vector field γ_n is computed as the image of vectors $\boldsymbol{\gamma}_n = \mathbf{K}_\Lambda \boldsymbol{\alpha}$. Along the lines of Section 2.2.3, this matrix multiplication can be performed as a cost $N_{\text{grid}} \log(N_{\text{grid}})$ using discrete convolutions and FFTs. Eventually, the algorithm may be written using only images of vectors and basic linear algebra, as summarized in Algorithm 2.

3.2.2.2 Speed of convergence

By construction, $\gamma_n(x_i) = 0$ for all estimated points $(x_i)_{i<n}$ (i.e. the residue R_n is orthogonal to the $W^*_{\mathcal{X}_{n-1}}$). This shows that the estimated points are all distinct, as long as $\boldsymbol{\gamma}_n \neq 0$. Therefore, the algorithm finishes at most when the total number of points in the grid is reached. However, in the cases of practical interest, much fewer points are needed to achieved a good approximation of the current. Indeed, the following proposition shows that the convergence of the residuals has an exponential speed.

Lemma 3.5. *Let Λ be a finite subset of $\mathbb{R}^3$ and W_Λ be the finite-dimensional subspace of the RKHS W as defined in Chapter 2. We assume that the kernel K^W is symmetric and*

Algorithm 2 Orthogonal Matching Pursuit for Currents (discrete version)

Input: an image of vectors $\boldsymbol{\gamma}$, a threshold $\eta > 0$, a linearly-spaced grid Λ
Store the image of vectors $\mathbf{k}$ ($\mathbf{k}(i) = K^W(x_0^\Lambda, x_i^\Lambda)$ for all $i \in \Lambda$) and its FFT.
$\boldsymbol{\gamma}_0 = \boldsymbol{\gamma}$, $n = 0$
while $\max_{i\in\Lambda} |\boldsymbol{\gamma}_n(i)| \geq \eta$ **do**
  $i_{n+1} \leftarrow \arg\max_{i\in\Lambda} |\boldsymbol{\gamma}_n(i)|$, $x_{n+1} = x^\Lambda_{i_{n+1}}$
  Find $(\alpha_i^{n+1})_{1\leq i\leq n+1}$ by solving $\sum_{p=1}^{n+1} \left(K^W(x^\Lambda_{i_q}, x^\Lambda_{i_p})\alpha_p^{n+1}\right)_k = \gamma(i_q)_k$
  Image of vectors $\boldsymbol{\alpha} = 0$
  for $k = 1$ to $n+1$ **do**
    $\boldsymbol{\alpha}(i_k) \leftarrow \boldsymbol{\alpha}(i_k) + \alpha_k^{n+1}$
  end for
  $\boldsymbol{\gamma}_{n+1} = \boldsymbol{\gamma} - \mathbf{k} \star \boldsymbol{\alpha}$ (computed by FFT's)
  $n \leftarrow n+1$
end while
Output: list of $(x_i, \alpha_i^n)_{i=1\ldots n}$

positive definite. Let d_Λ be the constant:

$$d_\Lambda = \sqrt{\frac{\lambda_{min}}{N}}, \tag{3.2.7}$$

where λ_{min} is the smallest eigenvalue of the block-matrix $\mathbf{K}_\Lambda$ whose block (i,j) is given by $K^W(x_i, x_j)$ for $x_i, x_j \in \Lambda$.

Then, for all $\gamma \in W_\Lambda$ we have the control:

$$\|\gamma\|_{\infty,\Lambda} \geq d_\Lambda \|\gamma\|_W \tag{3.2.8}$$

Proof. Due to the conditions on the kernel, $\mathbf{K}_\Lambda$ is a finite positive real matrix. In particular it is invertible and it is diagonal in an orthonormal basis with positive eigenvalues.

Let γ be a vector field in W_Λ. As shown in Section 2.2, γ is characterized by N momenta (x_i^Λ, α_i) located at the grid nodes, encoded as a Np-dimensional vector: $\boldsymbol{\alpha}$ (N denotes the number of points of Λ). In particular, $\|\gamma\|_W^2 = \boldsymbol{\alpha}^t \mathbf{K}\boldsymbol{\alpha} = \left|\mathbf{K}^{1/2}\boldsymbol{\alpha}\right|_2^2$ and $\|\gamma\|_{\infty,\Lambda} = |\mathbf{K}\boldsymbol{\alpha}|_\infty$.

Then, standard computations of the matrix norms show that:

$$\left|\mathbf{K}_\Lambda^{1/2}\boldsymbol{\alpha}\right|_2 \leq \underbrace{\left|\mathbf{K}_\Lambda^{-1/2}\right|_2}_{=\lambda_{\min}^{-1/2}} \underbrace{|\mathbf{K}_\Lambda\boldsymbol{\alpha}|_2}_{\leq\sqrt{N}|\mathbf{K}_\Lambda\alpha|_\infty} \leq \sqrt{\frac{N}{\lambda_{\min}}}\,|\mathbf{K}_\Lambda\boldsymbol{\alpha}|_\infty\,, \tag{3.2.9}$$

where $\lambda_{\min} > 0$ denotes the smallest eigenvalues of the matrix $\mathbf{K}_\Lambda$. ■

Proposition 3.6. *Let Λ be a finite subset of $\mathbb{R}^3$ with N points and $\gamma \in W_\Lambda$. Let γ_n be the residual vector field (in W_Λ) after $n \leq N$ steps of the algorithm 2 applied to the image of vectors $\boldsymbol{\gamma}$, the sampling of γ on the grid. We assume that the kernel K^W is symmetric, definite positive and translation invariant. Then there is a constant C such that*

$$\|\gamma_n\|_W \leq C^{n/2}\|\gamma\|_W \tag{3.2.10}$$

Proof. We follow the sketch of the proof of the convergence of the continuous version of the algorithm. The (n+1)*th* estimated point x_{n+1} is such that $|\gamma_n(x_{n+1})| = \|\gamma_n\|_{\infty,\Lambda}$. We denote $E_n = W_{\mathcal{X}_n} = \text{Span}\{K(x_i,.)\alpha; 1 \leq i \leq n, \alpha \in \mathbb{R}^p\}$. For any $\alpha \in \mathbb{R}^p$, $\gamma_n + K^W(x_{n+1},.)\alpha \in E_{n+1}$ and therefore $\|\gamma_{n+1}\|_W^2 \leq \left\|\gamma_n + K^W(x_{n+1},.)\alpha - \gamma\right\|_W^2$. The minimization of the right-hand side of this equation with respect to α leads to:

$$\|\gamma_{n+1}\|_W^2 \leq \|\gamma_n\|_W^2 - K^W(0)^2 \|\gamma_n\|_{\infty,\Lambda}^2 \tag{3.2.11}$$

Thanks to the lemma 3.5, $\|\gamma_n\|_{\infty,\Lambda}^2 \geq d_\Lambda^2 \|\gamma_n\|_W^2$. This shows that: $\|\gamma_{n+1}\|_W^2 \leq (1 - K^W(0)^2 d_\Lambda^2) \|\gamma_n\|_W^2$. Eventually,

$$\|\gamma_n\|_W \leq \left(1 - K^W(0)^2 d_\Lambda^2\right)^{n/2} \|\gamma\|_W \tag{3.2.12}$$

■

Remark 3.7. If Λ is a linearly spaced grid with periodic boundary conditions, then the matrix $\mathbf{K}$ is circulant and their eigenvalues can be found easily as the Fourier coefficient of its first row $\mathbf{k}$. As we showed in Section 2.2.4 for a Gaussian kernel with standard deviation λ_W, the smallest eigenvalue of this matrix $\lambda_{\min}$ decays exponentially, such that:

$$d_\Lambda^2 = \lambda_{\min}/N \sim C' \frac{\lambda_W^3}{V} \frac{\Delta^{3/2}}{V^{1/2}} \exp\left(-C \frac{\lambda_W^2}{\Delta^2} \left(\frac{N-1}{N}\right)^{2/3}\right),$$

where Δ is the grid step, C, C' a two numerical constants and $V = N\Delta^3$ is the volume delimited by the convex hull of the grid.

This shows that d_Λ^2 tend to zero as the grid step tend to zero ($\Delta \to 0$ with $V = N\Delta^3$ constant). As long as W_Λ converges to W, the speed of convergence decreases: $(1 - K^W(0)^2 d_\Lambda^2) \to 1$. At the limit, we end up with $\|\gamma_n\|_W \leq \|\gamma\|_W$ which does not give any information about the speed of convergence. In this case, however, we proved that the algorithm still converges (see Theorem 3.1). It is likely that the convergence speed in infinite dimension would be less than exponential.

This remark is valid for any kernels whose Fourier transform tends to zero at infinity. In this case, d_Λ tends to zero faster than the speed of decay of the Fourier transform of the kernel. □

Remark 3.8. The previous remark is counterintuitive. One expects that, with more basis elements, it would possible to find a basis vector which has a larger correlation with the residual. This would increase the decay of the residual and make the convergence faster. This argument is true. The problem is that the the convergence speed in Eq. (3.2.10) is an upper-bound which applies *uniformly* for all n. This uniform upper-bound does not reflect what really happens in practice. During the first steps of the algorithm, the estimated momenta reveals the major patterns of the signal: the correlation between the Dirac delta current and the current is very high. Since each Dirac delta current is the best explanation of the signal in a neighborhood of size λ_W, the first estimated momenta are at a distance roughly of λ_W one from the others. This process goes on until the current is covered by patches of size λ_W. This convergence is very fast and in practice much faster than the exponential decay of Eq. 3.2.10. During these first steps, the more basis vectors

(i.e. the finer the grid), the faster the decay of the residual. The problem appears once the current is covered by patches of size λ_W. Then, the residual contains very small and local variations which, in some sense, can be considered as noise (this local errors could be compared to the errors introduced by the sampling for instance). To correct for these local errors, the algorithm starts estimating some momenta in-between the patches. This re-introduces redundancy in the signal decomposition, something we precisely want to avoid! And the the finer the grid, the more redundancy the algorithm can re-introduce, the slower the convergence. This second convergence speed is the limiting factor which makes the upper-bound in Eq. 3.2.10 to degenerate at the limit $\Delta \to 0$.

These two different behaviors of the algorithm are clearly visible in the graph of figure 3.9. The left-hand part of the graph corresponds to the first steps of the algorithm with a very fast rate of decay of the residual (the curves are drawn from right to left as long as the algorithm runs). The upper part of the graph corresponds to the second behavior when the algorithm starts re-introducing redundancy in the signal decomposition. Our empirical value of $\tau = 5\%$ seems to be a good cut-off to keep only the most salient patterns of the current. □

Remark 3.9. During the iterations, we compute the momenta α_i by computing the orthogonal projection of γ on the finite-dimensional space spanned by the functions $K(.,x_i)\varepsilon_k$ for the set of estimated positions $\mathcal{X}_n = \{x_i\}_{1\leq i\leq n}$. This requires to solve a linear system by inverting at step n, the np-by-np block matrix whose block (i,j) is $K^W(x_i,x_j)$, namely inverting the matrix $\mathbf{K}_{\mathcal{X}_n}$ with the notations of Chapter 2. As we saw in Section 2.2.4, this matrix has a bad conditioning as soon as two estimated points are at a distance much smaller than λ_W. As discussed in the previous remark, this cannot occur during the first iterations since estimated points tend first to move apart one from the others. In practice, a bad-conditioning of this matrix has never been observed, even if we run the algorithm far below the threshold $\tau = 5\%$, like for plotting the curves of figure 3.9. As we shall see in the Section 3.5, even random Gaussian currents can be efficiently approximated by the matching pursuit algorithm without facing bad conditioning issues. If a bad conditioning of the matrix $\mathbf{K}_{\mathcal{X}_n}$ occurs anyway, this would mean that the matching pursuit fails for such a signal, the purpose of the matching pursuit being precisely to give a robust estimation of $\mathbf{K}_\Lambda^{-1}\gamma$ for a fine grid Λ by computing $\mathbf{K}_{\mathcal{X}_n}^{-1}$ for a good choice of point set $\mathcal{X}_n$. In case of failure, alternative approaches may be investigated like conjugate gradient for instance. □

3.3 A new tool for processing currents

3.3.1 A robust implementation of the deconvolution problem

In this section, we consider the matching pursuit in the perspective of the computational framework set up in Chapter 2. In Section 2.5, we showed that any standard operations on currents can be approximated efficiently using discrete spaces based on linearly-spaced grids. We also outlined that the implementation of a robust deconvolution scheme were missing in this framework. Indeed, there was no (numerically stable) way to map vector fields either in discrete or in continuous spaces back into the spaces of currents. The

matching pursuit introduced in this chapter can be seen as this missing robust deconvolution scheme, as illustrated in Fig. 3.1. Indeed, the matching pursuit estimates a solution of the deconvolution problem with an increasing number of momenta. Using a infinite number of momenta will completely solve the deconvolution problem, whereas the first momenta gives only an approximation of this solution, which is far enough in practice as we shall see in the next section.

The continuous version of the matching pursuit implements the map $\mathcal{L}_W$. The discrete version of the matching pursuit implements the multiplication by the matrix $\mathbf{K}_\Lambda^{-1}$ or equivalently the inverse of the convolution by the matrix $\mathbf{k}$ with the notations of Chapter 2. Thanks to these two algorithms, there is now a numerically stable way to map any of continuous or discrete spaces of currents or vector fields to any other of these spaces. These new possibilities will be used in Chapter 5 for the estimation of a template from a collection of shape. To implement this estimation, we will introduce the adjoint push-forward action which is defined only in the space of vector fields. Then, we will need to map the result of this operation back to the space of currents.

3.3.2 Three representations of a current

This completed computational framework offers a new representation for currents. The natural representation of a current is given as a list of oriented points called momenta resulting usually from the approximation of meshes or polygonal lines as a finite sum of Dirac delta currents. Therefore, a current is naturally encoded as a list of momenta in continuous coordinates. The usual operations in the space of currents: addition, scaling, subtraction or diffeomorphic deformations of currents preserve this structure of list. We denote N_T the total number of momenta in the decomposition of T. As discussed in Section 3.1, this representation of the current may be very redundant at the scale λ_W. This may cause a problem of memory space to store the current, of computational time to process the current or to interpret the current, especially when the current is the empirical mean or modes of a set of currents.

The computational framework illustrated in Fig. 3.1 leads to three solutions to this problem:

- We can project the current into a linearly spaced grid, as explained in Chapter 2. As the grid step tend to zero, this defines a collection of approximations which converges to the true current. This representation has N_{grid} terms and is stored as an image of vectors.

- We can compute the vector field associated to the current and apply the continuous version of the matching pursuit to give a sparse representation of the current in continuous coordinates. This sparse representation converges to the true current as the number of momenta increases. However, this representation is difficult to compute in practice. Indeed, there is no simple way to deal with a dense vector field and the implementation of the continuous version of the matching pursuit raises computational issues as discussed in the previous section.

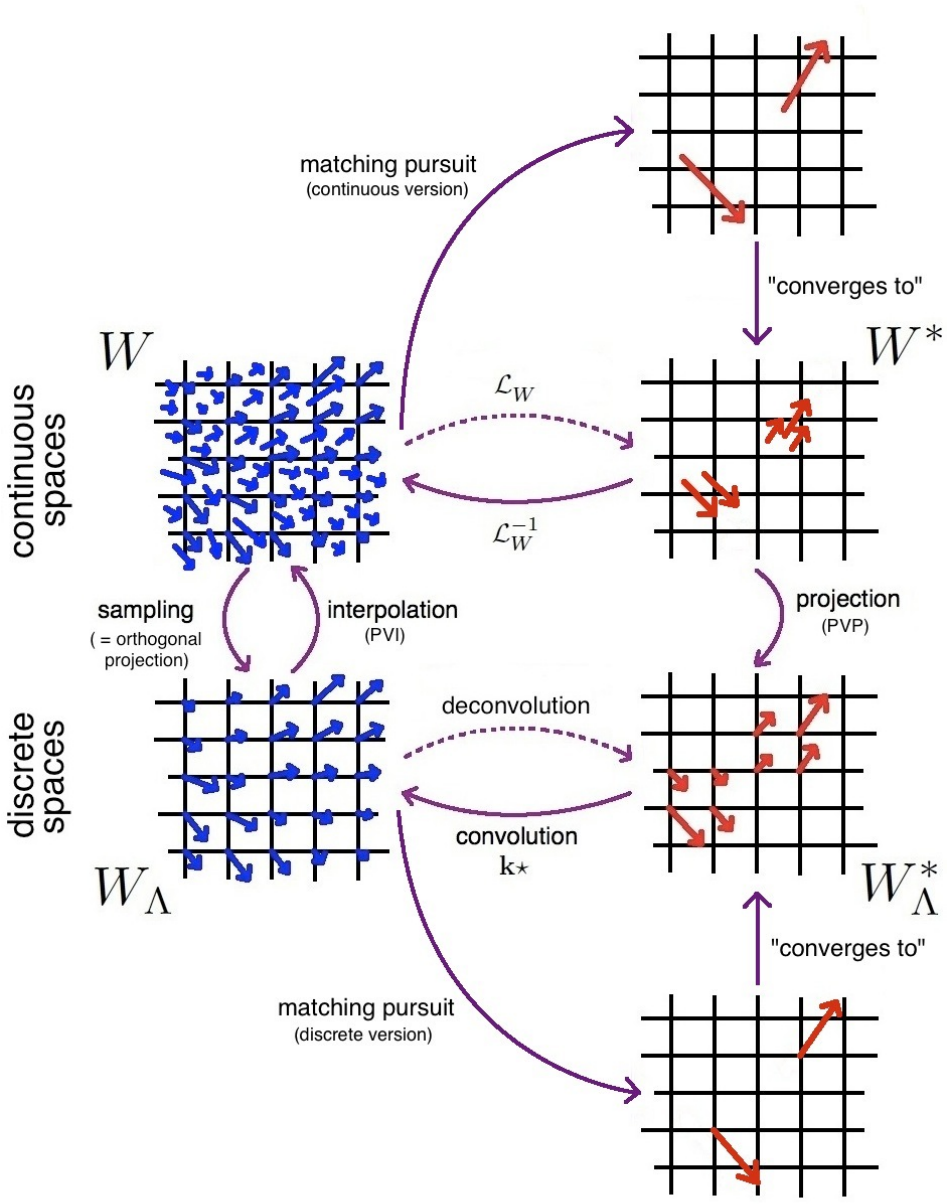


Figure 3.1: The matching pursuit implements a robust approximation of the deconvolution problem to map the spaces of vector fields to the space of currents (in the continuous or discrete domain). This completes the computational framework of Chapter 2 and leads to 4 representation of currents (right column). However, only the 3 representations at the bottom are used in practice.

- To workaround the previous issue, we can project the current on a linearly spaced grid, compute the vector field associated to the projected current via a convolution in the discrete spaces and use the discrete version of the matching pursuit to give a

sparse representation of the current in the discrete space. This sparse representation does not converge to the initial current but to the projection of the initial current on the discrete space of currents. It is stored as a list of momenta (the locations of the momenta can be stored as indices of grid nodes)

In addition to the original representation, this leads to four different representations of a current, whereas only three have a practical interest. Depending on the initial number of momenta, their redundancy at a given scale λ_W and their spreading in space, these three representations may vary drastically in size. The representation with grids has the advantage to have the right form to perform operations on currents in the discrete approximation spaces. One can apply directly convolution and matrix multiplication to compute the standard operations on currents. It is adapted for being used routinely in intensive computations. Usually, the sparse representation has fewer momenta than the number of grid nodes, which is a good representation when memory allocation for large grids makes the algorithm to slow down eventually. Moreover, the desired numerical precision completely determines the grid step and therefore the size of the grid-based representation. By contrast, there is no simple way to predict the size of the sparse representation for a fixed numerical precision.

Note that the sparse representation converges to the projection of the initial current on the discrete space of currents determined by the grid. The difference between the initial projection and its projection is of order Δ^2/λ_W^2 using the partial volume projection of Chapter 2. The total approximation error between the sparse representation computed by matching pursuit and the initial current is therefore the sum of the error of the matching pursuit (set as 5% of the variance of the set of currents) and the projection error of order Δ^2/λ_W^2. The first error is usually much greater than the second one. Therefore, the application of the discrete version of the matching pursuit algorithm leads to an approximation of the same order as 5% of the variance of the set of currents. The following experiments show that this approximation is a good trade-off between precision and compactness.

3.3.3 In which case using the matching pursuit?

In Chapter 2, we discussed how the approximation spaces could be used to implement standard operations on currents like the norm of a current for instance. We compared three different methods in Section 2.5.1: the exact computation, the projection of every currents in the approximation space or using an approximation of the vector field associated to a current. Now, one could imagine to use the matching pursuit to give a new implementation of such operations.

The sparse approximation of a current minimizes the number of momenta of its representation for a fixed numerical precision. Therefore, using this sparse representation instead of the original ones speeds up the projection of the current on the approximation spaces, as well as the interpolation of the image of vectors (in the discrete space of vector fields) at the location of the momenta (in approach ♯3 in Section 2.5.1). If the currents have a large number of momenta concentrated in a small domain of the space, then the space and time complexity of the computations can be determined by this number of momenta (see Table 2.5.1.3). In this case precisely, one can expect that the matching pursuit enables to

achieved a very high compression ratio of the currents (usually above 90% in the practical cases, as we will show in the next section). Therefore, using the sparse representation instead of the original representation of the currents can make a real difference.

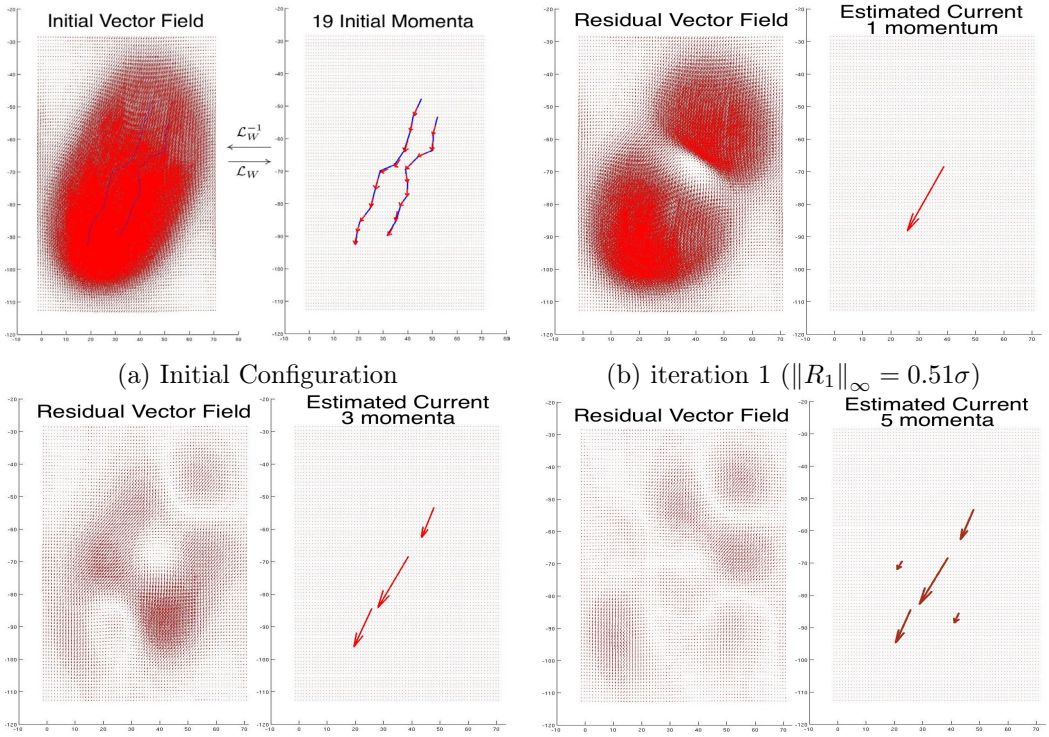
This conclusion, however, has to be discussed carefully. First, if the number of momenta decreases drastically in the sparse representation, it is likely that the exact computation with the double sums may the fastest solution, since no convolution will be required at a cost $N_{\text{grid}} \log(N_{\text{grid}})$ which may be larger than the complexity of the exact computation of order N_{mom^2}. Second, one must be aware that the computation of the sparse approximation takes much more time than any convolution. Using it *routinely* does not means necessarily an improvement in terms of processing time. By contrast, using it as a pre-processing can help to increase to computation time of the algorithms.

For instance, there are three cases where the computation of the matching pursuit leads to computational improvements. First, if we plan to perform intensive computations with a given set of currents, we may compute their sparse representation once for all before running the computations, like running a bench of statistical tests for instance. Second, if we want to register two sets of shapes which have a fine spatial resolution compared to the precision of the matching we are looking for, computing the sparse representation as a preprocessing enables to increase drastically the speed of the registration. Third, the currents may be given directly in a sparse form. This is the case for the template computed in Chapter 5 which has necessarily a sparse representation as a result of the estimation algorithm. In the other cases, one must be aware that the primary goal of the sparse representation is not to perform routinely operations on currents but rather to implement the deconvolution problem robustly and to compress currents for statistical purposes and for registration. In general, the matching pursuit should be used as a pre-processing to speed up the algorithms or as a post-processing to store and interpret the results.

3.4 Application to shape compression and interpretation of statistics

3.4.1 Matching Pursuit on a simulated example

We show here how the matching pursuit algorithm helps to approximate statistics on currents, and to give an interpretable representation of such statistics. We choose the mean of two 2D curves as a simulated example. In Fig. 3.2, we show the initial mean in the space of currents and the first and third iterations of the algorithm, which builds iteratively an approximation of the mean. The approximation error tends to zero as the algorithm goes on. In Fig. 3.3 we show how the number of momenta needed to represent the mean of the two curves varies with respect to the standard deviation of the kernel (λ_W) for the a fixed approximation error. The greater λ_W, the closer the two curves in the space of currents (i.e. their differences become small perturbations at the scale λ_W, the smaller the number of Delta Dirac currents needed to represent the mean for the same accuracy (i.e. the faster the terms of the series decrease).



(a) Initial Configuration

(b) iteration 1 ($\|R_1\|_\infty = 0.51\sigma$)

(c) iteration 3 ($\|R_3\|_\infty = 0.11\sigma$)

(d) iteration 5 ($\|R_5\|_\infty = 0.06\sigma$)

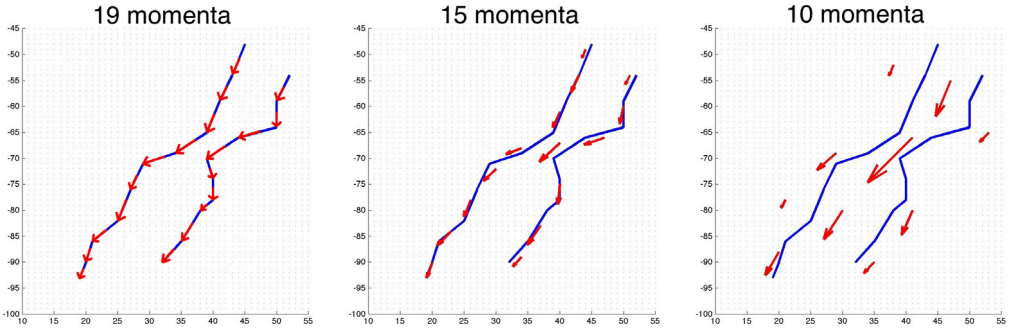
Figure 3.2: A sparse deconvolution scheme for currents: (a) the initial configuration. right: two curves in blue and their mean in red: the collection of all tangents scaled by 0.5, seen as momenta in the space of currents W^*. left: the Gaussian convolution ($\mathcal{L}_W^{-1}$) of the initial momenta gives the dual representation of the mean as a dense vector field in W ($\lambda_W = 15$). (b) (resp. (c) and (d)): first (resp. third and fifth) iteration of the matching pursuit algorithm: estimated momenta on the right panel, residual vector field on the left panel (what remains to be explained by the forthcoming momenta). The momenta converge to the true solution while the residual vector field tends to zero. $\|R_n\|_\infty$ denotes the norm of the residual vector field at step n and σ the standard deviation: $\|L - L'\|_\infty / \sqrt{2}$

We compute also the difference between both lines ($L_1 - L_2$) and use the matching pursuit to approximate this difference. The result is shown in Fig. 3.4.

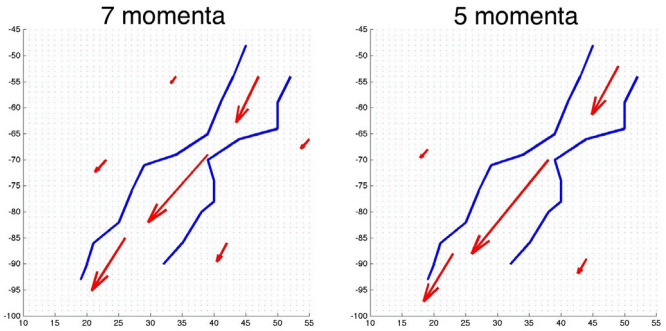
3.4.2 Compressed mean and modes of anatomical data

The matching pursuit enables to approximate easily statistics of currents like mean and principal modes, since such statistics are naturally defined as linear combinations of the input currents.

We assume that we have a collection of N discrete shapes (set of curves or surfaces): $T_1, \dots T_N$ which are modeled as currents. As shown in Chapter 1, such currents can be approximated by the union of their tangents or normals: $T_i = \sum_p \delta_{x_p^i}^{\alpha_p^i}$ where x_p^i is the



(a) Initial mean current (b) $\lambda_W = 5$ (c) $\lambda_W = 10$



(d) $\lambda_W = 15$ (e) $\lambda_W = 20$

Figure 3.3: Impact of λ_W. (a)- Two curves in blue and their mean in the space of currents in red. (b) to (d) the approximation of the mean of two curves for an increasing size of the kernel λ_W and a fixed precision $\tau = 5\%$ (defined as the ratio between the residual norm and the standard deviation $\|L - L'\|_\infty / \sqrt{2}$). The greater λ_W, the closer the two curves in the space of currents, the more redundant the initial momenta at the scale λ_W, the sparser the estimated decomposition.

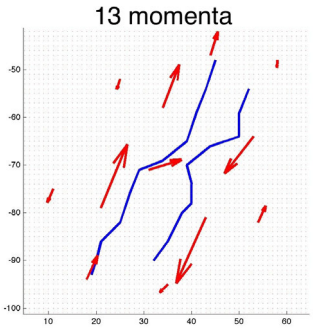


Figure 3.4: Difference between two lines computed in the space W and then approximated by a matching pursuit algorithm ($\lambda_W = 10$ and $\tau = 5\%$)

center of each segment (resp. mesh cell) of the curve (resp. surface) T_i, and α_p^i its segments (resp. its normals). Since the space of currents W^* is a vector space, one may compute directly the empirical mean as $\overline{T} = \sum_i T_i/N = \sum_i \sum_p \delta_{x_p^i}^{\alpha_p^i/N}$. This is simply the union of all the tangents (resp. normals) in the database, scaled by $1/N$.

Since the space of currents is provided with an inner-product, one may compute the

N-by-N empirical covariance matrix Γ: $\Gamma_{ij} = \left\langle T_i - \overline{T}, T_j - \overline{T} \right\rangle_{W^*}$ (computed in practice thanks to the inner-product between two Dirac delta currents in Eq. (1.2.8)). Let $V^1, \ldots, V^N$ be the eigenvectors of Γ. The n*th* principal mode of the population of currents is therefore given by: $\overline{T} + \sum_i V_i^n (T_i - \overline{T})$. Expanding this expression leads to a double sum of the form: $\sum_i \sum_p \delta_{x_p^i}^{w_p^{i,n} \alpha_p^i}$ for some weights $w_p^{i,n}$. Section 5.4.1 will detail the computations of statistics of currents.

These expressions of the empirical mean and modes are exact and can be used directly to give quantitative measures of the variability of the population. However, they are all given as a weighted sum of all the input Dirac delta currents. They have as many terms as the total number of tangents and normals within the database. This number increases as the number of subjects in the database increase, while at the same time the estimated mean and modes converge to their true values. This number increases also as the sampling of the shapes becomes finer, while the discrete shapes converge in the space of currents to continuous geometrical objects. This representation is therefore not stable from a computational point of view. If one wants to deform the mean to a new subject (like for an atlas to subject registration for instance), this representation may cause dramatic computational issues. Moreover, from a visualization point of view, the representation of the mean and principal modes as the union of scaled tangents and normals is particularly difficult to interpret.

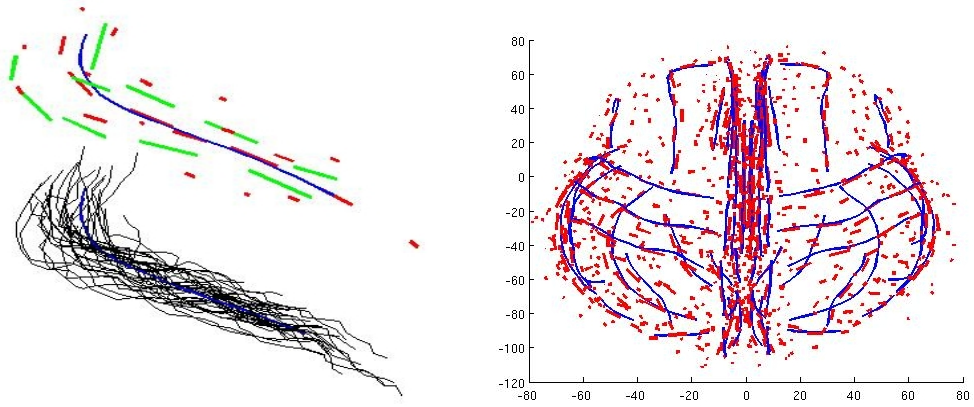
Although the representation of the mean and principal modes is exact, it is often far from being optimal. It may be highly redundant at the scale λ_W, especially if the shapes are at a distance of order λ_W one from the others. This is particularly visible in its representation in terms of vector field $\gamma = \mathcal{L}_W^{-1}(\overline{T})$. The vector field γ results from the convolution of every tangent or normal with the Gaussian kernel K^W with standard deviation λ_W. Two tangents or normals closer than λ_W contribute to the same "mode" of the Gaussian distribution γ. This vector field γ integrates precisely the redundancy of the mean current (or principal modes) at the scale λ_W (See Fig. 3.2-a).

Using the matching pursuit algorithm as explained in 3.3 allows us to give a sparse representation of the mean and modes at any desired accuracy. We illustrate this approach on a set of sulcal lines and a set of meshes of subcortical structures of the brain.

3.4.2.1 Statistics on sulcal lines

The sulci are the fissures on the brain surface and they are often used to measure anatomical differences between subjects [Thompson 1996a]. We perform here statistics on a set of 70 sulci delineated in $N_{obs} = 34$ subjects. These data were provided by Paul Thompson (Laboratory of NeuroImaging, University of California, Los Angeles) as part of the collaborative project `Brain-Atlas`. Before computing statistics, the set of curves of each subject were aligned together using affine transformation.

For each sulcal line, we approximate the mean current $\overline{L} = \frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} L_i$ for the scale of kernel $\lambda_W = 12$mm and sparsity parameter $\tau = 5\%$. Results are shown in Fig. 3.5 for the Sylvian Fissure of the right hemisphere and for all 70 sulci. The initial number of momenta for the mean fissure was $N_T = 899$ (i.e. the number of segments of all lines) whereas the final approximation needs only $N_{mom} = 54$ momenta. In this case, the compression ratio



Sylvian Fissure (right hemisphere) All 70 sulci (top view of the brain)

Figure 3.5: Statistics for 70 sulci in 34 subjects ($\lambda_W = 12\text{mm}, \tau = 5\%$). Left: The Sylvian fissure of the 34 subjects (black) with the sample mean (red) and first eigenmode at $+\sigma$ (green) of this set of curves. The first mode shows mostly the spreading of the set of lines. Right: Mean currents (red) compared to the mean lines (blue) computed from B-spline parameterization of curves [Fillard 2007c]. Results are visually in good agreement.

is of 94%. Considering all sulci, the compression rate is on average: $94.8\% \pm 0.02$. The grid has a step $\Delta = 2\text{mm}$ and typically $N_{grid} = 10^5$ points. Our mean is visually in good agreement with other mean curves computing from B-spline representation [Fillard 2007c].

Then, we compute the eigenmodes of the lines sets. We find the eigenvectors (V_k) of the $N_{obs} \times N_{obs}$ matrix $\Sigma = \left(\left\langle L_i - \overline{L}, L_j - \overline{L} \right\rangle_{W^*}\right)_{i,j=1\ldots N_{obs}}$. The k^{th} eigenmode is given by the linear combination of input currents: $m_k = \overline{L} \pm \sum_i (V_k)_i (L_i - \overline{L})$ (see Section 5.4.1 and Remark 5.9 for more details on PCA in the space of currents). We approximate the first eigenmode at $+\sigma$ of the Sylvian Fissure of the right hemisphere (Fig. 3.5-top): this mode captures mainly the spreading of the lines set around its mean.

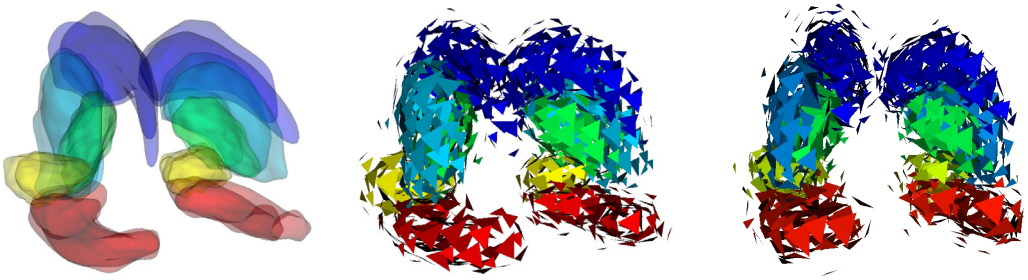
3.4.2.2 Statistics on surfaces of brain structures

Ten deep brain structures were segmented in a population of 50 autistics patients and 7 controls (Caudate, Putamen, Globus Pallidus, Amygdala and Hippocampus for each hemisphere) [Hazlett 2005]. As a result, we have 57 sets of 10 meshes. These meshes have been registered rigidly. For each structure, we approximate the mean of the autistic and the control shapes via the matching pursuit algorithm (see Fig. 3.6 and 3.7). Note that for surfaces, we represent the estimated momenta (normal of an infinitesimal mesh cell) as equilateral triangles whose normals is the momenta. The difference between the mean of autistic shapes and the mean of control shapes is still a current that we approximate: the arrows of Fig. 3.8 are the 10 first estimated momenta of this difference, suggesting that the autistic mean is more curved at the Hippocampus' extremity and thicker in the middle.

Even if the approximations do not look like a surface, they can be more easily interpreted than the collection of all normals of the 50 (or 7) instances. In some sense, this

representation of the estimated means are "optimal" since they try to minimize the number of momenta for a fixed approximation error. The compression ratio between the total number of normals in the database N_T and the number of estimated momenta N_{mom} for the 10 structures is on average of $99.96\% \pm 10^{-4}$ in the autistic group. The grid has a step $\Delta = 1$mm and for one structure we have the following typical values: $N_T = 50 \times 3000 = 1.5 \star 10^5$, $N_{grid} = 3 \star 10^5$ and $N_{mom} = 100$. Fig. 3.9 shows that the quality of approximation remains good until very high compression ratio.

This sparse representation of the mean current is particularly useful for further computations like the registration of this mean to the set of structures of one subject or the comparison with the mean of a population of control subjects. For instance, the deformation of a mean obtained from 3 shapes as in Fig .3.6a, which was previously taking 10 hours, is now taking about 5 minutes (using the same code as in [Vaillant 2005, Durrleman 2007]). For the full set of 50 instances, representing the mean requires 1.2 Kb in our framework, versus 8 Mb originally. Deforming the former still requires 5 minutes while it is not feasible to deform the latter without high performance computing.



(a) Structures of 2 autistics (b) Mean of 50 autistics (c) Mean of 7 controls

Figure 3.6: Sparse approximation of the mean current for 10 meshes segmented in 50 autistics patients (b) and 7 control subjects (c) with $\lambda_W = 5$mm, $\tau = 5\%$ (the diameter of the data is 60mm).

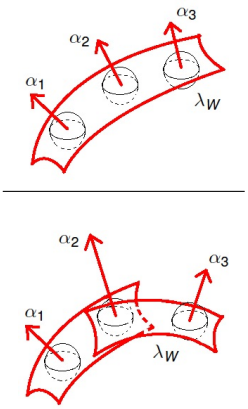


Figure 3.7: The matching pursuit applied on surfaces essentially divides the space into patches of size λ_W^3 and approximate the surface enclosed in the patch by one momentum. The magnitude of the momentum encodes the total surface area within the patch (taking orientation into account: two pieces of surfaces with opposite orientation would cancel out). These estimated momenta are represented by equilateral triangles in Fig. 3.6.

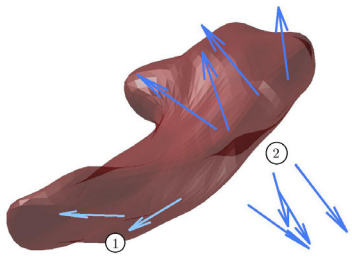


Figure 3.8: The arrows represent the difference between the mean of autistics and the mean of controls (shown in Fig. 3.6. They are superimposed with the Hippocampus of a control. This shows that the mean from autistics is more curved at hippocampus' extremity (area 1) and thicker in area 2.

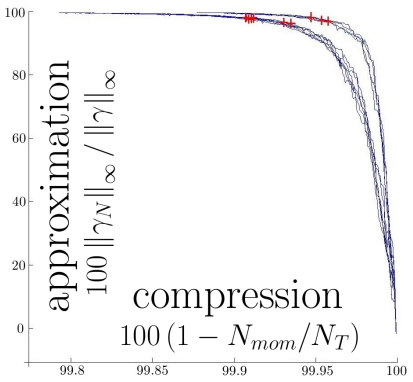


Figure 3.9: This graph shows the evolution of the approximation error when the number of momenta is increasing during the application of matching pursuit to approximate the mean of the 50 autistics meshes (see Fig. 3.6). Very high compression rate can be achieved while the approximation error remains small. Red points correspond to the approximation error equal to $\tau = 5\%$ of the variance of the structures. This empirical value seems to be a good threshold under which some significant patterns of the signal have been left aside and above which the addition of new momenta does not help to give a better approximation of the signal (see remark 3.8).

3.4.2.3 Compression of white matter fiber bundles

The extraction and analysis of the white matter fiber bundles which connect two different functional area of the cortex is of great interest in Computational Anatomy (see Chapter 7 for more details). Each extracted fiber bundle consists of several pieces of curves. As shown in Fig. 3.10, the density of the fibers within the bundle may be very high.

We can model each bundle as a single current, which adds the contribution of every single fibers, provided that the fibers have been consistently oriented (otherwise the contribution of a fiber with inverse orientation will be removed from the bundle). The size of the Gaussian kernel λ_W determines the spatial scale under which geometrical variations are considered as noise. Since the density of the fibers in a section of size λ_W of the bundle is very high, the bundle modeled as current may be efficiently approximated using the matching pursuit algorithm. All fibers that go through a local section of size λ_W contribute to the same mode of the dual vector field associated to the bundle. They can be replaced efficiently by a single momenta whose length account for the local redundancy of the fibers in the section (see Fig. 3.11). The sparse approximation of the bundle is shown in Fig.3.10.

In case of fiber bundles, the structure of vector space has a simple interpretation. Addition or subtraction of Delta Dirac currents means increasing or decreasing the local density

of fibers (by adding or removing infinitesimal element of fibers). Decreasing the local density is achieved by adding some momenta with the direction opposite to the one of the bundle. During the application of the matching pursuit algorithm, the subtraction is integrated into the computation of the vector field and the sparse representation of the current really looks like a bundle with smaller density. This gives eventually an interpretable interpretation of the statistics on such fiber bundles. This will be also illustrated in Chapter 7 (in particular in Fig. 7.8)

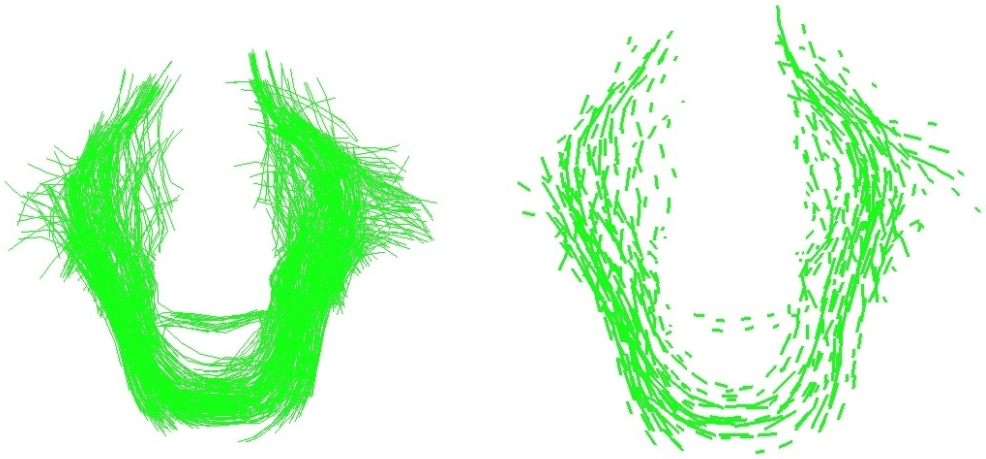


Figure 3.10: A white matter fiber bundle (top) made of several individual pieces of curves. Its sparse representation (bottom) integrates the local redundancy of the information at a scale $\lambda_W = 3$mm (the diameter of the data is 100mm). The length of the estimated momenta encodes precisely this local redundancy. For visualization purpose, these momenta are scaled by 0.1.

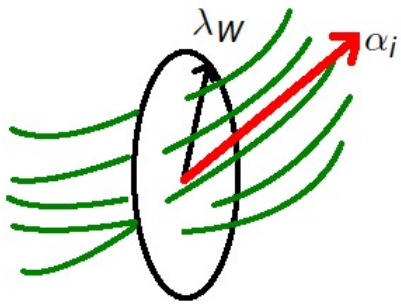


Figure 3.11: The matching pursuit applied on set of curves (like the fiber bundles in Fig 3.10) essentially approximates the curves which go through a patch of size λ_W by one momentum. The momentum encodes the sum the tangents (taking orientation into account: two curves which go through the patch in opposite direction would cancel out). If the curves represent wires which send information at a constant rate, the momentum represents the total information which goes through the patch.

3.5 Application to deconvolution: simulation of noise of currents

3.5.1 Discrete Approximations with grids of increasing resolution

In Chapter 2, we introduced grids Λ to approximate the *infinite* dimensional space of currents W^* with *finite* dimensional subspaces W^*_Λ. In Section 1.5.3, we showed that the projections of general random Gaussian currents on such subspaces admit a probability density function. More precisely, on a linearly spaced grid Λ, a random Gaussian currents have the form: $T = \sum_{x_i \in \Lambda} \delta_{x_i}^{\alpha_i}$, which is parametrized by the Np-dimensional vector $\boldsymbol{\alpha}$ (i.e. the concatenation of every α_i). In order to reproduce the Gaussian law on currents, this vector must be simulated according to a Gaussian distribution in $\mathbb{R}^{Np}$ proportional to $\exp(-\|T\|^2_{W^*}/2) = \exp(-\boldsymbol{\alpha}^t \mathbf{K}_\Lambda \boldsymbol{\alpha}/2)$, where the Np-by-Np matrix $\mathbf{K}_\Lambda$ is the metric on $\mathbb{R}^{Np}$ induced by the kernel K^W (see Section 2.2.2). This leads to:

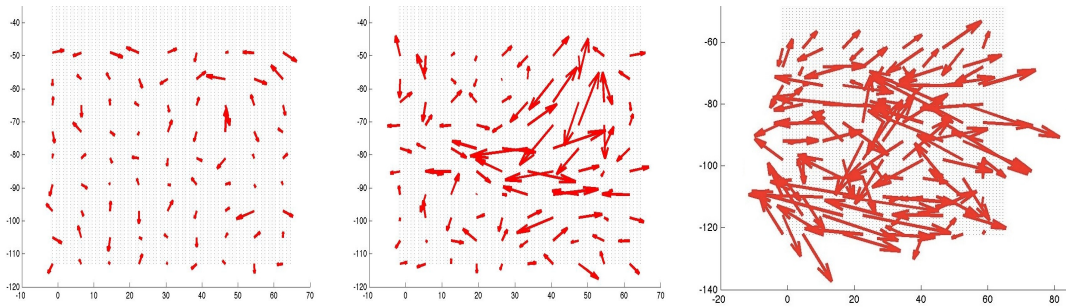
$$p(\boldsymbol{\alpha}) = \frac{1}{(2\pi)^{Np/2} |\mathbf{K}_\Lambda|^{1/2}} \exp\left(-\frac{\boldsymbol{\alpha}^t \mathbf{K}_\Lambda \boldsymbol{\alpha}}{2}\right) \tag{3.5.1}$$

In Section 2.2.4, we computed the spectrum of the matrix $\mathbf{K}_\Lambda$. This helps us to better understand what the noise en currents really is. The covariance matrix of the momenta in Eq. (3.5.1) is given by $\mathbf{K}_\Lambda^{-1}$. Proposition 2.9 shows that this matrix is diagonal in a Fourier basis with eigenvalues (we suppose here the kernel is Gaussian, but any kernel with light tails would lead to similar results) which behave like:

$$\sqrt{\frac{V}{\Delta^3}} \exp\left(C' \frac{\lambda_W^2}{\Delta^2} \left(\frac{i^2}{n_x^2} + \frac{j^2}{n_y^2} + \frac{k^2}{n_z^2}\right)\right), \tag{3.5.2}$$

for $C' > 0$ a positive numerical constant. $\mathbf{K}_\Lambda^{-1}$ is therefore a high-pass filter. This is not surprising since multiplying with this matrix is equivalent to applying the differential operator $\mathcal{L}_W$ on these finite dimensional subspaces of currents. This shows that the smaller Δ ($\Delta^3 \ll V$ and $\Delta \ll \lambda_W$), the closer the finite dimensional approximation to the "true" Gaussian current, the more concentrated the spectrum at the highest frequencies. As illustrated by the simulations in Fig. 3.12, the smaller Δ, the more oscillatory patterns in the simulations of the current.

From a numerical point of view, these simulations are ill-posed. This prevents us from visualizing the "true" Gaussian current by projecting it on grids with finer and finer resolution. Indeed, the conditioning number of the matrix $\mathbf{K}_\Lambda$ is the ratio between its largest and smallest eigenvalue. This number is equal to $\hat{k}(0,0,0)/\hat{k}(n_x-1, n_y-1, n_z-1) \sim \exp(C'\lambda_W^2/\Delta^2)$. This number growths exponentially as the ratio $(\Delta/\lambda_W)^2$ tends to zero, making impossible numerical simulations of the noise directly in the space of currents. Actually, with $\lambda_W = 10$ as in Fig. 3.12, the value $\Delta = 6$ is close to the limit under which the inversion of the spectrum of $\mathbf{K}_\Lambda$ is not more possible numerically.



Grid step $\Delta = 8$ Grid step $\Delta = 7$ Grid step $\Delta = 6$

Figure 3.12: The oscillatory nature of the noise in the space of currents. We simulate here a Gaussian noise directly in the space of currents W^* (momenta α simulated via Eq. (3.5.1)), while the resolution of the grid is increasing (i.e. while the finite dimensional sub-space gets closer to the space of currents). This simulation shows that the momenta oscillate at higher and higher frequencies while the points get closer and closer. The size of the kernel used in this experiment is $\lambda_W = 10$

3.5.1.1 Using Matching Pursuit as a robust deconvolution scheme

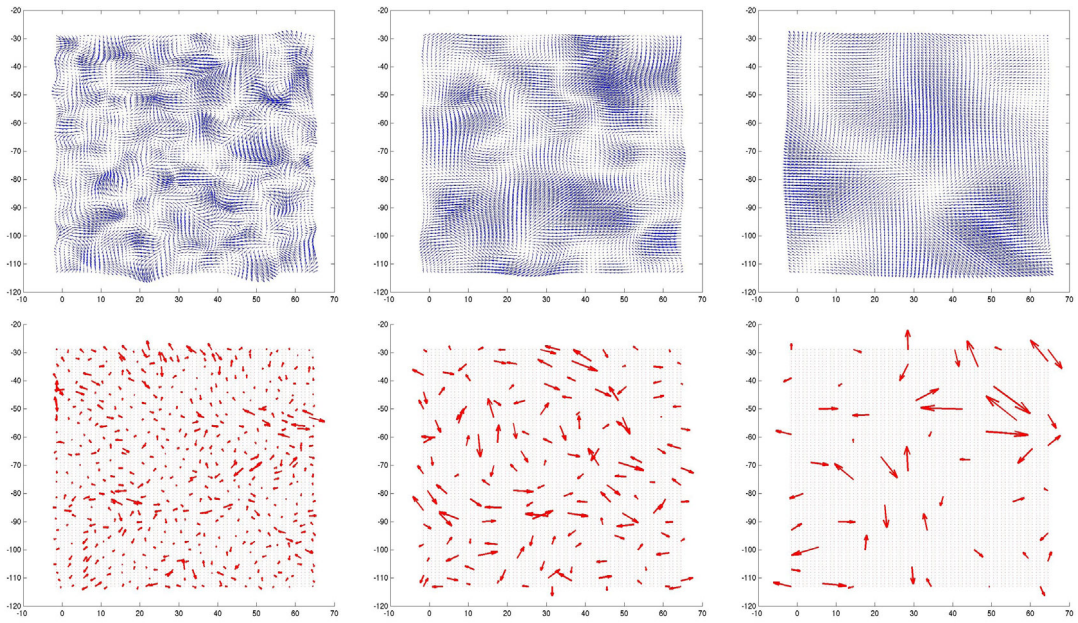
Simulated directly in the space of currents, the Gaussian noise looks like a distribution with oscillatory patterns at a possibly infinite frequency. This approach does not help to interpret the nature of this Gaussian noise. By contrast, the dual representation of this noise in the space of vector field W, which integrates (by a convolution) every oscillatory patterns, presents a clearly interpretable structure. This structure was also present, but invisible, in the direct simulation of the currents.

Indeed, we can simulate the variables $\boldsymbol{\gamma} = \mathbf{K}_\Lambda \boldsymbol{\alpha}$ instead of the momenta $\boldsymbol{\alpha}$. A change of variable within Eq. (3.5.1) shows that Gaussian image of vectors $\boldsymbol{\gamma}$ follows the distribution:

$$p(\boldsymbol{\gamma}) = \frac{|\mathbf{K}_\Lambda|^{1/2}}{(2\pi)^{Np/2}} \exp\left(-\frac{\boldsymbol{\gamma}^t \mathbf{K}_\Lambda^{-1} \boldsymbol{\gamma}}{2}\right) \tag{3.5.3}$$

Now, the covariance matrix of $\boldsymbol{\gamma}$ is $\mathbf{K}_\Lambda$ which is a good conditioned low-pass filter. The spectrum of the covariance matrix decreases like a Gaussian function with variance $1/\lambda_W^2$. Two vectors at a distance smaller than λ_W are correlated. The greater λ_W, the fewer high frequencies, the more regular the vector field γ. Now, it is possible to approximate the true underlying Gaussian vector field by simulating its sampling $\boldsymbol{\gamma}$ on a grid with a decreasing step.

Once the noise is simulated in the space of vector fields W, one has to find a way to get back this noise in the space of currents. Of course, the linear equation $\boldsymbol{\gamma} = \mathbf{K}_\Lambda \boldsymbol{\alpha}$ cannot be solved by inverting the badly conditioned matrix $\mathbf{K}_\Lambda$. Instead, we apply the matching pursuit to the simulated vector field. This gives an approximation of the solution of the linear equation at any desired accuracy. The result is therefore an approximation of the noise in the space of currents which corresponds to the simulated vector field. In Fig. 3.13, one sees that the sparse representation of the noise highlights clearly the underlying structure



$\lambda_W = 5$ $\lambda_W = 10$ $\lambda_W = 20$

Figure 3.13: Simulation of random Gaussian vector field (top row) and their approximation at 95% (bottom row). The matching pursuit is used here as a deconvolution tool to visualize currents when only their representation in terms of vector fields (in W) can be simulated. Whereas the noise in the space of currents is a very irregular and oscillatory object by nature (See Fig. 3.12), its analysis via the matching pursuit (which exhibits an adapted basis on which to decompose the current) reveals a much more intuitive structure.

of the Gaussian noise in the space of currents: momenta are drawn at a distance of roughly λ_W with arbitrary directions.

Remark 3.10. To simulate random Gaussian image of vectors in practice, we do not need to factorize the N-by-N matrix $\mathbf{K}_\Lambda$ (for a scalar kernel, otherwise $\mathbf{K}_\Lambda$ is Np-by-Np), with $N \geq 10^3$. We use the fact that for linearly spaced grids with periodic boundary conditions, the multiplication with $\mathbf{K}_\Lambda$ is equivalent to a discrete convolution with $\mathbf{k}$. Therefore, we simulate the stationary Gaussian noise by computing the convolution between a transfer function ρ and an image of white noise (both of size N). More precisely, let g be a n_x-by-n_y-by-n_z matrix of independent zero-mean Gaussian variables with unit variance. We look for a matrix ρ, such that the covariance of the image $\rho \star g$ is given by $\mathbf{k}$. Since g is

second-order stationary random process, so is $\rho \star g$. Then ρ must satisfy:

$$\begin{aligned}
&\mathbf{Cov}(\rho \star g(i,j,k), \rho \star g(0,0,0)) \\
&\quad = \mathbf{E}\left(\sum_{i',j',k'} \rho(i-i', j-j', k-k') g(i',j',k') \sum_{p',q',r'} \rho(-p',-q',-r') g(p',q',r') \right) \\
&\quad = \sum_{i',j',k'} \rho(i-i', j-j', k-k') \rho(i',j',k') \\
&\quad = \rho \star \check{\rho}(i,j,k) = \mathbf{k}(i,j,k)
\end{aligned} \tag{3.5.4}$$

where $\check{\rho}_{i,j,k} = \rho_{n_x-i, n_y-j, n_z-k}$. A ρ which satisfies the last equation exists. Indeed, in the Fourier domain, this condition becomes: $\hat{\mathbf{k}} = |\hat{\rho}|^2$. Since $\mathbf{k}$ is symmetric (due to the boundary conditions: $\check{\mathbf{k}} = \mathbf{k}$), $\hat{\mathbf{k}}$ is a real matrix. Moreover, since K^W is a positive kernel, then for any finite set of momenta (α_i) and points x_i: $\sum_{i,j} \alpha_i^t K^W(x_i, x_j) \alpha_j \geq 0$ (See definition B.5). Using the grid nodes as points x_i, this proves that $\sum_{i,j,k} \boldsymbol{\alpha}(i,j,k) \mathbf{k} \star \boldsymbol{\alpha}(i,j,k) \geq 0$. Applying this equation with $\boldsymbol{\alpha}$ the eigenvectors of $\mathbf{K}_\Lambda$ associated to the eigenvalues $\hat{\mathbf{k}}(i,j,k)$ (see Section 3.5.1) leads to: $\hat{\mathbf{k}}_{i,j,k} \geq 0$. Eventually, the FFT of the kernel $\mathbf{k}$ is real and non-negative. The inverse FFT of its square root satisfies the required conditions for the transfer function ρ. □

3.5.2 Discussion: currents versus shapes

The usual geometrical objects like curves or surfaces are *embedded* in the space of currents, thus meaning that the space of currents contains these objects but also objects of a very different kind, like a combination of Dirac delta currents which do not approximate a curve or a surface for instance. In this section, we discuss the structure of a set of "acceptable shapes" (still to be defined!) in the space of currents.

3.5.2.1 Currents are not shapes

The geometrical objects, once modeled as currents, differ from our intuitive idea of a geometrical shape. The modeling based on currents is blind to the topology of the shapes, such as the connectivity between consecutive points on curves or the meshing of a surface, for instance. A shape modeled as currents is divided into a set of unconnected momenta (oriented points). This may be an advantage for some applications like for the fiber bundles in Chapter 7 for instance, for which the only reliable information is the local orientation of the fibers and not the long-range interactions between points.

However, this lack of topological modeling can make the interpretation of currents in terms of surfaces or curves more difficult. In Fig. 3.6, a mean current is presented as a set of small triangles. There is no guarantee that this set of small triangles is the representation of a smooth surface in the space of currents. In other words, there is no guarantee that we can build a mesh which "interpolate" the triangles. If such a mesh exists, there is no obvious way to compute it. The difference between two surfaces also belong to the space of currents and can still be represented as a set of momenta (visualized as small triangles). In this case, there is no reason that this current comes from a mesh. Similarly, statistics

on currents like modes of variations can be represented as set of momenta thanks to the matching pursuit algorithm but may not be represented as curves or meshes. This lack of meshing information could be limitation to use statistics on currents to constrain the segmentation of anatomical structures in new images.

Furthermore, a current is a global object. Once oriented points have been gathered as a single current (as explained in Section 1.2.4), there is no way to make a point to play a particular role. This is precisely a feature that motivates this modeling: we want to avoid to introduce arbitrary point correspondences. The counterpart is that currents cannot be used to assess local properties of shapes. For instance, the norm on currents is a global dissimilarity measure which does not highlight where locally the differences occur. Nevertheless, the matching pursuit applied to the vector field of the difference between two currents can be used to highlight such local effects, as illustrated in Fig. 3.4 and Fig. 3.8 for instance.

3.5.2.2 The space of shapes is curved in the space of currents

As an embedding space, the space of currents is much larger than a set of "acceptable shapes". The previous experiments suggest that the space of "acceptable shapes" is curved in the space of currents. Indeed, when we performed linear combination of geometrical shapes in the space of currents, as in Section 3.4.2 for the mean and principal modes, we end up with currents which do not look like geometrical shapes. This is particularly visible for the mean meshes computed in Fig. 3.6. This tends to prove that the curvature of the space of "acceptable shapes" in the space of currents is not negligible. Mean and principal modes belong to the convex hull of the input shapes in the linear space of currents, which seems not to include (part of) the space of "acceptable shapes". This problem arises each time a curved manifold is embedded into a vector space. Intrinsic statistics, like in [Pennec 2006a] for instance, could be performed only if we manage to have a precise (and tractable!) definition of the space of "acceptable shapes". This curvature effect can be minimized by performing such linear computations on shapes which are close one to the others in the space of currents, for instance after shapes have been registered as in Chapter 5. In this case, the convex hull approximates the tangent-space of the space of "acceptable shapes", and the linear combination of shapes in the space of currents do not deviate too much from the space of "acceptable shapes". Nevertheless, we will show in our applications, and especially in Chapter 7, that statistics performed directly in the embedding space of currents may lead to relevant anatomical findings.

3.5.2.3 The space of shapes is of negligible measure

The simulations of noise in Section 3.5 tend to prove that the space of "acceptable shapes" has a negligible measure (in the sense of the Gaussian measure) in the space of currents. None of the simulated random currents look like a random curve of random surface. This is even more visible when we try to simulate noisy shapes in the space of currents. In Fig. 3.14, we add a Gaussian noise to a line L, with a very high signal to noise ratio. As a result, the noisy curve (in the space of current) is close to the initial line. However, the result does not look like a noisy curve: random Dirac delta currents were spread over the

entire domain whereas one expects that a noise added to the curve would move randomly the points of the curve. In this experiment, we explored the unit ball centered at the line L by simulating random currents in this ball (more precisely, in a ball of radius 1/SNR). This shows obviously that a reasonable set of curves (here we can think of all possible sets of 7 connected points with bounded angles for instance) is of null measure in the space of currents. In other words, the simulation of a Gaussian variation around the line L has no chance to look like the sampling of a real curve.

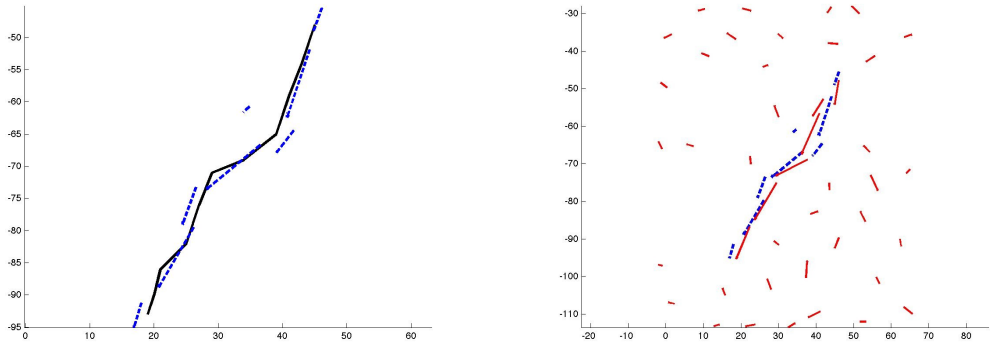


Figure 3.14: Noise ε with variance 1 added to a curve L: $\tilde{L} = L + \sigma\varepsilon$. First, the noise is simulated in the space W as a dense vector field which is added to the vector field $\mathcal{L}_W^{-1}(L)$. Then, the matching pursuit gives the set of segments shown here. The dissimilarity measure is equal to $\left\|\tilde{L} - L\right\|^2_{W^*} / \|L\|^2_{W^*} = 0.001$. This noisy current is about 10 times closer than the noisy curves shown in Fig. 3.15. However, this does not look like an expected random variation of a curve due to the non-localized nature of the noise of currents. Random variations around this line is unlikely to stay on the space of "acceptable shape". *(In this experiment $\lambda_W = 10$)*

3.5.2.4 The currents metric induced in the space of shapes

The space of "acceptable shapes" seems to be a curved manifold with null measure in the space of currents. This does not mean, however, that the norm of currents (which defines both the metric on the space of currents and the likelihood of random Gaussian current) is not adapted to measure shape variations. Actually, if we deal only with geometrical data, we use only the metric on the space of "acceptable shapes" which is induced from the metric in the space of currents. In the following experiments, we show that the restriction of the norm of currents to the space of real curves has good properties.

In Fig. 3.15, we move randomly the points of the curve L and measure the distance between the noisy curve and the original one in the space of currents. The graph of this figure shows that for a small level of noise, the metric on currents does not capture shape difference, thus highlighting its robustness to noise. This effects is due mostly to the smoothing effect of the kernel. The metric starts to be sensitive to the shape variations from a SNR of 0.02.

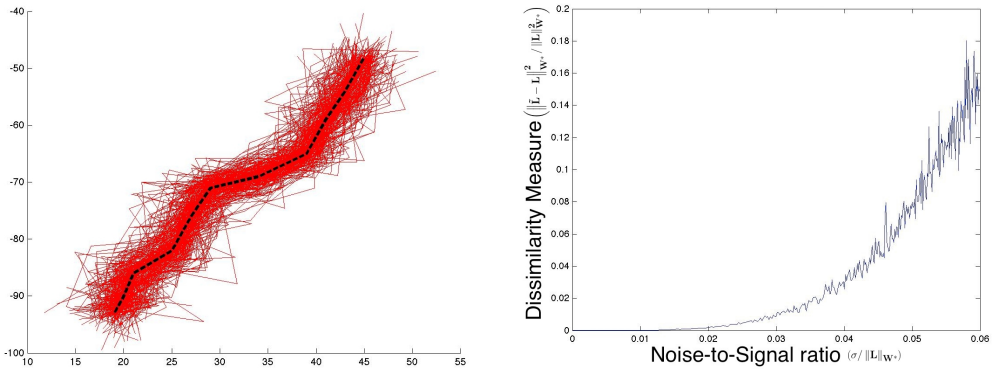


Figure 3.15: We build noisy curves $\tilde{L}$ from the curve L by moving the points of the line according to a Gaussian displacement (zero-mean, standard deviation σ) for $\sigma / \|L\|_{W^*} = 0, \ldots, 0.06$ (Left panel). On the right panel, we measure the discrepancy in the space of currents between noisy curves and the original one: $\left\|\tilde{L} - L\right\|_{W^*} / \|L\|_{W^*}$: all these lines belong to a ball of center L and radius $0.44 \|L\|_{W^*}$, whereas a random Gaussian variation in the space of currents is unlikely to produce such noisy curves. The flat aspect of the curve for high signal-to-noise ratio shows the robustness to noise of the metric between currents. *(In this experiment $\lambda_W = 10$)*

In Fig. 3.16, we translate the curve L along the x-axis and measure the distance between the translated curve and the original one in the space of currents. This graph emphasizes a typical range of sensitivity of the metric determined by the scale λ_W. Far below the scale λ_W, shape variations are considered as noise. When the lines are at a distance much greater than λ_W, they are orthogonal in the space of currents ($\langle L, \tau_* L \rangle_{W^*} \sim 0$ since $K^W(x, x+\tau) \sim 0$ when $|\tau| \gg \lambda_W$) and the distance between them becomes constant. More discussion on the effects of the size of the kernel λ_W to measure dissimilarity between lines can be found in Chapter 6.

These experiments show that metric on the space of currents is adapted to measure shape variations, even if these variations are unlikely to result from random variations in the space of currents. This metric can be used therefore 'as is' as long as we constrain the shapes to stay on the subspace of "acceptable shapes". This is the case for measuring the discrepancy between a diffeomorphic deformation of a curve and another curve as in Chapter 4. The deformation is driven by the discrepancy measure in the space of currents but moves the points without changing the topological structure of the curves (or surfaces). Therefore, it minimized the distance between source and target in the space of current, while enforcing the deformed curve not to deviate from the space of "acceptable curves".

3.5.2.5 Currents: a set of sensors to probe the geometry

The previous geometrical considerations help to better understand how shapes are modeled as currents. By contrast, they give few insights about what currents really are, espe-

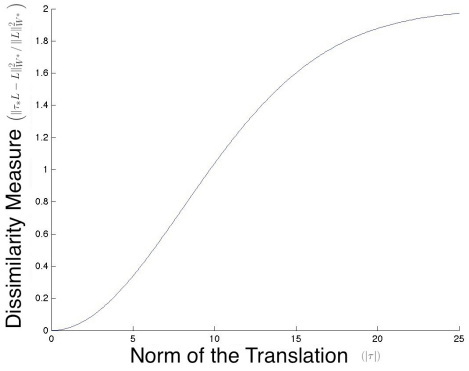


Figure 3.16: We translate the curve of Fig. 3.15 along the x-axis and measure the dissimilarity between translated curve and the original one. This dissimilarity metric is weakly sensitive to small variations with respect to λ_W ($\lambda_W = 10$ in this experiment). This is due to the smoothing effect of the kernel, which sets a typical "level of noise" on geometrical objects. For large variations with respect to λ_W, the two lines become orthogonal and the metric does not capture variations anymore. The band-width of the kernel λ_W determines a typical range of variations which are captured by the metric on currents.

cially when they are not modeling a particular shape.

One can imagine a current as a set of sensors (like a CCD captor) which patches the entire space, each sensor having a typical diameter of λ_W. Each local sensor averages the geometrical information in the neighborhood of size λ_W and the current combine nicely the information of every sensor. This combination of local measures explains why currents are insensitive to the topology of shapes and long-range interaction between points. The distance in the space of currents compares the response between each possible pair of sensors.

In this context, a noise on currents is interpreted as a noise on each sensor. This is particularly visible in Fig. 3.14: the noise adds a small segment in arbitrary direction in every patches of size λ_W, as if the response of each sensor has been corrupted by a random geometrical information.

This helps also to understand why the analysis based on currents cannot retrieve any geometrical details at a smaller scale than λ_W. For instance, the matching pursuit leads roughly to one momenta in each patch of size λ_W: it tries to estimate the response of each sensor which, once combined, retrieves the actual data. Estimating details beyond the scale of the sensor is impossible: numerically this leads to an ill-posed deconvolution problem.

This explains also why we need to set up grids whose steps are smaller than the scale λ_W and also why setting finer and finer grids does not help to really increase the precision of the sparse representation of a current.

Chapter 4

Flows of space deformations for the registration and the temporal regression of geometrical data

This chapter is based on the registration framework of currents of [Glaunès 2005, Glaunès 2008]. The extension of this registration scheme for the temporal regression of longitudinal data has been published in [Durrleman 2009d, Durrleman 2009b]. The application of the registration in lung imaging has been published in [Gorbunova 2009].

Contents

In the previous chapters, we introduced the currents for modeling shapes and defined numerical tools for processing shape statistics. However, in the perspective of Grenander's approach [Grenander 1998, Mumford 2002], we want to base the measure of shape dissimilarities also on the deformations which map one shape to another. One of the main purpose of the thesis is precisely to combine the statistics on shapes with the statistics on deformations. Before defining and estimating such statistical models, we need to define shape deformations in a way which is compatible with the modeling of shapes as currents.

Several deformation frameworks have been proposed in the field of medical imaging, mostly for the registration of images [Christensen 1996, Rueckert 1999, Ashburner 1999, Shen 2002, Avants 2004, Ashburner 2007, Yeo 2009a, Vercauteren 2009]. To use the deformations which result from registration in a statistical context, we would like that these deformations:

- *have a tangent-space representation*: large deformations such as diffeomorphisms usually do not build a vector space but a curved manifold instead. To compute statistics on such objects, it would be beneficial to use their tangent-space representation of the deformations in the perspective of [Vaillant 2004, Pennec 2006a] (as outlined in [Mumford 2007]): statistics are performed on the tangent-space and logarithm/exponential map are used to compute the tangent-space representation from the deformation and vice versa. For this purpose, the tangent-space should be provided with a metric which allows to easily compute statistics.

- *are geodesic*: the tangent-space representation can be used only if the deformations which result from registration are geodesic on the "manifold of deformations". Therefore, one must have shown that the registration algorithm converges to a geodesic deformation.

- *can be processed independently of a particular registration algorithm*: the deformation which results from registration should be parameterized in way which allows the composition or the inversion of deformations, even if these deformations result from the registration of different shapes. For instance, one should be able to use a given deformation to warp any shape in space. This excludes the registration frameworks which result only on a correspondence field between the voxels of the source and the target image.

- *are dense*: the anatomical curves and surfaces will be used to drive the registration of two different anatomies. The resulting deformation should be a dense deformation of the underlying biological material, which enables to align the anatomical landmarks. This excludes the registration frameworks which provide only the deformation of each shape individually without integrating these spatial constraints consistently into a dense $3D$ deformation.

Among other possible choices, it seems that the Large Diffeomorphic Deformation Metric Mapping (LDDMM) framework [Trouvé 1995, Trouvé 1998, Dupuis 1998] is particularly adapted for statistical purposes. First, is has been already shown that this

deformation framework could be used for the registration of shapes modeled as currents [Vaillant 2005, Glaunès 2008]. Second, it has been used also to define tangent-space statistics [Vaillant 2004], since the resulting deformations have been proved to be geodesic. Third, the metric on the tangent-space uses the mechanism of RKHS. This offers a way to use the computational framework defined for currents in Chapter 2 and 3 to deal efficiently with statistics on deformations. Fourth, computing with such deformations is particularly easy: the composition, the inversion and the application of the deformation to any shapes are all computed at the cost of the integration of an ODE.

The purpose of this chapter is therefore to introduce the LDDMM framework for the registration of currents in the perspective of [Vaillant 2005]. Applying this registration framework on anatomical data with more that 10^4 points requires a particular optimization scheme. The original implementation of J. Glaunès uses the Fast Gauss Transform. As mentioned in Chapter 2, setting the parameters of this optimization scheme can be done only manually. This prevents us from using this algorithm *routinely* to run many registrations, like a bench of template-to-subject registrations for instance. In this chapter, we will show how to use the computational framework defined in Chapter 2 to provide a more stable and more robust implementation of the registration algorithm.

Moreover, the LDDMM framework is based on flows of diffeomorphisms. As we will show in this chapter, this offers a natural way to extend the usual 3D registrations for the temporal regression of time-indexed shapes, a key feature for defining statistical model of shape evolution, as we shall see in Chapter 9. In this chapter, we will illustrate this temporal regression method on a set of fossil endocasts of bonobos which are associated to different dental ages.

Eventually, we will show how this registration framework integrates consistently different anatomical landmarks into a single deformation of the underlying 3D domain. We will use this algorithm to drive the registration of images of the lungs by the positions of the blood vessels and the surface of the lungs. This will be compared with a registration driven by the intensities of images. This integrative power of the method will be used also in Chapter 6 for the registration of a set of sulcal lines and in Chapter 7 for the registration of white matter fiber bundles.

4.1 Flows of diffeomorphisms for registration and temporal regression

4.1.1 Shape registration based on currents

The problem of shape registration may be formulated as the search of an "optimal" deformation (in a sense to be defined) which enables to minimize the dissimilarity between the deformed source shape and the target shape. In the previous chapters, we showed how geometrical data can be modeled as currents. The norm between currents provides a metric on the space of shapes. To take advantage of the properties of the currents, we need to define a registration scheme, which is compatible with the framework based on currents. This means that the deformation of geometrical data (such as curves or surfaces) modeled as

currents should remain a current and that the registration should minimize the discrepancy between the deformed source shape and the target measured in the space of currents.

The resulting registration framework will inherit therefore from the main properties of currents: the registration can be performed even in absence of point correspondence between structures. One can constrain the registration with correspondences at the anatomically relevant level (correspondence between points, curves or set of curves for instance), as emphasized in Section 1.2.4. This avoids to introduce arbitrary correspondence which may bias the computation of the shape alignment. Moreover, the registration will be robust to the change of topology of the shapes, such as curve interruption for instance.

4.1.2 Why diffeomorphisms?

Once the measure of dissimilarity has been given, we have to define the class of deformations to be used for the registration. Several choices are possible. The deformations with the fewest degrees of freedom are the rigid-body transformations (translation and rotation) and scaling. In 3D, the group of rigid-body transformations is of dimension 6. If we add the scaling, it becomes of dimension 7. These groups of linear transformations can be extended to the more general affine deformation group with 12 degrees of freedom. In these cases, optimizing a criterion over the whole space of possible transformations is particularly easy due to the small number of parameters to be optimized. However, these deformations are linear and, as such, may be unable to capture several local variations of shapes such as a torque of one part of the shape and an elongation of another part. Such variations, though, are likely to describe interesting anatomical features. Therefore, we must enlarge the space of possible deformations to capture relevant anatomical variations.

The diffeomorphisms are the non-linear extension of the invertible linear transformations (isomorphisms), which play a key role in differential geometry since they describe a local change of coordinates. A diffeomorphism is a smooth mapping of the space into itself, invertible with smooth inverse. As non-linear deformations, it is particularly well suited to capture local smooth variations. The use of such one-to-one deformations for shape registration assumes one-to-one correspondence between different anatomies. Of course, this assumption is not completely realistic from an anatomical point of view. Indeed, the inter-subject variability is likely to involve topological changes or matter creation/deletion which cannot be captured by diffeomorphisms. However, the purpose of registration, here, is not to find the true transformation (provided that this makes any sense!), but to fit a model with its own limitations which best describes the data. In this setting, the registration is a trade-off between the regularity of the deformation and the fidelity-to-data. If the difference between the sets of shapes involves non-diffeomorphic variations, then they will be captured by the residual shape that remains after registration (difference between the deformed source and the target). In Chapter 5, we will set up a statistical framework which account for both the diffeomorphic variations of shapes (captured by the registration) and the non-diffeomorphic variations (captured by the residual), so that we will not leave aside any information. If one does not want to use this registration framework for statistical purposes but just for aligning shapes (for segmentation purposes via atlas to subject registration for instance), then one may be aware that large misalignments may appear in area where

non-diffeomorphic variations occur.

Contrary to the linear transformations which have a finite dimensional parameterization, diffeomorphisms have an infinite number of degrees of freedom. This raises computational issues for dealing with such deformations. In particular, optimizing a registration criterion over the whole group of diffeomorphisms might not be possible. By contrast, we can define smaller group of diffeomorphisms, which still are of infinite dimension, but which allow to be processed via discrete parameterizations. The LDDMM framework is based on the group of diffeomorphisms set up in [Trouvé 1995, Dupuis 1998] which is constructed via integration of time-varying vector fields which belong to RKHS. As we shall see in this chapter, the registration of discrete structures (unstructured point sets, polygonal lines or surfaces) in this setting leads to a finite parameterization of the diffeomorphisms, which depends on the structure to be matched (see Section 4.2.2.2).

Affine transformations are particular cases of diffeomorphisms. In the framework presented here, we can adapt the regularity (i.e. the scale at which the motion of the points are correlated) of the diffeomorphisms from an locally rigid transformation (spatial scale tends to infinity) to irregular deformations which move every point almost independently (spatial scale tends to zero). This gives a very flexible way to adapt the regularity of the deformation to our needs.

4.1.3 Dense deformation for the integration of spatial constraints

Many shape registration frameworks look for a displacement field which is defined on the points of shapes only. Each shape to be matched leads to a correspondence field, without guaranteeing the spatial consistence of the different displacement fields if there are several geometrical primitives to be matched. However, in Computational Anatomy, the geometrical primitives to be matched have been extracted from images and are often some geometrical landmarks of specific structures or organs. Therefore, one would like to find the deformation of the underlying biological material which enables to align the extracted geometrical landmarks. What we are looking for is a dense deformation of the space which is constrained by the discrete shapes instead of the discrete deformation which cannot be extrapolated to the whole image domain. In the deformation framework presented here, one precisely looks for a diffeomorphism of the underlying 3D space which transports the shapes from the coordinate system of the source to the coordinate system of the target. The set of shapes to be registered are used as constraints to find the best diffeomorphism which will best align the two sets of shapes. Even if there are several geometrical primitives to be matched, the registration of two sets of geometrical data leads to a single deformation of the space. This allows, for instance, to use the resulting deformation to deform other geometrical structures or the underlying image itself.

However, it not always possible to find a diffeomorphic deformation which perfectly align all the spatial constraints, since the two sets of shapes might not be diffeomorphic (i.e. a different number of curves to be matched for instance) for instance. Therefore, it is important to ensure that the estimated deformation (returned by the registration algorithm) is really a diffeomorphism as the best fit of a diffeomorphic model to the data. This will decompose in a robust way the difference between the two sets of shapes into a

diffeomorphic variation and a non-diffeomorphic perturbation (in the residual which remain after registration).

The regularity of the deformation is a parameter of the method which determines the scale at which the diffeomorphism integrates the spatial constraints consistently. All pieces of shapes located in a neighborhood of the size of this parameter will be deformed consistently (i.e. locally an almost rigid deformation), whereas the pieces of shapes located at a distance much further may be deformed independently, while ensuring the diffeomorphic property though. This integrating property of the diffeomorphisms will be of great interest in Chapter 6 for defining the registration of the cortex surface knowing only the position of some sulcal constraints. In some sense, this is similar to the poly-affine framework proposed in [Arsigny 2005] which defines piecewise affine transformations in a way which ensures the resulting deformation to be a diffeomorphism. However, in the setting proposed here, we do not need to specify the partitioning of the space into small regions beforehand. We specify only the typical size of the regions on which the diffeomorphism is very regular.

4.1.4 Flows of diffeomorphisms: from registration to regression

In our framework, we build diffeomorphisms by integrating time-varying vector fields. As a consequence, we do not only define a diffeomorphism ϕ but also a differentiable flow of diffeomorphism ϕ_t where t is a continuous parameter within the interval $[0, T]$. The flow starts at time $t = 0$ with ϕ_0 being the identity mapping. It finishes at time $t = T$ where $\phi_{t=T} = \phi$ the desired transformation. This means that for any point x of the space, $\phi_t(x)$ is the path of this point which leads to the final position $\phi_T(x)$ which is in correspondence with x. This differs from other diffeomorphic frameworks which gives only the correspondence field $x \to \phi(x)$. As we shall see below, knowing the whole trajectory $\phi_t(x)$ for any point x allows us to easily compute operations on diffeomorphisms. To compute the inverse of the deformation for instance, it suffices to follow the path backward!

Moreover, flows of diffeomorphisms allow us to deal in a single framework with both the shape registration and the temporal shape regression. For the registration of a source shape S to a target shape S', we want to minimize the discrepancy between the deformed source $\phi_T(S)$ and the target S'. This is done by finding the flow of diffeomorphism $(\phi_t)_{0 \leq t \leq T}$ such that the final diffeomorphism ϕ_T minimize the fidelity-to-data term and such that the flow between $\phi_0 = \mathrm{Id}$ and ϕ_T is the shortest possible (i.e. geodesic in a sense to be defined). This framework extends directly for shape regression, as emphasized in Fig. 4.1. Let S_i be a set of shapes associated to some time t_i (or any other scalar parameter) and S_0 an initial shape. Then the regression of this data distributed over time can be expressed as a continuous shape evolution $S(t) = \phi_t(S_0)$ where ϕ_t is a flow of diffeomorphism such that the discrepancy between $S(t_i) = \phi_{t_i}(S_0)$ and S_i is minimized for every time-points t_i. The regularity constraint on the flow of diffeomorphism leads to a flow which is geodesic within each interval $[t_i, t_{i+1}]$.

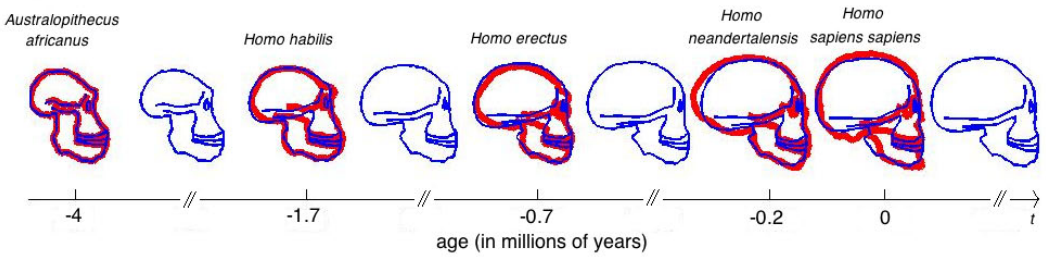


Figure 4.1: Example of shape regression: profiles of hominid skulls have been provided in red. We choose the australopithecus profile as the baseline S_0. The temporal regression computes a *continuous* flow of shapes $S(t)$ (here in blue) such that the deforming shape matches the observations at the corresponding time-points. The regression $S(t)$ can be computed by applying a flow of diffeomorphisms ϕ_t to the baseline S_0: $S(t) = \phi_t(S_0)$. In this framework, temporal regression is a natural extension of the registration between pairs of shapes.

4.2 A diffeomorphic deformation framework

As discussed above, the group of diffeomorphisms set up in [Trouvé 1998] seems particularly adapted to our needs. We recall here the definition and the main properties of this group of diffeomorphism which will be used in the next chapters. We refer the reader to [Trouvé 1998, Dupuis 1998, Joshi 2000, Miller 2002, Miller 2006] for more details on the theory.

4.2.1 A particular group of diffeomorphisms

4.2.1.1 Constructing diffeomorphisms by integrating time-varying vector fields

The concept of diffeomorphisms emerged naturally in the context of differential equations. Given an initial condition x_0, the integration of a differential equation of the type $\dot{x} = f(t, x)$ leads to a trajectory $x(t)$. The solution of this differential equation may be written in a general form as a mapping which maps every possible initial conditions x to the position of this particle at time t, which is denoted $\phi_t(x)$. Under some conditions (satisfied by usual mechanical system), the mapping $x \to \phi_t(x)$ is a diffeomorphism. And the mapping $t \to \phi_t$ is called a flow of diffeomorphisms.

The main idea is therefore to construct 3D-diffeomorphisms via the integration of time-varying vector fields. Given v_t a vector field (i.e. a mapping from $\mathbb{R}^3$ to $\mathbb{R}^3$) for every time $t \in [0, T]$, we define the following flow equation:

$$\begin{cases} \frac{\partial \phi_t(x)}{\partial t} = v_t(\phi_t(x)) \\ \phi_0(x) = x \end{cases} \tag{4.2.1}$$

or, equivalently, the integral equation:

$$\phi_t(x) = x + \int_0^t v_s(\phi_s(x))ds \tag{4.2.2}$$

In these equations, v_t plays the role of a speed vector field in the Eulerian coordinates: a particle that goes through the position x at time t has the velocity $v_t(x)$. A particle which starts from position x at time $t = 0$ is located therefore at $\phi_t(x)$ at time t. $t \to \phi_t(x)$ is the trajectory of the particle.

The question is to know under which conditions on the time-varying vector field this equation admits a solution $\phi_t : \mathbb{R}^3 \to \mathbb{R}^3$ for all $t \in [0, T]$ and under which condition this solution is a diffeomorphism of the space $\mathbb{R}^3$. Sufficient condition is to impose firstly that every mapping $x \to v_t(x)$ is differentiable (in space) and that both v_t and its Jacobian matrix tends to zero at infinity and, secondly, that the mapping $t \to v_t$ is L^1 (in time). Note that the regularity constraint is much stronger in space than in time. This is not surprising: even if the speed oscillates at a high frequency, the fact that two particles, which are close one to the other, have a correlated speed guarantees the invertibility of the flow (and enables to avoid tearing for instance). That's why a strong control of the spatial variations of the speed vector field is required.

More precisely, we define V the space of vector-field as:

Definition 4.1. *Let V be a Hilbert space of vector fields (mapping from $\mathbb{R}^3$ to $\mathbb{R}^3$) such that for any $v \in V$:*

- *v is $\mathcal{C}^1(\mathbb{R}^3, \mathbb{R}^3)$ (differentiable with continuous differential)*
- $v(x) \xrightarrow[|x|\to\infty]{} 0$
- $|d_x v| \xrightarrow[|x|\to\infty]{} 0$

Moreover, we suppose that there is a constant c_V such that for any $v \in V$:

$$\sup_{x\in\mathbb{R}^3} \left(|v(x)| + |d_x v|\right) \leq c_V \|v\|_V \tag{4.2.3}$$

This allows us to give some conditions under which the flow equation leads to diffeomorphisms of $\mathbb{R}^3$:

Theorem 4.2. *Let V be a Hilbert space of vector fields which satisfies the conditions of Definition 4.1. Let v be a mapping from $[0, T]$ to V such that:*

$$\int_0^T \|v_t\|_V^2 \, dt < \infty. \tag{4.2.4}$$

Then there is a flow ϕ_t^v which satisfies the integral equation:

$$\phi_t^v(x) = x + \int_0^t v_s(\phi_s^v(x)) ds \tag{4.2.5}$$

for all $t \in [0, T]$.

Moreover, for all $t \in [0, T]$, $x \to \phi_t^v$ is a $\mathcal{C}^1$-diffeomorphism of $\mathbb{R}^3$.

The proof of this theorem can be found in [Glaunès 2005] for instance.

Remark 4.3. In order to show the dependency of the resulting diffeomorphism on the time-varying speed vector field $v = (v_t)_{t\in[0,T]}$, we write the flow as: ϕ_t^v. □

Remark 4.4 (Acceptable vector fields can be L^1 as well)**.** The theorem still holds if the condition $\int_0^T \|v_t\|_V^2 \, dt < \infty$ (i.e. $v \in L_V^2$) is replaced by $\int_0^T \|v_t\|_V \, dt < \infty$ (i.e. $v \in L_V^1$). This last condition is weaker since we have $\int_0^T \|v_t\|_V \, dt \leq T^{1/2} \left(\int_0^T \|v_t\|_V^2 \, dt \right)^{1/2}$ thanks to the Cauchy-Schwarz inequality. However, the resulting group of diffeomorphism is the same under both conditions. □

Let V be a vector space which satisfies the conditions of Definition 4.1. We denote by $\mathcal{G}_V$ the set of all diffeomorphisms which results from the integration of a L_V^2 time-varying vector field at time $t = T$:

$$\mathcal{G}_V = \left\{ \phi_T^v; \quad v \in L_V^2 \right\} \tag{4.2.6}$$

By unicity of the solution of the flow equation (see [Glaunès 2005]), one can show that $\phi_1^v \circ \phi_1^{v'} = \phi_1^w$ where $w \in L_V^2$ is defined by:

$$w_t = \begin{cases} v_{2t} & \text{if } t \in [0, T/2] \\ v'_{2t-T} & \text{if } t \in [T/2, T] \end{cases}$$

The speed vector field w is the succession of the two speed vector fields v and v' which have been accelerated twice, so that the trajectory of a particle during $t \in [0, T]$ driving by w is the same than the succession of the two trajectories $\phi_t^v(x)$ during $t \in [0, T]$ and then $\phi_t^{v'}(x)$ during $t \in [0, T]$.

Moreover, the inverse of a diffeomorphism ϕ_T^v in $\mathcal{G}_V$ is still in $\mathcal{G}_V$. To compute $(\phi_T^v)^{-1}$, we just need to integrate the flow equation backwards:

$$(\phi_1^v)^{-1} = \phi_1^{v'} \text{ with } v'_t = -v_{T-t}$$

This shows that $\mathcal{G}_V$ is a subgroup of the group of diffeomorphism.

We notice that these equations are of high practical interest. Computing the composition and the inversion of diffeomorphisms can be done explicitly in this framework, simply by integrating ODE. This differs from other frameworks in which diffeomorphisms are given as correspondence fields $x \to \phi(x)$. Computing the inverse may be difficult and may require to optimize a criterion of the form $\|\phi \circ \psi - \mathrm{Id}\|$. Moreover, there is no guarantee that the estimated inverse map ψ belong to the same group of deformation as ϕ. By contrast, diffeomorphisms in the LDDMM framework are always given with their parameterization v which allows efficient computations within this group of diffeomorphisms.

Now, we introduce the notion of length of a path in the group of diffeomorphisms. The following proposition shows that this length does not depend on the time-parameterization of the flow of diffeomorphisms:

Proposition 4.5 (Length does not depend on time-parameterization)**.** *Let $(v_t)_{t\in[0,T]} \in L_V^2$ be a time-varying vector-field. This vector field draws a path on $\mathcal{G}_V$ between* Id *and ϕ_T: $(\phi_t^v)_{t\in[0,T]}$ via the flow equation in Eq. (4.2.2) and Theorem 4.2. The length of this path defined as:*

$$l\left((\phi_t^v)_{t\in[0,T]}\right) = \int_0^1 \|v_t\|_V \, dt \tag{4.2.7}$$

does not depend on the parameterization of the flow.

Proof. First, we notice that the Cauchy-Schwartz inequality implies (as in Remark 4.4) that the squared-integrable vector field (v_t) is also integrable. Let ψ be a change of parameterization of the flow (i.e. an increasing $1D$-diffeomorphism such that $\psi([0,T]) = [0,T]$). Then $x(t) = \phi_t^v(x)$ becomes $\tilde{x}(t) = x(\psi(t))$. The velocity field $\tilde{v}_t$ which parameterizes this new flow is given by:

$$\tilde{v}_t(x) = \frac{\tilde{x}(t)}{dt} = \frac{x(t)}{dt}\psi'(t) = \psi'(t)v_{\psi(t)}(x). \tag{4.2.8}$$

Therefore, $\tilde{v}_t = \psi'(t)v_{\psi(t)}$ and a change of variable $u = \psi(t)$ in the following integral leads to:

$$\int_{0,T} \|\tilde{v}_t\|_V \, dt = \int_{\psi^{-1}(0)=0}^{\psi^{-1}(1)=1} |\psi'(u)| \, \|v_u\|_V \, \frac{du}{|\psi'(u)|} = \int_0^T \|v_t\|_V \, dt \tag{4.2.9}$$

which proves the invariance of the length of the path in $\mathcal{G}_V$ with respect to a change of parameterization. ■

4.2.1.2 Vector fields belong to a RKHS

In Definition 4.1, the condition in Eq. (4.2.3) is a control of the spatial regularity of the vector field and of its derivative. In particular, this condition ensures also that:

$$\|v\|_\infty \leq c_V \, \|v\|_V \tag{4.2.10}$$

This proves that V is a RKHS, according to Proposition B.4. We denote K^V its kernel. Since the vector fields of the form $K^V(.,x)\alpha$ belong to V, the kernel is $\mathcal{C}^1$ and both the kernel and its differential tend to zero at infinity.

Conversely, if V is a RKHS such that its kernel is $\mathcal{C}^1$, tend to zero at infinity as well as its derivative **and** is twice differentiable with a bounded second differential, then one can show that V satisfies the conditions of Definition 4.1 (see [Glaunès 2005] for instance). For instance, the RKHS with a Gaussian kernel K^V with any standard deviation λ_V satisfies Definition 4.1. This condition allows us to check easily which kernels can be used to generate diffeomorphisms.

4.2.1.3 Metric and geodesic

For any diffeomorphism $\phi^v \in \mathcal{G}$, the total kinetic energy required to deform the space from its initial state $\phi_0^v = \mathrm{Id}$ (Id denotes the identity mapping, no deformation) to ϕ_T^v is given by: $\int_0^T \|v_t\|_V^2 \, dt$. We can define a metric between ϕ^v and the identity mapping as the minimum kinetic energy needed to map Id to ϕ_T^v:

$$d(\mathrm{Id}, \phi)^2 = \min \left\{ \int_0^T \|v_t\|_V^2 \, dt; \quad v \in L_V^2, \phi_T^v = \phi \right\} \tag{4.2.11}$$

for all $\phi \in \mathcal{G}_V$.

We can extend d on $\mathcal{G}_V \times \mathcal{G}_V$ by right-invariance: for all $\phi, \psi \in \mathcal{G}_V$,

$$d(\phi, \psi) = d(\mathrm{Id}, \psi \circ \phi^{-1}) \tag{4.2.12}$$

It is proven in [Trouvé 1995] that d is a distance on $\mathcal{G}_V$. For any $\phi \in \mathcal{G}_V$, there is a unique $v \in L^2_V$ which achieves the minimum in the definition of $d(Id, \phi)$. The associated flow of diffeomorphisms ϕ^v_t is the geodesic path between Id and $\phi = \phi^v_T$.

Moreover, in some particular cases, it is shown that the v which makes ϕ^v_t the geodesic path between Id and ϕ follows Euler-Lagrange equations. By analogy with mechanics, the minimization of the energy ($\min_{x,\dot{x}} E(x, \dot{x})$) leads to equations of motion ($\ddot{x} = f(x, \dot{x})$). As a consequence, the evolution of the system under the principle of energy minimization is entirely determined by its initial position and initial velocity. In our framework, the initial position is always the identity map $\phi_0 = \mathrm{Id}$. Therefore a geodesic starting at the identity map is entirely determined by the initial speed vector field v_0. This is called the "tangent-space representation" of the diffeomorphism. In some sense, the integration of the Euler-Lagrange equations plays the role of the exponential map in Riemannian geometry and is called "geodesic shooting" in this context. Similarly, the initial vector fields v_0, which is called the tangent-space representation of the diffeomorphism, plays the role of the logarithm in Riemannian geometry. This tangent-space representation of the diffeomorphisms will be used to define tangent-PCA on the space of diffeomorphisms like in Chapter 5 and in [Vaillant 2004]. In the next section, we will write the Euler-Lagrange equations in case of the momenta which parameterize the vector fields have a discrete support (in Eq. (4.2.22)). See [Miller 2006] for the derivation of the Hamiltonian system in other cases. We also refer the reader to [Mumford 2002, Michor 2007, Vialard 2009] for more details on the Hamitonian approach.

Remark 4.6 (Minimizing length or energy?)**.** Here, we define the geodesic distance as the minimal kinetic energy needed to connect to diffeomorphisms. However, it seems more natural to define the geodesic distance between Id and ϕ as the minimum length of every paths connecting the two diffeomorphisms, with the length as introduced in 4.5. This leads to the definition:

$$d(\mathrm{Id}, \phi) = \inf \left\{ \int_0^T \|v_t\|_V \, dt; \quad v \in L^1_V, \phi^v_T = \phi \right\} \tag{4.2.13}$$

where compared to Eq. (4.2.11), the L^2-metric has been replaced by the L^1-metric in the space of varying vector fields.

It is proven in [Glaunès 2005] that these two distances are equal: this change of the metric in the space of vector fields does not change the metric on the space of diffeomorphisms. However, the uniqueness of the minimizing vector field holds only for minimal path for the kinetic energy (L^2 metric), namely the vector field which satisfies the Euler-Lagrange equations. On the contrary, there is no uniqueness when using the L^1 metric. Indeed, it has been shown in Proposition 4.5 that any re-parameterization of the vector field which minimizes the length of the path between both diffeomorphisms leads to the same length and therefore to the same distance. Contrary to the L^2 metric, the L^1 metric does not favor any parameterization of the extremal path. This is very intuitive: independently of the speed at which one goes along the path, one covers always the same distance. But if one accelerates and slows down on the way, one would need more energy to cover the same distance. As a consequence, to minimize the kinetic energy, one needs to cover the distance at a constant speed. This is exactly what we will prove in Proposition 4.11.

This difference between the L^1 and L^2 metric will play an important role in Chapter 9 (in particular Section 9.5.2) in which we will allow to re-parameterize the geodesic path between diffeomorphisms. □

4.2.1.4 Diffeomorphisms and currents

Eventually, we still need to show that this group of diffeomorphisms is compatible with the framework of currents, namely that the push-forward action of this group of diffeomorphisms on the space of currents can be defined. In this case, a sub-manifold, which is modeled as a current, remains a current after being deformed by a diffeomorphism in $\mathcal{G}_V$. This is a key feature to use the metric on currents to drive the registration of geometrical data.

According to Defn. 1.15, the push-forward action of a diffeomorphism ϕ on a current T can be defined if $\sup_{x\in\mathbb{R}^3} |d_x\phi| < \infty$. The following proposition, which is proved in [Trouvé 2005b] for instance, shows that any diffeomorphisms in $\mathcal{G}_V$ satisfies this condition.

Proposition 4.7. *Let $v \in L^2([0,1],V)$, such that $\int_0^1 \|v_t\|_V^2\, dt < \infty$, then the diffeomorphism (ϕ_t^v) for any $t \in [0,1]$ (in the sense of Theorem 4.2) satisfies:*

$$\sup_{x\in\mathbb{R}^3} |d_x\phi_t^v| < \infty. \tag{4.2.14}$$

4.2.2 Minimization of point-based matching criteria

4.2.2.1 Existence of the minimum

In this section, we show how this diffeomorphic deformation framework can be used to solve efficiently some minimization problems like registration or temporal regression. The minimization problems we are addressing are typically of the form of a trade-off between a fidelity-to-data term and regularity term. For registration problem, one wants to find a deformation ϕ_T, such that the deformation of a source S_1 at the final time T matches the target S_2. This means that the data-to-fidelity term (the distance between $\phi_T(S_1)$ and S_2) depends on ϕ_T only, and not on the whole flow $(\phi_t)_{t\in[0,T]}$. Therefore, it is of the form: $A(\phi_T)$, where A is a mapping from $\mathcal{G}_V$ to $[0,+\infty[$. For regression problem, one wants that the deformation of the source S_1 matches some targets S_{t_i} at several time-points t_i. This is done by minimizing a least-square criterion which is the sum of the distance between $\phi_{t_i}(S_1)$ and S_{t_i}. The fidelity-to-data is therefore of the form $\sum_i A_i(\phi_{t_i})$, where A_i maps $\mathcal{G}_V$ to the non-negative real numbers. In particular, this last form includes the registration as a particular case.

We add a regularity term to these least-square criterion. The purpose is to find a trade-off between the regularity of the deformation and the precision of the matching. As regularity term, we use the total kinetic energy needed to deform the space from the identity mapping to ϕ_T: $d(\mathrm{Id},\phi_T)^2$.

The typical minimization problem can be written therefore as:

$$\min_{\phi\in\mathcal{G}_V} J(\phi) \quad \text{with } J(\phi) = \sum_i A_i(\phi_{t_i}) + \gamma d(id,\phi_T)^2 \tag{4.2.15}$$

In [Glaunès 2005], it has been shown that this minimization problem is equivalent to the minimization of the criterion over all possible time-varying vector fields $v \in L^2_V$:

$$\min_{v \in L^2_V} J(v) \quad \text{with } J(v) = \sum_i A_i(\phi^v_{t_i}) + \gamma \int_0^T \|v_t\|^2_V \, dt \tag{4.2.16}$$

The existence of a minimum of this functional has been proven, provided that $\gamma > 0$ and the functions $A_i : \mathcal{G}_V \to [0, \infty]$ are weakly continuous. This means that for all sequence v_n in L^2_V which weakly converges[1] to v, then $A_i(\phi^{v_n})$ converges to $A_i(\phi^v)$[2].

4.2.2.2 Parameterization of the minimizing diffeomorphisms

In practice, the fidelity-to-data terms measure the similarity between the moving source: $\phi_{t_i}(S)$ and target shapes S_{t_i}. Therefore, each $A_i(\phi^v_{t_i})$ can be written as $A_i(\phi^v_{t_i}(S))$. The source can be curves, surfaces or volumes: in any case it is a subset of $\mathbb{R}^3$. Then, we can prove that the time-varying vector field v which minimizes the criterion $J(v)$ can be parameterized by a set of momenta located at the points of S. When S is approximated by a finite set of Dirac delta currents in the framework of Proposition 1.11, the minimizing vector field is parameterized by a finite set of vectors. This is obviously of the uttermost interest from a computational point of view.

Proposition 4.8. *Let $S \subset \mathbb{R}^3$ and $v \in L^2_V$. For any $t \in [0,1]$, we denote $V_{\phi^v_t(S)}$ the sub-RKHS of V spanned by the points of $\phi^v_t(S)$, in the sense of Definition 2.1.*

We define $v^S \in L^2_V$ such that for every $t \in [0,T]$, v^S_t is the orthogonal projection of v_t on $V_{\phi^v_t(S)}$.

Then, the criterion $J(v) = \sum_i A_i(\phi^v_{t_i}(S)) + \gamma \int_0^T \|v_t\|^2_V \, dt$ is such that:

$$J(v^s) \leq J(v) \tag{4.2.17}$$

Proof. By definition of the orthogonal projection on $V_{\phi^v_t(S)}$, we have:

$$v^S_t(\phi^v_t(x_s)) = v_t(\phi^v_t(x_s))$$

for all points $x_s \in S$ and all $t \in [0,T]$ (see Section 2.3.2).

Therefore, the function $f(t) = \phi^v_t(x_s)$ for any point $x_s \in S$ satisfies:

$$\begin{aligned} f'(t) &= v_t(\phi^v_t(x_s)) \\ &= v^S_t(\phi^v_t(x_s)) = v^S_t(f(t)) \end{aligned} \tag{4.2.18}$$

with $f(0) = x_s$.

By definition of $\phi^{v^S}_t$, the function $f_S(t) = \phi^{v^S}_t(x_S)$ satisfies the same ODE: $f'_S(t) = v^S_t(f_S(t))$ with the same initial condition: $f_S(0) = x_S$. By unicity of the solution of the ODE, we have:

$$\forall t \in [0,T], \phi^v_t(x_s) = \phi^{v_S}_t(x_s) \tag{4.2.19}$$

[1] For a Hilbert space H, $v_n \in H$ weakly converges to $v \in H$ if and only if for all $u \in H$, $\langle v_n, u \rangle \to \langle v, u \rangle$

[2] Contrary to what the name suggests, the weak continuity is a stronger condition than the continuity. Indeed, A_i is continuous if for any sequence v_n which simply converges to v, then $A_i(\phi^{v_n})$ converges to $A_i(\phi^v)$. For the weak continuity, $A_i(\phi^{v_n})$ must converge to $A_i(\phi^v)$ for any sequence v_n weakly convergent to v. The property must be satisfied for a larger ensemble of sequences, since for sequences the convergence implies the weak convergence.

Therefore, we have that $A_i(\phi^v_{t_i}(S)) = A_i(\phi^{v^S}_{t_i}(S))$.

Moreover, for all $t \in [0,T]$, $\left\|v^S_t\right\|_V \leq \|v_t\|_V$ (v^S_t is the orthogonal projection of v_t). This proves that $\int_0^T \left\|v^S_t\right\|^2_V dt \leq \int_0^T \|v_t\|^2_V dt$ and therefore that $J(v^S) \leq J(v)$. ■

The consequence of this proposition is that we can limit the search for the minimum of J over the time-varying vector fields which are such that $v_t \in V_{\phi^v_t(S)}$. If S is given as finite set of points (polygonal lines or meshes): $S = \{x_1, \ldots, x_P\}$, then V_S and $V_{\phi^v_t(S)}$ are finite-dimensional and the minimizing vector-field is given as:

$$v_t(x) = \sum_i K^V(x, x_i(t))\alpha_i(t) \quad \text{where } x_i(t) = \phi^v_t(x_i) \tag{4.2.20}$$

This equation shows that the minimizing diffeomorphism is completely determined by the time-varying momenta $\alpha_i(t)$. Indeed, since $\frac{dx_i(t)}{dt} = v_t(x_i)$, one has:

$$\frac{dx_i(t)}{dt} = \sum_j K^V(x_i(t), x_j(t))\alpha_j(t), \tag{4.2.21}$$

for all points x_i. This gives a set of P differential equations. Once the time-varying momenta $\alpha_i(t)$ are given, the integration of this system of differential equations gives the trajectory of every points x_i: $x_i(t) = \phi^v_t(x_i)$ for all $t \in [0,T]$. Once these trajectories are computed, the velocity field v_t can be computed at any arbitrary point x via Eq. (4.2.20). Then, solving the ODE $dx(t)/dt = v_t(x(t))$ leads to the trajectory of any points x: $x(t) = \phi^v_t(x)$. As a consequence, the minimizing diffeomorphism depends only on the discrete set of time-varying momenta $\alpha_i(t)$ (L^2 functions). Once the time interval is discretized, we can derive a gradient descent scheme to optimize the criterion J numerically. The computation of the gradient will be done in the next section and the details of the numerical implementation will be explained in Section 4.4.4.2.

Remark 4.9. Using the vocabulary and notations introduced in Appendix B, we denote V^* the dual space of the RKHS V and $\mathcal{L}_V$ the isometric mapping from V to V^*. Then the minimizing time-varying vector field is such that for every time t,

$$v_t = \mathcal{L}_V^{-1}\left(\sum_p \delta^{\alpha_p(t)}_{x_p(t)}\right),$$

where $\delta^{\alpha_p(t)}_{x_p(t)}$ are Dirac delta currents in the space V^*. □

4.2.2.3 Minimizing diffeomorphisms and geodesics

In the criterion (Eq. (4.2.16)), the fidelity-to-data term $\sum_i A_i(\phi^v_{t_i}(S))$ depends on the value of time-varying vector field v only at some time points t_i. By contrast, the regularity term depends on the time-varying vector field at every time $t \in [0,T]$. As a consequence, to find the minimum of the criterion, we need first to optimize the positions of the moving shape at the time-points t_i so that they are close to the constraint $S(t_i)$ (the closeness depending on the trade-off γ between fidelity-to-data and regularity) and, second, to optimize the flow (ϕ^v_t) in-between these time points.

If we fix the points $x(t_i)$ at the time-points t_i, then the fidelity-to-data term is fixed and the criterion will decrease if we choose the velocity fields in-between the time points t_i which minimize the integrals: $\int_{t_i}^{t_{i+1}} \|v_t\|_V^2 \, dt$ in every time-interval $[t_i, t_{i+1}]$. As a consequence, the minimizing time-varying velocity field is geodesic in every time-interval $[t_i, t_{i+1}]$. For registration, there is only one time point $t = T$ and the flow which minimize the criterion is geodesic in the whole interval $[0, T]$. For general regression, the flow is piecewise geodesic. Note that the criterion does not penalize discontinuities of the velocity field. The velocity field may be discontinuous at the time points t_i. By integration, this leads to trajectories of points which are continuous and piecewise differentiable.

As a consequence, the velocity-field in the interval $[t_i, t_{i+1}[$ can be deduced from the velocity-field at time t_i via the Euler-Lagrange equations discussed in Section 4.2.1.3. In the case of a finite number of points, the initial vector field v_{t_i} is given in the form of Eq. (4.2.20), i.e. parametrized by a finite number of momenta $(x_p(t_i), \alpha_p(t_i))$. In this case, it is proven in [Miller 2006] that the geodesic evolution of v_t is such that v_t has the form $\sum_p K^V(., x_p(t))\alpha_p(t)$ at every time t for some time-varying momenta $(x_p(t), \alpha_p(t))$ and such that these momenta are solution of the Euler-Lagrange equations:

$$\begin{cases} \frac{d\alpha_p(t)}{dt} = -\left(d_{x_p(t)} v_t\right)^t \alpha_p(t) \\ \frac{dx_p(t)}{dt} = v_t(x_p(t)) \end{cases} \tag{4.2.22}$$

Solving this coupled system of differential equations in the interval $[t_i, t_{i+1}]$ leads to the geodesic momenta $(x_p(t), \alpha_p(t))$ which completely parametrize the flow of diffeomorphism in the interval $[t_i, t_{i+1}]$.

We will use these equations in Chapter 5 to compute statistics on the space of diffeomorphisms. Indeed, we can compute mean and principal modes of the vectors $\alpha_i(0)$ (which result from a template to subjects registration for instance). These statistics are given as initial momentum $(x_p(0), \alpha_p(0))$ located at the point of the template. They parametrize a diffeomorphism in $\mathcal{G}_V$ which can be computed explicitly by integrating these Euler-Lagrange equation in the interval $[0, T]$. The initial speed vector field v_0 (and its dual representation $(x_p(0), \alpha_p(0))$) is the representation of the diffeomorphism in the tangent space of $\mathcal{G}_V$ at the identity mapping (the logarithm of the diffeomorphism if we use a Riemannian terminology). Such PCA is called therefore tangent-PCA [Pennec 2006a, Vaillant 2004].

Remark 4.10. Since the solution of the minimization problem is piecewise geodesic, we can theoretically parametrize the minimizing vector fields with the momenta at time-points t_i only. Then, we can use the Euler-Lagrange equations to compute the momenta at any time t. This has the advantage to enforce the geodesic property during the optimization of the criterion. In this work, following the lines of [Glaunès 2005], we choose to discretize the time interval $[0, T]$ into several time steps $k\delta t$ and to parametrize the minimizing momenta as $\alpha_i(k\delta t)$. The piecewise geodesic property of the minimizing diffeomorphism is not guaranteed during the minimization procedure but is satisfied at the minimum. In practice, we checked that the momenta returned by the algorithm satisfy the geodesic property (i.e. are solution of the Euler-Lagrange equations). From a computational point of view, this alternative minimization has the advantage to decrease the dimension of the parameter space (initial momenta versus time-varying momenta) but it requires to integrate more differen-

tial equations (two ODEs for the integration of the Euler-Lagrange equations versus one ODE for the integration of the gradient). Moreover, this alternative approaches involves second-order derivative of the kernel, whereas the approach proposed here involves only the first derivative of the kernel. □

From these equations, we can show easily that the norm of the velocity field is constant along a geodesic paths. This can be seen as a conservation law along geodesics.

Proposition 4.11 (Conservation of the speed along geodesics)**.** *Let V be a RKHS of kernel K^V. Let the set of time-varying momenta $(x_p(t), \alpha_p(t))$ satisfy the Euler-Lagrange equations in Eq. (4.2.22) (i.e. follow a geodesic path) with:*

$$v_t(x) = \sum_p K^V(x, x_p(t))\alpha_p(t). \tag{4.2.23}$$

Then, the norm of v_t: $\|v_t\|_V$ is constant over time.

Proof. According to Remark 4.9, the norm of the velocity field satisfies in the RKHS V:

$$\begin{aligned} \|v_t\|_V^2 &= \mathcal{L}_V(v_t)(v_t) = \sum_p \delta_{x_p(t)}^{\alpha_p(t)}(v_t) \\ &= \sum_p \alpha_p(t)^t v_t(x_p(t)) \end{aligned} \tag{4.2.24}$$

Therefore, we have:

$$\begin{aligned} \frac{d\|v_t\|_V^2}{dt} &= \sum_p \frac{d\alpha_p(t)}{dt}^t v_t(x_p(t)) + \alpha_p(t)^t d_{x_p(t)} v_t \frac{dx_p(t)}{dt} \\ &= \sum_p -\alpha_p(t)^t d_{x_p(t)} v_t(v_t(x_p(t))) + \alpha_p(t)^t d_{x_p(t)} v_t(v_t(x_p(t))) \\ &= 0 \end{aligned} \tag{4.2.25}$$

by application of Eqs. (4.2.22). This achieves the proof. ■

4.3 Registration and regression of currents

In Chapter 1 (Section 1.3.3), we introduced the action of a diffeomorphism on a current. This action is compatible with the geometric transportation of shapes: the deformed shape $\phi(S)$ is modeled as the current $\phi_* \tilde{S}$ where $\tilde{S}$ is the current associated to S. In the following, we will write S instead of $\tilde{S}$.

This action allows us to write the dissimilarity metric between the deformed source S_1 at time t_i and the target S_i as $\|(\phi_{t_i})_* S_1 - S_i\|_{W^*}^2$. In the sequel, we will deal with discrete structures: S is represented as finite set of points $\{x_1, \ldots, x_P\}$ and therefore the fidelity-to-data term depends on the variables $(x_1(T), \ldots, x_P(T))$ for registration and $(x_1(t_i), \ldots, x_P(t_i))$ for several time points t_i for temporal regression:

$$A = \sum_i A_i(x_1(t_i), \ldots, x_P(t_i)) \quad \text{where } A_i(x_1(t_i), \ldots, x_P(t_i)) = \|(\phi_{t_i})_* S - S_i\|_{W^*}^2 \tag{4.3.1}$$

Note that each A_i may involve a sum of squared norm in the most general form of Eq. (1.2.22) (to account to possible anatomical labels of the structures to be matched). In this case, however, each fidelity-to-data term is still of the form $A_i(x_1(t_i), \ldots, x_P(t_i))$ and the following computation of the gradient still holds in this case.

4.3.1 Registration

For registration, the criterion has the general form:

$$J(v) = A\left(x_1(T), \ldots, x_N(T)\right) + \gamma \int_0^T \|v_t\|_V^2 \, dt \tag{4.3.2}$$

where A is a non-negative function and $x_i(t) = \phi_t^v(x_i)$. γ is the usual trade-off between regularity and fidelity-to-data.

4.3.1.1 Gradient of the registration criterion

We can apply the results of Section 4.2.2. The minimizing time-varying velocity field is parametrized by momenta $(x_i(t), \alpha_i(t))_{t\in[0,T]}$:

$$v_t(x) = \sum_i K^V(x, x_i(t))\alpha_i(t) \tag{4.3.3}$$

The criterion J depends therefore on the L^2 functions $\alpha_i(t)$ (from $[0,T]$ to $\mathbb{R}^3$). The gradient of J with respect to the variable $\alpha_i(t)$ is therefore a L^2-function of the time t. It is denoted as $\nabla_{\alpha_i(t)} J(t)$. Its expression is given in the following proposition.

Proposition 4.12. *Let J be the criterion:*

$$J\left((\alpha_i(t))_{1\leq i\leq P}\right) = A(x_1(T), \ldots, x_P(T)) + \gamma \int_0^T \sum_{i,j} \alpha_i(t)^t K^V(x_i(t), x_j(t))\alpha_j(t) dt \tag{4.3.4}$$

We suppose that the kernel is scalar and symmetric (i.e. there is a scalar function k^V such that $K^V(x,y) = k^V(x,y)\mathrm{I}_3$ and $k^V(x,y) = k^V(y,x)$).

Then the gradient of J with respect to the L^2 functions $\alpha_i(t)$ is given as a vector of P-L^2 functions $(\nabla_{\alpha_i(t)} J(t))_{i=1,\ldots,P}$, which satisfy:

$$\frac{d}{d\tau} J\left(\alpha_1(t), \ldots, \alpha_i(t) + \tau\tilde{\alpha}_i(t), \ldots, \alpha_P(t)\right) = \int_0^1 \left(\sum_{j=1}^P K^V\left(x_i(t), x_j(t)\right) \nabla_{\alpha_j(t)} J(t) \right)^t \tilde{\alpha}_i(t) dt \tag{4.3.5}$$

where the gradient is given as[3]*:*

$$\nabla_{\alpha_i(t)} J(t) = 2\gamma\alpha_i(t) + \eta_i(t) \tag{4.3.6}$$

where $\eta_i(t)$ is the solution of the backward integral equation:

$$\begin{aligned} \eta_i(t) &= \nabla_i A + \\ &\int_t^T \sum_j \Big(\alpha_i(u)^t \eta_j(u) + \alpha_j(u)^t \eta_i(u) + 2\gamma\alpha_i(u)^t \alpha_j(u) \Big) \nabla_1 k^V\big(x_i(u), x_j(u)\big) du, \end{aligned} \tag{4.3.7}$$

[3] In this expression of the gradient, we consider that the metric on the space of time-varying momenta is given by the time-varying $3P$-by-$3P$ matrix $\mathbf{K}_{\Lambda_t}$, with the notations of Definition 2.5 and for $\Lambda_t = \{x_1(t), \ldots, x_P(t)\}$.

where we denote $\nabla_i f$ the ith derivative of a scalar function f which depends on 3D-vectors (x_i) ($\nabla_i f$ is therefore a 3D vector).

Proof. A direct proof can be found in [Glaunès 2005]. However, this is a particular case of Proposition 4.14 which is proved in the next section. ■

A gradient descent starts from $\phi_t = Id$ for all t, which means $\alpha_i(t) = 0$ for all i and t. Then, the gradient is computed via Eq. (4.3.6) and (4.3.7). To compute η_i we need to compute the integral equation (4.3.7) backward in time (i.e. from $t = T$ to $t = 0$).

4.3.1.2 Differentiation of the fidelity-to-data term

We still need to compute the fidelity-to-data term when it is of the form $A((x_i(T)_{1\leq i\leq P})) = \|(\phi_T^v)_* S - S'\|^2_{W^*}$. By definition of the push-forward action on currents (in Definition 1.15), the current $(\phi_T^v)_*(S)$ corresponds to the sub-manifold $S(T)$ embedded in the space of currents, where $S(T)$ results from the displacement of the points of S by ϕ_T^v (i.e. the points $x_i(T)$). The shape $S(T)$ build faces (the mesh cells if $S(T)$ is a set of surfaces or the segments if $S(T)$ is a set of polygonal lines) with center of mass c_i. We denote η_i the normal or the tangent of the ith face. Then, $S(T)$ is approximated by $\sum_i \delta^{\eta_i}_{c_i}$ along the lines of Proposition 1.11. Note that the points c_i and vectors η_i depends on the points $x_i(T)$. The target shape S' is also approximated as a finite set of Dirac delta currents $S' = \sum_j \delta^{\eta'_j}_{c'_j}$.

The fidelity-to-data term is given as:

$$\begin{aligned} A &= \left\| \sum_i \delta^{\eta_i}_{c_i} - \sum_j \delta^{\eta'_j}_{c'_j} \right\|^2_{W^*} \\ &= \sum_{i,j} (\eta_i)^t K^W(c_i, c_j)\eta_j - 2\sum_{i,j} (\eta'_i)^t K^W(c'_i, c_j)\eta_j + Cte, \end{aligned} \tag{4.3.8}$$

where $Cte = \|S'\|^2_{W^*}$ is a constant with respect to the points $x_i(T)$.

The fidelity-to-data term A can now be differentiated with respect to the positions $x_i(T)$ to give $\nabla_i A$ in Eq. (4.3.7). In the sequel, we write simply x_i instead of $x_i(T)$. For curves, centers and faces have the form $c_i = (x_i + x_{i+1})/2$ and $\eta_i = (x_{i+1} - x_i)/2$. For surfaces, we have $c_i = (x_i + x_{i+1} + x_{i+2})/3$ and $\eta_i = (x_{i+1} - x_i) \times (x_{i+2} - x_i)/2$ (where $\times$ denotes the cross-product). The detail of the differentiation of A with respect to the points can be found in [Vaillant 2005, Glaunès 2005] for the surfaces and [Glaunès 2008] for the curves. The idea is to compute $\nabla_{x_i(f)} A$ the gradient of A with respect to the point x_i which is induced by a variation of a given face $f = (c, \eta)$ only. Then, $\nabla_{x_i} A$ is given as the sum of the $\nabla_{x_i(f)} A$ for every face which has x_i as vertex.

We suppose that K^W is a scalar kernel of the form: $K^W(x, y) = k^W(x, y)\mathrm{Id}_3$ for a positive scalar function $k^W(x, y)$ such that $k^W(x, y) = k^W(y, x)$. Then, we have for surfaces:

$$\begin{aligned} \nabla_{x_i(f)} A = &\sum_k e_i \times k^W(c_k, c_f)\eta_k - \sum_j e_i \times k^W(c'_j, c_f)\eta'_j \\ &+ \frac{2}{3}\sum_k \eta_f^t \eta_k \nabla_1 k^W(c_k, c_f) - \frac{2}{3}\sum_j \eta_f^t \eta'_j \nabla_1 k^W(c'_j, c_f), \end{aligned} \tag{4.3.9}$$

where c_f and η_f are the center and the normal of the face $x_i(f)$ and e_i the edge of the face f opposite to the vertex $x_i(f)$.

And for curves, we have:

$$\begin{aligned}\nabla_{x_i(f)} A =& 2\varepsilon \sum_k k^W(c_k, c_f)\eta_k - 2\varepsilon \sum_j k^W(c'_j, c_f)\eta'_j \\ &+ \sum_k \eta_f^t \eta_k \nabla_1 k^W(c_k, c_f) - \sum_j \eta_f^t \eta'_j \nabla_1 k^W(c'_j, c_f),\end{aligned} \tag{4.3.10}$$

where $\varepsilon = 1$ if $x_i(f)$ is the ending point of the face (oriented tangent) f and $\varepsilon = -1$ if $x_i(f)$ is the starting point of the oriented tangent f.

Remark 4.13. In this section, we differentiated the fidelity-to-data term with respect to the vertices of the m-mesh with a chain rule: a variation of the position of the vertices induces a variation of the position of the center of mass of the mesh cell, which induces, in turn, a variation of the current. In particular, this approach imposes that the m-mesh remains a m-mesh during deformation. However, one could also consider that a m-mesh is decomposed into a set of mesh cells and that the deformation can move each cell independently (in this case we consider the fidelity-to-data term as a function of the position of the center of mass of the cells). As a result, the m-mesh would not be a m-mesh after registration but a set of disconnected cells.

The alternative approach suggested here has the advantage to be closer to the modeling of currents. It would move every mesh cell or every segment independently, without preserving the topology of the shapes (i.e. the connectivity between points). This would be particularly adapted for the registration of set of Dirac delta currents returned by the Matching Pursuit Algorithm for instance which are not provided with any connectivity. By contrast, for pairwise registration of surfaces, one would prefer to enforce the moving shape to stay a mesh and use the norm on currents only to drive the deformation (see the discussion in Section 3.5.2). From a computational point of view, the alternative method is slightly more complex, since it requires to compute the Jacobian of the diffeomorphism (to deform infinitesimal tangents or normals).

If we still want to use the proposed algorithm to register currents which result from the application of the Matching Pursuit algorithm, we need to take care of the amplitude of the Dirac delta currents to be matched. Indeed, instead of matching the Dirac delta current δ_x^η, the algorithm will match the segment $[x - \eta/2, x + \eta/2]$ in case of 1-current or a small triangle whose normal is given by η in case of 2-currents. If the magnitude of η is large with respect to λ_V, then the deformation of the segment $[\phi(x - \eta/2), \phi(x + \eta/2)]$ will be a bad approximation of the deformed Dirac delta current $\phi_* \delta_x^\eta$. To avoid this, we write: $\delta_x^\eta = \frac{1}{a}\delta_x^{\eta'}$, where $\eta' = a\eta$ and a a coefficient such that $|\eta'| \ll \lambda_V$. Then, in the example of curves, the input of the registration algorithm will be the segment $[x - \eta'/2, x + \eta'/2]$ associated to the weight $1/a$. In this case, the deformation of the segment: $(\phi(x - \eta'/2) - \phi(x + \eta'/2))/a$ would be good approximation of $d_x\phi(\eta) = a d_x\phi(\eta')$. And the registration will be consistent with the framework based on currents. □

4.3.2 Regression of longitudinal data

For temporal regression, the criterion has the general form:

$$J(v) = \sum_{t_i} A_i\left(x_1(t_i), \ldots, x_N(t_i)\right) + \gamma \int_0^T \|v_t\|_V^2 \, dt \tag{4.3.11}$$

where A_i is non-negative functions and $x_i(t) = \phi_t^v(x_i)$. γ is the usual trade-off between regularity and fidelity-to-data.

The typical solution of this temporal regression problem is illustrated in Fig. 4.1, where $A_i = \left\|(\phi_{t_i}^v)_* S_0 - S_i\right\|_{W^*}^2$, S_0 a baseline and (t_i, S_i) a set of time-indexed shapes.

According to the results of Section 4.2.2, the minimizing velocity field is parameterized via time-varying momenta $(x_i(t), \alpha_i(t))$:

$$v_t = \sum_i K^V(., x_i(t))\alpha_i(t) \tag{4.3.12}$$

and the resulting flow of diffeomorphism is geodesic in every interval $[t_i, t_{i+1}]$, as well as the intervals $[0, t_0]$ and $[t_n, T]$.

The gradient of the regression criterion $J((\alpha_i(t))_{t\in[0,T]})$ is given in Proposition 4.14. We compute the derivatives of the A_i as in Section 4.3.1.2.

Proposition 4.14. *Let J be a regression criterion of the form:*

$$\begin{aligned} J\left((\alpha_i(t))_{t\in[0,T],1\le i\le P}\right) = &\sum_{t_i} A_i(x_1(t_i), \ldots, x_P(t_i)) \\ &+ \gamma \int_0^T \sum_{p,q=1}^P \alpha_p(t)^t K^V(x_p(t), x_q(t))\alpha_q(t) dt \end{aligned} \tag{4.3.13}$$

where $x_i(t) = \phi_t^v(x_i) = x_i + \int_0^t v_s(x_i(s))ds$ with

$$v_t(x) = \sum_{p=1}^P K^V(x, x_p(t))\alpha_p(t) \tag{4.3.14}$$

for $(\alpha_p(t))$, P L^2 functions from $[0,T]$ to $\mathbb{R}^3$.

We suppose moreover that the kernel K^V is scalar and symmetric (i.e. there is a scalar function k^V such that $K^V(x,y) = k^V(x,y)\mathrm{I}_3$ and $k^V(x,y) = k^V(y,x)$).

Then the gradient of J with respect to the L^2 functions $\alpha_p(t)$ is given as a vector of P-L^2 functions $(\nabla_{\alpha_p(t)} J(t))_{p=1,\ldots,P}$, which satisfy:

$$\frac{d}{d\tau} J\left(\alpha_1(t), \ldots, \alpha_p(t) + \tau\tilde{\alpha}_p(t), \ldots, \alpha_P(t)\right) = \int_0^1 \left(\sum_{q=1}^P K^V\left(x_p(t), x_q(t)\right) \nabla_{\alpha_q(t)} J(t)\right)^t \tilde{\alpha}_p(t) dt \tag{4.3.15}$$

where the gradient is given as[4]*:*

$$\nabla_{\alpha_p(t)} J(t) = 2\gamma\alpha_p(t) + \eta_p(t) \tag{4.3.16}$$

[4] In this expression of the gradient, we consider that the metric on the space of time-varying momenta is given by the time-varying $3P$-by-$3P$ matrix $\mathbf{K}_{\Lambda_t}$, with the notations of Definition 2.5 and for $\Lambda_t = \{x_1(t), \ldots, x_P(t)\}$.

where $\eta_p(t)$ is the solution of the linear set of backward integral equations:

$$\eta_p(t) = \sum_{t_i} (\nabla_p A_i) \mathbf{1}_{\{t \leq t_i\}} + \int_t^T \sum_q \left(\alpha_p(u)^t \eta_q(u) + \alpha_q(u)^t \eta_p(u) + 2\gamma \alpha_p(u)^t \alpha_q(u) \right) \nabla_1 k(x_p(u), x_q(u)) du \tag{4.3.17}$$

where $\mathbf{1}_{\{t \leq t_i\}} = 1$ if $t \leq t_i$ and 0 otherwise and $\nabla_i f$ the ith derivative of a scalar function f which depends on 3D-vectors (x_i) ($\nabla_i f$ is therefore a 3D vector).

Proof. For the sake of simplicity, we introduce matrix notations[5]: $\mathbf{x}_t$ (resp. $\boldsymbol{\alpha}_t$) denotes the $3P$ vector $(x_p(t))_{p=1\dots P}$ (resp. $(\alpha_p(t))_{p=1\dots P}$) and $\mathbf{K}_{\Lambda_t}$ the $3P$-by-$3P$ matrix $(K^V(x_p(t), x_q(t)))_{p,q}$. To make explicit the dependency of this matrix with respect to the positions $\mathbf{x}_t$, we will write it as $\mathbf{k}(\mathbf{x}_t, \mathbf{x}_t)$ in the following. The norm of the speed vector v_t is written: $\|v_t\|_V^2 = (\boldsymbol{\alpha}_t)^t \mathbf{k}(\mathbf{x}_t, \mathbf{x}_t) \boldsymbol{\alpha}_t$. By extension, we denote also $\mathbf{k}(x, \mathbf{x})\boldsymbol{\alpha} = \sum_p K(x, x_p)\alpha_p$. For A, a function from $\mathbb{R}^3$ to $\mathbb{R}$, we denote by $d_x A$ its Jacobian matrix at point x, so that for any vector V: $d_x A(V) = (\nabla_x A)^t V$. In turn, $\nabla_{\mathbf{x}} A$ denotes the $3P$ vector $(\nabla_{x_1} A, \dots, \nabla_{x_P} A)$.

With these notations, the criterion to be minimized becomes:

$$J\left((\boldsymbol{\alpha}_t)_{t\in[0,T]}\right) = \sum_{t_i} A_i(\mathbf{x}_{t_i}) + \gamma \int_0^T (\boldsymbol{\alpha}_t)^t \mathbf{k}(\mathbf{x}_t, \mathbf{x}_t) \boldsymbol{\alpha}_t dt \tag{4.3.18}$$

We compute the variation of the criterion J with respect to a variation of the momenta: $\boldsymbol{\alpha}^\varepsilon = \boldsymbol{\alpha} + \varepsilon\tilde{\boldsymbol{\alpha}}$. These momenta yield to velocity fields v_t^ε and points trajectory $\mathbf{x}_t^\varepsilon$. We denote $\tilde{\boldsymbol{\alpha}}$ (resp. $\tilde{v}$ and $\tilde{\mathbf{x}}$) the variation with respect to ε of the momenta (resp. the velocity field and the positions): $\partial\boldsymbol{\alpha}^\varepsilon/\partial\varepsilon$ (resp. $\partial v^\varepsilon/\partial\varepsilon$ and $\partial\mathbf{x}^\varepsilon/\partial\varepsilon$).

Since $v_t(x) = \mathbf{k}(x, \mathbf{x}_t)\boldsymbol{\alpha}_t$, we have :

$$\tilde{v}_t(x) = \partial_1(\mathbf{k}(\mathbf{x}_t, x)\boldsymbol{\alpha}_t)\tilde{\mathbf{x}}_t + \mathbf{k}(x, \mathbf{x}_t)\tilde{\boldsymbol{\alpha}}_t \tag{4.3.19}$$

where ∂_1 denotes the derivative of the $3P$-by-$3P$ matrix with respect to the $3P$-vector $\mathbf{x}$.

Thanks to the flow equation (Eq. (4.2.2)), $\mathbf{x}_t = \mathbf{x} + \int_0^t v_s(\mathbf{x}_s)ds$. The variations $\tilde{\mathbf{x}}_t$ satisfy therefore:

$$\tilde{\mathbf{x}}_t = \int_0^t (\partial_1 + \partial_2)(\mathbf{k}(\mathbf{x}_s, \mathbf{x}_s)\boldsymbol{\alpha}_s)\tilde{\mathbf{x}}_s + \mathbf{k}(\mathbf{x}_s, \mathbf{x}_s)\tilde{\boldsymbol{\alpha}}_s ds \tag{4.3.20}$$

The time-varying vectors $\tilde{\mathbf{x}}_t$ are solution of an inhomogeneous ordinary differential equation, which can be solved by the method of variation of parameters. Let R_{st} be the operator which gives the solution of the homogeneous equation:

$$\frac{dR_{st}}{dt} = (\partial_1 + \partial_2)(\mathbf{k}(\mathbf{x}_t, \mathbf{x}_t)\boldsymbol{\alpha}_t)R_{st} \tag{4.3.21}$$

so that the variations $\tilde{\mathbf{x}}_t$ are written as: $\tilde{\mathbf{x}}_t = \int_0^t R_{st}\mathbf{k}(\mathbf{x}_s, \mathbf{x}_s)\tilde{\boldsymbol{\alpha}}_s ds$.

[5]These are the same matrix notations as in Section 2.2.2, the set Λ now depends on t is given by the point positions $x_i(t)$.

We can now write the variation of the criterion J in Eq. (4.3.18) with respect to the variation $\boldsymbol{\alpha}^\varepsilon$:

$$\partial_\varepsilon J(\boldsymbol{\alpha}^\varepsilon) = \sum_i (\nabla_{\mathbf{x}_{t_i}} A_i)^t \tilde{\mathbf{x}}_{t_i} + 2\gamma \int_0^T (\tilde{\boldsymbol{\alpha}}_t)^t \mathbf{k}(\mathbf{x}_t, \mathbf{x}_t)\boldsymbol{\alpha}_t dt + \gamma \int_0^T (\boldsymbol{\alpha}_t)^t \partial_\varepsilon(\mathbf{k}(\mathbf{x}_t^\varepsilon, \mathbf{x}_t^\varepsilon)\boldsymbol{\alpha}_t)dt \tag{4.3.22}$$

Substituting $\tilde{\mathbf{x}}$ in this equation and permuting the two integrals (Fubini's theorem applies since every functions are in $L^2([0,T],\mathbb{R}^{3P})$), the third term becomes:

$$\gamma \int_0^T \left(\int_t^T R_{ts}^t (\partial_1 + \partial_2)(\mathbf{k}(\mathbf{x}_s, \mathbf{x}_s)\boldsymbol{\alpha}_s)^t \boldsymbol{\alpha}_s ds \right)^t \mathbf{k}(\mathbf{x}_t, \mathbf{x}_t)\tilde{\boldsymbol{\alpha}}_t dt. \tag{4.3.23}$$

The contribution of every A_i to first term can be written as:

$$(\nabla_{x_i} A_i)^t \tilde{\mathbf{x}}_{t_i} = \int_0^T (\nabla_{\mathbf{x}_{t_i}} A_i)^t R_{st_i} \mathbf{k}(\mathbf{x}_s, \mathbf{x}_s)\tilde{\boldsymbol{\alpha}}_s \mathbf{1}_{\{s \leq t_i\}} ds \tag{4.3.24}$$

where $\mathbf{1}_{\{t \leq t_i\}} = 1$ if $t \leq t_i$ and 0 otherwise (as a function of t).

The variation of the criterion is therefore:

$$\partial_\varepsilon J^\varepsilon = \int_0^T \mathbf{k}(\mathbf{x}_t, \mathbf{x}_t)(2\gamma\boldsymbol{\alpha}_t + \boldsymbol{\eta}_t)^t \tilde{\boldsymbol{\alpha}}_t dt \tag{4.3.25}$$

where $\boldsymbol{\eta}_t = \gamma \int_t^T (R_{ts})^t ((\partial_1 + \partial_2)\mathbf{k}(\mathbf{x}_s, \mathbf{x}_s)\boldsymbol{\alpha}_s)^t \boldsymbol{\alpha}_s ds + \sum_i (R_{tt_i})^t \nabla_{\mathbf{x}_{t_i}} A_i \mathbf{1}_{\{t \leq t_i\}}$.

The gradient of J as a L^2-function from $t \in [0,T]$ to $\mathbb{R}^{3P}$ can now be written as: $(\nabla J)_t = \mathbf{k}(\mathbf{x}_t, \mathbf{x}_t)(2\gamma\boldsymbol{\alpha}_t + \boldsymbol{\eta}_t)$. Since we provided the space of time-varying momenta with the metric $t \to \mathbf{k}(\mathbf{x}_t, \mathbf{x}_t)$, the gradient is given by:

$$(\nabla J)_t = 2\gamma\boldsymbol{\alpha}_t + \boldsymbol{\eta}_t \tag{4.3.26}$$

In order to compute the gradient, we still need to compute $\boldsymbol{\eta}_t$. For this purpose, we write the homogeneous equation (Eq. (4.3.21)) in its integral form: $R_{st} = \mathrm{Id} + \int_s^t R_{rt}(\partial_1 + \partial_2)(\mathbf{k}(\mathbf{x}_r, \mathbf{x}_r)\boldsymbol{\alpha}_r)dr$. This allows us to write $\boldsymbol{\eta}_t$ in the form (once the two integrals have been permuted thanks to Fubini's theorem):

$$\begin{aligned}\boldsymbol{\eta}_t = {} & \textstyle\sum_i (\nabla_{\mathbf{x}_{t_i}} A_i)\mathbf{1}_{\{t \leq t_i\}} + \int_t^T (\partial_1 + \partial_2)(\mathbf{k}(\mathbf{x}_u, \mathbf{x}_u)\boldsymbol{\alpha}_u)^t \\ & \Bigg(\gamma\boldsymbol{\alpha}_u + \underbrace{\sum_i (R_{ut_i})^t \nabla_{\mathbf{x}_{t_i}} A_i \mathbf{1}_{\{t \leq t_i\}}\mathbf{1}_{\{u \leq t_i\}} + \gamma \int_u^T (R_{us})^t (\partial_1 + \partial_2)(\mathbf{k}(\mathbf{x}_s, \mathbf{x}_s)\boldsymbol{\alpha}_s)^t \boldsymbol{\alpha}_s ds}_{(\star)} \Bigg) du\end{aligned} \tag{4.3.27}$$

Now, we notice that $t \leq u$ within the integral, which implies that $\mathbf{1}_{\{t \leq t_i\}}\mathbf{1}_{\{u \leq t_i\}} = \mathbf{1}_{\{u \leq t_i\}}$. Hence, $(\star)$ is precisely equal to $\boldsymbol{\eta}_u$. Therefore, $\boldsymbol{\eta}_t$ is the solution of the integral equation (integrated upstream in time):

$$\boldsymbol{\eta}_t = \sum_i \nabla_{\mathbf{x}_{t_i}} A_i \mathbf{1}_{\{t \leq t_i\}} + \int_t^T \left((\partial_1 + \partial_2)\mathbf{k}(\mathbf{x}_u, \mathbf{x}_u)\boldsymbol{\alpha}_u\right)^t (\gamma\boldsymbol{\alpha}_u + \boldsymbol{\eta}_u)du \tag{4.3.28}$$

Unsurprisingly, if there is only one time point $t_1 = T = 1$, we retrieve the same gradient as in [Vaillant 2005, Glaunès 2008] for a pairwise 3D registration. For several time-points,

we solve this equation from $t = T$ to $t = 0$. The successive contributions $(\nabla_{\mathbf{x}_{t_i}} A_i)$ are added as long as t reaches 0.

To retrieve Eq. (4.3.17), we rewrite this last equation with coordinates (here we use block-matrix and consider the coordinate as 3D-vectors). The $3P$ vector $\nabla_{\mathbf{x}_{t_i}} A_i$ is equal to $\nabla_{x_1(t_i)} A_i, \ldots, \nabla_{x_P(t_i)} A_i$. The kth coordinate of the $3P$-vector $\mathbf{k}(\mathbf{x}, \mathbf{y})\boldsymbol{\alpha}$ is given as: $(\mathbf{k}(\mathbf{x}, \mathbf{y})\boldsymbol{\alpha})_k = \sum_p k(x_k, y_p)\alpha_p$ (and this for generic $3P$-vectors $\mathbf{x}, \mathbf{y}, \boldsymbol{\alpha}$). We have therefore:

$$\begin{aligned} \partial_{x_i} \left(\mathbf{k}(\mathbf{x}, \mathbf{y})\boldsymbol{\alpha}\right)_j &= \sum_{p=1}^{P} \alpha_p \left(\nabla_1 k(x_i, y_p)\right)^t \delta(i-j) \\ \partial_{y_i} \left(\mathbf{k}(\mathbf{x}, \mathbf{y})\boldsymbol{\alpha}\right)_j &= \alpha_i \left(\nabla_2 k(x_i, y_i)\right)^t \end{aligned} \tag{4.3.29}$$

Therefore, for a generic $3P$-vector $\boldsymbol{\beta}$, we have:

$$\Big(\left(\partial_1 + \partial_2\right) \left(\mathbf{k}(\mathbf{x}, \mathbf{y})\boldsymbol{\alpha}\right)^t \boldsymbol{\beta} \Big)_k = \sum_{p=1}^{P} \alpha_p^t \beta_k \nabla_1 k(x_k, y_p) + \alpha_k^t \beta_p \nabla_2 k(x_p, y_k) \tag{4.3.30}$$

Now, we can apply this equation with $\mathbf{y} = \mathbf{x}$ and $\boldsymbol{\beta} = \gamma\boldsymbol{\alpha} + \boldsymbol{\eta}$ and combine it with Equation (4.3.28). Noticing that for a symmetric kernel, we have $\nabla_1 k(x, y) = \nabla_2 k(y, x)$, we get eventually Eq. (4.3.17). ■

We start the gradient descent by setting $\alpha_p(t) = 0$ for all t and p ($\phi_t = \mathrm{Id}$, for all $t \in [0, T]$). Then, the gradient is computed via Equations (4.3.16) and (4.3.17). The integration of Eq. (4.3.17) is performed upstream in time. The initial conditions at $t = T$ is given by $\nabla_{x_p(T)} A_T$ (which may be equal to 0 if there is no target at time $t = T$). Then the ODE is integrated for decreasing time t. As soon as a new time point t_i is reached, a new contribution $\nabla_{x_p(t_i)} A_i$ is added to η_u. As a consequence, $(\nabla J)_p(t)$ (and therefore the momenta $\alpha_p(t)$ and the vector field v_t) at time t depend on all the data which appear later than t. Once the vector field is computed, the positions $x_p(t)$ are computed by the integration of the flow equation ($x_p(t) = x_p + \int_0^t v_s(x_p(s))ds$) downstream in time (the initial condition is given at time $t = 0$ by $x_p(0) = x_p$). These positions at time t depend on the vector field v_t for all time earlier than t. As a result, the positions $x_p(t)$ depend on all the data in past and future. This regression fits the best trajectory ($\phi_t(S_0)$) to all the data *globally*, namely by taking into account all the matching constraints simultaneously. This differs, for instance, from pairwise registrations between consecutive time-points, although both techniques result in a piecewise geodesic flow.

The flow of diffeomorphisms (ϕ_t^v) can be extended at all times by assuming $v_t = 0$ (and hence ϕ_t^v constant) outside the time interval $[0, T]$. This property will be used for spatiotemporal registration in Chapter 9.

Remark 4.15 (boundary conditions)**.** In Definition 4.1, we imposed that the vector fields in V tend to zero at infinity. This defines the boundary conditions of the deformation. In areas which are located much further than λ_V from the data, the deformations resulting from a registration or a regression are locally similar to the identity mapping (i.e. no deformation). □

Remark 4.16 (Unit of the trade-off γ)**.** In the registration and regression criterion, we introduce the trade-off γ between the fidelity-to-data and regularity. The fidelity-to-data

term is the squared norm of the current. It is of dimension L^2 (L denotes length here) of curves and $(L^2)^2 = L^4$ for surfaces. The regularity term ($\int_0^T \|v_t\|^2 dt$) is the integral over time of a squared speed. It has the dimension $(LT^{-1})^2 T = L^2 T^{-1}$ (where T denotes time). For registration, the time interval is normalized $T = 1$ and we can suppose that the regularity term is of dimension L^2 (time is no physical sense here: it is an artificial variable). For regression, by contrast, the unit of time is given by the data, namely the age at which the subjects have been scanned. As a consequence the dimension of γ, which is the ration between the dimension of the fidelity-to-data and the one of the regularity term is given by the following table:

registration		regression	
curves	surfaces	curves	surfaces
No dim.	L^2	T	$L^2 T$

□

4.4 Numerical implementation

4.4.1 A gradient descent scheme on the time-varying momenta

We recapitulate here the different steps which lead to the computation of the gradient of the criterion and therefore to the implementation of the algorithm. In this section, we suppose that the kernel is of the form: $K^V(x,y) = k\left(\frac{|x-y|^2}{\lambda_V^2}\right)$ for a scalar function k (k' denotes its derivative).

Let the source data be the set of points $(x_i)_{i=1,\dots,P}$. The variables of the criterion in Eq. (4.3.13) are the P L^2-functions $\alpha_i(t)$. Once these momenta are given, we can compute the path of the points x_i through time via the set of P differential equations:

$$x_i(t) = x_i + \int_0^1 \sum_{j=1}^{P} k\left(\frac{|x_i(u) - x_j(u)|^2}{\lambda_V^2}\right) \alpha_j(u) du \tag{4.4.1}$$

Once these positions have been computed, we can compute the auxiliary variables $\eta_i(t)$ by integrating this set of differential equations upstream in time (from $t = T$ to $t = 0$):

$$\begin{aligned}\eta_i(t) = \sum_{t_k} \nabla_i A_k \mathbf{1}_{\{t \le t_k\}} + \frac{2}{\lambda_V^2} \int_t^T \sum_{j=1}^{P} k'\left(|x_i(u) - x_j(u)|^2\right) \\ \left(\alpha_i(u)^t \eta_j(u) + \alpha_j(u)^t \eta_i(u) + 2\gamma \alpha_i(u)^t \alpha_j(u)\right)\left(x_i(u) - x_j(u)\right) du,\end{aligned} \tag{4.4.2}$$

and then compute the gradient via:

$$\nabla_{\alpha_i(t)} J = 2\gamma \alpha_i(t) + \eta_i(t) \tag{4.4.3}$$

For computational purposes, we divide the time interval $[0,T]$ into N_{time} time-steps (denoted $t_k = kT/N_{\text{time}}$). The criterion becomes then a function of the PN_{time} 3D-vectors: $\alpha_{i,k} = \alpha_i(t_k)$, which can be stored as a 3-by-P-by-N_{time} matrix. Then we use a centered Euler method with prediction/correction scheme first to integrate Eq. (4.4.1) to give the

positions of points along time $x_{i,k} = x_i(t_k)$ and then to integrate Eq. (4.4.2) to compute the auxiliary variables $\eta_{i,k} = \eta_i(t_k)$. Eventually, the gradient of J at time t_k ($\nabla_i J(t_k)$) is a set of PN_{time} of 3D-vectors which is computed via Eq. (4.4.3). Both the auxiliary variable and the gradient can be stored as 3-by-P-by-N_{time} matrices. The matrix of the gradient $\nabla_i J(t_k)$ is used to update the current values of the momenta $\alpha_{i,k}$.

We initialize the algorithm by setting all the momenta to zero: $\alpha_{i,k} = 0$, which means there is no deformation. We used then a gradient descent with adaptive time step [Nocedal 2000]. We stop the algorithm when $(J^n - J^{n+1})/(J^0 - J^n)$ is below some positive threshold, where J^n denotes the value of the criterion after n step of the gradient descent. To evaluate the criterion, we only need to know the current positions of the moving points $x_{i,k}$ which are given by Eq. (4.4.1).

Note that a Runge-Kutta scheme for integrating equations (4.4.2) and (4.4.3) cannot be used easily since it requires to interpolate the values of the momenta in-between the time-points t_k. Using Euler scheme instead allows us to store the momenta, the positions and the gradient as matrices of the same size. The Euler method with prediction/correction scheme is of order 2 whereas the usual Runge-Kutta method is of order 4 (and simple Euler scheme of order 1).

Remark 4.17. This gradient descent use the time-varying momenta $\alpha_i(t)$ as variables. As a consequence, the constructed diffeomorphism is not necessarily geodesic (or piecewise geodesic) at each step of the gradient descent. However, when the minimum is reached, the diffeomorphism must be piecewise geodesic according to the discussion in Section 4.2.2.3. In practice, we use the initial momenta returned by the algorithm and use the geodesic shooting (see next section) to compute the geodesic time-varying momenta. Then we compare these time-varying momenta with the those returned by the gradient descent. In our experiments, this difference was numerically negligible (i.e. smaller than the approximation error induced by the optimization scheme used (see Section 4.4.4)). □

4.4.2 Flow and geodesic shooting

The output of the registration algorithm is the time-varying momenta $\alpha_i(t)$ sampled at some time-points t_k for all points x_i of the source shape. The algorithm returns also the trajectory of the points of the source shape $x_i(t)$ (such that $x_i(0) = x_i$). These momenta $(x_i(t), \alpha_i(t))$ parameterize the resulting diffeomorphism. They allow us to perform any operation involving the diffeomorphism.

Let S' be another structure. This may be any set of points: another shape or a the set of voxels of an underlying image for instance. To compute the deformation of this shape according to the diffeomorphism (i.e. $\phi_t^v(S')$), we compute the trajectory of any point y of S' by integrating the equation:

$$y(t) = y + \int_0^t \sum_{j=1}^{P} K^V\left(y, x_j(u)\right) \alpha_j(u) du \tag{4.4.4}$$

The parameterization of the inverse of the diffeomorphism is given by the time-varying

momenta $(\tilde{x}_i(t), \tilde{\alpha}_i(t))$, such that:

$$\begin{aligned} \tilde{x}_i(t) &= x_i(T-t) \\ \tilde{\alpha}_i(t) &= -\alpha_i(T-t) \end{aligned} \tag{4.4.5}$$

Therefore, if one like to compute $\phi^{-1}(S')$ instead of $\phi(S')$, one needs to integrate Eq. (4.4.4) with the time-varying momenta $(\tilde{x}_i(t), \tilde{\alpha}_i(t))$. We notice that the inversion of a diffeomorphism in this framework is straightforward and has exactly the same complexity as for computing the diffeomorphism itself.

If we want to compute the geodesic diffeomorphism which has $(x_i(0), \alpha_i(0))$ as initial momenta (which result from some statistics on the tangent-space for instance), we need to integrate the Euler-Lagrange equations. These equations are a system of two coupled differential equations:

$$\begin{cases} \alpha_i(t) = \alpha_i(0) + \displaystyle\int_0^t \sum_{j=1}^{P} \alpha_i(u)^t \alpha_j(u) \nabla_1 K^V(x_i(u), x_j(u)) du \\ x_i(t) = x_i(0) + \displaystyle\int_0^t \sum_{j=1}^{P} K^V(x_i(u), x_j(u)) \alpha_j(u) du \end{cases} \tag{4.4.6}$$

The integration of these equations leads to the time-varying momenta $(x_i(t), \alpha_i(t))$ of the geodesic diffeomorphism. We notice that integrating these equations is only twice the cost of the integration of the flow equation (4.4.4).

All these differential equations can be integrated with a centered Euler method with prediction/correction scheme which requires to know the momenta $(x_i(t), \alpha_i(t))$ at the discretization time-points only. This method returns the sampling of the dense trajectories at the same discretization points.

The computationally most expensive step when solving these partial differential equations is the computations of the sums over the P points. Compared to this cost, the cost of the numerical scheme to integrate the ODEs is negligible. In the next section, we will provide some optimization routines which allow us to solve these equations in few seconds for 10^5 points. By comparison, the gradient descent requires to solve almost 10 of such ODEs. Therefore, once the registrations have been computed, using the resulting deformations (to apply it to other structures like in Chapter 6, to compute statistics in the tangent-space like in Chapter 7 and 8 or to invert it) can be done at almost no additional costs.

4.4.3 Parameters

The whole framework depends mostly on 3 parameters to be set by the users: the kernel of the deformation K^V, the kernel of the currents K^W and the trade-off between regularity and fidelity-to-data γ. The kernels are usually chosen in a class of possible kernels which implies setting few parameters. For instance, they are of the form $K(x, y) = f(|x-y|^2 / \lambda^2)\mathrm{Id}_3$ for a fixed positive function f. The spatial scale λ_V determines the scale above which points move in an uncorrelated manner. The spatial scale λ_W determines the scale under which shapes variations are considered as noise.

Some parameters need also to be set. The optimization routines presented in the next section require to set one or two parameters. The gradient descent with adaptive step size

requires also some parameters like the initial step size and stopping criterion. However, these parameters does not depend on the input data and may be fixed once for all (except possibly the parameters of FGT optimization as explained below). The last parameter is the number of discretization steps within the interval $[0, T]$ to solve numerically the differential equations.

Remark 4.18. As discussed in Section 3.5.2 and in Fig. 3.16, the spatial scale on currents λ_W should not be too small with respect to the typical distance between the source and the target data. Otherwise, the norm on currents will be insensitive to the shape variations (source and target will be almost orthogonal). If this is not possible (one wants to compare shapes at a scale λ_W which is much smaller than the typical distance between parts of shapes to be matched), then multi-scale registration should be investigated. We start by registering shapes with a large λ_W, then we decrease λ_W and we initialize the registration with the deformation obtained in the previous step. Note, however, that we will not need to use such a strategy in our practical example. Surprisingly, even a large λ_W enables a matching with a great accuracy (i.e. the Gaussian function is not as flat at the origin as we think!). □

Remark 4.19. A RKHS with Gaussian kernel of large variance, which contains locally rigid deformations, is included into a RKHS with smaller variance, which contains deformations with possible small-scale variations. Therefore, one would like to set the spatial scale λ_V as small as possible, so that the corresponding RKHS will contain all the deformations (the ones with large-scale and small-scale variations). However, this strategy does not work, since a rigid deformation requires much more energy in a RKHS with small variance than in a RKHS with large variance. Therefore, if one looks for a large-scaled deformation with a kernel of small variance, the registration algorithm will be very likely to be trapped in a local minima. □

4.4.4 Optimization strategies

4.4.4.1 Kernel multiplication: the limiting factor

During the gradient descent, we need to compute both the criterion and the gradient. Given a set of N time-varying momenta $(x_i(t), \alpha_i(t))$, the computation of the criterion, at each time point t, requires to compute N times a sum of the form $\sum_{j=1}^{N} k\left(|y - x_j(t)|^2\right) \alpha_j(t)$ for $y = x_1(t), \ldots, x_N(t)$ and then to integrate the flow equation. The computational time of the integration scheme is negligible compared to the one of the computation of the sum. This sum can be computed independently for each coordinate. This leads to 3 sums to be computed N times. The complexity of the computation of the criterion is therefore of $3N^2$.

To compute the gradient, one essentially needs to compute the auxiliary variables $\eta_i(t)$. The limiting factor is the computation of N sums of the form:

$$\sum_{j=1}^{N} k'\left(|x_i - x_j|^2\right) \left(\alpha_i^t \beta_j + \alpha_j^t \beta_i\right) (x_i - x_j), \tag{4.4.7}$$

where $\beta_i = \gamma \alpha_i + \eta_i$, for $i = 1, \ldots, N$.

For computational purposes, we re-write this sum to be in the same form as the sum for the computation of the criterion. Noticing that $\alpha_i^t \beta_j x_j$ can be written as $(x_j \beta_j^t)\alpha_i$, we write:

$$\sum_{j=1}^{N} k'\left(|x_i - x_j|^2\right)\left(\alpha_i^t \beta_j\right)(x_i - x_j) = \alpha_i^t \left(\sum_{j=1}^{N} k'(|x_i - x_j|^2)\beta_j\right) x_i - \left(\sum_{j=1}^{N} k'(|x_i - x_j|^2) x_j \beta_j^t\right) \alpha_i \tag{4.4.8}$$

These sums involve the derivative of the kernel instead of the kernel itself. However, this does not change the complexity of the computations and the same optimization methods can be used[6].

Computing the sums coordinates by coordinates, the first term of Eq. (4.4.8) has a complexity of $3N^2$ (the vector β_j has 3 coordinates) and the second term has a complexity of $9N^2$ (the matrix $x_j^t \beta_j$ has 9 coordinates). The complexity of this equation is therefore of $12N^2$. Since there are 2 terms like this one in Equation (4.4.7), the total complexity of the computation of the gradient is eventually of $24N^2$, namely eight times the computation of the criterion.

The same arguments can be used for the computation of the fidelity-to-data term (given in Eq. (4.3.8)) within the criterion and the derivative of the fidelity-to-data term (given in Eq. (4.3.9) for surfaces and Eq. (4.3.10) for curves) within the derivative of the criterion. These equations have the same form than the ones studied here, but the kernel in the space of vector field V is replaced by the kernel on currents in W^*. Therefore, the computation the fidelity-to-data term has a complexity of $6NN_T$ and the computation of its derivative has a complexity of $24NN_T$, where N is the number of momenta of the moving source (same as above) and N_T be the number of momenta of the target shape.

The complexity of the whole gradient descent scheme is determined by the computation of these sums. Actually, the registration algorithm spends more than 90% of the time to compute such sums. It is crucial therefore to propose fast computations of sums of the form $\sum_{j=1}^{N} k(y, x_j)a_j$ for a given positive function k and N scalars a_i, when N is typically between 10^3 and 10^5.

4.4.4.2 Two optimization methods

In Chapter 2, we precisely introduced a numerical framework to compute such sums and compared it with the multipole approximations used in the original implementation of the algorithm in [Glaunès 2005].

Fast Gauss Transform

[6] Indeed, the theorem of Schoenberg [Schoenberg 1938] states that the function $-k'(|x-y|^2)$ defines a positive symmetric scalar kernel (this is trivial for the Gaussian kernel for which $k'(u) = -k(u)$). Therefore, any optimization scheme designed for a generic positive symmetric scalar kernel can be used both for the criterion (involving k) and the gradient (involving $-k'$)

As mentioned in Section 2.5.2, we can used the Fast Gauss Transform to compute the sums. The idea is to gather points into clusters and to use the Taylor expansion of the kernel at the center of the clusters. To set up this approximation framework, we need to specify mostly the number of clusters and the order at which the Taylor expansion is truncated. Empirically, a relative approximation error up to 5% in the computation of the sums is acceptable. However, there is no simple way to find the parameters which will lead to an approximation error smaller than 5%. Therefore we estimate the parameters by comparing an approximated sum with the exact result before running the registration.

The problem with this approach is that the approximation error depends on the distribution of points. The parameters are estimated for the original distribution of points. However, during the gradient descent, the positions of points change and there is no guarantee that the approximation error remain below 5%. Since the parameters are set manually, there is no simple way to re-estimate these parameters (the clustering of points especially) during the iterations of the gradient descent.

The main advantage of this method is its low space complexity: $\mathcal{O}(N)$ (or $\mathcal{O}(N+N_T)$) as shown in Section 2.5.2. Actually, this approximation does not require any additional memory space than what is needed to store the data. The time complexity is also of $\mathcal{O}(N)$ (or $\mathcal{O}(N+N_T)$). As shown in [Vaillant 2005, Glaunès 2008], this implementation of this approximation scheme has been applied successfully to register two sets of anatomical data. However, due to the lack of automatic parameters estimation, this method is not well adapted to perform routinely registrations on a large collection of shapes, like template-to-subjects registration in a statistical context for instance.

Grid-based numerical scheme

In Chapter 2, we provide another approximation scheme based on linearly-spaced grids to compute the sums. Using the notations of Remark 4.9, we need to compute the sums: $\mathcal{L}_V^{-1}\left(\sum_{j=1}^{N}\delta_{x_j}^{a_j}\right)(y_k)$, for N (or N_T) distinct points y_k. First, we set a linearly spaced grid which embed the points x_i and y_k. Then, we project the momenta (x_i, a_i) on the grid points using a partial volume projection, we compute the map $\mathcal{L}_V^{-1}$ via circular convolutions and FFT and eventually interpolate the values of the image of vectors at the points y_k.

The relative approximation error is entirely determined by the ratio Δ/λ_V between the grid step and the rate of decay of the kernel K^V (the standard deviation of a Gaussian kernel for instance) for the deformation term and Δ/λ_W for the fidelity-to-data term. In contrast to FGT, the approximation error is independent of the distribution of the points. This make possible to set a "default value" which guarantees a fixed approximation error. Empirically, we find that the value $\Delta/\lambda = 0.2$ (for λ_V or λ_W) corresponds to an approximation error below 5%. This approximation error remains constant during the gradient descent and is the same for any input data.

The complexity of the method is of $\mathcal{O}(N_{\text{grid}}\log(N_{\text{grid}}))$, where N_{grid} is the number of points of the grid. This number depends on the spreading of the points in space. However, it is insensitive to the local redundancy of the points in a neighborhood λ_V (or λ_W).

The main advantages of this method is the stability of the approximation error and the fact that parameters can be set automatically. By contrast, its main drawback is the spatial complexity of the method. The memory needed to store the grid is usually much larger

than the memory space needed to compute the exact sum or the FGT approximation. This is particularly critical when the points tends to spread in a large domain of space.

Remark 4.20. It would have been also possible to compute the sparse approximation of the current $\sum_j \delta_{x_i}^{a_i}$ in the space V^*, as explained in Chapter 3. This would give a new set of $\tilde{N}$ momenta $(\tilde{x}_i, \tilde{a}_i)$ so that sum $\sum_{j=1}^{N} k(y, x_j)a_i$ by may approximated by the sum $\sum_{j=1}^{\tilde{N}} k(y, \tilde{x}_j)\tilde{a}_i$ at any desired accuracy. If the distribution of the points x_i (the points of the source shape) is highly redundant at the scale λ_V, the number of estimated momenta $\tilde{N}$ must be much smaller than N and a direct computation may be possible. However, to run the matching pursuit algorithm (Algorithm 2), one needs to project the momenta on a linearly spaced grid as a pre-processing. Therefore, this method would not solve the spatial complexity problem of the grid-based numerical scheme. Moreover, it is not clear whether the following steps: the projection of the momenta on the grid, the application the matching pursuit and the computation the approximated sum need necessarily less time than the steps of required by the grid-based numerical scheme: the projection of the momenta on the grid, the computation of a circular convolution and the interpolation of the values on the grid nodes at points y_k. We must admit, though, that we do not test it empirically. We think that the matching pursuit algorithm would be more adapted as a post-processing, once the registrations have been performed. It can be used to give a visual interpretation of the velocity fields which drive the deformation. It may help to speed-up the computations of statistics on the initial speed vector fields as well as to give more interpretable results of these statistics. □

Remark 4.21. A way to accelerate the computations would be to decrease the number of momenta of the source shape. This could be done by approximating the source current in the space W^* by running the matching pursuit in this space. This is what we will do for registering the white fiber bundles in Chapter 7 which have a very redundant representation both at scale λ_V and λ_W. In this case, we consider that the input data are the approximated currents.

However, there is no guarantee that difference between the registration applied to the approximated current and to the original current is small. In particular, the parameterization of the minimizing vector-field (one momentum at each point of the source shape) is different. This is probably of little importance if the scale of currents λ_W is much smaller than the scale of deformation λ_V. In this case, there will be enough points in every neighborhood of size λ_V so that any vector field in V can be approximated by a vector field parametrized on the estimated source points. However, further investigations are needed in order to quantify this approximation error. □

4.5 Combined curve and surface registration of the lungs

The work presented here results from a collaboration with Vladlena Gorbunova (University of Copenhagen) and as been published in [Gorbunova 2009]. This section is a reproduction of this article. We only re-write some parts to avoid redundancy and to use consistent notations with the rest of the chapter.

The purpose of this section is to use the registration of currents to align 3D images of the lung of the same subject at end inhale and end exhale phases. For this purpose, we extract distinctive anatomical structures from images: the pulmonary vessel tree centerlines and the lung surface. Since no point correspondence can be drawn between the structures, they are modeled as currents. These features are then used to drive the registration between the image at end inhale phase to the image at end exhale phase. The resulting deformation is dense: it is applied afterwards to the underlying images. Finally, we compare this alignment with the one obtained from a pure intensity-based registration.

We conducted experiments on five pairs of images. To evaluate the registration, we used a set of 300 anatomical landmarks marked on every images. Using both vessel centerlines curves and lung surfaces yields better alignment (median error of 1.85 mm) than using only curves (2.37 mm) or surfaces (3.53 mm). The combined method achieves overall registration accuracy comparable to that of intensity-based registration. Nevertheless, the largest registration errors do not occur at the same locations for the two methods. This suggests that low dimensional geometrical features capture sufficient information to drive a reliable registration, while results can still be improved by combining intensity and feature based registration approaches into one framework.

4.5.1 Feature-based registration in lung imaging

Registration of chest CT scans is an important subject within pulmonary image analysis. The general task of registration is to establish a point-to-point correspondence between two images. Registration of lung CT images can be used in various clinical applications, such as lung cancer radiotherapy planning and quantitative analysis of disease progression.

Image registration methods can be separated into two general groups: intensity-based and feature-based methods. Intensity-based methods integrate spatial information over the entire image domain, whereas feature-based methods require a representation of the image data in terms of distinctive geometrical structures. Feature-based methods offer more robust registration when image intensity is changed because of pathology, image artifacts or differences in scan protocol for instance. Generally, segmentation of geometrical structures in lungs is less sensitive to intensity changes, since the method incorporates geometrical regularity constraints or prior anatomical knowledge. Moreover, segmentation of distinctive lung structures may be either corrected manually or delineated by a professional.

The most distinctive anatomical structures in lung CT images are vessels, airways, lobe fissures and lung surfaces. Lungs surface and lobe fissures define large-scale deformations of the lungs and provide an insight into the global motion of the lungs, while small-scale deformations are influenced by vessels and airway tree motion.

Feature-based registration relies on various geometrical structures, e.g., points, curves or surfaces. Thin-plate spline image registration [Rohr 2001, Johnson 2002, Bookstein 1991] is the standard method for matching points under the assumption that deformations are small. For large deformations, a diffeomorphic point matching approach was developed in [Joshi 2000] and was later adapted for surface matching in [Vaillant 2005] and curve matching [Glaunès 2008, Durrleman 2008c], as explained in the previous sections of this chapter.

Several surface-based registration methods were previously developed for lung CT images [Vik 2008, Li 2008, Betke 2003]. The outer surface of the lungs together with the outer surface of vessels were used in an algorithm similar to iterative closest point methods in [Vik 2008]. Lung surface was used to register CT lung images [Betke 2003] and to constrain intensity-based registration with a deformation field obtained from surface matching procedure [Li 2008]. The two main advantages of using the currents to model the features are: (1) no point correspondence is required between structures, thus making the registration less sensitive to the segmentation method used and the sampling noise and (2) the unified representation of curves and surfaces in a single framework to drive the registration with different kind of geometrical features.

The low dimensional geometrical features, such as curves and surfaces contain much fewer points compared to dense intensity images. Feature-based registration can be therefore more efficient from a computational point of view. This dimension of the features can be even more reduced using the sparse representation of currents introduced in Chapter 3. Compared to intensity-based methods, the feature-based methods select and extract meaningful information from the images. Some of the structures, like the tree-centerlines of the lung, may be almost invisible in images. They are too thin to influence an intensity-based registration using a L^2 metric between images. By contrast, the feature-based methods used this anatomically relevant information as a much harder constraint to drive the registration. As a consequence, we usually observe better alignment with feature-based registration in areas which contain lots of reliable geometrical structures.

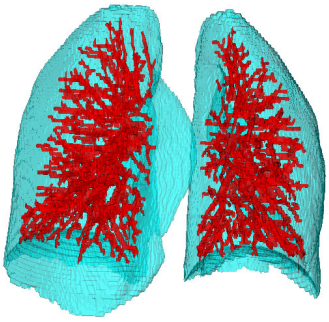
4.5.2 Segmentation of anatomical structures of the lung

In this study, we extract both the vessels and the surface of the lung from the images, as shown in Fig 4.2(a). The lung fields and vessels are segmented with the algorithm described in [Lo 2008]. A sparse triangulation of the lung surface was computed via the marching cube algorithm [iso2mesh]. For each face, the corresponding normals were computed and oriented to point outwards of the surface. We normalize the normal of each cell to 1. This leads to a description of the surface as a finite sum of Dirac delta currents: $S = \sum_s \delta_{c_s}^{n_s}$, where c_s is the center of each face and n_s the oriented normal of this face. Fig. 4.2(b) shows an example of the constructed current for a lung surface.

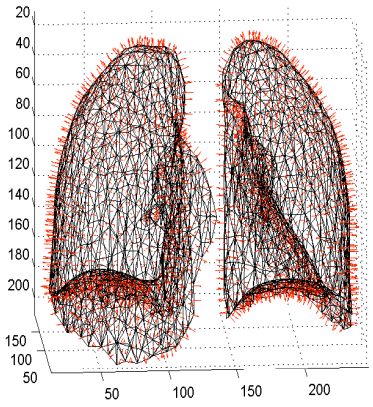
Vessel tree was segmented as follows: lung image was thresholded with a fixed intensity value $t_v = -600HU$, then a local analysis of Hessian matrix was performed in order to remove non-tube like structures. Large vessels segmented near the hilum area were omitted from the vessel tree segmentation. For more details on vessels segmentation algorithm we refer the reader to [Lo 2008]. Centerlines were extracted from the segmented vessel tree using a 3D thinning algorithm [Wang 2007].

The tangential direction of a centerline was computed via local principal component analysis. For each centerline point we extracted neighboring centerline points, applied PCA to the point cloud, and assigned the first principal component to the tangential direction at the centerline. For centerlines sufficiently far from vessel bifurcation and neighboring vessel, the principal direction points to a tangential direction of the centerline. For centerlines close to the bifurcation the principal direction points between the two splitting vessel centerlines.

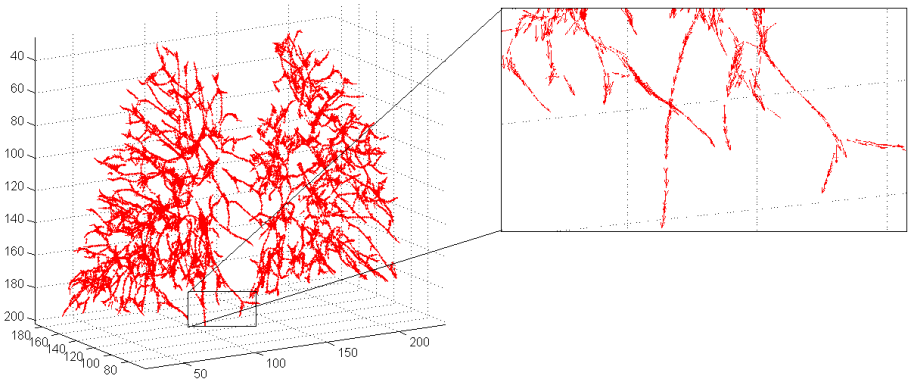
This is consistent with the framework of currents, were the addition of two Dirac delta currents which point in each direction of the bifurcation results in a Dirac delta current which point in-between the two directions. The orientation for the positive direction was set to point outwards from the center of the image. Since the norm of the tangential vectors results of a PCA, its norm is not a reliable geometrical information. To take into account only the direction of the centerlines, we normalize each tangential vector to 1. This construction leads to a description of the centerlines as $L = \sum_l \delta_{x_l}^{\tau_l}$, where x_l is the centerline points and τ_l the unit tangential direction at x_l. Since the tree-structure of the centerlines is not consistent across the subjects, we handle all centerlines as a single current L. Indeed, there is no anatomical reason to label the centerlines. Fig. 4.2(c) shows an example of the constructed current for a segmented vessel tree and a zoom-in into a bottom part of the image.



(a) Example of segmented lung surface and lung vessel tree



(b) Momenta corresponding to a lung surface.



(c) Momenta corresponding to a vessel tree centerlines.

Figure 4.2: Example of segmented lungs surface and vessel tree 4.2(a); triangulation of the lungs surface (black mesh) with the corresponding momenta (red vectors) 4.2(b); momenta corresponding to the vessel tree centerlines (red vectors) with a zoom-in 4.2(c).

4.5.3 Current-based Registration

To define similarity metric for the centerlines and for the surfaces, we introduce two RHKS of currents: W_L^* for the centerlines whose kernel is Gaussian with variance λ_L^2 and W_S^* for the surfaces whose kernel is Gaussian with variance λ_S^2. These parameters determine the typical scales at which the metric is sensitive to shape variations. Much smaller than this scale, shape variations are considered as noise. See Section 3.5.2 for more details on these parameters.

The space of diffeomorphisms is generated by integrating velocity fields which belong to the RKHS V whose kernel is Gaussian with variance λ_V^2. This scale determines the typical distance under which the points of the space move consistently.

The registration of the centerlines only from the source curves L_1 (considered as a single current) and the target curves L_2 (considered as a single current) is achieved via the minimization of the criterion:

$$J\left((v_t)_{t\in[0,1]}\right) = \left\|{\phi_1^{(v_t)}}_* L_1 - L_2\right\|_{W_L^*}^2 + \gamma_\phi \int_0^1 \|v_t\|_V^2 \, dt \tag{4.5.1}$$

The registration of the surface only from the source surface S_1 to the target surface S_2 is achieved via the minimization of the criterion:

$$J\left((v_t)_{t\in[0,1]}\right) = \left\|{\phi_1^{(v_t)}}_* S_1 - S_2\right\|_{W_S^*}^2 + \gamma_\phi \int_0^1 \|v_t\|_V^2 \, dt \tag{4.5.2}$$

The joint registration of both centerlines and surfaces from source data (L_1, S_1) to target (L_2, S_2) is defined by the minimization of the criterion:

$$J\left((v_t)_{t\in[0,1]}\right) = \left\|{\phi_1^{(v_t)}}_* L_1 - L_2\right\|_{W_L^*}^2 + \gamma_{SL}\left\|{\phi_1^{(v_t)}}_* S_1 - S_2\right\|_{W_S^*}^2 + \gamma_\phi \int_0^1 \|v_t\|_V^2 \, dt \tag{4.5.3}$$

4.5.4 Experiments

In order to quantify the accuracy of the proposed registration method with a ground truth, we used images from a publicly available dataset [Castillo 2009]. For each image pair, 300 manually placed corresponding landmarks were provided. Five pairs of images, where each pair consists of images extracted at end exhale and end inhale phases of 4D CT image, were used in our experiments. In-plane resolution of the images varied from 0.97×0.97 mm to 1.16×1.16 mm and slice thickness was 2.5 mm.

4.5.4.1 Parameter Settings

Vessel trees were segmented using the algorithm as in [Lo 2008] with the intensity threshold -600 HU, ratio of Hessian eigenvalues was set to $m_1 = 0.75$, $m_2 = 0.5$. For every centerline point we extracted a neighboring centerline points from the cube neighborhood of $7 \times 7 \times 7$ voxels size and computed the principal direction of the centerlines. A regular surface triangulation was constructed with a marching cube algorithm with further simplification of the mesh [iso2mesh].

In our experiments, end inhale phase of 4D-CT image was registered to end exhale phase. The following internal parameters of image registration were selected manually.

The accuracy of feature alignment λ_L was set to 5 mm for the curves and $\lambda_S = 10$ mm for surfaces. The spatial variability of deformation velocity field λ_V was set to 25 mm for both types of features. The weight coefficients in the cost functions (4.5.1), (4.5.2) and (4.5.3) were set to $\gamma_\phi = 10^{-4}$ for the regularizer and $\gamma_{SL} = 0.01\text{mm}^{-2}$ for the trade-off between surface and curve term.

4.5.4.2 Results

We evaluated four registration methods, as follows: combined curve- and surface-based registration with cost function (4.5.3); curve-based registration with cost function (4.5.1); surface-based registration with cost function (4.5.2); and a free-form B-Spline intensity-based method as in [Gorbunova 2008]. We compared registration accuracy of the four methods based on the alignment of 300 landmarks distributed uniformly in lung area, Fig. 4.3(b) shows an example of the spatial distribution of landmarks within the lungs.

The overall accuracy of the image registration methods was defined as the mean Euclidean distance between landmarks, target registration error (TRE), in millimeters. The mean and the standard deviation of TRE for the four methods is reported in Table 1. We performed Wilcoxon rank-sum test on TRE distribution to compare the combined curve- and surface-based registration with the curve-based and surface-based methods individually. Box-plots in Fig. 4.3(a) show the overall accuracy of the four image registration methods on a complete set of landmarks over all five cases.

Correlation between TRE for the intensity-based and combined curve- and surface-based registration was $\rho = 0.5$, varying from $0.17 - 0.59$ for the five cases. Overall, for 35.5% cases of landmarks the combined curve- and surface-based registration method performed better than intensity-based method.

4.5.5 Discussion

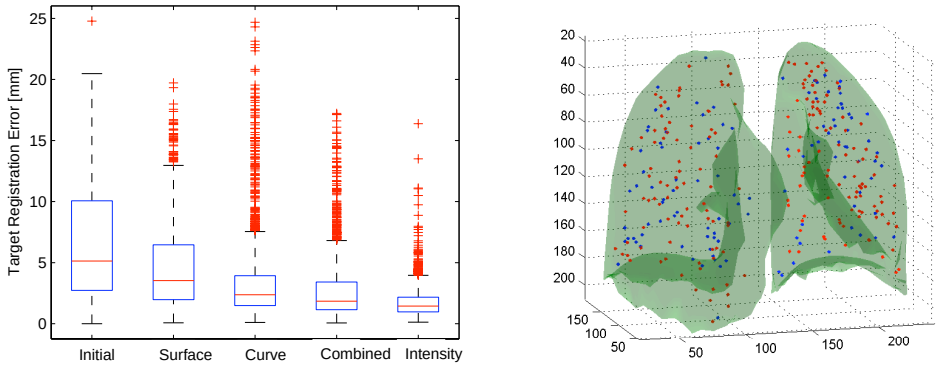
Fig. 4.3(a) shows that the curve-based method alone provides good registration accuracy for the majority of landmarks. However, there are many outliers present with errors of up to 2.5 cm. Within our framework, points located much further than the typical scale of deformations λ_V are not affected by the deformations, which might cause landmarks distant to the vessel centerlines to be misaligned.

Surface-based registration result in a slight overall improvement in TRE compare to the initial configuration. By contrast, incorporating both surfaces and curves into feature-based registration results in more accurate registration (1.85 mm) compared to both curve-based (2.37 mm) and surface-based (3.53 mm) methods.

The median of TRE for the combined curve- and surface-based registration was 1.85 mm compared to 1.44 mm for the intensity-based method. Several reasons may lead to larger TRE for the combined curve- and surface-based method, such as inconsistency in segmentations of vessels in the two images. Ambiguous segmentation of lung surface near the hilum may leads to large registration errors in this area. Fig. 4.4(b) shows a difficult

Image Registration Accuracy in mm [m ± std]						
N	Before	Combined	Surface	Curve	Intensity	%
1	3.89 ± 2.78	1.47 ± 0.72	$2.45 \pm 1.56^{*}$	$2.24 \pm 1.41^{*}$	1.23 ± 0.61	37.7
2	4.34 ± 3.90	2.19 ± 1.98	$3.63 \pm 2.94^{*}$	2.32 ± 2.06^{ns}	1.26 ± 0.67	39.0
3	6.94 ± 4.05	3.30 ± 3.05	$5.31 \pm 3.26^{*}$	$3.03 \pm 2.79^{*}$	1.86 ± 1.11	25.0
4	9.83 ± 4.86	3.34 ± 2.67	$5.98 \pm 3.74^{*}$	$5.28 \pm 4.52^{*}$	2.15 ± 1.48	36.0
5	7.48 ± 5.51	3.83 ± 3.54	$5.80 \pm 4.37^{*}$	$4.40 \pm 4.42^{*}$	2.32 ± 1.82	40.0
All 5 cases						
	6.50 ± 4.83	2.83 ± 2.72	$4.63 \pm 3.58^{*}$	$3.45 \pm 3.48^{*}$	1.76 ± 1.31	35.5
median	5.13	1.85	3.53	2.37	1.44	

Table 4.1: Registration error at the landmark positions in [mm] for the four registration methods. The mean (m) and the standard deviation (std) are reported. Statistical comparison of combined curve- and surface-based registration method was performed against the surface-based and curve-based methods. The notations of statistical significance level are as follows: * corresponds to $p < 0.05$ and ^{ns} to $p > 0.05$. The most right column indicates percentage of landmarks where the combined curve- and surface-based registration outperforms the intensity-based registration.

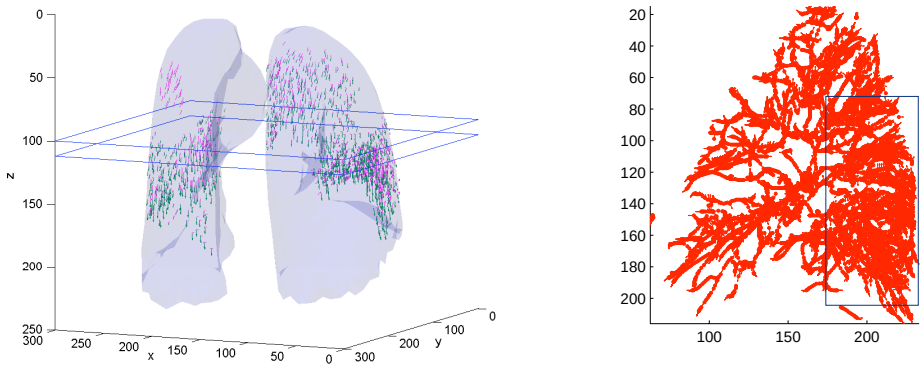


(a) Box-plot of target registration errors

(b) Distribution of landmarks

Figure 4.3: Target registration errors (TRE) is shown in 4.3(a), as follows, before registration was applied (Initial), after surface-based (Surface), after curve-based (Curve), after combined curve- and surface-based (Combined) and after intensity-based registration (Intensity). Example 4.3(b) shows the spatial distribution of landmarks in the lungs. The landmarks, better aligned with the combined feature-based method are shown in red and with the intensity-based method in blue.

case in the data with irregular centerlines in the back of the lungs. Our results show that the proposed feature-based registration method is robust to inconsistent segmentation and outliers in segmented features and capable of handling imperfect segmentations. Of course,



(a) Deformation vectors for the combined curve- and surface-based (magenta) and intensity-based (green) methods

(b) Example of segmentation with many outliers

Figure 4.4: (a) An example of discrepancy in deformation fields between the feature-based and intensity-based registration methods. (b) An example of a misleading segmentation for the back of the lung.

registration of lung images based on such geometrical structures like vessels centerlines and lung surfaces can be naturally improved by including airways and lung fissures into the presented framework.

In order to understand where are the main differences between the feature-based and intensity-based method, we visualized discrepancy between the two deformation fields in Fig. 4.4(a). For illustration purpose, we sparsely selected points where the orientation between deformation vectors were above 60° and with the magnitude of discrepancy vectors more than 3 mm and plotted inside the lung area. Interestingly, the discrepancy between the feature- and intensity-based methods were localized.

We further investigate image slices located at the areas where the discrepancy between the two methods was largest (blue cut planes in Fig. 4.4(a)). Fig. 4.5 shows the difference image with the moving image subtracted from the fixed image for both registration methods. Overall, lung surfaces and small vessels were aligned more accurately with the feature-based registration method.

Another important feature of currents is the possibility to weight each segment of the vessel centerlines. For the task of registration of repeated lung CT images, the current for a small vessel could be given more weight than for a large vessel, leading to more accurate registration of small vessels. This is an important advantage of current-based registration over intensity-based method where small vessels with low contrast to surrounding lung tissue have negligible impact on the overall cost function. In this paper we used equal weights for all currents and normalized the length to 1. These weights might be adapted according to some reliable anatomical knowledge.

On average, 35.5% of landmarks were aligned better with the curve- and surface-based registration. The low correlation coefficient (0.5) suggests that the two registration methods align landmarks differently and may be combined into a more robust registration method.

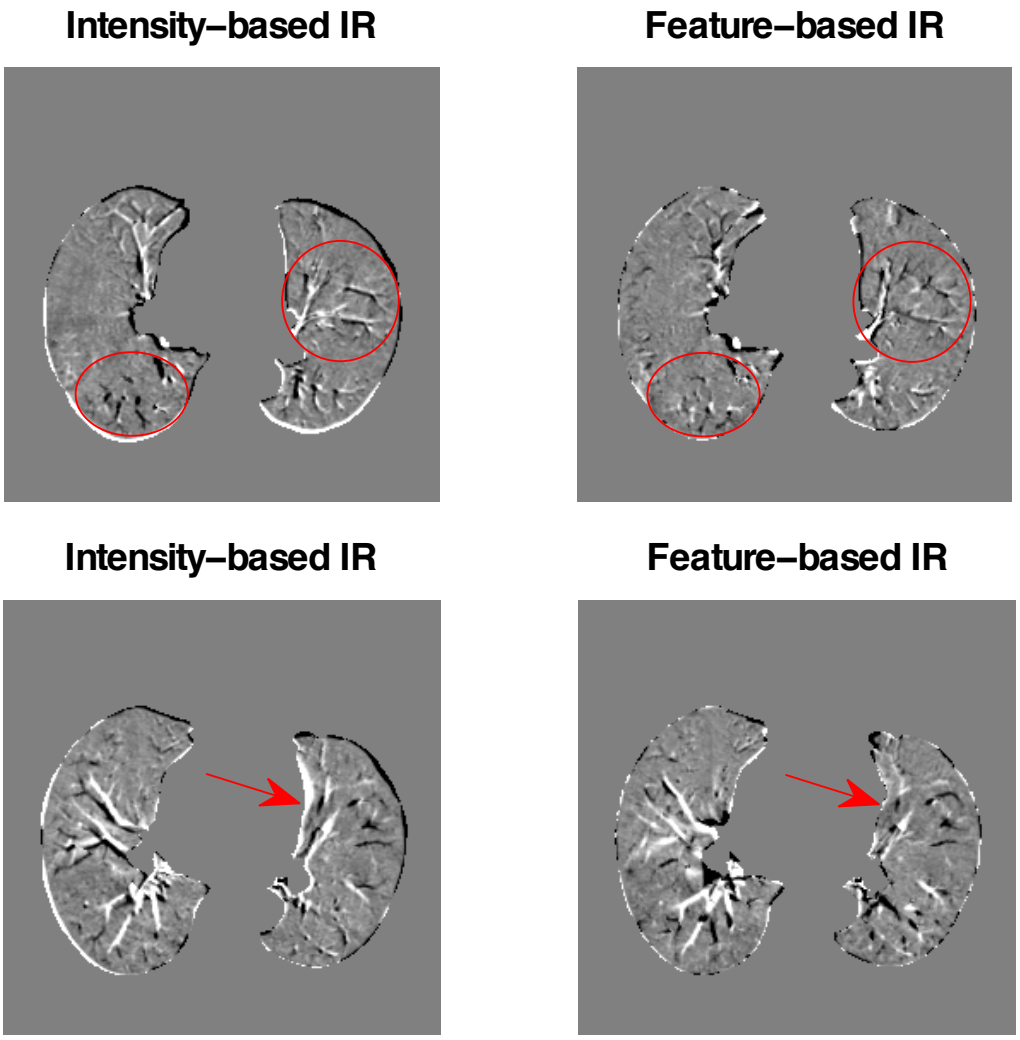


Figure 4.5: Visual comparison of the combined feature-based and intensity-based registration methods. Slice cuts from the difference image between fixed and deformed image for the intensity- and combined feature-based registration methods were extracted on the same level as the plane cuts in Fig. 4.4(a). In general, the currents-based registration aligns the vessels and lung surface better, as can be seen in the areas indicated with the red circles and arrows.

4.6 Analysis of endocast growth of bonobos and chimpanzees

In this section, we want to use the regression model to estimate a typical growth scenario of the endocasts of two species of the genus *Pan*: bonobos (*Pan paniscus*) and chimpanzees (*Pan troglodytes*), which are the two human closest living relatives. According to phylogenetic studies [Won 2005], these two species share a common ancestor, which used to live at least one million years ago. By comparison, the common ancestor of humans, chimpanzees

	Bonobos	Chimpanzees
infant	4	2
child	8	6
child/ young juvenile	3	4
young juvenile	11	10
old juvenile	7	13
sub-adult	9	10
adult	18	14
TOTAL	**60**	**59**

Table 4.2: Number of samples for each species and for each dental age.

and bonobos used to live at least ten million years ago.

From an evolutionary point of view, we expect that the analysis of the difference between the skulls of chimpanzees and bonobos would give a better insight into the possible morphological variations induced by two millions years of evolution. Focusing on skull *growth* should also help to better understand the relationship between the morphological traits and the behavioral differences between both species. Indeed, body growth is considered to be retarded in bonobos compared to chimpanzees, a feature corresponding to their delay in motor development during the first years of postnatal life [Kuroda 1989]. Bonobos appears as "juvenilized" versions of chimpanzees. The adult bonobo skull shows a decreased facial prognathism and teeth with a reduced sexual dimorphism [Shea 1989]. Bonobos are also characterized by a longer dependency of the child on the mother [de Waal 1995, Kano 1992]. Eventually, in evolutionary studies, the human traits are usually compared with the ones of chimpanzees. Such comparisons could be strengthen by a comparison with bonobos traits, provided that the difference between the two species would be better understood.

The main issue for such skull growth comparison is the lack of reliable data. Whereas chimpanzees have been studied quite intensively, data on brain growth in bonobos are totally inadequate. As part of the collaborative project `ARC 3D-Morphine`, we used a database of endocast of 59 chimpanzees and 60 bonobos. The endocast is a mould of the endocranium. The surface of this mould provides a replica of the inner surface of the skull. In this study, CT-scans of dry skull representing wild-shot individuals with approximately equal numbers of male and female have been acquired. The 3D-images have been segmented, so that the endocasts are given as a surface meshes, as illustrated in Fig. 4.6 and 4.7.

The analysis of the teeth emergence of the skulls provide an estimate of the age of the samples, which we call here "dental age". It has been observed in [Kinzey 1984] that the sequences teeth emergence in bonobos and chimpanzees are essentially identical. As a consequence, each skull is associated to one the 6 dental ages defined in [Shea 1989]: infant, child, young juvenile, old juvenile, sub-adult and adult. To refine the classification, some skulls have been associated the intermediate class 'child/young juvenile' by the experts. The number of samples within each class is detailed in Table 4.2. Our purpose is to perform a temporal shape regression of the endocasts of each species with respect to the dental age.

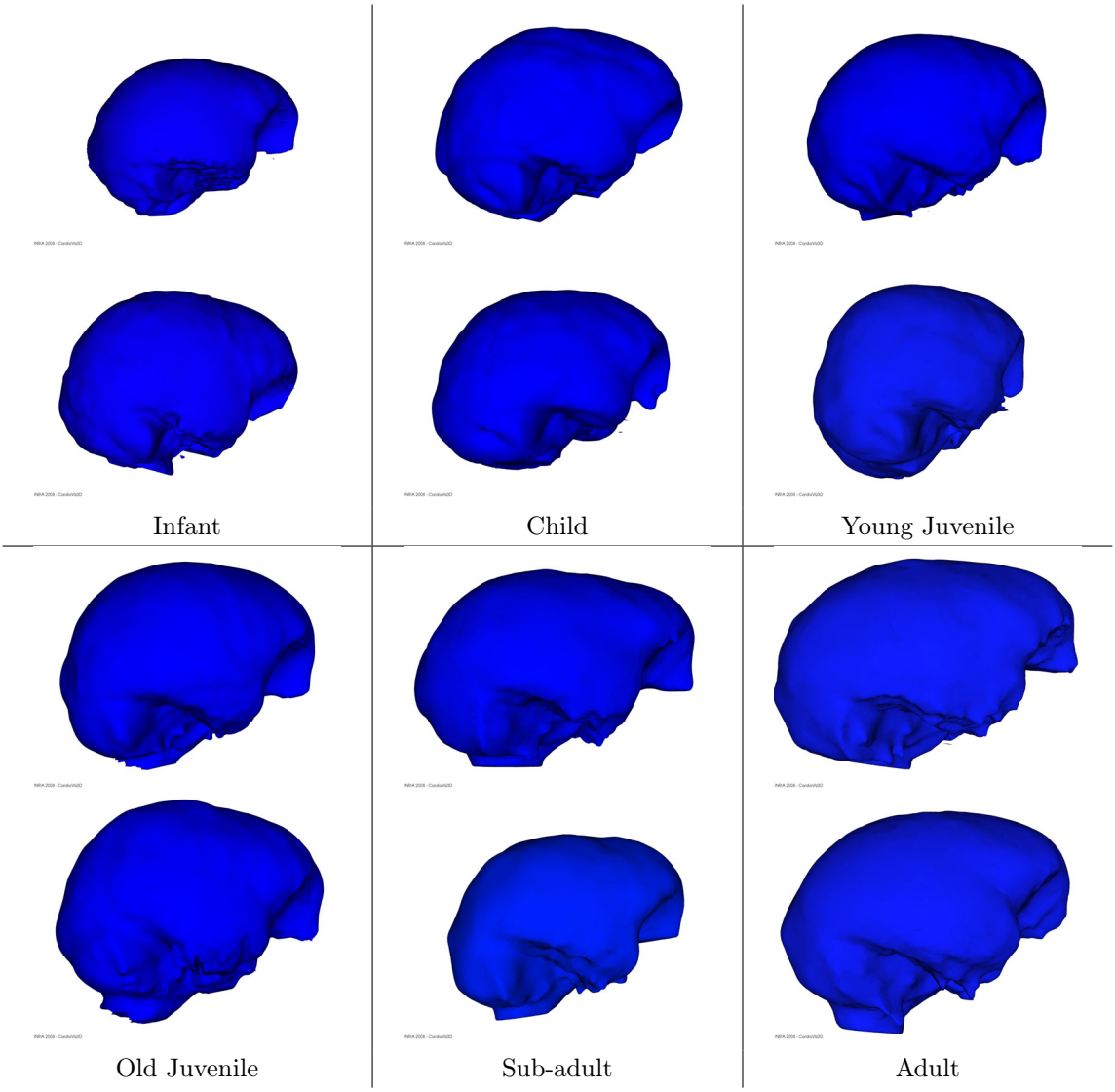


Figure 4.6: Endocasts of Bonobos associated with their respective dental age. 12 samples among 60.

4.6.1 Temporal regression of endocasts

To apply the temporal regression framework introduced in this chapter, we assume that each of the dental ages last the same amount of time. We divide each of the 6 ages into 5 time steps. As a consequence, the period between birth and adulthood has been divided into 30 time-steps.

For each species, we have n_i endocasts for each dental age t_i. We denote these meshes $S_{t_i}^k$ for $t_i = 5, 10, 15, 20, 25, 30$ and $k = 1, \ldots, n_i$. We choose the smallest endocast within the child class as a baseline and associate it to the time point $t = 2$. We denote this baseline S_0. We perform therefore a regression of the endocast with respect to the dental age by minimizing the cost function:

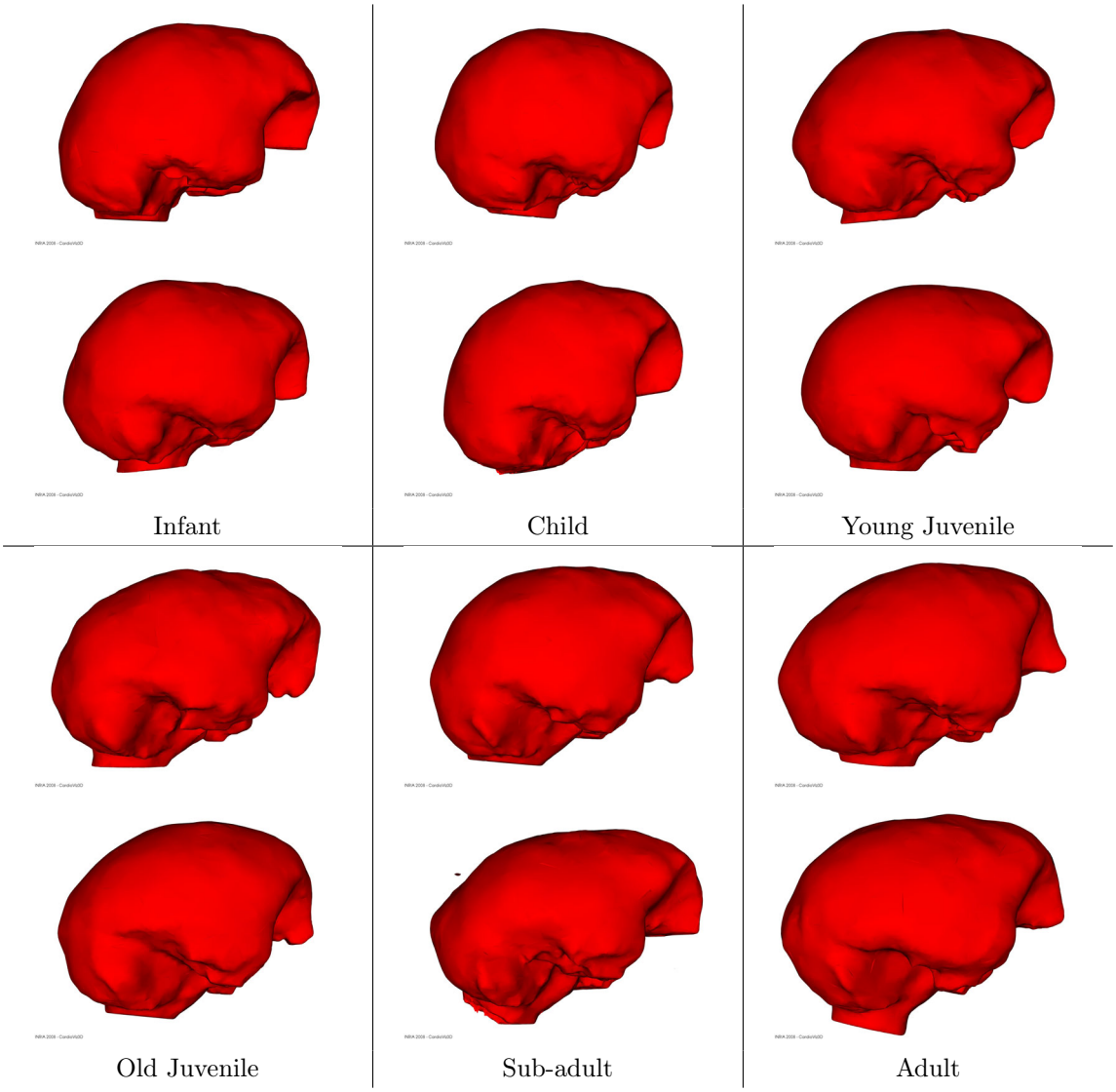


Figure 4.7: Endocasts of Chimpanzees associated with their respective dental age. 12 samples among 59.

$$J\left((v_t)_{t\in[2,31]}\right) = \sum_{\substack{t_i=5,10,15,20,25,30 \\ k=1,\dots,n_i}} \left\| {\chi_{t_i}^{(v_t)}}_* S_0 - S_{t_i}^k \right\|_{W^*}^2 + \gamma_\chi \int_2^{31} \|v_t\|_V^2 \, dt, \tag{4.6.1}$$

where W^* is the space of currents associated to the RKHS with Gaussian kernel of variance λ_W^2 and V the RKHS of vector fields with Gaussian kernel of variance λ_V^2. χ_t^v is the continuous flow of diffeomorphism within time interval $[2, 31]$ which results from the integration of the time-varying vector field v_t in the RKHS V.

4.6.2 Experimental results

We set the typical spatial interaction between currents $\lambda_W = 20$ mm, the spatial scale of deformation consistency $\lambda_V = 50$ mm and the trade-off between fidelity-to-data and regularity $\gamma_\chi = 10^{-4}$ units of time. The diameter of the endocasts are typically between 60 and 70 mm. For the integration of the ODE, the time interval has been discretized into 31 time-steps.

In Figure 4.8, we show the result of the temporal regression for each species. This growth scenario reveals that the endocast growth is anisotropic, with a main elongation in the posterior/anterior part of the skull and a slight diminution in the superior/inferior part. As a consequence the skull which has an almost spherical geometry at birth becomes more and more elongated (ellipsoidal shape). These two typical growth scenarios seem to differ a lot from infancy to childhood. This is mainly due the lack of data in infancy. We have only two samples of infant chimpanzees which have a large skulls compared to the infant bonobos and compared to child chimpanzees. To have a more precise idea of the endocast growth of chimpanzees in infancy, we expect to scan more chimpanzees skulls in the future.

From this geometrical regression of shape, we can deduce an estimation of the evolution of the endocranial volume during growth for each species. The result is shown also in Figure 4.8. It reveals that this shape regression also performs a regression of the endocranial volume. However, the regression proposed here is driven by the shape of the observations, not only by their volume. One intriguing feature is the apparent decrease in endocranial volume for bonobos at sub-adulthood. This feature is also visible when looking at the endocranial volume distribution in the original data used to compute the regression (see graph in Fig. 4.8): the mean of the volume at sub-adulthood is smaller than the one of old juveniles. However, the Mann-Whitney U test gives a p-value of 0.47 when comparing the volume distribution of old juveniles and sub-adults: the median of the two distributions are not proved to be statistically different. The same conclusion applies when comparing the distribution of sub-adults with the one of adults. Actually, the test run for every pair of consecutive distributions shows a significant increase of volume only between infancy and childhood for bonobos (p-value 9.10^{-3}) and between sub-adulthood and adulthood for chimpanzees (p-value 0.02).

These experiments show how the proposed methodology can be used to address practical questions. However, the results are still to be strengthen before being interpreted. We are mostly wondering how far the estimated growth scenario is influenced by the inter-individual variability due to the small amount of samples. Future work should focus on determining confidence interval of the estimated scenario. Permutation testing should also help to detect the presence of outliers and the robustness of the estimated scenario with respect to the choice of the samples.

The impact of the choice of the baseline also must be investigated. Here, we choose the smallest endocast within the group of children as baseline, since this endocast is likely to be similar to the endocast of an early infant. This choice, however, is arbitrary. We need to investigate whether another choice impact the whole growth scenario or only the scenario at the earliest time-steps.

Eventually, the assumption that each age lasts the same amount of time is arguable, especially for the adulthood which may last much more time than the childhood. This issue can be addressed easily by re-parameterizing the time-interval. Such time re-parameterization will be precisely introduced in Chapter 9. One must be aware, however, that changing the duration of each ages will not affect the geometrical changes of the estimated growth scenario but only the speed at which these geometrical changes occur. This point will be explained with more details in Section 9.5.2.

Eventually, one would like to compare the two growth scenarios more quantitatively. Chapter 9 will propose a spatiotemporal registration framework which will allow the comparison of two temporal shape evolution.

4.7 Conclusion

In this chapter, we showed how dense diffeomorphic deformations can be estimated from geometrical data modeled as currents, either for the registration of two sets of data or for the temporal regression of a collection of time-indexed shapes. We take advantage of the numerical framework of Chapter 2, to give a fast and robust implementation of the estimation of such deformations. The parameters of this implementation can be set automatically, independently of the input data, while a fixed approximation error is guaranteed. This offers a way to perform *routinely* registration or temporal regression on a large set of data, like for template-to-subjects registration for instance.

Moreover, the resulting deformations belong to a particular group of diffeomorphisms in which the usual operations like inversion, composition or application of the deformation to another data are computed at the cost of an integration of an ODE. This makes this deformation framework adapted to process a large set of diffeomorphisms. The estimated diffeomorphisms are also geodesic or piecewise geodesic. This leads to a tangent-space representation of the diffeomorphisms which is particularly well suited for computing statistics on deformations. The metric on the tangent-space is the one of a RKHS, thus inheriting of the computational tools introduced in Chapter 2 and 3 for processing, compressing and visualizing statistics.

On the one hand, the deformation framework defined in this chapter allows us to process routinely registration and to easily compute statistics on deformations. On the other hand, the previous chapters propose a computational framework to process large collection of shapes and to deal with shape statistics. This offers a way to the estimation of more sophisticated statistical model of shapes which combines both the statistics on deformations and the statistics on shapes. The definition and estimation of such a model is precisely the purpose of the forthcoming chapters.

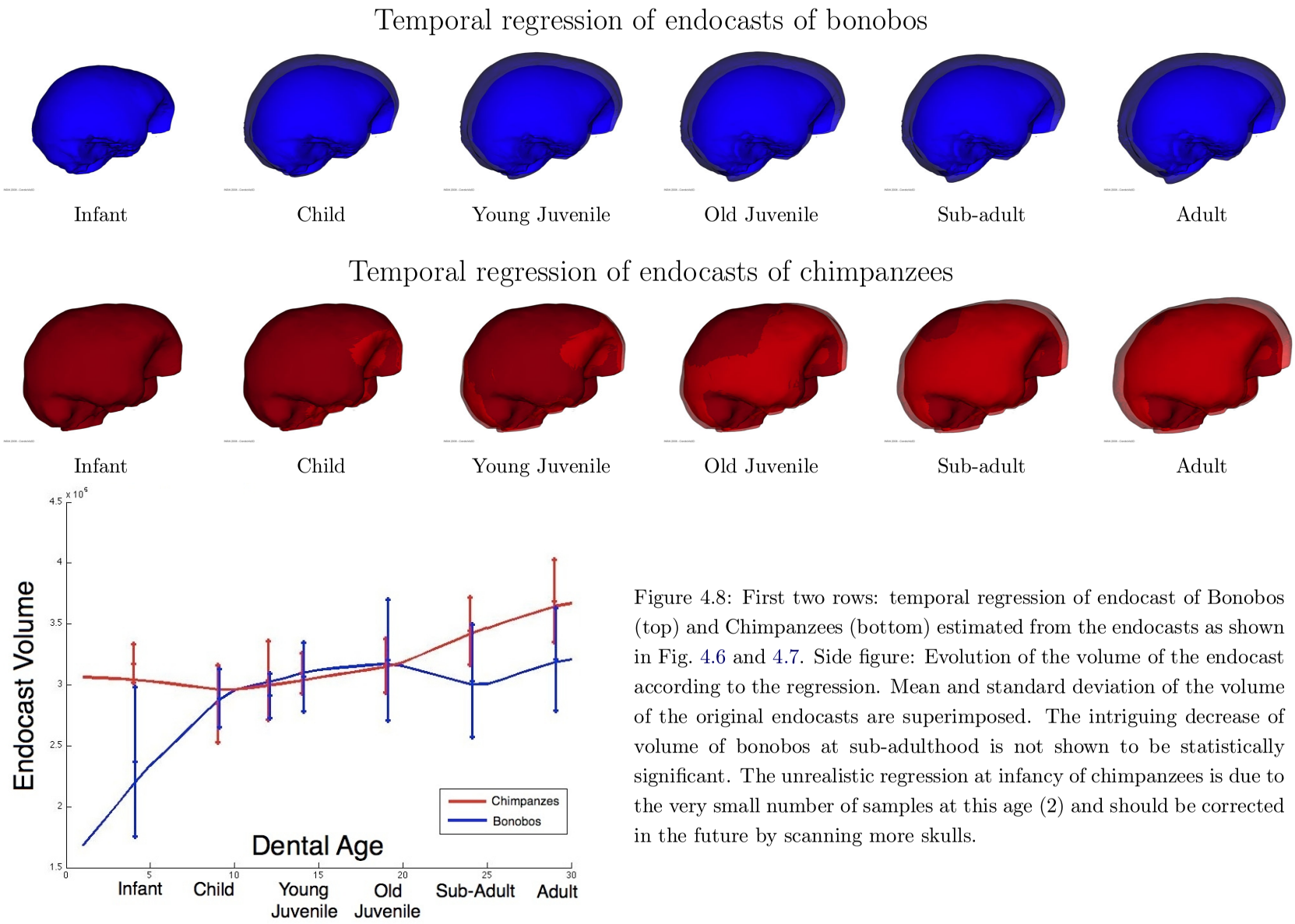


Figure 4.8: First two rows: temporal regression of endocast of Bonobos (top) and Chimpanzees (bottom) estimated from the endocasts as shown in Fig. 4.6 and 4.7. Side figure: Evolution of the volume of the endocast according to the regression. Mean and standard deviation of the volume of the original endocasts are superimposed. The intriguing decrease of volume of bonobos at sub-adulthood is not shown to be statistically significant. The unrealistic regression at infancy of chimpanzees is due to the very small number of samples at this age (2) and should be corrected in the future by scanning more skulls.

Part II

Statistical models of sets of curves and surfaces

Chapter 5

Atlas construction for the measure of anatomical variability

This chapter has been partly published in [Durrleman 2008a, Durrleman 2009c]

Contents

5.1 From shape statistics to atlas estimation

The primary goal of Computational Anatomy is to understand how anatomical structures vary in a population. This statistical analysis may be used to classify populations, to

find discriminative features between subgroups (e.g. pathologic vs. control) or as a prior to drive the segmentation of anatomical structures in new images, like in [Yeo 2008a] for instance. Given a set of shapes, the question is to find a common prototype shape and to describe how this prototype varies in the studied population.

In the previous chapters, we focused on the modeling of geometrical data based on currents. The purpose of this chapter is to define generative statistical models which combine the diffeomorphic deformations of Chapter 4 with the statistics on currents introduced in Chapter 3. Such models will allow us to root the variability measures of anatomical structures into rigorous statistical estimations.

Several frameworks have been already proposed in the medical imaging field to build atlases, namely the estimation of a mean anatomy using registration like in [Joshi 2004, Avants 2004, Marsland 2004a, Chui 2004, Zollei 2005] for instance. Most of these methods deal with images and not with shapes. Moreover, they are not always based on a rigorous statistical model. This makes the use of such templates in a statistical context particularly difficult. In this chapter, we propose a generic method for constructing atlases which builds on similar ideas but roots them into a consistent framework. The expected benefits are twofold. First, we argue that statistics based on a rigorous model are much more likely to highlight significant features, as this will be illustrated in Fig. 5.8 for instance. Second, such an approach makes explicit the assumptions on which the model is based, thus making much clearer the impact of each parameter and therefore the interpretation of the results.

5.1.1 A Gaussian shape model

As shown in Chapter 1, the space of currents is a Hilbert space: it is a vector space and is provided with an inner-product. This allows us to define directly statistics of a set of shapes such as mean and principal modes of variations. Indeed, given a set of N shapes $(T_i)_{1\leq i\leq N}$, we can define the empirical mean as $\overline{T} = \sum_i T_i/N$ and the empirical covariance matrix Γ whose elements Γ_{ij} are $\left\langle T_i - \overline{T}, T_i - \overline{T}\right\rangle_{W^*}$. The eigenvectors of this matrix enables to compute the principal modes of variations as a linear combination of the input shapes: $m = \sum_i w_i T_i$ (for some weights w_i given by the coordinates of the eigenvectors of the sample covariance matrix Γ: see Remark 5.9 page 183 for more details). These statistics (mean and modes) are defined in the space of currents. We can use the Matching Pursuit Algorithm introduced in Chapter 3 to give a sparse representation of these statistics which leads to interpretable results.

These computations rely on the assumption that the input observations T_i are generated as random Gaussian perturbations of an unknown mean shape $\overline{T}$. Mathematically speaking, this statistical model is written as: $T_i = \overline{T} + \varepsilon_i$ where ε_i are independent and identically distributed zero-mean Gaussian random variables (introduced in Section 1.5.3 in the continuous setting and in 2.3.3 in a discrete setting). The sample mean and the principal modes computed in the space of currents leads to an estimation of the true mean shape $\overline{T}$ and of the covariance matrix of the underlying Gaussian law of the variables ε_i.

This statistical model is simple and its parameters (mean and covariance) can be inferred directly from the actual observations using only the tools previously introduced (computation of linear combination of currents, inner-product between pair of currents and matching

pursuit to have compact and interpretable representation of the statistics).

However, this model used 'as is' has little chance to highlight interesting geometrical features. Indeed, the estimated statistics can be corrupted by some large-scale transformations, such as translation or rigid-body transformations between observations. These large-scale variations bias the statistics and are very likely to hide the true anatomical variability. This is the reason why one would like first to align the input observations (so that non meaningful large-scale effects will be discarded) and then perform statistics on what remains after registration. This is the strategy proposed in voxel-based morphometry [Ashburner 2000, Ashburner 2001, Good 2001] and in the so-called "statistical shape models" [Cootes 1995, Davies 2002, Davies 2008], in the vein of Kendall's theory of shapes [Kendall 1989].

Such a statistical analysis makes a strong assumption on the nature of the transformations which do not carry meaningful anatomical information and which should be discarded. Whereas it is clear that translations is a pure artifact of the acquisition protocol and the choice of the origin, the question of the size (which can be discounted by scaling) is much more subtle. It is often very unclear whether the pathology which one wants to characterize does or does not affect the size the structures to be analyzed. In one case, the effects of the pathology of the organ may disappear if we scale the input shapes. In the other case, if we keep the original size of the shapes, the informative variations may be hidden by the uninformative variations of the size. The question is even more trickier for affine transformation or for general diffeomorphisms. As a consequence, it is not reasonable to decide *a priori* whether the deformations does or does not contain meaningful anatomical information. Computing statistics only on the residual that remain after registration relies on arguable assumptions.

5.1.2 The importance of deformations for statistics

Conversely, Grenander's approach, which inspired later the tensor-based morphometry [Toga 1999, Gaser 1999], proposes to characterize the differences between two shapes with the deformation of one shape to the other [Thompson 1917, Grenander 1994]. Indeed, it has been shown that the deformations of a common template to each input shapes is able to capture significant information and give some insights about effects of pathologies on the anatomy [Csernansky 1998, Thompson 2004, Marsland 2004b, Narr 2007]. However, the statistics of shapes which rely on the deformations only is not satisfactory either, since such statistical estimations depend also on the choice of the type of deformations. Deformations with very few degrees of freedom, like rigid-body transformation for instance, will not be able to capture some local variations of the anatomy. On the contrary, one can consider a much larger set of deformations, like general diffeomorphisms with many more degrees of freedom. With such deformations, one can expect to warp one shape onto the other with a great accuracy, meaning that the residual which remain after registration becomes very small: the deformation capture all the differences between the two shapes. However, in this case, the deformations are likely to over-fit the observations and the analysis of the set of deformations is likely to detect artifacts rather than common anatomical variations across the population. There is therefore a trade-off to find between the accuracy of the matching

and the regularity of the deformation. The accuracy should be great enough (resp. the deformation should be irregular enough) to capture meaningful anatomical variations, but small enough (resp. the deformation regular enough) to highlight common features in the population and to prevent over-fitting.

Nevertheless, even if we are able to define an optimal trade-off between matching accuracy and regularity of the deformation, the statistics should not be performed on the deformations only. The optimal deformation has some regularity and, as such, has some constraints. For instance, diffeomorphisms are not able to capture topological changes, matter creation or deletion, etc. Some significant anatomical information may be contained in the residuals which remain after registration.

5.1.3 A joint model: deformations and residuals

As a consequence, the optimal trade-off between matching accuracy and regularity of the deformation should be considered as an optimal decomposition of the anatomical variability into, first, a geometrical part which is captured by smooth deformations and contained variations like torque, stretching, shrinking, etc. and, second, a "texture" part which contains everything that is not captured by the deformations: uninformative noise but also meaningful change of topology for instance. The statistical analysis of the anatomical variability should focus jointly and consistently on both the deformations and the residual shape. This recalls the image decomposition into geometry (contours, homogeneous area, etc..) and photometry (illumination, texture, etc..) in the field of image processing [Meyer 2001].

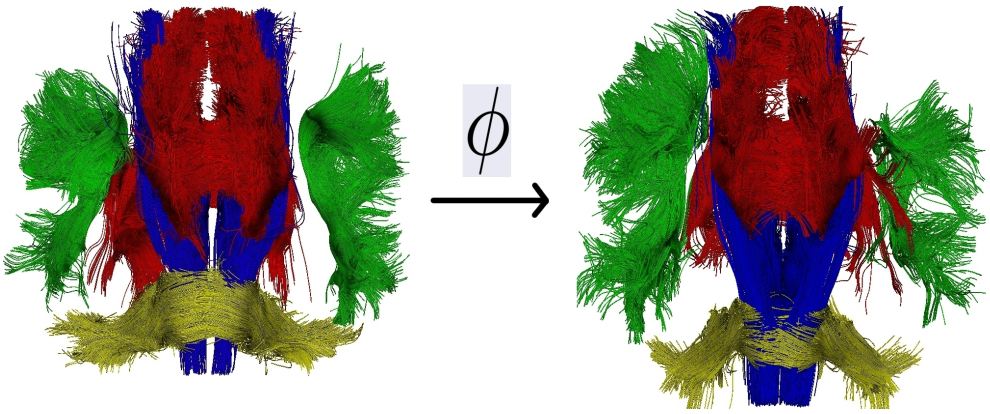
To combine the two different point of views, we propose in this chapter a new statistical model for shapes: each observation T_i is considered as an instance of a random deformation of an unknown template plus a random perturbation (in the space of currents). A Maximum A Posteriori (MAP) approach enables to estimate consistently the unknown template, the deformations of this template to each observations and the residual perturbations. Statistics may be performed via the tangent-space representation of the diffeomorphisms for the deformation (see Chap. 4) and in the space of currents for the residues. In the sequel, we call "atlas" the set of (1) the estimated template, (2) the deformations and (3) the residues jointly inferred from a population of shapes.

This joint model contains the two previous models as extremal cases. Highly constrained deformations leads to large residuals which contain almost all the variability. By contrast, deformation with many degrees of freedom gather all the variability in the deformation part and leave almost nothing in the residual. However, we believe that the best decomposition (in a precise statistical sense to be defined) is a trade-off between these two extremal cases. This chapter is precisely a first step toward the automatic estimation of such a trade-off.

It is clear that this decomposition in terms of geometry and texture depends on the parameters of the statistical model, namely the maximum number of degrees of freedom of the deformations, the trade-off between regularity and accuracy during registration and the choice of the metric in the space of currents. In this work, the atlas is estimated once these parameters are fixed. However, a more general MAP estimation would consist in the estimation of these parameters along with the estimation of the atlas. This would be a way to define the optimal decomposition into geometry and texture. We leave such an

estimation for future work.

This joint model has the advantage to take all information into account. It does not make strong assumptions on the nature of the variability one is looking for. The subsequent statistical analysis of the deformations and the residues will determine where the most significant differences occur. This model is generative: once the atlas is built, one can generate new instances according the estimated variability model. This helps to better interpret the variability that the model captured. This allows us also to compare new observations with the estimated variability and, for instance, detect pathologies as large deviations from the normal variability.



$$\mathcal{O}_2 = \phi_* \mathcal{O}_1 + \varepsilon$$

Figure 5.1: The registration tries to find the best smooth deformation (ϕ) which maps the anatomical structures on the left ($\mathcal{O}_1$) to those on the right ($\mathcal{O}_2$). This deformation captures the geometrical differences between the anatomies (torque, elongation, shrinking, etc.) whereas it cannot capture non-diffeomorphic differences such as a difference of the density of fibers within each bundle, which are left in the residual shape (ε). Actually, the registration decompose the variability into a geometrical part and a "texture" part. Anatomical statistical models should account for both kind of variability.

5.2 Forward versus backward atlas construction

5.2.1 Lagrangian or Eulerian noise?

In this section, we want to give a precise mathematical definition of our statistical model. According to the previous section, this model considers each input shapes (the observations T_i) as a random deformation of an unknown template plus a random perturbation. The deformations capture the smooth geometrical variations of the template in the population. The residual perturbation models everything that cannot be captured by smooth deformations like change of topology or change of matter density for instance, and also like physical

or numerical noise.

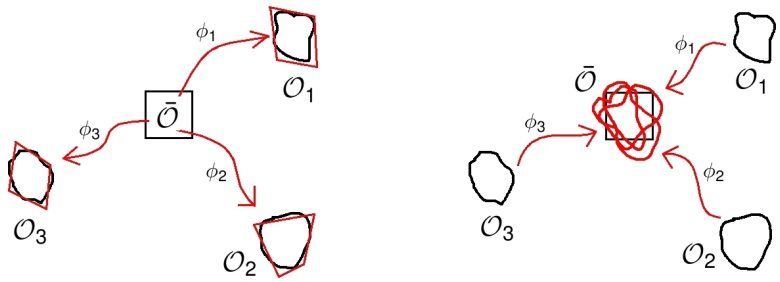
However, there are at least two ways to define such a statistical model. A common practice in the field of medical imaging is to co-register every observation into a common reference frame, the unknown space of the template and then perform statistics in this common space. This leads to the definition of a *backward* model (see Fig. 5.2): the template is equal to a deformation of the observation up to a random perturbation. In other words, the observations result from a random deformation of the template which have been *previously* perturbed by a random shape variation (in the space of currents). Mathematically, this model can be written as:

$$\phi_{i*}T_i = \overline{T} + \varepsilon_i \quad \Longleftrightarrow \quad T_i = {\phi_i^{-1}}_* \overline{T} + {\phi_i^{-1}}_* \varepsilon_i \tag{5.2.1}$$

where $\overline{T}$ denotes the unknown template, ϕ_i the random deformations and ε_i i.i.d. random variables in the shape space.

An alternative model, as pioneered in [Allassonnière 2007, Allassonnière 2008, Ma 2008], considers that the observations result from a random deformation of an unknown template, perturbed *afterwards* by a random shape variation (see Fig. 5.2). Mathematically, this model, which is called *forward* model, can be written as:

$$T_i = \phi_{i*}\overline{T} + \varepsilon_i \tag{5.2.2}$$



forward scheme backward scheme

Figure 5.2: In the forward scheme, the physical observations ($\mathcal{O}_i$) are seen as noisy deformation (ϕ_i) of unknown template ($\overline{\mathcal{O}}$). In the backward scheme, the template is an average of deformed observations. In the forward scheme the noise is removed from the observations whereas it is pulled back in the common frame with the backward scheme.

In the *backward* model, the perturbation ε_i is identically distributed in the *Lagrangian* coordinates. It is added in the coordinate system of the template, which moves under the action of the deformation. The perturbation is seen as an intrinsic property of the template. Writing this model in the fixed Eulerian coordinate (the coordinate system in which every observation is given), as in Eq. (5.2.1)-right, we see that the perturbations in this coordinate system are written as $\phi_{i*}\varepsilon_i$ and therefore depend on the deformation ϕ_i: these variables are no more *identically* distributed among the subjects but depend on the subject-specific deformation. The backward model has been explicitly formalized in such a statistical way

in [Glasbey 2001]. Several atlas construction methods rely implicitly on this model, like in [Glaunès 2006, Sabuncu 2008] for instance.

By contrast, in the *forward* model, the perturbation ε_i is identically distributed in the *Eulerian* coordinates. The noise is added in the image space (i.e. the space of the acquisition), after applying the deformation which carries the shape from the template space to the subject's space. The perturbation is considered as an intrinsic property of the subject itself.

In more concrete words, in the *backward* scheme, one needs to compute the dissimilarity $\left\|\phi_{i*}T_i - \overline{T}\right\|$. In the *forward* scheme, one needs to compute the dissimilarity $\left\|T_i - \phi_{i*}\overline{T}\right\|$. The two approaches would be equivalent if the norm used were left-invariant with respect to the action of diffeomorphism (i.e. $\|\phi_* T\| = \|T\|$ for any T). However, the usual norms on images or geometrical data do not satisfy this invariance rule. Note that some methods try to combine both approaches in a more symmetric way like in [Sabuncu 2009] for instance, although such methods have not been given a precise statistical sense.

5.2.2 Eulerian perturbations are required to account for noise

There are several reasons which explain why a perturbation attached to the subject's space is required. From a physical point of view, the noise which is added by the sensors is defined in the coordinate system of the subject and there is no reason for this noise to depend on a deformation which have been introduced for modeling purposes and which has nothing to do with the acquisition process.

In the proposed statistical models, the template $\overline{T}$ is supposed to be an "ideal" representation of the continuous underlying biological material (a L^2 function defined in the continuous domain as an ideal image for instance). The observations, by contrast, are given intrinsically as discrete and sampled objects. The difference between the continuous deformed template and the discrete observations is a perturbation, which can be called "sampling noise". Again, this sampling noise has the same law for every subject and has no reason to depend from an unobserved deformation. In other words, in the backward model, the sampled observations are deformed back to the template space: an extrinsic extrapolation scheme is required to compare this sampled deformed observations to the continuous template. By contrast, in the forward model, the continuous template is deformed (and the deformations act naturally on continuous objects) and then needs only to be sampled to compare the observations.

To summarize, in order to model the physical and the sampling noise of our data, we need to account for an i.i.d. perturbation in the Eulerian coordinates as in the forward model. By contrast, there is no consensus whether the possible anatomical variations, like a change of density of white matter fiber bundles for instance (see Chap. 7), are better described as i.i.d. perturbations in the template space or in the subject's space. In practice, modeling this variability by a unimodal Gaussian variable is probably a much stronger limitation for an accurate estimation of the true underlying variability than the possible impact of the deformation on the law of these perturbations.

5.2.3 Toward a mixture model

The proponents of the backward model claim that the perturbations defined in the template space may model some modes of variations which may be related to different pathologies (or different subtypes of a pathology). For instance, a change of topology may characterize a pathology with respect to a control case (one versus two connected components for instance). In this case, it would be more natural to describe this effect as an i.i.d. perturbation in the template space and not in the subject's space where the change of topology in the deformed configuration would be less accurately described by a single i.i.d. variable.

We can workaround this issue in two different ways. First, we can define a combined statistical model as $T_i = \phi_{i*}(\overline{T} + \eta_i) + \varepsilon_i$ where two i.i.d. perturbations, one in each coordinate system, are taken into account: η_i and ε_i. The main problem of such a model is to find a way to constraint the decomposition of the residual variability in the two terms η_i and ε_i. It is not clear, even intuitively, what this decomposition should be on some simple examples. This idea would lead to the estimation of an atlas in the context of the metamorphoses, where changes of photometry is allowed along with the geometrical deformations [Trouvé 2005b].

A more direct way would consist in estimating not only one template but several templates, each one modeling a different pathologies or subtypes of a pathology. In [Allassonnière 2009], the estimation of an atlas with several components (the number of components being not imposed *a priori*) is proposed in the forward setting for images and small deformations (displacement fields instead of diffeomorphisms). Extending this algorithm to account for large deformations and for currents should be investigated in the future.

In [Hufnagel 2008], a model with two perturbations is proposed. Deformations are supposed to be affine and only the perturbation in the Lagrangian coordinates is estimated. The authors claim that the modes of variations in the template space enables to describe the different types of variations which occur in a population. In this case, we believe that these modes are needed because the affine transformation has not enough degrees of freedom to describe all possible anatomical variations. By contrast, deformations with many more degrees of freedom, like diffeomorphisms for instance, should be able to capture the anatomical variations without introducing modes of variability of the template. In this case, statistics on deformations would retrieve the different types of pathologies. In Section 5.5, we will show on a simulated example how statistics on diffeomorphisms can be used to classify two populations in a forward model (without introducing any modes of variation in the template space). We believe that Lagrangian perturbations are justified only if the difference between types cannot be captured by the deformations allowed in the model.

5.2.4 Forward model is better adapted to statistical inference

In practice, we use a training set of several subjects to build an atlas and then we use this atlas to compare new subjects with the estimated variability model. The training step consists in estimating the template, the deformations of the template and the residuals. A

model of variability (i.e. mean and modes for instance) is estimated for both the deformations and the residual perturbation. The test step consists in computing the likelihood of any new subjects with respect to the estimated variability model, for diagnosis purpose for instance. In this section, we assume that we can define probabilities on objects T (images, curves, surfaces, etc.) and on deformations ϕ.

The construction of the atlas requires at least to compute the probability of having the template given a training database of T_i: $p(\overline{T}|T_i)$. Once the atlas is built, one would like to know how a new observation T_{new} is compared to the estimated variability model: one needs to compute the likelihood of this observation given the template $p(T_{\text{new}}|\overline{T})$. Because ϕ_i acts differently in the forward model (Eq. (5.2.2)) and in the backward model (Eq. (5.2.1)), the computational cost of these two steps varies significantly. In the backward scheme, computing $p(\overline{T}|T_i)$ is much simpler than computing $p(T_{\text{new}}|\overline{T})$ which depends on the Jacobian of the deformations ϕ_i. It is exactly the reverse for the forward scheme: computing the atlas is more difficult than to compare a new observation to the estimated variability. We argue that it is better to spend more time to build the atlas (which is done once for all) and to keep simple the test of any new available data: the forward model seems better suited even from a computational point of view. The backward scheme seems simpler if we do not compute joint statistics on the deformation ϕ and the residual perturbation ε.

From a theoretical point of view, the forward model is also better understood. For instance, the consistency of the Maximum A Posteriori (MAP) template estimation, when the number of available observations is growing, is proved for images and small deformations in [Allassonnière 2007]. Such proofs for the backward model seem currently out of reach.

For all these reasons, we base here our statistical estimations on the *forward model*. In the sequel, we show how the atlas building step, which is the most critical step in this paradigm, is possible in case of curves and surfaces. For this purpose, we take advantage of the Matching Pursuit algorithm for currents, introduced previously in Chapter 3. Numerical simulations on simulated examples will also show how the atlas may be used for group classification in the context of supervised learning. The next chapters will show how this atlas construction method can be used to describe the anatomical variability in real applications.

5.3 Joint estimation of the template, the deformations and the residues

From this section onwards, we focus on the atlas estimation in a forward setting when the observations are given as geometrical data. We model these data as currents and derive the equations for the construction of the atlas.

5.3.1 A Heuristic Maximum A Posteriori in infinite dimension

In Eq. (5.2.2) ($T_i = \phi_{i*}\overline{T} + \varepsilon_i$), T_i are the observations, $\overline{T}$ is unknown, ϕ_i are hidden variables and ε_i are independent and identically distributed Gaussian variables with known variance σ_W^2. To define statistics on deformations, we take advantage of the tangent-space

representation of the diffeomorphisms. In Chapter 4 (Section 4.2.2.3), we showed that the geodesic diffeomorphisms are completely determined by their initial speed vector fields v_0 in a RKHS V, in case of the dual representation of v_0 has a discrete support (then the discrete momenta follow the Euler-Lagrange equations in Eq. (4.2.22)). Let us assume in this section that this property holds for any initial vector field $v_0 \in V$, namely that every geodesic diffeomorphisms have a tangent-space representation.

Let us assume that we can define Gaussian probability density functions (pdf) on the space of Currents W^* and on the space of initial vector fields V: $p_\varepsilon(\varepsilon) = C_\varepsilon \exp(-\|\varepsilon\|^2_{W^*}/2\sigma_W^2)$ and $p_\phi(v) = C_\phi \exp(-\|v\|^2_V/2\sigma_V^2)$. In that case, a Maximum A Posteriori (MAP) estimation for independent observations and uniform prior on $\overline{T}$ maximizes $\prod_{i=1}^{N} p(T_i|\overline{T})$ over $\overline{T}$. Formally,

$$\begin{aligned} p(T_i|\overline{T}) &= \int p_\varepsilon(T_i|\overline{T}, v_0^i) p_\phi(v_0^i) dv_0 \\ &= \int p_\varepsilon(T_i - \phi^{v_0^i}{}_* \overline{T}) p_\phi(v_0^i) dv_0. \end{aligned} \tag{5.3.1}$$

Since the term within the integral depends on v_0 by a geodesic shooting of diffeomorphisms, there are no closed forms for this likelihood. A usual approximation consists in replacing the integral by the maximum of the distribution within the integral (i.e. its first mode). This means that $p(T_i|\overline{T})$ is replaced by a Dirac measure located at $\arg\max_{v_0^i} p_\varepsilon(T_i - \phi_i^{v_0^i}{}_* \overline{T}) p_\phi(v_0^i)$. This approximation, called Fast Approximation with Mode (FAM), leads to:

$$\left(\overline{T}, \left(\phi^{v_0^i}\right)_{1\leq i\leq N}\right) = \underset{\overline{T}, (v_0^i)_{1\leq i\leq N}}{\arg\min} \left\{ \frac{1}{\sigma_W^2} \sum_{i=1}^{N} \left\| T_i - \phi^{v_0^i}{}_* \overline{T} \right\|^2_{W^*} + \frac{1}{\sigma_V^2} \sum_{i=1}^{N} \left\| v_0^i \right\|^2_V \right\}. \tag{5.3.2}$$

However, the Gaussian variables have no pdf in infinite dimensional Hilbert spaces, such as the space of currents W^* and the space of initial vector fields V. In the next section, we will give a rigorous sense of this heuristic MAP derivation, as the limit of the likelihood derived in a finite-dimensional setting using the approximation spaces introduced in Chapter 2.

Note also that the approximation with mode (FAM) could be avoided. For instance, a sampling of the posterior can be estimated by Markov Chain Monte Carlo (MCMC) approaches, as shown in [Allassonnière 2009] in case of images and small deformations. Extending this work in case of diffeomorphisms and currents is left for future investigations.

5.3.2 A rigorous MAP derivation using approximation spaces

5.3.2.1 A finite-dimensional statistical model

In Chapter 2, we introduced finite-dimensional approximation spaces of currents. In particular, we showed in Section 2.3.3 that the trace of the Gaussian currents in such approximation spaces have probability density function. In this setting, we can therefore derive rigorously a MAP estimation. This requires first to define a proper statistical model in finite dimension.

Let V be a RKHS of vector field and W^* the space of currents as the dual space of the RKHS W. Let Λ_0 be a linearly spaced grid of points and $V^*_{\Lambda_0}$ and $W^*_{\Lambda_0}$ the discrete spaces associated to V^* (i.e. the dual space of V) and W^* in the sense given in Chapter 2.

We define the template $\overline{T}$ as a current in the discrete space $W^*_{\Lambda_0}$. We define η_i, a sequence of N i.i.d. Gaussian variables in the discrete space $V^*_{\Lambda_0}$ with variance σ_V^2 (see Section 2.3.3). We denote $v_i = \mathcal{L}_V^{-1}(\eta_i)$ their associated vector fields.

Let ε_i a sequence of N i.i.d. Gaussian variables in the continuous space of currents W^* with variance σ_W^2, as defined in Section 1.5.3. Then, we define the following statistical model:

$$T_i = {\phi^{v_i}}_* \overline{T} + \varepsilon_i. \tag{5.3.3}$$

In this model the observations T_i, the deformed template ${\phi^{v_i}}_*\overline{T}$ as well as the perturbations ε_i belong the continuous space of currents W^*, whereas the parameters $\overline{T}$ and (v_i) belong to discrete spaces induced by the *fixed* grid Λ_0.

To infer this statistical model in a discrete setting, we introduce *another* linearly spaced grid Λ with step Δ (which is supposed to tend to zero). Instead of maximizing the likelihood of the observation T_i, we want to maximize the likelihood of its orthogonal projection on W^*_Λ $(\Pi_\Lambda(T_i))$ over the parameters $\overline{T}$ and v_i:

$$\max_{\substack{\overline{T}\in W^*_{\Lambda_0} \\ v_i \in V_{\Lambda_0}}} p\left(\Pi_\Lambda(T_i)\right).$$

As shown in Section 2.3.3, the probability density function of the vector fields v_i is given as:

$$p_\phi(v_i) = p(\eta_i) \propto \exp\left(-\frac{\|v_i\|_V^2}{\sigma_V^2}\right), \tag{5.3.4}$$

since the norm on the approximation space V_{Λ_0} coincide with the norm on V (where $\propto$ means "proportional to").

Similarly, the posterior is given by $p_\varepsilon(\Pi_\Lambda(T_i)|v_i, \overline{T})$ where p_ε denotes the pdf of the Gaussian currents in the discrete space W^*_Λ, as in Section 2.3.3. Therefore, we have:

$$p_\varepsilon\left(\Pi_\Lambda(T_i)|v_i,\overline{T}\right) = p_\varepsilon\left(\Pi_\Lambda\left(T_i - {\phi^{v_i}}_*\overline{T}\right)\right) \propto \exp\left(\left\|\Pi_\Lambda\left(T_i - {\phi^{v_i}}_*\overline{T}\right)\right\|^2_{W^*}/\sigma_W^2\right). \tag{5.3.5}$$

To derive the MAP estimation, we still need to define a prior on the template shape $\overline{T}$. Let $\mathcal{B}_{W^*}(R)$ be the ball in W^* of center 0 and radius R and $\mathcal{B}^{\Lambda_0}_{W^*}(R) = W^*_{\Lambda_0} \cap \mathcal{B}_{W^*}(R)$. This space is the set of discrete currents in $W^*_{\Lambda_0}$ whose norm is smaller than R. This space is isometric to the set of momenta $\boldsymbol{\alpha}$ (pN dimensional vector where N is the number of grid nodes and p the dimension of the momenta ($p = 1$ or $p = 3$)) such that $\boldsymbol{\alpha}^t \mathbf{K}_{\Lambda_0} \boldsymbol{\alpha} \leq R$. This compact and discrete set of currents can be provided with the uniform measure. This is the prior that we choose for $\overline{T}$.

Now, we derive a rigorous MAP estimation of the atlas $(\overline{T}, v_i)$, using the same Fast Approximation with Modes as in the previous section. This leads to the minimization of the log-likelihood:

$$L^\Lambda\left((v_i)_{i=1,\dots,N}, \overline{T}\right) = \sum_{i=1}^N \left\|\Pi_\Lambda\left(T_i - {\phi^{v_i}}_*\overline{T}\right)\right\|^2_{W^*} + \gamma \|v_i\|_V^2, \tag{5.3.6}$$

over the vector fields $v_i \in V_{\Lambda_0}$ and the template $\overline{T} \in W^*_{\Lambda_0}$ such that $\left\|\overline{T}\right\|_{W^*} \leq R$.

The application of Prop. 4.8 about the parameterization of minimizing diffeomorphisms shows that the vector fields v_0 which minimize L^Λ are parameterized by momenta in the support of $\overline{T}$ which is included in Λ_0, since se supposed that $\overline{T} \in W^*_{\Lambda_0}$. Therefore, the fact that the vector field $v_0 \in V_{\Lambda_0}$ needs to be parameterized by momenta in Λ_0 is not a constraint which we need to enforce: it is automatically satisfied at the minimum.

This discussion suggests to use the registration scheme of Chapter 4 to minimize L^Λ with respect to the vector fields v_0. However, this minimization problem is not exactly a registration problem as stated in Chapter 4 due to the presence of the projection operator Π_Λ in the fidelity-to-data term. This minimization problem requires to adapt the registration scheme of Chapter 4 to take into account the projection of the fidelity-to-data term on the grid Λ. Moreover, this projection Π_Λ makes also difficult the minimization of L^Λ with respect to the template $\overline{T}$. Indeed we need to infer the momenta of $\overline{T}$ on the grid Λ_0, knowing only the projection of the deformed template $\phi_*\overline{T}$ on the grid Λ.

By contrast, the heuristic log-likelihood L derived in Eq. (5.3.2) looks similar to the rigorously derived log-likelihood L^Λ, except that it does not involve the projection on the grid Λ:

$$L\left((v_i)_{i=1,\ldots,N}, \overline{T}\right) = \sum_{i=1}^{N} \left\|T_i - {\phi^{v_i}}_*\overline{T}\right\|^2_{W^*} + \gamma \left\|v_i\right\|^2_V . \tag{5.3.7}$$

In this heuristic log-likelihood, we consider also that $\overline{T} \in W^*_{\Lambda_0}$. We consider also that $v_i \in V_{\Lambda_0}$, which is, as discussed above, the parameterization of the vector fields which enable to minimize L. Since L does not involve the discretization grid Λ, it seems much easier to minimized. In particular, as we shall see in the next section, minimizing this heuristic likelihood with respect to the initial vector field can be done using the registration scheme of Chapter 4.

Intuitively, it seems that the rigorously derived likelihood L^Λ tends to the heuristic likelihood L, as the step of the discretization grid Λ tends to zero (i.e. the grid becomes finer and finer). In the next section, we will precisely establish that the minima of the rigorously derived likelihood L^Λ tends to the minima of the heuristic likelihood L. This will justify the minimization of the heuristic likelihood L instead of the rigorously derived one.

Remark 5.1. In this section, we introduced *two* grids:

- A *fixed* grid Λ_0, which define the finite-dimensional space to which the variables of the model (template and momenta of the initial speed vector field) belong.
- A grid Λ, whose step tends to zero, on which we project the observations T_i.

In the next section, we will study the convergence of the likelihood of the observations when the step of the grid Λ tends to zero, namely when the projected observations tend to the original ones.

It could be interesting to also study the convergence of the statistical estimation of the template and the initial speed vector fields when the step of the grid Λ_0 tends to zero. At the limit, the template and initial vector field are not constrained to belong to finite

dimensional spaces anymore: they could be any arbitrary currents and vector fields. The main issue here is that these variables in infinite dimensional spaces do not admit probability density function. To do statistics in such spaces, one cannot use likelihood ratios anymore but probability measures on neighborhoods of the variables instead. Such an advanced statistical analysis is out of the scope of this work but could be worth being investigated in the future. □

Remark 5.2. In the previous chapters, we used approximations spaces to solve problems which were not tractable in the continuous case. By contrast, it appears here that the continuous formulation is better suited from a computational point of view than the derivation in the discrete setting. To justify this approach, we also need to prove that the heuristic likelihood L is the limit of the likelihood L^Λ, which has been derived rigorously in a discrete setting, when the step of the grid Λ tends to zero. □

5.3.2.2 Heuristic MAP as the limit of finite-dimensional MAP

In the following lemma and proposition, we establish a uniform control of the heuristic likelihood L by the discrete likelihood L^Λ.

Lemma 5.3. *Let Λ be any finite set of points in $\mathbb{R}^d$, $T \in W_\Lambda^*$ and $v \in V$ such that $\|v\|_V \leq R$. Then, ϕ^v and $d\phi^v$ are uniformly bounded and*

$$M\left(\phi^v{}_* T\right) \leq C M(T), \tag{5.3.8}$$

where C is a constant independent of v, T and the grid Λ. $M(T)$ is the mass-norm of T as in Defn. 1.3.

Proof. Since $T \in W_\Lambda^*$, $T = \sum_{i \in \Lambda} \delta_{x_i}^{\alpha_i}$. For each Dirac delta currents, we have $\phi_* \delta_x^\alpha = \delta_{\phi(x)}^{\phi_* \alpha}$ where we denote $\phi_* \alpha = d_x \phi(\alpha)$ if α is a 1-vector (tangent of a curve) and $\phi_* \alpha = d_x \phi(u) \times d_x \phi(v)$ if $\alpha = u \times v$ is a 2-vector (normal of a surface) with u and v orthogonal (see Appendix A).

If α is a 1-vector, $|\phi_* \alpha| \leq \|d\phi\|_\infty |\alpha|$. If α is a 2-vector, $|\phi_* \alpha| \leq \|d\phi\|_\infty^2 |\alpha|$. Therefore, we have, denoting $\phi = \phi^v$ and applying Prop. 1.12:

$$\begin{aligned} M\left(\phi_* T\right) &= M\left(\sum_{i \in \Lambda} \delta_{\phi(x_i)}^{\phi_* \alpha_i}\right) = \sum_{i \in \Lambda} |\phi_* \alpha_i| \\ &\leq \max\left(\|d\phi\|_\infty, \|d\phi\|_\infty^2\right) \sum_{i \in \Lambda} |\alpha_i| \\ &= \max\left(\|d\phi\|_\infty, \|d\phi\|_\infty^2\right) M(T). \end{aligned} \tag{5.3.9}$$

In [Trouvé 2005a] (Lemma 11), it is shown that there is a numerical constant C such that:

$$\|\phi^v\|_\infty \leq C \|v\|_V \exp(C \|v\|_V) \tag{5.3.10}$$

This result is essentially an application of the Gromwall lemma. This shows the result for bounded vector fields ($\|v\|_V \leq R$). ■

Proposition 5.4 (Control of the heuristic MAP)**.** *Let L and L^Λ be as defined in Eq. (5.3.7) and Eq. (5.3.6) respectively. Let W be a RKHS with translation-invariant scalar kernel ($K(x,y) = k((x-y)/\lambda_W)\mathrm{Id}$) and Λ a linearly spaced grid with step Δ. Then, there is a constant C such that:*

$$\forall v^i \in V_{\Lambda_0}, \quad \left\|v^i\right\|_V \leq R, \quad \forall \overline{T} \in W^*_{\Lambda_0}, \quad M(\overline{T}) \leq R,$$
$$\left|L^\Lambda\left(v^i, \overline{T}\right) - L\left(v^i, \overline{T}\right)\right| \leq C \left|k_W^{(p)}(0)\right|^{1/2} \left(\frac{\Delta}{\lambda_W}\right)^{p/2} \tag{5.3.11}$$

where $k^{(p)}(0)$ denote the first non-zero derivative of the function k at zero. C is a constant independent of the v^i's and of $\overline{T}$.

Proof. We denote $U_i = T_i - \phi_{i*}\overline{T}$, so that we have:

$$\begin{aligned}
\left|L^\Lambda\left(v^i, \overline{T}\right) - L\left(v^i, \overline{T}\right)\right| &\leq \sum_{i=1}^N \left| \|\Pi_\Lambda\left(U_i\right)\|^2_{W^*} - \|U_i\|^2_{W^*} \right| \\
&\leq \sum_{i=1}^N \left| \langle \Pi_\Lambda(U_i) + U_i, \Pi_\Lambda(U_i) - U_i \rangle_{W^*} \right| \\
&\leq \sum_{i=1}^N 2 \left\|U_i\right\|_{W^*} \left\|\Pi_\Lambda(U_i) - U_i\right\|_{W^*}
\end{aligned} \tag{5.3.12}$$

We have $M(U_i) \leq M(T_i) + CM(\overline{T})$ by application of Lemma 5.3. Therefore, since $\|U_i\|_{W^*} \leq C_W M(U_i)$ (see Prop. 1.25) and since $M(\overline{T}) \leq R$, then $\|U_i\|_{W^*}$ is bounded independently of ϕ and $\overline{T}$.

Since Π_Λ denotes the orthogonal projection on W^*_Λ, we have for all $\overline{U}_i \in W^*_\Lambda$, $\|U_i - \Pi_\Lambda(U_i)\|_{W^*} \leq \left\|U_i - \overline{U}_i\right\|_{W^*}$. Then, using for $\overline{U}_i$ the partial volume projection of U_i on the grid Λ as defined in Chapter 2, we have according to Proposition 2.20:

$$\left\|U_i - \Pi_\Lambda(U_i)\right\|_{W^*} \leq CM(U_i) \left|k^{(p)}(0)\right|^{1/2} \left(\frac{\Delta}{\lambda_W}\right)^{p/2} \tag{5.3.13}$$

Since we already proved that $M(U_i) \leq M(T_i) + CR$, this shows the result. ■

Now, we can prove that minimization of the heuristic likelihood L is equivalent to the minimization of the rigorously derived likelihood L^Λ in the limit $\Delta \to 0$, where Δ denotes the step of the grid Λ. The following proposition shows that the global minima of L^Λ converge to the global minima of L as the step of the grid Λ tends to zero.

Lemma 5.5. *Assume that the kernel K of W is weakly continuous. Then, for all $v \in V^*_{\Lambda_0}$, such that $\|v\|_V \leq R$ and for all $\overline{T} \in W^*_{\Lambda_0}$, we have:*

$$C_1 \left\|\overline{T}\right\|_{W^*} \leq \left\|{\phi^v}_* \overline{T}\right\|_{W^*} \leq C_2 \left\|\overline{T}\right\|_{W^*}, \tag{5.3.14}$$

where C_1, C_2 are two positive constants independent of v and $\overline{T}$.

Moreover, the function $L\left((v_i)_{i=1,\dots,N}, \overline{T}\right)$ is continuous over $(V_{\Lambda_0})^N \times W^*_{\Lambda_0}$.

Proof. With the notations of the proof of Lemma 5.3, we have $\left\|\phi^v{}_*\overline{T}\right\|^2_{W^*} = \sum_{i,j\in\Lambda_0}(\phi_*\alpha_i)^t K(\phi(x_i),\phi(x_j))\phi_*\alpha_j = \boldsymbol{\alpha}^t\mathbf{K}^{\phi^v}\boldsymbol{\alpha}$, where $\mathbf{K}^\phi$ is a symmetric positive definite matrices whose block (i,j) is given by $d_{x_i}\phi^t K(\phi(x_i),\phi(x_j))d_{x_j}\phi$ for 1-current and $|d_{x_i}\phi|\, d_{x_i}\phi^{-1}K(\phi(x_i),\phi(x_j))\,|d_{x_j}\phi|\, d_{x_j}\phi^{-t}$ for 2-currents.

It is proven in [Glaunès 2005] (Theorem 4) for instance, that the maps $v\to\phi^v$ and $v\to d\phi^v$ are weakly continuous (when we restrict ϕ and $d\phi$ on a compact subset of $\mathbb{R}^3$, here the convex hull of the grid Λ_0). Therefore the map $v\to\mathbf{K}^{\phi^v}$ is also weakly continuous. Since the vector fields v_i are bounded, this proves that the set of $\mathbf{K}^{\phi^v}$ for all possible v such that $\|v\|_V\leq R$ is compact (in infinite dimension, the unit ball is compact for the weak topology). Let $\mu_{\min}=\min_{\|v\|_V\leq R}\left|\mathbf{K}^{\phi^v}\right|$ and $\mu_{\max}=\max_{\|v\|_V\leq R}\left|\mathbf{K}^{\phi^v}\right|$. Therefore for all $\boldsymbol{\alpha}$ and all v ($\|v\|_V\leq R$), we have:

$$\mu_{\min}\boldsymbol{\alpha}^t\boldsymbol{\alpha}\leq\boldsymbol{\alpha}^t\mathbf{K}^{\phi^v}\boldsymbol{\alpha}\leq\mu_{\max}\boldsymbol{\alpha}^t\boldsymbol{\alpha}.$$

Let $\lambda^{id}_{\min}$ and $\lambda^{id}_{\min}$ the smallest and largest eigenvalues of $\mathbf{K}^{\mathrm{Id}}$. Then,

$$\frac{\mu_{\min}}{\lambda^{\mathrm{Id}}_{\max}}\boldsymbol{\alpha}^t\mathbf{K}^{\mathrm{Id}}\boldsymbol{\alpha}\leq\boldsymbol{\alpha}^t\mathbf{K}^{\phi^v}\boldsymbol{\alpha}\leq\frac{\mu_{\max}}{\lambda^{\mathrm{Id}}_{\min}}\boldsymbol{\alpha}^t\mathbf{K}^{\mathrm{Id}}\boldsymbol{\alpha},$$

which shows the first point for $C_1=\sqrt{\mu_{\min}/\lambda^{\mathrm{Id}}_{\max}}$ and $C_2=\sqrt{\mu_{\max}/\lambda^{\mathrm{Id}}_{\min}}$.

To prove the continuity of L, we essentially need to prove that the map: $(v,\overline{T})\to\left\|\phi^v{}_*\overline{T}\right\|_{W^*}$ is continuous. Let $v_n,\overline{T}_n$ a sequence in $V_{\Lambda_0}\times W^*_{\Lambda_0}$ which converges to $v,\overline{T}$. Then,

$$\begin{aligned}\left|\left\|\phi^{v_n}{}_*\overline{T}_n\right\|_{W^*}-\left\|\phi^v{}_*\overline{T}\right\|_{W^*}\right|&\leq\left\|\phi^{v_n}{}_*\overline{T}_n-\phi^v{}_*\overline{T}\right\|_{W^*}\\&\leq\left\|\phi^{v_n}{}_*\overline{T}_n-\phi^{v_n}{}_*\overline{T}\right\|_{W^*}+\left\|\phi^{v_n}{}_*\overline{T}-\phi^v{}_*\overline{T}\right\|_{W^*}.\end{aligned}$$

Thanks to the first point we showed, the first term satisfies $\left\|\phi^{v_n}{}_*(\overline{T}_n-\overline{T})\right\|_{W^*}\leq C_2\left\|\overline{T}_n-\overline{T}\right\|_{W^*}$ which tends to 0 as n tends to infinity (since the v_n converges to v, the sequence is bounded). The second term tends also to zero since $v\to\phi^v$ and $v\to d\phi^v$ are continuous (see [Glaunès 2005] (Theorem 4) for instance) and since for $\overline{T}\in W^*_{\Lambda_0}$, $\left\|\phi_*\overline{T}\right\|^2_{W^*}=\boldsymbol{\alpha}^t\mathbf{K}^\phi\boldsymbol{\alpha}$ depends continuously on ϕ and $d\phi$. This shows the continuity of L. ■

Proposition 5.6 (Convergence of the minima of L^Λ to the minima of L)**.** *Let Λ^n be a sequence of linearly spaced grids whose step Δ_n tends to zero. Let $((v_i^{(n)})_{i=1,\dots,N},\overline{T}^{(n)})$ be a sequence of minima of L^{Λ^n}. Then, this sequence is bounded in the space $(V_{\Lambda_0})^N\times W^*_{\Lambda_0}$.*

Therefore, we can find a subsequence $((v_i^{(n_k)})_{i=1,\dots,N},\overline{T}^{(n_k)})$ which converges to $((v_i^\infty)_{i=1,\dots,N},\overline{T}^\infty)$ in $(V_{\Lambda_0})^N\times W^*_{\Lambda_0}$. And this limit is one minimum of L.

Proof. First, we prove that the sequence of vector fields are bounded. By definition of the minimum, $L^{\Lambda^n}(v_i^{(n)},\overline{T}^{(n)})\leq L^{\Lambda^n}(0,0)\leq\sum_{i=1}^N\|T_i\|^2_{W^*}$ which is a constant with respect to $v^{(n)},T^{(n)}$ and Λ_n. Since, $\sum_{i=1}^N\left\|v_i^{(n)}\right\|^2_V\leq L^{\Lambda^n}(v_i^{(n)},\overline{T}^{(n)})$, every sequence $v_i^{(n)}$ (for $i=1,\dots,N$) is bounded: there is a constant R such that $\left\|v_i^{(n)}\right\|_V\leq R$.

Second, we prove that the sequence of template is bounded. Since the vector fields are bounded, we can applied Lemma 5.5: $\left\|\overline{T}^{(n)}\right\|_{W^*}\leq C_1^{-1}\left\|\phi^{v_i^{(n)}}{}_*\overline{T}^{(n)}\right\|_{W^*}\leq L^{\Lambda^n}(v_i^{(n)},\overline{T}^{(n)})$.

This last term is bounded, like for the vector fields $v_i^{(n)}$, by $L^{\Lambda_n}(0,0)$. Therefore, the sequence $\overline{T}^{(n)}$ is bounded.

Now, we can derive a subsequence $((v_i^{(n_k)})_{i=1,\dots,N}, \overline{T}^{(n_k)})$ which converges to $((v_i^\infty)_{i=1,\dots,N}, \overline{T}^\infty)$. Moreover,

$$\begin{aligned}\left|L\left(v_i^\infty, \overline{T}^\infty\right) - L^{\Lambda_n}\left(v_i^{(n)}, \overline{T}^{(n)}\right)\right| \leq \\ \left|L\left(v_i^\infty, \overline{T}^\infty\right) - L\left(v_i^{(n)}, \overline{T}^{(n)}\right)\right| + \left|L\left(v_i^{(n)}, \overline{T}^{(n)}\right) - L^{\Lambda_n}\left(v_i^{(n)}, \overline{T}^{(n)}\right)\right|\end{aligned} \quad (5.3.15)$$

The first term tends to 0 since L is continuous (thanks to Lemma 5.5). The second term tends to 0 thanks to the uniform bound proven in Proposition 5.4.

Assume now that $(v_i^\infty, \overline{T}^\infty)$ does not minimize L. Then, there is $(v_i, \overline{T})$ such that $L(v_i^\infty, T^\infty) - L(v_i, \overline{T}) = \delta > 0$. Thanks to Eq. (5.3.15), we can find N such that for all $n \geq N$, $\left|L(v_i^\infty, \overline{T}^\infty) - L^{\Lambda_n}(v_i^{(n)}, \overline{T}^{(n)})\right| \leq \delta/3$. This means that $L(v_i^\infty, T^\infty) \leq L^{\Lambda_n}(v_i^{(n)}, T^{(n)}) + \delta/3$ and therefore that:

$$L(v_i, \overline{T}) \leq L^{\Lambda_n}(v_i^{(n)}, \overline{T}^{(n)}) - \delta + \delta/3 \quad (5.3.16)$$

Thanks to the uniform bound of Prop. 5.4, we can choose $N' \geq N$, such that for all $n \geq N'$, $\left|L^{\Lambda^n}(v_i, \overline{T}) - L(v_i, \overline{T})\right| \leq \delta/3$. Combined with Eq. (5.3.16), this result shows that:

$$L^{\Lambda^n}(v_i, \overline{T}) \leq L(v_i, \overline{T}) + \delta/3 \leq L^{\Lambda_n}(v_i^{(n)}, \overline{T}^{(n)}) - \delta/3, \quad (5.3.17)$$

which is contrary to the fact that $(v_i^{(n)}, \overline{T}^{(n)})$ is one global minimum of L^{Λ_n}. Eventually, this proves that the limit $(v_i^\infty, \overline{T}^\infty)$ is one minimum of L. ■

This proposition justifies to minimize the continuous likelihood L instead of L^Λ in the following.

Remark 5.7 (Uniform prior on the template)**.** In this MAP derivation, we assume a uniform prior on the template in a ball in the space of currents. This prior is not realistic in the sense that it does not favor currents which "look like" usual curves or surfaces. Indeed, as mentioned in Section 3.5.2, the set of "acceptable shapes" is likely to be of negligible measure in the space of currents. One way to end up with a template which is close to usual curves or surfaces could be to change this prior and to penalize the template which deviate too much from a given set of shapes (still to be defined). □

5.3.3 An alternated minimization procedure

We propose to minimize the likelihood L in Eq. (5.3.2) alternatively with respect to the template and to the deformations. When the template $\overline{T}$ is fixed, each term of the likelihood can be minimized separately: for a given observation T_i and knowing the template $\overline{T}$, minimizing

$$\frac{1}{\sigma_W^2}\left\|T_i - \phi_i^{v_0^i}{}_*\overline{T}\right\|_{W^*}^2 + \frac{1}{\sigma_V^2}\left\|v_0^i\right\|_V^2, \quad (5.3.18)$$

with respect to v_0^i is a registration problem. The single difference with the registration criterion of Chapter 4 is that the variable is here the initial vector field v_0 instead of the

flow $(v_t)_{t\in[0,1]}$. However, as we mentioned in Section 4.2.2.3, the minimizing flow of vector field is geodesic and, at the minimum, $\int_0^1 \|v_t\|_V^2\, dt = \|v_0\|_V^2$. Therefore, the registration of $\overline{T}$ to T_i in the framework of Chapter 4 leads also to a minimum of the criterion in Eq. (5.3.18). As a consequence, the N registrations of the template $\overline{T}$ to each observation T_i (which can be computed in parallel) minimize the likelihood L with respect to the deformation parameters v_0^i.

When the deformations ϕ_i are fixed for every $i = 1 \dots N$, minimizing Eq. (5.3.2) with respect to the template $\overline{T}$ leads to the minimization of the convex function:

$$J(\overline{T}) = \frac{1}{2}\sum_{i=1}^{N} \left\|\phi_{i*}\overline{T} - T_i\right\|_{W^*}^2 \tag{5.3.19}$$

If $\phi_i = \mathrm{Id}$ (i.e. no deformation) for all i, the minimum is reached at the empirical mean: $\overline{T} = \frac{1}{N}\sum_i^N T_i$. For arbitrary deformations, there is no closed form and we use a gradient descent scheme. Let $\delta\overline{T}$ be a variation of the template in the space of currents. Then, by linearity of the push-forward action on currents (see Eq. (1.3.21)), we have

$$J\left(\overline{T} + \delta\overline{T}\right) = J\left(\overline{T}\right) + \sum_{i=1}^{N} \left\langle \phi_{i*}\delta\overline{T}, \phi_{i*}\overline{T} - T_i\right\rangle_{W^*} + o\left(\left\|\delta\overline{T}\right\|_{W^*}\right)$$

Therefore the gradient of Eq. (5.3.19) is given by:

$$\nabla_{\overline{T}} J = \sum_{i=1}^{N} \phi_{i*}^{\dagger}(\phi_{i*}\overline{T} - T_i) \tag{5.3.20}$$

where $\phi_*^\dagger$ is the adjoint action of ϕ_* defined by:

$$\left\langle \phi_*^\dagger T, T'\right\rangle_{W^*} = \left\langle T, \phi_* T'\right\rangle_{W^*}, \tag{5.3.21}$$

for any currents T and T'. This would be a matrix transpose if the deformation ϕ were a linear operator on W^* (like for an affine transformation of a set of points for instance). In this non-linear setting, the computations are slightly more complex. From a computational point of view, we must explicit $\phi_*^\dagger$ and express it in terms of known operations on currents.

Applying the definition of ϕ_* and the isometric mapping $\mathcal{L}_W$, we have for all currents T and T':

$$\begin{aligned}
\left\langle \phi_*^\dagger T, T'\right\rangle_{W^*} &= \left\langle T, \phi_* T'\right\rangle_{W^*} && \text{Def. of } \phi_*^\dagger \\
&= \phi_* T'\left(\mathcal{L}_W^{-1}(T)\right) && \text{Def. of } \langle .,.\rangle_{W^*} \text{ and } \mathcal{L}_W \text{ (Eq. (B.3.7))} \\
&= T'\left(\phi^* \mathcal{L}_W^{-1}(T)\right) && \text{Def. of } \phi_* \text{ and } \phi^* \text{ (see Definition 1.15)} \\
&= \left\langle \mathcal{L}_W(\phi^*\mathcal{L}_W^{-1}(T)), T'\right\rangle_{W^*} && \text{Def. of } \langle .,.\rangle_{W^*} \text{ and } \mathcal{L}_W \text{ (Eq. (B.3.7))}
\end{aligned}$$

These equations show that:

$$\phi_*^\dagger T = \mathcal{L}_W(\phi^*\mathcal{L}_W^{-1}(T)) \tag{5.3.22}$$

The application of this formula to δ_x^α as a 1-current (see Eq. (1.4.6)) shows that:

$$\phi_*^\dagger \delta_x^\alpha = \mathcal{L}_W\Big((d_x\phi)^t K^W(\phi(.), x)\alpha\Big) \tag{5.3.23}$$

Its application to δ_x^α as a 2-current (see Eq. (1.4.10)) shows that:

$$\phi_*^\dagger \delta_x^\alpha = \mathcal{L}_W \Big(|d_x\phi| \, d_x\phi^{-1} K^W(\phi(.), x)\alpha \Big) \tag{5.3.24}$$

Neither the dense vector field $(d_x\phi)^t K^W(\phi(.), x)\alpha$ or $|d_x\phi| \, d_x\phi^{-1} K^W(\phi(.), x)\alpha$ can be expressed simply as a linear combination of basis vector fields $K^W(., y)\beta$ (because of ϕ inside the kernel). Therefore, there is no simple closed form to compute the mapping $\mathcal{L}_W$ in these cases. To implement this ill-posed deconvolution problem, one needs a robust numerical scheme. In this thesis, we will use naturally the Matching Pursuit Algorithm introduced in Chapter 3.

Now, we can make explicit the computation of the gradient in Eq. (5.3.19). The input shapes T_i are sampled objects which are approximated as finite set of Dirac currents. As it will appear from this minimization procedure, the template will also always remain a finite set of Dirac currents at every iteration. Therefore, the current $\phi_{i*}\overline{T} - T_i$ is of the form $\sum_k \delta_{y_k^i}^{\beta_k^i}$ which gives finally $\phi_i^* \mathcal{L}_W^{-1}(\phi_{i*}\overline{T} - T_i) = \sum_k (d_x\phi_i)^t K^W(\phi_i(x), y_k^i)\beta_k^i$ in case of curves. For surfaces, $d_x\phi_i^t$ must be replaced by $|d_x\phi_i| \, d_x\phi_i^{-1}$. Finally, the dual representation of the gradient in the space of vector field W is given at any point x by:

$$\mathcal{L}_W^{-1}(\nabla_{\overline{T}} J)(x) = \sum_{i=1}^N (d_x\phi_i)^t \left(\sum_k K^W(\phi_i(x), y_k^i)\beta_k^i \right), \tag{5.3.25}$$

in case of curves and

$$\mathcal{L}_W^{-1}(\nabla_{\overline{T}} J)(x) = \sum_{i=1}^N |d_x\phi_i| \, d_x\phi_i^{-1} \left(\sum_k K^W(\phi_i(x), y_k^i)\beta_k^i \right), \tag{5.3.26}$$

in case of surfaces.

Thus, we see that we know how to compute the dual representation of the gradient (a vector field) but not the gradient itself (a current). We use now the Matching Pursuit algorithm of Chapter 3 to perform the deconvolution of the vector field $\mathcal{L}_W^{-1}(\overline{T} - \tau\nabla_{\overline{T}} J)$ (where τ is the adaptive step of the gradient descent) to give an approximation of the updated template at any desired accuracy. This approximation is given as a finite set of Dirac delta currents. As a consequence, the template remains a finite set of Dirac currents at each iteration. Moreover, the number of Dirac currents tends to be minimized, thus leading to a template in a sparse form.

We initialize the algorithm by setting $\phi_i = \mathrm{Id}$, $\overline{T} = \frac{1}{N}\sum_{i=1}^N T_i$ and by computing the vector field $\gamma_{\overline{T}} = \mathcal{L}_W^{-1}(\overline{T})$ associated to $\overline{T}$ via a Gaussian convolution (computed with FFT's on images of vectors as explained in Chapter 2). The current $\overline{T}$ is encoded as a list of momenta (couple (position, vectors)) that approximates small segments or small triangles. The dense vector field $\gamma_{\overline{T}}$ is discretized at a the points of a fixed grid: $\Lambda = \{x_p\}$ and is therefore encoded as an image of vectors.

The alternate minimization procedure for the template estimation is written as in Algorithm 3. The step which centers the template according to the new deformation via the minimization of $J(\overline{T}) = \sum_{i=1}^N \left\| \phi_{i*}\overline{T} - T_i \right\|_{W^*}$ is detailed in Algorithm 4.

Algorithm 3 Atlas Construction

Input: N shapes T_i (list of oriented points (tangents or normals): $T_i = (x_i^p, \alpha_i^p)$).

List $\overline{T} = (\overline{x}_k, \overline{\alpha}_k) \leftarrow$ concatenation of all $(x_i^p, \alpha_i^p/N)$ (encodes $\sum_{i=1}^N T_i/N$)
$\overline{T} \leftarrow$ sparse approximation of $\overline{T}$ using the Matching Pursuit (Algorithm 2)
repeat
  for $i = 1 \dots N$ **do** {Register Template to Subjects}
    $\phi_i \leftarrow$ registration of $\overline{T}$ to T_i.
  end for{Re-center the template}
  $\overline{T} \leftarrow$ CenterTemplate$(\overline{T}, \{\phi_i\}, \{T_i\})$ (Algorithm 4)
until convergence

Output: One template $\overline{T}$, N deformations ϕ_i, N residues $\phi_{i*}\overline{T} - T_i$ (stored as a list of momenta)

As a result, the algorithm returns an unbiased template $\overline{T}$ as well as the deformations of this template to every observation T_i. The residues are given in the space of currents by $T_i - \phi_{i*}\overline{T}$. The methodology developed in Section 5.4.1 can be used therefore to perform statistics on such residual currents.

Remark 5.8. In the backward scheme, Eq. (5.3.19) would be:

$$J(\overline{T}) = \frac{1}{2}\sum_{i=1}^{N} \left\|\overline{T} - \phi_{i*}T_i\right\|_{W^*}^2,$$

whose minimum has the closed form

$$\overline{T} = \frac{1}{N}\sum_{i=1}^{N} \phi_{i*}T_i$$

This shows that the minimum is the empirical mean of the observations pulled back into the current template configuration $(\phi_i)_*T_i$. In particular, no gradient descent would be required. By contrast, in our forward setting, the estimation of the template involves the Jacobian matrix of the deformations (via $\phi_*^\dagger$) and no closed form for the minimum is provided. However, computing the likelihood of any new observations $p(T_i|\overline{T})$ will be straightforward in the forward setting, whereas the computation of this likelihood in the backward setting will require to account for the Jacobian of the deformations. □

5.4 Statistical analysis of deformations and residual currents

Once the atlas is constructed along the lines of the previous section, one would like to analyze the variability of the population by performing PCA on the residual currents and on the deformations.

Algorithm 4 Template centering

{Implement the gradient descent to minimize $J(\overline{T}) = \sum_{i=1}^{N} \left\| \phi_{i*}\overline{T} - T_i \right\|_{W^*}$}
Input: an initial template $\overline{T}$, N deformations $\{\phi_i\}_{1\leq i\leq N}$, N currents $\{T_i\}_{1\leq i\leq N}$

Set a grid Λ which covers the currents (T_i) and $\phi_{i*}\overline{T}$.
Image of vectors $\gamma_{\overline{T}} \leftarrow \mathcal{L}_W^{-1}(\overline{T})$: $\forall x_p \in \Lambda, \gamma_{\overline{T}}(x_p) = \sum_k K^W(x_p, \overline{x}_k)\overline{\alpha}_k$, computed via convolution and FFT (see Section 2.2.3)
repeat {Gradient descent}
  Image of vectors grad $= 0$
  for $i = 1 \dots N$ **do**
    Deform $\overline{T}$ with ϕ_i: $\phi_{i*}\overline{T} = \begin{cases} (\phi_i(\overline{x}_k), d_{\overline{x}_k}\phi_i(\overline{\alpha}_k)) & \text{(curves)} \\ (\phi_i(\overline{x}_k), \left|d_{\overline{x}_k}\phi_i\right| d_{\overline{x}_k}\phi_i^{-t}(\overline{\alpha}_k)) & \text{(surfaces)} \end{cases}$
    Concatenate the list $(x_i^p, -\alpha_i^p)$ with the previous one to give (y_k^i, β_k^i) (encodes $\phi_{i*}\overline{T} - T_i$).
    Deform Λ with ϕ_i
    for all $x_p \in \Lambda$ **do**
      Compute $d_{x_p}\phi_i$ by a finite difference scheme
      Compute $G = \sum_k K^W(\phi_i(x_p), y_k^i)\beta_k^i$ (convolution via FFT (see Section 2.2.3))
      $\text{grad}(x_p) \leftarrow \text{grad}(x_p) + \begin{cases} 2(d_{x_p}\phi_i)^t G & \text{(curves)} \\ 2\left|d_{x_p}\phi_i\right|(d_{x_p}\phi_i)^{-1}G & \text{(surfaces)} \end{cases}$
    end for
  end for
  $\gamma_{\overline{T}} \leftarrow \gamma_{\overline{T}} - \tau\text{grad}$
  Deconvolution of $\gamma_{\overline{T}}$ to give the new $\overline{T} = (\overline{x}_k, \overline{\alpha}_k)$ via Matching Pursuit in Algorithm 2).
until convergence

Output: the new $\overline{T}$.

The residual currents are given as $R_i = \phi_{i*}\overline{T} - T_i$: a weighted sum of the momenta of the template and each input shape. This representation may not be optimal and may be difficult to interpret. Indeed, each residual has a different support: the union of the support of $\phi_i * \overline{T}$ and the support of T_i. Moreover, these two supports are close one from the other and this decomposition of the residual in terms of Dirac delta currents may be highly redundant at the scale λ_W. The sparse representation of the currents introduced in Chapter 3 is required to give a more compact and more interpretable representation of mean and modes of the set of residuals.

By contrast, the initial speed vector field of the deformations have all the same support, namely the points of the template. Note that this is a direct consequence of the forward setting. Therefore, in the framework of Chapter 2, every initial vector field belong to the same finite-dimensional RKHS (induced by the points of the template). A PCA can be applied directly in this finite-dimensional RKHS, which leads naturally to a compact

representation of the statistics on vector fields.

5.4.1 Mean and modes in the space of currents

After the construction of the atlas, each residue is given by: $R_i = \phi_{i*}\overline{T} - T_i$ and written as the sum of the deformed Dirac delta currents of the template $\overline{T}$ and the opposite of the Dirac delta currents of the input shape T_i. Let's write this sum: $R_i = \sum_k \delta_{x_k^i}^{\alpha_k^i}$.

We take advantage of the fact that the space of currents W^* is a vector space provided with an inner-product to perform directly a PCA of the residues.

The empirical mean of the residual currents is given by: $\overline{R} = \sum_i R_i/N = \sum_i \sum_p \delta_{x_p^i}^{\alpha_p^i/N}$, which is the union of the momenta of every observation scaled by $-1/N$ and of the momenta of the every deformed template scaled by $1/N$. Since the support the deformed template $\phi_{i*}\overline{T}$ is close to the support of the observation T_i, by definition of the registration, this representation of the empirical mean residual current has a very redundant representation at the scale λ_W. We use therefore the matching pursuit algorithm introduced in Chapter 3 to give a more compact and more easily interpretable representation of this empirical mean.

The N-by-N empirical covariance matrix Γ is given by: $\Gamma_{ij} = \left\langle R_i - \overline{R}, R_j - \overline{R} \right\rangle_{W^*}$ (computed in practice thanks to the inner-product between two Dirac delta currents in Eq. (1.2.8)). Let $V^1, \ldots, V^N$ be the eigenvectors of Γ. The nth principal mode of the residues is given by: $\overline{R} + \sum_i V_i^n (R_i - \overline{R})$. Expanding this expression leads to a double sum of the form: $\sum_i \sum_p w_p^{i,n} \delta_{x_p^i}^{\alpha_p^i}$ for some weights $w_p^{i,n}$. This support of this current is still the union of the support of the deformed template and the observation. Therefore this decomposition is still redundant at the scale λ_W and can be simplified using the sparse representation of Chapter 3.

These statistics can be used to interpret the anatomical variability captured by the residual currents and therefore drive the search for new anatomical knowledge. Moreover, since they have an adapted decomposition, they can be used easily to give more quantitative measures of the variability. For instance, they can be used to compute the projection of each observation to the first mode and analyze the correlation between modes and observations.

Remark 5.9. If $F_1, \ldots, F_n$ denote n feature vectors of dimension p, the usual PCA consists in finding the eigenvectors and eigenvalues of the p-by-p matrix AA^t, where A is the p-by-n matrix $(F_1|\ldots|F_n)$. The p-dimensional eigenvectors of this matrix are called the principal modes. However, when the dimension of the space p is much greater than the number of observations n, we use the property that the n-by-n matrix A^tA has the same non-zero eigenvalues than the matrix AA^t (the $p-n$ other eigenvalues of AA^t are null). If $v \in \mathbb{R}^n$ is an eigenvector of A^tA associated to the eigenvalue λ, then $Av = \sum_{i=1}^n v_i F_i \in \mathbb{R}^p$ is an eigenvector of AA^t associated to the same eigenvalue (conversely if $v \in \mathbb{R}^p$ is an eigenvector of AA^t, then A^tv is an eigenvector of A^tA). This shows that the principal modes are always given as linear combination of the input vectors F_i.

This shows also that we can perform PCA in infinite dimensional spaces: we can limit the search for the modes in the finite dimensional subspace spanned by the input vectors. In this case, only the second form A^tA can be computed. The $(i,j)th$ term of this matrix is given by this inner-product $F_i^tF_j$. This is what we compute here, except that we use the

metric induced by the kernel K^W instead of the Euclidean inner-product on the feature vectors. This is equivalent to finding the eigenvectors and eigenvalues of the matrix $A^t K A$ where K is the p-by-p matrix of the inner-product between every pair of Dirac delta currents (and therefore equivalent to replacing the matrix A by $K^{1/2}A$). □

Remark 5.10. We compute the k*th* mode of currents as $m^k = \overline{R} + \sum_i V_i^k (R_i - \overline{R})$, whereas one usually scales the sum with the eigenvalue λ^k which corresponds to the eigenvector V^k, so that the norm of the mode is equal to the standard deviation of the mode. This normalization is not required here, since we performed the PCA by finding the eigenvectors of a n-by-n matrix of the form $A^t A$ (see remark above). Since the eigenvectors $V^k \in \mathbb{R}^n$ are of unit norm ($\sum_{i=1}^n (V_i^k)^2 = 1$), we verify that: $\left\|m^k - \overline{R}\right\|_{W^*}^2 = \sum_i \sum_j V_i^k V_j^k \left\langle R_i - \overline{R}, R_j - \overline{R} \right\rangle_{W^*} = \sum_i \sum_j V_i^k V_j^k \Gamma_{ij} = (V^k)^t \Gamma V^k = \lambda^k$. The squared norm of the principal mode (with respect to the metric induced by the kernel K^W) already equals the variance of the mode. These remarks holds also for the following principal modes of deformations. □

5.4.2 Mean and modes of deformations

To compute statistics on deformations, we use the tangent-space representation of the diffeomorphisms as in [Vaillant 2004]. As explained in Chapter 4, each deformation is entirely determined by its initial speed vector field v_0^i. All these vector fields belong to a common RKHS V: the tangent space of the group of diffeomorphism at the identity transformation. By computing mean and modes of deformations, we mean computing the mean and mode of the initial vector fields in the common subspace V and then perform a geodesic shooting of the mean and the modes.

Moreover, since each diffeomorphism is computed by registering the same template to each observations, every initial speed vector field v_0^i is such that $\mathcal{L}_V(v_i^0) = \sum_{x_k \in \overline{T}} \delta_{x_k}^{\alpha_k^i}$. The dual representation of the vector fields share the same support: the set of point of the template denoted $\mathcal{X}_T = \{x_k\}_{k=1,\dots,M}$. Using the notations of Chapter 2, these dual representations belong to the same finite-dimensional RKHS $V_{\mathcal{X}_T}$. Since the points of the template results from the application of a matching pursuit (see Algorithm 4), they are "optimally" distributed in some sense, and the metric $\mathbf{K}_{\mathcal{X}_T}$ on this RKHS has a good conditioning. We remark however, that the template results from an application of the matching pursuit in the space W^*, thus limiting its redundancy at the scale λ_W. This does not mean that the representation is not redundant at the scale λ_V if this scale is greater than the scale λ_W. In practice, nevertheless, this representation is sparse enough and we do not need to perform a Matching Pursuit in the space V^*.

The same computations as for the currents apply here, provided that the inner-product in W^* is replaced by the inner-product in V^* and that we take advantage of the finite-dimensional representation of the vector fields.

Let $\boldsymbol{\alpha}^i$ be the $3M$ column vector $(\alpha_k^i)_{k=1,\dots M}$ and $\mathbf{K}_{\mathcal{X}_T}$ the $3M$-by-$3M$ block matrix whose block (i,j) is given by $K^V(x_i, x_j)$ as in Chapter 2 (K^V denotes here the kernel of the RKHS V). The mean momenta is given by: $\overline{\boldsymbol{\alpha}} = \sum_i \boldsymbol{\alpha}^i / N$, which leads to the mean

speed vector field via:

$$\overline{v}^0 = \mathcal{L}_V^{-1}\left(\sum_k \delta_{x_k}^{\overline{\boldsymbol{\alpha}}_k}\right) = \sum_k K^V(.,x_k)\left(\sum_i \alpha_k^i/N\right).$$

To compute the principal modes, we build the N-by-N matrix Γ^V such that

$$\Gamma_{ij}^V = \left\langle v_i^0 - \overline{v}^0, v_j^0 - \overline{v}^0 \right\rangle_V = (\boldsymbol{\alpha}^i - \overline{\boldsymbol{\alpha}})^t \mathbf{K}_{\mathcal{X}_T}(\boldsymbol{\alpha}^j - \overline{\boldsymbol{\alpha}}).$$

Let V^n be the eigenvectors of the matrix Γ^V. The nth principal mode is given by

$$\mathbf{m}^{\alpha,n} = \overline{\boldsymbol{\alpha}} + \sum_i V_i^n(\boldsymbol{\alpha}^i - \overline{\boldsymbol{\alpha}}),$$

which leads to the vector field: $m^{V,n} = \sum_k K^V(.,x_k)\mathbf{m}_k^{\alpha,n}$. A geodesic shooting of these principal modes gives a diffeomorphism $\phi^{m^{V,n}}$ which is called the nth mode of deformations.

5.5 Atlas construction on simulated 2D-curves

We illustrate here the template estimation procedure on a simulated example of 2D curves and show how this atlas can be used for group comparison.

5.5.1 Construction of a simulated database

We construct a database of random 2D-shapes, as illustrated in Fig. 5.3. From a template shape S, we construct two mean shapes $\phi^{\alpha_0}(S)$ and $\phi^{-\alpha_0}(S)$ where ϕ^{α_0} is a small-scale deformation which is parametrized by momenta α_0 whose support is close to the upper-right part of the shape (green squares in Fig. 5.3). For this deformation, the spatial scale of the Gaussian kernel is $\lambda_V^{\text{small}} = 0.12$. Then, we simulate random large-scale deformations ϕ^{α}, where the support of the momenta α is spread over the space (yellow diamonds in Fig. 5.3). The random momenta are simulated as random variables in the RKHS whose kernel is Gaussian with standard deviation $\lambda_V^{\text{large}} = 0.4$. This leads to two sets of shapes (one set is "centered" around the shape $\phi^{\alpha_0}(S)$, the other around $\phi^{-\alpha_0}(S)$) as shown in Fig. 5.4. From a visual inspection of the shapes, it is difficult to know whether the two classes are different and to guess the possible differences between them. To illustrate how our statistical modeling can be used for such shape comparison, we divide the simulated database into a set of 80 training data (40 per class) and 200 test data (100 per class).

5.5.2 Atlas estimation

We applied our atlas construction method (Algorithms 3 and 4) to construct an atlas from the 80 samples of the set of training data. For this construction, we consider the data as 0-currents (tangents are not taken into account). The parameters used for this estimation are the spatial scale of currents $\lambda_W = 0.05$, the spatial scale for deformations $\lambda_V = 0.4$, the trade-off between regularity and fidelity-to-data $\gamma = 10^{-4}$ and the sparsity number $\tau = 5\%$.

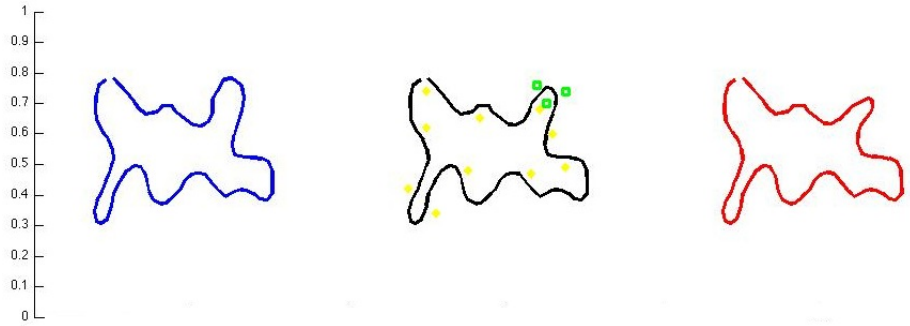


Figure 5.3: True templates used to create the data. From the mean template (black) we simulate a small-scale diffeomorphism (parametrized by momenta at the green squares). This diffeomorphism is applied to the black template to give the template of the red class (right). The inverse of the diffeomorphism leads to the template of the blue class (left). Finally, random large-scale diffeomorphisms (parametrized by momenta at the yellow diamond) are applied to the blue and red template. This constructs a database. Some samples are shown in Fig. 5.4

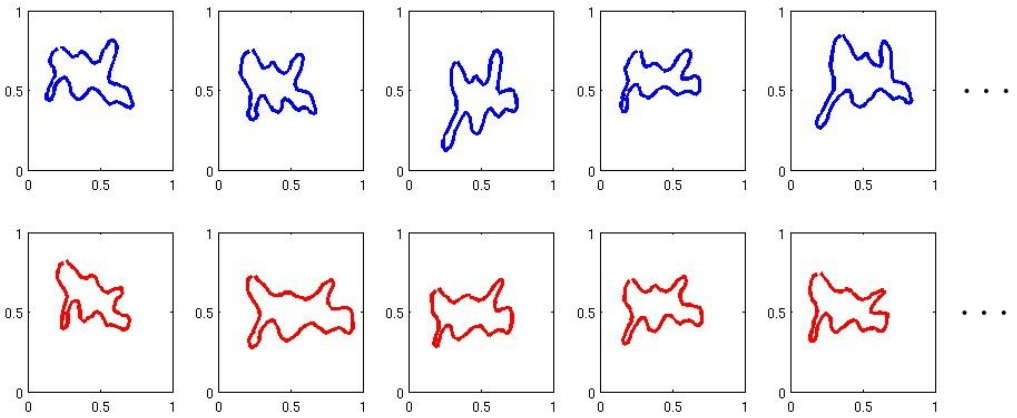
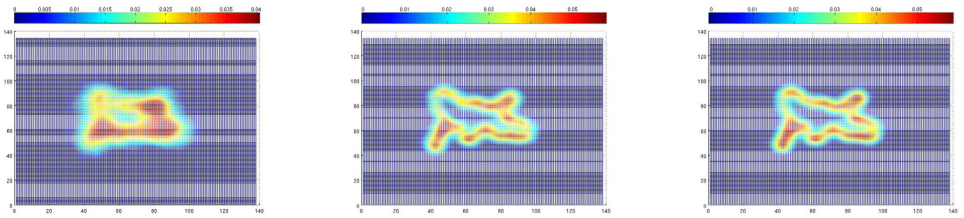


Figure 5.4: Five samples among the 40 samples for each class. Our goal is to show that our atlas estimation allows us to find the geometrical discriminative features between both classes.

The iterations of the algorithm are shown in Fig. 5.5. Two iterations only were needed until the convergence criterion of the adaptive gradient descent was reached. As a result, we end up with a template $\overline{T}$ (a set of unconnected weighted points) and deformations of this template to each shape in the database. The deformations are parametrized by momenta located at the points of the template.

Now, we can compare the estimated template $\overline{T}$ and the true template S used to construct the database. Measured in the space of currents, the difference between the two templates is equal to 0.24 times the standard deviation of the observations. With 80 sam-

ples, the t-statistics is equal to $\sqrt{80} * 0.24 = 2.15$, which is below the usual 3σ threshold to decide statistical significance[1].



a- mean current b- iteration 1 c- iteration 2

Figure 5.5: Atlas construction from 80 samples. At each iteration the algorithm register the template (shown here) to each sample and update the template according to the deformations. The figures show the dual representation of the template in the space W. As 0-currents, the shapes are modeled as point sets (and not oriented points): the dual representation is therefore a scalar field (shown here) instead of a vector field. Colors correspond to the magnitude of this scalar field. The initial template is the empirical mean current (a). The 2 iterations until convergence are shown in (b) and (c). Along with the iterations, the bias is removed from the template. As a result, the template appear to be less and less blurred.

5.5.3 Statistical Analysis of the deformations

In this section, we turn to the statistical analysis of the shapes. We observed that the residual in the space of currents does not capture discriminative features between classes. This is expected since data differs one from the others by smooth deformations. As a consequence, we focus on the statistics on deformations. In our forward setting, the tangent-spaces of the all the diffeomorphisms coincide. This makes the statistics on the initial speed vector particularly easy compared to the backward setting, for which initial speed vector field must be transported from each subject's space to the common template space before comparison.

5.5.3.1 Dimensionality reduction

Each deformation is parametrized by momenta α_i located at the points of the estimated template, or equivalently the associated vector-field sampled at the same points γ_i (according to Chapter 2, the two representations are deduced one from the others via the multiplication by the matrix $\mathbf{K}_{\mathcal{X}_T}$, where $\mathcal{X}_T$ denotes the set of points of the template). For each subject, the concatenation of the vectors γ_i leads to a feature vector $\boldsymbol{\gamma}$. For each subject s, this feature vector is denoted $\boldsymbol{\gamma}^s$.

[1]Note that computing a rigorous Hotelling T^2 test in this context is particularly difficult due to the infinite number of degrees of freedom of the currents.

The estimated template has 66 points. In $2D$, this leads to feature vector of dimension $2 * 66 = 132$, which is greater than the number of available observations: 80. This is a standard situation in medical imaging. To reduce the dimension of the feature vectors, we perform a non-centered Principal Component Analysis (PCA) and keep only the first 26 modes. This number of modes was chosen so that the classification results on the training data (as explained in the next section) are optimal.

As a consequence, we project all the feature vectors on the common sub-space determined by these 26 first directions. Now, the feature vectors have 26 components.

5.5.3.2 Population separation

Let $\boldsymbol{\mu}_b$ and $\boldsymbol{\mu}_r$ be the sample mean of the feature vectors for the blue class and the red class respectively. An Hotelling T^2 test is performed to decide whether the two populations are distinct (null hypothesis: $(\boldsymbol{\mu}_b = \boldsymbol{\mu}_r)$) and if the whole population is centered (null hypothesis: $((\boldsymbol{\mu}_b + \boldsymbol{\mu}_r)/2 = 0)$). The value of the Fisher-Snedecor statistics corresponding the Hotelling statistics is 14.4 for the difference of both means and 1.80 for the total sample mean. Statistics greater than 2.1 correspond to a probability of acceptance of the null hypothesis less than 1%. As a consequence, these feature vectors separate the two populations in two clusters with different means. Moreover, we cannot reject the hypothesis that the global population is centered.

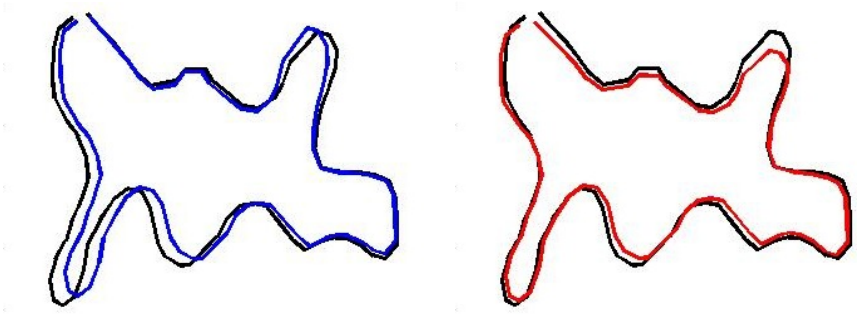
For each mean, we compute the diffeomorphism $\phi^{\mathbf{K}_{\mathcal{X}_T}\boldsymbol{\mu}_b}$ and $\phi^{\mathbf{K}_{\mathcal{X}_T}\boldsymbol{\mu}_r}$ via geodesic shooting of the vector fields $\mathbf{K}_{\mathcal{X}_T}\boldsymbol{\mu}_b$ and $\phi^{\mathbf{K}_{\mathcal{X}_T}\boldsymbol{\mu}_r}$ respectively (see Section 4.4.2). Then, we apply these diffeomorphisms to the template, as illustrated in Fig. 5.6. Note that for visualization purposes, we used here the true template (S) instead of the estimated one ($\overline{T}$). This figure shows that the difference between the two population at the first order is a torque of the upper-right part of the shape.

Now, we focus on three different methods to learn how to separate the two classes: a method based on Mahalanobis distances, a linear discriminative analysis and a classifier based on Support Vector Machines (SVM). This classification methods should determine whether the discriminative features captured by our model are reproducible across the populations.

Separation based on Mahalanobis distance On the common sub-space of dimension 26, we compute the sample mean $\boldsymbol{\mu}_b$ and $\boldsymbol{\mu}_r$ and the sample covariance matrix Σ_b and Σ_r for each class. Then, given any feature vector $\boldsymbol{\gamma}$, we can classify it according to the maximum of likelihood principle for the estimated Gaussian laws (i.e. a Neyman-Pearson test). The statistic is given by:

$$T(\boldsymbol{\gamma}) = (\boldsymbol{\gamma} - \boldsymbol{\mu}_b)^t \Sigma_b^{-1} (\boldsymbol{\gamma} - \boldsymbol{\mu}_b) - (\boldsymbol{\gamma} - \boldsymbol{\mu}_r)^t \Sigma_r^{-1} (\boldsymbol{\gamma} - \boldsymbol{\mu}_r). \tag{5.5.1}$$

If $T(\boldsymbol{\gamma}) \leq \eta$ for a given threshold η, then $\boldsymbol{\gamma}$ is classified as "blue" (it is closer to the blue population according to the Mahalanobis distance) and as "red" otherwise. For our experiments, there is a value of the threshold η which perfectly discriminate between the two populations. This means that there is a quadric surface which separates the two sets of feature vectors.



Mean of the blue class Mean of the red class

Figure 5.6: The true template (in black) and the means of the blue and the red class. This shows that the difference between both classes is mainly a torque at the upper-part of the shape (taking the blue class as a reference).

Fisher's linear discriminant Similarly, we can use a Fisher's linear discriminant rule to find a hyper-plane to separate the two populations. The discriminative direction is chosen so that the ratio between the inter-class variance and the intra-class variance is maximized. This leads to the statistics:

$$T(\boldsymbol{\gamma}) = (\boldsymbol{\mu}_r - \boldsymbol{\mu}_b)^t (\Sigma_r + \Sigma_b)^{-1} \left(\boldsymbol{\gamma} - \frac{\boldsymbol{\mu}_b + \boldsymbol{\mu}_r}{2} \right). \tag{5.5.2}$$

If $T(\boldsymbol{\gamma}) \leq \eta$ for a threshold η, then $\boldsymbol{\gamma}$ classified as "blue" and "red" otherwise. For this classifier also, there is a value of the threshold η which enables to separate perfectly the two classes. The hyper-plane which separate the two populations is orthogonal to the direction $(\Sigma_r + \Sigma_b)^{-1}(\boldsymbol{\mu}_r - \boldsymbol{\mu}_b)$.

Support Vector Machine In addition, we train a Support Vector Machine on the feature vectors. We used a linear kernel and a constant C equals to infinity. These parameters allows also to perfectly separate the two populations.

5.5.3.3 Classification of new data

Our statistical modeling does not only lead to descriptive measure of the variability, but can be used also to predict the class of new observations. This is of the uttermost interest in order to evaluate how well the estimated discriminative features can be generalized for a whole population.

For this purpose, we decompose any test shape (we have 100 test data per class) on our estimated atlas. The registration of the estimated template ($\overline{T}$) to each test shape leads to initial speed vectors at the points of the template. We project these feature vectors on the common subspace of dimension 26 which was estimated in the previous section. Then we classify these projected feature vectors according to our three classification methods: the difference of Mahalanobis distances, the linear discriminant analysis and the SVM methods. For the first two methods, we can compute the rate of false positive (red shapes classified

as blue) and the rate of true negative (blue shapes classified as blue) for every possible threshold η. This leads to ROC curves as shown in Fig. 5.7 along with one result for the SVM method.

These ROC curves shows that our atlas can be used to predict the class of new observation. It appears that the Fisher's linear discriminant is the method which enables to achieve the best classification ratio. To compare the methods quantitatively, we define the mean classification error as the minimal probability for a new data to be misclassified, this new data having 50% chance to be blue and 50% chance to be red. The Mahalanobis classifier achieves a mean error of 6%, the SVM of 3% and the Fisher's linear discriminant of 1%. This shows that our statistical atlas was able to capture at least 99% of the true geometrical variability which explains the difference between the two classes. The method does not over-fit the training data, but detect reproducible features across the population. Moreover, as shown in Fig. 5.6, our modeling leads to a geometrical interpretation of the variability which have been captured, namely a torque of the upper right part of the shape torque (if we consider the blue class as reference). Such features were almost impossible to guess from a simple visual inspection of the initial data in Fig. 5.4.

Eventually, we test whether our statistical atlas estimation could have been replaced by a more pragmatic approach. At least two alternatives are possible: we can choose randomly one shape in the database as a template or choose the empirical mean of the database in the space of currents as a template. This last template is the initialization of our atlas construction: using it saves the time of template centering in Algorithm 4. For each of these alternative templates, we run exactly the same statistical analysis as above. This leads to classification results for the Mahalanobis approach and the SVM which are presented in Fig. 5.8. This shows undoubtedly that our unbiased atlas estimation leads to better classification results than the other approaches, thus meaning a better description of the discriminative features between populations. This shows that rooting the statistical estimations in a rigorous statistical model may lead to slightly more complex algorithms, but with the benefit of more significant features.

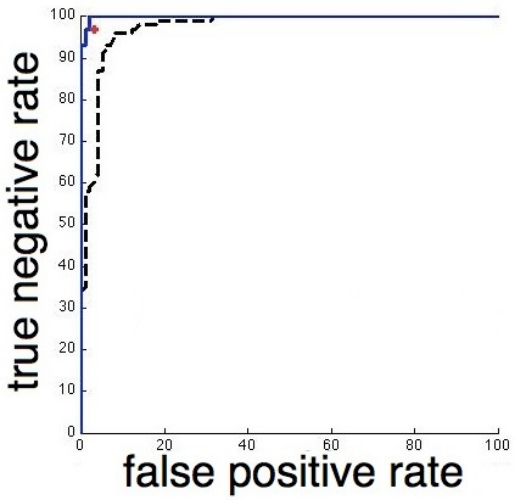


Figure 5.7: Comparison of the classification result for 3 methods on the test data: the difference of Mahalanobis distance (dashed black curve), the linear discriminant analysis (blue curve) and the SVM method (red cross). These methods lead to a mean classification error of 6%, 1% and 3% respectively. This shows that the discriminant features detected on the training data generalize well to the whole population.

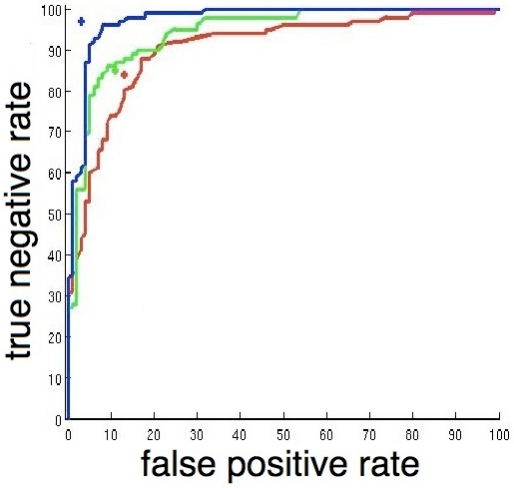


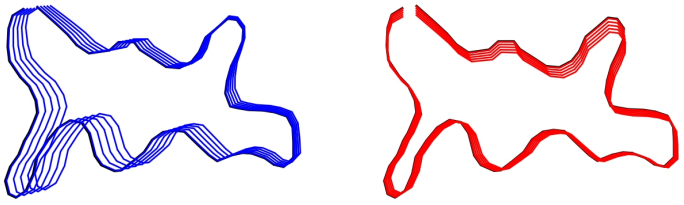
Figure 5.8: Group classification based on different templates: a particular shape of the training dataset (red), the mean current (green) and our unbiased template (blue) is used as template. The prediction is based on the difference of Mahalanobis distances for each class (ROC curves) or on SVM (crosses). In this case, our unbiased template estimated *consistently* with the deformations has the better discriminative power.

To go one step further, we use the classifier based on the Mahalanobis distances with only one mode per class ($N_{\text{modes}} = 1$). We choose these two modes so that the mean classification error was minimized. This allows us to achieve a classification error of 9% (compared to the 3% error rate achieved with the same method using all 26 modes). In some sense, these two modes are the second order features (after the mean) which best discriminate between the two populations. deformation of the template according to these two modes gives a visual representation of this discriminative feature. As shown in Fig. 5.9, this emphasizes the torque of the upper-right part of the shape (taking the blue class as reference).

At this stage, many other statistical computations may be conducted to quantify and describe the differences between both classes. Our purpose here was only to illustrate the potential of our method for the analysis of anatomical shapes. We have proposed a consistent framework to estimate template, deformations of this template within the population and residual shapes (which were not used in this section, but will be used in Chapter 7). The results on simulated examples tend to show that such an atlas may put into evidence significant anatomical differences between populations, although these differences are hidden by the normal variability within the populations.

5.6 Perspectives

In this chapter, we proposed an algorithm to construct atlases from a set of shapes. The purpose of such atlases is to infer from the observations a prototype shape called template and measure how this template varies in the population. This variability is described by two terms: a geometrical and a texture part. This method should be seen as a first attempt to analyze the geometrical variability of 3D geometrical structures, without imposing strong priors on the nature of the variability one is looking for. This framework is very versatile: it can be applied to every objects that can be modeled as currents: not only curves and surfaces, but also unstructured point sets, volumes, 2D or 3D images as shown in Chapter 1.



$7th$ mode of the blue class $15th$ mode of the red class

Figure 5.9: The two most discriminative modes for each class. They highlight a torque at the upper-right part of the shape (if consider the blue class as reference). These two modes, in addition to the means of Fig. 5.6, enables classify correctly 91% of the test data. This shows how the proposed statistical analysis may detect reproducible features across the observations, even if the discriminant signal may be hidden by a high variability within the populations. The signal highlighted here cannot be detected visually from the input shapes in Fig. 5.4.

The results on real anatomical data in the next chapters will reveal the potential of the method to capture significant anatomical features. This framework, however, raises several questions from a computational and statistical point of view.

5.6.1 Convergence and consistence

The estimation of the atlas consists in an alternated minimization procedure: (1) the registration of the template to every subject and (2) the update of the template according to the deformation. It has been proved in [Glaunès 2005] that the registration algorithm converges to a (possibly local) minimum of the matching criterion. The criterion for the update of the template is convex and the gradient descent converges therefore to the minimum of the criterion. Thus, each step converges individually. However, it has been shown in [Allassonnière 2007] that such an alternated minimization does not converge to a maximum of the likelihood of the observations. This seems to be particularly critical for low signal to noise ratio. It has been shown in [Allassonnière 2008, Allassonnière 2009] that stochastic algorithms using Monte Carlo Markov Chain methods may be used to achieve one local maximum of the observed likelihood (i.e. which integrates over the hidden variables such as the deformations). However, adapting such algorithm in our case remains challenging due to the very high dimension of the spaces we are dealing with. This computational burden becomes even more critical if one wants to estimate the hyper-parameters (parameters of the kernel for instance) along with the atlas.

Besides the convergence of the algorithm, one wants to study the consistency of the estimation. The question is to know whether the estimated atlas (and in particular the template) converges when the number of subjects increases. In this forward setting, the consistence of the atlas has been proved in [Allassonnière 2007] for images and small deformations. Adapting such proofs for non-linear deformations and currents remains challenging.

5.6.2 Toward multinomial priors or mixture models

In this framework, the atlas is estimated in the least square sense, namely by considering that the covariance of the Gaussian priors is of the form $\sigma^2 k(x,y)\mathrm{I}_3$ where σ^2 is a *fixed* variance and $k(x,y)\mathrm{I}_3$ and *isotropic* kernel for both the initial speed vector fields and for the currents. However, the subsequent PCA shows precisely a strong anisotropy of the covariance of the deformations or of the residual currents. In general, the normal anatomical variability could be described with much fewer directions that the dimension of the space of initial vector speed or currents (of very high or infinite dimension). One would like that the algorithm estimates also these directions, instead of considering the metric on the space of vector fields and currents as priors. Nevertheless, such approach is challenging due to the very high number dimension of the spaces (typically 10^6 or higher) compared to the number of available data (of order 10^3 at most). One could also imagine to use the result of the PCA as a prior for a re-estimation of the atlas. The convergence and relevance of such an alternated procedure has still to be investigated.

In the example presented in Section 5.5, we estimate a common template for the data of the two classes and perform two distinct PCA on the deformations. Estimating a common template has the advantage to use the total number of available data for the construction of the atlas. This is particularly critical when only few samples per class are available. However, to better discriminate between the two classes, one would like to replace the unimodal Gaussian prior by a bivariate Gaussian law. More generally, one would like to use a multivariate Gaussian laws with only a soft prior on the number of modes. This would be useful to detect possible consistent subtypes within a given group. One may also extend the framework in order to allow several number of templates along the lines of [Allassonnière 2009, Sabuncu 2008]. However, one needs to investigate the relevance of such approaches with a very limited number of data.

A tremendous amount of work has been published in the literature on the topic of mixture of models or multivariate Gaussian prior in a Bayesian framework. However, most of this work does not directly apply in this context due to first the highly non-linear nature of the problem (at least as far as the deformations are concerned) which induces high computational costs and second the very high dimension of the variables to be estimated compared to the number of available data.

Notice also that other alternative approaches are possible. Instead of estimating a deterministic atlas, one may estimate a probability map of different classes of tissues, like in [Ashburner 2009] for instance.

5.6.3 Choice and estimation of the parameters

The decomposition between geometry and texture depends highly on the parameters of the algorithm, in particular the spatial scale of the currents λ_W, the scale of the deformation λ_V, the trade-off between regularity and fidelity-to-data γ and the sparsity number τ.

The spatial scale of the currents λ_W gives the typical scale at which anatomical features are taken into account. Far below this scale, geometric variations are considered as noise. Therefore, it must be compared with the size the features of interest of the input shapes. As a consequence, the estimated template is the representation of a shape with no geometrical

details at a scale much smaller than λ_W. Moreover, λ_W is also used to drive the registration of the shapes. As show in Fig. 3.16 the pieces of shapes to be compared should not be located much further than λ_W one from the other so that the norm of currents is sensitive the shape dissimilarity. If there is a big difference between the distance between shapes and the scale at which one want to compare these shapes, then a multi-scale analysis has to be investigated. In our experiments, though, this was not the case once shapes have been rigidly registered (as a preprocessing).

The scale of the deformation λ_V measures the typical scale at which a structure may deform. It gives the typical distance at which points move consistently and therefore controls the rigidity of the deformation (or its number of degrees of freedom). It must be compared with the diameter of the data. If λ_V is large compared to the diameter of the data, the deformations would move the shape almost rigidly. On the contrary, small values of λ_V will favor uncorrelated motion of small parts of the shape while increasing the cost of a rigid motion.

The trade-off γ is the ratio σ_W^2/σ_V^2 in Eq. (5.3.2): it balances the fidelity-to-data against the regularity of the deformation. It controls the decomposition of the variability into smooth variations and residual perturbations which remain after registration. If γ tends to infinity, deformations are more and more constrained to remain close to the identity (no deformation), more and more variability are captured by the residuals, less and less by the deformations.

The sparsity threshold τ is a fraction of the standard deviation of the input shape, which determines the approximation error as explained in Chapter 3. It controls the numerical precision of the estimation of the atlas. As a fraction of the standard deviation, this value does not depend much on the data and can be fixed for every application. Usually, it is set to 5%. A value of 1% may be used, but it increases the number of points of the template and therefore slows down the computation of the atlas. The smaller, the more precise the computations, the more points in the representation of the template, the slowest the computation of the atlas.

For a given choice of these parameters, the algorithm will determine the boundary between geometry and texture which maximizes the likelihood of the observations. One would like that the algorithm estimates also these parameters, in order to find the optimal decomposition between geometry and texture in the sense of a maximum likelihood. A first attempt has been done in this direction in [Allassonnière 2007] for images and small deformations, which needs to be extended for large deformation and currents.

In this current stage of the work, it is up to the user to estimate manually the values of these parameters, depending on the data and the applications. These values may be assessed according to the size of the data and the size of the features of interest. In the next chapters, we will give the typical values of these parameters for each experiment. In practice, these parameters were adjusted by computing registrations between pairs of input data. More discussions about the impact of each of these parameters on the registrations may be found in Chap. 6. One could also run the construction of the atlas for a large range of possible parameters and select the parameters that enables to achieve the best classification ratio for instance. But this would require to have large data-sets with clearly

distinct classes. Moreover, the number of parameters to be adjusted and the computational time for the atlas construction limit the use of this approach.

5.6.4 Dimension vs. number of samples

The main challenge for the statistical analysis of the deformations and the residues is definitely the estimation of variables whose dimensions may be several thousand times higher than the number of available observations. Statistical studies of convergence and consistence usually focuses on the limit when the ratio between the number of samples and the dimension of the space tend to infinity. Here, we have to deal with the opposite limit: this ratio tends to zero.

Some work has been done in this context. In [Hall 2005], the authors noticed that the simulations of random Gaussian variables tend to lie deterministically at the vertices of a regular simplex as the dimension of the space tends to infinity. In [Marron 2007], the authors proposed a distance weighted discrimination methods which may overcome the limitations of Support Vector Machines in the paradigm of high-dimension low-sample size statistical analysis. In [Blanchard 2008], the authors proposed a strategy to estimate a sequence of subspaces with increasing dimension for binary classification. They show, in particular, that the convergence rate of their method is better than the one obtained via Support Vector Machines.

We perform some preliminary experiments on the simulated example of Section 5.5. We test a multi-scale analysis of the deformation by performing successive kernel-PCA with a Gaussian kernel K^V with decreasing standard deviation λ_V. This analyses the initial vector fields in different frequency bands. A PCA can be performed in each frequency band. This method seems to have two main advantages: first it shows at which scale the discriminative feature occurs and second the classification power of the method in these particular subspaces seems to be much better than the initial classification method which does not select any particular frequency bands. Future work should focus on giving a theoretical justification of such a method.

Chapter 6

From sulcal lines registration to variability measures of the cortex surface

In this chapter, we focus on the estimation of the variability of the cortex surface from the localization of the sulcal lines. We emphasize here the ability of diffeomorphic registration of currents to consistently integrate spatial constraints. Using the terminology of Chapter 5, we estimate the variability of the cortex using the forward model, assuming that the template is given as a prior. We focus on statistics on deformations, since the residual currents do not highlight anatomically meaningful information.

This chapter is self-contained and can be read independently from the rest of this thesis. It has been published in Medical Image Analysis [Durrleman 2008c]

Contents

Abstract

In the context of computational anatomy, one aims at understanding and modelling the anatomy of the brain and its variations across a population. This geometrical variability is often measured from precisely defined anatomical landmarks such as sulcal lines or meshes of brain structures. This requires first to compare geometrical objects without introducing too many non realistic priors and, second, to retrieve the variability of the whole brain from the variability of the landmarks.

We propose, in this paper, to infer a statistical brain model from the consistent integration of variability of sulcal lines. The similarity between two sets of lines is measured by a distance on currents that does not assume any type of point correspondences and it is not sensitive to the sampling of lines. This shape similarity measure is used in a diffeomorphic registrations which retrieves a single deformation of the whole 3D space. This diffeomorphism integrates the variability of all lines in a as spatially consistent manner as possible.

Based on repeated pairwise registrations on a large database, we learn how the mean anatomy varies in a population by computing statistics on diffeomorphisms. Whereas usual methods lead to descriptive measures of variability, such as variability maps or statistical tests, our model is generative: we can simulate new observations according to the learned probability law on deformations. In practice, this variability captured by the model is synthesized in the principal modes of deformations. As a deformation is dense, we can also apply it to other anatomical structures defined in the template space. This is illustrated the action of the principal modes of deformations to a mean cortical surface.

Eventually, our current-based diffeomorphic registration (CDR) approach is carefully compared to a pointwise line correspondences (PLC) method. Variability measures are computed with both methods on the same dataset of sulcal lines. The results suggest that we retrieve more variability with CDR than with PLC, especially in the direction of the lines. Other differences also appear which highlight the different methodological assumptions each method is based on.

6.1 Introduction

From the ever growing databases of medical images, there is considerable interest in extracting the most relevant information to characterize normal anatomical variability within a group of subjects as well as between different groups, to detect anatomical abnormalities, to classify new images according to their pathologies, and for understanding disease progression. However, modeling the individual anatomy and its normal variability across a population is difficult as there are commonly no physical models for comparing different subjects, and anatomical shapes are complex and require large number of degrees of freedom to model adequately. Moreover, anatomical landmarks such as curves or surfaces embedded in $\mathbb{R}^3$ as well as deformations of the 3D space do not belong to usual vector spaces. Defining statistical models is therefore difficult and specific tools have to be developed to accurately measure anatomical variations. If anatomic variation were better understood, tools encoding these variations could have a significant impact in neuroscience to minimize the influence of the anatomical variability in functional group analysis, and in clinical medical image analysis to better drive the personalization of generic models of the anatomy (called also template, atlas or prototype in the literature).

Instead of analyzing the anatomical variability directly in the 3D intensity space, it is often preferable to extract precisely defined anatomical landmarks such as sulcal lines [Thompson 1996a, Mangin 2004a], cortical surface models [Fischl 2001, Tosun 2005], or models of some sub-cortical structures [Hazlett 2005, Vaillant 2007]. The data to be

analyzed are thus curves, surfaces or volumes represented by structured or unstructured point sets. The first attempts to compare such shapes was based on defining correspondences between points [Zhang 1994, Chui 2003]. However, the sampling of two different geometric subjects can be so different that such a correspondence assumption introduces a bias that often hides the "real" underlying geometrical variability. To overcome this difficulty, some authors proposed to measure variations of some features extracted *a priori* such as length, area, volume, complexity, principal modes of variation of the cloud of sampled points, etc. See [Paus 2001] for instance. Although these approaches, that derive scalar measures from structure models, are relatively easy to set up from a computational point of view, they fail to capture fine geometrical variations between subjects like for instance a twisting of the extremity of a sulcus, which cannot be readily described by a set of *a priori* selected features. Also, data analysis often proceed by computing dense displacement fields that encode variations in shape and volume among individuals, often based on deformable registration of shapes. The deformation that maps one shape onto another has been proved to be useful for measuring significant anatomical variability among different subjects [Fillard 2007c, Vaillant 2007, Ashburner 1998, Durrleman 2007]. Due to the presence of noise and of sampling effects, it may be advantageous to allow a trade off between the regularity of the deformation and the precision of the matching, instead of exact matching. This raises the need to develop a consistent deformation framework and a shape similarity measure that does not rely on point correspondences nor on features selected *a priori*.

In this perspective, one interesting framework consists of modeling geometrical primitives as currents [Vaillant 2005, Vaillant 2007, Durrleman 2007] (see Chapter 1). The idea is to characterize shapes via vector fields, which are used to probe them. For instance, a surface is characterized by the flux of any vector field through it, a line by the path integral of any vector field along it. Conversely, associating a flux to any vector field specifies an object which is more general than a surface or a curve and which is called a current. This way of embedding shapes in a Hilbert space allows one to define algebraic operations such as addition or averaging, and to measure distance between geometrical primitives via an inner product that does not assume a specific type of point correspondence. Discrete and continuous objects are handled in the same setting, offering a way to measure the sampling quality and to guarantee numerical stability. This framework has been used to compute and visualize mean lines and surfaces, and to perform principal component analysis on datasets of such primitives, suggesting the efficiency and generality of the approach [Durrleman 2008b] (see Chapter 1 and 3).

However this similarity measure is too weak to capture the broad range of possible differences between shapes: it is beneficial to couple it with a deformation framework. The large deformation diffeomorphic metric mapping (LDDMM) framework [Trouvé 1998, Grenander 1998, Dupuis 1998, Miller 2002, Miller 2006] is ideal for this task as shown in [Vaillant 2005] and [Glaunès 2006] although it might be possible to adapt other diffeomorphism registrations method proposed for images (e.g. [Ashburner 2003, Avants 2006]). The deformation that matches a pair of shapes is sought within a group of regular diffeomorphisms in order to optimize a trade-off between the regularity of the de-

formation and matching accuracy, as measured by the dissimilarity measure on currents [Vaillant 2005, Durrleman 2007] (see Chapter 4). As a result, the registration decomposes the differences between two shapes into (1) a deformation that captures a "global" misalignment and (2) a residual term (representing the difference between the registered shape and the target shape) that contains possible non-diffeomorphic variations as well as physical and numerical noise. In the approach followed here, we perform our statistical analysis of shape on the deformation term only. Our results on a dataset of sulcal lines show that this method can be used to detect and characterize anatomical variability within a group of subjects.

Moreover, this diffeomorphic framework enables to register multiple objects in a spatially consistent way. Indeed, a dataset of anatomical landmarks often consists of a set of several shapes (e.g. a set of sulcal lines or set of meshes representing different subcortical structures for instance [Mangin 2004a, Duchesnay 2007, Gorczowski 2007]. If one set of manifolds, such as a distributed set of sulcal landmark curves, is embedded in another manifold which also varies, such as the cortical surface, one often aims to measure the variability not only of the embedded landmarks but also of the whole underlying brain surface or 3D brain volume. The framework based on currents enables precisely to define a distance between multiple objects sets even if they are not labeled or if all subjects have not the same number of objects. (In these cases the distance will just be less precise than for labeled objects). The diffeomorphic framework in turn finds a single deformation of the underlying image domain that integrates the variability of all objects in as consistent manner as possible. By contrast, several methods such as in [Fillard 2007c] analyze the variability of each shape individually; there is a need for an extrinsic extrapolation scheme to retrieve a variability in the space between the objects. The approach proposed in [Hellier 2003, Cachier 2001] makes a model of deformation that has constraints on sulci, cortex, and whole brain, all within a single optimization framework. Earlier work like [Thompson 1996c] just used the matching of low order manifolds first, and used these as hard constraints or boundary conditions on subsequent mappings one dimension higher. Altough there are many other registration frameworks in the literature, we focus in this paper on the current-based diffeomorphic registrations (CDR) to build brain variability models.

This paper aims to present and discuss such a framework, based on diffeomorphic registration of currents, in the case of curves. We apply the method on a dataset of labeled sulcal lines to infer the variability of the brain surface within a population. In Section 2, we explain the framework for registration of sulcal lines. How this differs from the point-wise line correspondence (PLC) approach, proposed in [Fillard 2007c], is discussed in depth from a methodological point of view. In Section 3, we perform a statistical analysis of the underlying brain surface based on these registrations. This allows us to measure and visualize how the brain surface varies in a population. A comparison with the results obtained in [Fillard 2007c] on the same database illustrates some of the different methodological assumptions each method is based on.

6.2 Registering sets of 3D curves

Registering a set of 3D curves L_0 onto another set of 3D curves L_1 can be formulated as the task of looking for the most regular deformation ϕ that transports all curves of L_0 and best matches the curves of L_1. We follow here the approach introduced in [Glaunès 2005]: the unknown deformation is sought in a subgroup of diffeomorphisms and its regularity is measured based its distance to the identity (i.e. no deformation), the similarity measure is computed by embedding the curves into a space of currents. As it is common practice in deformable image registration, we find the registrations by minimizing a cost function that balances the regularity of the deformation against the matching fidelity.

6.2.1 Non-parametric representation of curves as currents

The space of currents is a vector space that may be equipped with a norm that measures geometrical similarity between curves. In this space, curves could be discrete or continuous and may consist of several different parts. All these objects are handled in the same setting and inherit many interesting mathematical properties: linear operations, distance, convergence, etc. Moreover, this definition of distance between curves does not make any assumption about point correspondences, even implicitly. This framework differs therefore from usual methods such as that in [Joshi 2000] where landmark matchings are performed or those in [Chui 2003, Granger 2002, Cachier 2001] where curves are considered as unstructured point sets and different kind of "fuzzy" correspondences are assumed. We recall here how to build a space of currents and how to compute explicitly a similarity measure on curves. For more details on the theory we refer the reader to [Vaillant 2005, Glaunès 2005, Durrleman 2008b] and references therein (see Chapter 1).

In the framework of currents, curves are seen via the way they integrate vector fields. Any continuous curves or any finite set of polygonal lines (denoted here generally L) can be characterized by the path integral of any vector field ω along it:

$$\forall \omega \in W, \quad L : \omega \longrightarrow \int_L \omega(l)^t \tau(l) dl \tag{6.2.1}$$

where $\tau(l)$ is the unit tangent vector (defined almost everywhere) of L at point l and W is a *test* space of smooth vector fields (See Fig. 6.1). More generally, a current L is defined as a continuous linear mapping from the test space W to $\mathbb{R}$. As a set of mappings, the space of current (denoted W^*) is a **vector space**: $(L_1 + L_2)(\omega) = L_1(\omega) + L_2(\omega)$ and $(\lambda L)(\omega) = \lambda(L(\omega))$. For curves, this means that the path integral along two curves is the sum of the path integral along each curve: the addition corresponds therefore to the union of the two curves. Scaling a curve means scaling the path integral along the curve.

Suppose now, that we can provide the test space W with a norm $(\|.\|_W)$ that measures the regularity of the vector fields in W. We can define then a **norm of a current** L as the supremum path integral of any regular vector field (i.e. $\|\omega\|_W \leq 1$) along L: $\|L\|^2_{W^*} = \sup_{\|\omega\|_W \leq 1} |L(\omega)|$. The distance between two curves ($\|L - L'\|^2_{W^*} = \sup_{\|\omega\|_W \leq 1} |L(\omega) - L'(\omega)|$) is therefore obtained for the vector field that best separate the two lines, in the sense that the difference between the path integrals along both curves is

the largest possible. This distance between curves is geometric: *it does not depend on how curves are parametrized and does not assume any point-correspondences between curves.*

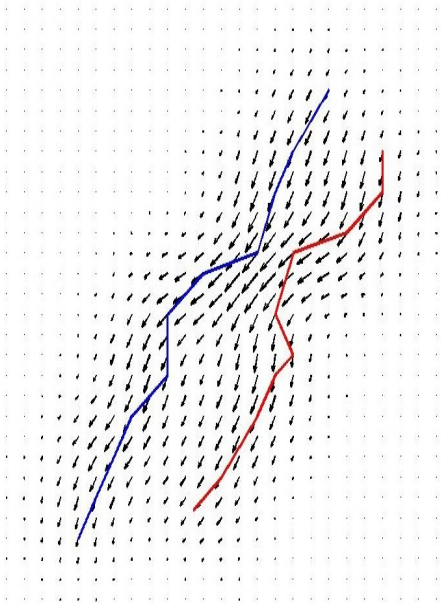


Figure 6.1: Measure of dissimilarity between lines modeled as currents: given two lines L and L' (in red and blue) we compute the difference between the path integral of a vector field ω (here drawn in black) along both lines. The maximum difference obtained when ω varies among all possible regular vector fields (i.e. $\|\omega\|_W \leq 1$) is a measure of the geometrical dissimilarity between the two lines. In this way, we define a distance between shapes without assuming point correspondences. The more we allow the test vector fields ω to have high spatial frequencies, the more finely we measure geometrical differences. In this application, manual delineation of the sulci is typically accurate to within a 1-2 mm Hausdorff distance to a gold standard developed from multiple raters, so the matching of features at a slightly coarser scale than this is empirically reasonable.

For computational purposes, we suppose, from now onward, that W is a reproducible kernel Hilbert space (r.k.h.s.) with kernel K^W (see [Aronszajn 1950, Saitoh 1988] for details): vector fields in W are convolutions between any square-integrable vector fields and the kernel. This framework is general and includes for instance radial basis functions. In this setting, the vector space of currents is a **dense span** of the set of all delta Dirac currents δ_x^α, which is defined by: $\delta_x^\alpha(\omega) = \omega(x)^t\alpha$ for any $\omega \in W$. A Dirac current may be seen as an oriented segment α entirely concentrated at one point x. Although a curve has an infinite set of tangents, polygonal lines may be approximated in the space of currents by a finite sum $\sum_k \delta_{c_k}^{\tau_k}$ where c_k is the center of the k^{th} segment and τ_k the tangent of the line at c_k.

In this setting, it has been shown [Vaillant 2005, Glaunès 2005] that the norm on W^* comes from **an inner product** $\langle .,. \rangle_{W^*}$. On basis elements, this inner product is $\left\langle \delta_x^\alpha, \delta_y^\beta \right\rangle_{W^*} = \alpha^t K^W(x,y)\beta$. The inner product between two polygonal lines $L = \sum_{i=1}^n \delta_{c_i}^{\tau_i}$ and $L' = \sum_{j=1}^m \delta_{c'_j}^{\tau'_j}$ (*where n is not necessarily equal to m*) is therefore given by:

$$\langle L, L' \rangle_{W^*} = \sum_{i=1}^{n}\sum_{j=1}^{m} (\tau_i)^t K^W(c_i, c'_j)\tau'_j \tag{6.2.2}$$

This enables to compute *explicitly* the distance between two curves:

$$d^2(L, L') = \|L' - L\|^2_{W^*} = \|L\|^2_{W^*} + \|L'\|^2_{W^*} - 2\langle L, L' \rangle_{W^*} \tag{6.2.3}$$

In our applications, we choose K^W to be isotropic and Gaussian: for all points $x, y \in \mathbb{R}^3$, $K^W(x,y) = \exp(-\|x-y\|^2/\lambda_W^2)\mathrm{Id}$.

We observe that the distance between two curves (Eq. (6.2.3)), induced by the Hilbertian inner product Eq. (6.2.2), measures geometrical differences both in pose and shape (See Fig. 6.1). If the points on one curve are at a distance much larger than λ_W from the points on the other curve, then curves are considered as orthogonal ($\langle L, L' \rangle_{W^*} \sim 0$) and their distance is large whatever their respective shapes are. By contrast, if two parts of the curves are located within an area of size λ_W, local alignment of the tangent vectors is taken into account by the inner product within the sums in Eq. (6.2.2), thus measuring shape variations. Furthermore within this area, variations at a scale much smaller than λ_W are not taken into account thanks to the smoothing effect of the kernel. Such variations are considered as noise. Finally, this distance captures first misalignment and then shape dissimilarities until a noise level quantified by λ_W is reached. Used as a data fidelity term, this distance integrates a denoising process, to some extent, into the modeling, preventing systematically overfitted registrations.

6.2.2 Diffeomorphic registration

We use here the large deformation framework founded in the paradigm of Grenander's group action approach for modeling objects (see [Grenander 1994, Miller 2006, Glaunès 2006, Marsland 2004a]). This framework enables to find a *globally consistent* deformation of the underlying space that best matches the sets of lines. This differs from [Fillard 2007c] where each line is registered individually without assuming spatial consistency of the displacement field between lines.

We build our deformations as diffeomorphisms ϕ_1^v, solutions at time $t = 1$ of the flow equation:

$$\frac{\partial \phi_t(x)}{\partial t} = v_t(\phi_t(x)) \tag{6.2.4}$$

with initial condition $\phi_0 = \mathrm{id}_{\mathbb{R}^3}$ (i.e., $\phi_0(x) = x$: no deformation). The time-varying vector field $v = (v_t)_{t\in[0,1]}$ is the speed field in the Lagrangian coordinates. We suppose, from now onwards, that at every time t, v_t belongs to a r.k.h.s. V with kernel K^V. We denote $\|.\|_V$ the norm on this space that measure the spatial regularity of the vector field. To measure the regularity of the final diffeomorphism, we integrate the regularity of this speed field along time [Grenander 1998, Miller 2002]: $v \in L^2([0,1],V)$: $\|v\|^2_{L^2([0,1],V)} = \int_0^1 \|v_t\|^2_V \, dt$.

Our registration problem is to map a set of n labeled sulcal lines $L_0 = \cup_{i=1}^n L_{0,i}$ to another labeled set $L_1 = \cup_{i=1}^n L_{1,i}$. We must find therefore a time-varying vector fields $(v_t)_{t\in[0,1]}$ that minimizes the following cost function J:

$$J(v) = \sum_{i=1}^{n} \|\phi_{1*}^v L_{0,i} - L_{1,i}\|^2_{W^*} + \gamma \|v\|^2_{L^2([0,1],V)} \tag{6.2.5}$$

where γ is a trade-off between the regularity of the deformation and the fidelity to data.

$\phi_* L$ represents the geometrical transportation of the curve L by the deformation ϕ. This formulation is compatible with our framework based on currents. The path integral of ω along a deformed curve $\phi(L)$ equals the path integral along L of the pulled-back vector

field: $\phi^*\omega(x) = (d_x\phi)^t\omega(\phi(x))$. This is a change of variables formula within Eq. (6.2.1), from which we deduce a general action of diffeomorphism on any currents: $\phi_* L(\omega) = L(\phi^*\omega)$. In particular, on basis element, this gives: $\phi_*\delta_x^\alpha = \delta_{\phi(x)}^{d_x\phi(\alpha)}$: an infinitesimal segment α at point x is transported into $\phi(x)$ and deformed by the Jacobian matrix: $d_x\phi$. Combined with Eq. (6.2.2) and (6.2.3), this makes computable the fidelity to data term in Eq. (6.2.5), once a deformation is given.

To minimize the cost function in Eq. (6.2.5), we take advantage of a dimensionality reduction property. Although the vector fields v_t are dense, it has been shown (for instance in [Miller 2002, Vaillant 2005]) that, in case of discrete curves, $L_0 = \sum_{i=1}^N \delta_{c_i}^{\tau_i}$, the minimum of Eq. (6.2.5) is achieved for a vector field v_t which interpolates the trajectories of the points of L_0:

$$\forall x \in \mathbb{R}^3, v_t(x) = \sum_{i=1}^{N} K^V(x, c_i(t))\alpha_i(t) \tag{6.2.6}$$

where the momenta $(\alpha_i(t))$ is a set of N vectors (in 3D) for each time t and $c_i(t) = \phi_t(c_i)$ are the trajectories of the points of L_0. Based on Eq. (6.2.4) evaluated at $x = c_i$, these trajectories depend only on the momenta $\alpha_i(t)$. This means that the minimizing dense vector field v_t is entirely determined by a set of $3N$ parameters for each time t. Once the time interval $[0, 1]$ is discretized, the cost function Eq. (6.2.5) depends on a finite set of parameters and may be therefore minimized by a standard gradient descent scheme. All computational details of this gradient descent can be found in [Vaillant 2005, Glaunès 2005].

For K^V we choose an isotropic and Gaussian kernel with standard deviation λ_V. This parameter controls the regularity of the speed vector field v_t at each time t and hence the regularity of the final diffeomorphism. λ_V defines roughly the scale of the diffeomorphism's spatial consistency (called also rigidity). This is then the scale at which the underlying deformation tries to integrate the geometrical information. If λ_V is much smaller than the distance between lines, the final deformation can vary dramatically in space, each piece of lines are then matched almost independently and the deformation is negligible outside the data. On the contrary, the greater λ_V, the more consistently the deformation tries to explain the variation of each lines, jointly with less and less precise matching.

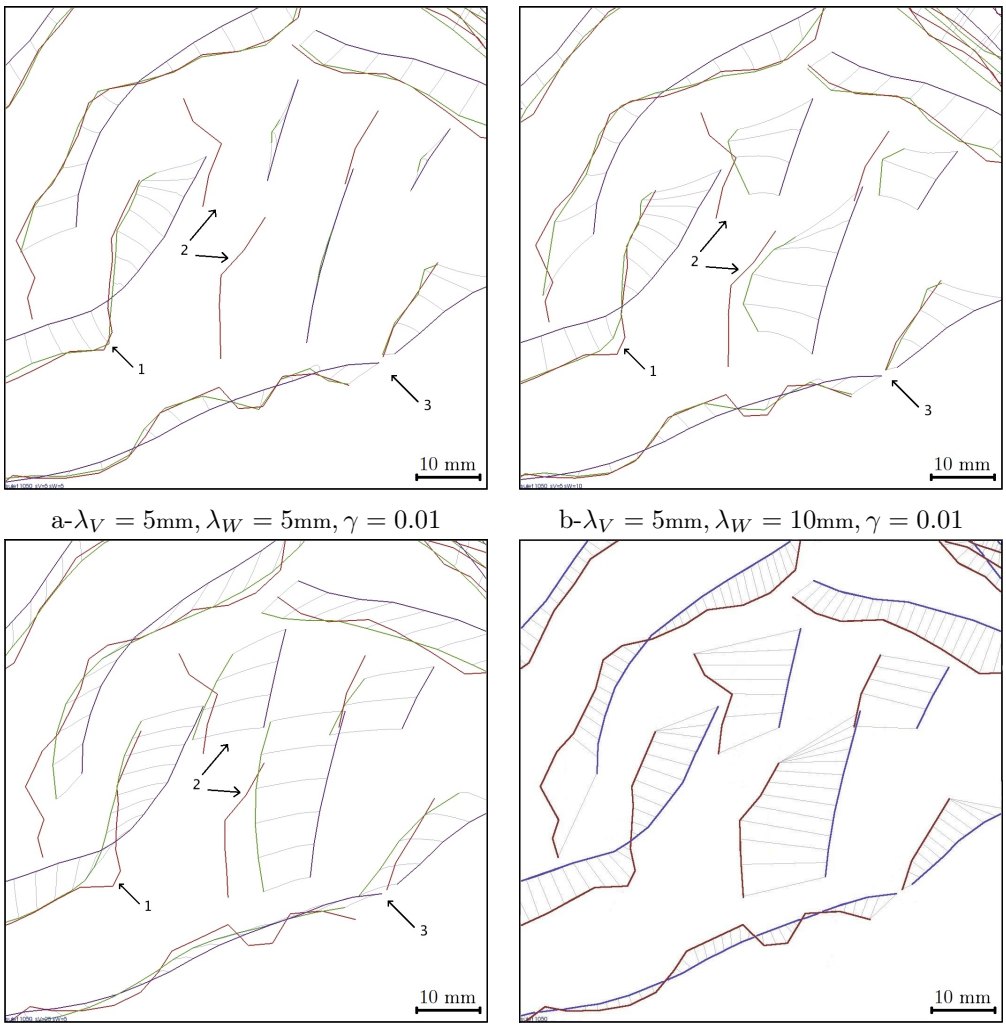
6.2.3 Registration results

As part of a collaborative project involving the Asclepios-LONI associated team `Brain-Atlas`, we used a dataset consisting of cortical sulcal landmarks (72 per brain) delineated in a large number of subjects scanned with 3D MRI (age: 51.8 +/- 6.2 years). In order to compare our measures of variability with the ones of [Fillard 2007c], we used the same set of 72 mean lines that the authors of [Fillard 2007c] computed from the same dataset. For 34 subjects in the database, we register this set of mean lines onto every individual subject's set of sulcal lines. The registrations were computed by J. Glaunès' algorithm detailed in [Glaunès 2005]. This algorithm depends on the 3 parameters: λ_V, λ_W and γ. To understand the impact of these parameters and the specificity of this current-based diffeomorphic registrations (CDR) with respect to a pointwise line correspondence (PLC) method [Fillard 2007c], we ran the registration algorithm for several different sets

Figure 6.2: Registration of the mean lines set (in blue) towards one subject's lines set (in red). A unique deformation transports all the mean blue lines to the registered green lines. The spatial consistency constraint as well as the smoothing effect of the norm of currents prevents overfitted registrations from occurring. The residuals (i.e., the difference between the registered green lines and target's red lines) contains physical and numerical noise as well as possible non-diffeomorphic variations. They are considered here as noise: the statistics on brain shape will rely only on the diffeomorphism. A movie of this deformation can be seen at first author webpage: www-sop.inria.fr/asclepios/personnel/Stanley.Durrleman.

of parameters.

Figure 6.3-a,b and c show for different parameter values the registrations in the superior temporal area of the cortex in the right hemisphere (view from inside), the brain faces to the left of this figure, and region surrounding the Sylvian fissure, on the lateral surface of the right hemisphere, is magnified. From Fig. 6.3-a to 6.3-b, λ_W is doubled: greater variations are considered as noise and the matching is less precise (area 1). However, when λ_W is too small, lines with few sampled points are not matched correctly (area 2). Small curves have small weight in the data fidelity term and matching them is not worth the cost of the deformation: the algorithm is locally in a minimum. Greater λ_W makes distance loss larger due to the curve's motion to its target. The local minimum issue is avoided. In both cases (a and b), the deformation kernel's size is very small: $\lambda_V = 5$mm whereas the diameter of the brain is about 120mm. The speed vector field is highly irregular and each curve is matched almost independently. This is particularly obvious in area 3, which is close to the supramarginal and angular gyri of the temporo-parietal cortex, where the speed vector field varies dramatically between two close points that belong to two different curves. There is almost no deformation between the curves. By contrast, in figure 6.3-c, $\lambda_V = 25$mm and the deformation tends to explain the sulcal lines variability with a consistent deformation of the underlying space. This makes the speed vector field more regular, as in area 3. Whereas λ_W is the same as in figure 6.3-a, small curves are now matched since they are "pushed" by the large curves in the surroundings. To match the larger curves, the space must deform consistently in this area with the effect of moving small curves to their target. This global constraint also leads to larger residual matching errors than in figure 6.3-a (area 1). Such residual errors contain both geometrical noise on lines (quantified by λ_W) and some variability that cannot be explained consistently with other curves in a neighborhood of size λ_V, which is regarded as noise in the model. Besides λ_V and λ_W, γ refines the compromise between the regularity of the deformation and the precision of the matching.



a-$\lambda_V = 5\text{mm}, \lambda_W = 5\text{mm}, \gamma = 0.01$ b-$\lambda_V = 5\text{mm}, \lambda_W = 10\text{mm}, \gamma = 0.01$

c-$\lambda_V = 25\text{mm}, \lambda_W = 5\text{mm}, \gamma = 0.01$ d-PLC method

Figure 6.3: Registration of the same subject for three different sets of parameters (a,b,c) and with a pointwise line correspondences approach (d). In these figures, the superior temporal area of the cortex is magnified (arrow 1 points the extremity of the Sylvian Fissure). The parameters influence the precision of the matching (like in area 1), the regularity of the deformation field (area 3) and the way the deformation integrates geometrical information (area 2).

Figure 6.3-d shows, in the same anatomical region, how a pointwise line correspondence method (PLC) set up in [Fillard 2007c] handles the same data. Lines are registered individually and point correspondences are assumed between source and target lines. Extremal points are supposed to be matched. In between points are matched via a closest neighbor procedure after B-spline smoothing and resampling. As no correlations between curves are assumed and no correspondence field computed outside the data, this matching can be seen as an approximation of our registrations when λ_V tends to zero. Since the correspondences

between points do not take noise into account, it is also the limit as λ_W tends to zero but avoiding the local minimum issue. The different way that PLC handles lines matching have important consequences. For instance, the tangential variation of two curves is mainly captured at the extremities of the curves, and minimized elsewhere. With the current approach, these variations are captured more geometrically all along the curves. Moreover, the PLC approach does not take noise into account (in the sense that noisy point correspondence field computed from the manual traces is regarded as true), and does not model any deformation of the space between the curves. As we will see in the next section, PLC approach needs afterwards to handle two additional processing, denoising and extrapolation of the variability measures, to compute brain shape statistics. This method consists therefore of 3 distinct processing steps: matching, denoising, extrapolation, each with its own assumptions and parameters. By contrast, the approach proposed here based on currents' diffeomorphic registrations (CDR) integrates denoising, matching and extrapolation within a single consistent framework. Matching based on currents avoids the need to define a principle for enforcing specific point correspondences. Denoising is performed jointly with the matching while minimizing the cost function. Extrapolation is performed on the basis of a deformation of the underlying biological material. The whole framework is explicitly controlled by 3 parameters λ_V, λ_W and γ that effect a compromise between the different processing steps. This method, however, discounts variability that is not compatible with the modeling. Residual matching errors may hide non-diffeomorphic variations between subjects although one would like to take them into account. Setting the 3 parameters is difficult since they are not independent and determine jointly the final residual matching errors. After extensive experiments, we choose the typical coherence scale of diffeomorphisms $\lambda_V = 25$mm, the typical noise scale on lines $\lambda_W = 5$mm and the tradeoff $\gamma = 0.01$ by visually inspecting the results. Changing these parameters would smoothly affect the typical correlation size of the following variability maps. These values highlight the specificity of our framework, that will explain, in turn, the different variability maps retrieved by our CDR method and the PLC approach on the same dataset as in [Fillard 2007c].

6.3 Statistics on deformations

6.3.1 Tangent-space representation of diffeomorphisms

To compute statistics on deformations, we take advantage of an additional property of the minimizing diffeomorphisms. It has been shown (in [Miller 2006] for instance) that the diffeomorphism retrieved by the minimization of Eq. (6.2.5) is geodesic: among all time-varying vector-fields that enables to go from Id to ϕ_1, the minimizer of Eq (6.2.5) has the smallest norm in $L^2([0,1],V)$. As a consequence, the momenta $(\alpha_i(t))_i$ solve the Euler-Lagrange equations [Miller 2002, Miller 2006]: they are entirely determined by their initial values: $\alpha_i(0)$. This is the usual tangent-space representation as highlighted in [Vaillant 2004] or in [Pennec 2006b] for finite dimensional manifolds. This representation enables to *generate* randomly deformations of L_0: given any set of N vectors α_i^0, we can solve the Euler-Lagrange equations (partial differential equations) to give the momenta at every time: $\alpha_i(t)$. We deduce then from Eq. (6.2.6) evaluated at every $x = c_i$, the speed

of the trajectories of the points c_i of L_0: $v_t(c_i)$. Integrating the flow equation Eq. (6.2.4) enables then to compute the whole trajectories $\phi_t(c_i)$. The generated deformation does not act only on the line L_0 but it is a diffeomorphism of the whole 3D space. Based on the interpolation property Eq. (6.2.6), we can compute the speed (and then the trajectory by Eq. (6.2.4)) of any point x of the space, thus computing the entire diffeomorphism. The purpose of our forthcoming statistical estimations is to learn a law on the momenta α_i^0, so that diffeomorphisms simulated according to this law model the variability within the studied population.

From the previous 34 registrations of the mean lines to each subject's lines, we store 34 sets of initial momenta: (α_i^s) for $i = 1 \dots N$ and $s = 1 \dots 34$ where N is the total number of points c_i within the set of mean lines: $L_0 = \sum_{i=1}^{N} \delta_{c_i}^{\tau_i}$. We define an inner product (resp. a norm) on this space of momenta as the inner product (resp. norm) of its associated dense vector field $v_0(x) = \sum_{i=1}^{N} K^V(x, c_i)\alpha_i(0)$ based on Eq. (6.2.6). Since V is a r.k.h.s. with kernel K^V, the inner product between two sets of $3N$ momenta (from the registration of two different subjects) $(\alpha_i^p)_i$ and $(\alpha_i^q)_i$ is equals to $\sum_{i,j=1}^{N} (\alpha_i^p)^t K^V(c_i, c_j)\alpha_j^q$. In the sequel, we denote this inner product between two vectors in $\mathbb{R}^{3N}$: $\langle \alpha^p, \alpha^q \rangle_{V^*}$. Our statistics on diffeomorphisms are then reduced to statistics in $\mathbb{R}^{3N}$ provided with this inner product.

6.3.2 Mean of deformations

Since the mean lines we used as a template were obtained in [Fillard 2007c], they are not consistent with our registration framework. The deformations are not centered around the identity (i.e. no deformation), so the vectors in $\mathbb{R}^{3N}$ do not have zero-mean. To measure the induced bias, we compute the mean of the initial momenta at each sample: $\overline{\alpha}_i = \frac{1}{34} \sum_{s=1}^{34} \alpha_i^s$. The norm of this bias is given by $\|\overline{\alpha}\|_{V^*}^2 = \sum_{i,j=1}^{N} \overline{\alpha}_i K^V(c_i, c_j)\overline{\alpha}_j$. Numerically we find in our experiments: $\|\overline{\alpha}\|_{V^*} / \sqrt{\frac{1}{34} \sum_{s=1}^{34} \|\alpha^s - \overline{\alpha}\|_{V^*}^2} = 0.39$. This means that the bias is less than 0.4 times the standard deviation, far below the usual 3σ threshold to decide a statistical significance.[1]

We now substract the mean field to each subject's field so that the analyzed data are centered for the following computations of second order statistics.

6.3.3 A Gaussian model on deformations

To compute the covariance structure of the set of deformations, we perform a Principal Component Analysis (PCA) on the set of vectors $\alpha^s \in \mathbb{R}^{3N}$ for each subject s. For this purpose, we build the 34×34 symmetric matrix whose coefficients are $\langle \alpha^p, \alpha^q \rangle_{V^*}$. If $V^1 \in \mathbb{R}^{34}$ is the first eigenvector of this matrix, the first mode of initial momenta is given by: $m_i = \sum_{s=1}^{34} V_s^1 \alpha_i^s$ (the normalization factor has been set to 1 so that $\|m\|_{V^*}^2 = \lambda_1$, i.e. the eigenvalue corresponding to V^1). Given this first mode of initial momenta m, we follow the procedure explained in section 6.3.1, to *generate* the diffeomorphism determined by m. We call this deformation, the first mode of deformation. It illustrates, to the first order, how

[1] Performing a real statistical test would imply the estimation of the number of degrees of freedom (since the initial momenta are not independent) as well as the curvature of the space [Bhattacharya 2003, Oller 1995, Pennec 1999]. This is particularly difficult due to the infinite dimension of the space.

the mean anatomy varies within the population. Since the diffeomorphism is dense, we can apply it not only on the mean lines points but also on a mean cortical surface to which the mean lines are close (Fig. 6.4-middle). This deformation shows the variability between 0 and σ (Fig. 6.4-right). Repeating the procedure for $-\alpha$ give the first mode of deformation between 0 and $-\sigma$ (Fig. 6.4-left). Complete movie of the first deformation mode is available at first author's webpage[2]. This illustrates the generative property of the modelling: the lines and surfaces build by this procedure do not belong to the original database.

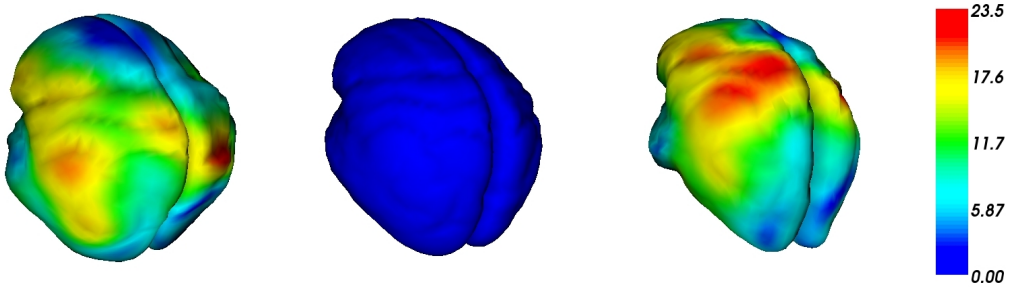


Figure 6.4: First mode of deformation obtained by a PCA on the initial vector speed fields. Original mean brain surface (Center) and its deformation at $-\sigma$ (Left) and $+\sigma$ (Right). Colors measure the displacement of each point along the deformation process (in mm).

These results show that we learn here a complete statistical model of the whole brain surface deformation constrained by the sulcal lines. This differs from other methods that measure only the variability of the sulcal lines. This is possible due to integrative power of the proposed approach: the diffeomorphisms integrate the information of all sulcal constraints to find the most acceptable deformation of the brain volume. As for the modes of deformation, this enables to generate new observations (new brain surfaces) according to the learned probability law on deformations. This determine at least visually how these new observations compare with the original data and therefore understand the variability that the model captured. For this reason, such models are called *generative* models: we not only learn how to factorize an observation into a deformation and residual noise, but also how to reconstruct it. Such models offer an approach to classify new subjects according to characteristics such as gender, handedness, pathologies, etc., and identify systematic differences in anatomy that correlate with these features. Given a previously unseen individual anatomy, we can decompose it into a global deformation driven by its sulcal lines position and a residual noise. Our statistical model tells us how probable such a deformation may be within a given population. Other methods try to retrieve similar correlations but with descriptive statistics such as statistical test for instance [Narr 2007, Hamilton 2007, Luders 2004]. In the PLC approach [Fillard 2007c] no deformation is computed outside the lines. In this framework, the lines variability is computed from the displacement field at each mean lines samples positions. Then an extrinsic extrapolation scheme enables to retrieve a variability

[2]www-sop.inria.fr/asclepios/personnel/Stanley.Durrleman/

measure on the whole brain surface. In this aspect, the link between the observations and the mean surface is broken by the succession of distinct processing: the statistical model is learned on the lines only and the extrapolation scheme does not directly infer a probability model on the brain surface. PLC approach deduces from the observations variability maps on the brain surface (see next section) but the way to reconstruct surfaces from these maps is missing. There are other methods that extrapolate sulcal line deformations to a full cortical surface, based on covariant partial differential equations that are invariant to the surface parametrization [Thompson 2002], based on harmonic mappings that minimize a surface-to-surface deformation energy [Shi 2007, Wang 2005]. Some of these methods even extend the surface deformation to the full volume, using interpolation [Thompson 1996c]. In each of these cases, the differential operator governing the mapping may be regarded, after suitable normalization, as the exponent of a Bayesian prior on the space of allowable deformations, so in a sense there is an assumed probability law that captures the variability and spatial covariance of the mappings in between the explicitly defined landmarks, even when a partial differential equation or variational method is used to interpolate the mappings.

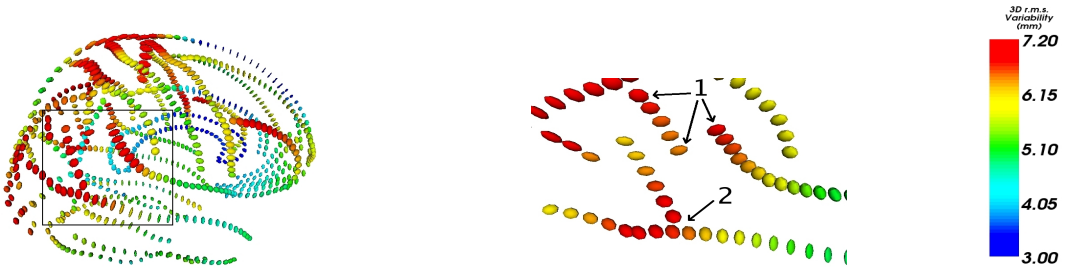
6.3.4 Comparison of variability maps

To compare our CDR-based variability measures with those computed with a PLC approach on the same database we create variability maps similar to those in [Fillard 2007c]: in absence of generative models, PLC approach performs such descriptive statistics. At each point x of the mean surface, we computed the covariance matrix of the 34 initial speed vectors $v_0(x)$, computed with Eq. (6.2.6) for each set of initial momenta. These 3×3 matrices (called also tensors) show how locally one point is varying among the studied population, as proposed in [Thompson 1996b, Thompson 1998]. We notice that this variance contains less information than the former principal component analysis. Here each point are considered independently whereas the PCA takes into account the correlations of all points' motion together. Moreover, due to the diffeomorphic approach, the initial vector field is dense and no extrinsic extrapolation scheme is required to compute the covariance matrices at each point of the mean surface. By contrast, PLC approach computes these 3×3 matrices from the correspondence fields at the mean lines samples. These tensors are then extrapolated to the whole brain surface using a log-Euclidean framework [Pennec 2006b, Arsigny 2006] without any guarantee that the obtained variability measures are compatible with an underlying deformation of the brain surface. This gives an aggregated measure at the population level that is not based on individual deformation mappings. The two approaches are based on radically different assumptions and the following variability maps show how these different models influence the results. Differences stem from the different way lines are matched, noise is removed and variability is extrapolated to the brain surface.

6.3.4.1 Regularity of the variability

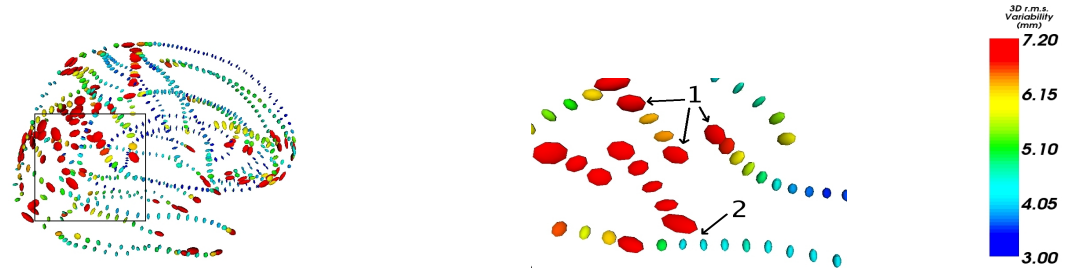
The figure (6.5) shows the covariance matrices built from the initial vector speed at the mean lines points (c_i) in our Current-based Diffeomorphic Registrations (CDR) (Fig. 6.5-a) and from the correspondence field in the PLC approach (Fig. 6.5-b). We notice that the

point matching method leads to irregular tensor fields at extremities of the lines and between lines, whereas the global regularity constraint of the diffeomorphism in CDR imposes the retrieved variability to be spatially smoother. CDR thus discounts any variability contained in the residual matching errors, which is considered here as noise. In PLC, the tensor field is denoised separately by removing "unreliable" large tensors at the end of lines before the extrapolation step.



Whole Brain Detail

a - Current-based Diffeomorphic Registrations (CDR)



Whole Brain Detail

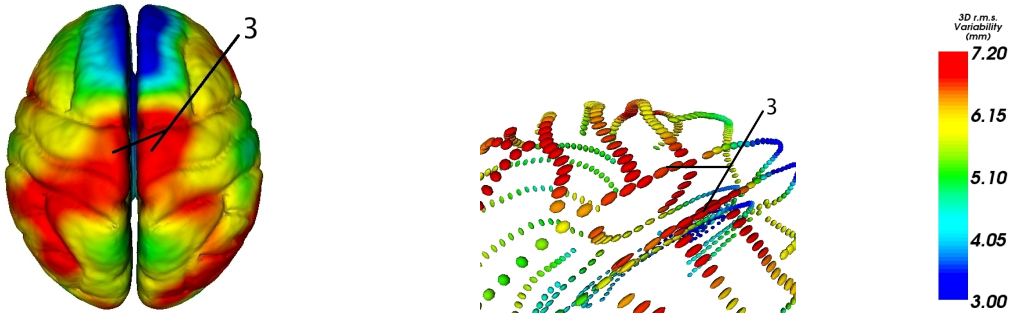
b - Pointwise Line Correspondences (PLC)

Figure 6.5: At each sampling point, ellipsoids represent the square root of the empirical covariance matrix of the initial speed vectors (left hand side) or displacement field (right hand side). With PLC method, extremal points are supposed to be matched: this induces a high variability at the extremities of lines (area 1, right). This is avoided by the current approach (area 1, left). With PLC, each line is registered individually: the variance can vary dramatically where lines cross (area 2, right). This situation can occur where a sulcus has a branch, in "Y"-shape configuration, and the junction may not be considered by the PLC approach. The global regularity constraint of CDR leads to smoother results (area 2, left).

6.3.4.2 Variability in the direction of the lines

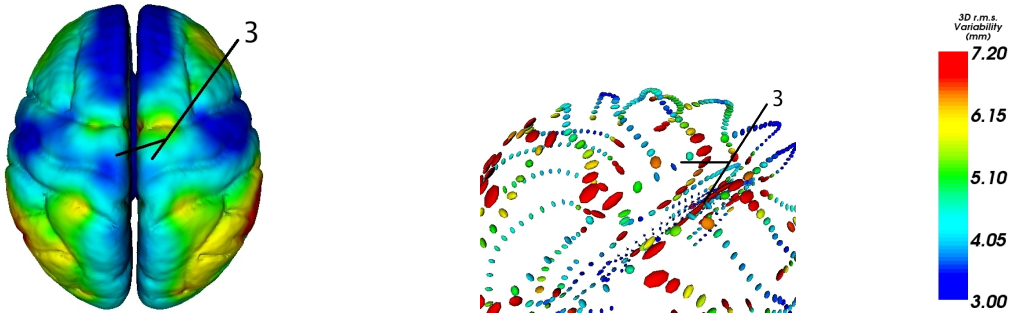
One drawback of PLC's method as underlined in [Fillard 2007c] is the fact that it systematically under-estimates the variability in the direction of the lines. This variability is indeed essentially captured at the extremities of the lines and minimized in between. As a pragmatic solution, in the PLC approach, the large extremal tensors were removed

before extrapolating the variability, and the final measures minimize the variability in the direction of the lines. This aperture problem is particularly visible on the top of the brain as shown in figure 6.6. By contrast, the models based on currents (CDR) manage to represent a larger part of this variability. This effect is of particular importance since this tangential variation is one of the major variation trends within the population as shown in Fig. 6.4. Anatomically, any lateral splaying of the central and pre-central sulci (at the top of the brain) is usually a sign atrophy, consistent with widening of the interhemispheric fissure. If this variation is discounted, for example by discarding tensors at the extreme points of sulci, any future registration approach that uses the tensor fields to model variation will underestimate the true anatomic variation in these areas. Otherwise, the variability which is orthogonal to the direction of the lines is in good agreement for the most part.



Variability map Covariance matrices

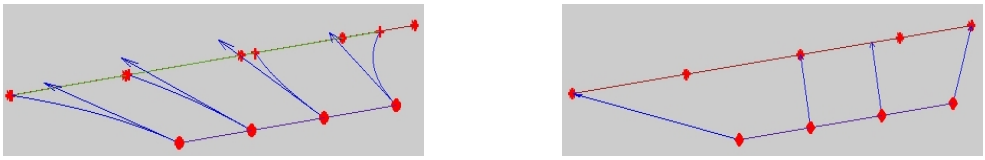
a - Current-based Diffeomorphic Registrations (CDR)



Variability map Covariance matrices

b - Pointwise Line Correspondences (PLC)

Figure 6.6: In the variability maps, a variability in the direction of the lines is retrieved in area 3 (extremities of central sulci) by CDR and not by PLC. The covariance matrices in this region show that the variability is mainly longitudinal. Since, in PLC method, large tensors at the endpoints of the sulcal lines are removed before the extrapolation, the variability in the direction of the lines is missed and the total variability is unreasonably small.



a - CDR b - PLC

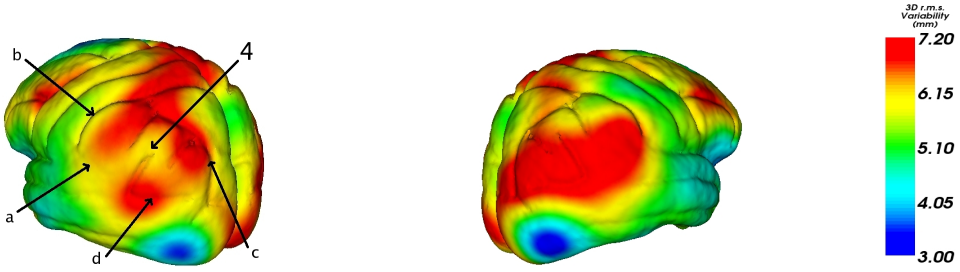
Figure 6.7: The registration method influences the way tangential variability is taken into account. With point correspondences the tangential variability is mainly captured at the endpoints of the lines and minimized in between. The approach based on current (CDR) retrieves a tangential component of variability all along the lines.

6.3.4.3 Distinction between correlated and anti-correlated motions

In our CDR framework the tensors at every points of the mean surface are computed from the initial speed vectors at these points that are interpolations of the initial speed vectors at the mean lines samples (see Eq. (6.2.6)): the interpolation is performed *before* computing the variability measures. By contrast, in the PLC method the covariance are computed on the mean samples and *then* extrapolated to the brain surface. As shown figure 6.9 this difference theoretically enables CDR to distinguish between areas where points are deviating from the mean anatomy in a correlated or anti-correlated manner. This is a possible explanation of the different variability maps retrieved in area 4 of figure 6.8.

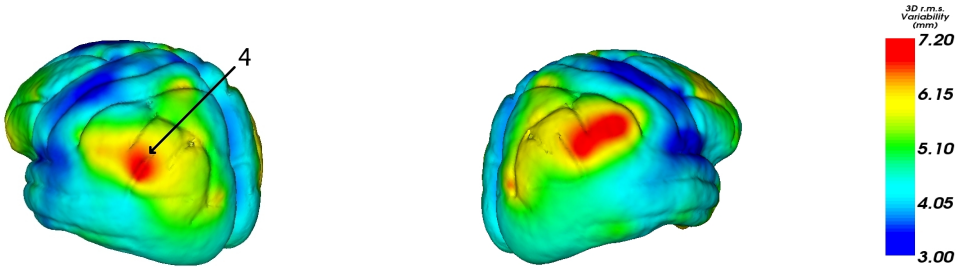
6.4 Discussion and Conclusion

In this paper we present a methodological framework to build global brain shape statistics by measuring and consistently integrating the variability of anatomical constraints such as sulcal lines. This framework is based on two methodological tools: lines are modeled as currents, and multiple object sets are matched by a single diffeomorphic deformation. By modeling lines as currents, we are able to measure geometrical dissimilarity between curves without assuming point correspondences between objects. Discrete and continuous lines are handled in the same setting, thus guaranteeing numerical stability and nice convergence properties. This distance is also robust to noise, preventing small perturbations of lines from hiding true underlying geometrical differences. On the other hand, the diffeomorphic framework consistently integrates the geometrical variations of a set of currents into a single deformation. We avoid modeling the variability of each objects independently. On the contrary, we try to explain the variability of the sulcal constraints by a global deformation of the underlying image domain. By inferring a Gaussian model on such deformations, we can define a generative statistical model that roots the variability measures into a rigorous model for individuals. Principal trends of variations within the database can be highlighted by looking at the deformation of a mean brain surface. Such statistical models offer an approach for classifying new observations according to their pathologies, gender,etc. The synthesis of the geometrical variations into principal modes of variations



Left Hemisphere Right Hemisphere

a - Current-based Diffeomorphic Registrations (CDR)



Left Hemisphere Right Hemisphere

b - Pointwise Line Correspondences (PLC)

Figure 6.8: Area 4 is surrounded by 4 major sulci: the Sylvian fissure (a), postcentral sulcus (b), intraparietal sulcus (c) and superior temporal sulcus (d). In the left hemisphere the first two vary, with respect to the sample mean, mostly in a decorrelated manner with respect to the last two sulci whereas their respective motions are much more correlated on the right hemisphere. The CDR approach tries to combine the motion of all lines and therefore leads to a small variability in area 4 (the perisylvian cortex) in the left hemisphere and to a large one in the right hemisphere. In these perisylvian areas, the variability is likely to differ by hemisphere as the right hemisphere perisylvian sulci are torqued forward and at a higher angle of elevation than their counterparts in the left hemisphere [Thompson 1998]. With PLC method this asymmetry in the magnitude of anatomical variability is not retrieved directly.

may also make it easier to identify spatially correlated anatomical patterns and may lead to new scientific findings.

This framework however raises several questions. Our statistical modeling focuses on the deformation term whereas there is obviously no ground truth regarding anatomical homology between brains. Even so, a diffeomorphism can capture a large part of the geometrical variability between shapes and sets of shapes. It is clear that some of the "true" underlying variability is not captured by the diffeomorphism and remains unmodeled in the residual matching errors. These residuals contain numerical and physical noise, possible non-diffeomorphic variations (changes in topology or folding patterns for instance) as well as variations that are not compatible with the variations of other objects in the surroundings. Our statistical model focuses here on the deformation term only and our results indicate that it can model a great part of the variability. However, a more complete statistical framework

a - CDR b - PLC

Figure 6.9: Extrapolation schemes in the simple case of anti-correlated vectors. Right: In PLC framework the tensors are computed at the sample points and then extrapolated in the middle point: the tensor in the middle is similar to the two others. Left: In CDR approach one first extrapolates the vector field and then computes at each point the covariance matrix. Since the vectors are anti-correlated, the field is close to zero at the center and the variability measured at this point is negligible.

would take into account the matching residuals as well. A given observation would be therefore decomposed into a deformation and residuals and the statistical model would say how probable such a decomposition might be. Building such a statistical framework is beyond the scope of the present paper, but it must be the topic of further investigation.

Some comment is also necessary regarding whether the norm on currents is appropriate for the data, as this model does not take explicitly curvature into account and every points on lines play the same role. When anatomical curves are matched using a smooth registration field, [Leow 2005] have previously explored the case where the curves are modeled as level sets of an implicit functions, and no explicit point correspondence is enforced, allowing the mapping to relax along sulcal lines. They investigated the matching of anatomic structures by directly constructing their implicit level set representations and the proposed matching cost functionals were shown to be closely related to the Hausdorff metric. With this type of mapping, brain structural variability was reduced by 10% in most regions and up to 40% in other regions; greatest reduction was observed in the temporal and frontal lobes, while a lesser reduction was observed in areas with greatest anatomic variability. Arguably, this results in mappings that have less distortion while still matching homologous gyral anatomy in detail from one subject to another. Contrary to some other norms that explicitly take into account surface curvature [Fischl 2004], or differential invariants within a curve (such as torsion) derived from the Frenet frames of the curves being matched [Guéziec 1994], our norm in this paper, and the one in the paper by [Leow 2005] do not consider that there are particular points of anatomical interest along the curves that can be identified as reliable landmarks. With some minor exceptions (such as the genu of the central sulcus [Vaillant 1999, Goualher 2000], curvature, at least at the finer scale of indentations within sulci, is not a reliable guide to functional or structural homology in the human cortex, and using points of maximum curvature to guide correspondence may be problematic. Our approach is somewhat more agnostic in regard to point matching. At least for the human cortex, details visible on MRI along the length of a sulcus would not be reliable features for anatomical matching, although this does not preclude their identification in future, e.g. using other modalities such as DTI.

The third issue concerns the choice of sulcal lines as constraints to retrieve the global

variability of the brain. The question of the anatomical significance of the sulcal lines is often raised in the literature [Toga 2007, Thompson 2003]. The sulci that we use here as topological constraints for cortical matching are essentially those which have been chosen to have functional significance, occur consistently in large numbers of normal individuals, and are not so variable in their incidence and relation to other sulci that it would preclude their reliable identification in large numbers of subjects. Moreover, the sulcal lines are labeled and supposed to be in a single part. However, the framework based on currents is not limited to such databases: lines could be discontinuous and consists of several parts, which may indeed be more accommodating of interrupted sulci, which are known to occur [Mangin 2004a, Duchesnay 2007]. Even if the lines are not labeled, a matching based on currents is still possible: we then consider all the lines as a single current. Preliminary experiments on the bottom lines of the sulci automatically extracted from a few subjects [Rivière 2002] raises however several problems. The geometry of the lines can be so complex that it is not possible to define a global orientation of the sulci. The variability is too high to retrieve sensible correspondences in the absence of any priors on sulcal labels. The quality of the lines extraction (and possibly the quality of the labelling itself) does not enable to find reasonable matchings between subjects. The geometry of the sulci depends actually on the process of extraction (whatever it is manual or automatic). In order to better constrain the registrations, one needs to account for more information than only the most probable lines. One would like for instance to use probability maps of the presence of the sulci in order to account for variability of the extraction itself. Sulcal ribbons could also contain more reliable geometrical information. A similar framework (but that does note require orientation of lines) has been used in [Auzias 2008] with the sulci of [Rivière 2002]. In the future, alternative cortical landmarks, including perhaps the endpoints of fiber paths inferred from DTI, may also supplant or partially replace the reliance on sulci as a guide to anatomical homology in the human cortex [Cathier 2006].

The comparison with a pointwise line correspondence (PLC) method is also difficult to interpret. The comparison remains here largely qualitative and at a methodological level. Even so, it is clear that each result is biased by the assumptions on which the variability measures are based. Each approach reveals new features from the database such as the principal modes of deformation, or unexpected patterns of anatomical correlation at distant points [Fillard 2007a]. A fair comparison between both methods would rely on objective statistical performance metrics, such as their respective predictive power for instance. In future, we will design studies that aim to predict extrinsic information about the subjects (e.g. sex, handedness, disease subtype, prognosis), from the information encoded in the cortical deformations. In a sense, the best model of anatomical variability is one that allows most reliable inferences and predictions regarding population. However, such a deep comparison is beyond the scope of this paper. Our purpose was here to present a general framework to compute statistics of brain shapes, to highlight its strength and limitations and finally to show its feasibility and relevance for addressing a range of statistical problems in the field of computational anatomy.

Chapter 7

Description of the anatomical variability of white matter fiber bundles

In this chapter, we focus on the white matter fiber bundles of the brain. In the framework of Chapter 5, we estimate a set of prototype fiber bundles along with their variability in the population. Statistics on both deformations and residual highlight interesting anatomical information.

This chapter is self-contained and can be read independently from the rest of this thesis. It has been published in the proceedings of IPMI [Durrleman 2009a].

Contents

Abstract

The purpose of this chapter is to measure the variability of a population of white matter fiber bundles without imposing unrealistic geometrical priors. In this respect, modeling fiber bundles as currents seems particularly relevant, as it gives a metric between bundles which relies neither on point nor on fiber correspondences and which is robust to fiber interruption or reconnection. First, this metric is included in a diffeomorphic registration scheme which consistently aligns sets of fiber bundles. The fiber bundles, which are extracted from the image of tensors, contain more anatomical information than the tensors themselves and thus better constrain the registration. In particular, the boundaries of the bundles, which are not visible in images, help to solve the aperture problem which appears when fiber

matchings are constrained by tensors only. Second, the measure of variability of a population of fiber bundles is based on a statistical model which considers every bundle as a random diffeomorphic deformation of a common template plus a random non-diffeomorphic perturbation. Thus, the variability is decomposed into a geometrical part and a "texture" part. Our results on real data show that both parts may contain interesting anatomical features.

7.1 Introduction

The primary goal of Computational Anatomy is to study the variability of anatomical structures in populations. This analysis can be used to classify populations (e.g. pathologic versus control), or to drive the segmentation of anatomical structures in new images. Variability measures usually rely on correspondences determined by registration. More generally, shape differences can be captured by the geometrical deformation of one structure onto another. These deformations are used to learn how a prototype structure (called also atlas or template) deform within a population. This requires to define a proper registration method and to infer a statistical model on the deformations between a template (to be estimated) and each subject in the population.

While such statistical analysis have already been proposed for sulcal lines [Durrleman 2008c, Fillard 2007c] (see Chapter 6), and subcortical structures [Vaillant 2004, Durrleman 2009c], much fewer tools are available for white matter fiber bundles obtained in diffusion MRI. These structures are of great importance as they may contribute to map brain connections between functional areas, or to understand effects of neurological pathologies (like Alzheimer's disease) onto the brain white matter. Most recent approaches are based on the nonlinear registration of fractional anisotropy (FA) maps [Goodlett 2008, Smith 2006] or tensor images [Zhang 2007, Yeo 2008b, Cao 2005, Yeo 2009b]. The deformation resulting from this image registration is then applied to fibers. In these methods, one may question if the fiber bundles are correctly aligned since the boundaries of the bundles are not visible in FA or tensor images. Fibers which belong to the same bundle connect specific brain regions together and therefore should be preserved during registration. Such information is not visible in the image of tensors which gives only a local information on the underlying biological tissue. By contrast, the extraction of fiber bundles, as a global feature, selects relevant anatomical information in those images. For this reason, we propose here to use directly fiber bundles as constraints to drive the registration.

Recent approaches measuring variability of fiber bundles rely on point or fiber correspondences between bundles [Corouge 2006, Ziyan 2007, Batchelor 2006]. However, tractography algorithms were never shown to produce stable and reproducible results. Thus, the comparison of bundles should not rely on individual fibers or points but rather on the global shape of the bundles. Furthermore, tractography might be valid only locally: the true neuronal pathway may correspond to the union of several pieces of fibers and one should not blindly consider sets of connected points produced by tractography as true fibers. The solution of considering a bundle as an unconnected cloud of points is not satisfactory either,

since it does not take into account the local orientation of the bundles encoded by the tangents of the fibers. Therefore, an ideal framework for fiber bundles should be robust to *curve connectivity and sampling*, and should take into account the *local orientation* of the bundle. Similarly, a distance between bundles (used as a dissimilarity measure during registration) should rely *neither on point nor on fiber correspondences*. In this paper, we propose to use the framework based on currents, that precisely models curves as a set of unconnected oriented points. This framework is robust to fiber interruption and provides a dissimilarity metric on curve sets that does not assume any kind of correspondences. Conversely, it is sensitive to the local fiber orientation and to the point density: a single fiber will be unlikely to influence registration, which makes currents naturally robust to outliers. Finally, currents are compatible with the diffeomorphic registration method of [Glaunès 2008], and therefore can be used for pairwise registration of fiber bundles.

Once a registration framework of fiber bundles modeled as currents is defined, it can be used to define a statistical model of variability. From the perspective of the deformable models, we consider the bundles of different subjects as random diffeomorphic deformations of an unknown template perturbed by non-diffeomorphic variations (called residues in the sequel). Following the lines of [Durrleman 2008a, Durrleman 2009c] (see Chapter 5), we jointly estimate this prototype bundle along with its deformations onto each subject's anatomy. In a second time, statistical analysis of bundles is achieved by a principal component analysis (PCA) of the diffeomorphic deformations and the residues to extract their principal modes of variations. The former accounts for the smooth variations of the template within the population: stretching, shrinking, dilation or torsion, while the later accounts for all variations that cannot be captured by regular diffeomorphisms, called *texture* in the sequel: fiber creation or topology changes. This model is not based on strong assumptions and can therefore retrieve a large range of geometrical variations.

The paper is organized as follows. Sec. 7.2 shows how currents are used to model fiber bundles and how they are interfaced with a diffeomorphic registration scheme. The statistical model is developed in Sec. 7.3. In Sec. 7.4, we evaluate the method on real data. We compare pairwise registrations of 5 fiber bundles with the alignment obtained from FA and tensors images. Finally, we build the atlas from 6 subjects and analyze the variability of the corticobulbar tract.

7.2 Fiber Bundles Registration based on Currents

7.2.1 Fiber Bundles Modeled as Currents

Currents are geometrical objects originally introduced in medical image analysis to model curves and surfaces [Vaillant 2005]. In this section, we recall the properties which are relevant for our topic and refer the reader to [Vaillant 2005, Durrleman 2008b, Durrleman 2009c] and to Chapter 1 for more details.

In the framework of currents, a set of fibers is characterized by the way it integrates vector fields. Given ω a square integrable $3D$ vector field, a bundle B made of several fibers

F_i integrates ω thanks to:

$$B(\omega) = \sum_{F_i \in B} \int_{F_i} \omega(x)^t \tau_i(x) dx, \tag{7.2.1}$$

where $\tau_i(x)$ is the oriented tangent vector of the fiber F_i at point x. A fiber bundle may be seen as a set of wires sending information in one direction at a constant rate. Eq. (7.2.1) computes the total rate of information that goes through the orthogonal sections (i.e. equipotential surfaces) of ω. To characterize a fiber bundle, we measure how this quantity varies while the equipotentials of ω varies. For this purpose, we define the test space W, in which ω varies, as the set of the convolutions between any square integrable vector fields and a smoothing kernel. This excludes from W the vector fields with too high spatial frequencies. Formally, W is the reproducing kernel Hilbert space (r.k.h.s.) whose kernel K^W is Gaussian: $K^W(x,y) = \exp(-\|x-y\|^2/\lambda_W^2)\mathrm{I}_3$ for any points (x,y)[1]. The standard deviation, λ_W, is the typical scale at which the vector fields ω varies spatially. In this setting, any set of smooth curves is a continuous linear mapping from W to $\mathbb{R}$. The space of currents W^* is the space containing such objects.

W^* is a vector space. The addition of two pieces of curves is simply the union of them. In Eq. (7.2.1), each fiber F_i or each piece of these fibers can be seen as a current individually: the union of them (their addition) is still a current. For instance, Eq. (7.2.1) would not change if each F_i were split into a collection of small segments. *It does not depend on the connectivity of the fibers within the bundle.*

Any current in W^* may be decomposed into an *infinite* sum of delta Dirac currents, which play the role of basis vectors. A Dirac current δ_x^τ is defined by: $\delta_x^\tau(\omega) = \tau^t \omega(x)$. It models an oriented point and encodes the direction τ of the fiber bundle at point x. Each segment of the polygonal lines returned by tractography is approximated by a Dirac current δ_x^τ where x is the center of the segment and τ its direction. This approximation converges in the space of currents as the sampling of the curves becomes finer. In this sense, the modeling is *weakly sensitive to the sampling* of the fibers. As a consequence, a bundle B is approximated by a *finite* sum over all segments within the bundle: $\sum_i \delta_{x_i}^{\tau_i}$.

In addition, the space of currents W^* is provided with a norm and an inner product, which define a distance between two bundles B and B' as: $\|B - B'\|_{W^*} = \sup_{\|\omega\|_W \leq 1} |B(\omega) - B'(\omega)|$. Following our analogy, this measures the rate of information along the wires of B and the wires of B' that goes through the orthogonal sections of the same ω. We look for the regular ω ($\|\omega\|_W \leq 1$) which makes this difference the largest possible, i.e. that captures the more differences between the two structures. This geometric distance compares bundles globally, *without assuming any kind of fiber or point correspondences between them.* The smaller the standard deviation λ_W, the smaller the scale at which ω varies, the finer the geometrical details captured by this distance.

This distance has a closed form. On the Dirac currents, the inner product is given by $\left\langle \delta_x^\alpha, \delta_y^\beta \right\rangle_{W^*} = \alpha^t K^W(x,y)\beta$. By linearity, the inner product between two bundles $B =$

[1] K^W is the Green's function of L^tL for some differential operator L. The inner product in W is defined then by $\langle \omega, \omega' \rangle_W = \langle L(\omega), L(\omega) \rangle_{L^2}$. See [Saitoh 1988] for more details.

$\sum_{i=1}^{n} \delta_{c_i}^{\tau_i}$ and $B' = \sum_{j=1}^{m} \delta_{c'_j}^{\tau'_j}$ is given by:

$$\langle B, B' \rangle_{W^*} = \sum_{i=1}^{n} \sum_{j=1}^{m} \tau_i^t K^W(c_i, c'_j) \tau'_j \tag{7.2.2}$$

This inner product (and hence the distance $\|B - B'\|_{W^*}$) does not require any condition on the curves sampling (n may not equal m, for instance). It compares all pairs of tangents (τ_i, τ'_j) weighted by a function of their distance $\|c_i - c'_j\|$.

Since the space of currents is a vector space provided with an inner product, we can directly compute the mean and the covariance matrix of a population of fiber bundles. However, this statistical analysis would not be relevant with unregistered fiber bundles. This will be used, instead, to perform PCA on the residuals that remain after registration.

7.2.2 Spatially Consistent Registration of Fiber Bundles

Our goal is to align two sets of fiber bundles segmented in images of two different subjects. The algorithm introduced in [Vaillant 2005] finds precisely a *consistent deformation of the underlying* $3D$ *space* that best matches two sets of labeled currents. The deformations are chosen as $3D$ diffeomorphisms (smooth deformations with smooth inverses), solution at time $t = 1$ of the flow equation: $\frac{\partial \phi_t(x)}{\partial t} = v_t(\phi_t(x))$, with initial condition $\phi_0 = \mathrm{Id}$ (no deformation). The time-varying vector field $(v_t)_{t\in[0,1]}$ is the speed vector field of the deformation, which is supposed to belong to a r.k.h.s. V with Gaussian kernel K^V. The standard deviation of K^V, λ_V, determines the typical scale of the deformation: the greater, the smoother the deformation. The regularity of the final deformation ϕ_1^v is measured by integrating the norm of the speed vector field over time: $d_V^2(\mathrm{Id}, \phi_1^v) = \int_0^1 \|v_t\|_V^2 \, dt$. The registration consists therefore in minimizing:

$$J(v) = \sum_{i=1}^{N} \left\| \phi_{1*}^v B_i - B'_j \right\|_{W^*}^2 + \gamma d_V^2(\mathrm{Id}, \phi_1^v) \tag{7.2.3}$$

where γ is the usual trade-off between fidelity to data and regularity. $\phi_* B$ denotes the geometrical transportation of the fiber bundle B by the diffeomorphism ϕ: each point x moves to $\phi(x)$ and each tangent τ is transformed into $d_x\phi(\tau)$, where $d_x\phi$ is the Jacobian matrix of ϕ. This geometrical transportation is conveyed in the space of currents thanks to $(\phi_* B)(\omega) = B((d\phi)^t \omega \circ \phi)$, which results simply from a change of variable in Eq. (7.2.1). On Dirac currents, we have $\phi_* \delta_x^\tau = \delta_{\phi(x)}^{d_x\phi(\tau)}$.

It is proved that the speed vector field minimizing Eq. (7.2.3) *is parametrized* by momenta $\alpha_k(t)$ at the points $x_k(t)$ of the moving bundle B: $v_t(x) = \sum_k K^V(x, x_k(t))\alpha_k(t)$ [Vaillant 2005, Miller 2002]. Once time is discretized, Eq. (7.2.3) is therefore minimized via a gradient descent on the parameters: $(\alpha_k(t_p))$. Moreover, the resulting diffeomorphisms are geodesic. Thanks to Euler-Lagrange equations [Miller 2002], they are entirely determined by their initial momenta $\alpha_k(0)$: *the tangent-space representation* of the diffeomorphism. The metric on this tangent space is given by $\|\alpha(0)\|^2 = \|v_0\|_V^2 = \alpha(0)^t k^V \alpha(0)$ where k^V is the matrix $(K^V(x_i, x_j))_{i,j}$. From now on, we denote ϕ^α the diffeomorphism ϕ_1^v with initial momenta α.

Applying this registration framework directly to fiber bundles, which may have up to 10^5 segments, raises computational issues. The computation of the data fidelity term in Eq. (7.2.3) ($\|\phi_{1*}^v B - B'\|_{W^*}$) requires to compare every segments of B with every segments of B', as shown in Eq. (7.2.2). Hopefully, this complexity is reduced thanks to the approximation scheme of [Durrleman 2008b] (sparse representation of Chapter 3).

7.3 A Statistical Model of Fiber Bundles

In this section, we show how to use the modeling based on currents and the previous registration tool to define a statistical model on fiber bundles. Following [Allassonnière 2007, Durrleman 2009c] (see Chapter 5), we model our observations as deformations of an unknown prototype bundle (also called template or atlas) perturbed by non-diffeomorphic variations (the residues). Formally, we consider the bundles $(B_i)_{i=1\dots N}$ (the same bundle for N different subjects) as instances of the following process:

$$B_i = \phi_{i*}\overline{B} + \varepsilon_i \tag{7.3.1}$$

where the bundles B_i are seen as currents, ϕ_i are diffeomorphisms that deform the unknown template $\overline{B}$ supposed to be a current as well. ε_i are the residual perturbations which account for everything that cannot be captured by a regular deformation. The ε_i's are supposed to be i.i.d. zero-mean Gaussian random variables in the space of currents. To infer a random model on the deformations ϕ_i, we use their tangent-space representation: an instance ϕ^α is simulated by shooting geodesically an instance of the momenta α (vector of finite dimension).

We estimate the template $\overline{B}$, the deformations ϕ_i and the residues ε_i with a Maximum A Posteriori approach with an approximation, as in [Durrleman 2008a, Durrleman 2009c] (see Chapter 5). As a result, we minimize:

$$\min_{\overline{B},\alpha_i} \left\{ \sum_{i=1}^{N} \left\| {\phi^{\alpha_i}}_*\overline{B} - B_i \right\|_{W^*}^2 + \gamma d_V^2(\mathrm{Id}, \phi^{\alpha_i}) \right\} \tag{7.3.2}$$

We start by setting $\phi_i = \mathrm{Id}$ (i.e. $\alpha_i = 0$, no deformation) and $\overline{B} = \sum_{i=1}^N B_i/N$, the empirical mean in the space of currents. Then, we minimize the functional by considering that $\overline{B}$ and the α_i's are fixed alternatively. The first step consists in registering $\overline{B}$ to each B_i, leading to initial momenta (α_i). The second step consists in updating $\overline{B}$ by minimizing $J(\overline{B}) = \sum_{i=1}^N \left\| {\phi^{\alpha_i}}_*\overline{B} - B_i \right\|_{W^*}^2$. This last minimization benefits from the approximation scheme of [Durrleman 2008b, Durrleman 2009c] (see Chapter 3).

Eventually, the algorithm returns an unbiased template $\overline{B} = \sum_{k=1}^{n_B} \delta_{x_k}^{\tau_k}$ and the deformations ϕ^{α_i} of $\overline{B}$ to each B_i. We perform a PCA on the momenta (α_i), and another PCA on the residual perturbations $\varepsilon_i = {\phi^{\alpha_i}}_*\overline{B} - B_i$. We shoot geodesically in the direction (resp. in the opposite direction) of the first mode of momenta, m_α (resp. $-m_\alpha$), to give the first mode of deformation at $+\sigma$ (resp. $-\sigma$): $\phi^{\pm m_\alpha}$. The PCA on residues is performed in the space of currents. This leads to the mean $\bar{\varepsilon} = \sum_i \varepsilon_i/N$ and the first mode at $\pm\sigma$: $m_\varepsilon = \bar{\varepsilon} \pm \sum_i E_i(\varepsilon_i - \bar{\varepsilon})$, where E is the first eigenvector of the covariance matrix $(\langle \varepsilon_i - \bar{\varepsilon}, \varepsilon_j - \bar{\varepsilon} \rangle_{W^*})_{i,j}$. As linear combinations of the input currents, the mean and

the first mode can be approximated using the scheme of [Durrleman 2008b] (i.e. sparse representation of Chapter 3) for a better visualization.

This joint statistical modeling accounts for both diffeomorphic and non-diffeomorphic variability. It is not biased by arbitrary point or fiber correspondences between different subjects. It does not impose strong prior on the nature of the variability. For instance, it does not assume that fibers of a bundle come from a mean line whose samples have been randomly moved, as in [Corouge 2006]. The major prior of our model consists in where to put the separation between the geometrical part (captured by the diffeomorphisms) and the texture part (contained in the residues). This separation is determined by the regularity parameters: λ_V, λ_W and the trade-off γ. In this paper, we set these parameters manually, whereas they could be set automatically along the lines of [Allassonnière 2007] for instance.

7.4 Experiments

Six brain DTI data sets acquired on a 1.5T GE scanner on healthy volunteers were used in this study. Image dimensions are $128 \times 128 \times 30$, and resolution is $1.8 \times 1.8 \times 4$mm. 25 non-collinear diffusion gradients and a b-value of $1000 s.mm^{-2}$ were used. Fiber tractography was performed using MedINRIA [2], which includes a robust tensor estimation and a streamline tractography algorithm using log-Euclidean tensor interpolation [Fillard 2007b]. Manual segmentation of five fiber bundles was done: the entire corpus callosum, the corticospinal and the corticobulbar tracts, and the left and right arcuate fasciculi (Fig. 7.6-a).

First, we evaluate the methodology developed in Sec. 7.2 by registering the bundles of two subjects and comparing the result with FA and tensor registration (Sec. 7.4.1). Second, our framework for atlas construction is evaluated with the construction of a diffeomorphic atlas of the five bundles of our data set (Sec. 7.4.3) and the statistical analysis of the corticobulbar tract (Sec. 7.4.4).

7.4.1 Fiber Bundle Registration

Diffeomorphic registration of fiber bundles using currents is compared to non-linear registration of FA [Vercauteren 2007] and tensor [Yeo 2008b] images. Pairwise registrations of 5 fiber bundles segmented in 2 subjects are conducted. For FA and tensor-based registration, deformation fields were computed between images and applied to bundles afterwards: bundles were not tracked again after registration. Note that the three methods produce diffeomorphic transformations and can be compared. The parameters were adjusted to produce deformations of about the same smoothness. Concerning our registration scheme, we set the regularity of the deformation $\lambda_V = 20$mm, the spatial scale of the currents $\lambda_W = 5$mm and the trade-off between regularity and fidelity-to-data: $\gamma = 10^{-4}$. For clarity purpose, we present registration results of two bundles only: the corpus callosum (CC) (Fig. 7.2) and the corticospinal tract (CST) (Fig. 7.3), since they highlight the most striking differences between methods.

[2] http://www-sop.inria.fr/asclepios/software/MedINRIA/

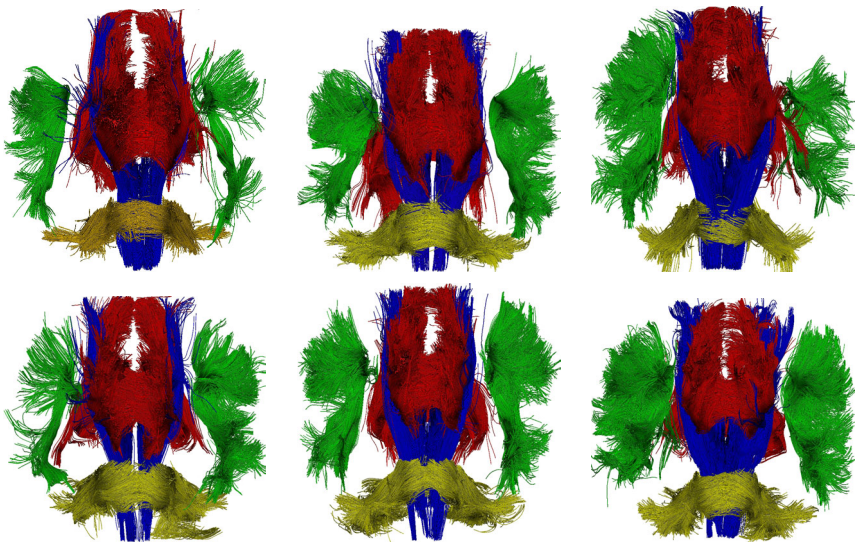


Figure 7.1: Five fiber bundles extracted in six subjects using MedINRIA. **Blue:** the corticospinal tract. **Yellow:** the corticobulbar tract. **Red:** the callosal fibers. **Green:** the left and right arcuates.

Fig. 7.2 a) shows two misaligned corpus callosum. Fig. 7.2b,c presents the registration of those bundles computed using respectively FA and tensor images. The registration of the fiber bundles with our method (Fig. 7.2d shows a greater overlap, synonym of a better alignment. Local improvements are noticeable in the left and right parts of the genu. This result shows that *the bundle information acts as a stronger prior* to align fiber tracts than the tensor image. Moreover, one can still notice few red fibers not aligned with the blue bundle in the exterior of the tract, which illustrates *the robustness of our methodology to outliers.*

Registration of two cortico-spinal tracts shows similar effects, especially in the anterior part expanded in a green square in Fig. 7.3. In those regions, multiple bundles may coexist whereas FA and tensor images are uniform, as shown in Fig. 7.4. Therefore, image-based registration is unable to correctly align the bundles, since bundle boundaries are not visible in images. This is an *aperture problem* inherent to FA and tensor-based methods. The selection of the fibers of each bundle is an extra information brought either by experts or by automatic bundling methods with anatomical priors. This information, which is not present in FA or tensor images, better constraint the registrations.

Registering labelled images of tensors would be a way to take such information into account in image registration. The global image would be divided into clusters which gather the tensors that correspond to the same bundle. However, such an approach would not be possible without defining a proper metric between labelled image of tensors. In particular, an extrapolation scheme would be required for the comparison of two clusters which do not overlap. Nevertheless, in our approach, we based our comparison on some

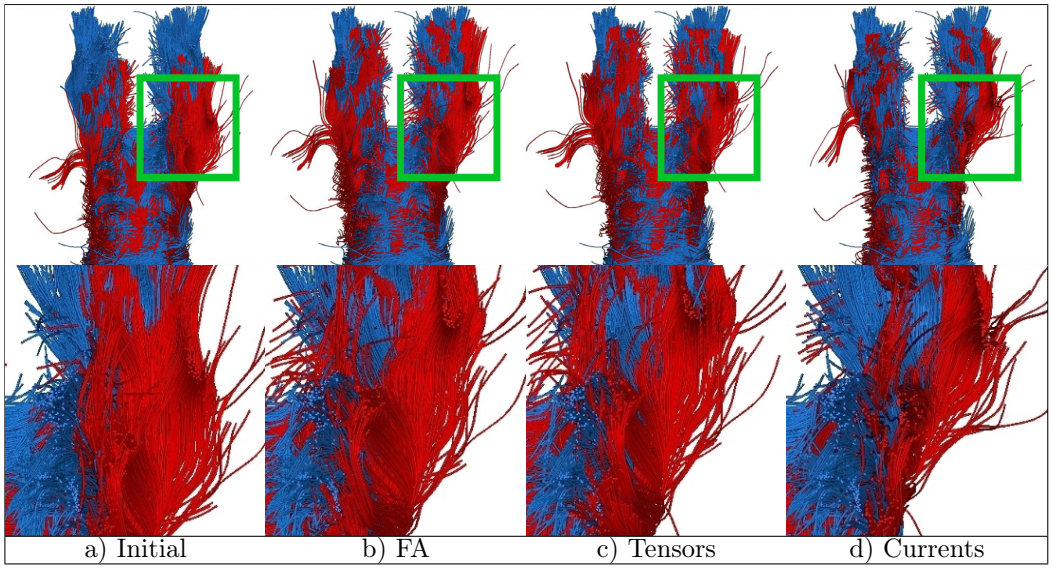


Figure 7.2: Diffeomorphic registration of two corpus callosum fibers. Bottom images are a close-up on the green squared region. Initial tracts (a) and registered tracts using FA (b), tensors (c) or bundles (d) as constraints. Overlap of blue and red fibers is greater using currents, especially in the left and right parts of the genu of the corpus callosum.

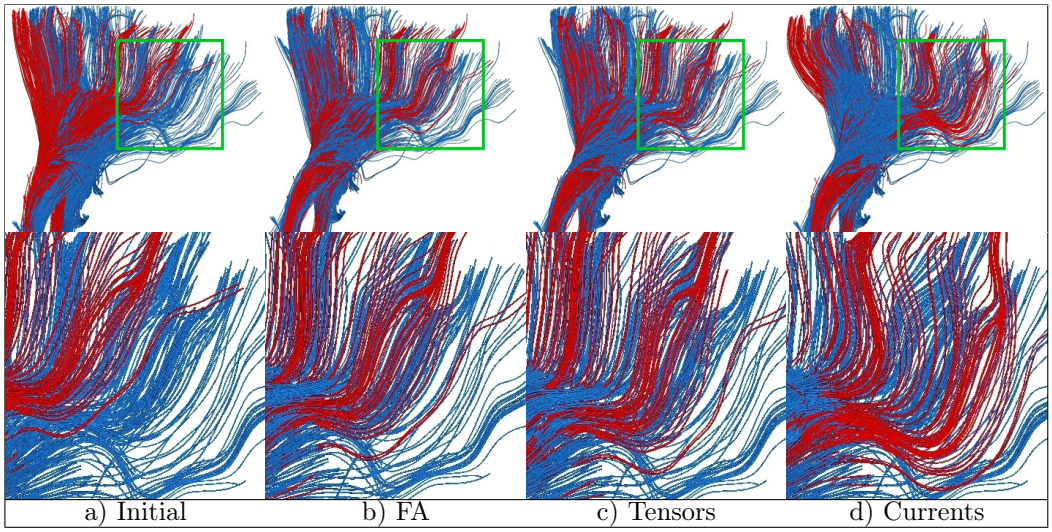


Figure 7.3: Registration of two corticospinal tracts. Bottom images are a close-up on the green squared region. Initial tracts (a) and registered tract using FA (b), tensors (c) or bundles (d) as constraints. Currents better warp the red fibers in the anterior part of the tract, highlighting the aperture problem inherent to FA and tensor-based approaches. Registration in the posterior part is mainly constrained by the corpus callosum which strongly pushes the fibers toward the posterior part of the brain.

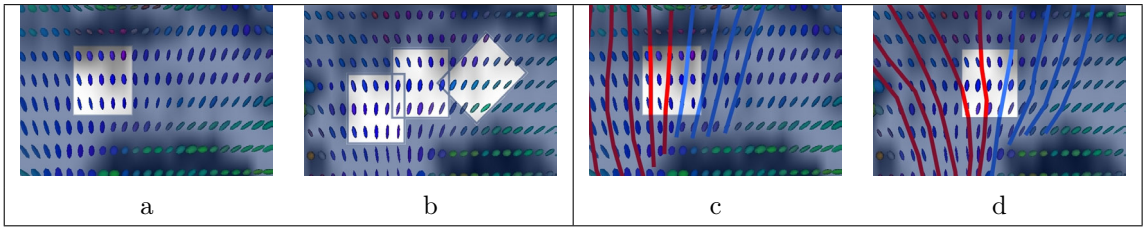


Figure 7.4: Illustration of the aperture problem in FA and tensor registration. **a and b**: Tensor fields of two subjects overlapped with FA images are shown (sagittal slice, inside the corona radiata). Without any prior, it is impossible to determine whether the rectangle in image **a** matches with any in image **b**: this is the aperture problem. **c and d**: two schematic fiber bundles in red and blue were added. It becomes clearer that the rectangle of image **c** has a unique correspondence in image **d**. The aperture problem is partly solved using the bundles as priors.

particular information extracted from the tensors: the estimated direction of the bundle and not the tensor itself. We believe that this extracted information is of much more relevant from an anatomical point of view than the tensor itself. The fiber bundles reflect a *global* feature of the anatomy whereas the tensors measure only the *local* diffusion property of the underlying biological tissue. For instance, the direction of the fibers are not necessarily aligned with the largest eigenvector of the tensor since some global regularity constraints are introduced during tractography. Tensor-based registrations may be corrupted by some non-reliable information (from an anatomical point of view), like the eigenvectors of the tensors at fiber crossing.

Of course, this framework depends on the assumptions on which the tractography is based and the selection of the fibers made by experts or some automatic methods. However, better understanding how the bundles vary within a normal population should help, in turn, to better constraint the tractography and selection methods. The subsequent statistical analysis could precisely be used first to better highlight the artifacts introduced by these extraction methods, then to remove (or at least better measure) these artifacts and eventually to incorporate more reliable anatomical priors during tractography, fiber selection and registration.

We remark also in Fig. 7.3, that the posterior parts of the cortico-spinal tracts seems not to be properly aligned. This is due to the presence of the corpus callosum in this region which strongly pushes the fibers toward the posterior part of the brain. Since the corpus callosum has much more fibers than the cortico-spinal tract, it acts as a much stronger constraint during the registration (which estimates a single deformation which aligns all bundles together). We can overcome this issue by associating a weight to every bundle during registration. This would weight the relative importance of each bundle (or possibly sub-parts of the bundles) in the cost function. However, how to determine such weights on an anatomical basis still remains an open question.

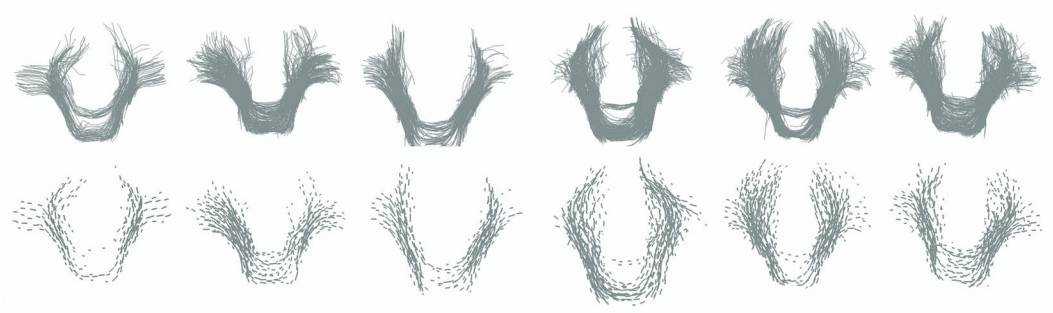


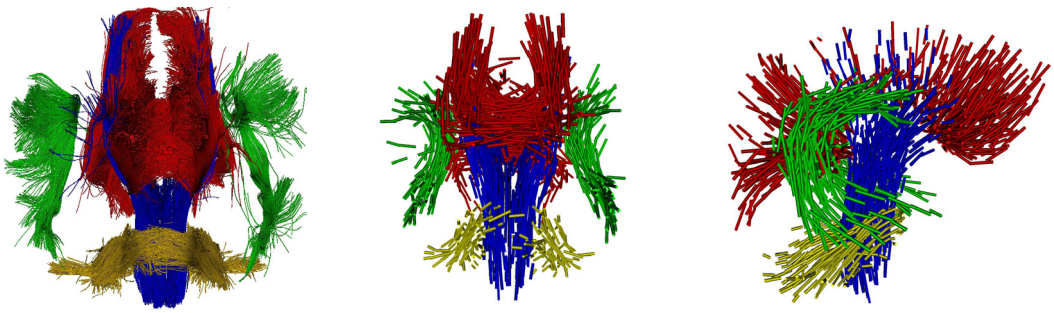
Figure 7.5: Original corticobulbar tract (top row) and their sparse representation for the scale $\lambda_W = 3$mm (the diameter of a tract is typically of 100mm). The sparse representation integrates the local redundancy of the tract at the scale λ_W and encodes the local density of fibers in the length of the segment. Here, for visualization purpose, the segments have been scaled by 1/10. The compression ratio is greater that 85% while the approximation error remains below 5% of the variance.

7.4.2 Sparse representation of the data

To efficiently run the pairwise registrations of fiber tracts and the construction of the atlas, we used the Matching Pursuit algorithm introduced in [Durrleman 2008b, Durrleman 2009c] (see Chapter 3). This algorithm provides a sparse representation of the current by integrating the local redundancy of the initial decomposition. The spatial scale of currents λ_W determines the scale under which the geometrical variations are considered as noise and therefore averaged. If many fibers go through a local section of size λ_W, then this set of fibers may be efficiently replaced by a single fiber which size is scaled to account for the local redundancy. In other words, a set of wires which goes through this section is replaced by a single wire which is scaled so that the total rate of information that goes through the section remains unchanged. As shown in Fig. 7.5 for the corticobulbar tract, this approximation leads to a much sparser representation of the current. Compression ratio is greater than 85% while the approximation error is below 5% of the variance of the population of corticobulbar bundles. This figure shows also that the fiber bundles modeled as currents take into account the geometrical information only up to the scale λ_W. As a consequence, the template estimated from these currents cannot contained geometrical information at a scale smaller than λ_W, as we shall see in the next section.

7.4.3 Fiber Atlas Construction

As explained in Sec. 7.3, we estimate a template such that the input data result from random deformations of this template added with random perturbations in the space of currents. As emphasized in Section 7.2.2, there is only one global deformation acting on all 5 bundles together, and 5 independent perturbations for each bundle. The template consists of five prototype bundles shown in Fig. 7.6. It has been computed by fixing the parameters of currents $\lambda_W = 5$mm, of deformations $\lambda_V = 20$mm and the trade-off $\gamma = 10^{-4}$.



(a) One subject (b) template (occipital view) (c) template (lateral view)

Figure 7.6: **Template of five bundles**: the corticospinal tract (blue), the corticobulbar tract (yellow), the callosal fibers (red), the left and right arcuates (green). (a): one subject among the six of the data set. (b,c) the estimated unbiased atlas. Original data result from a random deformation of the template, plus a random perturbation.

7.4.4 Variability Analysis of the Corticobulbar Tract

We show here how the first mode of deformations and the first mode of residues describe the variability of the corticobulbar tract within the studied population. The "geometrical" variability is captured by the deformations. As a result of the MAP estimation, the deformations appear to be centered: the norm of the mean parameters is 0.42 times the standard deviation, not significantly different from 0 (For the 6 subjects, the t-statistics is equal to $\sqrt{6} * 0.42 = 1.03$ with a p-value greater than 30%). The first mode of the deformations at $\pm\sigma$ is shown in Fig. 7.7. It shows the variability of the template which was captured by the regular diffeomorphisms. The main variations are a torque of the frontal part of the bundle, as well as a stretching/shrinking of its lateral parts. Further investigation should determine whether this torque is related to the well-known brain torque.

The variability in terms of "texture" is captured by the residual perturbations. The residues are centered: the mean current is 0.36 times its standard deviation. The first residual mode m_ε is shown in Fig. 7.8. It shows an asymmetry in the number of fibers in each lateral part of the bundle. This result shows, undoubtedly, that the variability left aside from the diffeomorphisms is not pure noise, *but still contains some interesting anatomical features.* In our case, further investigation is needed to determine whether this fiber creation/deletion effect is due to a true anatomical variability or to an artifact of the tracking algorithm. In any case, this shows how our modeling analyzes all the geometrical information without imposing strong priors on the kind of the variations we are looking for.

Our statistical model is generative: we can combine the deformation and the texture variability to simulate new data. This procedure is shown in Fig. 7.9: the template is deformed according to the first mode of deformation, and the first texture mode is added to the result. This new synthetic bundle can be compared to the input data in order to get an insight into the common features which has been detected in the population. This generative aspect of the model allows us to compare the data of a new subject with the estimated normal variability in the population and possibly detect abnormalities via large

deviations from this normal variability.

def. mode at $-\sigma$ $\phi^{-m_v}{}_* \overline{B}$ | template $\overline{B}$ | def. mode at $+\sigma$ $\phi^{+m_v} * \overline{B}$

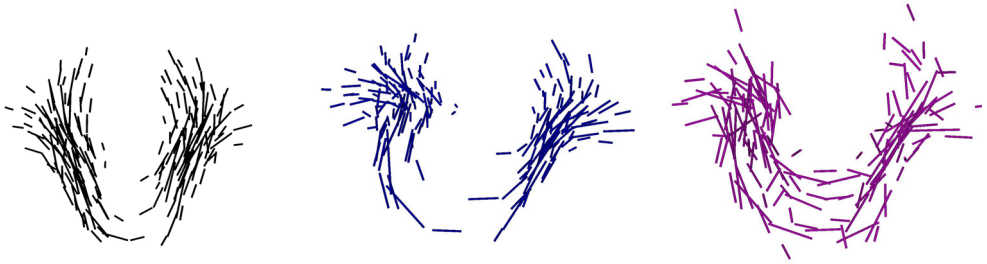
Figure 7.7: **First mode of deformation** of the corticobulbar tract at $\pm\sigma$. The diffeomorphic variability of the population around the prototype bundle (middle) is mainly a torque at the basis of the bundle and a stretching/shrinking effect of the left and right parts of the bundle.

texture mode at $-\sigma$ $\overline{B} - m_\varepsilon$ | template $\overline{B}$ | texture mode at $+\sigma$ $\overline{B} + m_\varepsilon$

Figure 7.8: **First "texture" mode** at $\pm\sigma$. This mode captures the residual variability of the population, once the diffeomorphic variability has been discounted. This mode is added to (c) or removed from (a) the prototype bundle (b). The mode at $+\sigma$ (resp. $-\sigma$) shows that the left (resp. right) part of the bundle becomes thicker, while its right (resp. left) part becomes thiner. Further investigation should determine whether this asymmetry is a true anatomical feature or an artifact of the tracking method.

7.5 Discussion and Conclusion

In this paper, we proposed a novel framework for the statistical analysis of fiber bundles using currents. Our methodology does not impose point-to-point or fiber-to-fiber correspondences, a crucial feature in regards to the variability of tractography algorithms outputs. It is also robust to outliers and weakly dependent of the fiber sampling. Diffeomorphic reg-



template $\overline{B}$ | def. mode at $+\sigma$ $\phi^{+m_v}{}_*\overline{B}$ | def. mode + texture mode $\phi^{+m_v}{}_*\overline{B} + m_\varepsilon$

Figure 7.9: **Simulation of a new corticobulbar tract** according to the variability inferred from the real input data. The template is deformed according to the first mode of deformation and then the first mode of the residues is applied to the result. This synthetic bundle compares with the original data. Such simulations show the common features that the model captures in the population.

istrations produce smooth and invertible deformation fields which match consistently a set of fiber bundles of one subject onto another. This registration scheme is further extended by a statistical model of fiber bundles, which estimates a template and its variability in a population. The variability is, in turn, decomposed into a geometrical part accounting for smooth deformations and a texture part which accounts for non-diffeomorphic changes in a population.

Pairwise registration shows that FA and tensor-based registration are less adapted than currents for fiber bundles registration, as bundle boundaries are not visible in those images (aperture problem). A misalignment of bundles may result in a loss of statistical power in group comparisons since contributions of several unrelated fiber bundles may be confounded. The method presented here optimally uses the strong prior that fibers belong to consistent bundles and ensure a proper alignment of those for statistical comparisons. This statistical analysis was evaluated on five fiber bundles extracted in six subjects. It shows consistent results with known anatomical variability (brain torque), which is put in evidence for the first time on fibers. Even with such a small dataset, our analysis managed to decompose the variability into two parts that are likely to describe relevant anatomical features.

The method, however, raises several questions. First, it is sensitive to the total number of fibers of a bundle, or fiber density, that may vary between two subjects. These variations may be caused by the tractography algorithm itself, as fiber density is generally an arbitrary parameter set by the method (some methods initiate fibers at the center of every voxels, some others over-sample voxels to produce more fibers). One solution would consist in normalizing this density by relying the fiber density to physical properties of the neural fibers, like the neural flux transported by the bundle. More generally, the framework based on currents enables to associate an attribute or a weight to each oriented point. Such a weight can be used to give more or less importance to some bundles (or some part of the bundles) during registration. For instance, using the FA as attributes would enable to relax

the geometrical constraint in regions where the FA is low (i.e. where the tracking returns less reliable fibers).

Our method requires all fibers in a bundle, and by extension all bundles of different subjects, to have a consistent orientation: all fibers should start at the same cortical region and end in the same region. We cannot reasonably assume that tractography algorithms produce consistently oriented fibers. In this work, reorientation was performed with an empirical procedure. For larger datasets, an automatic reorientation procedure has to be included, or the modeling based on currents has to be adapted to account for non-oriented curves.

Future work will focus on evaluating the method on much larger dataset to strengthen the interpretation of our results. Automatic bundling can be used to produce a complete set of anatomically relevant fiber bundles, as in [Maddah 2007, El Kouby 2005]. Our statistical analysis could be used then to detect and measure the artifacts or bias introduced by such automatic procedure. Once this bias has been corrected, these measures of variability could be used to constrain template-to-subject registrations by favoring deformation that are likely to occur in a normal population. Such registration can be used then to automatically select bundles in new subjects.

Our statistical measures both on the deformations and the residues could be useful to characterize pathologies which are known to affect fibers bundles like autism, to detect such pathologies in new patients, to find consistent sub-groups within populations, to detect abnormalities via deviations from the estimated variability. One could also combined this anatomical analysis with functional data. In particular, one could investigate possible correlations between the anatomical modes of variations and the activations of some functional areas.

Acknowlegments We would like to thank Denis Ducreux, MD, Bicêtre Hospital, Paris, for providing the brain datasets. This work was partially supported by European IP project Health-e-child (IST-2004-027749) and Microsoft Research.

Chapter 8

Prediction of remodeling of the right ventricle in Tetralogy of Fallot

In this chapter, we show how our statistical analysis can be used in clinical applications for the prediction of disease evolution and therapy planning. We use the atlas construction method of Chapter 5 on the surface of the right ventricle of hearts suffering from repaired Tetralogy of Fallot. The resulting modes of deformations were found clinically relevant by experts, as they exhibit known effects of the disease on the shape of the ventricle. We analyze the correlations between the modes of deformations and some clinical parameters. This allows us to propose a prediction of the remodeling of the shape of the right ventricle, which was successfully applied to two new patients.

This work results from a close collaboration with Tommaso Mansi. This chapter has been written by T. Mansi and published in the proceedings of MICCAI [Mansi 2009].

Contents

Abstract

Patients with repaired Tetralogy of Fallot commonly suffer from chronic pulmonary valve regurgitations and extremely dilated right ventricle (RV). To reduce risk factors, new pulmonary valves must be re-implanted. However, establishing the best timing for re-intervention is a clinical challenge because of the large variability in RV shape and in

pathology evolution. This study aims at quantifying the regional impacts of growth and regurgitations upon the end-diastolic RV anatomy. The ultimate goal is to determine, among clinical variables, predictors for the shape in order to build a statistical model that predicts RV remodeling. The proposed approach relies on a *forward* model based on currents and LDDMM algorithm to estimate an unbiased template of 18 patients and the deformations towards each individual shape. Cross-sectional multivariate analyses are carried out to assess the effects of body surface area, tricuspid and transpulmonary valve regurgitations upon the RV shape. The statistically significant deformation modes were found clinically relevant. Canonical correlation analysis yielded a generative model that was successfully tested on two new patients.

8.1 Introduction

Tetralogy of Fallot (ToF) is a severe congenital heart defect that requires surgical repair early in infancy. Yet, pulmonary valves may be damaged by the surgery, causing chronic regurgitations. As a result, the right ventricle (RV) dilates extremely, its shape is altered and the cardiac function is impaired: new valves must be implanted in adulthood to reduce risk factors [Geva 2006]. Understanding and quantifying RV remodeling in repaired ToF patients is crucial for patient management and therapy planning. However, high variability in pathology course and in RV anatomy makes difficult the decision of optimal timing for re-intervention [Geva 2006].

Contrary to the left ventricle, whose shape and deformations under pathological conditions are well documented, RV anatomy is complex and can vary tremendously among ToF patients. Several studies investigated possible correlations between clinical parameters in ToF [Geva 2006]. However, few works have quantified the anatomical alterations of the RV and their evolution due to the disease [Sheehan 2008, Zhang 2006]. In [Sheehan 2008], the authors measure the most striking differences in RV shape with respect to normals, quantifying some features of the complex RV remodeling observed in ToF. However, only one-dimensional indices are considered despite the availability of 3D segmentations. In [Zhang 2006], the authors present a 4D Active Appearance Model of the beating heart to segment RV in MRI. New indices based on the shape modes are proposed to classify patients from normal. Yet, the authors do not correlate their model with clinical features of ToF.

The clinical challenges raised by ToF encourage applying image-based shape analysis techniques to model the RV anatomical alterations due to pathological factors. These techniques generate a representative template of a population of interest and assess how it deforms within this population [Guimond 2000, Joshi 2004, Allassonnière 2007, Durrleman 2008a]. Yet, correlating shape with clinical variables require a rigorous framework: Biases may appear if the template is not defined in a consistent way, which may yield drastic differences in the statistical conclusions. Two strategies are available to create the template. The *backward* approach consists in modeling the template as the average of the deformed observations plus some residuals [Guimond 2000, Joshi 2004]. Such a template can be computed efficiently but the model parameters, especially the residuals, are

more difficult to identify. The *forward* approach consists in modelling the observations as deformations of the template plus some residuals [Allassonnière 2007, Durrleman 2008a] (see Chapter 5). Computing the template is more complex but model parameters can be faithfully estimated from images and clinical data.

In view of assisting the cardiologists in establishing the best time for re-intervention, we aim at statistically predict the RV remodeling in ToF. As a first step, we propose in this work to quantify the regional impacts of growth and regurgitations upon the end-diastolic RV anatomy of a cohort of 18 young ToF patients. The main deformation modes are estimated using the *forward* approach and analyzed through cross-sectional multivariate methods. We then derive a generative model of RV remodeling and test it on two new patients.

8.2 Methods

The right ventricle (RV) of multiple patients is segmented from cine-MRI as described in Sec. 8.3.1. To analyze this population of shapes, an unbiased template is first built. This template serves as reference atlas to determine the deformations towards each individual shape. Then, Principal Component Analysis (PCA) is applied on these deformations to extract the main deformation modes. The importance of each mode is statistically assessed with respect to child growth and valvar regurgitation severity, yielding a generative model of RV remodeling. Finally, we investigate how this model can predict the evolution of shape with respect to body surface area.

8.2.1 Unbiased Template of the Right Ventricle in Tetralogy of Fallot

The RV template is created using the *forward* strategy proposed by [Durrleman 2008a] (see Chapter 5). This approach is particularly suited for our purposes as 1) it is non-parametric, shapes are represented by currents; 2) model parameters are well-defined and can be estimated from clinical data, thus enabling statistical analyses; 3) template and deformations are computed simultaneously and consistently and 4) new patients can be integrated in the study without re-estimating the template.

The RV surfaces $\mathcal{T}^i$, or shapes, are modeled as the sum of a diffeomorphic deformation ϕ^i of the template $\overline{\mathcal{T}}$ and a residual term ϵ^i standing for the shape features that cannot be represented by the template (topology changes, acquisition artifacts, etc.): $\mathcal{T}^i = \phi^i_* \overline{\mathcal{T}} + \epsilon^i$. Currents are used to represent the shapes, the residuals and the deformations in the same common framework. They enable the usual operations (mean, variance...) on shapes as they form a vector space. Intuitively, currents can be seen as the flux of any vector field $\omega \in W$ across the shapes. W is a vector space of infinite dimension generated by a Gaussian kernel $K_W(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x}-\mathbf{y}\|^2/\lambda_W^2)$ that defines an inner product in W (W is a Reproducible Kernel Hilbert Space, RKHS). More precisely, a triangle centered at $\mathbf{x}$ with normal α is represented by the Dirac delta current $\delta^\alpha_{\mathbf{x}}$. Therefore, a discrete mesh is encoded by the sum of the currents of its triangles $\mathcal{T}^i = \sum_k \delta^{\alpha^i_k}_{\mathbf{x}^i_k}$. The residuals ϵ^i are modeled as a

Gaussian distribution on the α_k^i. The deformation ϕ^i that registers the template $\overline{\mathcal{T}}$ to the current $\mathcal{T}^i$ is estimated using the Large Deformation Diffeomorphic Mappings (LDDMM) framework [Vaillant 2005]. ϕ^i is parametrized by a smooth initial vector speed $\mathbf{v}_0^i$, which also belongs to a Gaussian RKHS V with variance λ_V^2. Moreover, this initial speed vector field is completely defined by the *moment* vectors β^i centered at the same point location as the template moments: $\mathbf{v}_0^i(\mathbf{x}) = \sum_k K_V(\mathbf{x}_k, \mathbf{x})\beta_0^i(\mathbf{x}_k)$. Finally, the template $\overline{\mathcal{T}}$ and the deformations ϕ^i towards each patient are estimated by means of an alternate two-step strategy, initialized with the mean current of the population.

8.2.2 Characterizing Deformation Modes of RV Shapes in ToF

In this work we analyse the deformations ϕ^i only as we mainly focus on the regional changes of the RV anatomy due to ToF. Principal Component Analysis (PCA) is performed directly on the moments β_i to extract the main deformation modes. The elements of the covariance matrix Σ are given by $\Sigma_{ij} =< \mathbf{v}_0^i - \overline{v_0}, \mathbf{v}_0^j - \overline{v_0} >_V = \sum_{\mathbf{x}_k,\mathbf{x}_l}(\beta^i(\mathbf{x}_k) - \overline{\beta}(\mathbf{x}_k))^t K_V(\mathbf{x}_k, \mathbf{x}_l)(\beta^j(\mathbf{x}_l) - \overline{\beta}(\mathbf{x}_l))$, $\mathbf{x}_k$ being the position of the k^{th} Dirac delta current of $\overline{\mathcal{T}}$. Then, the moment vector γ^m of the initial speed vector $\mathbf{u}_0^m$ related to the mth deformation mode is given by $\gamma^m = \overline{\beta} + \sum_i \mathbf{V}^m[i](\beta^i - \overline{\beta})$. In this equation, $\mathbf{V}_m$ is the mth eigenvector of Σ when the eigenvalues are sorted in decreasing order. Finally, the RV shape of each patient i is characterized by the shape vector $\mathbf{s}^i = \{s_m^i\}_{m=1..M}$ computed from the M first deformation modes, $s_m^i =< \mathbf{v}_0^i, \mathbf{u}_0^m >_V = \sum_{\mathbf{x}_k,\mathbf{x}_l} \beta^i(\mathbf{x}_k)^t K_V(\mathbf{x}_k, \mathbf{x}_l)\gamma^m(\mathbf{x}_l)$.

8.2.3 Can We Predict the Shape from Clinical Parameters?

First, cross-sectional analysis of the impact of growth on RV shape was performed. Multiple linear regression between the shape vectors $\mathbf{s}^i$ and body surface area (BSA) was carried out to exhibit the effects of BSA on each deformation mode. An optimal set of modes was selected by iteratively removing the modes with lowest significance, until the p-value of the regression overall significance stopped decreasing. Canonical Correlation Analysis (CCA) was then applied to quantify the amount of variation of each mode when BSA varies. Denoting R the overall correlation coefficient between BSA and shape vectors and ρ the correlation vector relating BSA to each deformation mode, the moments μ of the generative deformation Φ are $\mu = R \sum_k \rho[k]\gamma^k$. Deforming the template $\overline{\mathcal{T}}$ with Φ enables quantifying the average RV remodeling observed in our population.

Second, we assessed the impact of tricuspid and transpulmonary regurgitations on each deformation mode. As regurgitations were quantified by ordinal indices and only 18 subjects were available, we choose to perform two independent and component-by-component analyses to maximize statistical power. Rank-based Kruskal-Wallis analysis of variance was applied. If an effect was found for some deformation modes, post-hoc two-sample rank-based Wilconxon tests were used to determine which levels differed.

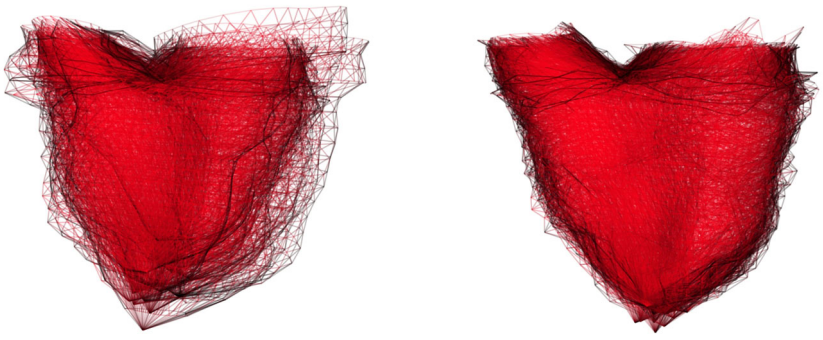
All the statistical tests were carried out using the shape vectors $\mathbf{s}^i$ (Sec. 8.2.2). The level of significance was set at $p < 0.1$ and multiple comparisons were corrected using Bonferroni adjustment.

8.3 Experiments and Results

8.3.1 Data Collection

Subjects and Image Preparation We selected 18 young patients (8 males, mean age $\pm$ SD $= 15 \pm 3$) with repaired Tetralogy of Fallot (ToF). Body-surface area (BSA) was reported for each patient (Dubois formula, mean $\pm$ SD $= 1.53 \pm 0.3$). Steady-State Free Precession cine MRI of the heart were acquired with a 1.5T MR scanner (Avanto, Siemens). Images were acquired in the short-axis view covering entirely both ventricles (10-15 slices; isotropic in-plane resolution: 1.1x1.1mm to 1.7x1.7mm; slice thickness: 6-10mm; 25-40 phases). Images were made isotropic through tri-cubic resampling.

Surface Meshes Preparation End-diastolic RV endocardium was segmented on the MRI cardiac sequence by fitting an anatomically accurate geometrical model. Its position, orientation and scale in the image was determined using minimal user interaction. Then, local boundaries were estimated by training a probabilistic boosting tree classifier with steerable features [Zheng 2007]. To reduce positioning effects in the shape analysis, the RV meshes were rigidly registered to a representative patient of the dataset by using GMM-Reg[1] [Jian 2005]. The results were visually inspected and remaining undesirable rotations were corrected manually (Fig. 8.1, left panel).



Rigid alignment Non-linear registration to the template

Figure 8.1: 3D RV meshes of 18 young ToF patients. *Left panel:* The meshes were rigidly registered to a representative patient of the dataset. Observe the extreme variability in shape. *Right panel:* The same meshes registered to the template using the non-linear deformations estimated during the template creation.

8.3.2 Statistical Shape Model of the Right Ventricles

Building the template $\overline{\mathcal{T}}$ required setting two parameters (Sec. 8.2.1): λ_V, which defines the "stiffness" of the non-linear deformations (the higher is λ_V, the more global are the transformations); and λ_W, which characterizes the resolution of the currents representation

[1] http://code.google.com/p/gmmreg/

(low λ_W values enable analyzing subtle shape features). As we were mainly interested in the regional ToF alterations (dilation, valve enlargement, regional bulging), these parameters were set to $\lambda_W = \lambda_V = 15$mm, about the diameter of the RV outflow tract. Lower values would have been inappropriate as the image slice thickness was approximately 10mm.

One iteration of the alternate minimization was needed to reach convergence. Yet, the resulting template $\overline{T}$ was well centered (mean over standard deviation of the deformations was 0.8). The first 10 deformation modes were selected, representing more than 90% of the total energy (Fig. 8.2).

Interestingly, the age of the closest patient to the template was 17 while his BSA 1.76. Both indices were close to the observed mean, suggesting that in our population, the mean shape was consistent with the mean BSA and age. Furthermore, this patient only suffers from trace valvar regurgitations, which is not surprising as no evident pathological bulging were visible in the template.

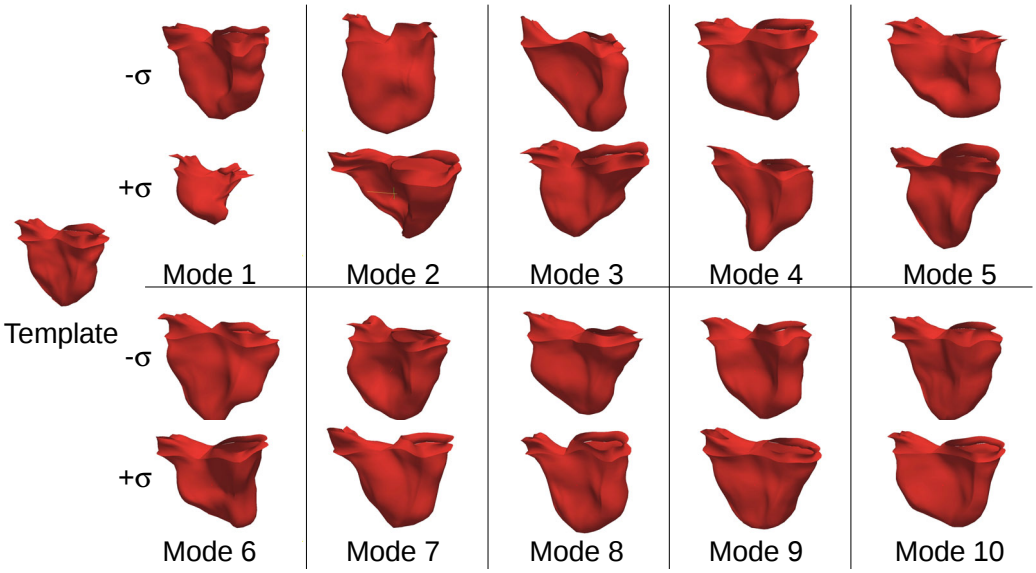


Figure 8.2: 10 first deformation modes extracted by PCA on a population of 18 patients suffering from repaired Tetralogy of Fallot. (Note: for visualization purposes, we match a subject to the estimated template so that the template and deformation modes can be seen as continuous meshes.)

8.3.3 Statistical Model of RV Remodeling in ToF Patients

Patient growth was quantified by body surface area (BSA) index (correlation with age in the data set: $R^2 > 0.5$, $p < 0.001$). Table 8.1 reports the regression coefficients a_l of the multiple linear regression between BSA and shape vectors $\mathbf{s}$, $\text{BSA} = a_0 + \sum_{l=1}^{10} a_l \mathbf{s}[l]$, the related t-values and the overall model significance. The sign of the a_l relates to the direction of the deformation modes (negative a_l meaning backwards deformations). Model reduction discarded all the non-significant modes (Table 8.1). The remaining modes were

found clinically pertinent by an expert after visual inspection (Fig. 8.2). Mode 1 clearly represented the overall RV dilation. Mode 2 seemed to model the dilation of the tricuspid annulus and of the inflow tract. Mode 3, 6, 7 and 9 exhibited a dilation of a specific RV region: apex (mode 3), basal area under the tricuspid valve (mode 6), apical area of the outflow tract (mode 7) and outflow tract (mode 9), reflecting possible direct impact of regurgitations on the neighboring tissues.

Canonical Correlation Analysis (CCA) provided a generative model of the RV remodeling observed in our population. Overall correlation coefficient with BSA was $R = 0.87$, suggesting a strong correlation between these deformation modes and growth. The correlation vector of the deformation modes was $\rho = \{-0.56, 0.45, -0.35, -0.33, -0.33, -0.37\}$. When BSA increases by 0.86, each deformation mode m varies by its related coefficient $\rho[m]$. The model was found clinically realistic by an expert (Fig. 8.3). As BSA increased, RV volume increased, RV free-wall and valves dilated, and septum was more concave. Indeed, dilation of the valves reduces the remaining pulmonary obstructions, thus decreasing the RV pressure. As a result, left-ventricle pushes the septum towards the right ventricle, making it more concave. However, as regurgitations are still present, the RV still dilates by pushing the RV free wall outwards.

	$\mathbf{a_1}$	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
Coef. $\times 10^{-5}$	**-2.9**	**6.4**	**-7.6**	4.6	-1.0	**-11.1**	**-11.9**	7.0	**-20.1**	-15.4
t-values	-3.28	2.64	-2.04	1.13	-0.19	-1.93	-1.92	0.84	-2.15	-1.43
Coef. $\times 10^{-5}$	**-2.9**	**6.4**	**-7.6**			**-11.1**	**-11.9**		**-20.0**	
t-values	-3.27	2.63	-2.03			-1.92	-1.92		-2.14	

Table 8.1: Linear regression coefficients a_l between shape modes and BSA, before model reduction (top row) with overall significance $R^2 = 0.84, p = 0.04$ and after model reduction (bottom row) with overall significance $R^2 = 0.75, p = 0.006$. In bold the significant coefficients ($p < 0.1$). After model reduction, coefficients stay unchanged, confirming the stability of the statistical tests.

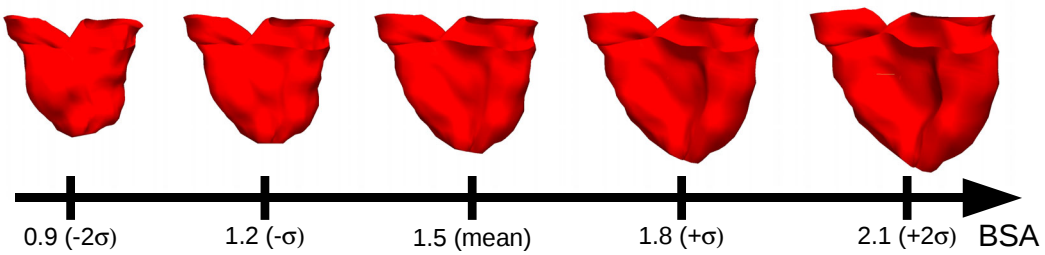


Figure 8.3: Mean RV remodeling observed in our population when body surface area (BSA, in m^2) increases. RV dimensions globally increase while valves dilate. Simultaneously, RV free wall becomes rounder and septum more concave.

8.3.4 Quantifying the Impact of Valvar Regurgitations on RV Shape

Color Doppler ultrasound (sweep speeds: 50-100 mm/s) was used to quantify tricuspid (TriReg) and transpulmonary valve (TPVReg) regurgitations. To assess the effects of TPVReg, patients were grouped into two different groups: trace TPVReg and severe TPVReg. Kruskal-Wallis analysis revealed a significant effect on deformation mode 2 ($p < 0.1$), which was confirmed by visual inspection: this mode exhibited an elongation of the RV outflow tract (Fig. 8.2).

Evaluation of TriReg classified the patients into 3 groups: none, trace or mild tricuspid regurgitations. Kruskal-Wallis analysis showed a significant impact of TriReg on three deformation modes: 3 ($p < 0.05$), 6 ($p < 0.1$) and 8 ($p < 0.1$). However, pair-wise Wilconxon tests showed that only mode 3 had two significant different levels (trace TriReg versus mild TriReg, $p < 0.1$). Visually, deformation mode 3 exhibited a deformation of the tricuspid annulus, from circular to triangular-shape, and a dilation of the RV inflow tract.

Interestingly, two deformation modes involved in the statistical model of RV growth were also related to the regurgitations. This may suggest possible cross-effects between growth and regurgitations on these specific shape variations.

8.3.5 Validating the Generalization of the Statistical Models

Generalizing a statistical model of RV remodeling is crucial for patient management and therapy planning. We thus tested the robustness of our model on two new patients with matched age (13 and 16). The template was registered to the patients and the related shape vectors $\mathbf{s}$ were computed. BSA were calculated from the optimal linear model estimated in Sec. 8.3.3. Results successfully compared with measured values (patient 1: estimated BSA: 1.61, measured BSA: 1.49; patient 2: estimated BSA: 1.29, measured BSA: 1.16). This suggests that the deformation modes involved in the RV remodeling model could be generalized, constituting potential quantitative parameters of remodeling in ToF.

8.4 Discussion and Future Works

In this study we investigated the impact of growth and valvar regurgitations upon the end-diastolic RV anatomy of patients with repaired ToF. End-diastolic time point was chosen as it is the time when the effects of the pathology are the most evident [Geva 2006, Sheehan 2008]. Multivariate statistical analyses provided a generative model of the observed RV remodeling. This model and the significant deformation modes were found clinically pertinent as they exhibited remarkably realistic changes in RV anatomy. To design the model, the deformation modes and their directions were statistically determined to limit the effects of PCA rotatability. Furthermore, the effects of regurgitation severity were analyzed on a component-by-component basis to preserve the statistical power of the tests due to the ordinal nature of the data. The groups were not sufficiently populated to apply more comprehensive statistics. Incorporating more patients is now required to confirm these findings and avoid possible over-interpretations. Various types of RV remodeling could be

identified (aneurysmal, with stiff myocardium, etc.), which may constitute new criteria for valve replacement decision. Future works also include analyzing the 4D cardiac motion. To the best of our knowledge, this study constitutes a first attempt at correlating 3D shape parameters to clinical measurements in ToF. These analyses may yield quantitative image-based predictors about RV anatomy and remodeling in ToF.

Acknowledgements This work has been partly funded by the European Commission, IST-2004-027749 Health-e-Child Integrated Project. The authors warmly thank Pr. G. Pongiglione, Istituto G. Gaslini, Genova, for his helpful comments about the models.

Part III

Statistical models of shape evolution

Chapter 9

Spatiotemporal variability analysis via mean scenario estimation and 4D-registration

This chapter has been partly published in [Durrleman 2009d, Durrleman 2009b].

Contents

9.1 Spatiotemporal variability of longitudinal data

In Chapter 5, we introduced statistical models to measure the anatomical variability of 3D anatomical structures. Given a set of anatomical shapes, we estimated a mean 3D anatomy and analyzed how this mean model varies in the population of shapes. This assumes implicitly that the 3D shapes are comparable. This means, in particular, that the anatomical shapes have been segmented in subjects of about the same age. Otherwise, it would be very difficult to distinguish between the inter-subject variability (the variations of the anatomy across subjects at a given age) from the variability due to the age (the variations of the anatomy of a given subject over time).

For this reason, there is an ever-growing interest in Computational Anatomy to analyze the variability of anatomical data both in terms of inter-individual differences and in terms of evolution of the structure over time. Indeed, pathologies may affect the evolution of an organ (during growth or aging) more than the shape of this organ at a given age. Even in a normal population, the anatomical variability may be due to both a change of morphology and a change of the speed of evolution of the structure. Let us consider girls and boys at adolescence for instance. There are obvious morphological differences between these two groups at any given age. But it is also clear that this variability is changing over time since the difference between boys and girls is increasing over time. The sexual dimorphism seems less pregnant between infants than between adults. In these cases, one wants to analyze how the growth process differs according to gender, or how the inter-subject variability evolves over time. Moreover, it is also common knowledge that the morphological changes at adolescence start earlier for girls than for boys[1]. In addition to the morphological differences between boys and girls, there are differences of the speed of the maturation process. As a consequence, we call "spatiotemporal" variability, the variability of an anatomical structure both in terms of geometrical changes over time *and* in terms of change of speed of evolution.

To analyze the spatiotemporal variability in a population, we need to compare the *evolution* of 3D-shapes over time and across subjects. For this purpose, we need *longitudinal* data. By longitudinal data, we mean a collection of subjects which have been scanned several times each (at least twice). A synthetic example of such a longitudinal database is presented in Fig. 9.1. In this chapter, we will define statistical models of spatiotemporal variability and infer the parameters of these models (mostly mean and covariance) from longitudinal data sets. These models will describe the variability of the studied population both in terms of morphological changes and in terms of change of speed of evolution.

To design such statistical models, we face two problems. First, the number of available scans per subjects is usually very small in most longitudinal studies in brain imaging. Nowadays, the publicly available database contain usually about 3 or 4 scans per subjects, even if we expect larger database to come in the future. Second, subjects are scanned a different number of times and at ages which do not necessarily correspond. To address these issues, we may consider a regression model which infers a continuous evolution from data sparsely distributed in time. This helps first to compare the data of one subject at a given age with the anatomy of another subject who has not been scanned at this age and, second, to estimate how the anatomy is varying at this age to detect possible change of the dynamics of evolution between subjects.

For the same reasons as in Chapter 5, we want to base our spatiotemporal variability analysis into a *generative* statistical model. This means that we consider the data as instances of a random process. Knowing a prototype evolution of the shape and its normal variability in the population, the statistical framework must answer these two questions: given a structure of subject scanned at time t, (1) how the prototype evolution can be derived to predict the evolution of the structure in the future? and (2) how the variability of this prototype evolution can predict the shape of this structure in another subject? As we shall see in Section 9.2.2, there are different ways to answer these questions, namely different

[1] Note that this is simply a pedagogic example. It is not supported by any rigorous scientific results.

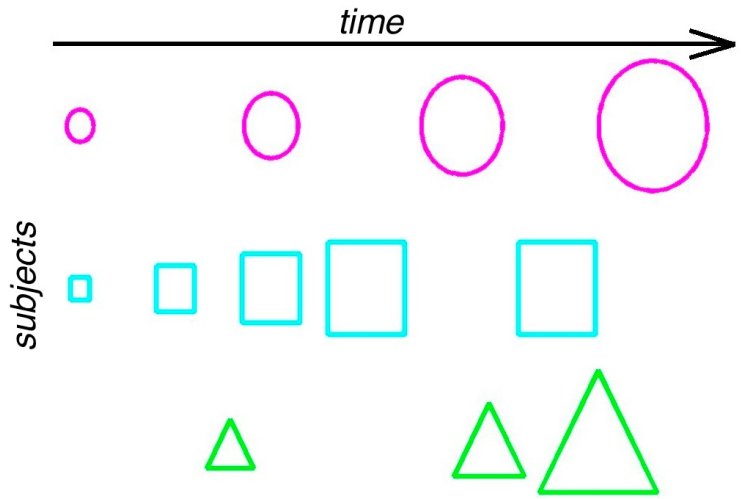


Figure 9.1: Synthetic example of a longitudinal database. The horizontal axis represents time. The aim of the spatiotemporal variability analysis is to describe the variability of this population in two ways: the one is the geometrical (anatomical) variability: there is a circle, a square and a triangle; the other is the variability in terms of change of dynamics of evolution: the square growths first at a faster rate than the circle and then slows down. To define such a model, we face two challenges: first, we have only few samples per subjects, sparsely distributed in time and, second, data have been acquired at time which do not correspond.

ways to generate instances from a prototype evolution. Depending on the hypotheses we use, the answers to these questions lead to different statistical models.

9.2 Methodology for statistics on longitudinal data

In this section, we aim at defining a general methodology for the spatiotemporal statistical analysis of longitudinal data. For this purpose, we start by a survey of the current literature on the topic. This will highlight the main tools that need to be linked consistently in a statistical framework. Then, we will show that there are two natural ways to define such a statistical model. They are based on different assumptions and therefore lead to different statistical estimations.

9.2.1 Several attempts to design 4D statistical analysis

9.2.1.1 4D analysis meant as regression or tracking

The first kind of 4D-analysis proposes to estimate a continuous sequence from a set of time-indexed shapes or images of the same subject. In [Craene 2009], the authors propose to use Large Diffeomorphic Free Form Deformations to estimate a dense deformation between the first and the last image of a sequence of cardiac images. This registration is driven by a global similarity criterion using Mutual Information between the time-indexed data

and the estimated continuous sequence. This method compares with the shape regression we proposed in Chapter 4, except that we used a different (but comparable) deformation framework and a different metric on data (Mutual Information versus distance in the space of currents). From the same perspective, it is proposed in [Davis 2007] to perform the regression of a sequence of images via a generalization of the kernel regression method to Riemannian manifolds.

All these methods are pure regression methods. If several subjects were scanned several times, these regression methods return a single evolution, the most probable evolution in some sense. However, it does not take into account that the scans come from different subjects. By contrast, in [Thompson 2000, Gogtay 2008], registrations between baseline and follow-up scans of the same subject are performed and scalar features are extracted such as the rate of growth of a particular anatomical structure. Then, these features are compared across subjects. However, we prefer here not to base our statistical model on some features, arbitrarily extracted from the data. In [Khan 2008], the authors propose to perform a regression of the image sequence of every subject separately (in a very similar framework as the one of Chapter 4) and then to average the time dependent velocity field of each regression to estimate a typical scenario of evolution. This approach makes sense only if each sequence is registered in the same reference frame, but no indication of how to perform registration of time-indexed sequences of images is given.

9.2.1.2 3D-registration of 4D-sequences

In cardiac motion analysis, the problem of registering time-indexed sequences has been investigated. For instance in [Chandrashekara 2003], the motion of the heart of each subject is tracked along time using Free Form Deformations. Then the end-diastolic anatomy is registered into the anatomy of a reference subject. This registration is used to transport the velocity field of the tracking of each subject into the common reference frame. Then, statistics are derived from these registered velocity fields. To overcome the problem of the choice of an anatomy of reference, one can use a template shape instead of a particular subject. For instance, in the context of longitudinal studies in brain imaging, the authors of [Qiu 2008, Qiu 2009] choose a template configuration which has been estimated from the baseline images of every subjects. Then, parallel transport is used to map the estimated evolution of each subject (estimated from its consecutive images) to the reference frame of the template. A similar approach has been proposed in lung imaging in [Ehrhardt 2008]. These methods suppose that the inter-subject variability can be captured considering only the baseline images. Using these deformations for registering the whole time-indexed sequence of images is arguable, since they do not take into account anatomical features which may appear later in the sequence.

This issue has been addressed in [Peyrat 2008] which proposed to register a time-indexed sequence by computing deformations between any successive scan of the same subject and between any pair of scans of two different subjects at the same time-point. This is done via a multi-channel co-registration of every image of the sequence simultaneously. Such an approach takes all temporal information into account and therefore leads to a much more robust registration scheme. However, this method supposes that every time-indexed

sequence has exactly the same number of images, which are acquired at time points which correspond across subjects. By contrast, in longitudinal studies, only few scans per subject are available, and the number of scans may vary drastically for different subjects. More importantly, using such a registration scheme in a statistical framework would be difficult. Indeed, to measure the inter-subject variability, one has to estimate the variability (the variance) of the inter-subject deformations independently at any time-point of the sequence. This means that the number of variables to be estimated tends to infinity as the temporal sampling of the sequence becomes finer and finer. It is not very reasonable not to take into account correlations between consecutive time points and across subjects. By contrast, we believe that the future evolution of one subject can be predicted (even with uncertainty) knowing how the anatomy varies for several other subjects. A generative statistical model should therefore explain how shape evolution can be derived from subjects to subjects.

In [Gerig 2006], a template image is built at every time-point independently. Then, the deformations between the baseline scan and the follow-up are compared to the deformation between the baseline atlas and the follow-up atlas. This approach enables to compare the anatomy of different subjects at the same age and the average change of anatomy over ages. In other words, it compares how the intra-subject differences vary across subjects. However, since the template image is build at each time point independently, it is not clear whether the difference between baseline atlas and follow-up atlas is the average of every subject's evolution. Moreover, such a method requires that the distribution of ages in the longitudinal database is very peaked at two distinct ages. An approach based on similar ideas has been also proposed in [Aljabar 2008].

9.2.1.3 Taking into account temporal re-alignment

The methods cited previously propose a way to combine the intra-individual evolution with the inter-subject variability: $4D$ image sequence are dealt with a set of $3D$ deformations. In particular, the age, at which the subjects are scanned, is considered as absolute time, which is measured independently of the subjects. This neglects the possible developmental delays between subjects (or some anomaly in the cardiac motion), a key feature that we precisely want to detect. A real $4D$-registration scheme should register a sequence both in space (usual geometrical variations of the anatomy) and in time (change of the speed of evolution).

In [Declerck 1998], a deformation of the 4D domain is provided via 4D planispheric transformations for the registration of the heart motion. In [Perperidis 2005], spatiotemporal deformations are computed. The temporal part is a 1D function showing the change of cardiac dynamics between the source and the target subject. This temporal alignment is performed jointly with the usual registration of the anatomy. This method requires a fine temporal sampling of the sequence for each subject. In particular, no regression function is provided to account for possible lack of data. Whereas this is a not so strong assumption in cardiac imaging, this is a clear limitation in pediatric or aging studies in brain imaging. These methods focus mostly on pairwise registrations and it is difficult to figure out how to use these deformations consistently into a statistical framework.

9.2.1.4 Requirements for a spatiotemporal statistical model

This short review of the literature shows that many aspects of the design of a 4D statistical analysis have been addressed separately by different authors, in different contexts, with different tools. This shows that a statistical framework for the variability analysis of longitudinal data should include:

- An *evolution function* which shows how the anatomy of a subject can be computed at any time-point. There should be a way to estimate this evolution function from data sparsely distributed in time (i.e. a set of few time-indexed shapes). The evolution function should be used to compare the anatomy of one subject scanned at a given age with another subject which has been scanned at another age.

- A *spatiotemporal deformation* framework which enables to deform evolution functions both in space (geometrical deformations of the $3D$-space which accounts for morphological variability) and in time (deformation of the time interval, which accounts for different speeds of evolution).

A statistical framework should combine in a consistent way the two previous functions in order to explain how the evolution of different anatomies derived the ones from the others. As we shall see in the next section, there are at least two natural ways to define such a statistical model.

9.2.2 Two possible generative models for longitudinal data

9.2.2.1 Spatiotemporal variations of a typical growth model

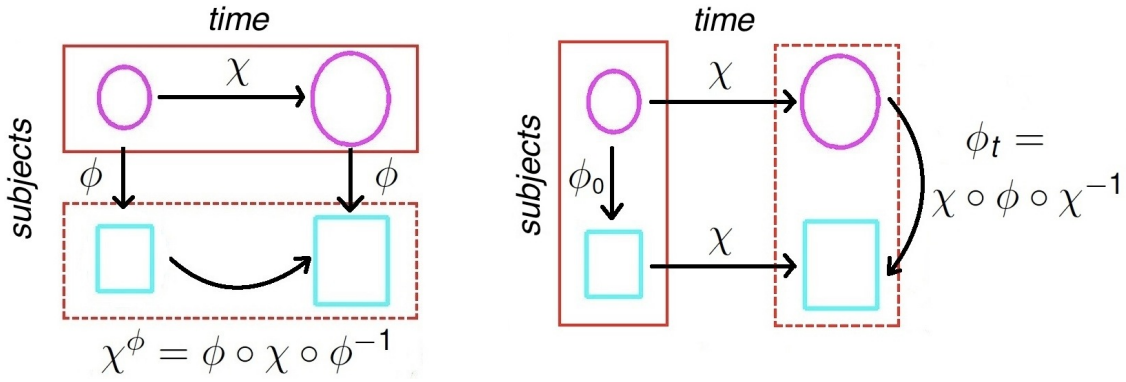
In this section, we want to design a statistical model which explains how the observed longitudinal data can be derived from the knowledge of a typical scenario of evolution. This should answer these two fundamental questions: given the anatomy of the subject at time t, how can we estimate the future evolution of this subject? and how can we estimate the appearance of this anatomy in another typical subject?

Let $M(t)$ be a smooth random process defined in the continuous time interval. It models the prototype evolution of the anatomy over time. For each time t, the law of the variations of this process describes the inter-subject variability at this given time-point. This variability allows us to derive the evolution of a particular subject $S(t)$ from the prototype scenario $M(t)$. One may consider that any evolution $S(t)$ is possible for each subject and that the inter-subject variability varies arbitrarily over time. This leads to a framework similar to the one presented in [Peyrat 2008], which is not constrained enough for statistical estimations, as we saw in the previous sections. To constrain the model, we can consider that:

- the inter-subject variability is stationary and the evolution function $S(t)$ is specific to every subject: the "subject-specific" approach.

- the inter-subject variability depends on time and the evolution function $S(t)$ is identical for every subject: the "time-specific" approach.

The schematic differences between both approaches are highlighted in Fig. 9.2 and Table 9.1. We describe now these two approaches with more details.



subject-specific approach time-specific approach

Figure 9.2: Two subjects have been scanned twice (one subject per rows, one time-point per column). There are two different ways to describe the variability of this population. **In the subject-specific approach**, one considers that one subject is "circle" and the other is "square": the difference between both subjects is a single function ϕ, which maps circles to squares. The evolution of the first subject is described by a function χ which maps a small circle to big circle. As a consequence, the evolution of the second subject is described by another function χ^{ϕ} which maps a small square to a big square. **In the time-specific approach**, one describes the evolution by a universal function χ, which is a scaling of the space in this example. At the first-time point, one describes the difference between subjects by a function ϕ which maps small circle to small square. At a later time, the inter-subject variability has changed according to the scaling χ: now the difference between both subjects is another function ϕ_t which maps a big circle to a big square. *In the subject-specific approach, one estimates how the intra-subject variations change from subjects to subjects. In the time-specific approach, one estimates how the inter-subject variations evolve over time.*

9.2.2.2 Subject-specific approach

In the subject-specific approach, we consider that a particular reference frame is attached to each subject. The evolution of each subject is described within its *own* reference frame. Since the whole evolution is described in the same reference frame, this reference frame is atemporal. We assume that there is a particular reference frame in which the evolution is written by the evolution function $M(t) = \chi_t(M_0)$, where χ_t is a deformation of the space and M_0 a prototype shape. This is the prototype scenario of evolution, the analogous in $4D$ of the template shape estimated in $3D$.

We introduce 3D-functions ϕ which map the reference frame of the mean scenario of evolution to the reference frame of each subject. These functions can be seen as a change of coordinates, written in Eulerian coordinates in the reference frame of mean scenario. Since the reference frames are atemporal, these functions ϕ do not depend on time. As

a consequence, the evolution function $M(t)$ has a different expression in each coordinate system: the evolution of a subject S is given as $S(t) = \phi(M(t))$. As illustrated in Fig. 9.2, the evolution function $\chi(t)$ is transported in the subject's space by the change of coordinate function ϕ: $\chi^S(t) = \phi(\chi(t))$. The evolution map is specific to each subject.

We can include time as an additional variable, so that the reference frame of each subject is described by 3 spatial coordinates and 1 temporal coordinate. This means that both the anatomy and the age is relative to the subject. This specific time variable can be called the "physiological age" of the subject, as if each subject has his own biological clock. In the reference frame of the prototype, the time would be the absolute age, computed from the date of birth. The $3D$ warp $\phi(x, y, z)$ becomes a deformation of the underlying $4D$ space: $\Phi(x, y, z, t)$.

The most general form of a $4D$-deformation is $\Phi(x, y, z, t) = (\phi(x, y, z, t), \psi(x, y, z, t))$, where $\phi(x, y, z, t)$ denotes the 3 spatial coordinates of $\Phi(x, y, z, t)$ (the morphological deformation) and $\psi(x, y, z, t)$ its temporal coordinate (the time warp).

If $\psi(x, y, z, t)$ depends on the spatial variables (x, y, z), this means that some parts of the anatomy of one subject may be delayed with respect to another subject, whereas another part of the anatomy may be in advance. Different parts of the anatomy evolve at different speeds. In this chapter, for the sake of simplicity, we will assume that ψ depends only on the time variable t, thus assuming that all points of the anatomy of a given subject have always the same physiological age over time. This assumption is likely to be valid in most of the longitudinal studies. Eventually, the time warp $\psi(t)$ maps the absolute age in the reference frame of the template to the physiological age of a given subject.

In this subject-specific approach, we assume that the morphological deformation is independent of time, thus leading to a map ϕ of the form: $\phi(x, y, z)$. In this setting, the $4D$ deformations have the form:

$$\Phi(x, y, z, t) = (\phi(x, y, z), \psi(t)).$$

The morphological deformation ϕ is used to measure the geometrical variability. The time warp ψ is used to detect possible developmental delays between subjects.

By contrast, in [Peyrat 2008], ϕ depends on time $(\phi(x, y, z, t))$ (without specifying a particular form of this dependency) but no developmental delays are allowed: $\psi(t) = t$. As we mentioned in Section 9.2.1.2, such spatial deformations ϕ are not constrained enough for statistical purposes. Moreover, introducing a temporal warp ψ different from the identity map makes the model even less constrained and therefore even more difficult to estimate.

9.2.2.3 Time-specific approach

In the time-specific approach, we suppose that there is a common reference frame at reference time $t = 0$ (the "origin of the world") in which the anatomy of every subject is described. The evolution function $\chi(t)$ changes the geometry of this reference frame over time. This evolution function is written in *Eulerian* coordinates with respect to this common reference frame. At each time, there is one single reference frame: it is universal. The same function χ_t applies for each subject, so that the evolution of any subject S is given by $S(t) = \chi_t(S_0)$, where S_0 the anatomy of the subject at the reference time $t = 0$.

	subject-specific approach	**time-specific approach**
Mean model	Mean scenario of evolution: $M(t)$	A template (prototype) at time $t = 0$: M_0
Evolution function (χ_t)	relative to the template space	universal (independent of the subject)
inter-subject variability (ϕ)	atemporal	relative to a reference time $t = 0$
Given the mean model, how to derive a shape $S(t)$at time t?	$S(t) = \phi\,(M(t))$	$S(t) = \chi_t\,(\phi(M_0))$
Given $S(t = 0)$, how to derive $S(t)$ at a later time $t > 0$?	$S(t) = \phi \circ \chi_t \circ \phi^{-1}(S(t = 0))$ ϕ maps template- to subject-space	$S(t) = \chi_t(S(t = 0))$
Given inter-subject difference (ϕ) at $t = 0$, how to derive $S_2(t)$ from $S_1(t)$?	$S_2(t) = \phi(S_1(t))$	$S_2(t) = \chi_t \circ \phi \circ \chi_t^{-1}(S_1(t))$
Typical examples from the literature	[Qiu 2009] [Ehrhardt 2008] [Perperidis 2005]	[Habas 2009] [Aljabar 2008] [Gerig 2006]

Table 9.1: Summary of the main differences between subject-specific and time-specific approach.

In the common reference frame at $t = 0$, the differences between a template anatomy and each subject's anatomy is described by 3D maps ϕ. These functions describe the inter-subject variability at time $t = 0$. They are written in Lagrangian coordinates in the reference frame attached to time $t = 0$. As a consequence these functions (i.e. the inter-subject variability) have a different expression at every time. At a later time t, the template-to-subject registration is given by: $\phi_t = \chi_t \circ \phi \circ \chi_t^{-1}$. The inter-subject variability is transported along with the evolution function χ_t. The evolution function is independent of the subject and the inter-subject variability is specific to time.

Note that we can include possible development delays in this framework also. If $\chi(t)$ is a universal function which carries the anatomies over time, we can imagine that every subject follow this universal scenario with its own speed. There is a subject specific time warp ψ, so that the evolution of this subject is given by $\chi(\psi(t))$. However, we must admit that this time-realignment fits less naturally in this time-specific framework than for the subject-specific framework. In particular, it is not clear to us how to distinguish a developmental delay from a variation of the inter-subject variability in this setting.

This time-specific approach defines also a deformation of the underlying $4D$-space $\Phi(x, y, z, t)$. Contrary to the subject-specific approach, the morphological deformation ϕ depends on time in this approach. But this dependency has a particular form, which is

given by the evolution function χ_t. More precisely the $4D$ deformations have the form

$$\Phi(x, y, z, t) = (\phi(x, y, z, \psi(t)), \psi(t)),$$

where the geometrical part has the form:

$$\phi(x, y, z, t) = \chi_t \circ \phi_0 \circ \chi_t^{-1}(x, y, z).$$

The subject-specific and the time-specific approaches give two different ways to derive the evolution of an anatomy from a prototype scenario of evolution. This gives two different ways to extract significant information from longitudinal data. Depending on the nature of the variability and the anatomical features one wants to detect, one approach may be more adapted than the other.

9.2.2.4 Which method for which problem?

The subject-specific approach focuses on the variations of the evolution of the anatomy across subjects. The time-specific approach focuses on the evolution of the inter-subject variability over time. Theses two approaches are based on different assumptions. They lead to different statistical estimations.

We think that the subject-specific approach is adapted when one wants to analyze the change of anatomy with respect to some characteristics of the subjects such as pathologies, subtypes of pathologies or some physiological parameters which are not supposed to change over time like gender or blood type for instance. In this case, the inter-subject variability is not supposed to change over time and one focuses on how the evolution of the anatomy is derived from subjects to subjects.

The time-specific framework seems well adapted to describe the variability of evolution as an evolution of the inter-subject variability over time. At the time origin (birth for instance), the variability is relatively small (or does not show particular modes). Then, as time goes on, the variability is magnified in many directions, possibly highlighting several significant modes. Note that if some inter-subjects differences are constant over time, they will be described by many variables (inter-subject differences at each time) and will be therefore more difficult to detect.

Let us take the same example as in the introduction about boys and girls at adolescence. If one wants to measure how some anatomical structures evolve differently (both in shape and in speed) for boys and for girls, then a subject-specific approach is adapted. This may show, for instance, that the morphological changes occur earlier for girls than for boys. By contrast, if one wants to show how the sexual dimorphism becomes more and more significant from childhood to adulthood, then a time-specific approach would be more adapted.

Therefore, the choice of the model should be determined by the applications and the anatomical features that one wants to detect. In this chapter, we would like to show how pathologies such as autism affect the development of deep brain structures. Here the difference between autistics and control subjects is established from a psychological diagnosis. The difference of shape evolution with respect to the pathology will be described more naturally in the subject-specific approach. This will also focus on the developmental

delays between subjects, a key feature we want precisely to detect. Moreover, we will use a database with only 2 acquisitions per subjects. Using a time-specific approach will raise robustness issues.

Before showing how we can implement the subject-specific approach for concrete applications, we show that the discussion between subject-specific and time-specific approaches may lead to a classification of the literature on longitudinal studies.

9.2.2.5 Back to the literature

The papers which were referenced in Section 9.2.1 can be classified according to the previous discussion: some are based on the assumptions of the subject-specific methodology, others follow the assumptions of time-specific methodology.

In [Chandrashekara 2003, Qiu 2009, Ehrhardt 2008], the idea is to estimate a scenario of evolution for each subject (using regression for instance) and to transport this scenario into a common reference frame for comparison. Such models fall naturally in the category of subject-specific model. In [Perperidis 2005], the spatial deformation does not depend on time: it is based also on a subject-specific framework. In [Gogtay 2008], the authors compute the differences between consecutive scans of the same subject and compare them within a population: even if the statistical model is only descriptive and not generative, the underlying framework is subject-specific.

By contrast, in [Gerig 2006, Aljabar 2008] the strategy adopted is to estimate an atlas for every age (usually the baseline and the follow-up): the inter-subject variability is supposed to depend on the age. Then, one compares the evolution between one shape and its follow-up with the evolution between the baseline template and the follow-up template. These models are based therefore on the assumptions of the time-specific approach. In [Habas 2009], the authors compute a probabilistic atlas of a fetal brain at different time points and use a regression function to interpolate between these probabilistic maps. This falls typically in the category of time-specific approaches.

9.3 A subject-specific approach using currents

From now on, we will focus on the subject-specific methodology. This framework seems better suited for the estimation of a typical scenario of growth, to couple spatial deformation with possible time re-alignment and therefore to detect possible developmental delays, a feature which motivated our investigations.

The subject-specific methodology, as described in the previous section, is very general. It can be applied on any kind of data (images, shapes, etc.) with very different tools, as the survey of the literature highlighted. The purpose of this section is to give an instance of this methodology using currents introduced in Chapter 1 for modeling shapes and large deformations introduced in Chapter 4 for capturing shape variations. Using currents enables to define a metric between shapes, which does not assume point correspondence between structures. Using the large deformations in the LDDMM framework allows us to root the statistical analysis of deformations into a rigorous framework. This is an extension in the subject-specific setting of the atlas construction method introduced in Chapter 5.

As emphasized in the previous section, designing a spatiotemporal variability model requires three ingredients. First, one needs an individual growth model in order to infer a continuous evolution of the shapes of one subject from its successive observations. This will allow us to compare the anatomy of one subject with the anatomy of another subject who has not been scanned at the same age. Moreover, this will allow us to analyze how the anatomy is varying locally around this age and therefore to detect possible developmental delays. We will based this individual growth model on the regression scheme introduced in Chapter 4.

Second, we will focus on spatiotemporal registrations between two subjects scanned several times, but not necessarily the same number of times. First, we compute the individual growth function of the source subject. Then, we want to estimate a spatiotemporal deformation of this continuous evolution, so that the sampling of this deformed evolution matches the target data. In the subject-specific approach, the spatiotemporal deformations of a continuous evolution consists of two deformations: (1) a morphological deformation (of the 3D space) which changes the geometry of every frame of the evolution *independently of the time points* and (2) a time warp (deformation of the time interval) which changes the dynamics of the evolution *without changing the geometry of shapes* (i.e. accelerate or slow down the speed of evolution).

Third, we define a statistical model for longitudinal data based on these spatiotemporal registrations. From a longitudinal database, we estimate first a template shape, second a mean scenario of evolution of this template (an individual growth model in the coordinate system of the template) and, third, the spatiotemporal registrations of this mean scenario to each subject in the population. This set of template, mean scenario and spatiotemporal deformations is called a *spatiotemporal atlas*. It is estimated via a Maximum A Posteriori approach, so that the shapes of each subject are seen as temporal samples of a spatiotemporal deformation of the mean scenario. Then, statistics can be performed on the spatiotemporal deformations to measure the spatiotemporal variability of the population: statistics on the geometrical part measure the anatomical variability independently of the age of the subjects, statistics on the temporal part measure the variability in terms of speed of evolution.

9.3.1 Growth model for individual shape evolution

In this section, our purpose is to fit a continuous shape evolution to a set of shapes (S_i) of the same subject acquired at different time points (t_i). Without loss of generality, we can assume that all time-points fall into the interval $[0, T]$. We write the individual growth model as $\chi_t(M_0)$, where M_0 is a baseline shape, t varies continuously in the time-interval $[0, T]$ and $(\chi_t)_{t\in[0,T]}$ a flow of diffeomorphisms as defined in Section 4.2. In particular, $\chi_0 = \mathrm{Id}$, which leads to $\chi_0(M_0) = M_0$.

The regression $(M(t) = \chi_t(M_0))$ is defined so that it minimizes the discrepancy between $M(t_i)$ and the shapes S_i at the time-points t_i, while a constraint controls the regularity of the regression. This is achieved by minimizing:

$$J(\chi) = \sum_{t_i} \mathrm{d}(\chi_{t_i}(M_0), S_i)^2 + \gamma^{\chi}\mathrm{Reg}(\chi) \tag{9.3.1}$$

where d is a similarity measure between shapes, $\mathrm{Reg}(\chi)$ a regularity term and γ^{χ} a trade-off between regularity and fidelity to data.

Defining d as the distance between currents, and $\mathrm{Reg}(\chi)$ as the total kinetic energy of the flow of diffeomorphism in the framework of Chapter 4, this equation is exactly the regression criterion as stated in Section 4.3.2. In this framework, the velocity field of χ_t is supposed to belong to a RKHS. We denote the kernel of this RKHS K^{χ}. In our application, we set K^{χ} as a Gaussian kernel: $K^{\chi}(x,y) = \exp(-|x-y|^2/\lambda_{\chi}^2)$. The standard deviation λ_{χ} is the spatial scale at which the points of M_0 move consistently. This is the regularity of the growth model. The parameter γ^{χ} is the usual trade-off between regularity and fidelity-to-data.

If there is only one data S_1 at time $t_1 = T$, this is the registration of M_0 to S_1. The result is a geodesic flow of diffeomorphism between $t = 0$ and $t = T$ that maps M_0 close to S_1. With several data at successive time points, the result is a flow of diffeomorphism which is geodesic only between successive time points (i.e. piecewise geodesic). We recall that the resulting regression function fits every data globally (i.e. taking into account all constraints simultaneously) and as such, differs from pairwise registration between consecutive shapes (see Section 4.3.2). For instance, if the trade-off γ tends to infinity (no fidelity-to-data term) the regression is a constant map. As γ increases, the constant map becomes a piecewise geodesic deformation which matches the data with increasing accuracy. This framework allows us also to perform the regression even if several data are associated to the same time-point as shown in Fig. 4.8 for instance.

Let $t_{\min}$ and $t_{\max}$ be the earliest and latest time-point in the database, respectively. The function χ_t is constant in the interval $[t_{\max}, T]$ (i.e. the minimizing velocity field vanishes in this interval). Setting a large time-interval $[0, T]$ allows us to extrapolate the evolution function outside the range of the ages of the data $[t_{\min}, t_{\max}]$. This is a useful property in order to compare this evolution with data of a new subject who has been scanned at ages outside the interval $[t_{\min}, t_{\max}]$.

This continuous shape evolution will be used in two ways: first, for pairwise registration in Section 9.3.2, one needs to perform the regression of the shape of the source subject, in order to compare the target's shape with the source at time-point where the source may have not been scanned and, second, for atlas estimation in Section 9.3.3, the mean scenario will be set as the continuous shape evolution of a template shape. In the first application, we set the baseline M_0 to the shape associated to the earliest time-point. In the second application, we set the baseline M_0 as the template shape (to be estimated).

9.3.2 Spatiotemporal pairwise registration

9.3.2.1 Spatiotemporal deformations and matching criterion

In this section, we suppose that we have two subjects S and T which have been scanned several times each (but not necessarily the same number of times and at possibly different ages). Let S_{t_i} (resp. T_{t_j}) denotes the segmented shapes of subject S (resp. T) at ages t_i

(resp. t_j). We define a time-interval of interest which contains every t_i's and t_j's. Without loss of generality, we can assume that this time interval of interest is of the form $[0, T]$.

We infer an individual growth model $S(t)$ from the data of the source subject S_{t_i}, using the growth model of the previous section (Sec 9.3.1). As a result, the continuous shape evolution $S(t)$ is of the form: $S(t) = \chi_t(S_0)$ for $t \in [t_{min}, T]$, where the baseline S_0 is the shape $S_{t_{min}}$ which corresponds to the earliest time point t_{min} within the t_i's. In the framework of Chapter 4, we denote v_t^χ the velocity fields which parameterize the flow of diffeomorphisms χ_t. We set $v_t^\chi = 0$ in the time-interval $[0, t_{min}]$ (as a consequence of the regression function, we also have $v_t^\chi = 0$ in the time-interval $[t_{max}, T]$), so that $S(t)$ is extended in a continuous shape evolution in the whole interval $[0, T]$. In particular, $S(t)$ is constant in the interval $[0, t_{min}]$ and $[t_{max}, T]$ (see Section 9.5.4 for alternative growth model with non-constant extrapolation).

Our goal is to defined a spatiotemporal deformation of the continuous evolution $S(t)$ into $S'(t)$ so that the deformed shapes $S'(t_j)$ at the time-points of the target t_j match the shape T_{t_j} (thanks to the continuous regression, we can define $S(t_j)$ for the target time point t_j even if the source has not been scanned at this age.) For this purpose, we introduce two functions: the morphological deformation ϕ and the time warp ψ. The morphological deformation is a 3D-function, which maps the geometry of the source into the geometry of the target (change of reference frame). It accounts for the geometrical variations between source and target. The time warp ψ maps the time-points t within the time interval $[0, T]$ to $\psi(t)$. This 1D-deformation models the change of the dynamics of the evolution of the source with respect to the evolution of the target. In the subject-specific paradigm, we define the spatiotemporal deformation of the continuous evolution $S(t)$ as:

$$S'(t) = \phi\left(S\left(\psi(t)\right)\right). \tag{9.3.2}$$

Using the fact that $S(t) = \chi_t(S_0)$, this becomes[2] $S'(t) = \phi(\chi_{\psi(t)}(S_0))$.

Since the shapes $S'(t_j)$ must match the target's shape T_{t_j}, we define the 4D-registration criterion as:

$$J(\phi, \psi) = \sum_{t_j} \mathrm{d}(\phi(S(\psi(t_j))), T_{t_j})^2 + \gamma^\phi \mathrm{Reg}(\phi) + \gamma^\psi \mathrm{Reg}(\psi), \tag{9.3.3}$$

where d is a distance between shapes, $\mathrm{Reg}(\phi)$ and $\mathrm{Reg}(\psi)$ the measure of regularity of the deformation ϕ and ψ and γ^ϕ, γ^ψ the usual trade-offs between regularity and fidelity to data.

As we did in Chapter 5, we model the inter-subject variability by diffeomorphisms (see Remark 9.1 for a discussion on this choice). Therefore the morphological deformation ϕ is a $3D$-diffeomorphism like for the usual $3D$ registration. In the framework of Chapter 4, we construct $3D$-diffeomorphism by integrating a varying vector field v_u^ϕ for the parameter[3] u in the interval $[0, 1]$:

$$\frac{\partial \phi_u(x)}{\partial u} = v_u^\phi(\phi_u(x)), \tag{9.3.4}$$

with the initial condition $\phi_0(x) = x$ (i.e. identity map). The morphological deformation ϕ is set to the final diffeomorphism ϕ_1.

[2] We notice that in the time-specific paradigm, this would be $S''(t) = \chi_{\psi(t)}(\phi(S_0))$.

[3] In Chapter 4, the parameter u was denoted t and we called v a time-varying vector field. However, in 4D-registration, the variable t and name "time" have another meaning.

In order to avoid time reversal, the 1D-time warp ψ must be strictly order-preserving, so that we can assume in a very realistic way that ψ is a 1D-diffeomorphism of $\mathbb{R}$ (see Remark 9.1). In the setting of Chapter 4, we construct such diffeomorphism by integrating a real-valued function v_u^ϕ (the speed profile of the time warp) for a parameter $u \in [0,1]$:

$$\frac{\partial \psi_u(t)}{\partial u} = v_u^\psi(\psi_u(t)), \tag{9.3.5}$$

with the initial condition $\phi_0(t) = t$ (i.e. identity map). The time warp ψ is set to the final diffeomorphism ψ_1.

As a consequence, the criterion in Eq. 9.3.3 is of the form $J(v^\phi, v^\psi)$. Now, we assume that for every parameter u, the velocity fields v_u^ϕ and v_u^ψ belong to a RKHS with kernel K^ϕ and K^ψ respectively. We denote $\{x_p\}$ the set of points of the discrete shape S_0. The source trajectory $S(t) = \chi_t(S_0)$ is described by the moving points $x_{p,t}$, where p denotes the *discrete* index of the points of S_0 and t the *continuous* time variable within the interval $[0,T]$. Therefore the fidelity-to-data term in Eq. (9.3.3) depends on the variables $\phi_1\left(x_{p,\psi_1(t)}\right)$ (see Fig. 9.3 for an illustrative scheme of the notations). The application of Proposition 4.8 to both v^ϕ and v^ψ leads to the following parameterization of the minimizing velocity fields:

$$v_u^\phi(x) = \sum_{p,t_j} K^\phi(x, \phi_u(x_{p,\psi_1(t_j)}))\alpha_{p,j}(u) \tag{9.3.6}$$

and

$$v_u^\psi(t) = \sum_j K^\psi(t, \psi_u(t_j))\beta_j(u) \tag{9.3.7}$$

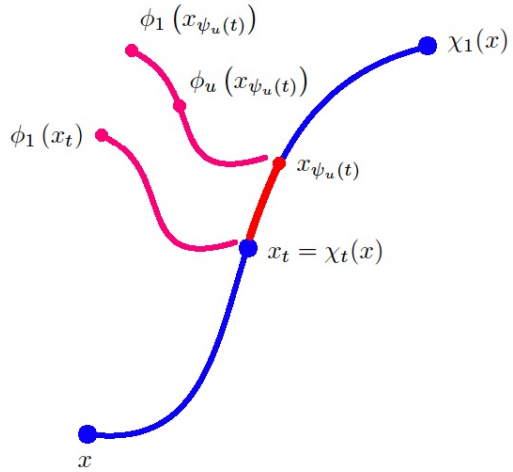


Figure 9.3: Illustrative scheme for the notations: x denotes a generic point of the source shape, $x_t = \chi_t(x)$ the continuous evolution of the source point, $(\psi_u)_{u\in[0,1]}$ is a flow of $1D$-diffeomorphism which moves the time-labels along the time-axis, $(\phi_u)_{u\in[0,1]}$ is a flow of $3D$-diffeomorphism which moves the points of the source evolution, *independently* of the time.

This means that, for every parameter u, the minimizing v_u^ϕ is parameterized by the varying momenta $(\phi_u(x_{p,\psi_1(t_j)}), \alpha_{p,j}(u))$. The momenta are located at the points of the moving shape $\phi_u(S(t))$ at the registered time points $\psi_1(t_j)$. The minimizing v_u^ψ is parametrized by momenta $(\psi_u(t_j), \beta_j(u))$. The 1D-functions $\beta_j(u)$ for $u \in [0,1]$ are scalar coefficients which parametrize the 1D-speed profile v_u^ψ. The integration of this speed profile according to the flow equation in Eq. (9.3.5) leads to a monotonic smooth scalar function.

As a consequence, the criterion in Eq. (9.3.3) depends the discrete set of L^2 functions $\alpha_{p,j}(u)$ and $\beta_j(u)$, where p is the index of the points of S_0 and j the index of the time-points of the target data T_{t_j}.

For the sake of simplicity, we introduce the notations $x_{p,t}(u)$ and $t_j(u)$ such that:

$$\begin{array}{ccc} \phi_u(x_{p,t}) = x_{p,t}(u) & x_{p,t}(0) = x_{p,t} & x_{p,t}(1) = \phi(x_{p,t}) \\ \psi_u(t) = t(u) & t(0) = t & t(1) = \psi(t) \end{array} \tag{9.3.8}$$

The regularity parameters are given by:

$$\mathrm{Reg}(\phi) = \int_0^1 \left\| v_u^\phi \right\|_{V^\phi}^2 du = \int_0^1 \sum_{p,j,p',j'} \alpha_{p,j}(u)^t K^\phi(x_{p,\psi_1(t_j)}(u), x_{p',\psi_1(t_{j'})}(u)) \alpha_{p',j'}(u) du \tag{9.3.9}$$

and

$$\mathrm{Reg}(\psi) = \int_0^1 \left\| v_u^\psi \right\|_{V^\psi}^2 du = \int_0^1 \sum_{j,j'} \beta_j(u)^t K^\psi(t_j(u), t_{j'}(u)) \beta_{j'}(u) du \tag{9.3.10}$$

Remark 9.1 (On the assumption of diffeomorphic maps)**.** In our framework, we suppose that the evolution function χ, the morphological deformation ϕ and the time warp ψ are all diffeomorphisms. However, this choice is more or less arbitrary depending on the function.

The evolution function χ maps the anatomy of a subject over time. Setting χ as a diffeomorphism assumes that the structure of a subject evolve smoothly over time. This includes feature like growth, atrophy, dilatation in some directions, torque, etc. However, this cannot model a tearing of the structure or the creation of another disconnected component of the structure over time. This assumption is realistic in many cases. For instance, it is reasonable to model the cardiac motion as diffeomorphic over one cycle: this is an elastic deformation of the biological material. Similarly, brain structures evolve often in a diffeomorphic way, at least at the resolution of the segmented shapes.

The morphological deformations ϕ model the geometrical inter-subject variability. Assuming that two different anatomies derive one from the other in a diffeomorphic way is a non-realistic assumption. We already discussed this choice in Sections 4.1.2 and 5.1.2: assuming that ϕ is diffeomorphic allows us to decompose the inter-subject variability into two terms: a diffeomorphic geometrical variability and a non-diffeomorphic variability in terms of "texture". We showed in Chapter 5 and Chapter 7 how to include the non-diffeomorphic residuals into the statistical analysis and how it may lead to relevant results. In this case, assuming that ϕ is diffeomorphic does not limit the anatomical variations that the model can capture.

The time warp ψ model the change of speed of evolution between subjects. Since every subject evolves in the same temporal direction, namely from birth to death, this time warp must be strictly order preserving. This holds even if a structure evolves in such a way that the later evolution of the structure looks like its earlier evolution: the age of the structure still increases. Moreover, assuming the evolution of a structure is at least differentiable (so that we can speak about the *speed* of evolution) as well as its inverse is a very realistic hypothesis. Therefore, the time warp ψ are intrinsically diffeomorphic. □

9.3.2.2 Exact differentiation of the criterion

The criterion J depends on the L^2 functions $\alpha_{p,j}(u)$ and $\beta_j(u)$. The differentiation of J with respect to the spatial momenta α is performed as for a 3D registration. The differentiation of J with respect to the temporal momenta β is slightly more complex, since the spatial regularity term depends also on the temporal deformation. Anyway, the computation of the gradient is based on the application of Proposition 4.12, as shown in the following proposition.

Proposition 9.2 (Differentiation of the 4D-registration criterion)**.** *Let $\{x_p\}$ be a discrete set of points and $(x_{p,t})$ a smooth trajectory of each of these points for $t \in [0,T]$. Let J be the function of the L^2 functions $\alpha_{p,j}(u)$ and $\beta_j(u)$:*

$$\begin{aligned} J(\alpha_{p,j}(u),\beta_j(u)) &= A\left(\{x_{p,t_j(1)}(1)\}_{p,j}\right) \\ &+ \gamma^\phi \int_0^1 \sum_{p,j,p',j'} \alpha_{p,j}(u)^t K^\phi(x_{p,t_j(1)}(u), x_{p',t_{j'}(1)}(u))\alpha_{p',j'}(u)du \\ &+ \gamma^\psi \int_0^1 \sum_{j,j'} \beta_j(u)^t K^\psi(t_j(u),t_{j'}(u))\beta_{j'}(u)du, \end{aligned} \tag{9.3.11}$$

for A a non-negative function and where for all $u \in [0,1]$:

$$\begin{aligned} t_j(u) &= t_j + \int_0^u \sum_k K^\psi\left(t_j(s),t_k(s)\right)\beta_k(s)ds \\ x_{p,t_j(1)}(u) &= x_{p,t_j(1)} + \int_0^u \sum_{p',j'} K^\phi\left(x_{p,t_j(1)}(s), x_{p',t_{j'}(1)}(s)\right)\alpha_{p',j'}(s)ds \end{aligned} \tag{9.3.12}$$

Let us assume that the kernels K^ϕ and K^ψ are scalar and symmetric (i.e. of the form $k(x,y)\mathrm{I}_3$ with the scalar function $k(x,y) = k(y,x)$). Let $\Lambda_u^\phi = \{x_{p,t_j(1)}(u)\}_{p,j}$ and $\Lambda_u^\psi = \{t_j(u)\}_j$. With the notations of Definition 2.5, let $u \to \mathbf{K}_{\Lambda_u^\phi} \otimes \mathbf{K}_{\Lambda_u^\psi}$ be the metric on the space of L^2-functions $(\alpha_{p,j}(u),\beta_j(u))$. Then, the gradient of J with respect to this metric is given by:

$$\nabla_{\alpha_{p,j}(u)} J(u) = 2\gamma^\phi \alpha_{p,j}(u) + \eta^\phi_{p,j}(u) \tag{9.3.13}$$

where $\eta^\phi_{p,j}(u)$ is the solution of the integral equation:

$$\begin{aligned} \eta^\phi_{p,j}(u) = \nabla_{x_{p,t_j(1)}(1)} A + \int_u^1 \sum_{p',t'_j} \Big(&\alpha_{p,j}(s)^t \eta^\phi_{p',j'}(s) + \\ &\alpha_{p',j'}(s)^t \eta^\phi_{p,j}(s) + 2\gamma^\phi \alpha_{p,j}(s)^t \alpha_{p',j'}(s)\Big) \nabla k^\phi\left(x_{p,t_j(1)}(s), x_{p',t'_j(1)}(s)\right) ds \end{aligned} \tag{9.3.14}$$

and by:

$$\nabla_{\beta_j(u)} J(u) = 2\gamma^\psi \beta_j(u) + \eta^\psi_j(u) \tag{9.3.15}$$

where $\eta^\psi_j(u)$ is the solution of the integral equation:

$$\begin{aligned} \eta^\psi_j(u) = \frac{\partial A}{\partial t_j(1)} + \gamma^\phi \frac{\partial}{\partial t_j(1)} \left(\int_0^1 \left\| v^\phi_u \right\|^2_{V^\phi} du \right) + \\ \int_u^1 \sum_{j'} \left(\beta_j(s)^t \eta^\psi_{j'}(s) + \beta_{j'}(s)^t \eta^\psi_j(s) + 2\gamma^\psi \beta_j(s)^t \beta_{j'}(s) \right) \nabla_1 k^\psi\left(t_j(s), t_{j'}(s)\right) ds \end{aligned} \tag{9.3.16}$$

where

$$\frac{\partial \int_0^1 \left\|v_u^\phi\right\|^2 du}{\partial t_j(1)} = 2\int_0^1 \sum_{p,q,i} \alpha_{p,i}(u)^t \alpha_{q,j}(u) \nabla_1 k^\phi \left(x_{p,t_i(1)}(u), x_{q,t_j(1)}(u)\right)^t \frac{\partial x_{q,t_j(1)}(u)}{\partial t_j(1)} du. \quad (9.3.17)$$

Proof. The differentiation of the gradient results from successive applications of Proposition 4.12.

Let us start with the differentiation with respect to the spatial momenta $\alpha_{p,j}(u)$. The third term of the criterion (i.e. $\int_0^1 \left\|v_u^\psi\right\|_{V^\psi}^2 du$) does not depend on the momenta $\alpha_{p,j}(u)$. Therefore, we can apply directly Proposition 4.12, which leads to Eq. (9.3.13) and (9.3.14).

The differentiation of J with respect to the temporal momenta β is slightly more complex, since the norm of the velocity field v_u^ϕ depends on the deformation ψ via the variables $t_j(1) = \psi_1(t_j)$ in the terms $x_{p,t_j(1)}$. However, the term $\mathrm{Reg}(\phi) = \int_0^1 \left\|v_u^\phi\right\|^2 du$ depends only on the final temporal diffeomorphism of the flow $\psi_1(t_j)$ and not on the intermediate temporal diffeomorphisms $\psi_u(t_j)$. Therefore, this term can be considered as part of the fidelity-to-data term in the differentiation of the criterion with respect to the temporal momenta. Let $\tilde{A}$ be the extended fidelity-to-data term:

$$\tilde{A}(\psi_1(t_1), \ldots, \psi_1(t_n)) = A\left(\{x_{p,t_j(1)}(1)\}_{p,j}\right) + \gamma^\phi \int_0^1 \left\|v_u^\phi\right\|_{V^\phi}^2 du$$

Now, we can apply Proposition 4.12 to the criterion $J(\beta_j(u)) = \tilde{A}(\{\psi_1(t_j)\}) + \gamma^\psi \int_0^1 \left\|v_u^\psi\right\|_{V^\psi}^2 du$. This leads to the gradient $\nabla_{\beta_j(u)} J(u) = 2\gamma^\psi \beta_j(u) + \eta_j^\psi(u)$ where the $\eta_j^\psi(u)$ satisfy the integral equation:

$$\eta_j^\psi(u) = \frac{\partial \tilde{A}}{\partial t_j(1)} + \int_u^1 \sum_{j'} \Big(\beta_j(s)^t \eta_{j'}^\psi(s) + \beta_{j'}(s)^t \eta_j^\psi(s) + 2\gamma^\psi \beta_j(s)^t \beta_{j'}(s)\Big) \nabla_1 k^\psi\left(t_j(s), t_{j'}(s)\right) ds. \quad (9.3.18)$$

The differentiation of the extended fidelity-to-data term $\frac{\partial \tilde{A}}{\partial t_j(1)}$ is decomposed into two terms. The differentiation of the true fidelity-to-data term A and the derivation of the regularity term of ϕ which leads to Eq. (9.3.17). ■

9.3.2.3 Differentiation of the fidelity to data term

In the differentiation with respect to the spatial momenta (Eq. (9.3.14)), we need to compute the gradient of the fidelity-to-data term A with respect to the final spatial positions $\phi_1(x_{p,t_j(1)})$ (i.e. the end point of the deformation of the points of $S(\psi(t_j))$): $\nabla_{x_{p,t_j(1)}(1)} A$. For the spatiotemporal registration of currents, the fidelity-to-data term is written as:

$$A = \sum_{t_j} \left\|\phi_* S(\psi(t_j)) - T_{t_j}\right\|_{W^*}^2,$$

where ϕ_* denotes the push-forward action of ϕ on the current $S(\psi(t))$, the mesh of the moving source S at time-point $\psi(t_j)$ and $\|.\|_{W^*}$ denotes the norm in the space of currents W^*. In this case, the differentiation is performed as explained in Section 4.3.1.2.

In the differentiation with respect to the temporal momenta (Eq. (9.3.16)), we need to differentiate the same fidelity-to-data term with respect to the final time-points $\psi(t_j)$. For each time-point t_j, we need to derive a function of the form:

$$\left.\frac{d\left\|\phi_* S(t) - T\right\|^2_{W^*}}{dt}\right|_{t=\psi(t_j)}.$$

The space of currents is a vector space provided with a norm, which comes from an inner-product. Therefore, one has:

$$\frac{d}{dt}\left\|\phi_* S(t) - T\right\|^2_{W^*} = \left\langle \phi_* S(t) - T, \frac{d}{dt}\phi_* S(t)\right\rangle_{W^*}.$$

Since the push-forward action is linear (see Section 1.3.3), one has: $\frac{d}{dt}\phi_* S(t) = \phi_* \frac{dS(t)}{dt}$. This requires to compute the derivative of the moving mesh $S(t)$ in the linear space of currents. If we write the mesh as a linear combination of Dirac delta currents $S(t) = \sum_p \delta^{\zeta_{p,t}}_{x_{p,t}}$, then the differentiation would involve the derivative of the Dirac delta currents, which would make the estimation of the gradient particularly difficult from a computational point of view. To overcome this issue, we take into account the fact that the input of the algorithm is a fine temporal sampling of the moving source $S(t)$ (at least, the regression function χ_t allows us to compute this moving shape at any arbitrary time-point t). Therefore, we can estimate the derivative $\frac{d\phi_* S(t)}{dt}$ easily with a centered Euler scheme, which leads eventually to:

$$\frac{d}{dt}\left\|\phi_* S(t) - T\right\|^2_{W^*} \sim \frac{1}{\delta t}\left\langle \phi_* S(t) - T, \phi_* S(t+\delta t) - \phi_* S(t-\delta t)\right\rangle_{W^*}. \tag{9.3.19}$$

This estimation involves the computation of the difference of two inner-products between currents which have been already computed and stored.

By contrast, there is no simple way to estimate the derivatives of $\partial \int_0^1 \left\|v^\phi_u\right\|^2_{V^\phi} du/\partial\psi(t_j)$ in Eq. (9.3.16) with a Euler scheme. The explicit differentiation leads to Eq. (9.3.17). In this equation, we still need to compute: $\partial x_{p,t_j(1)}(u)/\partial t_j(1)$, where $x_{p,t_j(1)}(u) = \phi_u(x_{p,t_j(1)})$. If we stored the regression function χ_t (at least its parameterizing momenta) which generated the continuous evolution $S(t)$ from the source data ($S(t) = \chi_t(S_0)$), then we can compute this term explicitly as:

$$\frac{\partial x_{p,t_j(1)}(u)}{\partial t_j(1)} = d_{x_{p,t_j(1)}}\phi_u\left(v^\chi_{t_j(1)}\right).$$

However, if do not want to compute the Jacobian matrix of ϕ_u, we can approximate this term by a centered Euler scheme (once again a fine temporal sampling of the positions $x_{p,t}$ is stored during the registration):

$$\frac{\partial x_{p,t_j(1)}(u)}{\partial t_j(1)} \sim \frac{x_{p,t_j(1)+\delta t} - x_{p,t_j(1)-\delta t}}{2\delta t} \tag{9.3.20}$$

Remark 9.3. Note that since the growth model χ_t is piecewise geodesic, the evolution $S(t)$ generated by χ_t is not differentiable at points t_j: the continuous $S(t)$ may have different left and right derivatives. This point will be discussed in Section 9.5.1. We will provide another optimization scheme that allows to take into account the fact the growth model is differentiable only in some directions. □

Remark 9.4. Note that we minimize J with respect to the geometrical and the temporal parameters *jointly*. We do not performed alternated minimization. □

9.3.2.4 A suboptimal approach

The gradient of the criterion J involves the differentiation of $\mathrm{Reg}(\phi)$ with respect to the temporal momenta β as computed in Eq. (9.3.17). This term is a coupling between temporal and geometrical parameters and can be seen as a correction of the derivative of the fidelity-to-data term. However, this correction may not be worth either the computational cost or the accumulation of numerical errors implied by the integration of this equation. We propose therefore a suboptimal approach, which consists in a different parameterization of the vector field v_u^ϕ. In Eq. (9.3.6), the momenta $\alpha_{p,j}(u)$ which parameterize v_u^ϕ are located at the positions $x_{p,\psi_1(t_j)}(u)$. One may look instead for a vector field whose momenta are located at the positions $x_{p,t_j}(u)$, namely by replacing $t_j(1) = \psi(t_j)$ by $t_j(0) = t_j$. In this case, Eq. (9.3.6) becomes:

$$v_u^\phi(x) = \sum_{p,t_j} K^\phi(x, x_{p,t_j}(u))\alpha_{p,j}(u) \tag{9.3.21}$$

We perform the gradient descent on the same variables $\alpha_{p,j}(u)$. As a consequence, the equations (9.3.13) and (9.3.15) remain the same. In Eq. (9.3.14), $t_j(1)$ and $t'_j(1)$ must be replaced by t_j and t'_j respectively. In Eq. (9.3.16), the second term vanishes since the regularity term $\mathrm{Reg}(\phi)$ does not depend on the temporal parameters β anymore. Therefore, the Eq. (9.3.17) does not need to be implemented.

This solution is not optimal. However, we believe (without any theoretical justification) that the change of the location of the momenta $\psi(t_j) \to t_j$ can be balanced by the magnitude of the momenta, so that this solution is not so far from the optimal one. At least, the numerical experiments in the next sections tend to prove the relevance of this approximation. Moreover, some recent (and yet unpublished) development suggests that a simpiler differentiation of the criterion would be possible, which avoids the need for approximation schemes.

9.3.3 Spatiotemporal atlas construction

9.3.3.1 Generative model and MAP estimation

In this section, we use the spatiotemporal registration to define and estimate a generative statistical model for longitudinal data. We suppose that N subjects $(S^i)_{i=1,\dots,N}$ have been scanned several times at time-points $(t_j^i)_j$. The segmentation of these scans leads to shapes S_j^i for the jth scan of the ith subject. We define a time-interval of interest of the form $[0, T]$ which contains every time-points t_j^i.

In the subject-specific approach, we assume the successive scans of each subject result from a temporal sampling of a random spatiotemporal deformation of a prototype scenario of evolution (called also a mean scenario of evolution). We assume that the prototype scenario of evolution has the form of a individual growth model of a prototype shape M_0, to be estimated. Formally, we assume that there is an evolution function χ_t for $t \in [0, T]$

such that $\chi_0 = \mathrm{Id}$, so that the mean scenario of evolution is written as: $M(t) = \chi_t(M_0)$ which satisfies $M(0) = M_0$.

For each subject i, we define a spatiotemporal deformation of the mean scenario as: $S^i(t) = \phi^i\left(M\left(\psi^i(t)\right)\right)$ for all $t \in [0, T]$. ϕ^i is the morphological deformation for subject i and ψ^i its time-change function. These two functions model how the anatomy of the subject and the dynamics of evolution can be derived from the prototype scenario of evolution. Eventually, we suppose that the observed shapes S^i_j for the subject i are samples from $S^i(t)$ at time point t^i_j, up to a random Gaussian noise:$S^i_j = S^i(t^i_j) + \varepsilon^i_j$. Formally, this leads to

$$S^i_j = \phi^i\left(\chi_{\psi^i(t^i_j)}(M_0)\right) + \varepsilon^i_j \tag{9.3.22}$$

for each subject i and every scan j of this subject.

The Gaussian variables ε^i_j are independent and identically distributed over the subject-index i and the time-index j. Actually, these random variables model physical and numerical noise of the image and the segmentation (see Section 5.2.2): they have no reason to be correlated even for successive scans of the same subject. Note that in the subject-specific setting, the spatiotemporal deformations (ϕ^i, ψ^i) depend on the subject only.

Our purpose is to estimate the parameters of this model from a longitudinal dataset. These parameters are the prototype shape M_0, the growth function χ_t and the set of spatiotemporal deformations ϕ^i, ψ^i for every subject i. These parameters are called a "spatiotemporal atlas". A Maximum A Posteriori Approach in the same setting as in Section 5.3.1 shows that the spatiotemporal atlas minimizes the following criterion:

$$\begin{aligned} &J\left((\psi^i)_{i=1,\dots,N}, (\phi^i)_{i=1,\dots,N}, \chi, M_0\right) = \\ &\quad \sum_{i=1}^{N}\left\{\sum_{t^i_j} \mathrm{d}(\phi^i(\chi_{\psi^i(t^i_j)}M_0), S^i(t^i_j))^2 + \gamma^\phi \mathrm{Reg}(\phi^i) + \gamma^\psi \mathrm{Reg}(\psi^i) + \gamma^\chi \mathrm{Reg}(\chi)\right\} \end{aligned} \tag{9.3.23}$$

Remark 9.5. Here, we combine a subject-specific approach with the forward setting defined in Chapter 5. Indeed, we deform the prototype scenario to the subjects and not the the subjects to the template. Therefore, every deformation share the same space of initial momenta in the reference frame of the template, a key feature for defining statistics on deformations, as we shall see in Section 9.3.3.5. □

9.3.3.2 Parameterization of deformations

We use the deformation framework of Chapter 4, to construct the diffeomorphisms χ, ϕ^i, ψ^i. As a consequence, every deformation satisfies a flow equation:

$$\begin{aligned} \frac{\partial \chi_t}{\partial t} &= v^\chi_t \circ \chi_t, \quad && t \in [0, T] \\ \frac{\partial \phi^i_u}{\partial u} &= v^{\phi^i}_u \circ \phi^i, \quad && u \in [0, 1], i = 1, \dots, N \\ \frac{\partial \psi^i_u}{\partial u} &= v^{\psi^i}_u \circ \psi^i, \quad && u \in [0, 1], i = 1, \dots, N \end{aligned} \tag{9.3.24}$$

We suppose that the prototype shape M_0 is given by a discrete set of points $\{x_p\}$: these points are one of the parameters of the atlas to be estimated. Here, we assume only that

M_0 has a finite number of points. Like in Chapter 5 for the template estimation in 3D, the initialization of M_0 as a mean mesh in the space of currents has the form of a finite set of Dirac delta currents. As we shall see below, the atlas estimation in Algorithm 5 will show that M_0 will keep this form along the iterations.

In this case, the application of Proposition 4.8 leads to the parameterization of the time-varying velocity fields by momenta as follows:

$$\begin{aligned} v_t^{\chi}(x) &= \sum_p K^{\chi}\left(x, x_p(t)\right) \alpha_p^{\chi}(t) \\ v_u^{\phi^i}(x) &= \sum_{p,t_j^i} K^{\phi}\left(x, x_{p,t_j^i}^i(u)\right) \alpha_{p,j}^i(u), \text{ for } i = 1, \ldots, N \\ v_u^{\psi^i}(t) &= \sum_{t_j^i} K^{\psi}\left(t, t_j^i(u)\right) \beta_j^i(u), \text{ for } i = 1, \ldots, N \end{aligned} \tag{9.3.25}$$

where K^{χ}, K^{ϕ} and K^{ψ} denotes 3 kernels of the RKHS V^{χ}, V^{ϕ} and V^{ψ} respectively and where we denote $\chi_t(x_p) = x_p(t) = x_{p,t}$, $\phi_u^i(x_{p,t}) = x_{p,t}^i(u)$ and $\psi_u^i(t) = t^i(u)$ for all $t \in [0,T]$ and $u \in [0,1]$. For the parameterization of the v^{ϕ^i}'s, we use the approximation of Section 9.3.2.4, namely by replacing $t_j^i(1)$ by t_j^i.

Now, we can written the criterion of the spatiotemporal atlas estimation as:

$$\begin{aligned} &J\left(\{\alpha_{p,j}^i(u)\}, \{\beta_j^i(u)\}, \{\alpha_p^{\chi}(t)\}, M_0\right) = \\ &\sum_{i=1}^N \left\{ \sum_{t_j^i} \mathrm{d}(\phi^i(\chi_{\psi^i(t_j^i)} M_0), S_j^i)^2 + \gamma^{\phi} \int_0^1 \left\| v_u^{\phi^i} \right\|_{V^{\phi}}^2 du + \gamma^{\psi} \int_0^1 \left\| v_u^{\psi^i} \right\|_{V^{\psi}}^2 du + \gamma^{\chi} \int_0^T \|v_t^{\chi}\|_{V^{\chi}}^2 dt \right\} \end{aligned} \tag{9.3.26}$$

When shapes are embedded in a space of currents W^*, then the fidelity-to-data term is written as:

$$A\left(\{x_p\}\right) = \sum_{i=1}^N \sum_{t_j^i} \left\| {\phi^i}_* \left({\chi_{\psi^i(t_j^i)}}_* M_0 \right) - S_j^i \right\|_{W^*}^2$$

9.3.3.3 An alternated minimization procedure

Since we based our 4D statistical models on similar assumptions as the 3D statistical model of Chapter 5, its estimation relies partly on the same algorithms. In particular, we will take advantage of the template-to-subject registrations and the algorithm to center the template according to deformations (Algorithm 4).

To minimize the criterion for spatiotemporal atlas estimation, we adopt a 3-steps alternated minimization procedure:

- If the template M_0 and the growth function χ are fixed, the criterion is divided into N independent functions to be minimized: each function is minimized when the deformations (ϕ^i, ψ^i) are N spatiotemporal registrations of the mean scenario $\chi_t(M_0)$ to the set of data S_j^i for each subject i. These spatiotemporal registrations are performed as in Section 9.3.2 (with the approximation of Section 9.3.2.4).

- If we fix the N spatiotemporal deformations (ϕ^i, ψ^i) and the growth function χ, the criterion to be minimized with respect to M_0 is reduced to:

$$J(M_0) = \sum_{i,j} \mathrm{d}(\Phi_{i,j}(M_0), S_j^i)^2,$$

  where we denote $\Phi_{i,j} = \phi^i \circ \chi_{\psi^i(t_j^i)}$. These deformations are $3D$-diffeomorphisms. This criterion is exactly the criterion in Eq. (5.3.19) for the template centering in the forward setting of Chapter 5. Therefore, the update of the prototype shape M_0 can be performed as in Algorithm 4, when the deformations are given by $\Phi_{i,j}$.

- If the template M_0 and the N spatiotemporal deformations (ϕ^i, ψ^i) are fixed, the criterion to be minimized becomes:

$$\sum_{i,j} \mathrm{d}(\phi^i(\chi_{\psi^i(t_j^i)} M_0), S^i(t_j^i))^2 + \gamma^\chi \mathrm{Reg}(\chi).$$

  This is not exactly the regression problem stated in Section 9.3.1 and 4.3.2 because of the deformation ϕ^i in the matching term. To turn it into a regression problem, we approximate the matching term $\mathrm{d}(\phi^i(\chi_{\psi^i(t_j^i)} M_0), S^i(t_j^i))$ by $\mathrm{d}(\chi_{\psi^i(t_j^i)}(M_0), (\phi^i)^{-1}(S^i(t_j^i)))$ (subject's shapes are matched back to the mean anatomy). This approximation is valid only for diffeomorphisms ϕ^i whose Jacobian is close to the identity, since the metric on currents d is not left-invariant. As a result, the evolution function χ_t performs the temporal regression of the set of shapes $(\phi^i)^{-1}(S_j^i)$ located at time-points $\psi^i(t_j^i)$. This regression problem has been solved in Section 4.3.2. Further investigations are needed in order to perform this regression in the forward setting (i.e. without this approximation), to be fully consistent with the previous step of the minimization procedure.

To initialize the minimization, we set M_0 as the mean current of the earliest data ((S_1^i) for every subject i) and set the diffeomorphisms χ, ϕ^i, ψ^i to identity. Normally, if we set all diffeomorphisms to the identity, the prototype M_0 which minimize the criterion is equal to the mean of every shape in the database. However, since M_0 is the initial shape of the mean scenario $M(t)$ (i.e. $M(0) = M_0$), we believe that the mean of the shapes associated to the earliest time-points is closer to the minimum of the criterion than the average of all shapes.

The whole minimization procedure is summarized in Algorithm 5.

9.3.3.4 Parameters

This statistical estimation depends mostly on eight parameters. There are 3 kernels of 3 distinct RKHS: K^χ, K^ϕ and K^ψ. We use Gaussian kernels determined by their standard deviations: λ_χ, λ_ϕ and λ_ψ respectively. They determine the degree of smoothness (i.e. the scale at which points have a correlated speed) of the mean scenario of evolution, the morphological deformations and the time-change functions. The first one compares with the scale of the geometrical variations of the structure over time for a typical subject (scale of the intra-subject variability). The second one compares with the scale of geometrical

Algorithm 5 Spatiotemporal Atlas Construction

Input: A set of time-indexed shapes $\{(S_j^i, t_j^i)\}_{j=1,\dots,n^i}^{i=1,\dots,N}$, where S_j^i is the jth scan (out of n^i) of the ith subject (out of N) at age t_j^i.

$M_0 \leftarrow \frac{1}{N}\sum_{i=1}^{N} S_1^i$

$M_0 \leftarrow$ sparse approximation of M_0 using the Matching Pursuit (Algorithm 2)

$M(t) \leftarrow \chi_t(M_0)$ the regression of every shapes S_j^i at time points t_j^i

repeat

  {Template-to-subject registration}

  for $i = 1 \dots N$ **do**

    $(\phi^i, \psi^i) \leftarrow$ spatiotemporal registration of $M(t)$ to S_j^i for $j = 1, \dots, n^i$

  end for

  {Center the template}

  $\Phi_{i,j} \leftarrow \phi^i \circ \chi_{\psi^i(t_j^i)}$ for all $i = 1, \dots, N$ and $j = 1, \dots, n^i$

  $M_0 \leftarrow \text{CenterTemplate}(M_0, \{\Phi_{i,j}\}, \{S_j^i\})$ (Algorithm 4)

  {Update the mean scenario}

  $M(t) \leftarrow \chi_t(M_0)$ the regression of every shapes $(\phi^i)^{-1}(S_j^i)$ at time points $\psi^i(t_j^i)$

until convergence

Output: One mean scenario $M(t) = \chi_t(M_0)$, N spatiotemporal deformations (ϕ^i, ψ^i)

variations between different subjects (geometrical inter-subject variability). The third one compares with the typical time-scale at which the dynamics of evolution changes from subjects to subjects.

The user must also set the 3 trade-offs between regularity and fidelity to data: $\gamma^\chi, \gamma^\phi, \gamma^\psi$. In addition, one needs to set the metric d between shapes. In the frameworks of currents, this metric depends on a kernel K^W. We choose a Gaussian kernel with standard deviation λ_W. This parameter set the scale under which the shape variations are considered as noise.

Eventually, one needs also to set the length of the time-interval of interest $[0, T]$. See Section 9.5.3 for a discussion on the impact of this parameter on the atlas estimation.

The dimension of the trade-off γ^χ and γ^ϕ have the same dimension as the trade-off γ in Remark 4.16 in the regression and registration case respectively. The dimension of γ^ψ is the one of the ratio between the fidelity-to-data term and the regularity term $\text{Reg}(\psi)$. This last term is of dimension T^2 (squared speed in the time-domain for a parameter interval of unit size). Therefore, γ^ψ is of dimension L^2T^{-2} for curves and L^4T^{-2} for surfaces.

9.3.3.5 Measures of the spatiotemporal variability

The construction of the spatiotemporal atlas leads to a mean scenario $M(t)$ and a spatiotemporal deformation of this mean scenario to each subject. These spatiotemporal deformations capture the spatiotemporal variability of the mean scenario in the population. This variability is decomposed into two parts: the geometrical variability which describes how the anatomy of the subjects is varying independently of the age and the temporal

variability which describes the variations of the dynamics of the evolution between different subjects. The former has been captured by the $3D$-diffeomorphisms ϕ^i, the later by the $1D$-diffeomorphisms ψ^i.

Since we perform template-to-subjects registration, we can find a common parameterization of the spatiotemporal deformations in the reference frame of the template like in Chapter 5. The diffeomorphisms ϕ^i are the final diffeomorphism of a flow parameterized by momenta $\alpha^i_{p,j}(u)$ at points $x_{p,t^i_j}(u)$. Since the flow is geodesic, the final diffeomorphism is entirely determined by the initial momenta: $\alpha^i_{p,j}(0)$ at points x_{p,t^i_j}. The entire mean scenario is described in the same reference frame, therefore all the Dirac delta currents $\delta^{\alpha^i_{p,j}(0)}_{x_{p,t^i_j}}$ belong to the same RKHS V^ϕ. Therefore, we can perform statistics on the set of N currents:

$$A_i = \sum_{j=1} \delta^{\alpha^i_{p,j}(0)}_{x_{p,t^i_j}},$$

in V^ϕ, as explained in Section 5.4.1. This leads to mean current $\overline{A}$ and principal modes of variations $\tilde{A}$. The geodesic shooting of these initial momenta at parameter $u = 1$ gives the mean and principal modes of the geometrical deformation of the mean scenario. These deformations give a description of the geometrical variability of the mean scenario in the population.

Similarly, the time warps ψ^i ($1D$-diffeomorphisms) are final diffeomorphisms of geodesic flows. They are parametrized by their initial momenta $\beta^i_j(0)$ located at time-points t^i_j. The Dirac delta currents $\delta^{\beta^i_j(0)}_{t^i_j}$ all belong to the same RKHS V^ψ. Therefore, the statistics are performed on the set of N currents:

$$B_i = \sum_{j=1} \delta^{\beta^i_j(0)}_{t^i_j}.$$

The geodesic shooting of the mean current and the principal modes of these currents leads to the mean time-change function and first modes of variations of the time-change functions. These statistics are still $1D$-diffeomorphisms and, as such, are monotonic functions. We notice that this differs from computing the point-by-point mean of the real-values functions $\psi^i(t)$. In other words, we compute the intrinsic statistics on the manifold of 1D-diffeomorphisms. Figure 9.8 shows such statistics on $1D$-diffeomorphisms.

9.4 Estimation of 4D statistics from anatomical curves and surfaces

9.4.1 Joint morphological deformation and time warp between 2D profiles of hominid skulls

In this section, we use the same database of 2D profiles of hominids skulls as in Figure 4.1, namely the segmentation of 2D-images of profiles of hominids[4]. As a result, we

[4]source: www.bordalierinstitute.com

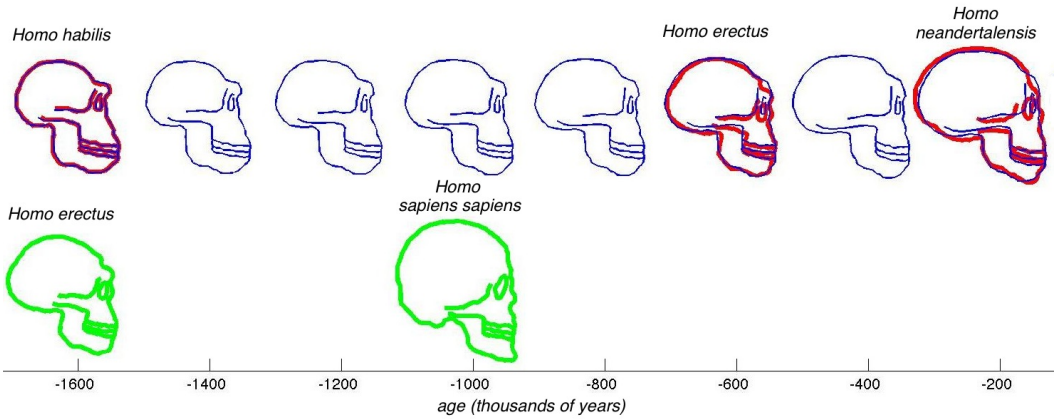


Figure 9.4: Illustrative pairwise registration. We divide the database in two to compare the evolution {Homo habilis-erectus-neandertalensis} (red shapes) to the evolution {Homo erectus-sapiens sapiens} (green shapes). To compare the relative morphological changes and speed of evolution between the earlier and the later evolution, we enforce them (by a temporal translation) to start at the same time. Then, we perform a shape regression of the source shapes (earlier evolution) to end up with a continuous growth model of the source. Now, we run the spatiotemporal registration of the blue source to the green target. Results are shown in Fig. 9.5 and 9.6.

have five profiles which consist of six lines each. Each profile correspond to an hominid (australopithecus, homo habilis, homo erectus, homo neandertalensis and homo sapiens sapiens) and is associated to an age (in millions of years). The *regression* of the 5 profiles has been performed in Fig. 4.1. Our purpose is to use this database to give an illustrative example of the spatiotemporal pairwise *registration*.

We want to compare the evolution {Homo habilis-Homo erectus-Homo neandertalensis} (called earlier evolution) with the evolution {Homo erectus-Homo sapiens sapiens} (called later evolution). In particular, we would like to measure the anatomical differences of the skull between these two periods and the change of the dynamics of evolution between the earlier and the later evolution. Therefore, we artificially divide the database into two groups, considered as two different subjects: homo habilis, homo erectus and homo neandertalensis as the source subject and homo erectus, homo sapiens sapiens as the target subject (see Fig. 9.4). In order to compare both evolution, we translate the target back to 4 millions of years, so that both evolution start at the same time (this can be seen as a "rigid" temporal alignment as a pre-processing).

The first step of the registration consists in performing the regression of the source data. This leads to a continuous source evolution $S(t)$ shown in blue in the first row of Fig. 9.4. Then, we compute the spatiotemporal deformation of this continuous source evolution so that the deformed scenario matches the target data. The result is a morphological deformation ϕ and a time warp ψ. The deformation of the source evolution along these two deformations is shown in Figure 9.5.

The morphological deformation shows that the jaw was less prominent and the skull

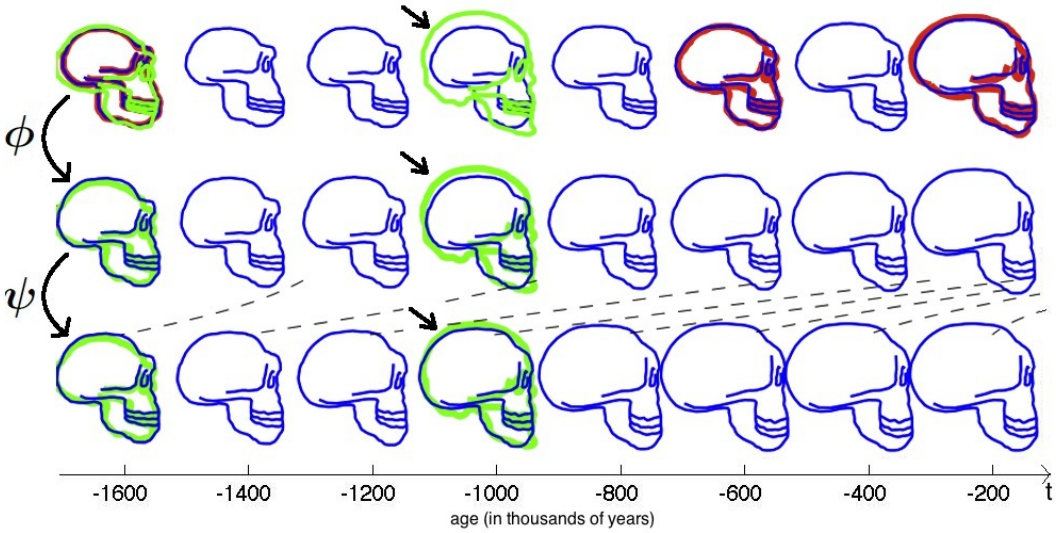
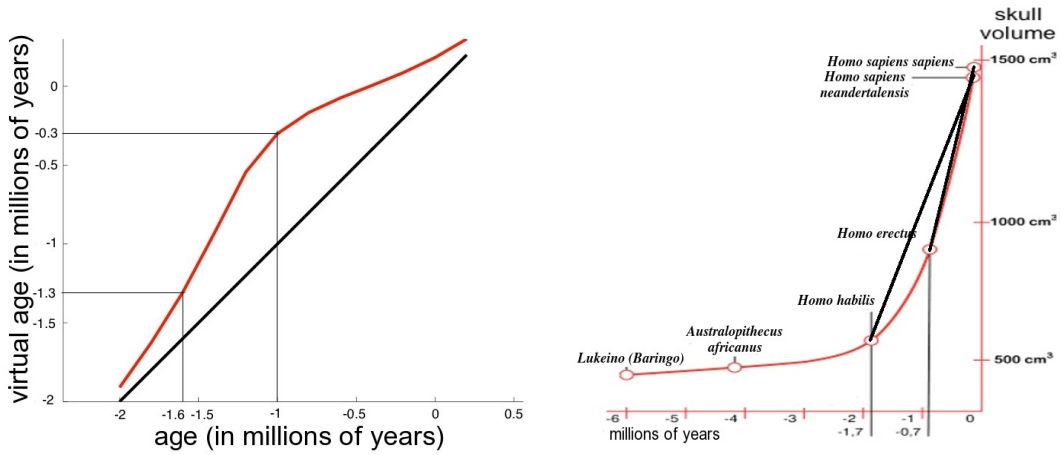


Figure 9.5: **Top row:** The input data as prepared in Fig. 9.4: the continuous source evolution (blue) is superimposed with the target (green) **Middle row:** The geometrical part ϕ of the spatiotemporal registration is applied to each blue frame. This shows morphological changes between source and target independently of the time: the skull is larger, rounder and the jaw less prominent during the later evolution compared to the earlier evolution. (time frame corresponding to target shape is made bold for a better comparison) **Bottom row:** The time warp ψ is applied to the evolution of the second row. The blue shapes are moved along the time axis (as shown by dashed black lines), but they are not deformed. This time re-alignment shows an acceleration of the evolution during the later evolution compared to the earlier evolution. Black arrows show that a better alignment is achieved when one accounts both for morphological changes *and* a change of the evolution speed. Note that here, the morphological part and temporal part are shown separately for clarity purposes although they are computed simultaneously (as minimizers of a combined cost function).

larger and rounder during the later evolution than during the earlier evolution. The time warp is plotted in Fig. 9.6a. The graph of Fig. 9.6 shows an almost linear increase in speed between the earlier and later evolution. The slope of the curve is of 1.66 during this period, thus meaning that the later evolution evolves at a speed 1.66 times faster than the earlier evolution. This value is compatible with the growth speed of the skull during these periods (See Fig. 9.6b): between Homo erectus and Homo sapiens sapiens the skull volume growths at a rate of $(1500-900)/0.7 = 860\text{cm}^3$ per millions of years, whereas between Homo habilis and Homo neandertalensis, it growths at $(1500-600)/1.7 = 530\text{cm}^3$ per millions of years, namely 1.62 times faster.



a- time warp b- skull volume evolution

Figure 9.6: a- time warp $\psi(t)$ of the registration in Fig. 9.5 (in black the reference $\psi(t) = t$). The slope greater than 1 between the two target data shows an acceleration of the evolution of the target compared to the evolution of the source. Note that the speed reduction after -1 millions of years (after which there is no more target data) is due to the boundary conditions: $\psi(t) = t$ at $t \to \pm\infty$. b- the graph of the skull volume over the human evolution as found in the literature (source: www.bordalierinstitute.com). This curve shows that the increase in skull volume between Homo erectus and Homo sapiens sapiens was 1.62 times faster than between Homo habilis and Homo neandertalensis (ratio between the slope of the two straight lines). This value is compatible with the acceleration measured by the time warp in a: the slope of the curve between target data is of 1.66.

9.4.2 Measure of developmental delays in autism from surfaces of deep brain structures

In this section, we are interested in the analysis of a longitudinal database of deep brain structures segmented from images of autistics, developmental delays and controls children [Hazlett 2005]. Each child has been scanned twice: a baseline at about age 3 years and a follow-up at about age 5 years (the distribution of the ages is shown in Fig. 9.9). In this application, we focus on the amygdala of the right hemisphere, since the volume of this structure is most variable among the other structures (caudate, putamen, hippocampus and corpus pallidus). The longitudinal database consists therefore of 24 meshes of amygdala: 12 subjects divided on 3 groups of 4 subjects (autistics, developmental delays and controls), each subject having two meshes (a baseline and a follow-up). As a pre-processing, we align all the meshes together via rigid transformations.

From these 12×2 meshes, we estimate a spatiotemporal atlas: a prototype shape, a mean scenario of evolution of this prototype shape and 12 spatiotemporal deformations of this mean scenario to each subject. We set the time-interval of interest to $[0.5, 7.1]$ years with a time-step of 0.2 years. The standard deviation of the Gaussian kernels were set to

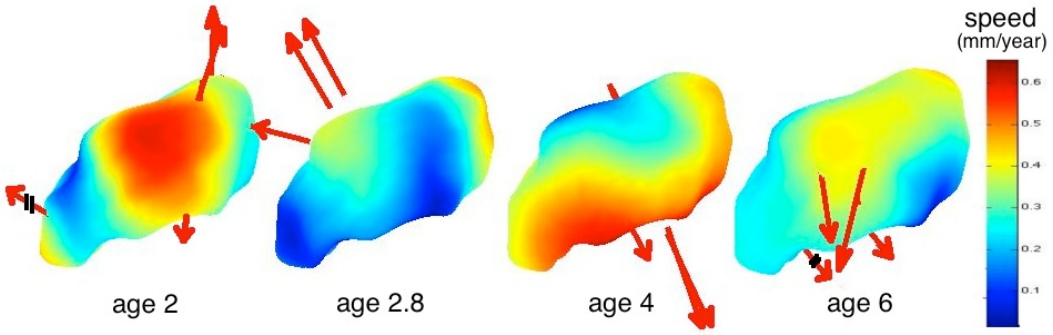


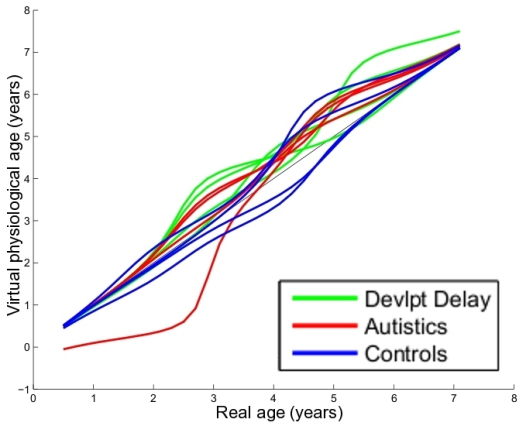
Figure 9.7: Mean Scenario of the right Amygdala (right lateral part). Arrows show the first momenta of the difference between the mesh at age $t + 0.2$ and age t approximated by the matching pursuit (as in Fig. 3.8). From age 2 to 2.8, the evolution is mainly a torque at the posterior part; then the structure becomes thicker, mostly at the superior part between age 2.8 and 4 and at the inferior between age 4 and 6; from age 6 the evolution is a mainly a torque at the anterior part. (Note: a representation of the template as a mesh has been computed by matching a subject's mesh back to the template.)

$\lambda_\chi = 15$ mm for the regression function, $\lambda_\phi = 15$ mm for the morphological deformation and $\lambda_\psi = 1$ year for the time warp. The typical scale on currents λ_W is set to 3 mm. The trade-offs were set to $\gamma^\chi = 10^{-3}$ mm^{2}year, $\gamma^\phi = 10^{-3}$ mm^2 and $\gamma^\psi = 10^{-6}$ mm^{4}year^{-2}. An amygdala is typically 10 mm large.

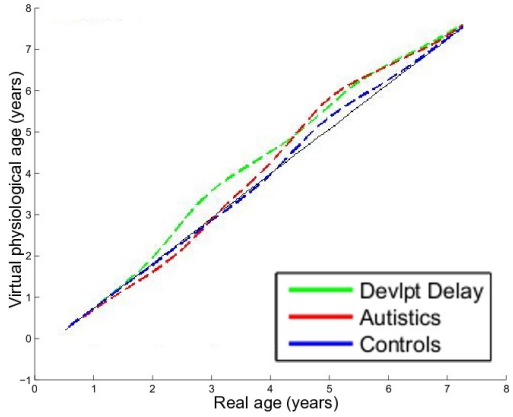
Significant samples of the estimated mean scenario of evolution are shown in Fig. 9.7. This mean scenario of evolution shows that the prototype growth of the structure is much more complex than a pure volume scaling over time. Indeed, the visual inspection of the movie of this scenario shows mainly 4 phases during growth in the time-interval of interest: first, we observe a torque at the posterior part of the structure, then the superior part of the structure is expanding before an expansion of the anterior part. Eventually, the growth is mostly a torque at the anterior part of the structure.

Now, we analyze the spatiotemporal variability of the mean scenario in the population. This variability is decomposed into a geometrical part captured by the morphological deformations ϕ^i and a temporal part captured by the time warp ψ^i. Preliminary tests do not show any correlations between the morphological deformations and the class of the subject (autistics, developmental delays and controls). The mean initial momenta of the morphological deformations of each group do not differ significantly from zero.

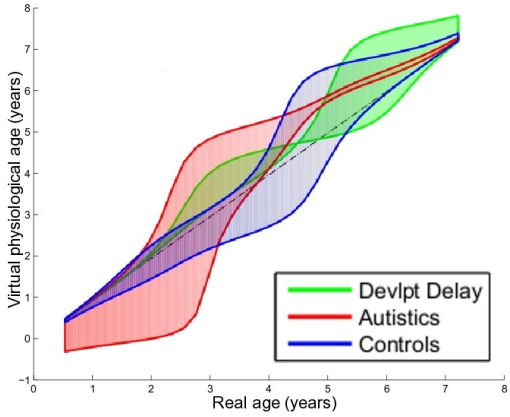
The time warps are plotted in Fig. 9.8 for every subjects. When the curve is above the $y = x$ axis (resp. below the $y = x$ axis), the evolution of the subject is in advance (resp. is delayed) with respect to the mean scenario. A slope greater than 1 (resp. smaller than 1) denotes an acceleration (resp. a speed reduction) of the evolution of the subject compared to the evolution of the mean scenario. From theses curves, we cannot conclude that an autistic or a developmental delay patient is systematically delayed or in advance compared to controls, even at a given age. However, the curves seem all to have the same pattern, namely an acceleration of the evolution with respect to the mean scenario. This visual



(a) - time warps for the 12 subjects



(b) mean time warp



(c) first mode of variations

Figure 9.8: **Temporal deformation of the mean scenario** (a) time warps for the 12 subjects, (b) averaged time warp in each class, (c) first mode of variation at $\pm\sigma$ of the time warps for each class. Autistics and controls show the same evolution pattern, namely a reduction of speed with respect to the mean scenario (slope smaller than 1) and then a quick acceleration (slope greater than 1). This pattern for the autistics group seems to occur later than for the control group. The developmental delays presents also such pattern but at an arbitrary age. Note: mean and standard deviation are computed within the space of 1D diffeomorphism.

pattern is confirmed by the mean time warps of each class and even more clearly by the first mode of variation of the curves plotted in Fig. 9.8b,c: it shows that the acceleration occurs between age 2.5 and 3.5 years for the autistics and between age 4 and age 5 years for the controls. The developmental delays also have such a pattern but it occurs at a very variable age.

These preliminary results on both the geometrical part and the temporal part of the variability suggest that the discriminative information between classes might not be inferred from the anatomical variability at a given age, but rather from variations of the growth process. Of course, these results have to be strengthened by more in-depth statistical

analysis and using much larger database, since this analysis may be biased by some outliers. In a general manner, the more time-points per subjects, the more constrained the mean scenario estimation. The more subjects, the more robust the statistics.

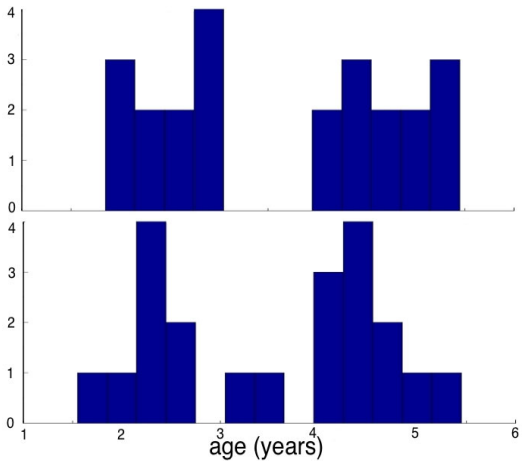


Figure 9.9: Top: the histogram of ages in the database. Bottom: histogram of the ages matched back in the reference frame of the mean scenario (via the maps ϕ_i^{-1}). This shows that the distribution of ages in the mean scenario is more picked, suggesting that the model is mostly determine by the anatomy at age 2.5 and age 4.5. Two data have been shifted at the center of the distribution, showing a temporal shift with respect to the mean scenario. The original histogram shows also that this database would not be adapted for a time-specific analysis: considering that every baseline have been acquired at age 2 and every follow-up at age 4 would hide the developmental delays between subjects.

9.5 Discussion

This chapter defines a consistent statistical framework for the analysis of longitudinal data. This work is still in its early developments and raises several questions. In this section, we discuss the most obvious theoretical issues.

9.5.1 Dealing with the discontinuities of the regression function

As mentioned in Remark 9.3, the evolution $S(t)$ generated by the regression function χ_t have different left and right derivatives at the target time-points t_j. This problem raises in Eq. (9.3.16), where we need to compute the gradient of the registration with respect to the temporal parameters. For registration as in Section 9.3.2, this is an issue mainly during the first step of the gradient descent, since then $\psi = \mathrm{Id}$. Otherwise, one needs to derive $S(t)$ only at time-points $\psi(t_j)$ which have little chance to be one of the $\{t_j\}$'s. For atlas estimation as in Section 9.3.3, however, the problem is more important: at the convergence, the mean scenario is derived precisely at the point where it is discontinuous. The bad estimation of the derivatives may cause the gradient to vanish although we are not at a critical point of the criterion.

Indeed, we propose so far to estimate the derivatives with a centered Euler scheme. This implicitly smoothes the vector field v_u^χ by averaging the left and right values at points t_j. However, there is no guarantee that the criterion may not decrease when using only the left

(or only the right) derivative. In this section, we propose an alternative optimization scheme which first takes into account the fact that the left and right derivatives of the fidelity-to-data term in Eq. (9.3.16) may be different by computing only directional derivatives, second performs an exhaustive search of all the possible combinations of left and right derivatives at points t_j and, third, stops only at critical point of the criterion.

We focus only on the differentiation of the registration criterion with respect to the temporal momenta $\beta_j(u)$, since the non-differentiability of the source does not affect the geometrical momenta (see Proposition 9.2 and Eq. (9.3.16)). In this section, we use the same matrix notations as in the proof of Proposition 4.14. Let N be the number of target time-points t_j. Then we denote $\boldsymbol{\beta}(u)$ the N-dimensional vector whose coordinates are $\beta_j(u)$. Let $\delta\boldsymbol{\beta}$ be a perturbation of the criterion $J(\boldsymbol{\beta})$ (defined in Eq. (9.3.11)). Then, this perturbation induces a variation δt_j of the final time-points $t_j(1) = \psi_1(t_j)$. We denote $\boldsymbol{\varepsilon}_j(\delta\boldsymbol{\beta}) = \text{sign}(\delta t_j(1))$, the sign of this perturbation: $\boldsymbol{\varepsilon}_j > 0$ means that the variation of the momenta tends to increase the deformed time-point $t_j(1)$. The rigorous differentiation of J in the direction of $\delta\boldsymbol{\beta}$ (directional gradient) can be done along the lines of the proof of Proposition 4.14. This gives:

$$\frac{dJ(\boldsymbol{\beta}+\xi\delta\boldsymbol{\beta})}{d\xi} = \int_0^1 \left(\nabla^{\boldsymbol{\varepsilon}(\delta\boldsymbol{\beta})}_{\boldsymbol{\beta}(u)} J(u)\right)^t \mathbf{k}(\mathbf{t}_u,\mathbf{t}_u)\delta\boldsymbol{\beta}(u)du = \left\langle \nabla^{\boldsymbol{\varepsilon}(\delta\boldsymbol{\beta})}_{\boldsymbol{\beta}(u)} J(u), \mathbf{K}_{\Lambda^\psi_u}\delta\boldsymbol{\beta}\right\rangle_{L^2}, \quad (9.5.1)$$

where $\mathbf{t}_u$ is the N-dimensional vector $(t_j(u))_j$ and $\nabla^{\boldsymbol{\varepsilon}(\delta\boldsymbol{\beta})}_{\boldsymbol{\beta}(u)} J(u) = 2\gamma^\psi\boldsymbol{\beta}(u) + \boldsymbol{\eta}^\psi_{\boldsymbol{\varepsilon}(\delta\boldsymbol{\beta})}(u)$ with the vector $\boldsymbol{\eta}^\psi_{\boldsymbol{\varepsilon}(\delta\boldsymbol{\beta})}$ satisfying the integral equation:

$$(\eta^\psi_{\boldsymbol{\varepsilon}(\delta\boldsymbol{\beta})})_j(u) = \frac{\partial A}{\partial t_j(1)}\bigg|_{\boldsymbol{\varepsilon}(\delta\boldsymbol{\beta})_j} + \frac{\partial}{\partial t_j(1)}\left(\int_0^1 \|v^\phi_u\|^2_{V^\phi}\,du\right)\bigg|_{\boldsymbol{\varepsilon}(\delta\boldsymbol{\beta})_j} +$$
$$\int_u^1 \sum_{j'} \left(\beta_j(s)^t(\eta^\psi_{\boldsymbol{\varepsilon}(\delta\boldsymbol{\beta})})_{j'}(s) + \beta_{j'}(s)^t(\eta^\psi_{\boldsymbol{\varepsilon}(\delta\boldsymbol{\beta})})_j(s) + 2\gamma^\psi\beta_j(s)^t\beta_{j'}(s)\right)\nabla_1 k^\psi\left(t_j(s),t_{j'}(s)\right)ds, \quad (9.5.2)$$

This equation is similar to Eq. (9.3.16), except that the partial derivatives (first and second term) are the left derivative if $\boldsymbol{\varepsilon}_j = -1$ and the right derivative if $\boldsymbol{\varepsilon}_j = +1$. This means that if the variation of the momenta $\delta\boldsymbol{\beta}$ is in a direction such that it tends to increase the deformed time-point $\psi(t_j)$, then the directional gradient requires to compute the right derivative and the left derivative in the opposite case.

Therefore, the direction $\delta\boldsymbol{\beta}$ which enables to achieve the maximum decay of the criterion satisfies:

$$\max_{\|\delta\boldsymbol{\beta}\|_{\mathbf{K}_{\Lambda^\psi_u}}=1} -\left\langle \nabla^{\boldsymbol{\varepsilon}(\delta\boldsymbol{\beta})}_{\boldsymbol{\beta}(u)} J(u), \delta\boldsymbol{\beta}\right\rangle_{\mathbf{K}_{\Lambda^\psi_u}}, \quad (9.5.3)$$

where for all N-dimensional vectors X, Y, we denote $\langle X,Y\rangle_{\mathbf{K}_{\Lambda^\psi_u}} = \int_0^1 X(u)^t\mathbf{K}_{\Lambda^\psi_u}Y(u)du$ the L^2-metric regularized by the kernel K^ψ.

This optimization problem can be turned into a constrained maximization problem:

$$\max_{\substack{\|\delta\boldsymbol{\beta}\|_{\mathbf{K}_{\Lambda^\psi_u}}=1 \\ \boldsymbol{\varepsilon}_j=\text{sign}(\delta t_j(1))}} -\left\langle \nabla^{\boldsymbol{\varepsilon}}_{\boldsymbol{\beta}(u)} J(u), \delta\boldsymbol{\beta}\right\rangle_{\mathbf{K}_{\Lambda^\psi_u}}. \quad (9.5.4)$$

Using the Lagrange multipliers $\boldsymbol{\lambda}$ (N-dimensional vector) and μ (real number) this maximization problem becomes:

$$\max_{\boldsymbol{\varepsilon}} \max_{\delta\boldsymbol{\beta}} \min_{\substack{\lambda_j \geq 0 \\ \mu \geq 0}} \left\{ - \left\langle \nabla^{\boldsymbol{\varepsilon}}_{\boldsymbol{\beta}(u)} J(u), \delta\boldsymbol{\beta} \right\rangle_{\mathbf{K}_{\Lambda^{\psi}_u}} + \boldsymbol{\lambda}^t \left(\boldsymbol{\varepsilon}.\delta\mathbf{t}_1\right) + \mu \left(1 - \|\delta\boldsymbol{\beta}\|^2_{\mathbf{K}_{\Lambda^{\psi}_u}}\right) \right\}, \tag{9.5.5}$$

where '.' denotes the element-wise multiplication of vectors and the constraint $\boldsymbol{\varepsilon}_j = \text{sign}(\delta t_j(1))$ written as $(\boldsymbol{\varepsilon}.\delta\mathbf{t}_1)_j \geq 0$

According to the proof of Proposition 4.14, the vector $\delta\mathbf{t}_1$ is given by:

$$\delta\mathbf{t}_1 = \int_0^1 R^1_u \mathbf{k}(\mathbf{t}_u, \mathbf{t}_u) \delta\boldsymbol{\beta}(u) du \tag{9.5.6}$$

where the varying N-by-N matrix R^1_u satisfies:

$$R^1_u = \mathrm{I} + \int_u^1 R^1_s \left((\delta_1 + \delta_2) \mathbf{k}(\mathbf{t}_s, \mathbf{t}_s) \boldsymbol{\beta}(s) \right) ds \tag{9.5.7}$$

as shown in Eq. (4.3.21). Therefore, $\boldsymbol{\lambda}^t \left(\boldsymbol{\varepsilon}.\delta\mathbf{t}_1\right) = \left(\boldsymbol{\varepsilon}.\boldsymbol{\lambda}\right)^t \delta\mathbf{t}_1$ satisfies:

$$\left(\boldsymbol{\varepsilon}.\boldsymbol{\lambda}\right)^t \delta\mathbf{t}_1 = \int_0^1 \left(\boldsymbol{\varepsilon}.\boldsymbol{\lambda}\right)^t R^1_u \mathbf{k}(\mathbf{t}_u, \mathbf{t}_u) \delta\boldsymbol{\beta}(u) du = \left\langle (R^1)^t (\boldsymbol{\varepsilon}.\boldsymbol{\lambda}), \delta\boldsymbol{\beta} \right\rangle_{\mathbf{K}_{\Lambda^{\psi}_u}}. \tag{9.5.8}$$

We define now the function $\boldsymbol{\lambda}^{\boldsymbol{\varepsilon}}(u) = (R^1_u)^t (\boldsymbol{\varepsilon}.\boldsymbol{\lambda})$. Using Eq. (9.5.7), this function satisfies the integral equation:

$$\boldsymbol{\lambda}^{\boldsymbol{\varepsilon}}(u) = \boldsymbol{\varepsilon}.\boldsymbol{\lambda} + \int_u^1 \left((\partial_1 + \partial_2) \mathbf{k}(\mathbf{t}_s, \mathbf{t}_s) \boldsymbol{\beta}(s) \right)^t \boldsymbol{\lambda}^{\boldsymbol{\varepsilon}}(s) ds. \tag{9.5.9}$$

We notice that $\boldsymbol{\lambda}^{\boldsymbol{\varepsilon}}(u)$ transport the final condition $\boldsymbol{\varepsilon}.\boldsymbol{\lambda}$ at $u = 1$ back to parameter $u = 0$ with the same differential equation as for $\boldsymbol{\eta}^{\psi}_u$.

Eventually, the criterion to be optimized is given as:

$$\max_{\boldsymbol{\varepsilon}} \max_{\delta\boldsymbol{\beta}} \min_{\substack{\lambda_j \geq 0 \\ \mu \geq 0}} \left\{ - \left\langle \nabla^{\boldsymbol{\varepsilon}}_{\boldsymbol{\beta}(u)} J(u), \delta\boldsymbol{\beta} \right\rangle_{\mathbf{K}_{\Lambda^{\psi}_u}} + \langle \boldsymbol{\lambda}^{\boldsymbol{\varepsilon}}, \delta\boldsymbol{\beta} \rangle_{\mathbf{K}_{\Lambda^{\psi}_u}} + \mu \left(1 - \|\delta\boldsymbol{\beta}\|^2_{\mathbf{K}_{\Lambda^{\psi}_u}}\right) \right\}, \tag{9.5.10}$$

As a common practice in constrained optimization problem ("minimax" problem), we could permute the minimum and maximum operator [Hiriart-Urruty 1996]. Provided that the conditions of such an inversion are satisfied, the criterion to be optimized becomes:

$$\max_{\boldsymbol{\varepsilon}} \min_{\substack{\lambda_j \geq 0 \\ \mu \geq 0}} \max_{\delta\boldsymbol{\beta}} \left\{ - \left\langle \nabla^{\boldsymbol{\varepsilon}}_{\boldsymbol{\beta}(u)} J(u), \delta\boldsymbol{\beta} \right\rangle_{\mathbf{K}_{\Lambda^{\psi}_u}} + \langle \boldsymbol{\lambda}^{\boldsymbol{\varepsilon}}, \delta\boldsymbol{\beta} \rangle_{\mathbf{K}_{\Lambda^{\psi}_u}} + \mu \left(1 - \|\delta\boldsymbol{\beta}\|^2_{\mathbf{K}_{\Lambda^{\psi}_u}}\right) \right\}, \tag{9.5.11}$$

The maximum over $\delta\boldsymbol{\beta}$ is achieved when $-\nabla^{\boldsymbol{\varepsilon}}_{\boldsymbol{\beta}(u)} J(u) + \boldsymbol{\lambda}^{\boldsymbol{\varepsilon}}(u) - 2\mu\delta\boldsymbol{\beta}(u) = 0$ which leads to:

$$\delta\boldsymbol{\beta} = \frac{1}{2\mu} \left(-\nabla^{\boldsymbol{\varepsilon}}_{\boldsymbol{\beta}} J + \boldsymbol{\lambda}^{\boldsymbol{\varepsilon}} \right). \tag{9.5.12}$$

Then, the criterion to be optimized becomes

$$\mu + \frac{1}{4\mu} \left\| -\nabla^{\boldsymbol{\varepsilon}}_{\boldsymbol{\beta}} J + \boldsymbol{\lambda}^{\boldsymbol{\varepsilon}} \right\|^2_{\mathbf{K}_{\Lambda^{\psi}_u}} \tag{9.5.13}$$

which is minimal for $\mu = \frac{1}{2}\left\|-\nabla^{\boldsymbol{\varepsilon}}_{\boldsymbol{\beta}} J + \boldsymbol{\lambda}^{\boldsymbol{\varepsilon}}\right\|_{\mathbf{K}_{\Lambda^{\psi}_u}}$. This gives $(\delta\boldsymbol{\beta}, \boldsymbol{\varepsilon})$ the steepest direction of the criterion:

$$\begin{cases} \boldsymbol{\varepsilon} = \arg\max_{\boldsymbol{\varepsilon}} \left\{ \min_{\lambda_j \geq 0} \|\mathbf{G}_{\boldsymbol{\beta},\boldsymbol{\varepsilon},\boldsymbol{\lambda}}\|_{\mathbf{K}_{\Lambda^{\psi}_u}} \right\} \\ \delta\boldsymbol{\beta}(u) = \dfrac{\mathbf{G}_{\boldsymbol{\beta},\boldsymbol{\varepsilon},\boldsymbol{\lambda}}(u)}{\|\mathbf{G}_{\boldsymbol{\beta},\boldsymbol{\varepsilon},\boldsymbol{\lambda}}\|_{\mathbf{K}_{\Lambda^{\psi}_u}}} \\ \mathbf{G}_{\boldsymbol{\beta},\boldsymbol{\varepsilon},\boldsymbol{\lambda}}(u) = -\nabla^{\boldsymbol{\varepsilon}}_{\boldsymbol{\beta}} J(1) + \boldsymbol{\lambda}.\boldsymbol{\varepsilon} + \displaystyle\int_u^1 \left((\partial_1 + \partial_2)\mathbf{k}(\mathbf{t}_s, \mathbf{t}_s)\boldsymbol{\beta}(s)\right)^t \mathbf{G}_{\boldsymbol{\beta},\boldsymbol{\varepsilon},\boldsymbol{\lambda}}(s) ds \end{cases} . \quad (9.5.14)$$

Assume that there is no integration and that we can deal each component of $\mathbf{G}$ independently, then for a fixed $\boldsymbol{\varepsilon}$, the minimum of $\left|-\nabla^{\boldsymbol{\varepsilon}}_{\boldsymbol{\beta}} J(1)_j + \boldsymbol{\lambda}_j.\boldsymbol{\varepsilon}_j\right|$ is equal to 0 if $\nabla^{\boldsymbol{\varepsilon}}_{\boldsymbol{\beta}} J(1)_j$ has the same sign as $\boldsymbol{\varepsilon}_j$ (achieved for $\boldsymbol{\lambda}_j = \left|-\nabla^{\boldsymbol{\varepsilon}}_{\boldsymbol{\beta}} J(1)_j\right|$) or the minimum is equal to $-\nabla^{\boldsymbol{\varepsilon}}_{\boldsymbol{\beta}} J(1)_j$ if it has the opposite sign of $\boldsymbol{\varepsilon}_j$ (minimum achieved for $\boldsymbol{\lambda}_j = 0$). This means that j$th$ component of the gradient vanishes, if the perturbation of the criterion in the direction of $\boldsymbol{\varepsilon}_j$ does not enables to decrease the criterion. As a consequence, if for every possible $\boldsymbol{\varepsilon}$ (i.e. for any of the 2^N combinations of left or right derivatives at each of the N time-points t_j) the gradient $\mathbf{G}$ vanishes, then there is no direction which can decrease the criterion: we are at a critical point of the criterion where every directional derivatives vanishes.

Here, the optimization is slightly more complex since the different components are mixed via the non-diagonal metric $\mathbf{K}_{\Lambda^{\psi}_u}$ and one has to integrate an ODE backward in time. We propose to follow this strategy: first, we focus on $|\mathbf{G}_{\boldsymbol{\beta},\boldsymbol{\varepsilon},\boldsymbol{\lambda}}(1)|$. If there is a choice of $\boldsymbol{\varepsilon}$, such that this quantity optimized over the $\boldsymbol{\lambda}$ (as explained in the previous paragraph) does not vanish, then this choice of $\boldsymbol{\varepsilon}$ gives a direction of descent. In this case, one integrates the ODE to give the whole $\mathbf{G}_{\boldsymbol{\beta},\boldsymbol{\varepsilon},\boldsymbol{\lambda}}(t)$ for $t \in [0,1]$. The optimization of the squared norm of this quantity over $\boldsymbol{\lambda}$ is a quadratic problem. Its solution gives the direction of descent $\delta\boldsymbol{\beta}$. If for all choice of $\boldsymbol{\varepsilon}$ the final condition $\mathbf{G}_{\boldsymbol{\beta},\boldsymbol{\varepsilon},\boldsymbol{\lambda}}(1)$ vanishes, then by linearity of the ODE, $\mathbf{G}_{\boldsymbol{\beta},\boldsymbol{\varepsilon},\boldsymbol{\lambda}}(t)$ vanishes over $[0,1]$. In this case, this means that there is no direction which enables to decrease the criterion: the algorithm stops at a critical point.

This strategy may be computationally expensive and this is the main reason why we do not implement it. From a theoretical point of view, we still need to prove that the permutation of the min and max is allowed, which should be feasible. At least, this discussion shows that a gradient descent algorithm must be possible even in presence of discontinuities of the criterion and in a way which guarantees that the algorithm stops only at critical points of the criterion.

9.5.2 Which time-parameterization for the mean scenario?

In pairwise registration, we consider the continuous shape evolution of the source (computed via regression) as an input: the spatiotemporal deformation does not change this scenario. By contrast, during the estimation of the atlas, one estimates the mean scenario (via regression) jointly with the spatiotemporal deformations of the mean scenario. In particular, the time-change functions can be seen as a time re-parameterization of the mean

scenario. It seems that a global (i.e. independent of the subject) re-parameterization of the mean scenario would lead to the same result. However, the following discussion shows that the parameterization of the scenario is not an extra degree of freedom of the modeling. On the contrary, the L^2 penalty term in the criterion favors a particular parameterization of the scenario.

The estimation of the atlas is performed by minimizing the criterion:

$$J(\psi^i, \chi) = \sum_{i=1}^{N} \sum_{t_j^i} \mathrm{d}\left(\chi_{\psi^i(t_j^i)}(M_0), S_j^i\right) + \gamma^{\psi}\mathrm{Reg}(\psi^i) + \gamma^{\chi}\mathrm{Reg}(\chi) \tag{9.5.15}$$

where we remove the geometrical diffeomorphisms ϕ^i which do not play any role in this discussion. Let us denote t_k the successive time-points $\psi^i(t_j^i)$ (in addition, we denote also $t_0 = 0$ and $t_{n+1} = T$) and set $\gamma^{\psi} = 0$ so that we do not penalize any $1D$-diffeomorphism. This leads to the simplified criterion:

$$J(\chi) = \sum_{t_k} \mathrm{d}\left(\chi_{t_k}(M_0), S_k\right) + \int_0^T \|v_t^{\chi}\|_{V^{\chi}}^2 \, dt. \tag{9.5.16}$$

Now, assume that we have a time re-parameterization function ψ_0 (independent of the subject), which change the mean scenario $\chi_t(M_0)$ into $\chi_{\psi_0(t)}(M_0)$. If we change the time-points t_k into $\psi_0^{-1}(t_k)$, then the fidelity-to-data is invariant under this time re-parameterization. Therefore, the question becomes: which parameterization of the mean scenario (and in particular which positions of the time-points t_k) makes the regularity term $\int_0^T \|v_t^{\chi}\|_{V^{\chi}}^2 \, dt$ minimal?

As we show in Section 4.2.2.3, the minimizing flow of diffeomorphism $(\chi_t)_{t\in[0,T]}$ is geodesic in each interval $[t_k, t_{k+1}]$ (for $k = 0, \ldots, n$). As a consequence, the minimizing flow is such that the norm of the velocity is constant in each interval (see Proposition 4.11). We denote $v_k = \|v_t^{\chi}\|$ for every $t \in [t_k, t_{k+1}]$. Therefore, the regularity term is written as:

$$\sum_{k=0}^{n} (t_{k+1} - t_k) v_k^2 \tag{9.5.17}$$

We denote $\delta_k = (t_{k+1} - t_k)$ and $l_k = \delta_k v_k$ the length of the geodesic which is invariant under time re-parameterization of the geodesic path (see Proposition 4.5). Therefore the regularity term becomes: $\sum_k l_k^2/\delta_k$. One must find the time points (i.e. the variables δ_k) which minimize this regularity term subject that $\sum_k \delta_k = T$. The solution of this constrained minimization problem leads to $\delta_k = v_0 l_k$ for a constant v_0. Combined with $\delta_k v_k = l_k$ this condition implies that $v_k = v_0$ for all k. This means that the minimizing parameterization of the piecewise geodesic path χ_t is such that *the velocity field has a constant norm all along this path.* This time parameterization moves the time-points t_k such that the condition: $(t_{k+1} - t_k)v_0 = l_k$ is satisfied. l_k can be seen as the intrinsic distance between two consecutive shapes S_k and S_{k+1}. The optimal parameterization adjusts the time-points t_k such that each distance l_k is covered at a constant speed. If two consecutive shapes have a very large difference, the time-interval tends to be larger. By contrast, the time parameterization tends to get closer consecutive shapes with little difference. This

can be stated as "the shape evolution sweeps out equal distance during equal intervals of time."[5]

We notice that this minimal parameterization is a direct consequence of the fact that we use a L^2 norm as a regularity term. If we use a L^1-metric instead (i.e. $\int_0^T \|v_t^\chi\|\,dt = \sum_k \delta_k v_k$), no particular parameterization would be favored since every parameterization would lead to the same regularity term. See Remark 4.6.

This fact can be proved experimentally. We use the same set of 5 profiles of hominid skulls as in Figure 4.1 and Section 9.4.1. The initial regression (without allowing any time re-parameterization) is performed as in Fig. 4.1. Then, instead of minimizing the pure regression criterion, we minimize the criterion:

$$J(\chi, \psi_0) = \sum_{t_k} \mathrm{d}\left(\chi_{\psi_0(t_k)}(M_0), S_k\right) + \int_0^T \|v_t^\chi\|_{V^\chi}^2\,dt. \tag{9.5.18}$$

which allows a time re-parameterization of the regression function (note that here we do not penalize the irregular time warps). Results are shown in Figure 9.10: they show that the time warp $\psi_0(t)$ moves the data time points so that the regression function goes from one shape to another at a constant speed.

This time re-parameterization has a direct impact for the construction of the atlas which minimizes Eq. (9.5.15). Indeed, assume that ψ^i, χ are the diffeomorphisms which minimize the criterion in Eq. (9.5.15), then the change $\chi_t \to \chi_{\psi_0(t)}$ and $\psi^i \to \psi_0^{-1} \circ \psi^i$ keep the regularity term invariant, whereas it minimizes the regularity term when ψ_0 is the optimal parameterization of the mean scenario. This means that the minimizing time warp ψ^i returned by the atlas estimation are of the form:

$$\psi^i = \psi_0^{-1} \circ \tilde{\psi}^i \tag{9.5.19}$$

where the time-change function ψ_0 is independent of the subject i and $\tilde{\phi}^i$ a 1D-diffeomorphism. This is a systematic bias of the time warp.

As a consequence, the statistics on the time-change functions must be invariant under the "addition" of such a bias. For instance, it does not make sense to test whether the mean of the time-change functions is significantly different from 0. One should test instead whether the means of two sub-groups are significantly different (the difference being considered as the composition with the inverse).

9.5.3 How to choose the time interval of interest?

One important parameter of the method is the choice of the time interval of interest $[0, T]$. The length of this interval and its position with respect to the data time-points are arbitrary parameters. We list here the impact of the choice of this parameter on the method.

For atlas construction, the time $t = 0$ corresponds to the time-point of the estimated template M_0. Let $t_{\min}$ be the earliest time point in the database. The evolution of the

[5] Remember Kepler's second law of planetary motion: "The line joining a planet and the Sun sweeps out equal areas during equal intervals of time", thus meaning that the planet accelerates as it goes closer to the Sun and slows down as it goes further away.

mean scenario in the interval $[0, t_{\min}]$ depends strongly on the shape difference between the template and the earliest shapes in the database as well as the length of the interval $[0, t_{\min}]$. This raises naturally the question: how the atlas adjust the estimation of the template with respect to the length of the time-interval $[0, t_{\min}]$?

The template M_0 is such that $\sum_{i,j} \left\| \phi^i \circ \chi_{\psi^i(t^i_j)}(M_0) - S^i_j \right\|^2$ is minimized. This function does not depend on the position of $t_{\min}$ with respect to 0 (we neglect here the impact that the mean scenario between 0 and $t_{\min}$ may have on the spatiotemporal deformations (ϕ^i, ψ^i)). Let us assume therefore that the template M_0 is fixed. Let us also assume that the mean scenario after $t_{\min}$ depends only on the data and not on the origin of the time-interval: in particular the mean scenario at time $t = t_{\min}$ is given by $S_{\min}$. Therefore the time origin which minimizes the criterion of the atlas construction should minimize $\int_0^{t_{\min}} \|v^\chi_t\|^2_{V^\chi}\, dt$: the kinetic energy of the geodesic path which connects M_0 to $S_{\min}$. Since we suppose that these two shapes do not depend on the time origin, the length of this geodesic is a constant: $t_{\min} v_0 = C$ (where v_0 denotes the (constant) norm of the geodesic v^χ in the interval $[0, t_{\min}]$). The kinetic energy is therefore given by $t_{\min} v_0^2 = C^2/t_{\min}$. This shows that the greater $t_{\min}$, the smaller the kinetic energy between 0 and $t_{\min}$, the more negligible the part of the regularity term $\mathrm{Reg}(\chi)$ between $[0, t_{\min}]$ compared to the regularity term between $t_{\min}$ and T. This suggests that the choice of the time-origin has less and less effect on the method as this origin is far from the first data time point. However, more experiments have to be performed to verify this fact and to show if our assumptions are realistic in practice (in particular the fact that the template does not change much if we only change the time-origin). Note that, by contrast, the value T has a little impact since v^χ_t vanishes on the interval $[t_{\max}, T]$.

In our method, we build the time warp as $1D$-diffeomorphism of $\mathbb{R}$ although we are using them as diffeomorphism of the time interval of interest $[0, T]$, thus assuming that $\psi(0) = 0$ and $\psi(T) = T$. This means that we approximate the interval $\psi([0, T])$ by $[0, T]$. Due to our boundary conditions ($\psi(t) \sim t$ as $t \to \pm\infty$) this approximation is more and more precise as the time-origin 0 and T go far away from the data time points (t^i_j). This is of little importance in theory, since all our framework can be extended directly to account for the time-interval of interest tending to $(-\infty, \infty)$ (v^χ_t has a compact support therefore it is integrable over the whole real line). By contrast, this is of great importance in practice. During the application of the atlas estimation, the values of $\psi(t)$ which are greater than T (resp. smaller than 0) are approximated by T (resp. by 0). This means that if the evolution of the subject i is in advanced with respect to the mean scenario near the end of time interval, then $\psi^i(t)$ tends to be much larger than T for t close to T. Because of the thresholding effect, there is nothing to prevent $\psi(t)$ to tend to ∞ (which results experimentally in $\psi(t) = T$ in a large time-interval $[T - \Delta T, T]$). This has been observed in practice when the time interval has been chosen to short compared to the spreading of the data time point. This can be avoided by choosing the interval borders far from the data time point. Another solution would be to redefine the time-interval during the atlas estimation. But this would also modify the weight of $\mathrm{Reg}(\chi)$ in the criterion as discussed in the previous paragraph and therefore would require to re-set the parameter γ^χ.

9.5.4 Toward a better individual growth model

In this framework, we use the regression model of Chapter 4 for the individual growth function $(\chi_t)_{t\in[0,T]}$. This models has two major drawbacks: first, it is piecewise geodesic, meaning that it is not differentiable on some points and, second, its extension for $t > t_{\max}$ and $t < t_{\min}$ is a constant function (where $t_{\max}$ and $t_{\min}$ denote the earliest and latest time point in the regression respectively.)

The non-differentiability of the regression model has been already discussed in Section 9.5.1, where we indicate how the minimization of the criterion can be performed even in presence of discontinuities. This method remains theoretical since it involves heavy computations via an exhaustive search of any combination of left and right derivatives. It would be beneficial to use in the future a regression model which is at least once differentiable.

Extrapolating the regression outside the time-interval of interest by a constant evolution is also arguable. Indeed, assume that a target shape is associated to a time-point which is later than the latest time-point of the source. This target shapes falls in the time-interval where the source evolution is constant. As a consequence, the differentiation of the matching term with respect to this time-point in Eq. (9.3.19) is equal to zero. Since the evolution of source is locally flat at this time-point, there is no reason to move this target shape along the time-axis to have a better matching with the source shapes. This does not mean that this time point will not move, since it may be influenced by other target shapes which may tend to accelerate or slow down the evolution of the target at this time point. This problem can be avoided by choosing the subject with the most widely spread time points as source subject. In that case, the target time points will fall within the non-flat part of the source evolution. Note that this is always the case in atlas to subject estimation, since the mean scenario of evolution is a regression of every shape in the database.

Even if we can overcome the drawback of having a constant evolution outside the time interval of interest, we think that another extrapolation method should be used, especially if we would like to use the regression also to predict the future evolution from past observations. A constant evolution is not the most natural way of extending the past evolution. Following the geodesic path instead seems a more natural idea. This suggests to use the Hamiltonian formulation for solving the regression problem, namely using the geodesic shooting to compute the evolution for $t > t_{\max}$.

Such an Hamiltonian approach has been followed in [Vialard 2009] to propose a new growth model which is differentiable everywhere and which is defined at any time in the future without assuming a constant extrapolation outside a given interval. The idea is the following: a geodesic flow of deformations follow the Euler-Lagrange equations as in Eq. (4.2.22):

$$\begin{cases} \frac{d\alpha_p(t)}{dt} = -\left(d_{x_p(t)}v_t\right)^t \alpha_p(t) \\ \frac{dx_p(t)}{dt} = v_t(x_p(t)) \end{cases} \tag{9.5.20}$$

This is an Hamiltonian system where the first equation gives the derivative of the velocity (i.e. the acceleration) and the second equation the derivative of the position (i.e. the velocity). We can generate perturbation of this Hamiltonian system by introducing an

external force ε in the acceleration term as follows:

$$\begin{cases} \frac{d\alpha_p(t)}{dt} = -\left(d_{x_p(t)} v_t\right)^t \alpha_p(t) \quad + \boldsymbol{\varepsilon}_{\mathbf{p}}(\mathbf{t}) \\ \frac{dx_p(t)}{dt} = v_t(x_p(t)) \end{cases} \tag{9.5.21}$$

The integration of this Hamiltonian system leads to a differentiable motion (even if the external force is discontinuous). However, the resulting flow of diffeomorphism is neither geodesic nor piecewise geodesic: it is a perturbation of a geodesic path.

How to estimate the external force so that this evolution performs a shape regression (in a sense of the minimization of a regression criterion) is a question to be addressed in the future. We can guess already that, as time tends to infinity (t outside the time interval of interest), the force tends to zero and the evolution follows a geodesic path. The relevance of the geodesic shooting for predicting future evolution must be also discussed.

9.5.5 $(d+1)$-currents in $4D$ for the comparison of shape evolutions

In the registration framework, we compare the source with the target by comparing the target shapes with the corresponding shape in the source evolution. As a consequence, one needs to perform the regression of the source only and not of the target. However, one can imagine to perform the regression of the source and the target and to compare both evolution.

The evolution of a d dimensional sub-manifold in $3D$ ($d = 0$ of points, $d = 1$ for curves, $d = 2$ for surfaces and $d = 3$ for volumes) can be seen as a $(d + 1)$ dimensional manifold in $4D$. Therefore, such shape evolution can be modeled as 4-currents along the lines of Chapter 1 and Section 1.3 in particular. A 4-dimensional kernel (which can be of the form $K^\phi \otimes K^\psi$) would define a RKHS for 4-currents and a norm between shape evolutions. Moreover, the spatiotemporal deformation (ϕ, ψ) defines a deformation of the underlying $4D$ space: $(\phi, \psi)(x, y, z, t) = (\phi(x, y, z), \psi(t))$. The application of this deformation on a shape evolution $S(t)$ leads to $(\phi, \psi)(S(t)) = \phi(S(\psi(t)))$ which is the same action as in our subject-specific framework. These $4D$-deformations can be extended as push-forward action on 4-currents along the lines of Section 1.3. This would lead to the matching criterion:

$$J(\phi, \psi) = \|(\phi, \psi)_* S(t) - T(t)\|^2_{W^*} + \gamma \mathrm{d}(\mathrm{Id}, (\phi, \psi))^2, \tag{9.5.22}$$

where $\mathrm{d}(\mathrm{Id}, (\phi, \psi))^2$ would be written as $\gamma^\phi \mathrm{d}(\mathrm{Id}, \phi)^2 + \gamma^\psi \mathrm{d}(\mathrm{Id}, \psi)^2$ in case of separable kernel $K = K^\phi \otimes K^\psi$ (i.e. no correlation between space and time).

This criterion looks similar to the criterion we proposed here except that we compare *every* instant of the target evolution with *every* instant of the deformed source evolution (instead of comparing only at the target time points) and that we compare not only the geometry of the shapes at each instant but also the derivative of this shape evolution at each instant. Indeed, the normal of the $(d + 1)$-dimensional manifold has a component which point in the direction of the local evolution the tangent-space of the d-dimensional manifold. Currently, we are considering the shape evolution as a collection of d-currents, which is different from considering the shape evolution as a $(d + 1)$-currents as explained in Remark 1.19.

However, the registration algorithm of Chapter 4 and the computational framework of currents introduced in Chapter 2 and 3 holds only for currents of dimension or co-dimension 0 or 1. This means that in $4D$ we can easily deal with evolution of points (dimension 1), evolution of surfaces (co-dimension 1) and evolution of volumes (co-dimension 0). However, the evolution of curves is of dimension 2 in a space of 4 dimension. The 2-differential form in $\mathbb{R}^4$ cannot be represented by a scalar or vector field. Dealing with such cases would require to define a new computational framework and to find new algorithms.

9.6 Conclusion

This chapter is an attempt to define a consistent framework for the spatiotemporal analysis of longitudinal data. It is based on a regression scheme which enables to infer a mean evolution from data sparsely distributed in time. It defines a spatiotemporal registration which compares to the set of data of two different subjects both in terms of an atemporal morphological change and a change of the dynamics of evolution. From a longitudinal data set (several subjects scanned several times), the method enables to estimate a mean scenario of evolution and to measure the spatiotemporal variations of this mean scenario in the population. This framework is rooted into a rigorous mathematical framework which allows us to perform statistics on both the morphological deformations and the time warps. The preliminary experiments we conducted suggest that this statistical framework can be used to characterize pathology in terms of change of dynamics of evolution instead of a pure geometrical variability at a given age.

This work is still in its early developments. In the one hand, more experiments should help to better understand the strength and limitations of the method. In the other hand, the methodology raises many theoretical questions. In particular, we discussed some of these questions:

- The more obvious problem of this methodology is the non differentiability of the regression function at some time-point. To overcome this issue, we proposed an alternative optimization strategy which computes only directional derivatives. This method has the advantage to stop only at critical point of the criterion by contrast to the approach we proposed in Section 9.3.2. However, due to its computational complexity, we did not implement it.

- The atlas estimation retrieves a mean scenario of evolution. At the first glance, it seems that the parameterization of this mean scenario is arbitrary (any time re-parameterization may be balanced by the time warps). However, we showed that the quadratic penalty term of the regression function imposes a particular time parameterization of the mean scenario. Having shown that, we conclude that statistics on time warps must be invariant under a global re-parameterization of the mean scenario.

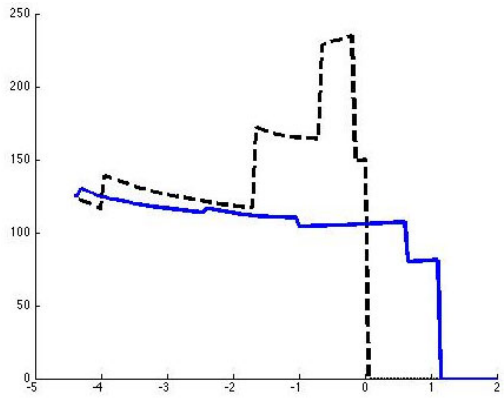
- The global framework depends on parameters which are not necessarily independent. We discussed the choice of the time-interval of interest and its possible impact on the estimation of the prototype shape.

- As a perspective for future work, we discussed the possibility to change the regression function with another growth model along the lines of [Vialard 2009]. In particular, this would avoid the discontinuity of the mean scenario of evolution and would allows us to extrapolate this mean scenario in a non-constant way both in the past and future.

- We claimed in Chapter 1 that currents are able to model a wide range of geometrical objects. We discussed the possibility to use 4-dimensional currents to model shape evolution.

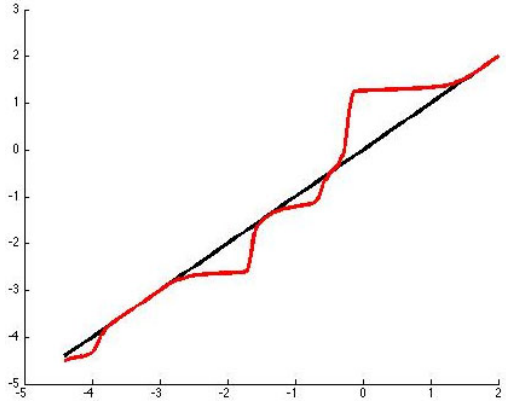
(a) original regression (top row) and regression with optimized parameterization (bottom row)

(original skulls in bold red, each blue frame last 0.1 million of years)



(b) norm of the velocity over time

(original: dashed line, re-parameterized: solid line)



(c) time warp ψ_0

Figure 9.10: Effect of time parameterization on the regression function: optimizing both the regression function and its time-parameterization leads to an optimal parameterization of the regression. This parameterization is given by the time-change function ψ_0 in (c): its tends to move the data time-points (a) so that the velocity along the regression function has a constant norm (b). From a geometrical point of view, the two regression functions are the same: one is simply accelerated or slowed down so that the shape changes are smoothed over time.

Conclusion and perspectives

Conclusion

This thesis aims to designing generic statistical models for extracting significant geometrical information from 3D or longitudinal anatomical data. For this purpose, we developed new tools for processing shapes in the framework of currents, new statistical models for measuring the variability of shapes and their evolution as well as algorithms to implement the estimation of such statistical models from anatomical data. We showed that this methodology can be used to address concrete anatomical problems in brain, heart and lung imaging.

The toolbox of currents for processing anatomical shapes

This work is rooted into the framework of currents for modeling shapes. This allows us to process in a single framework various kinds of geometrical data such as individual curves, bundles of curves or surfaces, for instance. A metric between sets of such shapes is derived, which takes into account correspondence between anatomical structures at any desired level: correspondence between points, between curves or between set of curves for instance.

Whereas this framework has been already used for registering set of anatomical data [Glaunès 2005], new tools were required for dealing with currents in a statistical context. First, we defined finite-dimensional approximation spaces, which give a way to derive robust and stable algorithms on currents. The approximation error is controlled by a single parameter (the step of the underlying grid) independently of the currents to be processed. This tool particularly was used to give a more robust and more generic implementation of the registration algorithm. Second, we introduced a matching pursuit algorithm to give a sparse representation of currents at any desired accuracy. This representation integrates the local redundancy of the signal at a given spatial scale. It is therefore particularly adapted for processing shape statistics.

Statistical models of anatomical variability

Once we completed the toolbox of currents, we focused on the definition, the discussion and the implementation of statistical models of anatomical data. Our purpose was to extract common anatomical features in a collection of shapes which have been segmented in several subjects. We rooted the extraction of significant features into a rigorous statistical model which considers the data as random deformations of an unknown template perturbed by random Gaussian currents. The deformations capture the geometrical variability such as torque or shearing effect for instance. The residual perturbations capture the non-diffeomorphic variations like creation of extra structures or change of matter density for example. This gives a description of the anatomical variability which takes all the available geometrical information into account.

The estimation of the statistical model requires to solve a deconvolution problem. We provide an approximation of the solution at any desired accuracy using the sparse representation of currents. Tests of the resulting algorithm for atlas construction on simulated data showed the potential of the statistical estimations to be used for group comparison and shape classification.

Addressing concrete anatomical problems

Then, we focused on the application of the proposed methodology to address concrete anatomical problems. First, the ability of the registration method to align consistently anatomical curves and surfaces is assessed on images of the lungs. Remarkably, registering the vessels and the lung surface alone enables to achieve results which are comparable to the registration of image intensities.

Second, we derived variability measures of the cortex surface from the localization of sulcal constraints in several subjects. This work showed how the diffeomorphic registrations of currents can be used to find a deformation of the underlying biological material, which best aligns the spatial constraints in a least squares sense. A comparison with the state-of-the art method showed that the integration of spatial constraint by a single deformation of the underlying biological tissue provides complementary measures of variability.

Third, we described the anatomical variability of white matter fiber bundles. We estimated consistently a template of fiber bundles along with its deformations in the population. The subsequent statistical analysis on the deformations and the residual perturbations in the space of currents showed that both the geometrical part and the "texture" part of the variability contains interesting anatomical features. This application is probably the one for which the framework of currents is more adapted and which shows the most competing advantages.

Fourth, we estimated the anatomical variability of the right ventricle of the heart of patients suffering from Tetralogy of Fallot. The retrieved modes of deformations have been given a precise anatomical meaning by experts. Then, correlations with clinical parameters allowed us to propose a personalized remodeling scenario of the ventricle for each patient. This scenario has been used successfully to predict the evolution of the shape of the ventricle for two patients.

Statistical analysis of shape evolution

Eventually, we extended our statistical models to analyze the spatiotemporal variability of longitudinal data (several subjects scanned several times). First, we extended the registration scheme of currents to a regression framework which estimates a continuous shape evolution from a set of time-indexed shapes. Then, we proposed a four-dimensional registration scheme which decomposes the difference between two sets of time-indexed shapes as a morphological deformation (independent of time) and a time warp which captures the change in the dynamics of the evolutions. Finally, we proposed a statistical model which infers from longitudinal data a mean scenario of evolution and its spatiotemporal variability in the studied population. This variability is decomposed into morphological variations and

variations of the speed of evolution. Such a statistical analysis performed on deep brain structures suggests that pathologies such as autism could be characterized more by delays in the growth of the anatomical structures than by a change of the shape of the structure at a given age.

Perspectives

In each chapter, we discussed the strength and limitations of the proposed contributions. Here, we discuss some points which appeared in several chapters and indicate some perspectives worth to be investigated in the future.

Generative models and measure of significance

In the third part, we focused mainly on the estimation of generative statistical models from a set of anatomical data. The estimation of such models allows us to reveal common anatomical features in the studied population. These features are represented as modes of deformations or modes of the residual shapes in the space of currents. This offers a way to *interpret* the variability that the model captured. At this point, our atlas construction method is considered as a way to probe a set of anatomical data and to extract significant information from it. Once this information has been extracted, one should focus on testing precise hypotheses to measure how significant or how discriminative the features are.

First, we can use the interpretation of the variability to select some anatomical parameters which are likely to be significant. For instance, the analysis of the variability of the white matter fiber bundles in Chapter 7 reveals a strong asymmetry in terms of density of fibers between the two branches of the cortico-bulbar tract (see Fig. 7.8). Then, one could measure carefully the local density of the fibers in each subject and test if this density differs significantly in each branch. The feature vector to be tested is of small dimension. It can be used to classify subjects or to find consistent clusters in the population. Particular feature vectors have to be designed manually for each problem.

Second, we can use directly the retrieved modes of deformation and modes of residuals to design statistical tests. For this purpose, one needs to select the significant modes, or the combination of modes which gives significant or discriminative statistics. Usual classification method, such as Fisher discriminant analysis can be used to find a discriminative hyperplane, like in Chapter 5 where the deformation modes were able to classify simulated data with an error below 1%. In Chapter 8, we proposed a way to select the combination of modes which show the strongest correlations with clinical parameters for the prediction of the remodeling of the heart.

The main limitation of this approach is that the dimension of the modes is usually much greater than the number of available data. Moreover, a mode usually combines several local effects, like a torque of one part of the template and an elongation of another part. These effects might also occur at different spatial scales. A way to find feature vectors of smaller dimension could be to decompose the modes into a set of localized space-frequency atoms. These filtered signals select the information contained in the modes, have intrinsically fewer degrees of freedom and therefore have more chance to be statistically significant. The use

of the matching pursuit introduced in Chapter 3 could be investigated in this context. The framework of RKHS also allows us to use several kernels with a decreasing spatial parameters (λ_V or λ_W) which define a decomposition of the space of L^2-signals into various frequency bands. Preliminary experiments on the simulated example of Chapter 5 show that the discriminative power of the method is increased if the modes are filtered beforehand. This helps also to highlight more precisely the most discriminative features. Future work should focus on better understanding this approach and, in particular, on determining for which parameters (dimension of the modes, number of subjects and signal to noise ratio) it may lead to highly significant statistics.

Shapes versus currents

The framework of currents does not lead to a precise definition of shapes. Indeed, it considers shapes via their response to stimulation by a vector field. It has the advantage to compare shapes globally and independently of their parameterization. In particular, it addresses the point correspondence issue. However, this framework embeds objects which may be very different from what we usually call "shapes". As we discussed in Chapter 3, a linear combination of shapes in the space of currents usually leads to a current does not look like a shape, like in Fig. 3.6 which shows the empirical mean of surfaces as a set of triangles (representing a linear combination of Dirac delta currents).

Two points of view are possible. On the one hand, one may use the modeling of currents to derive algorithms, while enforcing the data to be processed to stay in the space of shapes. On the other hand, one may fully embrace Federer's point of view and go beyond the intuitive idea of shapes: *"one must abandon the idea of describing all the competing surfaces by continuous maps from a single predetermined parameter space. One should rather think of surfaces as m-dimensional mass distributions, with tangent m-vectors attached."*, as outlined in Chapter 1. These two points of view, which seem incompatible, coexist in this thesis.

The registration algorithm of Chapter 4 uses the metric on currents to drive the deformation of shapes, in a way which preserves the representation of surfaces as meshes and the representation of curves as polygonal lines. As explained in Remark 4.13, we differentiated the norm of currents with respect to the position of the vertices of the mesh cells, so that the deformation of a mesh remains a mesh. By contrast, we could consider a mesh as a collection of unconnected mesh cells and differentiate the norm of current with respect to the center of mass of the cells. This has the advantage to be more consistent with the approximation of meshes as currents in the framework of Prop. 1.11, although the deformation of a mesh would result in collection of unconnected cells, namely a current which went away from the "space of acceptable shapes". The first approach seems more adapted to register surfaces like the endocasts of bonobos in Section 4.6. The second one seems more adapted to register the white matter fiber bundles in Chapter 7, which are given as a set of Dirac delta currents (set of unconnected oriented points) after having been approximated by the matching pursuit algorithm.

Similarly, in the framework of Chapters 5 and 9 we estimate a template directly in the space of currents, without constraining the template to be similar to a shape. Therefore, the estimated template is a combination of Dirac delta currents which looks like the current

represented in Fig. 3.6. We believe that this is a good representation for giving quantitative measures of variability. However, for pure visualization purposes, we usually map the mesh or the curves of one subject to the template (as we did in Chapters 5 and 8 for instance). By contrast, one could imagine to estimate the template by minimizing the same likelihood as we proposed in this thesis, but adding the constraint that the template should be the smooth deformation of a predetermined mesh. Such ideas have been developed in [Ma 2008] for instance. Of course, such an approach raises the question of the choice of the initial mesh.

We could also imagine to automatically reconstruct a mesh from a current, for instance via a minimal surface problem. Such an approach is usually possible only if we specify the topology of the surface to be reconstructed. However, if we average a surface with two connected components with surface with one connected component, how many components should have the average surface? Similarly, the problem of curves reconstruction leads to different answers whether we want to reconstruct a single sulcal lines or a whole fiber bundle. Therefore, it is difficult to imagine a generic answer to this reconstruction problem. A specific solution has to be designed in each case.

Modeling shapes as currents is particularly adapted to emphasize local properties of shapes, like the local direction of the information flux in the fiber bundles for instance. In this case, individual points of shapes do not play any particular role and long-range interactions between them are not taken into account. By contrast, this model of shapes is less adapted to describe global properties of shapes like total length or area, the number of connected components, connections between points. In this last case, currents could still be used, but at some point, one may need to project currents back into the "space of shapes". This arises, for instance, if one wants to use the estimated variability to constrain the segmentation of anatomical structures in new images.

Integration of anatomical, functional and physiological data

In this thesis, we dealt with the measure of the variability of sulcal lines, of white matter fiber bundles and of deep brain structures. These anatomical data are located at different locations in the brain: the cortex surface, the white matter and the grey matter. We showed that all these data can be processed in a single unifying framework. Therefore, applying the framework of currents to the combination of all these data can be done straightforwardly. This would lead to a model of variability of the whole brain. The more anatomical data, the more constrained the model, the more likely the statistical estimations to reveal interesting anatomical features. Nevertheless, to define such a model, we need to have access to a database in which various kind of anatomical data have been segmented in the same subjects.

In addition to the anatomical data, one should also focus on functional data. Actually, one the main goal of the brain image analysis is to understand how the anatomy correlates with the function. A first step in this direction could be the definition of the registration of the cortex surface along with the activation map drawn at the surface of the cortex. The framework of currents seems particularly adapted to deal with such kind of data. In Chapter 1, we showed that a map of attributes can be taken into account along with the

geometry. As we remarked, the coupling between attributes and geometry could lead to some ambiguities. However, we think that the framework of currents could be adapted to overcome this issue. Similarly, it would be of great interest to constraint such registration of activation map with the position of white fiber bundles, since such bundles are supposed to connect different functional areas of the brain. This could be a starting point to bridge the gap between anatomy and function in brain imaging.

More generally, as we discussed in the introduction, it would be beneficial to couple the geometrical variability with physiological or clinical parameters. Indeed, the ultimate goal of this approach is to explain the observed anatomical variability by a biological and physical model. A first step in this direction has been made in Chapter 8 for the prediction of the remodeling of the right ventricle of the heart. The correlations between the geometrical variability and the clinical parameters have been computed once the variability model has been estimated. One must think now to integrate the biological variables to constrain the estimation of the variability model. This would help, beyond the description and prediction of the anatomical variability, to understand the causes of this variability and therefore to offer a way to treat pathological cases for instance.

Valorization

Scientific Publications

This thesis leads to the following publications:

Articles in journal

- [Durrleman 2008c] S. Durrleman, X. Pennec, A. Trouvé, P. Thompson and N. Ayache. Inferring brain variability from diffeomorphic deformations of currents: an integrative approach. Medical Image Analysis 2008. *This paper is included in Chapter 6.*

- [Durrleman 2009c] S. Durrleman, X. Pennec, A. Trouvé and N. Ayache. Statistical models of sets of curves and surfaces based on currents. Medical Image Analysis, 2009. *This paper is included in Chapter 3 and 5.*

Peer-reviewed conference papers

- [Durrleman 2007] S. Durrleman, X. Pennec, A. Trouvé and N. Ayache, Measuring Brain Variability via Sulcal Lines Registration: a Diffeomorphic Approach, MICCAI 2007. *This paper has been extended to a journal version in [Durrleman 2008c].*

- [Durrleman 2008b] S. Durrleman, X. Pennec, A. Trouvé and N. Ayache, Sparse Approximation of Currents for Statistics on Curves and Surfaces, MICCAI 2008. *This paper has been awarded "young investigator award" in the category "statistical shape analysis". This paper has been extended to a journal version in [Durrleman 2009c].*

- [Durrleman 2009a] S. Durrleman, P. Fillard, X. Pennec, A. Trouvé and Nicholas Ayache, A Statistical Model of White Matter Fiber Bundles based on Currents, IPMI 2009. *This paper is included in Chapter 7. An extended journal version is in preparation.*

- [Durrleman 2009d] S. Durrleman, X. Pennec, A. Trouvé, G. Gerig and N. Ayache. Spatiotemporal Atlas Estimation for Developmental Delay Detection in Longitudinal Datasets, MICCAI 2009. *This paper is included in Chapter 9. An extended journal version is in preparation.*

- [Mansi 2009] T. Mansi, S. Durrleman, B. Bernhardt, M. Sermesant, H. Delingette, I. Voigt, Ph. Lurz, A. M. Taylor, J. Blanc, Y. Boudjemline, X. Pennec and N. Ayache. A Statistical Model of Right Ventricle in Tetralogy of Fallot for Prediction of Remodelling and Therapy Planning, MICCAI 2009. *This paper is included in Chapter 8.*

Workshop papers

- [Durrleman 2008a] S. Durrleman, X. Pennec, A. Trouvé and N. Ayache, A Forward Model to Build Unbiased Atlases from Curves and Surfaces. International Workshop on the Mathematical Foundations of Computational Anatomy (MFCA) 2008. *This paper has been extended to a journal version in [Durrleman 2008c].*

- [Gorbunova 2009] V. Gorbunova, S. Durrleman, P. Lo, X. Pennec and M. de Bruijne, Curve- and Surface-based Registration of Lung CT images via Currents. Second International Workshop on Pulmonary Image Analysis, 2009. *This paper is included in Chapter 4.*

Award

The paper "Sparse Approximation of Currents for Statistics on Curves and Surfaces" has been awarded "Young Investigator Award" at the conference MICCAI in 2008 (New-York).

Software

The algorithms presented in this thesis for the sparse representation of currents (Algorithm 2) and for the construction of atlas of curves and surfaces (Algorithm 3 and 4) along with the registration of currents optimized in the framework of Chapter 2 has been included into a toolbox called "exoShape". This software has been protected by the "Agence de Protection des Programmes (APP)" under Inter Deposit Digital Number FR.001.410028.000.S.A.2009.000.21000.

Appendix

Appendix A

Forms and differential forms

In this Appendix, we recall the definitions and the main basic properties of the multivectors, the m-forms and the differential m-forms on $\mathbb{R}^d$. We present here the material which is needed for the definition of currents in the framework of this thesis. Some properties are given without any proofs. For more detailed presentation, we refer the reader to any handbook of differential geometry such as [Lang 1962, Sternberg 1964, do Carmo 1994].

Contents

A.1 Wedge product and m-multivectors

A.1.1 Definitions

The wedge product is a generalization of the cross-product, which extends the usual measure of areas and volumes in 3D. Theoretically, the wedge product $u \wedge v$ between u and v, two vectors in $\mathbb{R}^d$, is equal to the tensor product $u \otimes v$ up to any linear combination of the form $x \otimes x$. The set of all wedge product between any pair of vector of $\mathbb{R}^d$ is called the exterior algebra over $\mathbb{R}^d$ and is denoted $\Lambda(\mathbb{R}^d)$. Formally, we have this abstract (and non tractable) definition:

Definition A.1 (exterior algebra over $\mathbb{R}^d$)**.** *The exterior algebra $\Lambda(\mathbb{R}^d)$ is defined as the quotient algebra of the tensor algebra by the two-sided ideal I generated by all elements of the form $x \otimes x$ such that $x \in \mathbb{R}^d$.*

All what we need to know about the wedge product is the two following properties (which is a definition of the wedge product, in some sense): the wedge product is a bilinear operation and vanishes if two vectors are equals:

$$\begin{cases} (\lambda u + v) \wedge w = \lambda(u \wedge v) + u \wedge w \\ u \wedge u = 0 \end{cases} \tag{A.1.1}$$

for all $u, v \in \mathbb{R}^p$ and $\lambda \in \mathbb{R}$.

As a direct consequence of these properties, we have that: $u \wedge v = -v \wedge u$. Indeed, we have $(u+v) \wedge (u+v) = 0 = u \wedge v + v \wedge u$.

Then, we extend the wedge product between two vectors to the wedge product between any family of m-vectors via the associativity law: $u \wedge v \wedge w = (u \wedge v) \wedge w$. This leads to the definition of the mth exterior power of $\mathbb{R}^d$:

Definition A.2 (mth exterior power of $\mathbb{R}^d$)**.** *We call the* mth *exterior power of* $\mathbb{R}^p$ *the vector space spanned by the vectors of the kind* $u_1 \wedge \ldots \wedge u_m$ *for all* $u_i \in \mathbb{R}^d$*. We denote this space* $\Lambda^m \mathbb{R}^p$*. The vectors in* $\Lambda^m \mathbb{R}^p$ *are called* m*-multivectors.*

As a consequence of this definition, the m-multivector $u_1 \wedge \ldots \wedge u_m$ is totally antisymmetric. This means that it vanishes as soon as two u_i are equals. More generally, we have for any permutation of $\{1, \ldots, m\}$ σ:

$$u_{\sigma(1)} \wedge \ldots \wedge u_{\sigma(m)} = \text{sign}(\sigma) u_1 \wedge \ldots \wedge u_m, \tag{A.1.2}$$

where $\text{sign}(\sigma)$ denotes the signature of the permutation σ.

Moreover, we have the following property:

Proposition A.3. *Let* $(u_i)_{i=1\ldots m}$ *be* m *vectors in* $\mathbb{R}^d$ *and* A *a* m*-by-*m *matrix. Let* $v_i = \sum_{j=1}^{m} A_{ij} u_j$*, then*

$$v_1 \wedge \ldots \wedge v_m = |A|\, u_1 \wedge \ldots \wedge u_m, \tag{A.1.3}$$

where $|A|$ *denotes the (signed) determinant of the matrix* A*.*

Proof. By linearity, we have:

$$\begin{aligned} v_1 \wedge \ldots \wedge v_m &= \left(\sum_{j=1}^{m} A_{1j} u_j \wedge \ldots \wedge \sum_{j=1}^{m} A_{mj} u_j \right) = \\ &= \sum_{p \in \mathcal{P}_m} A_{1p(1)} \ldots A_{mp(m)} \left(u_{p(1)} \wedge \ldots \wedge u_{p(m)} \right) \\ &= \left(\sum_{p \in \mathcal{P}_m} \text{sign}(\sigma) A_{1p(1)} \ldots A_{mp(m)} \right) u_1 \wedge \ldots \wedge u_m \\ &= |A|\, u_1 \wedge \ldots \wedge u_m, \end{aligned} \tag{A.1.4}$$

by definition of the determinant ($\mathcal{P}_m$ denotes the set of $m!$ permutations of $\{1, \ldots, m\}$). ■

A.1.2 Euclidean basis for multivectors

Let $(\epsilon_i)_{i=1\ldots d}$ be the canonical basis of $\mathbb{R}^d$, so that each vector u_i is decomposed into $\sum_{k=1}^{d} u_i^k \epsilon_k$ Thanks to the linearity and the alternating properties of the wedge product we

have:

$$\begin{aligned} u_1 \wedge \ldots \wedge u_m &= \left(\sum_{k_1}^{d} u_1^{k_1} \epsilon_{k_1} \right) \wedge \ldots \wedge \left(\sum_{k_m=1}^{d} u_m^{k_m} \epsilon_{k_m} \right) \\ &= \sum_{p \in C_m^d} \sum_{\sigma \mathcal{P}_m} u_1^{\sigma(p(1))} \ldots u_m^{\sigma(p(m))} \epsilon_{\sigma(p(1))} \wedge \ldots \wedge \epsilon_{\sigma(p(m))} \\ &= \sum_{p \in C_m^d} \left(\sum_{\sigma \in \mathcal{P}_m} \text{sign}(\sigma) u_1^{\sigma(p(1))} \ldots u_m^{\sigma(p(m))} \right) \epsilon_{p(1)} \wedge \ldots \wedge \epsilon_{p(m)}, \end{aligned} \tag{A.1.5}$$

where C_m^d denotes the set of all subsets of m elements in $\{1, \ldots, d\}$ and $\mathcal{P}_m$ the set of all permutations of $\{1, \ldots, m\}$. This shows that the vectors $\varepsilon_{i_1} \wedge \ldots \wedge \varepsilon_{i_m}$ for $1 \leq i_1 < \ldots < i_m \leq d$ spanned the vector space $\Lambda^m \mathbb{R}^d$. One can easily show that these vectors are linearly independent. Therefore, the space $\Lambda^m \mathbb{R}^d$ is of dimension $\binom{d}{m}$. Then we write any m-multivectors on $\mathbb{R}^d$ as:

$$u = \sum_{1 \leq i_1 < \ldots < i_m \leq d} u_{i_1 \ldots i_m} \epsilon_{i_1} \wedge \ldots \wedge \epsilon_{i_m} \tag{A.1.6}$$

We provide $\Lambda^m \mathbb{R}^p$ with the standard Euclidean inner-product and norm:

$$|u|^2 = \sum_{1 \leq i_1 < \ldots < i_m \leq d} \left(u_{i_1 \ldots i_m} \right)^2 \tag{A.1.7}$$

Of course, this definition does not depend on the choice of the basis.

A.1.3 Particular cases

We study now some particular cases of interest:

- if $m > d$, $\Lambda^m \mathbb{R}^d = \{0\}$,
- if $m = 0$, $\Lambda^0 \mathbb{R}^d$ is of dimension 1: this is the space of scalars $\mathbb{R}$ itself,
- if $m = 1$, $\Lambda^1 \mathbb{R}^d$ is of dimension d: this is the vector space $\mathbb{R}^d$ itself,
- if $m = d - 1$, $\Lambda^{d-1} \mathbb{R}^d$ is of dimension d. The decomposition of a $d-1$-multivector $u_1 \wedge \ldots \wedge u_{d-1}$ on the basis $(\tilde{\varepsilon}_i^d = \varepsilon_1 \wedge \ldots \varepsilon_{i-1} \wedge \varepsilon_{i+1} \ldots \wedge \varepsilon_d)_{i=1,\ldots,d}$ (which denotes the set of d-multivectors $\varepsilon_1 \wedge \ldots \wedge \varepsilon_d$ in which the vector ε_i is missing) leads to:

$$u_1 \wedge \ldots \wedge u_{d-1} = \sum_{i=1}^{d} \eta_i \tilde{\varepsilon}_i^d, \tag{A.1.8}$$

where

$$\eta_i = \sum_{\sigma \in \mathcal{P}_m ; \sigma(d) = i} \text{sign}(\sigma) u_1^{\sigma(1)} \ldots u_{d-1}^{\sigma(d-1)} \tag{A.1.9}$$

This shows that the vector η is such that for any vector $\alpha \in \mathbb{R}^d$:

$$\begin{aligned}\eta^t\alpha = \sum_{i=1}^{d} \eta_i\alpha_i &= \sum_{i=1}^{d} \sum_{\sigma\in\mathcal{P}_m;\sigma(d)=i} \operatorname{sign}(\sigma) u_1^{\sigma(1)} \dots u_{d-1}^{\sigma(d-1)}\alpha_i \\ &= \sum_{i=1}^{d} \sum_{\sigma\in\mathcal{P}_m;\sigma(d)=i} \operatorname{sign}(\sigma) u_1^{\sigma(1)} \dots u_{d-1}^{\sigma(d-1)}\alpha_{\sigma(d)} \\ &= \sum_{\sigma\in\mathcal{P}_m} \operatorname{sign}(\sigma) u_1^{\sigma(1)} \dots u_{d-1}^{\sigma(d-1)}\alpha_{\sigma(d)} = \det(u_1,\dots,u_{d-1},\alpha)\end{aligned} \tag{A.1.10}$$

Therefore, any $d-1$-multivector $u_1 \wedge \ldots \wedge u_{d-1}$ is associated to a vector η such that $\eta^t\alpha = \det(u_1,\dots,u_{d-1},\alpha)$ for every vector α (see below the instance in 3D)

- If $m = d$, $\Lambda^d\mathbb{R}^d$ is of dimension 1: it is spanned by the vector $\epsilon_1 \wedge \ldots \wedge \epsilon_d$. Thanks to Eq. (A.1.5), we have:

$$\begin{aligned}u_1 \wedge \ldots \wedge u_d &= \sum_{\sigma\in\mathcal{P}_d} \operatorname{sign}(\sigma) u_1^{\sigma(1)} \dots u_d^{\sigma(d)} \epsilon_1 \wedge \ldots \wedge \epsilon_d \\ &= \det(u_1,\dots,u_d)\epsilon_1 \wedge \ldots \wedge \epsilon_d\end{aligned} \tag{A.1.11}$$

  All d-multivector are proportional to the basis vector $\epsilon\wedge\ldots\wedge\epsilon_d$. There is a one-to-one map between d-mutlivectors in $\mathbb{R}^d$ and the determinant of the vectors.

Let u and v be two vectors in dimension 3. Then the 2-multivector $u \wedge v$ is given in coordinates as:

$$u \wedge v = (u^2v^3 - u^3v^2)\epsilon_1 \wedge \epsilon_2 + (u^3v^1 - u^1v^3)\epsilon_3 \wedge \epsilon_1 + (u^1v^2 - u^2v^1)\epsilon_1 \wedge \epsilon_2. \tag{A.1.12}$$

We notice that the coordinates of $u \wedge v$ in the canonical basis of $\Lambda^2\mathbb{R}^3$ are precisely the coordinates of the cross product between u and v: $u \times v$. Any 2-multivector $u \wedge v$ in dimension 3 can be mapped isometrically to $u \times v \in \mathbb{R}^3$. Moreover, we all know that $(u \times v)^t w = \det(u, v, w)$.

A.2 m-forms as antisymmetric tensors

We define now the forms on the space of m-multivectors: the m-forms.

Definition A.4 (m-forms)**.** *A m-form ω on $\mathbb{R}^d$ is an linear map from $\Lambda^m\mathbb{R}^d$ to $\mathbb{R}$: $\omega : (u_1 \wedge \ldots \wedge u_m) \longrightarrow \omega(u_1 \wedge \ldots \wedge u_m) \in \mathbb{R}$, where every u_i is a vector in $\mathbb{R}^d$. We denote by $(\Lambda^m\mathbb{R}^d)^*$ the space of m-forms on $\mathbb{R}^d$.*

If we write $\omega(u_1,\dots,u_m) = \omega(u_1 \wedge \ldots \wedge u_m)$, we see that ω can be written as a m-covariant tensor. Due to the symmetries of the wedge product, this m-covariant tensor is totally antisymmetric (i.e. alternated forms). For example, in 3D, $(u, v, w) \to \det(u, v, w)$ is a 3-form and $u, v \longrightarrow (u \times v)^t z$ for a fixed vector z is a 2-form.

As the dual space of $\Lambda^m\mathbb{R}^d$, the space of m-forms in $\mathbb{R}^d$ is of dimension $\binom{d}{m}$. As an alternated tensor, ω is decomposed into:

$$\omega = \sum_{1\le i_1<\ldots<i_m\le d} \omega_{i_1\ldots i_m} dx_{i_1} \wedge \ldots \wedge dx_{i_m}, \tag{A.2.1}$$

where dx_i denotes the dual basis of $\mathbb{R}^d$ (i.e. $dx_i(\epsilon_j) = \delta_{i,j}$) and $dx_1 \wedge \ldots \wedge dx_m$ the antisymmetric part of the tensor $dx_1 \otimes \ldots \otimes dx_m$. In particular, $dx \wedge dy = dx \otimes dy - dy \otimes dx$.

The space of m-forms inherits from the same properties as the space of m-multivectors:

- If $m = 0$, ω is simply a constant mapping on $\mathbb{R}$.
- If $m = 1$, ω is a linear form on $\mathbb{R}^d$: for all $u \in \mathbb{R}^d$, $\omega(u) \in \mathbb{R}$. Thanks to the Riesz representation theorem, this linear form can be represented by the inner-product with a fixed vector $\overline{\omega}$:
$$\omega(u) = \overline{\omega}^t u. \tag{A.2.2}$$
- If $m = d - 1$, the space of $d-1$-forms is also of dimension d. With the notations of Section A.1.3, we have for any $d-1$-multivectors $u_1 \wedge \ldots \wedge u_{d-1} = \sum_{i=1}^{d} \eta_i \tilde{\varepsilon}_i^d$. Therefore, by linearity a $d-1$-form satisfies:
$$\omega(u_1 \wedge \ldots \wedge u_{d-1}) = \sum_{i=1}^{d} \eta_i \omega(\tilde{\varepsilon}_i^d) = \eta^t \overline{\omega}, \tag{A.2.3}$$
where $\overline{\omega}$ denotes the vector whose coordinates equal $\omega(\varepsilon_1 \wedge \ldots \varepsilon_{i-1} \wedge \varepsilon_{i+1} \ldots \wedge \varepsilon_d)$ for $i = 1$ to d. Therefore, a $d-1$-form also can be represented by an inner-product such that:
$$\omega(u_1 \wedge \ldots \wedge u_{d-1}) = \eta^t \overline{\omega} = \det(u_1, \ldots, u_{d-1}, \overline{\omega}), \tag{A.2.4}$$
according to Section A.1.3.
- If $m = d$, all d-forms are proportional to the determinant (the space of d-form in dimension d is of dimension 1). Indeed, every d-multivector $u_1 \wedge \ldots \wedge u_d$ is equal to $\det(u_1, \ldots, u_d)(\varepsilon_1 \wedge \ldots \wedge \varepsilon_d)$. Therefore every d-forms in dimension d is written as:
$$\omega(u_1 \wedge \ldots \wedge u_d) = \overline{\omega} \det(u_1, \ldots, u_d), \tag{A.2.5}$$
for a given scalar $\overline{\omega} = \omega(\varepsilon_1 \wedge \ldots \wedge \varepsilon_d)$.

We define the Euclidean norm of a m-form ω as the spectral norm (which corresponds to the Euclidean norm on $(\Lambda^m \mathbb{R}^d)^*$):

Definition A.5. *Let ω a m-form in $\mathbb{R}^d$. The norm of ω is defined as:*

$$|\omega|_{(\Lambda^m \mathbb{R}^d)^*} = \sup_{|u_1 \wedge \ldots \wedge u_m| = 1} |\omega(u_1 \wedge \ldots \wedge u_m)| = \left(\sum_{1 \leq i_1 < \ldots < i_m \leq m} (\omega_{i_1 \ldots i_m})^2 \right)^{1/2}, \tag{A.2.6}$$

where $\omega_{i_1 \ldots i_m}$ are the coordinates of the m-forms in the basis $dx_{i_1} \wedge \ldots \wedge dx_{i_m}$.

A.3 Differential forms as multi-covariant tensor fields

A.3.1 Definition

Like we extend the concept of vectors to vector fields on a smooth sub-manifold, we extend the concept of m-forms to differential m-forms. Each point x of a manifold is associated to a m-form $\omega(x)$ whose input vectors are chosen in the tangent-space of the manifold at point x. This leads to the following definition:

Definition A.6 (differential m-forms)**.** *A differential m-form on $\mathbb{R}^d$ (or on an open subspace of $\mathbb{R}^d$) maps every $x \in \mathbb{R}^d$ to $\omega(x)$ a m-form in $(\Lambda^m \mathbb{R}^d)^*$. We denote $\mathcal{C}^0(\mathbb{R}^d, (\Lambda^m \mathbb{R}^d)^*)$ the space of the differential m-forms which are continuous and tend to zero at infinity. It is provided with the norm:*

$$\|\omega\|_\infty = \sup_{x \in \mathbb{R}^d} \sup_{|u_1 \wedge \ldots \wedge u_m| \leq 1} |\omega(x)(u_1 \wedge \ldots \wedge u_m)| \,. \tag{A.3.1}$$

If $m = 0$, a differential 0-form is simply a scalar function on $\mathbb{R}^d$.

If $m = 1$, a differential 1-form is a vector field on $\mathbb{R}^d$.

If $m = d-1$, the $d-1$ differential form can be associated to a vector field on $\mathbb{R}^d$ thanks to the isometric mapping between the $d-1$-form on $\Lambda^{d-1}\mathbb{R}^d$ and the vectors on $\mathbb{R}^d$.

If $m = d$, the d differential forms are all of the form $\omega = \overline{\omega}(x) \det$ where $\overline{\omega}(x)$ is a scalar function on $\mathbb{R}^d$ and det denotes the determinant form on $\mathbb{R}^d$.

A.3.2 Integration of differential forms on a colored sub-manifold

In order to model sub-manifolds of $\mathbb{R}^d$ as currents, we need to define the integration of differential m-forms on this manifold.

Definition A.7. *Let $\mathcal{M}$ be an oriented sub-manifold of dimension m in $\mathbb{R}^d$ and I a integrable function on $\mathcal{M}$ with respect to the Lebesgue measure on $\mathcal{M}$. Let $\omega \in \mathcal{C}^0(\mathbb{R}^d, (\Lambda^m \mathbb{R}^d)^*)$ be a m-differentiable form (Note that the degree of ω equals the dimension of the sub-manifold).*

For all $x \in \mathcal{M}$, we denote by $u_1(x), \ldots, u_m(x)$ a positively oriented basis of the tangent-space of $\mathcal{M}$ at point x (defined almost everywhere). Then, we define the integral of ω on $(\mathcal{M}, I)$ as:

$$\int_{\mathcal{M}} I\omega = \int_{\mathcal{M}} I(x)\omega(x) \left(\frac{u_1(x) \wedge \ldots \wedge u_m(x)}{|u_1(x) \wedge \ldots \wedge u_m(x)|} \right) d\lambda(x), \tag{A.3.2}$$

where the integral on the right hand denotes the usual Lebesgue integral of a scalar function on $\mathcal{M}$ $d\lambda$ the usual Lebesgue measure on $\mathcal{M}$.

Proposition A.8. *The definition of the integral in Eq. (A.3.2) does not depend on the choice of the positively oriented basis of the tangent-space of $\mathcal{M}$ at point x.*

Proof. Let A be a m-by-m matrix which change the basis $u_1(x), \ldots, u_m(x)$ to the basis $v_1(x) = \sum_{k=1}^{m} A_{1k} u_k, \ldots, v_m(x) = \sum_{k=1}^{m} A_{mk} u_k$. Since the change of basis is supposed not to change the orientation, the determinant of A is positive. Thanks to Proposition A.3, we have that:

$$\begin{aligned} \frac{u_1(x) \wedge \ldots \wedge u_m(x)}{|u_1(x) \wedge \ldots \wedge u_m(x)|} &= \frac{|A| \, (v_1(x) \wedge \ldots \wedge v_m(x))}{\left||A|\right| \, |v_1(x) \wedge \ldots \wedge v_m(x)|} \\ &= \frac{v_1(x) \wedge \ldots \wedge v_m(x)}{|v_1(x) \wedge \ldots \wedge v_m(x)|}, \end{aligned} \tag{A.3.3}$$

since $|A| = \left||A|\right|$ (the absolute value of the determinant of A). ■

Remark A.9. This definition still holds if $\mathcal{M}$ is of dimension 0. In this case, $\mathcal{M}$ is a discrete set of points. The Lebesgue measure on $\mathcal{M}$ must be replaced by the measure $\sum_{x\in\mathcal{M}} \delta_x$ which counts the number of elements in $\mathcal{M}$. An integrable function on $\mathcal{M}$ is therefore a function which satisfies: $\sum_{x\in\mathcal{M}} I(x) < \infty$. The integral of a 0-form on $\mathcal{M}$ is simply the integral of a scalar function on $\mathcal{M}$. □

To compute the integral in Eq. (A.3.2) in practice, we need to write it with local charts. Let $\{U_i, \pi_i\}$ be an atlas of $\mathcal{M}$ and χ_i a partition of unity of the open cover $\{U_i\}$. This means that $\mathcal{M}$ is parametrized locally (on $U_i \subset \mathbb{R}^m$) by a piecewise differentiable chart $\pi_i : U_i \to \mathcal{M}$. We suppose moreover that every chart are positively oriented.

Let $x = \pi_i(\mathbf{p})$ be a point on $\mathcal{M}$ for $\mathbf{p} \in U_i$. We can choose $u_k(x) = \frac{\partial \pi_i(\mathbf{p})}{\partial p_k}$ as the positively oriented basis of the tangent plane of $\mathcal{M}$ at point $\pi_i(\mathbf{p})$. These vectors are considered in $\mathbb{R}^d$. However, they all belong to the tangent-space of dimension m. Let Π_x denote the orthogonal projection on this tangent-space. Therefore,

$$\begin{aligned}\left|\frac{\partial \pi_i(\mathbf{p})}{\partial p_1} \wedge \ldots \wedge \frac{\partial \pi_i(\mathbf{p})}{\partial p_m}\right| &= \left|\Pi_x\left(\frac{\partial \pi_i(\mathbf{p})}{\partial p_1}\right) \wedge \ldots \wedge \Pi_x\left(\frac{\partial \pi_i(\mathbf{p})}{\partial p_m}\right)\right| \\ &= \det\left(\Pi_x\left(\frac{\partial \pi_i(\mathbf{p})}{\partial p_1}\right), \ldots, \Pi_x\left(\frac{\partial \pi_i(\mathbf{p})}{\partial p_m}\right)\right) \\ &= \det\left(\frac{\partial \pi_i(\mathbf{p})}{\partial p_1}, \ldots, \frac{\partial \pi_i(\mathbf{p})}{\partial p_m}\right) \\ &= |d_{\mathbf{p}}\pi_i|\end{aligned} \tag{A.3.4}$$

since the magnitude of a d-multivector in dimension d is equal to the determinant of these vectors (See Section A.1.3). Since the charts are positively oriented, this determinant is positive.

Moreover, the Lebesgue measure written in the charts π_i is equal to: $d\lambda(x) = |d_x\pi_i|\, d\mathbf{p}$ for $x = \mathbf{p}$. Therefore, in the charts π_i, the norm of the multivector and the normalizing factor of the Lebesgue measure cancel ($|d_x\pi_i|$ in the numerator and denominator). The integral in Eq. (A.3.2) finally is written as:

$$\int_{\mathcal{M}} I\omega = \sum_i \int_{U_i} \chi_i(\mathbf{p}) I(\pi_i(\mathbf{p}))\omega(\pi_i(\mathbf{p}))\left(\frac{\partial \pi_i}{\partial p_1}(\mathbf{p}) \wedge \ldots \wedge \frac{\partial \pi_i}{\partial p_m}(\mathbf{p})\right) d\mathbf{p}, \tag{A.3.5}$$

where the integrals of the right-hand side denotes the usual Lebesgue integral on open subset of $\mathbb{R}^d$. Proposition A.8 shows that this expression is independent of the choice of the basis.

We remark that the argument of ω within the integrals written in local charts is *not* normalized in Equation (A.3.5). If $\mathcal{M}$ is a surface parametrized by $S(u, v)$, then the argument of ω is the non-normalized normal $\frac{\partial S}{\partial u} \times \frac{\partial S}{\partial v}$. By contrast, in the intrinsic formulation in Eq. (A.3.2), the argument of ω is the *unit* normal of S (the Lebesgue measure on $\mathcal{M}$ taking care of the right scaling of the normal).

A.3.3 Change of variable formula

Let $\mathcal{M}$ be a sub-manifold of $\mathbb{R}^d$ and I an integrable function on $\mathcal{M}$. This function plays the role of an image (i.e. a map of colors) drawn on the manifold. The purpose of this

section is to define the geometrical transport of such a colored manifold and to compute the integration of a differential form on the transported manifold, namely by the definition of a proper change of variable formula.

For $\mathcal{M}$ a sub-manifold of $\mathbb{R}^d$ and ϕ a diffeomorphism of $\mathbb{R}^d$, then we define $\phi(\mathcal{M})$ the geometrical transport of $\mathcal{M}$, namely the set of points $\phi(x)$ for all $x \in \mathcal{M}$. Since ϕ is a diffeomorphism, the regularity of $\phi(\mathcal{M})$ is the same as the regularity of the original sub-manifold $\mathcal{M}$.

If I is an image drawn on $\mathcal{M}$, then we define the transport of I by the diffeomorphism ϕ as $I \circ \phi^{-1}$. This means that the intensities on the manifold are carried along the deformation without any change. This action $((\phi, I) \to I \circ \phi^{-1})$ is the usual transport of intensities for image registration.

This leads to the following definition:

Definition A.10 (geometric transport of colored sub-manifolds)**.** *Let $\mathcal{M}$ be a rectifiable sub-manifold of $\mathbb{R}^d$ and I a scalar function on $\mathcal{M}$. Let ϕ be a diffeomorphism of $\mathbb{R}^d$. We define the geometrical transport of the couple $(\mathcal{M}, I)$ as:*

$$\phi\,(\mathcal{M}, I) = (\phi(\mathcal{M}), I \circ \phi^{-1}). \tag{A.3.6}$$

Our purpose now is to compute the integration of a m-differential form ω over the couple $\phi\,(\mathcal{M}, I)$: $\int_{\phi(\mathcal{M})} I \circ \phi^{-1}\omega$. If $u_1(x), \ldots, u_m(x)$ is a positively oriented basis of the tangent-space of $\mathcal{M}$, then $d_x\phi(u_1(x)), \ldots d_x\phi(u_m(x))$ is a basis of the tangent-space of $\phi(\mathcal{M})$, where $d_x\phi$ is a d-by-d Jacobian matrix of ϕ at point x. Therefore, the integral $\int_{\phi(\mathcal{M})} I \circ \phi^{-1}\omega$ is written as (using the linearity of the form $\omega(x)$):

$$\begin{aligned}\int_{\phi(\mathcal{M})} I \circ \phi^{-1}\omega &= \int_{\mathcal{M}} I \circ \phi^{-1}(\phi(x))\omega(\phi(x)) \left(\frac{d_x\phi(u_1(x)) \wedge \ldots \wedge d_x\phi(u_m(x))}{|d_x\phi(u_1(x)) \wedge \ldots \wedge d_x\phi(u_m(x))|} \right) d\lambda^{\phi}(\phi(x)) \\ &= \int_{\mathcal{M}} I(x)\omega(\phi(x))\,(d_x\phi(u_1(x)) \wedge \ldots \wedge d_x\phi(u_m(x)))\,\frac{d\lambda^{\phi}(\phi(x))}{|d_x\phi(u_1(x)) \wedge \ldots \wedge d_x\phi(u_m(x))|},\end{aligned} \tag{A.3.7}$$

where $d\lambda^{\phi}$ denotes the Lebesgue measure on $\phi(\mathcal{M})$.

We can restrict the tangential map $d_x\phi$ to map the tangent-space of $\mathcal{M}$ at x to the tangent-space of $\mathcal{M}$ at $\phi(x)$. We denote $d_x\tilde{\phi}$ this m-by-m matrix. Therefore, $|d_x\phi(u_1(x)) \wedge \ldots \wedge d_x\phi(u_m(x))| = \left|d_x\tilde{\phi}(u_1(x)) \wedge \ldots \wedge d_x\tilde{\phi}(u_m(x))\right| = \left|d_x\tilde{\phi}\right| |u_1(x) \wedge \ldots \wedge u_m(x)|$. Moreover, the Lebesgue measure on $\phi(\mathcal{M})$) is given by $d\lambda(\phi(x)) = \left|d_x\tilde{\phi}\right| d\lambda(x)$, so that the factor $d_x\tilde{\phi}$ in the numerator and denominator cancels:

$$\begin{aligned}\int_{\phi(\mathcal{M})} I \circ \phi^{-1}\omega &= \int_{\mathcal{M}} I(x)\omega(\phi(x))\,(d_x\phi(u_1(x)) \wedge \ldots \wedge d_x\phi(u_m(x)))\,\frac{d\lambda(x)}{|u_1(x) \wedge \ldots \wedge u_m(x)|} \\ &= \int_{\mathcal{M}} I(x)\phi^*\omega(x) \left(\frac{u_1(x) \wedge \ldots \wedge u_m(x)}{|u_1(x) \wedge \ldots \wedge u_m(x)|} \right) d\lambda(x),\end{aligned} \tag{A.3.8}$$

where we denote $\phi^*\omega(x)(u_1 \wedge \ldots \wedge u_m) = \omega(\phi(x))(d_x\phi(u_1) \wedge \ldots \wedge d_x\phi(u_m))$.

This justifies the introduction of the pullback action of a diffeomorphism on a differential m-form:

Definition A.11 (pullback action on differential forms)**.** *Let ω be a m-differential form on $\mathbb{R}^d$ and ϕ a diffeomorphism of $\mathbb{R}^d$ such that $\sup_{x\in\mathbb{R}^d} |d_x\phi| < \infty$. We define $\phi^*\omega$ a m-differential form (of the same regularity as ω) as:*

$$\phi^*\omega(x)(u_1 \wedge \ldots \wedge u_m) = \omega(\phi(x))\left(d_x\phi(u_1) \wedge \ldots \wedge d_x\phi(u_m)\right), \tag{A.3.9}$$

for all points $x \in \mathbb{R}^d$ and every vectors $u_i \in \mathbb{R}^d$. The differential form $\phi^*\omega$ is called the pullback action of the diffeomorphism ϕ on the differential form ω.

We can verify easily that the vector field $\phi^*\omega$ still belong to our space of differential m-forms $\mathcal{C}^0(\mathbb{R}^d, \Lambda^m\mathbb{R}^d)$ (since we suppose that $\sup_{x\in\mathbb{R}^d} |d_x\phi| < \infty$). Moreover, the pullback action is really an action of the group of diffeomorphism on the space of differential form, namely that $(\phi \circ \psi)^*\omega = \phi^*(\psi^*\omega)$ for all diffeomorphism ϕ and ψ.

Proposition A.12. *Let $\mathcal{M}$ be a sub-manifold of dimension m in $\mathbb{R}^d$ and I an integrable function on $\mathcal{M}$. Let ϕ be a diffeomorphism of $\mathbb{R}^d$. Then:*

$$\int_{\phi(\mathcal{M})} I \circ \phi^{-1}\omega = \int_{\mathcal{M}} I\phi^*\omega. \tag{A.3.10}$$

Proof. This is exactly what we proved in Equation (A.3.8). ■

We can write the pullback action on a differential m-forms on $\mathbb{R}^d$, ω, in some particular cases of interest according to the dimension m:

- If $m = 0$, then $\omega(x)$ is a scalar field and $\phi^*\omega = \omega \circ \phi$.
- If $m = 1$, then $\omega(x)$ is represented by a vector field $\overline{\omega}(x)$: $\omega(x)(u) = \overline{\omega}(x)^t u$. Therefore,

  $$\phi^*\omega(x)(u) = \omega(\phi(x))(d_x\phi(u)) = \overline{\omega}(\phi(x))^t d_x\phi(u) = \left(d_x\phi^t\overline{\omega}(\phi(x))\right)^t u. \tag{A.3.11}$$

  The vector field associated to $\phi^*\omega$ is $\phi^*\overline{\omega}(x) = d_x\phi^t\overline{\omega}(\phi(x))$.
- If $m = d-1$, the $\omega(x)$ is represented by a vector field $\overline{\omega}(x)$ such that $\omega(x)(u_1 \wedge \ldots \wedge u_{d-1}) = \det(\overline{\omega}(x), u_1, \ldots, u_{d-1})$. Therefore,

  $$\begin{aligned}\phi^*\omega(x)(u_1 \wedge \ldots \wedge u_{d-1}) &= \omega(\phi(x))(d_x\phi(u_1) \wedge \ldots \wedge d_x\phi(u_{d-1}))\\ &= \det\left(\overline{\omega}(\phi(x)), d_x\phi(u_1), \ldots, d_x\phi(u_{d-1})\right)\\ &= |d_x\phi| \det\left(d_x\phi^{-1}\overline{\omega}(\phi(x)), u_1, \ldots, u_{d-1}\right)\\ &= \det\left(|d_x\phi|\, d_x\phi^{-1}\overline{\omega}(\phi(x)), u_1, \ldots, u_{d-1}\right),\end{aligned} \tag{A.3.12}$$

  so that the vector field associated to $\phi^*\omega$ is $|d_x\phi|\, d_x\phi^{-1}\overline{\omega}(\phi(x))$.
- If $m = d$, the $\omega(x)$ is represented by a scalar field $\overline{\omega}(x)$ such that $\omega(x)(u_1 \wedge \ldots \wedge u_d) = \overline{\omega}(x)\det(u_1, \wedge \ldots \wedge, u_d)$. Therefore,

  $$\begin{aligned}\phi^*\omega(x)(u_1 \wedge \ldots \wedge u_d) &= \overline{\omega}(\phi(x)) \det\left(d_x\phi(u_1) \wedge \ldots \wedge d_x\phi(u_d)\right)\\ &= |d_x\phi|\, \overline{\omega}(\phi(x)) \det(u_1, \wedge \ldots \wedge, u_d),\end{aligned} \tag{A.3.13}$$

  so that the scalar field associated to $\phi^*\omega$ is $|d_x\phi|\, \overline{\omega}(x)$.

Appendix B

Construction of RKHS and their dual spaces

Contents

A Hilbert space is a vector space provided with an inner-product which is topologically complete (i.e. in which every Cauchy sequence converges). These spaces play a tremendous role in almost all area of science, since they are the more natural extension of the usual Euclidean spaces $\mathbb{R}^n$. The structure of vector space and the inner-product allows us to perform standard computation in a way similar to linear algebra. Roughly speaking, Hilbert spaces give a rigorous framework to work with infinite-dimensional vectors and matrices. The completeness enables to consider such infinite-dimensional vectors as the limit of finite-dimensional vectors.

Among the functional Hilbert spaces (i.e. Hilbert spaces of functions), the reproducing kernel Hilbert spaces (RKHS) are of great practical interest. They make the framework of Hilbert spaces even more similar to the finite-dimensional Euclidean spaces. In $\mathbb{R}^n$, any symmetric definite-positive matrix K defines a metric. The inner product between two vectors X and Y is given simply by: $\langle X, Y \rangle_K = X^t K Y$. This can be seen as the Euclidean inner-product between $K^{1/2}X$ and $K^{1/2}Y$ (as if the matrix $K^{1/2}$ maps $\mathbb{R}^n$ with the metric K to the Euclidean space $\mathbb{R}^n$). Similarly, a RKHS is entirely determined by the choice of a function K, called a kernel. The squared root of the kernel (in a sense to be defined) maps the space of L^2 functions to the RKHS. Computations in the RKHS involves only standard operations with the kernel, as we shall see in this appendix. Besides the computational benefit, the framework of RKHS offers a way to adapt the metric to any particular applications, since defining a metric is equivalent to choosing a single function K.

The purpose of this appendix is to recall the basic properties of RKHS. We emphasize two important aspects. First, RKHS are build as a completion of the linear span of some

basis vectors. This gives a way for the definition of finite-dimensional approximation spaces like in Chapter 2. Second, there is a canonical isometric mapping between a RKHS and its dual space. This isometric mapping, which is directly linked to the kernel, plays a central role in the computation of standard operations on currents and vector fields. This allows us to define a whole computational framework for dealing with currents, as introduced in Chapter 2 and 3.

The material presented in this appendix results from standard mathematical constructions. We introduce here only what is needed in the framework of this thesis and we refer the reader to [Aronszajn 1950, Schwartz 1964, Saitoh 1988] for more details on the theory.

B.1 Where does it come from?

There are two different (but equivalent) ways to construct RKHS. The first way comes from the theory of differential equations: under some hypotheses, it is possible to express the solution of a differential equation as a convolution with a kernel. In this case, the space of solution is naturally a RKHS. The second construction starts choosing a kernel K and then builds a RKHS so that its kernel is given by K. In this appendix, we will present the second construction in details, since it is better suited for computational purposes. However, in this section, we will give a sketch of the construction from the point of view of the differential operators, since it gives better insights into the emergence of such spaces and their "reproducing property".

The rigorous framework of this construction is the one of the Friedrichs' extensions in functional analysis, as introduced in [Zeidler 1991]. Here, we recall simply the main steps of the construction.

Let L be a linear, self-adjoint operator which maps a space E, which is dense in the space of L^2 functions, to L^2. We suppose moreover that L is such that $\|u\|_{L^2}^2 \leq C \langle Lu, u \rangle_{L^2}$ for every $u \in L^2$. For instance, the Laplacian operator ($Lu = -\Delta u$) defined on the space E of twice differentiable functions with compact support satisfies these requirements. The differential operator L defines an inner-product on E by: $\langle u, v \rangle_E = \langle Lu, v \rangle_{L^2}$ for all functions u, v in E.

The space E provided with this inner-product is not yet a Hilbert space, since it is not topologically complete (the limit of Cauchy sequence in E may not be in E). Nevertheless, we can build the completion of E to give the Hilbert space W (one adds to E every limit of the Cauchy sequence of E), still included in L^2. In W, we have still $\langle u, v \rangle_W = \langle Lu, v \rangle_{L^2}$.

Under some assumptions (in particular that the evaluation functionals $\delta_x(u) = u(x)$ are continuous on W), one can prove that the differential operator is invertible, that L^{-1} maps the space L^2 to W and that L^{-1} admits a Green function K. This Green function satisfies, for every function $h \in L^2$:

$$L^{-1}h(x) = \int K(x,y)h(y)dy = \langle K(x,.), h \rangle_{L^2} . \tag{B.1.1}$$

Combining this equation with the definition of the inner-product in W leads to:

$$L^{-1}h(x) = \left\langle K(x,.), L^{-1}h \right\rangle_W . \tag{B.1.2}$$

This last equation shows that the function $h'(x) = L^{-1}h(x)$ in W satisfies the "reproducing property": $h'(x) = \langle K(x,.), h' \rangle_W$.

Applying this equation to the Green function K itself, called kernel is this context, leads to: $K(x,y) = \langle K(x,.), K(.,y) \rangle_W$. That's why K is called "auto-reproducing kernel".

Eventually, this construction shows that we can build a Hilbert space W of solutions of the differential equation $h = Lh'$. Such solutions satisfy the reproducing property, meaning that their evaluation on a point x is given by a convolution with a kernel K. Such Hilbert space W are then called reproducing kernel Hilbert spaces (RKHS).

This approach is usually followed in the field of fluid mechanics, for which the differential operator L is given by the laws of mechanics. Then, the Green function is defined implicitly. In our case, however, we prefer to control the kernel K which determines the metric and leaves the differential operator implicit. From a numerical point of view, it is better to write the operations on the RKHS with the kernel K which is a regularizing convolution instead of L which is a numerically unstable differential operator.

B.2 Construction of RKHS

In this section, we show how to construct a RKHS whose kernel is a given function K. First, we give rigorous definition of kernels and RKHS for scalar and vectorial functions.

B.2.1 Kernels and RKHS

Definition B.1 (auto-reproducing kernel Hilbert space (RKHS) (scalar case))**.** *Let W be a Hilbert space of scalar field on $\mathbb{R}^d$ (i.e. mapping from $\mathbb{R}^d$ to $\mathbb{R}$). W is a RKHS if the evaluation functions (linear forms on W) $\delta_x : W \to \mathbb{R}$ defined by:*

$$\delta_x(\omega) = \omega(x) \tag{B.2.1}$$

are continuous.

If W is a RKHS, then the Riesz representation theorem guarantees that for all $x \in \mathbb{R}^d$ there is a function $K_x \in W$ such that:

$$\omega(x) = \delta_x(\omega) = \langle K_x, \omega \rangle_W \tag{B.2.2}$$

We denote $K(x,y)$ the scalar mapping from $\mathbb{R}^d \times \mathbb{R}^d$ to $\mathbb{R}$: $K(x,y) = K_x(y)$. From the previous equation, we get $K_x(y) = \langle K_x, K_y \rangle_W = \langle K_y, K_x \rangle_W = K_y(x)$. This shows that K is symmetric: $K(x,y) = K(y,x)$.

Since we deal with vector fields, we give the slightly more general definition:

Definition B.2 (auto-reproducing kernel Hilbert space (RKHS) (vectorial case))**.** *Let W be a Hilbert space of mapping from $\mathbb{R}^d$ to $\mathbb{R}^p$. W is a RKHS is the evaluation functions (linear forms on W) $\delta_x^\alpha : W \to \mathbb{R}$ defined by:*

$$\delta_x^\alpha(\omega) = \omega(x)^t \alpha \tag{B.2.3}$$

are continuous for all point $x \in \mathbb{R}^d$ and all vectors $\alpha \in \mathbb{R}^p$ (i.e. each coordinates are continuous).

If W is a RKHS, then the Riesz representation theorem guarantees that for all points $x \in \mathbb{R}^d$ and all vectors $\alpha \in \mathbb{R}^p$ there is a function $K_x(\alpha) \in W$ such that:

$$\omega(x)^t \alpha = \delta_x^\alpha(\omega) = \langle K_x(\alpha), \omega \rangle_W \tag{B.2.4}$$

Applying this equation with $\alpha + \lambda\beta$ (for α, β two vectors an λ a real) shows that the mapping $\alpha \to K_x(\alpha)$ is linear. We denote therefore $K(x,y)$ the p-by-p matrix such that $K(x,y)\alpha = K_x(\alpha)(y)$ for all vectors α. Eventually, we have: $\alpha^t K(x,y)\beta = \langle K_x(\alpha), K_y(\beta) \rangle_W = \langle K_y(\beta), K_x(\alpha) \rangle_W = \beta^t K(y,x)\alpha$. This shows that $K(x,y) = K(y,x)^t$.

This shows that any RKHS contains a function K (i.e. a kernel as this will be shown in Theorem B.6) which satisfies the reproducing property in Eq. (B.2.2) and Eq. (B.2.3). The following proposition shows that this is actually a characterization of the RKHS.

Proposition B.3. *Let W be a Hilbert space which contains vector fields of the form $K(x,.)\alpha$ where K is a function from $\mathbb{R}^d \times \mathbb{R}^d$ to the space of p-by-p matrices. If every $\omega \in W$ satisfy the "reproducing property":*

$$\omega(x)^t \alpha = \langle \omega, K(.,x)\alpha \rangle_W \tag{B.2.5}$$

for all $x \in \mathbb{R}^d$ and all $\alpha \in \mathbb{R}^p$, then W is a RKHS.

Proof. Thanks to the Cauchy-Schwarz inequality, the evaluation functional verify:

$$|\delta_x^\alpha(\omega)| = \left|\omega(x)^t \alpha\right| \leq \|\omega\|_W \, \|K(.,x)\alpha\|_W \tag{B.2.6}$$

and therefore are continuous. ■

The following proposition gives an important example of RKHS, which is used to give a generic definition of the space of currents in Chapter 1.

Proposition B.4. *If W is a Hilbert space continuously embedded in the space of continuous mapping from $\mathbb{R}^d$ to $\mathbb{R}^p$ which tend to zero at infinity (i.e. such that for every $\omega \in W$, $\|\omega\|_\infty \leq C_W \|\omega\|_W$) for a fixed constant C_W, then W is a RKHS.*

Proof. If the condition is satisfied, then for every point x and vector α, $|\omega(x)^t\alpha| \leq \|\omega\|_\infty |\alpha| \leq C |\alpha| \|\omega\|_W$: the evaluation functions in Eq. (B.2.3) are continuous.

The condition means in particular that small errors measured in W are numerically small. ■

B.2.2 To each kernel its RKHS

Neither the definition of RKHS nor the propositions in the previous section give a practical way to construct RKHS. In this section, we show that, given a positive kernel K, there is a generic way to construct a RKHS whose kernel is K. The RKHS is therefore entirely determined by its kernel and we every operations in the RKHS can be written with the kernel.

First, we give the definition of positive kernel:

Definition B.5 (positive kernels)**.** *A positive definite scalar kernel K on $\mathbb{R}^d$ is a scalar function on $\mathbb{R}^d \times \mathbb{R}^d$ such that*

- $K(x,y) = K(y,x)$ *for all* $x, y \in \mathbb{R}^d$
- $\sum_{i,j} a_i K(x_i, x_j) a_j \geq 0$ *for all finite set of reals* (a_i) *and points* (x_i) *in* $\mathbb{R}^d$
- *If* $\sum_{i,j} a_i K(x_i, x_j) a_j = 0$ *when the* (x_i) *are all distinct, then all* $a_i = 0$.

If only the first two properties are satisfied, K is a positive semi-definite kernel.

A positive definite vectorial kernel K on $\mathbb{R}^d$ is a mapping $\mathbb{R}^d \times \mathbb{R}^d$ to the space of p-by-p matrix, such that

- $K(x,y) = K(y,x)^t$ *for all* $x, y \in \mathbb{R}^d$
- $\sum_{i,j} a_i^t K(x_i, x_j) a_j \geq 0$ *for all finite set of vectors* (a_i) *in* $\mathbb{R}^d$ *and points* (x_i) *in* $\mathbb{R}^d$
- *If* $\sum_{i,j} a_i^t K(x_i, x_j) a_j = 0$ *when the* (x_i) *are all distinct, then all* $a_i = 0$.

If only the first two properties are satisfied, K is a positive semi-definite kernel.

The following theorem shows that a unique RKHS corresponds to any positive kernel K. The idea is to build the vector space spanned by the vector fields of the form $K(x,.)\alpha$ and to make this space complete by adding to it the limit of every Cauchy sequence. This construction allows us in Chapter 1 to process in the same setting discrete meshes (finite linear combination of $K(x,.)\alpha$) and the continuous surfaces (limit of such finite combination).

Theorem B.6. *We have the two properties:*

- *The kernel of a RKHS is a positive semi-definite kernel,*
- *If K is a positive semi-definite kernel, then it exists a unique RKHS W such that K is its kernel.*

Proof. We prove the previous theorem in the vectorial case. It can be easily simplified to apply in the scalar case. If W is a RKHS and K its kernel, then

$$\left\| \sum_i K(.,x_i) a_i \right\|_W^2 = \sum_{i,j} a_j K(x_j, x_i) a_i \geq 0 \tag{B.2.7}$$

for all finite set of (x_i) and (α_i). K is positive semi-definite kernel.

Conversely, let K be a positive semi-definite kernel and E the vector space spanned by the function of the type $K(x,.)\alpha$ for all points x and vector α. Note that these vectors do not build a basis of E since the kernel is supposed to be only positive semi-definite. We provide E with the bilinear form defined on the $K(x,.)\alpha$ elements by: $\langle K(x,.)\alpha, K(y,.)\beta \rangle_E = \alpha^t K(x,y)\beta$. This bilinear form does not depend on the decomposition of the vectors $\omega \in E$. If a vector $\omega \in E$ has two different decompositions ω and $\tilde{\omega}$, one wants to prove that $\langle \omega, \omega' \rangle_E = \langle \tilde{\omega}, \omega' \rangle_E$. Assume that $\omega = \sum_i K(x_i,.)\alpha_i = 0$, then for any y and β, $\langle \omega, K(y,.)\beta \rangle_E = \beta^t \sum_i K(x_i, y)\alpha_i = \beta^t \omega(y) = 0$. By linearity, we get that $\langle \omega, \omega' \rangle_E = 0$ for every $\omega' \in E$.

We prove now that this bilinear form is an inner-product on E. Due to the definition of a positive kernel, this bilinear form is symmetric and positive. Let $\omega \in E$ such that

$\langle \omega, \omega \rangle_E = 0$. By linearity, the reproducing property which is satisfied for every $K(.,x)\alpha$ extends to every $\omega \in E$: $\omega(x)^t\alpha = \langle K(x,.)\alpha, \omega \rangle_E$. This implies thanks to the Cauchy-Schwarz inequality that: $|\omega(x)| = \sup_{|\alpha|=1} |\omega(x)^t\alpha| \leq \sup_{|\alpha|=1} \alpha^t K(x,x)\alpha \, \langle \omega, \omega \rangle_E = 0$. And $\omega = 0$.

E is therefore provided with an inner-product and satisfies the reproducing property. However, E is not Hilbert, since it is not complete. We build from E the space W which contains E and the limits of any Cauchy sequences of E.

Let ω_n be a Cauchy sequence in E. From the Cauchy-Schwarz inequality, we get:

$$|\omega_p(x) - \omega_q(x)| \leq \|\omega_p - \omega_q\|_E \sqrt{\sup_{|\alpha|=1} \alpha^t K(x,x)\alpha} \tag{B.2.8}$$

Therefore, $\omega_p(x)$ is a Cauchy sequence in $\mathbb{R}^d$ and hence converges. Let $\omega(x)$ be its limit. We define now W as the set of functions ω which are limits from Cauchy sequence in E: $W = \big\{\omega; \quad \exists(\omega_n) \in E(\text{Cauchy}), \forall x \in \mathbb{R}^d, \omega(x) = \lim_{n\to\infty} \omega_n(x)\big\}$. For any Cauchy sequence ω_n in E, $\|\omega_n\|_E$ is a Cauchy sequence in $\mathbb{R}$ and therefore converges. This allows us to provide W with the norm (and inner-product): $\|\omega\|_W = \lim_{n\to\infty} \|\omega_n\|_E$. Nevertheless, we have to check that this definition does not depend on the Cauchy sequence used to approximate ω. For this purpose, assume that ω_n is a Cauchy sequence in E, such that $\omega_n(x)$ converges to 0 for all x. We will prove that $\|\omega_n\|_E$ will converge to 0. Indeed, $\omega_n(x)^t\alpha = \langle \omega_n, K(x,.)\alpha \rangle_E \to 0$. By linearity, for all $\omega' \in E$, $\langle \omega_n, \omega' \rangle_E \to 0$. Then, since ω_n is a Cauchy sequence, there is an integer n such that for all $n \geq p$,

$$\|\omega_n\|_E - 2\langle \omega_p, \omega_n \rangle_E \leq \|\omega_p - \omega_n\|_E^2 \leq \varepsilon \tag{B.2.9}$$

for all $\varepsilon > 0$. Since $\langle \omega_p, \omega_n \rangle_E \to_{n\to\infty} 0$, for n large enough, $\|\omega_n\|_E \leq 2\varepsilon$ and therefore $\|\omega_n\|_E$ tends to 0.

Now, we prove that the construction of the Hilbert space W leads to a RKHS of kernel K. By definition of $\omega(x)$, we have $\omega(x)^t\alpha = \lim_{n\to\infty} \omega_n(x)^t\alpha$. We have $\omega_n(x)^t\alpha = \langle \omega_n, K(x,.)\alpha \rangle_W$ which converges to $\langle \omega, K(x,.)\alpha \rangle_W$ by definition of the norm in W. Therefore K is the kernel of the RKHS W.

We still need to prove that W is the unique RKHS whose kernel is K. If $\tilde{W}$ is a RKHS of kernel K, the every function of the type $K(x,.)\alpha$ are in $\tilde{W}$, and by linearity E is included in $\tilde{W}$. Let $\omega \in W$ as a limit of the Cauchy sequence ω_n in E. Due to the reproducing property, the inner-product $\langle .,. \rangle_W$, $\langle .,. \rangle_{\tilde{W}}$ and $\langle .,. \rangle_E$ all coincide on E. Therefore, ω_n is also a Cauchy sequence in $\tilde{W}$. This sequence converge pointwise to ω, the same limit as in W since this pointwise convergence does not depend on the Hilbert inner-product. Therefore $\omega \in \tilde{W}$ and W is a closed subset $\tilde{W}$. To prove the equality of the two spaces, we show that the orthogonal subspace of W in $\tilde{W}$ is equal to $\{0\}$. Let $\tilde{\omega} \in \tilde{W}$, such that for all $\omega \in W$, $\langle \omega, \tilde{\omega} \rangle_{\tilde{W}} = 0$. Then, $\tilde{\omega}(x)^t\alpha = \langle \tilde{\omega}, K(x,.)\alpha \rangle_{\tilde{W}} = 0$ and therefore $\tilde{\omega} = 0$. ■

A direct consequence of this proof is the following corollary:

Corollary B.7 (dense vector space in the RKHS)**.** *The span of vector fields of the form $K(x,.)\alpha$ for every $x \in \mathbb{R}^d$ and $\alpha \in \mathbb{R}^p$ is dense in W.*

This corollary offers a way to define a approximation spaces of the space W by limiting the point x to belong to a particular discrete subset (see Chapter 2).

B.2.3 Choice of the kernel

The previous theorem shows that the choice of the kernel determines the RKHS and especially its metric. The choice of this metric is therefore crucial and must be adapted to every particular problems. Here, we give some examples of parametric kernels. They are translation-invariant isotropic scalar kernels, which means of the form $K(x,y) = k(|x-y|)\mathrm{I}_p$. The following functions k lead to positive kernels, as shown in [Glaunès 2005]:

- Gaussian kernel: $k(x) = \exp\left(\frac{-x^2}{\lambda_W^2}\right)$
- Cauchy kernel: $k(x) = \left(1 + \frac{x^2}{\lambda_W^2}\right)^{-1}$
- Sobolev kernel: k is the inverse Fourier transform of $(1+x^2)^{-s}$ for $s > d + 1/2$

See [Glaunès 2005] and the theorem of Bochner for more details on translation-invariant kernels. In particular, it is shown that the Sobolev spaces $\mathbb{H}^s(\mathbb{R}^d, \mathbb{R}^m)$ are RKHS if $s > d + 1/2$.

However, how to choose the "best" kernel according a particular application is still an open question. Through the applications of chapter 6, 7 and 8, we will give some clue to adjust kernel's parameters in different context.

From now on, we consider only symmetric kernel so that we do make differences between $K(x,y)$ and $K(y,x)$.

B.3 A RKHS is isometric to its dual space

B.3.1 W^*: dual space of RKHS W

Let W be a RKHS of kernel K. We denote W^* the dual space of W (i.e. the space of continuous linear forms on W). This means that $T : W \to \mathbb{R}$ is in W^* if there is a constant C_T such that for all ω, $|T(\omega)| \leq C_T \|\omega\|_W$).

By definition of a RKHS in B.2, the evaluation functional δ_x^α are continuous linear forms on W. They belong therefore to W^*. They will play the role of Dirac delta currents in Chapter 1.

As a vector space of linear maps, W^* is provided with the operator norm:

$$\|T\|_{W^*} = \sup_{\|\omega\|_W \leq 1} |T(\omega)| \tag{B.3.1}$$

B.3.2 Isometric mapping $\mathcal{L}_W$

One of the key property of the RKHS is that there is a canonical isometric map between a RKHS W and its dual space W^*. This isometric map is used intensively throughout the thesis.

Definition B.8. *Let $\mathcal{L}_W$ be the mapping:*

$$\begin{array}{rccc} \mathcal{L}_W : & W & \longrightarrow & W^* \\ & \omega & & \mathcal{L}_W(\omega) \end{array} \tag{B.3.2}$$

where $\forall \omega' \in W, \mathcal{L}_W(\omega)(\omega') = \langle \omega, \omega' \rangle_W$. $\mathcal{L}_W(\omega)$ is continuous thanks to the Cauchy-Schwarz inequality and therefore belongs to W^*.

Proposition B.9. *$\mathcal{L}_W$ is an isometric mapping between W and W^*.*

Proof. The following equalities apply for all $\omega \in W$:

$$\begin{aligned} \|\mathcal{L}_W(\omega)\|_{W^*} &= \sup_{\|\omega'\|_W=1} |\mathcal{L}_W(\omega)(\omega')| \\ &= \sup_{\|\omega'\|_W=1} |\langle \omega, \omega' \rangle_W| = \|\omega\|_W \end{aligned} \tag{B.3.3}$$

■

This proposition shows that the operator norm on the dual space W^* (see Eq. (B.3.1)) derives from an inner-product. Indeed, the norm on W comes from the inner-product. Since, $\langle \omega, \omega' \rangle_W = (\|\omega + \omega'\|_W^2 - \|\omega - \omega'\|_W^2)/4$ and $\|\mathcal{L}_W(\omega)\|_{W^*} = \|\omega\|_W$, we have:

$$\langle T, T' \rangle_{W^*} = \left\langle \mathcal{L}_W^{-1}(T), \mathcal{L}_W^{-1}(T') \right\rangle_W \tag{B.3.4}$$

The isometric map $\mathcal{L}_W$ carries the Hilbert structure in W to W^*. This makes W^* a Hilbert space.

Moreover, let $T \in W^*$, then by definition of $\mathcal{L}_W$, the vector field $\mathcal{L}_W^{-1}(T)$ satisfies $T(\omega) = \mathcal{L}_W(\mathcal{L}_W^{-1}(T))(\omega) = \left\langle \mathcal{L}_W^{-1}(T), \omega \right\rangle_W$. Using the isometric map, we obtain these two equalities:

$$T(\omega) = \langle T, \mathcal{L}_W(\omega) \rangle_{W^*} = \left\langle \mathcal{L}_W^{-1}(T), \omega \right\rangle_W \tag{B.3.5}$$

In particular, this allows us to show that the vector field which achieves the supremum in the definition of the norm in W^* in Eq. (B.3.1) is given by $\mathcal{L}_W^{-1}(T) / \left\|\mathcal{L}_W^{-1}(T)\right\|_W$. Indeed we have:

$$\begin{aligned} \|T\|_{W^*} &= \sup_{\|\omega\|_W=1} |T(\omega)| \\ &= \sup_{\|\omega\|_W=1} \left|\left\langle \mathcal{L}_W^{-1}(T), \omega \right\rangle_W\right| \end{aligned} \tag{B.3.6}$$

whose supremum is achieved for $\omega = \pm \mathcal{L}_W^{-1}(T) \left\|\mathcal{L}_W^{-1}(T)\right\|_W$.

In Eq. (B.3.5), we write $T(\omega)$ via the map $\mathcal{L}_W$. Actually, any operations on W and W^* can be expressed using this map. In particular, the inner-product in these two spaces are given as:

$$\begin{aligned} \langle \omega, \omega' \rangle_W &= \mathcal{L}_W(\omega)(\omega') \\ \langle T, T' \rangle_{W^*} &= T\left(\mathcal{L}_W^{-1}(T')\right) \end{aligned} \tag{B.3.7}$$

The first equality is a direct consequence of Definition B.8. The second one results from the application of Eq. (B.3.5) with $\omega = \mathcal{L}_W^{-1}(T')$.

B.3.3 Link between $\mathcal{L}_W$ and the kernel

We have just shown that the metric on W and W^* can be expressed via the isometric map $\mathcal{L}_W$. Actually, this isometric map is closely related to the kernel K of the RKHS W. This allows us to express the metric in terms of operations with the kernel. Therefore, once the kernel is chosen, any operations in the RKHS will have a closed form.

First, we apply the previous equations in the particular case when $T \in W^*$ is an evaluation functional δ_x^α, as defined in Eq. (B.2.2). For every $\omega \in W$, we have by application of Eq. (B.3.5): $\delta_x^\alpha(\omega) = \left\langle \mathcal{L}_W^{-1}(\delta_x^\alpha), \omega \right\rangle_W$. Moreover, thanks to the reproducing property satisfied in the RKHS W, we have: $\delta_x^\alpha(\omega) = \omega(x)^t\alpha = \langle \omega, K(x,.)\alpha \rangle_W$. This proves that:

$$\mathcal{L}_W^{-1}(\delta_x^\alpha) = K(x,.)\alpha \tag{B.3.8}$$

This equation shows that the kernel K may be seen as the Green function of the mapping $\mathcal{L}_W$ (which is implicitly a differential operator).

The application of Eq. (B.3.4) and the reproducing property leads to the explicit computation of the inner-product between evaluation functionals:

$$\left\langle \delta_x^\alpha, \delta_y^\beta \right\rangle_{W^*} = \langle K(.,x)\alpha, K(.,y)\beta \rangle_W = \alpha^t K(x,y)\beta \tag{B.3.9}$$

By linearity, the inner-product between $T = \sum_{i=1}^n \delta_{x_i}^{\alpha_i}$ and $U = \sum_{j=1}^m \delta_{y_j}^{\beta_j}$ is given by:

$$\langle T, U \rangle_{W^*} = \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_i^t K(x_i, y_j)\beta_j \tag{B.3.10}$$

This equation may be written in a matrix form:

$$\langle T, U \rangle_{W^*} = \boldsymbol{\alpha}^t \mathbf{K} \boldsymbol{\beta}, \tag{B.3.11}$$

where $\boldsymbol{\alpha}$ (resp. $\boldsymbol{\beta}$) denotes the nd (resp. md) dimensional vector obtained by the concatenation of every vectors α_i (resp. β_j). $\mathbf{K}$ denotes the nd-by-md block matrix whose block (i,j) is given by the d-by-d matrix $K(x_i, y_j)$ (for $i = 1, \ldots, n$ and $j = 1, \ldots, m$). This shows that the map $\mathcal{L}_W^{-1}$ is computed via the matrix $\mathbf{K}$ when applied to finite linear combination of evaluation functionals.

This way to compute the metric on W^* in a matrix form involving only the kernel K is the core of the numerical framework for computing with currents, as introduced in Chapter 2 and 3. Indeed, by construction of the RKHS the span of the functions $K(x,.)\alpha$ is dense in W (see Corollary B.7). By isometry, the span of the evaluation functionals δ_x^α is a dense vector space in W^*. This means that we can always approximate a current in W^* as a finite linear combination of evaluation functionals and use this matrix form to compute the metric in W^*. The true map $\mathcal{L}_W^{-1}$ can be considered then as a multiplication with the matrix $\mathbf{K}$ whose dimensions tend to infinity.

Bibliography

[Aljabar 2008] P. Aljabar, K.K. Bhatia, M. Murgasova, J.V. Hajnal, J.P. Boardman, L. Srinivasan, M.A. Rutherford, L.E. Dyet, A.D. Edwards and D. Rueckert. *Assessment of brain growth in early childhood using deformation-based morphometry.* NeuroImage, vol. 39, no. 1, pages 348 – 358, 2008. 249, 253, 255

[Allassonnière 2007] S. Allassonnière, Y. Amit and A. Trouvé. *Towards a coherent statistical framework for dense deformable template estimation.* Journal of the Royal Statistical Society Series B, vol. 69, no. 1, pages 3–29, 2007. 168, 171, 192, 194, 222, 223, 234, 235

[Allassonnière 2008] S. Allassonnière, E. Kuhn and A. Trouvé. *Construction of Bayesian Deformable Models Via a Stochastic Approximation Algorithm: A Convergence Study.* Bernoulli Journal, 2008. In revision. 168, 192

[Allassonnière 2009] S. Allassonnière and E. Kuhn. *Stochastic Algorithm For Bayesian Mixture Effect Template Estimation.* ESAIM Probability and Statistics, 2009. In Press. 170, 172, 192, 193

[Aronszajn 1950] N. Aronszajn. *Theory of Reproducing Kernels.* Transactions of the American Mathematical Society, no. 68, pages 337–404, 1950. 202, 308

[Arsigny 2005] Vincent Arsigny, Xavier Pennec and Nicholas Ayache. *Polyrigid and Polyaffine Transformations: a Novel Geometrical Tool to Deal with Non-Rigid Deformations - Application to the registration of histological slices.* Medical Image Analysis, vol. 9, no. 6, pages 507–523, 2005. 122

[Arsigny 2006] Vincent Arsigny, Pierre Fillard, Xavier Pennec and Nicholas Ayache. *Log-Euclidean Metrics for Fast and Simple Calculus on Diffusion Tensors.* Magnetic Resonance in Medicine, vol. 56, no. 2, pages 411–421, 2006. 210

[Ashburner 1998] John Ashburner, Chloe Hutton, Richard Frackowiak, Ingrid Johnsrude, Cathy Price and Karl Friston. *Identifying global anatomical differences : deformation-based morphometry.* Human Brain Mapping, vol. 6, no. 5-6, pages 348–357, 1998. 199

[Ashburner 1999] J. Ashburner, J. Andersson and K. Friston. *High-dimensional Image Registration using Symmetric Priors.* NeuroImage, vol. 9, pages 619–628, 1999. 118

[Ashburner 2000] John Ashburner and Karl J. Friston. *Voxel-Based Morphometry–The Methods.* NeuroImage, vol. 11, no. 6, pages 805 – 821, 2000. 165

[Ashburner 2001] John Ashburner and Karl J. Friston. *Why Voxel-Based Morphometry Should Be Used.* NeuroImage, vol. 14, no. 6, pages 1238 – 1243, 2001. 165

[Ashburner 2003] J. Ashburner and K.J. Friston. *Morphometry*. In R.S.J. Frackowiak, K.J. Friston, C. Frith, R. Dolan, K.J. Friston, C.J. Price, S. Zeki, J. Ashburner and W.D. Penny, editors, Human Brain Function. Academic Press, 2nd édition, 2003. 199

[Ashburner 2007] J. Ashburner. *A fast diffeomorphic image registration algorithm*. NeuroImage, vol. 38(1), pages 95–113, 2007. 118

[Ashburner 2009] J. Ashburner and K.J. Friston. *Computing average shaped tissue probability templates*. NeuroImage, vol. 45(2), pages 333–341, 2009. 193

[Auzias 2008] Guillaume Auzias, Joan-Alexis Glaunès, Arnaud Cachia, Pascal Cathier, Eric Bardinet, Olivier Colliot, J. F. Mangin, Alain Trouvé and Sylvain Baillet. *Multi-Scale Diffeomorphic Cortical Registration Under Manifold Sulcal Constraints*. In IEEE International Symposium on Biomedical Imaging (ISBI), Nano to Macro, pages 1127–1130, 2008. 216

[Avants 2004] Brian Avants and James Gee. *Geodesic estimation for large deformation anatomical shape averaging and interpolation*. NeuroImage, vol. 23, pages 139–150, 2004. 118, 164

[Avants 2006] Brian B. Avants, C. L. Epstein and James C. Gee. *Geodesic Image Normalization and Temporal Parameterization in the Space of Diffeomorphisms*. In Medical Imaging and Augmented Reality, volume 4091 of *Lecture Notes in Computer Science*, pages 9–16. Springer, 2006. 199

[Batchelor 2006] P. G. Batchelor, F. Calamante, J.-D. Tournier, D. Atkinson, D. L. G. Hill and A. Connelly. *Quantification of the shape of fiber tracts*. MRM, vol. 55, no. 4, pages 894–903, 2006. 218

[Betke 2003] Margrit Betke, Harrison Hong, Deborah Thomas, Chekema Prince and Jane P Ko. *Landmark detection in the chest and registration of lung surfaces with an application to nodule registration*. Medical Image Analysis, vol. 7, no. 3, pages 265–281, 2003. 148

[Bhattacharya 2003] Rabi Bhattacharya and Vic Patrangenaru. *Large sample of theory of intrinsic and extrinsic sample means on manifolds*. Annals of Statistics, vol. 31, no. 1, pages 1–29, 2003. 208

[Blanchard 2008] Gilles Blanchard and Laurent Zwald. *Finite Dimensional Projection for Classification and Statistical Learning*. IEEE transactions on information theory, vol. 54, no. 9, pages 4169–4182, 2008. 195

[Bookstein 1991] Fred L. Bookstein. Morphological tools for landmark data; geometry and biology. Cambridge University press, 1991. 147

[Cachier 2001] Pascal Cachier, Jean-Francois Mangin, Xavier Pennec, Denis Rivière, Dimitri Papadopoulos-Orfanos, Jean Régis and Nicholas Ayache. *Multisubject Non-rigid Registration of Brain MRI Using Intensity and Geometric Features*. In W.J. Niessen and M.A. Viergever, editors, 4th Int. Conf. on Medical Image Computing

and Computer-Assisted Intervention (MICCAI'01), volume 2208 of *Lecture Notes in Computer Science*, pages 734–742, 2001. 200, 201

[Cao 2005] Yan Cao, Michael I Miller, Raimond L. Winslow and Laurent Younes. *Large Deformation Diffeomorphic Metric Mapping of Vector Fields*. IEEE Transactions on Medical Imaging, vol. 24, no. 9, 2005. 218

[Cartan 1970] Henri Cartan. *Les travaux de Georges de Rham sur les variétés différentiables*. Essays on Topology and Related Topics, pages 1–70, 1970. 20

[Castillo 2009] Richard Castillo, Edward Castillo, Rudy Guerra, Valen E Johnson, Travis McPhail, Amit K Garg and Thomas Guerrero. *A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets*. Physics in Medicine and Biology, vol. 54, no. 7, pages 1849–1870, 2009. 2, 150

[Cathier 2006] P. Cathier and J.-F. Mangin. *Registration of cortical connectivity matrices*. In IEEE Computer Society Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA), 2006. 216

[Chandrashekara 2003] R. Chandrashekara, A. Rao, G.Ivar Sanchez-Ortiz, R. H. Mohiaddin and D. Rueckert. *Construction of a Statistical Model for Cardiac Motion Analysis Using Nonrigid Image Registration*. In Information Processing in Medical Imaging, volume 2732 of *Lecture Notes in Computer Science*, pages 599–610. Springer, 2003. 248, 255

[Charpiat 2005] Guillaume Charpiat, Olivier D. Faugeras and Renaud Keriven. *Approximations of Shape Metrics and Application to Shape Warping and Empirical Shape Statistics*. Foundations of Computational Mathematics, vol. 5, no. 1, pages 1–58, 2005. 11

[Christensen 1996] G. Christensen, R. Rabbit and M. Miller. *Deformable Templates Using Large Deformation Kinematics*. IEEE Transactions on Image Processing, vol. 5, no. 10, pages 1435–1447, 1996. 118

[Chui 2003] Haili Chui and Anand Rangarajan. *A new point matching algorithm for nonrigid registration*. Computer Vision and Image Understanding, vol. 89, no. 2-3, pages 114–141, 2003. 11, 199, 201

[Chui 2004] H. Chui, A. Rangarajan, J. Zhang and C.M. Leonard. *Unsupervised learning of an Atlas from unlabeled point-sets*. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 26, no. 2, pages 160–172, 2004. 164

[Cohen-Steiner 2003a] David Cohen-Steiner and Jean-Marie Morvan. *Approximation of Normal Cycles*. Technical report 4723, INRIA, 2003. 21, 45

[Cohen-Steiner 2003b] David Cohen-Steiner and Jean-Marie Morvan. *Approximation of the Curvature Measures of a Smooth Surface endowed with a Mesh*. Technical report 4867, INRIA, 2003. 21

[Cootes 1995] T.F. Cootes, C.J. Taylor, D.H. Cooper and J. Graham. *Active shape models: their training and application.* Computer Vision and Image Understanding, vol. 61, pages 38–59, 1995. 11, 165

[Cootes 2008] T.F. Cootes, C.J. Twining, K.O. Babalola and C.J. Taylor. *Diffeomorphic statistical shape models.* Image and Vision Computing, vol. 26, no. 3, pages 326–333, 2008. 11

[Corouge 2006] Isabelle Corouge, P.Thomas Fletcher, Sarang Joshi, Sylvain Gouttard and Guido Gerig. *Fiber Tract-Oriented Statistics for Quantitative Diffusion Tensor MRI Analysis.* Medical Image Analysis, no. 10, pages 786–798, 2006. 218, 223

[Cotter 2008] Colin J. Cotter. *The variational particle-mesh method for matching curves.* Journal of Physics A: Mathematical and Theoretical, vol. 41, no. 34, page 344003 (18pp), 2008. 57

[Cotter 2009] Colin J. Cotter and Darryl .D. Holm. Discrete momentum maps for lattice EPDiff, pages 247–278. 2009. 57

[Craene 2009] Mathieu De Craene, Oscar Camara, Bart H. Bijnens and Alejandro F. Frangi. *Large Diffeomorphic FFD Registration for Motion and Strain Quantification from 3D-US Sequences.* In Proc. of Functional Imaging and Modeling of the Heart, volume 5528, pages 437–446. LNCS, 2009. 247

[Cremers 2006] D. Cremers. *Dynamical Statistical Shape Priors for Level Set-Based Tracking.* Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 28, no. 8, pages 1262–1273, August 2006. 11

[Csernansky 1998] J.G. Csernansky, S. Joshi, L. Wang, J.W. Haller, M. Gado, J.P. Miller, U. Grenander and M.I. Miller. *Hippocampal Morphometry in Schizophrenia by High Dimensional Brain Mapping.* In Proceedings of the National Academy of Science, volume 95, pages 11406–11411, 1998. 165

[Davies 2002] R.H. Davies, C.J. Twining, R.F. Cootes, J.C. Waterton and C.J. Taylor. *A minimum description length approach to statistical shape modeling.* IEEE Transactions on Medical Imaging, vol. 21, no. 5, page 525, 2002. 165

[Davies 2008] Rhodri Davies, Carole Twining and Chris Taylor. Statistical models of shape, optimisation and evaluation. Springer, 2008. 165

[Davis 1997] G.M. Davis, S. Mallat and M. Avellaneda. *Greedy adaptive approximations.* Journal of Constructive Approximation, vol. 13, no. 1, pages 57–98, 1997. 90

[Davis 2007] B.C. Davis, P.T. Fletcher, E. Bullitt and S. Joshi. *Population Shape Regression From Random Design Data.* In Proc. of ICCV, pages 1–7, Oct. 2007. 248

[de Waal 1995] Franz B. M. de Waal. *Bonobo sex and society.* Scientific American, vol. 272, pages 82–88, 1995. 155

[Declerck 1998] J. Declerck, J. Feldman and N. Ayache. *Definition of a 4D continuous planispheric transformation for the tracking and the analysis of LV motion.* Medical Image Analysis, vol. 4, no. 1, pages 1–17, 1998. 249

[do Carmo 1994] Manfredo do Carmo. Differential forms and applications. Springer-Verlag, 1994. 297

[Donoho 2006] D. L. Donoho, I. Drori, Y. Tsaig and J. L. Starck. *Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit.* Technical report, 2006. 89

[Duchesnay 2007] E. Duchesnay, A. Cachia, A. Roche, D. Rivière, Y. Cointepas, D. Papadopoulos-Orfanos, M. Zilbovicius, J.-L. Martinot and J.-F. Mangin. *Classification from cortical folding patterns.* IEEE Transactions on Medical Imaging, vol. 26, no. 4, pages 553–565, 2007. 200, 216

[Dupuis 1998] P Dupuis, U Grenander and M Miller. *Variational problems on flows of diffeomorphisms for image matching.* Quaterly of Applied Mathematics, vol. 56, no. 3, pages 587–600, 1998. 118, 121, 123, 199

[Durrleman 2007] Stanley Durrleman, Xavier Pennec, Alain Trouvé and Nicholas Ayache. *Measuring Brain Variability via Sulcal Lines Registration: a Diffeomorphic Approach.* In Nicholas Ayache, Sébastien Ourselin and Anthony Maeder, editors, Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI), volume 4791 of *Lecture Notes in Computer Science*, pages 675–682. Springer, 2007. 106, 199, 200, 293

[Durrleman 2008a] Stanley Durrleman, Xavier Pennec, Alain Trouvé and Nicholas Ayache. *A Forward Model to Build Unbiased Atlases from Curves and Surfaces.* In X. Pennec and S. Joshi, editors, Proc. of the International Workshop on the Mathematical Foundations of Computational Anatomy (MFCA-2008), September 2008. 163, 219, 222, 234, 235, 294

[Durrleman 2008b] Stanley Durrleman, Xavier Pennec, Alain Trouvé and Nicholas Ayache. *Sparse Approximation of Currents for Statistics on Curves and Surfaces.* In Dimitris Metaxas, Leon Axel, Gábor Székely and Gabor Fichtinger, editors, Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI), Part II, volume 5242 of *LNCS*, pages 390–398. Springer, 2008. 87, 199, 201, 219, 222, 223, 227, 293

[Durrleman 2008c] Stanley Durrleman, Xavier Pennec, Alain Trouvé, Paul Thompson and Nicholas Ayache. *Inferring brain variability from diffeomorphic deformations of currents: an integrative approach.* Medical Image Analysis, vol. 12/5, no. 12, pages 626–637, 2008. 147, 197, 218, 293, 294

[Durrleman 2009a] Stanley Durrleman, Pierre Fillard, Xavier Pennec, Alain Trouvé and Nicholas Ayache. *A Statistical Model of White Matter Fiber Bundles based on Currents.* In Jerry L. Prince, Dzung L. Pham and Kyle J. Myers, editors, Proc. of Information Processing in Medical Imaging (IPMI), LNCS. Springer, 2009. 217, 293

[Durrleman 2009b] Stanley Durrleman, Xavier Pennec, Guido Gerig, Alain Trouvé and Nicholas Ayache. *Spatiotemporal Atlas Estimation for Developmental Delay Detection in Longitudinal Datasets*. Research Report RR-6952, INRIA, 2009. 117, 245

[Durrleman 2009c] Stanley Durrleman, Xavier Pennec, Alain Trouvé and Nicholas Ayache. *Statistical models of sets of curves and surfaces based on currents*. Medical Image Analysis, vol. 13, no. 5, pages 793–808, 2009. 87, 163, 218, 219, 222, 227, 293

[Durrleman 2009d] Stanley Durrleman, Xavier Pennec, Alain Trouvé, Guido Gerig and Nicholas Ayache. *Spatiotemporal Atlas Estimation for Developmental Delay Detection in Longitudinal Datasets*. In Guang-Zhong Yang, David Hawkes, Daniel Rueckert, Alison Noble and Chris Taylor, editors, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009, volume 5761 of *Lecture Notes in Computer Science*, pages 297–304. Springer, 2009. 117, 245, 293

[Ehrhardt 2008] J. Ehrhardt, R. Werner, A. Schmidt-Richberg, B. Schulz and H. Handels. *Generation of a Mean Motion Model of the Lung Using 4D-CT Image Data*. In Proc. of Eurographics Workshop on Visual Computing for Biomedicine, pages 69–76. Eurographics Association, 2008. 248, 253, 255

[El Kouby 2005] V. El Kouby, Y. Cointepas, C. Poupon, D. Rivière, N. Golestani, C. Pallier, J.-B. Poline, D. Le Bihan and J.-F. Mangin. *MR Diffusion-Based Inference of a Fiber Bundle Model from a Population of Subjects*. In MICCAI, volume 3749 of *LNCS*, pages 196–204. Springer-Verlag, 2005. 231

[Federer 1969] Herbert Federer. Geometric measure theory. Springer-Verlag, 1969. 21, 45

[Fillard 2007a] P. Fillard, X. Pennec, P.M. Thompson and N. Ayache. *Evaluating Brain Anatomical Correlations via Canonical Correlation Analysis of Sulcal Lines*. In Proc. of MICCAI'07 Workshop on Statistical Registration: Pair-wise and Group-wise Alignment and Atlas Formation, Brisbane, Australia, 2007. 216

[Fillard 2007b] Pierre Fillard, Vincent Arsigny, Xavier Pennec and Nicholas Ayache. *Clinical DT-MRI Estimation, Smoothing and Fiber Tracking with Log-Euclidean Metrics*. IEEE Trans. on Medical Imaging, vol. 26, no. 11, pages 1472–1482, 2007. 1, 223

[Fillard 2007c] Pierre Fillard, Vincent Arsigny, Xavier Pennec, Kiralee M. Hayashi, Paul M. Thompson and Nicholas Ayache. *Measuring Brain Variability by Extrapolating Sparse Tensor Fields Measured on Sulcal Lines*. NeuroImage, vol. 34, no. 2, pages 639–650, January 2007. 105, 199, 200, 203, 204, 206, 207, 208, 209, 210, 211, 218

[Fischl 2001] B. Fischl, A. Liu and A.M. Dale. *Automated Manifold Surgery: Constructing Geometrically Accurate and Topologically Correct Models of the Human Cerebral Cortex*. I.E.E.E. Transactions in Medical Imaging, vol. 20, no. 1, pages 70–80, 2001. 198

[Fischl 2004] B. Fischl, A. van der Kouwe, C. Destrieux, E. Halgren, F. Ségonne, D.H. Salat, E. Busa, L.J. Seidman, J. Goldstein, D. Kennedy, V. Caviness, N. Makris,

B. Rosen and A.M. Dale. *Automatically parcellating the human cerebral cortex.* Cerebral Cortex, vol. 14, no. 1, pages 11–22, 2004. 215

[Fletcher 2004] P.T. Fletcher, C. Lu, S.M. Pizer and S.C. Joshi. *Principal Geodesic Analysis for the Study of Nonlinear Statistics of Shape.* Transactions on Medical Imaging, vol. 23, no. 8, pages 995–1005, 2004. 11

[Frank 2002] J. Frank, G. Gottwald and S. Reich. *A Hamiltonian particle-mesh method for the rotating shallow-water equations.* Lecture Notes in Computational Science and Engineering, vol. 26, pages 131–142, 2002. 57

[Frank 2003] Jason Frank and Sebastian Reich. *Conservation Properties of Smoothed Particle Hydrodynamics Applied to the Shallow Water Equation.* BIT Numerical Mathematics, vol. 43, no. 1, 2003. 57

[Friedman 1981] J.H. Friedman and W. Stuetzle. *Projection pursuit regression.* Journal of American Statistical Association, vol. 76, 1981. 89

[Gaser 1999] Christian Gaser, Hans-Peter Volz, Stefan Kiebel, Stefan Riehemann and Heinrich Sauer. *Detecting Structural Changes in Whole Brain Based on Nonlinear Deformations - Application to Schizophrenia Research.* NeuroImage, vol. 10, no. 2, pages 107 – 113, 1999. 165

[Gerig 2006] G. Gerig, B. Davis, P. Lorenzen, Shun Xu, M. Jomier, J. Piven and S. Joshi. *Computational Anatomy to Assess Longitudinal Trajectory of Brain Growth.* 3D Data Processing, Visualization, and Transmission, Third International Symposium on, pages 1041–1047, 2006. 249, 253, 255

[Geva 2006] T. Geva. *Indications and timing of pulmonary valve replacement after tetralogy of Fallot repair.* In Seminars in Thoracic and Cardiovascular Surgery: Pediatric Cardiac Surgery Annual, volume 9, pages 11–22. Elsevier, 2006. 234, 240

[Glasbey 2001] C. A. Glasbey and K. V. Mardia. *A penalised likelihood approach to image warping.* Journal of the Royal Statistical Society, Series B, vol. 63, pages 465–492, 2001. 169

[Glaunès 2005] J. Glaunès. *Transport par difféomorphismes de points, de mesures et de courants pour la comparaison de formes et l'anatomie numérique.* PhD thesis, Université Paris 13, http://cis.jhu.edu/ joan/TheseGlaunes.pdf, September 2005. 11, 13, 16, 21, 32, 56, 84, 85, 86, 117, 124, 125, 126, 127, 129, 131, 134, 144, 177, 192, 201, 202, 204, 287, 313

[Glaunès 2006] Joan Glaunès and Sarang Joshi. *Template estimation from unlabeled point set data and surfaces for Computational Anatomy.* In X. Pennec and S. Joshi, editors, Proc. of the International Workshop on the Mathematical Foundations of Computational Anatomy (MFCA-2006), 2006. 169, 199, 203

[Glaunès 2008] Joan Glaunès, Anqi Qiu, M.I. Miller and Laurent Younes. *Large Deformation Diffeomorphic Metric Curve Mapping.* International Journal of Computer

Vision, vol. 80, no. 3, pages 317–336, 2008. 13, 32, 84, 117, 119, 134, 138, 145, 147, 219

[Gogtay 2008] N. Gogtay, A. Lu, A.D. Leow, A.D. Klunder, A.D. Lee, A. Chavez, D. Greenstein, J.N. Giedd, A.W. Toga, J.L. Rapoport and P.M. Thompson. *3D Growth Pattern Abnormalities Visualized in Childhood-Onset Schizophrenia using Tensor-Based Morphometry.* Proceedings of the National Academy of Sciences, 2008. 248, 255

[Good 2001] Catriona D. Good, Ingrid Johnsrude, John Ashburner, Richard N. A. Henson, Karl J. Friston and Richard S. J. Frackowiak. *Cerebral Asymmetry and the Effects of Sex and Handedness on Brain Structure: A Voxel-Based Morphometric Analysis of 465 Normal Adult Human Brains.* NeuroImage, vol. 14, no. 3, pages 685–700, 2001. 165

[Goodlett 2008] C.B. Goodlett, P.T. Fletcher, J.H. Gilmore and G. Gerig. *Group statistics of DTI fiber bundles using spatial functions of tensor measures.* In Proc. of MICCAI, volume 3750 of *LNCS*, pages 1068–75, 2008. 218

[Gorbunova 2008] Vladlena Gorbunova, Pechin Lo, Haseem Ashraf, Asger Dirksen, Mads Nielsen and Marleen de Bruijne. *Weight preserving image registration for monitoring disease progression in lung CT.* In Dimitris Metaxas, Leon Axel, Gábor Székely and Gabor Fichtinger, editors, Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI), volume 5242 of *Lecture Notes in Computer Science*, pages 863–870, 2008. 151

[Gorbunova 2009] Vladlena Gorbunova, Stanley Durrleman, Pechin Lo, Xavier Pennec and Marleen de Bruijne. *Curve- and Surface-based Registration of Lung CT images via Currents.* In The Second International Workshop on Pulmonary Image Analysis, 2009. 117, 146, 294

[Gorczowski 2007] Kevin Gorczowski, Martin Styner, Ja-Yeon Jeong, J. S. Marron, Joseph Piven, Heather Cody Hazlett, Stephen M. Pizer and Guido Gerig. *Statistical Shape Analysis of Multi-Object Complexes.* In Computer Vision and Pattern Recognition CVPR, pages 1–8. IEEE, 2007. 200

[Goualher 2000] G. Le Goualher, A.M. Argenti, M. Duyme, W.F.C. Baare, H.E. Hulshoff Pol, C. Barillot and A.C. Evans. *Statistical Sulcal Shape Comparisons: Application to the Detection of Genetic Encoding of the Central Sulcus Shape.* NeuroImage, vol. 11, no. 5, pages 564–574, 2000. 215

[Granger 2002] Sébastien Granger and Xavier Pennec. *Multi-scale EM-ICP: A Fast and Robust Approach for Surface Registration.* In A. Heyden, G. Sparr, M. Nielsen and P. Johansen, editors, European Conference on Computer Vision (ECCV 2002), volume 2353 of *Lecture Notes in Computer Science*, pages 418–432. Springer, 2002. 201

[Greengard 1991] L. Greengard and J. Strain. *The fast Gauss transform.* SIAM Journal on Scientific and Statistical Computing, vol. 12, no. 1, pages 79–94, 1991. 84

[Grenander 1994] U. Grenander. General pattern theory: a mathematical theory of regular structures. Oxford University Press, 1994. 8, 12, 165, 203

[Grenander 1998] U Grenander and I Miller M. *Computational Anatomy: An Emerging Discipline*. Quarterly of Applied Mathematics, vol. LVI, no. 4, pages 617–694, 1998. 118, 199, 203

[Guéziec 1994] A. Guéziec and N. Ayache. *Smoothing and Matching of 3-D Space Curves*. The International Journal of Computer Vision, vol. 12, no. 1, pages 79–104, 1994. 215

[Guimond 2000] A. Guimond, J. Meunier and J.-P. Thirion. *Average Brain Models: A Convergence Study*. Computer Vision and Image Understanding, vol. 77, no. 2, pages 192–210, 2000. 234

[Habas 2009] Piotr A. Habas, Kio Kim, François Rousseau, Orit A. Glenn, A. James Barkovich and Colin Studholme. *A Spatio-temporal atlas of the human fetal brain with application to tissue segmentation*. In Proc. of Medical Image Computing and Computer-Assisted Intervention (MICCAI), volume 5761 of *LNCS*, pages 289–296. Springer, 2009. 253, 255

[Hall 2005] Peter Hall, J. S. Marron and Amnon Neeman. *Geometric representation of high dimension, low sample size data*. Journal Of The Royal Statistical Society Series B, vol. 67, no. 3, pages 427–444, 2005. 195

[Hamant 2008] Olivier Hamant, Marcus G. Heisler, Henrik JÃ¶nsson, Pawel Krupinski, Magalie Uyttewaal, Plamen Bokov, Francis Corson, Patrik Sahlin, Arezki Boudaoud, Elliot M. Meyerowitz, Yves Couder and Jan Traas. *Developmental Patterning by Mechanical Signals in Arabidopsis*. Science, vol. 322, no. 5908, pages 1650–1655, 2008. 5

[Hamilton 2007] L.S. Hamilton, K.L. Narr, E. Luders, P.R. Szeszko, P.M. Thompson, R.M. Bilder and A.W. Toga. *Asymmetries of cortical thickness: Effects of handedness, sex, and schizophrenia*. NeuroReport, vol. 18, no. 14, pages 1427–1431, 2007. 209

[Hazlett 2005] H.C. Hazlett, M. Poe, G. Gerig, R.G. Smith, J. Provenzale, A. Ross, J.H. Gilmore and J. Piven. *Magnetic resonance imaging and head circumference study of brain size in autism*. The Archives of General Psychiatry, vol. 62, pages 1366–1376, 2005. 1, 105, 198, 272

[Hellier 2003] P. Hellier and C. Barillot. *Coupling dense and landmark-based approaches for non rigid registration*. I.E.E.E. Transactions on Medical Imaging, vol. 22, no. 2, pages 217–227, 2003. 200

[Hiriart-Urruty 1996] Jean-Baptiste Hiriart-Urruty and Claude Lemarechal. Convex analysis and minimization algorithms, part i. Springer, 1996. Second edition. 277

[Hodgkin 1952] A.L. Hodgkin and A.F. Huxley. *A quantitative description of membrane current and its application to conduction and excitation in nerve*. The journal of physiology, vol. 117, no. 4, pages 500–544, 1952. 7

[Huber 1985] P.J. Huber. *Projection pursuit.* The Annals of Statistics, vol. 13, pages 435–525, 1985. 89

[Hufnagel 2008] H. Hufnagel, X. Pennec, J. Ehrhardt, N. Ayache and H. Handels. *Generation of a Statistical Shape Model with Probabilistic Point Correspondences and EM-ICP.* International Journal for Computer Assisted Radiology and Surgery (IJCARS), vol. 2, no. 5, pages 265–273, 2008. 11, 170

[Hunter 2003] Peter J. Hunter and T. Borg. *Integration from Proteins to Organs: The Physiome Project.* vol. 4, no. 3, pages 237–243, 2003. 7

[iso2mesh] iso2mesh. *A Matlab/Octave-based mesh generator. http://iso2mesh.sf.net.* 148, 150

[Jian 2005] B. Jian and BC Vemuri. *A robust algorithm for point set registration using mixture of Gaussians.* In Proc. ICCV 2005, volume 2, pages 1246–1251, 2005. 237

[Johnson 2002] H. J. Johnson and G. E. Christensen. *Consistent Landmark and Intensity-Based Image Registration.* IEEE Transactions on Medical Imaging, vol. 21, pages 450–461, 2002. 147

[Jones 1987] Lee. K. Jones. *On a conjecture of Huber concerning the convergence of projection pursuit regression.* The Annals of Statistics, vol. 15, no. 2, pages 880–882, 1987. 89

[Joshi 2000] Sarang Joshi and Michael Miller. *Landmark Matching via Large Deformation Diffeomorphisms.* IEEE Transaction on Image Processing, vol. 9, no. 8, pages 1357–1370, 2000. 123, 147, 201

[Joshi 2004] Sarang Joshi, Brad Davis, Matthieu Jomier and Guido Gerig. *Unbiased diffeomorphic atlas construction for computational anatomy.* NeuroImage, vol. 23, pages 151–160, 2004. 164, 234

[Joshi 2007] Shantanu H. Joshi, Eric Klassen, Anuj Srivastava and Ian Jermyn. *A Novel Representation for Riemannian Analysis of Elastic Curves in R^n.* In Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2007. 11

[Kano 1992] T. Kano. *The last ape: Pygmy chimpanzee behavior and ecology.* Stanford University Press, 1992. 155

[Kendall 1989] David D. Kendall. *A Survey of the Statistical Theory of Shapes.* Statistical Science, vol. 4, no. 2, pages 87–89, 1989. 12, 165

[Khan 2008] A.R. Khan and M.F. Beg. *Representation of time-varying shapes in the large deformation diffeomorphic framework.* pages 1521–1524, May 2008. 248

[Kinzey 1984] W. G. Kinzey. The dentition of the pygmy chimpanzee, pan paniscus, pages 65–88. Plenum (New York), 1984. 155

[Kuroda 1989] S. Kuroda. Developmental retardation and behavioural characteristics of pygmy chimpanzees, pages 184–193. Harvard University Press, 1989. 155

[Lang 1962] Serge Lang. Introduction to differential manifolds. Interscience, 1962. 297

[Leow 2005] A. Leow, C.L. Yu, S.J. Lee, S.C. Huang, R. Nicolson, K.M. Hayashi, H. Protas, A.W. Toga and P.M. Thompson. *Brain Structural Mapping using a Novel Hybrid Implicit/Explicit Framework based on the Level-Set Method.* NeuroImage, vol. 24, no. 3, pages 910–927, 2005. 215

[Leventon 2000] M. Leventon, O. Faugeras, E. Grimson and W. Wells. *Level Set Based Segmentation with Intensity and Curvature Priors.* In Mathematical Methods in Biomedical Image Analysis, MMBIA 2000, 06 2000. 11

[Leventon 2003] M. Leventon, E. Grimson, O. Faugeras, S. Wells and R. Kikinis. *Knowledge-based segmentation of medical images.* Geometric Level Set Methods in Imaging, Vision, and Graphics, 2003. 11

[Levine 2000] H.A. Levine, B.D. Sleeman and M. Nilsen-Hamilton. *A mathematical model for the roles of pericytes and macrophages in angiogenesis. I. the role of protease inhibitors in preventing angiogenesis.* Mathematical Bioscience, vol. 168, 2000. 7

[Li 2008] Pan Li, Urban Malsch and Rolf Bendl. *Combination of intensity-based image registration with 3D simulation in radiation therapy.* Physics in Medicine and Biology, vol. 53, pages 4621–4637, 2008. 148

[Lo 2008] Pechin Lo, J Sporring, H Ashraf, J.J.H Pedersen and M de Bruijne. *Vessel-guided airway segmentation based on voxel classification.* In First International Workshop on Pulmonary Image Analysis, 2008. 148, 150

[Luders 2004] E. Luders, K.L. Narr, P.M. Thompson, D.E. Rex, L. Jancke and A.W. Toga. *Gender Differences in Cortical Complexity.* Nature Neuroscience, vol. 7, no. 8, pages 799–800, 2004. 209

[Ma 2008] Jun Ma, Michael I. Miller, Alain Trouvé and Laurent Younes. *Bayesian template estimation in computational anatomy.* NeuroImage, vol. 42, no. 1, pages 252 – 261, 2008. 168, 291

[Maddah 2007] M. Maddah, W. M. Wells, S. K. Warfield, C.-F. Westin and W. E. L. Grimson. *Probabilistic Clustering and Quantitative Analysis of White Matter Fiber Tracts.* In Proc. of IPMI'07, volume 4584 of *LNCS*, pages 372–383, 2007. 231

[Mallat 1993] S. Mallat and Z. Zhang. *Matching Pursuits with time-frequency dictionaries.* IEEE Transactions on Signal Processing, vol. 41, no. 12, pages 3397–3415, 1993. 89

[Mangin 2004a] J.-F. Mangin, D. Rivière, A. Cachia, E. Duchesnay, Y. Cointepas, D. Papadopoulos-Orfanos, D. L. Collins, A. C. Evans and J. Régis. *Object-based morphometry of the cerebral cortex.* IEEE Transactions on Medical Imaging, vol. 23, no. 8, pages 968–982, 2004. 198, 200, 216

[Mangin 2004b] J.-F. Mangin, D. Rivière, A. Cachia, E. Duchesnay, Y. Cointepas, D. Papadopoulos-Orfanos, P. Scifo, T. Ochiai, F. Brunelle and J. Régis. *A framework to study the cortical folding patterns.* NeuroImage, vol. 23, no. Supplement 1, pages 129–138, 2004. Mathematics in Brain Imaging. 10

[Mansi 2009] Tommaso Mansi, Stanley Durrleman, Boris Bernhardt, Maxime Sermesant, Hervé Delingette, Ingmar Voigt, Philipp Lurz, Andrew M Taylor, Julie Blanc, Younes Boudjemline, Xavier Pennec and Nicholas Ayache. *A Statistical Model of Right Ventricle in Tetralogy of Fallot for Prediction of Remodelling and Therapy Planning.* In Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI'09), Lecture Notes in Computer Science. Springer, 2009. In press. 233, 293

[Marron 2007] Steve Marron, Michael J. Todd and A.H.N. Jeongyoun. *Distance-Weighted Discrimination.* Journal of the American Statistical Association, vol. 102, no. 480, pages 1267–1271, 2007. 195

[Marsland 2004a] Stephen Marsland and Carole Twining. *Constructing Diffeomorphic Representations for the Groupwise Analysis of Non-Rigid Registrations of Medical Images.* IEEE Transactions on Medical Imaging, vol. 23, no. 8, pages 1006–1020, 2004. 164, 203

[Marsland 2004b] Stephen Marsland and Carole Twining. *Constructing Diffeomorphic Representations for the Groupwise Analysis of Non-Rigid Registrations of Medical Images.* Transactions on Medical Imaging, vol. 23, no. 8, pages 1006–1020, 2004. 165

[Marsland 2008] Stephen Marsland, Carole Twining and Chris Taylor. *A Minimum Description Length Objective Function for Groupwise Non-Rigid Image Registration.* Image and Vision Computing, vol. 26, no. 3, pages 333–346, 2008. 11

[Mayneord 1932] W.V. Mayneord. *On a low of growth of Jensen's rat sarcoma.* American Journal of Cancer, vol. 16, 1932. 7

[Meijering 2002] Erik Meijering. *A Chronology of Interpolation: From Ancient Astronomy to Modern Signal and Image Processing.* In Proceedings of the IEEE, pages 319–342, 2002. 79

[Meyer 2001] Yves Meyer. Oscillating patterns in image processing and nonlinear evolution equations, volume 22 of *University Lecture Series.* American Mathematical Society, Providence, RI, 2001. The fifteenth Dean Jacqueline B. Lewis memorial lectures. 166

[Michor 2006] Peter W. Michor and David Mumford. *Riemannian geometries on spaces of plane curves.* J.EUR.MATH.SOC, vol. 8, pages 1–48, 2006. 11

[Michor 2007] Peter W. Michor and David Mumford. *An overview of the Riemannian metrics on spaces of curves using the Hamiltonian approach.* Applied and Computational Harmonic Analysis, vol. 23, no. 1, pages 74 – 113, 2007. Special Issue on Mathematical Imaging. 11, 127

[Miller 2002] I Miller M, A Trouvé and L Younes. *On the Metrics and Euler-Lagrange Equations of Computational Anatomy.* Annual Review of Biomedical Engineering, vol. 4, pages 375–405, 2002. 123, 199, 203, 204, 207, 221

[Miller 2006] M. Miller, A. Trouvé and L. Younes. *Geodesic Shooting for Computational Anatomy.* Journal of Mathematical Imaging and Vision, vol. 24, no. 2, pages 209–228, 2006. 123, 127, 131, 199, 203, 207

[Mio 2007] Washington Mio, Anuj Srivastava and Shantanu H. Joshi. *On Shape of Plane Elastic Curves.* International Journal of Computer Vision, vol. 73, no. 3, pages 307–324, 2007. 11

[Morgan 1987] Frank Morgan. Geometric measure theory. Acad. Press, 1987. 21

[Mulder 2008] Bela Mulder. *Developmental Biology: on Growth and Force.* Science, vol. 322, no. 5908, pages 1643–1644, 2008. 5

[Mumford 2002] David Mumford. *Pattern theory: the mathematics of perception.* In Proceedings of the International Congress of Mathematicians (Beijing), volume 1, pages 401–422. Higher Ed. Press, 2002. 118, 127

[Mumford 2006] David Mumford. *Empirical Statistics and Stochastic Models for Visual Signals.* In New Directions in Statistical Signal Processing: From Systems to Brain. MIT Press, 2006. 8

[Mumford 2007] David Mumford. *What's an Infinite Dimensional Manifold and How Can it Be Useful in Hospitals ?* Talk at the department of mathematics, University of Coimbra, 2007. http://www.dam.brown.edu/people/mumford/Papers/Talks/CoimbraB.pdf. 118

[Narr 2007] K.L. Narr, R.M. Bilder, E. Luders, P.M. Thompson, R.P. Woods, D. Robinson, P. Szeszko, T. Dimtcheva, M. Gurbani and A.W. Toga. *Asymmetries of Cortical Shape: Effects of Handedness, Sex and Schizophrenia.* NeuroImage, vol. 34, no. 3, pages 939–948, 2007. 165, 209

[Needell 2008] D. Needell and J. A. Tropp. *CoSaMP: Iterative signal recovery from incomplete and inaccurate samples.* Apr 2008. 89

[Nocedal 2000] Nocedal and Wright. Numerical optimization. Springer-Verlag, 2000. 141

[Oller 1995] J. Oller and J. Corcuera. *Intrinsic analysis of statistical estimation.* Annals of Statistics, vol. 23, no. 5, pages 1562–1581, 1995. 208

[Pardo 2004] X. M. Pardo, V. Leboran and R. Dosil. *Integrating prior shape models into level-set approaches.* Pattern Recognition Letters, vol. 25, no. 6, pages 631 – 639, 2004. 11

[Pati 1993] Y.C. Pati, R. Rezaifar and P.S. Krishnaprasad. *Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition.* In Conf. Record of the 27th Asilomar Conference on Signals, Systems and Computers, volume 1, pages 40–44, November 1993. 89

[Paus 2001] T. Paus, D.L. Collins, A.C. Evans, G. Leonard, B. Pike and A. Zijdenbos. *Maturation of white matter in the human brain: A review of magnetic resonance studies.* Brain Research Bulletin, vol. 54, no. 3, pages 255–266, 2001. 199

[Pennec 1999] Xavier Pennec. *Probabilities and Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements.* In A.E. Cetin, L. Akarun, A. Ertuzun, M.N. Gurcan and Y. Yardimci, editors, Proc. of Nonlinear Signal and Image Processing (NSIP'99), volume 1, pages 194–198. IEEE-EURASIP, 1999. 208

[Pennec 2006a] Xavier Pennec. *Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements.* Journal of Mathematical Imaging and Vision, vol. 25, no. 1, pages 127–154, July 2006. 113, 118, 131

[Pennec 2006b] Xavier Pennec, Pierre Fillard and Nicholas Ayache. *A Riemannian Framework for Tensor Computing.* International Journal of Computer Vision, vol. 66, no. 1, pages 41–66, 2006. 207, 210

[Perperidis 2005] Dimitrios Perperidis, Raad H. Mohiaddin and Daniel Rueckert. *Spatio-temporal free-form registration of cardiac MRI sequences.* Medical Image Analysis, vol. 9, no. 5, pages 441 – 456, 2005. 249, 253, 255

[Peyrat 2008] Jean-Marc Peyrat, Hervé Delingette, Maxime Sermesant, Xavier Pennec, Chenyang Xu and Nicholas Ayache. *Registration of 4D Time-Series of Cardiac Images with Multichannel Diffeomorphic Demons.* In Dimitris Metaxas, Leon Axel, Gabor Fichtinger and Gábor Székely, editors, Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI'08), volume 5242 of *Lecture Notes in Computer Science*, pages 972–979, New York, USA, September 2008. Springer-Verlag. 248, 250, 252

[Pizer 2003] S.M. Pizer, P.T. Fletcher, S.C. Joshi, A. Thall, J.Z. Chen, Y. Fridman, D.S. Fritsch, A.G. Gash, J.M. Glotzer, M.R. Jiroutek, C.L. Lu, K.E. Muller, G. Tracton, P.A. Yushkevich and E.L. Chaney. *Deformable M-Reps for 3D Medical Image Segmentation.* International Journal of Computer Vision, vol. 55, no. 2-3, pages 85–106, 2003. 11

[Qiu 2008] A. Qiu, L. Younes, M. Miller and J.G. Csernansky. *Parallel transport in diffeomorphisms distinguishes the time-dependent pattern of hippocampal surface deformation due to healthy aging and the dementia of the Alzheimer's type.* NeuroImage, vol. 40, pages 68–76, 2008. 248

[Qiu 2009] Anqi Qiu, Marilyn Albert, Laurent Younes and Michael I. Miller. *Time sequence diffeomorphic metric mapping and parallel transport track time-dependent shape changes.* NeuroImage, vol. 45, no. 1, Supplement 1, pages S51 – S60, 2009. 248, 253, 255

[Regis 2005] J. Regis, J.-F. Mangin, T. Ochiai, V. Frouin, D. Riviere, A. Cachia, M. Tamura and Y. Samson. *"Sulcal Root" Generic Model: a Hypothesis to Overcome the Variability of the Human Cortex Folding Patterns.* Neurol. Med. Chir., vol. 45, no. 1, pages 1–17, 2005. 7, 10

[Rivière 2002] D. Rivière, J.-F. Mangin, D. Papadopoulos-Orfanos, J.-M. Martinez, V. Frouin and J. Régis. *Automatic recognition of cortical sulci of the human brain using a congregation of neural networks.* Medical Image Analysis, vol. 6, no. 2, pages 77–92, 2002. 216

[Rohr 2001] K. Rohr, H. S. Stiehl, R. Sprengel, T. M. Buzug, J. Weese and M. H. Kuhn. *Landmark-based elastic registration using approximating thin-plate splines.* IEEE Transactions on Medical Imaging, vol. 20, no. 6, pages 526–534, 2001. 147

[Rueckert 1999] D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach and D.J. Hawkes. *Nonrigid registration using free-form deformations: application to breast MR Images.* IEEE Transactions on Medical Imaging, vol. 18, no. 8, pages 712–721, 1999. 118

[Sabuncu 2008] M.R. Sabuncu, S.K. Balci and P. Golland. *Discovering Modes of an Image Population through Mixture Modeling.* In Proc. of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), LNCS 5342, pages 381–389. Springer, 2008. 169, 193

[Sabuncu 2009] M.R. Sabuncu, B.T.T. Yeo, K. Van Leemput, T. Vercauteren and P. Golland. *Asymmetric Image-Template Registration.* In Proc. of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), LNCS 5761, pages 565–573. Springer, 2009. 169

[Saitoh 1988] S. Saitoh. Theory of reproducing kernels and its applications, volume 189 of *Pitman Research Notes in Mathematics Series.* Wiley, 1988. 202, 220, 308

[Schoenberg 1938] I.J. Schoenberg. *Metric spaces and completely monotone functions.* Annals of Math. (2), vol. 39, no. 4, pages 811–841, 1938. 144

[Schwartz 1964] Laurent Schwartz. *Sous espaces hilbertiens d'espaces vectoriels topologiques et noyaux associés (noyaux reproduisants).* J. Analyse Math., vol. 13, pages 115–256, 1964. 308

[Schwartz 1966] Laurent Schwartz. *Théorie des Distributions.* Publications de l'Institut de Mathématiques de l'université de Strasbourg, Hermann, 1966. 20, 32, 44

[Shea 1989] B.T. Shea. *Heterochrony in human evolution: The case for neoteny reconsidered.* Yearb. Phys. Anthropol., vol. 32, pages 69–101, 1989. 155

[Sheehan 2008] F.H. Sheehan, S. Ge, G.W. Vick III, K. Urnes, W.S. Kerwin, E.L. Bolson, T. Chung, J.P. Kovalchin, D.J. Sahn, M. Jerosch-Herold*et al. Three-Dimensional Shape Analysis of Right Ventricular Remodeling in Repaired Tetralogy of Fallot.* The American Journal of Cardiology, vol. 101, no. 1, page 107, 2008. 234, 240

[Shen 2002] Dinggang Shen and Christos Davatzikos. *HAMMER: Heirarchical Attribute Matching Mechanism for Elastic Registration.* IEEE Transactions on Medical Imaging, vol. 21, no. 11, pages 1421–1439, 2002. 118

[Shi 2007] Y.G. Shi, P.M. Thompson, I.D. Dinov, S.J. Osher and A.W. Toga. *Direct Cortical Mapping via Solving Partial Differential Equations on Implicit Surfaces.* Medical Image Analysis, vol. 11, no. 3, pages 207–223, 2007. 210

[Smith 2006] S.M. Smith, M. Jenkinson, H. Johansen-Berg, D. Rueckert, T.E. Nichols, C.E. Mackay, K.E. Watkins, O. Ciccarelli, M.Z. Cader, P.M. Matthews and T.E.J. Behrens. *Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data.* NeuroImage, vol. 31, pages 1487–1505, July 2006. 218

[Sternberg 1964] Shlomo Sternberg. Lectures on differential geometry. Prentice-Hall, Inc, 1964. 297

[Thévenaz 2000] Philippe Thévenaz, Thierry Blu and Michael Unser. *Interpolation revisited.* IEEE Transactions on Medical Imaging, vol. 19, pages 739–758, 2000. 79

[Thompson 1917] D'Arcy W. Thompson. *On Growth and Form.* 1917. 5, 12, 165

[Thompson 1996a] P.M. Thompson, C. Schwartz, R.T. Lin, A.A. Khan and A.W. Toga. *3D Statistical Analysis of Sulcal Variability in the Human Brain.* Journal of Neuroscience, vol. 16, no. 13, pages 4261–4274, 1996. 104, 198

[Thompson 1996b] P.M. Thompson, C. Schwartz and A.W. Toga. *High-Resolution Random Mesh Algorithms for Creating a Probabilistic 3D Surface Atlas of the Human Brain.* NeuroImage, vol. 3, no. 1, pages 19–34, 1996. 210

[Thompson 1996c] P.M. Thompson and A.W. Toga. *A Surface-based Technique for Warping 3-Dimensional Images of the Brain.* I.E.E.E. Transations on Medical Imaging, vol. 15, no. 4, pages 1–16, 1996. 200, 210

[Thompson 1998] P.M. Thompson, J. Moussai, A.A. Khan, S. Zohoori, A. Goldkorn, M.S. Mega, G.W. Small, J.L. Cummings and A.W. Toga. *Cortical Variability and Asymmetry in Normal Aging and Alzheimer's Disease.* Cerebral Cortex, vol. 8, no. 6, pages 492–509, 1998. 210, 214

[Thompson 2000] Paul M. Thompson, Jay N. Giedd, Roger P. Woods, David MacDonald, Alan C. Evans and Arthur W. Toga. *Growth Patterns in the Developing Human Brain Detected By Using Continuum-Mechanical Tensor Maps.* Nature, vol. 404, no. 6774, 2000. 248

[Thompson 2002] P.M. Thompson, K.M. Hayashi, G. de Zubicaray, A.L. Janke, S.E. Rose, J. Semple, D.M. Doddrell, T.D. Cannon and A.W. Toga. *Detecting Dynamic and Genetic Effects on Brain Structure using High-Dimensional Cortical Pattern Matching.* In Proc. International Symposium on Biomedical Imaging (ISBI), pages 473–476, 2002. 210

[Thompson 2003] P.M. Thompson and A.W. Toga. *Cortical Diseases and Cortical Localization.* Nature Encyclopedia of the Life Sciences, 2003. (Review Article). 216

[Thompson 2004] P.M. Thompson, K.M Hayashi, E.R. Sowell, N. Gogtay, J.N. Giedd, J.L. Rapoport, G.I. de Zubicaray, A.L. Janke, S.E. Rose, J. Semple, D.M. Doddrell, Y.L. Wang, T.G.M. van Erp, T.D. Cannon and A.W. Toga. *Mapping Cortical Change in Alzheimer's Disease, Brain Development, and Schizophrenia.* NeuroImage, vol. 23, no. 1, pages 2–18, 2004. Special Issue on Mathematics in Brain Imaging. 165

[Toga 1999] Arthur W. Toga and Paul Thompson. *Brain Warping.* In in Brain Warping, pages 1–26. Academic Press, 1999. 165

[Toga 2007] A.W. Toga and P.M. Thompson. *What is Where and Why it is Important.* NeuroImage, 2007. Peer-Reviewed Invited Commentary on a paper by Devlin J, Poldrack R "In Praise of Tedious Anatomy", Feb. 9 2007. 216

[Tosun 2005] D. Tosun and J.L. Prince. *Cortical Surface Alignment Using Geometry Driven Multispectral Optical Flow.* In Information Processing in Medical Imaging, volume 3565 of *Lecture Notes in Computer Science*, pages 480–492. Springer, 2005. 198

[Trouvé 1995] Alain Trouvé. *An approach of pattern recognition through infinite dimensional group actions.* Technical report, 1995. 118, 121, 127

[Trouvé 1998] Alain Trouvé. *Diffeomorphisms groups and pattern matching in image analysis.* International Journal of Computer Vision, vol. 28, pages 213–221, 1998. 118, 123, 199

[Trouvé 2005a] Alain Trouvé and Laurent Younes. *Local Geometry of Deformable Templates.* SIAM Journal on Mathematical Analysis, vol. 37, no. 1, pages 17–59, 2005. 175

[Trouvé 2005b] Alain Trouvé and Laurent Younes. *Metamorphoses through Lie group action.* Foundations of Computational Mathematics, vol. 5, no. 2, pages 173–198, 2005. 128, 170

[Twining 2005] Carole J. Twining, Tim Cootes, Stephen Marsland, Vladimir Petrovic adn Roy Schestowitz and Chris J. Taylor. *A Unified Information-Theoretic Approach to Groupwise Non-Rigid Registration and Model Building.* In Information Processing in Medical Imaging (IPMI). Lecture Notes of Computer Vision, Springer-Verlag, 2005. 11

[Vaillant 1999] Marc Vaillant and Christos Davatzikos. *Hierarchical Matching of Cortical Features for Deformable Brain Image Registration.* In Information Processing in Medical Imaging, volume 1613 of *Lecture Notes in Computer Science*, pages 182–195. Springer, 1999. 215

[Vaillant 2004] M. Vaillant, M.I. Miller, L. Younes and A. Trouvé. *Statistics on diffeomorphisms via tangent space representations.* NeuroImage, vol. 23, pages 161–169, 2004. 118, 119, 127, 131, 184, 207, 218

[Vaillant 2005] Marc Vaillant and Joan Glaunès. *Surface Matching via Currents.* In Proceedings of Information Processing in Medical Imaging, volume 3565 of *Lecture Notes*

in Computer Science, pages 381–392. Springer, 2005. 11, 13, 14, 21, 32, 106, 119, 134, 138, 145, 147, 199, 200, 201, 202, 204, 219, 221, 236

[Vaillant 2007] M. Vaillant, A. Qiu, J. Glaunès and M.I. Miller. *Diffeomorphic Metric Surface Mapping in Subregion of the Superior Temporal Gyrus*. NeuroImage, vol. 34, no. 3, pages 1149–1159, 2007. 198, 199

[Vercauteren 2007] Tom Vercauteren, Xavier Pennec, Ezio Malis, Aymeric Perchant and Nicholas Ayache. *Insight Into Efficient Image Registration Techniques and the Demons Algorithm*. In Proc. of IPMI'07, volume 4584 of *LNCS*, pages 495–506, 2007. 223

[Vercauteren 2009] Tom Vercauteren, Xavier Pennec, Aymeric Perchant and Nicholas Ayache. *Diffeomorphic Demons: Efficient Non-parametric Image Registration*. NeuroImage, vol. 45, no. 1, Supp.1, pages S61–S72, March 2009. 118

[Vialard 2009] François-Xavier Vialard. *Hamiltonian Approach to Shape Spaces in a Diffeomorphic Framework: From the Discontinuous Image Matching Problem to a Stochastic Growth Model*. PhD thesis, Ecole Normale Supérieure de Cachan, http://tel.archives-ouvertes.fr/docs/00/40/03/79/PDF/ThesisFXV.pdf, May 2009. 127, 282, 285

[Vik 2008] Torbjorn Vik, Sven Kabus, Jens von Berg, Konstantin Ens, Sebastian Dries, Tobias Klinder and Cristian Lorenz. *Validation and Comparison of Registration Methods for Free-Breathing 4D Lung-CT*. In J. M. Reinhardt and J. P. W. Pluim, editors, SPIE (Medical Imaging), volume 6914, 2008. 148

[Wang 2005] Y.L. Wang, M.C. Chiang and P.M. Thompson. *Automated Surface Matching using Mutual Information Applied to Riemann Surface Structures*. In Medical Image Computing and Computer Assisted Interventions (MICCAI), volume 3750 of *Lecture Notes in Computer Science*, pages 666–674, 2005. 210

[Wang 2007] T Wang and A Basu. *A note on 'a fully parallel 3D thinning algorithm and its applications'*. Pattern Recognition Letters, vol. 28, no. 4, pages 501–506, 2007. 148

[Wintgen 1982] P. Wintgen. *Normal cycle and integral curvature for polyhedra in Riemannian manifolds*. Differential Geometry, 1982. 21

[Won 2005] Y.J. Won and J. Hey. *Divergence population genetics of chimpanzees*. Mol. Biol. Evol., vol. 22, pages 297–307, 2005. 154

[Yang 2003] C. Yang, R. Duraiswami, N.A. Gumerov and L. Davis. *Improved fast gauss transform and efficient kernel density estimation*. In Ninth IEEE International Conference on Computer Vision, volume 1, pages 664–671, 2003. 84

[Yeo 2008a] B.T.T. Yeo, M.R. Sabuncu, R. Desikan, B. Fischl and P. Golland. *Effects of registration regularization and atlas sharpness on segmentation accuracy*. Medical Image Analysis, vol. 12(5), pages 603–615, 2008. 164

[Yeo 2008b] B.T.T. Yeo, T. Vercauteren, P. Fillard, X. Pennec, P. Golland, N. Ayache and O. Clatz. *DTI Registration with Exact Finite-Strain Differential*. In ISBI'08, pages 700–703, 2008. 218, 223

[Yeo 2009a] B.T.T. Yeo, M.R. Sabuncu, T. Vercauteren, N. Ayache, B. Fischl and P. Golland. *Spherical Demons: Fast Diffeomorphic Landmark-Free Surface Registration*. IEEE Transactions on Medical Imaging, 2009. In Press. 118

[Yeo 2009b] B.T.T. Yeo, T. Vercauteren, P. Fillard, J-M. Peyrat, X. Pennec, P. Golland, N. Ayache and O. Clatz. *DT-REFinD: Diffusion Tensor Registration with Exact Finite-Strain Differential*. IEEE Transactions on Medical Imaging, 2009. In Press. 218

[Zeidler 1991] Eberhard Zeidler. Applied functional analysis: Application to mathematical physics. Springer, 1991. 308

[Zhang 1994] Z. Zhang. *Iterative Point Matching for Registration of Free-Form Curves and Surfaces*. International Journal of Computer Vision, vol. 13, no. 2, pages 119–152, 1994. 199

[Zhang 2006] H. Zhang, N. Walker, S.C. Mitchell, M. Thomas, A. Wahle, T. Scholz and M. Sonka. *Analysis of four-dimensional cardiac ventricular magnetic resonance images using statistical models of ventricular shape and cardiac motion*. In Proc. SPIE 2006, volume 6143, page 614307, 2006. 234

[Zhang 2007] H Zhang, P A Yushkevich, D Rueckert, and J C Gee. *Unbiased white matter atlas construction using diffusion tensor images*. In Proc. of MICCAI'07, volume 4792 of *LNCS*, pages 211–218, 2007. 218

[Zheng 2007] Yefeng Zheng, Adrian Barbu, Bogdan Georgescu, Michael Scheuering and Dorin Comaniciu. *Fast Automatic Heart Chamber Segmentation from 3D CT Data Using Marginal Space Learning and Steerable Features*. In Proc. ICCV 2007, pages 1–8, 2007. 237

[Ziyan 2007] Ulas Ziyan, Mert R. Sabuncu, Lauren J. O'Donnell and Carl-Fredrik Westin. *Nonlinear Registration of Diffusion MR Images Based on Fiber Bundles*. In Proceedings of MICCAI'07, volume 4791 of *LNCS*, pages 351–358, 2007. 218

[Zollei 2005] Lilla Zollei, Erik Learned-Miller, Eric Grimson and William Wells. *Efficient Population Registration of Data*. In Computer Vision for Biomedical Image Applications, volume 3765 of *Lecture Notes in Computer Science*, pages 291–301. Springer, 2005. 164