N° ENSC-2009

THÈSE DE DOCTORAT
DE L'ÉCOLE NORMALE SUPÉRIEURE DE CACHAN

Présentée par
Monsieur Nicolas CHARON

pour obtenir le grade de

DOCTEUR DE L'ÉCOLE NORMALE SUPÉRIEURE DE CACHAN

Domaine : **MATHEMATIQUES**

Sujet de la thèse :

Analysis of geometric and functional shapes with extensions of currents. Applications to registration and atlas estimation.

Rapporteurs :

David Mumford	Professeur émérite, Brown University
Gabriel Peyré	Chargé de Recherches, CNRS

Thèse présentée et soutenue à Cachan le 14 novembre 2013 devant le jury composé de :

Joan Glaunès	Maître de conférences, Université Paris Descartes
Sarang Joshi	Maître de conférences, University of Utah
Stephane Mallat	Professeur, Ecole Normale Supérieure
Michael Miller	Professeur, John Hopkins University
Gabriel Peyré	Chargé de Recherches, CNRS (Rapporteur)
Alain Trouvé	Professeur, ENS Cachan (Directeur)

Centre de Mathématiques et de Leurs Applications
(ENS CACHAN/CNRS/UMR 8536)
61, avenue du Président Wilson, 94235 CACHAN CEDEX (France)

Il faut de temps en temps nous reposer de nous-mêmes, en nous regardant de haut, avec une distance artistique, pour rire, pour pleurer sur nous ; il faut que nous découvrions le héros et aussi le fou que cache notre passion de la connaissance ; il faut, de-ci, de-là, nous réjouir de notre folie pour pouvoir rester joyeux de notre sagesse.

—Le Gai Savoir—

Remerciements

Je ne saurais entamer ces remerciements autrement qu'en témoignant avant tout mon immense reconnaissance et admiration à Alain Trouvé, qui fut d'abord l'un de mes professeur au cours de mes études à l'ENS Cachan puis bien sûr mon directeur de thèse lors de mes quatre années passées au CMLA. A son contact, j'ai pu apprécier et progresser dans certaines qualités essentielles, à mon avis, au travail de chercheur en mathématiques ; parmi elles je citerai l'enthousiasme, l'imagination, la vivacité, la rigueur, la persévérance et la patience. Je le remercie pour sa disponibilité durant toute ma thèse et son encadrement très riche en apprentissages dont je garderai indiscutablement un excellent souvenir.

Je souhaite ensuite remercier les deux rapporteurs de ma thèse, David Mumford et Gabriel Peyré, pour avoir accepté de relire mon manuscrit et m'avoir fait part de leur appui et de leurs remarques sur ce travail. Je remercie également les membres du jury : Joan Glaunès, Sarang Joshi, Stéphane Mallat, Michael Miller et Gabriel Peyré d'être présent aujourd'hui.

Cette thèse est bien sûr le fruit de multiples collaborations et discussions avec de nombreux doctorants et professeurs que je tiens aussi à mentionner. En premier lieu, Stanley Durrleman et Joan Glaunès qui ont été les pionniers de l'utilisation des courants en anatomie numérique, sur quoi ce travail s'est grandement appuyé. Je les remercie tous deux d'avoir continué à interagir sur des sujets aussi bien théoriques que numériques tout au long de ma thèse. Merci également à Benjamin Charlier pour notre très active et intense collaboration pendant ma dernière année de thèse, notamment dans l'élaboration des algorithmes à l'origine des applications de ce travail, et sa grande patience en dépit de mes fréquentes crises de nerfs dans les périodes de débugage ! Merci aussi à José Braga et Jean-Luc Kahn de nous avoir fourni les données de surfaces d'oreilles internes, Faisal Beg, Marinko Sarunic et Evgeniy Lebed de Simon Fraser University pour les données de couches rétiniennes, Sigolhène Meilhac et Jean-François Le Garrec de l'Institut Pasteur pour celles des divisions cellulaires ainsi que Bertrand Thirion de Neurospin pour les données d'IRMf.
Au cours de ces quatre ans de thèse, j'ai pu également profiter de l'expérience de plusieurs personnes lors de discussions diverses, variées et parfois animées, parmi lesquelles Irène, Barbara, Giacommo, Sylvain, Sarang Joshi, François-Xavier Vialard, Stéphanie Allassonnière, Stefan Sommer ainsi que Jimmy Lambolley et Thierry De Pauw pour son excellent cours de master en théorie géométrique de la mesure et les quelques messages que nous avons échangé depuis sur le sujet. Je remercie au passage toute l'équipe de Darryl Holm à Imperial College puis celle de Michael Miller au CIS pour avoir organisé les 'Shape meeting' et offert une occasion tout à la fois enrichissante et décontractée de réunir notre communauté. Merci enfin à tous les professeurs de l'ENS Cachan dont j'ai eu l'honneur et plaisir de suivre les cours ainsi qu'aux nombreux enseignants qui m'ont transmis le goût des mathématiques avant l'ENS, je pense en particulier à Mmes Hiribarren et Lucaroni au lycée, Mme Camez et Mr Gonnord en prépa.

Outre la recherche, j'ai énormément apprécié mes trois années d'enseignement en tant que moniteur à l'ENS Cachan, grâce notamment à Frédéric Pascal avec qui

nous avons assuré l'enseignement du calcul différentiel et de toutes ses joyeusetés, ainsi qu'à Vincent Beck et Ayman Moussa pour m'avoir fourni leurs feuilles d'exercices du temps où eux-mêmes m'avaient enseigné ces matières. Je salue au passage tous mes anciens étudiants avec qui j'ai eu la chance de passer de très bons moments pendant et parfois même en dehors des cours.

Cette thèse doit aussi beaucoup au cadre dans lequel elle s'est déroulée et, à ce titre, je voudrais remercier toutes les personnes que j'ai eu l'occasion de cotoyer ici. A commencer par Micheline, Virginie et Véronique pour leur efficacité à régler tous mes déboires administratifs (et pour instaurer un peu de discipline dans l'utilisation de la cuisine !), Sandra pour tenir notre bibliothèque, Christine Rose à l'EDSP, Christophe, Nicolas et Atman pour l'informatique et la maintenance des serveurs de calculs que j'ai beaucoup mis à contribution au cours de ma thèse. Tous mes remerciements vont enfin à l'ensemble des thésards, post-docs et AGPR du CMLA que j'ai croisé pendant mon doctorat pour avoir su insuffler, grâce aux pause-café, sorties et autres sacro-saintes 'traditions', une vitatilité et une ambiance extrêmement agréable au labo. Adina, Anne, Ariane, Aude, Ayman, Barbara, Benjamin C, Benjamin M, Benoît, Bruno, Carlo, Cyril, Claire B, Claire R, Enric, Eric, Frédérique, Giacommo, Gaël, Irène, Ives, José, Julie D, Julien, Kleber, Lara, Laurent, Loïc, Marc, Marie, Mauricio, Mathieu, Miguel, Milana, Morgan, Neus, Nicola, Nicolas L, Rachel, Rafa, Saad, Samy, Sandrine, Tristan, Yohann, Zhongwei (la liste est déjà longue mais je m'excuse d'avance si j'ai omis quelqu'un) : merci à vous tous !

Sur le plan plus personnel, je voudrais aussi m'adresser à l'ensemble de mes amis sans la présence et le soutien desquels je n'aurai certainement pas pu en arriver là aujourd'hui. D'abord à tous mes camarades de promo, en particulier mes deux compères de toujours Popoff et Knarond, Agathe, Arthur, Baptiste, Cécilou, Gaëlle, Irène, JSK, Jimmy, Julie, Quentin, Romain et Sandra. Je garde un souvenir ému de tous ces moments inoubliables, souvent cocasses et jamais ennuyeux passés en votre compagnie, que ce soit à Paris, au chalet du lézard de La Clusaz, à Pralognan, en Normandie, en Bretagne ou en Corse, à Lille, à Nice, sur l'île de Ré et en Provence. Je pense aussi à tous mes autres amis de l'ENS, Emilien et Lucie en particulier pour les nombreux voyages et fêtes que j'ai pu faire avec vous (et souvent chez vous), Pierre-Antoine (bientôt le 9b+, accroche-toi), Ariane pour avoir bien voulu accompagner un débutant comme moi aux soirées salsa, les clubbers de Saint-Quentin Maxime & Maxime, Benjamin Butin et Vivien qui, de par leur jeu idoine, ont su me recadrer inlassablement au tennis pendant toutes ces années, Romain de la prestigieuse Heyman & Versachs, Sarah, Sorin, Jean, Jules, Pierre L, Arthur L, Thomas P et son aisance inégalée dans les cheminées de Bleau, Gabriel et nos quelques belles randos dans les Alpes, le groupe des skieurs de Chatel Jack, Marie, Camille, Valérian, Rodrigo et JR. Une pensée également pour le groupe des Rennais Julie, Mama, Romain B et Elise avec qui je me suis rapidement intégré, ainsi que la Drôme team qui se reconnaîtra : Lucie, Alice, Emilie, Elise, Ileyk, Marie, Mickaël, Valérie et Tigrane, merci pour ces super vacances d'été passées avec vous ! Merci aussi à mes amis couchsurfers Mi, Amina, Raphy, Nitesh, Lucie et Emilia, ainsi qu'aux Toulousains que j'ai encore l'occasion de voir : Mathieu, Guillaume, Marion, Jérôme et toute la 'Debiez family'. Et, pour finir, merci à toutes les autres personnes que j'ai eu la chance de rencontrer ici ou ailleurs : Clélia, Valentin, Sarah, Line, Justine, Lucie et Yann, Axel, Pauline, Adrien, Marie C...

Enfin, je réserve une place très spéciale à toute ma famille qui m'a toujours témoigné un soutien sans faille dans tous les moments les plus difficiles. Merci à ma petite soeur Elsa, merci à ma mère à qui je dois mon premier intérêt pour les maths et d'avoir persisté dans cette voie par la suite, à mon père dont j'espère pouvoir un jour égaler la culture générale et le sens de l'humour, et mes grands-parents auxquels je dois aussi beaucoup et qui auraient vraiment souhaité pouvoir être là aujourd'hui. Je ne pourrai jamais vous remercier assez pour ce que vous avez tous fait et continuez de faire pour moi. En modeste signe de reconnaissance, cette thèse vous est dédiée.

Résumé en français

Cette thèse s'articule autour de problématiques liées au récent domaine de l'anatomie numérique dont l'objet est de fournir des cadres à la fois mathématiques et numériques pour estimer la variabilité statistique au sein de populations de formes anatomiques. De tels modèles constituent une étape importante pour de nombreuses applications, notamment la détection automatique de pathologies en imagerie biomédicale. En termes mathématiques, ces problèmes ont pu être formalisés en grande partie via les travaux de Grenander et la notion d'espace de formes (ensemble d'objets géométriques sur lequel agit transitivement un groupe de transformations), puis par l'introduction du modèle LDDMM pour définir des groupes de 'grandes déformations' et des métriques invariantes à droite sur ces groupes.

Dans ce travail, on s'intéresse dans un premier temps au cas d'espaces de formes donnés par un ensemble de courbes, de surfaces ou sous-variétés avec pour premier objectif de définir une représentation et des termes d'attache aux données adéquats pour mesurer la dissimilarité entre objets dans les problèmes de recalage par grande déformation. Ce problème a été une première fois abordé, en anatomie numérique, par Glaunès & Vaillant (2005) dans le cas d'objets orientés en adaptant le concept de courant de la théorie géométrique de la mesure. La première contribution de ce travail est de proposer une extension de ces idées pour des formes géométriques non-orientées via la représentation des varifolds, issue à nouveau de la théorie géométrique de la mesure. On montre en particulier l'intérêt de ces représentations non-orientées dans plusieurs situations bien distinctes.

Dans un second temps, ce travail se penche sur l'étude d'objets géométrico-fonctionnels aussi baptisés 'formes fonctionnelles', c'est à dire de fonctions ou de signaux définis sur des supports géométriques variables entre les individus. Ces types de données sont de plus en plus fréquents en anatomie numérique mais, à notre connaissance, la plupart des approches standards commencent par normaliser la géométrie des objets pour se ramener au cas d'images plus classiques définies sur un support fixe. L'un des problèmes importants sur lequel se concentre cette thèse est de donner un cadre pour modéliser et analyser géométrie et signal conjointement. Pour cela, on définit notamment la notion de métamorphoses géométrico-fonctionnelles pour généraliser celles de grande déformation et de métriques invariantes à droite à ce contexte ainsi que la notion de courant fonctionnel pour mesurer la dissimilarité entre deux formes fonctionnelles. Ceci débouche assez naturellement sur un tout nouveau cadre mathématique et algorithmique permettant d'étendre les outils usuels de recalage.

La dernière partie de la thèse s'intéresse à la situation plus générale de l'estimation et l'analyse d'atlas pour des populations de telles structures. On propose en particulier une formulation mathématique bien posée pour de tels problèmes ainsi qu'un algorithme d'estimation simultanée géométrie/fonction de ces atlas. L'étape qui suit est naturellement l'analyse statistique de la variabilité des atlas : dans ce but, nous adaptons les outils classiques de PCA pour l'analyse des modes propres d'une part et de LDA pour la classification d'autre part. Toutes ces méthodes sont enfin illustrées sur des applications à quelques jeux de données synthétiques et d'autres issues de l'imagerie biomédicale.

Contents

1 Toward a framework for fshapes **13**
- 1.1 Introduction . . . 14
- 1.2 Computational Anatomy and deformation groups . . . 14
  - 1.2.1 General issues . . . 14
  - 1.2.2 Principle of shape analysis through group actions . . . 16
  - 1.2.3 Large deformations . . . 17
  - 1.2.4 The kernel trick . . . 19
  - 1.2.5 Geodesic equations in the landmark case . . . 21
- 1.3 Inexact registration and shape dissimilarities . . . 22
  - 1.3.1 Inexact matching formulation . . . 22
  - 1.3.2 Geodesic shooting algorithm . . . 23
  - 1.3.3 Measuring shape dissimilarities . . . 24
  - 1.3.4 The representation by currents . . . 26
- 1.4 Problematics and contributions of this work . . . 33
  - 1.4.1 Extension to geometrico-functional data . . . 33
  - 1.4.2 A non-oriented setting for shape dissimilarity . . . 39
  - 1.4.3 Organization of the chapters . . . 40

2 Functional currents **43**
- 2.1 Functional shapes . . . 44
  - 2.1.1 State of the art . . . 44
  - 2.1.2 Functional shapes . . . 46
  - 2.1.3 Representation by currents . . . 48
- 2.2 Functional currents : mathematical considerations . . . 51
  - 2.2.1 Definition and basic properties of functional currents . . . 51
  - 2.2.2 Reproducing Kernel Hilbert Space structures on functional currents . . . 55
  - 2.2.3 Properties of kernel norms on fcurrents . . . 60
- 2.3 Algorithms based on functional currents . . . 71
  - 2.3.1 Compression with Matching Pursuit on fcurrents . . . 72
  - 2.3.2 Registration of functional shapes . . . 75
  - 2.3.3 Numerical considerations . . . 82
- 2.4 Conclusion . . . 87

3 Varifolds **89**
- 3.1 Orientation and limitations of currents . . . 90
  - 3.1.1 Orientation ambiguity . . . 90
  - 3.1.2 Shape annihilation . . . 91
  - 3.1.3 Non-oriented representations in geometric measure theory . . . 93

3.2 Varifolds . . . 94
3.2.1 Definitions and basic properties . . . 94
3.2.2 Non-oriented shapes as varifolds . . . 97
3.3 Kernel metrics on varifolds . . . 100
3.3.1 Kernels on the Grassmann manifold . . . 100
3.3.2 Kernels on varifolds . . . 103
3.4 Properties of RKHS norms on varifolds . . . 107
3.4.1 Dependency on the scales of the kernels . . . 107
3.4.2 Control of the volume by W'-norm . . . 108
3.4.3 Metric variation formula . . . 112
3.5 Registration of unoriented shapes . . . 115
3.5.1 Description of the algorithm . . . 115
3.5.2 Discrete expressions for curves and surfaces . . . 116
3.5.3 Numerical complexity . . . 118
3.5.4 Simulations and results . . . 119
3.6 Discussion, extensions and perspectives . . . 121
3.6.1 Scale and noise . . . 122
3.6.2 Functional varifolds . . . 126
3.6.3 Oriented varifolds . . . 127
3.6.4 Conclusion . . . 127

4 Statistical Atlas estimation and analysis 129
4.1 Forward atlas models . . . 130
4.2 Atlas estimation for functional shapes . . . 132
4.2.1 A first model . . . 132
4.2.2 Estimation of functional residuals . . . 142
4.2.3 Results on the retina dataset . . . 147
4.2.4 A well-posed continuous framework for atlas estimation . . . 151
4.3 Statistical analysis of momenta and residuals . . . 153
4.3.1 Dimensionality reduction . . . 153
4.3.2 Generalized LDA on Hilbert spaces . . . 157
4.3.3 Modes and classification for functional atlases . . . 159
4.3.4 Classification results on the OCT dataset . . . 162
4.4 Atlas on varifolds . . . 168
4.4.1 Algorithm . . . 168
4.4.2 Works on embryonic mouse heart . . . 168
4.4.3 Template estimation on Cochleae dataset . . . 172
4.5 Conclusion and extensions . . . 175

Conclusion and and discussion 177

A Existence of geometrico-functional atlases 181
A.1 Existence of the template fshape . . . 181
A.2 Existence of full fshape atlases . . . 187

Chapter 1

Toward a framework for fshapes analysis in computational anatomy

Contents

1.1 Introduction 14
1.2 Computational Anatomy and deformation groups 14
1.2.1 General issues 14
1.2.2 Principle of shape analysis through group actions 16
1.2.3 Large deformations 17
1.2.4 The kernel trick 19
1.2.5 Geodesic equations in the landmark case 21
1.3 Inexact registration and shape dissimilarities 22
1.3.1 Inexact matching formulation 22
1.3.2 Geodesic shooting algorithm 23
1.3.3 Measuring shape dissimilarities 24
1.3.4 The representation by currents 26
1.4 Problematics and contributions of this work 33
1.4.1 Extension to geometrico-functional data 33
1.4.2 A non-oriented setting for shape dissimilarity 39
1.4.3 Organization of the chapters 40

1.1 Introduction

The work presented in this thesis can be articulated, for its essential part, around a recent problematic that arose in the field of computational anatomy following the increasing appearance of a new type of data structures that can be described as *geometrico-functional*. Computational anatomy has focused historically on the study of geometrical shapes as sets of points, curves and surfaces on the one side or classical images on the other. Yet the case of signals supported on variable geometrical shapes, as for instance activation maps on cortical surfaces in functional MRI, has been rarely addressed in its generality. We call functional shapes or **fshapes** such kind of structures made of a shape that carries a certain signal function.
The ultimate goal of computational anatomy, as we shall explain with more details in the next section, is to analyze the anatomical variability among a population of different subjects. Several frameworks have been already developed to achieve this for usual shapes. In this thesis, we intend to provide or at least initiate a complete theoretical and numerical framework for treating fshapes. The main challenge is to define a setting that can analyze both geometrical and signal variability together, as opposed to previous approaches that rather intend to normalize geometries of the subjects in the first place to then compare the signals. One of the most crucial point is thus the question of representation of the set of fshapes and the definition of appropriate metrics that can account for the geometrico-functional variability.
The objective of this first chapter is double. We introduce, in the two first sections, the general context and framework of computational anatomy. This is also the occasion to present all the mathematical concepts on which this work rely in large part. In the first section, we present what is related to the construction of large deformation groups including the geodesic formalism for diffeomorphisms and reproducing kernels for the definition of Hilbert metrics on vector fields. The second section deals with the issues of shape dissimilarity measures for inexact matching. In particular, we make a brief summary of the theory of *currents*, that shall be a central concept in this thesis. In the second place (section 1.4), we analyze the specific problematics related to functional shapes themselves by initiating the construction of a Riemannian framework for a simple class of fshapes and present the challenging issue of dissimilarity measures for these objects, that constitute the topic of the next chapters.

1.2 Computational Anatomy and deformation groups

1.2.1 General issues

One of the major origin of the present work is the development of the recent and very active discipline named *Computational Anatomy*. If the term was originally coined by Grenander and Miller in [45], this domain can be rooted to earlier works in biology related to morphogenesis that we can trace up to the pioneer study of D'Arcy Thompson [80]. The problematics were to understand and model the anatomical variations between species and try to explain physically the evolution in the shapes of certain body structures. In that purpose, D'Arcy Thompson was the first to introduce the idea of simple geometrical transformations underlying the shape variability as we show on the examples of figure 1.1.
With the more recent development of a wide variety of non-invasive acquisition techniques provided for instance by microscopy, coherence tomography, magnetic res-

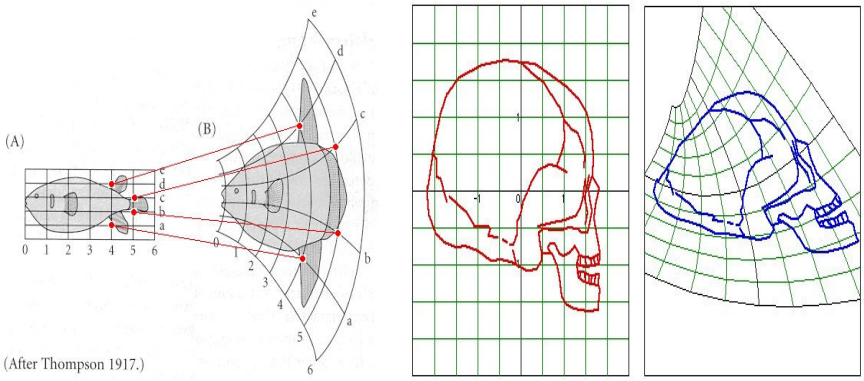


Figure 1.1: Two illustrations from D'Arcy Thompson's works.

onance imaging (MRI), diffusion tensor imaging (DTI), the accessible anatomical structures as well as the number of different subjects acquired for each structure has literally exploded, particularly during the past decade. A lot of different anatomical parts can be now imaged, giving to scientists a growing number of exploitable databases. Interestingly, the development of data acquisition technology came along with new challenges and perspectives at many levels. From the dataset point of view, the emergence of these new techniques has resulted in a diversification in the variety of shapes that are to be analysed, going from 2D or 3D images, landmarks, curves, surfaces, fiber sets... This involves particular modelling issues that we shall discuss later on. On the other hand, the access to certain complex anatomical structures, the brain is one of the most emblematic example, has evidenced that shape variability is not only present at the inter-specie level but can be also very pronounced among individuals of the same specie. The study of this *natural anatomical variability* has thus drawn sustained attention from researchers for the past decades.

Another important reason is that such questions have become of considerable interest in the field of medical imaging. Indeed, these new technologies have enabled more advanced studies of abnormal developments of certain organs, paving the way to early diagnosis of diseases caused or conveyed by unexpected deformations. Yet, in order to make such detections not simply qualitative and more automatized, the central issue is thus to estimate numerically what a pathological shape variation means. This is one the fundamental problematics of computational anatomy : formally, it consists in estimating statistical models of shape variability for different groups of subjects and propose statistical tests for discriminating 'natural' from 'pathological' variations. It is clearly a vast and difficult problem in general. Indeed, even the simpler question of computing shape variation between two given subjects, which is called *shape registration* or matching, has already given birth to many theoretical and numerical issues, that we shall elaborate in the following. Despite such difficulties, the existing methods, developed in the field of computational anatomy, have already proved successful in countless applications : let's quote, among many others, the study of the shape of Hippocampus in relation to the evolution of Alzheimer disease [24, 26, 75, 81], similar works on the planum temporale for schizophrenia [69], Down syndrome [33], the analysis of brain connectivity based on DTI imaging [30, 85], or studies of heart shapes and malformations [13, 56, 60].

1.2.2 Principle of shape analysis through group actions

Going back to the intuition of D'Arcy Thompson, the idea of seeing shape differences through certain transformations of the space was first formalized mathematically and in its full generality by Grenander in [44]. This formalism allowed the prolific interactions between anatomy on the one hand and pattern theory and statistics on the other, giving birth to computational anatomy as we know it.
To briefly present this formalism, we could say that, in this point of view, a *shape space* is a homogeneous space consisting of a set M of 'shapes' living in a domain $\mathbb{R}^n$ (it can be either a set of images, curves, submanifolds etc...) which is acted on the left by a group G of transformations of the space $\mathbb{R}^n$. The fact that we consider homogeneous spaces imposes that the action of G on M is transitive. Another way to say so is that we restrict to one particular orbit $G.m_0$ with $m_0 \in M$. Thus all shapes in the shape space are obtained by transformations of one particular m_0 (often called the *template*) by elements of the group G. The group G may be basically any subgroup of the group of bijections of $\mathbb{R}^n$ into itself. We generally assume more regularity by choosing subgroups of $\mathrm{Diff}(\mathbb{R}^n)$, the set of C^1-diffeomorphisms of $\mathbb{R}^n$. Here are two classical examples of shape spaces.

Example 1.2.1. *Landmarks. We consider the space*

$$M = \{m = (x_1, .., x_p) \in (\mathbb{R}^n)^p \ / \ \forall i \neq j, \ x_i \neq x_j\}$$

i.e the space of all distinct p landmarks in $\mathbb{R}^n$*. We define the left group action of* $G = Diff(\mathbb{R}^n)$ *on* M *by :*

$$\forall g \in G, \forall m \in M, \ g.m = (g(x_1), ..., g(x_p))$$

For $n \geq 2$*, the action of* G *on* M *is transitive which makes* M *a shape space in the previous sense. Following the situation of landmarks, more sophisticated shape spaces can be also obtained by quotienting* M *with the group of isometries of* $\mathbb{R}^n$*. These are known as the Kendall spaces and are studied in [51].*

Example 1.2.2. *Images. Let* $\Omega \subset \mathbb{R}^n$ *be a fixed domain of* $\mathbb{R}^n$ *and* $M = L^2(\Omega, \mathbb{R})$ *the set of continuous images on* Ω*. One can define a left action of the group* $G = Diff(\Omega)$ *on* M *by setting :*

$$\forall g \in G, \forall I \in M, \ g.I = I \circ g^{-1}$$

Then for any $I_0 \in M$*, the orbit* $G.I_0$ *is a homogeneous space.*

Now, in this modelling, a natural way to compare two shapes $m_0, m_1 \in M$ would be to evaluate the amount of deformation needed to deform m_0 on m_1. Mathematically, assuming that G is equipped with a distance d_G, one wants to induce from d_G a distance on M. This is not totally trivial because for m_0 and m_1 in M, there can be several elements $g \in G$ such that $g.m_0 = m_1$. In fact, if G_{m_0} denotes the *isotropy subgroup* of m_0, the set of such g is $g_0 G_{m_0}$ where $g_0 \in G$ is one transformation such that $g_0.m_0 = m_1$. Yet, as shown for example in [86] chapter 16, the following holds :

Theorem 1.2.1. *If* M *is a homogeneous space under the action of* G *and if* d_G *is a right-equivariant distance on* G*, i.e* $d_G(g_1 g, g_2 g) = d_G(g_1, g_2)$ *for all* $g, g_1, g_2 \in G$ *then :*

$$d_M(m_0, m_1) \doteq \inf_{g \in G} \{d_G(Id, g) / \ g.m_0 = m_1\} \tag{1.1}$$

defines a pseudo-distance on M*. It is a distance if, in addition,* G_{m_0} *is closed in* G *for the topology induced by* d_G*.*

The previous theorem implies that the construction of a distance on the shape space can be reduced to the one of defining right-invariant distances on the group G. From eq.(1.1), we see that the induced distance on M then consists in finding the 'best' deformation g (in the sense that g is closest to identity for d_G) that transforms m_0 into m_1. This also means that, formally, the study of shape variability in the set M, which has a priori no structure, can be done instead in the group G equipped with its metric d_G.

1.2.3 Large deformations

As a consequence of the previous section, we are now lead to focus on the question of the transformation group G and its metric. A very convenient setting for defining right-invariant metrics on a group is to assume that G is in addition a Lie group and to consider the special case of Riemannian metrics. Indeed, one can then define a metric on the Lie algebra of G, i.e the tangent space $T_{\mathrm{Id}}G$, and extend it to all tangent spaces by right-invariance. This is a classical construction in the case of finite-dimensional Lie groups like for instance the group of isometries, similitudes in $\mathbb{R}^n$. Nevertheless, such simple groups are clearly not rich enough in the perspective of applications to computational anatomy : one needs group of deformations that can account for complex and localized changes in shapes, which are called rightly *large deformations*.
The proper setting is thus to consider potentially infinite-dimensional subgroups of $\mathrm{Diff}(\mathbb{R}^n)$ and define right-invariant metrics on them. The general study of the geometry of infinite-dimensional Lie groups, as addressed for instance in [9, 62], is however much more technical than the finite-dimensional case and do not enter in the scope of this thesis. Yet, in [14, 27], was first introduced **Large Deformation Diffeomorphic Metric Mapping** (LDDMM), which provides one possible and more elementary tool to build such groups of diffeomorphisms and their right-invariant metrics. Since we shall be leaning on this deformation model at several points in this dissertation, we briefly recap the essential features of LDDMM. More detailed and complete presentations can be found in e.g [14], [27] or [86].
The principle of LDDMM is to consider diffeomorphisms that are obtained as *flows of time-varying vector fields* in $\mathbb{R}^n$. One starts by considering the space $C^1_0(\mathbb{R}^n, \mathbb{R}^n)$ of C^1 vector fields u such that u and du vanish at infinity, equipped with the norm

$$\|u\|_{1,\infty} \doteq \|u\|_\infty + \|du\|_\infty \,.$$

Now, following the notations of [86], one introduces the space χ^1 of time varying vector fields on the time interval $[0,1]$, that are assumed to be L^2 with respect to t. For any $v \in \chi^1$ and $t \in [0,1]$, we shall denote by v_t the vector field at time t. We have therefore, by hypothesis :

$$\int_0^1 \|v_t\|^2_{1,\infty} dt < \infty \,.$$

Note that this implies in addition, thanks to Cauchy-Schwarz inequality, that $\int_0^1 \|v_t\|_{1,\infty} dt < \infty$. Now, the first fundamental result, shown in [86] chapter 12, is the following :

Proposition 1.2.1. *Let $v \in \chi^1$. For all $x \in \mathbb{R}^n$, the differential system :*

$$\begin{cases} \dot{y}(t) = v_t(y) \\ y(0) = x \end{cases} \tag{1.2}$$

has a unique solution defined at all time $t \in [0,1]$. *Moreover, the flow application* ϕ_t^v, *where* $\phi_t^v(x)$ *is the solution of eq.*(1.2) *at time* t, *is a* C^1*-diffeomorphism of* $\mathbb{R}^n$.

Proposition 1.2.1 shows that we can generate diffeomorphisms by integration of vector fields in χ^1. From a more geometrical point of view, the Banach space $(C_0^1(\mathbb{R}^n,\mathbb{R}^n), \|.\|_{1,\infty})$ can be thought formally as the model of tangent space to Id to such diffeomorphisms. By analogy with the Riemannian idea in finite dimension, we need to introduce in addition a Hilbert structure on this set of vector fields. This is formalized by the definition below :

Definition 1.2.1. *Let* V *be a set of vector fields on* $\mathbb{R}^n$. *We say that* V *is an admissible space of vector fields if* V *is a Hilbert space and if we have a continuous injection* $V \hookrightarrow C_0^1(\mathbb{R}^n,\mathbb{R}^n)$ *i.e there exists a constant* $c_V > 0$ *such that :*

$$\forall u \in V,\ \|u\|_{1,\infty} \leq c_V \|u\|_V\,.$$

For such V, we introduce, as previously, the Hilbert space of time-varying vector fields $L_V^2 \doteq L^2([0,1],V) \subset \chi^1$. By definition, for all $v \in L_V^2$,

$$\int_0^1 \|v_t\|_V^2 dt < \infty\,.$$

We then define :

$$G_V = \{\phi_1^v \ /\ v \in L_V^2\}$$

The result of proposition 1.2.1 implies that G_V is a set of C^1-diffeomorphisms of $\mathbb{R}^n$. It is also easily verified [86] that G_V is a group. Now, a right-equivariant metric on G_V may be obtained as follows :

Theorem 1.2.2. *Let* V *be a space of admissible vector fields and* G_V *the corresponding group of diffeomorphisms. For all* $\phi \in G_V$, *let's define :*

$$d(Id,\phi) \doteq \inf\left\{\left(\int_0^1 \|v_t\|_V^2 dt\right)^{\frac{1}{2}}\ /\ \phi_1^v = \phi\right\}.$$

Then the previous infimum is reached for a certain $v \in L_V^2$ *and setting for all* $\phi, \phi' \in G_V$, $d(\phi,\phi') \doteq d(Id, \phi' \circ \phi^{-1})$, d_G *is a right-invariant metric on* G_V *and* (G_V, d_G) *is a complete metric space.*

This is proved by theorems 12.16 and 12.21 in [86]. The term $E(v) \doteq \int_0^1 \|v_t\|_V^2 dt$ will be called the **energy** : physically, it measures the amount of deformation induced on the whole space. From a Riemannian perspective, it corresponds to the square of the total length of the path $t \mapsto \phi_t^v$ in G_V. d_G can be thus interpreted formally as a *geodesic distance* on the Lie group G_V (cf [86]). The existence of a minimizing v in theorem 1.2.2 shows that there exists geodesics between any two diffeomorphisms in G_V. For such a minimizing v, we know also ([41] chapter 1) that the quantities $\|v_t\|_V$ are conserved, that is to say :

$$E(v) = E(v_0) = \|v_0\|_V^2$$

If we now summarize the two last subsections, using both theorems 1.2.1 and 1.2.2, we are able to compare objects on the original shape space M through the distance induced by d_G on M. This distance eventually writes :

$$d_M(m_0,m_1) = \inf_{v \in L_V^2}\left\{\left(\int_0^1 \|v_t\|_V^2 dt\right)^{\frac{1}{2}}\ /\ \phi_1^v.m_0 = m_1\right\}.$$

As we see, computing the distance between two shapes m_0 and m_1 amounts in solving a constrained optimization problem over vector fields in L^2_V, which is the **exact matching problem** between m_0 and m_1.

1.2.4 The kernel trick

To make the optimization problems underlying the distances d_G and d_M more explicit, one need first to address the issue of building admissible spaces of vector fields as introduced by definition 1.2.1. To achieve that, we make, in this section, a little parenthesis and remind the important and classical notion of **reproducing kernels**, that we shall extensively make use of in this work. Reproducing kernels were originally introduced in functional analysis by Aronszajn [10] and have become since then a widespread tool in the field of machine learning [47]. In our context of use, good presentations of kernels can be found either in [41] and [28] or in [86].

Reproducing kernels : general properties

As opposed to the majority of other applications of kernels that generally consider only scalar-valued ones, we shall need the more general definition of vector-valued kernels :

Definition 1.2.2. *Let A be a set and E an euclidean space of finite dimension. We say that K is a positive symmetric kernel on A with values in E (or E-positive kernel on A in short) if $K : A \times A \to \mathcal{L}(E)$ and satisfies :*
1. For all $x, y \in A$, $K(y,x) = K(x,y)^T$
2. For all $N \in \mathbb{N}^*$, $x_1, ..., x_N \in A$ and $\alpha_1, ..., \alpha_N \in E$, we have :

$$\sum_{i,j=1}^{N} \alpha_i.(K(x_i, x_j)\alpha_j) \geq 0\,.$$

An equivalent definition is to say that for all N and $x = (x_1, ..., x_N) \in A^N$, the operator $K(x,x)$ on E^N given by the block matrix $(K(x_i, x_j))_{i,j=1,..,N}$ is symmetric positive. In the specific case where $E = \mathbb{R}$, the previous definition gives again the more usual one of (scalar) positive kernels on A. Throughout the text, we shall reserve capital letters for vector-valued kernels and small ones for scalar kernels. At the same time, we define the following notion :

Definition 1.2.3. *Let H be a Hilbert space of functions defined on a set A with values in an euclidean space E. Then, H is said to be a Reproducing Kernel Hilbert Space (RKHS) if for all $x \in A$ and $\alpha \in E$, the functional $\delta_x^\alpha : f \in H \mapsto f(x).\alpha$ is a continuous linear form on H.*

As one could expect, the two notions of positive kernel and RKHS are closely related. One of the link is easy to evidence. If H is a RKHS as in definition 1.2.3, then, thanks to Riesz-Frechet theorem, we know that for all $x \in A$ and $\alpha \in E$, there exists an element $\mathbf{K}_H \delta_x^\alpha \in H$ such that :

$$\forall f \in H, \delta_x^\alpha(f) = \langle \mathbf{K}_H \delta_x^\alpha, f \rangle_H$$

In addition since $\mathbf{K}_H \delta_x^\alpha$ is in H, it is itself a function from A to E that we can write in the form $\mathbf{K}_H \delta_x^\alpha = K(x,.)\alpha$. The functions $K(x,.)\alpha$ are called the fundamental

functions of the RKHS. Applying the previous equation to $f = K(y,.)\beta$ we also get :

$$\langle K(x,.)\alpha, K(y,.)\beta\rangle_H = \alpha.(K(x,y)\beta)\,. \tag{1.3}$$

It is then straightforward to check that the function K, that we call the *reproducing kernel* of H, defines an E-positive kernel on A. In fact, the following central theorem, due to Aronszajn, shows that we have a converse property, namely :

Theorem 1.2.3. *To any positive kernel K on A with values in E, there corresponds a unique RKHS H of functions from A to E whose reproducing kernel is K.*

We refer the reader to [10] or [41] chapter 2 for the proof of this result. The RKHS H is basically constructed, up to some technicalities, as the completion of the vector space spanned by the fundamental functions $K(x,.)\alpha$ for all $x \in A$ and $\alpha \in E$. In particular, this vector space is dense in H. The Hilbertian metric $\langle .,.\rangle_H$ can be computed in closed form on fundamental functions exactly as in eq.(1.3).

RKHS of admissible vector fields

We now come back more specifically to the construction of spaces of admissible vector fields. We shall assume, from now on, that kernels are defined on $A = \mathbb{R}^n$ and taking values in $E = \mathbb{R}^n$. With theorem 1.2.3, the RKHS generated by such kernels are Hilbert spaces of vector fields in $\mathbb{R}^n$. One only needs the additional continuous injection property for V. The answer is provided by the following result, that can be found in [41] :

Proposition 1.2.2. *Let $p \in \mathbb{N}$. Assume that K is a positive kernel such that K is continuously differentiable of order $2p$ and such that all the derivatives of K up to the order p are bounded. Assume in addition that for any $x \in \mathbb{R}^n$, the function $K(x,.)$ and all its derivatives up to order p vanish at infinity. Then, the RKHS associated to K is continuously embedded into $C_0^p(\mathbb{R}^n, \mathbb{R}^n)$.*

In particular, if a positive kernel K satisfies the hypotheses of the previous proposition for $p = 1$, we can conclude that its corresponding reproducing space V is continuously embedded into $C_0^1(\mathbb{R}^n, \mathbb{R}^n)$ and thus V is automatically an admissible space of vector fields.
The construction of such kernels is well-detailed in [41]. Certain Sobolev spaces are examples of admissible RKHS. Another interesting class is provided by radial scalar kernels, which have the general expression :

$$K(x,y) = h(|x-y|).\mathrm{Id}_{\mathbb{R}^n}$$

with $h : \mathbb{R}_+ \to \mathbb{R}$. Schoenberg theorem (cf [86] chapter 13) shows that K is a positive kernel when the function h is the Laplace transform of a positive Borel finite measure on $\mathbb{R}_+$. The Cauchy kernel given by :

$$K(x,y) = \frac{1}{1 + \frac{|x-y|^2}{\sigma^2}}.\mathrm{Id}$$

is for instance one possibility. Yet the most widely used in applications remains the Gaussian kernel :

$$K(x,y) = e^{-\frac{|x-y|^2}{\sigma^2}}.\mathrm{Id}\,.$$

Proposition 1.2.2 shows indeed that its associated RKHS is continuously embedded into $C_0^p(\mathbb{R}^n, \mathbb{R}^n)$ for all p. The parameter σ controls the scale at which two points interact with each other. There are in addition many building tools to create new kernels from existing ones such as making sums, tensor products (cf lemma 2.2.1 in chapter 2)... The properties of sums of kernels with respect to diffeomorphism groups, for instance, are studied in [19]. Higher order kernels are also considered and used in [74].
As we shall see in the coming sections, the interest of using kernels and RKHS in this context is multiple : it gives Hilbert metrics on vector fields that can be conveniently computed in practice but the essential point is that the use of RKHS as admissible spaces of vector fields provide a very powerful reduction principle to parametrize geodesics in the group G_V, as we detail below in the case of landmarks.

1.2.5 Geodesic equations in the landmark case

At this point, even though we have been able to formulate the shape comparison problem by defining geodesic distances in groups of diffeomorphisms, the optimization problem giving these geodesics is still on the set of all possible time-varying vector fields, which is infinite-dimensional. As one could intuit, such geodesics could be described by less degrees of freedom, at least because, by analogy with the finite-dimensional Riemannian setting, they should be determined only by the vector field at time 0 thanks to the corresponding Euler-Lagrange equations. This has been formalized precisely, in the general case, by the interpretation of the equations as an optimal control problem : we refer to [65, 79] or [77] for more details.
To simplify a little the presentation, we shall restrict ourselves to the simpler case where the shape m_0 is made of a finite number of points $q = (q_i)_{i=1,..,p}$, which is the landmark situation described above. This shall be generic enough for the cases we consider in this work since, numerically, all shapes (unlabeled point sets, curves, surfaces,...) are indeed made of finite number of points (with possible additional mesh structure). Yet, the vector fields that describe geodesics are a priori still living in the infinite-dimensional space L^2_V. Let's call K_V the positive kernel on vector fields. The great reduction principle provided by the use of RKHS to model vector fields is that the optimal vector fields generating geodesics are necessarily of the form :

$$v_t = \sum_{i=1}^{p} K_V(q_{t,i}, .)p_{t,i} \tag{1.4}$$

where $q_{t,i} \doteq \phi_t^v(q_i)$ and the $p_{t,i}$ are auxiliary variables in $\mathbb{R}^n$ called the **momenta** of the deformation. This is proved for example in [41], and it is essentially a particular instance of the general kernel reduction principle on optimization problems from machine learning. Thus the differential equation that governs the time evolution of q can be written in a condensed way as :

$$\dot{q}_t = K_V(q_t, q_t)p_t$$

where $K_V(q_t, q_t)$ is by convention the $3p \times 3p$ matrix made of the blocks $K_V(q_{t,i}, q_{t,j})$ for $i, j = 1, .., p$.
As we mentioned earlier, the geodesics on G_V should be in addition determined only by the initial velocity field v_0 or equivalently by the **initial momentum** p_0. Indeed, it is well-known from [65] that the momenta $p_{t,i}$ are governed by the differential

equations :

$$\dot{p}_{t,i} = -(d_{q_{t,i}} v_t)^* p_{t,i}$$

which we can write in short :

$$\dot{p}_t = -(d_{q_t} v_t)^* p_t$$

Interestingly, the set of coupled differential equations on q and p can be interpreted as a Hamiltonian system of equations by introducing the following *reduced Hamiltonian* :

$$H_r(p,q) = \sum_{i,j=1}^{p} p_i^T K_V(x_i, x_j) p_j = p^T K_V(x,x) p \,. \tag{1.5}$$

Observe that $H_r(p_t, q_t)$ equals also the previously defined energy of the vector field v_t at time t. The dynamics of the position and momentum variables along a geodesic can be described by the Hamiltonian system :

$$\begin{cases} \dot{q}(t) = \dfrac{\partial H_r}{\partial p} \\ \dot{p}(t) = -\dfrac{\partial H_r}{\partial q} \,. \end{cases} \tag{1.6}$$

Note that the Hamiltonian is always conserved during the evolution of p and q, which is exactly the previous conservation property of energy along geodesics that writes :

$$\forall t \in [0,1],\ H_r(p_t, q_t) = H_r(p_0, q_0) = E(v_0) \,.$$

The Hamiltonian formulation thus completes the theoretical description of geodesics in G_V by showing in a precise way how all geodesics can be completely parametrized by a couple (p_0, q_0) of initial momentum and position. As we shall see in the next section, this is the fundamental idea behind geodesic shooting procedures to solve registration problems in practice. Moreover, this parametrization with initial momentum also enables the statistical analysis of deformations as developed in [79] and that we shall exploit in chapter 4.

1.3 Inexact registration and shape dissimilarities

1.3.1 Inexact matching formulation

The important assumption in Grenander's approach of shape spaces, which has not been discussed so far, is that the action of the deformation group is transitive on the set of shapes. In terms of metric, this hypothesis enables to nearly forget in a sense what the shapes in M really are (landmarks, images, curves...) because the metric is actually computed instead in the group of diffeomorphisms acting on M. Nevertheless, the transitiveness of the action is clearly not a reasonable assumption in most practical situations. First, because the groups of deformations we build with the previous framework do not contain all diffeomorphisms of the space. And second because one could consider shapes that need not even be diffeomorphic between each other as a result of noise for instance or because of varying number of connected components (as for instance when shapes are fiber bundles). Thus, the exact matching situation considered until now appears insufficient in general to efficiently estimate shape variations on real datasets. Let's also mention that, even if two shapes do belong to the same orbit, exact matching could be still inappropriate.

The diffeomorphism that maps one on the other might not be relevant from the statistical analysis point of view because it could be essentially driven by the matching of high-frequency variations or noise that may have no anatomical meaning at all.
For all these reasons, we generally replace the previous constrained optimization problem involved in the computation of d_M by an **inexact registration** in the shape space M. It simply consists in relaxing the constraint $\phi_1^v.m_0 = m_1$ by solving instead :

$$\inf_{v \in L_V^2} J(v) \doteq \int_0^1 \|v_t\|_V^2 dt + A(\phi_1^v.m_0, m_1)\,. \tag{1.7}$$

In that expression, A stands for what is generally called the **data attachment term**, i.e a function that measures the closeness between elements in M. As opposed to the previous distance in M based on deformation, the term A aims at driving the registration by measuring the residual dissimilarity between the deformed source shape $\phi_1^v.m_0$ and the target m_1. From a formal point of view, the optimization problem described by eq.(1.7) looks similar to many others that involve the sum of an attachment term and a regularization one (which here constrains the deformation energy).
Setting aside for the moment the construction of these data attachment functionals, a fundamental question is the existence of a solution to this minimization problem. This is no more guaranteed in this context. Fortunately, the results shown in [41] chap.1 give a condition on A for the existence of such a minimizer. The general result is the following :

Theorem 1.3.1. *If, for all m_0 and m_1 in M, the functional $v \mapsto A(\phi_1^v.m_0, m_1)$ is weakly continuous from L_V^2 to $\mathbb{R}$, then the inexact matching problem given by eq.*(1.7) *has a solution.*

The complementary issue of uniqueness of the minimizer is however totally unknown at this degree of generality.

1.3.2 Geodesic shooting algorithm

For a given data attachment functional A satisfying the conditions of proposition 1.3.1, an optimal solution v of eq.(1.7) do not map exactly m_0 on m_1 anymore but on $\phi_1^v.m_0$, hopefully close enough (in the sense given by A) to the target m_1. Yet the presence of the energy term in the optimized functional guarantees that, in the space of diffeomorphisms, we still obtain a geodesic path $t \mapsto \phi_t^v$. Thus, we can restrict the optimization only to geodesic trajectories and, as explained previously, the dynamics of the deformation can be then described by the initial momentum p_0 and the Hamiltonian system of equations. It results in particular that the functional $J(v)$ can be written instead as a functional of the initial momentum p_0. The principle of geodesic shooting algorithms is to optimize directly on p_0 and 'shoot' the source shape with the geodesic equations.
To be more explicit, let's assume, as in section 1.2.5, that the source shape consists in a finite set of points i.e that we have $q_0 = (q_{0,i})_{i=1,..,p}$. The data attachment term $A(\phi_1^v.m_0, m_1)$ is then a function of the final point configuration q_1 obtained by shooting through eq.(1.6), that we write $g(q_1)$. Then the functional $J(p_0)$ equals :

$$J(p_0) = H_r(p_0, q_0) + g(q_1)$$

The simplest way to solve this optimization problem is to apply a gradient descent scheme on p_0, as described in [65] and [7]. The gradient with respect to p_0 of

$H_r(p_0, q_0)$ is straightforward from eq.(1.5). The gradient of the term $g(q_1)$ is obtained by first computing the gradient of the attachment term with respect to q_1, i.e $\nabla_{q_1} g$, and pull it back on the shape q_0 by **backward integration of the linearized adjoint Hamiltonian system**, as explained in [5, 7]. Namely, one introduces the auxiliary variables P_t and Q_t that satisfies :

$$\begin{cases} \dot{Q}_t = -(\partial^2_{q,p} H_r(q_t, p_t))^* Q_t + (\partial^2_{q,q} H_r(q_t, p_t))^* P_t \\ \dot{P}_t = -(\partial^2_{p,p} H_r(q_t, p_t))^* Q_t + (\partial^2_{q,p} H_r(q_t, p_t))^* P_t \\ Q_1 = -\nabla_{q^1} g, \ P_1 = 0 \end{cases} \tag{1.8}$$

and the gradient of J with respect to p_0 is then given by :

$$\nabla_{p_0} J = K_V(q_0, q_0) p_0 - P_0 \,.$$

Thus, at each step, the gradient can be computed by flowing forward the Hamiltonian system of eq.(1.5) with the current initial momentum p_0 to get the final point positions q_1, then computing $\nabla_{q_1} g$ and flowing backward in time the adjoint equations eq.(1.8). This makes a gradient descent algorithm on p_0 completely achievable as soon as the dissimilarity function is fixed and that one can compute its differentials. The registration algorithms built in chapter 2 and 3 are based on this principle.
Let's also mention that some additional equations can be also derived by expressing that the variation of $J(v)$ with respect to any perturbation of v in L^2_V (not only in the space of geodesic trajectories) vanishes when the minimum is reached. In particular (cf [5] chap. 2), one can show that the optimal momentum at time 1 p_1 satisfies a *transversality equation* which is :

$$p_1 + \nabla_{q^1} g = 0 \,. \tag{1.9}$$

This transversality equation can be used as an alternative to gradient descent algorithms : for instance, Allassonnière et al. propose in [5] a Newton algorithm to compute iteratively p_1 with eq.(1.9).

1.3.3 Measuring shape dissimilarities

We finally reach the question of defining these dissimilarity functions on shapes, which is not specific, by the way, to the LDDMM framework, but is a requirement to most matching algorithms. It must be emphasized that this is not to be necessarily identified to the problem of building metrics on shape spaces. Here, the dissimilarity must be thought as related to a notion of residual noise on shapes that links the observed shapes with the ideal ones all living in a common orbit. In fact, the function A need not be a distance in all cases. If one intends to give a kind of axiomatic approach for the construction of such A's, one would impose that A satisfies the following general principles :

- A does not depend on a particular representation of shapes (for instance parametrization for curves or surfaces).
- $A(m_1, m_0) = A(m_0, m_1)$ for all $m_0, m_1 \in M$.
- $A(m_0, m_1) = 0$ if and only if $m_0 = m_1$.
- A smoothly behave with respect to smooth deformations of shape, for instance $\phi \mapsto A(\phi.m_0, m_1)$ is differentiable on any diffeomorphism group G_V.

The case where A is a metric on M is thus a stronger assumption but usually quite convenient. Anyhow, the choice of A must be done in accordance to the nature of

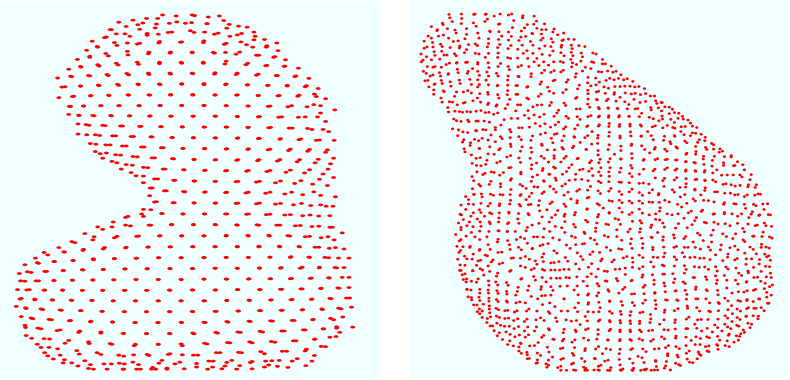


Figure 1.2: Two distributions of unlabeled points in $\mathbb{R}^2$.

the shapes contained in the space M. In the rest of this section, we briefly review the most usual types of shapes that can be encountered in applications.
The case of *landmarks* does not pose any particular difficulty in that respect. Indeed, by definition, landmarks are labeled set of points so that correspondences are known. Thus, one can measure the dissimilarity between landmarks just by the L^2 distance. If $x = (x_1, .., x_p)$ and $y = (y_1, .., y_p)$ belong to $M = (\mathbb{R}^n)^p$ then :

$$A(x, y) = \sum_{k=1}^{p} |x_k - y_k|^2$$

where $|.|$ is the L^2 norm in $\mathbb{R}^n$. One can also extend this L^2 distance by changing the metric through a covariance matrix as proposed in [50].
For *images*, the situation is quite similar. If $\Omega \subset \mathbb{R}^n$ is the domain of the images then a natural metric to compare two images I_0 and I_1 is the L^2 distance on Ω, i.e :

$$A(I_0, I_1) = \int_\Omega (I_0 - I_1)^2(x)dx$$

It has been widely used as an attachment term in most of LDDMM-based registration applications to images.
More difficulties appear however in the case of *unlabeled sets of points*. This corresponds to the case where shapes are distribution of points in $\mathbb{R}^n$ as in figure 1.2 but, unlike landmarks, no correspondences are known. Nor do the distributions necessarily have the same number of points. The definition of relevant attachment functionals in that case is thus not a trivial matter and has fostered efforts in several directions. In the field of computational anatomy, it has been addressed in particular by Glaunès in [42], where authors model such point distributions as Borel measures and define kernel Hilbert metrics on these. This can be actually viewed as a particular case of currents that we present in more details in the following.
Even more challenging is the case of *curves, surfaces* and more generally of *submanifolds*. Indeed, the previous requirement of parametrization-invariance on A makes it a particularly delicate issue and has lead to many interesting approaches that it would be hard to present in an exhaustive way. We could mention, among others, the parametrization-invariant quotient Riemannian setting on the space of plane curves introduced by Michor & Mumford in [63], dissimilarities based on the Hausdorff distance as in [23] for compact subsets of $\mathbb{R}^2$ and $\mathbb{R}^3$, or approximations of the Gromov-Hausdorff distance [61, 18]. Most of these frameworks are however

specific to certain class of data or are very sensitive to the topology of shapes for some of them, which can represent an obstacle in using them for attachment metrics on certain real datasets that are likely to present many topological irregularities.
In contrast, the approach of *currents* initiated in [43, 28] and inherited from geometric measure theory, has the advantage of providing a common framework to embed practically any case of submanifolds in a functional space, which enables the definition of inner products via the use of reproducing kernels. In terms of applications, such metrics present the great interest that the resulting attachment terms and their variations are easily computable and robust to small topological changes as disconnections or holes. We recap the main definitions and properties related to currents in the following section.

1.3.4 The representation by currents

Currents were historically introduced as a generalization of distributions by L. Schwartz and then G. De Rham in [70]. The theory was later developed and connected to geometric measure theory in large part by H. Federer [37]. At first, these results found interesting applications in the calculus of variations as well as differential equations. However, the use of currents in the field of computational anatomy is more recent, initially proposed in [41]. Without pretending to make a complete or self-contained presentation of the vast theory of currents, we try to outline, in this section, the minimum theoretical background needed for the rest of this thesis.

Rectifiable sets

We start by the notion of rectifiable set. From now on, we fix $E = \mathbb{R}^n$ the embedding space of shapes. Rectifiable subsets can be thought as the proper generalization of submanifolds in the context of geometric measure theory. We shall use the classical notation $\mathcal{H}^d$ for the d-dimensional **Hausdorff measure** on E. We remind that $\mathcal{H}^d$ is defined as an outer measure on E that basically measures the d-dimensional volume of a subset in E. In particular, we have $\mathcal{H}^n = \lambda^n$, the usual Lebesgue measure. If X is a p-dimensional submanifold of E then $\mathcal{H}^d(X)$ is rigorously the d-volume of X if $p = d$, vanishes if $p < d$ and equals $+\infty$ whenever $d < p$.
Now, let d be an integer between 0 and n. With a slight abuse of definitions, we will call a d-dimensional rectifiable subset of E what is actually defined as a countable $\mathcal{H}^d$-rectifiable subset in [37] 3.2.14, i.e :

Definition 1.3.1. *We say that X is a d-dimensional rectifiable subset of E if $\mathcal{H}^d(X) < \infty$ and if there exists a countable family $(f_i)_{i\in\mathbb{N}}$ of Lipschitz maps $f_i : \mathbb{R}^d \to E$ such that :*

$$\mathcal{H}^d\left(X \setminus \bigcup_{i\in\mathbb{N}} f_i(\mathbb{R}^d)\right) = 0\,.$$

In other terms, a rectifiable subset is almost everywhere covered by a countable union of images of Lipschitz maps from $\mathbb{R}^d$ to E. Note that this is analogous to the definition of a d-dimensional submanifold as a graph. The analogy can be pushed further since, by Rademacher theorem, one knows that a Lipschitz function $f : \mathbb{R}^d \to E$ is differentiable λ^d almost everywhere on $\mathbb{R}^d$ and thus (cf [67]) an equivalent definition is given by :

Property 1.3.1. *A subset $X \subset E$ is d-rectifiable if and only if $\mathcal{H}^d(X) < \infty$ and $\mathcal{H}^d$-almost all of X is contained in a countable union of d-dimensional C^1 submanifolds.*

A second very useful structural property is that a rectifiable subset X, as a submanifold, has $\mathcal{H}^d$-almost everywhere on X a d-dimensional tangent subspace. We refer to [37] 3.2.16 and 3.2.19 for the associated definitions of approximate tangent cones and the precise statement of this result. In this context, an *orientation* of a rectifiable subset X is simply a choice of an orientation for each of these tangent spaces, such that $x \mapsto T_x X$ is $\mathcal{H}^d$-measurable on X. We will say then that X is an *oriented rectifiable subset*. Note that, in this setting, unorientable submanifolds can be still given an orientation in the sense of rectifiable subsets but, as we shall examine more carefully in chapter 3, such orientations induce cancellations effects when represented as currents.
Rectifiable subsets obviously include submanifolds and submanifolds with boundary. From the structural point of view of measure, these objects actually behave much like submanifolds. Yet, this notion encompasses a much wider variety of interesting cases. For instance, all polyhedral curves, surfaces, etc.., which are the actual objects in computational geometry, are still rectifiable. Despite the previous properties, note however that rectifiable subsets may actually 'look' extremely irregular since the category of rectifiable subsets also contain certain fractal shapes which may not be a submanifold at any of their points (cf the example related to figure 3.12.2 in [67]).

Currents

We now come to the concept of currents strictly speaking. As mentioned earlier, currents were originally introduced as extended distributions that were meant to carry information of local geometry. We will adopt the following notations and definitions more or less similar to [28] :
- $\Lambda^p E$ $(0 \leq p \leq n)$: p-times exterior power of E, which is a vector space of dimension $\binom{n}{p}$ spanned by the set of simple p-vectors $\xi_1 \wedge .. \wedge \xi_p$. We remind that, by construction, the wedge product operation is bilinear and antisymmetric. Note also that for $p = 1$ or $p = n - 1$, $\Lambda^p E = E$.
- $\Lambda^p E$ is equipped with the euclidean metric given for two simple p-vectors $\xi = \xi_1 \wedge .. \wedge \xi_p$ and $\eta = \eta_1 \wedge .. \wedge \eta_p$ by the determinant of the Gram matrix $\langle \xi, \eta \rangle = \det(\langle \xi_i, \eta_j \rangle)_{i,j}$. In particular, $|\xi|$ gives the volume of the corresponding parallelotope.
- $\Lambda^p E^*$ is the dual vector space and can be also identified ([28] appendix A) to the space of alternated p-forms on E.
- $\Omega_0^p(E) \doteq C_0^0(E, \Lambda^p E^*)$: the set of continuous p-dimensional differential forms on E vanishing at infinity. If $dx^1, ..., dx^n$ is a basis of E^*, any $\omega \in \Omega_0^p(E)$ can be written under the general form :

$$\forall x \in E, \quad \omega_x = \sum_{1 \leq i_1 < ... < i_p \leq n} a_{i_1,...,i_p}(x) dx^{i_1} \wedge ... \wedge dx^{i_p}$$

where $a_{i_1,...,i_p}$ are continuous functions, vanishing at infinity.
- $\Omega_0^p(E)$ is equipped with the infinite norm

$$\|\omega\|_\infty = \sup_{x \in E} \sup_{|\xi|=1} \omega_x(\xi)$$

which makes it a Banach space.
- The previous definitions can be extended to k-times continuously differentiable differential forms. We will call d the classical exterior differentiation operator on differential forms and denote by $\Omega_{0,k}^p(E)$ the space of k-times continuously differentiable ω such that ω and its differentials up to order k vanish at infinity.

This leads now to the following definition of p-currents :

Definition 1.3.2. *The space of p-currents on E is the space $\Omega_0^p(E)'$ of all continuous linear forms on the space of differential forms.*

We draw attention on the fact that this definition of currents we adopt in our framework is actually a little more restrictive than the classical one of [37, 67], where authors define currents as elements of the dual of the smaller space of C^∞ compactly-supported differential forms. One reason for this slight nuance in the definition is essentially that it defines currents as a strict generalization of finite signed Borel measures, already introduced in computational anatomy for unlabeled point sets [41]. Indeed, in the particular case $p = 0$, the previous definition is exactly the one of usual distributions on E (dual of the space $C_0(E, \mathbb{R})$). Just as for distributions, simplest examples of currents are given by **Dirac currents** δ_x^ξ with $x \in E$, $\xi \in \Lambda^p E$ such that for any differential form $\omega \in \Omega_0^p(E)$, we have $\delta_x^\xi(\omega) = \omega_x(\xi)$.
Now, the relationship between shapes and currents lies fundamentally in the fact that *every d-dimensional oriented submanifold X of E of finite volume can be represented by an element of $\Omega_0^d(E)'$*. It is indeed a classical result from integration theory ([16]) that any d-dimensional differential form can be integrated along X and thus :

$$T_X : \ \omega \mapsto \int_X \omega \tag{1.10}$$

defines an element of $\Omega_0^d(E)'$. If, for instance, X is given through a parametrization given by a certain smooth immersion $\gamma : U \to E$ with U an open subset of $\mathbb{R}^d$, then we have explicitly :

$$T_X(\omega) = \int_U \omega_{\gamma(u)} \left(\frac{\partial \gamma}{\partial u_1} \wedge ... \wedge \frac{\partial \gamma}{\partial u_d} \right) du_1 ... du_d \, .$$

Observe that a simple application of the change of variables formula shows that this expression is indeed independent of any positive parametrization. The representation of X by the current T_X is thus **geometric** in the sense that it only depends on the shape itself and not its parametrization.
As one can imagine, the space of d-currents is much larger and contains many other interesting objects, among them are the oriented rectifiable subsets of E, for which one has a formula similar to eq.(1.10). Indeed, as explained previously, for any d-dimensional oriented rectifiable subset X, there exists for $\mathcal{H}^d$-almost all $x \in X$ an oriented tangent space and thus a unit d-vector $\tau(x) = \tau_1(x) \wedge ... \wedge \tau_d(x)$, where $\tau_1(x), ..., \tau_d(x)$ is an orthonormal basis of the tangent space. Then, similarly to the submanifold case, to X corresponds the d-current :

$$T_X(\omega) \doteq \int_X \omega_x(\tau(x)) d\mathcal{H}^d(x) \, . \tag{1.11}$$

Again, this expression does not depend on the choice of the oriented orthonormal basis $\tau_1, ..., \tau_d$ at each point $x \in X$. As easily checked, if two rectifiable subsets X and Y are such that $T_X = T_Y$, then X and Y coincide up to set of zero Hausdorff measure.
The *transport* of currents by diffeomorphisms of E can be easily expressed through the pull-back and push-forward operations. These are defined for all $T \in \Omega_0^d(E)'$ and $\phi \in \mathrm{Diff}(E)$ by the relations :

$$\left(\phi_\sharp T\right)(\omega) = T\left(\phi^\sharp \omega\right) \tag{1.12}$$

where $\phi^\sharp\omega$ is the pull-back of a differential form by ϕ defined for all $x \in E$ and $\xi = \xi_1 \wedge ... \wedge \xi_p \in \Lambda^p E$:

$$\left(\phi^\sharp\omega\right)_x (\xi) = \omega_{\phi(x)}(d_x\phi(\xi_1) \wedge ... \wedge d_x\phi(\xi_p)) \tag{1.13}$$

$d_x\phi$ being the notation we use for the differential of the diffeomorphism at point x. The reason of such definitions is that it gives a compatibility between the transport of shapes and the one of currents :

Property 1.3.2. *For any rectifiable subset X of E and diffeomorphism $\phi \in Diff(E)$, we have :*

$$\phi_\sharp T_X = T_{\phi(X)}\,.$$

This is proved by [37] 4.1.30.
Finally, this formalism adapts with no changes to the case of discrete geometry. Any discrete shape given as a set of points with a mesh (thus a polyhedron) can be transcribed into a current by approximating each cell of the mesh with one Dirac current located at the center with a simple d-vector encoding the local volume element, as explained with more details in [28] chapter 1. It results that discrete d-dimensional shapes can be always approximated by finite sums of Dirac currents of the form

$$T_X \approx \sum_{i=1}^{m} \delta_{x_i}^{\xi_i}\,.$$

The approximation error can be basically controlled by the maximum diameter of the mesh cells (cf [28] 1.3.2).
Thus the introduction of currents allows us to embed all rectifiable sets of given dimension into one common functional space on which one can define more easily a metric structure.

Metrics on currents

As the dual of the space of differential forms, the space of currents naturally inherits its operator norm, which is called the **mass norm** in [37, 67]. For any current $T \in \Omega_0^d(E)'$, it equals :

$$M(T) \doteq \sup_{\|\omega\|_\infty \leq 1} |T(\omega)|\,. \tag{1.14}$$

From eq.(1.11) and (1.14), it is straightforward that, for any rectifiable set X, we have in particular :

$$M(T_X) = \mathcal{H}^d(X)$$

Yet the mass norm does not induce a satisfying distance between shapes as noticed in [28] 1.5.1 : for instance, if X and Y are two distinct submanifolds in E then one easily sees that $M(T_X - T_Y) = \mathcal{H}^d(X) + \mathcal{H}^d(Y)$ no matter how close in distance the points of X might be from the ones of Y.
This drawback results essentially from the fact that the set of test functions ω in eq.(1.14) is too big. For that reason, the alternative **flat norm** is often considered in geometric measure theory. It is defined by :

$$F(T) \doteq \sup_{\substack{\|\omega\|_\infty \leq 1 \\ \|d\omega\|_\infty \leq 1}} |T(\omega)| \tag{1.15}$$

It restrains the test differential forms to be in $\Omega^d_{0,1}(E)$ and adds a control on $d\omega$. It is argued in [67] that the flat norm is much better fitted to shape comparison because the distance is basically controlled by the $(d+1)$-volume delimited by the two shapes. Unfortunately, the flat norm has no closed form in general which makes it virtually impossible to use in any numerical purpose.
To overcome this issue, the idea of introducing once again Reproducing Kernel Hilbert Spaces was proposed in [41]. Instead of considering linear forms on $\Omega^d_0(E)$, one intends to restrict to a Hilbert space W embedded into $\Omega^d_0(E)$. A convenient setting for building such Hilbert spaces of differential forms is the one of reproducing kernels presented in section 1.2.4. We shall use the identification of $\Lambda^d E^*$ with $\Lambda^d E$ based on the inner product defined previously on $\Lambda^d E$ to consider indifferently those two spaces. That being said, with the terminology of section 1.2.4, we are lead to consider in this case $\Lambda^d E$-kernels on E and we have the exact same property as in proposition 1.2.2 :

Proposition 1.3.1. *Assume that K is a $\Lambda^d E$-positive kernel on E such that K is continuously differentiable of order $2k$ and such that all the derivatives of K up to the order k are bounded. Assume in addition that for any $x \in E$, the function $K(x,.)$ and all its derivatives up to order k vanish at infinity. Then, W, the RKHS associated to K, is continuously embedded into $\Omega^d_{0,k}(E)$.*

If the previous assumptions on the kernel are fulfilled for $k = 0$, W is then continuously embedded into the space of differential forms $\Omega^d_0(E)$. We then obtain the dual Hilbertian structure on W' that induces a distance on the space of d-currents provided W is dense in $\Omega^d_0(E)$. One can also show a few straightforward control properties of such metrics. Indeed, the continuous embedding $W \hookrightarrow \Omega^d_0(E)$ gives the existence of a constant $c_W > 0$ such that for all $\omega \in W$:

$$\|\omega\|_\infty \leq c_W \|\omega\|_W$$

Now, for any $T \in W'$, since $\|T\|_{W'} = \sup_{\|\omega\|_W=1} |T(\omega)|$, it results easily that :

$$\forall T \in W', \quad \|T\|_{W'} \leq c_W M(T) \tag{1.16}$$

and in the special case of currents T_X where X is a rectifiable subset, we see that :

$$\|T_X\|_{W'} \leq c_W \mathcal{H}^d(X) \tag{1.17}$$

A similar control can be also obtained by the flat norm if one makes the stronger assumption that K satisfies the hypotheses of proposition 1.3.1 for $k = 1$, implying that $W \hookrightarrow \Omega^d_{0,1}(E)$. In that case, there is a constant $\tilde{c}_W$ such that $\|\omega\|_\infty + \|d\omega\|_\infty \leq \tilde{c}_W \|\omega\|_W$ and just as before :

$$\forall T \in W', \quad \|T\|_{W'} \leq \tilde{c}_W F(T) \tag{1.18}$$

Last but not least, the fact that W is a RKHS implies in particular that the Dirac currents δ^ξ_x belong to W' and the reproducing kernel property gives in addition that for all $x_1, x_2 \in E$ and $\xi_1, \xi_2 \in \Lambda^d E$:

$$\begin{aligned} \langle \delta^{\xi_1}_{x_1}, \delta^{\xi_2}_{x_2} \rangle_{W'} &= \langle K(x_1,.)\xi_1, K(x_2,.)\xi_2 \rangle_W \\ &= \langle \xi_1, K(x_1,x_2)\xi_2 \rangle \end{aligned} \tag{1.19}$$

As a consequence, in the discrete setting where shapes are represented as finite sums of Dirac, we see that the RKHS distances can be computed in closed form as double sums of kernel evaluations. Thus, we obtain a corresponding attachment terms between any shapes X and Y by setting

$$A(X,Y) = \|T_X - T_Y\|^2_{W'} .$$

It can be both easily computed and differentiated with respect to the point positions of the shapes (we refer to [41] for the whole extended discrete expressions). The numerical issues behind such computations are addressed in [28] and will be also discussed later on in chapter 2. With these attachment terms, the registration problem between curves, surfaces or submanifolds can be therefore treated by the geodesic shooting algorithm presented in 1.3.2.

Cartan's formula and differentiation of the metrics

To conclude this quick presentation of currents, let us mention the important requirement of differentiability of the previous metrics on shapes that was evoked above and will be a recurrent issue in this work. Besides the computations of these gradients for discretized shapes, that do not pose any special theoretical issue, it seems also interesting to understand, in the continuous setting, how these metrics behave with respect to deformations of shapes. In the context of currents, this can be answered based on Cartan's formula for differential forms.
Cartan's formula is a classical result from differential geometry that gives an expression of the Lie derivative of differential forms with respect to vector fields. For a given stationary, smooth and compactly supported vector field v on E, we introduce, similarly to section 1.2.3, its flow at time t that we will denote ϕ_t. Now, if $\omega \in \Omega^d_{0,1}(E)$ is a C^1 d-differential form on E, the *Lie derivative* of ω in the direction of v is the d-differential form $\mathcal{L}_v\omega$ that one can define by :

$$\mathcal{L}_v\omega \doteq \left.\frac{d}{dt}\right|_{t=0} (\phi_t)^\sharp\omega \,. \tag{1.20}$$

Cartan's formula basically connects this notion of derivative with the more usual exterior differentiation d. For this, one needs to introduce the contraction operator on differential forms. For a vector field v as previously, the contraction operation $\imath_v : \Omega^{d+1}_0(E) \to \Omega^d_0(E)$ is obtained by :

$$\forall \omega \in \Omega^{d+1}_0(E), \quad (\imath_v\omega)_x(\xi) \doteq \omega_x(v(x) \wedge \xi) \tag{1.21}$$

for any $x \in E$ and $\xi \in \Lambda^d E$. Now the result is the following :

Proposition 1.3.2. *For all $\omega \in \Omega^d_{0,1}(E)$ and smooth compactly supported vector field v of E,*

$$\mathcal{L}_v\omega = \imath_v(d\omega) + d(\imath_v\omega) \,.$$

The proof can be done in an elementary way by differentiating directly $(\phi_t)^\sharp\omega$ using eq.(1.13) and reason by induction on the dimension d as done for instance in [71]. Other proofs based on more abstract constructions can be also found in [52]. This result on differential forms gives an interesting corollary in terms of currents associated to oriented submanifolds :

Theorem 1.3.2. *Let X be a compact oriented submanifold (possibly with boundary), v a compactly supported smooth vector field and ϕ_t its flow. Then for any C^1 d-differential ω on E :*

$$\left.\frac{d}{dt}\right|_{t=0} T_{\phi_t(X)}(\omega) = \int_X \mathcal{L}_v\omega = \int_X \imath_v(d\omega) + \int_{\partial X} \imath_v\omega\,.$$

Proof. It is almost straightforward from Cartan's formula. Indeed, thanks to property 1.3.2, we have :

$$T_{\phi_t(X)}(\omega) = \left(\phi_t^\sharp T\right)(\omega) = T_X((\phi_t)^\sharp\omega) = \int_X (\phi_t)^\sharp\omega\,.$$

Since X is compact, one can differentiate inside the integral and get :

$$\left.\frac{d}{dt}\right|_{t=0} T_{\phi_t(X)}(\omega) = \int_X \mathcal{L}_v\omega = \int_X \imath_v(d\omega) + \int_X d(\imath_v\omega)\,.$$

From Stokes' theorem, the second term in the sum equals :

$$\int_X d(\imath_v\omega) = \int_{\partial X} \imath_v\omega$$

which concludes the proof. □

Roughly speaking, this result expresses how the current associated to a shape X varies when X is deformed in the direction of vector field v. It can be thus interpreted as the gradient of terms $T_X(\omega)$ with respect to X. In particular, gradients of the previously defined RKHS metrics are also contained in theorem 1.3.2 ; if the kernel is such that $W \hookrightarrow \Omega^d_{0,1}(E)$, since for any current $T \in W'$, $\langle T_X, T\rangle_{W'} = T_X(\mathbf{K}_W T)$ (we remind that $\mathbf{K}_W$ is the natural isometry between W' and W), we have then $\langle T_X, T\rangle_{W'} = T_X(\omega)$ where $\omega = K_W T \in W$ is a C^1 d-differential form thanks to the embedding of W. Thus the previous theorem applies to express variations of norms and distances between currents.
The interest of such formulas is that they give qualitative properties about the gradients of the metrics we are using, which do not appear obvious at all when writing the discrete expressions of gradients used in the numerical part. In this case, we first see the presence of a boundary term signifying a particular behavior for points on the boundary whereas, in the inside of X, one can see that the first integral in the formula depends in fact only in the component of v normal to X at each point of X, showing incidentally that the gradient is orthogonal to the shape at all points of X, let alone on the boundary. We refer to the following chapters where these remarks will be treated with more detail.

In summary, we briefly went over the essential advantages of currents and RKHS norms on currents in building data attachment measures for datasets consisting of submanifolds. Except for the problem of non-invariance to orientation changes that shall be the main topic of chapter 3, the resulting distances comply to most the requirements we mentioned. Such attachments have been used in several registration algorithms like the previous one of geodesic shooting and has been already applied to datasets of curves [36, 31], fiber bundles [30] and surfaces [43, 35].

1.4 Problematics and contributions of this work

Having outlined the general framework of computational anatomy as well as some of the existing mathematical tools related to the thesis, we now introduce, in this section, the essential issues that are addressed by this work.

1.4.1 Extension to geometrico-functional data

The central question studied here is the extension of all the previous modelling and setting for shape analysis to the case of structures that mix both geometry and signal, which we have named functional shapes (or fshapes in short). As we will see in chapter 2, such type of objects are of growing frequency in computational anatomy and more particularly in medical imaging since the apparition of new data acquisition technologies that extract the shape of the anatomy while measuring additional functional signal data on these shapes, as for instance fMRI datasets for the brain or thickness maps for various organs. Contrarily to usual images, geometrico-functional datasets have varying geometrical support in addition to the variability of the signal itself, which is challenging at several levels for one to propose a coherent framework of analysis in the spirit of LDDMM.

Functional shapes and metamorphoses

In this introduction, we will focus on the simplest case of *functional shapes*, where the signals carried by shapes are real-valued functions. This notion shall be introduced in its full generality in chapter 2, notably to encompass situations where signals are vectors or tensors for instance. In this case, a functional shape can be thought naively as a real function on a shape X. If we adopt the definitions of the previous section, we could say more precisely that a d-dimensional functional shape is a couple (X, f) where X is a d-dimensional rectifiable subset of the vector space E and f is a real-valued function defined on X. We can assume also the function f to belong to the space $L^2(X)$ i.e that

$$\int_X f^2(x) d\mathcal{H}^d(x) < +\infty \,.$$

Note that the function f is not assumed to be defined on E but only on the geometrical support given by the shape X. This is contrasting with the case of images where the support is fixed once and for all. Here, we see that the spaces on which the functions live are not comparable from one functional shape to another (except if the supports are identical).
Diffeomorphisms of E can act on these functional shapes by simple transport of the geometrical support. If $\phi \in \mathrm{Diff}(E)$, $\phi.(X, f)$ is the functional shape of support $\phi(X)$ and signal $f \circ \phi^{-1}$, which we write :

$$\phi.(X, f) = (\phi(X), f \circ \phi^{-1})$$

This generalizes the transport of images described in example 1.2.2. In that situation, the Riemannian setting of LDDMM presented in section 1.2 applies almost straightforwardly. Nevertheless, this model is limited to purely geometrical changes of the functional shape and cannot account for some residual signal variations that are often necessary to consider in several applications. For classical images, such generalized transformations consisting of both a deformation and residual displacements

was introduced through the concept of *metamorphosis* in [78]. It can be generalized to functional shapes in a natural way. For a fshape (X, f), we consider deformations (ϕ, ζ) where $\phi \in \mathrm{Diff}(E)$ and $\zeta \in L^2(X)$ that acts on (X, f) by the formula :

$$(\phi, \zeta).(X, f) = (\phi(X), (f + \zeta) \circ \phi^{-1}) \,. \tag{1.22}$$

As in the case of images, this action is not transitive on the set of all functional shapes and one has to restrict to specific orbits. We introduce, as in 1.2.3, a group G_V of diffeomorphisms of E, V being a Hilbert space of vector fields continuously embedded into $C_0^1(E, E)$. For a fixed rectifiable subset X_0 of dimension d, we call $\mathcal{X}_0 \doteq G_V.X_0$ the orbit of all the $\phi(X_0)$ for $\phi \in G_V$ and define the following subset of functional shapes :

$$\mathcal{F} = \{(X, f) \mid X \in \mathcal{X}_0,\ f \in L^2(X)\} \,.$$

Then it's straightforward that for all (X, f) and (X', f') in $\mathcal{F}$, there exists $(\phi, h) \in G_V \times L^2(X)$ such that $(X', f') = (\phi, h).(X, f)$.
Now, as for metamorphoses, we can consider these deformations of functional shapes as resulting from integration of infinitesimal variations. Let (X, f) be a functional shape in $\mathcal{F}$. For the geometric part, diffeomorphisms ϕ are again generated as the flow ϕ_t^v of a time-varying vector field $v \in L_V^2$ while the residual signals can be constructed as $\zeta_t^h = \int_0^t h_s ds$, $h_s \in L^2(X)$ being the instantaneous speed at time s. We equip these deformations with the following energy measure :

$$E_X(v, h) \doteq \frac{\gamma_V}{2} \int_0^1 |v_t|_V^2 dt + \frac{\gamma_f}{2} \int_0^1 \int_{\phi_t^v(X)} |h_t \circ (\phi_t^v)^{-1}|^2(x) d\mathcal{H}^d(x) \tag{1.23}$$

where γ_V and γ_f are just ponderation parameters between the geometrical deformation part and the residual one. Note that the inside term in the second integral can be also rewritten, thanks to the area formula [37] corollary 3.2.20 (or simply a change of variables in the case where X is a submanifold) :

$$\int_{\phi_t^v(X)} |h_t \circ (\phi_t^v)^{-1}|^2(x) d\mathcal{H}^d(x) = \int_X |h_t|^2(x) |d_x \phi_t^v.\xi(x)| d\mathcal{H}^d(x) dt$$

where $\xi(x) = \xi_1(x) \wedge ... \wedge \xi_d(x)$ is the unit d-vector representing the tangent space to X at x and $|d_x \phi_t^v.\xi(x)|$ is the norm of the d-vector $d_x \phi_t^v(\xi_1) \wedge ... \wedge d_x \phi_t^v(\xi_d)$. Thus the energy equals also :

$$E_X(v, h) = \frac{\gamma_V}{2} \int_0^1 |v_t|_V^2 dt + \frac{\gamma_f}{2} \int_0^1 \int_X |h_t|^2(x) |d_x \phi_t^v.\xi(x)| d\mathcal{H}^d(x) dt \,. \tag{1.24}$$

It is also a classical control result that there exists an increasing function $C : \mathbb{R}_+ \to \mathbb{R}_+$ such that, for all $x \in E$, $t \in [0, 1]$ and $\xi = \xi_1 \wedge ... \wedge \xi_d \in \Lambda^d E$, we have

$$|d_x \phi_t^v(\xi_1) \wedge ... \wedge d_x \phi_t^v(\xi_d)| \leq C \left(\int_0^1 \|v_t\|_{1,\infty} dt \right) |\xi_1 \wedge ... \wedge \xi_d| \,.$$

We refer to [86] for instance or to 2.2.3 in chapter 2. Since $\|v\|_{1,\infty} \leq c_V \|v\|_V$, we eventually obtain the existence of an increasing function C such that for all (v, h), we have the following bound on the energy :

$$E_X(v, h) \leq \frac{\gamma_V}{2} \int_0^1 \|v_t\|_V^2 dt + \frac{\gamma_f}{2} . C \left(\int_0^1 \|v_t\|_V^2 dt \right) \int_0^1 \|h_t\|_{L^2(X)}^2 dt \,. \tag{1.25}$$

A Riemannian framework for fshapes metamorphoses

We can now extend the previous idea of geodesic distances to this situation of fshapes' metamorphoses. We restrict, as previously, to a set $\mathcal{F}$ for which $\mathcal{X}_0 = G.X_0$ and X_0 is also assumed to be compact. It implies that for all $X \in \mathcal{X}_0$, X is compact. We then define, for all (X, f) and (X', f') in $\mathcal{F}$:

$$d_{\mathcal{F}}((X,f),(X',f')) \doteq \inf_{(v,h)} \left\{ E_X(v,h)^{\frac{1}{2}} \mid (\phi_1^v, \zeta_1^h).(X,f) = (X',f') \right\}. \tag{1.26}$$

Theorem 1.4.1. *The function $d_{\mathcal{F}}$ given by eq.*(1.26) *defines a distance on $\mathcal{F}$ ie. is symmetric, satisfies the triangle inequality and $d_{\mathcal{F}}((X,f),(X',f')) = 0$ if and only if $X = X'$ and $f = f' \in L^2(X)$.*

Proof. A first point to be verified is that the distance $d_{\mathcal{F}}$ is finite between any pair of fshapes. Indeed, by definition there exists between any X and $X' \in \mathcal{X}_0$ a path $t \to \phi_t^v.X$ with $v \in L^2([0,1], V)$. Now, if $h_t = (f' \circ \phi_1^v - f)$, we get $(v,h) \in L^2([0,1], V \times L^2(X))$ and $f' = (f + \zeta_1^v) \circ (\phi_1^v)^{-1}$ so that

$$d_{\mathcal{F}}((X,f),(X',f')) \leq E_X(v,h)^{1/2} < \infty .$$

The symmetry comes from a usual process of *time reversal*.
Let $(X,f),(X',f') \in \mathcal{F}$ and $(v,h) \in L^2([0,1], V \times L^2(X))$ such that

$$X' = \phi_1^v.X, \;\; f' = (f + \zeta_1^h) \circ (\phi_1^v)^{-1} . \tag{1.27}$$

If we define

$$\tilde{v}_t \doteq -v_{1-t} \text{ and } \tilde{h}_t \doteq -h_{1-t} \circ (\phi_1^v)^{-1}$$

for any $t \in [0,1]$, then for $X_t \doteq \phi_t^v.X$ (so that $X_0 = X$ and $X_1 = X'$) we have

$$\phi_t^v.X_0 = X_t = \phi_{1-t}^{\tilde{v}}.X_1$$

and $(\tilde{v}, \tilde{h}) \in L^2_V \times L^2(X')$ with

$$E_X(v,h) = E_{X'}(\tilde{v}, \tilde{h}) . \tag{1.28}$$

Since one easily checks that $\zeta_1^{\tilde{h}} = -\zeta_1^h \circ (\phi_1^v)^{-1}$ so that if $f' = (f + \zeta_1^h) \circ (\phi_1^v)^{-1}$ we have $f = (f' + \zeta_1^{\tilde{h}}) \circ (\phi_1^{\tilde{v}})^{-1}$ with $X = \phi_1^{\tilde{v}}.X'$ This gives immediately the symmetry.
As for the triangular inequality, it follows from a usual process of *path concatenation*.
Let (X,f), (X,f') and (X'',f'') be three fshapes in $\mathcal{F}$ such that $d_{\mathcal{F}}((X,f),(X',f')) > 0$ and $d_{\mathcal{F}}((X',f'),(X'',f'')) > 0$. One easily checks that for $(v,h) \in L^2([0,1], V \times L^2(X))$ and $(v',h') \in L^2([0,1], V \times L^2(X'))$ with

$$(X',f') = (\phi_1^v.X, (f + \zeta_1^h) \circ (\phi_1^v)^{-1}) \text{ and } (X'',f'') = (\phi_1^{v'}.X', (f' + \zeta_1^{h'}) \circ (\phi_1^{v'})^{-1})$$

then denoting for $\alpha, \beta > 1$ such that $1/\alpha + 1/\beta = 1$ and $s \in [0,1]$

$$\mathrm{Cat}_\alpha((v',h'),(v,h))_s \doteq \beta(v'_{\beta(s-1/\alpha)}, h'_{\beta(s-1/\alpha)} \circ \phi)\mathbf{1}_{s \geq 1/\alpha} + \alpha(v_{\alpha s}, h_{\alpha s})\mathbf{1}_{0 \leq s < 1/\alpha} \tag{1.29}$$

with $\phi = \phi_1^v$, we have for $(\tilde{v}, \tilde{h}) \doteq \mathrm{Cat}_\alpha((v',h'),(v,h))$ that $(\tilde{v}, \tilde{h}) \in L^2([0,1], V \times L^2(X))$ and

$$(X'',f'') = (\phi_1^{\tilde{v}}.X, (f + \zeta_1^{\tilde{h}}) \circ (\phi_1^{\tilde{v}})^{-1}) \tag{1.30}$$

so that

$$d_{\mathcal{F}}((X,f),(X'',f'')) \leq E_X(\tilde{v},\tilde{h})^{1/2}\,. \tag{1.31}$$

However, for $\alpha_* = (E_X(v,h)^{1/2} + E_{X'}(v',h')^{1/2})/E_X(v,h)^{1/2}$, we check easily that

$$E_X(\tilde{v},\tilde{h})^{1/2} = E_X(v,h)^{1/2} + E(v',h')^{1/2} \tag{1.32}$$

and the triangle inequality follows immediately.
Now, if $d_{\mathcal{F}}((X,f),(X',f')) = 0$ then there exists a sequence ϕ_n such that $\phi_n.X = \phi_n(X) = X'$ and $\phi_n \to \mathrm{Id}$ on X. In particular, X' is dense in X. As X' is compact, $X \subset X'$. By symmetry we get $X = X'$. Moreover, there exists a sequence $\zeta_n \in L^2(X)$ such that $\zeta \to 0$ in $L^2(X)$ and $f' = (f+\zeta_n)\circ\phi_n^{-1}$ with $\phi_n(X) = X$. We get

$$\int_X (f'-f)^2 \leq 2\int_X (f\circ\phi_n^{-1} - f)^2 + 2\int_X (\zeta_n\circ\phi_n^{-1})^2 \to 0$$

and the result is proved. □

The distance $d_{\mathcal{F}}$ extends to fshapes the metamorphosis metrics on images examined in [78]. A fundamental question is then the existence of minimizing geodesics between two functional shapes in the orbit $\mathcal{F}$. This is answered by the result below :

Theorem 1.4.2. *For any (X,f), and (X',f') in $\mathcal{F}$, there exists $(v,h) \in L^2([0,1],V\times L^2(X))$ such that $E_X(v,h)^{1/2} = d_{\mathcal{F}}((X,f),(X',f'))$. In particular, if*

$$(X_t,f_t) \doteq (\phi_t^v.X, (f+\int_0^t h_s ds)\circ(\phi_t^v)^{-1}) \tag{1.33}$$

the path $t\to(X_t,f_t)$ can be considered as a minimizing geodesic between (X,f) and (X',f').

Proof. It is sufficient to show that $(v,h)\to E_X(v,h)$ is lower semi-continuous for the weak convergence of the space $L^2([0,1],V\times L^2(X))$.
Indeed, if this is the case, then from any minimizing sequence (v_n,h_n) such that $E_X(v_n,h_n) \to d_{\mathcal{F}}((X,f),(X',f'))^2$ we deduce that v_n is bounded on $L^2([0,1],V)$ and :

$$\int_0^1\int_X |h_{n,t}(x)|^2 d\mathcal{H}^d(x)dt \leq \int_0^1 (\sup_{x\in X}|d_x\phi_t^{v_n}(\xi(x))|^{-1})\int_X (|h_{n,t}(x)|^2)|d_x\phi_t^{v_n}.\xi(x)|d\mathcal{H}^d(x)$$

Now, $\sup_{x\in X}|d_x\phi_t^{v_n}.\xi(x)|^{-1} = 1/\inf_{x\in X}|d_x\phi_t^{v_n}.\xi(x)|$ and, in a similar way as mentioned earlier, one can show that there exists a decreasing function $\tilde{C}:\ \mathbb{R}_+\to\mathbb{R}_+^*$ independent of v such that for all $x\in E$,

$$|d_x\phi_t^v.\xi(x)| \geq \tilde{C}\left(\int_0^1 \|v_t\|_V^2 dt\right).$$

It results that there is an increasing function C such that :

$$\begin{aligned}\int_0^1\int_X |h_{n,t}(x)|^2 d\mathcal{H}^d(x)dt &\leq C\left(\int_0^1 \|v_{n,t}\|_V^2 dt\right)\int_X (|h_{n,t}(x)|^2)|d_x\phi_t^{v_n}.\xi(x)|d\mathcal{H}^d(x)\\ &\leq C\left(\int_0^1 \|v_{n,t}\|_V^2 dt\right)E_X(v_n,h_n)\end{aligned}$$

and we get that the sequence (h_n) is bounded in $L^2([0,1], L^2(X))$. Hence, by weak compactness of strong balls in $L^2([0,1], V \times L^2(X))$, we can assume that, up to the extraction of a sub-sequence, (v_n, h_n) weakly converges towards $(v_\infty, h_\infty) \in L^2([0,1], V \times L^2(X))$ and by lower semi-continuity of E_X, we deduce that

$$E_X(v_\infty, h_\infty) \leq d_{\mathcal{F}}((X,f),(X',f'))^2 \,.$$

We only need to check that $(X', f') = (\phi_1^{v_\infty}(X), (f + \zeta_1^{h_\infty}) \circ (\phi_1^{v_\infty})^{-1})$. This last results follow from the fact if (v_n, h_n) weakly converges to (v_∞, h_∞) then $\phi_1^{v_n} \to \phi_1^{v_\infty}$ uniformly on any compact sets (which is a well-known result proved in [86] theorem 12.12 for instance) and $(f + \zeta_1^{h_n}) \circ (\phi_1^{v_n})^{-1}$ weakly converges to $(f + \zeta_1^{h_\infty}) \circ (\phi_1^{v_\infty})^{-1}$. We now come to the proof of the lower semi-continuity of E_X. Considering again a sequence (v_n, h_n) that weakly converges toward (v_∞, h_∞) in $L^2([0,1], V \times L^2(X))$, we first know from a classical result on weak convergence that $\int_0^1 \|v_{\infty,t}\|_V^2 dt \leq \liminf_{n\to\infty} \int_0^1 |v_{n,t}|_V^2 dt$. Besides, we have

$$\int_0^1 \int_X |h_{\infty,t}(x)|^2 |d_x \phi_t^{v_\infty}.\xi(x)| d\mathcal{H}^d(x) = \lim_{n\to\infty} \int_0^1 h_{n,t}(x) h_{\infty,t}(x) |d_x \phi_t^{v_\infty}.\xi(x)| d\mathcal{H}^d(x)$$

by definition of the weak convergence of (h_n). Now, since $d_x \phi_t^{v_n} \to d_x \phi_t^{v_\infty}$ uniformly on $t \in [0,1]$ and $x \in X$, we get

$$\begin{aligned}
&\int_0^1 \int_X |h_{\infty,t}(x)|^2 |d_x \phi_t^{v_\infty}.\xi(x)| d\mathcal{H}^d(x) \\
&= \lim_{n\to\infty} \int_0^1 \int_X h_{n,t}(x) h_{\infty,t}(x) (|d_x \phi_t^{v_n}.\xi(x)|)^{1/2} |d_x \phi_t^{v_\infty}.\xi(x)|^{1/2} d\mathcal{H}^d(x) \\
&\leq \liminf_{n\to\infty} \left(\int_0^1 \int_X |h_{n,t}(x)|^2 |d_x \phi_t^{v_n}.\xi(x)| d\mathcal{H}^d(x) \right)^{1/2} . \\
&\qquad \left(\int_0^1 \int_X |h_{\infty,t}(x)|^2 |d_x \phi_t^{v_\infty}.\xi(x)| d\mathcal{H}^d(x) \right)^{1/2}
\end{aligned}$$

so that

$$\int_0^1 \int_X |h_{\infty,t}(x)|^2 |d_x \phi_t^{v_\infty}.\xi(x)| d\mathcal{H}^d(x) \leq \liminf_{n\to\infty} \int_0^1 \int_X |h_{n,t}(x)|^2 |d_x \phi_t^{v_n}.\xi(x)| d\mathcal{H}^d(x) \,. \tag{1.34}$$

□

The results of the two last theorems show that comparison of functional shapes through metamorphoses can be done within a similar framework as the previous one for usual shapes : it involves geodesic distances on the space of geometrico-functional deformations. One could in addition, following the same path as in section 1.2.5, write the Hamiltonian equations related to the geodesic shooting in this setting. This framework generalizes, on the one hand, the situation where shapes are only deformed geometrically, corresponding to the case of an infinite penalty $\gamma_f \to \infty$ on the residual part of the energy in eq.(1.24). On the other hand, it's also a generalization of the metamorphosis idea for images, for which basically all shape supports equal a fixed domain of the space.

A tangential model

In the practical cases considered in this work, we shall be dealing with an earlier model that simplifies a little bit the previous metamorphosis framework on functional shapes. Instead of considering residual signals on X obtained by integrating velocities $\dot{\zeta}_t = h_t$ and compute an energy along time-dependent spaces $L^2(X_t)$, we can freeze the metric to its initial support X, namely replace the residual part of the energy by :

$$\int_0^1 \int_X |\dot{\zeta}_t(x)|^2 d\mathcal{H}^d(x) dt \,.$$

Contrarily to eq.(1.24), this energy does not take into account the variation of support of the signal ζ_t through the local change of volume $|d_x\phi_t^v.\xi(x)|$. The advantage is that for a fixed end-point condition $\zeta = \zeta_1$, the optimal path ζ_t for this cost is always $\zeta_t = t\zeta$ and it results that the energy can be simply written in that case :

$$\tilde{E}_X(v,\zeta) = \frac{\gamma_V}{2} \int_0^1 \|v_t\|_V^2 dt + \frac{\gamma_f}{2} \|\zeta\|_{L^2(X)}^2 \,. \tag{1.35}$$

Then, one can consider :

$$\tilde{d}_{\mathcal{F}}((X,f),(X',f')) \doteq \inf_{(v,\zeta)\in L_V^2 \times L^2(X)} \left\{ \tilde{E}_X(v,\zeta)^{\frac{1}{2}} \mid (\phi_1^v,\zeta).(X,f) = (X',f') \right\} . \tag{1.36}$$

Unlike the previous $d_{\mathcal{F}}$ function, $\tilde{d}_{\mathcal{F}}$ is no more a distance on $\mathcal{F}$ (it doesn't satisfy symmetry and triangular inequality). Yet it still gives a formulation for the exact matching problem between two fshapes (X,f) and (X',f') belonging to a common orbit $\mathcal{F}$, which amounts in the following constrained optimization :

$$\inf_{(v,\zeta)\in L_V^2 \times L^2(X)} \{ \tilde{E}_X(v,\zeta) \mid (X',f') = (\phi_1^v(X), (f+\zeta)\circ(\phi_1^v)^{-1}) \} \,. \tag{1.37}$$

The good property is that we still get the existence of a minimizer :

Theorem 1.4.3. *For any (X,f) and (X',f') in $\mathcal{F}$, there exists (v^*,ζ^*) in $L_V^2 \times L^2(X)$ that minimizes eq.(1.37).*

The proof can be done following the same arguments as in theorem 1.4.2 and is even simplified due to the fact that the function ζ is no more varying in time. As we see, the residual signal function ζ can be now searched directly as a stationary L^2 function on X.

Data attachment functionals

Until now, we have been considering the exact matching situation for which functional shapes are assumed to belong to the same orbit $\mathcal{F}$. In realistic cases, as already noticed in section 1.3, one has to introduce inexact registration to be able to treat shapes belonging to different orbits. In the context of the previous simplified tangential setting, inexact registration could be formulated as the optimization problem :

$$\inf \left\{ \int_0^1 \|v_t\|_V^2 + \gamma_f \|\zeta\|_{L^2(X)}^2 + A\left((\phi_1^v,\zeta).(X,f),(X',f')\right) \mid (v,\zeta) \in L_V^2 \times L^2(X) \right\} \tag{1.38}$$

where A is here a data attachment function on the set of functional shapes. This is the general registration setting that shall be implemented in chapter 2 (section 2.3.2). Yet, the remaining issue is now to define proper dissimilarity measures in the case of functional shapes. At first glance, this is clearly not an easy problem, in great part because the registration formulation that we propose here requires dissimilarity measures that must take into account geometry **and** signal simultaneously. In section 1.3, we addressed the question of data attachment functions in the case of geometrical objects like submanifolds on the one hand or in the case of images on the other, but the situation of signals that are supported by deformable geometrical support do involve specific challenges that, to our knowledge, have been rarely treated in that field. This is one of the subject that we propose to focus on in the present work and chapter 2 is dedicated to the question of representations and metrics on those functional shapes, enabling the extension of LDDMM registration algorithms to this new class of data. The framework that we develop, named **functional currents**, follows the ideas of data attachment resulting from currents' norms and defines its right extension to embed fshapes.

1.4.2 A non-oriented setting for shape dissimilarity

Still about data attachment metrics, we also address a second issue that deals with the orientation of shapes when compared through the RKHS norms on currents presented in 1.3.4. In the first place, this can be thought as a problem independent to the previous one of geometrico-functional datasets. Focusing on purely geometrical shapes, it has been indeed a consistent drawback, in some situations, that currents basically represent oriented rectifiable subset. The orientation is fundamental in that regard since, if X is for instance a connected submanifold with one orientation and $\check{X}$ is the same submanifold equipped with the opposite orientation, then in the space of currents, we have $T_{\check{X}} = -T_X$. In chapter 3, we shall detail the problems that this poses with respect to the representation of certain types of data. An important one is that it imposes some kind of *consistent orientation* of shapes between their different parts, to avoid non-canonical cancellation effects due to misorientation. Yet, in the general case of rectifiable subsets, it's quite obvious that the very notion of consistent orientation is not relevant at all, particularly in situations where shapes are made of many disconnected pieces that can technically cross with each other at many locations.

For these reasons, a non-oriented setting for building dissimilarity measures between shapes, that would preserve at the same time the essential interests of the currents' modelling of section 1.3.4, is undoubtedly a valuable tool to add in the arsenal of LDDMM methods. In the chapter 3 of this thesis, we examine several directions that were undertaken in the theoretical field of geometric measure theory to address this issue of orientation. We then focus on the particularly elegant concept of **varifolds**, introduced after F. Almgren in his pioneer work of [8]. Their definition generalizes the idea of measures but in a different way than currents : d-dimensional varifolds are still usual Borel finite measures but on the augmented space $E \times G_d(E)$, where $G_d(E)$ is the Grassmann manifold of d-dimensional subspaces of E. This Grassmann manifold is thus the set of all possible non-oriented tangent spaces of dimension d in E. Unlike currents, any (non-oriented) d-dimensional rectifiable subset of E can be represented in the space of varifolds, roughly speaking as a spatial distribution in E of elements in $G_d(E)$ representing the tangent space at each point. The essential contribution of chapter 3 is to define metrics in the space of varifolds that provides

data attachment terms between non-oriented shapes and study registration algorithm based on this new representation. As a dual of a space of functions, Hilbert metrics based on kernels can be indeed adapted for varifolds and their properties shall be thoroughly examined as well.

1.4.3 Organization of the chapters

This thesis is divided into four chapters. Chapters 2 and 3 are addressing the problems of dissimilarity measures and registration respectively for functional shapes and for non-oriented shapes. These are widely independent and can be read separately for the most part. Based upon the setting of functional currents and varifolds, the last chapter broadens the approach to group-study by introducing a framework for statistical atlas estimation and analysis of populations of subjects. The contents of each chapter can be summarized as follows :

• **Chapter 2** : this chapter introduces the concept of functional shapes within computational anatomy and gives a wider definition than the one considered in this introduction. In direct line of the fshape metamorphosis setting presented in 1.4.1, the main issue that remains for inexact registration is the construction of data attachment terms. The framework we propose in this chapter is an extension of currents, that we have designated by functional currents, which augments the space of usual currents with a component in the signal space. These spaces embed the set of oriented functional shapes. We then explain how, similarly to currents, one can build Hilbert structures on functional currents, given by reproducing kernels. The distances that are induced on fshapes allow simultaneous comparison of shape and signal without ever requiring point-to-point correspondences between the two objects. We show in addition a few properties of control for these metrics and generalize Cartan's variation formula to fcurrents (theorem 2.2.4). The last sections of chapter 2 focus on the discrete case and detail inexact registration algorithms on fshapes with a few results. The numerical issues related to kernel computations for diffeomorphisms and fcurrents' distances are also discussed.

• **Chapter 3** : after exposing the principal orientation issues coming along with currents, we turn to the alternative concept of varifolds introduced originally in the context of geometric measure theory. We explain how the space of varifolds can embed rectifiable subsets of E without the need of defining an orientation. The rest of the chapter is devoted to the construction of kernel norms on varifolds and studies the main properties of these compared to currents. In particular, we show that, under simple assumptions, norms of rectifiable subsets are bounded below by the actual volume of the set (theorem 3.4.1 and corollary 3.4.1), which prevents shape cancellation phenomena occurring with currents. We also derive a variation formula for varifold norms (theorem 3.4.2). The end of chapter 3 adapts the large deformation registration algorithm to the case of non-oriented shapes represented as varifolds and show, on a few simple examples, the potential interest of this framework compared to the previous one of currents. It concludes by merging this approach with the one of chapter 2 resulting in the definition of a functional varifold setting to treat non-oriented functional shapes.

• **Chapter 4** : in this chapter, we eventually come to the real heart of the problematic of statistical analysis on populations. The objective is to propose a coherent and robust framework to model statistically the variability among a group of functional

shapes both in their geometry and signal. For this purpose, we rely on the tools developed in the previous chapters and start by extending the simpler situation of registration of only two subjects to group-analysis. The framework that is developed estimates in a simultaneous minimization a template functional shape representing some kind of prototype fshape for the group, together with deformations and residual signals for each subject. In a second step, we address the problem of learning principal modes of variations for these sets of deformations and residuals and performing classification for multi-class datasets based on these. Thereby, we adapt some statistical learning methods such as PCA and LDA to this situation. Some results of atlas estimation on both geometric and geometrico-functional datasets from biology, as well as a few classification results, are also presented in this chapter.

Chapter 2

An extended framework for the analysis of geometrico-functional data

Contents

2.1 Functional shapes 44
2.1.1 State of the art 44
2.1.2 Functional shapes 46
2.1.3 Representation by currents 48
2.2 Functional currents : mathematical considerations . . . 51
2.2.1 Definition and basic properties of functional currents 51
2.2.2 Reproducing Kernel Hilbert Space structures on functional currents 55
2.2.3 Properties of kernel norms on fcurrents 60
2.3 Algorithms based on functional currents 71
2.3.1 Compression with Matching Pursuit on fcurrents 72
2.3.2 Registration of functional shapes 75
2.3.3 Numerical considerations 82
2.4 Conclusion 87

Results of this chapter were partially published in [21].

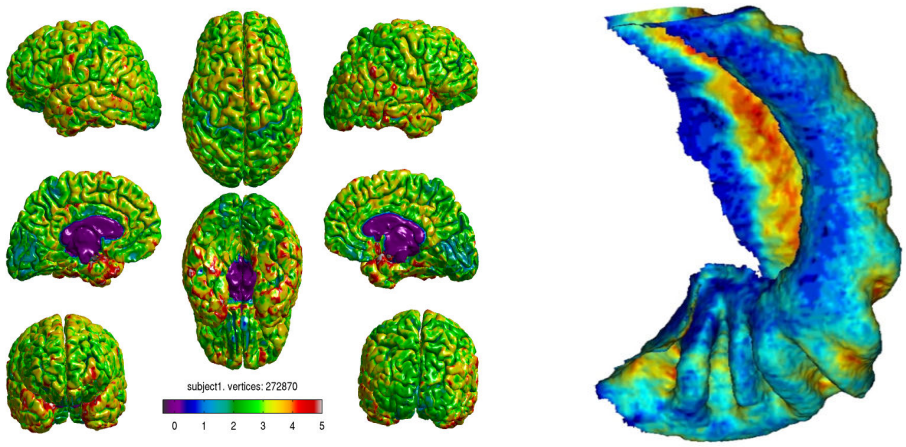


Figure 2.1: Cortical thickness estimation on several full brains (left figure) and more specifically on one hippocampus (right).

2.1 Functional shapes in computational anatomy : some new challenges

2.1.1 State of the art

In the introduction, the focus was made on the analysis of shape, which has caught the major part of the efforts in computational anatomy since its early stage. Meanwhile, the upcoming of new acquisition processes, in medical imaging amongst other, has generated new data structures that cannot be described as purely geometrical objects but instead as shapes that carry in addition some functional information. Such functions or signals can be for instance activation maps on the cortex in fMRI as a response to specific stimuli. It can represent thickness estimations on tissues, as cortical thickness in the study of Alzheimer's disease (figure 2.1) or thicknesses of retina layers studied for the evolution of glaucoma (cf chapter 4). Signals can be also directions in the space as for instance with cells' direction of division in particular membranes during growth (figure 2.2), or even diffusion tensors in brains and hearts as provided by the recent advances in DTI imaging [13, 54, 85].
The growing occurrence of such type of datasets has naturally attracted attention among researchers both from the signal processing community and computational anatomy, leading to the emergence of Computational Functional Anatomy (CFA) as first formalized in [64]. In this article, the authors mention the most important challenges that remain to be addressed in CFA, which are essentially the generalization of usual signal processing schemes (as smoothing, interpolation...) to signals that are no more supported by flat objects as with usual images, and, as a second issue, the comparison and statistical analysis of functional responses across individuals in a group. The fact that different subjects do not have the same geometrical support is a clear obstacle and makes this last problem particularly challenging since we do not have a point to point correspondence between the subjects to compare their signals. For this reason, a vast majority of approaches have been trying to build such correspondence maps between anatomical coordinates and transfer functional information using different transport actions according to the nature of signals, as explained in [64]. Other methodologies have been normalizing geometries of all subjects by mapping on a common coordinate system (e.g a sphere for cortex surfaces, [40, 72]) and

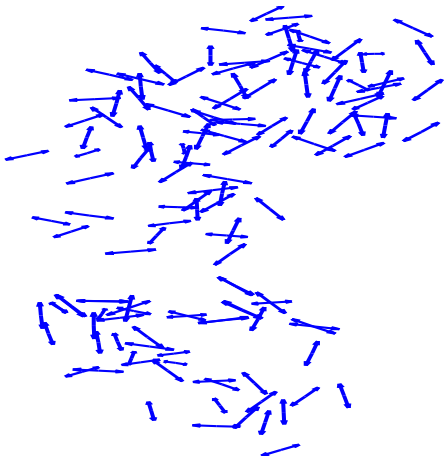


Figure 2.2: A set of cells' division directions inside a mouse embryo's heart (data courtesy of S. Meilhac' team in Institut Pasteur).

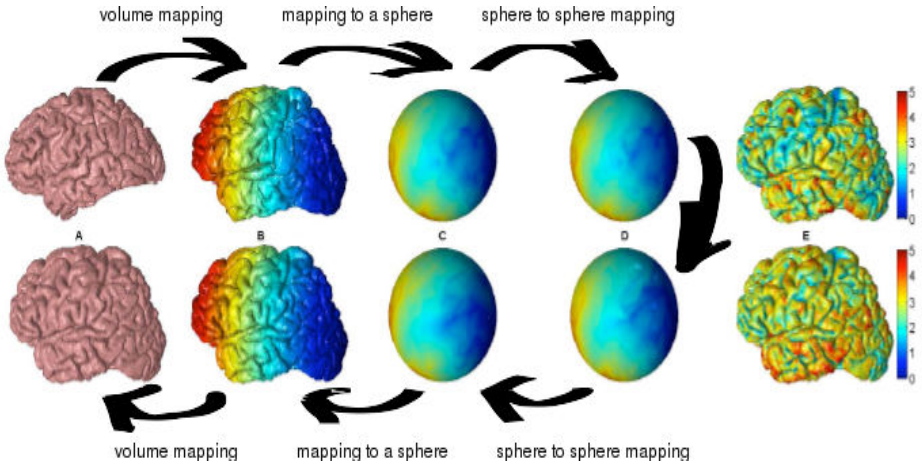


Figure 2.3: The different steps of SAVOR algorithm.

then finding correspondences from sphere to sphere based on functional informations and original curvature of shapes. This is basically the principle of SAVOR algorithm summed up in figure 2.3.

If important applications have been made thanks to the aforementioned works, we argue that enforcing common coordinate systems for the whole database makes them specific to a particular type of topology and thus the choice of this normalized coordinate system should be tailored to each dataset. It becomes consequently a delicate and not canonical step in the approach. Moreover, treating surfaces with small holes or fiber bundle datasets (as studied in DTI imaging) would become a dramatically hard problem due to the non-equivalent topologies across subjects. In addition, the mappings from subject to subject obtained on the normalized sphere model may have no relevant meaning in the original embedding space of shapes, which makes transformations themselves difficult to use in any statistical or classification purpose. The last thing that we can point out is that geometry and function are nearly entirely decorrelated during such processes. In fact, the question of providing a solid way of modelling geometry and signal together, to our knowledge, has not been really addressed yet.

The objective of this chapter is precisely to provide robust mathematical and numerical frameworks for representing and comparing 'functional shapes', the core idea being to avoid any precomputation of point to point correspondences between supports. As explained in the introduction chapter, currents' kernel metrics give distances between purely geometrical shapes that do not require such correspondences and they thus constitute interesting candidates. In the following subsections, we give a precise sense to the idea of 'functional shapes'. We will then examine two straightforward possible generalizations of currents to incorporate functional shapes and show why they are not totally satisfying in our context.

2.1.2 Functional shapes

We start by setting up mathematically the concept of functional shapes in its higher generality. Those can be thought intuitively either as shapes carrying signals or images with deformable geometrical support. Although most of the numerical applications will be dealing with real-valued signals, the theory that we intend to develop is expected to be general enough to model practically any sort of signal, from real-valued images to directions and tensors as mentioned earlier. Let E be the n-dimensional embedding vector space of shapes and M a Riemannian manifold that we shall call the *signal manifold*. For images, $M = \mathbb{R}$. In the case where signals are oriented directions, we would have $M = \mathbb{S}^{n-1}$ the unit sphere of E or $M = P(E)$ the projective space for unoriented directions. Signals could be also symmetric tensors in which case $M = S_n(\mathbb{R})$. We define functional shapes very naturally as :

Definition 2.1.1. *A functional shape of dimension d (or fshape in short) is a couple (X, f) where X is a d-dimensional rectifiable subset of E and $f : X \to M$ a measurable function on X taking values in the signal manifold M.*

Note that we do not assume a priori any particular regularity on the functions f. As for the geometrical support X, we only assume rectifiability so that it includes submanifolds of E as well as piecewise submanifolds and polygons and thus encompasses shapes both in discrete and continuous representations. Classical images are also special cases of functional shapes, for which X is a fixed flat domain of $\mathbb{R}^2$. For every functional shape (X, f), by analogy with usual functions, we can consider the **graph** $G_{(X,f)} = \{(x, f(x)),\ x \in X\}$ as a subset of $E \times M$. With no additional assumptions, this set has no particular structure. However, if we assume f to be locally Lipschitz continuous then it's easy to see from definition 1.3.1 that $G_{(X,f)}$ is a $\mathcal{H}^d$-rectifiable subset of $E \times M$. With stronger assumptions, namely if X is a submanifold and f is C^1, then $G_{(X,f)}$ is even a d-dimensional submanifold of $E \times M$. In addition to that, a fundamental question is to define a proper deformation model on functional shapes, i.e express the group action of deformations on a given class of fshapes. With respect to pure geometry, this does not pose any particular issue since a diffeomorphism $\phi \in \mathrm{Diff}(E)$ acts on a shape X by simply transporting X by ϕ, i.e $\phi.X = \phi(X)$. However, the action on functions offers far more possibilities depending on the nature of the signals. Such issues of functional transfer are discussed for instance in [64]. We review a few different situations. The most basic one is the case of real-valued signals on shapes that are simply transported with their support by the deformations, namely the deformed functional shape $\phi.(X, f)$ is given by :

$$\phi.(X, f) = (\phi(X), f \circ \phi^{-1}) \tag{2.1}$$

This corresponds to an action that deforms the geometric support by keeping original functional values. In the particular case where X is a fixed domain in $\mathbb{R}^2$ or $\mathbb{R}^3$ and we impose that ϕ preserves the domain ($\phi(X) = X$) then we can think of f as a usual image and eq.(2.1) is exactly the action of deformations on images (i.e $\phi.I = I \circ \phi^{-1}$), as introduced in original LDDMM algorithms [14]. One could also enrich the deformations' group by adding a residual function $h : E \to \mathbb{R}$ on X. In that case, the group of transformations would be $\mathcal{M}(E, \mathbb{R}) \rtimes \text{Diff}(E)$, with $\mathcal{M}(E, \mathbb{R})$ the set of real-valued Borel measurable functions, the product being given by $(h_2, \phi_2) \times (h_1, \phi_1) = (h_1 + h_2 \circ \phi_1, \phi_2 \circ \phi_1)$ and the action :

$$(\phi, h).(X, f) = (\phi(X), (f + h) \circ \phi^{-1}) \tag{2.2}$$

Such models are well-adapted to the real-valued cases in which there is not a direct prior relationship between geometry and signal but we see that they become irrelevant if signals are, for instance, tangent vectors to X at each point of X, because the transported signals would not be anymore tangent to the image shape $\phi(X)$. In that case, we have for all $x \in X$, $f(x) \in T_x X \subset E$ and we see that signals should be also transformed by the differential of ϕ :

$$\phi.(X, f) = (\phi(X), (d\phi * f) \circ \phi^{-1}) \tag{2.3}$$

where $d\phi * f(x) = d_x\phi(f(x))$. The case of tensor-valued functions, as provided by DTI, is more involved because it must account for the orientation transformation on the diffusion matrix that results from the space deformation. A model has been established for instance in [2, 3]. If $f(x) = D_x$ is a symmetric 3×3 matrix, then the action of a diffeomorphism ϕ is given by :

$$\phi.(X, f) = (\phi(X), (d\phi \star f) \circ \phi^{-1}) \tag{2.4}$$

where $d\phi \star D_x = \lambda_{1,x}\tilde{e}_{x,1}(\tilde{e}_{x,1})^T + \lambda_{2,x}\tilde{e}_{x,2}(\tilde{e}_{x,2})^T + \lambda_{3,x}\tilde{e}_{x,3}(\tilde{e}_{x,3})^T$ in which $\lambda_{i,x}$ are the eigenvalues of D_x, $e_{i,x}$ the corresponding eigenvectors and $\tilde{e}_{x,i}$ the Gram-Schmidt orthonormalization of the basis $(d_x\phi(e_{x,i}))$.
The previous list is of course not exhaustive but was meant to show that, unlike classical shapes, there seems to be numerous possibilities of deformation models on functional shapes, depending on each data type. To have a more unified vision of all the cases together, one could think of deformations of functional shapes as deformations of their *graph* in the space $E \times M$, considering for instance diffeomorphisms of $\text{Diff}(E \times M)$. However, we see that this group is too wide because the deformation of the graph of (X, f) is not necessarily a graph of another functional shape (for some deformations, one could create points with multiple functional images). To preserve the underlying graph structure, a sufficient condition is to restrict to a subgroup H of transformations in $E \times M$ that can be written

$$\psi(x, m) = (\psi^g(x), \psi^f(x, m))$$

where $\psi^g \in \text{Diff}(E)$ and ψ^f is a measurable application from $E \times M$ to M. In that case, the image of $G_{(X,f)}$ by ψ is the graph of a functional shape and the group action is :

$$\psi.(X, f) = (\psi^g(X), \psi^f((\psi^g)^{-1}, f \circ (\psi^g)^{-1})) \tag{2.5}$$

Considering different particular groups H with the action of eq.(2.5) enables to recover all the previous cases, as one can check easily.

2.1.3 Representation by currents

Starting from the idea of currents presented in the introduction, we evaluate two different ways of embedding functional shapes (at least in some special cases) in spaces of currents.

Colored currents

First attempts to include signals supported geometrically through the current representation were investigated in [28] with the idea of *colored currents*. This relies basically on the fact already mentioned that the set of d-currents is already a big space that contains a wider variety of objects than d-dimensional submanifolds. In particular, weighted submanifolds can be considered as currents : if f is a real-valued function on X then we can associate to (X, f) a d-current in E by considering the signal as a weight on X :

$$T_{(X,f)}(\omega) = \int_X f\omega$$

By the previous setting, we embed functional shapes in the space of currents on E itself. Although this seems to be the most straightforward way to apply currents to functional shapes, it quickly appears that such a representation suffers from several important drawbacks. First is the impossibility to generalize colored currents for signals that are not simply real-valued, particularly if the signal space is not a vector space. The second point arises when the previous equation is discretized into Dirac currents, which leads to an expression of the form $\sum_{k=1..N} f(x_k)\delta_{x_k}^{\xi_k}$. We notice an ambiguity appearing between the signal and the volume element ξ since for any $r \neq 0$,

$$f(x_k)\delta_{x_k}^{\xi_k} = rf(x_k)\delta_{x_k}^{\xi_k/r}$$

Separating geometry from signal in the discretized version appears as a fundamental difficulty. In addition, the energy of Dirac terms are proportional to the value of the signal at the corresponding point which induces an asymmetry between low and high-valued signals. In this setting, areas having very small signals become negligible in terms of current, which not justified in general and can drastically affect the matching of colored currents. We show a simple illustration of this issue when matching two colored ellipsoids with this approach in figure 2.4. Finally an additional limitation in using colored currents is the fact that there is no flexibility to treat the signals at different scale levels than geometry, making this approach highly sensitive to noise.

Currents in product space

Another possible and interesting way to represent a functional shape by a current is to consider the graph $G_{(X,f)}$ as a shape in the product space $E \times M$, extending the idea of seeing a 2D image as a 3D surface. Now, one can consider the graph as a d-current in the product space $E \times M$. This is not straightforward though. The case of a general signal manifold M is for instance more difficult to handle in this setting since it requires to consider differential forms defined on M : for simplicity, we will assume for the rest of this section that M is a vector space. A second point mentioned earlier is that we need to impose more regularity on the signal f, for instance piecewise locally Lipschitz continuous so that $G_{(X,f)}$ is rectifiable and can be represented as a current in $\Omega_0^d(E \times M)$. These difficulties apart, there still are some important elements to point out. The first one is the increase of dimensionality

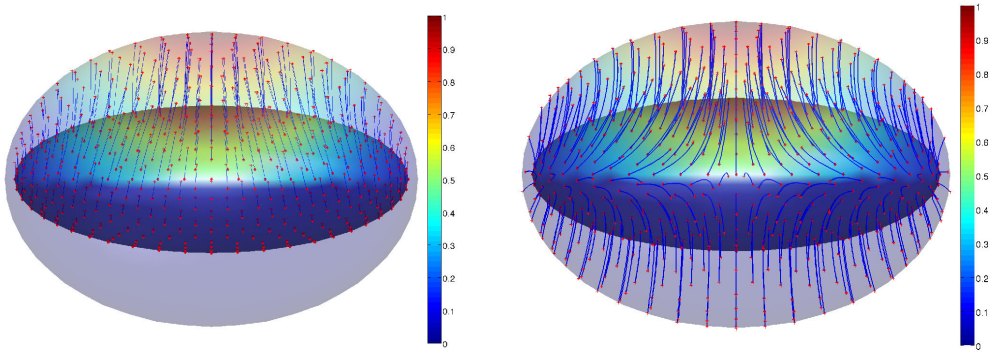


Figure 2.4: An example of registration between two ellipsoids. On the left, the matching is done by comparing functional shapes with the colored currents' representation. Values of the signals are two diffused stains both on the source ellipsoid (inside surface) and the target one (exterior shaded surface). We display in blue trajectories of the points. The points compounding to zero-valued area of the signal in the source shape are not matched to the corresponding points in the target surface. On the right, we show what should be the expected result. It is obtained through the approach of functional currents presented in 2.3.2.

of the approach, because, while we are still considering shapes of dimension d, the codimension is higher : the space of d-vectors characterizing local geometry $\Lambda^d(E \times M)$ is now of dimension $\binom{n+dim(M)}{d}$, with significant consequences from a computational point of view. From a more theoretical angle, we see that, in such an approach, geometrical support and signal play a symmetric role. In this representation, the modeled topology is no more the one of the original shape because we also take into account variations within the signal space. Whether this is a strength or a weakness is not obvious a priori and would depend on the kind of applications. What we can state is that this representation is not robust to topological changes of the shape : in practice, the connectivity between all points becomes crucial, which we illustrate in figure 2.5. Figure 2.6 shows a consequence of that if two curves carrying signals are matched under this model. It shows that it can actually lead to very unnatural registration in cases where some disconnections and discontinuities of signals appear. In the field of computational anatomy, when processing data such as fiber bundles, where connections between points of the fibers are not always reliable, we argue that this would be a clear drawback.

To summarize this section, we have investigated two direct ways to see a functional shape as a current. The colored current setting, although being very close to the modelling of purely geometrical shapes, is not acceptable mainly because it mixes geometry and signal in an inconsistent way. As for the second idea of immersing the functional shape in a product space, we have explained its limits both in terms of the difficulty in practical implementation, the increase of dimension and of the lack of robustness with respect to topology of the geometrical support. Somehow, the space of colored currents is not rich enough to encompass functional shapes whereas currents in the product is probably too big. These observations constitute our motivation to redefine a proper class of mathematical objects that would preserve the usefulness of currents while overcoming the previous drawbacks.

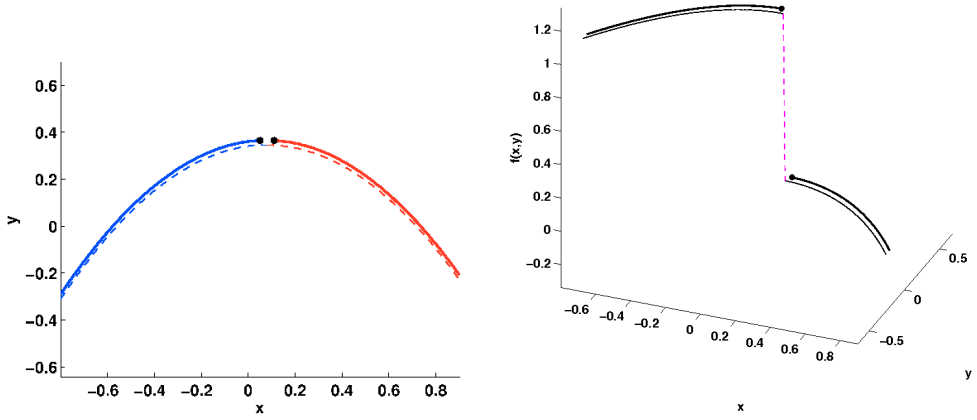


Figure 2.5: Product currents and topology. On the left, we show a disconnected 2D curve with signal values 0 in blue and 1 in red as well as the connected curve in dashed line. On the right hand side are the corresponding curves in the 3-dimensional geometry $\times$ signal space. What we want to emphasize here is the fact that no RKHS norm on product currents would provide a continuity of this representation with respect to connectivity : the difference between the two curves is the magenta dashed part which represents a pure variation in the signal domain.

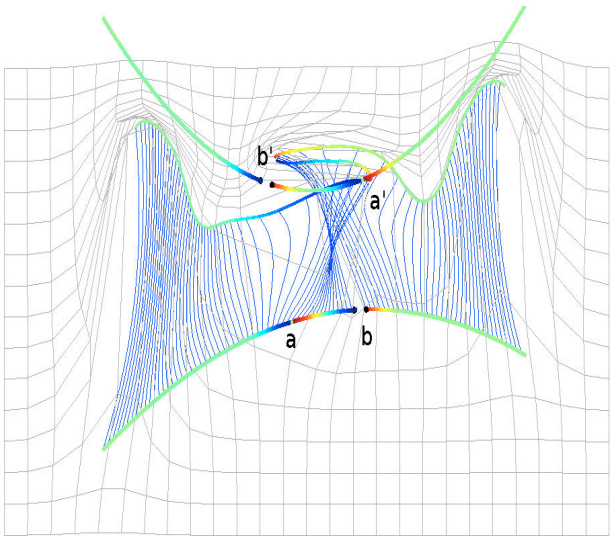


Figure 2.6: LDDMM matching of two planar curves with discontinuous signals and topological disconnections represented as currents in the product space $\mathbb{R}^2 \times \mathbb{R}$. Each curve has two points of functional discontinuity, one of them being also a disconnection of the geometrical support (b and b'). The resulting deformation is much perturbed by the disconnections since the algorithm intends to match connected part of the source shape on a connected part of the target shape.

2.2 Functional currents : mathematical considerations

In this section, we propose an extension of the notion of currents to represent functional shapes. The new mathematical objects we introduce, named *functional currents*, are **not** usual currents strictly speaking, as opposed to the methods presented in 2.1.3. They would rather derive from the very general concept of double current introduced originally by De Rham in [70].

2.2.1 Definition and basic properties of functional currents

As in the previous section, we will consider functional shapes (X, f), with X a d-dimensional rectifiable subset of the n-dimensional Euclidean space E and f a measurable mapping from X to the signal manifold M. In the theoretical framework that we present here, M can be any Riemannian manifold.

Functional p-forms and functional currents

We define the space of functional currents again as the dual of a space of continuous forms :

Definition 2.2.1. *We call a functional p-form on (E, M) an element of the space $C_0(E \times M, \Lambda^p E^*)$ which will be denoted by $\Omega_0^p(E, M)$ hereafter. We consider the uniform norm on $\Omega_0^p(E, M)$ defined by : $\|\omega\|_\infty = \sup_{(x,m)\in E\times M} \|\omega_{(x,m)}\|$. A functional p-current (or* fcurrent *in short) is defined as a continuous linear form on $\Omega_0^p(E, M)$ for the uniform norm. The space of functional p-currents will be therefore denoted $\Omega_0^p(E, M)'$.*

From the definition, one sees that functional currents are augmenting usual currents with positions in the signal domain M. A functional p-form is somehow a field of differential forms on the space E but spread on the product $E \times M$. Indeed a generic functional p-form can be written, in a coordinate system $(x_1, ..., x_n)$ of E,

$$\omega = \sum_{1\leq i_1<..<i_p\leq n} a_{i_1,...,i_p}(x, m)dx^{i_1} \wedge ... \wedge dx^{i_p} \tag{2.6}$$

Note, however, the important difference between the space of functional currents $\Omega_0^p(E, M)'$ and the space of currents in $E \times M$, $\Omega_0^p(E \times M)'$ discussed previously. With functional currents, the local geometry is still the one of the geometrical shape represented by an element of $\Lambda^p E$ as opposed to the product current setting that models the geometry of the lifted functional shape in $E \times M$, requiring the higher dimensional space of p-vectors $\Lambda^p(E \times M)$. Thus the space of functional currents is bigger than the one of currents on E but smaller than the space of currents on $E \times M$. Now, just as one can establish a correspondence between shapes and currents, to any functional shape we now associate a fcurrent.

Proposition 2.2.1. *Let (X, f) be a functional shape, with X an oriented rectifiable subset of dimension d and of finite volume and f a measurable function from X to M. For all $\omega \in \Omega_0^d(E, M)$, $x \mapsto \omega_{(x,f(x))}$ can be integrated along X. We set :*

$$C_{(X,f)}(\omega) \doteq \int_X \omega_{(x,f(x))}\,. \tag{2.7}$$

Then $C_{(X,f)} \in \Omega_0^d(E, M)'$ and therefore $(X, f) \mapsto C_{(X,f)}$ associates, to any functional shape, a functional current.

As for usual currents, an orientation on X is needed for the previous integral to have any meaning. We remind that X being rectifiable, there exists an oriented d-dimensional tangent space at almost every point $x \in X$ and, if we denote $\tau(x)$ the unit d-vector in $\Lambda^d E$ associated to the tangent space at x, then the integral in eq.(2.7) is :

$$C_{(X,f)}(\omega) \doteq \int_X \omega_{(x,f(x))}(\tau(x)) d\mathcal{H}^d(x)\,. \tag{2.8}$$

In the more restricted case where X is a submanifold, then the previous can also be written through local parametrization with a given partition of the unity of X. If $\gamma : U \to E$ is a parametrization of X with U an open subset of $\mathbb{R}^d$, then

$$C_{(X,f)}(\omega) = \int_U \omega_{(\gamma(u),f\circ\gamma(u))} \left(\frac{\partial \gamma}{\partial u_1} \wedge ... \wedge \frac{\partial \gamma}{\partial u_d} \right) du_1 ... du_d\,.$$

and it's an easy check that a reparametrization doesn't change the integral. Of course, as for regular currents, the previous correspondence between functional shapes and functional currents is not surjective. For instance, a sum of functional currents of the form $C_{(X,f)}$ do not generally derive from a functional shape. In the functional current framework, Dirac masses are naturally generalized by elementary functional currents or Dirac fcurrents $\delta^{\xi}_{(x,m)}$ for $x \in X$, $m \in M$ and $\xi \in \Lambda^p E$ such that $\delta^{\xi}_{(x,m)}(\omega) \doteq \omega_{(x,m)}(\xi)$. In the same way as explained in the previous section, one can give a discretized version of functional currents associated to (X, f) when a mesh is defined on X. $C_{(X,f)}$ is then approximated by a sum of Dirac fcurrents :

$$C_{(X,f)} \approx \sum_{k=1..N} \delta^{\xi_k}_{(x_k,m_k)} \tag{2.9}$$

In the particular case of a triangulated surface, the discretized version of the fcurrent can be simply obtained as explained for classical currents and by adding the 'interpolated value' of signal at each center point of triangles (or for a general signal manifold, the Frechet mean in M), as in the example of figure 2.7. In the following proposition, we try to give a more precise sense to eq.(2.9). We shall assume for simplicity that X is a C^1-submanifold of E which admits a parametrization $F : \Omega \to E$ with Ω an open subset of $\mathbb{R}^d$ and F an embedding. Namely, assume that we are given a d-dimensional mesh on a C^1-submanifold X whose faces will be denoted by $(\Delta_i)_{i=1,..,n}$ and that the signal . From the discrete representation, we can observe that functional currents have a very simple interpretation, which consists in attaching values of the signal f to the usual representation of X as a d-current. In fact, one could alternatively define functional forms as a tensor product space by making the identification $\Omega^d_0(E, M) = \Omega^d_0(E) \otimes C_0(M, \mathbb{R})$ and thus consider fcurrents as tensor products of d-currents on E and 0-currents (i.e measures) on M, which is closer to the notions presented in [70].

Diffeomorphic transport of fcurrents

What about diffeomorphic transport of functional currents ? This question cannot be addressed as simply as in the classical current setting in its full generality, which essentially results from the discussion in 2.1.2. As we explained, there are different possibilities both for the group of transformations that one can choose and the action of these deformations on the considered class of functional shapes. To remain general, suppose that a certain class of functional shapes together with a group $G \subset \mathrm{Diff}(E \times$

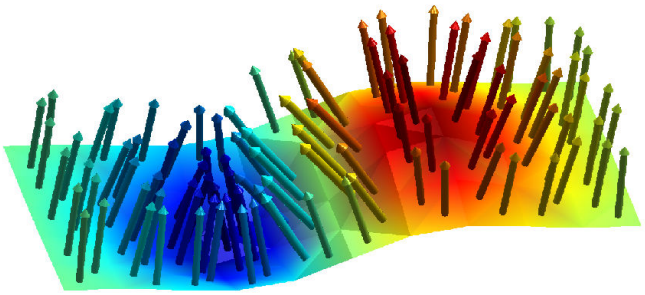


Figure 2.7: Dirac fcurrents computed from a triangulated functional surface.

M) and a left group action are fixed, we will note $\phi.(X, f)$ the action of $\phi \in G$ on a functional shape (X, f). Then,

Definition 2.2.2. *We call a deformation model on the space of functional currents a left action of the group G on $\Omega_0^d(E, M)'$ which is such that for any functional shape (X, f) and any diffeomorphism ϕ, if ϕ_* stands for the action on fcurrents, the following property holds :*

$$[\phi_* C_{(X,f)}](\omega) = C_{\phi.(X,f)}(\omega) \tag{2.10}$$

for all $\omega \in \Omega_0^d(E, M)$.

Eq.(2.10) is basically a commutation condition between the embedding of fshapes into fcurrents' space with the deformation transport. With usual currents, this always holds with the pull-back and push-forward actions. Here, it is necessary to adapt the definition of the action on fcurrents to be compatible with a given action on functional shapes.
In practical applications, this is usually not difficult. In the case of real valued signal data deformed as in eq.(2.1), the action of $\phi \in \mathrm{Diff}(E)$ on a functional current C can be defined in a very similar way :

$$\begin{cases} \phi_* C(\omega) \doteq C(\phi^* \omega), \ \forall \omega \in \Omega_0^d(E, M) \\ \\ \text{where for all } \forall x \in E, \ m \in M, \ \xi = \xi_1 \wedge ... \wedge \xi_p \in \Lambda^p E \\ \\ (\phi^* \omega)_{(x,m)}(\xi) \doteq \omega_{(\phi(x),m)}(d_x\phi(\xi_1) \wedge ... \wedge d_x\phi(\xi_d)) \, . \end{cases} \tag{2.11}$$

It can be easily checked from the previous equations that for all functional shapes (X, f), we have $\phi_* C_{(X,f)} = C_{(\phi(X), f \circ \phi^{-1})}$ as we expected under this model.
In the case of the group of deformations $\mathrm{Diff}(E) \times \mathcal{M}(E, \mathbb{R})$ acting as in eq.(2.2), we obtain the corresponding deformation model on fcurrents by setting :

$$\begin{cases} (\phi, h)_* C(\omega) \doteq C((\phi, h)^* \omega), \ \forall \omega \in \Omega_0^d(E, M) \\ \\ \text{where for all } \forall x \in E, \ m \in M, \ \xi = \xi_1 \wedge ... \wedge \xi_p \in \Lambda^p E \\ \\ ((\phi, h)^* \omega)_{(x,m)}(\xi) \doteq \omega_{(\phi(x), m + h(x))}(d_x\phi(\xi_1) \wedge ... \wedge d_x\phi(\xi_d)) \, . \end{cases} \tag{2.12}$$

These are essentially the two situations that shall be considered in numerical applications (section 2.3.2). Yet the same thing can be written for the general action on

fshapes given by eq.(2.5) which translates to functional currents by the relations :

$$\begin{cases} \psi_* C(\omega) \doteq C(\psi^* \omega), \ \forall \omega \in \Omega_0^d(E, M) \\ \\ \text{where for all } \forall x \in E, \ m \in M, \ \xi = \xi_1 \wedge ... \wedge \xi_p \in \Lambda^p E \\ \\ (\psi^* \omega)_{(x,m)}(\xi) \doteq \omega_{\psi(x,m)}(d_x \psi^g(\xi_1) \wedge ... \wedge d_x \psi^g(\xi_d)) \, . \end{cases} \tag{2.13}$$

Relationship with other currents

We mentioned earlier that the space of functional currents is 'bigger' than the one of usual currents, as we see easily if one takes the signal space as a singleton ($M = \{m\}$) and identifies $E \times M$ with E. A more rigorous way to say so is that we have a surjective application $s : \ \Omega_0^d(E, M) \to \Omega_0^d(E)$ which we obtain by fixing $m \in M$ and associating to $\omega \in \Omega_0^d(E, M)$,

$$s(\omega) = \omega_{(.,m)}(.) \in \Omega_0^d(E).$$

The dual application

$$j : \ C \in \Omega_0^d(E)' \longmapsto [C : \ \omega \mapsto C(s(\omega))] \in \Omega_0^d(E, M)'$$

is a one to one linear map.
However, as we could expect, it's a 'smaller' space than the one of currents in the product $E \times M$. Indeed, assuming, to simplify, that M is a vector space, we can embed $\Lambda^d E$ into $\Lambda^d(E \times M)$ by setting for $\xi_1 \wedge ... \wedge \xi_d \in \Lambda^d E, \ \tilde{l}(\xi_1 \wedge ... \wedge \xi_d) = (\xi_1, 0) \wedge ... \wedge (\xi_d, 0) \in \Lambda^d(E \times M)$. Thus, we can define the application :

$$\begin{aligned} \tilde{p} : \ \Omega_0^d(E \times M) &\longrightarrow \Omega_0^d(E, M) \\ \omega &\longmapsto \tilde{p}(\omega) / \ \tilde{p}(\omega)_{(x,m)}(\xi) = \omega_{(x,m)}(\tilde{l}(\xi)) \end{aligned}$$

On the other way, there is a surjective application $\tilde{l}$ from $\Lambda^d(E \times M)$ to $\Lambda^d E$ which is such that for all $(\xi_1, \eta_1), ..., (\xi_d, \eta_d) \in E \times M, \ l((\xi_1, \eta_1) \wedge ... \wedge (\xi_d, \eta_d)) = \xi_1 \wedge ... \wedge \xi_d$, which induces the application :

$$\begin{aligned} p : \ \Omega_0^d(E, M) &\longrightarrow \Omega_0^d(E \times M) \\ \omega &\longmapsto p(\omega) / \ p(\omega)_{(x,m)}(\xi) = \omega_{(x,m)}(l(\xi)) \end{aligned}$$

and we easily see that $\tilde{p} \circ p$ is the identity on $\Omega_0^d(E, M)$. It results that $\tilde{p}$ is a surjective application. We obtain therefore a dual application :

$$\begin{aligned} i : \ \Omega_0^d(E, M)' &\longrightarrow \Omega_0^d(E \times M)' \\ C &\longmapsto [i(C) : \ \omega \mapsto C(\tilde{p}(\omega))] \end{aligned}$$

that is one to one. In conclusion, we just proved that we have the following diagram :

$$\Omega_0^d(E)' \underset{j}{\hookrightarrow} \Omega_0^d(E, M)' \underset{i}{\hookrightarrow} \Omega_0^d(E \times M)'$$

2.2.2 Reproducing Kernel Hilbert Space structures on functional currents

In this section, we now address the fundamental question of *comparing* functional currents through appropriate metrics. For this purpose, we adapt the ideas of RKHS that were presented briefly for currents in the introduction chapter. In particular, we propose a generic way to build kernels that embeds functional currents into Hilbert spaces and provide closed form expressions of the dot product between Diracs.

Kernels' construction

As we have seen for currents, the theory of RKHS defines an inner product between currents through a certain kernel function satisfying some regularity and boundary conditions. We shall take advantage of the idea that functional d-currents can also be considered as the tensor product of d-currents on E and 0-currents on M. We remind that for a set A and a finite-dimensional vector space V, we say that K is a V-positive kernel on A if $K: A\times A \to \mathcal{L}(V)$, $K(a',a) = K(a,a')^*$ and K satisfies the positivity property : for all $a_1, ..., a_N \in A$ and $v_1, .., v_N \in V$, $\sum_{i,j=1}^{N} \langle K(a_i, a_j)v_i, v_j\rangle \geq 0$. Let's first mention the following very general result from kernel theory :

Lemma 2.2.1. *Let A and B be two sets and K_A a V_A-positive kernel on A, K_B a V_B-positive kernel on B where V_A and V_B are two finite dimensional vector spaces. Then, the tensor product $K_A \otimes K_B$ defined by :*

$$\langle K_A \otimes K_B\left((a,b),(a',b')\right) u_A \otimes u_B, v_A \otimes v_B\rangle = \langle K_A(a,a')u_A, v_A\rangle.\langle K_B(b,b')u_B, v_B\rangle$$

is a $V_A \otimes V_B$-positive kernel on $A \times B$.

Proof. Let $N \in \mathbb{N}^*$, $a_1, .., a_N \in A$ and $b_1, ..., b_N \in B$, $u_1, .., u_N \in V_A$ and $v_1, .., u_N \in V_B$. We have by definition of the kernel :

$$\begin{aligned}\sum_{i,j=1}^{N} \langle K_A \otimes K_B\left((a_i,b_i),(a_j,b_j)\right) u_i \otimes v_i, u_j \otimes v_j\rangle &= \sum_{i,j=1}^{N} \langle K_A(a_i,a_j)u_i, u_j\rangle.\langle K_B(b_i,b_j)v_i, v_j\rangle \\ &= Tr(M_A M_B)\end{aligned}$$

where $M_A = (\langle K_A(a_i,a_j)u_i, u_j\rangle)_{i,j}$ and $M_B = (\langle K_B(b_i,b_j)v_i, v_j\rangle)_{i,j}$ are both positive symmetric matrices (because K_A and K_B are positive kernels). It is then a well-known fact that $Tr(M_A M_B) \geq 0$ and thus $K_A \otimes K_B$ is a positive kernel. □

Note that if K_B is a real kernel (i.e $V_B = \mathbb{R}$) then $K_A \otimes K_B$ is a $V_A \otimes \mathbb{R} \approx V_A$-positive kernel on $A \times B$. Now, we apply lemma 2.2.1 to the case of functional currents.

Proposition 2.2.2. *Let $K_g : E\times E \to \mathcal{L}(\Lambda^d E)$ be a positive kernel on the geometrical space E and $k_f : M\times M \to \mathbb{R}$ a real positive kernel on the signal space M. We assume that both kernels are continuous, bounded and vanishing at infinity. Then $K_g \otimes k_f$ defines a $\Lambda^d E$-positive kernel on $E \times M$ whose corresponding reproducing Hilbert space W is continuously embedded into $\Omega_0^d(E, M)$.*

Proof. From the conditions on both kernels, we know that to K_g and k_f correspond two RKHS W_g and W_f that are respectively embedded into $\Omega_0^p(E)$ and $C_0(M, \mathbb{R})$ (cf section 1.2.4 in chapter 1). By applying lemma 2.2.1 with $A = E$, $B = M$,

$V_A = \Lambda^d E$ and $V_B = \mathbb{R}$, we get that $K \doteq K_g \otimes k_f$ defines a $\Lambda^d E$-positive kernel on $E \times M$. Moreover, since k_f is real-valued, we have the following expression for K :

$$K\left((x_1, m_1), (x_2, m_2)\right) = k_f(m_1, m_2).K_g(x_1, x_2) \tag{2.14}$$

The kernel K corresponds to a unique RKHS W that is the completion of the vector space spanned by all the functions $\{k_f(., m).K_g(., x)\xi\}$ for $x \in E,\ m \in M,\ \xi \in \Lambda^d E$. Since functions $k_f(., m)$ and $K_g(., x)$ are both continuous and vanishing at infinity from, this holds for $k_f(., m).K_g(., x)\xi$ as well, so that W is indeed embedded into $\Omega_0^d(E, M)$. It only remains to prove that the embedding is *continuous*, which reduces to dominate the uniform norm by $\|.\|_W$.
Let $\omega \in W$. For all $(x, m) \in E \times M$ and $\xi \in \Lambda^d E$ such that $|\xi| = 1$, we have

$$|\omega_{(x,m)}(\xi)| = |\delta^{\xi}_{(x,m)}(\omega)|\,. \tag{2.15}$$

Since W is a RKHS, all $\delta^{\xi}_{(x,m)}$ are continuous linear forms on W. In addition, the Riesz representation theorem provides an isometry $K_W : W' \to W$. Then :

$$\begin{aligned}
\langle \delta^{\xi_1}_{(x_1,m_1)}, \delta^{\xi_2}_{(x_2,m_2)} \rangle_{W'} &= \langle K_W(\delta^{\xi_1}_{(x_1,m_1)}), K_W(\delta^{\xi_2}_{(x_2,m_2)}) \rangle_W \\
&= \langle k_f(., m_1)K_g(., x_1)\xi_1, k_f(., m_2)K_g(., x_2)\xi_2 \rangle_W \\
&= k_f(m_1, m_2).\langle K_g(x_1, x_2)\xi_1, \xi_2 \rangle
\end{aligned} \tag{2.16}$$

Now, back to equation (2.15), we have :

$$\begin{aligned}
|\omega_{(x,m)}(\xi)| &\leq \|\delta^{\xi}_{x,m}\|_{W'}\ \|\omega\|_W \\
&\leq \sqrt{k_f(m, m).\langle K_g(x, x)\xi, \xi \rangle}\ \|\omega\|_W
\end{aligned}$$

Since we assume that $m \mapsto k_f(m, m)$ and $x \mapsto K_g(x, x)$ are bounded we deduce that $\sqrt{k_f(m, m).\langle K_g(x, x)\xi, \xi \rangle}$ is bounded with respect to x, m and ξ with $|\xi| = 1$. Hence, by taking the supremum in the previous equation, we finally get

$$\|\omega\|_\infty \leq C\|\omega\|_W$$

which means that the embedding is continuous. □

By duality, we have an application

$$\imath^* :\ \Omega_0^d(E, M)' \to W'$$

that sends any functional current into W' (which is also a Hilbert space) and we can thus compare functional shapes with the norm provided by the kernel. However, it's important to point out that $\imath^*$ is not necessarily an embedding. If for instance $M = \mathbb{S}^1$ and $\forall \theta_1, \theta_2 \in \mathbb{S}^1,\ k_f(\theta_1, \theta_2) = 1$, we see that for $x \in E,\ \xi \in \Lambda^d E$ and $\theta_1, \theta_2 \in \mathbb{S}^1$, we have $\imath^*(\delta^{\xi}_{(x,\theta_1)}) = \imath^*(\delta^{\xi}_{(x,\theta_2)})$ and $\imath^*$ is not injective. The fact that $\imath^*$ is an embedding is called the **C_0-universality** property of the kernel and, as proved in [20], it is equivalent to the property of W being dense in $\Omega_0^d(E, M)$. This is the case in particular if both W_g and W_f are respectively dense in $\Omega_0^d(E)$ and $C_0(M, \mathbb{R})$. In [20] are also given a few examples of universal kernels and building techniques in the general case. Gaussian kernels are the simplest examples, which explains their wide use in practical applications.
To sum up, proposition 2.2.2 provides a quite natural (but not unique) way to build

kernels for functional currents by making the tensor product of kernels defined separately in the geometrical domain (d-currents in E) and in the signal domain (0-currents in M). Note that non product kernels could also be used but the product situation corresponds to an independence assumption between shape and functional information which is natural when modeling the residual difference between two functional shapes as noise. Moreover, the use of product kernels leads to simpler and faster computational schemes, as we shall explain later on. We see that, eventually, everything relies on the specification of kernels on E and M.
Kernels on the vector space E obviously do not raise any additional difficulty in our approach compared to usual current settings. Among others, we have already mentioned radial scalar kernels defined for $x, y \in E$ by $K(x,y) = k(|x-y|).\mathrm{Id}_E$ where k is a function defined on $\mathbb{R}_+$ and vanishing at infinity. This family of kernels is the only one that induces a RKHS norm invariant for affine isometries. The most usual is the Gaussian kernel, which is also C_0 universal as we said earlier, defined by $K(x,y) = \exp\left(-\frac{|x-y|^2}{\sigma^2}\right)\mathrm{Id}_E$, σ being a scale parameter that can be interpreted as a range of interactions between point positions.
The definition of a kernel on the signal manifold M is often more involved. However, it is important to note that, in the setting of functional currents, this issue is drastically simplified because we only need to define **real-valued** kernels on M. This is contrasting, for instance, with the idea of product space currents of subsection 2.1.3, that would require kernels living in the exterior product of the fiber bundle of M. If M is an embedded manifold, one could simply use the following trivial property on kernels :

Property 2.2.1. *Let k be a real-valued positive kernel on a set A. Then, if B is a subset of A, the function $\tilde{k} : B \times B \to \mathbb{R}$ obtained by restriction is a positive kernel on B.*

In our case, if M is a submanifold of some vector space F, obtaining real-valued kernels on M becomes straightforward by restriction to M of kernels defined on F. For vector kernels, this process does not apply because of the non-equal tangent spaces at different points. A second important fact is that the C_0-universality property transfers by restriction, i.e :

Property 2.2.2. *If k is a real-valued positive C^0-universal kernel on a Banach space A and B a closed subset of A then the restricted kernel $\tilde{k}$ is also C^0-universal.*

This is proved in [20], corollary 3. In practical terms, prop 2.2.2 guarantees that we can obtain C^0-universal kernel on M if we restrict for instance a Gaussian kernel in F on M. In the case of an abstract signal manifold M for which we don't have a canonical embedding into a vector space, the resulting kernels would depend on the way that M is embedded into a determined space.

Functional currents' kernel metrics : a few examples

From eq.(2.16), we see that the metric between two Dirac fcurrents is a combination of geometry comparison through the positions of the two Diracs and the local volume elements, and signal proximity evaluated by the kernel k_f. More generally, one can write the distance on functional shapes that results from these metrics. If (X, f) and (Y, g) are two functional shapes to which we associate their functional currents'

representations $C_{(X,f)}$ and $C_{(Y,g)}$, then we have :

$$\begin{aligned}
&\|C_{(X,f)} - C_{(Y,g)}\|^2_{W'} = \|C_{(X,f)}\|^2_{W'} + \|C_{(Y,g)}\|^2_{W'} - 2\langle C_{(X,f)}, C_{(Y,g)}\rangle_{W'} \\
&= \int_{X\times X} k_f(f(x_1), f(x_2)).\langle K_g(x_1,x_2)\tilde{\xi}_X(x_1), \tilde{\xi}_X(x_2)\rangle d\mathcal{H}^d(x_1)d\mathcal{H}^d(x_2) \\
&+ \int_{Y\times Y} k_f(g(y_1), g(y_2)).\langle K_g(y_1,y_2)\tilde{\xi}_Y(y_1), \tilde{\xi}_Y(y_2)\rangle d\mathcal{H}^d(y_1)d\mathcal{H}^d(y_2) \\
&- 2\int_{X\times Y} k_f(f(x), g(y)).\langle K_g(x,y)\tilde{\xi}_X(x), \tilde{\xi}_Y(y)\rangle d\mathcal{H}^d(x)d\mathcal{H}^d(y) \qquad (2.17)
\end{aligned}$$

where $\tilde{\xi}_X(x)$ is the unit d-vector of $\Lambda^d E$ characterizing the oriented tangent space to X at x. We can also write the exact equivalent of eq.(2.17) for discrete shapes. Namely if $C_{(X,f)} = \sum_{i=1}^n \delta^{\xi_i^X}_{(x_i,f_i)}$ and $C_{(Y,g)} = \sum_{k=1}^m \delta^{\xi_k^Y}_{(y_k,g_k)}$, then :

$$\begin{aligned}
\|C_{(X,f)} - C_{(Y,g)}\|^2_{W'} = &\sum_{i=1}^n\sum_{j=1}^n k_f(f_i,f_j).\langle K_g(x_i,x_j)\xi_i^X, \xi_j^X\rangle \\
&+ \sum_{k=1}^m\sum_{l=1}^m k_f(g_k,g_l).\langle K_g(y_k,y_l)\xi_k^Y, \xi_l^Y\rangle \\
&- 2\sum_{i=1}^n\sum_{k=1}^m k_f(f_i,g_k).\langle K_g(x_i,y_k)\xi_i^X, \xi_k^Y\rangle \qquad (2.18)
\end{aligned}$$

This last equation gives a rather simple closed formula for computing distances between functional shapes based only on kernel evaluations. We also see that for sufficiently regular kernels, we obtain differentiable distances with respect to position and functional values, which shall be presented in more details in the next sections. Now, we try to give a little more insight on how these metrics behave by considering more specifically the case of real-valued signals $M = \mathbb{R}$. Let's also take the kernels k_f and K_g to be Gaussian kernels, i.e $k_f(f_1,f_2) = e^{-\frac{|f_2-f_1|^2}{\sigma_f^2}}$ and $\langle K_g(x_1,x_2)\xi_1,\xi_2\rangle = e^{-\frac{|x_2-x_1|^2}{\sigma_g^2}}\langle\xi_1,\xi_2\rangle$. The parameters σ_g and σ_f are the scales of the kernels respectively in the geometrical and signal domain. The interaction between two diracs decreases as they lie 'farther' from each other with respect to σ_g or if the difference between the signals they carry becomes larger compared to σ_f. In particular, if $\sigma_f \to +\infty$, we have $k_f(f_1,f_2) = 1$ for all f_1, f_2 and the distance we get is exactly the distance between the purely geometrical shapes X and Y given by currents. On the other hand, if we let $\sigma_g \to +\infty$, as if all points were seen by the kernel at the same location geometrically, eq.(2.17) and eq.(2.18) become somehow a metric between the signal distributions of the two objects. However, we see that it still takes into account some of the geometry through the local volume elements and just as for usual currents, there are some cancellation effects occurring essentially because of orientation, which causes these metrics to be sensitive only to boundaries of level sets. We shall be dealing in more details with these questions in the next chapter.
Another interesting subcase that gives a simpler vision of such metrics is the very special one of 2D **images**. If $\Omega \in \mathbb{R}^2$ is the fixed definition domain of images, one can represent an image $I : \Omega \to \mathbb{R}$ as a functional current $C_{(\Omega,I)} \in \Omega^2_0(\mathbb{R}^2,\mathbb{R})'$. In that case, the domain is flat, there are no currents' cancellation effects as we mentioned in the previous paragraph, and local geometry is described by the same 2-vector at each

point, which is simply a surface area and the kernel on the geometrical space can be considered as a real-valued kernel k_g. The distance of eq.(2.17) then becomes :

$$\|C_{(\Omega,f)} - C_{(\Omega,g)}\|^2_{W'} = \int_{\Omega\times\Omega} k_g(x,y).[k_f(I_1(x),I_1(y)) + k_f(I_2(x),I_2(y)) - 2k_f(I_1(x),I_2(y))]dxdy \tag{2.19}$$

- For two Gaussian kernels $k_g(x,y) = e^{-\frac{|x-y|^2}{\sigma_g^2}}$ and $k_f(g,h) = e^{-\frac{|g-h|^2}{\sigma_f^2}}$, we see that the distance we get combines geometric proximity between positions at scale σ_g and proximity of the intensities at scale σ_f, which is actually close to the idea of bilateral filtering in image processing [76].
- If one takes for k_f the linear kernel, $k_f(g,h) = gh$, and let $\sigma_g \to 0$, eq.(2.19) becomes :

$$\begin{aligned}\|C_{(\Omega,f)} - C_{(\Omega,g)}\|^2_{W'} &= \int_\Omega I_1(x)^2 + I_2(x)^2 - 2I_1(x)I_2(x)dx \\ &= \int_\Omega (I_1(x) - I_2(x))^2 dx\end{aligned}$$

  and this is the usual L^2-distance between images.
- Now, if we look at images at very large geometrical scale by letting $\sigma_g \to +\infty$, eq.(2.19) is a distance between the intensities' distributions of the two images ; more precisely, it can be interpreted as a metric between the measures $\mu_1 = \lambda \circ I_1^{-1}$ and $\mu_2 = \lambda \circ I_2^{-1}$, where λ is the Lebesgue measure on $\mathbb{R}^2$. Indeed, by a change of variables in the integrals :

$$\begin{aligned}\|C_{(\Omega,f)} - C_{(\Omega,g)}\|^2_{W'} &= \int_{\Omega\times\Omega} k_f(I_1(x),I_1(y))d\lambda(x)d\lambda(y) \\ &+ \int_{\Omega\times\Omega} k_f(I_2(x),I_2(y))d\lambda(x)d\lambda(y) \\ &- 2\int_{\Omega\times\Omega} k_f(I_1(x),I_2(y))d\lambda(x)d\lambda(y) \\ &= \int_{\mathbb{R}\times\mathbb{R}} k_f(u,v)d\mu_1(u)d\mu_1(v) + \int_{\mathbb{R}\times\mathbb{R}} k_f(u,v)d\mu_2(u)d\mu_2(v) \\ &- 2\int_{\mathbb{R}\times\mathbb{R}} k_f(u,v)d\mu_1(u)d\mu_2(v)\end{aligned}$$

  In the discrete setting, representing every pixel as a Dirac fcurrent $\delta_{(x_i,I(x_i))}$, these measures can be thought as intensity **histograms** counting the number of pixels having a particular intensity. Thus the metrics we obtain at infinite geometrical scales are measuring differences in the intensities histograms of the images. For instance, if k_f is Gaussian and we let $\sigma_f \to 0$, one can verify that the previous equation gives exactly the L^2-distance between the images' histograms.

In summary, these last examples show that functional currents' representation combined with the use of kernels provide a very wide range of possible metrics between functional shapes, from distances based on pure geometry on one extreme to distances that simply compare histograms of signals on the other. It also enables analysis at various scales both in the geometrical and functional domain as well as *multi-scale* metrics : one simple way being to make sums of Gaussian kernels with different σ_f and σ_g.

2.2.3 Properties of kernel norms on fcurrents

We now turn to general properties of the distances we have exhibited previously and show the theoretical benefits of this approach with respect to the original problem we set out to handle in this chapter. As in the previous section, we assume that two kernels K_g and k_f are given respectively on space E and manifold M, providing two RKHS W_g and W_f. We will also assume, in all this section, that both kernels satisfy the requirements of proposition 2.2.2.

Control by L^1-norm on a fixed geometrical support

We start by the following very simple lemma :

Lemma 2.2.2. *For all $x_1, x_2 \in E$, $\xi_1, \xi_2 \in \Lambda^d E$ and $m_1, m_2 \in$, the following inequality holds :*

$$\|\delta^{\xi_2}_{(x_2,m_2)} - \delta^{\xi_1}_{(x_1,m_1)}\|_{W'} \leq \|\delta_{m_1}\|_{W'_f}\|\delta^{\xi_2}_{x_2} - \delta^{\xi_1}_{x_1}\|_{W'_g} + \|\delta^{\xi_2}_{x_2}\|_{W'_g}\|\delta_{m_2} - \delta_{m_1}\|_{W'_f}\,. \quad (2.20)$$

Proof. This is just a consequence of the triangular inequality since :

$$\|\delta^{\xi_2}_{(x_2,m_2)} - \delta^{\xi_1}_{(x_1,m_1)}\|_{W'} \leq \|\delta^{\xi_2}_{(x_2,m_2)} - \delta^{\xi_2}_{(x_2,m_1)}\|_{W'} + \|\delta^{\xi_2}_{(x_2,m_1)} - \delta^{\xi_1}_{(x_1,m_1)}\|_{W'}$$

and by the expression of the kernel eq.(2.14),

$$\begin{aligned}\|\delta^{\xi_2}_{(x_2,m_2)} - \delta^{\xi_2}_{(x_2,m_1)}\|^2_{W'} &= \langle K_g(x_2,x_2)\xi_2, \xi_2\rangle.\,(k_f(m_2,m_2) - 2k_f(m_1,m_2) + k_f(m_1,m_1)) \\ &= \|\delta^{\xi_2}_{x_2}\|^2_{W'_g}\|\delta_{m_2} - \delta_{m_1}\|^2_{W'_f}\end{aligned}$$

and similarly we have $\|\delta^{\xi_2}_{(x_2,m_1)} - \delta^{\xi_1}_{(x_1,m_1)}\|_{W'} = \|\delta_{m_1}\|_{W'_f}\|\delta^{\xi_2}_{x_2} - \delta^{\xi_1}_{x_1}\|_{W'_g}$. □

Therefore, the RKHS distance between punctual fcurrents is dominated both with respect to the variation of their geometrical parts and of their functional values. This is the general idea we will formulate in a more precise way with the two following propositions. We denote by d_M the geodesic distance induced on M by its Riemannian structure. The next proposition examines the case where the geometrical support is a fixed support X and shows that the variation of the W'-norm is then dominated by the L^1 norm on X.

Theorem 2.2.1. *Let X be a d-dimensional oriented rectifiable subset of E of finite volume and $f_1 : X \to M$ and $f_2 : X \to M$ two measurable functions defined on X taking value in M. We assume that W_f is continuously embedded into $C^1_0(M, \mathbb{R})$. Then, there exists a constant β such that :*

$$\|C_{(X,f_1)} - C_{(X,f_2)}\|_{W'} \leq \beta \int_X d_M(f_1(x), f_2(x)) d\mathcal{H}^d(x)$$

Proof. We recall that for a d-dimensional oriented rectifiable subset there exists for $\mathcal{H}^d$-almost every point $x \in X$ a unit d-vector $\tilde{\xi}(x)$ and by definition :

$$C_{(X,f)}(\omega) = \int_X \omega_{(x,f(x))}(\tilde{\xi}(x)) d\mathcal{H}^d(x) = \int_X \delta^{\tilde{\xi}(x)}_{(x,f(x))}(\omega) d\mathcal{H}^d(x)$$

Now we have by triangular inequality on $\|.\|_{W'}$:

$$\|C_{(X,f_1)} - C_{(X,f_2)}\|_{W'} \leq \int_X \|\delta^{\tilde{\xi}(x)}_{(x,f_1(x))} - \delta^{\tilde{\xi}(x)}_{(x,f_2(x))}\|_{W'} d\mathcal{H}^d(x)\,. \quad (2.21)$$

From lemma 2.2.2 and eq.(2.20), $\|\delta^{\tilde{\xi}(x)}_{(x,f_1(x))} - \delta^{\tilde{\xi}(x)}_{(x,f_2(x))}\|_{W'} \leq \|\delta^{\tilde{\xi}(x)}_x\|_{W_g'}\|\delta_{f_2(x)} - \delta_{f_1(x)}\|_{W_f'}$. Now, for any m_1, $m_2 \in M$ and $h \in W_f$ we have :

$$\begin{aligned} |(\delta_{m_1} - \delta_{m_2})(h)| &= |h(m_1) - h(m_2)| \\ &\leq \|Dh\|_\infty d_M(m_1, m_2) \\ &\leq \text{Cte}\, \|h\|_{W_f} d_M(m_1, m_2) \end{aligned}$$

the last inequality resulting from the continuous embedding $W_f \hookrightarrow C_0^1(M, \mathbb{R})$. Therefore we get

$$\|\delta_{f_2(x)} - \delta_{f_1(x)}\|_{W_f'} \leq \text{Cte}\, d_M(f_1(x), f_2(x))\,.$$

Moreover, since we assume that the kernel K_g is bounded, we also have $\|\delta^{\tilde{\xi}(x)}_x\|_{W_g'} \leq$ Cte. Back to equation eq.(2.21), we get from the previous derivations the existence of a constant $\beta > 0$ such that :

$$\|C_{(X,f_1)} - C_{(X,f_2)}\|_{W'} \leq \beta \int_X d_M(f_1(x), f_2(x)) d\mathcal{H}^d(x)$$

which proves the stated result. □

A straightforward consequence of theorem 2.2.1 and the dominated convergence theorem is that if f_n is a sequence of functions on X that converges pointwisely to a function f, then $C_{(X,f_n)} \to C_{(X,f)}$. In other words, **pointwise convergence of the signal implies convergence for the W'-norm**. We shall see, with the following results, that the reverse domination does not hold in general.

Control by deformations' metrics

The purpose is now to link the W'-metrics with functional shapes' deformations. We show that we can basically control the distance in the space of fcurrents between a functional shape and its deformed version by the amount of deformation. Let's first focus the simplest transport model given by 2.1 and assume that diffeomorphisms ϕ are C^1 and such that $|\phi(x) - x| \to 0$ and $|d_x\phi - \text{Id}| \to 0|$ as $x \to \infty$. The central result of this section is the following :

Theorem 2.2.2. *Let X be a d-dimensional oriented rectifiable subset of E of finite volume and $f : X \to M$ a measurable function. Assume that W_g is continuously embedded into $C_0^1(E, \Lambda^d E^*)$. There exists a constant $\gamma > 0$ depending only on the kernels K_g and k_f such that, for all deformation ϕ :*

$$\|C_{(X,f)} - C_{(\phi(X),f\circ\phi^{-1})}\|_{W'} \leq \gamma \mathcal{H}^d(X) \|\phi - Id\|_{1,\infty}$$

Proof. For the entire proof, we shall use the same notation Cte for writing the successive different 'universal' constants (i.e not dependent on the shape (X, f) and the deformation ϕ). Starting as in theorem 2.2.1, we have :

$$\|C_{(X,f)} - C_{(\phi(X),f\circ\phi^{-1})}\|_{W'} \leq \int_X \|\delta^{\tilde{\xi}(x)}_{(x,f(x))} - \delta^{d_x\phi.\tilde{\xi}(x)}_{(\phi(x),f(x))}\|_{W'} d\mathcal{H}^d(x) \tag{2.22}$$

where for the volume element $\tilde{\xi}(x) = \tilde{\xi}_1(x) \wedge ... \wedge \tilde{\xi}_d(x)$, $d_x\phi.\tilde{\xi}(x)$ is the transported volume element by ϕ, equal to $d_x\phi.\tilde{\xi}(x) = d\phi_x(\tilde{\xi}_1(x)) \wedge ... \wedge d\phi_x(\tilde{\xi}_d(x))$. From eq.(2.20) we get :

$$\|\delta^{\tilde{\xi}(x)}_{(x,f(x))} - \delta^{d_x\phi.\tilde{\xi}(x)}_{(\phi(x),f(x))}\|_{W'} \leq \|\delta_{f(x)}\|_{W_f'} \|\delta^{d_x\phi.\tilde{\xi}(x)}_{\phi(x)} - \delta^{\tilde{\xi}(x)}_x\|_{W_g'}$$

and since the kernel k_f is bounded we have $\|\delta_{f(x)}\|_{W'_f} \leq \text{Cte}$. Moreover :

$$\|\delta^{d_x\phi.\tilde{\xi}(x)}_{\phi(x)} - \delta^{\tilde{\xi}(x)}_{x}\|_{W'_g} \leq \|\delta^{d_x\phi.\tilde{\xi}(x)}_{\phi(x)} - \delta^{\tilde{\xi}(x)}_{\phi(x)}\|_{W'_g} + \|\delta^{\tilde{\xi}(x)}_{\phi(x)} - \delta^{\tilde{\xi}(x)}_{x}\|_{W'_g}\,. \tag{2.23}$$

Focusing first on the second term, we can write for any $\omega \in W_g$,

$$\begin{aligned}|(\delta^{\tilde{\xi}(x)}_{\phi(x)} - \delta^{\tilde{\xi}(x)}_{x})(\omega)| &= |\omega_{\phi(x)}(\tilde{\xi}(x)) - \omega_x(\tilde{\xi}(x))| \\ &\leq |d\omega|_\infty.|\phi(x) - x|.|\tilde{\xi}(x)|.\end{aligned}$$

We have in addition $|\tilde{\xi}(x)| = 1$. W_g being continuously embedded in $C^1_0(E, \Lambda^d E^*)$, $|d\omega|_\infty \leq \text{Cte}\|\omega\|_{W_g}$. Thus, we deduce that $\|\delta^{\tilde{\xi}(x)}_{\phi(x)} - \delta^{\tilde{\xi}(x)}_{x}\|_{W'_g} \leq \text{Cte}\|\phi - \text{Id}\|_\infty$. As for the first term in eq.(2.22),

$$\|\delta^{d_x\phi.\tilde{\xi}(x)}_{\phi(x)} - \delta^{\tilde{\xi}(x)}_{\phi(x)}\|_{W'_g} = \|\delta^{d_x\phi.\tilde{\xi}(x)-\tilde{\xi}(x)}_{\phi(x)}\|_{W'_g} \leq \text{Cte}|d_x\phi.\tilde{\xi}(x) - \tilde{\xi}(x)|$$

since K_g is bounded and, by definition, $|d_x\phi.\tilde{\xi}(x) - \tilde{\xi}(x)| \leq |d\phi - \text{Id}|_\infty.|\tilde{\xi}| = |d\phi - \text{Id}|_\infty$. It results that we get eventually :

$$\|C_{(X,f)} - C_{(\phi(X),\psi\circ f\circ\phi^{-1})}\|_{W'} \leq \text{Cte}.\|\phi - \text{Id}\|_{1,\infty}\int_X d\mathcal{H}^d(x) = \text{Cte}.\mathcal{H}^d(X)\|\phi - \text{Id}\|_{1,\infty}$$

which concludes the proof. □

This is somehow the analogous to theorem 2.2.1 for variations of the geometrical support : it shows that the RKHS norm is Lipschitz regular with respect to deformations. This property is a natural one to expect from a distance in our context, but note that many usual metrics do not satisfy it. This is notably what happens for L^p distances. Indeed, if we specify the previous to the particular case where ϕ is a small deformation that leaves X globally invariant ($\phi(X) = X$) and we wish to compare the initial functional shape (X, f) with the deformed one $(\phi(X), f\circ\phi^{-1}) = (X, f\circ\phi^{-1})$, then by theorem 2.2.2, we know that, for any function f, the fcurrent's distance remains small if the deformation ϕ is small. It is no longer true if we compute instead $\int_X |f - f\circ\phi^{-1}|^p$, the L^p distance on X ($0 < p \leq \infty$). This is easily seen if we choose for X the unit circle $\mathbb{S}^1$ and consider crenelated signals as in figure 2.8. Introducing the operator $\tau_{d\theta}$ that acts on functional shapes by rotation of an angle $d\theta$, we see indeed that :

$$\sup_{f\in L^p(\mathbb{S}^1),\|f\|_{L^p}\leq 1}\int_{\mathbb{S}^1}|f - f\circ\tau^{-1}_{d\theta}|^p = 1$$

whereas, according to theorem 2.2.2

$$\sup_{f\in L^p(\mathbb{S}^1),\|f\|_{L^p}\leq 1}\|C_{(X,f)} - C_{(X,f\circ\tau^{-1}_{d\theta})}\|_{W'} = O(d\theta)\,.$$

This eventually shows that W' norm and L^p norm on a fixed geometrical support are **not** equivalent in general. This simple example also reveals the limits of some previous approaches for functional shape comparison that rely on estimating correspondence maps between the geometrical supports before comparing signals. RKHS norms on fcurrents successfully avoid such issues because shape and function are compared through one single norm that has enough spatial regularity to compare functional values at points in a certain neighborhood.

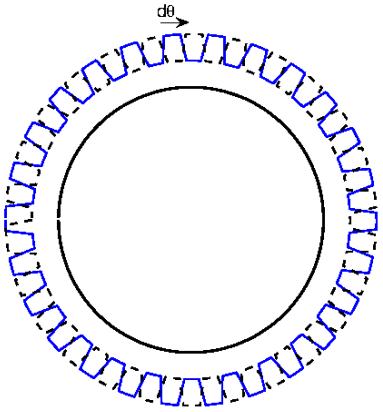


Figure 2.8: Comparison of the fcurrent's norm and the L^p norm on a fixed geometrical support : example of crenelations on the unit circle.

We can also draw a few more consequences of theorem 2.2.2. In particular, as already explained in chapter 1, we can model diffeomorphisms of E as flows of time-varying vector fields. Namely, we consider vector fields $v(t,x)$ such that for all $t \in [0,1]$, $v(t,.)$ is a C^1 vector field of E such that u and its first order derivatives are vanishing at infinity. We define $|v|_{1,\infty} = |v|_\infty + |dv|_\infty$. Then, χ^1 denotes the space of all such time-varying vector fields such that $\int_0^1 |v(t,.)|_{1,\infty} dt < +\infty$ and it is naturally equipped with the norm $\|u\|_{\chi^1} = \int_0^1 |u(t,.)|_{1,\infty} dt$. As already explained in chapter 1 and [86], any $u \in \chi^1$ gives a flow application $\phi_t^v : x \mapsto \phi_t^v(x)$ solving the ODE $\frac{dy}{ds} = v(s,y)$ with initial condition $y(0) = x \in E$, and for all $t \in [0,1]$, ϕ_t^v is a C^1-diffeomorphism of E such that $|\phi_t^v(x) - x| \to 0$ and $|d_x\phi_t^v - \mathrm{Id}| \to 0|$ as $x \to \infty$. We have in addition the following lemma (we refer to [41] or [86] for the details) :

Lemma 2.2.3. *For all $R > 0$ there is a constant $C(R) > 0$ such that, for all $u \in \chi^1$ with $\|u\|_{\chi^1} \leqslant R$ and $t \in [0,1]$:*

$$\|\phi_t^v - Id\|_\infty \leqslant C(R).\|v\|_{\chi^1}, \ \|d\phi_t^v - Id\|_\infty \leqslant C(R).\|v\|_{\chi^1} .$$

From theorem 2.2.2 and lemma 2.2.3, we obtain straightforwardly the following control :

Corollary 2.2.1. *Let X be a d-dimensional oriented rectifiable subset of E of finite volume and $f : X \to M$ a measurable function. Assume that W_g is continuously embedded into $C_0^1(E, \Lambda^d E^*)$. There exists a constant $\gamma > 0$ depending only on the kernels K_g and k_f such that, for all $v \in \chi^1$ with $\|u\|_{\chi^1}$ small enough :*

$$\|C_{(X,f)} - C_{(\phi_1^v(X), f\circ(\phi_1^v)^{-1})}\|_{W'} \leq \gamma \mathcal{H}^d(X)\|v\|_{\chi^1}$$

In addition, assuming sufficient spatial regularity on the vector field v, one can show the important property of weak continuity of W' norms with respect to v. It relies on the following lemma, that is proved in [41] chap. 1 or in [86] :

Lemma 2.2.4. *Let V be a Hilbert space of vector fields on E such that V is continuously embedded into $C_0^2(E,E)$. Then, if v_n is a sequence of vector fields in $L^2([0,1],V) = L^2_V$ such that $v_n \to v$ weakly in L^2_V, $\phi_t^{v_n}$ and $d\phi_t^{v_n}$ converge towards ϕ_t^v and $d\phi_t^v$ uniformly on every compact of E.*

We obtain therefore :

Corollary 2.2.2. *If V is a Hilbert space of vector fields continuously embedded into $C_0^2(E,E)$ then, for any fixed and bounded rectifiable set X of finite volume and measurable signal f on X, the application $v \mapsto C_{(\phi_1^v(X), f\circ(\phi_1^v)^{-1})}$ is weakly continuous from L_V^2 to W'.*

Proof. Let v_n is a sequence that weakly converges toward v in L_V^2. With lemma 2.2.4, we know that $\phi_1^{v_n}$ and $d\phi_1^{v_n}$ converge respectively to ϕ_1^v and $d\phi_1^v$ uniformly on every compact of E, i.e $\|\phi_1^{v_n} - \phi_1^v\|_{1,\infty} \to 0$ on a neighborhood of X. Then, thanks to theorem 2.2.2, we are allowed to conclude that $C_{(\phi_1^{v_n}(X), f\circ(\phi_1^{v_n})^{-1})} \xrightarrow{W'} C_{(\phi_1^v(X), f\circ(\phi_1^v)^{-1})}$, which proves the weak continuity. □

To conclude on this topic, we give a little more general result that encompasses theorems 2.2.1 and corollary 2.2.1 by controlling fcurrents' norms both with respect to variations of the support and the signal values. We consider the general action of applications $\psi(x,m) = (\psi^g(x), \psi^f(x,m))$ presented at the end of 2.1.2 and in eq.(2.5). As we did previously, we can generate such transformations by flowing vector fields of the form $(v^g(x), v^f(x,m))$. We assume that $v^g \in \chi^1$ and that v^f is a time-varying function such that for all $t \in [0,1]$, v_t^f is a locally Lipschitz continuous and bounded function from $E \times M$ to TM and that $\int_0^1 \|v_t^f\|_\infty dt < \infty$. We shall denote by

$$\|v\|_{\chi^0} \doteq \int_0^1 \|v_t\|_\infty$$

Then, the following holds :

Theorem 2.2.3. *Let X be a d-dimensional oriented rectifiable subset of E of finite volume and $f : X \to M$ a measurable function. Assume that W_g is continuously embedded into $C_0^1(E, \Lambda^d E^*)$ and that W_f is continuously embedded into $C_0^1(M, \mathbb{R})$. There exists a constant $\gamma > 0$ depending only on the kernels K_g and k_f such that, for all $v = (v^g, v^f)$, with v^g a χ^1 vector field of E and v^f a locally Lipschitz bounded vector field from $E \times M$ to TM, we have :*

$$\|C_{(X,f)} - C_{\psi.(X,f)}\|_{W'} \leq \gamma\, Vol(X) \left(\|v^g\|_{\chi^1} + \|v^f\|_{\chi^0} \right)$$

Proof. The proof essentially combines the two previous theorems. We remind that $\psi.(X,f) = (\psi^g(X), \psi^f((\psi^g)^{-1}), f \circ (\psi^g)^{-1})$. To lighten the expressions, we shall denote by $\psi^f.f$ the function defined on X by

$$(\psi^f.f)(x) \doteq \psi^f(x, f(x))$$

We have, by triangular inequality :

$$\begin{aligned} \|C_{(X,f)} - C_{\psi.(X,f)}\|_{W'} &\leq \|C_{(X,f)} - C_{(X,\psi^f.f)}\|_{W'} + \|C_{(X,\psi^f.f)} - C_{\psi.(X,f)}\|_{W'} \\ &\leq \|C_{(X,f)} - C_{(X,\psi^f.f)}\|_{W'} + \|C_{(X,\psi^f.f)} - C_{(\psi^g(X), \psi^f.f\circ(\psi^g)^{-1})}\|_{W'} \end{aligned} \tag{2.24}$$

Using theorem 2.2.2, the second term is dominated by :

$$\|C_{(X,\psi^f.f)} - C_{(\psi^g(X), \psi^f.f\circ(\psi^g)^{-1})}\|_{W'} \leq \text{Cte.Vol}(X) \|v^g\|_{\chi^1} \tag{2.25}$$

For the first term, applying theorem 2.2.1, we find that :

$$\|C_{(X,f)} - C_{(X,\psi^f.f)}\|_{W'} \leq \text{Cte} \int_X d_M((\psi^f.f)(x), f(x)) d\mathcal{H}^d(x)\,.$$

Using now usual result on the flow of the vector field $v = (v^g, v^f)$, we can bound, for all $(x,m) \in E \times M$, $d_M(\psi^f(x,m),m)$ by $\|(v^g,v^f)\|_{\chi^0} \leq \|v^g\|_{\chi^0} + \|v^f\|_{\chi^0}$ and thus we have :

$$\|C_{(X,f)} - C_{(X,\psi^f.f)}\|_{W'} \leq \text{Cte.Vol}(X).\left(\|v^g\|_{\chi^0} + \|v^f\|_{\chi^0}\right) \tag{2.26}$$

Since $\|v^g\|_{\chi^0} \leq \|v^g\|_{\chi^1}$, eq.(2.25) and (2.26) give the desired result.

□

Variations' formula for the metrics

We now go one step further and give a more precise description of the infinitesimal variation of the W'-metric with respect to the geometrical positions and signal values of the functional shapes. This shall lead to a generalization of Cartan's formula that we stated for usual currents in theorem 1.3.2 of section 1.3.4. In all this section, we will assume that M is a vector space and that *all differential forms have in addition C^1 regularity*. Similarly to 1.3.4, fixing a functional shape (X,f), we want to express the variation of $C_{(X,f)}(\omega)$ for any $\omega \in \Omega_0^d(E,M)$, with respect to X as well as the signal values f. Therefore, we introduce some smooth and compactly supported vector field v in $E \times M$ which, as previously, is such that $v(x,m) = (v_g(x), v_f(x,m))$ and its flow at time t, $\psi_t(x,m) = (\psi^g(x), \psi^f(x,m))$. What we look for is the variation of $C_{(X,f)}(\omega)$ when deforming (X,f) in the direction of v, i.e :

$$\left.\frac{d}{dt}\right|_{t=0} C_{\psi_t.(X,f)}(\omega) = \left.\frac{d}{dt}\right|_{t=0} C_{(X,f)}(\psi_t^*\omega)$$

where $\psi_t^*\omega$ is defined as in eq.(2.13). The simplest way to proceed is to use Cartan's formula for currents in $E \times M$ and then deduce an equivalent result for functional currents through the applications introduced in 2.2.1. Let's call T_G the current in $\Omega_0^d(E \times M)'$ associated to the graph G of (X,f) in $E \times M$. We assume that X is a C^1 compact oriented submanifold and that f is C^1 so that G is itself a compact oriented submanifold in $E \times M$. For any differential form $\eta \in \Omega_0^d(E \times M)$, we have $T_{\psi_t(G)}(\eta) = T_G(\psi_t^\sharp \eta) = \int_G \psi_t^\sharp \eta$ and by Cartan's formula :

$$\left.\frac{d}{dt}\right|_{t=0} T_{\psi_t(G)}(\eta) = \int_G \imath_v(d\eta) + \int_{\partial G} \imath_v(\eta) \tag{2.27}$$

where $\imath_v$ is once again the contraction operator on differential forms (in $E \times M$). To get back to functional forms and currents, we recall from 2.2.1 that we have an injective linear application $p :\ \Omega_0^d(E,M) \to \Omega_0^d(E \times M)$ such that $p(\omega)_{(x,m)}(\zeta) = \omega_{(x,m)}(l(\zeta))$ with l the projection of $\Lambda^d(E \times M)$ on $\Lambda^d(E)$ defined in 2.2.1.

Lemma 2.2.5. *With the previous notations, we have :*

$$C_{(X,f)}(\omega) = T_G(p(\omega)) \tag{2.28}$$

Proof. Up to using a partition of unity of X, we can assume, without loss of generality, that X admits a C^1-parametrization $\gamma : U \to E$ with U an open subset of $\mathbb{R}^d$. Then $\Gamma : U \to E \times M$ defined by $\Gamma(u) = (\gamma(u), f \circ \gamma(u))$ is a parametrization of G and :

$$\begin{aligned} T_G(p(\omega)) &= \int_U p(\omega)_{(\gamma(u), f\circ\gamma(u))} \left(\frac{\partial \Gamma}{\partial u_1} \wedge ... \wedge \frac{\partial \Gamma}{\partial u_d} \right) du_1 ... du_d \\ &= \int_U \omega_{(\gamma(u), f\circ\gamma(u))} \left(l \left(\frac{\partial \Gamma}{\partial u_1} \wedge ... \wedge \frac{\partial \Gamma}{\partial u_d} \right) \right) du_1 ... du_d \end{aligned}$$

By definition of l and Γ, we see that $l \left(\frac{\partial \Gamma}{\partial u_1} \wedge ... \wedge \frac{\partial \Gamma}{\partial u_d} \right) = \frac{\partial \gamma}{\partial u_1} \wedge ... \wedge \frac{\partial \gamma}{\partial u_d}$. And thus :

$$\begin{aligned} T_G(p(\omega)) &= \int_U \omega_{(\gamma(u), f\circ\gamma(u))} \left(\frac{\partial \gamma}{\partial u_1} \wedge ... \wedge \frac{\partial \gamma}{\partial u_d} \right) du_1 ... du_d \\ &= C_{(X,f)}(\omega) \end{aligned}$$

□

We have in addition the following commutativity property :

Lemma 2.2.6. *For any $\omega \in \Omega_0^d(E, M)$,*

$$p(\psi_t^* \omega) = \psi_t^\sharp(p(\omega)) \tag{2.29}$$

Proof. For any $(x, m) \in E \times M$ and $\zeta = \zeta_1 \wedge ... \wedge \zeta_d \in \Lambda^d(E \times M)$, we let $\xi_1 \wedge ... \wedge \xi_d = l(\zeta_1 \wedge ... \wedge \zeta_d)$. Then :

$$\begin{aligned} p(\psi_t^* \omega)_{(x,m)}(\zeta) &= (\psi_t^* \omega)_{(x,m)}(\xi_1 \wedge ... \wedge \xi_d) \\ &= \omega_{\psi_t(x,m)}(\partial_x \psi_t^g(\xi_1) \wedge ... \wedge \partial_x \psi_t^g(\xi_d)) \end{aligned}$$

Now, since $\partial_m \psi_t^g = 0$, we have in fact $\forall i \in \{1, .., d\}$, $\partial_x \psi_t^g(\xi_i) = d\psi_t^g(\zeta_i)$ and $\partial_x \psi_t^g(\xi_1) \wedge ... \wedge \partial_x \psi_t^g(\xi_d) = l(d\psi_t(\zeta_1) \wedge ... \wedge d\psi_t(\zeta_d))$, which gives :

$$\begin{aligned} p(\psi_t^* \omega)_{(x,m)}(\zeta) &= \omega_{\psi_t(x,m)}(l(d\psi_t(\zeta_1) \wedge ... \wedge d\psi_t(\zeta_d))) \\ &= \left(\psi_t^\sharp p(\omega) \right)_{(x,m)}(\zeta) \end{aligned}$$

□

With eq.(2.27),(2.28) and (2.29), we now deduce :

$$\begin{aligned} \left. \frac{d}{dt} \right|_{t=0} C_{\psi_t.(X,f)}(\omega) &= \left. \frac{d}{dt} \right|_{t=0} C_{(X,f)}(\psi_t^* \omega) \\ &= \left. \frac{d}{dt} \right|_{t=0} T_G(p(\psi_t^* \omega)) \\ &= \left. \frac{d}{dt} \right|_{t=0} T_G(\psi_t^\sharp p(\omega)) \\ &= \int_G \imath_v(d[p(\omega)]) + \int_{\partial G} \imath_v(p(\omega)) \end{aligned}$$

We have thus shown the following proposition :

Proposition 2.2.3. *For X a compact oriented submanifold of E, f a C^1 function on X and for any vector field of the form $v(x,m) = (v_g(x), v_f(x,m))$ whose flow ψ_t exists at time $0 < t < \epsilon$, we have for all C^1 functional form $\omega \in \Omega_0^d(E,M)$:*

$$\left.\frac{d}{dt}\right|_{t=0} C_{\psi_t.(X,f)}(\omega) = \int_G \imath_v(d[p(\omega)]) + \int_{\partial G} \imath_v(p(\omega)) \tag{2.30}$$

We see that the derivatives we look for eventually consist in taking the variation of the product current T_G associated to the graph but restricted to particular differential forms $p(\omega)$ of $E \times M$. We already can draw a few consequences from this, relying on what was previously done with Cartan's formula for currents (theorem 1.3.2). In that case, it shows that in $E \times M$, the gradient of fcurrents' metrics at each point in the interior of X is **orthogonal to the graph** of (X, f). It also involves a term on the boundary of G where points have gradients that are orthogonal to this boundary in $E \times M$. If f is a C^1 function, ∂G is simply the set of $\{(x, f(x))\}$ for $x \in \partial X$ but we see that this could be generalized to the case where for instance f is piecewise C^1, in which case some new boundaries appear on the zones of discontinuities of the signal.

The previous result is however intrinsically written in the context of currents in $E \times M$. Starting from eq.(2.30), we can actually go a little further by reinterpreting the previous terms on the graph G as integrals on X and separate the variations induced by the field v_g from the one induced by v_f. For any coordinate systems $(x_1, ..., x_n)$ and $(m_1, ..., m_p)$ of E and M, a functional form can be written :

$$\omega \doteq \sum_{1 \leq i_1 < ... < i_d \leq n} a_{i_1,...,i_d}(x,m) dx^{i_1} \wedge ... \wedge dx^{i_d}$$

and then $p(\omega)$ has the exact same expression in the coordinate system $(x_1, ..., x_n, m_1, ..., m_p)$ of $E \times M$. We shall denote in short $\omega = a_I(x,m)dx^I$ for the sum on all multi-indices in a given coordinate system. Now, since we have assumed that the function f is C^1 : $\partial G = \{(x, f(x)) \in E \times M / \ x \in \partial X\}$ which is the graph of the $(d-1)$-dimensional functional shape $(\partial X, f_{|\partial X})$. If $\eta : \ O \to E$ is a parametrization of ∂X (O being an open subset of $\mathbb{R}^{d-1}$), $\tilde{\eta} : \ u \mapsto (\eta(u), f \circ \eta(u)) \in E \times M$ is a parametrization of ∂G and :

$$\begin{aligned}
\int_{\partial G} \imath_v(p(\omega)) &= \int_O p(\omega)_{(\eta(u), f\circ\eta(u))}(v(\eta(u), f \circ \eta(u)) \wedge \frac{\partial \tilde{\eta}}{\partial u_1} ... \wedge \frac{\partial \tilde{\eta}}{\partial u_{d-1}}) \\
&= \int_O \omega_{(\eta(u), f\circ\eta(u))}(v_g(\eta(u)) \wedge \frac{\partial \eta}{\partial u_1} ... \wedge \frac{\partial \eta}{\partial u_{d-1}}) \\
&= \int_{\partial X} \imath_{v_g} \omega
\end{aligned}$$

The case of the integral on G is a little more difficult. Let's consider some local coordinate systems $(x_1, ..., x_n)$ of E such that $(dx^1, .., dx^d)$ is a basis of $T_x X^*$ and $(dx^{d+1}, .., dx^n)$ is a basis of $(T_x X^{\perp})^*$. In the same spirit, we will denote by $v_g^{\top}(x)$ the component of v_g tangent to X at x and $v_g^{\perp}(x)$ the normal component. We then decompose the vector field as

$$v = (v_g^{\perp}, 0) + (v_g^{\top}, 0) + (0, v_f)$$

By linearity of the contraction operator, we have :

$$\int_G \imath_v dp(\omega) = \int_G \imath_{(v_g^{\perp}, 0)} dp(\omega) + \int_G \imath_{(v_g^{\top}, 0)} dp(\omega) + \int_G \imath_{(0, v_f)} dp(\omega)\,. \tag{2.31}$$

Now, in the previous coordinate system, we can write the exterior derivative of $p(\omega)$ as a differential form of $E \times M$:

$$\begin{aligned} d[p(\omega)] &= \sum_{j=1}^{n} \frac{\partial a_I}{\partial x_j} dx^j \wedge dx^I + \sum_{l=1}^{p} \frac{\partial a_I}{\partial m_l} dm^l \wedge dx^I \\ &= d_x\omega + d_m\omega \end{aligned}$$

and introducing a parametrization of X, $\gamma : \ U \to E$ and $\Gamma : \ u \mapsto (\gamma(u), f \circ \gamma(u))$ the corresponding parametrization of G, we have, for the first integral of eq.(2.31) :

$$\int_G \imath_{(v_g^\perp, 0)} dp(\omega) = \int_U (d_x\omega + d_m\omega) \left((v_g^\perp(\gamma(u)), 0) \wedge \frac{\partial \Gamma}{\partial u_1} \wedge ... \wedge \frac{\partial \Gamma}{\partial u_d} \right) du$$

with

$$\frac{\partial \Gamma}{\partial u_i} = \left(\frac{\partial \gamma}{\partial u_i}, df \circ \frac{\partial \gamma}{\partial u_i} \right) . \tag{2.32}$$

We recall, by definition of exterior product of differential forms, that for all $l \in \{1, .., p\}$ and $u \in U$:

$$\begin{aligned} & dm^l \wedge dx^I \left((v_g^\perp(\gamma(u)), 0) \wedge \frac{\partial \Gamma}{\partial u_1} \wedge ... \wedge \frac{\partial \Gamma}{\partial u_d} \right) \\ &= dm^l(v_g^\perp(\gamma(u)), 0) dx^I \left(\frac{\partial \Gamma}{\partial u_1} \wedge ... \wedge \frac{\partial \Gamma}{\partial u_d} \right) \\ &+ \sum_{i=1}^{d} (-1)^i dm^l \left(\frac{\partial \Gamma}{\partial u_i} \right) dx^I \left(v_g^\perp \wedge \frac{\partial \Gamma}{\partial u_1} \wedge .. \widehat{\frac{\partial \Gamma}{\partial u_i}} .. \wedge \frac{\partial \Gamma}{\partial u_d} \right) \end{aligned}$$

The first term equals zero because $(v_g^\perp(\gamma(u)), 0)$ has only zero components in M. The second one is also null since all differential forms dx^I in our case have only components dx^i in $T_x X^*$ whereas $v_g^\perp(x) \in T_x X^\perp$ for all $x \in X$. Therefore :

$$\begin{aligned} \int_G \imath_{(v_g^\perp, 0)} dp(\omega) &= \int_U d_x\omega \left((v_g^\perp(\gamma(u)), 0) \wedge \frac{\partial \Gamma}{\partial u_1} \wedge ... \wedge \frac{\partial \Gamma}{\partial u_d} \right) du \\ &= \int_U d_x\omega \left(v_g^\perp(\gamma(u)) \wedge \frac{\partial \gamma}{\partial u_1} \wedge ... \wedge \frac{\partial \gamma}{\partial u_d} \right) du \\ &= \int_X \imath_{v_g^\perp} d_x\omega \end{aligned} \tag{2.33}$$

the second equality resulting from eq.(2.32) and the fact the $(d+1)$-differential form $d_x\omega$ has no component with respect to the functional coordinates $dm^1, .., dm^p$.
Now for the second integral of eq.(2.31), we write :

$$\int_G \imath_{(v_g^\top, 0)} dp(\omega) = \int_U (d_x\omega + d_m\omega) \left((v_g^\top, 0) \wedge \frac{\partial \Gamma}{\partial u_1} \wedge ... \wedge \frac{\partial \Gamma}{\partial u_d} \right) du \, .$$

On the first hand, we have :

$$\begin{aligned} d_x\omega \left((v_g^\top, 0) \wedge \frac{\partial \Gamma}{\partial u_1} \wedge ... \wedge \frac{\partial \Gamma}{\partial u_d} \right) &= \sum_{j=1}^{n} \frac{\partial a_I}{\partial x_j} dx^j \wedge dx^I \left((v_g^\top, 0) \wedge \frac{\partial \Gamma}{\partial u_1} \wedge ... \wedge \frac{\partial \Gamma}{\partial u_d} \right) \\ &= \sum_{j=1}^{n} \frac{\partial a_I}{\partial x_j} dx^j \wedge dx^I \left(v_g^\top \wedge \frac{\partial \gamma}{\partial u_1} \wedge ... \wedge \frac{\partial \gamma}{\partial u_d} \right) \end{aligned}$$

and since $\frac{\partial\gamma}{\partial u_1}, ..., \frac{\partial\gamma}{\partial u_d}$ is a basis of the tangent space $T_{\gamma(u)}X$, $v_g^\top$ is linearly dependent to these vectors. Thus $v_g^\top \wedge \frac{\partial\gamma}{\partial u_1} \wedge ... \wedge \frac{\partial\gamma}{\partial u_d} = 0$ for all u and

$$\int_G \imath_{(v_g^\top,0)} d_x\omega = 0\,.$$

On the other hand,

$$d_m\omega\left((v_g^\top,0) \wedge \frac{\partial\Gamma}{\partial u_1} \wedge ... \wedge \frac{\partial\Gamma}{\partial u_d}\right) = \sum_{l=1}^{p} \frac{\partial a_I}{\partial m_l} \underbrace{dm^l \wedge dx^I\left((v_g^\top,0) \wedge \frac{\partial\Gamma}{\partial u_1} \wedge ... \wedge \frac{\partial\Gamma}{\partial u_d}\right)}_{\Delta_l}\,.$$

Developing Δ_l gives :

$$\begin{aligned}
\Delta_l &= dm^l(v_g^\top,0)dx^I\left(\frac{\partial\Gamma}{\partial u_1} \wedge ... \wedge \frac{\partial\Gamma}{\partial u_d}\right) \\
&\quad + \sum_{i=1}^{d}(-1)^i dm^l\left(\frac{\partial\Gamma}{\partial u_i}\right) dx^I\left((v_g^\top,0) \wedge \frac{\partial\Gamma}{\partial u_1} \wedge ... \wedge \widehat{\frac{\partial\Gamma}{\partial u_i}} \wedge ... \wedge \frac{\partial\Gamma}{\partial u_d}\right) \\
&= \sum_{i=1}^{d}(-1)^i dm^l\left(\frac{\partial\Gamma}{\partial u_i}\right) dx^I\left((v_g^\top,0) \wedge \frac{\partial\Gamma}{\partial u_1} \wedge ... \wedge \widehat{\frac{\partial\Gamma}{\partial u_i}} \wedge ... \wedge \frac{\partial\Gamma}{\partial u_d}\right) \\
&= \sum_{i=1}^{d}(-1)^i df_l\left(\frac{\partial\gamma}{\partial u_i}\right) dx^I\left(v_g^\top \wedge \frac{\partial\gamma}{\partial u_1} \wedge ... \wedge \widehat{\frac{\partial\gamma}{\partial u_i}} \wedge ... \wedge \frac{\partial\gamma}{\partial u_d}\right)
\end{aligned}$$

One can write df_l in the same coordinate system as $df_l = \sum_{s=1}^{d} \frac{\partial f_l}{\partial x_s} dx^s$, which gives :

$$\begin{aligned}
\Delta_l &= \sum_{s=1}^{d} \frac{\partial f_l}{\partial x_s} \sum_{i=1}^{d}(-1)^i dx^s\left(\frac{\partial\gamma}{\partial u_i}\right) dx^I\left(v_g^\top \wedge \frac{\partial\gamma}{\partial u_1} \wedge ... \wedge \widehat{\frac{\partial\gamma}{\partial u_i}} \wedge ... \wedge \frac{\partial\gamma}{\partial u_d}\right) \\
&= \sum_{s=1}^{d} \frac{\partial f_l}{\partial x_s} dx^s \wedge dx^I\left(v_g^\top \wedge \frac{\partial\gamma}{\partial u_1} \wedge ... \wedge \frac{\partial\gamma}{\partial u_d}\right) - \sum_{s=1}^{d} \frac{\partial f_l}{\partial x_s} dx^s(v_g^\top) dx^I\left(\frac{\partial\gamma}{\partial u_1} \wedge ... \wedge \frac{\partial\gamma}{\partial u_d}\right) \\
&= -df_l(v_g^\top) dx^I\left(\frac{\partial\gamma}{\partial u_1} \wedge ... \wedge \frac{\partial\gamma}{\partial u_d}\right)\,.
\end{aligned}$$

Indeed $v_g^\top \wedge \frac{\partial\gamma}{\partial u_1} \wedge ... \wedge \frac{\partial\gamma}{\partial u_d} = 0$ since $v_g^\top \in T_{\gamma(u)}X$ and $\frac{\partial\gamma}{\partial u_1}, ..., \frac{\partial\gamma}{\partial u_d}$ is a basis of $T_{\gamma(u)}X$. Eventually, we obtain that :

$$\begin{aligned}
\int_G \imath_{(v_g^\top,0)} dp(\omega) &= \int_U -\sum_{l=1}^{n} \frac{\partial a_I}{\partial m_l} df_l(v_g^\top) dx^I\left(\frac{\partial\gamma}{\partial u_1} \wedge ... \wedge \frac{\partial\gamma}{\partial u_d}\right) \\
&= \int_U -\sum_{l=1}^{n} \frac{\partial a_I}{\partial m_l} dm^l(df(v_g^\top)) dx^I\left(\frac{\partial\gamma}{\partial u_1} \wedge ... \wedge \frac{\partial\gamma}{\partial u_d}\right) \\
&= \int_U -\sum_{l=1}^{n} \frac{\partial a_I}{\partial m_l} dm^l \wedge dx^I\left(df(v_g^\top) \wedge \frac{\partial\gamma}{\partial u_1} \wedge ... \wedge \frac{\partial\gamma}{\partial u_d}\right) \\
&= \int_X -\imath_{df(v_g^\top)} d_m\omega \qquad (2.34)
\end{aligned}$$

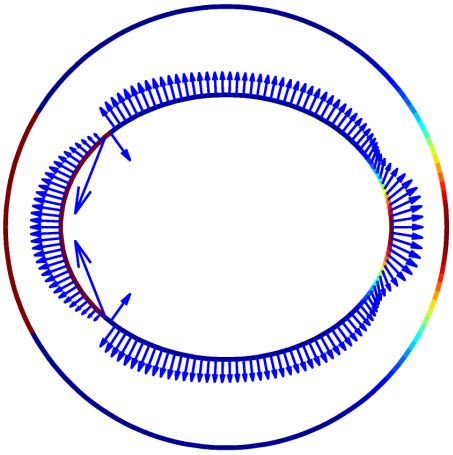


Figure 2.9: Gradient of the distance between two functional shapes. We represent the opposite of the gradient with respect to point positions. The gradient is orthogonal to the curve on the areas of constant signals but not in zones where signal is varying. Note also the particular behavior on boundaries created by discontinuities of signal.

Finally, for the last integral of eq.(2.31), with the same kind of reasoning, we obtain easily that :

$$\int_G \imath_{(0,v_f)} dp(\omega) = \int_U d_m\omega \left((0, v_f) \wedge \frac{\partial \Gamma}{\partial u_1} \wedge ... \wedge \frac{\partial \Gamma}{\partial u_d} \right) du$$

and just as previously we can write for all $u \in U$:

$$\begin{aligned}
& d_m\omega \left((0, v_f) \wedge \frac{\partial \Gamma}{\partial u_1} \wedge ... \wedge \frac{\partial \Gamma}{\partial u_d} \right) \\
&= \sum_{l=1}^{p} \frac{\partial a_I}{\partial m_l} dm^l \wedge dx^I \left((0, v_f) \wedge \frac{\partial \Gamma}{\partial u_1} \wedge ... \wedge \frac{\partial \Gamma}{\partial u_d} \right) \\
&= \sum_{l=1}^{p} \frac{\partial a_I}{\partial m_l} dm^l(v_f).dx^I \left(\frac{\partial \Gamma}{\partial u_1} \wedge ... \wedge \frac{\partial \Gamma}{\partial u_d} \right) \\
&= \sum_{l=1}^{p} \frac{\partial a_I}{\partial m_l} dm^l(v_f).dx^I \left(\frac{\partial \gamma}{\partial u_1} \wedge ... \wedge \frac{\partial \gamma}{\partial u_d} \right) \\
&= \sum_{l=1}^{p} \frac{\partial a_I}{\partial m_l} dm^l \wedge dx^I \left((0, v_f) \wedge \frac{\partial \gamma}{\partial u_1} \wedge ... \wedge \frac{\partial \gamma}{\partial u_d} \right) .
\end{aligned}$$

The second and the last equality result again from the fact that $dx^i(0, v_f) = 0$ for all $i \in \{1, .., n\}$. Consequently :

$$\int_G \imath_{(0,v_f)} dp(\omega) = \int_X \imath_{v_f} d_m\omega \,. \tag{2.35}$$

Grouping eq.(2.33) (2.34) and (2.35) together, we have eventually shown the following variation's result for the functional currents' metrics :

Theorem 2.2.4. *For any functional shape (X, f) with X a compact oriented submanifold of E and f a C^1 function, and for any vector field of the form $v(x, m) =$*

$(v_g(x), v_f(x,m))$ whose flow ψ_t exists at time $0 < t < \epsilon$, we have for all C^1 functional d-form ω :

$$\left.\frac{d}{dt}\right|_{t=0} C_{\psi_t.(X,f)}(\omega) = \int_X \left(\imath_{v_g^\perp} d_x\omega + \imath_{(v_f - df(v_g^\top))} d_m\omega\right) + \int_{\partial X} \imath_{v^g}\omega$$

For the correct understanding of the previous formula, let's remind that we have adopted the shortcut notations $\imath_u d_x\omega$ for $\imath_{(u,0)} d_x\omega$ (when u is a vector field with values in E) and $\imath_u d_m\omega$ for $\imath_{(0,u)} d_m\omega$ (when u is a vector field with values in M). This last result gives the variation based only on integrals on X and ∂X of several differential forms obtained by differentiating ω. It provides a much more precise description of the gradient of the metrics than proposition 2.2.3. Indeed, we first see that the variation with respect to the vector field v_g involves both its orthogonal and tangential components : thus the gradient of $C_{(X,f)}(\omega)$ with respect to displacements of X is not necessarily orthogonal to X as for usual currents. However, we see that the only term involving $v_g^\top$ actually depends only in $df(v_g^\top)$. In particular, it vanishes whenever $\nabla f = 0$ which implies that **the gradient of $C_{(X,f)}(\omega)$ will be orthogonal to X in regions where f is constant**. This is illustrated by figure 2.9. A second important element is the term on the boundary of X that only affects the variation with respect to the geometry since it only takes into account the component v_g. On points on the boundary, **the gradient is orthogonal to the boundary** but not necessarily to X. The presence of integrals on ∂X in the variation is of crucial importance numerically when computing gradients of fcurrents distances because they induce singularities on the boundary, as we study in more details in 4.2.1. As we mentioned earlier after proposition 2.2.3, we could generalize theorem 2.2.4 to piecewise C^1 functions f, in which case additional boundaries are created by discontinuities of the signal, inducing again gradient singularities at those points as one can see on figure 2.9.

2.3 Algorithms based on functional currents

We now illustrate how the theoretical concept of functional currents introduced above combined with kernel metrics offers a genuine solution to the simultaneous processing of the geometric and signal information. The algorithms we present will focus exclusively on real-valued signals although, as we mentioned, it could be adapted to different types of signal spaces. We have explained, from the theoretical point of view, how Hilbert norms based on kernels share interesting properties with respect to geometrical and functional perturbations. In addition, the Hilbert space structure opens the way for various processing algorithms. In this section, we intend to shed light on the potential of the proposed framework. The first application illustrates the full potential of the Hilbert structure with the design of redundancy reduction or compression algorithms for functional shape representations through matching pursuit schemes on functional currents. The second one, more directly related to computational anatomy's issues exposed in the beginning of the chapter, is the design of a large deformation matching algorithm for the simultaneous geometric and functional registration of functional shapes under diffeomorphic transport. We postpone applications to medical imaging datasets to chapter 4, which will generalize the framework to atlas estimation procedure for functional shapes.

2.3.1 Compression with Matching Pursuit on fcurrents

Let us start with the issue of the redundancy of fcurrent representations. The discretization of a functional shape in punctual fcurrents given by eq.(2.9) will provide a representation with a number of elements that corresponds to the initial sampling of the shape. For a surface, this corresponds to the number of triangles in the mesh. Generally, this representation could be clearly reduced since, for simple functional shapes with low-frequency geometries and signals, only a few terms should capture *most of* the shape. However, the quality of the approximation needs to be quantified in a meaningful way, especially when the functional part is also involved, through an appropriate norm for which we have a natural candidate given by the kernel Hilbert structure. This issue of redundancy reduction or compression is important for instance when making crude means of currents because without further treatment, the number of Dirac currents involved in the representation of the mean would increase dramatically. Yet, as we shall see in chapter 4, with more recent atlas estimation procedures, compression is far less decisive in practice. However, when considering higher order statistics for the analysis of variability around a mean functional shape through computations of W'-inner products in LDA or SVM for instance, having reduced first the number of Diracs in the representations of the data can lead to substantial gain in computational time. What we propose here is one way to address the problem by performing **compression in the space of functional currents** : it is important to keep in mind that this is not equivalent to more classical mesh reduction or vertices's decimation for instance because our output is a general fcurrent (a set of Diracs) that does not have an underlying shape structure. This loss of structure allows in compensation high compression rates with an accuracy that is directly related to the metric on fcurrents that we use.

As one can expect, the problem of redundancy reduction or compression is deeply simplified thanks to the Hilbert space structure that has been defined on functional currents. Indeed, classical matching-pursuit algorithms in general Hilbert spaces have already been studied by Mallat and Zhang in [59] and later adapted to currents in [35]. We can proceed in a similar way for functional currents. Considering a discretized fcurrent $C = \sum_{i=1..N} \delta^{\xi_i}_{(x_i,m_i)} \in W'$. N, the number of momenta, is automatically given by the mesh on the submanifold (number of edges for curves, number of triangles for surfaces,...). This submanifold might have some very regular regions with low-frequency geometrical and functional variations, which results in a very redundant representation by fcurrents due to the fact that many adjacent nodes present the same local geometry and signal. The goal of matching-pursuit is to find a more appropriate and reduced representation of C in terms of elementary functional currents. Given a certain threshold $\varepsilon > 0$, we want to find $\Pi_n(C)$ such that $C = \Pi_n(C) + R_n(C)$ and $\|R_n(C)\|_{W'} \leqslant \varepsilon$. $R_n(C)$ will be called the residual of the approximation. Somehow, this is linked to the problem of finding the best projection of C on a n-dimensional subspace of W'. This problem is however too time-consuming computationally for usual applications. Instead, matching pursuit is a greedy algorithm that constructs a family of approximating vectors step by step. The result is a suboptimal fcurrent that approximates the functional current C with a residual whose energy is below threshold. The algorithm basically proceeds as follows. We need to specify a 'dictionary' $\mathcal{D}$ of elements in W'. In our case, we typically consider the set of all elementary functional currents $\{\delta^{\xi}_{(x,m)}\}$ with ξ a unit vector

in $\Lambda^d E$. The first step of matching pursuit algorithm is to find $\delta^{\xi'_1}_{(x'_1,m'_1)} \in \mathcal{D}$ that is best correlated to C. In other words, we try to maximize, with respect to x, m, ξ, the quantity :

$$\langle C, \delta^{\xi}_{(x,m)} \rangle_{W'} = \xi^T \left(\sum_{i=1}^{N} K((x,m),(x_i,m_i))\xi_i \right) . \tag{2.36}$$

Since ξ is taken among unit vectors, the problem is equivalent to maximizing

$$| \sum_{i=1..N} K((x,m),(x_i,m_i))\xi_i | = |\gamma(x,m)|$$

with respect to (x,m) and take ξ as the unit vector in the direction of $\gamma(x,m)$. We get a first approximation of C :

$$C = \Pi_1(C) + R_1(C) .$$

where $\Pi_1(C) = \langle C, \delta^{\xi'_1}_{(x'_1,m'_1)} \rangle_{W'} \delta^{\xi'_1}_{(x'_1,m'_1)}$. The algorithm then applies the same procedure to the residual $R_1(C)$, which provides a second vector $\delta^{\xi'_2}_{(x'_2,m'_2)} \in \mathcal{D}$, and a residual $R_2(C)$. The algorithm is stopped when the RKHS norm of the residual decreases below the given threshold ε.

In most cases, it appears that the compression is better with the orthogonal version of the previous scheme, in which the family of vectors is orthonormalized at each step, in order to force the projection and the residual to be orthogonal in W'. The classical algorithm is based on a Gram-Schmidt orthonormalization at each step. In our case, following the ideas of [35], it is possible to orthogonalize more efficiently by keeping the values of (x'_i, m'_i) found during previous steps and simply modify the vectors ξ'_i. This is done by imposing the following orthogonality condition. Let (e_k) be the canonical basis of the vector space $\Lambda^d E$. If $C = \Pi_n(C) + R_n(C)$ and $\Pi_n(C) = \sum_{i=1..n} \delta^{\alpha^n_i}_{(x'_i,m'_i)}$, we will add the orthogonality constraint :

$$\delta^{e_k}_{(x'_i,m'_i)} \perp R_n(C) \Longleftrightarrow \langle C, \delta^{e_k}_{(x'_i,m'_i)} \rangle_{W^*} = \langle \Pi_n(C), \delta^{e_k}_{(x'_i,m'_i)} \rangle_{W^*}$$

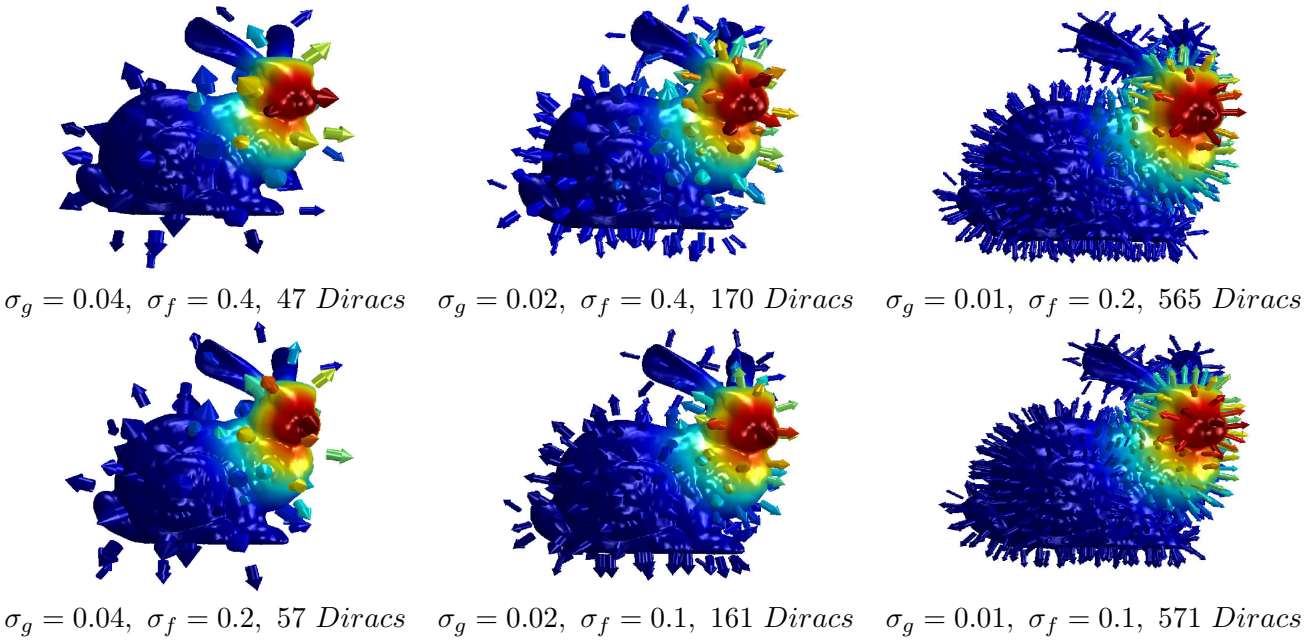
for all basis vectors e_k and for all $i \in \{1,..,n\}$. It is then straightforward to check that these conditions are equivalent to the following system of linear equations to find the α^n_i :

$$\forall i \in \{1,..,n\}, \ \sum_{j=1}^{n} \left(K((x'_i,m'_i),(x'_j,m'_j))\alpha^n_j \right)_k = \gamma(x'_i,m'_i)_k . \tag{2.37}$$

We can show that the norm of the residual $R_n(C)$ monotonically decreases to zero as $n \to \infty$. Hence the algorithm converges and eventually when the residual goes below the given threshold at a certain step n, we obtain a compressed representation of C with n orthogonal dirac fcurrents (with generally $n \ll N$, as we shall see on the coming examples).
Let's give a few comments on the computational complexity of the algorithm, although numerical issues related to fcurrents will be developed in 2.3.3. At each step, we have to solve a linear system of equations eq.(2.37) which is $3n \times 3n$ if the embedding space is $\mathbb{R}^3$: since n is the number of Diracs of the compressed version of C, it remains generally small. Instead, the more technical and time-consuming

part of the algorithm is the optimization of $|\gamma(x,m)|$ with respect to (x,m). Indeed, the optimum need not be one of the (x_i,m_i) and thus it requires to compute $\gamma(x,m)$ for points on a neighborhood in $E\times M$ of the functional shape (because using Gaussian kernels for instance makes γ decreases to zero when (x,m) goes far from the (x_i,f_i)). This reduces to the computation of sums of kernels for each point (x,m), which has a brute force complexity of n times the number of points of the chosen neighborhood. Usually, one chooses a grid of points in $E\times M$ around the functional shape whose steps depend on the scales of the kernels σ_g and σ_f. Such grids can have numerous points especially for small kernels which makes brute force computation nearly impossible in some cases. In [28, 35] was proposed, for usual currents, an alternative computation method, expressing everything as convolutions on the grid of points, which can be computed much more efficiently using FFTs. For functional currents, such methods still suffer both in memory and speed from the higher dimension of the space due to the adjunction of signals. A possible way to handle it is to make alternative maximization between x and m in eq.(2.36), which can be done by using successively purely geometrical and purely functional grids. We refer to following section 2.3.3 for a more complete discussion on the numerical topic.

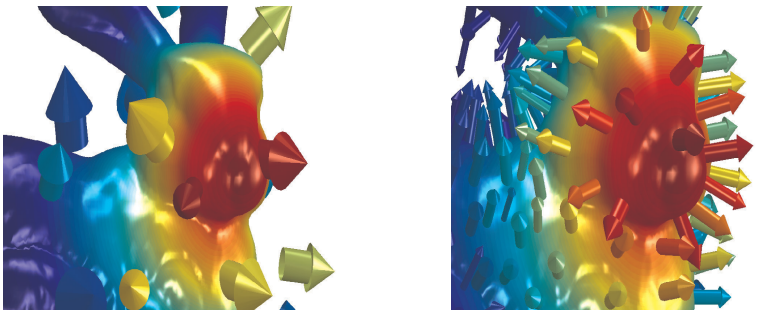
Here are now a few illustrative examples for real valued data on curves or surfaces. We will always consider kernels on fcurrents that are the tensor product of a Gaussian kernel in $\mathbb{R}^3$ of scale parameter σ_g with a real Gaussian kernel in the signal space of scale parameter σ_f. In figure 2.10 and 2.11, we emphasize the influence of both kernel sizes on the compression factor as well as on the precision of the functional values of the compressed shape. The bigger the parameter σ_g, the coarser the scale of representation is and fewer punctual fcurrents are therefore needed to compress shapes but more smaller features are lost. In figure 2.12, we focus more precisely on the compression's behavior when computing matching-pursuit on a simulated fiber bundle of 2D curves carrying different signals. The scale σ_g is the same for both figures but we show the results of matching-pursuit for two radically different values of σ_f. In both cases, matching pursuit provides an accurate approximation of the mean (accordingly to the kernel norm) with a very limited number of Diracs compared to the original sampling. However, note the important influence of σ_f : taking a larger value for this parameter means that the radius of averaging for the signal part is higher.
In conclusion, these first examples of functional shape processing were meant to highlight that the combination of the fcurrent's representation with the use of RKHS metrics provides an easy solution to address the issue of redundancy and compression. The method provides important compression factors and enables *scale analysis* on geometry and signal through the kernel parameters σ_g and σ_f.



$\sigma_g = 0.04,\ \sigma_f = 0.4,\ 47\ Diracs$ $\sigma_g = 0.02,\ \sigma_f = 0.4,\ 170\ Diracs$ $\sigma_g = 0.01,\ \sigma_f = 0.2,\ 565\ Diracs$

$\sigma_g = 0.04,\ \sigma_f = 0.2,\ 57\ Diracs$ $\sigma_g = 0.02,\ \sigma_f = 0.1,\ 161\ Diracs$ $\sigma_g = 0.01,\ \sigma_f = 0.1,\ 571\ Diracs$

Figure 2.10: Matching pursuit on a "painted" bunny with different parameters σ_g and σ_f. Geometrically, the surface has $0.16 \times 0.22 \times 0.12$ extension in the 3D space and the signal goes from value zero (blue) to one (red). The original sampling of the fcurrent representation has 69451 Diracs and we choose a stopping criterion for the algorithm of $\epsilon = 5\%$. The resulting Dirac fcurrents $\delta^{\xi_k}_{(x_k,m_k)}$ are here represented as colored vectors accordingly to the functional values m_k. Vectors are all of same length covering an area proportional to the norm of ξ_k. Notice that the sampling increases as σ_g is smaller while the vector's colors are more accurate when σ_f is smaller.



$\sigma_g = 0.04,\ \sigma_f = 0.4$ $\sigma_g = 0.01,\ \sigma_f = 0.1$

Figure 2.11: Close up on two of the previous results.

2.3.2 Registration of functional shapes

A second class of applications that can be easily derived from the framework of functional currents is the problem of registration between functional shapes. This is a recurrent issue in computational anatomy that is often a preliminary step for atlas estimation algorithms and statistical analysis of the variability (cf chapter 4). The distances resulting from kernel metrics on fcurrents that we have constructed can obviously play the role of a data attachment distances between the objects and be used in basically any inexact matching approach. The one we focus on, in this thesis,

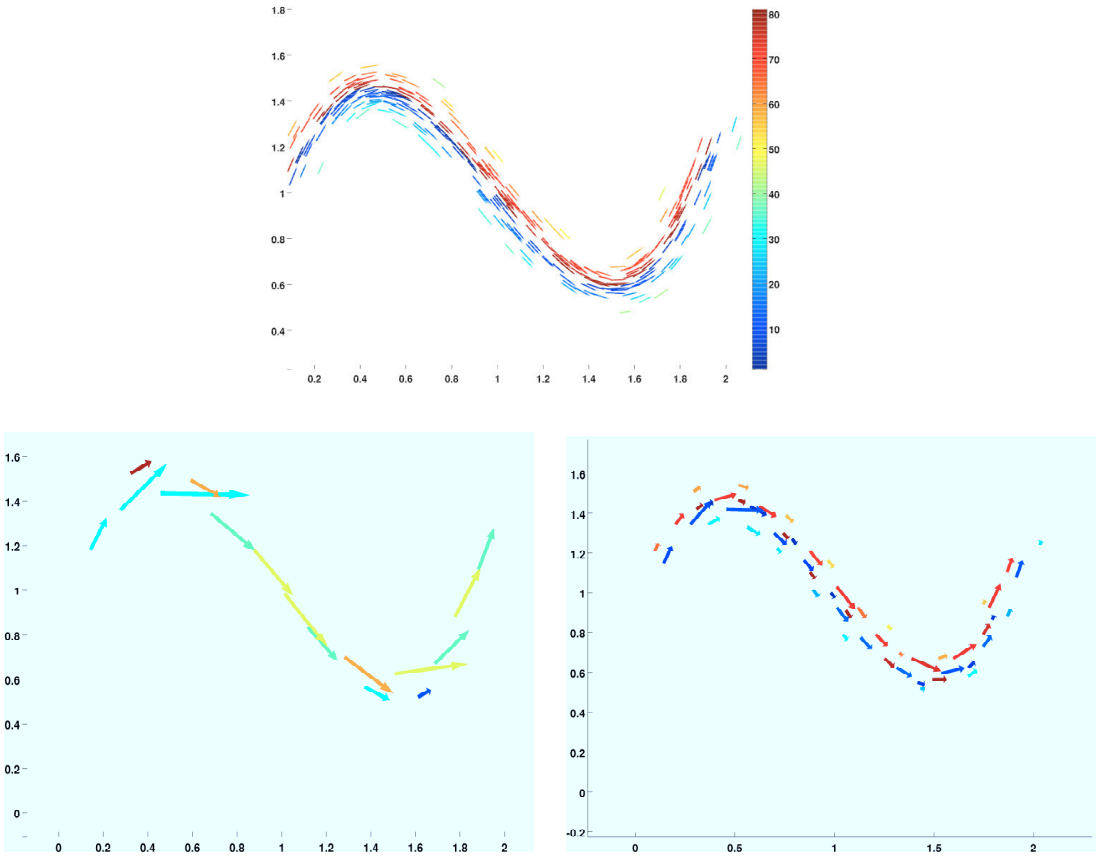


Figure 2.12: Matching pursuit on a 2D fiber bundle, each fiber carrying one value of signal represented by the color. On top, the initial object consisting of 300 fibers. Below, we show two results of matching-pursuit with the same σ_g but two different values for σ_f : $\sigma_f = 200$ for the left figure, $\sigma_f = 20$ for the right one.

is the LDDMM setting presented in the introduction. This will lead to a generalization of the curve and surface registration algorithm based on currents to the wider case of functional shapes. Although, as we mentioned several times, there are no theoretical limitations to apply fcurrents in various cases of signals, the numerical implementations that have been done are restricted to **real-valued signals** in basically two situations : purely diffeomorphic registration of the support (based on the deformation model of eq.(2.1)) and with adjunction of a residual signal (modeled as in eq.(2.2)). The essential practical difference with the previous algorithm on currents is the computation of the new data attachment distances and their gradients. We detail that in the following paragraph.

Distances and gradients in the discrete setting

Let's consider discrete functional shapes : a source (X, f) given by the set of points $(q, f) = (q_i, f_i)_{i=1,..,n}$, $q_i \in E$, $f_i \in M = \mathbb{R}$ and a mesh containing the list of faces (segments for curves, triangles for surfaces) which can be taken, in practice, to be a list of $(d+1)$-oriented simplexes $(\alpha_i^1, .., \alpha_i^{d+1})_{i=1,..,m}$ with α_i^k indexes in $\{1, .., n\}$, a target (Y, g) with points $(y_j, g_j)_{j=1,..,r}$ given also with mesh $(\beta_i^1, .., \beta_i^{d+1})_{i=1,..,s}$. From this, one has first to compute the fcurrent representations $C_{(X,f)} = \sum_{i=1}^m \delta^{\xi_i}_{(\tilde{q}_i, \tilde{f}_i)}$ and

$C_{(Y,g)} = \sum_{j=1}^{s} \delta^{\eta_j}_{(\tilde{y}_j,\tilde{g}_j)}$ of each shape. Indeed, the variables $\tilde{q}_i$, $\tilde{f}_i$ and ξ_i are themselves functions of the points (q_i) and signals (f_i) of (X, f) respectively as the centers, interpolated signal value and normal vector of the ith face. More explicitly, for any $i = 1, .., m$:

$$\begin{cases} \tilde{q}_i = \dfrac{1}{d+1}(q_{\alpha_i^1} + ... + q_{\alpha_i^{d+1}}) \\ \tilde{f}_i = \dfrac{1}{d+1}(f_{\alpha_i^1} + ... + f_{\alpha_i^{d+1}}) \\ \xi_i = \Lambda_{h=1}^{d}(q_{\alpha_i^h} - q_{\alpha_i^{d+1}}) \end{cases} \tag{2.38}$$

and similar equations hold for the target shape (Y, g). Now, the data attachment distance that shall be used in the algorithms is naturally $\|C_{(X,f)} - C_{(Y,g)}\|^2_{W'}$ which writes, as in eq.(2.18) :

$$\begin{aligned} \|C_{(X,f)} - C_{(Y,g)}\|^2_{W'} = &\sum_{i=1}^{m}\sum_{j=1}^{m} k_f(\tilde{f}_i, \tilde{f}_j).\langle K_g(\tilde{q}_i, \tilde{q}_j)\xi_i, \xi_j\rangle \\ &+ \sum_{k=1}^{s}\sum_{l=1}^{s} k_f(\tilde{g}_k, \tilde{g}_l).\langle K_g(\tilde{y}_k, \tilde{y}_l)\eta_k, \eta_l\rangle \\ &- 2\sum_{i=1}^{m}\sum_{k=1}^{s} k_f(\tilde{f}_i, \tilde{g}_k).\langle K_g(\tilde{q}_i, \tilde{y}_k)\xi_i, \eta_k\rangle \end{aligned} \tag{2.39}$$

Considering only the dependency with respect to the source fshape, the previous distance is a function of the variables $\tilde{q}_i, \tilde{f}_i, \xi_i$ that we will call $\tilde{A}$. By composition with eq.(2.38), we see that $\tilde{A}(\tilde{q}, \tilde{f}, \xi)$ is a function of q and f that we will denote $A(q, f)$. For registration algorithm, one also needs the **gradients of** A with respect to q and f. Those can be computed using chain rule differentiation since :

$$\begin{cases} \partial_{q_k} A = \sum\limits_{i=1}^{n} \left(\partial_{\tilde{q}_i}\tilde{A} \circ \partial_{q_k}\tilde{q}_i + \partial_{\xi_i}\tilde{A} \circ \partial_{q_k}\xi_i\right) \\ \partial_{f_k} A = \sum\limits_{j=1}^{n} \partial_{\tilde{f}_i}\tilde{A} \circ \partial_{f_k}\tilde{f}_i \end{cases} \tag{2.40}$$

Now, the gradients $\partial_{\tilde{q}_i}\tilde{A}$, $\partial_{\tilde{f}_i}\tilde{A}$ and $\partial_{\xi_i}\tilde{A}$ are computed easily by differentiating eq.(2.39), while the $\partial_{q_k}\tilde{q}_i$, $\partial_{f_k}\tilde{f}_i$ and $\partial_{q_k}\xi_i$ are obtained by differentiation of eq.(2.38). One notable advantage of expressing the gradients in this way is that it separates what involves the passage to fcurrents' variables from what involves only kernels' computations, allowing to write numerical codes which are very adaptable to various dimensions and to the use of different kernels.

Registration with geometrical deformations

Now that the data attachment distance and its gradients are explicit, the whole LDDMM machinery detailed in chapter 1 can be extended to functional shapes almost straightforwardly. In the first place, we shall be dealing with the situation where we intend to deform functional shapes only with diffeomorphisms of E, which is the model of transport given in eq.(2.1). In that case, the optimization problem in LDDMM writes in general :

$$\min_{v \in L^2_V} J(v) \doteq \int_0^1 \|v_t\|^2_V dt + \gamma \|C_{(\phi_1(X), f\circ(\phi_1)^{-1})} - C_{(Y,g)}\|^2_{W'} . \tag{2.41}$$

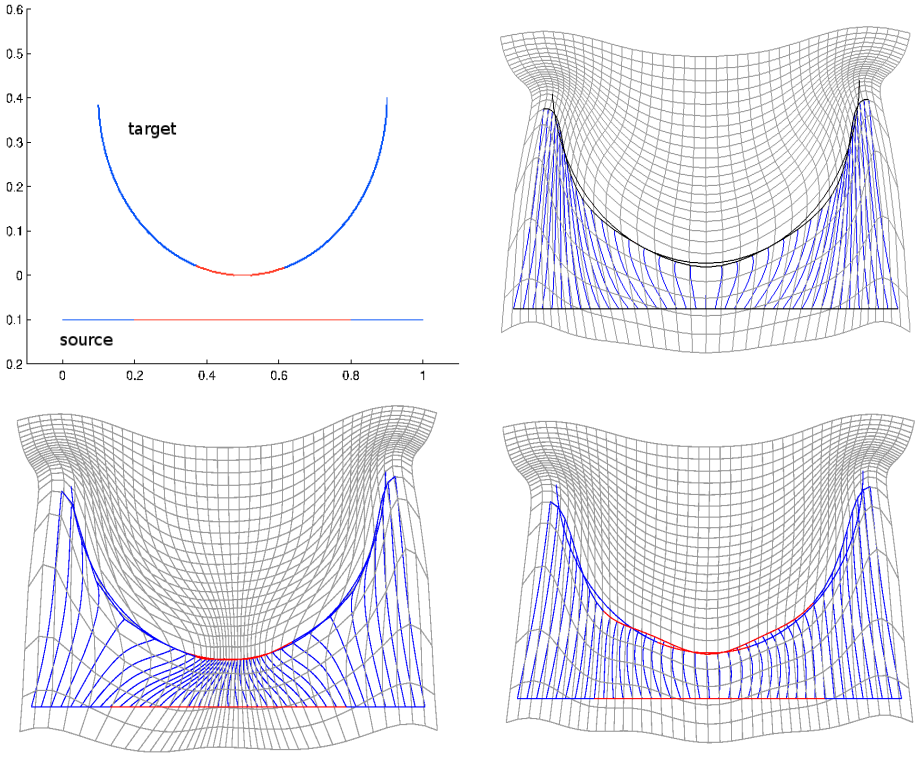


Figure 2.13: Example of registration of two functional curves (top left) with binary signal (blue is zero and red is one). On top right, we show the classical matching with currents on the purely geometrical curves. On bottom left, the same curves are matched with our extension of LDDMM to functional currents. In both cases, the deformed curve fits closely to the target one but note the difference of the deformation field for the functional current's approach. Finally, on the right, we show the result of matching we obtain again with fcurrents' LDDMM but with a big value of σ_f compared to the signal, in which case the matching is nearly similar to the current matching.

Based on the control result of previous sections, one can show existence of solutions to this variational problem :

Proposition 2.3.1. *If W_g is continuously embedded into $C_0^1(E, \Lambda^d E^*)$ and V is continuously embedded into $C_0^2(E, E)$ then, for any functional shapes (X, f) and (Y, g), there exists a minimum to the problem of eq.*(2.41).

Proof. Thanks to the result of proposition 1.3.1 from chapter 1, we only need to prove that the application $v \mapsto \|C_{(\phi_1^v(X), f\circ(\phi_1^v)^{-1})} - C_{(Y,g)}\|^2_{W'} \doteq A(\phi_1^v)$ is weakly continuous on L^2_V. But this is a direct consequence of corollary 2.2.2 from the previous section. □

Now, in the discrete setting, as we presented in 1.3.2, this problem can be rewritten as an optimization over the initial momentum p_0, i.e :

$$\min_{p_0} J(p_0) \doteq E(p_0) + \gamma \|C_{(\phi_1(X), f\circ(\phi_1)^{-1})} - C_{(Y,g)}\|^2_{W'} . \tag{2.42}$$

where $E(p_0)$ is the energy (or equivalently the reduced Hamiltonian) of the deformation which equals $H_r(q, p_0) = p_0^T K_V(q, q) p_0$ if the vector fields are generated by a kernel K_V, ϕ_1 is the deformation at time 1 obtained by flowing forward the geodesic shooting equations and γ is the trade-off parameter.
The dynamics of the deformation part (basically the forward and backward differential equations) remain totally identical to the usual currents' case because in this model of deformation, the transport of functional information is trivial (i.e f_t is constant and equal to f along the geodesic). Note that for more complicated actions like the one on vector fields or tensors, forward and backward equations would also affect the signal part. An adaptive step gradient descent is then done on the momentum $p_0 = (p_{0,i})_{i=1,..,n}$. The gradient with respect to p_0 is obtained by first flowing the forward Hamiltonian equation which gives the point positions q_t and the momentum p_t at all times $t \in [0, 1]$, the signals f_t remaining constantly equals to f. Then, we obtain :

$$\nabla_{p_0} J = K_V(q, q) p_0 - P_0$$

with P_0 an adjoint variable obtained by flowing backward in time the coupled equation :

$$\begin{cases} \dot{Q}_t = -(\partial^2_{q,p} H_r(q_t, p_t))^* Q_t + (\partial^2_{q,q} H_r(q_t, p_t))^* P_t \\ \dot{P}_t = -(\partial^2_{p,p} H_r(q_t, p_t))^* Q_t + (\partial^2_{q,p} H_r(q_t, p_t))^* P_t \\ Q_1 = -\nabla_{q^1} A(q_1, f),\ P_1 = 0 \end{cases} \tag{2.43}$$

where $\nabla_{q_1} A(q_1, f)$ is basically the gradient of the attachment distance function between the deformed source fshape and the target, computed as in 2.3.2.
The whole procedure was coded in MATLAB for curves and surfaces, with some parts in C and GPU, as we will explain in 2.3.3. We present a few results of this algorithm. As we can expect with this model, the resulting matching is driven both by the geometry of the shapes and by the functional values they carry, which we first show on the very simple example of figure 2.13. We also see in that figure that the attachment of the functionals is controlled by the size of the kernel k_f : indeed, at very coarse scale for σ_f, the functional are no more taken into account and the approach is thus identical to usual purely geometrical currents. Now, we draw attention to the important differences that exists when matching functional shapes with functional currents instead of the other possible spaces of currents discussed in 2.1.3. This is particularly easy to highlight with colored currents : if we refer again to

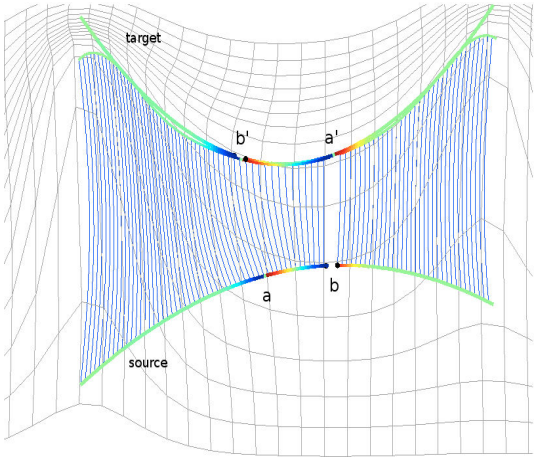


Figure 2.14: Back to the example of figure 2.6, the matching is here performed with the approach of functional currents.

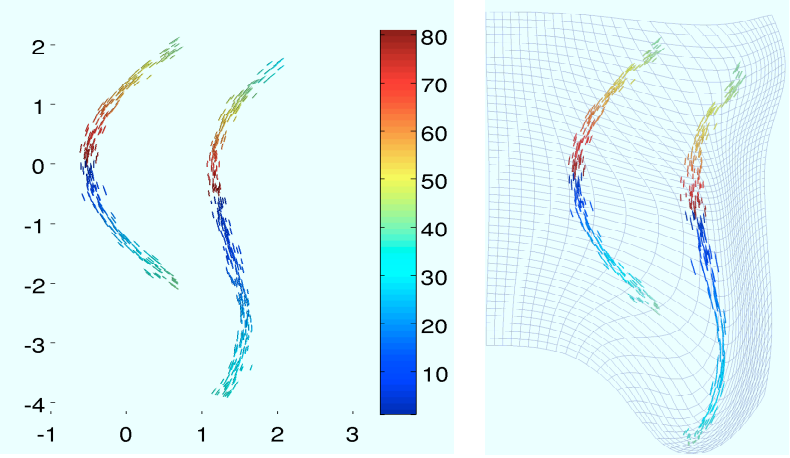
the surface example of figure 2.4, we can observe that functional currents introduce no bias between low and high signals as opposed to colored currents. It also gives significantly different results from currents in the product $E \times M$ in certain situations, as the example shown in figure 2.14. Functional currents provide robustness to disconnections of the shape and discontinuities of the signal. This is a particularly important property when working with fiber bundle type of data like the example of figure 2.15. Some additional results on medical imaging datasets will be studied in detail in chapter 4, that generalizes this setting to atlas estimation on a group of subjects.

Registration with additional functional residuals' estimation

The deformation model underlying the previous algorithm only transports functional values by diffeomorphism. It is clear that in many situations, some variations in the patterns of the signals might exist between the source and target functional shapes, which are not taken into account nor estimated by the algorithm. This can also deteriorate the geometrical deformations themselves since regions with different signal values would have a tendency to be mismatched. One solution to address such situations is to introduce an enriched deformation group that would involve, in complement to the diffeomorphism acting on the support, a residual function that adds to the source shape. We can model this exactly by the transport of functional shapes expressed in eq.(2.2) and the corresponding action on fcurrents provided by eq.(2.12). In this setting, the goal would be to estimate both the diffeomorphism ϕ as previously, and the residual function h, by solving therefore an optimization problem of the form :

$$\min_{p_0,h} \; E(p_0) + \gamma_1 \|C_{(\phi_1(X),(f+h)\circ(\phi_1)^{-1})} - C_{(Y,g)}\|_{W'} + \gamma_2 \int_X h^2(x) d\mathcal{H}^d(x) \qquad (2.44)$$

Observe that a penalty term on the residual function h is added to the criterion in order to impose a constraint on this function h we intend to estimate. This is important to avoid the situation where **all** the functional variability would be explained by the function h and not also by the deformation acting on the support (which cannot modify directly the values of the functional but can change the shape of the level



Source and target Fcurrent matching

Figure 2.15: Example of matching on the case of a fiber bundle with signal. On the left figure, the source and target functional shapes. On the left, the resulting matching with the deformed shape and the deformation grid for the functional currents' setting.

sets of signals). The additional weight parameter γ_2 precisely controls this. Here, we choose a rather weak penalty function that is the L^2-norm of h on the shape X. Different choices could be made to impose more regularity to the function h but will not be discussed here. Neither will be studied the theoretical existence of a minimum to eq.(2.44) in the general setting, which is not a priori obvious to deduce from the purely geometrical case and would probably require particular attention.
The numerical scheme we propose for optimizing eq.(2.44) is not fundamentally different from the previous one : we optimize simultaneously in both variables p_0 and h by adaptive multi-step gradient descent. The gradient with respect to p_0 is computed exactly as we did earlier since the last term does not depend in p_0. The only new component is the gradient with respect to h. In the discrete setting, if $h_i = h(q_i)$ and $\tilde{h}_i = \dfrac{1}{d+1}(h_{\alpha_i^1} + ... + h_{\alpha_i^{d+1}})$, then the penalty term on h is :

$$|h|_{L^2} = \sum_{i=1}^{m} \tilde{h}_i^2 \|\xi_i\| \tag{2.45}$$

and thus the gradient with respect to h is obtained as :

$$\nabla_{h_k} |h|_{L^2} = \sum_{i=1}^{n} 2\tilde{h}_i . \partial_{h_k} \tilde{h}_i \ \|\xi_i\| \tag{2.46}$$

Note that the terms in the previous sum are different from zero only if h_k is a vertex of the ith face of X so that the sum is not actually over all the faces. Registration with estimation of h can be thus performed as summed up in algorithm 1. This algorithm somehow generalizes the previous one because taking $\gamma_2 \to \infty$ would enforce the residuals to remain zero-valued on all the source shape. In figure 2.16 is shown one result of such a matching : we draw special attention on the balance between the deformation of existing signal on the source shape (as in the matching model of subsection 2.3.2) and the residual function that is added to X. This balance is controlled by the ponderation parmameters γ_1 and γ_2. Thanks to the use of fcurrents'

Algorithm 1 Registration of functional shapes with adjunction of residuals

Require: Source shape $(X, f) = (q_k, f_k)_{k=1,..,n}$, target shape (Y, g), gradient descent step δ.

1: set $p_0 = 0$ and $h = 0$.
2: **while** Convergence **do**
3: Flow (X, f) by forward integration with p_0 and store the deformed template at time 1 $(\phi^{p_0}, h).(X, f) = (q_{k,1}, f_k + h_k)$.
4: Compute the fcurrent representations of (Y, g) and $(\phi^{p_0}, h).(X, f)$.
5: Compute the gradients of the attachment term A with respect to the final point configuration $q_{k,1}$ eq.(2.40).
6: Apply the backward flow integration of eq.(2.43) and deduce P_0.
7: Compute $\nabla_{p_0} J = K_V(q, q)p_0 - P_0$.
8: Compute the gradient $\nabla_h A$ of the attachment term A with respect to the residual values h_k by eq.(2.40) and the gradient of the L^2 norm by eq.(2.46).
9: Update variables :

$$p_0 \leftarrow p_0 + \delta \nabla_{p_0} J$$
$$h \leftarrow h + \delta \nabla_h J$$

10: Compute new functional $J(p_0, h)$ and adapt step δ by line search.
11: **end while**
12: **return** initial momentum of the deformation p_0 and the functional residual h.

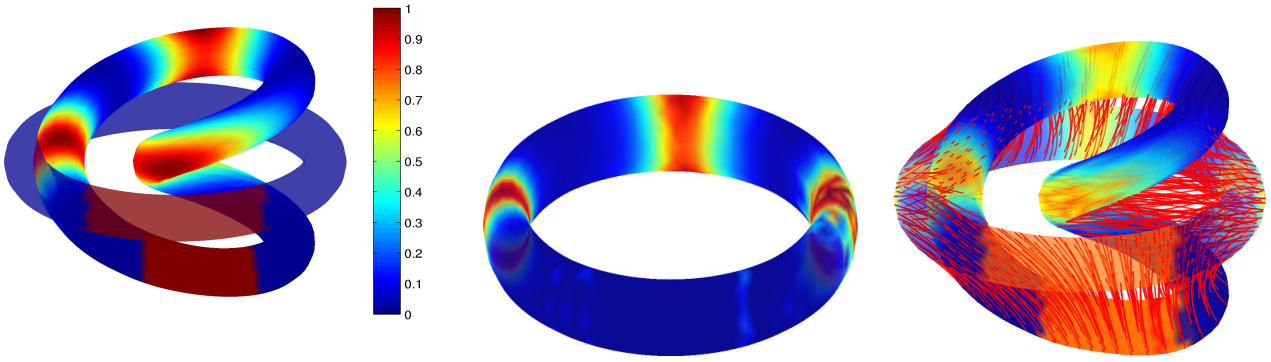
norms to drive the deformations, residuals can be concentrated in regions where the difference of the signals cannot be explained by deformations.

2.3.3 Numerical considerations

To conclude this chapter, we examine in depth the numerical issues that are specifically induced by the previous algorithms, especially when the number of points on the shapes gets large, and evaluate a few possible solutions to address them.

Kernel sums' computations : numerical schemes and theoretical complexities

The overwhelming part of numerical computations of the algorithms presented in the previous section is taken by evaluation of kernels at each descent step. This occurs at essentially two spots in the process. First for the dynamics of the deformations themselves because it is necessary to integrate forward and backward differential equations. As we explained in chapter 1, if we fix a kernel K_V for the vector fields in E (generically a Gaussian kernel of scale σ_V), at each time step, a displacement field generated by the momenta at every point of the source shape need to be computed at each of the source shape points. Let's consider for example the forward equation on q : at each time step, one has to compute the displacement field $\sum_{j=1}^{n} K_V(q_{t,i}, q_{t,j})p_{t,j}$ for each of the source point positions $q_{t,i}$. As we see, this has a n^2 complexity in the number of points of the source shape which has to be then multiplied by the number of time steps of the integration scheme.



Source and target | Estimated residual h | Deformed shape with residual

Figure 2.16: Example of matching with estimation of residual function between two colored surfaces.

A second bottleneck in terms of computational intensity comes with the fcurrent data attachment distance and its gradient. This is again an evaluation of kernels which is this time a tensor product kernel $K_g \otimes k_f$ living on the bigger product space $E \times M$. As we see with eq.(2.18), computing the distance is once again **quadratic** in the number of Diracs of the source and target. This number of Diracs corresponds typically to the number of segments if we are working on curves or the number of triangles for surfaces, which we can consider roughly of the same order as the number of points of the shapes.
In both cases, we see that the recurrent computations are essentially n kernel 'convolutions', that we can write in general :

$$\forall i = 1, .., n, \quad \sum_{j=1}^{n} K(a_i, a_j)\alpha_j \tag{2.47}$$

where a_i and α_i are either elements of E or $E \times M$. This quadratic complexity is a clear limitation of the approach when one has to work with highly-sampled shapes, which is rather frequent in medical imaging datasets. As we will highlight in the following, brute force computations of kernels in MATLAB on a standard PC becomes nearly intractable for shapes with more than 2000 points whereas many of the applications we deal with in this manuscript involve surfaces with more than 10000 points. This is the reason why several alternative numerical methods have been investigated in the past and are still explored at the present time. Those can consist either in computing eq.(2.47) using different numerical platforms than Matlab that would take profit in parallel computing for instance or in finding approximate but faster numerical schemes. We present a few possibilities : a comparative table of time computations will be given and commented in the next subsection.

- A first straightforward thing that we tried is to implement all kernel computations as C++ mex files that can be used in the MATLAB body program. This allows an easy use of parallel computing libraries in C that improve by a substantial factor the computational time. Another promising alternative is provided by the recent development of graphics cards. We thus experimented the computation of eq.(2.47) based on a mex CUDA implementation on GPU and included it also in our code.
- At the moment, all these methods are exact or nearly exact computation of kernel

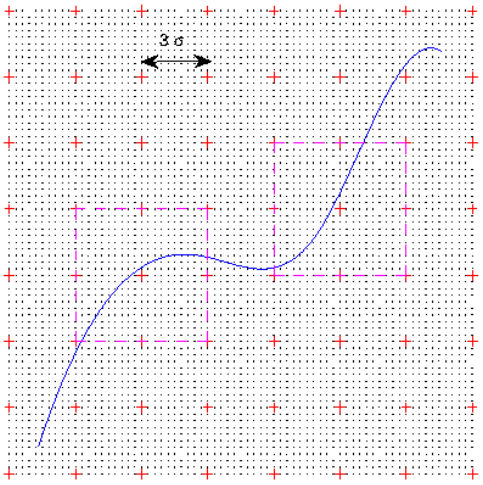


Figure 2.17: Illustration of the grid-partitioning strategy in 2D. Small black dots are the original grid points, red crosses correspond to all possible centers of blocks. They form a regular sur-grid of bigger step (usually of the order 3σ).

sums. Other methods intend to approximate eq.(2.47). Usually, such methods are designed for the case of radial scalar kernels. We will assume, to simplify the presentation, that we are working with a Gaussian kernel with a certain scale parameter σ. A first simple idea is to notice that if points are farther away from each other than 3σ then the value of the kernel is lower than 10^{-4} : thus kernels need not be evaluated between all pairs of points but only for points lying at a small enough distance. This suggests that we can save computations of exponentials especially when σ is small compared to the size of the objects. In addition to that, one could also pre-compute a table of values for the Gaussian on a discretized grid of distances and then access the corresponding box instead of recomputing it at each iteration of the loop. The search for points in a the desired neighborhood of a given position could be also improved computationally by using KD-tree algorithms.

– In the same vein, more sophisticated methods have been also developed for Gaussian kernels, such as the Fast Gauss Transform [84], which was applied later on to compute distances on currents in [41]. This method combines the idea of clustering in order to avoid computations of negligible kernel values with the principle of fast multiple methods (FMM) for accelerating exponentials' evaluations, leading to an algorithm that has a theoretical complexity of $O(np)$ where p is the number of clusters. Yet this number of clusters is directly correlated to the scale σ and to the dimension of the space. Since p is not set up automatically in the algorithm of [84], one difficulty of the method is to choose it properly to have both a good accuracy without slowing down computations too much. In practice, although it leads to fast computations especially for big σ and not too high dimension, it makes the quality of the approximation difficult to evaluate.
– Another interesting numerical scheme was investigated in [28], which basically projects all the momenta α_i located at points q_i on a fixed **linearly spaced grid** and then expresses eq.(2.47) as a convolution on this grid which can be

n	100	1000	2000	10000	50000	exact computation
MATLAB brute force computation	0.98s	98.2s	391.4s	∞	∞	yes
C++ parallelized computation	0.025s	0.42s	1.6s	40.8s	1030s	yes
Truncation and pre-computation of Gaussian	0.009s	0.52s	2.1s	53.3s	1931s	no
GPU	0.02s	0.03s	0.07s	1.16s	26.5s	yes
FGT	0.035s	0.36s	0.72s	3.94s	20.5s	no
Grid convolution	0.11s	0.11s	0.11s	0.16s	0.43s	no

Table 2.1: Computational time for 20 convolutions with $\sigma = 0.3$.

computed by FFT and inverse FFT. The projection step is linear in n while the computation of the convolution in the grid domain is of complexity $O(n_g \log n_g)$ where n_g is the number of points of the grid. Now, the grid is chosen as a linearly spaced set of points delimited by a rectangular parallelepiped containing all the points and whose sampling is chosen as a fraction of the scale of the Gaussian σ. The accuracy of the approximation can be then controlled by the grid step, as shown in [28] chapter 2. As for the number of points in the grid, we see that it grows as s^{-D} where s is the step of the grid and D the dimension of the ambient space. The interesting thing to observe is that, if we except the projection on the grid (which is anyhow not limiting), the complexity of the computation does not depend anymore on the number of points on the shapes but only on the grid size, which makes it particularly adapted when n is very large. However, the size of the grid itself grows exponentially with the dimension : in practice it remains affordable in dimension 3 or less but it becomes rapidly incredibly consuming both in time and in memory in dimension 4 and more. In the applications we have been considering so far (with $E = \mathbb{R}^3$), this method is well-adapted for the deformation part, but we see that fcurrent attachment distances involve computation of kernels living on $E \times M = \mathbb{R}^4$. In that case, introducing 4D grids that contain all the points $\{(x_i, f_i)\}$ produces huge n_g. As we see, this is essentially a problem of codimension because the points of the functional shapes lives basically on a 2-dimensional manifold in $\mathbb{R}^4$. A possible solution that we implemented to control the size of the grids is to make block computations based on a partition of the original grid as illustrated in figure 2.17, which restricts the convolutions to only overlapping blocks that actually contain points of the shape. Yet, even reduced, the curse of dimension is still occurring in each block, which makes this approach barely feasible for 4D but still unthinkable for higher dimensions.

Comparative computational times

To give a more concrete vision of the previous discussion, we show a few computational time for the different numerical methods that have been evoked previously. All computations have been performed on a standard quad core laptop with a NVIDIA NVS 5100 graphics card. We first show the elapsed time for the computation of 20 convolutions as in eq.(2.47) for a 3D Gaussian kernel and with an increasing number of points n. The positions q_i were picked randomly in the cube $[0, 1]^3$. In table 2.1, computations are first done for a big kernel ($\sigma = 0.3$). Brute-force computation in MATLAB becomes unfeasible with more than 2000 points. We also see that truncation of exponentials is clearly not efficient in that case whereas for a smaller scale $\sigma = 0.1$ (table 2.2), we can gain a factor 2 compared to the exact computation in

n	100	1000	2000	10000	50000	exact computation
MATLAB brute force computation	0.98s	98.1s	392s	∞	∞	yes
C++ parallelized computation	0.033s	0.42s	1.6s	41s	1039s	yes
Truncation and pre-computation of Gaussian	0.004s	0.25s	0.92s	23.2s	843s	no
GPU	0.02s	0.03s	0.078s	1.16s	26.7s	yes
FGT	0.06s	0.62s	1.7s	11.2s	58.9s	no
Grid convolution	1.20s	1.24s	1.20s	1.32s	1.7s	no

Table 2.2: Computational time for 20 convolutions with $\sigma = 0.1$.

n_g	32000	64000	260000	1000000
Time	0.13s	0.34s	0.8s	4.7s
Relative error	5%	2.9%	1%	0.4%

Table 2.3: Precision vs time with the grid approach (with $\sigma = 0.2$)

C++. Note the important acceleration provided by GPU computation in both cases that outperforms the previous methods as soon as n is bigger than 1000. It is particularly optimal in the range 1000-10000 points. With larger n, some approximative methods may become faster. The FGT, for instance, is quite good for large scale kernels since, to obtain a given precision, the required number p of clusters decreases with σ. For smaller kernels, the gain is not as clear. The choice of parameters (p in particular) is also a not trivial step with FGT. In the experiments we show here, we made several tries to adapt the number of clusters in order to obtain a relative precision around 1% for each case.
The grid convolution method is also very efficient. In those examples, we used grids sampled at $\sigma/5$ which gives $n_g \approx 32000$ in table 2.1 and $n_g \approx 260000$ in table 2.2. As already explained and confirmed by numerical simulations, the computational time is then almost independent of n, which is particularly interesting for highly sampled shapes (typically $n \geq 10000$). However, this is an approximation of the true result. In table 2.3, we show both time and precision for a computation of convolutions with $\sigma = 0.2$ using finer and finer grids. Getting a very good approximation requires more accurate grids and leads to longer computations. This is basically why such a method is not well-fitted for higher dimension than 3. In 4D for instance, with $\sigma = 0.2$, requiring a precision of the order of 1% necessitates the use and storage of grids with at least $n_g = 5 \times 10^6$ points, with the limitations that this induces in time and memory.

So which one is the best and how far can we go ?

It is still hard, at this point, to give a definite answer to the numerical issue. The positive aspect is the fact that there are several possible numerical schemes available and still some new ones appearing in the literature. Focusing only on the ones we have experimented on, the previous discussion tends to show that the choice of the numerical method should be probably tailored to the situation. Working on computers equipped with NVIDIA graphics cards, GPU computation is obviously a very solid candidate that is adapted to most cases since it provide exact computation of kernels with limited memory consumption and very competitive computational times, not mentioning that the times shown in the previous subsection were not ob-

tained with the most recent and powerful graphics card. Yet, this is still a quadratic scheme which becomes inevitably limited when the number of points exceed 10000. For such high resolution shapes, the grid strategy offers a good alternative for the computations of the diffeomorphisms' part (involving the kernel K_V which is always in 3D in our applications) because the computational time is then dependent on the size of the grid and not the sampling of the shape. However, it does not work as well with the fcurrent distance computation because of the dimension increase and the resulting grids' sizes. A satisfying solution in those cases is to use grid computations for diffeomorphisms (which is the part that involves in practice the larger number of convolutions) and GPU for the data attachment term. More recent versions of the FGT, as in [66, 73] for instance, could be also an interesting track but were not investigated here.
The implementation we have made integrates all the different methods that were mentioned, allowing therefore a relative adaptability to many cases. To give an overall idea, at the present state, a complete registration on a computational server (100 gradient descent steps) between surfaces with 1000 points takes a few minutes and less than half an hour with 10000 points using the most adapted method. The algorithm is also able to run in decent time for shapes with up to 50000 points.

2.4 Conclusion

We have presented in this chapter a way to generalize the notion of currents for the purpose of integrating functional shapes into a coherent and robust representation. Functional currents provide a framework to model geometrically-supported signals of nearly any type and regularity while preserving the advantages of currents to model the geometry. We have shown that RKHS metrics can be defined in spaces of functional currents, providing distances between functional shapes that allow the simultaneous comparison of geometry and signal without requiring any preliminary exact matching between the geometrical supports. This class of norms also shares important control properties both with respect to deformations and signal transformations, as well as robustness to small disconnections and changes of topology. The resulting Hilbert structure on fcurrents opens the way to a very wide class of applications. We detailed two possible numerical applications that are of usual interest in the field of computational anatomy : a matching pursuit scheme to address fcurrent compression and averaging and an adaptation of LDDMM algorithm for diffeomorphic registration of two functional shapes. A natural and expected extension of registration is the problem of atlas estimation on a group of functional shapes, which brings special problematics and will be the subject of chapter 4. From another angle, the question of *orientation* of shapes that is fundamental with currents is also transfered to functional currents : this issue is precisely the subject of the next chapter.

Chapter 3

Non-oriented shapes and varifolds

Contents

3.1 Orientation and limitations of currents 90
3.1.1 Orientation ambiguity 90
3.1.2 Shape annihilation 91
3.1.3 Non-oriented representations in geometric measure theory . 93
3.2 Varifolds 94
3.2.1 Definitions and basic properties 94
3.2.2 Non-oriented shapes as varifolds 97
3.3 Kernel metrics on varifolds 100
3.3.1 Kernels on the Grassmann manifold 100
3.3.2 Kernels on varifolds 103
3.4 Properties of RKHS norms on varifolds 107
3.4.1 Dependency on the scales of the kernels 107
3.4.2 Control of the volume by W'-norm 108
3.4.3 Metric variation formula 112
3.5 Registration of unoriented shapes 115
3.5.1 Description of the algorithm 115
3.5.2 Discrete expressions for curves and surfaces 116
3.5.3 Numerical complexity 118
3.5.4 Simulations and results 119
3.6 Discussion, extensions and perspectives 121
3.6.1 Scale and noise 122
3.6.2 Functional varifolds 126
3.6.3 Oriented varifolds 127
3.6.4 Conclusion 127

Results of this chapter were partially published in [22].

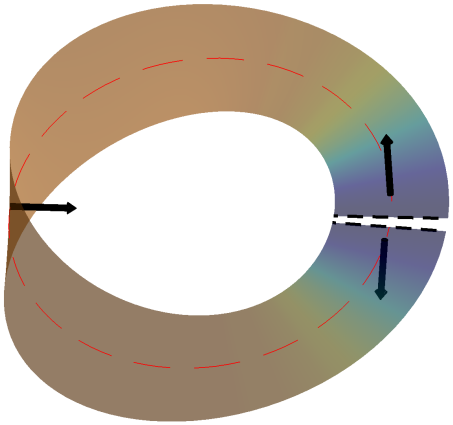


Figure 3.1: Orientation issues on non-orientable surfaces.

3.1 Orientation and limitations of currents

As we mentioned several times in the previous chapters, currents are intrinsically modelling **oriented** objects. This means that the process of representing a rectifiable subset X by its corresponding current C_X as in chapter 1 requires an orientation on X. If X is typically a connected submanifold, there are two possible orientations for X. Denoting X and $\check{X}$ the two resulting oriented submanifolds, we have by definition :

$$C_{\check{X}} = -C_X$$

When using kernel metrics on currents, the orientation is diffused in space leading thus to potential cancellations of some parts of shapes. Typically, for a Gaussian kernel of width σ,

$$\|\delta_x^\xi + \delta_y^{-\xi}\|_{W'} = 2|\xi|^2 \left(1 - e^{-|x-y|^2/\sigma^2}\right) \tag{3.1}$$

which gets small whenever $|x-y|$ is small compared to σ. In several situations, this orientation dependency and cancellation effect is an important drawback of currents as first pointed out in geometric measure theory [8, 67] and later on in [28, 41] for the use of currents in large deformations' settings. In the following, we list the different limitations that orientation induces in our context.

3.1.1 Orientation ambiguity

Difficulties can first occur with non-orientable shapes on which integration of differential forms like in equation 1.10 is not well-defined. If we consider for instance a Möbius band as in figure 3.1, one could still imagine that we cut the Möbius band at some location to obtain a usual orientable band that can be represented as a current. Yet, for the kernel metric, due to the presence of opposite orientations within a certain domain in space and eq.(3.1), we actually cancel a non-zero surface part of the shape around the cutting curve that depends on the typical width of the kernel. Thus there is no canonical and satisfying way to represent non-oriented shapes with currents.
This first drawback remains anecdotal though, since most shapes in computational anatomy are orientable. However, even if orientable, orientating shapes **consistently**

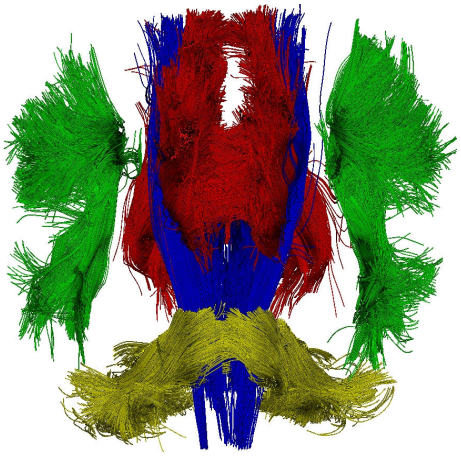


Figure 3.2: An example of white matter fiber bundle estimated from Diffusion Tensor Imaging (DTI) illustrating the potential difficulty of consistent orientation of all different fibers.

can be either a difficult or even ill-posed problem in certain datasets. This is notably the case for fiber bundles in the 2D or 3D space consisting of many different and possibly disconnected pieces of curves. Indeed, if a given shape has N connected components, there are 2^N different possible orientations. Using currents with such objects requires a way to propagate orientation from one part of the shape to all the rest. Technically, this leads to additional pre-processing which can be particularly tricky for highly disconnected bundle of curves with many different directions like the ones corresponding to white matter fibers estimated from DTI in brain imaging, we show the example of one subject in figure 3.2. This example also emphasizes another difficulty appearing with sets of curves crossing each other for which the very notion of 'consistent' orientation may become meaningless. In such cases, the orientation is clearly irrelevant and one would like to treat objects as sets of unoriented shapes.

3.1.2 Shape annihilation

Orientation may also represent an obstacle even in the simplest case of usual oriented and connected submanifolds because structures like sharp spines or tails naturally lead, when represented in a kernel Hilbert space of currents, to annihilation of some parts of the shapes. This comes again from equation (3.1), or, to state it in a more general way, from the fact that, while the W'-norm of a set is always dominated by its Hausdorff measure as we saw in section 1.3.4 eq.1.17, **there exists sets of given Hausdorff measure but with arbitrarily small W'-norm**. The plane curve of figure 3.3 is an example. Indeed, the 1-Hausdorff measure being the length of the curve : $\mathcal{H}^1(\gamma) = 2(1+\epsilon) \geq 2$. On the other hand, if γ is parametrized on an interval I, then we have :

$$\|C_\gamma\|^2_{W'} = \int_{I\times I} \gamma'(s)^T K(\gamma(s),\gamma(t))\gamma'(t) ds dt$$

If we assume, to simplify computations, that kernel K is a translation invariant kernel of the form $K(x,y) = k(x-y)\,\mathrm{Id}_{\mathbb{R}^2}$ with Lipschitz regularity, then, by expanding the integrals as the sum of the four pieces of curves γ_{v_1}, γ_{v_2}, γ_{h_1}, γ_{h_2}, we can eventually

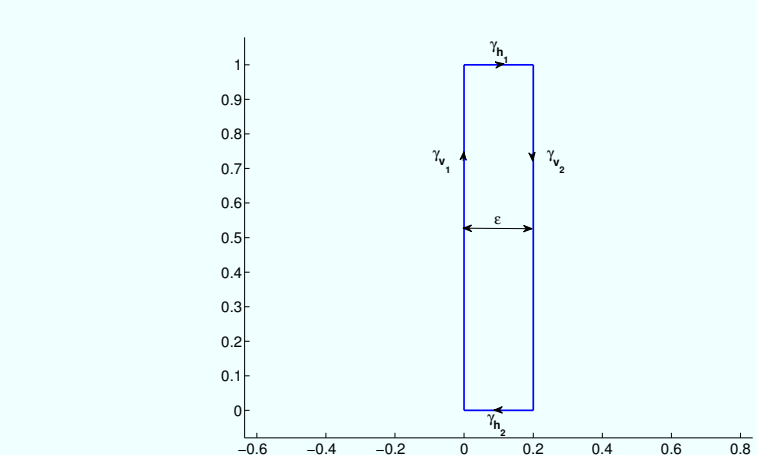


Figure 3.3: Counter-example to the reciprocal domination. For all ϵ, $\mathcal{H}^1(\gamma) = 2(1+\epsilon) \geq 2$ whereas $\|C_\gamma\|_{W'} \underset{\epsilon \to 0}{\longrightarrow} 0$

simplify the result thanks to the translation-invariance, which gives :

$$\|C_\gamma\|^2_{W'} = 2(\|C_{\gamma_{v_1}}\|^2_{W'} - \langle C_{\gamma_{v_1}}, C_{\gamma_{v_2}} \rangle_{W'}) + 2(\|C_{\gamma_{h_1}}\|^2_{W'} - \langle C_{\gamma_{h_1}}, C_{\gamma_{h_2}} \rangle_{W'}) \,.$$

Since $\|C_{\gamma_{h_1}}\|^2_{W'} = \|C_{\gamma_{h_2}}\|^2_{W'} = \iint_{[0,1]\times[0,1]} k(\gamma(s) - \gamma(t))\epsilon^2 ds dt$, it is straightforward that $\|C_{\gamma_{h_1}}\|^2_{W'} - \langle C_{\gamma_{h_1}}, C_{\gamma_{h_2}} \rangle_{W'} = O(\epsilon^2)$. Moreover,

$$\|C_{\gamma_{v_1}}\|^2_{W'} - \langle C_{\gamma_{v_1}}, C_{\gamma_{v_2}} \rangle_{W'} = \iint_{[0,1]\times[0,1]} [k(\gamma(s) - \gamma(t)) - k(\gamma(s) - \gamma(t) - (\epsilon, 0))] \, ds dt \,.$$

Since the kernel k is assumed to be Lipschitz,

$$[k(\gamma(s) - \gamma(t)) - k(\gamma(s) - \gamma(t) - (\epsilon, 0))] \leq \text{Cte}\, \epsilon$$

Therefore, we have eventually proved that $\|C_\gamma\|_{W'} = O(\sqrt{\epsilon})$ whereas for all ϵ, $\mathcal{H}^1(\gamma) \geq 2$.
If this annihilation can be beneficial in certain situations because it is way to smooth noise, it also means that some meaningful structures of shapes, modeled as elements of a RKHS of currents, will vanish in this representation. We will detail, in a coming section, what important limitations this can induce in matching certain examples of curves or surfaces having such kind of structures.
Another fundamental corollary to that can be summed up by the motto **'currents don't see anything at large scales'**. Indeed, if we consider a Gaussian kernel on the space E and make its characteristic scale tend to ∞, i.e $k_e(x, y) = Id_E$, then for any d-dimensional submanifold of X, the norm of the current C_X becomes :

$$\|C_X\|^2_{W'} = \iint_{X\times X} \langle \xi(x), \xi(y) \rangle = \langle \int_X \xi(x), \int_X \xi(x) \rangle$$

Now, from Stokes' formula, one can check that $\int_X \xi(x)$ is a term that only depends on the boundary of X. Therefore, at large scale, RKHS norms on currents are only sensitive to the boundaries of objects, and all shapes with no boundary vanish in W'.

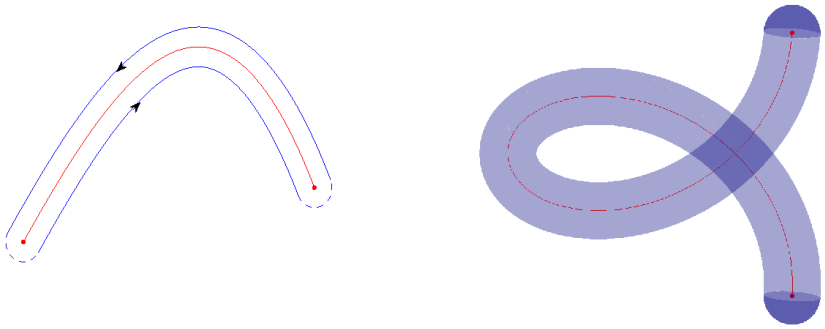


Figure 3.4: Normal bundles of curves with boundary both in 2D and 3D.

3.1.3 Non-oriented representations in geometric measure theory

The issue of orientation is not new in the field of geometric measure theory and has drawn attention since its early beginnings. After the definition of currents, several concepts have been proposed to account for non-orientability.
The closest to the idea of currents is to consider *flat chains and currents modulo* 2 as introduced in [37] section 4.2.26 and summed up in [67]. The basic idea is to define a quotient space by identifying rectifiable currents that differs by $2.Q$ where Q is some rectifiable current. In that quotient space, we easily see that $C_X = -C_X$ for any rectifiable subset X, and thus the representation is insensitive to orientation. Although several compactness, approximation and existence results were shown for these objects, they are essentially theoretical at the present time and do not seem to adapt easily to our type of applications.
Chronologically, the second approach that was proposed is the one of *varifolds*, as pioneered by F. Almgren in [8], later developed by W. Allard in [4] and used in various applications in calculus of variations [17, 55]. We shall focus on this particular one and present it in detail in this chapter. As we shall see in the following, the essential step is to define a model of non-oriented tangent spaces, which is the role of the Grassmann manifold.
Finally, we could also evoke the more recent notion of *normal cycle* presented in [39], which was essentially used in computational geometry as a way to define and compute curvature tensors for discrete polyhedral surfaces and show the convergence toward the continuous versions [25]. To present it briefly, the idea of normal cycles is to consider the unit normal bundle of a submanifold X which is defined as

$$NU(X) = \{(x, n(x))/\ x \in X, n(x) \in (T_x X)^{\perp}, \|n(x)\| = 1\}$$

and can be considered as a submanifold of dimension $n-1$ in the $2n$-dimensional vector space $E \times E$. The important point is that $NU(X)$ is always orientable (even if X is not) and can be canonically oriented. It can be roughly identified to a tubular envelope around X as on the curve examples of figure 3.4. This definition can be generalized to the case where X is a manifold with boundaries or for polyhedrons by adding *normal cones* : we refer to [25] for more details. Now the normal cycle associated to X is simply the representation as a current in $E \times E$ of the normal bundle $NU(X)$. In our context of work, this approach does not differ significantly from currents in the theoretical point of view : kernels could be defined as well and provide metrics between non-oriented shapes as with usual currents which could be used to drive matching algorithms. This is presently studied by J. Glaunès and has

been implemented in the planar curve case. Even though this model is rich and is not intractable for actual computations, it leads to an important increase in the dimension of the spaces, first because we are looking at currents in the bigger space $E \times E$ but also because the actual modeled object is no more X but its unit normal bundle $NU(X)$ that can be a submanifold of higher dimension (e.g in the 3D curve case). Thus, it seems unlikely, at the present time, that normal cycle representations could provide metrics with numerical complexities comparable to currents.

3.2 Varifolds

We now focus exclusively on varifolds. We shall explain in what respect varifolds offer in our sense a natural and convenient setting to embed non-oriented shapes in the context of computational anatomy both from the theoretical and numerical aspect. In this section, we present varifolds first as a theoretical object, following the original work of Almgren in [8] and later of Allard in [4]. We then explain in detail, in a spirit very similar to currents, the fundamental relationship between varifolds and unoriented rectifiable sets and show how this representation can be efficiently transcribed in a practical way. The construction of kernel metrics on varifolds is postponed to the next section.

3.2.1 Definitions and basic properties

Varifolds have been first introduced in the context of geometric measure theory as a way to address Plateau's problem of finding least area surfaces with a prescribed boundary. These developments clearly do not enter in the scope of this manuscript. Our purpose here is rather to connect conveniently varifolds to the framework and language of computational anatomy. Therefore, in the following, we will focus mainly on definitions of such objects and explain in what respect these definitions are computationally relevant. The basic idea behind varifolds is to represent any rectifiable set (orientable or not) as a distribution of unoriented tangent spaces spread in the embedding space E. As we shall see, varifolds encompass not only manifolds and rectifiable sets but more generally sets of directions in the space. Let's take again the conventions of the previous chapters, E being a vector space of dimension n and d an integer with $0 \leq d \leq n$. What we first need is a way to represent tangent spaces of dimension d in the space E.

Grassmann manifold

The proper space in that regard is precisely the Grassmann manifold defined as :

Definition 3.2.1. *The Grassmann manifold of dimension d in E, denoted $G_d(E)$, is the set of all d-dimensional subspaces of E. It can be identified to the quotient space of all families of d independent vectors of E by the equivalence relation obtained by identifying families that generate the same subspace.*

As a quotient space, one can show that $G_d(E)$ inherits a structure of compact Riemannian manifold of dimension $d(n-d)$ (cf [16] chap 4 and [83]). It is also a classical result that $G_d(E)$ is a homogeneous space under the action of the group $\mathrm{GL}(E)$ and the subgroup of direct isometries $\mathrm{SO}(E)$.

Remark 3.2.1. *Considering the application $G_d(E) \to G_{n-d}(E),\ V \mapsto V^\perp$, one sees easily that $G_d(E)$ and $G_{n-d}(E)$ can be identified with each other. In addition, in the particular cases where $d = 1$ or $d = n - 1$, the Grassmann manifold is nothing else than the real projective space of E.*

The geometry of the Grassmann manifold can be also described in a less abstract way by giving explicit local charts. If V is an element of $G_d(E)$, one can consider $\mathcal{U}_V = \{W \in G_d(E) \mid W \cap V^\perp = \{0\}\}$, which is an open neighborhood of V in the Grassmann manifold.

Lemma 3.2.1. *Every element $W \in \mathcal{U}_V$ can be written as $W = \{v + l(v),\ v \in V\}$ for a certain linear function $l \in \mathcal{L}(V, V^\perp)$ and l is uniquely determined by W.*

Proof. If $W \in \mathcal{U}_V$, for dimensionality reason, we have $W \oplus V^\perp = E$ and denoting h the linear projector on $V^\perp$ associated to this decomposition, we deduce that $W = \{u - h(u) \ / \ u \in E\}$. Now writing for all $u \in E,\ u = v + v^\perp$ with $v \in V$ and $v^\perp \in V^\perp$, we have

$$u - h(u) = v + v^\perp - h(v) - h(v^\perp) = v - h(v)$$

since $h(v^\perp) = v^\perp$. Thus $W = \{v + l(v) \ / \ v \in V\}$ with $l = (-h)_{|V}$ and l is indeed uniquely determined by W. □

Therefore, we have a bijective map $\psi_V \ : \ \mathcal{U}_V \to \mathcal{L}(V, V^\perp),\ W \mapsto l$.

Proposition 3.2.1. *The set of charts $(\mathcal{U}_V, \psi_V)$ defines an atlas on the manifold $G_d(E)$ and, as a consequence, there is a natural isomorphism between the tangent space $T_V G_d(E)$ and $\mathcal{L}(V, V^\perp)$.*

Proof. We refer the reader to [68] section 2.2 for the complete proof of that. □

In addition, the Grassmannian can be considered as a submanifold of $\mathcal{L}(E)$ by identifying any $V \in G_d(E)$ with the orthogonal projection p_V on V which gives the embedding :

$$\begin{array}{rcl} G_d(E) & \longrightarrow & \mathcal{L}(E) \\ V & \longmapsto & p_V \end{array} \tag{3.2}$$

We shall make use of that when defining kernels on $G_d(E)$ (section 3.3). It gives also a convenient way to identify tangent spaces and compute variations. Indeed, if V_t for $t \in]-\epsilon, \epsilon[$ is a smooth curve on the Grassmann manifold then p_{V_t} is a smooth curve on $\mathcal{L}(E)$. Let's denote $\delta p_V \doteq \left.\dfrac{d}{dt}\right|_{t=0} p_{V_t}$. Then, differentiating the relations $p_{V_t} \circ p_{V_t} = p_{V_t}$ and $p_{V_t} = p_{V_t}^*$, we find that :

$$\delta p_V = \delta_{p_V} \circ p_V + (\delta_{p_V} \circ p_V)^*$$

Moreover, for all $v, v' \in V$, we have :

$$\langle \delta p_V(v), v' \rangle = \langle \delta p_V(v) + \delta p_V^*(v), v' \rangle = \langle \delta p_V(v), v' \rangle + \langle v, \delta p_V(v') \rangle$$

And thus for all $v, v' \in V$, $\langle v, \delta p_V(v') \rangle = 0$. We deduce that $\delta p_V = p_{V^\perp} \circ \delta p_V$ and :

$$\delta p_V = p_{V^\perp} \circ \delta_{p_V} \circ p_V + (p_{V^\perp} \circ \delta_{p_V} \circ p_V)^* .$$

It results that we can identify any element δp_V of the tangent space $T_V G_d(E)$ with the application

$$p_{V^\perp} \circ \delta_{p_V} \circ p_V = (p_{V^\perp} \circ \delta_{p_V})_{|V} \in \mathcal{L}(V, V^\perp) .$$

Moreover, if $A(t)$ is a curve in $GL(E)$ with $A(0) = \mathrm{Id}$ and $A(t).V$ denotes the action of $GL(E)$ on $G_d(E)$, it is a straightforward verification that, under this identification :

$$\left.\frac{d}{dt}\right|_{t=0} A(t).V = p_V^\perp \circ A'(0)_{|V} . \tag{3.3}$$

Finally, let's mention that the Grassmann manifold can be also linked to the space of d-vectors used in the definition of currents. This is formulated by the Plücker embedding property :

Proposition 3.2.2. *The following application is an embedding :*

$$\begin{array}{rcl} i_P : \ G_d(E) & \longrightarrow & P\left(\Lambda^d E\right) \\ vect(v_1, ..., v_d) & \longmapsto & [v_1 \wedge ... \wedge v_d] \end{array}$$

where $P\left(\Lambda^d E\right)$ is the real projective space (i.e the set of all lines) of $\Lambda^d E$. It is even a homeomorphism in the cases $d = 1$ or $d = n - 1$.

As a result, we can think of an unoriented d-dimensional tangent space as an element of $P\left(\Lambda^d E\right)$, which is however a space of dimension $\binom{n}{d} - 1$, much bigger in general than $G_d(E)$.

Principle angles

A common and useful notion closely related to Grassmannian is the one of *principle angles* between subspaces. We very briefly recall the few properties that shall be interesting for us in the following. If $V, W \in G_d(E)$ are two d-dimensional subspaces of E, the d principle angles $\theta_1, .., \theta_d$ between them are defined recursively by the relations :

$$\begin{array}{rcl} \cos(\theta_k) & = & \max\limits_{v_k \in V} \max\limits_{w_k \in W} \langle \dfrac{v_k}{|v_k|}, \dfrac{w_k}{|w_k|} \rangle \ with \\ & & \forall i \in \{1, .., k-1\}, \ v_k \perp v_i, \ w_k \perp w_i \end{array}$$

As we see, the first principle angle is the smallest angle between all pairs of unit vectors in V and W. In the case of two lines, it is exactly the non-oriented angle between them. The next θ_i's are defined in a similar way by restricting to orthogonal vectors. The angles are all vanishing if and only if $V = W$ and we easily see that there are at most $\min(d, n-d)$ non-zero principle angles.
The central result that we borrow to [83] shows that the principle angles completely determine the relative position between the two subspaces. In addition, there exists orthonormal frames $(v_1, .., v_d)$ and $(w_1, .., w_d)$ of V and W such that $\cos(\theta_i) = \langle v_i, w_i \rangle$ and such that the different angle 2-planes $\mathrm{Span}\{v_i, w_i\}$ are mutually orthogonal to each other. Be careful that $\mathrm{Span}\{v_i, w_i\}$ is a degenerate 2-plane whenever $\theta_i = 0$.

Definition of varifolds

Following closely [4], varifolds are now defined precisely as distributions of unoriented tangent spaces in E :

Definition 3.2.2. *A d-dimensional varifold on E is a Borel finite measure (or distribution) on the product space $E \times G_d(E)$, i.e an element of $C_0(E \times G_d(E))'$.*

Note that this differs from the original definition of varifolds given by Almgren but his definition is equivalent to the previous one as explained in the preface of [8]. The relationship between the measure and the distribution point of view is, as usual, given by Riesz representation theorem :

$$\forall \omega \in C_0(E \times G_d(E)),\ \mu(\omega) = \int_{E \times G_d(E)} \omega(x, V) d\mu(x, V)$$

In particular, Diracs in the space of varifolds are of the form $\delta_{(x,V)}$ with $x \in E$ and $V \in G_d(E)$ and act on functions of $C_0(E \times G_d(E))$ by the relation :

$$\forall \omega \in C_0(E \times G_d(E)),\ \delta_{(x,V)}(\omega) = \omega(x, V)$$

In this context, a Dirac consists in the data of a position x in the space attached to a d-dimensional (non-oriented) subspace V that will play the role of tangent space in the case of rectifiable sets. Note that varifolds differ from the idea of measures as considered in [41], which are simply elements of the dual of $C_0(E)$. Somehow, varifolds enrich the measure representation with a notion of local directions.
We also define the additional notion of positive varifold that shall be useful for the following :

Definition 3.2.3. *A varifold μ is said to be positive if it is in addition a positive measure on $E \times G_d(E)$ or equivalently if for all positive function $\omega \in C_0(E \times G_d(E))$, $\mu(\omega) \geq 0$.*

For a positive varifold μ, we will also call the support of μ and denote $\mathrm{Supp}(\mu)$ the smallest closed subset $C \subset E \times G_d(E)$ such that

$$\mu((E \times G_d(E)) \backslash C) = 0\,.$$

3.2.2 Non-oriented shapes as varifolds

Rectifiable varifolds

In section 1.3.4, we have seen that oriented rectifiable sets of dimension d are canonically represented as d-currents. In the case of non-oriented shapes, the right notion is precisely the one of varifolds. Indeed, let X be a non-oriented rectifiable set of E of dimension d. To X, one can associate a varifold μ_X, which, in the measure point of view, is given by : $\mu_X(A) = \mathcal{H}^d(\{x \in X \mid (x, T_x X) \in A\})$ for all Borel subset $A \subset E \times G_d(E)$. Now, seen as a continuous linear form on $C_0(E \times G_d(E))$, μ_X writes, for all $\omega \in C_0(E \times G_d(E))$:

$$\mu_X(\omega) = \int_{E \times G_d(E)} \omega(x, V) d\mu_X(x, V)$$

and since a rectifiable subset of E has a tangent space application $x \mapsto T_x X$ defined for $\mathcal{H}^d$-almost all $x \in X$ which is also $\mathcal{H}^d$ measurable (cf section 1.17), we can write :

$$\mu_X(\omega) = \int_X \omega(x, T_x X) d\mathcal{H}^d(x)\,. \tag{3.4}$$

We will call such special varifolds *rectifiable varifolds* (note that we do not consider, in our definition, multiplicities or density functions as in [4]). This last integral can be also written with a parametrization if X is more simply a smooth compact submanifold of E given by a parametrization $\gamma : U \to E$ with U an open subset of $\mathbb{R}^d$. Then for all $\omega \in C_0(E \times G_d(E))$

$$\mu_X(\omega) = \int_U \omega(\gamma(u), T_{\gamma(u)}X)|\gamma'(u)|du \tag{3.5}$$

with the notation $\gamma'(u) = \Lambda_{i=1}^d \partial\gamma/\partial u_i \in \Lambda^d(E)$, $|\gamma'(u)|$ representing the local d-volume element. Note that the integral of eq.(3.5) is, as expected, independent of **any reparametrization** of X, positively or negatively oriented.

Polygonal meshes

Now, in discrete geometry, shapes are given as polygonal sets of points which are also encompassed in the category of rectifiable subsets and therefore representable as varifolds. In the same spirit as with currents and functional currents, any mesh set can be coded as a finite sum of dirac varifolds of the form $\sum_{i=1}^n m_i.\delta_{(p_i,V_i)}$. Again, n is the number of cells of the mesh, $p_i \in E$ are the centers of each cell, $V_i \in G_d(E)$ the unoriented tangent space to the shape at point p_i and $m_i \in \mathbb{R}_+^*$ the volume of the i-th cell. More specifically, let's examine the two most usual examples in practice.
First, consider a curve X given as a set of points $\{x_k\}_{k=1..N}$ and a set of edges $(f_i^1, f_i^2) \in \{1, .., N\}^2$ for $i = 1, .., n$. Then, with the previous conventions :

$$\begin{cases} p_i = \dfrac{x_{f_i^1} + x_{f_i^2}}{2} \\ \\ V_i = (x_{f_i^1} x_{f_i^2}) \\ \\ m_i = \|\overrightarrow{x_{f_i^1} x_{f_i^2}}\| \, . \end{cases} \tag{3.6}$$

Let us precise that here $(x_{f_i^1} x_{f_i^2})$ denotes the line in $G_1(\mathbb{R}^2)$ that is spanned by the vector $\overrightarrow{x_{f_i^1} x_{f_i^2}}$. Note also that all the previous equations remain unchanged if for any i we take (f_i^2, f_i^1) instead of (f_i^1, f_i^2) as a face, which is consistent with the idea of non-orientation.
In a similar way, if X is now for instance a triangulated surface in $E = \mathbb{R}^3$ given by a set of points $\{x_k\}_{k=1..N}$ with a set of n triangles $(f_i^1, f_i^2, f_i^3) \in \{1, .., N\}^3$ for $i = 1, .., n$ then we get :

$$\begin{cases} p_i = \dfrac{x_{f_i^1} + x_{f_i^2} + x_{f_i^3}}{3} \\ \\ V_i = \left[\overrightarrow{x_{f_i^1} x_{f_i^2}}, \overrightarrow{x_{f_i^1} x_{f_i^3}}\right] \\ \\ m_i = \dfrac{1}{2}.\|\overrightarrow{x_{f_i^1} x_{f_i^2}} \wedge \overrightarrow{x_{f_i^1} x_{f_i^3}}\| \, . \end{cases} \tag{3.7}$$

Here $\left[\overrightarrow{x_{f_i^1} x_{f_i^2}}, \overrightarrow{x_{f_i^1} x_{f_i^3}}\right]$ denotes the 2−dimensional space generated by $\overrightarrow{x_{f_i^1} x_{f_i^2}}$ and $\overrightarrow{x_{f_i^1} x_{f_i^3}}$. Again, all these equations are not dependent on the orientation given to each triangle.

Action of diffeomorphisms

A last important point for the following is to be able to express the transport of varifolds by diffeomorphism in a way that is compatible with the transport of shapes. Let's fix a varifold $\mu \in C_0(E \times G_d(E))'$ and $\phi \in \mathrm{Diff}(E)$, we define the transport $\phi_*\mu$ of μ by ϕ by pull-back and push-forward operations :

$$\forall \omega \in C_0(E \times G_d(E)),\ (\phi_*\mu)(\omega) = \mu(\phi^*\omega) \tag{3.8}$$

where $\phi^*\omega$ is the pull-back of ω by ϕ. For any $x \in E$ and $V \in G_d(E)$ a d-dimensional subspace with orthonormal basis $(u_1, ..., u_d)$, $\phi^*\omega$ is defined by the relation :

$$(\phi^*\omega)(x, V) = |d_x\phi(u_1) \wedge ... \wedge d_x\phi(u_d)|\ \omega(\phi(x), d_x\phi.V)\,. \tag{3.9}$$

In the last equation, $d_x\phi.V$ denotes the element of $G_d(E)$ that is the image of V by $d_x\phi$, the term $|d_x\phi(u_1) \wedge ... \wedge d_x\phi(u_d)|$ is the d-dimensional Jacobian of ϕ on the subspace V and represents the local change of d-dimensional volume of the transformation. With equations (3.8) and (3.9), it is straightforward to compute the transport of a dirac $\delta_{(x,V)}$:

$$\phi_*\delta_{(x,V)} = |d_x\phi(u_1) \wedge ... \wedge d_x\phi(u_d)|\ .\delta_{(\phi(x),d_x\phi.V)} \tag{3.10}$$

if $(u_1, ..., u_d)$ is an orthonormal basis of V. Now, with the previous definitions, we can show that the varifold representation of a rectifiable set commutes with the transport by diffeomorphism in the sense given by the following proposition :

Proposition 3.2.3. *If X is a d-dimensional rectifiable subset of E and $\phi \in$ Diff(E), then :*

$$\phi_*\mu_X = \mu_{\phi(X)}\,.$$

Proof. If X is a rectifiable subset of E, with eq.(3.4), we have for all ω :

$$\mu_X(\omega) = \int_X \omega(x, T_xX) d\mathcal{H}^d(x)$$

and thus

$$\begin{aligned}
\phi_*\mu_X(\omega) &= \mu_X(\phi^*\omega) \\
&= \int_X (\phi^*\omega)(x, T_xX) d\mathcal{H}^d(x) \\
&= \int_X \omega(\phi(x), d_x\phi(T_xX)) |d_x\phi.T_xX| d\mathcal{H}^d(x)
\end{aligned}$$

where $|d_x\phi.T_xX|$ is the d-dimensional Jacobian of ϕ along the tangent space T_xX, as previously. For almost all $x \in X$, $d_x\phi(T_xX)$ is a tangent space of $\phi(X)$ and by the generalization of the change of variables for rectifiable subsets (corollary 3.2.20 in [37]), we obtain :

$$\phi_*\mu_X(\omega) = \int_{\phi(X)} \omega(y, T_y\phi(X)) d\mathcal{H}^d(y)\,.$$

Thus, for any ω, $\phi_*\mu_X(\omega) = \mu_{\phi(X)}(\omega)$ and we have proved proposition 3.2.3. □

To conclude this section, varifolds offer a possible framework to represent non-oriented shapes both from the continuous setting and in the computational cases of meshed curves, surfaces... Up to this point, we have shown how to represent computationally a meshed shape as a finite set of diracs, each of them carrying a local information of position, unoriented tangent space and local volume and derive the equations of varifolds' transport by deformation. The remaining issue to examine is the question of the distance in the space of varifolds, which we address in the following.

3.3 Kernel metrics on varifolds

The use of reproducing kernels is a fundamental step when working with currents because it provides regularized metrics for the comparison of shapes which has the additional advantage of having an underlying Hilbert space structure. We have also mentioned several times that kernels are particularly well-fitted to compute distances between discretized shapes because of the simple expression of dot product between two diracs. All these reasons motivate a similar approach in the treatment of unoriented shapes through the varifold setting that has been presented. It's worth mentioning that other approaches could be possible, by working with more general Riemannian metrics on distribution of tangent spaces, generalizing for instance what is done for the Grassmann manifold in [1]. We argue however that RKHS metrics are very convenient from a computational point of view, especially in our applications to shape matching and analysis. In this section, we propose a generic and simple way to build relevant reproducing kernels on the space of varifolds, by basically making tensor products of kernels on E and on $G_d(E)$.

3.3.1 Kernels on the Grassmann manifold

The first step is to work on the Grassmannian. Since it is a rather abstract manifold, the construction of kernels is not a trivial problem and has been the subject of several recent works related to machine learning, as for instance in [46, 82]. We propose an alternative way that exploits the different embeddings of $G_d(E)$ into other spaces in which kernels are more easily defined. We will essentially consider two possibilities : a construction based on the identification with orthogonal projectors and a second one using the Plücker embedding.

Embedding in $\mathcal{L}(E)$

As explained in section 3.2, one can identify any $V \in G_d(E)$ with the orthogonal projection $p_V \in \mathcal{L}(E)$. As a consequence, one can induce straightforwardly a kernel on $G_d(E)$ by restriction of a positive kernel defined on $\mathcal{L}(E)$. Since $\mathcal{L}(E)$ is a finite-dimensional vector space, there are no difficulties in defining such kernels. Typical choices are given by :

$$k_P(V,W) = \langle p_V, p_W \rangle^k \tag{3.11}$$

with $k \in \mathbb{N}^*$ and $\langle .,. \rangle$ the usual Frobenius metric in $\mathcal{L}(E)$. More generally, any function $P(\langle p_V, p_W \rangle)$, P being a polynomial with positive coefficients, is a positive kernel on $\mathcal{L}(E)$. We shall call these **polynomial kernels** on the Grassmann manifold and **linear kernel** the one obtained for $k = 1$ in eq.(3.11), i.e :

$$k_L(V,W) = \langle p_V, p_W \rangle \tag{3.12}$$

Other important possibilities are the kernels induced by **Gaussians** in $\mathcal{L}(E)$, namely :

$$k_G(V,W) = e^{-\frac{|p_V - p_W|^2}{\sigma^2}} \tag{3.13}$$

Such kernels allow the comparison of subspaces with respect to a certain scale given by parameter σ. Many other could be defined with this method.
It's also possible to express these kernels using the principle angles introduced in section 3.2.1. If we denote $(v_1,..,v_d)$ and $(w_1,..,w_d)$ the orthonormal frames of V and W such that $\cos(\theta_i) = \langle v_i, w_i\rangle$ and $\text{Span}(v_i,w_i)\perp\text{Span}(v_j,w_j)$ for $j \neq i$ then we can write :

$$\begin{aligned}\langle p_V, p_W\rangle &= \langle \sum_{i=1}^{d} v_i^* \otimes v_i, \sum_{j=1}^{d} w_j^* \otimes w_j\rangle \\ &= \sum_{i,j=1}^{d} \langle v_i, w_j\rangle^2 \\ &= \sum_{i=1}^{d} \cos^2(\theta_i)\end{aligned}$$

It results that the kernel of eq.(3.11) has the following expression :

$$k_P(V,W) = \left(\sum_{i=1}^{d} \cos^2(\theta_i)\right)^k \tag{3.14}$$

The Gaussian kernel also has the alternative expression :

$$k_G(V,W) = e^{-\frac{2}{\sigma^2}\cdot\sum_{i=1}^{d}(1-\cos^2(\theta_i))} = \prod_{i=1}^{d} e^{-\frac{2}{\sigma^2}\sin^2(\theta_i)} \tag{3.15}$$

This formulation with principle angles offers, in our sense, a nicely interpretable way to understand the behavior of such kernels with respect to the subspaces' relative position. In particular, one can better understand the meaning of the width σ for the Gaussian kernel. The general construction we have proposed thus provides a wide variety of induced kernels on $G_d(E)$ that are effectively computable (either by the expression of the projection matrix or by computing the principle angles).

Embedding in $P(\Lambda^d E)$

Following a similar path, kernels on the Grassmannian could be also defined based on the Plücker embedding mentioned in 3.2.1, that embeds $G_d(E)$ into the projective space $P(\Lambda^d E)$. The Plücker application associates to any $V = \text{Span}\{v_1,..,v_d\}$ the line in the exterior product generated by $\xi_V = v_1 \wedge ... \wedge v_d$ which we denoted by $[\xi_V]$. The idea would be again to induce a kernel on $G_d(E)$ from a kernel on $P(\Lambda^d E)$. This is however not as trivial as previously since $P(\Lambda^d E)$ is still not a vector space. But we can build kernels on this projective space based on the following proposition :

Proposition 3.3.1. *Let k_0 be a real positive kernel on the vector space $\Lambda^d E$ such that for all $\xi, \xi' \in \Lambda^d E$, $k_0(-\xi,-\xi') = k_0(\xi,\xi')$. Then the application :*

$$k([\xi],[\xi']) = \frac{1}{2}\left[k_0\left(\frac{\xi}{|\xi|},\frac{\xi'}{|\xi'|}\right) + k_0\left(\frac{\xi}{|\xi|},-\frac{\xi'}{|\xi'|}\right)\right] \tag{3.16}$$

defines a real positive kernel on $P(\Lambda^d E)$.

Proof. The first thing to notice is that the application of eq.(3.16) is well-defined on $P(\Lambda^d E)$, since it does not depend on the representants ξ and ξ' in the equivalence classes $[\xi], [\xi']$. The next point is that the application $k_0(\frac{\xi}{|\xi|}, \frac{\xi'}{|\xi'|})$ is by restriction a positive kernel on the unit sphere of the vector space $\Lambda^d E$. Let's call H the corresponding RKHS. H is a Hilbert space of functions on the sphere and we remind that, by definition of the RKHS,

$$k_0\left(\frac{\xi}{|\xi|}, \frac{\xi'}{|\xi'|}\right) = k_0\left(\frac{\xi'}{|\xi'|}, \frac{\xi}{|\xi|}\right) = \left\langle k_0\left(\frac{\xi}{|\xi|}, .\right), k_0\left(\frac{\xi'}{|\xi'|}, .\right)\right\rangle_H .$$

Now for any $p \in \mathbb{N}^*$ and any family $(\xi_i)_{i=1,..,p} \in \Lambda^d E$ and $(\alpha_i)_{i=1,..,p} \in \mathbb{R}$, we have :

$$\begin{aligned}
&\sum_{i,j=1}^{p} k([\xi_i],[\xi_j])\alpha_i\alpha_j \\
&= \frac{1}{2}\sum_{i,j=1}^{p}\left[k_0\left(\frac{\xi_i}{|\xi_i|}, \frac{\xi_j}{|\xi_j|}\right) + k_0\left(\frac{\xi_i}{|\xi_i|}, -\frac{\xi_j}{|\xi_j|}\right)\right]\alpha_i\alpha_j \\
&= \frac{1}{4}\sum_{i,j=1}^{p}\left[k_0\left(\frac{\xi_i}{|\xi_i|}, \frac{\xi_j}{|\xi_j|}\right) + k_0\left(-\frac{\xi_i}{|\xi_i|}, \frac{\xi_j}{|\xi_j|}\right) + k_0\left(\frac{\xi_i}{|\xi_i|}, -\frac{\xi_j}{|\xi_j|}\right) + k_0\left(-\frac{\xi_i}{|\xi_i|}, -\frac{\xi_j}{|\xi_j|}\right)\right]\alpha_i\alpha_j \\
&= \frac{1}{4}\sum_{i,j=1}^{p}\left\langle k_0\left(\frac{\xi_i}{|\xi_i|}, .\right) + k_0\left(-\frac{\xi_i}{|\xi_i|}, .\right), k_0\left(\frac{\xi_j}{|\xi_j|}, .\right) + k_0\left(-\frac{\xi_j}{|\xi_j|}, .\right)\right\rangle_H \alpha_i\alpha_j \\
&= \frac{1}{4}\left\langle \sum_{i=1}^{p}\alpha_i\left[k_0\left(\frac{\xi_i}{|\xi_i|}, .\right) + k_0\left(-\frac{\xi_i}{|\xi_i|}, .\right)\right], \alpha_j\left[\sum_{j=1}^{p} k_0\left(\frac{\xi_j}{|\xi_j|}, .\right) + k_0\left(-\frac{\xi_j}{|\xi_j|}, .\right)\right]\right\rangle_H \\
&= \left\|\sum_{i=1}^{p}\alpha_i\left[k_0\left(\frac{\xi_i}{|\xi_i|}, .\right) + k_0\left(-\frac{\xi_i}{|\xi_i|}, .\right)\right]\right\|_H^2 .
\end{aligned}$$

The last expression is thus positive. □

Given a kernel as in proposition 3.3.1, we obtain by induction a kernel on the Grassmann manifold. This gives again a lot of possibilities. Among them, one could take $k_0(\xi, \xi') = \langle \xi, \xi' \rangle^{2k}$ for $k \in \mathbb{N}^*$, which satisfies the hypotheses of the proposition and get a kernel on $G_d(E)$ that writes :

$$k_{CB}(V, W) = \left\langle \frac{\xi_V}{|\xi_V|}, \frac{\xi_W}{|\xi_W|} \right\rangle^{2k} \tag{3.17}$$

Another interesting possibility is to take $k_0(\xi, \xi') = e^{\frac{-|\xi/|\xi| - \xi'/|\xi'||^2}{\sigma^2}} + e^{\frac{-|\xi/|\xi| + \xi'/|\xi'||^2}{\sigma^2}}$ and we obtain :

$$k_{G2}(V, W) = e^{-\frac{2}{\sigma^2}\left(1 - \left\langle \frac{\xi_V}{|\xi_V|}, \frac{\xi_W}{|\xi_W|}\right\rangle\right)} + e^{-\frac{2}{\sigma^2}\left(1 + \left\langle \frac{\xi_V}{|\xi_V|}, \frac{\xi_W}{|\xi_W|}\right\rangle\right)} \tag{3.18}$$

Again, these expressions can be written as functions of the principle angles between the subspaces. Introducing, as in the previous subsection, the orthonormal bases $v_1, ..., v_d$ and $w_1, ..., w_d$ of V and W such that $\cos(\theta_i) = \langle v_i, w_i \rangle$ and $\mathrm{Span}(v_i, w_i) \perp \mathrm{Span}(v_j, w_j)$ for $j \neq i$, we have :

$$\langle \wedge_{i=1}^d v_i, \wedge_{j=1}^d w_j \rangle = \det(\langle v_i, w_j \rangle)_{i,j} = \prod_{i=1}^{d} \cos(\theta_i)$$

and it results that :

$$k_{CB}(V,W) = \prod_{i=1}^{d} \cos^{2k}(\theta_i) \tag{3.19}$$

and

$$k_{G2}(V,W) = e^{-\frac{2}{\sigma^2}\left(1-\prod_{i=1}^{d}\cos(\theta_i)\right)} + e^{-\frac{2}{\sigma^2}\left(1+\prod_{i=1}^{d}\cos(\theta_i)\right)} \tag{3.20}$$

Note that the kernel k_{CB} of eq.(3.19) gives, for $k = 1$, the classical *Cauchy-Binet* kernel on the Grassmannian, studied for instance in [46].

Which kernels to choose ?

As we see, despite not having defined kernels directly on $G_d(E)$ as an abstract manifold, the two approaches we presented provide a very wide variety of possibilities that depend obviously on the chosen embedding of the Grassmann manifold. From a purely theoretical point of view, it's not easy to answer yet on which one of these embeddings should be favored.
For high dimension and codimension, we can state that the dimension of the space $P(\Lambda^d E)$ grows much faster than $\mathcal{L}(E)$ with d and n and so the Grassmannian represent a very small dimensional submanifold in this big space. A side effect possibly resulting from it, if we look at the expression of the kernel of eq.(3.19), is that two subspaces V and W are orthogonal for the corresponding metric as soon as **only one** of their principle angle is a right angle. We could argue that, in such situations, this kernel is somehow too 'concentrated' and decays too fast, if we compare it for instance to the kernel k_P of eq.(3.14) obtained with the embedding into $\mathcal{L}(E)$. This kind of behavior in higher dimensions was observed in classifications' experiments based on such kernels in [46].
There is a second argument pleading rather in favor of the first approach, that shall appear more clearly in the next sections when examining the C_0-universality property of the kernels. Indeed, the definition by quotient with the projective space makes such properties, at least from our experience, significantly more difficult to deal with.
In a more practical point of view, the cases we shall be looking at in this manuscript are focused on curves and surfaces in $\mathbb{R}^2$ or $\mathbb{R}^3$ and the Grassmann manifolds involved are essentially $G_1(\mathbb{R}^2)$ and $G_1(\mathbb{R}^3) = G_2(\mathbb{R}^3)$ (with the identification formulated in the remark of 3.2.1). Thus there is only one principle angle to consider in all these cases and the kernels k_P and k_{CB} for instance are identical. In the applications presented in this chapter, we will use the kernels k_L and k_G.

3.3.2 Kernels on varifolds

Tensor products' construction

The issue of kernels on $G_d(E)$ being addressed, we now move to the case of varifolds. Since we are considering functions defined on a product space, a natural way to build appropriate kernels is the tensor product trick. The result of lemma 2.2.1 in chapter 2 shows that the tensor product of a real-valued kernel on a set A and real-valued kernel on a set B gives a real-valued positive kernel on $A \times B$. In the case of varifolds, the following holds :

Proposition 3.3.2. *Assume that we are given a positive real kernel k_e on the space E such that k_e is continuous, bounded and for all $x \in E$, the function $k_e(x,.)$ vanishes at infinity. Assume that a second kernel k_t is defined on the manifold $G_d(E)$ and*

is also continuous. Then the RKHS W associated to the positive kernel $k_e \otimes k_t$ is continuously embedded into the space $C_0(E \times G_d(E))$.

Proof. By definition :

$$k\left((x,V),(y,\tilde{V})\right) = k_e(x,y)\, k_t(V,\tilde{V}) \tag{3.21}$$

and so, thanks to the assumptions on the kernels, $k((x,V),.)$ is continuous on $E \times G_d(E)$ and belongs to $C_0(E \times G_d(E))$. The vector space W_0 generated by these functions is thus included in $C_0(E \times G_d(E))$. Moreover, if $\omega \in W_0$:

$$\omega(x,V) = \delta_{(x,V)}(\omega) = \langle k((x,V),.),\omega\rangle_W$$

With Cauchy-Schwarz inequality : $|\omega(x,V)| \leq \|k((x,V),.)\|_W.\|\omega\|_W$. In addition, $\|k((x,V),.)\|_W = \sqrt{k((x,V),(x,V))}$ and both kernels k_e and k_t are bounded so that k is also bounded. We conclude that $|\omega|_\infty \leq \sqrt{|k|_\infty}.\|\omega\|_W$. Thus Cauchy sequences in W_0 for the W-norm are also Cauchy sequence for the infinite norm. It results that all their limits belong to $C_0(E \times G_d(E))$ and therefore W is included in $C_0(E \times G_d(E))$. The previous inequality then holds for all $\omega \in W$, which shows that the inclusion embedding $\imath : W \hookrightarrow C_0(E \times G_d(E))$ is indeed continuous. □

Consequently, there exists a continuous mapping i^* of the space of varifolds $C_0(E \times G_d(E))'$ into the dual of W. Just as for currents, if we introduce the isometry $K_W : W' \to W$ defined by $\langle K_W\mu,\omega\rangle_W = \mu(\omega)$ for all $\mu \in W'$ and $\omega \in W$. By the reproducing kernel property, we know that $K_W\delta_{(x,V)} = k((x,V),.)$. Then for all $x_1, x_2 \in E$ and $V_1, V_2 \in G_d(E)$,

$$\begin{aligned}\langle \delta_{(x_1,V_1)}, \delta_{(x_2,V_2)}\rangle_{W'} &= \langle K_W\delta_{(x_1,V_1)}, K_W\delta_{(x_2,V_2)}\rangle_W \\ &= K_W\delta_{(x_2,V_2)}(x_1,V_1)\end{aligned}$$

and thus we have the following expression for the inner product between two Diracs :

$$\langle \delta_{(x_1,V_1)}, \delta_{(x_2,V_2)}\rangle_{W'} = k_e(x_1,x_2)\, k_t(V_1,V_2)\,. \tag{3.22}$$

For any varifolds μ and μ' in $C_0(E \times G_d(E))'$, the inner product between them writes :

$$\begin{aligned}\langle \mu,\mu'\rangle_{W'} &= \langle K_W\mu, K_W\mu'\rangle_W \\ &= \mu(K_W\mu') \\ &= \int_{E\times G_d(E)} K_W\mu'(x,V)d\mu(x,V) \\ &= \int_{E\times G_d(E)} \langle K_W\mu', K_W\delta_{(x,V)}\rangle_W d\mu(x,V) \\ &= \int_{E\times G_d(E)} \left(\int_{E\times G_d(E)} K_W\delta_{(x,V)}(x',V')d\mu'(x',V')\right) d\mu(x,V)\end{aligned}$$

and gives finally :

$$\langle \mu,\mu'\rangle_{W'} = \iint_{(E\times G_d(E))^2} k_e(x,x')\, k_t(V,V')d\mu(x,V)d\mu'(x',V')\,. \tag{3.23}$$

In the particular case of rectifiable varifolds μ_X and μ_Y, we even obtain thanks to eq.(3.23) and (3.4) :

$$\langle \mu_X, \mu_Y \rangle_{W'} = \int_X \int_Y k_e(x,y)\, k_t(T_xX, T_yY) d\mathcal{H}^d(x) d\mathcal{H}^d(y)\,. \tag{3.24}$$

We have therefore a generic way to build kernels for varifolds which are *separable* since such kernels are tensor products of a kernel on the ambient space E with a kernel on the set of all tangent spaces $G_d(E)$. Building kernels on the euclidean space E raises no difficulties (examples were already given in chapter 1) and we have presented a way to build kernels on $G_d(E)$.

Dual embedding and C_0-universality

However, at this stage, the metric on the space of varifolds remains only a pseudo-distance because it is still unclear whether the dual application $i^* : C_0(E \times G_d(E))' \to W'$ is always an embedding. As we already evoked in chapter 2 section 2.2.2, in general, this is actually not the case : although $W \hookrightarrow C_0(E \times G_d(E))$ is an embedding, the dual application need not be injective. For instance, choosing the linear kernel k_L of eq.(3.12) on the Grassmannian makes i^* not injective. Indeed, let's place ourselves in the case $E = \mathbb{R}^2$ with an orthonormal basis $(e_1, e_1^\perp)$ and $(e_\theta, e_\theta^\perp)$ the rotated vectors by a certain angle θ. Let x be a fixed point in E. For any kernel $k_e \otimes k_L$, with k_e a kernel on E, we have :

$$\begin{aligned}
&\|(\delta_{(x,e_1)} + \delta_{(x,e_1^\perp)}) - (\delta_{(x,e_\theta)} + \delta_{(x,e_\theta^\perp)})\|^2_{W'} \\
&= \|(\delta_{(x,e_1)} + \delta_{(x,e_1^\perp)})\|^2_{W'} + \|(\delta_{(x,e_\theta)} + \delta_{(x,e_\theta^\perp)})\|^2_{W'} \\
&\quad - 2\langle (\delta_{(x,e_1)} + \delta_{(x,e_1^\perp)}), (\delta_{(x,e_\theta)} + \delta_{(x,e_\theta^\perp)}) \rangle_{W'} \\
&= k_e(x,x)\left[4 - 2\langle p_{e_1}, p_{e_\theta}\rangle - 2\langle p_{e_1^\perp}, p_{e_\theta}\rangle - 2\langle p_{e_1}, p_{e_\theta^\perp}\rangle - 2\langle p_{e_1^\perp}, p_{e_\theta^\perp}\rangle\right] \\
&= k_e(x,x)\left[4 - 2\cos^2(\theta) - 2\sin^2(\theta) - 2\sin^2(\theta) - 2\cos^2(\theta)\right] \\
&= 0
\end{aligned}$$

An therefore $i^*(\delta_{(x,e_1)} + \delta_{(x,e_1^\perp)}) = i^*(\delta_{(x,e_\theta)} + \delta_{(x,e_\theta^\perp)})$ for all θ.
The fact that i^* is an embedding is called the **C_0-universality** property of the kernel and, as proved in [20], it is equivalent to the property of W being dense in $C_0(E \times G_d(E))$. In the previous counterexample, the lack of injectivity basically comes from the presence of several directions of $G_d(E)$ located at one single point x. When representing submanifolds or reunion of submanifolds, it is not something that could happen except on isolated points. In fact, we can show result of 'weak injectivity' that applies for a very wide range of kernels. It is stated in the following :

Proposition 3.3.3. *Let $k = k_e \otimes k_t$ be a kernel as in proposition 3.3.2. Assume that kernel k_e is C_0-universal and that the kernel k_t is such that $k_t(V,V) > 0$ for all $V \in G_d(E)$. Let $X = \bigcup_{i=1}^N X_i$ and $Y = \bigcup_{j=1}^M Y_j$ be two finite union of compact d-dimensional submanifolds of E. If $\|\mu_X - \mu_Y\|_{W'} = 0$ then $X = Y$.*

Proof. We will denote by W_e and W_t the RKHS associated to kernels k_e and k_t. Let's start by the case where X is a single submanifold of E. If $\|\mu_X - \mu_Y\|_{W'} = 0$ and $X \neq Y$, one can find $x_0 \in \mathring{X} \backslash Y$ and there exists $r > 0$ such that $B(x_0, r) \cap Y = \emptyset$. Let $a \in C_0(E, \mathbb{R})$ be any continuous function such that the support of a is included

in $B(x_0, r)$. Let $\omega_t = k_t(T_{x_0}X, .) \in W_t$. For all functions $\omega_e \in W_e$, $\omega_e \otimes \omega_t \in W$ and thus $(\mu_X - \mu_Y)(\omega_e \otimes \omega_t) = 0$, i.e

$$\int_X \omega_e(x)\omega_t(T_xX)d\mathcal{H}^d(x) - \int_Y \omega_e(y)\omega_t(T_yY)d\mathcal{H}^d(y) = 0$$

Now, since W_e is dense in $C_0(E, \mathbb{R})$, we can approximate uniformly a by functions in W_e. By uniform convergence in the previous integral and the fact that $a = 0$ in Y, we get :

$$\int_X a(x)\omega_t(T_xX)d\mathcal{H}^d(x) = 0$$

This holds for all continuous functions a supported in $B(x_0, r)$, which is clearly impossible since $x \mapsto \omega_t(T_xX)$ is continuous and

$$\omega_t(T_{x_0}X) = k_t(T_{x_0}X, T_{x_0}X) \neq 0\,.$$

If X is now a finite reunion of compact submanifolds and $X \neq Y$ then, as previously, we can find $x_0 \in \bigcup_{i=1}^N \mathring{X}_i$ with $B(x_0, r_0) \cap Y = \emptyset$. $B(x_0, r_0) \cap X$ is itself a non-empty finite reunion of submanifolds. In order to get a similar proof as for the one submanifold case, we need to show the following : there exists a point $\hat{x} \in B(x_0, r_0) \cap X$, $\rho > 0$ with $B(\hat{x}, \rho) \subset B(x_0, r_0)$ and $j \in \{1, .., N\}$ such that for all function ω supported in $B(\hat{x}, \rho) \times G_d(E)$, $\mu_X(\omega) = \mu_{X_j}(\omega)$. This can be proved recursively.
Assuming the result for $N - 1$, we can find x_1, r_1,j_1 with $B(x_1, r_1) \subset B(x_0, r_0)$ and $\mu_{\bigcup_{i=1}^{N-1} X_i} = \mu_{X_{j_1}}$ for functions supported in $B(x_1, r_1)$. Now, there are two distinct cases : either $B(x_1, r_1) \cap X_{j_1} = B(x_1, r_1) \cap X_N$ or we can assume for instance that there exists $x_2 \in B(x_1, r_1) \cap X_{j_1}$, $r_2 > 0$ such that $B(x_2, r_2) \subset B(x_1, r_1)$ and $B(x_2, r_2) \cap X_N = \emptyset$. In the first case, since X_N coincides with X_{j_1} on $B(x_1, r_1)$, we have still $\mu_{\bigcup_{i=1}^{N} X_i} = \mu_{X_{j_1}}$ for functions supported in $B(x_1, r_1) \times G_d(E)$ and we take $\hat{x} = x_1$, $\rho = r_1$, $j = j_1$. In the second case, one can take $\hat{x} = x_2$, $\rho = r_2$, $j = j_1$ and the result holds as well. The rest of the proof is then exactly similar to the case of a unique submanifold. □

Even though there is no general injectivity for the dual application, proposition 3.3.3 ensures that the metrics we use are at least distances on the set of finite reunion of submanifolds, which is usually enough for the applications we aim at. We also believe that the result could be generalized to any rectifiable subsets of E (for which we have for $\mathcal{H}^d$-almost all x one tangent space direction) but the proof does not not seem to follow directly from the previous one : the passage from finite reunion to countable reunion is for instance not obvious.
In the specific case of the Gaussian kernel on $G_d(E)$ given in section 3.3.1, we can actually recover the injectivity of i^* on $C_0(E \times G_d(E))'$ itself, which is stated in the next proposition :

Proposition 3.3.4. *If $k = k_e \otimes k_G$ with k_e a C_0-universal kernel on E and k_G the restriction of a Gaussian kernel on $\mathcal{L}(E)$ given in eq.*(3.13)*, then k is a C_0-universal kernel on $E \times G_d(E)$.*

Proof. The proof is mainly based on two results about RKHS. The first one is the very well-known property that a Gaussian kernel on a finite dimensional vector space is C_0-universal. In our context, the Gaussian kernel K defined on $\mathcal{L}(E) \times \mathcal{L}(E)$ by

$K(l_1, l_2) = e^{-\frac{|l_1 - l_2|^2}{\sigma^2}}$ is thus C_0-universal (cf [20]). Now, the kernel k_G that we have defined in eq.(3.13) is the restriction of K to the subset of all orthogonal projections on d-dimensional subspaces (which is identified to $G_d(E)$). This subset is closed in $\mathcal{L}(E)$ and using property 2.2.2 from chapter 2 section 2.2.2, we deduce that k_G is then a C_0-universal kernel. If $W_e \subset C_0(E, \mathbb{R})$ and $W_t \subset C_0(G_d(E), \mathbb{R})$ are the RKHS corresponding to kernels k_e and k_G, it results that we have W_e dense in $C_0(E, \mathbb{R})$ (because k_e is assumed to be C_0-universal) and W_t dense in $C_0(G_d(E), \mathbb{R})$. It is then clear that W is dense in $C_0(E \times G_d(E), \mathbb{R})$. □

In summary, we have explained in this subsection how to define Hilbert metrics on varifolds that are computed from tensor products between kernels on E and kernels on the Grassmann manifold. This does not provide necessarily real distances on varifolds because of the possible non-injectivity of the application between the space of varifolds and the dual of the RKHS W. Yet, we have shown that for a very wide class of such kernels, the resulting distances are separating finite unions of submanifolds and we have argued in favor of Gaussian kernels on the Grassmann manifold for which we obtain the real injectivity and thus distances on the space of all varifolds. We will elaborate a little more on the interest of such Gaussian kernels in specific situations in the next section (cf figure 3.5).

3.4 Properties of RKHS norms on varifolds

3.4.1 Dependency on the scales of the kernels

One of the major advantage of kernel metrics is that they allow analysis at different scales. In the case of varifolds, this can be controlled both in the space E and the Grassmannian by the widths σ_e and σ_t of the Gaussians if one takes Gaussian kernels. The parameter σ_e can be thought as representing the typical interaction distance on the Diracs' positions in E whereas σ_t controls the desired correlation in the directions of the subspaces. An interesting fact is that for very large σ_e (basically all points seen at the same position by the kernel), the resulting metric on varifolds do not become necessarily trivial as with currents (cf the discussion of section 3.1.2). Indeed for $\sigma_e \to \infty$, the W'-distance becomes a distance between the distributions of tangent space directions depending on the kernel k_t. More specifically :

$$\begin{aligned} \|\mu_X\|^2_{W'} &\to \iint_{X \times X} k_t(T_x X, T_y X) d\mathcal{H}^d(x) d\mathcal{H}^d(y) \\ &= \iint_{G_d(E) \times G_d(E)} k_t(u, v) d\nu_X(u) d\nu_X(v) \end{aligned}$$

where $\nu_X \doteq \mathcal{H}^d \circ Tan^{-1}$ is the image measure of $\mathcal{H}^d$ by the application $Tan : X \to G_d(E)$, $x \mapsto T_x X$. In other words, at infinite scale for k_e, $\|\mu_X\|^2_{W'}$ represents a metric on the distribution of the non-oriented tangent spaces to the shape X, which we can see, in a certain way, as a histogram metric on $G_d(E)$. Therefore, even at infinite scale for k_e, something of the shape still remains in the varifold representation contrarily to currents. In those situations, the choice of the kernel on the Grassmann part of varifolds is quite decisive. In particular, the C_0-universality issue discussed previously becomes important. We show an illustration of such phenomenon in figure 3.5 for the linear kernel k_L which is not C_0-universal and the Gaussian kernel k_G, which was proved to be. The figure confirms what was already noticed at the beginning of

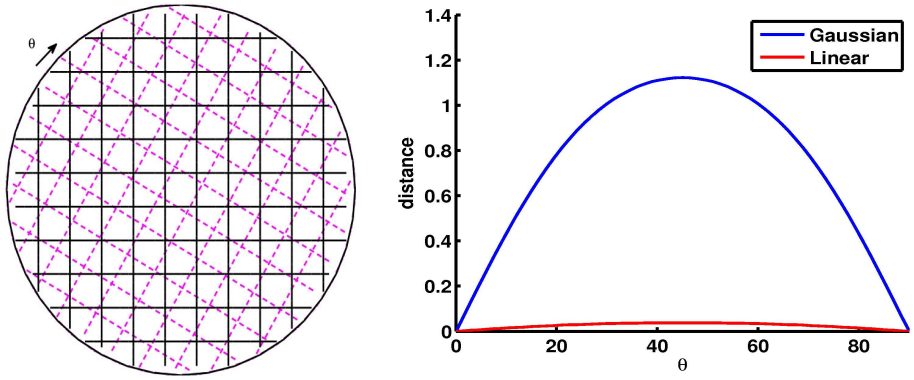


Figure 3.5: Computation of varifolds' distance between a 2D grid-like shape (in black) and its rotated version (in magenta) for various rotation angles and for both the linear kernel eq.(3.12) and the Gaussian kernel of eq.(3.13) on the Grassmann manifold, all in the case of large-scale kernel on the space E. The graph displays the ratio between the varifold distances and the norm of the original shape as a function of the rotation angle. With the linear kernel, the values remain very small for all angles, and it is nearly unable to distinguish both shapes whereas the Gaussian kernel shows an expected behavior with a maximum distance for a 45° angle.

section 3.3.2 about the linear kernel. The conclusion that we can be drawn in general is that Gaussian kernels on $G_d(E)$ provide metrics that are obviously more effective in situations of multiple directions crossing at nearby points, as might appear when treating fibers or tree-like structures.
On the other hand, if we let $\sigma_t \to \infty$, we see that $k_t(u,v) = 1$ for all $u, v \in G_d(E)$ and the RKHS norm we find only takes into account the position of the Diracs : these are exactly the measure distances between shapes introduced by Glaunès in [41] chap 4.

3.4.2 Control of the volume by W'-norm

We now come to the problem raised by our discussion related to figure 3.3. We show in which precise sense RKHS metrics on varifolds successfully avoid the cancellation phenomenon of currents. This is formulated in the theorem below and its corollary :

Theorem 3.4.1. *Let μ be a positive d-dimensional varifold of E such that $Supp(\mu) \subset B(0,1) \times G_d(E)$. Let k be a reproducing kernel on $E \times G_d(E)$ and W its RKHS, defined like in section 3.3.2 by $k((x,u),(y,v)) = k_e(x,y).k_t(u,v)$ where k_e is a reproducing kernel on E and k_t a reproducing kernel on the Grassmannian. We make the following additional assumptions :*
1. k_e is a radial scalar kernel so that we can write $k_e(x,y) = h_e(|x-y|)$. h_e is assumed to be a continuous positive function, with $h_e(0) > 0$.
2. k_t is a continuous and positive function on $G_d(E) \times G_d(E)$ such that for all $u \in G_d(E)$, $k_t(u,u) > 0$.
Then there exists a constant Cte independent of $\mu \in W'$ such that :

$$\|\mu\|_{W'} \geq Cte\, \mu(E \times G_d(E))\,.$$

Proof. For any varifold $\mu \in W'$, we have :

$$\|\mu\|^2_{W'} = \iint_{(E\times G_d(E))^2} k_e(x,y).k_t(u,v) d\mu(x,u) d\mu(y,v)\,. \tag{3.25}$$

For the proof of theorem 3.4.1, we shall first examine the case of a constant kernel for k_e.

Step 1: We first assume that $k_e(x,y) = 1$ for all x, y. Let's denote by $p : E \times G_d(E) \to G_d(E)$ the application $(x,u) \mapsto u$. We introduce the image measure $\nu \doteq \mu \circ p^{-1}$ defined on $G_d(E)$. Then :

$$\|\mu\|^2_{W'} = \iint_{G_d(E)\times G_d(E)} k_t(u,v) d\nu(u) d\nu(v)\,.$$

Note that $\nu(G_d(E)) = \mu(p^{-1}(G_d(E))) = \mu(E \times G_d E)$. Now, the compact group $SO(E)$ of direct isometries of E acts transitively on $G_d(E)$ by the relation $g.\mathrm{Span}(e_1,..,e_d) \doteq \mathrm{Span}(g(e_1),..,g(e_d))$, making $G_d(E)$ a homogeneous space. Since $SO(E)$ is a compact group, we can consider its unique bi-invariant Haar measure which we denote λ with the convention $\lambda(SO(E)) = 1$. If we fix a particular element $u_0 \in G_d(E)$, λ induces in turn a measure on $G_d(E)$ defined for all B by $\lambda_{Gr}(B) = \lambda(\{g \in SO(E) \mid g.u_0 \in B\})$. One can check that the right-invariance of λ implies that λ_{Gr} does not depend on the choice of $u_0 \in G_d(E)$. In the same way, thanks to the left-invariance of λ, λ_{Gr} is invariant by the action of G i.e $\lambda_{Gr}(g.B) = \lambda_{Gr}(B)$ for all g.

A second important element is that, as a homogeneous space under the action of a compact Lie group (on which there exists a bi-invariant metric), $G_d(E)$ can be equipped by projection with a left-invariant distance with respect to the action of $SO(E)$, which we will denote d. Let's now consider a continuous, positive L^1 (for the measure λ_{Gr}) function ϕ on $G_d(E)$. Since $\lambda_{Gr}(G_d(E)) = \lambda(SO(E)) = 1 < +\infty$, ϕ is also in L^2 and we assume that $\|\phi\|_2 = 1$. We can assume in addition that the support $\mathrm{Supp}(\phi)$ of ϕ is included in a certain ball of radius δ centered at u_0. We also introduce the regularization function ψ on $G_d(E) \times G_d(E)$ defined by :

$$\psi(u,v) = \int_{SO(E)} \phi(g.u)\phi(g.v) d\lambda(g)\,.$$

It's obvious that ψ is also a positive continuous function. Now, the definition and invariance properties of λ_{Gr} show that for all $u \in G_d(E)$, if $\tau_u : g \mapsto g.u$ then, $\lambda \circ \tau_u^{-1} = \lambda_{Gr}$ and it results that :

$$\int_{SO(E)} \phi(g.u)^2 d\lambda(g) = \int_{G_d(E)} \phi(v)^2 d\lambda_{Gr}(v) = \|\phi\|_2^2 = 1\,,$$

From Cauchy-Schwarz inequality, we have on the other hand

$$\begin{aligned}\forall (u,v) \in G_d(E)^2,\ \psi(u,v) &\leq \left(\int_{SO(E)} \phi(g.u)^2 d\lambda(g)\right)^{\frac{1}{2}} \left(\int_{SO(E)} \phi(g.v)^2 d\lambda(g)\right)^{\frac{1}{2}} \\ &\leq \|\phi\|_2^2\end{aligned}$$

therefore $\psi(u,v) \leq 1$. In addition, if $d(u,v) \geq 2\delta$, we have $d(g.u, g.v) \geq 2\delta$ for all g since d is left-invariant. Therefore $d(g.u,u_0) + d(g.v,u_0) \geq d(g.u,g.v) \geq 2\delta$. It results that either $d(g.u,u_0) \geq \delta$ or $d(g.v,u_0) \geq \delta$ and anyhow $\phi(g.u)\phi(g.v) = 0$. Consequently, ψ vanishes for $d(u,v) \geq 2\delta$. Now, since k_t is positive and $0 \leq \psi \leq 1$

$$\|\mu\|^2_{W'} \geq \iint_{G_d(E)\times G_d(E)} k_t(u,v)\psi(u,v) d\nu(u) d\nu(v)\,.$$

Moreover, k_t is also continuous and $k_t(u,u) > 0$ for all u so, by compactness of $G_d(E)$, there exists $\alpha > 0$ and an open domain $D \subset G_d(E) \times G_d(E)$ of the form $D = \{(u,v) \mid d(u,v) < \epsilon\}$ such that for all $u, v \in D$, $k_t(u,v) \geq \alpha$. We can also assume, by choosing appropriately the function ϕ, that $\delta \leq \frac{\epsilon}{2}$ and therefore $\psi(u,v) = 0$ outside D. It results that :

$$\|\mu\|^2_{W'} \geq \alpha \iint_{G_d(E)\times G_d(E)} \psi(u,v) d\nu(u) d\nu(v)\,. \tag{3.26}$$

Using Fubini's theorem in the last integral, we can write

$$\begin{aligned}
&\iint_{G_d(E)\times G_d(E)} \psi(u,v) d\nu(u) d\nu(v) \\
&\qquad = \int_{SO(E)} \left(\iint_{G_d(E)\times G_d(E)} \phi(g.u)\phi(g.v) d\nu(u) d\nu(v) \right) d\lambda(g) \\
&\qquad = \int_{SO(E)} \left(\int_{G_d(E)} \phi(g.u) d\nu(u) \right)^2 d\lambda(g) \\
&\qquad \geq \left(\int_{SO(E)} \int_{G_d(E)} \phi(g.u) d\nu(u) d\lambda(g) \right)^2
\end{aligned} \tag{3.27}$$

the last estimate resulting from Cauchy-Schwartz inequality and the fact that $\lambda(SO(E)) = 1$. Using again Fubini's theorem,

$$\int_{SO(E)} \int_{G_d(E)} \phi(g.u) d\nu(u) d\lambda(g) = \int_{G_d(E)} \left(\int_{SO(E)} \phi(g.u) d\lambda(g) \right) d\nu(u)\,.$$

Making the change of variable $\tau_u :\ g \mapsto g.u$ in the inside integral and using the same argument as previously, we obtain that

$$\int_{SO(E)} \phi(g.u) d\lambda(g) = \int_{G_d(E)} \phi(w) d\lambda_{Gr}(w) = \|\phi\|_1 > 0$$

Inserting the previous in equation (3.27),

$$\iint_{G_d(E)\times G_d(E)} \psi(u,v) d\nu(u) d\nu(v) \geq \|\phi\|_1^2 \nu(G_d(E))^2 = \|\phi\|_1^2 \mu(E \times G_d E)^2\,.$$

So that we eventually obtain :

$$\|\mu\|^2_{W'} \geq \alpha \|\phi\|_1^2 . \mu(E \times G_d E)^2$$

and since $\beta \doteq \alpha\|\phi\|_1^2$ is a constant that does not depend on μ, this concludes the proof in that case.

Step 2: We now move to the proof for a general kernel k_e. From the hypotheses, the support of μ is included in $B(0,1) \times G_d(E)$ and k_e is a continuous and positive radial scalar kernel on E. Thus, there exists a real number $\delta < 1$ such that for any $x, y \in E$ with $|x - y| \leq \delta$, we have $k_e(x,y) \geq h_e(0)/2 \doteq \kappa$. Let's now consider a partition of the space with a set of cubes $\{C_i\}_{i=1,..,M}$ of diameter smaller than δ such

that $X \subset \bigcup_{i=1}^{M} C_i$. Then the subsets $S_i \doteq (C_i \times G_d(E)) \cap \mathrm{Supp}(\mu)$ form a partition of $\mathrm{Supp}(\mu)$. Moreover,

$$\begin{aligned}
\|\mu\|_{W'}^2 &= \iint_{(E\times G_d(E))^2} k_e(x,y)\, k_t(u,v) d\mu(x,u) d\mu(y,v) \\
&\geq \sum_{i=1}^{M} \iint_{S_i\times S_i} k_e(x,y)\, k_t(u,v) d\mu(x,u) d\mu(y,v) \\
&\geq \kappa \sum_{i=1}^{M} \iint_{S_i\times S_i} k_t(u,v) d\mu(x,u) d\mu(y,v)\,.
\end{aligned}$$

We can apply the result of step 1 to each of these integrals, thus obtaining

$$\begin{aligned}
\|\mu\|_{W'}^2 &\geq \beta\kappa \sum_{i=1}^{M} \mu(S_i)^2 \\
&\geq \beta\kappa . \frac{1}{M} \left(\sum_{i=1}^{M} \mu(S_i) \right)^2 \\
&\geq \frac{\beta\kappa}{M} \mu(E\times G_d E)^2\,,
\end{aligned}$$

the second inequality being a discrete Cauchy-Schwartz and the last one resulting from the fact that $\{S_i\}$ is a partition of $\mathrm{Supp}(\mu)$. This ends the proof because the constants β and κ are both independent of μ, and so is M which only depends on the kernel k_e. □

The hypotheses of theorem 3.4.1 on the kernels are mostly technical but are not particularly restrictive in practice since one can check easily that all the examples of kernels given in section 3.3.1 comply to the requirements on k_t. As a direct corollary, we also have the following result :

Corollary 3.4.1. *Let X be a rectifiable subset of E included in the unit ball. We make the same hypotheses on the kernel as in theorem 3.4.1. Then there exists a constant Cte independent of X such that*

$$\|\mu_X\|_{W'} \geq Cte\, \mathcal{H}^d(X)\,.$$

Proof. This is essentially a special case of theorem 3.4.1 for rectifiable varifolds. Indeed, the support of μ_X is included in $B(0,1)\times G_d(E)$ and thanks to the theorem, $\|\mu_X\|_{W'} \geq \mathrm{Cte}\, \mu_X(E\times G_d(E))$. In addition, from the very definition of μ_X, $\mu_X(E\times G_d(E)) = \mathcal{H}^d(X)$. □

This result is theoretically essential because it shows that pathological cases as the one of figure 3.3 cannot happen when shapes are represented as varifolds. In practical applications, it ensures the consistency of the kernel norm we use with the actual volume of the shapes so that artificial elimination of mass during registration process or template estimation are not likely to occur in this setting. We will show more specific examples of this in the next section. Note that when X is of codimension one and is defined as the boundary of a bounded domain, the notion of volume of a shape X is the area of the boundary of the domain and not the volume of the domain. In particular, the corollary says that when a plain shape is represented by its boundary, its norm for generic varifold kernel is larger than the area of its boundary.

3.4.3 Metric variation formula

A last fundamental issue to mention is the question of the variation of kernel norms with respect to the geometrical support of shapes, since registration and template estimation algorithms rely on the computations of gradients of such norms. We shall detail the technical discrete computations of such gradients in the next section but it is also very valuable to have a theoretical and general interpretation of how the metric on shape varies when the support is deformed. Namely, if X is a shape (here we will assume compact oriented submanifold) and a RKHS W on varifolds is set, we want to express the variation with respect to X of terms like $\langle \mu_X, \mu' \rangle_{W'}$ with μ' a certain element of W'. Since $\langle \mu_X, \mu' \rangle_{W'} = \mu_X(K_W \mu')$ (cf section 3.3.2), we can look equivalently for variations of $\mu_X(\omega)$ for a fixed $\omega \in W$.
What we wish to derive is thus a result corresponding to Cartan's formula for currents. The proper formulation is to express the derivative of $\mu_X(\omega)$ for small variations of X obtained by flowing a vector field from X. With some hypotheses on X and W, it is possible to derive a formula that generalizes in our context the notion *first variation of a varifold* studied in [4]. The result is summed up below :

Theorem 3.4.2. *Let X be an orientable compact submanifold and μ_X its associated varifold. Let v be any C^1 vector field with compact support defined on E and consider the associated one-dimensional subgroup $t \mapsto \phi_t$ generated by the flow of v. Then, if $X_t \doteq \phi_t(X)$ is the transported manifold, we have for any C^1 function $(x, V) \mapsto \omega(x, V)$ on $E \times G_d(E)$:*

$$\left.\frac{d}{dt}\right|_{t=0} \mu_{X_t}(\omega) = \int_X \left(\frac{\partial \omega}{\partial x} - \mathrm{div}_X \left(\frac{\partial \omega}{\partial V} \right) - \omega H_X | v^\perp \right) + \int_{\partial X} \langle \nu, \omega v^\top + \left(\frac{\partial \omega}{\partial V} | v^\perp \right) \rangle$$

where $v^\top$ and $v^\perp$ denote the tangential and normal part of v along X, ν is the unit outward normal along ∂X, and H_X the mean curvature vector to X.

Proof. Even though the objective is similar to Cartan's formula for currents, the proof is quite different in the details. We assume that X is a smooth compact orientable submanifold of E so that we can consider the canonical volume form σ_X on X. Given a C^1 vector field v on E with compact support, we consider the 1-parameter group of diffeomorphisms ϕ_t with $\phi_0 = Id$ and $\partial_t|_{t=0} \phi_t = v$. The whole point of the theorem is to compute the variation of $\mu_X(\omega)$ when X is deformed by vector field v, that is to say :

$$\left.\frac{d}{dt}\right|_{t=0} \mu_{\phi_t(X)}(\omega) = \left.\frac{d}{dt}\right|_{t=0} \int_{X_t} \omega(x, T_x X_t) d\mathcal{H}^d(x) \tag{3.28}$$

where $X_t = \phi_t(X)$. Note that this generalizes the first variation of a varifold computed in [4], for which $\omega = \mathbb{1}$. Here, ω is any C^1 function on $E \times G_d(E)$. As already explained in section 3.2.2, we have $\mu_{\phi_t(X)}(\omega) = \mu_X(\phi_t^* \omega)$ and :

$$\left.\frac{d}{dt}\right|_{t=0} \mu_{\phi_t(X)}(\omega) = \int_X \left.\frac{d}{dt}\right|_{t=0} (\phi_t^* \omega)(x, T_x X) d\mathcal{H}^d(x) = \int_X \pounds_v \omega(x, T_x X) d\mathcal{H}^d(x) \tag{3.29}$$

where $\pounds_v \omega$ stands for the Lie derivative of $\omega(x, V)$ along v. Hereafter, to simplify the notation, for a function $(x, V) \mapsto f(x, V)$, we will simply write $\int_X f$ instead of $\int_X f(x, T_x X) d\mathcal{H}^d(x)$.
Now, $(\phi_t^* \omega)(x, T_x X) = |d_x \phi_t . T_x X| \omega(\phi_t(x), d_x \phi_t(T_x X))$. Let us recall that $J_t = |d_x \phi_t . T_x X|$ is the volume change in the direction of the tangent space $T_x X$. Taking

the derivative inside the integral at $t = 0$ leads to three terms : differentiate the function ω with respect to position, with respect to the tangent space direction and differentiate the volume change. The first term is the simplest one and an immediate computation shows that it equals $\left(\frac{\partial \omega}{\partial x}|v\right)$. The two others are more involved.

Derivative of the volume change : For any vector field u defined on X, we shall denote by $u^\top$ and $u^\perp$ the tangential and normal components of u with respect to the tangent space of X at each point. We also introduce the connection $\nabla_{\cdot}\cdot$ on the ambient space and an orthonormal frame of tangent vector fields $(e_i)_{i=1,..,d}$ on X. Now $J_t = \sqrt{\det(\langle d_x\phi_t(e_i), d_x\phi_t(e_j)\rangle)_{i,j}}$ so a simple calculation shows that :

$$\left.\frac{d}{dt}\right|_{t=0} J_t = \sum_{i=1}^{d} \langle e_i, \nabla_{e_i} v\rangle$$

Writing $v = v^\top + v^\perp$ provides a first term $\sum_{i=1}^{d}\langle e_i, \nabla_{e_i} v^\top\rangle$ which is the tangential divergence of the vector field $v^\top$ denoted usually $\mathrm{div}_X(v^\top)$. The second term becomes $\sum_{i=1}^{d}\langle e_i, \nabla_{e_i} v^\perp\rangle$. For all $i = 1, .., d$, we have $\langle e_i, v^\perp\rangle = 0$ so that after differentiation we find that $\langle e_i, \nabla_{e_i} v^\perp\rangle = -\langle \nabla_{e_i} e_i, v^\perp\rangle$. Therefore :

$$\begin{aligned}\sum_{i=1}^{d}\langle e_i, \nabla_{e_i} v^\perp\rangle &= -\sum_{i=1}^{d}\langle \nabla_{e_i} e_i, v^\perp\rangle \\ &= -\left\langle \left(\sum_{i=1}^{d} \nabla_{e_i} e_i\right)^\perp, v^\perp \right\rangle\end{aligned}$$

In this last expression, we recognize the **mean curvature vector** to the submanifold X, which is the trace of the Weingarten map and is denoted H_X. As a result, we find that :

$$\int_X \omega \left.\frac{d}{dt}\right|_{t=0} J_t = \int_X \omega \,\mathrm{div}_X(v^\top) - \int_X \omega\langle H_X, v^\perp\rangle$$

Now, we will show that the first term can be rewritten as a boundary integral. Indeed, if we denote by $\tilde{\omega}$ the function defined on X by $\tilde{\omega}(x) = \omega(x, T_xX)$, we have $\mathrm{div}_X(\tilde{\omega} v^\top) = \tilde{\omega}\,\mathrm{div}_X(v^\top) + \langle \nabla\tilde{\omega}|v^\top\rangle$. Applying the divergence theorem on the orientable manifold X gives :

$$\int_X \omega\,\mathrm{div}_X(v^\top) = -\int_X \langle \nabla\tilde{\omega}|v^\top\rangle + \int_{\partial X} \omega\langle \nu, v^\top\rangle$$

where ν is the unit outward normal to the boundary.

Derivative with respect to tangent spaces : We now come to the last term in equation (3.29), which is the variation of ω with respect to the tangent space in $G_d(E)$. As explained in section 3.2.1, we can identify the tangent space to $G_d(E)$ at V with the space of linear applications $\mathcal{L}(V, V^\perp)$. In addition, if we set $V_t = d_x\phi_t(T_xX)$, we have, thanks to eq.(3.3) :

$$\left.\frac{d}{dt}\right|_{t=0} V_t = p_{T_xX^\perp} \circ \nabla v|_{T_xX} \in \mathcal{L}(T_xX, (T_xX)^\perp)$$

which we will note more concisely $\left.\frac{d}{dt}\right|_{t=0} V_t = \nabla^\perp v$. The space $\mathcal{L}(T_xX, (T_xX)^\perp)$ being trivially isomorphic to $T_xX^\perp \otimes (T_xX)^*$, we can consider $\nabla^\perp v$ being an element

of that space and we introduce $\frac{\partial \omega}{\partial V}$ which we therefore identify to an element of $(T_xX^{\perp})^* \otimes T_xX$: $\frac{\partial \omega}{\partial V} = \sum_{j=d+1}^{n} \eta_j^* \otimes \alpha_j$ for $(\eta_{d+1}, .., \eta_n)$ an orthonormal frame of $T_xX^{\perp}$ and (α_j) vectors of T_xX (as usual η^* denotes the linear form $\langle \eta, . \rangle$). Then the variation we wish to compute is :

$$\left(\frac{\partial \omega}{\partial V}|\nabla^{\perp} v\right) = \left(\frac{\partial \omega}{\partial V}|\nabla v\right) = \sum_{j=d+1}^{n} \langle \eta_j, \nabla_{\alpha_j} v \rangle$$

If we introduce $\left(\frac{\partial \omega}{\partial V}|v\right) = \sum_{j=d+1}^{n} \eta_j^*(v)\alpha_j = \sum_{j=d+1}^{n} \langle \eta_j, v \rangle \alpha_j$ which is a tangent vector field on X, we have :

$$\mathrm{div}_X \left(\frac{\partial \omega}{\partial V}|v\right) = \sum_{i=1}^{d} \sum_{j=d+1}^{n} (\langle e_i, \nabla_{e_i} \alpha_j \rangle \langle \eta_j, v \rangle + \langle e_i, \langle \nabla_{e_i} \eta_j, v \rangle \alpha_j \rangle + \langle e_i, \langle \eta_j, \nabla_{e_i} v \rangle \alpha_j \rangle)$$

The last term in the sum is also $\sum_{j=d+1}^{n} \langle \eta_j, \nabla_{\alpha_j} v \rangle$, which is nothing else than $\left(\frac{\partial \omega}{\partial V}|\nabla v\right)$. As for the two other terms in the sum, it's easy to see that it equals :

$$\left(\sum_{i=1}^{d} \langle e_i, \nabla_{e_i} \sum_{j=d+1}^{n} \eta_j^* \otimes \alpha_j \rangle | v\right) = (\mathrm{div}_X \left(\frac{\partial \omega}{\partial V}\right)|v)$$

So get eventually that :

$$\left(\frac{\partial \omega}{\partial V}|\nabla v\right) = \mathrm{div}_X \left(\frac{\partial \omega}{\partial V}|v\right) - (\mathrm{div}_X \left(\frac{\partial \omega}{\partial V}\right)|v) \tag{3.30}$$

Integrating equation (3.30) over the submanifold X and using as previously the divergence theorem, we find that :

$$\int_X \left(\frac{\partial \omega}{\partial V}|\nabla v\right) = \int_{\partial X} \langle \nu, \left(\frac{\partial \omega}{\partial V}|v\right) \rangle - \int_X (\mathrm{div}_X \left(\frac{\partial \omega}{\partial V}\right)|v) \tag{3.31}$$

Synthesis : Grouping all the different terms obtained so far, we get the following :

$$\begin{aligned} \int_X \pounds_v \omega \;\; = \;\; & \int_X \left(\frac{\partial \omega}{\partial x} - \mathrm{div}_X \left(\frac{\partial \omega}{\partial V}\right)|v\right) - \int_X \langle \nabla \tilde{\omega} | v^{\top} \rangle + \omega \langle H_X | v^{\perp} \rangle \\ & + \int_{\partial X} \langle \nu, \left(\frac{\partial \omega}{\partial V}|v\right) + \omega v^{\top} \rangle \end{aligned}$$

We remind that $\tilde{\omega}(x) = \omega(x, T_xX)$ so $(\nabla \tilde{\omega}|v^{\top}) = \left(\frac{\partial \omega}{\partial x}|v^{\top}\right) + \left(\frac{\partial \omega}{\partial V}|\nabla v^{\top}\right)$ and applying the result of equation (3.30) to $v^{\top}$:

$$\left(\frac{\partial \omega}{\partial V}|\nabla v^{\top}\right) = \mathrm{div}_X \left(\frac{\partial \omega}{\partial V}|v^{\top}\right) - (\mathrm{div}_X \left(\frac{\partial \omega}{\partial V}\right)|v^{\top})$$

Noticing that $\left(\frac{\partial \omega}{\partial V}|v^{\top}\right) = 0$ by the expression of $\frac{\partial \omega}{\partial V}$ and using the equality $v = v^{\top} + v^{\perp}$ we find eventually that :

$$\int_X \pounds_v \omega = \int_X \left(\frac{\partial \omega}{\partial x} - \mathrm{div}_X \left(\frac{\partial \omega}{\partial V}\right) - \omega H_X | v^{\perp}\right) + \int_{\partial X} \langle \nu, \left(\frac{\partial \omega}{\partial V}|v\right) + \omega v^{\top} \rangle$$

which proves the result of theorem 3.4.2. □

We will just make a few qualitative comments since there are a few noteworthy consequences of theorem 3.4.2 to mention. The first important remark is that the variation of varifold metric is controlled only by the vector field v and not its derivatives. This is not straightforward since a varifold kernel $k((x,V),(x',V'))$ is encoding in an arbitrary non-linear way first order information through the inclusion of Grassmannian variables V and V'. Interestingly, the result is valid even if the dimension or the codimension of X is different from 1. In return, we see that the formula involves some terms on the **boundary** of X, one of them expressing the tangential extension of X along its boundary and a second one related to the variation in the tangent space direction on the boundary in the normal direction $v^\perp$. The presence of these boundary terms will appear of particular importance in chapter 4. In the interior of X, we see that the variation depends only on the orthogonal component $v^\perp$ of v, which, in other terms, shows that the gradient of the attachment distance is **orthogonal** to the shape. Note also that we have assumed X is oriented essentially because the proof makes use of the divergence theorem (that relies on Stokes theorem). Yet, since we work with varifolds, the integral does not depend on the orientation of X and so one could presumably remove this hypothesis by just considering partition of unity and local orientations.

3.5 Large deformation registration algorithm for unoriented shapes

At this point, we have defined a theoretical background to represent and compare unoriented shapes through varifolds and kernels on varifolds. What we have obtained is a distance (actually a whole class of distances provided by different kernels) between the objects that can serve as an attachment term in practically any matching process. As we have done so far in this thesis, we focus on the **LDDMM** model for large deformations, which was already exposed in the previous chapters. The purpose of the following is to derive the equations needed for the numerical implementation of LDDMM on varifolds and present a few examples of matchings based on this framework. In all this section, we will now restrict ourselves to the usual cases of curves and surfaces living in the 3-dimensional euclidean space $\mathbb{R}^3$.

3.5.1 Description of the algorithm

We have already presented LDDMM registration both in the context of currents in chapter 1 and of functional currents in chapter 2. In the case of varifolds, the principles are basically the same. We again consider a RKHS space V of admissible vector fields on E and flows of L^2 time-varying vector fields v equipped with the right-invariant metric $\int_0^1 |v_t|_V^2 dt$. In our situations, we shall focus on RKHS V that are generated by Gaussian kernels of certain scale σ_V or by sums of such kernels, as explained in more detail in chapter 1. The geodesics for that metric are parametrized by their initial momentum $p_0 \in V^*$. Then, considering two unoriented shapes X (the source) and Y (the target) and their representation μ_X and μ_Y in a RKHS of varifolds W', we can formulate similarly the registration problem between X and Y as the

optimization of :

$$\begin{cases} \min_{p_0} J(p_0) \doteq E(p_0) + \gamma \|\mu_{\phi_1(X)} - \mu_Y\|^2_{W'} \\ \\ \dfrac{\partial \phi_t}{\partial t} = v_t \circ \phi_t \end{cases} \tag{3.32}$$

Up to the change on the attachment term which is now a varifold kernel distance, this problem has an identical formulation with the geodesic shooting registration on finite number of points (cf introduction section 1.3.2). With the same notations, the gradient of J with respect to p_0 is therefore obtained again by first flowing the forward Hamiltonian equation which gives the point positions q_t and the momentum p_t at all times $t \in [0,1]$ and then by :

$$\nabla_{p_0} J = K_V(q,q)p_0 - P_0$$

with P_0 an adjoint variable obtained by flowing backward in time the coupled equations :

$$\begin{cases} \dot{Q}_t = -(\partial^2_{q,p} H_r(q_t,p_t))^* Q_t + (\partial^2_{q,q} H_r(q_t,p_t))^* P_t \\ \dot{P}_t = -(\partial^2_{p,p} H_r(q_t,p_t))^* Q_t + (\partial^2_{q,p} H_r(q_t,p_t))^* P_t \\ Q_1 = -\nabla_{q^1} A(q_1), \ P_1 = 0 \end{cases} \tag{3.33}$$

where $\nabla_{q_1} A(q_1)$ is basically the gradient of the attachment distance function between the deformed source shape $\phi_1(X)$ and the target Y. The only change with respect to previous algorithm on currents is the attachment quantity $A(q_1)$ for the computation of the total energy and its gradient $\nabla A(q_1)$ for the initialization of P in the backward equation. We detail those computations for the case of embedded curves and surfaces in $\mathbb{R}^3$ in the next subsection. The rest of the algorithm, similarly to LDDMM for landmarks or currents, is an adaptive step gradient descent scheme on p_0.

3.5.2 Discrete expressions for curves and surfaces

Let's now focus on the attachment distance term $A(q_1)$. In our context of unoriented shapes modeled as varifolds, we use distances provided by kernels and their associated RKHS presented in section 3.3. Those are the tensor product of a kernel k_e on the space $\mathbb{R}^3$ and a kernel k_t on the Grassmann manifold. If W is the RKHS corresponding to $k = k_e \otimes k_t$, the attachment distance we have is then :

$$A = \|\mu_{\phi_1(X)} - \mu_Y\|^2_{W'} \tag{3.34}$$

Since $\phi_1(X)$ is the transported source object by the vector field at time 1 (represented by the set of points q_1 with the previous notations), formally the problem reduces to express explicitly quantities like $\|\mu_X - \mu_Y\|^2_{W^*}$ where X and Y are two shapes of the same dimension and compute the gradient of such terms with respect to the points of X.

We assume now that X and Y are two unoriented d-dimensional shapes (possibly disconnected) given respectively as sets of vertex $(q_k)_{k=1,..,N}$, $(y_l)_{l=1,..,M}$ together with sets of unoriented simplexes $(f^1_i, .., f^d_i)_{i=1,..,n}$ and $(g^1_j, .., g^d_j)_{j=1,..,m}$. The computation of the varifold representations μ_X and μ_Y of each shape was explained in section 3.2.2 and provided by equations (3.6) for curves, (3.7) for surfaces. We

write $\mu_X = \sum_{i=1}^{n} l_i.\delta_{(p_i,U_i)}$ and $\mu_Y = \sum_{j=1}^{m} \lambda_j.\delta_{(q_j,V_j)}$. Then the attachment distance becomes :

$$\begin{aligned} A &= \|\mu_X\|^2_{W^*} - 2\langle \mu_X, \mu_Y \rangle_{W^*} + \|\mu_Y\|^2_{W^*} \\ &= \sum_{i,j=1..n} l_i l_j k_e(p_i,p_j).k_t(U_i,U_j) \\ &\quad -2 \sum_{i=1..n} \sum_{j=1..m} l_i \lambda_j k_e(p_i,q_j).k_t(U_i,V_j) \\ &\quad + \sum_{i,j=1..n} \lambda_i \lambda_j k_e(q_i,q_j).k_t(V_i,V_j) \end{aligned} \tag{3.35}$$

Thus, computing the distance between X and Y consists basically in computing the varifold representation of the shapes from their sets of points and meshes and compute kernel evaluations between point positions and tangents. It's important to note that, in the cases of curves and surfaces, the computation of terms like $k_t(U,V)$ is actually much simplified and does not require to go through projection matrices. The reason is that the dimension or codimension is always one and so the expression of the kernels with the principal angles in 3.3.1 show that one simply has :

$$\begin{aligned} k_P(U,V) &= \langle \xi_U, \xi_V \rangle^{2k} \\ k_G(U,V) &= e^{-\frac{2}{\sigma^2}(1-\langle \xi_U, \xi_V \rangle^2)} \end{aligned}$$

where ξ_U is the unit tangent vector to the segment in the case of curves or the unit normal to the triangle in the case of surfaces, no matter what orientation to ξ_U is given.

The more technical part is the computation of the gradient of the previous distance with respect to the (x_k)'s, the points of the first shape. In equation (3.35), the attachment is a function of the p_i, l_i and U_i but all these terms are themselves functions of the (x_k)'s, respectively as the centers, volumes and tangent directions to every cell of X. The resulting function that we have denoted $A((x_k))$ can be therefore differentiated as a composition :

$$\partial_{x_k} A = \sum_{i=1}^{n} (\partial_{p_i} A \circ \partial_{x_k} p_i + \partial_{l_i} A \circ \partial_{x_k} l_i + \partial_{U_i} A \circ \partial_{x_k} U_i) \tag{3.36}$$

The differentials of A with respect to the p_i, l_i and U_i are easily expressed from eq.(3.35), involving the differentials of the kernels k_e and k_t.

$$\begin{cases} \partial_{p_i} A = \sum_{j=1}^{n} l_i l_j [\partial_1 k_e(p_i,p_j) + \partial_2 k_e(p_j,p_i)].k_t(U_i,U_j) \\ \qquad -2 \sum_{j=1}^{m} l_i \lambda_j \partial_1 k_e(p_i,q_j).k_t(U_i,V_j) \\ \\ \partial_{U_i} A = \sum_{j=1}^{n} l_i l_j k_e(p_i,p_j).[\partial_1 k_t(U_i,U_j) + \partial_2 k_e(U_j,U_i)] \\ \qquad -2 \sum_{j=1}^{m} l_i \lambda_j k_e(p_i,q_j).\partial_1 k_t(U_i,V_j) \\ \\ \partial_{l_i} A = 2 \sum_{j=1}^{n} l_j k_e(p_i,p_j).k_t(U_i,U_j) \\ \qquad -2 \sum_{j=1}^{m} \lambda_j k_e(p_i,q_j).k_t(U_i,V_j) \end{cases} \tag{3.37}$$

where ∂_1 and ∂_2 denotes the derivative with respect to the first and second argument.

Now, the last step consists in computing derivatives of p_i, l_i and U_i with respect to the points of X. These depend on the dimension and codimension of the shape. In the case of curves and surfaces embedded in $\mathbb{R}^3$, it can be simply done by differentiating equations (3.6) for curves or equations (3.7) for surfaces. We give the details for these two cases below.

In the case of **curves**, we have $p_i = \frac{(x_{f_i^1} + x_{f_i^2})}{2}$, $l_i = |x_{f_i^2} - x_{f_i^1}|$. The tangent space direction can be represented as a single normalized vector $u_i = \frac{x_{f_i^2} - x_{f_i^1}}{|x_{f_i^2} - x_{f_i^1}|}$, which gives for all $i \in \{1, .., n\}$ and $k \in \{1, .., N\}$

$$\begin{cases} \partial_{x_{k,s}} p_i = \dfrac{1}{2}(\delta_{\{k=f_i^1\}} + \delta_{\{k=f_i^2\}}) \\ \\ \partial_{x_{k,s}} u_i = \dfrac{1}{|x_{f_i^2} - x_{f_i^1}|}\left(e_s - \dfrac{x_{f_i^2,s} - x_{f_i^1,s}}{|x_{f_i^2} - x_{f_i^1}|} u_i\right)\delta_{\{k=f_i^2\}} \\ \qquad\qquad - \dfrac{1}{|x_{f_i^2} - x_{f_i^1}|}\left(e_s - \dfrac{x_{f_i^2,s} - x_{f_i^1,s}}{|x_{f_i^2} - x_{f_i^1}|} u_i\right)\delta_{\{k=f_i^1\}} \\ \\ \partial_{x_{k,s}} l_i = \dfrac{x_{f_i^2,s} - x_{f_i^1,s}}{|x_{f_i^2} - x_{f_i^1}|}\delta_{\{k=f_i^2\}} - \dfrac{x_{f_i^2,s} - x_{f_i^1,s}}{|x_{f_i^2} - x_{f_i^1}|}\delta_{\{k=f_i^1\}} \end{cases} \tag{3.38}$$

In the previous equations, $x_{k,s}$ denotes the number s coordinate of x_k in the embedding space's canonical basis (e_s). These equations can be simply interpreted and implemented as the way to distribute the gradient computed with respect to the segments over the points of the shape X.

In the case of **triangulated surfaces** in $\mathbb{R}^3$, we have $p_i = \frac{(x_{f_i^1} + x_{f_i^2} + x_{f_i^3})}{3}$, $l_i = |(x_{f_i^2} - x_{f_i^1}) \wedge (x_{f_i^3} - x_{f_i^1})|$ and again the tangent space is simply encoded by the normalized normal vector $u_i = \frac{(x_{f_i^1} - x_{f_i^2}) \wedge (x_{f_i^2} - x_{f_i^3})}{|(x_{f_i^1} - x_{f_i^2}) \wedge (x_{f_i^2} - x_{f_i^3})|}$ so that for all $i \in \{1, .., n\}$ and $k \in \{1, .., N\}$

$$\begin{cases} \partial_{x_{k,s}} p_i = \dfrac{1}{3}(\delta_{\{k=f_i^1\}} + \delta_{\{k=f_i^2\}} + \delta_{\{k=f_i^3\}}) \\ \\ \partial_{x_{k,s}} u_i = \dfrac{1}{l_i}\left(e_s \wedge (x_{f_i^2} - x_{f_i^3}) - \left\langle e_s \wedge (x_{f_i^2} - x_{f_i^3}) \ , \ u_i \right\rangle u_i\right)\delta_{\{k=f_i^1\}} \\ \qquad\qquad + ... \\ \\ \partial_{x_{k,s}} l_i = \left\langle e_s \wedge (x_{f_i^2} - x_{f_i^3}) \ , \ u_i \right\rangle \delta_{\{k=f_i^1\}} + ... \end{cases} \tag{3.39}$$

We wrote "+..." to mean that similar terms are obtained for $k = f_i^1$ and $k = f_i^2$. Combining the last relations with equations (3.36) and (3.37) provides a full description of the gradient computation in practice.

3.5.3 Numerical complexity

From a purely computational point of view, what appears from the previous discrete expressions is that the difference between currents and varifolds, in the cases of 3D curves and surfaces, lie entirely in the kernels that operate on the tangent or normal vectors. Basically, this kernel in currents is always the linear one on the

tangents, i.e $\langle \xi, \xi' \rangle$ whereas, with varifolds offer many other possibilities like for instance $k_P = \langle \xi, \xi' \rangle^{2k}$ or $k_G = e^{-\frac{2}{\sigma^2}(1-\langle \xi, \xi' \rangle^2)}$. In practice, these two methods are thus very close together and it is easy to design one common code structure to include both.
Another consequence is the fact that the complexity for doing varifold registration is of the exact same order than the one of currents. Computing attachment distance functionals and its gradients by brute force method is also $O(N^2)$ for shapes with N points. Just as for currents' metrics, the essential part of computations are kernel convolutions (here in the product space $\mathbb{R}^3 \times G_1(\mathbb{R}^3)$). Thus the alternative numerical schemes and tricks for faster kernel evaluations, presented in chap 2. section 2.3.3, could also be transposed to varifolds as for instance FGT and GPU computations. We had shown empirically that GPU can perform extremely well up to the limit of 10000-20000 points, and this remains perfectly true for all varifold kernels as well.
For very high N, we had mentioned that the FFT convolution method on grids (as introduced and implemented in [28]) is probably the most adapted. For currents, computations of the energy and gradients can be expressed using only discrete convolutions in $\mathbb{R}^3$ of the form $\sum_j k_e(x_i, x)\xi_i$. This holds essentially because the kernel $\langle \xi, \xi' \rangle$ can be expanded in coordinates as a product of separate terms in ξ and ξ'. In the case of varifolds, this remains true for the polynomial kernels k_P and grid strategies are worth implementing in that case. However, it breaks down for the Gaussian kernel k_G (unless by truncating the Taylor series) for which one would need higher dimensional grids, usually too memory and time-consuming to be really interesting, as mentioned in chap 2. 2.3.3.
Finally, to extend a little bit the discussion to larger dimensions and codimensions (although these have not been subject to implementation yet), it's worth pointing out that the space $\Lambda^d E$, which is used to represent local geometry in the currents' setting, becomes very high dimensional when working with submanifolds embedded in a large vector space E. This space is much bigger than the space of all possible tangent spaces, which is not the case of the Grassmannian. We argue that this would result in a much more numerically intensive use of currents than varifolds in those cases.

3.5.4 Simulations and results

We now come to a few results of varifold LDDMM algorithm, in which we want to emphasize the benefit in situations that traditionally involve orientation issues when objects are matched with the currents' framework. In all the following experiments, we chose the Gaussian kernel $k_e(x, y) = e^{-\frac{|x-y|^2}{\sigma_e^2}}$ on the space E. There are also many possibilities for the kernel k_t on the Grassmann manifold : we have focused essentially on the linear kernel k_L and the Gaussian kernel k_G. Even though both kernels were proved to induce a distance on the set of reunion of submanifolds, we have also explained why the Gaussian kernel has better separation properties, as argued before, especially when the scale σ_e of the kernel on E is chosen to be large. Still, the polynomial kernel has the advantage of not introducing an additional scale parameter and also leads to fast numerical computations for shapes with high number of points.
The example of figure 3.6 shows a situation with pikes' structures that are typically not well matched using LDDMM with currents. This example can be also used to show the robustness of varifolds when we increase the scale parameter of the spatial

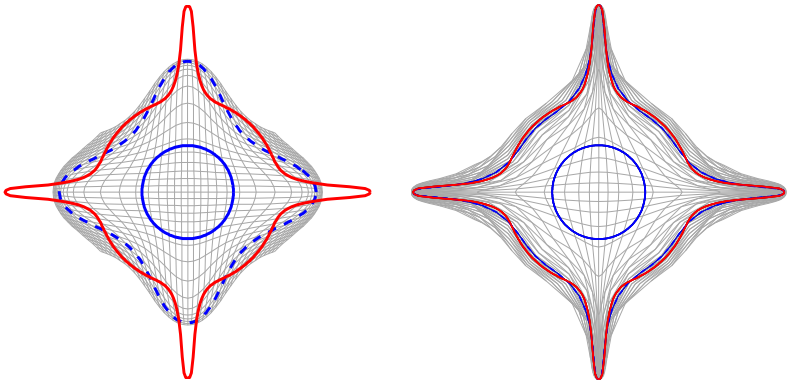


Figure 3.6: Example of registration between 2D curves. The source shape is the blue circle, the target is the red star with four narrow branches. On the left, the matching is performed with the approach of currents. On the right, with the approach of varifolds exposed in section 3.5.1 with the same parameters. We see that the branches are well recovered with varifolds whereas the current's metric is nearly insensitive to them.

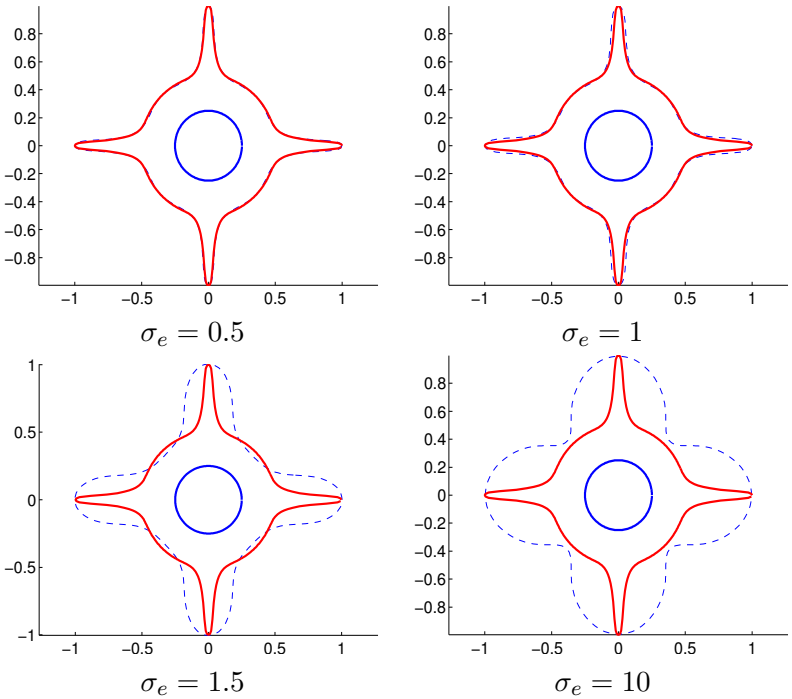


Figure 3.7: The same example for varifolds but performed at different scales σ_e.

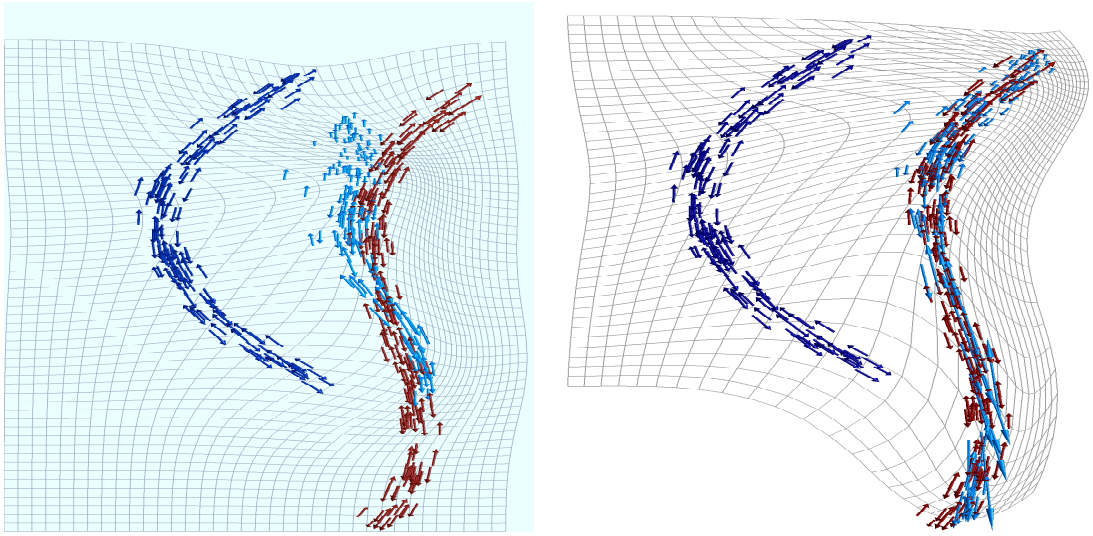


Figure 3.8: Registration result on fibers with the framework of currents (left) and of varifolds (right). As expected, varifolds can deal efficiently with any orientation of the different components.

kernel k_e, as illustrated in figure 3.7. It is particularly remarkable that even at scales much larger than the size of the object, the algorithm is still able to capture the overall description of the target shape, simply based on the distribution of tangent spaces (cf discussion of section 3.4.1).
In figure 3.8, we show an example where shapes are constituted of many disconnected pieces of curves which are not consistently oriented, and show the results of both methods (currents and varifolds). Other interesting situations are given by tree-like structures for which orientating the different branches consistently for registration can become a nearly intractable problem when faced with many intersecting branches. We give a simple insight of this phenomenon in figure 3.9.
We also present some results of surface matching done with our varifold algorithm. In figure 3.10, a surface registration is applied between a standard sphere and the Stanford bunny model with same parameters for both methods. We observe some mismatch artifacts in the currents's result : the ears are not fully recovered as well as some small details in the head and we see the apparition of undesirable membranes at the basis of the ears. This is due to a 'flattening effect' of currents that often squeeze both sides of the surface together in order to artificially eliminate unmatched parts of the shapes. This has been a consistent lack of currents for many applications. In contrast, we see that the varifold algorithm matches almost perfectly all parts of the shapes.

3.6 Discussion, extensions and perspectives

By way of conclusion to this chapter, we wish to provide a few additional comments and perspectives that naturally arise from all the previous material. We divide it into three main topics. We first elaborate a little more on the difference of currents and varifolds with respect to kernel scales and the issue of noise. We then propose, as a combination of the ideas of chapter 2 and 3, an extension of varifolds to model unoriented functional shapes. And finally, we briefly mention an alternative notion

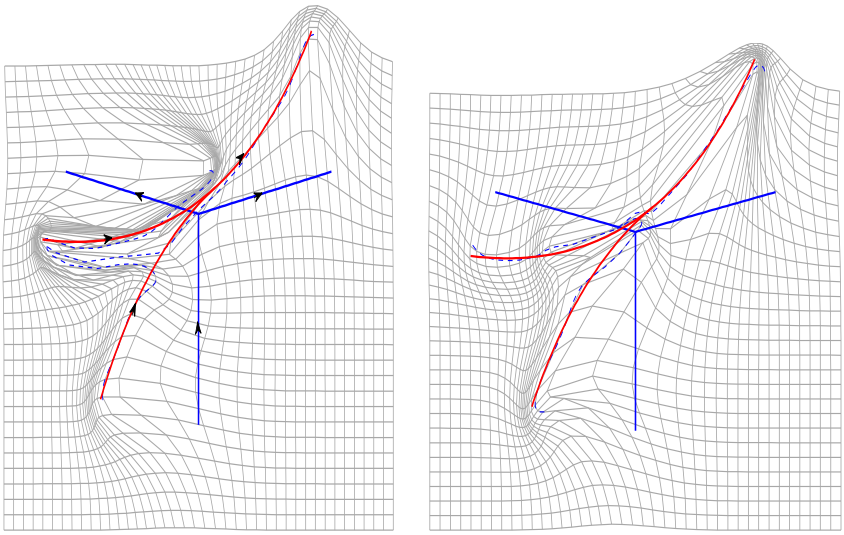


Figure 3.9: Example of registration on trees with currents (left) and varifold (right). Observe again that, with currents, consistent orientation between the source and the target is necessary to avoid the unnatural deformation shown on the figure.

of oriented varifold that could be interesting in future studies as well.

3.6.1 Scale and noise

An important difference between currents and varifolds that we have emphasized several times throughout this chapter is the behavior of kernel metrics when we increase the width σ_e of the Gaussian kernel on the space E. The cancellation effects of currents make such norms become more and more insensitive to anything else than the boundaries of the objects as noticed in 3.4.1. On the other hand, with varifolds, we obtain distances between the distribution of tangent spaces of the two shapes. We showed that, at infinite scale σ_e, what remains from a shape X is the image measure $\mathcal{H}^d \circ Tan^{-1}$. From the discrete point of view, these can be thought as *histograms* on $G_d(E)$. If we focus on the special case of planar curves, $G_1(\mathbb{R}^2)$ can be represented as the unit semi-circle (with identified extremities) and such histograms can be computed numerically and visualized, as we show on a few examples in figure 3.11. There is of course no one-to-one correspondence between histograms and the set of all shapes since these objects are totally blind to the localization of the tangent spaces in E. However, fixing a given histogram still imposes constraints on shape. When constraints are also added on the set of possible transformations, as we do in large deformation approaches for instance, it becomes a particularly interesting question to understand how much attachment distances based only on histograms could drive registrations ; a topic that we just skimmed through with the example shown in figure 3.7.
There is however a negative counterpart to the fact that varifolds remain sensitive to tangent spaces distributions even if the scale σ_e becomes large. This is the problem of noisy shapes that can involve important perturbations in the Grassmannian domain. On the opposite, currents have a regularizing effect that makes the approach robust to noise for well-chosen σ_e. We show the result of a matching performed under noise in figure 3.12. As we see, currents are able to ignore noise quite well since the high-frequency variations are cancelling for the chosen scale σ_e. For varifolds, things are

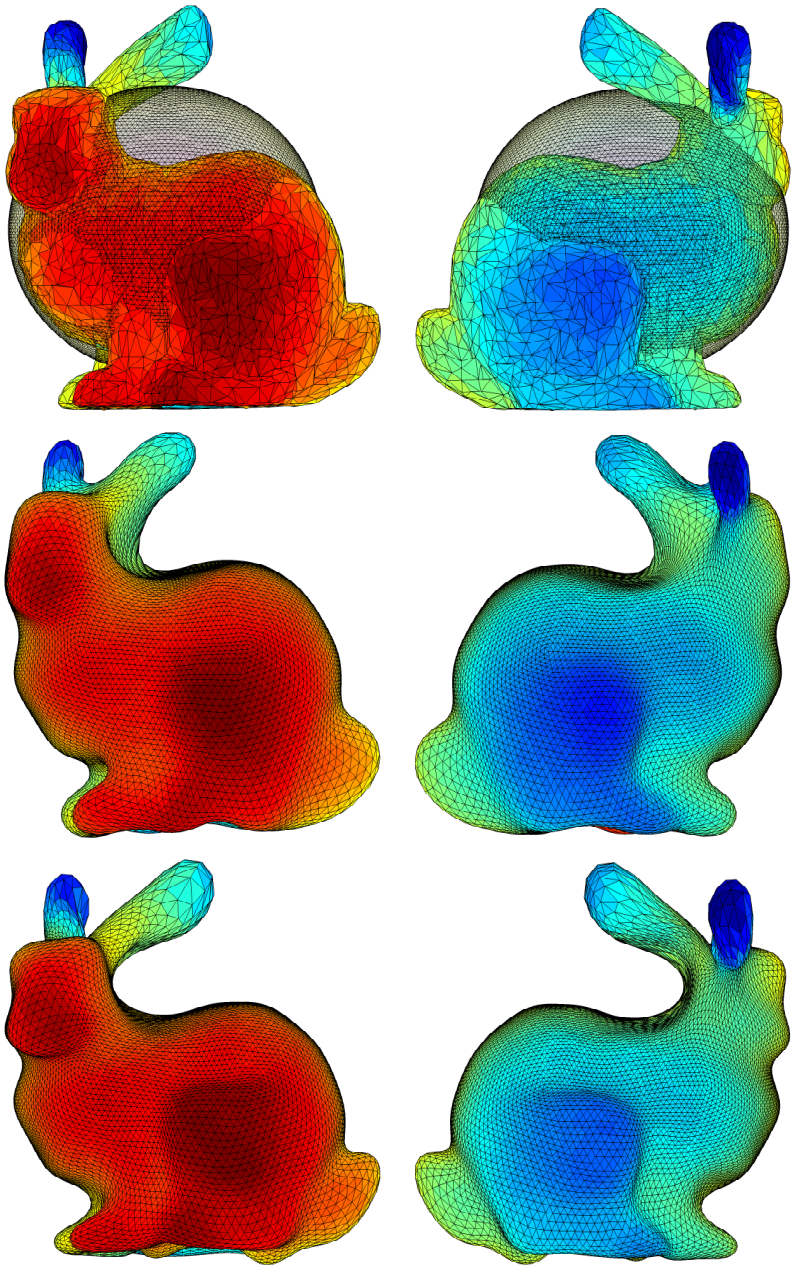


Figure 3.10: Surface matching results between a sphere and the Stanford bunny. At top row, the source sphere and the target bunny shape. In the middle row, the result after matching with currents. At bottom row, with the varifold approach.

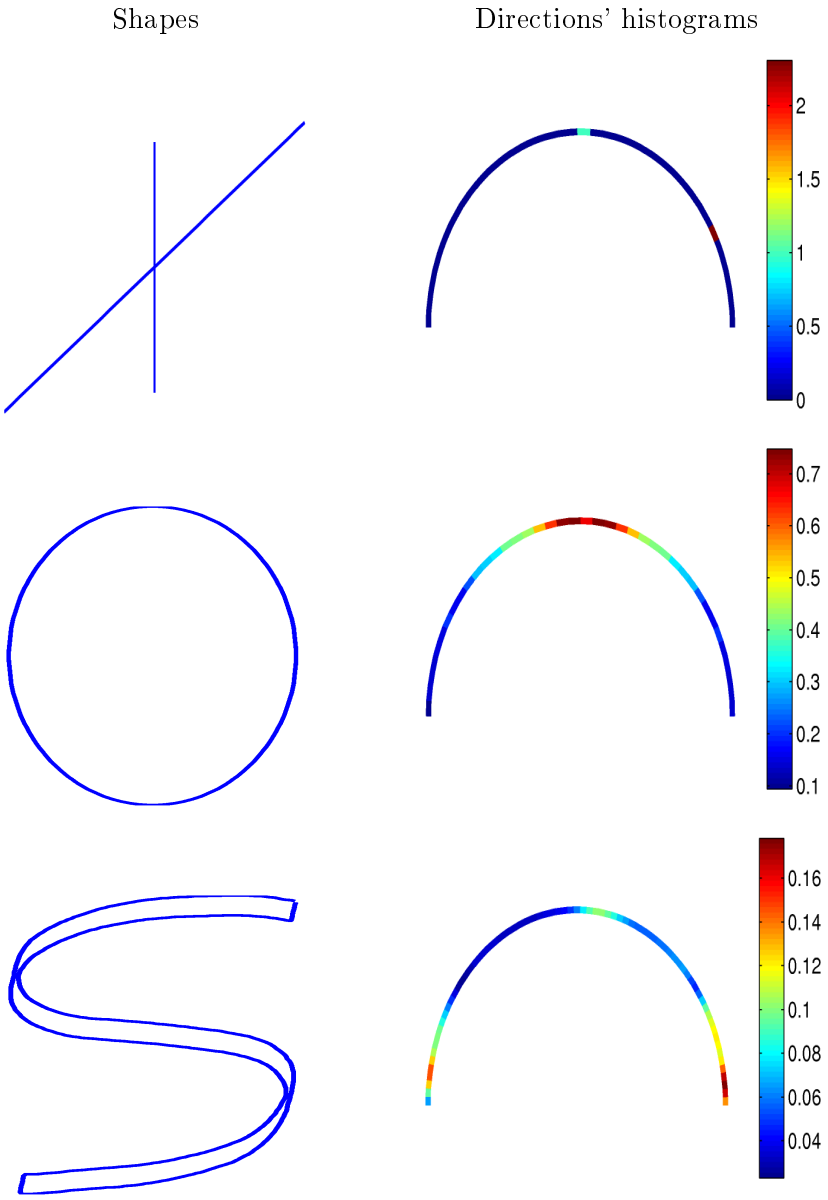


Figure 3.11: Several plane curves and the histograms of the tangent space directions. For visualization purposes, we have identified the Grassmannian $G_1(\mathbb{R}^2)$ with the unit semi-circle. The color on each point of the semi-circle represents the density (possibly singular) for the Hausdorff measure $\mathcal{H}^1$ of points having a tangent pointing in that particular direction. The sum of all these densities is precisely the total length of the curve.

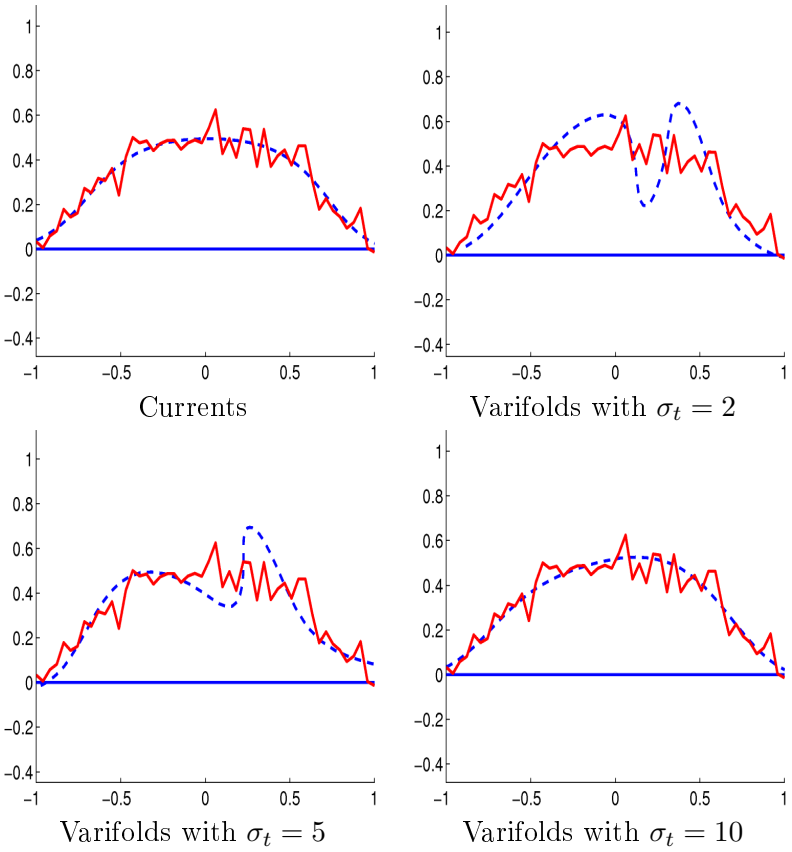


Figure 3.12: Registration of the blue segment on the noisy shape in red. We show the deformed source shape in dashed line. Current and varifold algorithms were applied with both a Gaussian kernel on E of common $\sigma_e = 0.4$. For varifolds, we used the Gaussian kernel k_G of scale σ_t on the Grassmann part and we show the behavior as we increase σ_t.

more complex and depend on the scale chosen for the Gaussian on $G_d(E)$. If σ_t is too small, the algorithm becomes sensitive to the distorted directions of the tangents and creates a prominence to account for these directions. One must therefore reduce this sensitivity by increasing the width σ_t in order to achieve robustness to noise. In practical applications, this could be a drawback of the varifold approach, typically if only some parts of shapes are affected by noise while others are smooth because it would enforce the choice of a large σ_t, providing thus only a gross analysis of tangent directions everywhere.
We could still argue that the issue of noise can be also treated in practice from the diffeomorphism point of view, for instance by imposing more regularity to the vector field v of the deformation. This can be done simply by reducing the tradeoff γ or by increasing the typical scale σ_V of vector fields. Yet the question of finding, in this context, representations and attachment metrics that could simultaneously avoid orientation issues while providing a natural robustness to noise is, to our knownledge, still an open question.

3.6.2 Functional varifolds

The varifold idea to represent non-oriented shapes can be also adapted to the case of shapes carrying signals, following a similar path as chapter 2 with currents. In fact, linking the ideas this chapter with the ones of chapter 2, it is even a quite easy extension of varifolds. We will just briefly sum up the main lines. With the notations and definitions of chapter 2, we are naturally led to introduce the following definition of functional varifolds :

Definition 3.6.1. *We say that μ is a d- dimensional functional varifold on (E, M), E being the embedding vector space and M the signal manifold, if μ is a Borel finite measure on the space $E \times G_d(E) \times M$ or equivalently if $\mu \in C_0(E \times G_d(E) \times M)'$.*

This is simply augmenting the varifold representation with a signal component. Now, any functional shape (X, f), with X a d-dimensional rectifiable subset, has an equivalent representation in terms of functional varifolds given by :

$$\mu_{(X,f)}(\omega) = \int_X \omega(x, T_x X, f(x)) d\mathcal{H}^d(x) \tag{3.40}$$

The discretized version of this relation for a polyhedral shape writes
$\sum_{i=1}^n m_i.\delta_{(x_i,V_i,f_i)}$ with the conventions already explained in previous sections. The definition of kernel metrics on functional varifolds is also very natural. One can just consider kernels of the form $k_e \otimes k_t \otimes k_f$, where k_e, k_t and k_f are real-valued positive kernels respectively on E, $G_d(E)$ and M satisfying enough regularity properties. We then obtain a Hilbert metric between functional shapes that writes :

$$\langle \mu_{(X,f)}, \mu_{(Y,g)} \rangle^2_{W'} = \iint_{X \times Y} k_e(x,y)\, k_t(T_x X, T_y Y)\, k_f(f(x), g(y)) d\mathcal{H}^d(x) d\mathcal{H}^d(y) \tag{3.41}$$

The rigorous study of the properties of these norms, as we did in chapter 2 section 2.2.3 for functional currents, is postponed to a future work, but does not seem, a priori, to bring additional difficulties.
The implementation of functional varifolds' metric for computing functional shapes' distances and registration is very convenient starting from the similar codes on varifolds. Indeed, all it requires is to add the ponderations coming from the kernel

k_f and the functional values and compute the new gradients with respect to point positions and signals. Thus this framework (in the case $M = \mathbb{R}$) was also included in our codes as an alternative to functional currents.
Functional varifolds, as far as numerical simulations can tell up to now, successfully combine the interest of non-orientation (compared to currents) with the possibility to use signals in registration and estimate functional residuals. To link that with the previous discussion on the dependency of the metrics with respect to kernels' scales, it also appears that functional varifolds, by playing with the three possible parameters σ_e, σ_t, σ_f (in the case of Gaussian kernels), encompass in a unified framework a very wide variety of metrics between objects : measures ($\sigma_t, \sigma_f \to \infty$), varifolds ($\sigma_f \to \infty$), directions' histograms ($\sigma_e, \sigma_f \to \infty$), signals' histograms ($\sigma_e, \sigma_t \to \infty$).

3.6.3 Oriented varifolds

Let's finally mention a last possible track that, we believe, could lead to interesting developments in shape modelling and analysis, which is called *oriented varifolds*. This concept, although less common in geometric measure theory, has a definition very close to varifolds. It makes use of the oriented Grassmann manifold which is defined as the set of all oriented d-dimensional subspaces of E, denoted generally $\widetilde{G}_d(E)$. In a similar way as the Plücker embedding, $\widetilde{G}_d(E)$ can be embedded into the unit sphere of $\Lambda^d E$. In particular, if $d = 1$ or $d = n-1$, oriented Grassmann manifold can be actually identified to the unit sphere of E , which makes it a particularly simple space to use in practice. Oriented varifolds are then defined similarly to varifolds as :

Definition 3.6.2. *An oriented varifold is a Borel finite measure on $E \times \widetilde{G}_d(E)$, i.e an element of the functional space $C_0(E \times \widetilde{G}_d(E))'$.*

Now, oriented varifolds provide a representation of any oriented rectifiable subsets X of E by integrating as usual any function of $C_0(E \times \widetilde{G}_d(E))$ on X, which associate to X the oriented varifold denoted $\widetilde{\mu}_X$. However, it is **not** equivalent neither to currents nor to varifolds in general. If X is for instance a connected oriented submanifold and $\check{X}$ the one with opposite orientation, in that setting, $\widetilde{\mu}_{\check{X}} \neq -\widetilde{\mu}_X$ and $\widetilde{\mu}_{\check{X}} \neq \widetilde{\mu}_X$.
Although it has not been carried out yet, one could obviously define kernels as we did with varifolds using in that case the embedding of $\widetilde{G}_d(E)$ into the sphere of $\Lambda^d E$ and obtain RKHS distances in that context. Such a framework, as we can expect, would not suffer from the cancellation issues appearing with currents (at the price of a lower robustness to noise). Compared to varifolds, it would likely improve results when working on shapes with a clearly defined and useful orientation information but it remains clearly ineffective in treating cases where orientation is not well-defined or relevant.

3.6.4 Conclusion

Having explored in the first place the orientation issues that come along with the use of currents in shape modelling, we have adapted in this chapter the alternative concept of varifolds, by representing shapes as spatially spread distributions of un-oriented tangent spaces themselves modeled as elements of the Grassmann manifold. As for currents, we embed a wide variety of objects like submanifolds and rectifiable subsets in one common functional space. The first contribution of our work was to define metrics that share theoretical requirements and enable convenient numerical computations. We have shown how this can be done by defining Hilbert structures

on varifolds through reproducing kernels and proposed a general construction process for such kernels. From the theoretical side, we proved that these metrics, as opposed to the previous representation by currents, do not artificially 'eliminate' shape volume due to orientation. We also computed a variation formula of the metric with respect to shapes. In terms of numerical applications, we have presented an adaptation of large deformation shape registration based on this framework. The results we presented are mainly focused on synthetic examples for the time being, but we will show a few applications to real datasets in chapter 4, in which we extend the varifold framework to atlas estimation.

As we started to mention in this last section, this work is also paving the way to several extensions and new issues. For instance, the ideas of functional varifolds or oriented varifolds can be almost straightforwardly implemented and now need to be experimented on larger sets of applications. Furthermore, the discussion on the relationship between varifolds and tangent spaces' histograms, that was only initiated in 3.6.1, would probably deserve more in-depth study, perhaps by linking that to shape comparison based on descriptors, for instance shape contexts in 2D and 3D shapes [15] or SIFT descriptors in images.

Chapter 4

Statistical Atlas estimation and analysis

Contents

4.1 Forward atlas models 130
4.2 Atlas estimation for functional shapes 132
4.2.1 A first model 132
4.2.2 Estimation of functional residuals 142
4.2.3 Results on the retina dataset 147
4.2.4 A well-posed continuous framework for atlas estimation . . 151
4.3 Statistical analysis of momenta and residuals 153
4.3.1 Dimensionality reduction 153
4.3.2 Generalized LDA on Hilbert spaces 157
4.3.3 Modes and classification for functional atlases 159
4.3.4 Classification results on the OCT dataset 162
4.4 Atlas on varifolds 168
4.4.1 Algorithm 168
4.4.2 Works on embryonic mouse heart 168
4.4.3 Template estimation on Cochleae dataset 172
4.5 Conclusion and extensions 175

4.1 Forward atlas construction in computational anatomy

In the previous chapter, we have presented a first range of applications related to the problem of diffeomorphic registration between two functional shapes. In the context of anatomical studies, the estimated deformations are likely to make possible the detection of abnormalities in the shape of subjects, leading to potential medical diagnosis for certain diseases. To have any meaning from the statistical point of view, registration need to be transposed to larger database of subjects. In this chapter, we discuss the problem of group-study, namely how to analyze in a statistical way potentially large groups of geometrico-functional datasets.

This problem has been extensively studied in the past, mostly in the case of 2D and 3D images [11, 6, 49] and more recently generalized to shapes like curves and surfaces [28, 32]. The main challenge, starting from registration of two subjects, is precisely to avoid the bias that would inevitably occur if we chose one specific subject in the database and register all other individuals on this one. Instead of this naive approach, recent methods are generally based on algorithms that try to estimate an **atlas** from the data. In shape analysis, an atlas is made of a *template shape*, or a mean shape, that captures the common features of all subjects, together with a set of deformations between template and individuals, which, properly analyzed, should transcribe the *shape variability* among the population. Technically speaking, given a population of N subjects $\{\mathcal{S}_i\}_{i=1,..,N}$, the output is a template $\overline{\mathcal{S}}$ and N deformations $\{\phi_i\}_{i=1,..,N}$. These deformations can either pull the subjects back on the template eq.(4.1), which is thus called the *backward model*, or transport the template toward the individuals eq.(4.2), which is known as the *forward model*.

$$\phi_i(\mathcal{S}_i) = \overline{\mathcal{S}} + \epsilon_i \iff \mathcal{S}_i = \phi_i^{-1}(\overline{\mathcal{S}}) + \phi_i^{-1}.\epsilon_i \,. \tag{4.1}$$

$$\mathcal{S}_i = \phi_i(\overline{\mathcal{S}}) + \epsilon_i \,. \tag{4.2}$$

In both these models, the ϵ_i are shape *residuals* for each subject, corresponding to the mismatch between the template and the subject. As already mentioned for registration, these are necessary to account for all that cannot be captured by the deformation model including noise, topological changes and high-frequency variations. Note however that in the backward model, the residuals are living on the template coordinate system and are transported by each deformation whereas in the forward model, they are intrinsic properties of the subject. This is summed up by Figure 4.1. It is argued in [28] that the forward model is better fitted to modelling considerations and more adapted to statistical purposes : it can be derived from a consistent maximum a posteriori (MAP) estimator and computing the likelihood of a new subject is much easier than in the case of the backward model. This approach has been implemented for 2D and 3D currents and found successful applications to atlas estimation of white matter fiber bundles [30] or deep brain structures' surfaces [36].

Before coming to that, we briefly sum up the general derivation of MAP estimator in the context of usual currents, following [6] and [28]. We insist on the fact that these are only formal derivations since they are done infinite dimensional vector spaces. However, as detailed in [28], it can be translated in terms of finite-dimensional models, in the case of usual applications. From a Bayesian point of view, given the shape observations $\mathcal{S}_i$ represented as currents in a RKHS W', template estimation

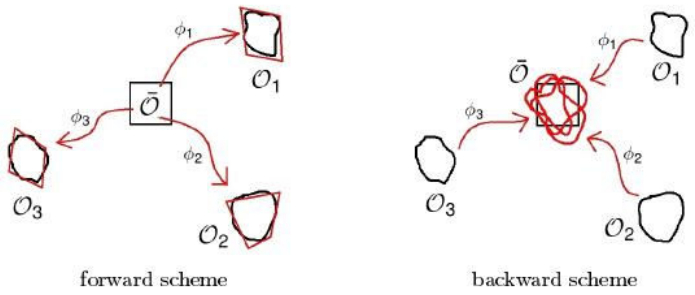


Figure 4.1: Forward and backward atlas estimation's models [28].

reduces to the problem of finding a template $\overline{\mathcal{S}}$ that intuitively maximizes the likelihood of the observations knowing the template, i.e for independent observations, $\prod_{i=1}^{N} p(\mathcal{S}_i|\overline{\mathcal{S}})$, where the deformations ϕ_i in eq.(4.2) are hidden variables. To give a precise meaning to that, one need to introduce probabilistic models and priors for the residuals and deformations. Residuals ϵ_i are considered as iid random variables on the space of currents with Gaussian probability density function :

$$p_r(\epsilon) = C_r . \exp(-\|\epsilon\|^2_{W'}/(2\sigma^2_W)) .$$

Deformations are totally determined by their initial velocity field in $v_0 \in V$ thanks to the geodesic equations, so that we can introduce a prior on deformation with a probability density function :

$$p_\phi(v_0) = C_\phi . \exp(-\|v_0\|^2_V/(2\sigma^2_V)) \tag{4.3}$$

Now, one can express formally the probability of an observation $\mathcal{S}_i$ knowing the template by using the Bayes rule :

$$\begin{aligned} p(\mathcal{S}_i|\overline{\mathcal{S}}) &= \int p(\mathcal{S}_i|\overline{\mathcal{S}}, v_0^i) p_\phi(v_0^i) dv_0^i \\ &= \int p_r(\mathcal{S}_i - (\phi^{v_0^i})_* \overline{\mathcal{S}}) p_\phi(v_0^i) dv_0 \end{aligned} \tag{4.4}$$

The term $p_r(\mathcal{S}_i - (\phi^{v_0^i})_* \overline{\mathcal{S}})$ depends on the vector field v_0 by a geodesic shooting procedure which provides no closed form for maximizing the previous integral. Hence, we apply a standard approximation, called Fast Approximation with Modes (FAM, cf [6]), replacing the integral of eq.(4.4) by the maximum of its integrand : $p(\mathcal{S}_i|\overline{\mathcal{S}}) \approx max_{v_0^i} p_r(\mathcal{S}_i - (\phi^{v_0^i})_* \overline{\mathcal{S}}) p_\phi(v_0^i)$. Then, the maximization of the likelihood or equivalently the minimization of the log-likelihood leads to the following optimization problem :

$$(\overline{\mathcal{S}}, v_0^i) = \text{argmin}_{(\overline{\mathcal{S}}, v_0^i)} \left(\sum_{i=1}^{N} \frac{1}{2\sigma_V^2} \|v_0^i\|^2_V + \frac{1}{2\sigma_W^2} \|\mathcal{S}_i - (\phi^{v_0^i})_* \overline{\mathcal{S}}\|^2_{W'} \right) \tag{4.5}$$

Again, this minimization is just formal at this point because it does not specify in what space the template shape $\overline{\mathcal{S}}$ is optimized on. Only in the case of discrete shapes can eq.(4.5) be actually given a real meaning. In that case, the template is made of a finite number of points and the variable $\overline{\mathcal{S}}$ lives in a finite-dimensional space. The resolution of this problem then gives both the template shape $\overline{\mathcal{S}}$ and the

deformations $\phi^{v_0^i}$ from the template to each of the observation, which (still in the discrete situation) are obtained by geodesic shooting from the initial vector fields v_0^i. Note that, for a fixed $\overline{\mathcal{S}}$, the previous problem consists exactly in solving N matching problems between $\overline{\mathcal{S}}$ and the $\mathcal{S}_i$. Conversely, for given $\phi^{v_0^i}$, optimizing with respect to $\overline{\mathcal{S}}$ uses the computation of the variation of the attachment distance given by currents. Thus, estimating an atlas in the context of currents can be done by a simultaneous gradient descent on the template and the deformations. This approach has been implemented on various shape datasets, as for instance in [33, 34].

4.2 Atlas estimation for functional shapes

Having outlined in the previous section the idea of atlas estimation of shapes represented by currents through forward scheme, we now intend to provide a similar framework in the context of functional shapes. Our approach makes profit on the metrics given by the functional currents' setting that was developed in chapter 2 and that demonstrated interesting properties on the simpler problem of registering two functional shapes. In the first place, we present a model that directly transcribes in terms of functional currents the atlas estimation problem of eq.(4.5) and we will show in what respect combining geometry and signal can provide more relevant templates. We shall then propose an enriched model where functional residuals on each subject are estimated in addition to the template and the deformations. Finally, in light of more recent but still prospective reflexions at the time, we discuss the possibility to formalize, in the continuous setting, a well-posed problem that could guarantee the existence of minimizers for atlases.

4.2.1 A first model

Formulation with MAP

We now consider the situation where we are given N functional shapes $\mathcal{S}_i = (X^i, f^i)$ as in chapter 2, X^i being d-dimensional oriented rectifiable subsets and f^i functions defined on X^i with values in a given signal space M. We intend to find, just as in section 4.1, a template that would be now a functional shape $(\overline{X}, \overline{f})$ and deformations $\phi_i \in \mathrm{Diff}(E)$ satisfying :

$$C_{(X^i,f^i)} = (\phi^{v^i})_* C_{(\overline{X},\overline{f})} + \epsilon_i \tag{4.6}$$

We remind that $(\phi^{v^i})_* C_{(\overline{X},\overline{f})}$ is the action of deformations on functional currents, depending on the nature of the signal. One noticeable advantage of the generalization of the concept of currents we have presented is that the formal MAP derivation of atlas estimation explained in section 4.1 can be adapted almost straightforwardly to all sorts of functional shapes. In the **discrete** case, where the template X is assumed to consist of a finite set of points and thus a finite set of signal values f, we also know from section 1.2.5 in chapter 1, that the deformations ϕ^{v^i} are parametrized by initial momenta p_0^i (which are themselves finite sets of covectors attached to each point of X) and thus the atlas estimation problem takes the following form :

$$(\overline{X}, \overline{f}, p_0^i) = \arg\min_{(X,f,p_0^i)} \left(\sum_{i=1}^{N} E(p_0^i) + \gamma \| C_{(X_i,f_i)} - (\phi^{p_0^i})_* C_{(X,f)} \|_{W'}^2 \right) \tag{4.7}$$

γ being a weight parameter between the deformation cost and the data attachment distance (or the ratio between the residual prior's variance and the deformation prior's variance), $\phi^{p_0^i}$ denotes the deformation obtained by geodesic shooting with initial momentum p_0^i. Note that, in this discrete setting, this minimization makes sense because all variables X,f and p_0^i live in finite-dimensional spaces. For a coherent formulation in the continuous setting, we refer to the later discussion of 4.2.4. We also emphasize the fact this atlas estimation formulation remains general enough to incorporate all functional shapes, provided that one can express the way diffeomorphisms act on those functional shapes. In that formalism, if the template fshape $(\overline{X}, \overline{f})$ can be interpreted somehow as a prototype or *mean* of the observations, note that it is not the actual mean in the space of fcurrents, i.e $\frac{1}{N}\sum_{i=1}^N C_{(X^i,f^i)}$. Indeed, this arithmetic mean is not rectifiable in general, so that there is no way to think of it as a functional shape. Instead, the approach formalized by eq.(4.7) optimizes on fcurrents that are already written under the form $C_{(X,f)}$.
Now, if we specify the previous to the case of real-valued signals $f^i : X^i \to \mathbb{R}$, which shall be the situation considered in our applications, as discussed in 2.2, we have $(\phi^{v_0^i})_* C_{(X,f)} = C_{\phi^{v_0^i}.(X,f)}$ with $\phi^{p_0^i}.(X,f) = (\phi^{p_0^i}(X), f \circ (\phi^{p_0^i})^{-1})$ and therefore eq.(4.7) writes :

$$\min_{(X,f,p_0^i)} J(X,f,(p_0^i)) \doteq \left(\sum_{i=1}^N E(p_0^i) + \gamma \| C_{(X^i,f^i)} - C_{(\phi^{p_0^i}(X), f\circ(\phi^{p_0^i})^{-1})} \|^2_{W'} \right) \quad (4.8)$$

Assuming that one can solve this optimization problem (which is the subject of next section), the output is a functional shape $(\overline{X}, \overline{f})$ that captures the common features of the set of observations both in shape and in signal as well as a set of deformations that supposedly explain the geometrical variation around this common template. The rest of the variability is all contained in the fcurrent residues $C_{(X^i,f^i)} - C_{(\phi^{p_0^i}(\overline{X}), \overline{f}\circ(\phi^{p_0^i})^{-1})}$ that combine higher-frequency geometrical variations not captured by the ϕ_i's as well as the variability of signals f_i around $\overline{f}$.

Atlas estimation algorithm

We now come into more details about the effective optimization of eq.(4.8). In [28], was proposed an alternated minimization procedure between the template shape and the deformations. This approach has however the inconvenient of updating the template part by a gradient computed in the space of currents, which first requires deconvolution schemes like Matching Pursuit for the computation and also looses the structure of rectifiable subset of the template. Instead, the algorithm we are going to develop is a single gradient descent scheme that **simultaneously** updates X, f and the deformation momenta p_0^i. To achieve this, we simply need to compute the gradient of the functional with respect to each of these variables.
As we said, the template X is assumed to be composed of a finite set of points $q = (q_k)_{k=1,..n}$ and all initial momenta can be identified to vectors of E attached to each point q_k, i.e $p_0^i = (p_{0,k}^i)_{k=1,..n}$ generating initial vector fields v_0^i by :

$$v_0^i = \sum_{k=1}^n K_V(q_k, .) p_{0,k}^i \,.$$

The gradient of the deformation energy is then almost straightforward since :

$$E = \sum_{i=1}^{N} H_r(q, p^i) = \sum_{i=1}^{N} \sum_{k,l=1}^{n} K_V(q_k, q_l).\langle p_k^i, p_l^i \rangle$$

and thus :

$$\begin{aligned} \nabla_{q_k} E &= \sum_{i=1}^{N} \partial_{q_k} H_r(q, p_i) = \sum_{i=1}^{N} \sum_{l=1}^{n} (\partial_1 K_V(q_k, q_l) + \partial_2 K_V(q_l, q_k)) \langle p_k^i, p_l^i \rangle \\ \nabla_{p_k^i} E &= \partial_{p_k} H_r(q, p_i) = \sum_{l=1}^{n} K_V(q_k, q_l) p_l^i \end{aligned} \tag{4.9}$$

Note that in most cases, K_V is symmetric in both its variables (as for all radial kernels for instance) so the two partial derivatives in the first equation above are actually equal and the gradient becomes particularly simple. The gradient of the attachment terms g_i is again slightly more technical but can be derived in the exact same way as for a registration problem. Indeed, g_i is a function of the points $q_{k,1}^i$, the final positions of the template points shot by momentum p_0^i and do not depend on the final momenta $p_{k,1}^i$. Therefore, $\nabla_{q_1^i} g_i = \nabla_{q_1} g_i(q_1^i)$ where $\nabla_{q_1} g_i$ was already expressed in section 2.3.2, and $\nabla_{p_1^i} g_i = 0$. These gradients being computed at the final state variables q_1^i and p_1^i, one need to pull them back on the initial template position and momenta, which, as already explained in chapter 2 section 2.3.2, is the role of the backward evolution equation, i.e for each $i \in \{1, , , N\}$, we have $\nabla_q g_i = Q_0^i,\ \nabla_{p^i} g_i = P_0^i$ with :

$$\begin{cases} \dot{Q}_t^i = -(\partial_{p,q}^2 H_r(q_t^i, p_t^i))^* Q_t^i + (\partial_{q,q}^2 H_r(q_t^i, p_t^i))^* P_t^i \\ \dot{P}_t^i = -(\partial_{p,p}^2 H_r(q_t^i, p_t^i))^* Q_t^i + (\partial_{q,p}^2 H_r(q_t^i, p_t^i))^* P_t^i \\ Q_1^i = -\nabla_{q^1} g_i(q_1^i),\ P_1^i = 0 \end{cases} \tag{4.10}$$

And it results that the gradient of the functional J with respect to variables q and p^i is given by :

$$\begin{aligned} \nabla_q J &= \sum_{i=1}^{N} (\partial_q H_r(q, p_0^i) - Q_0^i) \\ \nabla_{p_0^i} J &= \partial_p H_r(q, p_0^i) - P_0^i \end{aligned} \tag{4.11}$$

There still remains the computation of the gradient with respect to the functional values on the template, which we shall denote in short by $f_k = f(q_k)$. We see that only the attachment terms g_i depend on f and the gradient can be derived exactly as in chapter 2. Indeed, we remind that, in the model of transport we are considering, the functional values are not modified by the geodesic shooting so there is no backward integration needed. Now, taking similar notations and conventions, let's denote $C_{(X^i, f^i)} = \sum_{k=1}^{m_i} \delta^{\xi_k^i}_{(\tilde{x}_k^i, \tilde{f}_k^i)}$ and $C_{\phi^{p_0^i}.(X,f)} = \sum_{k=1}^{m} \delta^{\xi_k}_{(\tilde{q}_k, \tilde{f}_k)}$ where each dirac encodes one face of the template or the subjects. Considered as a function of the

auxiliary variable $\tilde{f}$, we have :

$$\begin{aligned} g^i &= \langle C_{(X,f)}, C_{(X,f)} \rangle_{W'} - 2\langle C_{(X,f)}, C_{(X^i,f^i)} \rangle_{W'} + \langle C_{(X^i,f^i)}, C_{(X^i,f^i)} \rangle_{W'} \\ &= \sum_{k,l=1}^{m} k_g(\tilde{q}_k, \tilde{q}_l) k_f(\tilde{f}_k, \tilde{f}_l) \langle \xi_k, \xi_l \rangle - 2 \sum_{k=1}^{m} \sum_{l=1}^{m_i} k_g(\tilde{q}_k, \tilde{x}_l^i) k_f(\tilde{f}_k, \tilde{f}_l^i) \langle \xi_k, \xi_l^i \rangle \\ &\quad + \sum_{k,l=1}^{m_i} k_g(\tilde{x}_k^i, \tilde{x}_l^i) k_f(\tilde{f}_k^i, \tilde{f}_l^i) \langle \xi_k^i, \xi_l^i \rangle \end{aligned}$$

And so, the gradient with respect to $\tilde{f}$ writes :

$$\begin{aligned} \nabla_{\tilde{f}_k} J &= \sum_{i=1}^{N} \nabla_{\tilde{f}_k} g^i, \quad \text{with} \\ \nabla_{\tilde{f}_k} g^i &= \sum_{l=1}^{m} k_g(\tilde{q}_k, \tilde{q}_l) [\partial_1 k_f(\tilde{f}_k, \tilde{f}_l) + \partial_2 k_f(\tilde{f}_l, \tilde{f}_k)] \langle \xi_k, \xi_l \rangle \\ &\quad - 2 \sum_{l=1}^{m_i} k_g(\tilde{q}_k, \tilde{x}_l^i) \partial_1 k_f(\tilde{f}_k, \tilde{f}_l^i) \langle \xi_k, \xi_l^i \rangle \end{aligned} \tag{4.12}$$

Since $\tilde{f}$ is basically the interpolated value of the function f at the center of each cell of the mesh on X, the gradient of J with respect to f can be simply obtained by distributing $\nabla_{\tilde{f}} J$ on the points of X.

Now, the atlas estimation is roughly a gradient descent on all variables (q, f, p^i) together, which we recap in algorithm 2. Practically, any type of gradient descent could be applied, the simplest one being an adaptive step scheme. However, in our situation, since we are optimizing on several sorts of variables (point positions, momenta and functional values) which are not necessarily homogeneous, a better approach is to use different steps for q, f and the p^i's, which can be adapted separately by doing a basic line search in order to find a better descent direction after each gradient computation. This allows some flexibility of the algorithm and prevents it from updating too fast one of the variables compared to the others. We also want to emphasize the fact that the energy and gradient we have to compute at each step are the combination of terms related to only one observation at a time. This has the important consequence that the computations done within the FOR loop in algorithm 2 can be parallelized straightforwardly, improving significantly time computations. Working for instance on a machine with as many processors as the number N of subjects do not make atlas estimation longer in time than a single registration between two subjects.

At this stage, algorithm 2 suffers from several drawbacks essentially because it does not take into account the dependency of energies and gradients in either the scale or the sampling of template and data. The other issue comes from boundary behavior that we have already outlined from a theoretical point of view in chapter 2 theorem 2.2.4. All these motivate the efforts that we present in the following to provide a more robust and stable numerical scheme.

Normalization with respect to scales

Despite the obvious invariance of our algorithm if we apply a common isometry of the geometrical space to all observations and initial template, it no longer remains true

Algorithm 2 Atlas estimation on functional shapes

Require: N functional shapes (X^i, f^i), initialization of the template fshape $(X, f) = (q_k, f_k)_{k=1,..,n}$ and the gradient descent step δ.

set $p^i = 0$ for all $i \in \{1, .., N\}$.
while Convergence **do**
  for i=1,..,N **do**
    Flow (X, f) by forward integration with p^i and store the deformed template at time 1 $\phi^{v_0^i}.(X, f) = (q_{k,1}^i, f_k)$.
    Compute the fcurrent representations of (X^i, f^i) and $\phi^{v_0^i}.(X, f)$.
    Compute the gradients of g^i with respect to the final point configuration $q_{k,1}^i$.
    Apply the backward integration of eq.(4.10) and deduce $\nabla_q J$ and $\nabla_{p_i} J$ as in eq.(4.11).
    Compute $\nabla_f J$ with eq.(4.12).
  end for
  Update variables :

$$q \leftarrow q + \delta \nabla_q J$$
$$f \leftarrow f + \delta \nabla_f J$$
$$p^i \leftarrow p^i + \delta \nabla_{p^i} J$$

  Compute new functional $J(q, f, (p^i))$ and possibly adapt step δ by line search.
end while
return template functional shape $(\overline{X}, \overline{f})$ and N initial momenta (p^i).

when it comes to rescaling. In this section, we precisely examine the effect of scaling both in the geometrical and the signal space on the functional J and its gradients. The objective is to make the energies and variables' updates in algorithm 2 invariant if for instance we apply any homothetic transformation to all the geometrical supports of the initial template and the data together or if we multiply all signals by a certain constant.

Let's consider a first homothety on the geometrical space, $x \mapsto \tilde{x} = \lambda_g x$ and homothety on the real-valued signals $f \mapsto \tilde{f} = \lambda_s f$. We assume that the parameters of kernels K_V and k_g are also rescaled by λ_g and that the one of kernel k_f is rescaled by λ_s so that we have : $\tilde{K}_V(\lambda_g x, \lambda_g y) = K_V(x, y)$, $\tilde{k}_g(\lambda_g x, \lambda_g y) = k_g(x, y)$ and $\tilde{k}_f(\lambda_s f, \lambda_s g) = k_f(f, g)$. That being set up, it is easy to see that the new momenta of the deformation have to be multiplied by the factor λ_g, i.e $\tilde{p} = \lambda_g p$ so that the displacement generated by $\tilde{p}$ is also rescaled by λ_g. The new energy of deformation is then $\tilde{E} = \sum_{i,j} \tilde{K}_V(\tilde{x}_i, \tilde{x}_j)\langle \tilde{p}_i, \tilde{p}_j \rangle = \lambda_g^2 E$. One can thus normalize the energy by the factor $1/\lambda_g^2$, which makes E independent of scale in the algorithm. Concerning the data attachment term, the dependency in λ_g is given by dot products between d-vectors proportional to the d-dimensional area of the shapes and therefore to λ_g^d. It results that the attachment term is proportional to λ_g^{2d} and has to be normalized by $1/\lambda_g^{2d}$ to be scale-invariant. Note incidentally that, without these normalizations, the balance between the regularity and the attachment terms is modified by scale if $d \neq 1$, which is not desirable in practice. As for the functional dependency, we see

that the attachment term is not affected by the scale λ_s.
Now, we examine the gradients' behavior. As we just said, we have, for the functional dependency of the criterion J properly renormalized, $\tilde{J}(\tilde{f}) = J(f) = J(\frac{\tilde{f}}{\lambda_s})$ and $\nabla_{\tilde{f}}\tilde{J} = \frac{1}{\lambda_s}\nabla_{\tilde{f}}J$. Therefore, the gradient on f must be multiplied by λ_s^2 in order to update the function f by a term which is proportional to the values of functional signals. In the same spirit, focusing on the dependency in the template point positions x, $\tilde{J}(\tilde{x}) = J(\frac{\tilde{x}}{\lambda_g}) \Rightarrow \nabla_{\tilde{x}}\tilde{J} = \frac{1}{\lambda_g}\nabla_{\tilde{x}}J$, so the gradient of J with respect to x has to be renormalized by a factor λ_g^2. Finally, we have again $\tilde{J}(\tilde{p}) = J(\frac{1}{\lambda_g}.\tilde{p})$ and just as previously, we find that the normalization factor in front of $\nabla_p J$ is also λ_g^2.
In practical applications of algorithm 2, one can have an estimation of typical scales λ_g and λ_f by simply looking at the average spatial and functional extensions of the observations. The normalizations that we have proposed then guarantee that using similar but rescaled data will provide comparable energies and energy decrease during gradient descent. Once again, we must insist that such a normalization is not absolutely fundamental in theory but it has the great advantage of normalizing the choice of some parameters in the algorithm, notably the balance between deformation cost and data attachment as well as the initial steps for the gradient descent, which could otherwise vary a lot according to the size of the observations.

Normalization with respect to template sampling

Using a similar method, we now focus on the behavior of algorithm 2 with respect to sampling of the discretized template. Namely, working for instance with surfaces, we wish to understand qualitatively the asymptotical behavior of the expressions of the functional J and the different gradients when the mesh on the template is refined.
Let $(X^{(n)}, f)$ be a sequence of discretized templates sampled from a continuous template fshape (X, f), with n the total number of points on $X^{(n)}$ such that $C_{(X^{(n)},f)} \xrightarrow{W'} C_{(X,f)}$ and $v_0^{(n)} \xrightarrow{V} v_0$. Since the initial vector field on $X^{(n)}$ is related to the momenta by the relation $v_0^{(n)}(x) = \sum_{k=1}^n K_V(q_k^{(n)}, x)p_k^{(n)}$, we see that $p^{(n)}$ is of asymptotical order $\frac{1}{n}$. The sequence of functionals

$$J(X^{(n)}, p_0^{(n)}) = \|v_0^{(n)}\|_V^2 + \|C_{(\phi^{v_0^{(n)}}(X^{(n)}),f)} - C_{(Y,f_Y)}\|_{W'}^2$$

then converges to its continuous counterpart $\|v_0\|_V^2 + \|C_{(\phi^{v_0}(X),f)} - C_{(Y,f_Y)}\|_{W'}^2$. The values of the deformation energies and attachment distances computed in the algorithm are therefore stable with the sampling of the template.
Let's now examine the gradients of the deformation energy E. The gradient with respect to p is given by :

$$\nabla_{p_l^{(n)}} E = 2\sum_{k=1}^n K_V(q_k^{(n)}, q_l^{(n)})p_k^{(n)} = 2v_0^{(n)}(q_k^{(n)})$$

which is asymptotically of order 1 (because it converges toward the vector field v_0). As noticed previously, the momenta $p_l^{(n)}$ are of order $\frac{1}{n}$ so that homogeneity requires to rescale the gradient $\nabla_{p^{(n)}} E$ by a factor $\frac{1}{n}$. In a similar way, we have

$$\nabla_{q_l^{(n)}} E = 2\langle\sum_{k=1}^n \partial_1 K_V(q_l^{(n)}, q_k^{(n)})p_k^{(n)}, p_l^{(n)}\rangle$$

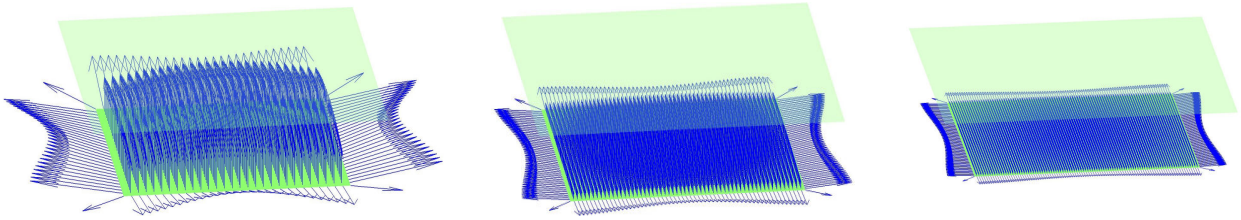


Figure 4.2: Gradient dependency in the number of points n. The figure shows the gradient of the fcurrent norm between two surfaces : the template (the underneath square) and a target (the above bigger square). The gradient is computed numerically for three different sampling of the template surface : 900 points (left), 3600 points (center) and 10000 points (right). We only display here the gradient with respect to the point positions represented as a vector at each of these points. Observe the global decrease when the sampling is higher as well as the boundary effect which makes the gradient higher on the boundary, particularly for high sampling.

and $\sum_{k=1}^{n} \partial_1 K_V(q_l^{(n)}, q_k^{(n)}) p_k^{(n)}$ converges toward $\int \partial_1 K_V(q, .)p,\, p_l^{(n)}$ is of order $\frac{1}{n}$ so $\nabla_{q^{(n)}} E$ is asymptotically proportional to $\frac{1}{n}$ and must be therefore normalized by a factor n.
Concerning the attachment term g, the gradients with respect to initial positions of the template is computed basically by flowing backward, with eq.(4.10), the gradients with respect to final variables q_1 and p_1. As stated previously, we have $\nabla_{p_1} g = 0$ and $\nabla_{q_1} g$ is the gradient of $\|C_{(\phi^{p_0}(X),f)} - C_{(Y,f_Y)}\|^2_{W'}$ with respect to the points of the shooted shape $\phi^{v_0}(X)$. In theorem 2.2.4 of chapter 2, we derived a general formula for the variation of such fcurrent norms relying on the generalization of Cartan's formula. As we explained, putting aside the boundary terms that we shall treat in the next section, the variation of the metric with respect to the displacement of an interior mesh cell is proportional to the d-volume of this cell. Thus for a shape sampled with n points, it decreases as $\frac{1}{n}$. It results that asymptotically $\nabla_{q_1} g \sim \frac{1}{n}$ and so is $\nabla_q g$ thanks to the linearity of the backward equation : this gives a renormalization by factor n for $\nabla_q g$. With Cartan's formula generalized to fcurrents (theorem 2.2.4), we have for the exact same reason that $\nabla_f g$ is at each position proportional to the d-volume of the corresponding cell and thus the normalization factor is again n.
To sum up the whole discussion, we conclude that consistency of the algorithm with respect to sampling of the template can be achieved by doing the following renormalization of gradients : for a template with n points, multiply $\nabla_q J$, $\nabla_f J$ by n and divide $\nabla_p J$ by n.

Boundary terms and regularization of gradients

Cartan's formula for functional currents reveals that the variation of W'-norms involves in addition integrals on the boundary of the submanifold X. These induce particular effects on the boundary of X in the template estimation process that must be accounted for when renormalizing gradients with respect to the template mesh accuracy.
Assuming that the $X^{(n)}$ are regularly discretized with n points, we may assume that the number of points on the boundaries $\partial X^{(n)}$ is of order $n^{1-1/d}$ and local $(d-1)$-volume elements on $\partial X^{(n)}$ are proportional to $n^{1/d-1}$. It results that, on the boundary points, $\nabla_{q_1} g$ depends on the size of the mesh asymptotically in $n^{1/d-1}$

as opposed to interior points where the dependency is in n^{-1}. There is therefore a multiplicative factor $n^{1/d}$ between the gradient on the boundary and in the interior of $X^{(n)}$, which is corroborated numerically by the simple example of figure 4.2. Now, in algorithm 2, $\nabla_{q_1} g$ serves as the initialization for the backward equation that computes $\nabla_q g$, and thus we get a similar behavior when we update the template X by the gradient $\nabla_q J$. Indeed, the gradient is computed for the L^2 metric on the set of points q, which results in updating the template point positions by simply adding the gradient. Therefore, these high boundary terms may cause, at certain steps of the gradient descent, undesired effects like self-crossing or changes in the topology of X.

One convenient solution to address such issues is to compute the gradient of X with respect to a more regular metric than L^2. Namely, one can consider a Gaussian kernel K_{reg} of scale $\lambda_{reg} > 0$ and compute $\tilde{\nabla}_q J = K_{reg}(q,q)\nabla_q J$. Then, $-K_{reg}(q,q)\nabla_q J$ is still a descent direction for the functional J, as the matrix $K_{reg}(q,q)$ is positive definite. However, this new gradient is **regularized** at a scale given by the regularization kernel and it can be thus considered as a sampling of a dense vector field belonging to the RKHS V_{reg}. The displacement of the template from its initial position in the gradient descent is then the integration of a sequence of vector fields living in V_{reg} so that, with sufficiently small steps δ in the gradient descent, it ensures that the final estimation of X is an element of the diffeomorphic orbit of its initial estimate. In more precise terms, one can actually show that the flow of the regularized gradients is indeed defined at all times, as stated in the following :

Property 4.2.1. *Let $q_s = (q_{k,s})_{k=1,..n}$ be the position of the template points at iteration s (which we consider as a continuum) and*

$$u_s(x) = -\sum_{k=1}^{n} K_{reg}(q_{s,k}, x)\nabla_{q_{s,k}} J$$

the vector field obtained by the regularized gradient at each iteration. Then the flow of the differential equation $\dot{y}(s) = u_s(y)$ exists for all $s \in \mathbb{R}_+$.

Proof. Due to the regularity of the regularization kernel K_{reg}, we see that the vector fields u_s are smooth for all s so that Cauchy-Lipschitz theorem gives the existence of maximal solutions. For a given initial condition x, the solution is defined on an interval $[0, s_f[$ and we only need to prove that $s_f = +\infty$. Let's denote in short $J(s)$ for $J(q_s, p^i_{0,s})$. Since we know that $|u_s|_\infty \leq \text{Cte.}|u_s|_{V_{reg}}$, this will be a consequence of the following estimate :

$$\int_0^{s_f} |u_s|^2_{V_{reg}} ds \leq J(0)\,. \tag{4.13}$$

We have indeed :

$$\begin{aligned}
|u_s|^2_{V_{reg}} &= \sum_{k=1}^{n} \left((\nabla_{q_{s,k}} J)^T \sum_{l=1}^{n} K_{reg}(q_{s,k}, q_{s,l}) \nabla_{q_{s,l}} J \right) \\
&= -\sum_{k=1}^{n} \left((\nabla_{q_{s,k}} J)^T u_s(q_{s,k}) \right) \\
&\leq -\sum_{k=1}^{n} (\nabla_{q_{s,k}} J)^T \frac{dq_{s,k}}{ds} \underbrace{-\sum_{i=1}^{N} (\nabla_{p^i_{0,s}} J)^T \frac{dp^i_{0,s}}{ds}}_{\geq 0} = -\frac{dJ}{ds}\,.
\end{aligned}$$

The positiveness of the second term is simply due to the fact that by definition $\frac{dp^i_{0,s}}{ds} = -\nabla_{p^i_{0,s}} J$. It results that $\int_0^{s_f} |u_s|^2_{V_{reg}} \leq J(0) - J(s_f) \leq J(0)$ which concludes the proof. □

The consequence is that regularization provides numerical stability of the template support by the algorithm. But incidentally, it constrains the template to live in a particular orbit and makes the algorithm dependent on the choice of the initialization. Even though smaller values of λ_{reg} allow more flexibility (while diminishing the stability), changes of topology as creation of holes or disconnections on X cannot be obtained by our approach and the initialization of the template is consequently crucial. We can argue however that topological changes in applications to computational anatomy are generally not desirable.
Concerning the momenta, the behavior is exactly dual. Indeed, p, as a covariable, lives in the dual of the space of vector fields V and need not be regular at all. In fact, once the minimum is reached, one has the transversality equation mentioned in the introduction (section 1.3.2) which imposes that :

$$p_1 = -\nabla_{q_1} g$$

This suggests that the momenta must be singular on the boundary of X (asymptotically in $n^{1-1/d}$ in the discrete setting), due once more to Cartan's formula for functional currents. Such singularities are difficult to create by the gradient descent process because $\nabla_p J$ is obtained numerically as a regularized vector field through the backward equation. It results that the framework will require numerous descent steps for surfaces with boundaries in order to reproduce this momentum's singularities, as we have also observed in practical applications. Thus, a way to increase the convergence speed of the algorithm is to simply extract the boundary of the template shape (which can be done only once at the beginning since the previous regularization on the template position's evolution ensures that boundary points will remain boundary points during the whole gradient descent) and, after each gradient computation, multiply $\nabla_p J$ on the boundary by a factor $n^{1/d}$. As we have confirmed on a few numerical examples, this straightforward procedure allows to significantly decrease the number of iterations for a predefined convergence threshold up to a factor 2 or 3.

First results and discussion

We now apply algorithm 2, with the additional normalizations and regularizations of gradients described in the previous subsections, for several simulated and real datasets involving sets of shapes carrying different real-valued signals. As mentioned, the algorithm estimates a template functional shape $(\overline{X}, \overline{f})$ that can be interpreted as a baseline of the data ensemble and a set of deformations mapping the template on each subject represented by their initial momenta (p^i). We first show, on the simple example of figures 4.3 and 4.4, the interest of doing geometrico-functional estimation together instead of separating both. This is precisely made possible with the approach of functional currents. As we see, those norms enforce momenta to match regions with similar signals on each other which has also important impact on the estimated template fshape itself. Indeed, the signal we obtain on the template is much closer to the expected pattern than the one given by separating geometry and signal (figure 4.4) which basically averages the signals located at corresponding

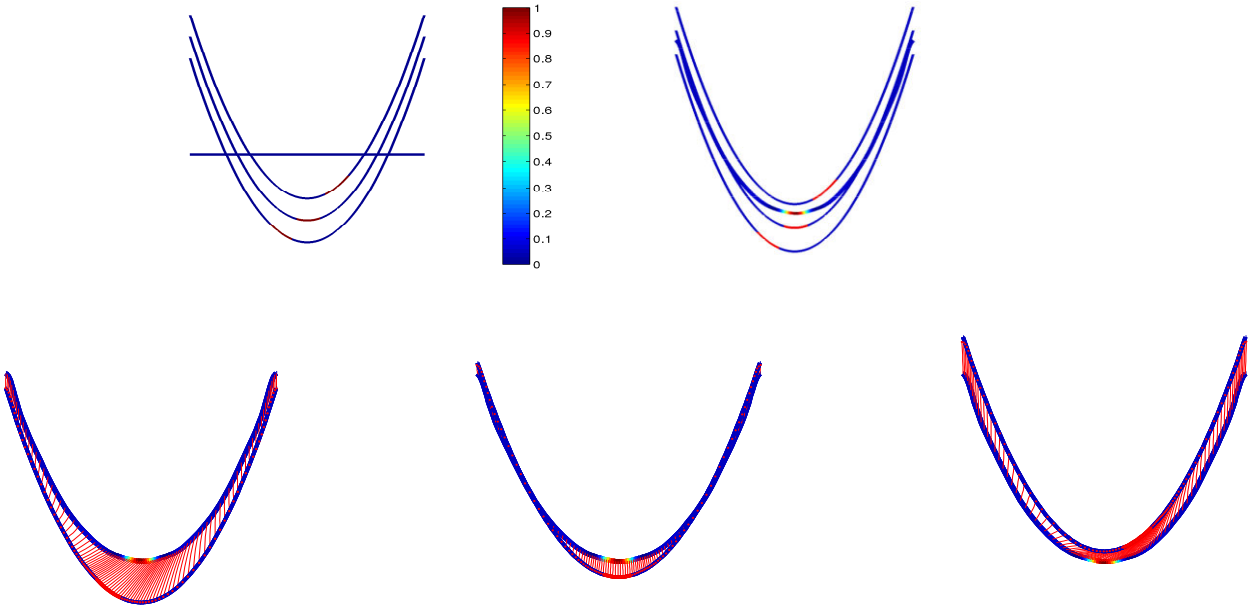


Figure 4.3: Template estimation on a set of three colored curves based on functional currents. The template and the deformations are estimated using both geometrical and functional information at the same time through the fcurrent norm. On the first row, we show the three subjects and the initialization of the template (on the left figure) and the template estimated at the end of the gradient descent (right figure). On the second row are displayed the trajectories corresponding to the deformations from the template to each one of the subjects.

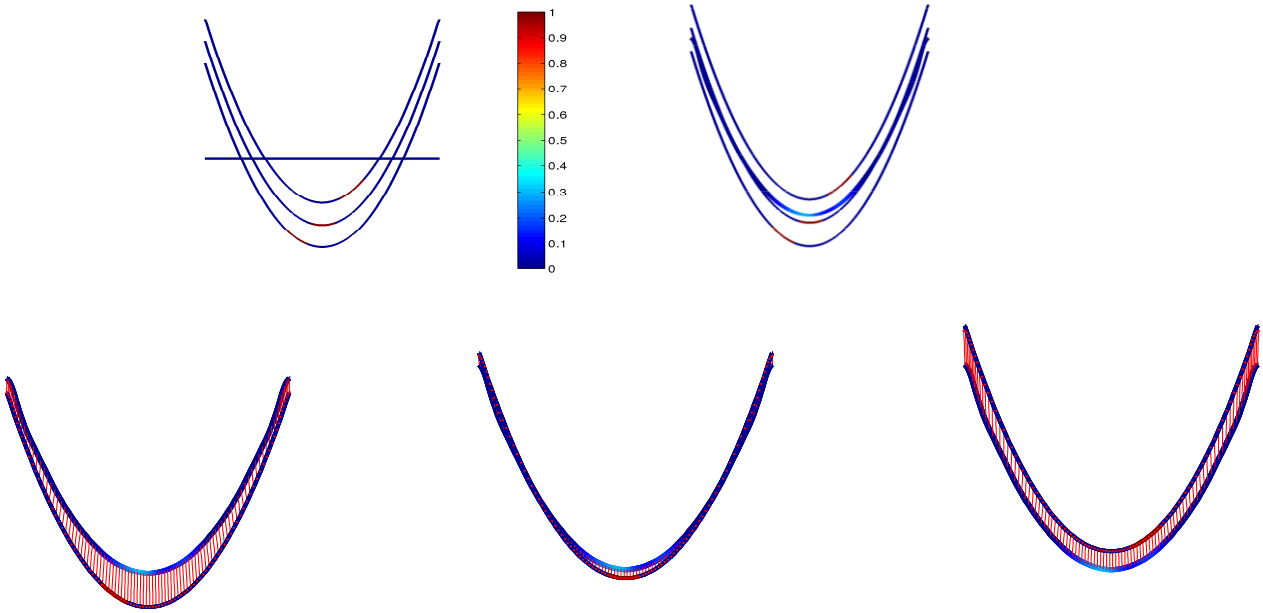


Figure 4.4: The same example for which the geometry and the function are computed separately, by first estimating a template shape by currents and then pulling back signals from the different subjects on the template. Observe that the resulting signal on the template is a more spread and averaged crenel.

positions but provided by purely geometrical matchings. This observation argues in favor of the algorithm we propose because, in many cases, it makes possible to learn signals with a better accuracy and resolution than all methods that intend to decouple the shape from the functional information.
On the example of figure 4.5, we show the interest of the gradient regularization for the template displacement $\nabla_q J$ developed in 4.2.1. Observe in particular, even on such a simple example, the change of topology of the template in the non-regularized case which is mainly a consequence of the higher gradients on the boundaries which can provoke undesired self-crossings.
The third example is a template computation on a set of 'cup' surfaces with Gaussian decreasing signals centered at the base. Notice the evolution of the template fshape throughout the gradient descent steps

4.2.2 Estimation of functional residuals

A richer model for functional variability

The atlas estimation procedure of 4.2.1 has the advantage of providing a geometrico-functional baseline for a given dataset, and deformations whose variability can be then statistically analysed based on the tangent space representation provided by the momenta, as pioneered in [79]. This shall be detailed in section 4.3. However, it only accounts for the geometrical variability of the dataset around the template. In generic situations, signals are also likely to vary across individuals and this variability can have a statistical significance. For example, in the case of the OCT dataset presented previously, the thickness is expected to decrease in several locations as a

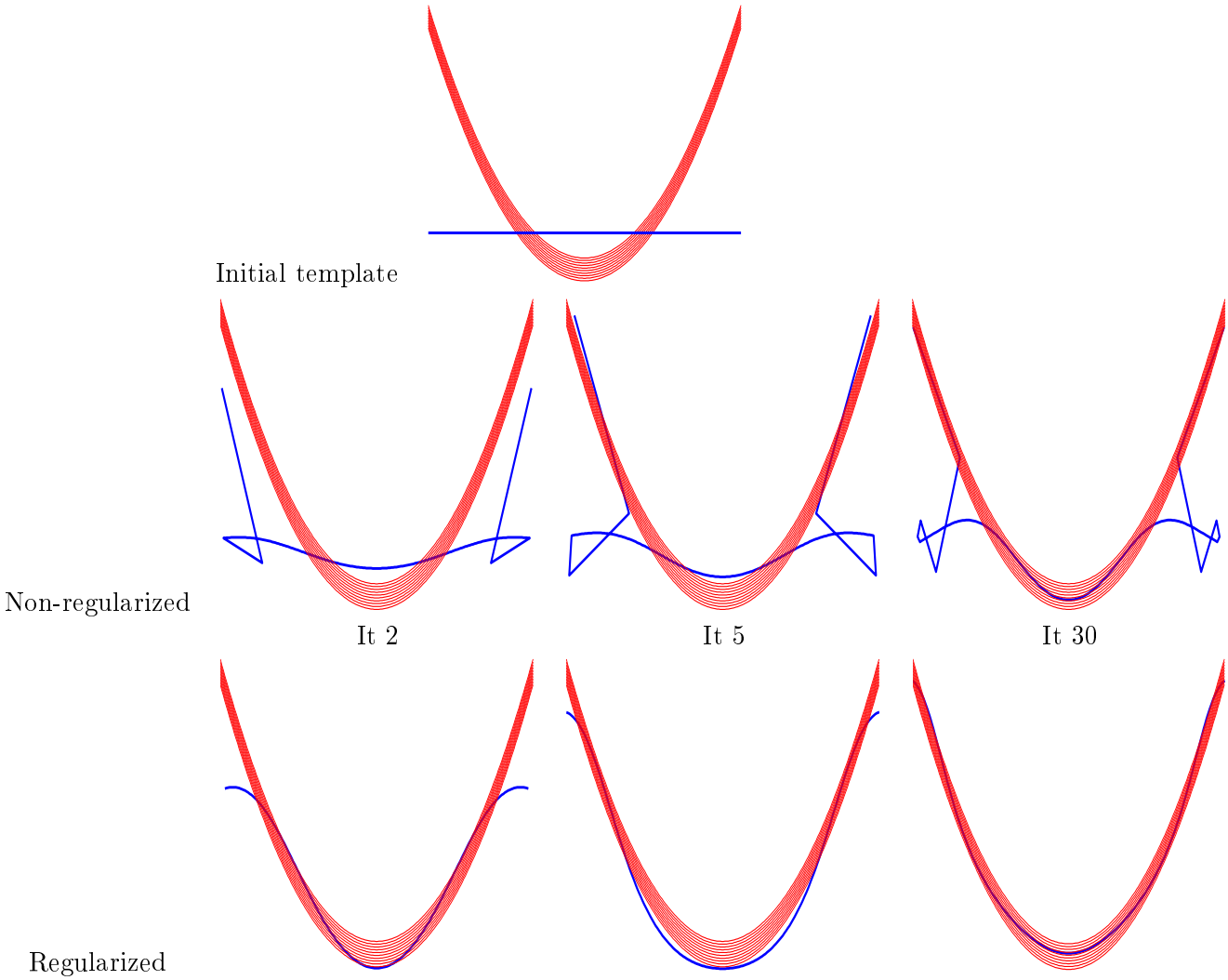


Figure 4.5: Several intermediate iterations in the template estimation algorithm both for non-regularized and regularized gradients. Figures show the dataset of ten close parabolas (which are pure geometrical shapes in that case) and the evolution of the template (in blue) at different steps of the gradient descent.

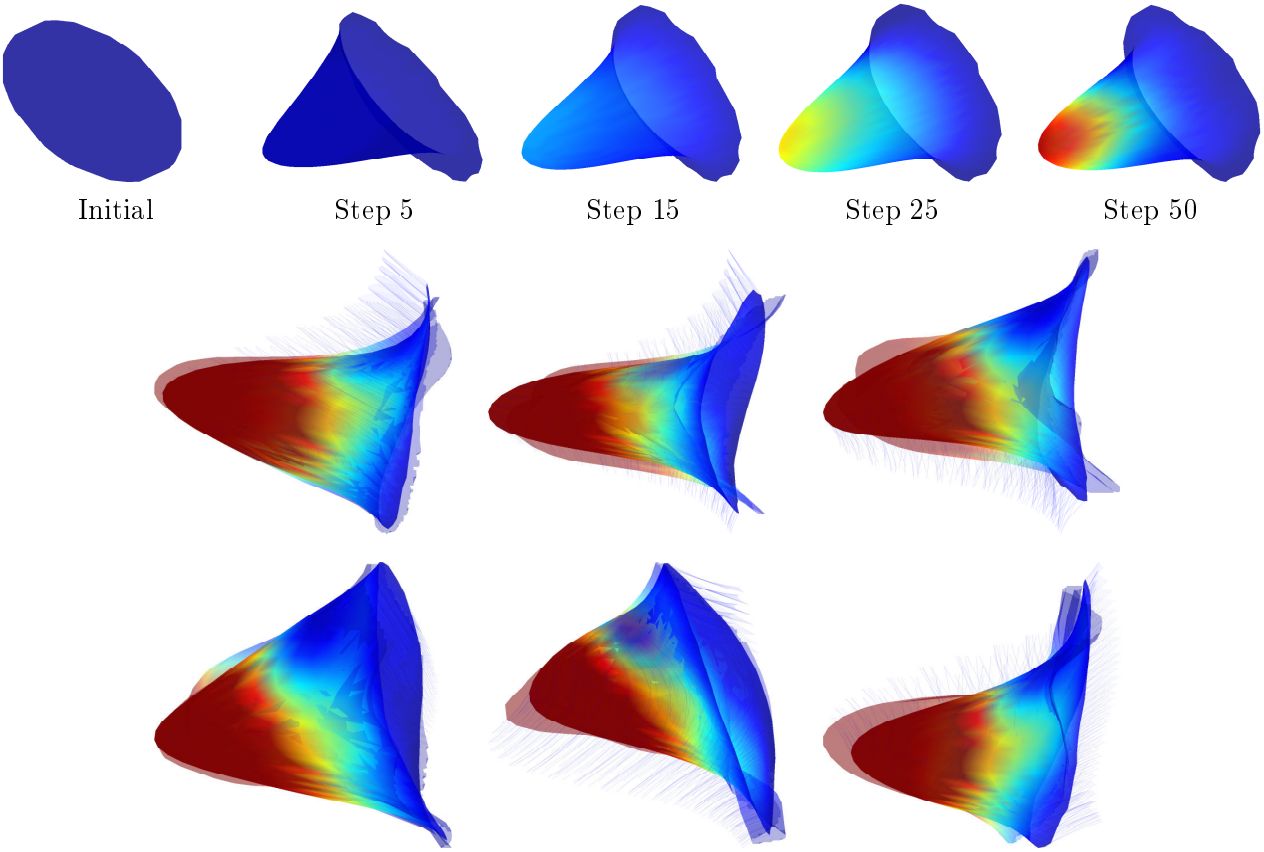


Figure 4.6: Template estimation on a set of 6 'cup' shapes with artificial Gaussian signals. On top, the evolution of the template during gradient descent. On the bottom, we show the final results of registration between the template and each of the subjects : the more transparent surface is the subject, the solid one is the deformed template.

consequence of the glaucoma so that inter-individual thickness variability can be a highly relevant biomarker to perform classification. With our previous model, this variability is contained in the residuals of matching to each subject, namely the fcurrents associated to $\mathcal{S}_i - (\phi^{v_0^i})_*\overline{\mathcal{S}}$, which could be also analysed from a statistical point of view, as elements in the infinite dimensional space of fcurrents. The essential drawback is that these terms do not only represent the functional variability but also include high-frequency geometrical variations not captured by the deformations as well as noise and artifacts of the datasets. In addition, such residuals are nearly impossible to visualize except as linear combinations of Dirac fcurrents and thus making the interpretability of discriminative directions for PCA or LDA practically intractable.
This motivates an adaptation of the previous framework that would estimate, in addition to the template and deformations, a *functional residual* for each subject, namely a function defined on the template that measures the gap between the baseline and subject's functional. This is precisely the model introduced in 1.4.1 and already implemented for registration in 2.3.2. In that case, a generative model that we can introduce is the following :

$$(X^i, f^i) = \left(\phi^i(\overline{X}), (\overline{f} + \zeta^i) \circ \phi^{i,-1}\right) + \epsilon_i \tag{4.14}$$

The data are considered as generated by a template functional shape with a submanifold $\overline{X}$ and a signal $\overline{f}$ on which is added residual functions $\zeta^i : \overline{X} \to \mathbb{R}$ and deformed by diffeomorphism ϕ^i. Note that we consider the residuals (ζ^i) to be defined on the template shape and thus added before deformation instead of after. The reason for this choice is essentially that residuals will then share a common geometrical support, enabling easier numerical estimation and a much nicer framework for statistics (cf section 4.3). Just as we did for the previous model, we can derive a formal MAP estimator associated to eq.(4.14). We represent functional shapes by fcurrents and assume Gaussian density priors on the residual fcurrents ϵ_i. We have now hidden variables that are both the deformations ϕ^i parametrized by the initial momenta p_0^i on which we assume a prior probability distribution as given by eq.(4.3), and the residuals ζ_i which we shall consider as L^2 functions on $\overline{X}$ with a prior distribution :

$$p_{res}(\zeta|\overline{X}) = C_{res}.\exp(-\|\zeta\|^2_{L^2(\overline{X})}/(2\sigma^2_{res})) \tag{4.15}$$

where we remind that $\|\zeta\|_{L^2(\overline{X})}$ stands for the L^2 norm on the submanifold $\overline{X}$. Then, one can write an equivalent MAP estimator in that context and using the Fast Approximation with Modes, it leads to a minimization problem of the form :

$$\min_{(X,f,p_0^i,\zeta^i)} \sum_{i=1}^{N} \left(E(p_0^i) + \gamma_{res}\|\zeta^i\|^2_{L^2(X)} + \gamma_W \|C_{(X^i,f^i)} - C_{\left(\phi^{p_0^i}(X),(f+\zeta^i)\circ(\phi^{p_0^i})^{-1}\right)}\|^2_{W'} \right) \tag{4.16}$$

where $\gamma_{res}, \gamma_W > 0$ are the weight of the respective terms. Compared to the previous model of eq.(4.8), we have additional variables ζ^i to estimate and the presence of an additional term in the functional that depends on the ζ^i's and the template shape X. Note that if X and f are fixed, then the problem given in eq.(4.8) reduces to solve the registration with residual of (X, f) on each (X^i, f^i), which is exactly what was studied in chapter 2 section 2.3.2. Algorithmically, the framework doesn't change much except for the computation of additional gradients, which we detail in the next subsection.

New algorithm

As in algorithm 2, we propose a gradient descent algorithm that simultaneously optimizes the functional with respect to all variables together. To achieve this, we only need to make explicit the computation of the new terms $E_f(\zeta^i) \doteq \|\zeta^i\|^2_{L^2(X)}$ and express the gradients with respect to the new variables ζ^i.
The computations of the L^2-norms $E_f(\zeta^i)$ are almost straightforward because, for a discretized shape X and signals ζ^i, these are basically the sum of the volumes of the mesh cells weighted by the interpolated values of the functions ζ^i to the center of the cells. Namely, with the notations of section 4.2.1, we have :

$$E_f(\zeta) = \sum_{l=1}^{m} \tilde{\zeta}_l^2 . |\xi_l| \tag{4.17}$$

where for each k, $\tilde{q}_k$ is the position of the center of the lth cell, $|\xi_l|$ equals its d-dimensional volume and $\tilde{\zeta}_l$ is the average of the cell vertices' signal values.
Now, differentiating eq.(4.17) with respect to the template point positions simply requires to compute gradients of the local volumes $|\xi_l|$ with respect to the q_k's, which is the same computation as we did for the gradient of the fcurrent attachment distances. The gradient with respect to the residual functions' values (ζ_k) is :

$$\nabla_{\zeta_k} E_f = \sum_{l=1}^{m} 2\tilde{\zeta}_l \nabla_{\zeta_k} \tilde{\zeta}_l . |\xi_l|$$

and gradients $\nabla_{\zeta_k} \tilde{\zeta}_l$ are simply computed by redistribution from centers to vertices. Finally, the gradient of the attachment distance terms $g^i = \|C_{(X^i,f^i)} - C_{(\phi^i(X),(f+\zeta^i)\circ\phi^{i,-1})}\|^2_{W'}$ with respect to the residuals ζ^i can be computed in the exact same way as with respect to f, cf eq.(4.12). In fact, we see that $\nabla_f g = \sum_{i=1}^{N} \nabla_{\zeta^i} g^i$.

The last important issue is the one of gradient normalizations and regularizations that has been thoroughly discussed in section 4.2.1. Of course, the different conclusions hold with this new model and we only need to examine the new penalization terms and the behavior with respect to the new variables ζ^i. As far as scales are concerned, the attachment term g and $\nabla_{\zeta_i} g$ can be treated exactly the same as with respect to f (cf 4.2.1). $E_f(\zeta^i)$ being the squared L^2 norm of ζ^i on X, rescaling by λ_g on the geometrical part and by λ_f on the signal part multiplies such terms by $\lambda_g^d \lambda_f^2$ and thus we achieve scale invariance by applying renormalization factors $1/(\lambda_g^d \lambda_f^2)$ on the E_f. For similar reasons, the gradient $\nabla_{\zeta^i} E_f$ is to be renormalized by $1/\lambda_g^d$ and the gradient $\nabla_q E_f$ by λ_g^2/λ_f^2.
On the dependency with respect to the template sampling, using the notations and reasoning of subsection 4.2.1, we easily see that $\nabla_{\zeta_k} E_f$ is proportional to the local volume element on $X^{(n)}$ which is asymptotically of order $1/n$ and thus these terms must be renormalized by a factor n. The gradient of the term E_f with respect to the positions of the template points q can be seen as a particular instance of Cartan's formula for functional currents (taking $\omega(x,m) = \zeta(x)$), and therefore it involves some boundary effects as in 4.2.1. A good solution is again to regularize the gradient by the kernel $K_{reg}(q,q)$. We sum up all the different steps of the full process in algorithm 3. Note that this new algorithm is a generalization of the previous one : taking the parameter $\gamma_{res} \to +\infty$ and a null gradient step in the residuals ζ^i, we exactly get back to algorithm 2.

Algorithm 3 Atlas estimation on functional shapes with functional residuals

Require: N functional shapes (X^i, f^i), initialization of the template fshape $(X, f) = (q_k, f_k)_{k=1,..,n}$ and the gradient descent step δ.

set $p^i = 0$ and $\zeta^i = 0$ for all $i \in \{1, .., N\}$.
while Convergence **do**
  for i=1,..,N **do**
    Flow (X, f) by forward integration with p^i and store the deformed template at time 1 $\phi^{v_0^i}.(X, f) = (q_{k,1}^i, f_k + \zeta_k^i)$.
    Compute the fcurrent representations of (X^i, f^i) and $\phi^{v_0^i}.(X, f + \zeta^i)$.
    Compute the gradients of g^i with respect to the final point configuration $q_{k,1}^i$.
    Apply the backward integration of eq.(4.10) and deduce $\nabla_q g^i$ and $\nabla_{p_i} g^i$.
    Compute $\nabla_q E$ and $\nabla_{p^i} E$ with eq.(4.9).
    Compute $\nabla_{\zeta_i} J = \nabla_{\zeta^i} g^i + \nabla_{\zeta^i} E_f$
  end for
  Deduce $\nabla_f J = \sum_{i=1}^N \nabla_{\zeta_i} J$.
  Normalize with respect to scale and sampling all the gradients.
  Apply regularization kernel to $\nabla_q J$.
  Update variables :

$$
\begin{aligned}
q &\leftarrow q + \delta \nabla_q J \\
f &\leftarrow f + \delta \nabla_f J \\
p^i &\leftarrow p^i + \delta \nabla_{p^i} J \\
\zeta^i &\leftarrow \zeta^i + \delta \nabla_{\zeta^i} J
\end{aligned}
$$

  Compute new functional $J(q, f, (p^i), (\zeta^i))$ and adapt step δ by line search.
end while
return template functional shape $(\overline{X}, \overline{f})$, N initial momenta (p^i) and functional residuals (ζ^i).

4.2.3 Results on the retina dataset

We have applied algorithms 2 and 3 to a dataset of retina membranes. This is a current joint project with Dr Faisal Beg, Dr Evgeniy Lebed, Dr Marinko Sarunic and Sieun Lee from Biomedical Research group in Simon Fraser University. The data was acquired using Optimal Coherence Tomography (OCT) coupled with compressive sampling for rapid image acquisition, as introduced in [87]. The raw data consist in 3D images of a portion of the retina for a certain number of subjects, which are then regularized. The next pre-processing step is the segmentation of the different layers and membranes in the retina which are basically (from the most superficial to the deepest) : the inner layer membrane (ILM), the nerve fiber layer surface (NFL), the Brook's membrane (BM) and the Choroid surface. This is shown in the images of figure 4.7. In addition to that, was computed two thickness maps : one representing the thickness between the ILM and NFL surfaces that we will call the NFL thickness and the Choroidal thickness between the BM and the Choroid. In this manuscript, we will focus only on the NFL and its thickness even though some work was also

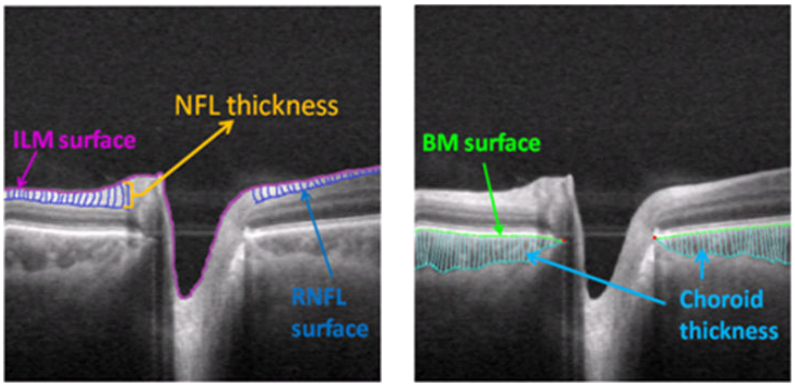


Figure 4.7: The different layers and membranes observed in the retina 3D images acquired by OCT.

carried on the Choroid as well as on the shape of the ILM. One of the reason is that the estimation of the thickness on the NFL is more accurate and provides patterns that are more easily interpretable.
Thus, restricted to the NFL, the dataset obtained after pre-processing is made of different subjects' surfaces on which we have a functional data representing the thickness of the layer at each point. The database we have been working on contain such functional surfaces for 30 different subjects put in the same coordinate system, some of which are shown in figure 4.8. The geometry of the surfaces is quite close to be planar in most subjects but the orientation varies from subject to subject. Another important element is the opening (through which the optical nerve passes in the retina) that also varies in size an position within the dataset. Thicknesses offer more or less common patterns as well but also significant local differences.
Now, the main clinical issue that makes the interest of this data is the study of the **glaucoma**, an ocular pathology that is reflected by retina and optical nerve damages leading progressively to partial or even complete visual field loss. The dataset we just presented was acquired for both controls and glaucomatous subjects. The question raised is the possibility to characterize the presence and stage of the disease by the shape of the retina layers. In our particular case, we would wish to understand **the effects of glaucoma both on the shape and the thickness of the NFL**.
One usual empirical hypothesis formulated by clinicians is that glaucoma induces a degeneration of tissues provoking some localized thickness decrease, possibly coming along with membranes' deformations. In order to assess it qualitatively in the first place, we applied **algorithm 2** to estimate a template for the control population and a template for the subjects with confirmed glaucoma. Both these template functional shapes are shown in figure 4.9. It shows indeed quite clearly an average loss of thickness for glaucomatous subjects compared to controls particularly in some areas around the opening as well as a rather significant change in the orientation of the membrane as we see on the images of the first row in figure 4.9 and some small differences in the shape of the opening.
This, however, constitute only a qualitative study at this point. For the thickness maps, it is only giving some mean thickness for both classes but not the intra-class and inter-class covariances. As for the shapes of the NFL surfaces, since we have two templates, the set of deformations' momenta are lying on two different surfaces accordingly to the class. Doing further statistics would then require to estimate likelihoods of deformations around each class, as introduced for grey-level images in

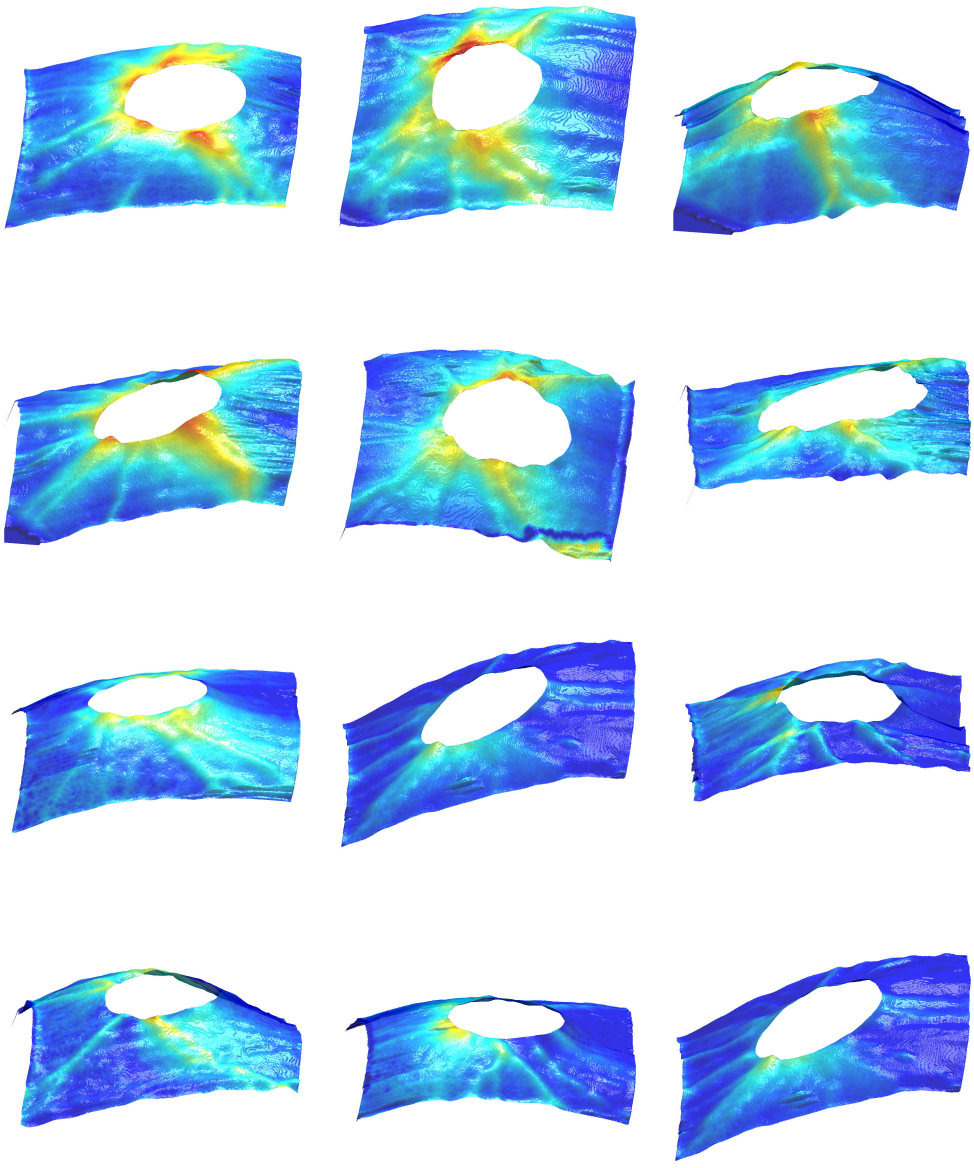
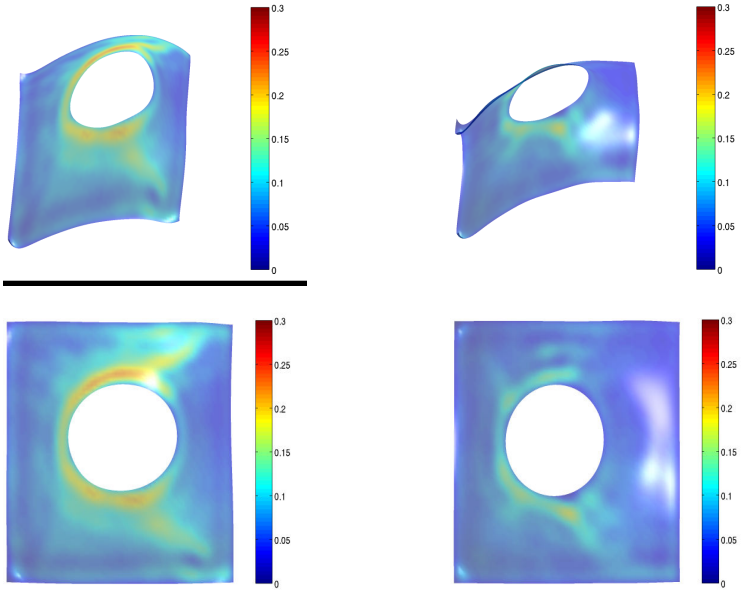


Figure 4.8: Twelve NFL surfaces with their thickness from the database provided by Biomedical Optics Research Group. The two first rows are normal subjects, the two last ones are subjects affected by Glaucoma. The chosen view is the same for all images as well as the color map (blue for small thickness, red for large).



Template for normal subjects Template for glaucomatous subjects

Figure 4.9: Two views of the templates for both classes of subjects estimated with algorithm 2. The initialization was, in both case, a square with a centered opening, with all functional values at zero.

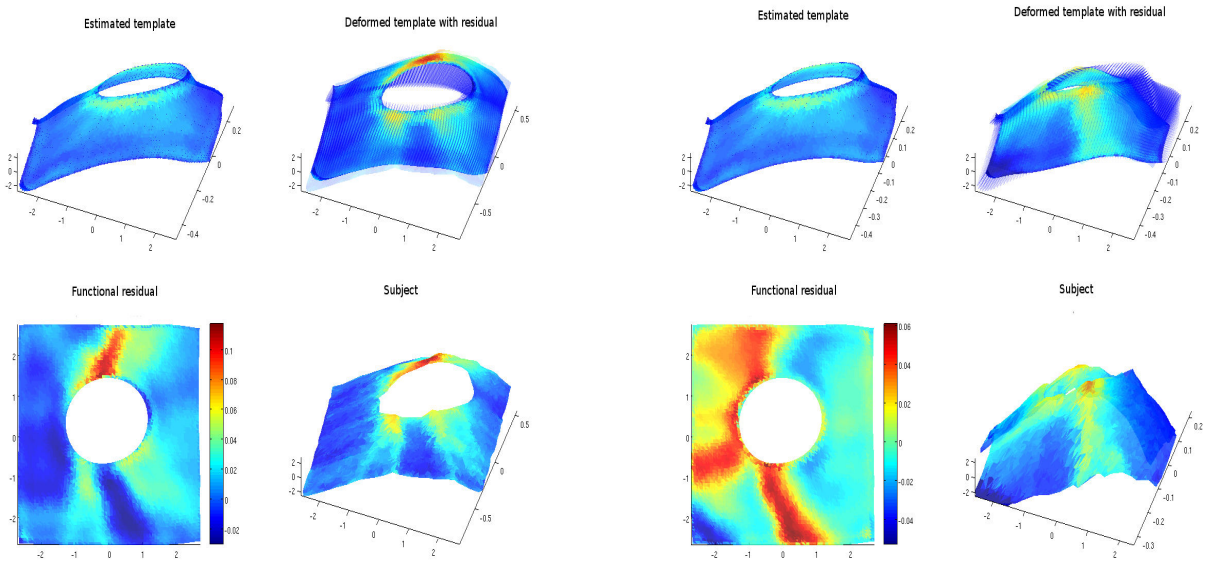


Figure 4.10: Deformation and functional residuals obtained by algorithm 3 for two subjects.

[6]. Instead, the approach we implemented with algorithm 3 is meant to simplify the following statistical analysis a little bit. It estimates only **one single template** functional shape and for each subject, a **momentum** p_0^i for the deformation of the template to the subject and a **residual thickness** ζ^i so that, now, all momenta and residuals for both classes are defined on this common template support and can be studied all together. This enables a deeper statistical analysis of the variability and the design of classification algorithms which we postpone the exposition, together with results on this dataset, to section 4.3.

4.2.4 A well-posed continuous framework for atlas estimation

As we have mentioned, the two previous models of atlas estimation that we derived are essentially valid in the discrete situation where all variables are assumed to live in finite-dimensional spaces. From the algorithmic point of view, this is of course a natural assumption and we have assessed on several examples that the algorithms seem to provide expected and stable results. Yet, the question of having an underlying continuous model that, among other properties, guarantees theoretically the **existence** of a minimizer for the atlas estimation problem remains a very important point for inquiry. It is quite clear that the formulation of eq.(4.16) does not provide a well-posed problem in the continuous situation, first because one needs to give a meaning to the minimization over the 'shape' X in that case. In this section, we intend to address these issues from the theoretical perspective : the corresponding implementations of such frameworks, although not very far from the ones previously presented, will be left for future work.

Hence the first point is to give a proper setting for the search of the template shape X, since the set of all possible shapes (for instance of all rectifiable subsets) lacks structure to consider the optimization over this whole set and it is very unlikely that the existence of a minimum for atlas estimation could be proved in this large space. Instead, we can follow the idea of *hypertemplate* introduced for images in [57] and later adapted to surface template estimation in [58]. Fixing a hypertemplate shape X_0, we shall look for templates that are constrained to live in the orbit of X_0 for a certain group of diffeomorphisms. Namely, let's consider a space of admissible vector fields V_0 and its associated group of diffeomorphisms G_{V_0}. Let $\mathcal{X}_0 \doteq G_{V_0}.X_0$ be the orbit of the hypertemplate under the action of G_{V_0}. Then (cf 1.2.3) we have the Riemannian distance $d_{\mathcal{X}_0}$ on $\mathcal{X}_0$ defined by :

$$d_{\mathcal{X}_0}(X_1, X_2) = \inf_{v^0 \in L^2_{V_0}} \left\{ \left(\int_0^1 \|v_t^0\|^2_{V_0} dt \right)^{\frac{1}{2}} \ / \ \phi_1^{v^0}.X_1 = X_2 \right\} .$$

In that setting, we can thus replace the optimization over X by an optimization over a vector field v^0 in $L^2_{V_0}$ that displaces the initial hypertemplate in the orbit. Note that the gradient regularization procedure that we applied in the previous algorithms by convolving gradients through a kernel at each gradient descent step can be seen as some kind of greedy way to also restrict the updates of the template in the orbit of the initialization (this is formulated by property 4.2.1). Here, we have in addition the metric $d_{\mathcal{X}_0}$ that can be used to penalize the displacements of the template from its initial position : such penalization should hopefully guarantee a compactness property for X needed for proving existence of solutions to the atlas problem.

Thus, we introduce the following functional :

$$J(X,f,(v^i),(\zeta^i)) = \frac{\gamma_0}{2} d_{\mathcal{X}_0}(X_0,X)^2 + \frac{\gamma_f}{2}\|f\|_{L^2(X)} + \frac{1}{2}\sum_{i=1}^{N}\left(\int_0^1 \|v_t^i\|_V^2 + \gamma_\zeta \|\zeta_i\|_{L^2(X)}^2 + \gamma_W \|C_{(X^i,f^i)} - C_{(\phi^{v_i}(X),(f+\zeta^i)\circ(\phi^{v_i})^{-1})}\|_{W'}^2\right) \tag{4.18}$$

Compared to the one of eq.(4.16), we essentially add penalty terms for the template displacement from its initialization and for the baseline signal f on X. The atlas estimation for the group of fshapes (X^i, f^i) writes :

$$\inf\{J(X,f,(v^i),(\zeta^i)) \mid X \in \mathcal{X}_0,\ f \in L^2(X),\ (v^i) \in (L_V^2)^N,\ (\zeta^i) \in (L^2(X))^N\} \tag{4.19}$$

Note that, here, the minimization over $X \in \mathcal{X}_0$ makes sense because it is equivalent to minimizing with respect to a time-varying vector field $v^0 \in L^2_{V_0}$ by setting $X = \phi^{v^0}(X_0)$. Proving existence of a minimizer to eq.(4.19) is however not immediate, the essential reason being that we have no weak semi-continuity in $L^2(X)$ of the application $f \mapsto \|C_{(X^i,f^i)} - C_{(\phi^{v_i}(X),(f+\zeta^i)\circ(\phi^{v_i})^{-1})}\|_{W'}^2$. Yet the result we have is summed up by the following theorem :

Theorem 4.2.1. *Assume that W is continuously embedded in $C_0^2(E\times\mathbb{R},\Lambda^d E^*)$, that X_0 and $(X^i)_{1\le i\le N}$ are finite volume bounded oriented d-dimensional rectifiable subsets and that $f^i \in L^2(X^i)$ for $1 \le i \le N$. Assume that γ_f/γ_W and γ_ζ/γ_W are large enough with $\gamma_f, \gamma_W, \gamma_\zeta \ge 0$ and $\gamma_0 > 0$. Then*

$$J(X,f,(\zeta^i),(v^i)) \doteq \frac{\gamma_0}{2} d_{\mathcal{X}_0}(X_0,X)^2 + \frac{\gamma_f}{2}\int_X |f(x)|^2 d\mathcal{H}^d(x) + \frac{1}{2}\sum_{i=1}^{N}\left(\|v^i\|_{L^2([0,1],V)}^2 + \gamma_\zeta \int_X |\zeta^i(x)|^2 d\mathcal{H}^d(x) + \gamma_W \|C_{X^i,f^i} - C_{\phi^{v^i}(X),f+\zeta^i}\|_{W'}^2\right) \tag{4.20}$$

achieves its minimum on $\{(X,f,\zeta^i,(v^i)) \mid X \in \mathcal{X}_0,\ f \in L^2(X),\ \zeta = (\zeta^i) \in L^2(X)^N,\ (v^i) \in L^2([0,1],V)^N\}$ and any minimizer $(X_*, f_*, (\zeta_*^i), (v_*^i))$ is such that f_* and ζ_*^i for $1 \le i \le N$ are in $L^\infty(X_*)$. Moreover, if X_0 is also a C^p submanifold and $W \hookrightarrow C_0^m(E\times\mathbb{R},\Lambda^d E^*)$ with $m \ge \max\{p,2\}$, then f_* and the ζ_*^i's are in $C^{p-1}(X_*)$.

The full proof of theorem 4.2.1 is quite technical and is presented in appendix A. The most subtle part is the existence of the fshape template (X, f) : the proof we derive follows basically the so-called *direct method* of geometric measure theory where we first show the existence of a minimizer in a proper space of currents or measures and then show that this solution does indeed result from a true fshape. We emphasize that the existence holds for the quite weak assumption of L^2 regularity on the signals but at the price of sufficient penalties on the L^2 norms. It's also worth noticing the regularizing effect of fcurrents' norms on the signal functions f_* and ζ_*^i since the solutions appear to be more regular as we increase the regularity of the fcurrents' kernel.

4.3 Statistical analysis of momenta and residuals

Starting from the atlases estimated by previous algorithms, the further step is to go beyond the simple qualitative observation of means by first analyzing the intra and inter-group variability and provide a quantitative classification criterion. In the case of geometrico-functional atlases given by algorithm 3, this can based on the study of deformations represented by the initial momenta p_0^i as well as on the functional residuals ζ_i. In both cases, the central issue is dimensionality reduction in the space of the subjects. Namely, one wants to learn a lower dimensional subspace that can best describe the principal modes of variation in the dataset and separate the different classes. From a purely geometrical point of view, the description of deformations by momenta in LDDMM gives a tangent space representation of diffeomorphisms, which was first coupled to statistical analysis in [79] in the case of images and later on applied to curves and surfaces datasets as well [30, 36]. The general idea in all cases is to perform Principal Component Analysis (PCA) and construct classifiers based on the directions provided by the PCA.
Following similar ideas, we will first recall some classical features of dimensionality reduction and classification in a generic finite dimensional vector space and evaluate the alternative possibility of linear discriminant analysis (LDA). We shall then focus on the specific cases of momenta and functional residuals and show a few results of classification based on such algorithms for the previous OCT dataset.

4.3.1 Dimensionality reduction

The problem of dimensionality reduction is an extremely wide topic in statistical learning and it is clearly out of the scope of this thesis to pretend to be exhaustive in any way. Our more humble purpose is to review briefly the learning methods that we actually experimented with in our applications. In this subsection, we first consider the generic framework of data reduction and place ourselves in the case of data lying in the finite dimensional euclidean space $\mathbb{R}^p$. The collection of 'subjects' in that setting is thus a set of points in $\mathbb{R}^p$ that we will denote $(x_i)_{i=1,..,N}$ and $\overline{x} = \frac{1}{N}\sum_{i=1}^N x_i$ the mean of the dataset. We will also denote X the centered data matrix $X = (x_i - \overline{x}) \in \mathbb{R}^{p\times N}$. Up to this global translation of the data, we shall assume from now on that the dataset is centered.

Principal Component Analysis

The most common and simple technique for analyzing such multivariate linear data is the PCA and has been extensively used in many situations. The idea behind the method is to find an orthonormal basis of a lower dimensional subspace on which the projection of the data points has a maximum variance. The PCA thus consists in computing iteratively the principal components $w_1, .., w_k, .. \in \mathbb{R}^p$. The first one is obtained as the unit vector maximizing the variance of the projections $x_i^T w_1$ i.e

$$w_1 = \arg\max_{\|w\|=1} \|X^T w\|^2 = \arg\max_{\|w\|=1} w^T (XX^T) w$$

The kth component is given by solving the previous problem on the new data matrix $X - \sum_{s=1}^{k-1} w_s w_s^T X$ obtained by just removing to each data point its projection on the subspace generated by the $(k-1)$ previous principal components. We will also call these principal components *principal modes* of variation since they encode the

directions of main variations in the dataset. One of the great advantage of PCA is that it can be also computed directly as the eigenvectors of the empirical covariance matrix XX^T in the decreasing order of the eigenvalues ([47]). Thus the matrix W of all the principal components is obtained by the eigenvalues decomposition of the symmetric matrix XX^T, which writes :

$$XX^T = W^T \Lambda W$$

where Λ is the diagonal matrix of the (positive) eigenvalues. Observe, since the rank of XX^T is less than $\min(N, p)$, that there are at most $\min(N, p)$ non-zero eigenvalues. However, the general paradigm is that the data can be represented with much less components and that only a few ones should be able to capture the essential part of the statistical variability. For instance, in the degenerate case where all the data lie in a low-dimensional subspace of $\mathbb{R}^p$, the first eigenvectors will precisely be vectors of that subspace. The interest of PCA is twofold : it is first a way to extract the directions of main variance of the dataset in the original space but it provides in addition a lower dimensional subspace on which to project and visualize the data.
Now, in situations when one has a learning set of data consisting in several subjects belonging to different known groups or classes, this reduced subspace can be then used typically to learn a classifier from the data, for instance using various clustering methods in the PCA subspaces.

Remark 4.3.1. *An important thing to note also is that PCA is based exclusively on the covariance structure within the dataset. The component vectors necessarily belong to the vector space spanned by the data points. In this respect, it is equivalent to perform PCA by working on the linear subspace generated by the data and replace the data matrix by $G^{\frac{1}{2}}$ where $G = X^T X \in \mathbb{R}^{N \times N}$ is the Gram matrix of the data. In that point of view, the vectors provided by PCA are to be thought as weights of a linear combination of the dataset. Moreover, we see that other metrics could be used instead of the euclidean one in the computation of the Gram matrix, which is for instance the basis of kernel PCA methods.*

Linear Discriminant Analysis

Nevertheless, it's important to point out that the vectors given by the PCA are only taking into account the global variance of the whole dataset. In the situations that we are interested in, our learning sets consist generally of subjects belonging to known classes (controls and glaucomatous in the example of the retina database). Since PCA is not supervised by the classes, it does consider the inter-class variance and for that reason, the resulting subspaces might be very poorly relevant to discriminate between groups. This is what we intend to show on figure 4.11 : none of the two first PCA axis is able to separate classes because the inter-class variation is basically overwhelmed by the greater inner variance of the two class clusters.
Linear Discriminant Analysis (or LDA), which is also very close to Fischer's discriminant analysis, was introduced in statistical learning precisely to overcome this kind of issues. These methods are well-detailed for instance in [47, 48]. We just briefly remind its essential lines.
Let's focus first on the case of two classes. We consider again the set of points $(x_i) \in \mathbb{R}^{p \times N}$ with global center $\overline{x}$ and we assume that these points belong to two classes 0 and 1 given by the sets $C_0 \subset \{1, ..., N\}$ and $C_1 = \{1, ..., N\} \backslash C_0$ of respective sizes N_0 and $N_1 = N - N_0$. We also call $\overline{x}_0$ and $\overline{x}_1$ the centers of each class. The

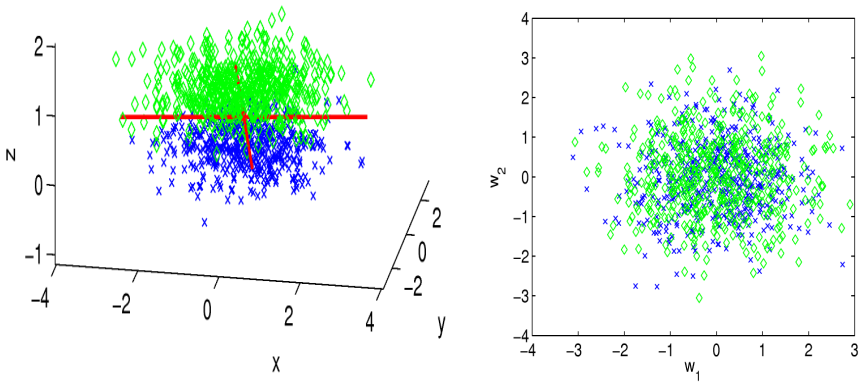


Figure 4.11: A simulated example of PCA on a set of points in $\mathbb{R}^3$. On the left figure, the points drawn using two elongated and translated normal distributions with colors corresponding to the two different classes. In red, we show the two first principal directions found by PCA. On the right figure are represented the points projected in the corresponding subspace.

goal of LDA is to find one discriminant axis between X_{C_0} and X_{C_1}. It introduces two matrices of $\mathbb{R}^{p\times p}$. The first one is the *within-class scatter matrix* and is defined by :

$$S_w = \sum_{i\in C_0} (x_i - \overline{x}_0)(x_i - \overline{x}_0)^T + \sum_{i\in C_1} (x_i - \overline{x}_1)(x_i - \overline{x}_1)^T .$$

The second one is the *between-class scatter matrix* :

$$S_b = N_0(\overline{x}_0 - \overline{x})(\overline{x}_0 - \overline{x})^T + N_1(\overline{x}_1 - \overline{x})(\overline{x}_1 - \overline{x})^T .$$

Remark 4.3.2. *Simple computations show that* $S_b \propto (\overline{x}_1 - \overline{x}_0)(\overline{x}_1 - \overline{x}_0)^T$ *and thus* S_b *is proportional to the orthogonal projection matrix on the line spanned by* $(\overline{x}_1 - \overline{x}_0)$.

In order to separate the two classes as much as possible, the principle of LDA is to find a direction v that maximizes the between-class variance of the projections on Span(v) while minimizing the within-class variance. Thus, it intends to solve :

$$v = \arg\max_{\|u\|=1} \frac{u^T S_b u}{u^T S_w u} . \tag{4.21}$$

We see that this problem makes sense only if S_w is a non-singular matrix. From the very definition of S_w, this is clearly not the case if the number of subjects is smaller than the dimension of the embedding space, i.e if $N < p$. We shall be dealing with that case in a following paragraph. If $p \ll N$, unless the data strictly lie in a lower dimensional subspace of $\mathbb{R}^N$, S_w is generally non-singular and, cf [12], v is the eigenvector corresponding to the largest eigenvalue λ of $S_b S_w^{-1}$, that is :

$$S_b v = \lambda S_w v$$

which amounts in solving a generalized eigenvalue problem between S_b and S_w. Since S_b is a rank one matrix, there is only one non-zero eigenvalue. In this simple case, one can actually express the true solution :

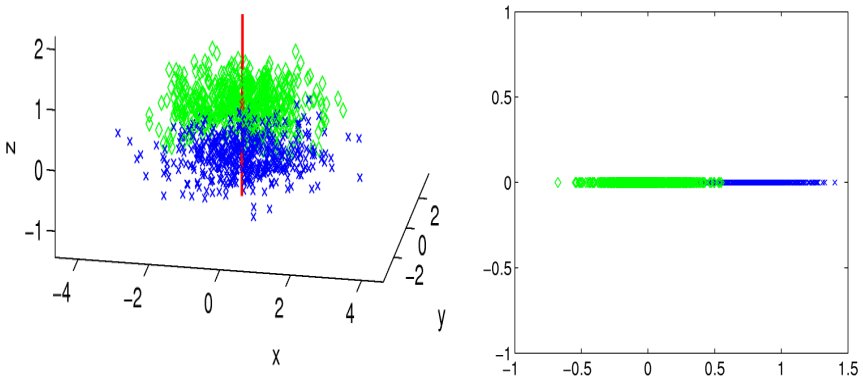


Figure 4.12: The LDA direction obtained on the example of figure 4.11 and the projection of points on this subspace.

Property 4.3.1. *The LDA vector v solving* (4.21) *is given explicitly by :*

$$v = \pm \frac{S_w^{-1}(\overline{x}_1 - \overline{x}_0)}{\|S_w^{-1}(\overline{x}_1 - \overline{x}_0)\|}$$

Proof. Since $S_w v = \frac{1}{\lambda} S_b v$ and $S_b v \propto (\overline{x}_1 - \overline{x}_0)$ thanks to the previous remark, we see that $S_w v$ belongs to the line spanned by the vector $(\overline{x}_1 - \overline{x}_0)$ and thus v is collinear to $S_w^{-1}(\overline{x}_1 - \overline{x}_0)$. □

In figure 4.12, we show the result of LDA on the example of the two translated Gaussian distributions of figure 4.11. As we expected, compared to PCA, the projections of the data points on the LDA axis are far better separated.
Even though the applications we show in this thesis are focused on the case of two classes, let's mention that LDA generalizes to an arbitrary number of classes $(C_j)_{j=1,..,k}$ in a quite natural way. With analogous conventions, one defines the matrices S_w and S_b by :

$$S_w = \sum_{j=1}^{k} \sum_{i \in C_j} (x_i - \overline{x}_j)(x_i - \overline{x}_j)^T$$

$$S_b = \sum_{j=1}^{k} N_j (\overline{x}_j - \overline{x})(\overline{x}_j - \overline{x})^T$$

Note that S_b is now a matrix whose rank is the dimension of the vector space spanned by the class centers $\overline{x}_i$ and thus S_b is a matrix of rank less or equal than $k-1$, since $\sum_{i=1}^{k} N_i(\overline{x}_i - \overline{x}) = 0$. This being set, similarly to the case $k = 2$, the multi-class LDA consists in finding the non-zero generalized eigenvalues between S_b and S_w. This provides several LDA vectors, that are all obtained as linear combination of the data points, and thus a discriminant subspace (of dimension $\leq k-1$) in which one can project and analyze the data with definitely better class separation properties.

Regularized LDA

As we mentioned, the previous computations are valid when S_w is non-singular but this hypothesis breaks down as soon as the number of data points is smaller than

the dimension N. In our cases of applications, we are typically confronted to that since the number of available subjects is rather small compared to the dimension of the features we want to analyze which is generally huge (for momenta or functional residuals, this is proportional to the number of points on the shapes).
To tackle this issue, some methods propose to first project the data on a lower dimensional subspace using typically PCA. However, we argue that this could be dangerous in that we do not really control how much between-class variance is lost by doing so. Other approaches like in [48] rely on generalized SVD's to solve the generalized eigenvalues problem. Yet the most common and simple alternative is to regularize the LDA [38] by replacing S_w by the matrix :

$$S_w^\varepsilon = S_w + \varepsilon \mathrm{Id}$$

for a certain $\epsilon > 0$, which automatically makes S_w^ε invertible and, similarly to property 4.3.1, gives $v = (S_w + \varepsilon \mathrm{Id})^{-1}(\overline{x}_1 - \overline{x}_0)$ in the two classes case. The drawback is the presence of the additional regularization parameter. The choice of ε can be done by optimizing it with respect to classification rates on cross-validation tests, which we shall explain a little more below.

4.3.2 Generalized LDA on Hilbert spaces

The previous LDA was presented in its classical setting, namely in the case of a finite-dimensional euclidean space. Nevertheless, the typical applications that we aim at in this chapter have to deal with momenta that belong to the dual V' of a space of vector fields or residual functions in the space $L^2(\overline{X})$ for instance. Those are manifestly Hilbert spaces of infinite dimensions. Even though one could argue that, in practice, the datasets are sampled with a finite number of points and that the momenta and functions can be represented in large dimensional vector spaces, the metric corresponding to V' for example is generally not the euclidean metric (it involves a kernel matrix). In this section, we show that LDA can be generalized in a natural way to Hilbert spaces and explain how it can be actually reduced to a usual regularized LDA in a certain finite-dimensional euclidean space.
Let's fix a Hilbert space H and its inner product $\langle ., . \rangle_H$. We take the same notations as in the previous subsection : $(x_i)_{i=1,..,N} \in H^N$ is the dataset and $\overline{x}$ its center, $C_j \subset \{1, .., N\}$ for $j = 1, .., k$ are the k different classes and $\overline{x}_j \in H$ are the class centers. We can then define, by analogy to usual LDA, the within-class and between-class operators :

$$S_w(x) = \sum_{j=1}^{k} \sum_{i \in C_j} \langle x_i - \overline{x}_j, x \rangle_H (x_i - \overline{x}_j)$$

$$S_b(x) = \sum_{j=1}^{k} N_j \langle \overline{x}_j - \overline{x}, x \rangle_H (\overline{x}_j - \overline{x}) \, .$$

Note that these definitions generalize the previous euclidean case. We introduce now the finite-dimensional subspace $H_0 = \mathrm{Span}(x_i | i = 1, .., N)$ spanned by the data points and denote by S_{w,H_0} and S_{b,H_0} the restriction of these operators to H_0. Note that, since both S_w and S_b take values in H_0, S_{w,H_0} and S_{b,H_0} are linear applications of H_0 to itself. In general, S_{w,H_0} is not invertible and we consider as previously the regularized matrix :

$$S_{w,H_0}^\varepsilon = S_{w,H_0} + \varepsilon \mathrm{Id}_{H_0} \in \mathcal{L}(H_0)$$

Now, we have the following property :

Property 4.3.2. *Let $L : H_0 \to \mathbb{R}^q$, $q \in \mathbb{N}^*$, be any isometry and let's denote by $\tilde{S}_w^\varepsilon, \tilde{S}_b$ the scatter matrices of the image data points $L(x_i) \in \mathbb{R}^q$. Then we have, for all $\epsilon \geq 0$:*

$$\tilde{S}_w^\varepsilon = LS_{w,H_0}^\varepsilon L^\dagger, \quad \tilde{S}_b = LS_{b,H_0}L^\dagger$$

where $L^\dagger : \mathbb{R}^q \to H_0$ is the adjoint of L. It results that, up to the change of variable $\hat{u} \mapsto u = L^\dagger \hat{u}$, to any eigenvalue and eigenvector $(\lambda_j, \hat{u}_j)$ of $(\tilde{S}_w^\varepsilon)^{-1} S_b$ corresponds the eigenvalue and eigenvector $(\lambda_j, L^\dagger \hat{u_j})$ of $(S_{w,H_0}^\varepsilon)^{-1} S_{b,H_0}$.

Proof. By definition, we have for all $a \in \mathbb{R}^q$ and $x \in H$:

$$\langle L^\dagger a, x \rangle_H = L(x)^T a$$

We deduce that for all $a, b \in \mathbb{R}^q$:

$$\begin{aligned}
b^T \tilde{S}_w(a) &= b^T \left(\sum_{j=1}^k \sum_{i \in C_j} (L(x_i) - L(\overline{x}_j))(L(x_i) - L(\overline{x}_j))^T a \right) \\
&= b^T \left(\sum_{j=1}^k \sum_{i \in C_j} (L(x_i) - L(\overline{x}_j)) \langle L^\dagger a, x_i - \overline{x}_j \rangle_H \right) \\
&= \sum_{j=1}^k \sum_{i \in C_j} b^T (L(x_i) - L(\overline{x}_j)) \langle L^\dagger a, x_i - \overline{x}_j \rangle_H \\
&= \sum_{j=1}^k \sum_{i \in C_j} \langle L^\dagger b, x_i - \overline{x}_j \rangle_H \langle L^\dagger a, x_i - \overline{x}_j \rangle_H \\
&= \left\langle L^\dagger b, \sum_{j=1}^k \sum_{i \in C_j} \langle L^\dagger a, x_i - \overline{x}_j \rangle_H (x_i - \overline{x}_j) \right\rangle_H \\
&= \langle L^\dagger b, S_{w,H_0}(L^\dagger a) \rangle_H \\
&= b^T (LS_{w,H_0} L^\dagger a)
\end{aligned}$$

which implies that $\tilde{S}_w = LS_{w,H_0}L^\dagger$. In addition, since L is assumed to be an isometry, we have that $LL^\dagger = \mathrm{Id}_{\mathbb{R}^q}$ and it follows that :

$$\tilde{S}_w^\varepsilon = LS_{w,H_0}^\varepsilon L^\dagger \, .$$

In the same way, one checks easily that $\tilde{S}_b = LS_{b,H_0}L^\dagger$. Now, we deduce :

$$(\tilde{S}_w^\varepsilon)^{-1} \tilde{S}_b = (L^\dagger)^{-1} \left(S_{w,H_0}^\varepsilon \right)^{-1} S_{b,H_0}) \, L^\dagger \, .$$

It results that the eigenvalues of $(\tilde{S}_w^\varepsilon)^{-1}\tilde{S}_b$ and $(S_{w,H_0}^\varepsilon)^{-1} S_{b,H_0}$ are the same and we obtain the eigenvectors of the second operator from the first one by the change of coordinates $\hat{u} \in \mathbb{R}^q \mapsto L^\dagger \hat{u} \in H_0$. □

Property 4.3.2 means basically that LDA in the Hilbert space H can be done equivalently by applying usual 'euclidean' LDA in $\mathbb{R}^q$ on the data points $L(x_i)$. The only remaining point to address is to make explicit such an isometry L. For that purpose, let's assume that the x_i are linearly independent so that they constitute a basis of H_0.

If H is typically a Hilbert space whose dimension is infinite or very big compared to N, this assumption is not aberrant since N random elements in H will be generically independent. Otherwise, one can just restrict to a sub-database made of independent x_i's. We then introduce the Gram matrix of the dataset $G = (\langle x_i, x_j \rangle_H)_{i,j=1,..,N}$, which is symmetric positive definite, and define an operator $L : H_0 \to \mathbb{R}^N$ as follows :

$$\begin{aligned} \forall x \in H_0,\, L(x) &= G^{-\frac{1}{2}}\gamma(x) \\ \text{with } \gamma(x) &= (\langle x, x_i \rangle_H)_{i=1,..,N}\,. \end{aligned} \tag{4.22}$$

Note that the image $L(x_i)_{i=1,..,N}$ of the data matrix by L is $G^{\frac{1}{2}}$. In addition :

Property 4.3.3. *The linear application $L : H_0 \to \mathbb{R}^N$ defined by eq.(4.22) is an isometry. Moreover, the adjoint $L^\dagger$ equals for all $a \in \mathbb{R}^q$:*

$$L^\dagger(a) = \sum_{i=1}^{N} \alpha_i x_i, \text{ with } \alpha = G^{-\frac{1}{2}}a \tag{4.23}$$

Proof. Let $y, z \in H_0$. By definition, we have :

$$L(y)^T L(z) = \gamma(y)^T G^{-1} \gamma(z)\,.$$

Now if we write $z = \sum_{i=1}^N \mu_i x_i$, we see that $\gamma(z) = G\mu$ and thus :

$$L(y)^T L(z) = \gamma(y)^T \mu = \sum_{i=1}^{N} \langle y, x_i \rangle_H \mu_i = \langle y, z \rangle_H\,.$$

L is therefore an isometry. A consequence is that we have also $LL^\dagger = \mathrm{Id}_{\mathbb{R}^N}$. If $a \in \mathbb{R}^N$, $L^\dagger(a) \in H_0$ which writes $L^\dagger(a) = \sum_{i=1}^N \alpha_i x_i$. Thus,

$$a = LL^\dagger(a) = G^{-\frac{1}{2}}\gamma(L^\dagger(a)) = G^{-\frac{1}{2}}G\alpha = G^{\frac{1}{2}}\alpha$$

and finally, we conclude that $\alpha = G^{-\frac{1}{2}}a$. □

In concrete terms, the previous property shows that the computation of the LDA discriminant vectors in the Hilbert space H can be actually done by first applying usual LDA in the euclidean space $\mathbb{R}^N$ for the data matrix $G^{\frac{1}{2}}$. The discriminant directions in H are then obtained by the adjoint operator $L^\dagger$ that can be computed explicitly using eq.(4.23). Any new data $x \in H$ can be also projected to this discriminant subspace by first computing $L(p_{H_0}(x)) = G^{-\frac{1}{2}}\gamma(x)$ and compute directly inner products with the discriminant vectors in $\mathbb{R}^N$.

4.3.3 Modes and classification for functional atlases

We specify and detail the previous PCA and LDA analysis to the case where the data is now a set of deformation momenta and functional residuals on the template shape as provided by the previous atlas estimation algorithm on functional shapes. We shall restrict to the case where the ambient space $E = \mathbb{R}^3$ as in our applications, although what we present remains valid in any dimension. From the theoretical point of view, we are manipulating objects that live in infinite dimensional Hilbert spaces (the dual V' of the RKHS of vector fields for momenta, the space of functions on the template shape in the case of residuals). In practice, these are sampled on the points of the template but still belong to very high dimensional vector spaces. The second point is the metric that is chosen to compute correlations between the data. We examine those issues for both features in the following.

Momenta analysis

In that case, the dataset to analyze is the set of initial momenta p_0^i for $i = 1, .., N$, N being the number of subjects and each of the p_0^i can be represented as a vector $p_{0,k}^i \in \mathbb{R}^3$ attached to every point $(q_k)_{k=1,..,n}$ of the estimated template shape and thus to a large vector of $\mathbb{R}^{3n}$. However, these momenta must be considered as elements of V' and consequently one has to evaluate the covariance not with respect to the L^2-metric in $\mathbb{R}^{3n}$ (which does not have any 'natural' meaning here) but with the RKHS metric on V', that is to say :

$$\langle p_0^i, p_0^j \rangle_{V'} = \langle K p_0^i, K p_0^j \rangle_V = \langle v_0^i, v_0^j \rangle_V$$

where we remind that $v_0^i(x) = \sum_{k=1}^n K_V(q_k, x) p_{0,k}^i$ is the initial velocity field corresponding to the momentum p_0^i. Thus, from the numerical point of view, using usual kernel computations, for all pairs of subjects $i, j \in \{1, .., N\}$ we have :

$$\langle p_0^i, p_0^j \rangle_{V'} = \sum_{k,l=1}^n (p_{0,k}^i)^T K_V(q_k, q_l) p_{0,l}^j \,. \tag{4.24}$$

These computations are of numerical complexity $O(n^2)$ for each inner product but can be also done using all the arsenal of fast kernel computations developed in section 2.3.3. Based on eq.(4.24), we can then compute the Gram matrix of the data

$$G = (\langle p_0^i, p_0^j \rangle_{V'})_{i,j=1,..,N} \,. \tag{4.25}$$

In our situations, we have generally $N \ll n$, and thus computing eigenvalues of G as needed for the PCA is numerically very fast compared to the computation of G.
Applying PCA in this setting just consists in computing the largest eigenvalues of the matrix G. It provides corresponding eigenvectors $w_s \in \mathbb{R}^N$ which are the principal modes of geometrical variability in the dataset. Coming back to the space of momenta, these modes are given by

$$p_s^{PCA} = \sum_{i=1}^N w_{s,i} p_0^i$$

Shooting the template using such momenta thus allows to visualize what these modes are in term of deformation, which we will show on the retina dataset in the next section.
The case of LDA can be treated as explained in subsection 4.3.2 (for $H = V'$). Adopting the same notations, we first compute the Gram matrix as in eq.(4.25) and its square root $G^{\frac{1}{2}}$, then apply usual LDA in $\mathbb{R}^N$. It provides a discriminant vector $a \in \mathbb{R}^N$ which, in the space of momenta, translates to :

$$p^{LDA} = L^\dagger(a) = \sum_{i=1}^N \alpha_i p_0^i, \;\; \alpha = G^{-\frac{1}{2}} a$$

Again, this discriminant direction can be visualized by shooting the template with initial momentum p^{LDA}. One can then project the data on the line $\mathrm{Span}(p^{LDA})$ by computing $\langle p_0^i, p^{LDA} \rangle_{V'}$, which can be done in the space $\mathbb{R}^N$ since :

$$\langle p_0^i, p^{LDA} \rangle_{V'} = \sum_{j=1}^N \alpha_j G_{ij} \,.$$

Any new data p_0 outside the training set can be also projected to this subspace, by first computing $L(p_0)$ and write :

$$\langle p_0, p^{LDA} \rangle_{V'} = \sum_{j=1}^{N} u_j L(p_0)_j$$

Classification for both PCA and LDA can be performed based on the projections of the data onto the low dimensional subspaces provided by the methods. If these subspaces are sufficiently discriminative, one can expect that the dataset splits into several clusters, each one corresponding to one class. The classification rule can then consist in simply finding the closest cluster center to the projection of a given data, possibly by computing Mahalanobis distances in order to take into account clusters' variances.

Residuals analysis

The treatment of functional residuals follows the exact same pattern as momenta. One only needs to introduce the right metric to compute covariance between subjects. These residuals $(\zeta^i)_{i=1,..,N}$ are all functions defined on the common template shape $\overline{X}$. For the template estimation procedure, we used the L^2 metric on $\overline{X}$ as a penalty term in the functional. This metric could be also used as the measure of correlation. With the notations of section 4.2.2, it writes :

$$\langle \zeta_i, \zeta_j \rangle_{L^2(\overline{X})} = \int_{\overline{X}} \zeta^i(x)\zeta^j(x) d\mathcal{H}^d(x) = \sum_{l=1}^{m} \tilde{\zeta}_l^i \tilde{\zeta}_l^j |\xi_l| \,. \tag{4.26}$$

This metric, however, could be argued to be too strong in this purpose because it compares values of residuals only at corresponding points on the template. One easy possibility to overcome that is to regularize the L^2 metric by a Gaussian $e^{-\frac{|x_1-x_2|^2}{\sigma^2}}$, which would give the following correlation measure between residuals :

$$\begin{aligned} \langle \zeta_i, \zeta_j \rangle &= \iint_{\overline{X}\times\overline{X}} e^{-\frac{|x_1-x_2|^2}{\sigma^2}} \zeta^i(x_1)\zeta^j(x_2) d\mathcal{H}^d(x_1) d\mathcal{H}^d(x_2) \\ &= \sum_{k,l=1}^{m} e^{-\frac{|x_k-x_l|^2}{\sigma^2}} \zeta_k^i \zeta_l^j |\xi_k||\xi_l| \,. \end{aligned} \tag{4.27}$$

A third possibility could be to lean again on the kernel norms on functional currents (or even on functional varifolds) that were used as data attachment distances in registration and atlas estimation algorithms. In that case, the Hilbert space H in which the PCA and LDA are done is the space W' by identifying a residual ζ with the fcurrent $C_{(X,f+\zeta)} \in W'$. Natural candidates for kernels are also products of a Gaussian $K_g(x_1, x_2) = e^{-\frac{|x_1-x_2|^2}{\sigma_g^2}} \mathrm{Id}$ with $k_f(f_1, f_2) = e^{-\frac{|f_1-f_2|^2}{\sigma_f^2}}$. The correlation measure between residuals is then :

$$\begin{aligned} \langle C_{(\overline{X},\zeta^i)}, C_{(\overline{X},\zeta^j)} \rangle_{W'} &= \iint_{\overline{X}\times\overline{X}} e^{-\frac{|x_1-x_2|^2}{\sigma_g^2}} e^{-\frac{|\zeta^i(x_1)-\zeta^j(x_2)|^2}{\sigma_f^2}} \langle \tilde{\xi}(x_1), \tilde{\xi}(x_2) \rangle d\mathcal{H}^d(x_1) d\mathcal{H}^d(x_2) \\ &= \sum_{k,l=1}^{m} e^{-\frac{|x_k-x_l|^2}{\sigma_g^2}} e^{-\frac{|\zeta_k^i-\zeta_l^i|^2}{\sigma_f^2}} \langle \xi_k, \xi_l \rangle \end{aligned} \tag{4.28}$$

where we adopted the same notations as in chapter 2, $\tilde{\xi}(x) = \xi(x)/|\xi(x)|$ is the unit d-vector at each point $x \in X$. Note that one can take other kernel widths σ_g and σ_f than the one used for the atlas estimation part, thus allowing a certain flexibility of the analysis with respect to various scales (at the price of additional parameters). Such metrics have the advantage of taking more into account the geometry of the support X on which all the ζ^i are defined. But there are two counterparts : first, the computations are more expensive compared to L^2 norm and second, the variation modes and discriminant directions are basically linear combination of the fcurrents $C_{(X,f+\zeta^i)}$ which cannot be interpreted and visualized directly as a signal defined on X due to the non-linearity of k_f.
Once the metric is chosen and the Gram matrix of the dataset computed, PCA and LDA can be performed exactly as previously explained for momenta. In both cases, we obtain modes of variations that are linear combinations of the ζ^i's and that can be represented as residual functions on the template. Classification is then done again by projecting the training and testing datasets on the lower-dimensional subspaces provided by PCA and LDA.

4.3.4 Classification results on the OCT dataset

We now show a few results of the previous framework on the retina thickness maps of section 4.2.3. We had obtained, with algorithm 3, an atlas of the NFL surfaces consisting of a functional surface baseline (the template) plus momenta p^i and residuals thicknesses ζ^i for each of the $N = 30$ subjects. Based on these, we intend to illustrate the statistical analysis of geometrical and functional variability developed in this section.
Starting with the analysis on momenta, in figure 4.13, we show the template deformed in the directions of the two first principal components. As we can see, the first component is essentially a dilatation/shrinkage of the whole geometrical support while the second one is more or less a displacement and distortion of the opening. As we could suppose, these are mainly explained by the variability caused both by acquisition (position and orientation of the eye with respect to the device during the measurement) and by global variations in the size of the retina and openings across individuals. Figure 4.3.4 shows the projection of the dataset on the plane generated by these principal components and confirms that these do not discriminate satisfyingly between controls and glaucomatous. One would obviously need to compute more PCA vectors to get discriminative directions but the first components are clearly not enough.
This is contrasting with the LDA which provides a component vector detecting the between-class variability. As for PCA, we show this discriminative direction on figure 4.15. It was obtained by regularized LDA with $\epsilon = 0.1$. Observe that the corresponding deformation of the template is very different to the ones provided by the first PCA components. The displacements are essentially concentrated around the ocular nerve opening and seem to correspond to an increase and decrease of the layer digging. As we see thanks to figure 4.3.4, this mode of variation is apparently more discriminant between glaucomatous and controls because it separates the classes in two nearly distinct clusters.
The same analysis can be performed for the residual thicknesses on the template shape. This time, the component vectors provided by PCA or LDA, as detailed previously, are represented by a function on the template shape which is added/subtracted to the baseline thickness. We have used the simplest correlation provided by the L^2

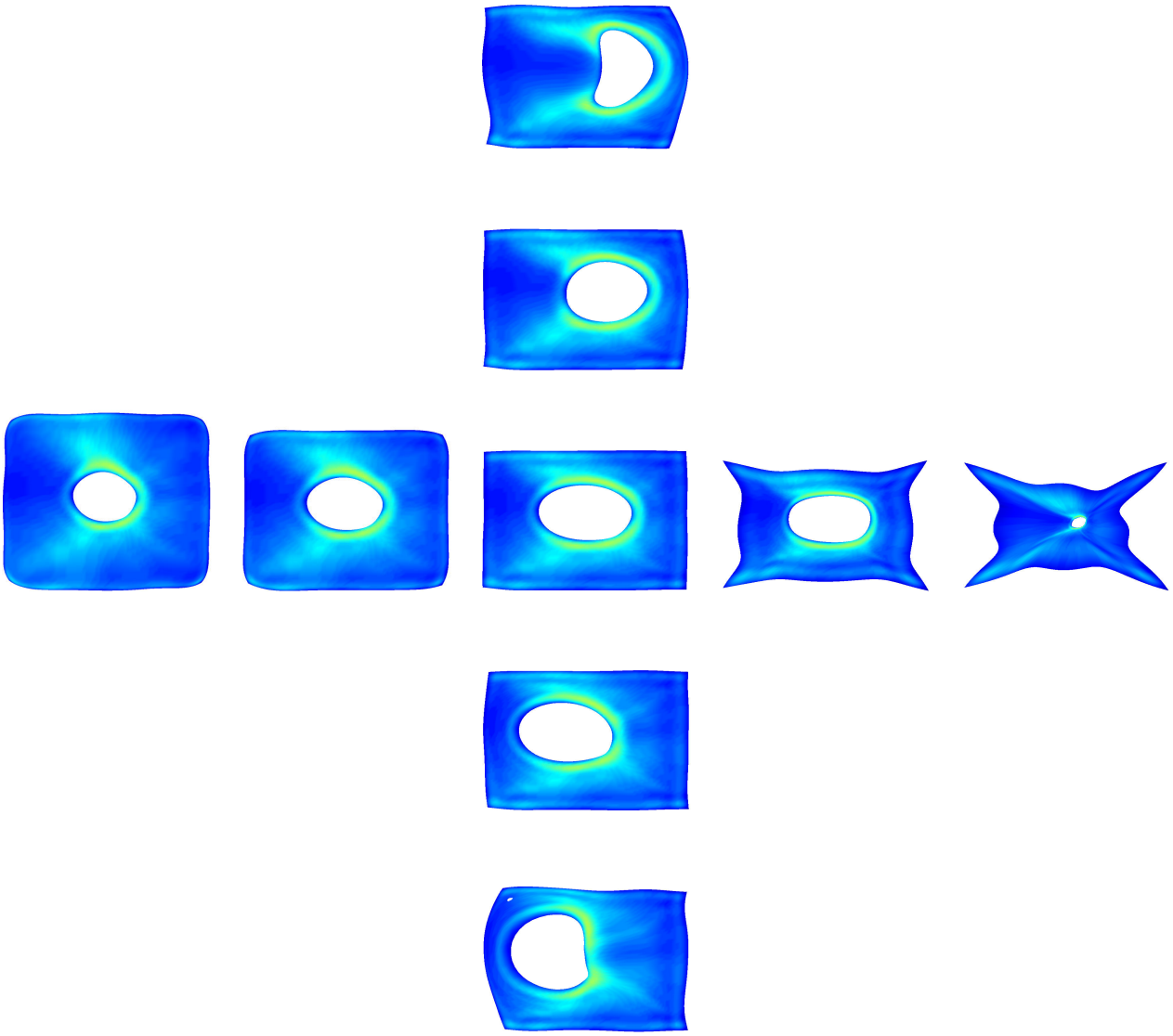


Figure 4.13: The first (middle row) and second (middle column) principal components of the momenta. What we show is the template shooted for both negative and positive times in the direction given by the PCA vectors, as explained in the previous section.

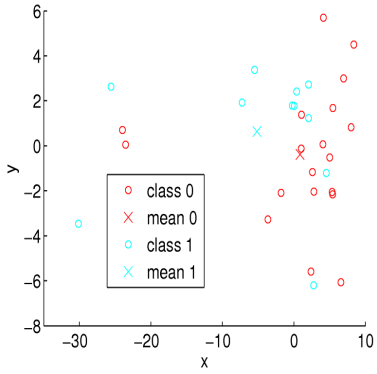


Figure 4.14: Projections and class means of the dataset on the plane given by the two first PCA vectors for the momenta. Class 0 corresponds to controls and 1 to glaucomatous subjects.

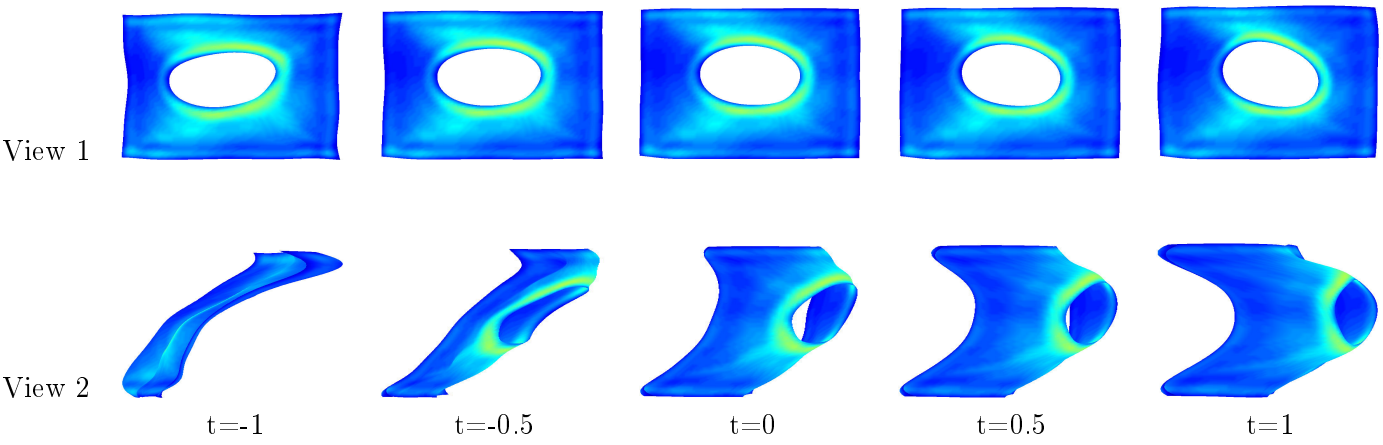


Figure 4.15: Resulting LDA direction on the momenta (obtained with a regularization parameter $\epsilon = 0.1$).

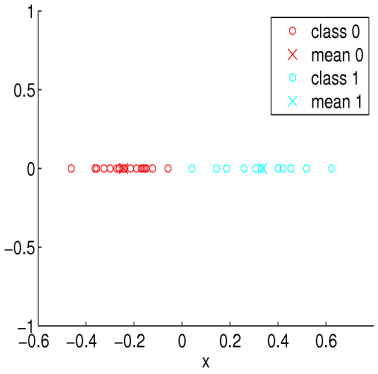


Figure 4.16: Projections and class means of the dataset on the one-dimensional subspace given by the LDA on momenta.

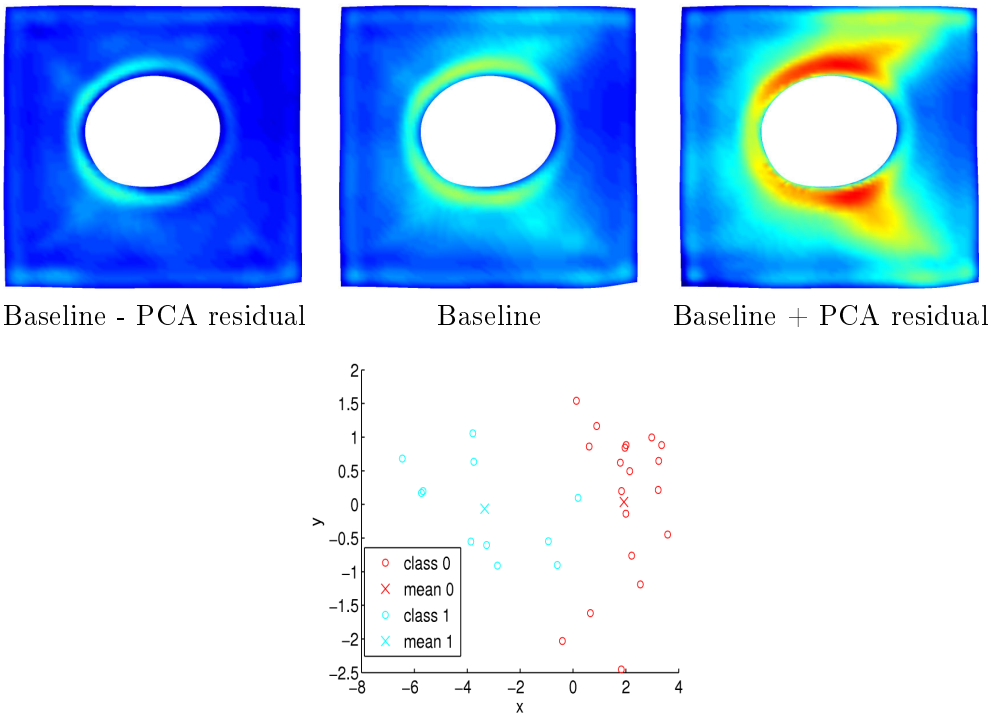


Figure 4.17: Results of the PCA on the thickness residuals. We computed two component vectors for the PCA. On the top figure, we represent the first one by showing the addition and subtraction of the residual to the baseline thickness. On the bottom figure are the projected coordinates of the residuals from the dataset on the two-dimensional subspace generated by the PCA vectors.

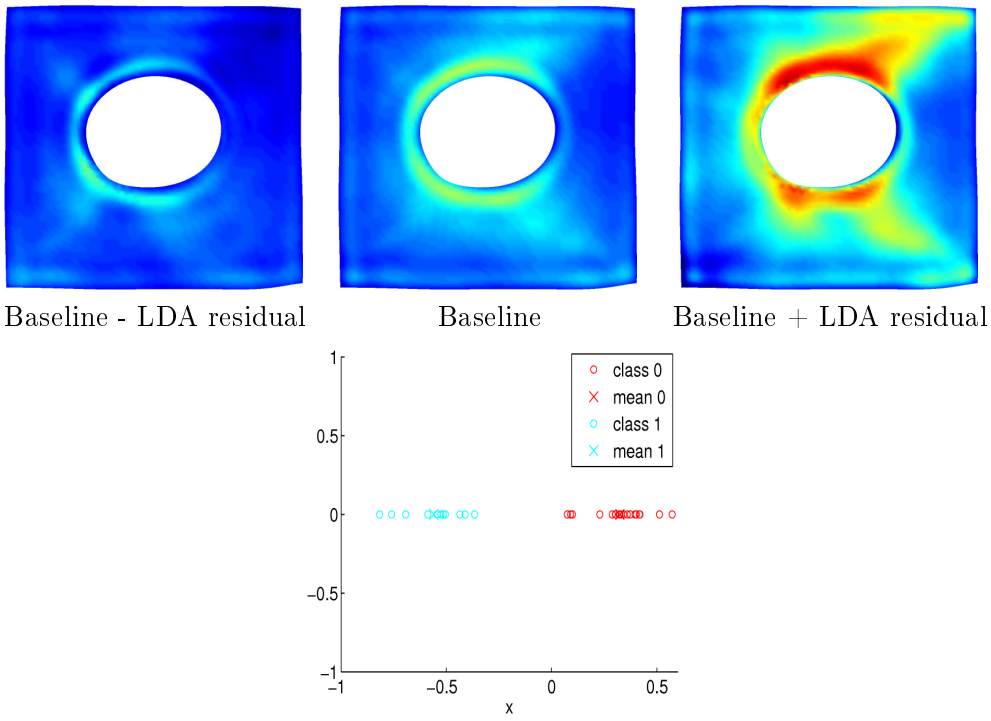


Figure 4.18: Results of the regularized LDA on the thickness residuals ($\epsilon = 0.1$). It provides one component vector which we represent on the top figure. The projections on this direction are shown on the lower figure.

metric on X (cf previous section). We display the results for the PCA in figure 4.17. The first component vector basically shows a thickness decrease for subjects with glaucoma on a C-shape concentrated mainly around the opening, which tends to confirm the overall observations of clinicians. As we see on the lower plot, the two first PCA vectors for thicknesses seem to separate the classes in a nicer way than momenta. This is even more obvious with the LDA. The variations of thicknesses in the direction of the LDA vector follow a quite similar pattern as the first PCA vector, as shown in figure 4.18. Yet one can remark once again the greater class separation property of the LDA.

These first sets of results suggest that applying LDA (rather than PCA) on momenta and thickness maps should provide a simple way to classify, at least with a reasonable efficiency, between controls and glaucoma. This could be assessed more rigorously by applying the prediction rule learned by LDA on a testing set of newly acquired subjects as we explained in the previous section. In the case of this particular application, we did not have yet new subjects available to work on (this is a current work from our collaborators' side in SFU). Instead, we tested the quality of prediction by cross-validation tests on the learning dataset itself. Namely, we applied leave-one out procedures which consist in excluding one (or more generally p) subject from the dataset and predict the class of these subjects by using the remainder. This is exactly as if one artificially separates the dataset into training and testing sets.

In practical terms, we randomly choose p subjects among the $N = 30$ available, estimate a template and LDA axis with the $N - p$ others, then project the p 'test' subjects on the LDA subspace and affect each of them to the class of the closest class mean. By repeating this procedure for numerous draws of p among N, one can

Data	p=1	p=2	p=3	p=4
Momenta	0.100	0.082	0.090	0.089
Thickness residuals	0.033	0.040	0.041	0.056
Momenta and thickness	0.000	0.008	0.021	0.032

Table 4.1: LDA classification error rates in cross validation ($N = 30$) for leave-p-out tests.

Data	p=1	p=2	p= 3	p=4
Momenta	0.150	0.174	0.183	0.207
Thickness residuals	0.000	0.008	0.018	0.035
Momenta and thickness	0.050	0.053	0.057	0.067

Table 4.2: LDA classification error rates on the old subject sub-dataset ($N = 20$).

then measure the number of false class predictions. Note that we also try several regularization parameters to optimize the classification rates. Thickness and momenta coordinates can be also combined in the classification : here we have tried a simple ℓ^1 distance to class centers in the 2D space provided by LDA coordinates for both momenta and residuals. The results are summed up in table 4.1. Classification based on momenta or thicknesses leads to misclassification rates of the order of 9%, which of course increases slowly as we remove more subjects from the training set (i.e when p increases). We remark that error rates are also slightly better using thicknesses. However, the misclassifications do not occur generally for the same subjects in the two methods. It results that the combination of the two provides very solid classification rates in that case since we get an error under 1% for $p = 1, 2$.

Naturally, these are still crude results at the present time. An important follow-up study is to interpret from a more clinical point of view these discriminative modes of variations found by LDA, especially the geometrical deformations. On the other hand, it's also interesting to analyze incidental parameters of the dataset in order to eliminate hidden effects that are not specifically correlated to the development of glaucoma. The subjects' age is typically an important one because it generally affects the shape of the retina. Since it is a known data, we also conducted the same classification experiments but restrained to only old subjects from the dataset (which gives 20 subjects, controls and glaucomatous). The cross-validation classification results are presented in table 4.2. We see that the classification performance based on geometry with the momenta is less accurate in that situation (around 17% error), whereas the one based on the thickness analysis improves quite spectacularly, suggesting that thickness decrease is probably a better biomarker of the disease than shape variations which could be partly caused by age. This is a study to be continued as a future work, one possible trail being to separate the dataset into more classes for instance by adding age and perform multi-class LDA : this would become particularly relevant statistically if the dataset is enriched with additional subjects' acquisitions.

4.4 Atlas estimation in the context of varifolds and applications

The atlas estimation procedures of section 4.2 can be readily transposed to the case of varifolds and functional varifolds that were introduced in chapter 3. All it needs basically is to replace fcurrents' norms by varifolds with the corresponding gradients for the update of the template position, momenta and residuals. Compared to the previous algorithms, the advantages are the one detailed in chapter 3 : the possibility to work with non-oriented manifolds and the greater stability of the estimation at bigger scales. In this section, we will essentially present two ongoing applications of that on real datasets of purely geometrical shapes : a first one concerns the study of cells' divisions in mice heart envelopes while the second will present atlases of Cochlea's surfaces.

4.4.1 Algorithm

The algorithms are very similar to the previous ones so we will just briefly highlight the few differences. If we consider pure geometries with no signals, atlas estimation in that context writes :

$$\min_{(X,p_0^i)} J(X,(p_0^i)) \doteq \left(\sum_{i=1}^{N} \|v_0^i\|_V^2 + \gamma \|\mu_{X^i} - \mu_{\phi^i(X)}\|_{W'}^2 \right) \tag{4.29}$$

μ_X denotes the varifold associated to X and $\|.\|_{W'}$ is a RKHS norm on varifolds as defined in chapter 3. Now the gradients of J with respect to X and the momenta p_0^i can be computed also by the same forward and backward dynamics. Only the initialization of the adjoint variables Q^i in the backward equation eq.(4.10) must be changed to $Q_1^i = -\nabla_q g_i(q_1^i)$ where g_i is the squared distance to the ith subject but now for the varifold metric. These gradients are directly derived from the computations presented in chapter 3. The normalizations of gradients presented in section 4.2.1 and 4.2.1 are still valid for varifolds. And finally, the variation formula shown in theorem 3.4.2 brings out singularities on the boundary for the gradient which can be tackled, as in 4.2.1, by regularization.
Now the functional component can be also added to the varifold framework, following what was exposed in chapter 3 section 3.6.2. It results that algorithm 3 is also transposable with minimal changes to varifolds, providing an alternative to fcurrents for estimating geometrico-functional atlases.

4.4.2 Works on embryonic mouse heart

As a first illustration of the interest of varifolds' atlas estimation from a more applied point of view in the field of biological imaging, we present a first series of results from an ongoing project in collaboration with S. Meilhac and J.F Le Garrec from the team *Molecular Genetics of Development* in Institut Pasteur. It consists in the analysis of cells' directions of divisions during the growth of hearts in mice embryos, as pioneered in [53]. The dataset itself is made of several subjects' *envelopes*, which are surfaces in the 3D space. These were acquired by confocal microscopy and delimit only a section on top of the heart of the embryo (until a depth limited by the acquisition device). The depth of acquisition is rather small and thus the surfaces are thin membranes. Inside these envelopes, were measured, at the time of embryo's

sampling, the *divisions axes*' (unoriented) directions of the cells which were in a mitosis process at this precise moment. We show in figure 4.19 the full dataset with the nine subjects envelopes and the corresponding division axes inside (except for a few artifacts located outside the envelopes on some subjects).

The biological issues at stake are several. As we can see, the number of measured divisions' axes within a subject is rather small in a given envelope. One of the question raised by biologists is to find if some regions show some meaningful correlations of the divisions' directions to understand the features of the heart's growth. However, estimating correlations on subjects separately would lead to very unstable and non-reproducible results due to the low number of axes for each one. Thus one needs a framework to gather the informations from the different subjects in order to have a denser distribution of axes. The second important issue, that we have just touched upon for the time being, is to study the influence of a mutation on a particular gene in the directions' repartition. The dataset contains actually three subjects of each class : wild species (with two non-mutated genes), heterozygous and mutants.

In the first place, we have investigated a way to gather together axes data from the different subjects in one common set of divisions. A reasonable way is to normalize geometries of the subjects by computing deformations between the envelopes and transport the directions through these deformations, i.e through the Jacobian of the diffeomorphisms as already described by eq.(2.3). Instead of choosing one of the subjects as a reference with the bias that this would introduce in the method, we propose instead to compute first a template of the envelope shapes. The essential difficulties here comes from a few artifacts on the envelopes' triangulations and from the fact that these membranes are usually very thin, which causes important cancellation problems when using usual currents (cf discussion of chapter 3 section 3.1.2). From this point of view, the framework of varifolds and the template estimation framework presented just previously seems more adapted. We initialize the template to an ellipsoid to avoid bias as much as possible and run the atlas estimation algorithm based on varifolds : this leads to the template shown in figure 4.20.

A second step is to compute the deformations from each subject to the estimated template (note that these are more or less the *inverses* of the maps ϕ^i given at the end of the atlas estimation). Finally, we pull the axes' directions of each subject back on the template by applying the differentials of the previous diffeomorphisms. In the end, we get a template envelope containing the contribution of the axes from the nine subjects together, which is represented in figure 4.21.

Since the total number of axes data transported in the template envelope is now more than 1000, statistics of spatial correlation between the directions makes far more sense than separate statistics on each subject. The result of figure 4.21 were analyzed by biologists using clustering methods combined with Rayleigh statistical tests on the set of directions : this study has evidenced three regions of the template that present significant correlation of the axes along a particular direction. Matching pursuit algorithm like the one presented for functional currents in chapter 2 was also adapted to varifolds and used to visualize the most significant correlations. Although quite preliminary at this stage, these first sets of results seemed encouraging enough to begin more precise investigations on the dependency of these regions and directions with respect to the gene mutation, which is currently at work but would most likely require the acquisition of additional subjects.

From the methodological point of view, we believe that this dataset offers some more perspectives and open questions that deserve to be mentioned. The varifold

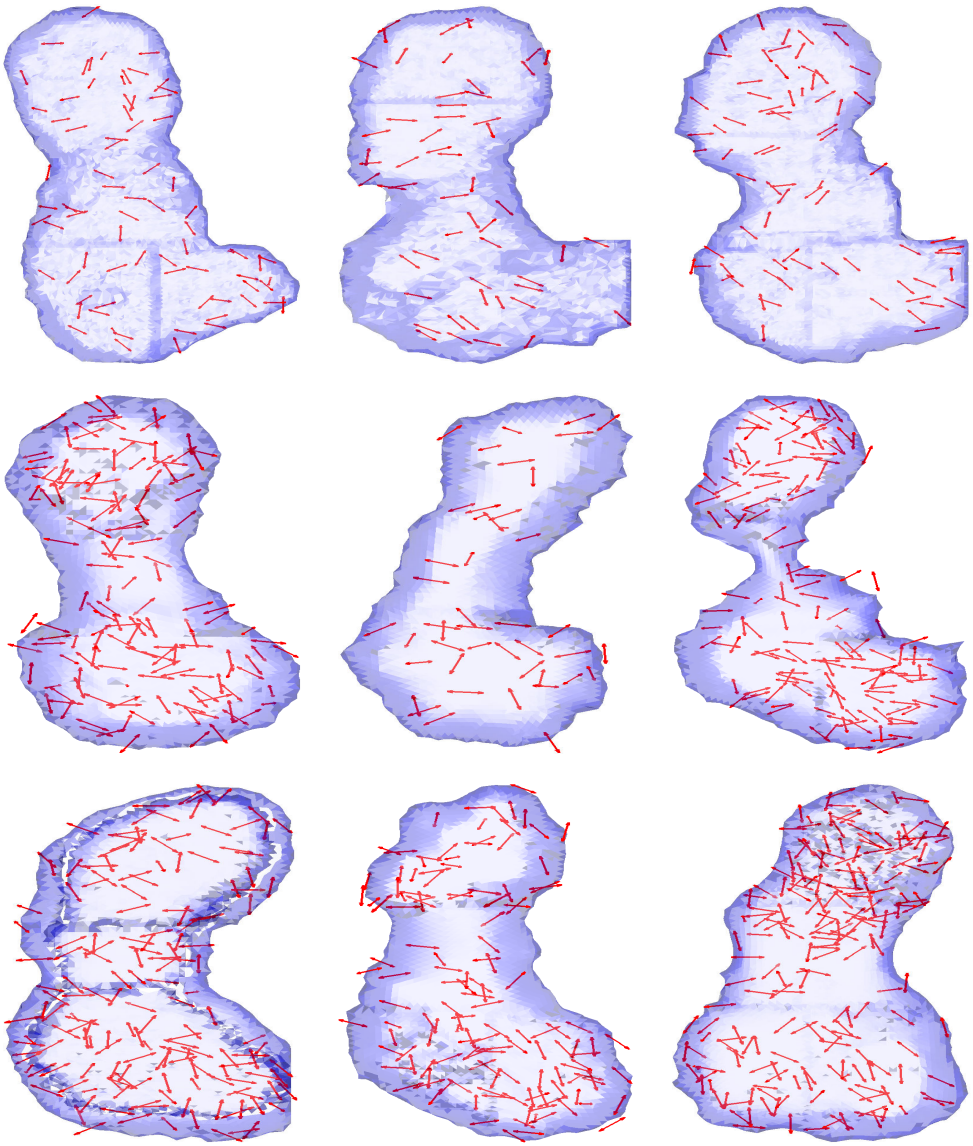


Figure 4.19: The nine different subjects of Pasteur dataset showing the triangulated envelopes' 3D surfaces seen from above and the division axes in red. On the first row are the three wild subjects, on the second the three heterozygous and on the last one the mutant subjects. In most subjects, one sees the development of two ventricles (up and down) separated by the inter-ventricular region.

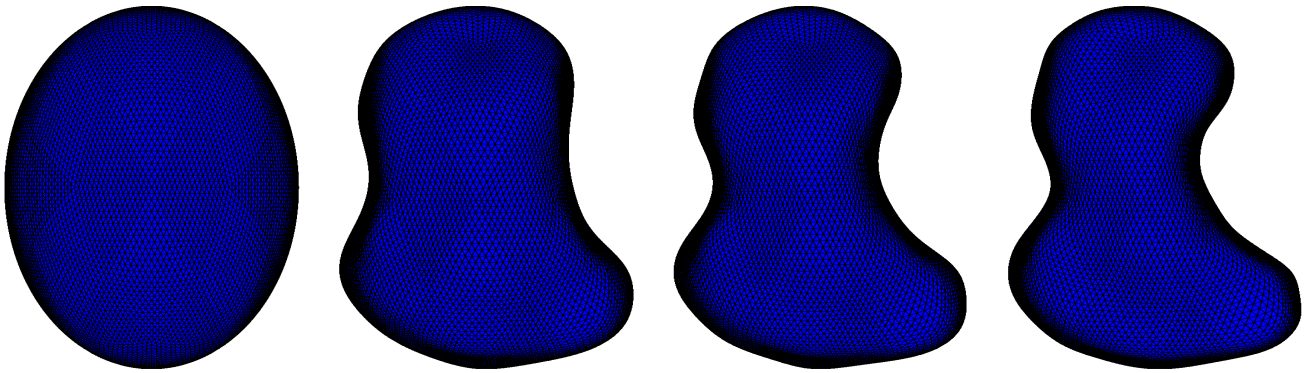


Figure 4.20: Evolution of the template envelope through the iterations of the atlas estimation process.

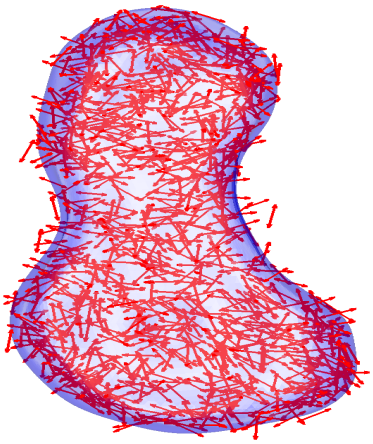


Figure 4.21: The template envelope with the transported division axes from the nine subjects.

representation was used here only on the envelopes, which provides perceptively better results of template and registration than currents. However, as we see on figure 4.19, on a few subjects, some boundaries are artificial because they correspond only to the limit of the acquisition box rather than to the real limit of the envelope. This issue of **partial observation** due to acquisition devices, even if not too determinant in that case, is a recurrent problem in many medical imaging datasets and is not yet diligently addressed by the present registration and template estimation algorithms. A second interesting point is the axes directions' analysis. Since these are basically distributions of non-oriented directions in the space, varifolds and kernel metrics offers again a natural representation and framework that could be used to estimate meaningful statistical correlations. One possibility, yet totally prospective at the moment, would be to estimate a template as an envelope shape containing a dense map of directions estimated from the database.

4.4.3 Template estimation on Cochleae dataset

We show now a second illustration of varifold atlas estimation on a completely different dataset. We gratefully thank professor Jose Braga from University Paul Sabatier in Toulouse, professor Jean-Luc Kahn, curator at the museum of 'Institut d'anatomie normale et pathologique' in Strasbourg and doctor Stanley Durrleman from 'Institut du Cerveau et de la Moelle épinière' in Paris for providing the material data used in this section. To briefly explain the context of this study and the overall project behind it, anthropologists are interested in understanding the evolution in the shape of the inner ear (or Cochlea) in a phylogenetic tree and relate some shape features of the Cochlea to other morphological characteristics of species such as the body mass index or the audiogram.

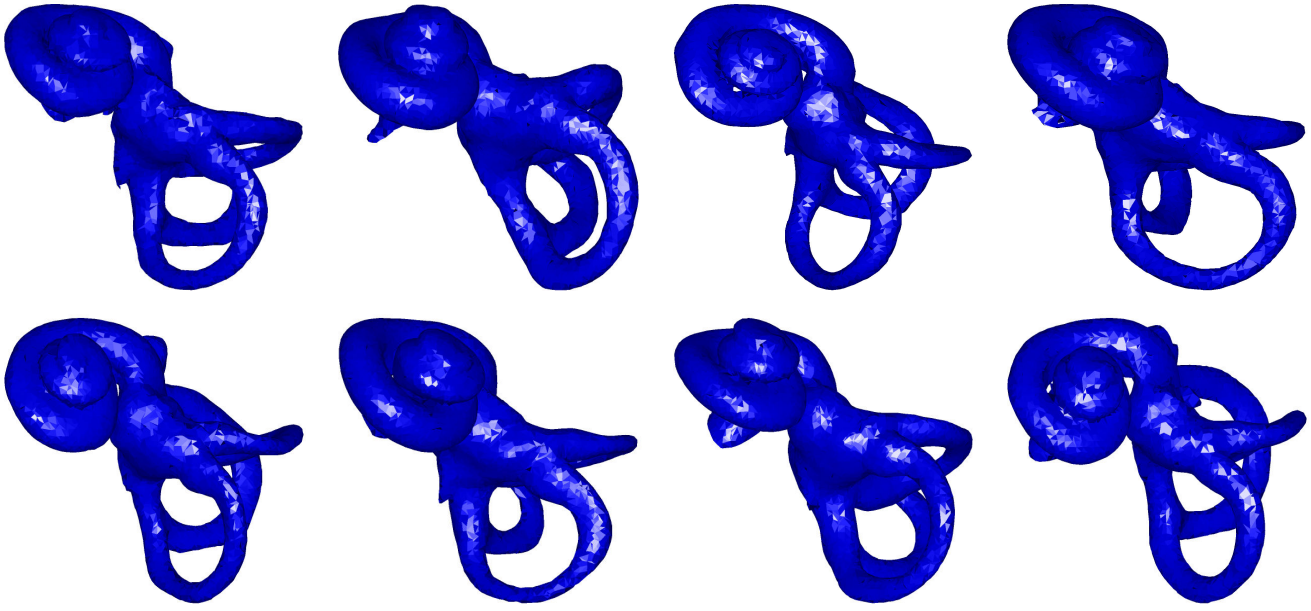


Figure 4.22: The eight different subjects of the Homo Sapiens Cochlea dataset. The surfaces we show here were subsampled from the original data to approximately 7500 points each, using vertex decimation filter in Paraview.

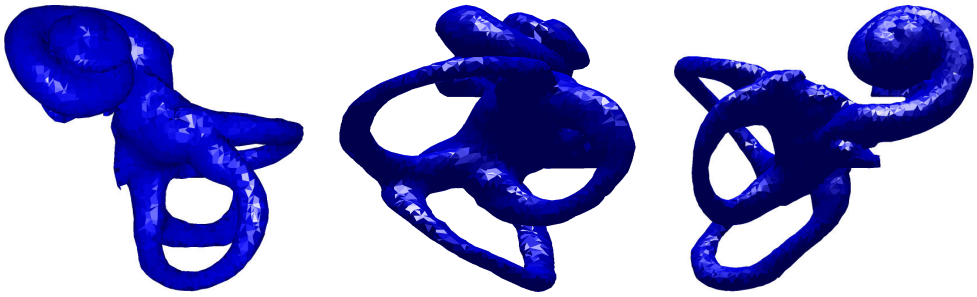


Figure 4.23: Three different views of the first subject. Note the presence of the 'snail' shape of the Cochlea and the presence of three canals (posterior, superior and horizontal).

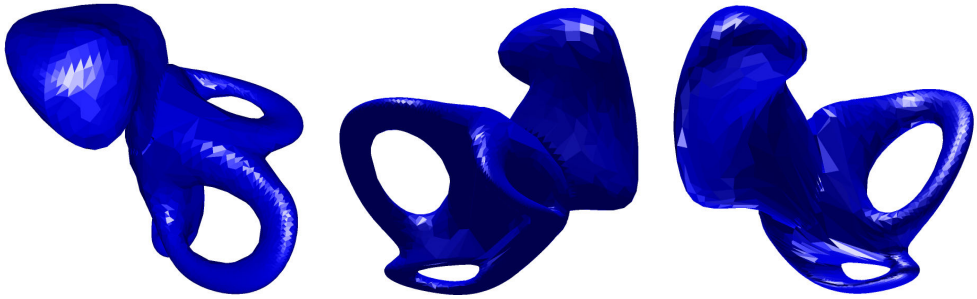


Figure 4.24: Template estimated with currents from different views.

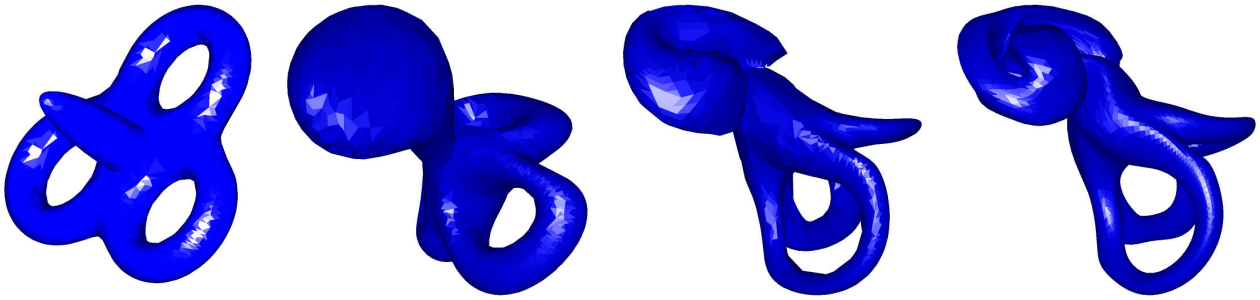


Figure 4.25: The initialization of the template shape (on the left) and the template shape estimated by the varifold algorithm (on the right), with several intermediate gradient descent steps.

The work we present here is less ambitious for the time being. We focused only on the Cochlea of Homo Sapiens for which the database contained the triangulated surfaces of eight subjects, which we show in figures 4.22 and 4.23. The inter-subject shape variability even among one particular specie is also an interesting and still not very well-known feature. The atlas estimation algorithms we have been investigating so far constitute a procedure to estimate this variability. Due to the complex geometry of the Cochlea shape, notably the presence of the three thin canals, the algorithms based on currents can very hardly recover a prototype shape from the dataset mainly because the cancellation effect of currents makes the kernel metrics insensitive to those pieces. Instead, the varifold atlas estimation algorithm is expected to be more robust in that regard. A special care for the initialization of the template is needed

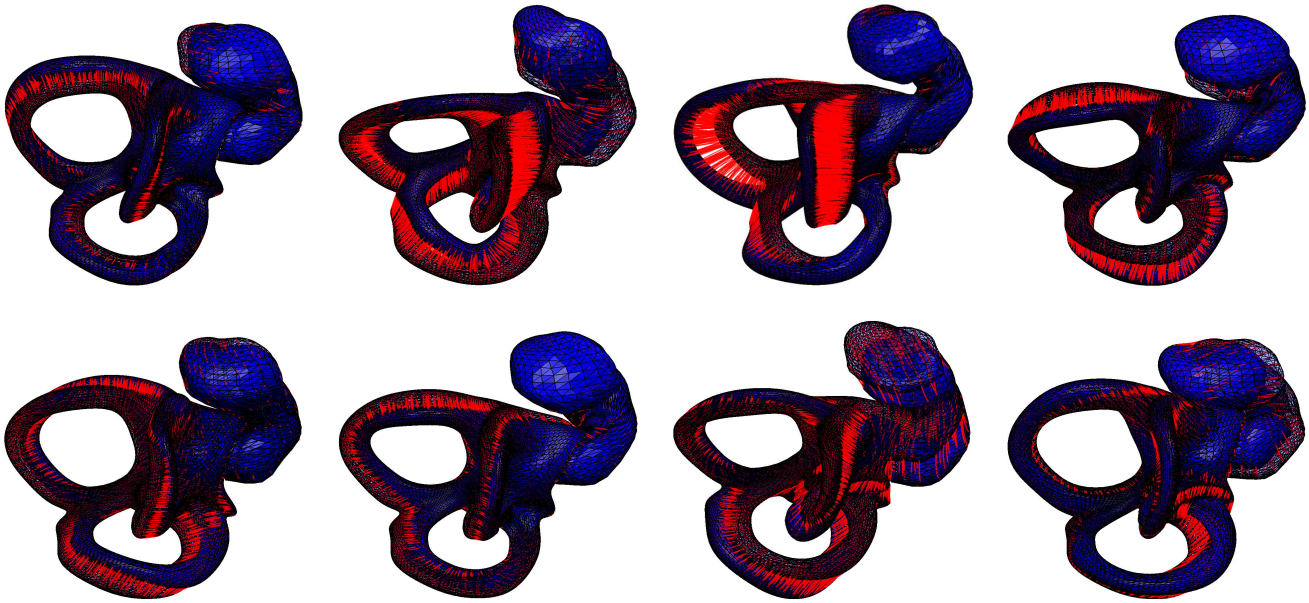


Figure 4.26: Estimated deformations between the template and the subjects from the dataset. The template is the transparent surface and its deformed version is the opaque one. Trajectories of the points are displayed in red.

in that specific dataset because of the particular topology of the inner ear. After reducing the number of points in the triangulations, we initialized the algorithm to the simple shape shown in figure 4.25, which has the topology of a triple torus but is not biased toward one of the individual. The estimated template shape is shown on the right of figure 4.25 and the deformations from the template to the subjects of the dataset in figure 4.26. As a comparison, we show the template estimated with the approach of currents with identical parameters for the kernels (figure 4.24). We notice a particularly unsatisfying estimation in the region of the canals, which is the most subject to orientation issues.

As we see, the template and deformations obtained with varifolds are able to capture some of the shape global features and variability between individuals, which could be studied in a deeper way using for instance the framework proposed in section 4.3. The layout of the three canals is obviously quite well recovered by the algorithm. However, the coiling number in the snail part, which is also an important biological marker, is still difficult to catch using this approach, essentially because it involves important deformations at very small scales. It can be recovered quite well the template is initialized to one subject of the database. To evaluate this dependency on the choice of the initial template, in figure 4.27, we show the obtained template for different initializations. The interesting point is that, for this dataset, we see a rather satisfying consistency in the estimated shape, even for different initializations so that the bias toward particular individuals does not seem a very limitating factor in this case.

Finally, to give a rough idea of computational times for such examples, let's mention that the previous atlases were obtained in about one hour on the CMLA computational server for Cochlea shapes subsampled to approximately 8000 points each.

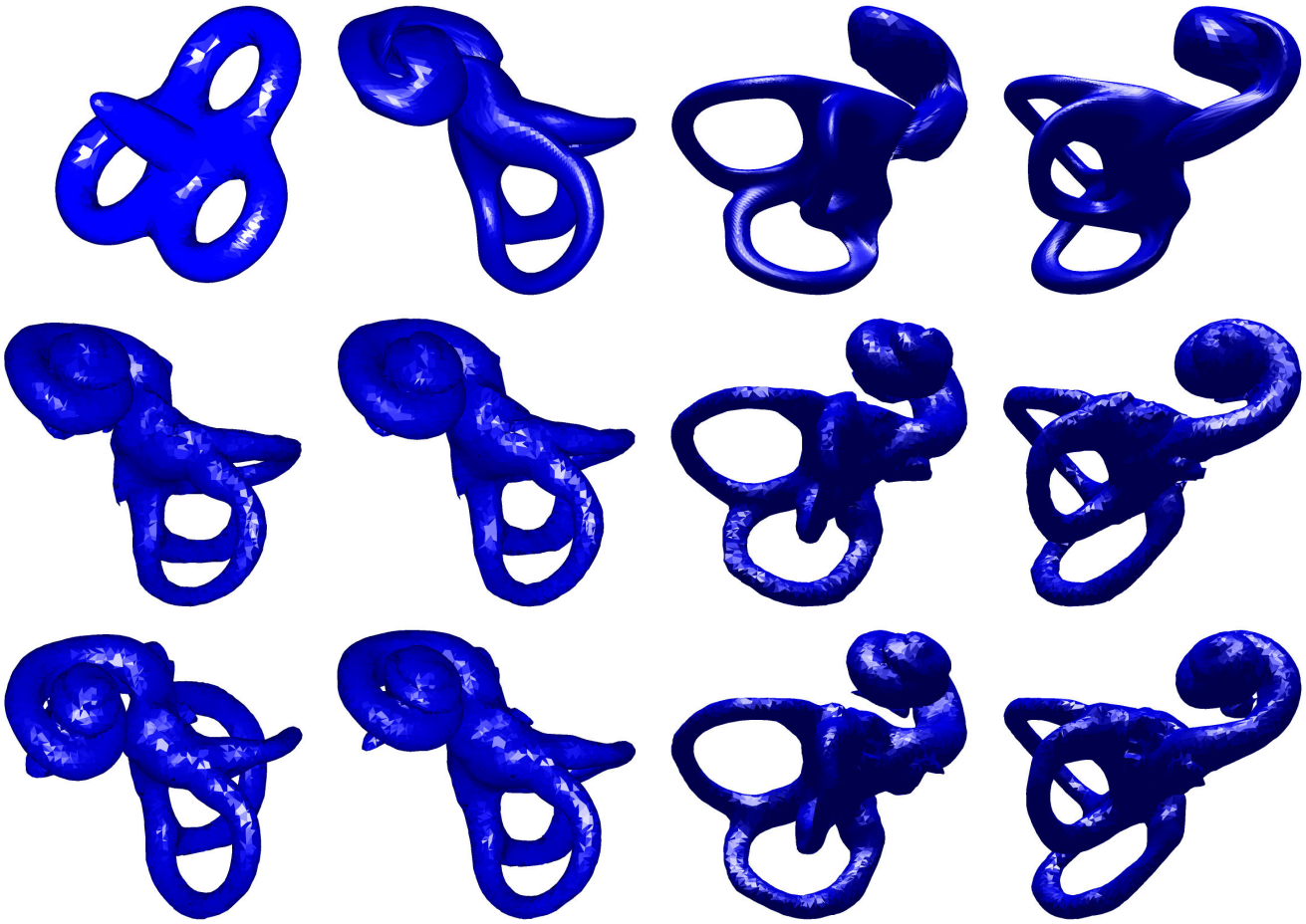


Figure 4.27: The estimated template at the convergence of the algorithm for different initializations. On the first row, the template of figure 4.25. On the second one, the template obtained by initializing with subject 1. The last row with initialization to subject 8.

4.5 Conclusion and extensions

In this chapter, we have focused on the problem of estimating anatomic and anatomo-functional atlases on a population of subjects. This is a way to extend the scope of registration between two individuals, in the perspective of studying group variability of a whole dataset.

In the first place, we proposed algorithms to compute atlases on geometrico-functional shapes carrying real-valued signals, based on the framework of functional currents of chapter 2. We showed that this can be done by estimating iteratively and simultaneously a template functional shape $(\overline{X}, \overline{f})$ from the dataset and a set of momenta p^i that deforms the template to each subject, which is the purpose of algorithm 2. This can be then extended by estimating in addition some residual signals ζ^i on the baseline shape, encoding the functional differences between the template and the different subjects : this resulted in algorithm 3. Both algorithms involved some important gradient normalizations and regularizations to be robust to rescaling, sampling changes or boundary effects. As already mentioned for registration, some natural and po-

tentially quite reachable extensions of such algorithms would consist in generalizing them to other signal spaces than $\mathbb{R}$, vectors or tensors being probably the most interesting cases for practical applications. Another algorithmic improvement could be also brought to the present code by incorporating the **control point** formulation of large deformations, as introduced in [29, 33, 34], which would be beneficial in that it would speed up computations related to diffeomorphisms and enable sparse representation of deformations.

From the theoretical point of view, we addressed in 4.2.4 the central issue of setting a general continuous framework for geometrico-functional atlas estimation, that ensures existence of solutions. We have shown that the existence holds with the rather weak L^2 assumptions on the signal parts, but provided penalty constants are chosen high enough. A pending question would be to examine the theoretical benefits of assuming more regularity on the signals (by working for instance on Sobolev spaces or with bounded variations' functions). Yet, the use of such spaces would obviously introduce new numerical challenges, all the more that the supports are, in our case, varying geometrical shapes. The second important point, that seems achievable in a near future, is to formulate and show the existence of such atlases in the slightly more complex Riemannian framework of metamorphosis that was presented in the first chapter.

In addition, we also examined the question of statistical interpretation and classification based on those atlases. This part is the most likely to lead to future improvements or extensions. Indeed, we chose to focus on simple statistical analysis techniques such as the PCA and its supervised version for multi-class datasets, the LDA. We proposed adaptations of these methods to study the statistical variability of shape and signal based on momenta and functional residuals. Even though more sophisticated statistical learning tools could be used instead, we showed that the combination of geometrico-functional atlas estimation and LDA analysis of momenta and thicknesses leads to interesting classification rates on the detection of glaucoma in the retina database. Undoubtedly, applications to other datasets of this kind enters in the range of our framework and shall be tested in a near future.

Finally, following the issues raised in chapter 3, the case of unoriented shapes was also treated in the atlas estimation point of view. This can be easily done in the varifold setting, and provides template estimation algorithms very close to currents and functional currents. We showed a few results and benefits of such methodologies in comparison to oriented ones on applications to embryonic mice heart membranes on the one hand and on the study of Cochlea shape on the other.

Conclusion and discussion

On the representation of non-oriented shapes

Chapter 3 was focused on the issue of orientation which has been one of the main drawback of the framework of currents in applications to computational anatomy. Imposing a consistent orientation of shapes, as we pointed out, is a clear limitation of this approach in treating several sorts of datasets, including fiber bundles and non-connected shapes. We have also evidenced the currents' cancellation phenomenon which has several negative consequences in practical cases of registration.
The approach that we have developed in this thesis, as an alternative or complement to currents, relies on another concept from geometric measure theory called varifolds. The challenge of chapter 3 was to show that varifolds can be also transposed to the large deformations' setting and to detail the strengths and weaknesses of this new framework. We have proved in particular that the space of varifolds embeds all non-oriented rectifiable subsets and provide representation and metrics that do not suffer from shape annihilation as currents. Registration algorithms based on varifolds were then implemented and tested on a few examples.
As briefly mentioned at the end of chapter 3, there are a few interesting open questions that could follow from this work. One of them would be to establish mathematically a relationship between varifolds and the other representation of non-oriented submanifolds given by normal cycles (cf section 3.1.3). Intuitively, normal cycles seem to be a higher-order representation since it models intrinsically variations of tangent spaces, but this remains to be evaluated in a more rigorous way.
A second promising track, in our point of view, is to go further on the study of the constraints on shapes that result from tangent direction histograms (cf 3.4.1), which, as we showed, corresponds to RKHS on varifolds with infinite scale σ_e. One of the theoretical question that could be examined for example is to relate a deformation path of a given shape in E with the corresponding path in the space of histograms : the real issue there being to find how much (or up to what transformations) a histogram path can determine a shape deformation path. This problematic is of course not only related to varifolds but, in our context, it would give a way to better understand the behavior of varifold-based registration algorithms accordingly to the scale parameters of kernels.

On fshapes' modelling and dissimilarity measures

The main problematic addressed by this dissertation consisted in proposing a framework that can extend the approach of computational anatomy to functional shapes. As we explained in chapter 1, one of the essential challenge was to have a representation that provides dissimilarity metrics for defining relevant data attachment

terms on the set of functional shapes. In chapter 2, we have proposed a possible framework to do so. The key idea was to define a proper generalization of currents, we named *functional currents*, that embeds functional shapes of given dimension into a common functional space which is the dual of a space of extended differential forms. Metrics were then defined on the space of functional currents by considering appropriate reproducing kernel Hilbert spaces.
On the modelling side, functional currents offer similar advantages to currents : the resulting metrics do not require point to point correspondences between objects, are robust to noise and topological perturbations and are controlled by geometrico-functional deformations, as shown by the results of 2.2.3. This theoretical setting is also general enough to treat a wide variety of possible signal types : real, vector or tensor-valued functions are among usual examples. Numerically, we have focused on the case of real-valued signals exclusively and proposed large deformation registration algorithms for functional shapes based on the data attachment metrics provided by fcurrents. As we mentioned at the end of chapter 3, the varifold setting can be also superimposed to all the previous in a very simple way, which results in the notion of *functional varifolds* that can have the additional robustness to orientation issues. A more extensive reflexion on the advantages and drawbacks of functional varifolds versus functional currents, notably on the question of noise, is still to be conducted.
A natural extension of this work, that is essentially a computational issue, is to add other possibilities of signals and group actions to the present code. This comes along with the efforts initiated in section 2.3.3 to improve the numerical complexity and speed for computing the different terms involved in our algorithms and hopefully enable numerical schemes to address applications with higher-dimensional signals.
Last but not least, we want to point out that the functional currents (or functional varifolds) setting is not limited to its use as an intermediate to define data attachment terms for registration or atlas estimation. Indeed, the Hilbert space structure that we obtain through kernels offers many other application possibilities. We have sketched some of these in this thesis, for instance with the Matching-Pursuit procedure of 2.3.1 for data compression. Yet, a still largely unexplored range of applications would be to exploit such metrics for doing direct statistics and classification on geometrico-functional datasets (after gross registration of the subjects), using either discriminant analysis as in section 4.3.1 or support vector machines. This is clearly a less accurate analysis than the atlas-based methods developed in chapter 4 that really estimate geometrico-functional variation from a template fshape but it has the interest of being faster to compute for very big datasets and could thus serve as a preliminary analysis.

On the estimation and statistical analysis of atlases

The final objective of computational anatomy is to study the variability of structures within a certain population of subjects and deduce classification rules to discriminate between different populations. This is the problem that we finally intended to address in the case of functional shapes in chapter 4. Having proposed a satisfying setting to compare fshapes together and estimate deformations between two subjects, we generalized this approach to the problem of atlas estimation and statistical interpretation.
Our first task in chapter 4 was to propose algorithms, adapted from the forward generative model used for images and currents, to compute an atlas from a population

of fshapes. In the simplest setting, these atlases consist in a template fshape together with a set of deformations mapping the template to the subjects. As we have detailed in 4.2.1, there are several algorithmic tricks that can be implemented to improve the estimation algorithm itself, notably concerning robustness to scale and sampling as well as boundary effects. The model can be then enriched by the additional estimation of functional residuals on the template to describe signal variability in an interpretable way. This setting was implemented as well in our current codes.
On the algorithmic side, there are many other tracks that could be followed to extend, speed up or simply improve the present version, some of which we have already experimented while others are still mere suggestions. For instance, *multi-shapes'* situations where all the objects to analyze have several distinct elements (as for example several surface layers with or without signals, or objects mixing curves and surfaces) are quite usual in applications. We have extended our atlas estimation schemes to account for these. Concerning speed and sparsity of representations, the *control points* setting of [29, 34] is probably a good idea to reduce computational times related to the diffeomorphism framework but has not been integrated in our algorithms yet. Another possibility is to work rather on the gradient descent scheme on which the algorithm is based. Computing functionals and gradients at each step is obviously linear in time with respect to the number of subjects in the population and it is clear that, for large datasets, parallelization will not be enough to limit it. One potential solution that remains to be tested for these applications is given by *stochastic gradient* algorithms, which only require such computations for one individual per descent step at the price of a slower convergence rate.
In the second part of chapter 4, we have addressed the next issue of statistical analysis of the variability in shape and function. We proposed an approach based on PCA and LDA in Hilbert spaces applied to the deformation momenta and on the functional residuals obtained by atlas algorithms, that estimates principal modes of variations and discriminant directions for classification. There is surely room for improvement at this level since we chose to focus on simple statistical learning tools whereas more sophisticated ideas from machine learning could be also investigated. Yet, as we evidenced with a few results on the retina dataset, this framework seems quite successful in finding meaningful variation patterns both on the geometric and functional level, and offers significant classification rates on cross-validation tests. Undoubtedly, these results need to be comforted in the future by similar studies carried on other datasets.

A complete and well-posed Riemannian framework for fshapes' group analysis ?

To give a final conclusion to the present work, we wish to come back on the original problematic that we had initiated in section 1.4.1 of chapter 1. We introduced the idea of a metamorphosis framework, in which two functional shapes within a fixed bundle are compared by the geodesic distance induced from the space of fshapes metamorphoses. As we explained, this gives a Riemannian formulation to the problem of exact registration, which amounts in computing a geodesic path in the metamorphoses space. Similarly, within this setting, estimating atlases is simply the computation of a mean in the sense of this geodesic distance.
However, this holds in the ideal situation of shapes belonging to a single orbit and we argued for the necessity of an inexact formulation involving dissimilarity mea-

sures between different orbits. Defining such dissimilarity measures in the context of functional shapes have been one of the important challenge addressed here. It appears therefore as an attainable goal to provide a general methodology for group analysis on functional shapes which would completely rely on a coherent Riemannian framework. What we have presented in this thesis, and notably in the last chapter, is a big step in this direction. Nevertheless, the story is not completed yet in so far as the framework for atlas estimation we considered and implemented was done under the simplified 'tangential model' (cf 1.4.1) of stationary residual functions on the template. A remaining step is to adjust our current framework to the real metamorphosis setting, which would affect essentially the dynamics of the geodesic shooting equations. A last point would be then to extend the existence result of theorem 4.2.1 to this setting.

Appendix A

Proof of theorem 4.2.1

In this appendix, we give the proof of the existence of geometrico-functional atlases expressed by theorem 4.2.1. We have decomposed it in several steps by first showing the existence of the template part (X, f) for a simplified functional and then generalize these arguments to the full atlas model.

A.1 Existence of the template fshape

In all this part, we will consider the simpler situation where all variables v^i and ζ^i in equation (4.18) are freezed and show that a minimum of the functional exists in the variables X and f. We start by introducing a few additional definitions.

A.1.1 Preliminary definitions

Functional currents and measures on $E \times \Lambda^d(E) \times \mathbb{R}$

For the sake of the proof, we shall have to consider, in the following, a slightly bigger space than functional currents which is the set of measures on the product space $F \doteq E \times \Lambda^d(E) \times \mathbb{R}$. Namely, we will call $\mathcal{M}(F)$ the set of finite positive Borel measures on F. For any $f \in L^2(X)$ we denote $\nu_{X,f}$ the finite Borel measure on F such that for any bounded and continuous $g \in C_b(F, \mathbb{R})$:

$$\nu_{X,f}(g) \doteq \int_X g(x, \xi(x), f(x)) d\mathcal{H}^d(x) \tag{A.1}$$

where $\xi(x)$ is the unique unit normed element in $\Lambda^d(E)$ associated with the oriented tangent space T_xX defined at almost every location x on X since X is rectifiable. Note that

$$\nu_{X,f}(|\xi|) = \mathcal{H}^d(X) < \infty\,. \tag{A.2}$$

Moreover, for any $\nu \in \mathcal{M}(F)$ such that $\nu(|\xi|) < \infty$, we can identify ν with a functional current $C_\nu \in W'$ such that for any $\omega \in W$, we have

$$C_\nu(\omega) = \int \omega_{(x,f)}(\xi) d\nu(x, \xi, f)\,. \tag{A.3}$$

One can check indeed that $|C_\nu(\omega)| \leq |\omega|_\infty \nu(|\xi|) \leq K\|\omega\|_W \nu(|\xi|)$ for a fixed $K > 0$ so that

$$\|C_\nu\|_{W'} \leq K\nu(|\xi|)\,. \tag{A.4}$$

Thus, C_ν is indeed a continuous linear form on W.

In addition, if ϕ is a diffeomorphism of E, one can define, for $\nu \in \mathcal{M}(F)$, its transport $\phi_*\nu$ by the following relation, for all $g \in C_b(F, \mathbb{R})$:

$$\phi_*\nu(g) \doteq \int |d_x\phi.\xi|.g(\phi(x), [d_x\phi.\xi], f)d\nu(x, \xi, f)\,. \tag{A.5}$$

where, for $\xi = \xi_1 \wedge ... \wedge \xi_d$, $d\phi_x.\xi$ is, as usual, the transported d-vector $d_x\phi(\xi_1) \wedge ... \wedge d_x\phi(\xi_d)$ and $[d_x\phi.\xi]$ its normed version. One can then easily check that this action commutes with the usual action on functional currents of eq.(2.11) i.e.

$$\phi_* C_\nu = C_{\phi_*\nu}$$

and that

$$\phi_*\nu_{X,f} = \nu_{\phi(X),f\circ\phi^{-1}}\,. \tag{A.6}$$

X supported measures

For a fixed finite-volume, bounded and oriented rectifiable subset X, we introduce a particular subset of $\mathcal{M}(F)$ that we define as the set $\mathcal{M}^X(F)$ of all Borel finite measures ν on $E \times \Lambda^d E \times \mathbb{R}$ such that :

$$\int h(x, \xi)d\nu(x, \xi, f) = \int_X h(x, \xi(x))d\mathcal{H}^d(x)\,. \tag{A.7}$$

for all continuous and bounded applications h on $E \times \Lambda^d(E)$. These are measures that have a particular marginal on $E \times \Lambda^d(E)$. Note that any $\nu_{X,f}$ for f a L^2 function on X belongs to $\mathcal{M}^X(F)$. In addition, $\mathcal{M}^X(F)$ is a closed subset for the usual weak convergence of measures defined by $\nu_n \rightharpoonup \nu_\infty$ if for any $g \in C_b(F, \mathbb{R})$

$$\nu_n(g) \to \nu_\infty(g)\,. \tag{A.8}$$

Perturbation

We introduce also a perturbation process on any measure $\nu \in \mathcal{M}(F)$ that shall be useful for the following. Let $a > 0$ to be fixed later and consider for any $t \in \mathbb{R}$ the function $\rho_t : \mathbb{R} \to \mathbb{R}$ such that

$$\rho_t(z) = z + t(\mathrm{sgn}(z)a - z)\mathbf{1}_{|z|>a} \tag{A.9}$$

where $\mathrm{sgn}(z)$ is the sign of z. We have $\rho_0 = \mathrm{Id}_{\mathbb{R}}$ and ρ_1 is a symmetric threshold at level a. Now for any $t \in \mathbb{R}$, we denote ν_t the new measure defined for any $g \in C_b(F, \mathbb{R})$ as:

$$\nu_t(g) = \int_F g(x, \xi, \rho_t(f))d\nu(x, \xi, f)\,. \tag{A.10}$$

Obviously $\nu_0 = \nu$ and ν_1 is such that $\nu_1(|f| > a) = 0$ so that $t \to \nu_t$ is an homotopy from ν to a measure under which the signal is a.e. bounded by a.

A.1.2 Existence with X fixed

This subsection is dedicated to the proof of the following intermediate proposition :

Proposition A.1.1. *Assume that W is continuously embedded in $C_0^2(E\times\mathbb{R},\Lambda^d E^*)$, that X and $(X^i)_{1\leq i\leq N}$ are finite volume bounded oriented d-dimensional rectifiable subsets and that $f^i\in L^2(X^i)$ for $1\leq i\leq N$. Assume that γ_f/γ_W is large enough with $\gamma_f,\gamma_W\geq 0$. Then*

$$J_X(f)\doteq\frac{\gamma_f}{2}\int_X|f(x)|^2d\mathcal{H}^d(x)+\frac{1}{2}\sum_{i=1}^N\gamma_W\|C_{X^i,f^i}-C_{X,f}\|_{W'}^2)$$

achieves its minimum on $L^2(X)$ and any minimizer f_* is such that $f_*\in L^\infty(X)$. Moreover, if we assume that X is a C^p submanifold and that $W\hookrightarrow C_0^m(E\times\mathbb{R},\Lambda^d E^*)$ with $m\geq\max\{p,2\}$, then $f_*\in C^{p-1}(X)$.

Existence of a measure minimizer

In the first place, we will show existence of a minimizer in the space $\mathcal{M}(F)$ of measures, and thus consider the extended functional $\tilde{J}$ defined for $\nu\in\mathcal{M}^X(F)$ by

$$\tilde{J}(\nu)\doteq\frac{\gamma_f}{2}\nu(|f|^2)+\frac{\gamma_W}{2}\sum_{i=1}^N\|C_{X^i,f^i}-C_\nu\|_{W'}^2 \tag{A.11}$$

where $\nu(|f|^2)$ is the notation we shall use for $\int|f|^2d\nu$.
Now let $\nu\in\mathcal{M}^X(F)$. Consider for $t\in\mathbb{R}$, $J_t\doteq\tilde{J}(\nu_t)$ where ν_t is the previously defined perturbation of ν and assume that $J_0<\infty$ (which is equivalent to say that $\nu(|f|^2)<\infty$). We recall that $\|C_{X^i,f^i}-C_{\nu_t}\|_{W'}^2=(C_{X^i,f^i}-C_{\nu_t})(\omega^i)$ where $\omega^i=K_W(C_{X^i,f^i}-C_{\nu_t})\in W$. Then one easily checks that $J_t<\infty$ and, since we assume that W is continuously embedded into $C_0^2(E\times\mathbb{R},\Lambda^d E^*)$, with existing derivative J_t' at any location t given by

$$J_t'=\nu\left(\frac{d}{dt}(\rho_t(f))\left(\gamma_f\rho_t(f)+\gamma_W\sum_{i=1}^N\frac{\partial\omega^i_{(x,\rho_t(f))}}{\partial f}(\xi)\right)\right) \tag{A.12}$$

Using again the continuous embedding of W into $C_0^2(E\times\mathbb{R},\Lambda^d E^*)$ we get for a constant $C>0$ that

$$\begin{aligned}\left|\frac{\partial\omega^i_{(x,\rho_t(f))}}{\partial f}(\xi)\right|&\leq C\|\omega^i\|_W\\&\leq K\left(\|C_{X^i,f^i}\|_{W'}+\|C_{\nu_t}\|_{W'}\right).\end{aligned}$$

Moreover, it is a trivial verification that we have, by analogy to eq.(1.17) in the case of currents, $\|C_{X^i,f^i}\|_{W'}\leq\text{cte.}\mathcal{H}^d(X^i)$. With similar arguments, $\|C_{\nu_t}\|_{W'}\leq\text{cte.}\nu_t(E\times\Lambda^d(E)\times\mathbb{R})$ and, since $\nu_t\in\mathcal{M}^X(F)$, we have $\nu_t(E\times\Lambda^d(E)\times\mathbb{R})=\mathcal{H}^d(X)$ and consequently $\|C_{\nu_t}\|_{W'}\leq\text{cte.}\mathcal{H}^d(X)$. Thus, there exists a constant $K>0$ such that :

$$\left|\sum_{i=1}^N\frac{\partial\omega^i_{(x,\rho_t(f))}}{\partial f}(\xi)\right|\leq K\sum_{i=1}^N\Big(\mathcal{H}^d(X^i)+\mathcal{H}^d(X)\Big) \tag{A.13}$$

Noticing now that $\frac{d}{dt}(\rho_t(f))\,\rho_t(f)\leq 0$, that $|\frac{d}{dt}(\rho_t(f))|=0$ for $|f|\leq a$ and that $|\rho_t(f)|\geq a$ for $|f|\geq a$ and $t\in[0,1]$, we get for $t\in[0,1]$

$$J_t'\leq\nu\left(-\left|\frac{d}{dt}(\rho_t(f))\right|\mathbf{1}_{|f|>a}\left(\gamma_f a-\gamma_W K\sum_{i=1}^N\Big(\mathcal{H}^d(X^i)+\mathcal{H}^d(X)\Big)\right)\right) \tag{A.14}$$

so that

$$\tilde{J}(\nu_1) \leq \tilde{J}(\nu_0) \text{ if } a \geq K\frac{\gamma_W}{\gamma_f}\sum_{i=1}^{N}\left(\mathcal{H}^d(X^i)+\mathcal{H}^d(X)\right). \tag{A.15}$$

An important consequence of (A.15) is that one can restrict the search of a minimum for $\tilde{J}$ on $\nu \in \mathcal{M}^X(F)$ such that

$$\nu(\mathbf{1}_{|f|>a}) = 0 \tag{A.16}$$

with $a = K\frac{\gamma_W}{\gamma_f}\sum_{i=1}^{N}\left(\mathcal{H}^d(X^i)+\mathcal{H}^d(X)\right)$. In particular, since $\nu \in \mathcal{M}^X(F)$, we will have

$$x \in X,\ |\xi| \leq 1 \ \text{ and } |f| \leq a\ \nu \text{ a.e.} \tag{A.17}$$

Since X is bounded, we can restrict the search of a minimum on a measure supported on a *compact* subset $C \subset F$ so that we introduce :

$$\mathcal{M}^{X,C}(F) \doteq \{\ \nu \in \mathcal{M}^X(F) \mid (x,\xi,f) \in C \ \ \nu \text{ a.e. }\}. \tag{A.18}$$

One can then check easily that $\tilde{J}$ is lower semi-continuous on the set $\mathcal{M}^{X,C}(F)$ for the weak convergence topology. In addition, $\mathcal{M}^{X,C}(F)$ is sequentially compact. Indeed, if ν_n is a sequence in $\mathcal{M}^{X,C}(F)$ then all ν_n are supported by the compact C and in particular (ν_n) is tight. Also, as we already noticed, there exists a constant cte independent of n such that $\nu_n(F) \leq \text{cte}\mathcal{H}^d(X)$ and thus the sequence is uniformly bounded for the total variation norm. It results, thanks to Prokhorov theorem, that there exists a subsequence of (ν_n) converging for the weak topology. These compactness and lower semicontinuity properties guarantee the existence of a minimizer ν_* of $\tilde{J}$ with $\nu_* \in \mathcal{M}^{X,C}(F)$ and

$$\tilde{J}(\nu_*) \leq \inf_{f\in L^2(X)} J_X(f). \tag{A.19}$$

Existence of a minimizing function f

At that point we do not have yet a minimizer of J_X. The problem is that if the marginal on $E \times \Lambda^d(E)$ of ν_* is the transport of $\mathcal{H}^d_{|X}$ under the application $x \mapsto (x,\xi(x))$, we cannot guarantee that ν_* does not weight multiple signal values in the fiber above a location $(x,\xi(x))$. We will show now that for γ_f/γ_W large enough there exists $f_* \in L^2(X)$ such that $\nu_* = \nu_{X,f_*}$ so that we will deduce

$$J_X(f_*) = \tilde{J}(\nu_*) \leq \inf_{f\in L^2(X)} J_X(f) \tag{A.20}$$

and the existence of a minimizer on $L^2(X)$.
Let $\delta f \in C_b(F,\mathbb{R})$ and for any $t \in \mathbb{R}$ consider the perturbation $\nu_t \in \mathcal{M}^X(F)$ of $\nu \in \mathcal{M}^{X,C}(F)$ such that for all $g \in C_b(F,\mathbb{R})$ we have :

$$\nu_t(g) \doteq \int g(x,\xi,f+t\delta f(x,\xi,f))d\nu(x,\xi,f). \tag{A.21}$$

Here again, the function $t \to \tilde{J}(\nu_t)$ is differentiable everywhere and we have for $\omega^i \doteq K_W(C_{X^i,f^i} - C_\nu)$

$$\frac{d}{dt}\tilde{J}(\nu_t)|_{t=0} = \nu\left(\left(\gamma_f f + \gamma_W\sum_{i=1}^{N}\frac{\partial\omega^i_{(x,f)}}{\partial f}(\xi)\right)\delta f(x,\xi,f)\right),$$

so that when $\nu = \nu_*$ we get

$$\begin{cases} \gamma_f f + \gamma_W A(x,\xi,f) = 0 \ \nu_* \text{ a.e.} \\ \\ \text{with} \\ A(x,\xi,f) \doteq \sum_{i=1}^N \frac{\partial \omega^i_{(x,f)}}{\partial f}(\xi)\,. \end{cases} \tag{A.22}$$

The partial derivative of $f \mapsto \gamma_f f + \gamma_W A(x,\xi,f)$ with respect to f equals $\gamma_f + \gamma_W \frac{\partial A}{\partial f}(x,\xi,f)$. Using as previously the continuous embedding $W \hookrightarrow C_0^2(E\times\mathbb{R}, \Lambda^d E^*)$, one can show that we have once again a certain constant K such that $\frac{\partial A}{\partial f}(x,\xi,f)$ can be bounded uniformly by $K\sum_{i=1}^N \left(\mathcal{H}^d(X^i) + \mathcal{H}^d(X)\right)$. It results that for γ_f/γ_W large enough and for all $(x,V) \in E \times \Lambda^d E$, $f \mapsto \gamma_f f + \gamma_W A(x,\xi,f)$ is a strictly increasing function going to $-\infty$ (resp $+\infty$) at $-\infty$ (resp $+\infty$) and thus there is a unique solution $\tilde{f}(x,\xi)$ to (A.22). Now, since the application $f \mapsto \gamma_f f + \gamma_W A(x,\xi,f)$ is also C^1 on $E \times \Lambda^d E \times \mathbb{R}$, we deduce from the implicit function theorem that $\tilde{f}$ is a C^1 function on $E \times \Lambda^d E$. Going back to the solution ν_*, we know that for ν_* almost every $(x,\xi,f) \in E \times \Lambda^d E \times \mathbb{R}$, we have $(x,\xi,f) \in C$ and $f = \tilde{f}(x,V)$, so that $|\tilde{f}| \leq a$ a.e. For any continuous and bounded function g :

$$\nu_*(g) = \int g(x,\xi,f) d\nu_* = \int g(x,\xi,\tilde{f}(x,\xi)) d\nu_*$$

and if we denote by $\tilde{g}(x,\xi) \doteq g(x,\xi,\tilde{f}(x,V))$ which is a continuous and bounded function on $E \times \Lambda^d E$, we have by definition of the space $\mathcal{M}^X(F)$ (eq.(A.7)) :

$$\nu_*(g) = \nu_*(\tilde{g}) = \int_X g(x,\xi(x),\tilde{f}(x,\xi(x))) d\mathcal{H}^d(x)\,. \tag{A.23}$$

Therefore, setting $f_*(x) = \tilde{f}(x,\xi(x))$ for $x \in X$, we see that $|f_*| \leq a$ so that $f_* \in L^\infty(X)$ and with eq.(A.23), we deduce that $\nu_* = \nu_{(X,f_*)}$ which shows that the solution of the optimization is the fcurrent associated to the fshape (X, f_*). Moreover, if we assume that $W \hookrightarrow C_0^m(E \times \mathbb{R}, \Lambda^d E^*)$ with $m \geq \max\{p, 2\}$ then A and $\tilde{f}$ are C^{p-1} functions and if X is a C^p submanifold then $x \mapsto \xi(x)$ is a C^{p-1} function on X so that f_* is also C^{p-1}, which concludes the proof of proposition A.1.1.

A.1.3 Existence with non-fixed X

We now consider the existence of a template when X is no more fixed and is estimated as well. In this case, as we mentioned in 4.2.4, it is enough to introduce a RKHS Hilbert space V_0 continuously embedded in $C_0^2(E,E)$, an initial hypertemplate X_0 and consider also an optimization of the template X in the orbit of X_0 under the action of $\phi^{v_0} \in G_{V_0}$, the group of diffeomorphisms associated with V_0 i.e. $X = \phi^{v_0}(X_0)$. To prove existence result, we will need to introduce a penalization depending on the distance between X_0 and X i.e. on $d_{G_{V_0}}(\mathrm{Id}, \phi^{v_0})$. A typical functional would be, if $\mathcal{X}_0$ is the orbit of X_0 under G_{V_0} and $d_{\mathcal{X}_0}(X,X') \doteq \inf_{\phi \in G_{V_0}, \phi(X)=X'} d_{G_{V_0}}(\mathrm{Id},\phi)$ is the induced distance between two templates in the orbit,

$$\begin{aligned} J_1(X,f) = &\frac{\gamma_0}{2} d_{\mathcal{X}_0}(X_0,X)^2 \\ &+ \frac{\gamma_f}{2}\int_X |f(x)|^2 d\mathcal{H}^d(x) + \frac{\gamma_W}{2}\sum_{i=1}^N \|C_{X^i,f^i} - C_{X,f}\|^2_{W'} \end{aligned} \tag{A.24}$$

Since a diffeomorphism $\phi^0 \in G_{V_0}$ is obtained as the flow of a time-varying vector field $v^0 \in L^2([0,1], V_0)$, we consider the minimization of the following functional :

$$\begin{aligned} J_2(v^0, f) = & \frac{\gamma_0}{2}\|v^0\|^2_{L^2_{V_0}} \\ & + \frac{\gamma_f}{2}\int_X |f(x)|^2 d\mathcal{H}^d(x) + \frac{\gamma_W}{2}\sum_{i=1}^N \|C_{X^i,f^i} - C_{X,f}\|^2_{W'} \end{aligned} \tag{A.25}$$

for $v^0 \in L^2_{V_0}$, $f \in L^2(X)$ with $X = \phi^{v_0}(X)$. However, it is more convenient for $X = \phi^{v_0}(X_0)$ to consider the change of variable $f \mapsto f_0 = f \circ \phi^{v_0}$ from $L^2(X) \to L^2(X_0)$ so that we keep working in a fixed space $L^2(X_0)$. Hence we end up with functional

$$\begin{aligned} J_3(v^0, f_0) = & \underbrace{\frac{\gamma_0}{2}\|v^0\|^2_{L^2_{V_0}}}_{\text{penalization on } X} + \frac{\gamma_f}{2} \underbrace{\int_{X_0} |f_0(x)|^2 |d_x\phi^{v_0}.\xi(x)| d\mathcal{H}^d(x)}_{=\int_X |f(x)|^2 d\mathcal{H}^d(x) \text{ for } f = f_0 \circ (\phi^{v_0})^{-1} \in L^2(X)} \\ & + \frac{\gamma_W}{2}\sum_{i=1}^N \|C_{X^i,f^i} - (\phi_1^{v_0})_* C_{X_0,f_0}\|^2_{W'} \end{aligned} \tag{A.26}$$

with $f_0 \in L^2(X_0)$. The existence of a minimizer $(v^0_*, f_{0,*})$ for J_3 gives immediately the existence of a minimizer (X_*, f_*) for J_1 with $X_* \doteq \phi_1^{v^0_*}(X_0)$, $f_* \doteq f_{0,*} \circ (\phi_1^{v^0_*})^{-1} \in L^2(X_*)$.
The existence result then becomes :

Proposition A.1.2. *Assume that W is continuously embedded in $C^2_0(E \times \mathbb{R}, \Lambda^d E^*)$, that X and $(X^i)_{1\le i\le N}$ are finite volume bounded oriented d-dimensional rectifiable subsets and that $f^i \in L^2(X^i)$ for $1 \le i \le N$. Assume that γ_f/γ_W is large enough with $\gamma_f, \gamma_W \ge 0$ and $\gamma_0 > 0$. Then*

$$\begin{aligned} J(X, f) \doteq & \frac{\gamma_0}{2} d_{\mathcal{X}_0}(X_0, X)^2 \\ & + \frac{\gamma_f}{2}\int_X |f(x)|^2 d\mathcal{H}^d(x) + \frac{\gamma_W}{2}\sum_{i=1}^N \|C_{X^i,f^i} - C_{X,f}\|^2_{W'} \end{aligned}$$

achieves its minimum on $\{(X, f \mid X \in \mathcal{X}_0,\ f \in L^2(X)\}$ and any minimizer (X_*, f_*) is such that $f_* \in L^\infty(X_*)$. Moreover, if X_0 is a C^p submanifold and $W \hookrightarrow C^m_0(E \times \mathbb{R}, \Lambda^d E^*)$ with $m \ge \max\{p, 2\}$, $f_* \in C^{p-1}(X_*)$.

The proof of Proposition A.1.1 can be in fact easily adapted to this new situation. We can consider the formulation of equation (A.26) with J_3 a functional on the vector field v^0 and the function $f \in L^2(X_0)$. With respect to v_0, thanks to the penalization in (A.26), we can restrict the search of a minimum on a closed ball B_0 of given radius b in $L^2_{V_0}$, which guarantees at the same time that the Jacobians $|d_x\phi^{v_0}.\xi(x)|$ are uniformly lower bounded. This closed ball is also compact for the weak topology in $L^2_{V_0}$. In addition, it follows from corollary 2.2.2 that $v^0 \mapsto \sum_{i=1}^N \|C_{X^i,f^i} - (\phi_1^{v_0})_* C_{X_0,f_0}\|^2_{W'} = \sum_{i=1}^N \|C_{X^i,f^i} - C_{\phi_1^{v_0}(X_0), f_0 \circ (\phi_1^{v_0})^{-1}}\|^2_{W'}$ is weakly continuous on $L^2_{V_0}$ and it is also classical that $v^0 \mapsto \|v^0\|_{L^2_{V_0}}$ is lower semi-continuous for the weak convergence topology. Therefore, for all fixed $f_0 \in L^2(X_0)$, $v^0 \mapsto J_3(v^0, f_0)$ is weakly lower semicontinuous on $L^2_{V_0}$. It results that we obtain existence of a minimizing vector field v^0 and, reasoning as in the previous subsection, one deduces easily the claim of proposition A.1.2.

A.2 Existence of full fshape atlases

We are now ready to generalize the previous results to the existence of complete atlases of fshapes' datasets that is the claim of theorem 4.2.1. In addition to the template (X, f), one wants to estimate simultaneously transformations from the template to each subject. These consist in deformations $(\phi^{v_i})_{i=1,..,N}$ obtained as flows of time varying vector fields (v^i) and residual signals (ζ^i) that are L^2 functions on X. The proof of the theorem can be done by leaning essentially on the same arguments as detailed in the previous subsections. We shall go rapidly through each of them by referring also to what was proved previously. First of all, exactly as in A.1.3, existence of a template shape X is guaranteed with the same compacity and lower semicontinuity arguments. Thus we may assume that X is fixed and we only have to show existence of minimizers to the simplified functional :

$$J_X(f,(\zeta^i),(v^i)) \doteq \frac{\gamma_f}{2}\int_X |f(x)|^2 d\mathcal{H}^d(x)$$
$$+\frac{1}{2}\sum_{i=1}^N \left(\|v^i\|^2_{L^2_V} + \gamma_\zeta \int_X |\zeta^i(x)|^2 d\mathcal{H}^d(x) + \gamma_W \|C_{X^i,f^i} - C_{\phi^{v^i}(X),f+\zeta^i}\|^2_{W'}\right)$$

Now, as for v^0, due to the presence of the penalizations $\|v^i\|^2_{L^2_V}$, one can assume that all vector fields v^i belong to a fixed closed ball B of radius $r > 0$ in L^2_V. As in the proof of proposition A.1.1, we first show existence of a minimizer in a space of measures. Namely, extending the definitions of the previous subsections, we introduce the space $\mathcal{M}^X(F)$ of measures ν on $F \doteq E \times \Lambda^d E \times \mathbb{R} \times \mathbb{R}^N$ that are supported on X, i.e such that for all continuous and bounded function h on $E \times \Lambda^d E$, we have :

$$\nu(h) = \int h(x,V) d\nu(x,\xi,f,(\zeta^i)) = \int_X h(x,\xi(x)) d\mathcal{H}^d(x)\,.$$

For a measure ν on F and a diffeomorphism ϕ, we denote by $\phi^*\nu$ the transport of ν by ϕ defined,by analogy with eq.(A.5) by :

$$(\phi_*\nu)(g) = \int |d_x\phi.\xi| g(\phi(x),[d_x\phi.\xi],f,(\zeta^i)) d\nu(x,\xi,f,(\zeta^i))\,.$$

We now introduce the extended functional :

$$\tilde{J}(\nu,(v^i)) \doteq \frac{\gamma_f}{2}\nu(|f|^2) + \frac{1}{2}\sum_{i=1}^N \left(\|v^i\|^2_{L^2_V} + \gamma_\zeta \nu(|\zeta^i|^2) + \gamma_W \|C_{X^i,f^i} - (\phi^{v^i})_* C_{\nu^i}\|^2_{W'}\right)$$

for $\nu \in \mathcal{M}^X(F)$, $(v^i) \in (L^2_V)^N$ and where, for all $i \in \{1,..,N\}$, ν^i is the measure on $E \times \Lambda^d(E) \times \mathbb{R}$ defined for all $g \in C_b(E \times \Lambda^d(E) \times \mathbb{R}, \mathbb{R})$ by :

$$\nu^i(g) = \int g(x,\xi,f+\zeta^i) d\nu(x,\xi,f,(\zeta^j)) \tag{A.27}$$

and C_{ν^i} the fcurrent defined for all $\omega \in W$ as in (A.3) :

$$C_{\nu^i}(\omega) = \int \omega_{(x,f+\zeta^i)}(\xi) d\nu^i(x,\xi,f,(\zeta^i))\,.$$

As previously, we can consider the perturbation function ρ_t acting on signals and the measures

$$\nu_t(g) \doteq \int g(x,\xi,\rho_t(f),(\rho_t(\zeta^i))) d\nu(x,\xi,f,(\zeta^i))\,.$$

Denoting $J_t = \tilde{J}(\nu_t, (v^i))$, we have, for $t \in [0,1]$,

$$J_t' = \nu\left(\frac{d}{dt}(\rho_t(f))\left(\gamma_f \rho_t(f) + \gamma_W \sum_{i=1}^N \frac{\partial \omega^i_{(\phi^{v^i}(x),\rho_t(f)+\rho_t(\zeta^i))}}{\partial f}(d_x\phi^{v^i}.\xi)\right)\right)$$
$$+ \nu\left(\sum_{i=1}^N \frac{d}{dt}(\rho_t(\zeta^i))\left(\gamma_\zeta \rho_t(\zeta^i) + \gamma_W \frac{\partial \omega^i_{(\phi^{v^i}(x),\rho_t(f)+\rho_t(\zeta^i))}}{\partial f}(d_x\phi^{v^i}.\xi)\right)\right) \quad (A.28)$$

where, for all $i \in \{1,..,N\}$, $\omega^i = K_W(C_{X^i,f^i} - (\phi^{v^i})_* C_{\nu_t^i})$. On the first hand, we know that there exists a constant cte such that for all i, $x \in E$ and $\xi \in \Lambda^d E$, $|\xi| = 1$, $|d_x\phi^{v^i}.\xi| \leq \text{cte}|d\phi^{v^i}|_\infty$. In addition, it is a classical result on flows of differential equations (cf [86]) that there exists a non-decreasing continuous function $C : \mathbb{R}^+ \to \mathbb{R}^+$, $C(0) = 0$, independent of $v \in L^2_V$ such that $|d\phi^v|_\infty \leq C(\|v\|_{L^2_V})$. Therefore, since for all i, $\|v^i\|_{L^2_V} \leq r$

$$\left|\frac{\partial \omega^i_{(\phi^{v^i}(x),\rho_t(f)+\rho_t(\zeta^i))}}{\partial f}(d_x\phi^{v^i}.\xi)\right| \leq C(\|v^i\|_{L^2_V}).\left|\frac{\partial \omega^i}{\partial f}\right|_\infty$$
$$\leq C(r).\left|\frac{\partial \omega^i}{\partial f}\right|_\infty .$$

Now, using the same controls as in the previous subsections, we have, on the other hand :

$$\left|\frac{\partial \omega^i}{\partial f}\right|_\infty \leq \text{cte}\|\omega^i\|_W$$
$$\leq \text{cte}(\|C_{X^i,f^i}\|_{W'} + \|(\phi^{v^i})_* C_{\nu_t^i}\|_{W'})$$
$$\leq \text{cte}(\mathcal{H}^d(X^i) + (\phi^{v^i})^* \nu_t^i(E \times \Lambda^d E \times \mathbb{R}))\,.$$

It is easy to check, using eq.(A.5), that

$$(\phi^{v^i})^* \nu_t^i(E \times \Lambda^d E \times \mathbb{R})) \leq \text{cte}|d\phi^{v^i}|_\infty \nu_t^i(E \times \Lambda^d E \times \mathbb{R})$$

and, using the fact that $\nu \in \mathcal{M}^X(F)$ as already argued in A.1.2, $\nu_t^i(E \times \Lambda^d E \times \mathbb{R}) = \mathcal{H}^d(X)$. Also, since all v^i's are bounded by r for the L^2_V-norm, the terms $|d\phi^{v^i}|_\infty$ can be also bounded by a common constant (cf lemma 2.2.3). It results, from all the previous inequalities, the existence of a constant $K > 0$ such that for all i, x, V, f, ζ^i :

$$\left|\frac{\partial \omega^i_{(\phi^{v^i}(x),\rho_t(f)+\rho_t(\zeta^i))}}{\partial f}(d_x\phi^{v^i}.\xi)\right| \leq K.(\mathcal{H}^d(X^i) + \mathcal{H}^d(X)) \quad (A.29)$$

Following the same path that previously lead to (A.14)

$$J_t' \leq \nu\left(-\left|\frac{d}{dt}(\rho_t(f))\right|\mathbf{1}_{|f|>a}\left(\gamma_f a - \gamma_W K \sum_{i=1}^N \left(\mathcal{H}^d(X^i) + \mathcal{H}^d(X)\right)\right)\right)$$
$$+ \sum_{i=1}^N \nu\left(-\left|\frac{d}{dt}(\rho_t(\zeta^i))\right|\mathbf{1}_{|\zeta^i|>a}\left(\gamma_\zeta a - \gamma_W K\left(\mathcal{H}^d(X^i) + \mathcal{H}^d(X)\right)\right)\right). \quad (A.30)$$

Just as in A.1.2, this implies that $\tilde{J}(\nu_1,(v^i)) \leq \tilde{J}(\nu_0,(v^i))$ as soon as :

$$\begin{cases} a \geq \frac{\gamma_W}{\gamma_f} \sum_{i=1}^N K\left(\mathcal{H}^d(X^i) + \mathcal{H}^d(X)\right) \\ \text{ and} \\ a \geq \max_i \frac{\gamma_W}{\gamma_\zeta} K\left(\mathcal{H}^d(X^i) + \mathcal{H}^d(X)\right) \end{cases}$$

Therefore, one may restrict the search of a minimum on a set of measures ν that are supported on a compact subset C of $E \times \Lambda^d E \times \mathbb{R} \times \mathbb{R}^N$, which space we shall denote again $\mathcal{M}^{X,C}(F)$. The rest of the proof is now very close to the one of A.1.2. Due to lower semi-continuity of the functional and the compacity of $\mathcal{M}^{X,C}(F)$ and B for the weak convergence topologies (respectively on the space of measures and on L^2_V), we obtain the existence of a minimizer $(\nu_*,(v^i)_*)$ for the functional $\tilde{J}$.
The last step is to prove that ν_*, which belongs a priori to the measure space $\mathcal{M}^{X,C}(F)$, can be written under the form $\nu_* = \nu_{X,f_*,(\zeta^i_*)}$, i.e that there exists functions f_* and ζ^i_* on X such that, for all continuous and bounded function g on F :

$$\nu_*(g) = \int_X g(x,\xi(x),f_*(x),(\zeta^i_*(x)))d\mathcal{H}^d(x) \tag{A.31}$$

Hence, we consider variations of the signals $(\delta f,(\delta\zeta^i))$ all belonging to the space $C_b(F,\mathbb{R})$ and the path $t \mapsto \nu_t$ defined by :

$$\nu_t(g) = \int g(x,\xi,f+t\delta f(x,\xi,f,(\zeta^i)),(\zeta^i + t\delta\zeta^i(x,\xi,f,(\zeta^i))))d\nu_*(x,\xi,f,(\zeta^i))\,.$$

Now, if $J_t \doteq \tilde{J}(\nu_t,v^i_*)$, expressing that $J'_t|_{t=0} = 0$ for all δf and $(\delta\zeta^i)$ gives the following set of equations :

$$(\gamma_f f,(\gamma_\zeta\zeta^i)) = -A(x,V,f,(\zeta^i))\ \nu_*\text{-a.e} \tag{A.32}$$

with $A(x,\xi,f,(\zeta^i)) \doteq \left(\sum_{i=1}^N \frac{\partial\omega^i}{\partial f}(\phi^{v^i_*}(x),d_x\phi^{v^i_*}.\xi,f+\zeta^i),(\frac{\partial\omega^i}{\partial f}(\phi^{v^i_*}(x),d_x\phi^{v^i_*}.\xi,f+\zeta^i))\right)$.
The derivatives $\partial_f A$ and $\partial_{\zeta^i} A$ can be shown again to be uniformly bounded in x,V,f,ζ^i. A previous argument provides the existence, for γ_f/γ_W and γ_ζ/γ_W large enough, of unique solutions $f = \tilde{f}(x,V)$ and $\zeta^i = \tilde{\zeta}^i(x,V)$ to (A.32). The rest of the proof is exactly the same as in A.1.2 : we set $f_*(x) = \tilde{f}(x,\xi(x))$ and $\zeta^i_*(x) = \tilde{\zeta}^i(x,\xi(x))$, which are again L^∞ functions on X. In addition, one shows easily that the minimizing measure ν_* equals $\nu_{X,f_*,(\zeta^i_*)}$ in the sense of (A.31). Finally, the regularity of f_* and ζ_* when X is a C^p submanifold is obtained again by applying the implicit function theorem to (A.32).

Bibliography

[1] P. A. Absil, R. Mahony, and R. Sepulchre. Riemannian geometry of grassmann manifolds with a view on algorithmic computation. *Acta Appl. Math.*, 80:199–220, Jan 2004.

[2] D. Alexander, J. Gee, and R. Bajcsy. Strategies for data reorientation during non-rigid warps of diffusion tensor images. In C. Taylor and A. Colchester, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI'99*, volume 1679 of *Lecture Notes in Computer Science*, pages 463–472. Springer Berlin Heidelberg, 1999.

[3] D. C. Alexander and J. C. Gee. Elastic matching of diffusion tensor images. *Computer Vision and Image Understanding*, 77(2):233 – 250, 2000.

[4] W. Allard. On the first variation of a varifold. *Annals of mathematics*, 95(3), 1972.

[5] S. Allassonnière. *Représentation et estimation statistique de modèles déformables pour la reconnaissance de formes et l'anatomie numérique.* PhD thesis, 2007.

[6] S. Allassonnière, Y. Amit, and A. Trouvé. Towards a coherent statistical framework for dense deformable template estimation. *Journal of the Royal Statistical Society*, 69(1):3–29, 2007.

[7] S. Allassonnière, A. Trouvé, and L. Younes. Geodesic Shooting and Diffeomorphic Matching Via Textured Meshes. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 3757 of *Lecture Notes in Computer Science*, pages 365–381. 2005.

[8] F. Almgren. *Plateau's Problem: An Invitation to Varifold Geometry.* Student Mathematical Library, 1966.

[9] V. Arnold. Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits. *Annales de l'Institut Fourier*, 16(2):319–361, 1966.

[10] N. Aronszajn. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68:337–404, 1950.

[11] B. Avants, C. Epstein, and J. Gee. Geodesic image normalization and temporal parameterization in the space of diffeomorphisms. *Lecture notes in computer science*, 4091:9–16, 2006.

[12] G. Baudat and F. Anouar. Generalized Discriminant Analysis using a kernel approach. *Neural computation*, 12:2385–2404, 2000.

[13] F. Beg, P. Helm, E. McVeigh, M. Miller, and R. Winslow. Computational cardiac anatomy using magnetic resonance imaging. *Magn. Reson. Med*, 52:1167–1174, 2004.

[14] M. Beg, M. Miller, A. Trouvé, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision*, 61(139-157), 2005.

[15] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):509–522, 2002.

[16] W. Boothby. *An introduction to differentiable manifolds and Riemannian geometry*. Academic Press, 1986.

[17] E. Bretin, S. Masnou, and E. Oudet. Phase-field approximations of the Willmore functional and flow. *preprint*, 2013.

[18] A. M. Bronstein, M. M. Bronstein, R. Kimmel, M. Mahmoudi, and G. Sapiro. A Gromov-Hausdorff Framework with Diffusion Geometry for Topologically-Robust Non-rigid Shape Matching. *International Journal of Computer Vision*, 89(2-3):266–286, 2010.

[19] M. Bruveris, L. Risser, and F. Vialard. Mixture of Kernels and Iterated Semidirect Product of Diffeomorphisms Groups. *Multiscale Modeling and Simulation*, 10(4):1344–1368, 2012.

[20] C. Carmeli, E. De Vito, A. Toigo, and V. Umanita. Vector valued reproducing kernel hilbert spaces and universality. *Analysis and Applications*, 8(01):19–61, 2010.

[21] N. Charon and A. Trouvé. Functional currents : a new mathematical tool to model and analyse functional shapes. *JMIV*, 2013.

[22] N. Charon and A. Trouvé. The Varifold Representation of Non-oriented Shapes for Diffeomorphic Rregistration. *SIAM journal on Imaging Sciences*, 6:2547–2580, 2013.

[23] G. Charpiat. *Distance-based shape statistics for image segmentation with priors*. PhD thesis, 2006.

[24] M. Chupin, E. Gérardin, R. Cuingnet, C. Boutet, L. Lemieux, S. Lehéricy, H. Benali, L. Garnero, and O. Colliot. Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. *Hippocampus*, 19(6):579–587, 2009.

[25] D. Cohen-Steiner and J. Morvan. Restricted delaunay triangulation and normal cycle. *Comput. Geom*, pages 312–321, 2003.

[26] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, M.-O. Habert, M. Chupin, H. Benali, and O. Colliot. Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. *NeuroImage*, 56(2):766–781, 2011.

[27] P. Dupuis, U. Grenander, and M. I. Miller. Variational Problems on Flows of Diffeomorphisms for Image Matching. 1998.

[28] S. Durrleman. *Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution*. PhD thesis, 2009.

[29] S. Durrleman, S. Allassonnière, and S. Joshi. Sparse adaptive parameterization of variability in image ensembles. *International Journal of Computer Vision*, 101(1):161–183, 2013.

[30] S. Durrleman, P. Fillard, X. Pennec, A. Trouvé, and N. Ayache. Registration, atlas estimation and variability analysis of white matter fiber bundles modeled as currents. *NeuroImage*, 2010.

[31] S. Durrleman, X. Pennec, A. Trouvé, and N. Ayache. Measuring brain variability via sulcal lines registration: a diffeomorphic approach. *Medical Image Computing and Computer-Assisted Intervention*, 10(Pt 1):675–682, 2007.

[32] S. Durrleman, X. Pennec, A. Trouvé, and N. Ayache. A forward model to build unbiased atlases from curves and surfaces. *Proc. of the International Workshop on the Mathematical Foundations of Computational Anatomy*, 2008.

[33] S. Durrleman, M. Prastawa, N. Charon, J. Korenberg, S. Joshi, G. Gerig, and A. Trouvé. Deformetrics : morphometry of shape complexes with space deformations. *preprint*, 2013.

[34] S. Durrleman, M. Prastawa, J. Korenberg, S. Joshi, A. Trouvé, and G. Gerig. Topology preserving atlas construction from shape data without correspondence using sparse parameters. *Lecture Notes in Computer Science*, 7512:223–230, 2012.

[35] S. Durrleman, A. Trouvé, and N. Ayache. Sparse approximations of currents for statistics on curves and surfaces. *Proc. Medical Image Computing and Computer Assisted Intervention*, sep 2008.

[36] S. Durrleman, A. Trouvé, N. Ayache, and X. Pennec. Statistical models of sets of curves and surfaces based on currents. *Medical Image Analysis*, 2009.

[37] H. Federer. *Geometric measure theory*. Springer, 1969.

[38] J. H. Friedman. Regularized Discriminant Analysis. *J. Amer. Statist. Assoc.*, 84:165–175, 1989.

[39] J. Fu. Curvature of singular spaces via the normal cycle. *Proceedings of Symposia in Pure Mathematics*, 54(211-221), 1993.

[40] E. Gibson, A. Khan, and M. Beg. A Combined Surface And VOlumetric Registration (SAVOR) Framework to Study Cortical Biomarkers and Volumetric Imaging Data. *MICCAI*, 5761:713–720, 2009.

[41] J. Glaunès. *Transport par difféomorphismes de points, de mesures et de courants pour la comparaison de formes et l'anatomie numérique*. PhD thesis, 2005.

[42] J. Glaunès, A. Trouvé;, and L. Younes. Diffeomorphic matching of distributions: A new approach for unlabelled point-sets and sub-manifolds matching. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2:712–718, 2004.

[43] J. Glaunes and M. Vaillant. Surface matching via currents. *Proceedings of Information Processing in Medical Imaging (IPMI), Lecture Notes in Computer Science*, 3565(381-392), 2006.

[44] U. Grenander. *General pattern theory: A mathematical study of regular structures*. Clarendon Press Oxford, 1993.

[45] U. Grenander and M. I. Miller. Computational anatomy: an emerging discipline. *Q. Appl. Math.*, LVI(4):617–694, 1998.

[46] J. Hamm and D. Lee. Grassmann discriminant analysis : a unifying view on subspace-based learning. *Proceedings of the 25th International Conference on Machine Learning*, 2008.

[47] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning*. Springer, 2001.

[48] P. Howland, M. Jeon, and H. Park. Structure preserving dimension reduction for clustered text data based on the generalized singular value decomposition. *SIAM J. Matrix Anal. Appl*, 25(1):165–179, 2002.

[49] S. Joshi, B. Davis, M. Jomier, and G. Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *Neuroimage*, 23, 2004.

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

[51] D. G. Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, 16(2):81–121, 1984.

[52] S. Lang. *Differential and Riemannian Manifolds*. Springer, 1995.

[53] J. F. Le-Garrec, C. V. Ragni, S. Pop, A. Dufour, J. C. Olivio-Marin, M. E. Buckingham, and S. M. Meilhac. Quantitative analysis of polarity in 3d reveals local cell coordination in the embryonic mouse heart. *Development*, 140:395–404, Jan 2013.

[54] N. Lee and S. Mori. DTI analysis of pediatric populations using landmark-based approach. *MICCAI*, 2008.

[55] G. Leonardi and S. Masnou. Locality of the mean curvature of rectifiable varifolds. *Advances in Calculus of variations*, (17-42), 2009.

[56] H. Lombaert, J.-M. Peyrat, P. Croisille, S. Rapacchi, L. Fanton, F. Cheriet, P. Clarysse, I. Magnin, H. Delingette, and N. Ayache. Human atlas of the cardiac fiber architecture: Study on a healthy population. *IEEE Trans. on Medical Imaging*, 31(7):1436–1447, 2012.

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

[58] J. Ma, M. I. Miller, and L. Younes. A Bayesian Generative Model for Surface Template Estimation. *International Journal of Biomedical Imaging*, 2010.

[59] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on signal processing*, 41(3397-3415), 1993.

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

[61] F. Memoli. On the use of Gromov-Hausdorff Distances for Shape Comparison. *Symposium on Point Based Graphics*, pages 81–90, 2007.

[62] P. W. Michor and A. Kriegl. The Convenient Setting of Global Analysis. *Mathematical Surveys and Monographs*, 53, 1997.

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

[64] M. I. Miller and A. Qiu. The emerging discipline of computational functional anatomy. *NeuroImage*, 45:16–39, 2009.

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

[66] V. I. Morariu, B. V. Srinivasan, V. C. Raykar, R. Duraiswami, and L. S. Davis. Automatic online tuning for fast gaussian summation. In *Advances in Neural Information Processing Systems*, pages 1113–1120, 2008.

[67] F. Morgan. *Geometric measure theory, a beginner's guide.* Academic Press, 1995.

[68] P. Piccione and D. Tausk. *On the Geometry of Grassmannians and the Symplectic Group: the Maslov Index and its Applications.* UFF, 2000.

[69] A. Qiu, M. Vaillant, P. Barta, J. Ratnanather, and M. I. Miller. Region-of-interest-based analysis with application of cortical thickness variation of left planum temporale in schizophrenia and psychotic bipolar disorder. *Human Brain Mapping*, 29(8):973–985, 2008.

[70] G. D. Rham. *Variétés différentiables : formes, courants, formes harmoniques.* Hermann, 1955.

[71] J. Robbin and D. Salamon. *Introduction to differential topology.* UW Madison, 2011.

[72] M. Sabuncu and B. Singer. Function-based intersubject alignment of human cortical anatomy. *Cerebral Cortex*, pages 130–140, jan 2010.

[73] R. S. Sampath, H. Sundar, and S. K. Veerapaneni. Parallel fast gauss transform. In *Proceedings of the 2010 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 1–10. IEEE Computer Society, 2010.

[74] S. Sommer, F. Lauze, M. Nielsen, and X. Pennec. Sparse multi-scale diffeomorphic registration : The kernel bundle framework. *to appear in Journal of Mathematical Imaging and Vision*, 2013.

[75] P. M. Thompson, J. Moussai, S. Zohoori, A. Goldkorn, A. A. Khan, 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*, 8(6):492–509, 1998.

[76] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. *Proceedings of the 1998 IEEE International Conference on Computer Vision*, pages 839–846, 1998.

[77] A. Trouvé. Géométrie et espace de formes. *Masters degree course*, 2012.

[78] A. Trouvé and L. Younes. Metamorphoses through lie group action. *Foundation of computational mathematics*, 5:173–198, 2005.

[79] M. Vaillant, M. Miller, A. Trouvé, and L. Younes. Statistics of diffeomorphisms via tangent space representation. *Neuroimage*, 23(161-169), 2004.

[80] W. D'Arcy Thompson. *On Growth and Form.* 1917.

[81] L. Wang, F. Beg, T. Ratnanather, C. Ceritoglu, L. Younes, J. Morris, J. Csernansky, and M. Miller. Large Deformation Diffeomorphism and Momentum Based Hippocampal Shape Discrimination in Dementia of the Alzheimer type. *IEEE Transactions on Medical Imaging*, 26(4):462–470, 2007.

[82] L. Wolf and A. Shashua. Learning over sets using kernel principle angles. *Journal of Machine Learning*, 2003.

[83] Y.-C. Wong. Differential geometry of grassmann manifolds. 1967.

[84] C. Yang, R. Duraiswami, N. Gumerov, and L. Davis. Improved Fast Gauss Transform and efficient kernel density estimation. *Proceedings of the ninth IEEE International Conference on Computer Vision (ICCV)*, 0-7695-1950-4/03, 2003.

[85] B. Yeo, T. Vercauteren, P. Fillard, X. Pennec, P. Golland, N. Ayache, and O. Clatz. DTI Registration with Exact Finite-Strain Differential. *IEEE International Symposium on Biomedical Imaging*, pages 700–703, 2008.

[86] L. Younes. *Shapes and diffeomorphisms*. Springer, 2010.

[87] M. Young, E. Lebed, Y. Jian, P. J. Mackenzie, M. F. Beg, and M. V. Sarunic. Real-time high-speed volumetric imaging using compressive sampling optical coherence tomography. *Biomed. Opt. Express*, (2):2690–2697, 2011.