Geometric Statistics for Computational Anatomy

Nina Miolane

UNIVERSITÉ CÔTE D'AZUR

ECOLE DOCTORALE STIC

SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

THÈSE DE DOCTORAT

Présentée en vue de l'obtention du grade de

Docteur en Informatique

de l'Université Côte d'Azur

par

Nina MIOLANE

Statistiques géométriques pour l'anatomie numérique

Dirigée par: Xavier PENNEC

Préparée à INRIA Sophia Antipolis, ÉQUIPE ASCLEPIOS
en collaboration avec Susan HOLMES, Stanford University

Soutenue le 16 décembre 2016

Devant le jury composé de :

Nicholas AYACHE	-	Directeur de recherche, INRIA Asclepios, France. *Examinateur.*
Ian DRYDEN	-	Professeur, University of Nottingham, Grande Bretagne. *Rapporteur.*
Susan HOLMES	-	Professeur, Stanford University, États-Unis. *Co-superviseur.*
Stephan HUCKEMANN	-	Professeur, Göttingen Universität, Allemagne. *Examinateur.*
Sarang JOSHI	-	Professeur, University of Utah, États-Unis. *Rapporteur.*
Xavier PENNEC	-	Directeur de recherche, INRIA Asclepios, France. *Directeur de thèse.*
Alain TROUVÉ	-	Professeur, Ecole Normale Supérieure, France. *Rapporteur.*

UNIVERSITÉ CÔTE D'AZUR

DOCTORAL SCHOOL STIC

SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

P H D T H E S I S

to obtain the title of

PhD of Science

of Université Côte d'Azur

Specialty : Computer Science

Defended by

Nina Miolane

Geometric Statistics for Computational Anatomy

Thesis Advisor: Xavier Pennec

prepared at INRIA Sophia Antipolis, Asclepios Team
in collaboration with Susan Holmes, Stanford University

Defended on December 16th, 2016

Jury :

Nicholas Ayache	-	Research Director, INRIA Asclepios, France. *Examiner.*
Ian Dryden	-	Professor, University of Nottingham, United Kingdom. *Reviewer.*
Susan Holmes	-	Professor, Stanford University, United States. *Co-supervisor.*
Stephan Huckemann	-	Professor, Göttingen Universität, Germany. *Examiner.*
Sarang Joshi	-	Professor, University of Utah, United States. *Reviewer.*
Xavier Pennec	-	Senior Researcher, INRIA Asclepios, France. *Advisor.*
Alain Trouvé	-	Professor, Ecole Normale Supérieure, France. *Reviewer.*

To my parents Edith and Yves, and my brothers Léo and Tom...

Abstract: Geometric statistics for Computational Anatomy

This thesis develops Geometric Statistics for the quantitative analysis of organ shapes in Computational Anatomy. Statistics is a science that studies methods of inference from data which often belong to vector spaces, i.e. linear spaces. In contrast, *Geometric statistics* generalizes statistics for data belonging to manifolds, i.e. non-linear spaces. Non-linear data spaces emerge naturally in Computational Anatomy, where statistics aims to uncover the normal and pathological variability of the human anatomy. First, organ shapes - seen in medical images - can be modeled as deformations of the same template shape: they are elements of a Lie group of deformations, which is a manifold *with an additional group structure*. Second, organ shapes can be modeled as the equivalence classes of their 3D configurations under the action of a Lie group of transformations, e.g. of rotations and translations: they belong to a quotient space, which is a manifold *with an additional stratification*. Third, the medical images themselves can be represented as manifolds *with an additional structure of horizontal distribution*, when considering both the image's intensity and its gradient. Defining a metric or a distance to perform statistics on these manifolds might not be enough, as the metric or distance might not consistently account for the aforementioned additional structures. Theories of statistics on Riemannian manifolds and metric spaces need to be extended. This thesis advocates that Geometric statistics is a very natural extension that applies to manifolds *with additional geometric structures* in Computational Anatomy.

Our first research axis tackles *the definition of Geometric Statistics*. The model of organ shapes as deformations of a template leads to the challenging problem of defining statistics on Lie groups. We provide an algorithm that constructs a (pseudo-)Riemannian metric compatible with the group structure, i.e. bi-invariant, when it exists. We find that some Lie groups do not admit any such (pseudo-)metric. This result thus advocates for non-metric statistics on Lie groups. In our second axis, we *use Geometric Statistics for the analysis of existing algorithms in Computational Anatomy*. We reformulate the widespread algorithm of template shape computation with organ shapes modeled as elements of a quotient space. This geometric approach enables us to show its asymptotic bias. We illustrate our results on brain templates and suggest an improved algorithm. We then study algorithms that have a pre-processing registration step before a statistical analysis. We show again their bias and provide a geometric correction, by taking into account the quotient space stratification in addition to the Riemannian framework. Our third axis focuses on *the applications of Geometric Statistics to medical image processing*, like in-painting or anisotropic smoothing. Considering the image intensity with its gradient leads to statistics on manifolds with a horizontal distribution. We provide the mathematical basis to extend sub-Riemannian structures, currently used for 2D processing, to our 3D images. In all these aspects, this work develops Geometric Statistics well beyond Riemannian geometry.

Résumé : Les statistiques géométriques pour l'anatomie numérique

Cette thèse développe les statistiques géométriques pour l'analyse quantitative des formes d'organes en anatomie numérique. Les statistiques étudient les méthodes d'inférence à partir de données, qui appartiennent le plus souvent à des espaces vectoriels, i.e. des espaces linéaires. Les statistiques géométriques, elles, généralisent les statistiques pour des données appartenant à des variétés différentielles, i.e. des espaces non-linéaires. Les espaces de données non-linéaires apparaissent naturellement en anatomie numérique, où l'on cherche à dévoiler la variabilité normale et pathologique de l'anatomie humaine. Premièrement, les formes d'un organe - vues dans les images médicales - peuvent être représentées par des déformations du même organe de référence: donc par des éléments du groupe de Lie des déformations, qui est une variété différentielle avec une structure additionnelle de groupe. Deuxièmement, les formes d'un organe peuvent être représentées comme les classes d'équivalence de leur configuration tridimensionnelle sous l'action d'un groupe de Lie de transformations, par exemple de rotations et de translations: elles appartiennent donc à un espace quotient, qui est une variété différentielle avec une structure additionnelle de stratification. Troisièmement, les images médicales elles-mêmes peuvent être représentées comme des variétés différentielles avec une structure additionnelle de distribution horizontale, lorsque l'on considère à la fois l'intensité de l'image et son gradient. Définir une métrique ou une distance pour faire des Statistiques sur ces variétés différentielles n'est probablement pas suffisant : la métrique ou distance ne tient pas toujours compte des structures additionnelles mentionnées ci-dessus. Les théories de statistiques sur les variétés différentielles Riemanniennes et les espaces métriques doivent donc être généralisées. Cette thèse défend les statistiques géométriques comme une extension très naturelle qui s'applique aux variétés différentielles avec des structures géométriques additionnelles en anatomie numérique.

Premièrement, nous définissons les statistiques géométriques sur les groupes de Lie. Nous proposons une construction algorithmique de (pseudo-)métrique Riemannienne, compatible avec la structure de groupe, lorsqu'elle existe. Nous trouvons que certains groupes n'admettent pas de telle (pseudo-)métrique et défendons l'idée de statistiques non-métriques sur les groupes de Lie. Ensuite, nous utilisons les statistiques géométriques pour analyser l'algorithme de calcul d'organe de référence, reformulé avec des espaces quotient. Nous montrons son biais et suggérons un algorithme amélioré. Enfin, nous appliquons les statistiques géométriques au traitement d'images, en généralisant les structures sous-Riemanniennes, utilisées en 2D, au 3D. A travers tous ces aspects, notre travail développe donc les Statistiques Géométriques bien au delà de la Géométrie Riemannienne.

Acknowledgments

First of all, I am extremely grateful to my two supervisors who have guided me for the three years of this Ph.D. On the one hand, Xavier Pennec introduced me to the Medical Imaging community. He showed me how mathematical tools - that I had learned in Theoretical Physics - can be used to address cutting-edge problems in this exciting field. On the other hand, Susan Holmes made me discover the field of Statistics and showed me how to use our mathematical models to effectively answer practical useful questions in Computational Anatomy. I would like also to thank both INRIA, the Asclepios Team, and Susan Holmes' lab for offering me outstanding working conditions for these years. I thank their respective leaders, Nicholas Ayache and Susan Holmes.

I warmly thank Ian Dryden, Sarang Joshi and Alain Trouvé for the precious time they dedicated to review the manuscript. Their encouraging compliments are invaluable. I am also indebted to Nicholas Ayache and Stephan Huckemann who kindly accepted to join the defense committee in Nice. It is a great honor for me to have such an outstanding jury.

I am very thankful to all the people who went along with me during these years. In particular, I would like to thank the lab-mates at INRIA: Marmichouchou, Thomas, Sophie, Rocio, Krissy, Jan, Bishesh, Vikash, Flo, Florent, Mike, Loïc C., Hakim, Loïc L., Loïc D., Rochfifi, Rafifou, Shuman, Qiao, Nicolas C., Nicolas D., Chloé, Milkymat, Pawel, Aziz, Pauline, Mehdi and Clair for all the good times spent together running, partying, and also a bit working. I have a particular thank-you to Isabelle Strobant, for her patience and kind help when organizing my different - and often complicated - trips. I am grateful to the lab-mates at Stanford: Claire, Mika, Lucas, Kris, Julia, Christof, Lan and Ben. Then, I want to thank my house-mates at Valbonne: Zoom, Liony, Francis, Thomus, Clemencebenardcinema, Manon, Nico, Marmar, Charliiine and Schmitt for making my daily life the funniest Epic Sax Dream. I am also indebted to my house-mates at Palo Alto who I cannot all cite here, but a special shout-out to: Louis, Matteo, Damien, Jan, Montana, Mario, Megan, Will, Neekan, J-Lo, Marc and Angie. I am then very thankful to my friends in Paris, who supported me from the capital: Solène, Alexandra, Cécile, Sébastien, Tristan, PA, Barbu, Antoine, Guismo, Valentin, Gustaw, Sarah, Vincent, Navid, Margot and Béné. I also have special thoughts for Andrea and italian friends. Lastly, I would like to thank my parents Edith and Yves, and my brothers Léo and Tom.

Un grand merci à tous !

Contents

Chapter 1

Introduction

Fundamental research in anatomy seeks to understand both healthy and pathological structure and functioning of our organs, in order to help the diagnosis and treatment of diseases. The better our anatomical knowledge, the better our medicine. This introduction presents an historical overview of research in anatomy, and shows how successive methods or technologies used for anatomical investigations - dissections, medical imaging, computer science - have enabled progress in medicine. We then turn to Computational Anatomy - the main purpose of this thesis - and describe the geometric nature of the data spaces involved in this field. We argue for the need of a statistical theory beyond statistics on metric spaces and Riemannian manifolds, in order to handle them. Ultimately, we present the overview of our manuscript.

1.1 A brief history of medicine and anatomy

Medicine is the science and practice of the diagnosis, treatment, healing, and prevention of diseases. In the ancient world already, the Greek mythology had its own god of medicine: Asclepius, represented on Figure 1.1(left). His rod, a snake-entwined staff, was supposed to heal any disease and remains a symbol of medicine in today's western civilization, see Figure 1.1(middle).

Figure 1.1: Left: Statue of Asclepius in the Epidaure's sanctuary. Middle: The flag of the World Health Organization with Asclepius rod. Right: Portrait of Andreas Vesalius (1514-1564) from De humani corporis fabrica.

1.1.1 Before medical imaging

Medicine and Anatomy in the Antiquity The Greek physician Hippocrates (c. 460 - c. 370 BCE) is often considered the "father of Western medicine", because he was the first to argue that diseases were not caused by any action of the gods but rather by bad environmental factors, diet, and living habits:

> *It is thus with regard divine nor more sacred than other diseases, but has a natural cause from the originates like other affections. Men regard its nature and cause as divine from ignorance and wonder...*" - Hippocrates, On the Sacred Disease.

His argument implied that humans could understand and act on their own health. Hippocrates thus created Medicine as a science.

However, the Greeks knew almost nothing about human anatomy because of a taboo forbidding human dissections. Their anatomical knowledge essentially came from the work of Galen (129 AD - c. 200/ c. 216), who postulated theories by dissecting monkeys and pigs, but never human bodies. Some of these theories are now known to be incorrect, like Humorism, which is the belief that the balance four bodily fluids , known as humors, influences one's health. Even if incorrect, they remained uncontested for almost 1,300 years, until the Renaissance.

Medicine and Anatomy in the Medieval Age and Renaissance Anatomical knowledge was recognized as pivotal for medicine only during the Renaissance. Anatomy experienced a revolution led by Andreas Vesalius, see Figure 1.1 (right), who advocated that direct observation through human dissections was the only reliable resource, in a huge break with ancient taboos. He published a seven-volume De humani corporis fabrica (On the Fabric of the Human Body) showing printed descriptions and illustrations of human dissections.

Moreover, Vesalius defended the idea that humans were sharing a *common inner structure*, like a prototype anatomy. His illustrations of the human anatomy were meant to show this prototype, by a *mental averaging of the anatomies* he had seen during his dissections. Such a prototypical anatomy is still a central aspect of anatomy today, and especially in Computational Anatomy where it is called a *template anatomy*.

Human dissection can be seen as the first revolution in the history of anatomical research. It represented a fundamentally new tool for understanding the human body's interior. Medical imaging can be seen as the second revolution.

1.1.2 Relying on medical imaging

Medical imaging and computer science Medical imaging was invented in the end of the 19th century, with the discovery of X-rays by Rontgen [Roentgen 1896], Physics Nobel Prize in 1901. Today, medical imaging equipments are omnipresent in hospitals around the world, see Figure 1.2. Their number even increases: the number of IRMs in French hospitals has increased of 43% since 2010 and the number

of PET scans has increased by 49,4% in four years from 2010 to 2014, see Figure 1.3 [Chasseing 2016].

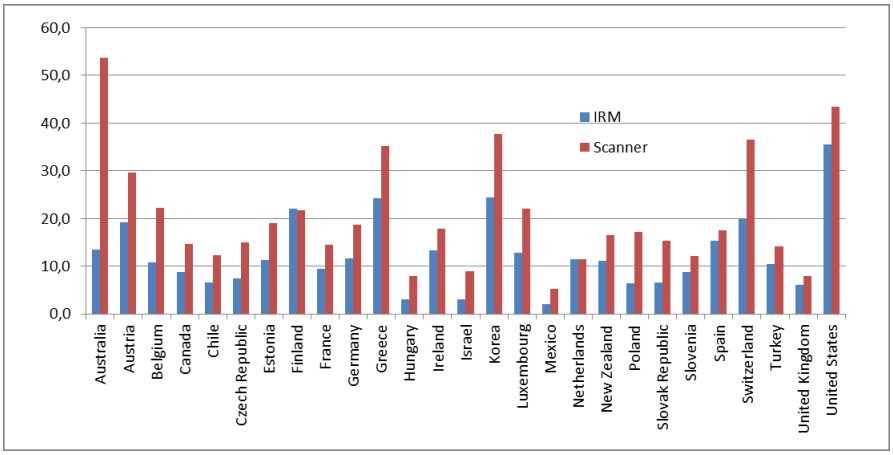


Figure 1.2: Number of IRMs and scanners (equipment) per million of inhabitants worldwide for the year 2013 [Chasseing 2016].

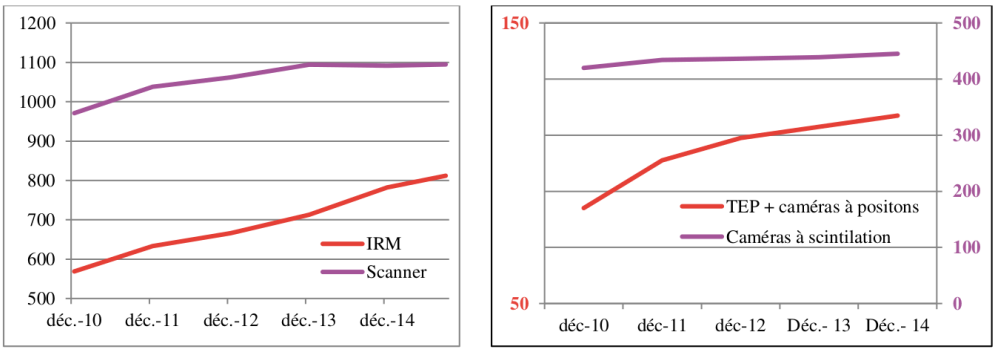


Figure 1.3: Evolution of the number of medical imaging equipment in France [Chasseing 2016].

These devices produce medical images in several modalities: X-ray radiography or X-ray CT scans, magnetic resonance imaging (MRI), medical ultrasonography or ultrasound or echography, endoscopy, elastography, tactile imaging, thermography, medical photography and nuclear medicine functional imaging techniques such as positron emission tomography (PET) and Single-photon emission computed tomography (SPECT). Medicine heavily relies on these images, as shown on Figure 1.4: for the countries members of the OECD (Organization for Economic Cooperation and Development) for instance, the number of IRMs exams is 52 for 1000 inhabitants on average and the number of scan exams is 120 per 1000 inhabitants on average in 2013.

Medical imaging noninvasively produce databases of thousands of tridimensional visual representations of the human organs, bones and tissues (anatomy) as well as

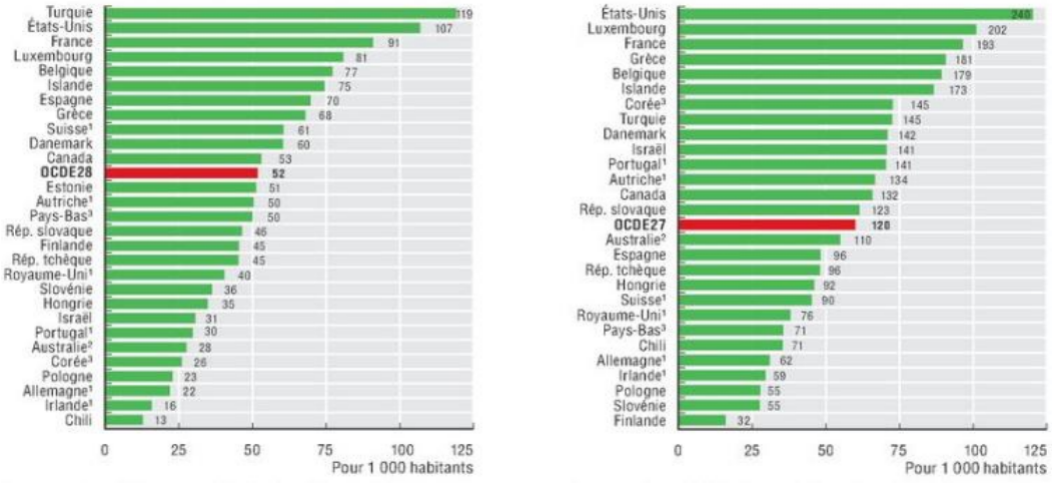


Figure 1.4: Number of IRMs (left) and scanners (right) exams in 2013 with respect to the population [Chasseing 2016].

visual representations of the function of some organs or tissues (physiology). For example, the UK Biobank [Sudlow 2015] collects medical images of multiple organs at a population level, together with information such as lifestyle, demographic and genomic data. There are currently around 10,000 participants in the study. Each image data of such database already contains an abundance of information, which is illustrated in the size it takes in a computer's memory. For example, tridimensional X-ray CT videos at a rate of 2-3 images per second for 40 seconds already represent 200 GB: only 5 such sequences can be stored on a 1 terabyte hard disk.

Clinicians often rely on these images to decide on diagnosis, treatment and medical intervention. But their eyes and memory may not be enough to fully analyze the complexity of information shown in the images. For example the intensity in X-Ray images contains quantified information, the amount of radiation absorbed by a given tissue, which the eye cannot quantitatively processed.

There is a need of automatic tools designed to assist the clinicians in the analysis of medical images. The invention of the computer in the beginning of the 20th century and the explosion of computer science thereafter provides the technology to do so. For example in the US, the resources allocated specifically for medical image computing are growing exponentially: as an example, the cloud computing market allowed specifically for medical images computing shows a growth of 27% in less than 10 years, see Figure 1.5. Today's increasing availability of high-performance computing - for example GPU (Graphics Processing Unit) computing - makes it possible to extract clinically relevant information from the medical images.

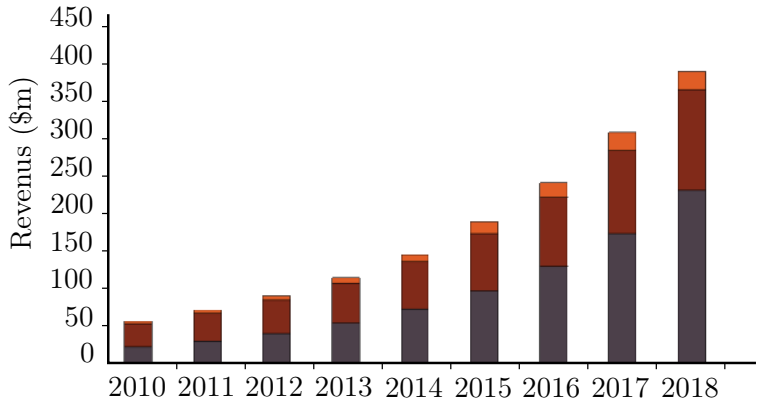


Figure 1.5: Market of Cloud Computing for medical imaging. Grey: Cloud based storage and Disaster Recovery Market. Red: Cloud based Sharing and Communication Software market. Orange: Picture Archiving and Communication System on cloud [GlobalData 2011].

1.2 Why Geometric Statistics for Computational Anatomy

1.2.1 What is Computational Anatomy?

Computational Anatomy aims to uncover the normal and pathological variability of the human anatomy, through the processing and analysis organ shapes shown on medical images [Miller 1997, Grenander 1993]. The field is multidisciplinary as it ranges from mathematics to statistics, to computer science, to anatomy and medicine. In particular, Computational Anatomy considers questions such as: how different is the shape of the brain for patients with a risk for depression? or: can schizophrenia affect the visual aspect of the cerebral cortex?

Computational Anatomy is one pillar of Computational Medicine, together with Computational Physiology, Computational Molecular Medicine and Computational Healthcare. As such, its direct application is the development of tools helping the daily medical practice. In fact, organ shapes often represent a statistically significant predictors for various clinical parameters: the brain shape is corelated to neurodegenerative diseases like Alzheimer [Gerber 2010, Chou 2008, Baron 2001, Lorenzi 2011, Apostolova 2007] but also to other neurological and neuropsychiatric illnesses, including epilepsy [Eriksson 2001] and schizophrenia [Kubicki 2007, Brignell 2010] for example.

Therefore, a computational representation of the anatomy can serve the diagnosis of diseases from medical images. We may characterize the brain morphology associated with a particular pathology such as Alzheimer's disease and learn how to discriminate between groups of healthy and diseased subjects [Lorenzi 2011]. Furthermore, Computational Anatomy contributes to the transition to a personalized medicine: for example, we can model the evolution of a brain shape through time

[Lorenzi 2010, Singh 2013], and especially the evolution of a diseases like Alzheimer's for example [Hadj-Hamou 2016, Khanal 2016].

1.2.2 Modeling the data in Computational Anatomy

Computational Anatomy relies on the statistical analysis of organ shapes shown in medical images. Organ shapes are complicated data which are usually modeled as elements of non-linear data spaces. More precisely, these data spaces are *manifolds with additional geometric properties* which we introduce here.

Manifolds with Lie group actions We first consider the raw data, which are the medical images. They are intensity functions I over a tridimensional domain and as such: elements of a vector space, which is a linear manifold. The vector space is equipped with an additional structure of Lie group action as soon as we consider registration of images. The registration of two images is the problem of finding correspondences between the anatomies they show, i.e. to look for an optimal way of geometrically deforming one image into the other [Brown 1992]. The geometric deformation belongs to a Lie groups and is said to act on the image.

We can also put landmarks on the organ shown in the medical image [Darmanté 2014], or consider the organ's surface, eventually represented with the framework of currents [Durrleman 2010]. The space of landmarks or the space of currents are vector spaces too. Likewise, geometric deformations can be used to register one set of landmarks to another [Joshi 2000], or one surface to another [Vaillant 2005]. In all these cases, the anatomical data are modeled as elements of a manifold with a Lie group action.

Lie groups We can deform one anatomy onto another through the process of registration. If we fix one anatomical image as a reference - called the template - we can represent all other images as deformations of this template [Amit 1991], i.e. as an element of a Lie group [Thompson 1992]. The same works with sets of landmarks or surfaces. In a longitudinal study, the template may be the first image of the time line and all other images are represented as elements of a Lie group.

The Lie groups may be finite-dimensional, like rigid or affine transformations [Ashburner 2003, Boisvert 2008] or infinite-dimensional, like Lie groups of diffeomorphisms [Trouvé 1998, Dupuis 1998, Beg 2005].

Quotient spaces We consider the raw data: the medical image, or alternatively sets of landmarks or surfaces. Some of the Lie group acting on these data represent information that have no impact for the analysis of the organ's shapes. For example rotations or translations of the anatomical shape may only denote the fact that patients were imaged in different positions. In fact, the shape of a set of landmarks, the shape of a surface, or the shapes in an image can be defined as the remainder after we have filtered out the position and the orientation of the object [Kendall 1984,

Small 1989, Bookstein 1997, Dryden 1998]. The resulting data space of (anatomical) shapes is thus a quotient space, called the *shape space* [Huckemann 2010].

An important literature has been developed for shape spaces of landmarks, especially in 2D and 3D [I. L. Dryden 1991, Dryden 1992, Goodall 1993, Le 1993]. In this case, the landmarks configurations can be represented as matrices and linear algebra operations are performed to handle these data: this line of work is known as Procrustean analysis [Goodall 1991, Dryden 1998, Gower 2004].

Horizontal distributions Lastly, we move away from models of shapes applied to the *analysis* of medical images: we consider models involved in the *processing* of images. We consider one image, and the data are now the voxels. By default, the voxels belong to the tridimensional domain of the image, which is included in the vector space $\mathbb{R}^3$: the voxels are elements of a linear space. Recent studies offer to embed this space into a larger space which is a manifold [Citti 2006, Petitot 2013]. Instead of considering only the position of the voxel in $\mathbb{R}^3$, we also consider the orientation of the intensity gradient at this position. The domain of the image is now a lifted space of positions and directions, i.e. it has a structure of horizontal distribution. Considering this extended domain may lead to powerful image processing, like in-painting [Boscain 2014] or anisotropic smoothing [Duits 2010a].

1.2.3 Analyzing the data in Computational Anatomy

Data in Computational Anatomy are often modeled as elements of manifolds. Efforts have thus being engaged to generalize usual statistics on linear spaces to a theory of statistics on manifolds [Bhattacharya 2012, Patrangenaru 2015].

Theories of statistics on manifolds A usual procedure - referred to as *extrinsic* - is to embed the manifold in a large enough vector space of parameters. Then, the Euclidean distance defined on the embedding space induces a distance on the embedded manifold which can be used in evaluating estimators on a manifold [Hendriks 1991]. This approach is often chosen in Directional Statistics, for example statistics on rotations in $\mathbb{R}^n$ [Jupp 1989, Mardia 1999], where the manifolds considered here are spheres and projective spaces which are easily embedded.

In contrast, the intrinsic approach does not rely on an embedding. An intrinsic distance may be determined by the experimental setting and defines a metric structure on the manifold [Fréchet 1948]. One can generalize definitions from statistics on Euclidean space to statistics on metric spaces, by replacing the Euclidean distance by the intrinsic distance: for example, the mean of a dataset may be defined as the global minimum of the sum of the squared intrinsic distances to the data, a definition known as the Fréchet mean [Fréchet 1948]. Several means may be defined as the local minima of the same sum, a definition known as the Karcher means [Karcher 1977]. Statistical asymptotic properties of such means have been studied in [Bhattacharya 2003, Bhattacharya 2005]. Statistics on metric spaces have been employed successfully, for example in Phylogenetics and its tree spaces [Billera 2001].

The intrinsic distance may also be induced by a stronger structure on the manifold: a Riemannian structure, where one has an inner product defined at each tangent space of the manifold [Postnikov 2001, Pennec 2006]. First, one may use the definitions of the statistical theory on the induced metric spaces. But one may also rely on the additional structures provided by the Riemannian framework. One may use the Levi-Civita connection associated to the Riemannian metric and define the mean as an Exponential Barycenter [Émery 1991, Corcuera 1998]. The mean can be seen as the 0-dimensional space which best represents the data set. One may consider higher dimensional subspaces that describe the dataset, and perform Principal Geodesic Analysis proposed in [Fletcher 2004] and implemented in [Sommer 2014], Geodesic Principal Component Analysis [Huckemann 2010] or Barycentric Subspaces [Pennec 2015]. Regression techniques may also be generalized to Riemannian manifolds, for example on manifolds defining brain shapes [Davis 2010].

Towards Geometric Statistics Shall we use the theories of statistics on metric spaces or on Riemannian manifolds for the data spaces of Computational Anatomy detailed in the previous subsection? These data spaces possess additional structures, like the action of a Lie group, a Lie group structure, an horizontal distribution. We shall acknowledge this additional structures to build, and use, a consistent new statistical theory.

We take one example to illustrate the need of enforcing the compatibility between the statistical theory and the structures on the manifolds. We consider the definition of the mean on Lie groups. We could construct a distance or a Riemannian metric on the group and apply statistics on metric spaces or Riemannian manifolds, using for example the Fréchet mean. To respect the Lie group structure, it would be judicious to have a notion of mean that is stable by the group composition, by the left and by the right: so that the mean of a translated dataset is the translation of the original dataset's mean. This property is however not enforced in general [Pennec 2012]. The Ph.D thesis describes these situations where the consistency between statistics and geometry needs to be studied.

1.2.4 What is Geometric Statistics?

Geometric Statistics is a theory of statistics on manifolds, that generalizes statistics on metric spaces and statistics on Riemannian spaces to take into account other additional geometric structures on manifolds. "Geometric Statistics" is an appellation towards which the community seems to converge to design these research efforts. This Ph.D thesis will present Geometric Statistics for Computational Anatomy. So let first present other theories that merge Statistics and Differential Geometry but still are different from Geometric Statistics, i.e. let first present what Geometric Statistics is not.

Geometric Statistics is different from *Information geometry* [Amari 1990], which applies techniques of differential geometry to probability theory. A probability distri-

bution is modeled as a point of a Riemannian manifold, called a *statistical manifold*. In contrast in Geometric Statistics, Differential Geometry is not used to represent a space of probability distributions of some data but rather to represent the space of the data themselves.

Geometric Statistics is also different from *Manifold Learning* [Lee 2007], which is a subfield of Machine Learning focusing on non-linear dimensionality reduction. Manifold Learning algorithms are based on the idea that the dimensionality of many data sets is only artificially high. They seek to learn lower dimensional manifolds of the high dimensional vector space of the data. The Differential Geometry involved is thus embedded in the Euclidean geometry. In contrast, Geometric Statistics model the data space as non-Euclidean from the very start.

Geometric Statistics is different from *Stochastic Geometry*, which studies random geometric patterns and especially the study of random point patterns [Matheron 1975]. And Geometric statistics is different from *Geometric probability* which studies probability functions defined on convex, or one studies the densities of lines in 2D or 3D that verify certain conditions [Santalo 2004]. In fact, Geometric Statistics may encompass Stochastic Geometry and Geometric Probability, by building a statistical framework on the manifold of point patterns or of lines.

1.3 Manuscript overview

This Ph.D thesis advocates for Geometric Statistics in Computational Anatomy. Geometric Statistics naturally emerges in the various subfields of Computational Anatomy and has three objectives. First, it defines a statistical framework for creating new algorithms analyzing anatomical shapes. Second, it provides a formal setting to analyze the statistical properties of computational methods already employed for many years in the field, but whose statistical consistency could not be evaluated. Third, it provides new tools for (medical) image processing.

This manuscript is divided into two parts. In the first part, we introduce elements from Statistics - especially Estimation theory - and Geometry independently - with the mathematical definitions of the geometric nature of our data spaces. In the second part, we show the applications of Geometric Statistics on different essential problems of Computational Anatomy varying both the geometric structures and the statistical reasoning to give a wide view of the field.

Our contributions are detailed in six core Chapters, which essentially correspond to the journal articles we have published so far, except for the last Chapter, which contains promising preliminary results that have been presented at a workshop of MICCAI'2015.

1.3.1 Part I: Foundations of Geometric Statistics

- **Chapter 2** introduces Estimation theory, as it is well-known in the context of linear spaces. We offer a condensed illustration of the statistical problems in estimation procedures, exemplified with the computation of the sample mean.

Computing the sample mean, in different geometric spaces, will indeed be one leading interest in this thesis.

- **Chapter 3** presents the mathematical framework for the geometric spaces that will be studied in the thesis: Riemannian (pseudo-)metrics, affine connection spaces, Lie groups, Lie group actions, quotient spaces, fiber bundles, etc. We also emphasize the links and interactions between these structures.

1.3.2 Part II: Geometric Statistics applied to Computational Anatomy

- **Chapter 4** turns to Lie groups and the definition of a consistent statistical framework on these. We investigate the relationship between the Riemannian and pseudo-Riemannian framework with the group structure. We derive an algorithm for deciding if one can construct a (pseudo-)Riemannian metric that is compatible (i.e. "bi-invariant") with the Lie group given as input. The algorithmic procedure gives such a metric as output, in case of existence. Our procedure shows that a bi-invariant (pseudo-) metric fails to exists on general Lie groups. As a consequence, the definition of a consistent mean - and more generally of statistics- on Lie groups cannot rely on the (pseudo-)Riemannian framework, that is so widespread in our community. We advocate for other methods to define consistent statistics on Lie groups, like for example affine connection spaces.

- **Chapter 5** considers Lie group actions and quotient spaces. We introduce a new approach to analyze the statistical properties of well-known algorithms of Computational anatomy, like the algorithm of template anatomy estimation. The approach relies on Geometric Statistics, as we reformulate the algorithm through the lenses of statistics on quotient spaces. The algorithm then boils down to the computation of the mean, on a stratified space which is a quotient space. Geometric Statistics thus enable us to demonstrate its asymptotic bias: even with an infinite number of data, the algorithm does not converge to the anatomy it is designed to estimate. Our result is the first proof of this effect that had been already observed experimentally. Our study shows that Geometric Statistics not only provides an elegant framework to analyze algorithms, it is also needed to avoid statistical caveats. The intrinsic geometry of a computational procedure in the data space needs to be taken into account, as it can create unexpected pathological behaviors in its daily use.

- **Chapter 6** exemplifies the previous results for neuroimaging studies, and more precisely for the computation of a brain template from MRIs. Computing the brain template is the first step in almost all neuroimaging pipelines. We illustrate the consequences of the results of Chapter 5 and 6 by quantifying spatially the systematic bias on the brain template. We also suggest a new algorithm that aims to control the bias by a topological method.

- **Chapter 7** applies tools of Geometric Statistics to other statistical procedures, beyond the computation of the template anatomical shape of Chapter 5. We consider unsupervised learning algorithms on shapes that are performed after a pre-processing step of data (images, meshes...) registration. We show that the registration step leads to bias in these algorithms, just as it led to bias for the computation of the template anatomy. We suggest a geometric correction of this bias, taking into account the stratified geometry of the data space. Again, one needs to take into account the additional structure of the manifold, in order to understand the algorithms and to adapt them when they fail.

- **Chapter 8** considers statistics on sub-Riemannian manifolds and their use for medical image processing. The sub-Riemannian framework enables to consider both the intensity on the image and the direction of the intensity gradient. This is very useful for anisotropic smoothing or in-painting techniques. Sub-Riemannian geometry has been successfully used in 2D computer vision. Our contribution is to give the mathematical basis for using the structures for 3D images, which is absolutely not straightforward as new geometric structures appear while going from the 2D to the 3D. We provide a mathematical toolbox that make possible the use of sub-Riemannian geometry for 3D medical images processing.

1.3.3 Publications

The publications and awards of the author are listed below.

Peer-reviewed journal papers

1. Miolane, N., Holmes, S., Pennec, X.: Topologically constrained template estimation via Morse-Smale complexes allows to control its statistical consistency. SIAM Journal on Applied Algebra and Geometry (2016). *To be submitted.* This paper presents the spatial quantification of the bias in the estimation of the template anatomy. This paper is presented in Chapter 6.

2. Miolane, N., Holmes, S., Pennec, X.: Template organ shape estimation in Computational Anatomy: Correcting an asymptotic bias. SIAM Journal of Imaging Sciences (2016). *Submitted.* This paper shows a novel approach using Geometric Statistics to study the properties of the template anatomy algorithm in Computational Anatomy. It completes conference paper "Biased estimators on quotient spaces" and is presented in Chapter 5.

3. Miolane, N., Pennec, X.: Computing bi-invariant pseudo-metrics on Lie groups for consistent statistics. Entropy 17(4), 1850–1881 (2015). `https://hal.inria.fr/hal-01133922`. This paper presents an algorithmic procedure to compute a bi-invariant pseudo-metric on a Lie group, when it exists. Such a pseudo-metric enables to apply Riemannian statistics on Lie group while respecting the group structure. It completes conference paper "Statistics on

Lie groups : a need to go beyond the pseudo-Riemannian framework" and is presented in Chapter 4.

4. Darmanté, H., Bugnas, B., Dompsure, R.B.D., Barresi, L., Miolane, N., Pennec, X., de Peretti, F., Bronsard, N.: Analyse biométrique de l'anneau pelvien en 3 dimensions – à propos de 100 scanners. Revue de Chirurgie Orthopédique et Traumatologique 100 (7, Supplement), S241 – (2014). http://www.sciencedirect.com/science/article/pii/S187705171400327X. This paper shows the applications of statistics on pelvis shapes in order to develop a new pelvis surgery. This paper is not presented in this thesis.

Peer-reviewed conference papers

1. Miolane, N., Pennec, X.: A survey of mathematical structures for extending 2D neurogeometry to 3D image processing. 18th International Conference on Medical Image Computing and Computer Assisted Intervention. Medical Computer Vision: Algorithms for Big Data. Volume 9601 of the series Lecture Notes in Computer Science, pp 155-167. (2015). https://hal.inria.fr/hal-01203518/document. This paper presents mathematical structures to perform new algorithms of 3D image processing. The paper is presented in Chapter 8.

2. Miolane, N., Pennec, X.: Biased estimators on quotient spaces. 2nd International Conference on Geometric Sciences of Information. Springer, 9389, pp.130-139, Lecture notes in computer science (LNCS). (2015). https://hal.inria.fr/hal-01203805/document. **(Oral presentation)**. This paper shows first computations of Differential Geometry to show the bias of template shape estimation. It has led to the submitted journal paper presented in Chapter 5.

3. Miolane, N., Pennec, X.: Statistics on Lie groups : a need to go beyond the pseudo-Riemannian framework. 34th International workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering. AIP Proceedings, 1641, pp.59–66. (2014). https://hal.inria.fr/hal-01091515. **(Oral presentation)**. This paper advocates for the need to go beyond the pseudo-Riemannian framework for defining consistent statistics on Lie groups. It has led to the journal paper presented in Chapter 4.

Invited talks

- Miolane, N.: Template shape estimation in Computational Anatomy. Stanford Statistics Seminar – June 7th, 2016. See: https://statistics.stanford.edu/events/statistics-seminar/all?field_event_academic_year_tid_1=601.

- Miolane, N.: Statistical properties of the Fréchet mean in quotient spaces. Applications to Computational Anatomy. Statistics Seminar of Université de Montpellier, France – November 2nd, 2015. See: http://www.i3m.univ-montp2.fr/themes-de-recherche/gta/index.php?option=com_content&view=article&id=59&catid=19&seance=3582.

- Miolane, N.: Noise Effects on Quotient Spaces M/G. Workshop on Infinite-Dimensional Riemannian Geometry with Applications to Image Matching and Shape Analysis. Erwin Schröedinger International Institute for Mathematics and Physics, Vienna – February 19, 2015. See: http://www.esi.ac.at/activities/events/2015/files/bmb_ws2_schedule.

Miscellaneous

- **Poster**: Miolane, N., Pennec, X., Holmes, S.: Towards a unified geometric Bayesian framework for template estimation in Computational Anatomy. International Society for Bayesian Analysis World Meeting. (2016). **(Young Researcher Travel Grant).**

- **Poster**: Miolane, N.: Statistics on Lie groups : can we obtain a consistent framework with pseudo-Riemannian metrics? Workshop on Geometrical Models in Vision, Institut Henri Poincaré, Paris (2014). http://gmvision.lsis.org/slides/miolane.pdf

- **Video**: Miolane, N., Khanal, B.: Statistics on Lie groups for Computational Anatomy. Educational challenge. 17th International Conference on Medical Image Computing and Computer Assisted Intervention, MIT Boston (2014). http://www.miccai.org/edu/index.html. **(First Popular Prize).**

1.3.4 Awards and fellowships

Awards

- 11th/1052, L'Oréal-Unesco fellowship "For Women in Science"– 2016

- Young Researcher Travel Award at International Society for Bayesian Analysis World Meeting, Sardegna, Italy – 2016

- 1st prize, Educational Challenge at the International Conference of Medical Image Computing and Computer Assisted Intervention (MIT, Boston, USA) – 2014. See: http://www.miccai.org/edu/videos.html#mec2014winners.

Fellowships

- Inria@SiliconValley postdoctoral fellowship – 2016 . See: https://project.inria.fr/siliconvalley/2016/09/13/inriasiliconvalley-post-doc-2016/.

- Applied Sciences Scholarship, France-Stanford Center for Interdisciplinary Studies – 2014. See: http://francestanford.stanford.edu/fellowships/fellowships_recipients/2014_2015.
- Inria-Cordi Fellowship for PhD student – 2013

Part I

Foundations of Geometric Statistics

Chapter 2

Elements of Statistics: estimation theory

This Chapter gives an overview of estimation theory in vector spaces. The notions introduced here represent the basics of Statistics used in this Ph.D thesis. They will be merged with the geometry of Chapter 3 in the remaining of the thesis, to define Geometric Statistics.

2.1 Introduction

The science of statistics is defined as the collection of data, their analysis and interpretation.

Historical landmarks of statistics The work of Al-Kindi in the 9th century is one of the first occurrence of statistics in the literature. One can read in his book: *"One way to solve an encrypted message, if we know its language, is to find a different plain text of the same language [...] and then we count the occurrences of each letter. We call the most frequently occurring letter the 'first', the next most occurring letter the 'second', the following most occurring the 'third', and so on [...]. Then we look at the cipher text we want to solve and we also classify its symbols. We find the most occurring symbol and change it to the form of the 'first' letter of the plain text sample, the next most common symbol is changed to the form of the 'second' letter, and so on..."*. In other words, Al-Kindi presented the first code breaking algorithm based on frequency analysis.

Nevertheless, statistics in their modern form - and in particular Estimation theory - are said to start in the 17th century. The mathematical foundations of Probability theory needed for its development were laid by Cardano, Pascal and Fermat. In the 18th century, Gauss and Legendre developed simultaneously the least squares method [Stigler 1981], today's widespread procedure for fitting an equation or a curve to data. These theories deal with data and parameters that are elements of Euclidean spaces. We can also cite Galton, Pearson, Gosset and Fisher among other pioneers of the field. They all contributed to make statistics a rigorous mathematical discipline used for analysis across many fields and not only in science. In the late 20th century and 21st century, statistical methods extensively use modern computers that have made practicable new methods.

Estimation theory Within statistics, Estimation theory gives a mathematical framework to answer questions like: what is the typical blood pressure of an individual in the population? how many individuals will develop this pathology? what is the healthy anatomy of the human brain? The goal of Estimation theory in Computational Medicine is to extract information from the data, in order to answer these questions.

Exact answers, of course, are almost impossible to obtain. But approximate answers may be accessible and Estimation theory provides a set of rules for obtaining them. Above all, the theory makes precise such assertions as "this approximated answer is likely to be close to the exact answer" or "this answer is better than that one". Asserting the reliability of an answer is crucial with respect to the applications of Computational Medicine, that is: patients' health.

As an introduction to some problems involved in Estimation theory, we present here a toy example. We will use this example throughout this chapter. We consider estimating the human typical blood pressure, which we call μ. We have at our disposal a collection of blood pressures x_i's among n subjects. An intuitive estimate $\hat{\mu}$ of μ is to compute the average - i.e. the *sample mean* - of the n observed blood pressures: $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Why is it reasonable to consider the sample mean $\hat{\mu}$ as an estimate of the human blood pressure μ? And how well does $\hat{\mu}$ evaluate the real typical blood pressure μ, how does it depend on the number of observed data n? Estimation theory brings quantified answers to these interrogations.

Outline of this Chapter This Chapter introduces elements of Estimation theory needed in this Ph.D thesis. Section 1 presents definitions on estimators and measures of their accuracy. Section 2 shows how to compute an estimator's measures of accuracy in practice. Section 3 provides methods to improve the accuracy of an estimator. The notions are illustrated with the example of the blood pressures given above. This Chapter is useful to understand the statistical issues arising in the usual estimators in Computational Medicine in the later chapters.

2.2 Theory on estimators

We refer to [Lehmann 1998] for additional details on this section.

2.2.1 First definitions

A model describing the observations We consider the toy example given in the introduction. The goal is to answer the following question: what is the human typical blood pressure? This question implicitly relies on the assumption that there exists a *unique blood pressure* - which we call μ- that is shared across (healthy) humans. The fact that we measure n *different* blood pressures x_i's for the n different subjects is explained by measurement errors, for example from imperfect measurement devices. In other words, the observed data x_i's are generated following a

model, described by the parameter μ and additional parameters representing the measurement error.

Specifically, one assumes that the data x_i's are observations of a random variable $\mathbf{x}$, which takes values in the *sample space* $\mathcal{X}$. In our example, the observations x_i's are real numbers $x_i \in \mathbb{R}$ and the sample space is the real line $\mathcal{X} = \mathbb{R}$. On Figure 2.1, the data x_i's are represented as white dots on the real horizontal line $\mathcal{X} = \mathbb{R}$. In contrast, the next Chapter introduces data that belong to sample spaces $\mathcal{X}$ that are not necessarily vector spaces.

One further assumes that the random variable $\mathbf{x}$ follows some probability distribution P, defined over the sample space $\mathcal{X}$, and that P belongs to a class $\mathcal{P}$. The elements of $\mathcal{P}$ are usually indexed by a *parameter* denoted θ:

$$\mathcal{P} = \{P_\theta, \theta \in \Omega\} \tag{2.1}$$

For example, one may assume that the blood pressures x_i's are generated from a Gaussian distribution of parameters $\theta = (\mu, \sigma)$, where μ is the mean and σ the standard deviation, see Figure 2.1 where the probability distribution P_θ is represented in green. A *statistical model* is defined as a set of assumptions concerning the generation of observed data x_i's and usually refers to the class $\mathcal{P}$.

The notation θ is standard for referring to the parameter characterizing a given P. The parameter θ can be a vector belonging to a finite-dimensional vector space, like in parametric and semi-parametric models, or to an infinite-dimensional, like in non-parametric models where θ can represent a function. In contrast, the next Chapter introduces parameters θ that do not necessarily belong to vector spaces.

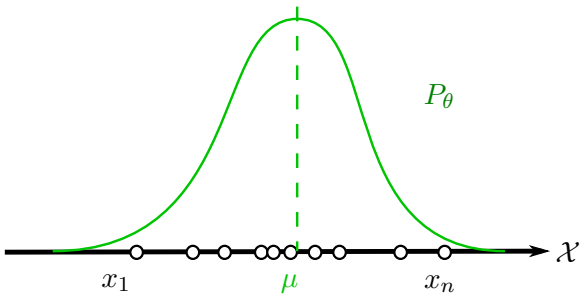


Figure 2.1: Data x_i's generated by underlying P_θ in the sample space $\mathcal{X}$.

Definitions: estimator, estimate, estimand The goal is to estimate the typical human blood pressure μ from the information available, that is the observed pressures x_i's. In general, an estimation procedure looks for the parameter θ of the generative model relying on the data x_i's. It defines a function $\hat{\theta}$ over the sample space $\mathcal{X}$, which is called an *estimator*. Note that $\hat{\theta}$ is the function and $\hat{\theta}(\mathbf{x})$ is a random variable, whose distribution is represented in orange in Figure 2.2 (b), the estimand being in green.

The parameter we want to estimate θ is called the *estimand*. $\hat{\theta}(x)$ is called the *estimate*. $\hat{\theta}(x)$ represents our best guess for θ given the data x and is illustrated on

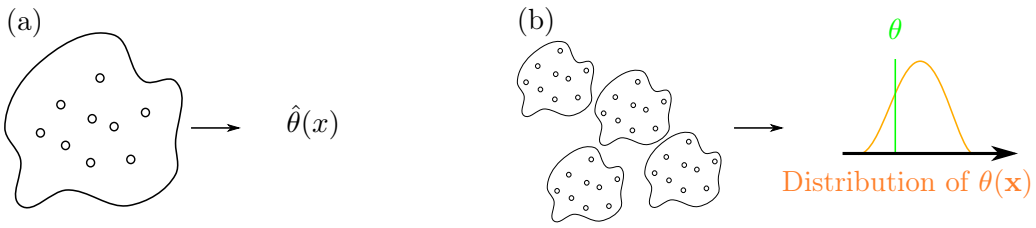


Figure 2.2: Data x_i's generated by underlying P_θ in the sample space $\mathcal{X}$.

Figure 2.2 (a). Both θ and $\hat{\theta}(x)$ belong to $\mathcal{X}$. All in all, an estimator $\hat{\theta}$ is a rule for calculating an estimate $\hat{\theta}(x)$ of an estimand θ based on observed data x. We note that the terms "estimator" and "estimate" are often used interchangeably in the literature. The goal of estimation theory is to construct and compare estimators.

In our example, the pressures x_i's are generated by a Gaussian distribution of mean μ and standard deviation σ. We seek to estimate the parameter μ, which is the estimand. To this aim, we can compute the estimate $\hat{\mu}(x) = \frac{1}{n}\sum_{i=1}^{n} x_i$. The estimator is the function $\hat{\mu}$, which is called the sample mean.

Example of estimators: Maximum-Likelihood The definition of an estimator gives no restrictions on which procedure one shall use to estimate θ. Statisticians have developed paradigms to construct estimators. One of them is *Maximum likelihood estimation (MLE).*

The paradigm of Maximum Likelihood selects for $\hat{\theta}(x)$ the set of values of the model parameters θ that maximizes the likelihood function:

$$L(\theta) = p(x_1, ..., x_n|\theta), \tag{2.2}$$

which is the probability that this parameter θ has generated the data x_i's, $i = 1...n$. The Maximum Likelihood method seeks to maximize the "agreement" of the model parameterized by θ with the observed data.

In our example, the estimator $\hat{\mu}$ chosen is the sample mean. In fact, the sample mean is the Maximum Likelihood estimator of the mean of a Gaussian distribution. So we are using the Maximum Likelihood paradigm assuming that the blood pressures are generated with our Gaussian distribution.

2.2.2 Estimators' properties

We assume that we have available different rules to construct different estimators of the same parameter θ. How can we compare them? What is the "best" estimator to use in a given situation? And does "best estimator" even make sense? This leads to the question of the accuracy of an estimator, with respect to the parameter it is designed to estimate.

To evaluate the estimator $\hat{\theta}$, we consider the quantity $\hat{\theta}(\mathbf{x})$ where x is the random variable that generates the observations x_i, $i = 1...n$. Thus $\hat{\theta}(\mathbf{x})$ is also a random

Table 2.1: Properties of an estimator $\hat{\theta}$ of a parameter θ, for $\theta \in V$ an Euclidean space.

Property of $\widehat{\theta}$	Euclidean definition
Error	$e = \widehat{\theta} - \theta$
Mean Squared Error	$\mathrm{MSE}(\widehat{\theta}) = \mathrm{E}[(\widehat{\theta} - \theta)^2]$
Std deviation/error	$d = \sqrt{\widehat{\theta} - \mathrm{E}[\widehat{\theta}]}$
Variance	$\mathrm{var}(\widehat{\theta}) = \mathrm{E}[(\widehat{\theta} - \mathrm{E}[\widehat{\theta}])^2]$
Bias	$B(\widehat{\theta}) = \mathrm{E}[\widehat{\theta}] - \theta$
Consistency	$\hat{\theta}_n \underset{n\to\infty}{\overset{P}{\to}} \theta$
Asymptotic Normality	$\sqrt{n}(\hat{\theta}_n - \theta) \underset{n\to+\infty}{\overset{D}{\to}} N(0, V)$, for some variance V.

variable and can consider its probability distribution function (p.d.f.). The properties of its p.d.f., with respect to the parameter θ, give indications on the accuracy of the estimator $\hat{\theta}$. The usual properties of estimators are summarized in Table 2.1 and we detail them below.

Evaluation criterion: Error and Mean square error of the estimator $\hat{\theta}$ We want to compare $\hat{\theta}$ to the parameter θ it is designed to estimate. The most intuitive idea is to consider their difference: $e = \hat{\theta} - \theta$, or $e(x) = \hat{\theta}(x) - \theta$ for a given set of observations x_i, $i = 1...n$. This is the error $e(x)$. However, this depends on the specific set of observations x and we would rather have a quantity that is independent of a sample. Therefore, we consider the random variable $\theta(\mathbf{x}) - \theta$ and, to get rid of the sign of the error, its squared version: $(\theta(\mathbf{x}) - \theta)^2$. Ultimately, as we seek one real number that can quantify the error of the estimator, we get to the definition of the Mean Square Error as: $\mathrm{MSE}(\widehat{\theta}) = \mathrm{E}[(\widehat{\theta}(X) - \theta)^2]$. The MSE indicates how far, on average, the estimates are from the single parameter being estimated. In our example of estimating the blood pressure μ, it indicates how far, on average, the sample means $\mu(x)$ computed from different samples x's are from the actual parameter μ.

We remark that the MSE is just a criterion indicating that $\hat{\theta}(x)$ is close to θ. Such a criterion is called the *risk* or the *loss function* associated with the estimator $\hat{\theta}$. We could as well have want another criterion, for example the following probability to be large: $P(|\hat{\theta} - \theta| < c)$, or the following expectation to be small: $\mathbb{E}(|\hat{\theta} - \theta|)^p$ or, for a positive estimator and estimand, the following to be small: $\mathbb{E}(|\frac{\hat{\theta}}{\theta} - 1|^p)$. But the MSE has good computational properties and is widely used in the literature.

Bias and variance of the estimator $\hat{\theta}$ We can consider other measures of accuracy of the estimator $\hat{\theta}$. The *bias of* $\hat{\theta}$ is the expected value of the error:

$B(\widehat{\theta}) = \mathrm{E}(\widehat{\theta}) - \theta$. But a large error for one estimate does not mean the estimator is biased. Moreover, we can have a null error for a given estimate of a biased estimator just by chance. The sample mean $\hat{\mu}$ is an unbiased estimator of the mean μ of a Gaussian distribution. In our blood pressure example, our estimator is thus unbiased.

The *standard error of* $\hat{\theta}$ is: $d = \sqrt{\widehat{\theta} - \mathrm{E}(\widehat{\theta})}$. Its square is the *variance of* $\hat{\theta}$. Both quantify the amount of dispersion of the estimates around the true value of the parameter θ. For our example of the blood pressures, the standard error is equal to σ/n, where n is the sample size.

Bias and variance are linked through the definition of the MSE. The MSE indeed decomposes as:

$$MSE(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2] = \Big(\mathbb{E}(\hat{\theta} - \theta)\Big)^2 + \mathbb{E}((\hat{\theta} - \mathbb{E}(\hat{\theta}))^2) \tag{2.3}$$

where the first term is the squared *bias of* $\hat{\theta}$ and the second term is the squared *standard error of* $\hat{\theta}$. We note that the variance is the MSE when the estimator $\hat{\theta}$ is unbiased. Bias and variance are represented on Figure 2.3. They refer to the first and second moments of the probability distribution of the estimator $\hat{\theta}$.

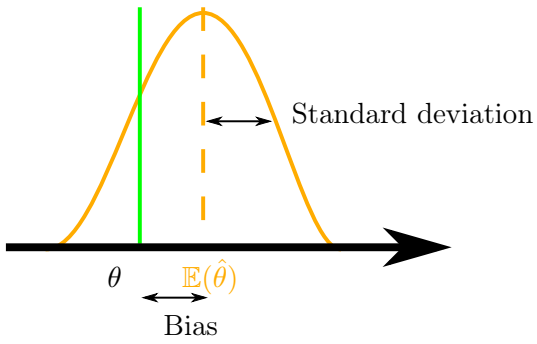


Figure 2.3: Probability distribution of $\hat{\theta}$ for n data.

The ideal situation is to have an unbiased estimator with low variance. Additionally, we want to limit the number of outliers, i.e. samples where the error is extreme. But unbiasedness is not essential. If one allows for a little bias, one may find an estimator with lower MSE and/or fewer outlier sample estimates. A rule-of-thumb states that the bias can be neglected as long at it is less that 0.25 of the variance [Efron 1993].

Skewness and kurtosis We can borrow definitions associated to p.d.f's to further characterize the estimator $\hat{\theta}$. We consider the third and fourth moments of the p.d.f. of $\hat{\theta}$, just as the bias and variance referred to first and second moments.

The *skewness of* $\hat{\theta}$ is the third standardized moment:

$$\text{Skew}[\hat{\theta}] = \frac{\text{E}\left[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^3\right]}{\text{E}\left[(\hat{\theta} - \text{E}[\hat{\theta}])^2\right]^{\frac{3}{2}}} \tag{2.4}$$

The skewness measures the asymmetry of the probability distribution of $\hat{\theta}$ about its mean $\mathbb{E}(\hat{\theta})$.

The *kurtosis of* $\hat{\theta}$ is the fourth standardized moment:

$$\text{Kurt}[\hat{\theta}] = \frac{\text{E}\left[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^4\right]}{\text{E}\left[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2\right]^2}, \tag{2.5}$$

where σ is the standard deviation. The kurtosis is a measure of the "tailedness" of the probability distribution of a real-valued random variable, and in particular of a real-valued estimator.

Asymptotic behaviors: consistency and asymptotic normality The properties above have been defined for a given size n of the data set and corresponding estimator $\hat{\theta}_n$. We consider now the behavior of the sequence $(\hat{\theta}_n)_n$ for $n \to +\infty$, i.e. we tackle the question: how does the estimator's accuracy depend on the number of observed data n?

A *consistent estimator* $\hat{\theta}$ is an estimator converging in distribution to a Dirac delta function centered at θ. The probability distributions of the $\hat{\theta}_n$'s become more and more concentrated, near the true value θ. In other words, the estimator finds the real parameter θ when there is an infinite number of data.The sample mean is a consistent estimator of the mean on a vector space, such that, for the blood pressure example, we can find the exact value of the human blood pressure if we measure the blood pressure of an infinite number of subjects. We remark that bias is related to consistency as follows: a sequence of estimators is consistent if and only if it converges to a value and the bias converges to zero. An estimator can be unbiased but not consistent, or biased but consistent.

An *asymptotically normal estimator* $\hat{\theta}$ is a consistent estimator whose distribution around the true parameter θ approaches a normal distribution with standard deviation shrinking in proportion to $1/\sqrt{n}$ as the sample size n grows. For example, the central limit theorem implies asymptotic normality of the sample mean $\hat{\mu}$ as an estimator of the true mean μ. For our example, this means that with a lot of subjects, the distribution of the estimates will look normal, centered around the human blood pressure. More generally, maximum likelihood estimators are asymptotically normal under some regularity conditions.

"Best" estimators? We have defined some accuracy measures to evaluate the quality of an estimator. Given a statistical problem and a parameter we wish to

estimate, can we answer the question: what is the best estimator? The definition of efficiency provides answers to this.

An *efficient estimator* $\hat{\theta}$ is an estimator that estimates the parameter θ in some "best possible" manner. This depends on the criterion chosen to evaluate the estimator, so usually the mean squared error MSE. If $\hat{\theta}_1$ and $\hat{\theta}_2$ are both estimators for the parameter θ, then $\hat{\theta}_1$ is said to dominate $\hat{\theta}_2$ if:

$$\mathrm{MSE}(\hat{\theta}_1)^2) \leq \mathrm{MSE}(\hat{\theta}_2)^2) \tag{2.6}$$

holds for all θ, with strict inequality holding somewhere. One usually says that a given estimator is "inefficient" if there exists another estimator that dominates it.

Let us consider only unbiased estimators. A *uniformly minimum-variance unbiased estimator* or *minimum-variance unbiased estimator* (UMVUE or MVUE) is an unbiased estimator that has lower variance than any other unbiased estimator for all possible values of θ. The MVUE is thus efficient among unbiased estimators. If an unbiased estimator of θ exists, then one can prove there is an essentially unique MVUE. For example, the sample mean is an efficient estimator for the mean of the Gaussian distribution. The sample mean is also a MVUE. As a consequence, it represents a notion of "best" possible estimator for the mean μ.

2.3 Estimating an estimator's properties

The measures of accuracy of the estimator $\hat{\theta}$ use the parameter θ in their definition. However, θ is not known since our goal is to estimate it. For example, we do not actually know the value of the parameter μ of the Gaussian distribution. So one cannot apply the above definitions of measure of accuracy directly. This section shows how to compute these measures in practice, and more precisely how to *estimate* these measures.

2.3.1 Using closed forms

Some estimators, like the sample mean $\hat{\theta} = \hat{\mu}$ for estimating the mean μ, are widespread in the literature. The computation of their measures of accuracy in practice can be done using closed forms.

The sample mean $\hat{\mu}$ is known to be unbiased. Its standard error is computed as follows:

$$\hat{d}(x) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2}$$

where $\{x_1, x_2, \ldots, x_n\}$ is the sample. This is actually a biased estimate of the standard deviation. Bessel's correction is the use of $n-1$ instead of n in the formula:

$$\hat{d}(x) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

This corrects the bias in the estimation of the standard error on the sample mean.

2.3.2 Resampling methods: bootstrap

But often, there is no closed form formula for computing - or estimating - the measures of accuracy of the estimator $\hat{\theta}$. We need an "universal" method that can estimate the accuracy of $\hat{\theta}$, that works for any $\hat{\theta}$. The "resampling methods" like the bootstrap [Efron 1993], which we introduce in this subsection. They are called "resampling methods" because we sample, often with replacement, from the original sample of the x_i's.

Plug-in principle: P_θ and $\hat{P}_\theta$ The bootstrap is an application of the plug-in principle, which we present here. We recall that P_θ is the probability distribution that generates the data x_i's:

$$P_\theta \to x = (x_1, ..., x_n) \tag{2.7}$$

The *empirical distribution* $\hat{P}_\theta$ is defined to be the discrete distribution which puts probability $1/n$ on each observed value x_i. It is a simple estimate of the distribution P. Alternatively, one can consider $P_{\hat{\theta}}$, which is the probability distribution having $\hat{\theta}$ as parameter and is also a simple estimate of P.

If one wants to estimate some aspects of P_θ, like its mean or the parameter θ, one can consider the corresponding aspect of $P_{\hat{\theta}}$ or $P_{\hat{\theta}}$. This is called the *plug-in principle*. For example, let us denote t the operation that extracts the parameter θ from the probability distribution P_θ. Then, the plug-in estimate of $\theta = t(P_\theta)$ is naturally $\hat{\theta} = t(\hat{P}_\theta)$. We consider our running example of estimating the mean of the Gaussian. The plug-in estimate of the mean $\theta = \mu$ of P_θ is the mean of $\hat{P}_\theta$, i.e. the sample mean.

Non-parametric and parametric bootstrap: sampling from $\hat{P}_\theta$ and $P_{\hat{\theta}}$ A non-parametric *bootstrap sample* is defined to be a sample generated by the empirical distribution $\hat{P}_\theta$, so that:

$$\hat{P}_\theta \to x^* = (x_1^*, ..., x_n^*) \tag{2.8}$$

It is common to use the star-notation to refer to the bootstrap sample. We note that creating a bootstrap sample from $\hat{P}_\theta$ can be done by drawing n times from the n data in x with replacements. The bootstrap sample has elements of the original sample, some appearing several times and some not appearing at all.

In contrast, a parametric *bootstrap sample* is defined to be a sample generated by the distribution $P_{\hat{\theta}}$, so that:

$$P_{\hat{\theta}} \to x^* = (x_1^*, ..., x_n^*) \tag{2.9}$$

In this case, one generates n data from $P_{\hat{\theta}}$.

Using bootstrap samples to estimate the accuracy of $\hat{\theta}$ We use the (parametric or non-parametric) bootstrap samples to estimate the measures of accuracy as follows. The estimate $\hat{\theta}(x)$ was computed from the data set x as: $\hat{\theta}(x) = s(x)$,

i.e. using a function s that takes the sample as input. A *bootstrap replication of* $\hat{\theta}$ is defined as: $\hat{\theta}^* = s(x^*)$. In other words, we apply the same function s we had use on the original data set, but we apply it to the bootstrap data set. We can perform B resampling, leading to B bootstrap samples, each of size n. Each time, we compute a bootstrap replication $\hat{\theta}^*$ of $\hat{\theta}$. The bootstrap procedure writes:

- For $b = 1...B$:
  - Generate b-th bootstrap sample:
    $\hat{P}_\theta \rightarrow x^{b*} = (x_1^{b*}, ..., x_n^{b*})$ (non-parametric)
    or
    $P_{\hat{\theta}} \rightarrow x^{b*} = (x_1^{b*}, ..., x_n^{b*})$ (parametric)
  - Compute b-th bootstrap replication: $\hat{\theta}^{*b} = s(x^{b*})$
- From the empirical distribution of $\hat{\theta}^*$'s, compute the bootstrapped desired property: bias, standard error, etc.

As an illustration, we estimate the properties of the sample mean as an estimator of the real mean parameter. To this aim, we resample B times, with replacement, from the original data set x_i, $i = 1...n$. This gives us B data sets each of size n, from which we compute the sample mean. We then use the distribution of the sample mean as an approximation of the distribution of the corresponding estimator.

If we repeat the bootstrap sampling an infinite number of times, i.e. for $B \rightarrow \infty$ *in the above procedure, we get the plug-in estimate of the distribution of the estimator* $\hat{\theta}$ and the plug-in estimates of various measures of accuracy that we could compute from it. But in practice, we do not repeat an infinite number of times the experiment so that B is finite. So we get an approximation of the plug-in estimate of the distribution of $\hat{\theta}$ and approximation of the plug-in estimates of the measures of accuracy. These are called *bootstrap estimates* for a number B of replications. The convergence of the bootstrap estimates to the ideal bootstrap estimate are linked to Monte-Carlo methods. In practice, the number of replications B needed will depend on the property of $\hat{\theta}$ we want to estimate. We give examples below.

Bootstrap estimation of MSE, standard error, and bias We have the *bootstrap estimate*, for a number B of replications, of the distribution of $\hat{\theta}$. We can compute the moments of this distribution, and for example its mean and its variance which lead to bootstrap estimates of the bias and standard error of $\hat{\theta}$. We note the subscript B that illustrates the fact that they are bootstrap estimates. We provide rules of thumb regarding the recommended number B of replications for each case, which comes from the convergence properties of the Monte-Carlo [Efron 1993].

The bootstrap estimate of the standard error of $\hat{\theta}$ is:

$$\hat{se}_B(\hat{\theta}) = \left(\sum_{b=1}^{B} \frac{(\hat{\theta}^{*b} - \text{mean}(\hat{\theta}^*))^2}{B-1} \right)^{1/2} \tag{2.10}$$

The recommended number of replications is around $B = 50$. Then, the bootstrap estimate for the bias is:

$$\hat{bias}_B = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_b^* - \hat{\theta} \tag{2.11}$$

which takes around $B = 3000$ to converge. One can also use:

$$\overline{\hat{bias}}_B = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_b^* - \overline{\theta}^* \tag{2.12}$$

where $\overline{\theta}^*$ is the estimator computed from the $\overline{P} * *$, the average of the bootstrap distributions. The recommended number of replications is $B = 400$. Ultimately, the bootstrap estimate of the MSE of $\hat{\theta}$ is:

$$\hat{\text{MSE}}_B(\hat{\theta}) = \frac{1}{B}\left(\sum_{b=1}^{B}(\hat{\theta}^{*b} - \hat{\theta}^2\right)^{1/2} \tag{2.13}$$

Nested bootstrap The bootstrap procedure gives estimations of an estimator's properties: an *estimate* of the bias $\hat{bias}_B$, an *estimate* of the standard error $\hat{se}_B(\hat{\theta})$, etc. How accurate are these estimations? To answer this question, we can apply the same reasoning as before. We compute bootstrap replications of $\hat{bias}_B$ and $\hat{se}_B(\hat{\theta})$! This leads to a nested bootstrap [de La Rosa 2006].

• For each $b = 1...B$:

- Take b-th bootstrap sample x^{b*} and b-th bootstrap replication $\hat{\theta}^{b*}$ computed above,
- For $c = 1...C$:
  - Compute b, c-th bootstrap sample:
    With replacement from x^{b*} (non-parametric) or from $\hat{\theta}^{b*}$ (parametric) to get: $x^{b,c*}$
  - Compute b, c-th bootstrap replication: $\hat{\theta}^{b,c*}$
- Compute the bootstrap distribution of $\hat{\theta}b*$ and bootstrapped desired property: $\hat{bias}^{b*}$

• This gives the bootstrapped distribution of the desired property: $\left(\hat{\text{bias}}^{b*}\right)_b$: one can compute the properties of this estimator.

2.4 Improving an estimator

We have seen how to compute measures of accuracy of an estimator. Now we present computational methods to improve the estimator by improving the corresponding measures of accuracy. In particular, we show how to reduce the bias and the variance of an estimator.

2.4.1 Bias correction

Assume we have an estimator known to be biased. We present one methods for correcting the bias of $\hat{\theta}$ [Efron 1987], which is a direct application of the bootstrap procedures above.

We estimate the bias with a parametric or non-parametric procedure. Then, we modify the estimator by subtracting the estimated bias from the original estimate. If in addition, we believe that the estimate of the bias is itself biased, we can perform a nested bias-correction. We first correct the bias in the bias' estimate and then correct the estimate $\hat{\theta}$ with the bias-corrected bias' estimate. However, while it will reduce the bias of $\hat{\theta}$ it will also increase the variance. So for it to be useful the improvement in bias must be large relative to the loss in the variance.

2.4.2 Bagging: variance reduction

Bootstrap aggregating, also called bagging, is a procedure meant to improve the variance of an estimator [Breiman 1996]. It is usually performed to avoid over-fitting for high variance low bias models like decisions trees or neural networks.

The procedure, very close to the bootstrap, is:

- For $b = 1...B$:
  - Generate b-th bootstrap sample:
    $\hat{P}_\theta \rightarrow x^{b*} = (x_1^{b*}, ..., x_n^{b*})$ (non-parametric) or $P_{\hat{\theta}} \rightarrow x^{b*} = (x_1^{b*}, ..., x_n^{b*})$ (parametric)
  - Compute b-th bootstrap replication: $\hat{\theta}^{*b} = s(x^{b*})$
- Aggregate the estimates to get $\hat{\theta} = \frac{1}{B}\sum_{b=1}^{B} \theta_b^*$.

Bagging averages B estimators θ_b^*'s computed from B different bootstrap samples. This average has smaller variance than just one sampling from the same distribution. To see it, we can consider a the standard Gaussian with 0 mean and $\sigma = 1$. One sample x has mean 0 and variance 1. When we sample B times and average the results, the mean is still be 0, but the variance becomes $1/B$.

2.5 Conclusion

In this chapter, we have presented an overview of Estimation theory. We have introduced the concepts in the case of data and parameters belonging to a Euclidean space. This chapter represents the basis of the vocabulary that will be used in this thesis. We will use it to evaluate the estimations procedures that are used in Computational Medicine and that show non-Euclidean behaviors.

We have especially emphasized examples linked to estimating the mean of a sample. We have detailed procedures to estimate the mean of a data set of blood pressures $x_i \in \mathbb{R}$ and introduced the generalization needed to estimate the mean of a data set of brain anatomies. Estimating the mean - which represents the easiest

estimation procedure - will be central in the context of Computational Medicine in the following of the thesis.

Chapter 3

Elements of Geometry: manifolds with additional structures

This Chapter introduces elements of Differential Geometry used in this Ph.D thesis. They describe some non-linear data spaces of Computational Anatomy. The Chapter can be read independently from Chapter 2. Chapter 2 defined the basics of the "Statistics" aspect of "Geometric Statistics". This Chapter now introduces the "Geometric" aspect.

While reading this Chapter, we invite the reader to think about the generalization of Estimation theory of Chapter 2 to non-linear data spaces. One can consider data x_i's belonging to a non-linear space $\mathcal{X}$. One can also take the model's parameter θ belonging to a non-linear space. The remaining of the Ph.D thesis will apply the statistical principles of Estimation theory to these non-linear data spaces, in order to investigate open issues in Computational Anatomy.

3.1 Introduction

Historical landmarks of Differential Geometry The theory emerged from the study of curves and surfaces in 2D and 3D Euclidean spaces in the 18th and 19th centuries. Since the late 19th century Differential Geometry has grown into a more abstract fields, concerned with structures on differentiable manifolds and among them: (pseudo-) Riemannian geometry, Finsler geometry, Symplectic geometry, Contact geometry, Complex and Káhler geometry, CR geometry, Lie groups, Bundles and connections, etc.

Differential Geometry has been extremely successful in describing phenomena in natural sciences, and for example in Physics. Riemannian geometry for example represents the mathematical foundations of Einstein's General Relativity in the 20th century. Symplectic geometry is used to describe Hamiltonian systems and Contact geometry has found applications in Thermodynamics. One may wonder which branches of Differential Geometry are the most suited to Computational Medicine and Computational Anatomy.

Differential Geometry for Computational Anatomy Differential Geometry appears naturally in Computational Anatomy where the data spaces can be described as manifolds. The data are organ shapes - as seen in medical images - which may be modeled in several ways: (i) as deformations of a reference shape i.e. as elements of a Lie group, which is a manifold with a group structure, and (ii) as the

equivalence class of their configuration in the 3D space under the action of rotations, translations i.e. as elements of a quotient space. Furthermore, the medical images themselves can be embedded in a sub-Riemannian manifold for better processing.

What do these data spaces have in common: the Lie groups, the manifolds with Lie group actions, and the sub-Riemannian spaces? They are all manifolds, i.e. non-linear spaces that *locally* look like vector spaces. Defining a theory of statistics on manifolds is still considered open research. A widespread approach is to define a Riemannian metric on the manifold and then define statistics on *Riemannian manifolds* [Pennec 2006]. Indeed, the Riemannian metric enables to define a Riemannian distance between the points on the manifold. The distance can be used to generalize definitions of Statistics to the manifold.

How do these data spaces differ? While being all manifolds, they have structural specifications of their own. They may be equipped with: a Lie group action, a stratification, an horizontal distribution, etc.

Outline of this Chapter This Chapter introduces elements of Differential Geometry needed in this Ph.D thesis. The first part of this chapter - Section 3.2 - thus focuses on the definition of manifold and of the (pseudo-) Riemannian metrics that we may consider on them. Then, Sections 3.3-3.7 define some additional structures on manifolds and their relations. For each structure, we give a formal and intuitive definition and we show the compatibility conditions arising while defining a Riemannian pseudo-metric on it.

3.2 Manifolds and Riemannian manifolds

Vector spaces, high-dimensional smooth surfaces and the data spaces encountered in Computational Anatomy are all "manifolds". This section gives some mathematical definitions about manifolds. We refer to [Nakahara 2003] and [Postnikov 2001] for formal details and the Einstein notation on indices.

3.2.1 Differentiable structure of manifolds

A *manifold* M of dimension m is a generalization of a vector space of dimension m: M is a space that looks like a m-dimensional vector space in the neighborhood of every point. A *differentiable manifold* M of dimension m is a type of m-dimensional manifold that is locally similar enough to a vector space to allow one to do calculus. Such calculus is performed through a chart, which is an invertible map ϕ between a subset of the manifold M and $\mathbb{R}^m$, such that both the map and its inverse preserve the desired structure. This type of calculus is called "Differential Calculus" and represents the basis of "Differential Geometry". The next paragraph briefly explains why one uses the term "Differential".

We can define the space $C^\infty(M)$ of smooth functions f on the manifold M, which associate a real number to a given point $X \in M$. But how do we define the differentiation of such a function? There may be several coordinate systems on the

manifold: how do we give a coordinate-independent definition of derivative? The tangent vectors of M, defined below, allow us to do exactly this.

We consider a chart on M, $\phi : U \to \mathbb{R}^m$ where U is an open subset of M containing a point X. We take two curves $\mathcal{C}_1 : [-1, 1] \to M$ and $\mathcal{C}_2 : [-1, 1] \to M$ that coincide at 0: $\mathcal{C}_1(0) = \mathcal{C}_2(0) = X$. Moreover $\phi \circ \mathcal{C}_1$ and $\phi \circ \mathcal{C}_2$ are assumed to be differentiable at 0. The curves are called equivalent at X if the ordinary derivatives of $\phi \circ \mathcal{C}_1$ and $\phi \circ \mathcal{C}_2$ at 0 coincide. One can show that this defines an equivalence relation on such curves. The equivalence classes are known as *the tangent vectors of M at X*. The space of tangent vectors at X is called the tangent space of M at X and is denoted $T_X M$.

More generally, $T_X M$ forms the algebra of derivations at X of smooth functions $f \in \mathcal{C}^\infty(M)$. Indeed, we have the fundamental isomorphism of Differential Geometry at each point $X \in M$:

$$\mathcal{D}_X(M) \simeq T_X M \tag{3.1}$$

where $\mathcal{D}_X(M)$ is the vector space of directional derivatives at X [Nakahara 2003]. This is the reason of the term "Differential" in "Differential Geometry".

The set of all tangent spaces of M, at any point of M, is called the *tangent bundle TM of M*. Then, we denote $\Gamma(M)$ the space of vector fields on M. In a local coordinate system, the action of a vector field $V \in \Gamma(M)$ at X on f is :

$$V[f]\Big|_X = \partial_X f\Big|_X = \frac{d}{d\lambda}(f(X + \lambda X|_X))\Big|_{\lambda=0}. \tag{3.2}$$

When composing the derivatives, $V[U[f]]|_X = \partial_V \partial_U f|_X$ involves second order derivations and does not behave like a vector (or a directional derivation). We usually remove the second order terms by subtracting $\partial_V \partial_U$ to define the *Lie Bracket of vector fields*:

$$[V, U][f] = V[U[f]] - U[V[f]] \tag{3.3}$$

which is now a derivation [Nakahara 2003].

The above structures on tangent vectors and tangent spaces are defined on any differentiable manifold. In particular they are defined on all Lie groups, manifolds with group actions, or bundles that we will encounter in the next sections.

3.2.2 Riemannian and pseudo-Riemannian structure on manifolds

So far, the manifold M is only equipped with a differentiable structure. Distance between points are not yet defined, and neither are angles between tangent vectors. One can imagine M as a soft and elastic high-dimensional smooth surface. Defining a (Riemannian) metric on M provides a way to define distances and angles. As such, the metric gives the manifold M some rigidity: the high-dimensional surface chooses a shape.

A *Riemannian metric* is a smooth collection of positive definite bilinear 2-forms on the tangent spaces of the manifold M, i.e. it defines the dot product of tangent vectors. This structure is the most commonly used in Computational Anatomy.

Then, by relaxing the positivity assumption on the metric, one defines a Riemannian pseudo-metric. A *Riemannian pseudo-metric* is a smooth collection of definite bilinear 2-forms on tangent spaces of the manifold M, i.e. non necessarily positive definite. We denote (p, q) its signature, $p + q = \dim M$, where p is the number of strictly positive eigenvalues, and q the number of strictly negative ones. We remark that a Riemannian metric is a Riemannian pseudo-metric of signature $(\dim M, 0)$.

A Riemannian (pseudo-)metric allows us to determine angle between tangent vectors (of the same tangent space) and their lengths. Moreover, the (pseudo-)metric induces a pseudo-distance along a curve $\mathcal{C}$ of the manifold M through integration of its tangent vector in affine parameterization s:

$$I[\gamma] = \int_\gamma ds \sqrt{< \gamma(\dot{x}), \gamma(\dot{s}) > |_{\gamma(s)}} \quad \text{(i.e. the length)}.$$

Then, a *Riemannian geodesic* γ of M is a curve minimizing the functional $I[\gamma]$ or, equivalenly, the functional [Nakahara 2003]:

$$K[\gamma] = \frac{1}{2} \int_\gamma ds < \gamma(\dot{x}), \gamma(\dot{s}) > |_{\gamma(s)} \quad \text{(i.e. the kinetic energy)}$$

The Riemannian (pseudo-)metric gives an intrinsic way of measuring length between two points on a manifold: we compute the length of the geodesic linking the two points, provided that it exists and is unique.

3.3 Lie groups

In this section, we turn to a first additional structure one may consider on a manifold. We add an algebraic structure on M, the structure of a group. This leads us to the definition of Lie groups.

Lie groups are extensively used in Physics where they describe symmetries in physical systems. For example, they express the relativity principle of General Relativity: this is the requirement that the equations describing the laws of physics have the same form in all admissible frames of reference. Going from one frame to another is expressed through a Lie group element. Lie groups are also used to represent elementary particle symmetries, where two different Lie groups describe quark fields or electron fields.

More recently, Lie groups have also been used to represent organ transformations or deformations in Computational Anatomy. For example, any rotation or translation of an imaged organ, due to the differences in imaging procedures, is an element of a Lie group. Smoothly deforming an organ shape to have it match another also requires an element of a Lie group. The structure of Lie groups will be used in particular in Chapters 4 and 5 of this Ph.D thesis.

3.3.1 Lie groups and its translations

A *Lie Group* G is a smooth manifold with a compatible group structure. It is provided with an identity element e, a smooth composition law $* : (g, h) \mapsto g*h \in G$

and a smooth inversion law $Inv : g \mapsto g^{(-1)} \in G$. Its tangent space at g is written T_gG.

There is a canonical way to define two diffeo/auto-morphisms families of G, called *the left and right translations*: $L_g : f \mapsto g * f$ and $R_g : f \mapsto f * g$ in addition to a third one called *the conjugations* $C_g = f \mapsto g * f * g^{(-1)}$ [Nakahara 2003].

3.3.2 Adding a metric structure

We consider a Riemannian pseudo-metric g^G defined on the Lie group G. As such pseudo-metric can be defined on any manifold, it can also be defined on G. We investigate the compatibility of g^G with the Lie group structure.

The pseudo-metric g^G is said *left-invariant pseudo-metric* if for all $U, V \in T_g\mathcal{G}$ and for all $g, h \in \mathcal{G}$ we have:

$$< DL_h(g)U, DL_h(g)V > |_{L_hg} =< U, V > |_g \tag{3.4}$$

where L_h is the left translation by h, i.e. if the left translations are isometries for this pseudo-metric. Similarly, g^G could be *right-invariant* and *bi-invariant*. Bi-invariant pseudo-metrics show the most compatibility with the group structure. Chapter 4 studies the bi-invariant property of pseudo-metrics and its importance for Statistics on Lie groups.

3.4 Quotient spaces

In this section, we turn to another additional structure on a manifold M: the structure of Lie group action. The manifold M is *not* the whole Lie group G like in the previous section, but G acts on M. Furthermore, we consider the quotient of the Lie group action which leads us to the structure of quotient space.

Such structures intuitively describe several situations in Computational Anatomy. For example rotating and translating an organ's surface involves *the action* of a rotation and a translation on the surface. Then, the shape of the organ's surface may be defined by the geometry that is preserved while rotating and translating the surface. The quotient of the space of surfaces by the action of rotations and translation defines the space of the surfaces' shapes. The structures of Lie group actions and quotient spaces will be used in particular in Chapters 5 and 6 of this Ph.D thesis.

3.4.1 Lie group actions and quotient spaces

A *Lie group (left) action of* G on a manifold M is a map:

$$\rho : G \times M \to M, (g, X) \mapsto g \cdot X \tag{3.5}$$

such that for all $X \in M, e \cdot X = X$, where e is the identity, and for all $(g, g') \in G^2$ and for all $X \in M, g' \cdot (g \cdot X) = (g'g) \cdot X$. Right group actions are defined similarly. We will consider proper and effective actions (for definitions, see [Huckemann 2010]).

The *orbit of* X under the action of G, written O_X, is defined as all points of M reachable the action of G on X, thus $O_X = \{g \cdot X \mid g \in G\}$. Each orbit O_X is a submanifold of M. The *isotropy group of* X or *stabilizer of* X, written G_X, is defined as the elements G that leave X unmoved: $G_X = \{g \in G \mid g \cdot X = X\}$. Each isotropy group G_X is a Lie subgroup of G. The set of orbits $\{O_X\}_X$ forms a partition of M. It is written $Q = M/G$ and is called the *quotient space of* M *by the action of* G.

3.4.2 Adding a metric structure

We consider a Riemannian metric g^M on the manifold M. The compatibility condition between the metric and the Lie group action is the following. The action of G on M is called *isometric with respect to* g^M if $d\rho_g : T_X M \to T_{g \cdot X} M$ leaves g^M invariant. Equivalently, g^M is said to be G*-invariant*. It means that the distance between two data does not change if we act on both data in the same manner. This compatibility condition can always be satisfied in theory: given the action of G of M, we can always find a g^M that is G-invariant.

The compatibility condition with respect to the Lie group action is important in order to induce a Riemannian metric g^Q on the quotient space $Q = M/G$. From now on, we assume that g^M is G-invariant. We consider d_M the distance associated to g^M. The metric g^M being G-invariant, d_M is also G-invariant. Let $\pi : M \to M/G$ be the canonical projection to the quotient space with quotient topology. The G-invariant distance d_M on M induces a quasi-distance d_Q on Q, interpreted as the quasi-distance between shapes. It is defined as:

$$d_Q(O_{X_1}, O_{X_2}) = \inf_{g \in G} d_M(g \cdot X_1, X_2), \tag{3.6}$$

for $X_1, X_2 \in M$.

3.5 Fiber bundles

We now consider the structure of fiber bundles over a manifold M. Particles fields and the corresponding mathematics of fiber bundles are the core of Relativistic Quantum Field theory. They represent some of the most advanced model used to describe nature, that has been confirmed by experiments. In Relativistic Quantum Field theory, a field of quarks can be roughly thought as follows: there is a (probability of finding a) particle at each point of space-time $\mathbb{R}^4$. The particles interact spatially, so there exists some correlation on finding a particle at two neighboring locations.

There are similarities between fields of elementary particles of Theoretical Physics and fields encountered in Medical Images. A tridimensional medical image can be seen as a field over a domain of $\mathbb{R}^3$, or over a grid of voxels which is a discretized version of $\mathbb{R}^3$. One may consider that there is a "particle" at each point of the grid, which can be an intensity (MRIs...) or a tensor (DTI images), a p-value

etc. These "particles" interact spatially in the sense that there is a correlation in the intensity of neighboring voxels. The structure of fiber bundles will be useful in Chapter 8.

3.5.1 Fiber bundles

A *fiber bundle* (P, M, π) *of base* M is a manifold P equipped with a differentiable projection π onto the manifold M, such that the reciprocal images $\pi^{-1}(X)$, for $X \in M$, are all diffeomorphic. Such a $F_X = \pi^{-1}(X)$ is called the *fiber* above X. For example, the tangent bundle TM of M is a fiber bundle of base M (see Figure 3.1, middle). In this case, the projection π is the operator which maps each tangent vector to its foot, i.e. each $v \in T_X M$ to X. Fiber bundles are generalizations of product spaces: fiber bundles have no projection onto the fiber F, only a projection onto the base M, whereas product spaces have both.

The projection $\pi : P \to M$ is often surjective and if not, we can restrict M to match π's image. However, π is not injective in general: the cardinal of the fiber is different from 1 in general. For each $X \in M$, we can choose an element of the fiber F_X to build a right inverse for π. We call this element $\sigma(X)$ and by construction we have: $\pi \circ \sigma = Id$. The map σ is called a *section* of the fiber bundle. There are different choices possible for defining a section. In the example of the tangent bundle, a section defines a vector field over M. There are different possible sections just as we have different vector fields on a given manifold M.

Examples We can consider the image's domain (in its continuous version) to be the base manifold M. Taking the fiber to be $\mathbb{R}$, we can build the trivial fiber bundle: $M \times \mathbb{R}$ which can be seen as the set of all possible images over the domain M. A section of this bundle gives an image. There are different sections possibles just as there are different images that can be defined on the same domain.

Alternatively we can choose the fiber to be $\mathrm{Sym}_3(\mathbb{R})$ i.e. the vector space of symmetric matrices over $\mathbb{R}$ in 3 dimensions. In this case, we can build the (trivial) fiber bundle $M \times \mathrm{Sym}_3(\mathbb{R})$. A section of this bundle represents a image of DTI: at each point of the domain, we have a symmetric matrix. The fiber bundle itself represents the set of all possible DTI images over this domain.

We can also consider the vector field defined by the intensity gradient of the given image. This vector field belongs to a vector bundle defined over the image's domain. In Physics, we can take the base manifold to be space-time $\mathbb{R}^4$ and the fiber to be $\mathbb{C}^3$. In this case, a section of the fiber can be thought of as a simplified model of a quark field.

These may seem an overly complicated representation of images. But this additional structure enables to describe operations on images in a compact way, as we shall see in the next sections.

3.5.2 Principal bundles

A *principal bundle* (P, M, π) *of structure group* G is a special type of fiber bundle, where each fiber looks like G: see Figure 3.1, right, where each fiber looks like $SO(2)$, i.e. S^1. For example the bundle of all local frames of the manifold M is a principal bundle of structure group GL_m. Thus, it can be thought about as a generalization of the direct product spaces $M \times G$. Formally, a principal bundle is a fiber bundle, equipped with a right action of the group G on P, which is transitive on the fibers, and such that the fibers are all homeomorphic to G.

Examples In Physics and more particularly in General Relativity, we can consider the principal bundle of all frames over space-time $\mathbb{R}^4$, i.e. FM. These frames are called *external frames*. In contrast, bundles of internal structure are defined by principal bundles that have nothing to do with the frame bundle FM. The external structure defines the properties of the field in space-time, and the internal structure represents the intrinsic properties like particle symmetries properties of the field.

We come back to the examples from Medical Imaging and take an image's domain as the base manifold M. Again, we can build several principal bundles with the same base manifold M. Taking the multiplicative group $(\mathbb{R}_+^*, .)$, we can form the trivial principal bundle $M \times \mathbb{R}_+^*$. Taking the additive group $(\mathbb{R}, +)$, we can form the trivial principal bundle $M \times \mathbb{R}$. We could also consider the general linear group in 3 dimensions GL_3 and create the trivial principal bundle $M \times GL_3$. These principal bundles do not define new types of images. They rather define objects that can interact with the previously defined images. They will define the images' properties, just as they defined the particles' intrinsic symmetries in Physics. This is the aim of the *associated fiber bundles* of the next section.

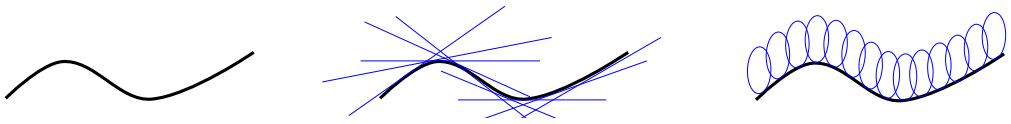


Figure 3.1: Left: a smooth curve $\mathcal{C}$ is an example of a 1-dim. manifold. Center: the tangent bundle of the curve $\mathcal{C}$, whose fibers are in blue. Right: a principal bundle with base $\mathcal{C}$ and structure group S^1, the fibers are in blue.

3.6 Connections

In this section we consider the structure of connections on the manifold M. We recall that, without any metric, the manifold M can be seen as a soft and elastic surface. The connection does not define distances between points of M, but it defines some notion of parallelism. A manifold M with a connection is still allowed to deform, but in a restricted way that preserves the parallelism. The connection is

a less restrictive structure than the metric as it does not completely fix the shape of M. We use connections on Lie groups in Chapter 4.

3.6.1 Affine connection spaces

If one wants to compare data in the tangent space $T_X M$ at point X of the group, with data at another tangent space, say $T_Y M$, one needs a specific mapping Π_X^Y between these two tangent spaces. This is the essence of parallel transport whose infinitesimal version is the connection.

Let γ be a curve in M joining $\gamma(0)$ to $\gamma(1)$. A parallel transport along this curve is a collection of mappings $\Pi(\gamma)_s^t : T_{\gamma(s)}M \mapsto T_{\gamma(t)}M$ such that:

- $\Pi(\gamma)_s^s = Id$ (identity transformation of $T_{\gamma(s)}M$),
- $\Pi(\gamma)_u^t \circ \Pi(\gamma)_s^u = \Pi(\gamma)_s^t$ i.e. consistency along the curve,
- The dependence of Π on γ, s and t is smooth.

The notion of (affine) connection is the infinitesimal version of parallel transport. The connection hence represents how a vector transforms when it is parallel transported to an infinitesimally close tangent space. Let $x = \dot{\gamma}(0)$ be the tangent vector at the initial point of the curve γ and U be a vector field. Then the quantity:

$$\lim_{t\to 0} \frac{\Pi(\gamma)_t^0 U|_{\gamma(t)} - U|_{\gamma(0)}}{t} = \frac{d}{dt}\Pi(\gamma)_t^0 U|_{\gamma(t)}\Big|_{t=0} \tag{3.7}$$

is independent of the curve γ [Postnikov 2001]. Hence we can forget about the previous expression of parallel transport and define an affine connection without specifying a curve γ.

An affine connection ∇ is a bilinear map:

$$\begin{aligned} \nabla : \Gamma(M) \times \Gamma(M) &\longmapsto \Gamma(M) \\ U, V &\longmapsto \nabla_U V \end{aligned}$$

with the following properties:

- $\nabla_{fU} V = f \nabla_U V$,
- $\nabla_U (fV) = \partial_U f V + f \nabla_U V$ i.e. Leibniz rule in the second variable for any $U, V \in \Gamma(M)$ and $f \in \mathcal{C}^\infty(M)$.

The torsion induced by ∇ on the manifold M is defined as:

$$T^\nabla(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y]. \tag{3.8}$$

If the torsion vanishes, the connection is said to be torsion-free or symmetric.

Examples In DTI images, the connection defines what happens to the 3D reference frame when we travel from one tensor to a neighboring one in the 3D image. For example, we can force the frame to be aligned with the principal axes of the tensors. In this case, we define a connection that in fact represents the structure of the fibers of the organ studied. The properties of this connection may give indication about the tissue that is being imaged. This directly links the Differential Geometry of medical imaging to Differential Geometry of continuous mechanics. The DTI imaging is a way to image the material connection, in this case the connection of the tissue.

3.6.2 Adding a metric structure

We consider a (pseudo-)metric g^M on the manifold M. We investigate the compatibility condition with respect to the connection.The connection ∇ is compatible with g^M if at any point $X \in M$ we have:

$$\nabla_V g^M = 0 \quad \forall U \in T_X M \tag{3.9}$$

i.e. if the pseudo-metric is parallel transported along any curves, or equivalently, if we have the product rule:

$$U < V, W >=< \nabla_U V, W > + < V, \nabla_U W > . \tag{3.10}$$

For a given pseudo-metric g^M, there is a unique symmetric (i.e. torsion free) connection compatible with it. It is called the Levi-Civita connection and can be computed with the Koszul formula as follow:

$$\begin{aligned} < \nabla_U V, W >= \frac{1}{2}(&U < V, W > + V < U, W > - W < U, V > \\ &- < U, [V, W] > - < V, [U, W] > + < W, [U, V] >) \end{aligned}$$

3.7 Sub-Riemannian manifolds

This section considers another metric structure on a manifold M, the structure defined on M by a sub-Riemannian metric.

An horizontal distribution on M is a subbundle HM of the tangent bundle of M. An horizontal distribution HM is called completely non-integrable if for $X \in M$, any tangent vector of $T_X M$ can be presented as a linear combination of vectors of the following types $U(X)$, $[U, V](X)$, $[U, [V, W]](X)$, etc, are in $T_X M$ where the vector fields $U, V, W, ...$ are in $H_X M$.

A sub-Riemannian manifold is a triple (M, H, g), where M is a differentiable manifold, H is a completely non-integrable horizontal distribution and g is a smooth section of positive-definite quadratic forms on H. Any sub-Riemannian manifold carries a natural intrinsic metric, called the metric of Carnot-Caratheodory [Postnikov 2001].

3.8 Conclusion

This Chapter has introduced the basic definitions of the main structures of Differential Geometry that can be used in Computational Medicine to describe organ shapes. We have chosen an intuitive approach and given examples with medical images. For each structure, we have given its relation and compatibility with the introduction of a Riemannian (pseudo-)metric.

The next part of this Ph.D thesis generalizes the statistics of Chapter 2 to the geometric structures seen in this Chapter.

Part II

Geometric Statistics applied to Computational Anatomy

Chapter 4

Geometric Statistics on Lie groups: analysis of anatomical deformations

This Chapter represents our first contribution with respect to the construction of Geometric Statistics. We consider the structure of Lie groups introduced in Chapter 3 and tackle the challenging problem of defining statistics on these spaces.

This Chapter has been published in the Entropy journal under the title: "Computing bi-invariant pseudo-metrics on Lie groups for consistent statistics". It completes the conference paper entitled "Statistics on Lie groups : a need to go beyond the pseudo-Riemannian framework" which has been presented at the workshop MaxEnt'2014..

4.1 Introduction

Modeling with Lie groups Data can be modeled as elements of Lie groups in many different fields: Computational Anatomy, Robotics, Paleontology etc. Indeed, Lie groups are continuous groups of transformations and thus appear naturally whenever one deals with articulated objects or shapes.

Regarding articulated objects, one can take examples in Robotics or in Computational Anatomy. In Robotics first, a spherical arm is obviously an articulated object. The positions of the arm can be modeled as the elements of the 3-dimensional Lie group of rotations $SO(3)$. In Computational Anatomy then, the spine can be modeled as an articulated object. In this context, each vertebra is considered as a orthonormal frame that encodes the rigid body transformation from the previous vertebra. Thus, as the human spine has 24 vertebrae, a configuration of the spine can be modeled as an element of the Lie group $SE(3)^{23}$, where $SE(3)$ is the Lie group of rigid body transformations in 3D, i.e. the Lie group of rotations and translations in $\mathbb{R}^3$ also called the Special Euclidean group.

Regarding shapes, the general model of d'Arcy Thompson suggests to represent shape data as the diffeomorphic deformations of a reference shape [Thompson 1992], thus as elements of an infinite dimensional Lie group of diffeomorphisms. This framework can be applied as well in Paleontology than in Computational Medicine. In Paleontology first, a monkey skull or a human skull can be modeled as the diffeomorphic deformation of a reference skull. In Computational Medicine then,

the shape of a patient's heart can be modeled as the diffeomorphic deformation of a reference shape. Obviously, many more examples could be given, also in other fields.

Statistics on Lie groups Once data are represented as elements of a Lie group, we may want to perform statistical analysis on them for prediction or quantitative modeling. Thus we want to perform statistics *on Lie groups*. How can we define an intrinsic statistical framework that is efficient on all Lie groups? How do we compute the mean or the principal modes of variation for a sample of Lie groups elements? In order to train our intuition, we consider finite dimensional Lie groups here.

To define a statistical framework, it seems natural to start with the definition of a mean. The definition of mean on a Lie group exemplifies the issues one can encounter while defining the whole statistical framework. We know that the usual definition of the mean is the weighted sum of the data elements of the sample. However, this definition is linear and Lie groups are not linear in general. Consequently, we cannot use this definition on Lie groups: we could get a mean of Lie group elements that is not a Lie group element. One can consider as an example the half sum of two rotations matrices that is not always a rotation matrix.

In fact, the definition of the mean on a Lie group should be *consistent with the group structure*. This consistence leads to several requirements of the mean, or properties. First, the mean of Lie group elements should be in the Lie group. Then, it seems natural to require that a left or right translation of the data set should translate its mean accordingly. The Figure (4.1) illustrates the case when this condition is fulfilled. Finally, the inversion of all data elements should lead to an inverted mean. A mean verifying all these properties is said *bi-invariant*.

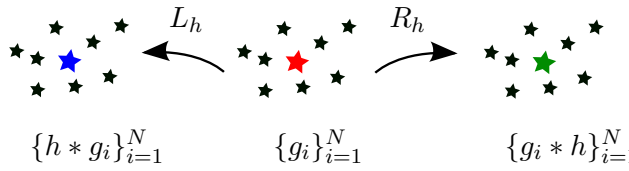


Figure 4.1: Left and right translation of a data set $\{g_i\}_{i=1}^N$ on the Lie group G. The initial data set $\{g_i\}_{i=1}^N$ has a mean represented in red. The left translated data set $\{h * g_i\}_{i=1}^N$ has a mean represented in blue. The right translated data set $\{g_i * h\}_{i=1}^N$ has a mean represented in green. We require that the mean of the (right or left) translated data set is the translation of the red mean, which is the case on this illustration: the blue mean is the left translation of the red mean, the green mean is the right translation of the red mean.

A naturally bi-invariant candidate for the mean on Lie groups is the *group exponential barycentre* [Pennec 2012] defined as follows. A group exponential barycenter m of the data set $\{g_i\}_{i=1,..,N}$ is a solution, if there are some, of the following *group*

barycenter equation:

$$\sum_{i=1}^{N} \mathrm{Log}(m^{(-1)} * g_i) = 0 \tag{4.1}$$

where Log is the group logarithm. As the group exponential barycenter is naturally bi-invariant, we call a group exponential barycenter a *bi-invariant mean*. The *local* existence and uniqueness of the bi-invariant mean have been proven if the dispersion of the data is small enough. "*Local*" means that the data are assumed to be in a sufficiently small normal convex *neighborhood* of some point of the Lie group.

Now we want to provide a computational framework for the bi-invariant mean that would set the foundations for computations on Lie groups statistics in general. For that, we are interested in characterizing the *global* domains of existence and uniqueness of the bi-invariant mean. By "*global domain*", we mean for example a ball of maximal radius such that any probability measure with support included in it would have a unique bi-invariant mean. Note that there is a priori no problem having several means, which can be called several "modes", or no mean at all. Our aim is rather to characterize the different situations that may occur: no mean, one unique mean, several means.

Using Riemannian and pseudo-Riemannian structures for statistics on Lie groups To this aim, we are interested in additional geometric structures on Lie groups that could help, by providing computational tools. For example we are interested in a distance on a Lie group, that could enable to *measure* radii of balls. Such a distance could obviously help characterizing balls of *maximal radius*.

But a Lie group is a group that carries an additional manifold structure and one can define a pseudo-metric on a manifold, making it a *pseudo-Riemannian manifold*. Thus, we can add a pseudo-metric on Lie groups which then induces a pseudo-distance. Could this additional pseudo-Riemannian structure help to define the statistical framework on Lie groups in practice?

We consider first the case of the Riemannian structure, i.e. when the pseudo-metric is in fact a metric (positive definite). Several definitions of the mean on *Riemannian* manifolds have been proposed in the literature: the *Fréchet mean*, the *Karcher mean* or the *Riemannian exponential barycentre* [Fréchet 1944, Fréchet 1948, Karcher 1977, Kendall 1990, Émery 1991, Corcuera 1998]. For example, the Riemannian exponential barycentres are defined as the critical points of the variance of the data, defined as: $\sigma^2(y) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{dist}(x_i, y)^2$, where $\{x_i\}_{i=1}^{N}$ are the data and dist the distance induced by the Riemannian metric. The Riemannian framework provides theorems for the *global* existence and uniqueness domains of this mean [Émery 1991, Corcuera 1998, Huiling 2004, Yang 2010, Afsari 2011], ensuring the computability of statistics on Riemannian manifolds. These represent exactly the kind of results we would like to have for the bi-invariant mean on Lie groups. Thus, one may wonder if we can apply this computational framework for statistics on Lie groups, and more particularly for the bi-invariant mean, by adding a Riemannian metric on the Lie group.

In fact, the notions of Riemannian mean and group exponential barycentre (or bi-invariant mean) coincide when the Riemannian metric is itself bi-invariant. In this case, the Riemannian geodesics coincide with the geodesics of the Cartan-Schouten connection [Sternberg 1964]. Thus, we can use the computational framework for Riemannian means only if we can add a bi-invariant metric on a Lie group.

But it is known that a Lie group does not have any bi-invariant Riemannian metric in general. The Lie group $ST(n)$ of Scalings and Translations of $\mathbb{R}^n$, the Heisenberg group H, the Lie group $UT(n)$ of Upper Triangular matrices of size $n \times n$ and the Lie group $SE(n)$ of rotations and translations of $\mathbb{R}^n$ do not have any bi-invariant metric while they admit a locally unique bi-invariant mean [Pennec 2012]. Therefore, if we want to characterize the bi-invariant mean with an additional geometric structure on Lie groups, we have to consider a structure that is more general than the Riemannian one.

The *pseudo*-Riemannian framework is a generalization of the Riemannian framework. Thus, it represents a tempting alternative for the characterization of the bi-invariant mean and for the definition of computational statistics on Lie groups in general. The pseudo-metric is not required to be positive definite anymore, only definite: the class of Lie groups that admit a bi-invariant *pseudo-metric* is larger than the class of those with a bi-invariant *metric*. Therefore, we could try to generalize the Riemannian statistical framework to a pseudo-Riemannian statistical framework and apply it for Lie groups. For instance, the mean on a pseudo-Riemannian manifold could still be defined as a critical point of the variance $\sigma^2(y) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{dist}(x_i, y)^2$, but dist would now be the pseudo-distance induced by the pseudo-metric. Of course, existence and uniqueness theorems would have to be re-established but we could get intuition from the Riemannian case.

In order to use the pseudo-Riemannian framework to characterize the bi-invariant mean, the first issue is: how many Lie groups do admit a bi-invariant pseudo-metric? Is it the case for the real Lie groups $ST(n)$, H, $UT(n)$ and $SE(n)$, which have a locally unique bi-invariant mean?

Lie groups and Lie algebras with bi-invariant pseudo-metrics If $\mathcal{G}$ is a connected Lie group, it admits a bi-invariant non degenerate symmetric bilinear form if and only if its Lie algebra admits a nondegenerate symmetric bilinear inner product, also called a bi-invariant pseudo-metric. Lie algebras with bi-invariant pseudo-metric were known to exist since the 1910's with the classification of simple Lie algebra [Cartan 1933] and the well-known Cartan-Killing form which is not degenerate in this case, but their specific study began in the 1950's with the works of [Tsou 1957, Tsou 1962]. Later, [Astrakhantsev 1978] started to study properties of these Lie algebras from their structural point of view and introduce the decomposability of indecomposability of these Lie algebras as a direct sum of ideals. However, the decomposition of [Astrakhantsev 1978] was not enough to characterize all Lie algebras with bi-invariant pseudo-metrics, as some authors [Medina 1985b, Keith 1984, Guts 1984] remark that the so-called oscillators algebra

arising in quantum mechanics carried a bi-invariant pseudo-metric without being decomposable in the sense of [Astrakhantsev 1978]. This leads, Medina and Revoy [Medina 1982, Medina 1985a] and Keith [Keith 1984] to build independently a classification of these Lie algebras, by showing that they all arise through direct sums and a structure called the double extension in [Medina 1982, Medina 1985a] and the bi-extension in [Keith 1984].

These results have been complemented by [Hofmann 1986], and then generalized by Bordemann to any non associative algebras with bi-invariant form through the T^*-extension structure [Bordemann 1997]. They have been completely described for certain dimensions in specific cases. The classification of the nilpotent quadratic Lie algebras of dimensions ≤ 7 is obtained in [Favre 1987], of the real solvable quadratic Lie algebras of dimensions ≤ 6 in [Campoamor-Stursberg 2008], the irreducible non solvable Lie algebras of dimensions ≤ 13 in [Benayadi 2014]. The specific cases of indecomposable quadratic Lie algebras with pseudo-metrics of different indices have been studied: bi-invariant pseudo-metrics of index 1 are described in [Medina 1985a, Hilgert 1985], of index 2 in [Kath 2004] and finally of general index in [Kath 2006]. The dimension of the space of bi-invariant pseudo-metrics has been studied in [Duong 2011] where bounds are provided.

Authors from other fields than pure algebra have also contributed to the study of bi-invariant pseudo-metrics. For example in Functionnal Analysis, Manin triple are a special type of Lie algebra with bi-invariant pseudo-metric that allow to interpret the solutions of the classical Yang-Baxter equation [Drinfeld 1987]. In this context, the Manin triple have been themselves classified for semi-simple Lie algebras in [Belavin 1998] and for complex reductive Lie algebra in [Delorme 2001].

Simultaneously, people started to gain interest in computational aspects on finite dimensional Lie algebras, implementing the identification of a Lie algebra from its structure constants given in any basis [Rand 1986, Rand 1988] or the Levi decomposition [Cohen 1997, Ronyai 1997]. The state-of-the-art regarding implementations on finite dimensional Lie algebra is summarized in [De Graaf 2000]. However, computations deal with the algebraic aspects of Lie algebras and, to the knowledge of the authors, do not consider metrics or pseudo-metrics.

Contributions and outline Our contribution is an algorithmic reformulation of a classification theorem for Lie algebras [Medina 1982, Medina 1985a] that answers these questions. More precisely, taking a Lie group $\mathcal{G}$ as input, the algorithm constructs a bi-invariant pseudo-metric on $\mathcal{G}$ in case of existence. Using this algorithm, we show that most Lie groups that have a locally unique bi-invariant mean do not possess a bi-invariant pseudo-metric. We conclude that, for the purpose of statistics on general real Lie groups -and more precisely for the computational framework of the bi-invariant mean- generalizing the Riemannian statistical framework to a pseudo-Riemannian framework may not be the optimal program.

The paper is organized as follows. In the first section, we introduce notions on quadratic Lie groups that will be useful for the understanding of the paper. In

the second section, we present the (tree-structured) algorithm that constructs bi-invariant pseudo-metrics on a given Lie group, in case of existence. In the third section, we apply the algorithm on $ST(n)$, H, $UT(n)$ and $SE(n)$ and show that most of them do not have any bi-invariant pseudo-metric.

4.2 Introduction to Lie groups with bi-invariant pseudo-metrics

Here, we define the algebraic and geometric notions which will be used throughout the paper.

4.2.1 Quadratic Lie groups and Lie algebras

In the following, we consider finite dimensional connected simply connected Lie groups over the field $\mathbb{F}$, where $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$.

Lie groups A *Lie Group* $\mathcal{G}$ is a smooth manifold with a compatible group structure. It is provided with an identity element e, a smooth composition law $* : (g, h) \mapsto g * h \in \mathcal{G}$ and a smooth inversion law $Inv : f \mapsto f^{(-1)} \in \mathcal{G}$. Its tangent space at g is written $T_g\mathcal{G}$.

The map $L_h : \mathcal{G} \ni g \mapsto h * g \in \mathcal{G}$ is the *left translation* by h and is a diffeomorphism of $\mathcal{G}$. Therefore, its differential (at g), $DL_h(g) : T_g\mathcal{G} \mapsto T_{L_h g}\mathcal{G}$ is an isomorphism that connects tangents spaces of $\mathcal{G}$. Similarly, one can define $R_h : \mathcal{G} \ni g \mapsto g * h \in \mathcal{G}$, the *right translation* by h.

A vector field X on $\mathcal{G}$ is *left invariant* if $(dL_h)(X(g)) = X(L_h(g)) = X(h * g)$ for each $g, h \in \mathcal{G}$. Similarly, one could define *right invariant vector fields.* The left invariant vector fields form a vector space which we denote $\Gamma(T\mathcal{G})^L$ and which is isomorphic to $T_e\mathcal{G}$. The Lie bracket of two left invariant vector fields is a left-invariant vector fields [Postnikov 2001].

Lie algebras As $\Gamma(T\mathcal{G})^L$ is closed under the Lie bracket of vector fields, we can look at $T_e\mathcal{G}$ as a Lie algebra. More precisely, we define $\mathfrak{g}$ the *Lie algebra of* $\mathcal{G}$ as $T_e\mathcal{G}$ with the Lie bracket induced by its identification with $\Gamma(T\mathcal{G})^L$. The Lie algebra essentially captures the local structure of the group. In the case of Lie algebras of matrices, the Lie bracket corresponds to the commutator. For a more complete presentation of Lie groups and Lie algebras, we refer the reader to [Bourbaki 1989].

Writing the expression of the Lie bracket $[,]_\mathfrak{g}$ on a given basis $\mathcal{B}_\mathfrak{g} = \{e_i\}_{i=1}^n$ of $\mathfrak{g}$, we define the *structure constants* f_{ijk} as:

$$[e_i, e_j]_\mathfrak{g} = f_{ijk} e_k \tag{4.2}$$

The structure constants f_{ijk} depend on the basis $\mathcal{B}_\mathfrak{g}$ chosen. They are always skew-symmetric *in the first two indices* but they may have additional symmetry properties if we write them in a well-chosen basis (see below). The structure constants f_{ijk}

completely determine the algebraic structure of the Lie algebra. Therefore, the structure constants are often the starting point, or the input, of algorithms on Lie algebras [Rand 1986, Rand 1988, Cohen 1997, De Graaf 2000]. It will also be the case for the algorithm we present in this paper.

Pseudo-metrics A *pseudo-metric* $<,>$ on $\mathcal{G}$ is defined as a smooth collection of definite inner products $<,>|_g$ on each tangent space $T_g\mathcal{G}$. Then, $\mathcal{G}$ becomes a *pseudo-Riemannian manifold*. A *metric* is defined as a pseudo-metric whose inner products are all positive definite. In this case, $\mathcal{G}$ is called a *Riemannian manifold*.

The *signature* (p, q) of a pseudo-metric is the number (counted with multiplicity) of positive and negative eigenvalues of the real symmetric matrix representing the inner product $<,>|_g$ at a point g and with respect to a basis of $T_g\mathcal{G}$. The signature is independent on the choice of the point g and on the basis at $T_g\mathcal{G}$. By definition, a pseudo-metric is definite, thus there is no null eigenvalues and we have $p+q=n$, where n is the dimension of $\mathcal{G}$. By definition, a metric is positive define and thus its signature is $(n, 0)$. Again, further details about such Differential Geometry can be found in [Postnikov 2001].

Quadratic Lie groups and algebras A *left-invariant pseudo-metric* is a pseudo-metric $<,>$ such that for all $X, Y \in T_g\mathcal{G}$ and for all $g, h \in \mathcal{G}$ we have:

$$< DL_h(g)X, DL_h(g)Y >|_{L_h g} =< X, Y >|_g \tag{4.3}$$

where L_h is the left translation by h. In other words, the left translations are isometries for this pseudo-metric. Similarly, we can define *right-invariant* and *bi-invariant* pseudo-metrics $<,>$. Note that any Lie group admits a left (or right) invariant pseudo-metric: we can define an inner product on the Lie algebra $\mathfrak{g} = T_e\mathcal{G}$ and propagate it on each tangent space $T_g\mathcal{G}$ through $DL_g(e)$ (or $DR_g(e)$). However, not any Lie group admits a bi-invariant pseudo-metric.

The Lie groups that admit a bi-invariant pseudo-metric are called *quadratic Lie groups*. The corresponding Lie algebras are called *quadratic Lie algebras*. Note that quadratic Lie groups or algebras are called differently in the literature. We find the appellation *metrizable* or *metrized* in [Tsou 1957, Tsou 1962, Astrakhantsev 1978], *metric* in [Kath 2004, Kath 2006], quasi-classical in [Campoamor-Stursberg 2008] and finally quadratic in [Favre 1987, Benayadi 2014].

Figure (4.2) shows a summary of the structures we just introduced.

We now recall that a non-degenerate bi-invariant inner product on a finite dimensional Lie algebra $\mathfrak{g}$ gives rise to a bi-invariant pseudo-metric on every Lie group whose Lie algebra is $\mathfrak{g}$ (see for example [Milnor 1976]). Therefore, we focus on Lie algebras from now on. We will still use the terms "pseudo-metric" or "metric" and the notation "$<,>$" in order to refer to the corresponding inner products on the Lie algebra $\mathfrak{g} = T_e\mathcal{G}$.

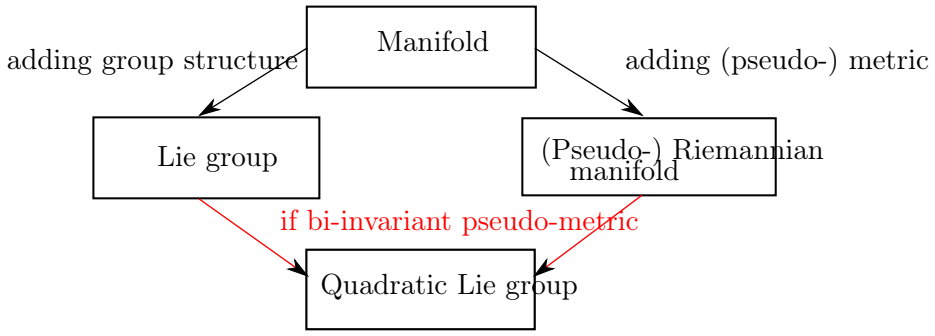


Figure 4.2: Algebraic and Geometric structures. If we require compatible algebraic and geometric structures on the manifold, we get a quadratic Lie group: a Lie group with a bi-invariant pseudo-metric.

Characterization of quadratic Lie algebras We give here different formulations of an equation characterizing a pair $(\mathfrak{g}, <,>)$ as a quadratic Lie algebra. A Lie algebra $\mathfrak{g}$ is quadratic if and only if it has a pseudo-metric $<,>$ verifying:

$$\forall x, y, t \in \mathfrak{g}, \qquad < [x, y]_{\mathfrak{g}}, t > + < y, [x, t]_{\mathfrak{g}} >= 0 \tag{4.4}$$

A proof for this characterization is given in [Postnikov 2001] and [Sternberg 1964].

First, taking advantage of the linearity in x, y, t we can rewrite Equation (4.4) on basis vectors. Let $\mathcal{B}_{\mathfrak{g}} = \{e_i\}_{i=1}^n$ be a basis of $\mathfrak{g}$, we consider: $x = e_i$, $y = e_j$ and $z = e_k$. Thus we can express the Lie bracket in terms of the structure constants and we get:

$$\forall i, j, k \in \{1, .., n\} \qquad f_{ijl} < e_l, e_k > + f_{jkl} < e_l, e_j >= 0 \tag{4.5}$$

In particular, we observe that the structure constants written in a basis orthonormal with respect to a bi-invariant *metric* are totally skew-symmetric. The structure constants written in a basis orthonormal with respect to a bi-invariant *pseudo*-metric will have additional symmetric properties as well.

Then, as we consider finite dimensional Lie groups, we can also rewrite Equation (4.4) in terms of matrices:

$$\forall x \in \mathfrak{g}, \quad A(x)^T.Z + Z.A(x) = 0 \tag{4.6}$$

where $A(x)$ is the matrix of the endomorphism denoted $[x, \bullet]$ defined as $y \mapsto [x, y]$, and Z a symmetric invertible (non necessarily positive) matrix representing $<,>$ on $\mathcal{B}_{\mathfrak{g}}$, the basis of $\mathfrak{g}$. Note that: $x \mapsto A(x)$ is itself linear.

Finally, taking advantage of the linearity again, and writing: $A(e_i) = A_i$, we can again reformulate Equation (4.4) and we get:

$$\forall i \in \{1, .., n\}, \quad A_i^T.Z + Z.A_i = 0 \tag{4.7}$$

which is now a linear system of n matrix equations. Note that Equation (4.5) corresponds to Equation (4.7) written in coordinates.

How to compute bi-invariant pseudo-metrics? Given a Lie algebra $\mathfrak{g}$ as input, we see now that the computation of bi-invariant pseudo-metrics on $\mathfrak{g}$ amounts to the resolution of the linear system of equations (4.7) for Z. The solutions of the linear system (4.7) form a vector space, which is called the *quadratic space* $\mathcal{Q}(\mathfrak{g})$ [Duong 2011]:

$$Q(\mathfrak{g}) = \{Z \in \mathrm{Sym}(n) \quad | \quad \forall i \in \{1, .., n\}, \quad A_i^T.Z + Z.A_i = 0\} \tag{4.8}$$

Obviously, the vector space $\mathcal{Q}(\mathfrak{g})$ contains invertible and non invertible solutions. Recalling the definition of a pseudo-metric, we emphasize that we will be interested in invertible solutions only.

In order to solve the system (4.7) for Z, i.e. to compute the quadratic space $Q(\mathfrak{g})$, we could adopt an analytic point of view. At i fixed, a single equation of the system (4.7) is a particular case of a Lyapunov equation that is studied in the context of control theory [Bartels 1972]. Thus, computational methods exist for studying *one* of our linear matrix equations [Kitagawa 1977]. For our purpose however, we want to *understand the structure* of a quadratic Lie group, in order to get an intuition for the generalization to infinite dimensional Lie groups of diffeomorphisms. Thus, we do not rely on an analytic point of view to solve the system (4.7).

We rather consider the whole system of equations (4.7) from an algebraic point of view. The pure algebraic point of view enables to solve the system (4.7) completely in most cases, like in the examples provided at the end of the paper. In the other cases, it leads to a smaller system of equations that can be solved analytically or computationally. Thus, the algebraic point of view provides not only a theoretical understanding of quadratic Lie groups, it also either solves the problem, or reduces the problem in order for the analytic point of view to solve it.

Therefore, we present in the next subsection the algebraic and geometric notions needed to set up, and later implement, the algebraic point of view.

4.2.2 Lie algebra representations

How can we *understand the structure* of a Lie algebra? An idea is to represent the Lie algebra elements as matrices acting on vectors. Then, the study of the behavior of these matrices helps to understand the Lie algebra as a whole. This is the purpose of the theory of Lie algebra representations, which we present briefly relying on [Cartan 1933, Medina 1985a, Bourbaki 1989, De Graaf 2000] in all this subsection.

Lie algebras representations A $\mathfrak{g}$-*representation* on the vector space V is a Lie algebra homomorphism $\eta : \mathfrak{g} \mapsto \mathfrak{gl}(V)$, that represents the elements of $\mathfrak{g}$ as matrices acting on the vector space V. The $\mathfrak{g}$-representations θ_1 and θ_2 are said to be *isomorphic* if there is an *isomorphism of representations* between them, i.e. an isomorphism of vector spaces $l : V_1 \mapsto V_2$ that verifies: $\theta_2(x) \circ l = l \circ \theta_1(x)$. We denote $\mathrm{Hom}_{\mathfrak{g}}(V_1, V_2)$ the vector space of isomorphisms of representations between V_1 and V_2.

In order to understand the representations of a Lie algebra $\mathfrak{g}$, and thus the Lie algebra $\mathfrak{g}$ itself, a strategy is to decompose the representations into smaller bricks, and then study those bricks. In this context, a $\mathfrak{g}$-*subrepresentation* of the $\mathfrak{g}$-representation V is a subspace of V stable by the elements of $\eta(\mathfrak{g})$. An *irreducible* $\mathfrak{g}$-*subrepresentation* is a $\mathfrak{g}$-subrepresentation without proper $\mathfrak{g}$-subrepresentation. An *indecomposable* $\mathfrak{g}$-*subrepresentation* is a $\mathfrak{g}$-subrepresentation that can not be decomposed into $\mathfrak{g}$-subrepresentations.

Note that irreducibility implies indecomposability but the converse is false: a $\mathfrak{g}$-representation can have a $\mathfrak{g}$-subrepresentation which does not have a supplementary that is also a $\mathfrak{g}$-subrepresentation (it would be "only" a vector space). Thus, it is not always possible to decompose a $\mathfrak{g}$-representation into irreducible $\mathfrak{g}$-subrepresentations, but only into indecomposable ones. In this context, a $\mathfrak{g}$-representation that can be decomposed into irreducible $\mathfrak{g}$-representations is called *completely reducible*.

Adjoint and co-adjoint representation We can choose the vector space V on which we represent $\mathfrak{g}$. Taking $V = \mathfrak{g}$, thus representing the Lie algebra on itself, we define the so-called *adjoint representation* of $\mathfrak{g}$, $\mathrm{ad} : \mathfrak{g} \ni x \mapsto \mathrm{ad}(x) = [x, \bullet]_{\mathfrak{g}} \in \mathfrak{gl}(\mathfrak{g})$. In its matrix version, we recognize the matrices A of the previous subsection. We see also that the set of matrices A_i defining the adjoint representation is equivalent to the set of structure constants of $\mathfrak{g}$.

We can rewrite again the Equation (4.4), but now in terms of the adjoint representation. We get:

$$\forall x, y, t \in \mathfrak{g}, \qquad < \mathrm{ad}(x).y, t > + < y, \mathrm{ad}(x).t >= 0 \tag{4.9}$$

Thus, the statement that $\mathfrak{g}$ is quadratic with bi-invariant pseudo-metric $<,>$ is equivalent to the requirement that all endomorphisms $\mathrm{ad}(x)$ are skew-symmetric endomorphisms with respect to $<,>$. Recalling the matrix version of Equation (4.4), that is Equation (4.6), we see that solving for a bi-invariant Z amounts to find a symmetric isomorphism of representations Z between the adjoint representation of $\mathfrak{g}$, written in its matrix form as $x \mapsto A(x)$, and the representation written in its matrix form as $x \mapsto -A(x)^T$.

If we choose to represent the Lie algebra $\mathfrak{g}$ on the dual vector space $\mathfrak{g}^*$, i.e. we choose $V = \mathfrak{g}^*$, we can define the *co-adjoint representation* θ: $\mathfrak{g} \ni x \mapsto \theta(x) \in \mathfrak{gl}(\mathfrak{g}^*)$ where $< \theta(x).f, t >=< f, \mathrm{ad}(x).t >$ for $f \in \mathfrak{g}^*$, $x, y \in \mathfrak{g}$ and $<,>$ the inner product used to define the dual basis. If we write $A(x)$ the matrix of the endomorphism $\mathrm{ad}(x)$, $T(x)$ the matrix of the endomorphism $\theta(x)$ and Z the inner product defining the dual basis, the previous definition states that Z is in fact an isomorphism of representation between the co-adjoint representation $x \mapsto T(x)$ and the representation: $x \mapsto A(x)^T$.

Now, if the inner product $<,>$ used to define the dual basis is bi-invariant, by identifying the vector spaces $\mathfrak{g}$ and $\mathfrak{g}^*$, we can again rewrite Equation (4.4) to get:

$$\forall x, y, t \in \mathfrak{g}, \qquad < \mathrm{ad}(x).y, t > + < \theta(x).y, t >= 0 \tag{4.10}$$

We conclude that the bi-invariance of the inner product implies the following relation between the adjoint and co-adjoint representations: $\mathrm{ad} = -\theta$. As Z (that represents $<,>$) is an isomorphism of representations between the co-adjoint and the representation $x \mapsto A(x)^T$, we recover that the statement of Z being a bi-invariant pseudo-metric on $\mathfrak{g}$ is equivalent to Z being a symmetric isomorphism of representations between $x \mapsto A(x)$ and $x \mapsto -A(x)^T$.

Some algebra vocabulary The adjoint representation is related to the structure constants of $\mathfrak{g}$ and thus completely characterizes $\mathfrak{g}$. Thus, it links the language of abstract algebras and the language of representations for $\mathfrak{g}$.

For the special case of the adjoint representation ad, $\mathfrak{g}$-subrepresentations are *ideals of* $\mathfrak{g}$, irreducible $\mathfrak{g}$-representations are *minimal ideals of* $\mathfrak{g}$, indecomposable $\mathfrak{g}$-representations are ideals of $\mathfrak{g}$ that can not be decomposed into a direct sum of ideals of $\mathfrak{g}$. We will use the two languages of ideals or of representations.

If the adjoint representation is itself irreducible but not 1-dimensional, $\mathfrak{g}$ is said to be *simple*. If the adjoint representation is completely reducible, $\mathfrak{g}$ is said to be *reductive*. If the adjoint representation is completely reducible without 1-dimensional subrepresentations, $\mathfrak{g}$ is *semi-simple*. If the adjoint representation is completely reducible with only 1-dimensional subrepresentations, $\mathfrak{g}$ is *abelian*. A reductive Lie algebra is thus the sum (in the sense of subrepresentations) of a semi-simple Lie algebra and an abelian Lie algebra.

Some vocabulary of geometry An ideal I of a Lie algebra B is said to be *isotropic* with respect to a pseudo-metric given on B if $I \cap I^\perp \neq \{0\}$. The ideal I is said to be *totally isotropic* if $I \subset I^\perp$. The intersection between I and $I^\perp$ represents the vectors that are orthogonal to themselves, and thus that have zero norm even if they are themselves non zero.

Thus, isotropic ideals appear only in a case of a pseudo-metric that is not a metric. From the intuition provided by Theoretical Physics, we can interpret the vectors in $I \cap I^\perp$ as photons: they have zero mass even if they have non-zero velocity.

4.2.3 Constructions with Lie algebra representations

We have seen that we can study the structure of a given Lie algebra by looking at its representations and more particularly at its adjoint representation. Here, we study decompositions of the adjoint representation that will be pertinent for the characterization of quadratic Lie algebras: the direct sum decomposition and the double extension decomposition. We show how these decompositions can be implemented in a computational framework. In this subsection, we use the notation $(B, [,]_B)$ to denote the Lie algebra, because this is the notation that we will use in the core of our algorithm (see Section 4.4).

Definition of direct sum $B = B_1 \oplus_B B_2$ is the *direct sum* of B_1, B_2 if:

- $B = B_1 \oplus B_2$ in terms of vector spaces,
- $[B, B_1]_B \subset B_1$, and $[B, B_2]_B \subset B_2$, making B_1 and B_2 subrepresentations of the adjoint representation of B, in other words: ideals of B.

This decomposition was first studied by [Astrakhantsev 1978]. We illustrate it with the matrices A representing the adjoint representation $b \mapsto [b, \bullet]_B$ of B, i.e. the matrices denoted: $b \mapsto A(b) = [b, \bullet]_B$. The direct sum of B is equivalent to the decomposition of the adjoint representation into the B-representations B_1 and B_2 i.e.:

$$A(b) = \begin{pmatrix} A(b_1) & 0 \\ 0 & A(b_2) \end{pmatrix} \tag{4.11}$$

on a basis respecting $B = B_1 \oplus_B B_2$. Note that we write $\oplus_B$ to emphasize the fact that this direct sum decomposition is more than the direct sum decomposition into vector spaces.

Direct sum decomposition and bi-invariant pseudo-metrics We have the following property: B being quadratic is equivalent to B_1 and B_2 being quadratic. Indeed, if $<,>_{B_1}$, $<,>_{B_2}$ are bi-invariant pseudo-metrics on B_1, B_2 and represented by the matrices Z_{B_1}, Z_{B_2} then:

$$Z_{B_1 \oplus_B B_2} = \begin{pmatrix} Z_{B_1} & 0 \\ 0 & Z_{B_2} \end{pmatrix} \tag{4.12}$$

is bi-invariant on B. And conversely, if $<,>_B$ is bi-invariant on B, its restrictions $<,>_B|_{B_1}$ and $<,>_B|_{B_2}$, are bi-invariant on B_1, B_2 [Medina 1982, Medina 1985a].

Computing the direct sum The direct sum decomposition of a Lie algebra B into indecomposable subrepresentations is unique, up to isomorphisms. In practice, writing $\mathcal{B}_B = \{e_k\}_{k=1}^{\dim(B)}$ a basis of B and $A_k = A(e_k)$, computing the direct sum decomposition of B into indecomposable B_i's amounts to the simultaneous bloc diagonalization of the matrices A_k.

Definition of double extension $B = W \oplus S \oplus S^*$ is the *double extension of W by a simple S* if:

- $B = W \oplus S \oplus S^*$ in terms of vector spaces,
- $(W, [,]_W)$ is a Lie algebra and $[S, W]_B \subset W$ makes W a S-representation,
- $(S, [,]_S)$ is a simple Lie subalgebra of B: $[s, s']_B = [s, s']_S$,
- S^* is the dual space of S and $[S, S^*]_B \subset S^*$ makes S^* the co-adjoint representation,

- $\forall w, w' \in W : \quad [w, w']_B = [w, w']_W + \beta(w, w')$ where $\beta : \Lambda^2 W \mapsto S^*$ is a (skew-symmetric) S-equivariant map, i.e. a map that commutes with the action of S.

This definition relies on the framework introduced in [Medina 1985a], or in [Keith 1984] under the appellation "bi-extension". Here, we can illustrate it with the matrices representing the adjoint representation $b \mapsto [b, \bullet]_B$ of B, i.e. the matrices denoted: $b \mapsto A(b)$. The double extension decomposition is equivalent to the following decomposition of the adjoint representation of B:

$$A(b) = \begin{pmatrix} [w, \bullet]_W + [s, \bullet]_B & [w, \bullet]_B & 0 \\ 0 & [s, \bullet]_S & 0 \\ \beta(w, \bullet) & [f, \bullet]_B & [s, \bullet]_B \end{pmatrix} \tag{4.13}$$

on a basis respecting $B = W \oplus S \oplus S^*$ and $b = w + s + f$. Note that, in the blocks of the matrix $A(b)$, we have identified endomorphisms with their corresponding matrices.

The definition of double extension uses a number of different notations. First, we recognize $\mathrm{ad}(s) = [s, \bullet]_S$ and $\mathrm{ad}(w) = [w, \bullet]_W$ to be respectively the adjoint representation of S (on S) and the adjoint representation of W (on W). But $[s, \bullet]_B$ is a S-representation *on* W that has nothing to do with the adjoint (the adjoint is a representation of a Lie algebra *on itself*).

Then, we should be careful with the structures that are manipulated. For example we can consider the vector space S^* as an abelian Lie subalgebra of B. But we can not consider W as a subalgebra of B. The skew-symmetric map β represents precisely the corresponding obstruction.

Double extension decomposition and bi-invariant pseudo-metrics We have the following property: B being quadratic is equivalent to W being quadratic. Indeed, if $<,>_W$ is bi-invariant on W, represented by Z_W, then:

$$Z_{W \oplus S \oplus S^*} = \begin{pmatrix} Z_W & 0 & 0 \\ 0 & 0 & \mathbb{I} \\ 0 & \mathbb{I} & 0 \end{pmatrix} \tag{4.14}$$

is bi-invariant on B. Conversely, if B is quadratic and written as a double extension of W with S simple (or 1-dimensional), then the restriction $<,>_W = <,>_B |_W$ is bi-invariant [Medina 1982, Medina 1985a]. Note here that we can write the $\mathbb{I}$-blocks because the basis of S and S^* are chosen to be duals of each other. If two different basis were chosen, the corresponding bi-invariant pseudo-metric on $B = W \oplus S \oplus S^*$ would have the form:

$$Z_{W \oplus S \oplus S^*} = \begin{pmatrix} Z_W & 0 & 0 \\ 0 & 0 & L \\ 0 & L^T & 0 \end{pmatrix} \tag{4.15}$$

with L an invertible matrix representing precisely the change of basis. More precisely, by computing Equation (4.6) on this last $Z_{W\oplus S\oplus S^*}$ while choosing $s \in S$, we show that L is necessarily an isomorphism of S-representations on S and I, i.e. $L \in \mathrm{Hom}_S(S, S^*)$. This remark will be used in practice in the algorithm (see Section 4.4).

Computing double extensions Contrary to *the* direct sum decomposition, the decomposition of a quadratic Lie algebra B as a double extension is not necessary unique. For example, given a quadratic indecomposable non simple B, we can build a double extension decomposition from each minimal ideal of B [Medina 1985a]. It proceeds as follows. We take a minimal ideal I of B and consider $I^\perp$ its orthogonal *with respect to a bi-invariant pseudo-metric* $<,>_B$. The decomposition:

$$B = W \oplus S \oplus S^* \qquad \text{where:} \qquad W = I^\perp/I, \quad S = B/I^\perp \text{ and } S^* = I$$

is a double extension of W with S simple (or 1-dimensional). Moreover, one can show that I and $I^\perp$ verify the following properties:

- I is abelian,
- $I^\perp$ is a maximal ideal,
- $I \subset I^\perp$ (total isotropy),
- $[I, I^\perp] = 0$ (commutativity),
- $\mathrm{codim}(I^\perp) = \dim(I)$.

These necessary conditions are taken from [Medina 1982, Medina 1985a, Astrakhantsev 1978].

In practice in our algorithm, we will have to build a double extension from a B *in order to compute a bi-invariant pseudo-metric on* B, if it exists (see Section 4.4). Therefore, even if we know an abelian minimal ideal I of B, we will not have its orthogonal $I^\perp$ needed for the construction shown above: we do not know any bi-invariant pseudo-metric as we want to build one! Thus, given an abelian minimal ideal I, we shall test all ideals J that could be an $I^\perp$ for a bi-invariant pseudo-metric, i.e. all ideals J that verify the necessary conditions listed above.

We show here that the only plausible ideals that can play the role of $I^\perp$ are, either $J = C_B(I)$ the centralizer of I in B in the case $C_B(I) \neq B$, or the maximal ideals of codimension 1 containing I in the case $C_B(I) = B$.

We have seen above that a first necessary condition for a J to be an $I^\perp$ is its commutativity with I: $[I, J] = 0$. We recall that the centralizer $C_B(I)$ of I in B is defined as the set of elements that commute with I. Thus: $J \subset C_B(I)$.

Another necessary condition for a plausible J is to be a maximal ideal. As I is an ideal, $C_B(I)$ is also an ideal. Thus J is a maximal ideal included in the ideal $C_B(I)$: we have necessarily $J = C_B(I)$ in the case $C_B(I) \neq B$. In this case, the

condition $I \subset J$ is fulfilled as I is abelian. The last necessary condition to check is $\text{codim}(C_B(I)) = \dim(I)$.

However if $C_B(I) = B$, then we shall look for maximal ideals of B. But in this case, I commutes with all elements of B and therefore I is necessarily of dimension 1 as a minimal ideal. Therefore we shall look for maximal ideals J of codimension 1. Adding the last necessary condition, we conclude that in the case $C_B(I) = B$, we shall consider only maximal ideals of codimension 1 containing I.

4.3 Structure of quadratic Lie groups

Here, we characterize the structure of quadratic Lie algebras, using the constructions defined in the previous section. We first present a reformulation of a classification theorem of quadratic Lie algebras. Then, we emphasize which Lie algebras we add by asking for a bi-invariant *pseudo-metric* instead of a bi-invariant *metric*. We finally investigate how we can go from a bi-invariant *pseudo-metric* to a bi-invariant *dual metric* on a special class of Lie algebra with bi-invariant pseudo-metrics.

4.3.1 A classification theorem

To characterize the structure of a quadratic Lie algebra, we use a reformulation of a classification theorem than can be found in [Medina 1985a] or [Keith 1984].

Theorem 4.3.1 (Classification of quadratic Lie algebras) *The Lie algebra $\mathfrak{g}$ is quadratic if and only if its adjoint representation decomposes into indecomposable subrepresentations B that are of the following types:*

- *Type (1): B is simple (or 1-dimensional),*
- *Type (2): $B = W \oplus S \oplus S^*$ is a double extension of a quadratic W by S simple (or 1-dimensional).*

This means that any quadratic Lie algebra writes $g = B_1 \oplus_{\mathfrak{g}} ... \oplus_{\mathfrak{g}} B_N$ where each B is of Type (1) or of Type (2). In particular, we can already conclude that any reductive (a fortiori, semi-simple or abelian) Lie algebra $\mathfrak{g}$ is quadratic. Moreover, if $\mathfrak{g}$ is quadratic but not reductive, then $\mathfrak{g}$ has non-irreducible indecomposable sub-representations, and these are necessarily double extensions of Type (2).

We recall that the notions of representation decomposition come from a simultaneous diagonalization of matrices. Therefore, they depend on the base field $\mathbb{F}$: a Lie algebra reductive in $\mathbb{R}$ is reductive in $\mathbb{C}$ but the converse is false. And thus, being quadratic also depends on the field we consider. A Lie algebra quadratic on $\mathbb{R}$ will be quadratic on $\mathbb{C}$ but the converse is false.

Elementary bi-invariant pseudo-metrics The previous characterization of quadratic Lie algebras in terms of their structure is useful in practice. It enables to construct a type of bi-invariant pseudo-metric $<,>_{\mathfrak{g}}$ that exists necessarily on a

quadratic $\mathfrak{g}$. We call this type of pseudo-metrics the *elementary bi-invariant pseudo-metrics* of $\mathfrak{g}$.

The elementary bi-invariant pseudo-metric $<,>_B$ of a 1-dimensional Lie algebra B is defined to be the multiplication. The elementary bi-invariant pseudo-metric $<,>_B$ of a simple Lie algebra B is defined to be the Killing form. Now let us define recursively the elementary bi-invariant pseudo-metrics of a general quadratic $\mathfrak{g}$.

Let us be given a quadratic Lie algebra $\mathfrak{g}$ on which we know an auxiliary bi-invariant pseudo-metric $<,>_\mathfrak{g}$ (not necessarily of the elementary type). First, we decompose the adjoint representation of $\mathfrak{g}$ into indecomposable subrepresentations B's: $\mathfrak{g} = B_1 \oplus_\mathfrak{g} ... \oplus_\mathfrak{g} B_N$. Then, we study separately the two cases: the B's of Type (1) and the B's of Type (2).

On the B's of Type (1), we define the elementary bi-invariant pseudo-metric $<,>_B$ as above: the multiplication if B is 1-dimensional or the Killing form if B is simple.

On the B's of Type (2), we build a double extension. To this aim, we consider a minimal ideal I and, *using the auxiliary bi-invariant pseudo-metric* $<,>_\mathfrak{g}$ *of* $\mathfrak{g}$, we compute $I^\perp$. We get the double extension $B = W \oplus S \oplus S^*$ with $W = I^\perp/I$, $S = B/I^\perp$ and $S^* = I$. We construct an elementary bi-invariant pseudo-metric $<,>_W$ on W recursively. We then define an elementary bi-invariant pseudo-metric $<,>_B$ on the double extension $B = W \oplus S \oplus S^*$ to be of the form of Equation (4.14).

Finally, we define the elementary bi-invariant pseudo-metric $<,>_\mathfrak{g}$ on the direct sum decomposition $\mathfrak{g} = B_1 \oplus_\mathfrak{g} ... \oplus_\mathfrak{g} B_N$ to be of the form of Equation (4.12). This construction defines (and proves the existence of) elementary bi-invariant pseudo-metrics on a quadratic $\mathfrak{g}$.

4.3.2 Riemannian and pseudo-Riemannian quadratic Lie groups

The previous characterization of quadratic Lie algebras can be refined to distinguish between quadratic Lie algebras that admit bi-invariant *metric* with respect to quadratic Lie algebras with bi-invariant *pseudo-metrics*. In other words, it answers the questions: which Lie algebras do we add by removing the positivity of the metric?

Studying the signature We recall from Section 4.2 that a metric on $\mathfrak{g}$ of dimension n has signature $(n, 0)$. Now, we take a quadratic $\mathfrak{g}$ that is decomposed into indecomposable pieces $\mathfrak{g} = B_1 \oplus_\mathfrak{g} ... \oplus_\mathfrak{g} B_N$, where the B_i are either simple (or 1-dimensional) or double extensions. The signature on the direct sum is the sum of the signatures on the B_i [Postnikov 2001]:

$$\text{sgn}_\mathfrak{g} = \text{sgn}_{B_1} + ... + \text{sgn}_{B_N} \tag{4.16}$$

Therefore, asking for a positive definite signature on $\mathfrak{g}$ is equivalent to asking for a positive definite signature on each of the B's.

If B is simple, it possess a bi-invariant metric if and only if it is compact. If B is a double extension, a bi-invariant pseudo-metric has necessary a non positive

definite signature of the form [Medina 1985a]:

$$\mathrm{sgn}_B = \mathrm{sgn}_W + (m, m) \tag{4.17}$$

where m is the dimension of the minimal ideal I used to build the double extension .

We conclude that $\mathfrak{g}$ admits a bi-invariant *metric* if and only if its indecomposable parts are simple compact or 1-dimensional, i.e. if and only if $\mathfrak{g}$ is reductive with compact simple parts.

Comparison The trees of Figure 4.3.2 and Figure 4.3.2 illustrate the comparison between Lie algebras with bi-invariant *metric* and Lie algebras with bi-invariant *pseudo-metric*.

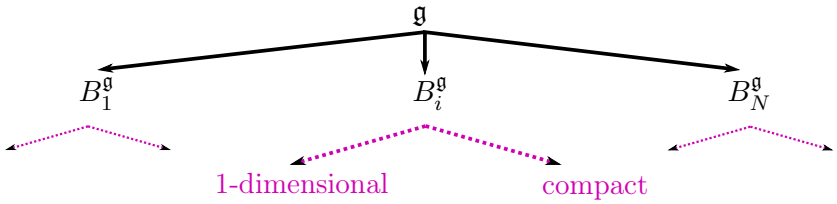


Figure 4.3: Structure of a Lie algebra with bi-invariant *metric*.

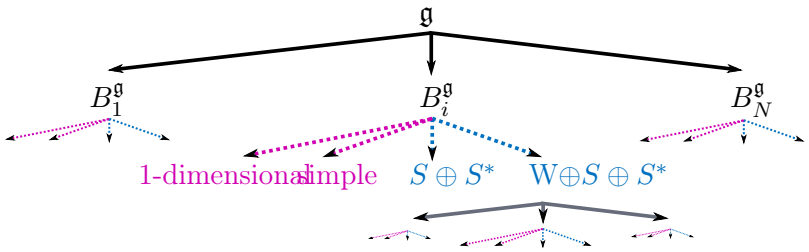


Figure 4.4: Structure of a Lie algebra with bi-invariant *pseudo-metric*.

Thus, going from Riemannian to pseudo-Riemannian enables to add the simple algebras that generalize the compact algebras and the double extension structures (in blue) with its recursive construction that is not present in the Riemannian case.

4.3.3 From a bi-invariant pseudo-metric to a bi-invariant dual metric?

We investigate here a special case of Lie algebras we gain by going from Riemannian to pseudo-Riemannian: the double extension of $W = \{0\}$ by a compact simple Lie algebra K, which is an example of Manin triple (see [Drinfeld 1987, Belavin 1998]). We will see in this subsection that we can view this case as a Riemannian case by changing the base field $\mathbb{F}$ (which is $\mathbb{R}$ or $\mathbb{C}$ for us) to its dual algebra $\mathbb{D}$. This development is a new contribution which is a justification and an extension of the dual quaternions for $SE(3)$.

Dual numbers and vectors Given a field $\mathbb{F}$, the *algebra* $\mathbb{D}$ *of dual numbers over this field* is defined as $\mathbb{D} = \mathbb{F} + \varepsilon\mathbb{F}$ where $\varepsilon^2 = 0$ and $\varepsilon \neq 0$ defines the multiplication [Grunwald 1906]. We can define a m-dimensional *dual vector space* $\mathbb{D}^m = \mathbb{F}^m + \varepsilon\mathbb{F}^m$, whose elements are *dual vectors*. Note here that the term "vector" is abusive in the sense that a vector space is usually defined on a field, not on an algebra. In the following, in order to study the properties of the dual vector space, we will use the *dual map* :

$$\begin{aligned} \psi : \quad & \mathbb{F}^m \oplus \mathbb{F}^m \mapsto \mathbb{D}^m \\ & x_0 + x_\varepsilon \mapsto x_0 + \varepsilon x_\varepsilon \end{aligned}$$

using the same notation ψ for mapping either to dual numbers or to dual vectors.

From the double extension $\mathfrak{g} = K \oplus K^*$ to its dual $\overline{\mathfrak{g}} = K + \varepsilon K^*$ Now we consider the double extension $\mathfrak{g} = K \oplus K^*$ where K is compact simple and $\dim(K) = m$ so that $\dim(\mathfrak{g}) = 2m$. We take the following elementary bi-invariant pseudo-metric on $\mathfrak{g}$:

$$Z_{K\oplus K^*} = \begin{pmatrix} \mathbb{I} & \mathbb{I} \\ \mathbb{I} & 0 \end{pmatrix} \tag{4.18}$$

As K and K^* have same $\mathbb{F}$-dimension m, we consider the dual space $\overline{\mathfrak{g}} = K + \varepsilon K^*$, of $\mathbb{D}$-dimension m. Its dual vectors write $\overline{x} = x_0 + \varepsilon.x_\varepsilon$ where $x_0 \in K$ and $x_\varepsilon \in K^*$.

Proposition 4.3.2 *The dual map:*

$$\begin{aligned} \psi : \quad & \mathfrak{g} = K \oplus K^* \mapsto \overline{\mathfrak{g}} \\ & x_0 + x_\varepsilon \mapsto x_0 + \varepsilon x_\varepsilon \end{aligned}$$

is an isomorphism of Lie algebras that respects the sum $K \oplus K^*$. The canonical inner product on $\overline{\mathfrak{g}}$ is bi-invariant and corresponds to the bi-invariant pseudo-metric $Z_{K\oplus K^*}$ above.

This can been shown as follows. First consider the Lie bracket on $\overline{\mathfrak{g}}$ inherited from ψ. We have:

$$\begin{aligned} [\psi(x), \psi(x')] &= [x_0 + \varepsilon x_\varepsilon, x'_0 + \varepsilon x'_\varepsilon] \\ &= [x_0, x'_0] + \varepsilon([x_0, x'_\varepsilon] + [x_\varepsilon, x'_0]) \qquad (\text{as } \varepsilon^2 = 0) \\ &= \psi([x, x']) \qquad (\text{definition of double extension}) \end{aligned}$$

which proves the isomorphism of Lie algebras.

We now show that the pseudo-metric $Z_{K\oplus K^*}$ on the Lie $\mathbb{F}$-algebra $\mathfrak{g}$ maps to the canonical metric $Z = \mathbb{I}$ on the Lie $\mathbb{D}$-algebra $\overline{\mathfrak{g}}$:

$$\begin{aligned} \psi(x)^T.\psi(x') &= (x_0 + \varepsilon.x_\varepsilon)^T.(x'_0 + \varepsilon x'_\varepsilon) \\ &= x_0^T.x'_0 + \varepsilon(x_\varepsilon^T.x'_0 + x_0^T.x'_\varepsilon) \\ &= \psi(x^T.Z_{K\oplus K^*}.x) \qquad (\text{using } \psi \text{ for dual numbers}) \end{aligned}$$

In others words, the spaces $\mathfrak{g}$ and $\overline{\mathfrak{g}}$ are isometric. But, again the term "isometric" is abusive, as we recall that $\mathfrak{g}$ and $\overline{\mathfrak{g}}$ are not defined on the same field - the latter being defined on an algebra.

Towards statistics on dual Riemannian manifolds We have shown that a double extension $\mathfrak{g} = K \oplus K^*$ of $W = \{0\}$ by a compact simple K, endowed with a bi-invariant *pseudo-metric* is isometrically isomorphic to a dual Lie algebra $\overline{\mathfrak{g}}$ with a bi-invariant *metric*. Thus, we could think of generalizing the theory of statistics on Riemannian manifolds to a theory of statistics on dual Riemannian manifolds. However, the fact that the space is defined on an algebra may cause some problems.

Generalization? One could wonder if we can use this construction for any general double extension. But we should note that this construction takes advantage of the fact that K^* is totally isotropic and abelian. The element ε such that $\varepsilon^2 = 0$ enables to represent the commutativity of K^* (Lie bracket is null) and the self orthogonality of K^* (inner product is null) at the same time. A general Lie algebra with bi-invariant pseudo-metric, is not necessarily decomposable into two subspaces of same dimension such that one of them is abelian and isotropic. For example, take a Lie algebra of odd dimension.

4.4 An algorithm to compute bi-invariant pseudo-metrics on a given Lie group

We go back to the general case of any quadratic Lie algebra over the field $\mathbb{F}$ ($\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$). We present in this section an algorithm that computes bi-invariant pseudo-metrics on a Lie algebra given as input.

Then, we show how one could generalize the algorithm to compute all bi-invariant pseudo-metrics on $\mathfrak{g}$. Finally, we apply the algorithm to some Lie groups known to possess a unique bi-invariant mean: we find that most of them are not quadratic.

4.4.1 The algorithm: computation of one bi-invariant pseudo-metric

For the computations, we will use matrix representations Z of pseudo-metrics $<,>$, where the basis will be specified. The input is $\mathcal{B}_\mathfrak{g} = \{e_i\}_{i=1}^n$, a basis of $\mathfrak{g}$ and the structure constants f_{ijk} on this basis. The output is a symmetric invertible matrix $Z_\mathfrak{g}$ on the basis $\mathcal{B}_\mathfrak{g}$, representing an elementary bi-invariant pseudo-metric, or a message of error: "The Lie algebra $\mathfrak{g}$ is not quadratic".

Core of the algorithm The core of the algorithm tests the structure of the Lie algebra given as input, to determine if it matches the characteristic tree-structure of quadratic Lie algebras described in the Section 4.3 (see Figure 4.3.2). Simultaneously with the progress through the tree, the algorithm tries to construct recursively an

elementary bi-invariant pseudo-metric $<,>_{\mathfrak{g}}$ by testing all possible candidates. If it succeeds, we return the bi-invariant elementary pseudo-metric, proving that $\mathfrak{g}$ is quadratic. If not, we conclude that $\mathfrak{g}$ is not quadratic and we return the error message. More precisely, the algorithm is divided in four steps as follows.

Step 1: Direct sum decomposition In this step, we decompose the adjoint representation of $\mathfrak{g}$ into indecomposable B's, in other words: we decompose $\mathfrak{g}$ as a direct sum of B's.

$$\mathfrak{g} = B_1 \oplus_{\mathfrak{g}} ... \oplus_{\mathfrak{g}} B_N \tag{4.19}$$

An implementation of this step can be found in [Rand 1988].

From now on, we work on the basis $\mathcal{B}'_{\mathfrak{g}}$ that respects the direct sum: $\mathfrak{g} = B_1 \oplus_{\mathfrak{g}} ... \oplus_{\mathfrak{g}} B_N$. The B's are indecomposable Lie algebras, thus we can take advantage of the classification theorem 4.3.1 of Section 4.3. In the following two steps, we test if each B is either of Type (1) (1-dimensional or simple) or of Type (2) (a double extension).

Step 2: Testing the Type (1) In this step, we test if the indecomposable B is of Type (1), i.e. if B is 1-dimensional or simple (see dichotomy of Theorem 4.3.1).

To test if B is 1-dimensional, we can obviously count the number of basis vectors of B in the basis $\mathcal{B}'_{\mathfrak{g}}$. If B is found 1-dimensional, we return the multiplication, which is an elementary bi-invariant pseudo-metric on B.

To test if B is simple, we use a function that computes the radical of the Levi decomposition of B [Levi 1905]. The indecomposable piece B is simple if and only if the radical is null. Such a function can be found in [Cohen 1997]. If B is found simple, we return the Killing form, which is an elementary bi-invariant pseudo-metric.

If B is neither 1-dimensional nor simple, we conclude that B is not of Type (1). We test in the following step if B is of Type (2).

Step 3: Testing the Type (2) In this step, we test if B is of Type (2), i.e. if B is a double extension of a quadratic W by a simple S (see dichotomy of Theorem 4.3.1). We recall that the double extension structure of B is not necessarily unique. Therefore it might seem that we need to test all possible candidates for a double extension structure of B, in order to answer if B is of Type (2). We proceed slightly differently.

As B is indecomposable and not of Type (1) (see previous steps), B being of Type (2) is equivalent to B being quadratic. More precisely, at this step of the algorithm, the following assertions are equivalents:

(a) B is of Type (2),

(b) B is quadratic,

(c) $\forall I$ minimal, I abelian, there is a double extension decomposition of B,

(d) $\exists I$ minimal, abelian such that there is a double extension decomposition of B.

Thus we will consider only one minimal ideal I of B and try to construct a double extension out of it, of the form: $B = W \oplus S \oplus I$. Note that this step will need

to call the algorithm recursively, to determine if the candidate for W in the double extension structure is quadratic or not. The details of this step are below.

Step 3.a First, we compute a minimal ideal I. More precisely, recalling the necessary conditions of the double extension structure of Section 4.2, we compute I an *abelian* minimal ideal, which is also a minimal abelian ideal. A function that finds a minimal abelian ideal of B can be derived from an algorithm of [Ceballos 2012], that computes all abelian ideals of B: we can choose one of minimal dimension among those.

Step 3.b Then, we compute $C_B(I)$, the maximal ideals J's and the corresponding candidates for the double extension structure of B. The computation of $C_B(I)$ is implemented in [Motsak 2006].

If $C_B(I) \neq B$, we take $J = C_B(I)$ and verify the condition $\text{codim}(J) = \dim(I)$. If the condition is not fulfilled, there is no double extension structure possible for B. Therefore, we conclude that B is not of Type(2).

If $C_B(I) = B$, we compute the maximal ideals J of B of codimension 1 containing I (see Section 4.2). If no such ideals are found, there is no double extension structure possible for B. Again in this case, we conclude that B is not of Type(2).

If J's are found, we compute the corresponding double extension candidates of B, one per J, as:

$$B = W \oplus S \oplus S^* \text{ where: } W = J/I,\ S = B/J \text{ and: } S^* = I. \tag{4.20}$$

We call the algorithm recursively on W, i.e. we determine recursively if W is quadratic. If there is no double extension candidate with a quadratic W, we conclude that B is not of Type (2). Otherwise, we keep the double extension candidates that have a quadratic W (with an elementary bi-invariant pseudo-metric Z_W).

Step 3.c Then, we try to compute an elementary pseudo-metric for all double extension candidates of the form: $B = W \oplus S \oplus S^*$ where $W = J/I$ is quadratic with corresponding Z_W, $S = B/J$ and $S^* = I$. Given a double extension candidate, we know from Section 4.2 that an elementary pseudo-metric on B has the form:

$$Z_{B=W\oplus S\oplus I} = \begin{pmatrix} Z_W & 0 & 0 \\ 0 & 0 & L \\ 0 & L^T & 0 \end{pmatrix} \tag{4.21}$$

where $L \in \text{Hom}_S(S, I)$.

Therefore we need to compute $\text{Hom}_S(S, I)$. We recall that S is simple thus its adjoint representation is irreducible. As we are in the case of a finite dimensional irreducible representation, we can apply Schur's Lemma. Its general form states that $\text{Hom}_S(S, S)$ is a associative division algebra over $\mathbb{F}$ ($= \mathbb{R}$ or $\mathbb{C}$), which is of finite degree because S is finite dimensional [Schur 1905]. When the base field is $\mathbb{F} = \mathbb{C}$, we use the fact that a finite-dimensional division algebra over an algebraically closed field is necessarily itself. Thus $\text{Hom}_S(S, S) = \mathbb{C}$ and $\dim_{\mathbb{C}}(\text{Hom}_S(S, S)) = 1$. When the base field is $\mathbb{F} = \mathbb{R}$, we use Frobenius theorem which asserts that the only

real associative division algebras are $\mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$, the field of quaternum numbers, [Frobenius 1878]. Thus $\mathrm{Hom}_S(S,S)$ is $\mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$, and $\dim_{\mathbb{R}}(\mathrm{Hom}_S(S,S))$ is 1, 2 or 4. Now if I and S are isomorphic, $\mathrm{Hom}_S(S,I)$ is isomorphic to $\mathrm{Hom}_S(S,S)$ and thus of maximal dimension 4 over $\mathbb{F}$. Otherwise, if I and S are not isomorphic, we have $\mathrm{Hom}_S(S,I) = \{0\}$.

The computation of $\mathrm{Hom}_S(S,I)$ is implemented in [Brooksbank 2008], more generally for any finite-dimensional modules of a finitely generated algebra.

Step 3.d To conclude Step 3, we determine if one of the possible elementary pseudo-metrics computed above is bi-invariant. To this aim, we plug the expression of $Z_{B=W\oplus S\oplus I}$ in Equation (4.7) and solve it for L. Thus, the initial system of Equations 4.7 has been reduced to an equation in maximum 1 (complex case) or in 4 (real case) parameters.

We run this step for each double extension candidate. If a bi-invariant elementary pseudo-metric Z_B is found on one of the candidates, we return Z_B. Otherwise, we conclude B is not of Type (2).

Step 4: Construction of a bi-invariant pseudo-metric on the whole $\mathfrak{g}$ In this step, we construct a bi-invariant (elementary) pseudo-metric on $\mathfrak{g}$, if it exists. If one B of the direct sum decomposition $\mathfrak{g} = B_1 \oplus_{\mathfrak{g}} ... \oplus_{\mathfrak{g}} B_N$ is neither of Type (1) nor of Type (2), we conclude from Theorem 4.3.1 that $\mathfrak{g}$ is not quadratic. We return the error message. Otherwise, we glue together the elementary bi-invariant pseudo-metrics Z_B's that have been returned on the B's.

More precisely, we follow the construction of Section 4.2 to build the elementary bi-invariant pseudo-metric $Z'_{\mathfrak{g}}$ on the basis $\mathcal{B}'_{\mathfrak{g}}$ of $\mathfrak{g}$ that respects the direct sum decomposition:

$$Z_{\mathfrak{g}=B_1\oplus_{\mathfrak{g}}...\oplus_{\mathfrak{g}}B_N} = \begin{pmatrix} Z_{B_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & Z_{B_N} \end{pmatrix} \tag{4.22}$$

Finally, we perform a change of basis from $\mathcal{B}'_{\mathfrak{g}}$ to $\mathcal{B}_{\mathfrak{g}}$ in order to return $Z_{\mathfrak{g}}$, an elementary bi-invariant pseudo-metric on the basis of the Lie algebra given as input.

Tree structure of the algorithm The algorithm has a natural tree structure presented in Figure 4.5. The bi-invariant pseudo-metric $Z_{\mathfrak{g}}$ is computed in a postfix manner. A tree level corresponds to a reduction of an adjoint representation: reduction of $\mathfrak{g}$ into B's for the first level, reductions of the W's into B's for the others. The arrows in dashes represent the cases we investigate to test if $\mathfrak{g}$ is quadratic. If B is not in one of such cases, then B is not quadratic, so neither is $\mathfrak{g}$ and we exit the algorithm.

In pseudo-code, the algorithm is written as follows.

This gives a bi-invariant pseudo-metric on the Lie algebra $\mathfrak{g}$. We can then make it a bi-invariant pseudo-metric on the Lie group $\mathcal{G}$ by propagating it through $DL_g(e)$ (or $DR_g(e)$) on all tangent spaces $T_g\mathcal{G}$ (see Section 4.2).

Input: $\mathcal{B}_{\mathfrak{g}} = \{e_i\}_i$ basis of $\mathfrak{g}$, Structure constants f_{ijk} on this basis.
Initialization $B = \mathfrak{g}$.
Core: Switch B:

- Case B is decomposable:
  - decompose into $B = B_1 \oplus_B ... \oplus_B B_N$,
  - call algorithm recursively on the B_i's;
  - return: $Z_B = \mathrm{Diag}(Z_{B_1}, \ldots, Z_{B_N})$.
- Case B is 1-dimensional: return $Z_B = \begin{pmatrix} 1 \end{pmatrix}$.
- Case B is simple: return $Z_B = Z_{\text{Killing}}$.
- Else:
  - compute I minimal abelian ideal; if no I exists: "EXIT".
  - compute its centralizer $C_B(I)$; if $codim(C_B(I)) \neq \dim(I)$: "EXIT".
  - compute $S = B/C_B(I)$, $W = C_B(I)/I$;
  - call algorithm recursively on $B = W$;
  - compute $\mathrm{Hom}_S(S, I)$;
  - solve Equation(4.7) for $L \in \mathrm{Hom}_S(S, I) = \{0\}$ by plugging: $Z_{B=W\oplus S\oplus I} = \begin{pmatrix} Z_W & 0 & 0 \\ 0 & 0 & L \\ 0 & L^T & 0 \end{pmatrix}$,
  - If there is no solution: "EXIT". Else, return the corresponding Z_B.

Output:

- if "EXIT": return the message "The Lie algebra $\mathfrak{g}$ is not quadratic";
- Else: return the elementary bi-invariant pseudo metric on $\mathfrak{g}$.

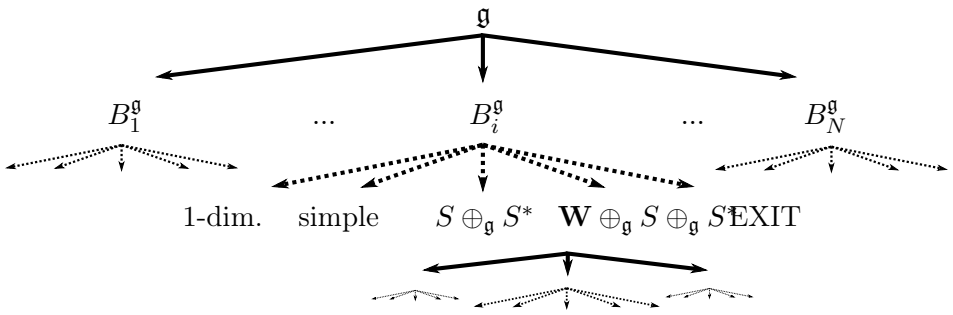


Figure 4.5: Tree structure of the algorithm.

All in all, the algorithm allows to compute *one* bi-invariant pseudo-metric of $\mathfrak{g}$, i.e. one invertible element of the quadratic space $\mathcal{Q}(\mathfrak{g})$. We can generalize the algorithm, in order to compute *all* bi-invariant pseudo-metrics of $\mathfrak{g}$, thus the whole quadratic space $\mathcal{Q}(\mathfrak{g})$. This is the purpose of the next subsection.

4.4.2 Generalization of the algorithm: computation of all bi-invariant pseudo-metrics

Here, we present how one should proceed in order to compute *all* bi-invariant pseudo-metrics of a given Lie algebra $\mathfrak{g}$, i.e. the whole quadratic space $\mathcal{Q}(\mathfrak{g})$. Note that the dimension of $\mathcal{Q}(\mathfrak{g})$ is unknown in the general case [Duong 2011]. However, the algorithmic procedure allows to compute the space anyway.

We follow the strategy of the previous algorithm: we decompose $\mathfrak{g}$ into indecomposable B's, we compute the quadratic spaces $\mathcal{Q}(B)$ for each of them and then glue these spaces together to get $\mathcal{Q}(\mathfrak{g})$.

Computing the quadratic space of indecomposable Lie algebras In this step, we compute the quadratic space for all indecomposable pieces B's of $\mathfrak{g}$, the simple (or 1-dimensional) and the double extensions.

The quadratic space of a 1-dimensional piece B is the weighted multiplication, so the whole base field $\mathbb{F}$:

$$\mathcal{Q}(B) = \{Z_B = \alpha \mathbb{I} \quad | \quad \forall \alpha \in \mathbb{F}\} = \mathbb{F} \tag{4.23}$$

The quadratic space of a simple piece B is the vector space spanned by the Killing form.

$$\mathcal{Q}(B) = \{Z_B = \alpha Z_{\text{Killing}} \quad | \quad \forall \alpha \in \mathbb{F}\} \tag{4.24}$$

The quadratic space of a double extension $B = W \oplus S \oplus S^*$, where the basis of

S and S^* are chosen duals, is given by:

$$\mathcal{Q}(B) = \left\{ Z_B = \begin{pmatrix} Z_W & M & N \\ M^T & \alpha Z_{Killing} & \beta\mathbb{I} \\ N^T & \beta\mathbb{I} & (0) \end{pmatrix} \;\middle|\; \begin{array}{c} \forall \alpha, \beta \in \mathbb{F}, \forall Z_W \in \mathcal{Q}(W), \\ \forall M, N \text{ solutions of equations derived from (4.7)} \end{array} \right\} \tag{4.25}$$

We leave to the reader the computations of the equations derived from Equation (4.7) that M and N are solving. Because of the dimension reduction, these equations can be solved in a lot of interesting cases. In our computations on selected Lie groups in the next subsection, N and N are vectors or scalars for example.

Computing the quadratic space of a direct sum The second step is the computation of the quadratic space of a direct sum $\mathfrak{g} = B_1 \oplus_{\mathfrak{g}} ... \oplus_{\mathfrak{g}} B_N$, given the quadratic spaces of each of its indecomposable pieces B_i. This gives:

$$\mathcal{Q}(\mathfrak{g}) = \left\{ Z_{\mathfrak{g}} \in \mathrm{Sym}(n) \quad | \quad \text{s.t. for } i \in \{1, ..N\} \text{ (block index):} \quad \begin{array}{rcll} Z_{\mathfrak{g}ii} & = & Z_{Bi} \in \mathcal{Q}(B_i) & \text{if } i = j \\ Z_{\mathfrak{g}ij} & = & M_{ij} \text{ if } i < j & \end{array} \right\} \tag{4.26}$$

where M_{ij} is a matrix that solves the following equation, derived from (4.7):

$$A(b_i)^T.M_{ij} + M_{ij}.A(b_j) = 0 \quad \forall b_i \in B_i, \forall b_j \in B_j \tag{4.27}$$

In summary, the problem of computing all bi-invariant pseudo-metrics of a given $\mathfrak{g}$ amounts to the resolution of a reduced number of algebraic equations of lower dimension.

4.4.3 Results of the algorithm on selected Lie groups

We run our algorithm *manually* to determine if a bi-invariant pseudo-metric exists on some real Lie groups for which there is a locally unique bi-invariant mean: $SE(n)$, $ST(n)$, H and $UT(n)$, for $n \in \mathbb{N}^*$ [Pennec 2012].

We run the computations manually and illustrate them, for each example, with the corresponding progress through the tree of the algorithm. The results show that most of these Lie groups are not quadratic.

Scalings and translations $ST(n)$ The Lie group $ST(n)$ comprises uniform scalings together with translations of $\mathbb{R}^n$. It is the semi-direct product $\mathbb{R}_+^* \ltimes \mathbb{R}^n$, its elements being written (λ, t). More precisely, $ST(n)$ is defined by its action on $\mathbb{R}^n$: $(\lambda, t).x = \lambda.x + t$. The group law and the group inversion are written as follows: $(\lambda_1, t_1) * (\lambda_2, t_2) = (\lambda_1.\lambda_2, \lambda_1 * t_2 + t_1)$ and $(\lambda, t)^{(-1)} = (1/\lambda, -t/\lambda)$.
The Lie algebra $\mathfrak{st}(n)$ comprises the $(\mu, u) \in \mathbb{R} \oplus \mathbb{R}^n$ with Lie Bracket:

$$[(\mu_1, u_1), (\mu_2, u_2)] = (0, \mu_2.u_1 - \mu_1.u_2). \tag{4.28}$$

Input We choose the basis $(D, \{P_a\}_{a=1}^n)$ defined as: $D = (1, 0)$ and $P_a = (0, e_a)$ with $(e_a)_{a=1}^n$ the canonical basis of $\mathbb{R}^n$. In this basis, the structure constants can be read in the following Lie brackets:

$$\begin{aligned}[P_a, P_b] &= 0,\\ [D, P_a] &= P_a,\\ [D, D] &= 0.\end{aligned}$$

Step 1 From the expression of the Lie brackets above, we can compute all ideals of $\mathfrak{st}(n)$ manually and find: Span(P_1), ..., Span(P_n) and their linear combinations. We remark that there is no ideal containing D. Thus $\mathfrak{st}(n)$ cannot be written as the direct sum of ideals, i.e. $\mathfrak{st}(n)$ is indecomposable.

Step 2 First, as $n \in \mathbb{N}^*$, we have $\dim(\mathfrak{st}(n)) > 1$. Thus $\mathfrak{st}(n)$ is not 1-dimensional. Then, as Span(P_1) for example is an ideal, $\mathfrak{st}(n)$ is not simple. We conclude that $\mathfrak{st}(n)$ is not of Type (1).

Step 3 We take $I = \text{Span}(P_1)$ which is obviously a minimal abelian ideal. From the commutation relations given by the Lie brackets, we see that $C_{\mathfrak{st}(n)}(I) = \text{Span}(\{P_a\}_{a=1}^n)$ and we are in the case $C_{\mathfrak{st}(n)}(I) \neq \mathfrak{st}(n)$. Thus there is only one double extension candidate, with $J = C_{\mathfrak{st}(n)}(I)$. We define $S = \mathfrak{st}(n)/J = \text{Span}(D)$ and $W = J/I = \text{Span}(P_2, ..P_n)$. We call the algorithm recursively on W, which decomposes into 1-dimensional ideals on which we return the multiplication.

The S-representation on S is the null representation: $[D, D] = 0$. The S-representation on I is the trivial representation: $[D, P_1] = P_1$. Hence, I and S are not isomorphic S-representations and $\text{Hom}_S(S, I)$ is zero. We conclude that $\mathfrak{st}(n)$ is not of Type (2).

Output We have found that $\mathfrak{st}(n)$ is indecomposable and neither of Type (1) nor of Type (2). Thus $\mathfrak{st}(n)$ is not quadratic: there is no bi-invariant pseudo-metric $<,>$ on $\mathfrak{st}(n)$.

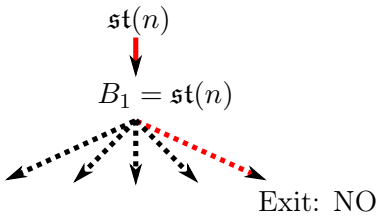


Figure 4.6: Schematical result for $ST(n)$. We see on the top level that $\mathfrak{st}(n)$ is indecomposable (it decomposes into itself). We see on the bottom level that $\mathfrak{st}(n)$ is neither 1-dimensional, nor simple, nor a double extension and therefore we exit the algorithm: $\mathfrak{st}(n)$ is not quadratic.

Heisenberg group H The Heisenberg group H comprises 3D upper triangular matrices M of the form:

$$M = \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}.$$

Thus, an element of this group can be written as $(x, y, z) \in \mathbb{R}^3$, with corresponding group law $(x_1, y_1, z_1) * (x_2, y_2, z_2) = (x_1 + x_2, y_1 + y_2, z_1 + z_2 + x_1 * y_2)$ and group inversion $(x, y, z)^{(-1)} = (-x, -y, -z + xy)$.

The Lie algebra $\mathfrak{h}$ comprises the nilpotent matrices:

$$N = \begin{pmatrix} 0 & p & c \\ 0 & 0 & q \\ 0 & 0 & 0 \end{pmatrix}.$$

Input A basis for $\mathfrak{h}$ is thus (P, Q, C) with clear notations. In this basis, the structure constants can be read in the following Lie brackets:

$$[C, P] = 0,$$
$$[C, Q] = 0,$$
$$[P, Q] = C.$$

Step 1 From the expression of the Lie brackets above, we can compute all ideals of $\mathfrak{h}$ manually and we find: $\mathrm{Span}(C)$, $\mathrm{Span}(C, P)$ and $\mathrm{Span}(C, Q)$. We remark that there is no ideal whose a supplementary is also an ideal. Thus $\mathfrak{h}$ is indecomposable.

Step 2 $\mathfrak{h}$ is obviously not 1-dimensional. Moreover, as $\mathrm{Span}(C)$ for example is an ideal, $\mathfrak{h}$ is not simple. We conclude that $\mathfrak{h}$ is not of Type (1).

Step 3 We take $I = \mathrm{Span}(C)$ which is a minimal abelian ideal of $\mathfrak{h}$. From the commutation relations given by the Lie brackets, we compute the commutator of I and we see that we are in the case $C_{\mathfrak{h}}(I) = \mathfrak{h}$. Thus we consider all maximal ideals of $\mathfrak{h}$ that are of codimension 1 and contain I. We get $J = \mathrm{Span}(C, P)$ or $J = \mathrm{Span}(C, Q)$, thus we have two double extension candidates. By symmetry in $P \leftrightarrow Q$ (see structures constants), we can consider $J = \mathrm{Span}(C, P)$ only, without lost of generality. We define $S = \mathfrak{h}/J = \mathrm{Span}(Q)$ and $W = J/I = \mathrm{Span}(P)$. We call the algorithm recursively on W. As W is 1-dimensional, W is quadratic and we return $Z_W = \begin{pmatrix} 1 \end{pmatrix}$.

The S-representation on S is given by the bracket $[Q, Q] = 0$: it is the null representation. The S-representation on I is given by the bracket $[Q, C] = 0$: it is also the null representation. The isomorphism of vector spaces L that maps C on Q is an isomorphism of representations whose matrix form is the identity in our basis. The dimension of $\mathrm{Hom}_S(S, I)$ is obviously 1.

Thus we plug:

$$Z_{W\oplus S\oplus I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

into Equation (4.6) to determine if it is bi-invariant. Computations show that it is not. We conclude that $\mathfrak{h}$ is not of Type (2).

Output We have found that $\mathfrak{h}$ is indecomposable and neither of Type (1) nor of Type (2). Thus $\mathfrak{h}$ is not quadratic: there is no bi-invariant pseudo-metric $<,>$ on $\mathfrak{h}$.

We try the algorithm on the general Heisenberg algebra $\mathfrak{h}_{2m+1}$ which is defined abstractly by the basis $\left\{C, \{P_i\}_{i=1}^m, \{Q_j\}_{j=1}^m\right\}$ and the Lie bracket:

$$\begin{aligned} [C, P_i] &= 0, \\ [C, Q_j] &= 0, \\ [P_i, Q_j] &= \delta_{ij} \end{aligned}$$

where δ is the Kronecker symbol. We are in the same situation as with $\mathfrak{h}$, except that W is abelian (but not necessarily 1-dimensional). We thus decompose W into abelian 1-dimensional ideals and we return the following elementary bi-invariant pseudo-metric:

$$Z_W = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

However, we exit the algorithm as previously. Thus the algorithm confirms that the general $\mathfrak{h}_{2m+1}$ has no bi-invariant pseudo-metric [Medina 1982].

The group of Scaled Upper Unitriangular Matrices $UT(n)$ The group $UT(n)$ comprises the upper triangular matrices M of the form: $M = \lambda.Id + N$ where $\lambda > 0$ and N an upper triangular nilpotent matrix.

The Lie algebra $\mathfrak{ut}(n)$ comprises the matrices of the form $X = \mu.Id + Y$ where $\mu \in \mathbb{R}$ and Y an upper triangular nilpotent matrix, the Lie bracket being the commutator of matrices.

Now $\mathfrak{ut}(n)$ is decomposable into the 1-dimensional Lie algebra generated by $\mathbb{I}$ and the Heisenberg algebra $\mathfrak{h}$. As $\mathfrak{h}$ has no bi-invariant pseudo-metric, neither does $\mathfrak{ut}(n)$.

Rigid Body Transformations $SE(n)$ The group of isometries $SE(n)$ comprises rotations together with translations of $\mathbb{R}^n$. It is the semi-direct product $SO(n) \ltimes \mathbb{R}^n$, its elements being written (R, t). More precisely, $SE(n)$ is defined by its action on $\mathbb{R}^n$ as $(R, t).x = R.x + t$. The group law and the group inversion are $(R_1, t_1) * (R_2, t_2) = (R_1.R_2, R_1 * t_2 + t_1)$ and $(R, t)^{(-1)} = (R^{(-1)}, R^{(-1)}.(-t))$.

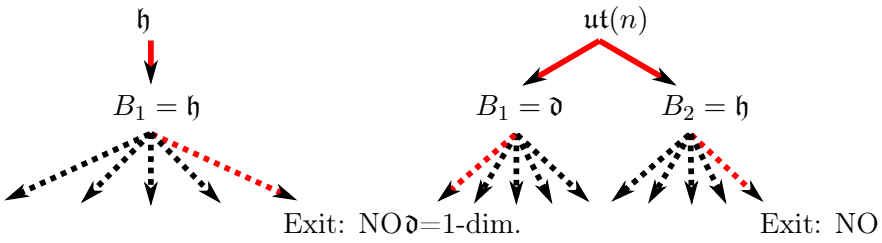


Figure 4.7: Schematical result for H and $UT(n)$. The top level indicates the direct sum decomposition step. Thus $\mathfrak{h}$ is indecomposable and $\mathfrak{ut}(n)$ decomposes into $\mathfrak{d}$ and $\mathfrak{h}$. The bottom level for $\mathfrak{h}$ indicates that $\mathfrak{h}$ is neither 1-dimensional, nor simple, nor a double extension and therefore we exit the algorithm: $\mathfrak{h}$ is not quadratic. The bottom level for $\mathfrak{ut}(n)$ indicates that $\mathfrak{d}$ is 1-dimensional and therefore quadratic but that $\mathfrak{h}$ is not quadratic: $\mathfrak{ut}(n)$ is not quadratic.

The Lie algebra $\mathfrak{se}(n)$ comprises the $(A, u) \in Skew(n) \oplus \mathbb{R}^n$ with Lie Bracket:

$$[(A_1, u_1), (A_2, u_2)] = (A_1.A_2 - A_2.A_1, A_1.u_2 - A_2.u_1) \tag{4.29}$$

Input We choose the basis: $(\{J_{ij}\}_{1 \leq i < j \leq n}, \{P_a\}_{a=1}^n)$ with $J_{ij} = e_i.e_j^T - e_j.e_i^T$ and $\{P_a\}_{a=1}^n$ the canonical basis of $\mathbb{R}^n$. In this basis, the structure constants can be read in the following Lie brackets:

$$\begin{aligned}
[J_{ij}, J_{kl}] &= \delta_{ik}.J_{jl} - \delta_{jk}.J_{il} + \delta_{jl}.J_{ik} - \delta_{il}.J_{jk}, \\
[J_{ij}, P_a] &= \delta_{aj}.P_i - \delta_{ai}.P_j, \\
[P_a, P_b] &= 0,
\end{aligned}$$

with δ the Kronecker symbol.

As preliminaries, we show that $P = \text{Span}(\{P_a\}_{a=1}^n)$ is the only proper ideal of $\mathfrak{se}(n)$. First, we see from the Lie brackets that P is a proper ideal of $\mathfrak{se}(n)$. Suppose that $\mathfrak{se}(n)$ has another proper ideal K. Then either $K \cap P$ is a proper ideal of $\mathfrak{se}(n)$ included in P or $K \subset \mathfrak{so}(n)$ is a proper ideal of $\mathfrak{se}(n)$. P does not contain any proper ideal of $\mathfrak{se}(n)$, because $\mathfrak{so}(n)$ acts transitively on P with the Lie bracket. We can show that $\mathfrak{so}(n)$ does not contain any proper ideal of $\mathfrak{se}(n)$ (considering independently the case $n = 4$). Thus P is the only proper ideal of $\mathfrak{se}(n)$.

Step 1 The Lie algebra $\mathfrak{se}(n)$ has only one ideal P. Thus $\mathfrak{se}(n)$ cannot be decomposed as a direct sum of ideals. We conclude that $\mathfrak{se}(n)$ is indecomposable.

Step 2 If $n = 1$, $\mathfrak{se}(1)$ is obviously 1-dimensional. We return the multiplication, which is a bi-invariant pseudo-metric on $\mathfrak{se}(1)$. Otherwise $\dim(\mathfrak{se}(n)) > 1$. As P is an ideal of $\mathfrak{se}(n)$, $\mathfrak{se}(n)$ is not simple. We conclude that $\mathfrak{se}(1)$ is quadratic with the multiplication as bi-invariant pseudo-metric and that $\mathfrak{se}(n)$ with $n > 1$ is not of Type(1). We go on with $n > 1$.

Step 3 We take $I = P$ and $J = C_{\mathfrak{se}(n)}(I) = P = I$. The necessary condition $\text{codim}(J) = \dim(I)$ is verified only for $n = 3$. We conclude that $\mathfrak{se}(n)$ is not of Type (2) if $n \neq 3$. We go on with $n = 3$. We compute $S = \mathfrak{se}(3)/P \sim \mathfrak{so}(3)$ and $W = P/P = \{0\}$.

In order to study the S-representations, we write the Lie bracket as:

$$\begin{aligned}[J_m, J_n] &= \varepsilon_{mnp}.J_p,\\ [J_m, P_a] &= \varepsilon_{map}.P_p,\\ [P_a, P_b] &= 0\end{aligned}$$

where we define $J_1 = J_{23}, J_2 = J_{31}$ and $J_3 = J_{12}$. The S-representation on S is the adjoint representation: $[J_m, J_n] = \varepsilon_{mnp}.J_p$. The S-representation on $I = P$ is given by: $[J_m, P_a] = \varepsilon_{map}.P_p$. It is also the adjoint representation. The isomorphism of vector spaces L that maps each P_a on J_a is an isomorphism of representations whose matrix form is the identity in our basis.

Hence we write $Z_{\mathfrak{se}(3)}$ on the decomposition $S \oplus I = \mathfrak{so}(3) \oplus P$ with basis $(\{J_a\}_{a=1}^3, \{P_a\}_{a=1}^3)$ and get:

$$Z_{\mathfrak{se}(3)} = \begin{pmatrix} 0 & \mathbb{I}_3 \\ \mathbb{I}_3 & 0 \end{pmatrix}. \tag{4.30}$$

We plug it into Equation (4.7). Running the computation shows that the pseudo-metric $Z_{\mathfrak{se}(3)}$ is bi-invariant on $\mathfrak{se}(3)$. $Z_{\mathfrak{se}(3)}$ is actually known as the Klein form [Karger 1985].

Output $\mathfrak{se}(1)$ is quadratic, we return the multiplication which is a bi-invariant pseudo-metric on $\mathfrak{se}(1)$. $\mathfrak{se}(3)$ is quadratic, we return the Klein form which a bi-invariant pseudo-metric on $\mathfrak{se}(3)$. Otherwise, $\mathfrak{se}(n)$ is indecomposable and neither of Type (1) nor of Type (2): it is not quadratic.

not quadratic.

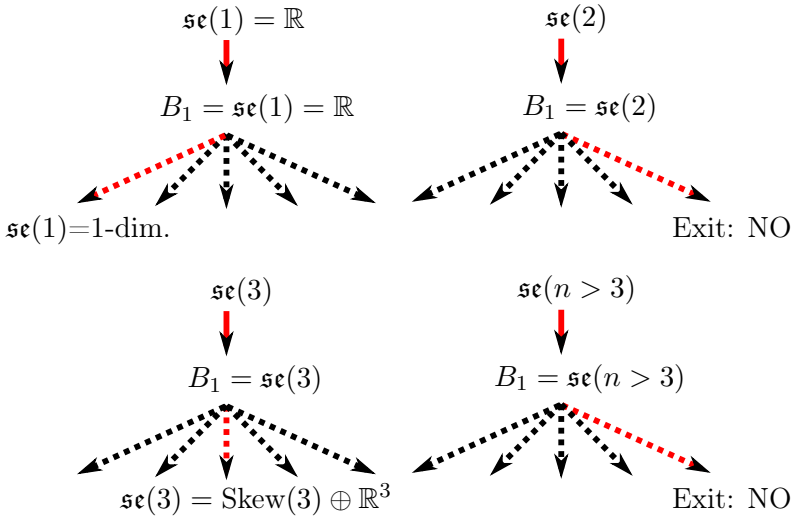


Figure 4.8: Schematical result for $SE(n)$. We recover the different cases depending on n.

We can build the whole quadratic space of $\mathfrak{se}(3)$. This gives the 2-dimensional

vector space:

$$\mathcal{Q}(\mathfrak{se}(3)) = \left\{ \begin{pmatrix} \alpha Z_{\text{Killing}} & \beta.\mathbb{I} \\ \beta.\mathbb{I} & 0 \end{pmatrix} \quad \middle| \quad \forall \alpha, \beta \in \mathbb{F} \right\} \tag{4.31}$$

Moreover, we have recognized in $\mathfrak{se}(3)$ the special case of a double extension $K \oplus K^*$ of $W = \{0\}$ by a compact Lie algebra $K = \mathfrak{so}(3)$. Therefore, the dual structure presented in Section 4.3 can be used in practice. We recall that we can represent the elements of $SO(3)$ as unit quaternions. Thus, we can represent the elements of $SE(3)$ as unit dual quaternions [Kenwright 2012]. A generalization of the theory of Riemannian statistics to a theory of *dual Riemannian statistics* would thus be useful for rigid body transformations, which are present in many different fields.

4.5 Conclusions

In this paper, we have presented an algorithmic method to compute a bi-invariant pseudo-metric on a Lie group, in the case of existence. The method allows to test simultaneously if the Lie group given as input is quadratic or not. We indicated how to compute all the bi-invariant pseudo-metrics on the given Lie group. First, the algorithm by itself represents a contribution to the field of computational Lie algebra.

Then, regarding statistics on Lie groups which was our original motivation, we see two consequences of this article. First, it enables to distinguish, from a practical point of view, Lie groups on which a future pseudo-Riemannian theory of statistics could be used and implemented. This is the case of $SE(3)$, the Lie group of rotations and translations of the 3D space which is found in various fields.

Second, this paper shows that a general Lie group with bi-invariant mean does not admit a bi-invariant metric. Therefore, if one wants to define a general theory of statistics that works for all Lie groups, one needs to find a geometric framework beyond the Riemannian and the pseudo-Riemannian ones.

Chapter 5

Geometric Statistics on quotient spaces: bias in anatomical template shape estimation

In Chapter 4, we have investigated the definition of a statistical theory on Lie groups. We have referred to the definition of the sample mean on Lie groups, one of the simplest estimators, in order to illustrate our point.

In this Chapter, we also consider a sample mean estimator, but not on Lie groups anymore. We consider the sample mean defined as the Fréchet mean on a quotient space. We use an generalization of estimation theory of Chapter 2 to study the properties of this sample mean and in particular its bias. We adapt techniques of Chapter 2 to quotient spaces, in order to improve this estimator.

This Chapter has been submitted to SIAM journal Imaging Science under the title: "Template shape estimation: correcting an asymptotic bias". It completes the conference paper "Biased estimators on quotient spaces" which has been presented to Geometric Science of Information 2015.

Introduction

The shape of a set of points, the shape of a signal, the shape of a surface, or the shapes in an image can be defined as the remainder after we have filtered out the position and the orientation of the object [Kendall 1984]. Statistics on shapes appear in many fields. Paleontologists combine shape analysis of monkey skulls with ecological and biogeographic data to understand how the *skull shapes* have changed in space and time during evolution [Elewa 2012]. Molecular Biologists study how *shapes of proteins* are related to their function. Statistics on misfolding of proteins is used to understand diseases, like Parkinson's disease [Li 2008]. Orthopaedic surgeons analyze *bones' shapes* for surgical pre-planning [Darmanté 2014]. In Signal processing, the *shape of neural spike trains* correlates with arm movement [Kurtek 2011]. In Computer Vision, classifying *shapes of handwritten digits* enables automatic reading of texts [Allassonnière 2015b]. In Medical Imaging and more precisely in Neuroimaging, studying *brain shapes* as they appear in the MRIs facilitates discoveries on diseases, like Alzheimer [Lorenzi 2011].

What do these applications have in common? Position and orientation of the skulls, proteins, bones, neural spike trains, handwritten digits or brains do not

matter for the studies' goal: only *shapes* matter. Mathematically, the study analyses the statistical distributions of *the equivalence classes of the data* under translations and rotations. They project the data in a quotient space, called the *shape space*.

The simplest - and most widely used - method for summarizing shapes is the computation of the mean shape. Almost all neuroimaging studies start with the computation of the mean brain shape [Evans 2012] for example. One refers to the mean shape with different terms depending on the field: mean configuration, mean pattern, template, atlas, etc. The mean shape is an average of *equivalence classes of the data*: one computes the mean after projection of the data in the shape space. One may wonder if the projection biases the statistical procedure. This is a legitimate question as any bias introduced in this step would make the conclusions of the study less accurate. If the mean brain shape is biased, then neuroimaging's inferences on brain diseases will be too. This paper shows that a bias is indeed introduced in the mean shape estimation under certain conditions.

Related work

We review papers on the shape space's geometry as a quotient space, and existing results on the mean shape's bias.

Shapes of landmarks: Kendall analyses The theory for shapes *of landmarks* was introduced by Kendall in the 1980's [Kendall 1977]. He considered shapes of k labeled landmarks in $\mathbb{R}^m$. The size-and-shape space, written $S\Sigma_m^k$, takes also into account the overall size of the landmarks' set. The shape space, written Σ_m^k, quotients by the size as well. Both $S\Sigma_m^k$ and Σ_m^k have a Riemannian geometry, whose metrics are given in [Le 1993]. These studies model the probability distribution of the data directly in the shape space Σ_m^k. They do not consider that the data are observed in the space of landmarks $(\mathbb{R}^m)^k$ and projected in the shape space Σ_m^k. The question of bias is not raised.

We emphasize that there is a distinction between "form" and "shape" in the literature. "Form" relates to the quotient of the object by rotations and translations only. "Shape" denotes the quotient of the object by rotations, translations, and scalings. Kendall shape spaces refer to "shape": the scalings are quotiented by constraining the size of the landmarks' set to be 1.

Shapes of landmarks: Procrustean analyses Procrustean analysis is related to Kendall shape spaces but it also considers shapes of landmarks [Goodall 1991, Dryden 1998, Gower 2004]. Kendall analyses project the data in the shape space by explicitly computing their coordinates in Σ_m^k. In contrast, Procrustean analyses keep the coordinates in $(\mathbb{R}^m)^k$: they project the data in the shape space by "aligning" or "registering" them. Orthogonal Procrustes analysis "aligns" the sets of landmarks by rotating each set to minimize the Euclidean distance to the other sets. Procrustean analysis considers the fact that the data are observed in the space $(\mathbb{R}^m)^k$ but does not consider the geometry of the shape space.

The mean "shape" was shown to be consistent for shapes of landmarks in 2D and 3D in [Lele 1993, Le 1998]. Such studies have a generative model with a scal-

ing component α and a size constraint in the mean "shape" estimation procedure, which prevents the shapes from collapsing to 0 during registration. In constrast, the mean "form" - i.e. without considering scalings - is shown to be inconsistent in [Lele 1993] with an reducto ad absurdum proof. However, this proof does not provide any geometric intuition about how to control or correct the phenomenon. More recently, similar inconsistency effects have been observed in [Du 2015], showing that implementing ordinary Procrustes analysis without taking into account noise on the landmarks may compromise inference. The authors propose a conditional scoring method for matching configurations in order to guarantee consistency.

Shapes of curves Curve data are projected in their shape space by an alignment step [Joshi 2006], in the spirit of a Procrustean analysis. The bias of the mean shape is discussed in the literature. Unbiasedness was shown for shapes of signals in [Kurtek 2011] but under the simplifying assumption of no measurement error on the data. Some authors provide examples of bias when there is measurement error [Allassonnière 2007]. Their experiments show that the mean signal shape may converge to pure noise when the measurement error on simulated signals increases. The bias is proven in [Bigot 2011] for curves estimated from a finite number of points in the presence of error. But again, no geometric intuition nor correction strategy is given.

Abstract shape spaces [Huckemann 2010] studies statistics on abstract shape spaces: the shapes are defined as equivalence classes of objects in a manifold M under the isometric action of a Lie group G. This unifies the theory for shapes of landmarks, of curves and of surfaces described above. [Huckemann 2010] introduces a generalization of Principal Component Analysis to such shape spaces and does not compute the mean shape as the 0-dimensional principal subspace. Therefore, the bias on the mean shape is not considered.

But in the same abstract setting, [Miolane 2015a] shows the bias of the mean shape, in the special case of a finite-dimensional flat manifold M. The authors emphasize how the bias depends on the noise σ on the measured objects, more precisely on the ratio of σ with respect to the overall size of the objects. [Allassonnière 2016] also presents a case study for an infinite dimensional flat manifold M quotiented by translations, where the noise σ is one of the crucial variables controlling the bias. However, the case of general curved manifolds M has not been investigated yet.

Contributions and outline

We are still missing a *global geometric* understanding of the bias. Which variables control its magnitude? Is it restricted to the mean shape or does it appear for other statistical analyses? How important is it in practice: do we even need to correct it? If so, how can we correct it? Our paper addresses these questions. We use a geometric framework that unifies the cases of landmarks, curves, images etc.

Contributions We make three contributions. First, we show that statistics on shapes are biased when the data are measured with error. We explicitly compute

the bias in the case of the mean shape. Formulated in the Procrustean terminology, our result is: the Generalized Procrustes Analysis (GPA) estimator of mean "form" is asymptotically biased, because we do not consider scalings. Second, we offer an interpretation of the bias through the geometry of the shape space. In applications, this aids in deciding when the bias can be neglected in contrast with situations when it must be corrected. Third, we leverage our understanding to suggest several correction approaches.

Outline The paper has four Sections. Section 1 introduces the geometric framework of shape spaces. Section 2 presents our first two contributions: the proof and geometric interpretation of the bias. Section 3 describes the procedures to correct the bias. Section 4 validates and illustrates our results on synthetic and real data.

5.1 Geometrization of template shape estimation

5.1.1 Two running examples

We introduce two simple examples of shape spaces that we will use to provide intuition.

First, we consider two landmarks in the plane $\mathbb{R}^2$ (Figure 7.1 (a)). The landmarks are parameterized each with 2 coordinates. For simplicity we consider that one landmark is fixed at the origin on $\mathbb{R}^2$. Thus the system is now parameterized by the 2 coordinates of the second landmark only, e.g. in polar coordinates (r, θ). We are interested in the shape of the 2 landmarks, i.e. in their distance which is simply r.

Second, we consider two landmarks on the sphere S^2 (Figure 7.1 (b)). One of the landmark is fixed at the north pole of S^2. The system is now parameterized by the 2 coordinates of the second landmark only, i.e. (θ, ϕ), where θ is the latitude and ϕ the longitude. The shape of the two landmarks is the angle between them and is simply θ.

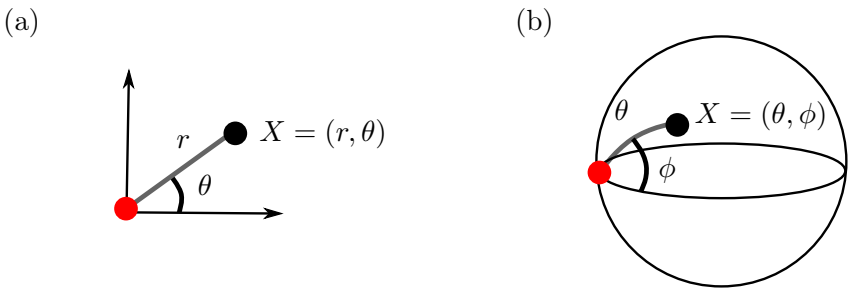


Figure 5.1: Two landmarks, one in red and one in black, on the plane $\mathbb{R}^2$ (a) and on the sphere S^2 (b). The landmark in red is fixed at the origin of the coordinates. The system is entirely represented by the coordinates X of the landmark in black.

5.1.2 Differential Geometry of shapes

5.1.2.1 The shape space is a quotient space

The data are objects $\{X_i\}_{i=1}^n$ that are either sets of landmarks, curves, images, etc. We consider that each object X_i is a point in a Riemannian manifold M. In this paper, we restrict ourselves to finite dimensional manifolds. We have $M = \mathbb{R}^2$ in the plane example: a flat manifold of dimension 2. We have $M = S^2$ in the sphere example: a manifold of constant positive curvature and of dimension 2.

By definition, the objects' shapes are their equivalence classes $\{[X_i]\}_{i=1}^n$ under the action of some finite dimensional Lie group G: G is a group of continuous transformations that models what does not change the shape. The action of G on M will be written with "$\cdot$". In our examples, the rotations are the transformations that leave the shape of the systems invariant. Let us take g a rotation. The action of g on the landmark X is illustrated by a blue arrow in Figures 7.2 (a) for the plane and (d) for the sphere. We observe that the action does not change the shape of the systems: the distance between the two landmarks is preserved in (a), the angle between the two landmarks is preserved in (d). The equivalence class of X_i is also called its orbit and written O_{X_i}. The equivalence class/orbit of X is illustrated with the blue dotted circle in Figure 7.2 (a) for the plane example and in Figure 7.2 (d) for the sphere example. The orbit of X in M is the submanifold of all objects in M that have the same shape as X. The curvature of the orbit as a submanifold of M is the key point of the results in Section 6.4.1.

The *shape space* is by definition the space of orbits. This is a quotient space denoted $Q = M/G$. One orbit in M, i.e. one circle in Figure 7.2 (b) or (e), corresponds to a point in Q. The shape space is $Q = \mathbb{R}_+$ in the plane example. This is the space of all possible distances between the two landmarks, see Figure 7.2 (c). The shape space is $Q = [0, \pi]$ in the sphere example. This is the space of all possible angles between the two landmarks, see Figure 7.2 (f).

5.1.2.2 The shape space is a metric space

We consider that the action of G on M is *isometric with respect to the Riemannian metric of* M. This implies that the distance d_M between two objects in M does not change if we transform both objects in the same manner. In the plane example, rotating the landmark X_1 and another landmark X_2 with the same angle does not change the distance between them.

The distance in M induces a quasi-distance d_Q in Q: $d_Q(O_{X_1}, O_{X_2}) = \inf_{g \in G} d_M(g \cdot X_1, X_2)$ [Huckemann 2010]. The distance between the shapes of X_1 and X_2 is computed by first registering/aligning X_1 onto X_2 by the minimizing g, and then using the distance in the ambient space M. In the plane example, the distance between two shapes is the difference in distances between the landmarks. One can compute it by first aligning the landmarks, say on the first axis of $\mathbb{R}^2$, then one uses the distance in $\mathbb{R}^2$.

5.1.2.3 The shape space has a dense set of principal shapes

The *isotropy group of* X_i is the subgroup of transformations of G that leave X_i invariant. For the plane example, every $X_i \neq (0,0)$ has isotropy group the identity and $(0,0)$ has isotropy group the whole group of 2D rotations. Objects on the same orbit, i.e. objects that have the same shape, have conjugate isotropy groups.

Principal orbits or principal shapes are orbits or shapes with smallest isotropy group conjugation class. In the plane example, $\mathbb{R}^2 \setminus (0,0)$ is the set of objects with principal shapes. Indeed, every X in $\mathbb{R}^2 \setminus (0,0)$ belongs to a non-degenerate circle centered at $(0,0)$ and has isotropy group the identity. The set of principal shapes corresponds to $\mathbb{R}_+^*$ in the shape space and is colored in blue on Figure 7.2 (c). *Singular orbits or singular shapes* are orbits or shapes with larger isotropy group conjugation class. In the plane example, $(0,0)$ is the only object with singular shape. It corresponds to 0 in $\mathbb{R}_+$ and is colored in red in Figure 7.2 (c).

Principal orbits form an open and dense subset of M, denoted M^*. This means that there are objects with non-degenerated shapes almost everywhere. In the plane example, $\mathbb{R}^2 \setminus (0,0)$ is dense in $\mathbb{R}^2$. In the sphere example, $S^2 \setminus \{(0,0),(\pi,0)\}$ is dense in S^2, where $(0,0)$ denotes the north pole and $(\pi,0)$ the south pole of S^2. Likewise, principal shapes form an open and dense subset in Q, denoted Q^*. In the plane example, $\mathbb{R}_+^*$ is dense in $\mathbb{R}_+$. In the sphere example, $]0,\pi[$ is dense in $[0,\pi]$.

The dense set M^* makes the projection in the quotient space a Riemannian submersion [Huckemann 2010], which we use to embed $Q^* = M^*/G$ in M^*. In other words, regular shapes of Q^* are embedded in the space of objects with regular shapes M^*. Moreover, the computations in Section 6.4.1 will be carried out on the dense set M^* of principal orbits. The curvature of these principal orbits - i.e. of the blue circles of Figures 7.2(b) and (e) - will be the main geometric parameter responsible for the asymptotic bias studied in this paper. We note that the curvature of principal orbits is closely related to the presence of singular orbits: principal orbits wrap around the singular orbits. In the plane example, any blue circle - i.e. any principal orbit - wraps around its center, the red dot $(0,0)$ - which is the singular orbit, see Figure 7.2(b).

We have focused on an intuitive introduction of the concepts. We refer to [Postnikov 2001, Alekseevsky 2003, Huckemann 2010] for mathematical details. From now on, the mathematical setting is the following: we assume a proper, effective and isometric action of a finite dimensional Lie group G on a finite dimensional Riemannian manifold M.

5.1.3 Geometrization of generative models of shape data

We recall that the data are the $\{X_i\}_{i=1}^n$ that are sets of landmarks, curves, images, etc. In the general case, one can interpret the data X_i's as random realizations of the generative model:

$$X_i = \mathrm{Exp}(g_i \cdot Y_i, \varepsilon_i) \qquad i = 1...n, \tag{5.1}$$

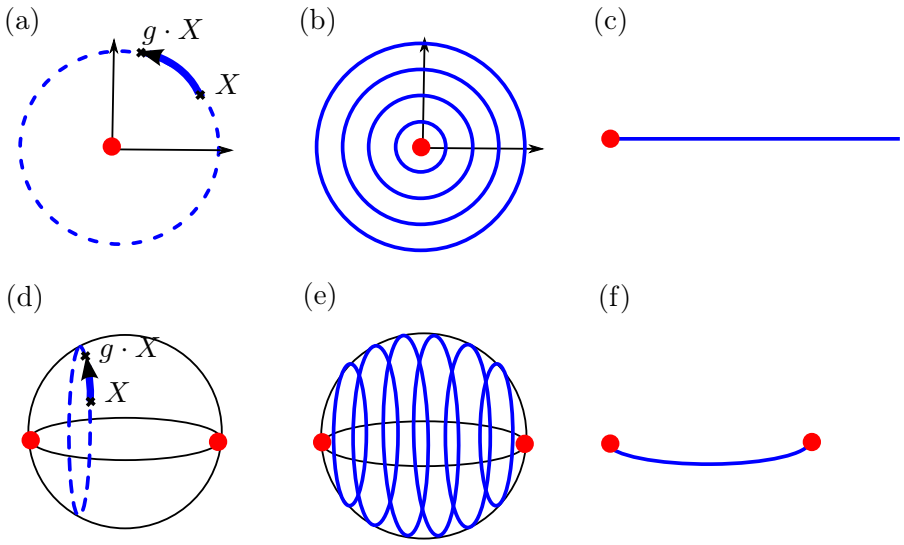


Figure 5.2: First line: Action of rotations on $\mathbb{R}^2$, with (a): action of rotation $g \in SO(2)$ on point $X \in \mathbb{R}^2$ and orbit of X in blue dotted line; (b) Stratification of $\mathbb{R}^2$ into principal orbit type (blue) and singular orbit type (red); (c) shape space $\mathbb{R}_+ = \mathbb{R}^2/SO(2)$ with a singularity (red dot). Second line: Action of $SO(2)$ on S^2 with (d): action of rotation $g \in SO(2)$ on point $X \in S^2$ and orbit of X in blue dotted line; (e) Stratification of S^2 into principal orbit type (blue) and singular orbit type (red) (f) shape space $[0, \pi] = S^2/SO(2)$ with two singularities (red dots).

where $\text{Exp}(p, u)$ denotes the Riemannian exponential of u at point p. The Y_i, g_i, ε_i are respectively i.i.d. realizations of random variables that are drawn independently.

In this paper as well as often in the literature [Allassonnière 2007, Allassonnière 2016, Bigot 2010, Bigot 2011, Kurtek 2011], we consider mainly the following simpler generative model:

$$X_i = \text{Exp}(g_i \cdot Y, \varepsilon_i) \qquad i = 1...n, \tag{5.2}$$

where Y is a parameter which we call the template shape. The following three step formulation of the generative models 5.1 and (7.2) gives technical details and their interpretation in terms of shapes.

Step 1: Generate the shape $Y_i \in M^*/G \subset M^*$ In the full generative model (5.1), we assume that there is a probability density of shapes in the Riemannian manifold $Q^* = M^*/G$, with respect to the measure on Q^* induced by the Riemannian measure of M^*. The Y_i's are i.i.d. samples drawn from this distribution. For example, it can be a Gaussian - or one of its generalization to manifolds [Pennec 2006] - as illustrated in Figure 7.3 on the shape spaces for the plane and sphere examples. This is the variability that is meaningful for the statistical study, whether we are analyzing shapes of skulls, proteins, bones, neural spike trains, handwritten digits or brains.

We mainly assume in this paper the simpler generative model (7.2) with parameter: the template shape $Y \in M^*/G$. In other words, we assume that the probability distribution is singular and more precisely that it is simply a Dirac at Y. This is the most common assumption within the model (5.1) [Allassonnière 2007, Allassonnière 2015a, Kurtek 2011, Bigot 2010]. We point out that Y is a point of the shape space M^*/G, which is embedded in the object space by $M^*/G \subset M^*$, see previous subsection.

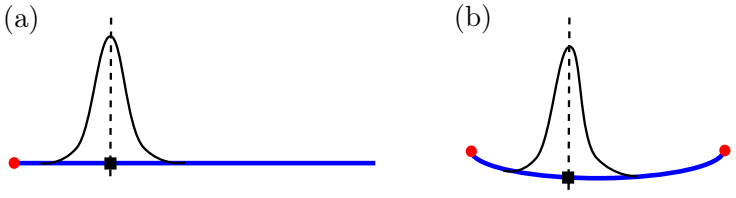


Figure 5.3: Step 1 of generative model of Equation (5.1) for the plane example (a) and the sphere example (b). The black curve illustrates the probability distribution function on shape space. This is a distribution on $r \in \mathbb{R}_+$ for the plane example (a) and on $\theta \in [0, \pi]$ for the sphere example. The black square represents its expectation. For the simpler generative model of Equation (7.2), the probability distribution boils down to a single point at Y i.e. at the black square.

Step 2: Generate its position/parameterization $g_i \in G$, to get $g_i \cdot Y \in M^*$ We cannot observe shapes in $Q = M/G$. We rather observe objects in M, that

are shapes posed or parameterized in a certain way. We assume that there is a probability distribution on the positions or parameterizations of G, or equivalently a probability distribution on principal orbits with respect to their intrinsic measure. We assume that the distribution does not depend on the shape Y_i that has been drawn. The g_i's are i.i.d. from this distribution. For example, it can be a Gaussian - or one of its generalization to manifolds [Pennec 2006] - as illustrated in Figure 7.4 on the shape spaces for the plane and sphere examples.

The drawn g_i is used to pose/parameterize the shape Y_i drawn in Step 1 (in the case of model of Equation (5.1)), where $Y_i = Y$ (in the case of model of Equation (7.2)). The shape is posed/parameterized through the isometric action of G on $Q^* \subset M^*$, to get the object $g_i \cdot Y_i \in M^*$, or the object $g_i \cdot Y \in M^*$ in the case of the simpler model of Equation (7.2).

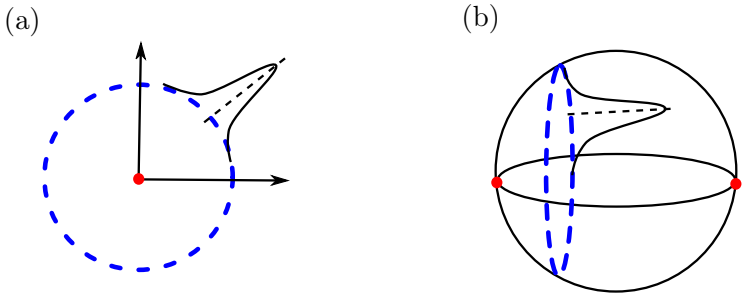


Figure 5.4: Step 2 of generative model of Equation (7.2) for the plane example (a) and the sphere example (b). The blue dotted curve illustrates the orbit of the shape drawn in Step 1. The black curve illustrates the probability distribution function on this orbit. This is a distribution in angle $\theta \in [0, 2\pi]$ for the plane example (polar coordinates) and in angle $\phi \in [0, 2\pi]$ for the sphere example (spherical coordinates).

Step 3: Generate the noise $\varepsilon_i \in T_{g_i \cdot Y_i} M$ The observed X_i's are results of noisy measurements. We assume that there is a probability distribution function on $T_{g_i \cdot Y_i} M$ representing the noise. We further assume that this is a Gaussian - or one of its generalization to manifolds [Pennec 2006] - centered at $g_i \cdot Y_i$, the origin of the tangent space $T_{g_i \cdot Y_i} M$, and with standard deviation σ, see Figure 7.5. The parameter σ will be extremely important in the developments of Section 6.4.1, as we will compute Taylor expansions around $\sigma = 0$.

Other generative models may be considered in the literature. We find in [Allassonnière 2015a] the model: $X_i = g_i \cdot \mathrm{Exp}(Y_i, \varepsilon_i)$, where the Riemannian exponential Exp is also performed in M through the embedding $Y_i \in Q \subset M$. In [Kurtek 2011], we find the model without noise: $X_i = g_i \cdot Y_i$.

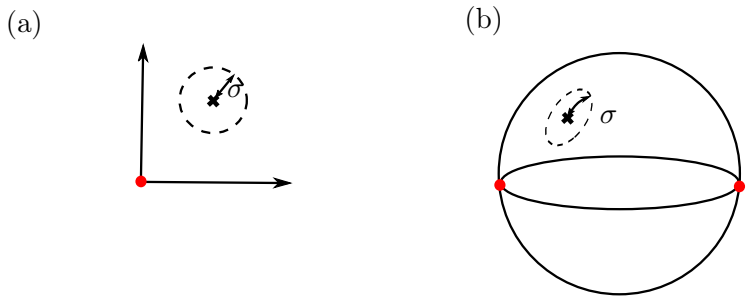


Figure 5.5: Step 3 of generative model of Equation (7.2) for the plane example (a) and the sphere example (b). The dotted curve represents the isolevel at σ of the Gaussian distribution function on the ambient space.

5.1.4 Learning the variability in shapes: estimating the template shape

Our goal is to unveil the variability *of shapes in* $Q = M/G$ while we in fact observe the noisy *objects* X_i*'s in* M. We focus on the case where the variability in the shape space is assumed to be a Dirac at Y. Our goal is thus to estimate the template shape Y, which is a parameter of the generative model.

Estimating the template shape with the Fréchet mean in the shape space We describe the procedure usually performed in the literature [Kurtek 2011, Allassonnière 2007, Allassonnière 2016, Bigot 2010, Bigot 2011]. One initializes the estimate with $\hat{Y} = X_1$. Then, one iterates the following two steps until convergence:

$$\begin{cases} (i) \quad \hat{g}_i = \underset{g \in G}{\operatorname{argmin}}\ d_M(\hat{Y}, g \cdot X_i), \quad \forall i \in \{1, ..., n\}, \\ (ii) \quad \hat{Y} = \underset{Y \in M}{\operatorname{argmin}} \sum_{i=1}^{n} d_M(Y, \hat{g}_i \cdot X_i)^2. \end{cases} \tag{5.3}$$

This procedure has a very intuitive interpretation. Step (i) is the projection of each object X_i in the shape space Q, as illustrated in Figure 7.6 (a)-(i) and (b)-(i) with the blue arrows. We assume that each minimizer $\hat{g}_i$ exists and is attained. In practice, this will often be the case and for example when the Lie group is compact, like the Lie group of rotations for example. We take X_1, X_2, X_3 three objects in $\mathbb{R}^2$ in Figure 7.6 (a)-(i) and on S^2 in Figure 7.6 (b)-(i). One filters out the position/parameterization component, i.e. the coordinate on the orbit. One projects the objects X_1, X_2, X_3 in the shape space Q using the blue arrows.

Step (ii) is the computation of the mean of the registered data $\hat{g}_i \cdot X_i$, i.e. of the objects' shapes, as illustrated in Figure 7.6(a)-(i) and (b)-(i) where $\hat{Y}$ is shown in orange. Again, we assume that the minimizer $\hat{Y}$ exists and is attained. In practice, this will be the case as we will consider a low level of noise in Step 3 of the generative model. The registered data $\hat{g}_i \cdot X_i$ will be concentrated on a small neighborhood

of diameter of order σ. As a consequence, their Fréchet mean in the Riemannian manifold Q^* is guaranteed to exist and be unique [Émery 1991].

The procedure of Equations (5.3) (i)-(ii) decreases at each step the following cost, which is bounded below by zero:

$$\text{Cost}(g_1, ..., g_n, Y) = \sum_{i=1}^{n} d_M^2(Y, g \cdot X_i). \tag{5.4}$$

Under the assumptions that both steps (i) and (ii) attained their minimizers, we are guaranteed convergence to a local minimum. We further assume that the procedure converges to the global minimum. The estimator computed with the procedure is then:

$$\hat{Y} = \underset{Y \in M}{\text{argmin}} \sum_{i=1}^{n} \min_{g \in G} d_M^2(Y, g \cdot X_i). \tag{5.5}$$

The term $\min_{g \in G} d_M^2(Y, g \cdot X_i)$ in Equation 6.5 is the distance in the shape space between the shapes of Y and X_i. Thus, we recognize in Equation 6.5 the Fréchet mean on the shape space. The Fréchet mean is a definition of mean on manifolds [Pennec 2006]: it is the point that minimizes the squared distances to the data in the shape space. All in all, one projects the probability distribution function of the X_i's from M to M/G and computes its "expectation", in a sense made precise later.

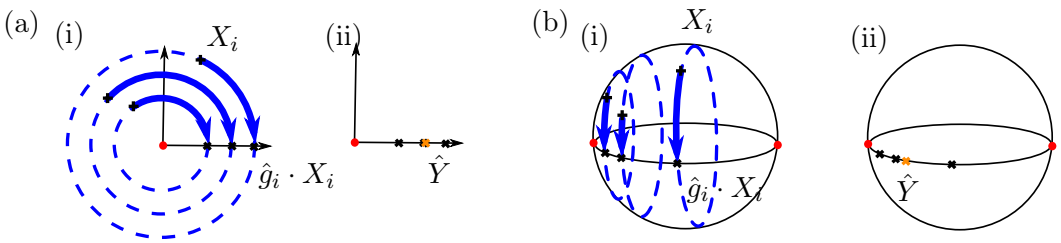


Figure 5.6: Steps (i) and (ii) of procedure of template shape estimation described in Equations (5.3) (i)-(ii) for the plane example (a) and the sphere example (b). The 3 black plus signs in $\mathbb{R}^2$ (a) or S^2 (b) represent the 3 data. The 3 dotted blue curves are their orbits. In Step (i), the X_i's are registered, i.e. their projected in the shape space: 3 curved blue arrows represent their registration with the minimizers $\hat{g}_i$. The 3 black crosses in $\mathbb{R}_+$ (positive x-axis) (a) or $[0, \pi]$ (b) represent the registered data. In Step (ii), the template shape estimate $\hat{Y}$ is computed as the Fréchet mean of the registered data and is shown in orange.

We implemented the generative model and the estimation procedure on the plane and the sphere in shiny applications available online: `https://nmiolane.shinyapps.io/shinyPlane` and `https://nmiolane.shinyapps.io/shinySphere`. We invite the reader to look at the web pages and play with the different parameters of the generative model. Figure 5.7 shows screen shots of the applications.

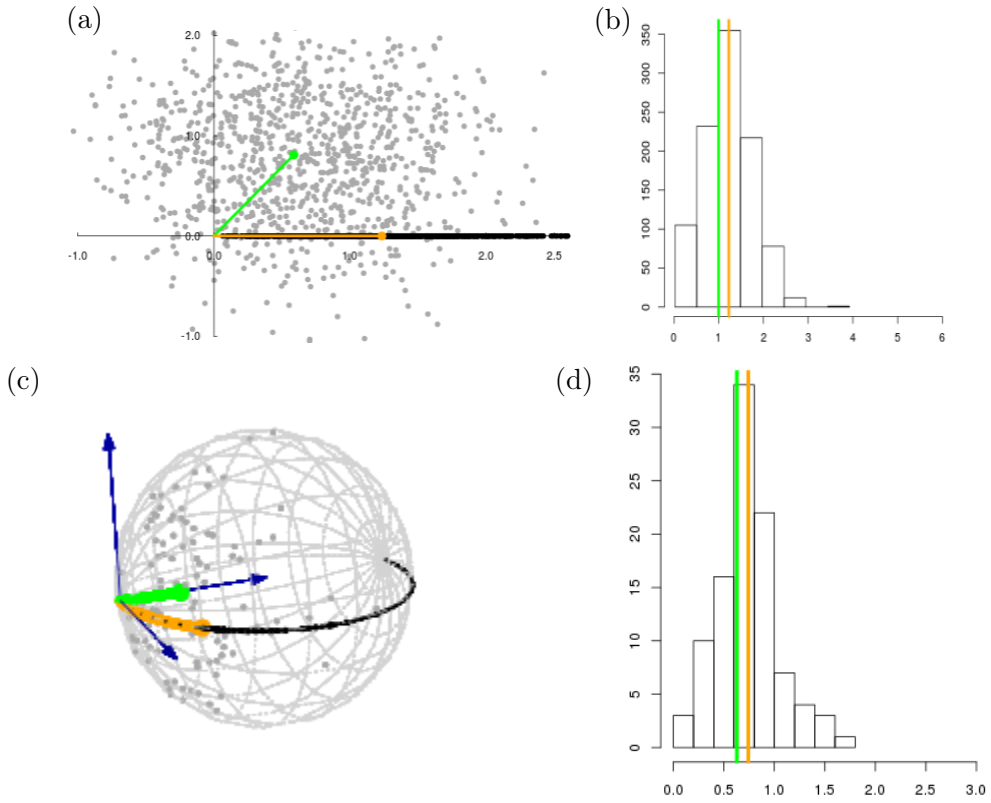


Figure 5.7: Screenshot of `https://nmiolane.shinyapps.io/shinyPlane` and `https://nmiolane.shinyapps.io/shinySphere`. Simulated data X_i's (grey points), template shape Y (green), registered data $\hat{g}_i \cdot X_i$ (black points), template shape estimate $\hat{Y}$ (orange). Induced distributions on the shapes, template shape Y (green), template shape estimate $\hat{Y}$ (orange).

Probabilistic interpretation of the procedure in Equations (5.3): an approximation of a Maximum-Likelihood Beside its intuitive interpretation, the procedure of template shape estimation of Equations (5.3) has a probabilistic interpretation. We have the generative model of the data X_i's: it is described in Equation (7.2) and Steps (1)-(3) of the previous subsections. Thus, one may consider the Maximum Likelihood (ML) estimate of Y, which is one of its parameters:

$$\begin{aligned}\hat{Y}_{ML} = \underset{Y \in Q}{\text{argmax}}\ \mathrm{L}(Y) &= \underset{Y \in Q}{\text{argmax}} \sum_{i=1}^{n} \log(P(X_i|Y)) \\ &= \underset{Y \in Q}{\text{argmax}} \sum_{i=1}^{n} \log\left(\int_{g \in G} P(X_i|Y,g).P(g)dg\right).\end{aligned}$$

In the above, $P(X_i|Y)$ is the probability distribution of the data in M as a function of the parameter Y. $P(g)$ is the probability distribution on the poses/parameterizations in G as described in Step 2 of the generative model given

in the previous subsection. Then, $P(X_i|Y, g)$ is the probability distribution of the noise as described in Step 3.

The g's are hidden variables in the model. The Expectation-Maximization (EM) algorithm is therefore the natural implementation for computing the ML estimator [Allassonnière 2007]. But the EM algorithm is computationally expensive, above all for tridimensional images. Thus, one can usually rely on an approximation of the EM, which is described in [Allassonnière 2007] as the "modal approximation" and used in [Allassonnière 2015a, Kurtek 2011, Bigot 2010].

We can check that this approximation is the procedure described in Equations (5.3). Step (i) is an estimation of the hidden observations g_i and an approximation of the E-step of the EM algorithm. Step (ii) is the M-step of the EM algorithm: the maximization of the surrogate in the M-step amounts to the maximization of the variance of the projected data. This is exactly the minimization of the squared distances to the data of (ii). We refer to [Allassonnière 2007] for details.

Purpose of this paper reformulated with the geometrization Our main result is to show that the procedure presented in Equations (5.3) (and illustrated on Figure 7.6) gives an asymptotically biased estimate $\hat{Y}$ for the template shape Y of the generative model presented in Equation (7.2) (and illustrated in Figures 7.3, 7.4 and 7.5). Figures 5.7 (a)-(d) present what is meant by *asymptotic bias*: the estimate $\hat{Y}$, of the procedure, is in orange and the template shape Y, of the generative model, is in green. The estimator $\hat{Y}$ (in orange) does converge when the number of data, i.e. the grey points in Figures 5.7(a)-(c), goes to infinity, *but $\hat{Y}$ does not converge to the template shape Y it is designed to estimate.* For Figures 5.7 (a)-(d), this means that even for an infinite number of grey points, the orange estimate will be different from the green parameter. We say that $\hat{Y}$ has an asymptotic bias with respect to the parameter Y.

Where does this asymptotic bias come from and why doesn't $\hat{Y}$ converge to Y? In a nutshell, the bias comes from the external curvature of the template's orbit and we explain and summarize this in Figure 6.7 and its caption. The full geometric answer with its technical details is provided in the next section.

5.2 Quantification and correction of the asymptotic bias

This section explains, quantifies and corrects the asymptotic bias of the template shape estimate $\hat{Y}$ with respect to the parameter Y. We start from the definition of the asymptotic bias of an estimator with respect to the parameter it is designed to estimate. More precisely we start from a generalization of this definition to Riemannian manifolds:

$$\text{Bias}(\hat{Y}, Y) = \mathbb{E}\left[\text{Log}_Y \hat{Y}\right]. \tag{5.6}$$

This is the asymptotic bias of the estimator $\hat{Y}$ with respect to the (manifold-valued) parameter Y, which generalizes the corresponding definition for linear spaces:

$$\text{Bias}(\hat{Y}, Y) = \mathbb{E}\left[\hat{Y} - Y\right] \tag{5.7}$$

In the Riemannian definition of the bias, $\text{Log}_Y \hat{Y}$ is the Riemannian logarithm of $\hat{Y} \in Q$ at $Y \in Q$, i.e. a vector of the tangent space of Q at the real parameter Y, denoted $T_Y Q$. The tangent vector $\text{Log}_Y \hat{Y}$ is illustrated on Figures 5.8 (a) and (b) for the plane and sphere examples. $\text{Log}_Y \hat{Y}$ represents how much one would have to shoot from Y to get the estimated parameter $\hat{Y}$. The norm of $\text{Log}_Y \hat{Y}$, computed using the metric of Q at Y, represents the dissimilarity between $\hat{Y}$ and Y.

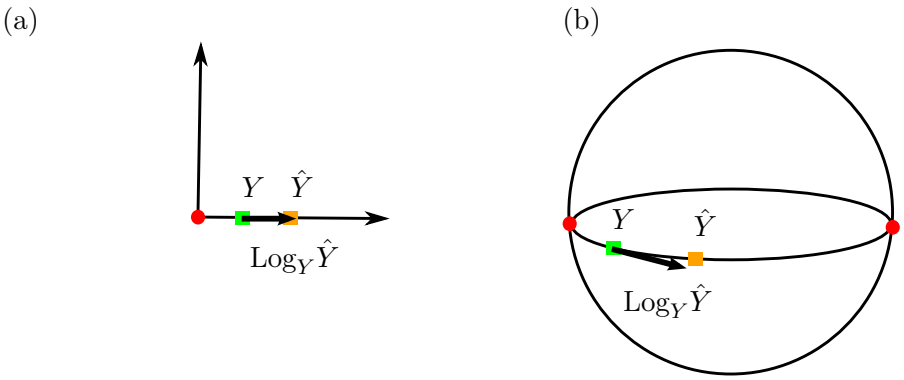


Figure 5.8: Illustration of the Riemannian definition of the asymptotic bias $\text{Log}_Y \hat{Y}$ for the plane example (a) and the sphere example (b). Log refers to the Riemannian logarithm [Postnikov 2001] and $\text{Log}_Y \hat{Y}$ is thus a tangent vector of the quotient space Q at Y. $\text{Log}_Y \hat{Y}$ represents how much one would have to shoot from Y to get the estimated parameter $\hat{Y}$. The norm of $\text{Log}_Y \hat{Y}$, computed using the metric of Q at Y, represents the distance or the dissimilarity between $\hat{Y}$ and Y, i.e. how far $\hat{Y}$ is from estimating Y.

We could also consider the variance of the estimator $\hat{Y}$. The variance is defined as $\text{Var}_n(\hat{Y}) = \text{E}[d_M(Y, E[Y])^2]$. In the limit of an infinite sample, we have: $\text{Var}_\infty(\hat{Y}) = 0$. This is why we focus on the asymptotic bias.

5.2.1 Asymptotic bias of the template's estimator on examples

We first compute the asymptotic bias for the examples of the plane and the sphere to give the intuition.

The probability distribution function of the X_i's comes from the generative model. This is a probability distribution on $\mathbb{R}^2$ for the plane example, parameterized in polar coordinates (r, θ) like Figure 7.1. So we can compute the projected distribution function on the shapes, which are the radii r here. This is done simply by integrating out the distribution on θ, the position on the circles. This gives a

probability distribution on $\mathbb{R}_+$ for the plane example. We write it $f : r \mapsto f(r)$. We remark that f does not depend on the probability distribution function on the θ_i's of Step 2 of the generative model. We can also compute $f : \theta \mapsto f(\theta)$ in the sphere example: we integrate over ϕ the probability distribution function on (θ, ϕ).

Figure 7.7 (a) shows f for the plane example, for a template $r = 1$. We plot it for two different noise levels $\sigma = 0.3$ and $\sigma = 3$. Note that here f is the Rice distribution. Figure 7.7 (b) shows f for the sphere example, for a template $\theta = 1$. We plot it for different noise levels and $\sigma = 0.3$ and $\sigma = 3$. In both cases, the x-axis represents the shape space which is $\mathbb{R}_+$ for the plane example and $[0, \pi]$ for the sphere example. The green vertical bar represents the template shape, which is 1 in both cases. The red vertical bar is the expectation of f in each case. It is $\hat{Y}$, the estimate of Y We see on these plots that f is not centered at the template shape: the green and red bars do not coincide. f is skewed away from 0 in the plane example and away from 0 and π in the sphere example. The skew increases with the noise level σ. The difference between the green and red bars is precisely the bias of $\hat{Y}$ with respect to Y.

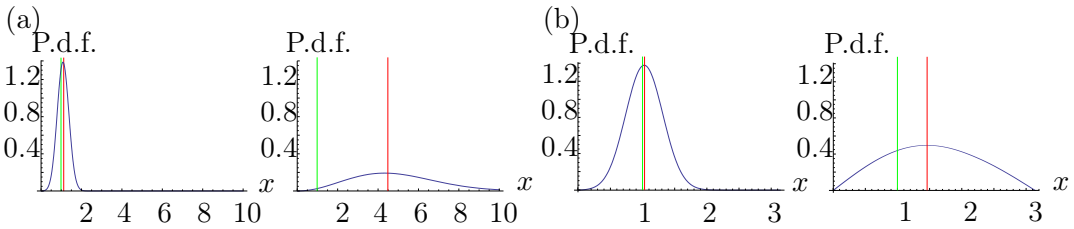


Figure 5.9: (a) Induced distributions on the distance r between two landmarks in $\mathbb{R}^3$ for real distance $y = 1$ (in green) and noise level $\sigma = 0.3$ and $\sigma = 3$. (b) Induced distributions on the angle x between the two landmarks on S^3, for real angle $y = 1$ and noise levels $\sigma = 0.3$ and $\sigma = 3$. In both cases the mean shape estimate $\hat{y}$ is shown in red.

Figure 5.10 shows the bias of $\hat{Y}$ with respect to Y, as a function of σ, for the plane (left) and the sphere (right). Increasing the noise level σ takes the estimate $\hat{Y}$ away from Y. The estimate is repulsed from 0 in the plane example: it goes to ∞ when $\sigma \to \infty$. It is repulsed from 0 and π in the sphere example: it goes to $\pi/2$ when $\sigma \to \pi$, as the probability distribution becomes uniform on the sphere in this limit. One can show numerically that the bias varies as σ^2 around $\sigma = 0$ in both cases. This is also observed on the shiny applications [RStudio, Inc 2013] at `https://nmiolane.shinyapps.io/shinyPlane` and `https://nmiolane.shinyapps.io/shinySphere`.

These examples already show the origin of the asymptotic bias of $\hat{Y}$, for low noise levels $\sigma \to 0$ or for high noise levels: $\sigma \to +\infty$ for the plane example and $\sigma \to \pi$ for the sphere example. As long as there is noise, i.e. $\sigma \neq 0$, there is *a bias that comes from the curvature of the template's orbit.* Figure 6.7 shows the template's orbit in blue, in (a) for the plane and (b) for the sphere. In both cases the black circle represents the level set σ of the Gaussian noise. In the plane example (a), the

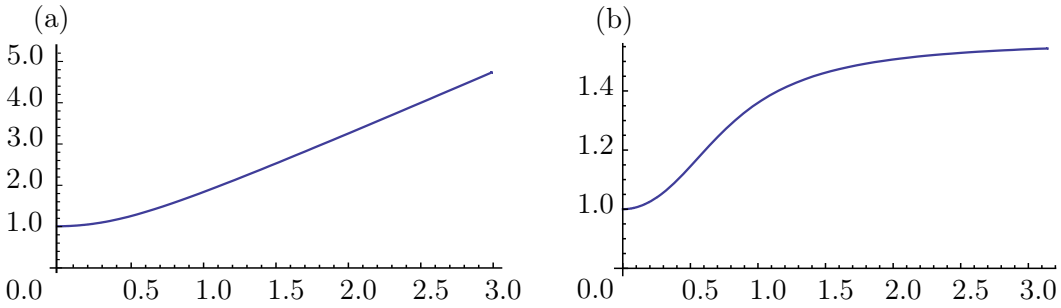


Figure 5.10: Asymptotic bias on the mean shape estimate $\hat{Y}$ with respect to the noise level σ for $r = 1$ in the plane example (a) and $\theta = 1$ in the sphere example (b). The bias is quadratic near $\sigma = 0$. Increasing σ takes the estimate $\hat{Y}$ away from 0 in shape space $Q = \mathbb{R}_+$ (a) and away from 0 and π in shape space $Q = [0, \pi]$ (b).

probability of generating an observation X_i outside of the template's shape orbit is bigger than the probability of generating it inside: the grey area in the black circle is bigger than the white area in the white circle. There will be more registered data that are greater than the template. Their expected value will therefore be greater than the template and thus biased. In the sphere example (b), if the template's shape orbit is defined by a constant $\theta < \pi/2$, the probability of generating an observation X_i "outside" of it, i.e. with $\theta_i > \theta$, is bigger than the probability of generating it "inside". There will be more registered data that are greater than the template θ and again, their expected value will also be greater than the template. Conversely, if the template is $\theta > \pi/2$, the phenomenon is inverted: there will be more registered data that are smaller than the template. The average of these registered data will also be smaller than the template. Finally, if the template's shape orbit is the great circle defined by $\theta = \pi/2$, then the probability of generating an observation X_i on the left is the same as the probability of generating an observation X_i on the right. In this case, the registered data will be well-balanced around the template $\theta = \pi/2$ and their expected value will be $\pi/2$: there is no asymptotic bias in this particular case. We prove this in the general case in the next section.

5.2.2 Asymptotic bias of the template's estimator for the general case

We show the asymptotic bias of $\hat{Y}$ in the general case and prove that it comes from the external curvature of the template's orbit. We show it for Y a principal shape and for a Gaussian noise of variance σ^2, truncated at 3σ. Our results will need the following definitions of curvature.

The *second fundamental form* h of a submanifold O of M is defined on $T_XO \times T_XO$ by $h(v, w) = (\nabla_v w)^\perp \in N_XO$, where $(\nabla_v w)^\perp$ denotes the orthogonal projection of covariant derivative $\nabla_v w$ onto the normal bundle. The *mean curvature vector* H *of* O is defined as: $H = Tr(h)$. Intuitively, h and H are mea-

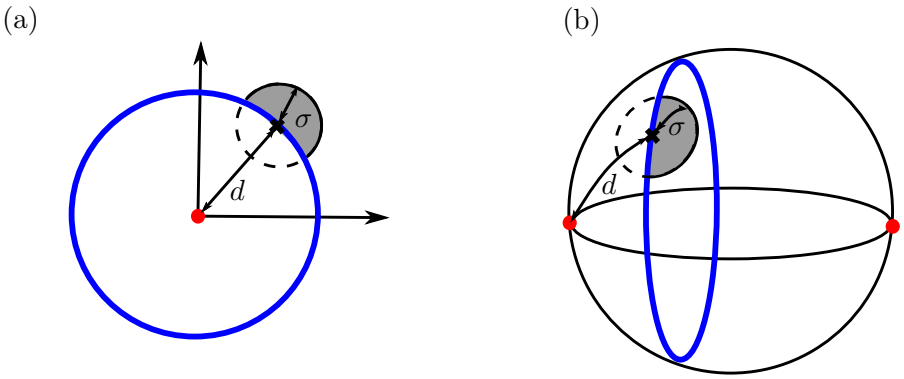


Figure 5.11: The external curvature of the template's orbit creates the asymptotic bias, in the plane example (a) and the sphere example (b). The blue curve represents the template's orbit. The ball of radius σ represents a level set of the Gaussian distribution of the noise in $\mathbb{R}^2$ (a) and S^2 (b). The grey-colored area represents the distribution of the noise that generates data outside the orbit of Y, in Step 3 of the generative model of Equation (7.2) and Figure 7.5. There is a higher probability that the data are generated "outside" the orbit. The template shape estimate is biased towards greater radii (a) or towards angles closer to $\pi/2$ (b).

sures of extrinsic curvature of O in M. For example an hypersphere of radius R in $\mathbb{R}^m$ has mean curvature vector $||H|| = \frac{m-1}{R}$.

Theorem 5.2.1 *The data X_i's are generated in the finite-dimensional Riemannian manifold M following the model: $X_i = Exp(g_i \cdot Y, \varepsilon_i), i = 1...n$, described in Equation (7.2) and Figures 7.3-7.5. In this model: (i) the action of the finite dimensional Lie group G on M, denoted $\cdot$, is isometric, (ii) the parameter Y is the template shape in the shape space Q, (iii) ε_i is the noise and follows a (generalization to manifolds of a) Gaussian of variance σ^2, see Section 6.2.*

Then, the probability distribution function f on the shapes of the X_i's, $i = 1...n$, in the asymptotic regime on an infinite number of data $n \to +\infty$, has the following Taylor expansion around the noise level $\sigma = 0$:

$$f(Z) = \frac{1}{(\sqrt{2\pi}\sigma)^q} \exp\left(-\frac{d_M^2(Y,Z)}{2\sigma^2}\right)\left(F_0(Z) + \sigma^2 F_2(Z) + \mathcal{O}(\sigma^4) + \varepsilon(\sigma)\right)$$

where (i) Z denotes a point in the shape space Q, (ii) F_0 and F_2 are functions of Z involving the derivatives of the Riemannian tensor at Z and the derivatives of the graph G describing the orbit O_Z at Z, and (iii) ε is a function of σ that decreases exponentially for $\sigma \to 0$.

Proof The detailed proof is in Appendix A.

The exponential in the expression of f belongs to a Gaussian distribution centered at Z and of isotropic variance $\sigma^2\mathbb{I}$. However the whole distribution f differs from the

Gaussian because of the Z-dependent term in the right parenthesis. This induces a skew of the distribution away from the singular shapes, as observed for the examples in Figure 7.7. This also means that the expectation of this distribution is not Z and that the variance is not the isotropic $\sigma^2\mathbb{I}$.

Theorem 5.2.2 *The data X_i's are generated with the model described in Equation (7.2) and Figures 7.3-7.5, where the template shape Y is a parameter and under the assumptions of Theorem A.3.1. The template shape Y is estimated with $\hat{Y}$, which is computed by the usual procedure described in Equations (5.3).*

In the regime of an infinite number of data $n \to +\infty$, the asymptotic bias of the template's shape estimator $\hat{Y}$, with respect to the parameter Y, has the following Taylor expansion around the noise level $\sigma = 0$:

$$Bias(\hat{Y}, Y) = -\frac{\sigma^2}{2} H(Y) + \mathcal{O}(\sigma^4) + \varepsilon(\sigma) \tag{5.8}$$

where (i) H is the mean curvature vector of the template shape's orbit which represents the external curvature of the orbit in M, and (ii) ε is a function of σ that decreases exponentially for $\sigma \to 0$.

Proof The proof is in Appendix A.

This generalizes the quadratic behavior observed in the examples on Figure 5.10. The asymptotic bias has a geometric origin: it comes from the external curvature of the template's orbits, see Figure 6.7.

We can vary two parameters in equation A.15: Y and σ. The external curvature of orbits generally increases when Y is closer to a singularity of the shape space (see Section 1) [Lytchak 2010]. The singular shape of the two landmarks in $\mathbb{R}^2$ arises when their distance is 0. In this case, the mean curvature vector has magnitude $|H(Y)| = \frac{1}{d}$: it is inversely proportional to d, the radius of the orbit. d is also the distance of Y to the singularity 0.

5.2.3 Limitations and extensions

Beyond Y being a principal shape Our results are valid when the template Y is a principal shape. This is a reasonable assumption as the set of principal shapes is dense in the shape space. What happens when Y approaches a singularity, i.e. when Y changes stratum in the stratified space Q? Taking the limit $d \to 0$ in the coefficients of the Taylor expansion is not a legal operation. Therefore, we cannot conclude on the Taylor expansion of the Bias for $d \to 0$. Indeed, the Taylor expansion may even change order for $d \to 0$. We take $M = \mathbb{R}^m$ with the action of $SO(m)$ and the template $Y = (0, ..., 0)$:

$$\text{Bias}(\hat{Y}, Y) = \sqrt{2}\frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})}\sigma. \tag{5.9}$$

The bias is linear in σ in this case.

Beyond $\sigma << 1$ The assumption $\sigma << 1$ represents our hope that the noise on the shape data is not too large with respect to the overall size of the mean shape. Nevertheless it would be very interesting to study the asymptotic bias for any σ, including large noises ($\sigma \to +\infty$). The distribution over the X_i's in M will be spread on the whole manifold M. We cannot rely on local computations on M (at the scale of σ) anymore. We have to make global assumptions on the manifold M.

The plane example is the canonical example of a flat manifold. The sphere example is the canonical example of manifold with constant (positive) curvature. The bias as a function of σ is plotted in Figure 5.10. It leads us to the conjecture that the estimate converges towards a barycenter of shape space's singularities when the noise level increases. Singularities have a repulsive action on the estimation of each template's shape. Such repulsive force acts on each estimators. As a result, the estimators of the mean shape finds an equilibrium position: the barycenter.

Beyond one Dirac in Q: several templates We have considered so far that there is a unique template shape Y: the generative model has a Dirac distribution at Y in the shape space. What happens for other distributions? We assume that there are K template shapes $Y_1, ..., Y_K$. Observations are generated in M from each template shape Y_k with the generative model of Section 2. Our goal is to unveil the structure of the shape distribution, i.e. the K template shapes here, given the observations in M. The distributions on shapes projected on the shape space is a mixture of probability density functions of the form of the density in Theorem A.3.1. Its modes are related to the template shapes. The K-means algorithm is a very popular method for data clustering. We study what happens if one uses K-means algorithms on shapes generated with the generative model above.

The goal is to cluster the shape data in K distinct and significant groups. One performs a coordinate descent algorithm on the following function:

$$J(c, \mu) = \sum_i d_Q(X_i, \mu_{c_i})^2. \tag{5.10}$$

In other words, the minimization of J is performed through successive minimizations on the assignment labels c's and the cluster's centers μ's. Given the c, minimizing J with respect to the μ's is exactly the simultaneous computation of K Fréchet means in the shape space. Meaningful well separated clusters (high inter-clusters dissimilarity) are chosen so that members are close to each other (high intra-cluster similarity). In other words, the quality of the clustering is evaluated by the following criterion:

$$D = \min_{\text{clusters } i,j} \frac{d_Q(c_i, c_j)}{\max_i \text{diam}(c_i)}, \tag{5.11}$$

which is the dissimilarity between clusters quotiented by the diameter of the clusters. In the absence of singularity in the shape space, the projected distribution looks like Figure 5.12 (a) and $D \propto \frac{1}{\sigma}$. The criterion is worse in the presence of singularities.

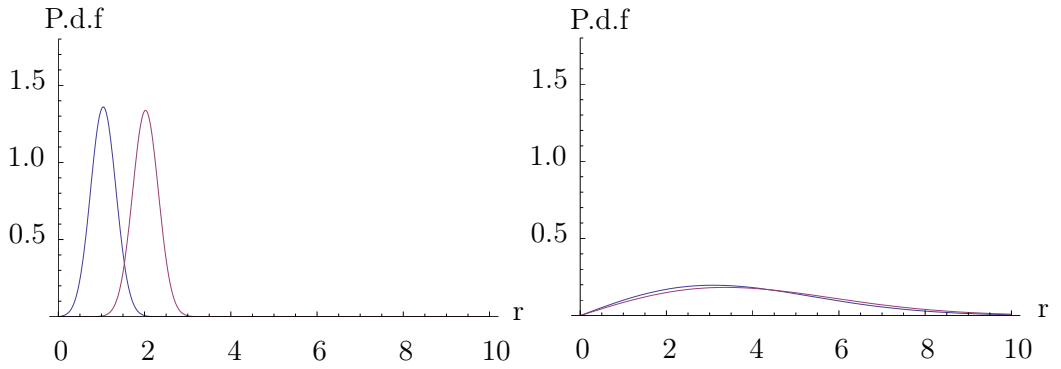


Figure 5.12: Two clusters of template shapes for the plane example: $r_1 = 1$ (blue) and $r_2 = 2$ (dark red). Noise levels: $\sigma = 0.3$ (left) and $\sigma = 3$ (right). The 2 clusters are hardly distinguishable when the noise increases.

Figure 5.12 illustrates this behavior for the plane example. We consider any two clusters i, j and call $\hat{Y}_i$, $\hat{Y}_j$ the estimated centroids. The criterion D writes:

$$D \equiv \frac{\hat{y}_i - \hat{y}_j}{\sigma} \underset{\sigma \to +\infty}{\sim} \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{2} m \Gamma\left(\frac{m}{2}\right)} \frac{y_i^2 - y_j^2}{\sigma^2} = O\left(\frac{1}{\sigma^2}\right).$$

Even in the best case with correct assignments to the clusters i and j, the K-means algorithm looses an order of validation when computed on shapes.

Beyond the finite dimensional case Our results are valid when M is a finite dimensional manifold and G a finite dimensional Lie group. Some interesting examples belong to the framework of infinite dimensional manifold with infinite dimensional Lie groups. This is the case for the LDDMM framework on images [Joshi 2006]. It would be important to extend these results to the infinite dimensional case.

We take $M = \mathbb{R}^m$ with the action of $SO(m)$. We have a analytic expression of f in this case [Miolane 2015a]. Figure 5.13 shows the influence of the dimension m for the probability distribution functions on the shape space and for the Bias. The bias increases with m. This leads us to think that it appears in infinite dimensions as well.

5.3 Correction of the systematic bias

We propose two procedures to correct the asymptotic bias on the template's estimate. They rely on the bootstrap principle, more precisely a parametric bootstrap, which is a general Monte Carlo based resampling method that enables us to estimate the sampling distributions of estimators [Efron 1979]. We assume that we know the variance $\hat{\sigma}^2$ from the experimental setting.

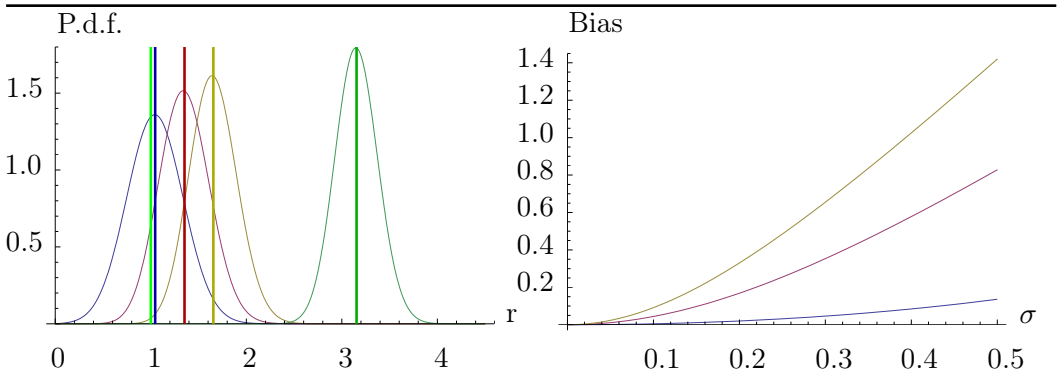


Figure 5.13: Probability distributions functions (noise $\sigma = 0.3$) and bias for $\mathbb{R}^m$ for $m = 2$, $m = 10$, $m = 20$ and $m = 100$. Template shape is $r = 1$.

5.3.1 Iterative Bootstrap

The first procedure is called an Iterative Bootstrap. Algorithm 5.3.1 gives the details. Figure 5.14 illustrates it on the plane example.

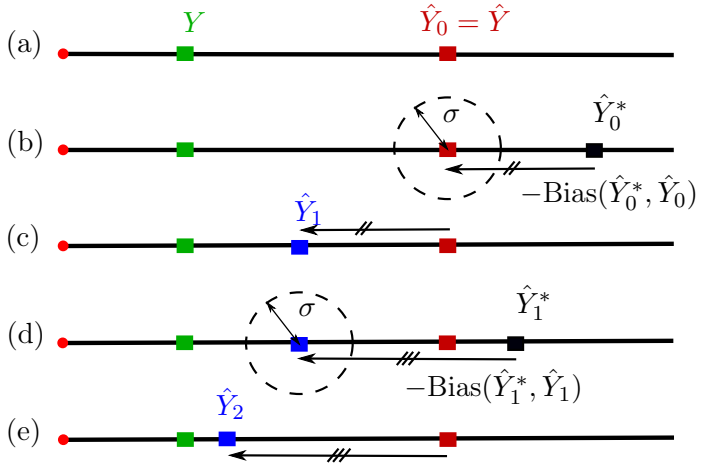


Figure 5.14: Algorithm 5.3.1 Iterative bootstrap procedure on the plane example for $n \to +\infty$. (a) Initialization, (b) Generate bootstrap sample from $\hat{Y}_0$ and compute the corresponding estimate $\hat{Y}_0{}^*$, compute the bias $\hat{Y}_0 - \hat{Y}_0^*$, (c) Correct $\hat{Y}_0$ with the bias to get $\hat{Y}_1$, (d) Generate bootstrap sample from $\hat{Y}_1$ and iterate as in (b), (e) Get $\hat{Y}_2$ etc.

Algorithm 5.3.1 starts with the usual template's estimate $\hat{Y}_0 = \hat{Y}$, see Figure 5.14 (a). At each iteration, we correct $\hat{Y}$ with a better approximation of the bias. First, we generate bootstrap data by using $\hat{Y}$ as the template shape of the generative model. We perform the template's estimation procedure with the Fréchet mean in the shape space. This gives an estimate $\hat{Y}_0^*$ of $\hat{Y}_0$. The bias of $\hat{Y}_0*$ with respect to $\hat{Y}_0$ is $\text{Bias}(\hat{Y}_0^*, \hat{Y}_0)$. It gives an approximation of the bias $\text{Bias}(\hat{\hat{Y}}, \hat{Y})$, see Figure 5.14

(b). We correct $\hat{Y}$ by this approximation of the bias. This gives a new estimate $\hat{Y}_1$, see Figure 5.14 (c). We recall that the bias $\text{Bias}(\hat{\hat{Y}}, \hat{Y})$ depends on Y, see Theorem A.4.1. $\hat{Y}_1$ is closer to the template Y than $\hat{Y}_0$. Thus, the next iteration gives a better approximation $\text{Bias}(\hat{Y}_1^*, \hat{Y}_1)$ of $\text{Bias}(\hat{\hat{Y}}, \hat{Y})$. We correct the initial $\hat{Y}$ with this better approximation of the bias, etc. The procedure is written formally for a general manifold M in Algorithm 5.3.1.

Algorithm 1 Corrected template shape estimation with **Iterative Bootstrap**

Input: Objects $\{X_i\}_{i=1}^n$, noise variance σ^2
Initialization:
$\hat{Y}_0 = \text{Frechet}(\{[X_i]\}_{i=1}^n)$
$k \leftarrow 0$
Repeat:
Generate bootstrap sample $\{X_i^{(k)^*}\}_{i=1}^n$ from $\mathcal{N}_M(Y_k, \sigma^2)$
$\widehat{Y_k} = \text{Fréchet}(\{[X^{(k)^*}]_i\}_{i=1}^n)$
$\text{Bias}_k = \text{Log}_{Y_k} \widehat{Y_k}$
$\hat{Y}_k = \text{Exp}_{\hat{Y}_0}\left(-\Pi_{\hat{Y}_k}^{\hat{Y}_0}(\text{Bias}_k)\right)$
$k \leftarrow k + 1$
until convergence: $||\text{Log}_{\hat{Y}_{k+1}} \hat{Y}_k|| < \varepsilon$
Output: $\hat{Y}_k$

In Algorithm 5.3.1, Π_A^B denotes the parallel transport from T_AM to T_BM. For linear spaces like $\mathbb{R}^2$ in the plane example, $\text{Log}_{P_1} P_2 = \overrightarrow{P_1P_2}$, $\text{Exp}_{P_1}(u) = P_1 + u$, For linear spaces and the parallel transport is the identity $\Pi_{P_1}^{P_2}(u) = u$. For other manifolds like S^2 in our sphere example, the parallel transport $\Pi_A^B(u)$ can theoretically be computed by solving the parallel transport equation at any point on the chosen curve linking A to B: $D_{t_{AB}} v = 0$ in v, where D is the covariant derivative in the direction t_{AB}, the tangent vector of the curve at the chosen point [Postnikov 2001]. In practice, the Schild's ladder [Ehlers 2012] or the Pole ladder [Marco 2013] can be used to compute an approximation of the parallel transport.

Algorithm 5.3.1 is a fixed-point iteration $Y^{(k+1)} = F(Y^{(k)})$ where:

$$F(X) = \text{Exp}_{\hat{Y}}(-\Pi_X^{\hat{Y}}(\text{Bias})) \qquad \text{where:} \qquad \text{Bias} = \text{Log}_X \hat{X}. \tag{5.12}$$

In a linear setting we have simply $F(X) = \hat{Y} - \overrightarrow{X\hat{X}}$. One can show that F is a contraction and that Y, the template shape, is the unique fixed point of F (using the local bijectivity of the Riemannian exponential and the injectivity of the estimation procedure). Thus the procedure converges to Y in the case of an infinite number of observations $n \to +\infty$. Figure 5.15 illustrates the convergence for the plane example, with a Gaussian noise of standard deviation $\sigma = 1$. The template shape $Y = 1.2$ was initially estimated at $\hat{Y} = 4.91$. Algorithm 5.3.1 corrects the bias.

Figures 5.16 and 5.17 show the iterations of Iterative Bootstrap for the plane and the sphere example.

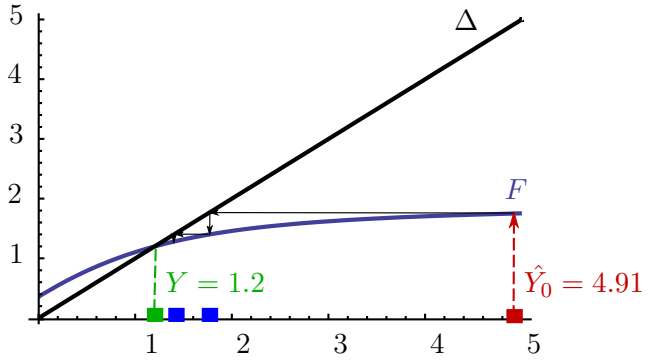


Figure 5.15: F of the fixed-point procedure and first 2 iterations for $\sigma = 1$, $m = 3$. Δ is the first diagonal. The initial estimate is biased $\hat{Y}_0 = 4.91$. The Iterative Bootstrap converges towards the template shape $Y = 1.2$.

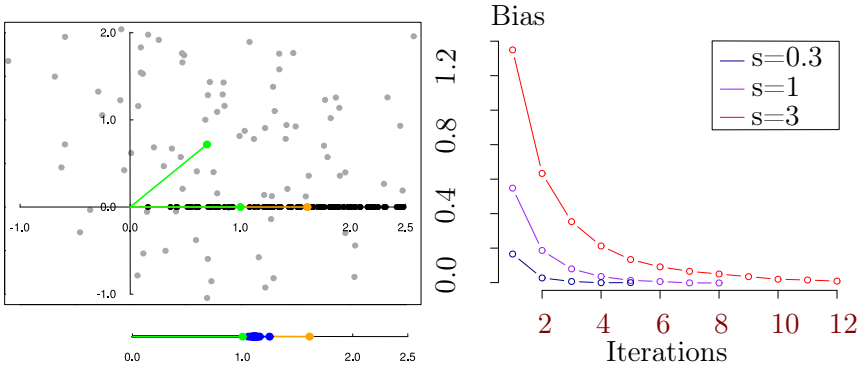


Figure 5.16: Left: Implementation of the plane example: the green point is the template shape Y, the grey points are the data X_i's generated with the model (7.2), the black points are the registered data $\hat{g}_i \cdot X_i$'s, the orange point is the template shape estimate $\hat{Y}$. The quotient space $\mathbb{R}_+$ is copied below, and the blue points show the iterations of the iterative bootstrap of Algorithm 5.3.1 that corrects the bias of $\hat{Y}$ as an estimate of Y: the blue points go from the orange point $\hat{Y}$ to the green point Y. Right: Convergence of the iterative bootstrap of Algorithm 5.3.1, for the plane example. The colors red, purple, blue represent different noises σ. The bias of $\hat{Y}$ as an estimator of Y is shown on the ordinate axis: it converges to 0 in a few iterations.

5.3.2 Nested Bootstrap

The second procedure is called the Nested Bootstrap. Algorithm 3 details it. Figure 5.18 illustrates it on the plane example.

Algorithm 3 starts like Algorithm 5.3.1 with $\hat{Y}_0 = \hat{Y}$, see Figure 5.18 (a). It also performs a parametric bootstrap with $\hat{Y}^{(0)}$ as the template, computes the bootstrap replication $\hat{Y}_0^*$ and the approximation $\text{Bias}(\hat{Y}_0^*, \hat{Y}_0)$ of $\text{Bias}(\hat{Y}, Y)$, see Figure 5.14 (b). Now Algorithm 3 differs from Algorithm 5.3.1. We want to know how biased is $\text{Bias}(\hat{Y}_0^*, \hat{Y}_0)$ as an estimate of $\text{Bias}(\hat{Y}, Y)$? This is a valid question as the bias de-

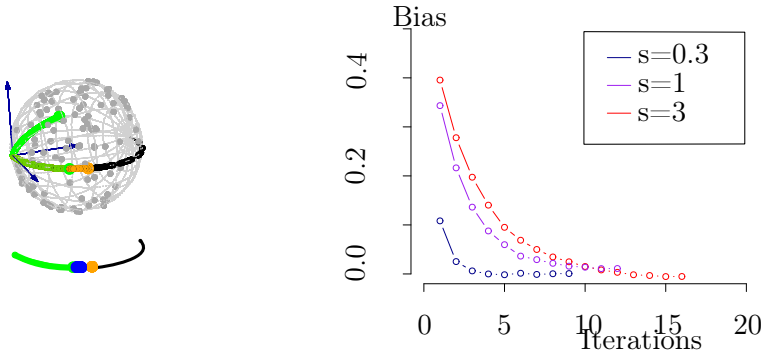


Figure 5.17: Left: Implementation of the sphere example: the green point is the template shape Y, the grey points are the data X_i's generated with the model (7.2), the black points are the registered data $\hat{g}_i \cdot X_i$'s, the orange point is the template shape estimate $\hat{Y}$. The quotient space $[0, \pi]$ is copied below, and the blue points show the iterations of the iterative bootstrap of Algorithm 5.3.1 that corrects the bias of $\hat{Y}$ as an estimate of Y: the blue points go from the orange point $\hat{Y}$ to the green point Y. Right: Convergence of the iterative bootstrap of Algorithm 5.3.1, for the sphere example. The colors red, purple, blue represent different noises σ. The bias of $\hat{Y}$ as an estimator of Y is shown on the ordinate axis: it converges to 0 in a few iterations.

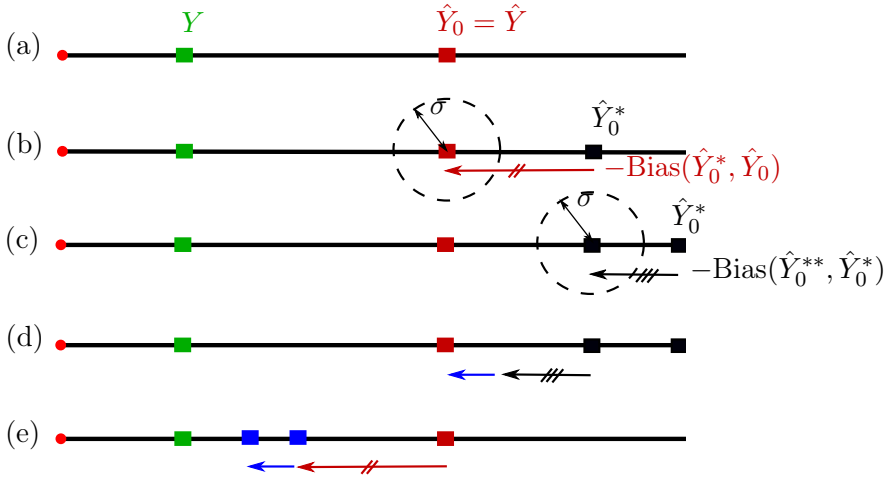


Figure 5.18: Algorithm 3 Nested Bootstrap on the plane example for $n \to +\infty$. (a) Initialization, (b) Generate bootstrap sample from $\hat{Y}_0$; compute the estimate $\hat{Y}_0{}^*$, compute the bias $\hat{Y}_{(0)} - \hat{Y}_0^*$, (c) Generate bootstrap sample from $\hat{Y}_{0^*}$; compute the estimate $\hat{Y}_0{}^{**}$, compute the bias $\hat{Y}_0^* - \hat{Y}_0^{**}$, (d) compute the blue arrow, i.e. the bias of $\text{Bias}(\hat{Y}_0^{**}, \hat{Y}_0^*)$ as an estimate of $\text{Bias}(\hat{Y}_0^*, \hat{Y}_0)$, (e) Correct $\hat{Y}$ with the bias-corrected bias.

pends on the template Y, see Theorem A.4.1. We want to estimate this dependence. We perform a bootstrap, nested in the first one, with $\hat{Y}^{(0)*}$ as the template. We

compute the estimate $\hat{Y}_0^{**}$ and the approximation $\text{Bias}(\hat{Y}_0^{**}, \hat{Y}_0^{*})$ of $\text{Bias}(\hat{Y}_0^{*}, \hat{Y}_0)$, see Figure 5.14 (c). We observe how far $\text{Bias}(\hat{Y}_0^{**}, \hat{Y}_0^{*})$ is from $\text{Bias}(\hat{Y}_0^{*}, \hat{Y}_0)$. This gives the blue arrow, which is the bias of $\text{Bias}(\hat{Y}_0^{**}, \hat{Y}_0^{*})$ as an estimate of $\text{Bias}(\hat{Y}_0^{*}, \hat{Y}_0)$, see Figure 5.14 (d). The blue arrow is an approximation of how far $\text{Bias}(\hat{Y}_0^{*}, \hat{Y}_0)$ is from $\text{Bias}(\hat{Y}, Y)$. We correct our estimation of the bias (in red) by the blue arrow. We correct $\hat{Y}$ by the bias-corrected estimate of its bias, see Figure 5.14 (e).

Algorithm 2 Corrected template shape estimation with **Nested Bootstrap**

Input: Objects $\{X_i\}_{i=1}^n$, noise variance σ^2
Initialization:
$\hat{Y}_0 = \text{Frechet}(\{[X_i]\}_{i=1}^n)$
Bootstrap:
Generate bootstrap sample $\{X_i^*\}_{i=1}^n$ from $\mathcal{N}_M(\hat{Y}_0, \sigma^2)$
$\hat{Y}_0^* = \text{Fréchet}(\{[X^*]_i\}_{i=1}^n)$
$\text{Bias} = \text{Log}_{\hat{Y}_0} \widehat{Y}_0^*$
Nested Bootstrap:
For each i:

- Generate bootstrap sample $\{X_i^{**}\}_{k=1}^n$ from $\mathcal{N}_M(\hat{Y}_0^*, \sigma^2)$
- $\hat{Y}_{0,i}^{**} = \text{Fréchet}(\{[X^{**}]_i\}_{k=1}^n)$

$\text{Bias(Bias)} = \text{Log}_{\hat{Y}_0} \widehat{Y}_0^* - \Pi_{\hat{Y}_0^*}^{\hat{Y}_0} \text{Log}_{Y_0^*} \widehat{Y}_0^{**}$
$\hat{Y}_1 = \text{Exp}_{\hat{Y}_0}(-\text{Bias} - \text{Bias(Bias)})$
Output: $\hat{Y}_1$

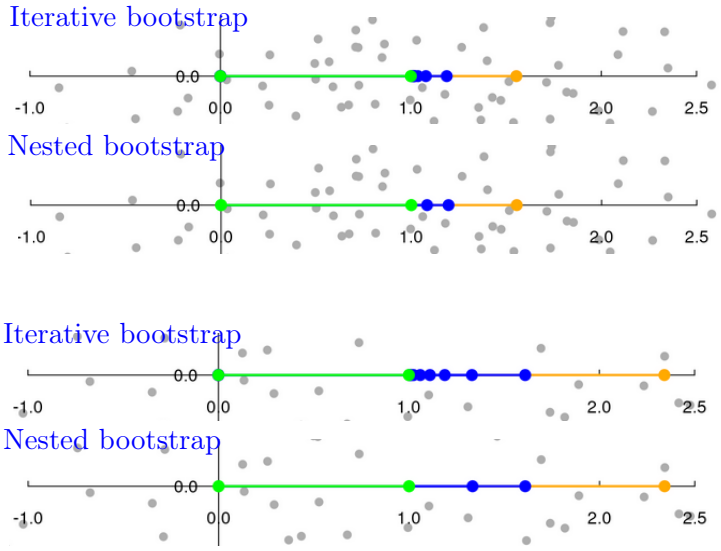
5.3.3 Comparison

One may use the Iterative Bootstrap or the Nested Bootstrap depending on the experimental setting. We illustrate them both on the plane example in Figure 5.19. Figure 5.19 (a) shows the performance of both algorithms for a signal-over-noise ratio (SNR) of 1: the template shape in green is a $r = 1$ and the standard deviation of the noise is $\sigma = 1$, so that $\text{SNR} = \frac{r}{\sigma} = 1$. Figure 5.19 (b) shows both algorithms for $\text{SNR} = \frac{r}{\sigma} = \frac{1}{3} = 0.33$. In all four experiments: the template shape is the green dot at $r = 1$, the template shape estimate is in orange, and the successive steps of the bootstrap algorithms are the blue dots: we have several blue dots for the Iterative Bootstrap, and two blue dots for the Nested Bootstrap.

The advantages of the Iterative Bootstrap are the following. It corrects the bias of $\hat{Y}$ perfectly in the case of a very large number of observations n, as we can see in Figures 5.19 (a) on top and (b) on top: the blue dots converge to the green dot for the two different SNRs. Thus, the Iterative Bootstrap can be used to experimentally compute the mean curvature vector H of each orbit of a group action. One probes the orbit's curvature by "feeling it" with a Riemannian Gaussian on M and projecting on the shape space. The drawbacks of the Iterative Bootstrap

are the following. It works only with very large n. It is not robust as it uses the generative model several times. If the generative model is far from being true, then the iterative bootstrap fails.

The advantages of the Nested Bootstrap are the following. It is a standard statistical procedure that is more robust with respect to variations of the generative model. Even if generative model is different from the one that we assume, the Nested Bootstrap performs well. Moreover, it does not need as much data as the Iterative Bootstrap. Its drawback is that it does not correct perfectly the bias, especially when the noise is large. This can be seen in Figures 5.3.1 (a) on bottom and (b) on bottom. While the Nested Bootstrap gets close to the green dot on Figure 5.19 (a) on bottom for the SNR $= 1$, it stays significantly far from the green dot on Figure 5.19 (b) bottom for the SNR $= 0.33$.



Signal-Noise ratio $= 1$

Signal-Noise ratio $= 0.33$

Figure 5.19: Comparison of the Iterative bootstrap and the Nested bootstrap on simulation with two different Signal-Noise ratio, which is SNR $= \frac{Y}{\sigma} = \frac{r}{\sigma}$, the ratio of the template Y, which is the radius r in the plane example, on the noise level σ. (a) shows SNR $= 1$ and (b) shows SNR $= 0.33$. In all four experiments: the template shape is the green dot at $r = 1$, the template shape estimate is in orange, and the successive steps of the bootstrap algorithms are the blue dots: we have several blue dots for the Iterative Bootstrap, and two blue dots for the Nested Bootstrap.

These simulations give a rule of thumb, i.e. some intuition, for when the bias needs to be corrected. They confirm what could already be observed in Figure 5.10. In Figure 5.10, the template is fixed at $r = 1$ or $\theta = 1$. A variation in the noise level σ corresponds to a variation in the SNR. In particular, we read the threshold $SNR = 1$ when $\sigma = 1$, i.e. when the noise σ is comparable to the distance of the template Y to the singularity. In both cases for $SNR > 1$, the template estimate is significantly different from the template as the bias is of the order of magnitude of

the template itself.

5.4 Applications to simulated and real data

5.4.1 Simulated triangles

We perform a simulation using the iterative bootstrap on triangles. We randomly generate $n = 10^5$ triangles in $\mathbb{R}^2$ through the generative model described in Equation (7.2) of Section 6.2. This is illustrated on Figure 5.20. We consider the isometric action of the Lie group $SO(2)$ of 2D rotations on $(\mathbb{R}^2)^3$, the space of 3 landmarks in 2D. For Step 1 of the generative model of Section 6.2, the template triangle is chosen arbitrarily and then fixed during the simulations. The template triangle is represented in green in Figure 5.20. For Step 2, we consider a Dirac distribution at the identity in the Lie group $SO(2)$. In other words, we do not rotate the triangles. At the end of this step, each of the 10^5 triangles is exactly the green triangle of Figure 5.20. This simpler model does not decrease the impact of the simulation: the noise of Step 3 is independent of the position of the triangle on their orbit, and Step (i) of the procedure is to quotient out the position of the orbit. For Step 3, we add bivariate Gaussian noise on each landmark, i.e. on each of the three points defining the green triangle. This gives a data set of 10^5 triangles. Some of them are represented in grey on Figure 5.20.

We then apply the procedure described in Equations (5.3) (i)-(ii) to estimate the (green) template triangle. In Step (i), we register the (grey) triangle data. This gives the registered the data, illustrated in black in Figure 5.20. We then compute the Fréchet mean of the black triangles by computing the Euclidean mean of each of their 3 landmarks. This gives the estimate of the template triangle, in orange on Figure 5.20.

The template estimate in orange is different from the template in green, even with a very high number of observations: $n = 10^5$. We apply the iterative bootstrap to correct this bias. The number of iterations required for the convergence of Algorithm 1 with respect to the noise level are shown in Figure 5.20. We observe the convergence in the three experiments for less than 10 iterations.

5.4.2 Real triangles: shape of the Optic Nerve Head

Now we go to real triangle data. We have 24 images of Rhesus monkeys' eyes, acquired with a Heidelberg Retina Tomograph [Patrangenaru 2015]. For each monkey, an experimental glaucoma was introduced in one eye, while the second eye was kept as control. One seeks a significant difference between the glaucoma and the control eyes. On each image, three anatomical landmarks were recorded: S for the superior aspect of the retina, N for the nose side of the retina, and T for the side of the retina closest to the temporal bone of the skull. The data are matrices $\{X_i\}_{i=1}^n$ where the landmark coordinates form the rows. For the ONH example, M is the space of 3 landmarks in 3D, $M = (\mathbb{R}^3)^3$ and the rotations act isometrically on each object X_i.

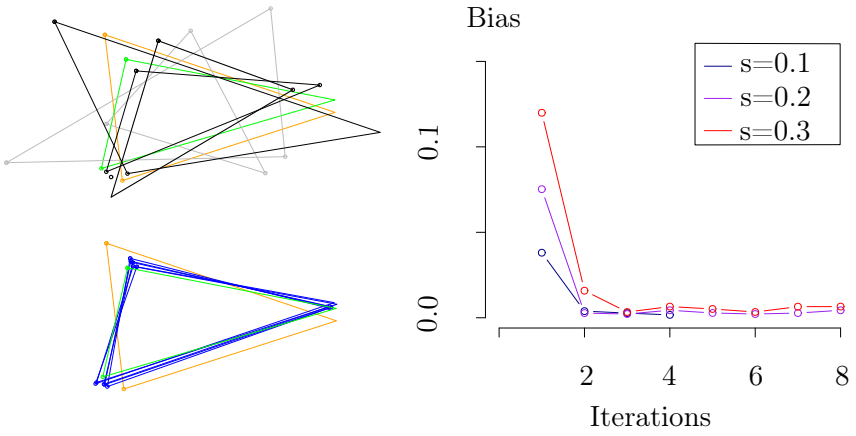


Figure 5.20: Left: Implementation for simulated triangles: the green triangle is the template shape Y, the grey triangles are (some of) the data X_i's generated with the model (7.2), the black triangles are (some of) the registered data $\hat{g}_i \cdot X_i$'s, the orange triangle is the template shape estimate $\hat{Y}$. The blue triangles show the iterations of the iterative bootstrap of Algorithm 5.3.1 that corrects the bias of $\hat{Y}$ as an estimate of Y: the blue triangles go from the orange triangle $\hat{Y}$ to the green triangle Y. Right: Convergence of the iterative bootstrap of Algorithm 5.3.1. The colors red, purple, blue represent different noises σ. The bias of $\hat{Y}$ as an estimator of Y is shown on the ordinate axis: it converges to 0 in a few iterations.

Analysis This simple example illustrates the estimation of the template shape. We use the following procedure to compute the mean shape for each group. We initialize $\hat{Y}$ with X_1 and repeat the following two steps until convergence:

$$(1) \quad \forall i \in \{1, ..., n\}, \quad \hat{R}_i = \underset{R \in SO(3)}{\operatorname{argmin}} ||\hat{Y} - X_i.R||^2, \quad \text{(register to the current mean shape)},$$

$$(2) \quad \hat{Y} = \frac{1}{n} \sum_{i=1}^{n} X_i.\hat{R}_i \quad \text{(update the mean shape estimate)}.$$

Figure 5.21 shows the mean shapes $\hat{Y}^{\text{control}}$ of the control group (left) and $\hat{Y}^{\text{glaucoma}}$ of the glaucoma group (right) in orange, while the initial data are in grey. The difference between the two groups is quantified by the distance between their means: $||\hat{Y}^{\text{control}} - \hat{Y}^{\text{glaucoma}}|| = 21.84\mu\text{m}$. We want to determine if this analysis presents a bias that significantly changes the estimated shape difference between the groups.

We use the nested bootstrap to compute an approximation of the asymptotic bias on each mean shape, for a range of noise's standard deviation in $\{100\mu\text{m}, 200\mu\text{m}, 300\mu\text{m}, 400\mu\text{m}\}$. The asymptotic bias on the template shape of the glaucoma group is $\{0.1\mu\text{m}, 0.11\mu\text{m}, 0.12\mu\text{m}, 0.13\mu\text{m}\}$ and of the control group is $\{0.27\mu\text{m}, 0.42\mu\text{m}, 0.55\mu\text{m}, 0.67\mu\text{m}\}$. The corrected template shape differences are $\{22.01\mu\text{m}, 22.08\mu\text{m}, 22.14\mu\text{m}, 22.18\mu\text{m}\}$. In particular, for $\sigma = 400\mu\text{m}$, we observe that the bias in the template shape are respectively $0.67\mu\text{m}$ for the healthy group and $0.13\mu\text{m}$ for the glaucoma group. This follows the rule-of-thumb: the bias is

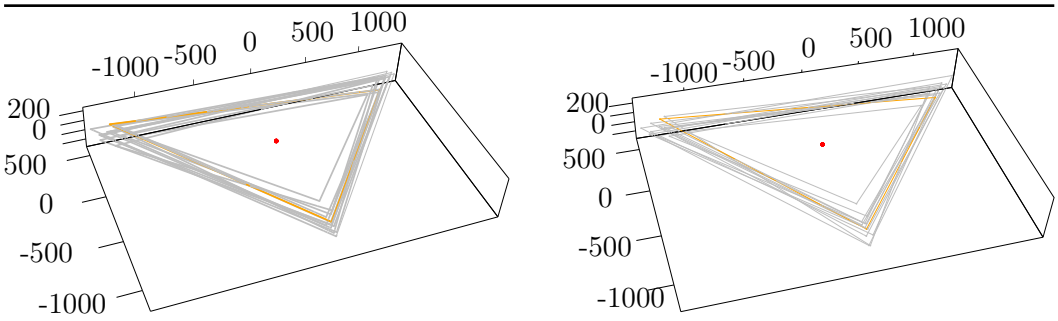


Figure 5.21: Triangles data in grey for the control group (left) and the glaucoma group (right). In orange, the estimated template shapes. Distances are measured in µm.

more important for the healthy group, for which the overall size is smaller than the glaucoma group, for a same noise level. The bias of the template shape estimate accounts for less than 1µm in this case, which is less than 0.1% of the shapes' sizes. This computation guarantees that this study has not been significantly affected by the bias.

5.4.3 Protein shapes in Molecular Biology

We estimate the impact of the bias on statistics on protein shapes. This subsection aims to suggest the potential importance of the results of this paper for Molecular Biology.

A standard hypothesis in Biology is that structure (i.e. shape) and function of proteins are related. Fundamental research questions about protein shapes include structure prediction - given the protein amino-acid sequence, one tries to predict its structure - and design - given the shape, one tries to predict the sequence needed.

One relies on experimentally determined 3D structures gathered in the Protein Data Base (PDB) [Berman 2000]. They contain errors on the protein's atoms coordinates. Average errors range from 0.01 Å to 1.76 Å, which is of the magnitude of the length of some covalent bonds. These values are averaged over the whole protein and in general, the main-chain atoms are better defined than the side-chain atoms or the atoms at the periphery. This is illustrated on Figure 5.22 where we have plot the B-factor (related to coordinates errors [Tickle 1998]) as a colored map on the atoms for proteins of PDB-codes 1H7W and 4HBB.

Protein's radius of gyration A biased estimate of a protein shape has consequences for studies on proteins folding. Stability and folding speed of a protein depend on both the estimated shape of the denatured state (unfolded state) and of the native state (folded state). One may study if compact initial states yield to faster folding. The protein compactness is represented by the protein's Radius of Gyration, defined as: $R_g^2 = \frac{1}{N}\sum_{\text{non H atoms i}}^{N}(r_i - R_C)^2$, where N is the number of

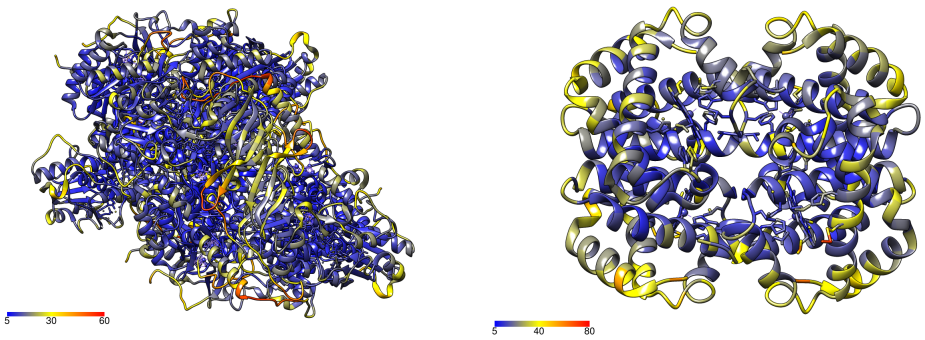


Figure 5.22: Errors on atoms coordinates represented by the B-factor, for proteins 1H7W (left) and 4HBB (right). Atoms at the periphery of the proteins tend to have more errors, which appear in yellow-red colors.

non-hydrogen atoms, r_i, R_C are resp. the coordinates of atoms and centers. Error on atoms coordinates give a bias on the estimate of the Radius of Gyration:

$$\mathbb{B}(R_g^2) = \sigma^2 \frac{3(N-1)}{N} = \bar{R_g}^2 \frac{(N-1)}{N} \frac{3}{\mathrm{SNR}^2}, \tag{5.13}$$

where we also express this bias with respect to an adaptation of the signal-noise-ratio introduced in Section 7.5: $\mathrm{SNR}^2 = \frac{R_g^2}{\sigma^2}$.

The radius of protein HJSJ (85 residues) is known to be around 10 Å. The error on R_g^2 is of 0.3% with an average error of positions on the atoms of 0.3Å. It is 8.6% for an error of 1.7Å. The error will be greater if one considers binding sites at the periphery of the proteins rather than the whole protein. Indeed sites' size is smaller and they have less atoms.

One could think about doing clustering on radii of Gyration using the K-means algorithm on shapes. The index D of Section 7.5 is:

$$D = \frac{R_1^{\sigma\,2} - R_2^{\sigma\,2}}{\sigma} = \frac{R_1^2 - R_2^2}{\sigma} + 3\sigma \left(\frac{N_1 - 1}{N_1} - \frac{N_2 - 1}{N_2} \right). \tag{5.14}$$

Clustering on radii of gyration may lead to a misleading indicator. D indicates that the clustering performs better that it actually does.

False positive probability in protein's motif detection The relation between a protein's shape and function is linked to its motifs, which define the supersecondary structure. Motifs have biological properties: for example the helixâĂŞturnâĂŞhelix motif [Brennan 1989]is responsible for the binding of DNA within several prokaryotic proteins. Automatic motif detection is another challenge in the study of protein shapes. We investigate the impact of bias on the false positive probability estimation in motif detection.

Let us consider a set $\{P_i\}_{i=1}^n$ of proteins each with N_i atoms. One is interested in the motifs of k atoms that can be detected in the protein's set, where $k < N$.

We define σ that represents an allowed error zone. The number of detected motifs increases if: (i) one decreases k, or (ii) one increases σ, or (iii) increases n. Thus how many detected motifs actually come from chance, with respect to the parameters k, σ, n? The false positives probability indicates when one detects truth and when one detects noise. The usual estimate of the false positive probability is $P = \frac{\mathcal{V}_0}{\mathcal{V}_l}$. Here $\mathcal{V}_0$ is the volume of the error zone allowed. $\mathcal{V}_l$ is the total volume of the protein [Pennec 1998b], thus the a ball of radius the Radius of Gyration. Thus $\mathcal{V}_l$ may be biased and overestimated. The probability of false positive is underestimated.

We consider the example of [Pennec 1998a]. One tries to find motifs between the tryptophan repressor of Escherichia coli (PDB code 2WRP) and the CRO protein of phage 434 (PDB code 2CRO). These two proteins are known to share the helix-turn-helix motif. The radius of Gyration of 2WRP is $R_g = 20\mathring{A}$, the total volume is: $\mathcal{V}_l = \frac{4}{3}\pi R_g^3 \simeq 33510\mathring{A}^3$. We assume an error zone that takes the form of a diagonal covariance matrix with standard deviations $\sigma = 0.35\mathring{A}$. We get the error zone volume: $\mathcal{V}_0 = \chi^3 \frac{4}{3}\pi\sigma^3 = 4.06\mathring{A}^3$, where $\xi^2 = 8$ comes from a convention about how much error is allowed: the covariance of the error within the error volume shall be less than ξ^2, see [Pennec 1998a] for details. The estimation of the false positive probability is: $P = 1.2 \times 10^{-4}$. We find that P is underestimated by 0.27% using the expression of the Radius of Gyration's bias.

5.4.4 Brain template in Neuroimaging

We apply the rule of thumb of Section 7.5 to determine when the bias needs a correction in the computation of a brain template from medical images. Here M and G will be infinite dimensional, so that the results of this paper do not apply directly. Nevertheless, this subsection allows us to gain intuition about how this paper may impact the field of neuroimaging.

In neuroimaging, a template is an image representing a reference anatomy. Computing the template is often the first step in medical image processing. Then, the subjects' anatomical shapes may be characterized by their spatial deformations *from the template*. These deformations may serve for (i) a statistical analysis of the subject shapes, or (ii) for automated segmentation by mapping the template's segmented regions into the subject spaces. In both cases, if the template is not centered among the population, i.e. if it is biased, then the analyzes and conclusions could be biased. We are interested in highlighting the variables that control the template's bias.

The framework of Large Deformation Diffeomorphic Metric Mapping (LDDMM) [Younes 2012] embeds the template estimation in our geometric setting. The Lie group of diffeomorphisms acts on the space of images as follows:

$$\rho : \mathrm{Diff}(\Omega) \times L_2(\Omega) \to L_2(\Omega), \qquad (\phi, I) \mapsto \phi \cdot I = I \circ \phi^{-1}. \tag{5.15}$$

The isotropy group of I writes: $G_I = \{\phi \in \mathrm{Diff}(\Omega) | I \circ \phi^{-1} = I\}$. Its Lie algebra $\mathfrak{g}_I$ consists of the infinitesimal transformations whose vector fields are parallel to the

level sets of I: $\mathfrak{g}_I = \{v | \forall x \in \Omega, \nabla I(x)^T.v(x) = 0\}$. The orbit of I is : $O_I = \{I' \in L_2(\Omega) | \exists \phi \in \text{Diff}(\Omega) \text{ s.t. } I' \circ \phi^{-1} = I\}$.

The "shape space" is by definition the space of orbits. Two images that are diffeomorphic deformations of one another are in the same orbit. They correspond to the same point in the shape space. Topology of an image is defined as the image's properties that are invariant by diffeomorphisms. Consequently, the shape space is the space of the images topology, represented by the topology of their level sets. We get a stratification of the shape space when we gather the orbits by orbit type. A stratum is more singular than another, if it has higher orbit type, i.e. larger isotropy group.

The manifold M has an infinite stratification. One changes stratum every time there is a change in the topology of an image's level sets. Singular strata are connected to simpler topology. "Principal" strata are connected to more complicated topology. Indeed, the simpler the topology of the level sets is, the higher is the "symmetry" of the image. Thus the larger is its isotropy group. Note that strata with smaller isotropy group (more detailed topology) do not represent "singularities" from the point of view of a given image and do not influence the bias. In fact, such strata are at distance 0: an infinitesimal local change in intensity can create a maximum or minimum, thus complexifying the topology.

Using the rule-of-thumb of Section 7.5, the template's bias depends on its distance d to the next singularity, at the scale of σ the intersubjects variability. The template is biased in the regions where the difference in intensity between maxima and minima is of the same amplitude as the variability. The template may converge to pure noise in these regions.

Conclusion

We introduced tools of statistics on manifolds to study the properties of template's shape estimation in Medical imaging and Computer vision. We have shown asymptotic bias by considering the shape space's geometry. The bias comes from the external curvature of the template's orbit at the scale of the noise on the data. This provides a geometric interpretation for the bias observed in [Allassonnière 2007, Allassonnière 2015a]. We investigated the case of several templates and the performance K-mean algorithms on shapes: clusters are less well separated because of each centroid's bias. The variables controlling the bias are: (i) the distance in shape space from the template to a singular shape and (ii) the noise's scale. This gives a rule-of-thumb for determining when the bias is important and needs correction. We proposed two procedures for correcting the bias: an iterative bootstrap and a nested bootstrap. These procedures can be applied to any type of shape data: landmarks, curves, images, etc. They also provide a way to compute the external curvature of an orbit.

Our results are exemplified on simulated and real data. Many studies use the template's shape estimation algorithm in Molecular Biology, Medical Imaging or

Computer vision. Their estimations are necessarily biased. But these studies often belong to a regime where the bias is not important (less than 0.1%). For example, the bias is important in landmark shapes analyses when the landmarks' noise is comparable to the template shape's size. Studies are rarely in this regime. We have considered shapes belonging to infinite dimensional shape spaces. Our results do not apply to the infinite dimensional case. We have used them to gain intuition about it. The bias might be more important in infinite dimensions and needs a correction as we have suggested.

Chapter 6

Geometric Statistics on topologies: controlling the bias in brain template estimation

---

This Chapter applies the analyses of Chapter 5 to the estimation of brain templates. As the Lie group involved is a group of diffeomorphisms, the quotient structure leads us to consider spaces of topology. This Chapter is a practical illustration of tools of Geometric Statistics on real datasets of Computational Anatomy.

This Chapter will be submitted to the SIAM Journal on Applied Algebra and Geometry under the title: "Topologically constrained template estimation via Morse-Smale complexes allows to control its statistical consistency".

6.1 Introduction

In neuroimaging, as well as in many other medical image analysis domains, a *template* is an image representing a reference anatomy. A brain template is computed from a database of brain images to serve as the image "prototype" for further statistical analyses.

Computation of a brain template Various methods exist to compute a brain template from a given database [Evans 2012]. A first practice selects one image from the database as the template. If the selected subject's anatomy is far from the population mean anatomy, the template is necessarily biased towards this specific individual. Thus, the template fails at being a prototype of the population. This is why researchers consider the computation of "unbiased template" that should better represent the mean anatomy.

Such an "unbiased" template is often constructed by performing an iterative averaging of intensities and deformations [Guimond 1998, Joshi 2004, Hadj-Hamou 2016]. One initializes with a template being one of the subject images. Then, during each iteration, one registers the subjects to the current template, and computes the mean deformation. The new template is computed as the mean intensity of the subjects images, deformed with the mean deformation. This procedure does not favor any subject's image , if it does not end in a local minimum. In this sense, the procedure is called "unbiased".

The computed template may look blurred or sharp depending on the design chosen for the registration in the above iterative procedure. If the algorithm is

designed using linear registration, the template may appear blurred. In contrast, if one uses diffeomorphic registration, the template is more likely to look sharp and the sharpness depends on the amount of regularization used [Evans 2012].

Purpose and desirable properties of the computed template Computing a template is often the first step in medical image processing because of its many applications. In general, the template is used as a standardized 3D coordinate frame where the subject brains can be compared. The subjects are then characterized by their *spatial diffeomorphic deformations from the template*. These deformations may then serve for a statistical analysis of the subject shapes, as in Deformation Based Morphometry (DBM) [Ashburner 1998]. One studies the normal and pathological variations of the subjects with respect to the template. The deformations also facilitate automated segmentation, by mapping the template's already segmented regions into each subject space. In other words, the template serves as prior knowledge of the brain anatomy [Despotovic 2015].

What are the desirable properties of the template, with respect to the applications mentioned above? First, the template should be representative of the population, to remove any bias toward a specific subject during the analysis [Bookstein 2001, Ashburner 2001, Baloch 2009]. Second, the template should be sharply defined, so that subtle anatomical structures can be easily observed or segmented. All in all, it is desirable that the template accurately estimates the brain anatomy shared by the subjects in the population.

The brain template, an inconsistent estimator of the unique brain anatomy However, it seems that the two desirable properties cannot be fulfilled simultaneously. Both intuition and recent statistical studies argue that there is a trade-off between a sharply-defined template and a template that accurately represents the brain anatomy shared among the subjects.

First, this trade-off can be understood intuitively. Consider a brain population divided into two groups that have different topologies in the following sense. The first group has subjects with three sulci - i.e. depressions or grooves in the cerebral cortex - in a specified brain region. The second group has subjects with only two sulci in the same region. A sharply defined template has to decide on a specific topology in this brain region, i.e. whether it shows two or three sulci. Therefore, it might not estimate correctly the brain anatomy of this population, which might be problematic for the following applications in neuroimaging. For example during a statistical analysis, registering the subjects with the three sulci topology to a template that has chosen a two sulci topology might not be reasonable [Baloch 2009]. Thus, a sharp template is only meaningful if the anatomical structures are representative of the population under study.

Second, the trade-off has been emphasized in recent studies, which have investigated the asymptotic bias of the template as an estimator of the shared anatomy of a given population. In the classical approach [Allassonnière 2007], an initial as-

sumption states that there is a unique (brain) anatomy shared by the subjects of the population. The subjects are then modeled, through a generative model, as random deformations of the unique brain anatomy with additional noise. The unique brain anatomy is then a parameter of this model. The template computation is interpreted as an estimation of this unique anatomy, and one can ask about its asymptotic bias: does the template converge to the unique brain anatomy if we compute it from a database with an infinite number of images?

This question has been investigated for signals, i.e. 1D images. Some authors prove the asymptotic unbiasedness of the template under the simplifying assumption of no measurement error on the observed signals [Kurtek 2011]. Other authors have already provided examples of asymptotic bias, and therefore inconsistency, when there is measurement error [Allassonnière 2007]. Their experiments show that the template signal may converge to pure noise when the measurement error on simulated signals increases. A bias is shown to occur in [Bigot 2011] for curves estimated from a finite number of points in the presence of noise.

Recently, the asymptotic bias of the template has been shown in the setting of Lie group actions [Miolane 2015a, Miolane 2016] and the authors provide a geometric interpretation for it. Their argument, shown in an abstract geometric context but adapted here to our brain images, is as follows. They look at the subspace defined by all brains with the same shape as the unique brain anatomy. They show that the curvature of this space, at the scale of the measurement noise, introduces the bias on the template brain's estimation.

Using topology to investigate the brain template's asymptotic bias Can we link: (i) the fact that a population with two groups of different brain topologies cannot be accurately represented by a sharp template, with (ii) the mathematical results on the template's bias as an estimate of the anatomy shared by the population? The framework of [Miolane 2015a] is based on the quotient of the space of the observed data by the action of a Lie group. The observed data in our case are the brain images, and the Lie group action is the action of diffeomorphisms on these images. The quotient of the images by the action of diffeomorphisms, i.e. what remains when one filters out any information that is invariant by diffeomorphic deformations, gives the topology of the images' level sets. This means that we could quantify the brain template's asymptotic bias *using a representation of its topology.*

This quantification could enable us to decide when and where a sharply defined template makes sense. We could want a sharp template where the intersubject anatomical variability is low and a fuzzier template when this variability is higher. Alternatively, we could consider computing several sharp brain templates using mixtures. This discussion boils down to the question: when is it reasonable to assume that a unique brain anatomy represents the whole subject population?

Furthermore, we could think about controlling the brain template's asymptotic bias by constraining its topology. Topological representations of images have been used with various objectives in the literature. For example, [Chung 2009]

uses a topological representation of images for classification of autism versus normals. Topological constraints have also been implemented for segmentation where the reconstruction of the cortical surface needs to match the brain anatomy [MacDonald 2000, Mangin 1995, Han 2003]. However, topological representation of images or topological constraints on images have not been used to study and enforce a statistical property, like the asymptotic unbiasedness.

Contributions and Outline We use a topological representation of images - the Morse-Smale complex - to investigate and control the asymptotic bias of the brain template estimation in neuroimaging. We make three main contributions in this paper. First, we link the framework of diffeomorphic deformations of (brain) images with Morse-Smale theory. We show how the Morse-Smale complex is a computational representation of geometric concepts involved in the template computation procedure. Second, we leverage this topological representation to analyze the template as an estimator of the brain anatomy. We quantify its asymptotic bias locally on the template image. This leads us to discuss the initial assumption of a unique anatomy. Third, we present an adaptation of the template computation algorithm that bounds the bias, through topological constraints, at the price of constructing a "smoother" template.

Section 6.2 presents the geometry and the topology of the template computation algorithm. In particular, this section emphasizes the geometric and topological variables that describe the bias of the brain template. Section 6.3 presents the chosen computational representation of the geometric and topological variables controlling the brain template's bias: the Morse-Smale complex. Section 6.4 leverages the previous computational model to identify spatially the biased regions of the template. These are interpreted as the brain regions where the assumption of a unique anatomy breaks down. A "smoother" template may be preferable there. We thus propose an adaptation of the template computation algorithm that builds a hierarchical brain template, with topological constraints bounding the asymptotic bias. In Section 6.5 our methodology is used on the Open Access Series of Imaging Studies (OASIS) database of T1-weighted MR brain images.

6.2 Geometry and topology in the template estimation problem

This theoretical section reviews the template estimation algorithm through the lens of geometry and topology. It gives the framework that formalizes the intuition given in the introduction, for example that a database with two groups showing different brain topologies cannot be accurately represented by a *sharp* template.

We refer to [Postnikov 2001, Alekseevsky 2003, Huckemann 2010] for mathematical details on what follows. The mathematical ingredients indeed come from well-known theories of Differential Geometry and Topology. We show how these geometric and topological theories combine to formalize the template computation

algorithm. We present below only the concepts needed for our development, i.e. the definitions that will then be used in Section 6.3.

6.2.1 Geometrization of the action of diffeomorphisms on images

We consider the template computation designed with non-linear registration of brain images using diffeomorphisms [Younes 2012]. We adopt the point of view of images as square-integrable functions I over the compact domain Ω, i.e. we write $I \in L_2(\Omega)$, where $L_2(\Omega)$ is a Hilbert space. The corresponding L_2 distance is invariant by volume preserving diffeomorphisms (volumorphisms). In order to illustrate the following mathematical concepts, we will often use the toy Hilbert space $\mathbb{R}^2$ where one point schematically represents one image, see Figure 6.1. Illustrations on $\mathbb{R}^2$ facilitate the representation of the geometry in the abstract space of images.

Action of diffeomorphisms on (brain) images A *diffeomorphism of* Ω is a differentiable map $\phi : \Omega \to \Omega$ which is a bijection whose inverse $\phi^{(-1)}$ is also differentiable. The set of diffeomorphisms of Ω forms an infinite dimensional Lie group, denoted $\mathrm{Diff}(\Omega)$. The Lie group of diffeomorphisms $\mathrm{Diff}(\Omega)$ acts on the space of images $L_2(\Omega)$:

$$\rho : \mathrm{Diff}(\Omega) \times L_2(\Omega) \to L_2(\Omega), \qquad (\phi, I) \mapsto \phi \cdot I = I \circ \phi^{-1} \tag{6.1}$$

The map ρ is differentiable and obeys the axioms of Lie group action: $Id \cdot I = I$ and $(\phi_1 \circ \phi_2) \cdot I = \phi_1 \cdot (\phi_2 \cdot I)$, where Id is the identity.

In contrast to the Lie group actions of [Miolane 2016], the action of the diffeomorphisms on the brain images is not isometric. In practice, it will be close to isometric because of the regularization used in the registration. In the following the diffeomorphisms and the images are considered to be smooth enough [Michor 2013]: the diffeomorphisms are taken to be C^∞ and the images are taken to be C^∞_C, which is C^∞ defined on the compact support Ω.

Figure 6.1(b) illustrates the above notions on the toy Hilbert space $\mathbb{R}^2$, where one point represents schematically one image I. The action of a diffeomorphism ϕ on I is represented by the blue curved arrow, i.e. by the action of a 2D rotation. The action transforms the image I into another image $\phi \cdot I$, i.e. into a different point in the Hilbert space. Figure 6.1(a) shows an image I and its diffeomorphic deformation. Both illustrations describe the registration of images that will be used in the template computation algorithm.

Orbit O_I of a (brain) image I The *orbit* O_I of a brain image I is defined as all images reachable through the action of diffeomorphisms on I:

$$O_I = \{I' \in L_2(\Omega) | \exists \phi \in \mathrm{Diff}(\Omega) \text{ s.t. } I' \circ \phi^{-1} = I\}. \tag{6.2}$$

The blue dotted circle on Figure 6.2(b) represents O_I the orbit of I. This orbit defines a submanifold of images: in this toy illustration, the submanifold is the blue

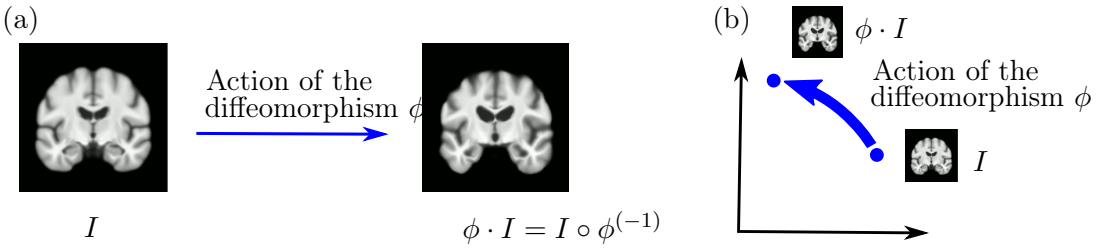


Figure 6.1: Action of a diffeomorphism ϕ on a brain image I. (a) the brain image before and after the action of ϕ, (b) schematic representation of the action of ϕ on the brain image I, represented as a dot in $\mathbb{R}^2$.

dotted circle. The four images of Figure 6.2(a) belong to the orbit of I and are represented as blue points on the dotted circle of Figure 6.2(b). The red point on Figure 6.2(b) represents $O_{I'}$, the orbit of another image I'. This orbit contains only one point and is a submanifold of dimension 0.

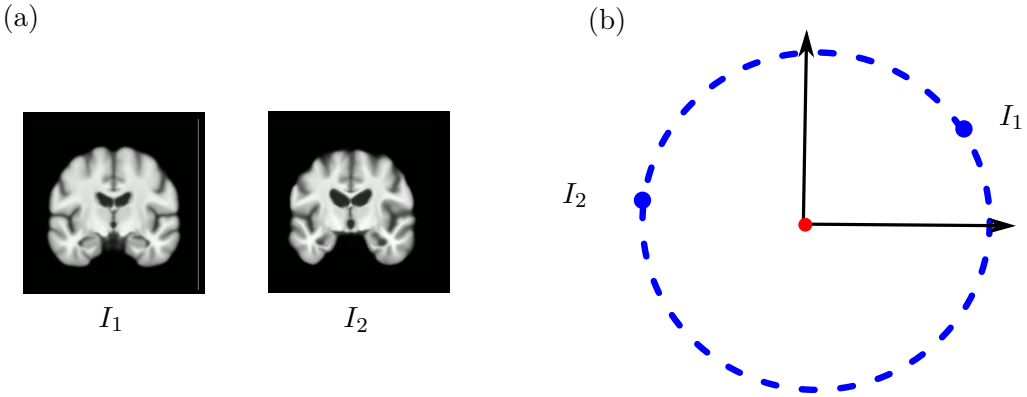


Figure 6.2: Orbit of a brain image. (a) Two images belonging to the same orbit: I_2 is a diffeomorphic deformations of I_1. (b) The orbit is represented as the blue dotted circle and the two images I_1, I_2 are points on this circle. Their isotropy groups are conjugate.

Isotropy group G_I of a (brain) image I The notion of "isotropy group" dictates the dimension of a given orbit, i.e. whether we have a 1-dimensional submanifold like the blue dotted circle or a 0-dimensional submanifold like the red point on Figure 6.2. The *isotropy group* G_I of a brain image I is defined as the subgroup of $\mathrm{Diff}(\Omega)$ formed by the diffeomorphisms that leave I unchanged:

$$G_I = \{\phi \in \mathrm{Diff}(\Omega) | I \circ \phi^{-1} = I\}. \tag{6.3}$$

It describes the intrinsic symmetry of the image I: the more symmetric is I, the larger its isotropy group. All images on the same orbit have conjugate isotropy

groups. Moreover, the isotropy group (also called the stabilizer) and the orbit of an image are linked by the orbit-stabilizer theorem: $O_I \sim \text{Diff}(\Omega)/G_I$. This means that the larger the isotropy group (and thus, the more symmetry the image has), the smaller the orbit.

The toy example of $\mathbb{R}^2$ where the diffeomorphisms are represented by rotations illustrates the notion, see Figure 6.3. The isotropy group of the image at the blue point on Figure 6.3(b) is the identity. Only the identity leaves this image at the same place. The isotropy group of the image represented by the red point is the whole Lie group of 2D rotations. Any rotation leaves this point invariant. Going back to real diffeomorphisms, the isotropy group of a uniform image, i.e. I constant map over Ω, is the whole group $\text{Diff}(\Omega)$. Figure 6.3(a) shows two brain images: the isotropy group of I_1 is larger than the isotropy group of I_2, in the sense of the inclusion. As a consequence, the orbit of I_1 has smaller dimension than the orbit of I_2.

Figure 6.3: (a) Two brain images with different isotropy groups: the left image shows more symmetry, thus a larger isotropy group, and the right image as more asymmetric details and hence smaller isotropy group. (b) The isotropy group is linked to the dimension of the image's orbit. I_2 has a smaller isotropy group, it is represented as a blue point which has a circle as orbit. I_1 has larger isotropy group: it is represented as the red point at $(0, 0)$ whose orbit is itself.

6.2.2 From geometry to topology

We have introduced notions of differential geometry, which we now link to the topology of the brain images of the database.

Topology of (brain) images The *topology of a brain image* I is defined as the topology of its level sets (surfaces of Ω with constant intensity), i.e. the level sets' properties that are preserved under smooth deformations [Gerber 2012]. It refers to properties such as the number of holes, or the number of connected parts, in the surface, see Figure 6.4(a). The topology is precisely the set properties conserved by

the action of diffeomorphisms on the brain image I: two images in the same orbit have same topology.

How is the topology embedded in the geometric framework of the previous subsection? Two images I and I' that are diffeomorphic deformations of each another, i.e. $\phi \cdot I' = I$ for a given ϕ, have the same topology. They are also on the same orbit O_I. Thus, we can consider that the orbit O_I itself represents the topology of image I (and I'). As a consequence, the set of orbits $Q = \{O_I | I \in L_2(\Omega)\}$ is the set of the topology of the images. In Differential Geometry, this set is called the *quotient space of* $L_2(\Omega)$ *by the action of Diff*(Ω). One point in the quotient space is an orbit in $L_2(\Omega)$, i.e. a set of images with a defined topology. We say that the quotient space Q is the space of the brain images' topologies.

Figure 6.4(b) shows how the space of images $\mathbb{R}^2$ is partitioned into orbits: blue circles and one red "singular circle", the red point at $(0, 0)$. Figure 6.4(b) also shows $\mathbb{R}_+$, the quotient space of $\mathbb{R}^2$ by the group of 2D rotations, which schematically represents the quotient space of the space of brain images $L_2(\Omega)$ by the Lie group of diffeomorphisms Diff(Ω). Each of the four blue circles in $\mathbb{R}^2$ becomes a blue point in the quotient space $\mathbb{R}_+$.

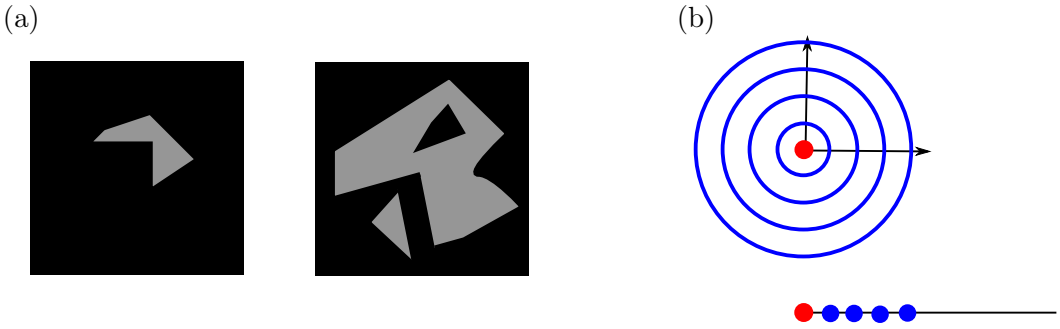


Figure 6.4: (a) Two images with different topologies: one cannot be diffeormorphically deformed to match the other. (b) Top: schematic representation of the space of images partitioned into orbits. Bottom: schematic representation of the quotient space of brain images $L_2(\Omega)$ by the Lie group of diffeomorphisms Diff(Ω).

Gathering brain images with similar topologies: orbit types By definition, two brain images are of the same *orbit type* if their isotropy groups are conjugate subgroups in the Lie group of diffeomorphisms. In particular, brain images that belong to the same orbit have same orbit type. The type corresponding to the smallest isotropy group - in the sense of the inclusion among the subgroups of the diffeomorphism group - is sometimes called the principal type [Alekseevsky 2003]. We use this appellation in this paper. For example, if an orbit has type the identity of the group of diffeomorphisms, then it is necessarily of principal type. Equivalently, the orbits of principal type are called *principal orbits*. Other orbits are called *singular orbits*.

The blue circles on Figure 6.4(b) have same orbit type: the images on these orbits have the identity $\{Id\}$ of the Lie group as isotropy group. They are of principal type since the identity $\{Id\}$ is the smallest subgroup - in the sense of the inclusion - of the group of rotations.

Stratification of the space of topologies In the space of brain images, we can gather orbits of same orbit type: we gather the blue circles of Figure 6.4(b) into $\mathbb{R}^2 \setminus \{(0,0)\}$ on the one hand, and keep the red dot $(0,0)$ on the other hand. In this sense, the orbit type itself is a submanifold of the space of brain images: $\mathbb{R}^2 \setminus \{(0,0)\}$ or $(0,0)$ in the schematic brain images space $\mathbb{R}^2$. Furthermore, these orbit type submanifolds form a *stratification*, meaning they fit together in a particularly nice way in the space of brain images.

The quotient space Q is also naturally partitioned into manifolds, and this partitioning is also a stratification. All in all, Q is *not* a manifold, but Q composed of manifold pieces, and those pieces being called strata. Furthermore, there is a partial ordering of the strata in the quotient space, using the inclusion [Hughes 1996].

Figure 6.4(b) shows the orbits grouped by orbit type: the color blue denotes one orbit type and the color red another orbit type. We see for example that $Q = \mathbb{R}_+$ is stratified into one stratum being $\mathbb{R}_+^*$ - corresponding to the stratum $\mathbb{R}^2 \setminus \{(0,0)\}$ in the space of brain images - and one stratum being $\{0\}$ - corresponding to the stratum $(0,0)$ in the space of brain images.

6.2.3 Geometry of generative model and estimation procedure

We introduce the procedure whose statistical analysis is the main purpose of this paper: the brain template estimation of neuroimaging.

Generative model The n brain images $I_1, ..., I_n$ are interpreted with a generative deformable model:

$$I_i = \phi_i \cdot T + \varepsilon_i, \qquad i = 1...n, \tag{6.4}$$

where each image I_i is a diffeomorphic deformation ϕ_i of a unique brain anatomy T, to which variability ε_i is added. The parameter T is called the template, it is the brain anatomy shared by the population. The transformations ϕ_i's and the noises ε_i's are i.i.d. realizations of random variables. The transformations ϕ_i's follow a law that is a generalization of the Gaussian for Lie groups, i.e. a Riemannian Gaussian on the Lie group of diffeomorphisms and the ε_i's represent Gaussian noise on the space of images. We denote σ^2 its variance. Definitions of distributions on finite dimensional Riemannian manifolds are taken from [Pennec 2006] and the Gaussian distributions for infinite dimensional spaces from [Lifshits 1995].

The model can be interpreted by a three step generative procedure illustrated schematically in Figure 6.5. First, there is only the template T. Second, the template T is deformed with the diffeomorphism ϕ_i and gives a brain image $\phi_i \cdot T$. Third, measurement noise is added through ε_i, which gives the brain image I_i actually observed in our database.

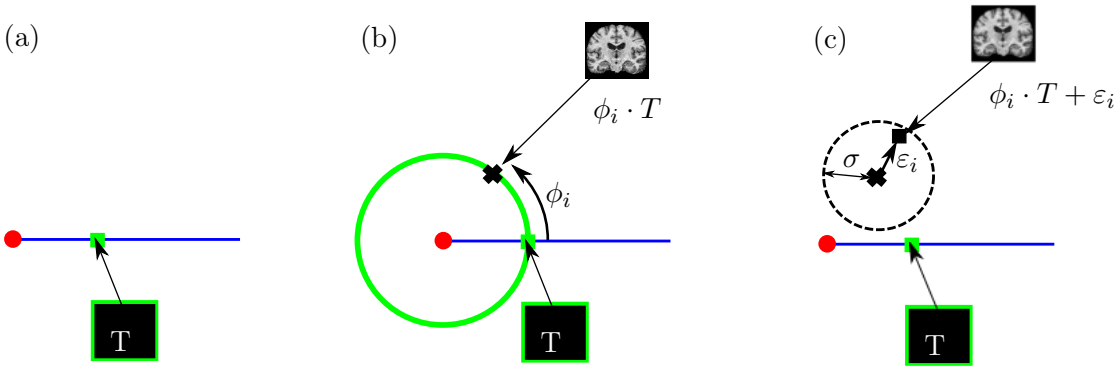


Figure 6.5: Schematic illustration of the generative model of the brain images data. As before, the space of brain images is represented by the plane $\mathbb{R}^2$. (a) First step of the generative model: generate a brain anatomy. One usually assumes that there is a unique brain anatomy, which is by definition the brain template T, in green. (b) Second step of the generative model: generate a deformation $\phi_i \in \mathrm{Diff}(\Omega)$ which is used to deform the template. The brain image $\phi_i \cdot I$ belongs to the template's orbit, represented by the green circle. (c) Third step of the generative model: generate noise ε_i in the space of images. The brain image $\phi_i \cdot T + \varepsilon_i$ does not belong to the template's orbit anymore.

Computing the template: an estimation procedure Computing the brain template amounts to invert the generative model: given the data, we want to estimate the parameter T. The transformations ϕ's are hidden variables of the model. The natural statistical procedure to estimate T in this context is the Expectation-Maximization (EM) algorithm [Dempster 1977]. The EM is an iterative procedure that maximizes the log-likelihood of the generative model with hidden variables. As such, we note that the EM gives a asymptotically unbiased and consistent estimation of the brain anatomy T.

But in practice, one does not use the EM algorithm because the EM is computationally too expensive, especially when dealing with tridimensional brain images. Thus, most neuroimaging pipelines rely on *an approximation of* the Expectation-Maximization algorithm to estimate the brain anatomy. *This approximation is the template estimation procedure described in the Introduction and the main purpose of this paper.* The procedure runs as follows. One initializes the estimate with $\hat{T} = I_1$, i.e. one of the brain images from the database. Then, one iterates the following two steps until convergence:

$$(1) \quad \hat{\phi}_i = \underset{\phi \in \mathrm{Diff}(\Omega)}{\mathrm{argmin}} \, d_{L_2(\Omega)}(\hat{T}, \phi \cdot I_i), \quad \forall i \in \{1, ..., n\},$$

$$(2) \quad \hat{T} = \underset{T \in L_2(\Omega)}{\mathrm{argmin}} \sum_{i=1}^{n} d_{L_2(\Omega)}(T, \hat{\phi}_i \cdot I_i)^2.$$

Step (1) is an estimation $\hat{\phi}_i$ of the hidden observations, the diffeomorphisms ϕ_i,

and an approximation of the E-step of the EM algorithm. In practice, each brain image I_i is registered to the current template estimate and the $\hat{\phi}_i$ is the result of this registration. Step (2) is the M-step of the EM algorithm: the maximization of the surrogate in the M-step amounts to the maximization of the variance of the projected data. This computes the updated template estimate, as the mean intensity of the subjects images I_i, deformed with the mean deformation of the $\hat{\phi}_i$'s. We remark that adding a regularization term in Step (2) gives the framework of "unbiased brain template computation" of [Joshi 2004].

Estimation procedure interpreted as the Fréchet mean in the quotient space The procedure of brain template estimation converges to a local minimum because it decreases at each step a cost bounded below by zero. The estimator of the brain anatomy computed with this procedure is:

$$\hat{T} = \underset{T \in L_2(\Omega)}{\operatorname{argmin}} \sum_{i=1}^{n} \min_{\phi \in \mathrm{Diff}(\Omega)} d^2_{L_2(\Omega)}(T, \phi \cdot I_i). \tag{6.5}$$

We analyze the expression of the brain anatomy estimate in Equation 6.5 within the geometric framework introduced previously. The term $\min_{\phi \in \mathrm{Diff}(\Omega)} d^2_{L_2(\Omega)}(T, \phi \cdot I_i)$ in Equation 6.5 is the distance in the quotient space between T and I_i. Thus, we recognize in Equation 6.5 the Fréchet mean on the quotient space. The Fréchet mean on manifolds [Pennec 2006] is the point that minimizes the sum of the squared distances to the data in the quotient space. All in all, one projects the probability distribution function of the I_i's from $L_2(\Omega)$ to $L_2(\Omega)/\mathrm{Diff}(\Omega)$ and computes its "expectation". Note that we represent the projection of the images in the quotient space, i.e. their orbit or equivalence class under the action of diffeomorphisms, by the registered representatives of these classes.

Figure 6.6 gives the geometric interpretation of the iterative procedure in the context of the geometric framework. The registration step amounts the alignment of the n subject images by transporting them on their orbit (see Figure 6.6(b)), i.e. projecting them in the quotient space (see Figure 6.6 (c)). The averaging step takes the Fréchet mean of the n images projected in the quotient space (see Figure 6.6 (d)). Iterating these steps until convergence creates the template's estimate $\hat{T}$.

Note on the terminology: template and template estimate We clarify the terminology used in this paper. On the one hand, there is the parameter T of the generative model, called the *template*, which represents the brain anatomy shared by the population. On the other hand, there is its *estimate* $\hat{T}$, which is the result of the procedure. In other papers, $\hat{T}$ may be called the template directly.

6.2.4 Geometry of the template estimator's evaluation

This subsection gives the definitions of the estimation's statistical properties that we will investigate for the template computation procedure, through the lense of the geometry.

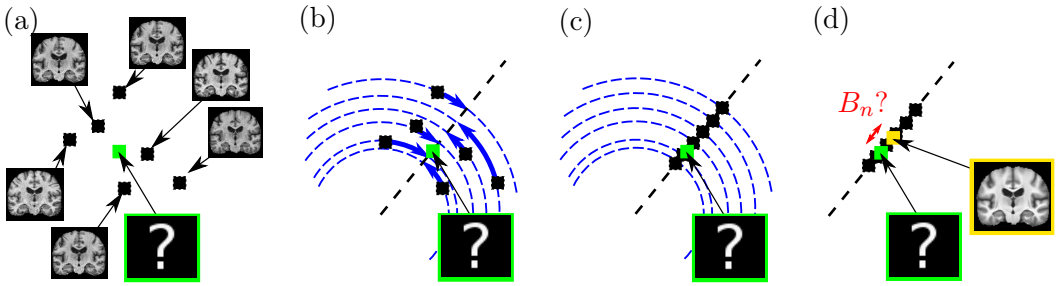


Figure 6.6: Geometrization of an iteration of the template's computation: (a) n subjects images (black squares); (b) the n images are registered, they travel on their orbit (the blue circles) to get aligned; (c) registered images; (d) the empirical brain template $\hat{T}$ (in yellow) is computed as the Fréchet mean of the n registered images. How far is it from the unique anatomy T of the generative model (in green): can we quantify B_n?

Definition of the asymptotic bias B_∞ We evaluate $\hat{T}$ as an estimator of the unique brain anatomy T (see Figure 6.6 (d)) given n observations I_i, $i = 1...n$. We consider two measures of the accuracy of this estimator: its variance V_n^2 and bias B_n, which are defined as:

$$V_n^2 = \mathbb{E}((\hat{T} - \mathbb{E}(\hat{T}))^2) \qquad \text{and} \qquad B_n = \mathbb{E}(\hat{T} - T).$$

In particular, the bias for n images B_n is illustrated on Figure 6.6. We are interested in the measures of accuracy in the asymptotic behavior $n \to +\infty$, i.e. when the number of brain images goes to infinity. In other words, we want to study the behavior of the procedure in an ideal situation. One would expect that the procedure converges to the template T, i.e. to the brain anatomy it is originally designed to estimate.

It is assumed that the estimator converges so that its variance is asymptotically zero: $V_\infty^2 = 0$. However, the authors of [Allassonnière 2015a, Miolane 2015a] show that the template estimate $\hat{T}$ is inconsistent by showing that its asymptotic bias is non zero: $B_\infty \neq 0$. In other words, for an infinite number of brain images $n \to +\infty$, the estimate converges, but not to the brain anatomy T of the generative model that it was designed to compute.

Asymptotic bias B_∞ and curvature of template's orbit This issue has a geometric origin that is explained in [Miolane 2015a, Allassonnière 2015a] using the framework introduced in the previous subsections. The authors of [Miolane 2015a] quantify the template's asymptotic bias B_∞ using the geometry of a general manifold with Lie group action, and induced quotient space. Their results involve a definition of orbits' curvature, the mean curvature vector H, which we introduce here briefly in the context of an abstract manifold M, which will later represent the space of brain images. The *second fundamental form* h of a submanifold O of M is defined on

$T_XO \times T_XO$ by $h(v,w) = (\nabla_v w)^\perp \in N_XO$, where $(\nabla_v w)^\perp$ denotes the orthogonal projection of covariant derivative $\nabla_v w$ onto the normal bundle. The *mean curvature vector* H *of* O is defined as: $H = Tr(h)$. Intuitively, h and H are measures of extrinsic curvature of O in M.

Now we can quantify the asymptotic bias B_∞ of the template estimation. It was observed experimentally that such a bias was dependent on the amount of measurement error on the observations [Allassonnière 2007]. Therefore, the authors of [Miolane 2015a] compute a Taylor expansion of the asymptotic bias B_∞ around the noise $\sigma = 0$, in the case of a finite dimensional manifold and isometric Lie group action [Miolane 2016]:

$$B_\infty = \frac{\sigma^2}{2} H(T) + \mathcal{O}(\sigma^3). \tag{6.6}$$

where $H(T)$ denotes the mean curvature vector of the template's orbit. There is no asymptotic bias - i.e. $B_\infty = 0$ and the template's estimation is consistent - when there is no measurement error $\sigma = 0$.

We remark also that the coefficient $H(T)$ depends on the template T that we are estimating. We investigate this dependency in the geometric framework, in particular with respect to the stratification of the quotient space of Subsection 6.2.2. We assume that there exists a fixed point o of the Lie group action, i.e. a point that is left unmoved by the whole Lie group. We fix this point o, and we consider the orbit O_T of the template T. As the action is isometric, we can easily prove that the orbit belongs to a geodesic sphere S_d with center o and radius d. Now, a geodesic sphere of radius d in a manifold - like a hypersphere of radius d in $\mathbb{R}^m$ - has mean curvature vector: $||H(T)|| = \frac{(m-1)}{d}$ and a closed hypersurface in a manifold of dimension m obtained by perturbing the geodesic sphere has: $||H(T)|| = \frac{(m-1)}{d} + O(1)$ [Pacard 2005, Lytchak 2010]. In other words, the distance d of the template to the singularity o also governs the asymptotic bias. If we write the template's bias in the units of d, then the asymptotic bias depends on $\left(\frac{\sigma}{d}\right)^2$. In other words, the distance of the template to the singularity o, at the scale of the noise σ governs the asymptotic bias B_∞.

Figure 6.7 shows the intuition behind the above development and why these geometric parameters - especially the mean curvature vector of the template's orbit $H(T)$- are responsible for the asymptotic bias B_∞. On Figure 6.7(a), $\mathbb{R}^2$ schematically represents the space of brain images - thus the black squares represent brain images from the database, the green square is the brain template - and the green circle is the orbit of the brain template. The dotted circles, that have their centers on the template's orbit, represent the probability distribution of the (2D isotropic) Gaussian noise in the generative model. More precisely, they represent the level set at σ of the noise distribution. The curvature $H(T)$ controls the area in grey on Figure 6.7, which is the area inside the Gaussian level set that is outside the template's orbit. This area is greater that the area inside the template's orbit. As a consequence, the probability that the brain images are generated "outside" the template's orbit is higher than the probability that they are generated inside the template's orbit.

Figure 6.7(b) shows the registration step of the brain template estimation: each brain image follows the curved blue arrow to be registered. In this step, there will be a higher probability that the registered images are away from the template T, as if repulsed from the singularity around which the orbits warp. When one averages the registered images, one sees that the template's estimate becomes biased as it will systematically give an image that is further away than T from the quotient space's singularity, i.e. from the red dot.

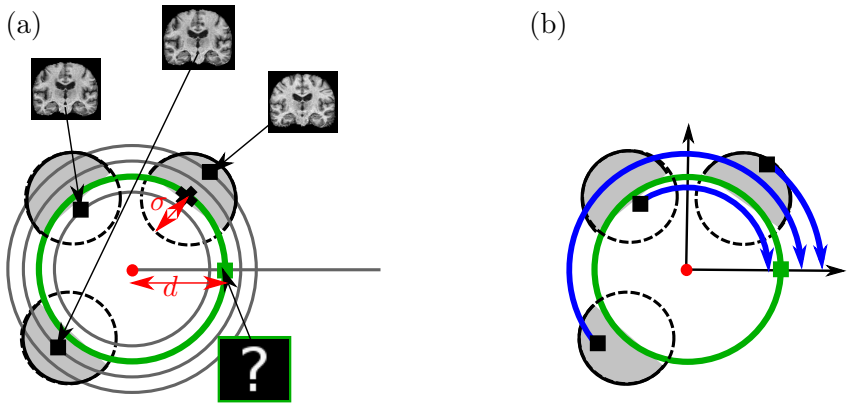


Figure 6.7: Schematic illustration of the asymptotic bias in the template computation algorithm of neuroimaging. (a) The images generated with the model described above have a higher probability to be "outside" (with respect to the curvature) of the template's orbit. (b) The registration during template's estimation aligns the images. The distribution of images is unbalanced with respect to the real template.

Quantifying the asymptotic bias B_∞ of the brain template in neuroimaging So far, we have explained the asymptotic bias B_∞ of the brain template estimate with differential geometry. How does it apply to neuroimaging? In [Miolane 2016], the the manifold M and the Lie group considered are finite dimensional. In neuroimaging, the manifold is the space of brain images $L_2(\Omega)$ and the Lie group is the group of diffeomorphisms $\mathrm{Diff}(\Omega)$, which are both infinite dimensional, so one could wonder if we can apply the geometry of [Miolane 2015a].

This paper assumes that we can, because there are indications in the literature that the asymptotic bias B_∞ appears in the same fashion in infinite dimensions. First, the work of [Miolane 2016] studies the behavior of the asymptotic bias when the dimension of the manifold m increases and shows that the asymptotic bias B_∞ of the template estimate increases. Then, the work of [Allassonnière 2015a] shows that there exists an asymptotic bias in infinite dimensions. Ultimately, [Bigot 2010] gives a lower bound of the asymptotic bias for shapes of curves in 2D, where terms depend on derivatives of the functions representing the curve. These derivatives can be interpreted as the derivative of the action of translations, which leads to the curvature of the orbit of the given function under the translations' action. As

a consequence, we assume that the intuition provided by [Miolane 2016] applies to neuroimaging. In particular, we assume that the asymptotic bias of the brain template depends on the ratio $\left(\frac{\sigma}{d}\right)^2$, where again: d is the distance in the quotient space from T to a singularity, and σ is the standard deviation of the noise on the data I_i in $L_2(\Omega)$.

6.3 Computational representation of geometry and topology

We want quantify the asymptotic bias B_∞ on the brain template estimation, using the ratio $\left(\frac{\sigma}{d}\right)^2$ introduced in the previous section. This section shows how the topology of brain images - represented by Morse-Smale complexes - enables to estimate the geometric parameters d and σ.

6.3.1 Definition of Morse-Smale complexes for (brain) images

We introduce a computational representation of images' topology, and more precisely of their intensity level sets' topology.

Morse-Smale (intensity) functions A real-valued smooth map $I : \Omega \to \mathbb{R}$ is a *Morse function* if all its critical points are non-degenerate (the Hessian matrix is non-singular) and no two critical points have the same function value. The intensity function I representing a bi- or tri-dimensional brain image is a Morse function, at least after a convolution with a smoothing Gaussian [Boscain 2012]. In the following, I represents a brain image. *Morse theory* traditionally analyzes the topology of a manifold by studying the Morse functions on that manifold. Here, the manifold is known: it is the image domain Ω. However, we are not interested in the topology of Ω but rather in the topology of the functions themselves, that is: we would like to know the distribution of their critical points. Figure 6.8 shows a 2D slice of a 3D brain image I, where the intensity is represented as the height, to better emphasize its maxima and minima: the maxima are in red and the minima in blue.

We introduce the notions of integral lines, ascending and descending manifolds that are needed to define Morse-Smale (intensity) functions. An *integral line* is a maximal path in the image domain Ω whose tangent vector correspond to the intensity gradient ∇I, the gradient of I, at every point. This notion comes from autonomous ordinary differential equation, where it represents the trajectory of a system verifying:

$$\frac{dx}{dt}(x) = \nabla I(x) \tag{6.7}$$

Each integral line starts and ends at critical points of I, where the gradient ∇I is zero. *Ascending $A(x_i)$ and descending $D(x_i)$ manifolds* of an extremum x_i are

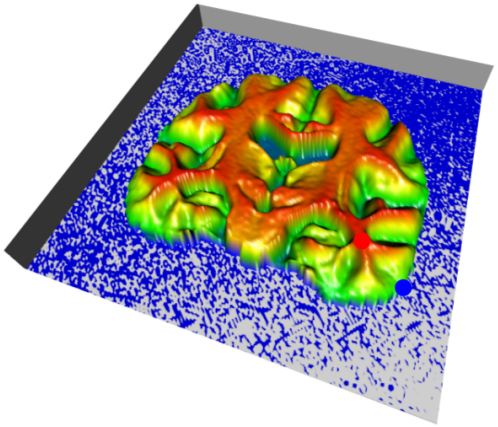


Figure 6.8: Intensity I on a 2D brain image visualized as height: maxima are in red, minima in blue. This is a Morse-Smale function $I : \Omega = \mathbb{R}^2 \to \mathbb{R}$.

defined as:

$$A(x_i) = \{x \in \Omega \quad | \quad \text{The integral line going through } x \text{ ends at } x_i\}$$
$$D(x_i) = \{x \in \Omega \quad | \quad \text{The integral line going through } x \text{ starts at } x_i\}$$

Take the two manifolds $A(x_i)$ and $D(x_i)$ in Ω and assume they intersect at a point $p \in \Omega$. Let T_A (resp. T_D) denotes the set of all vectors tangent to $A(x_i)$ (resp. $D(x_i)$ at p. If every vector in Ω is the sum of a vector in T_A and a vector in T_D, then $A(x_i)$ and $D(x_i)$ are said to intersect transversely at the point p. The intensity function I defining the brain image is *Morse-Smale* if the ascending and descending manifolds only intersect transversely. We assume in the following that all brain images are Morse-Smale.

Morse-Smale complex and persistence The *Morse-Smale complex* of a Morse-Smale function is the set of intersections $A(x_i) \cap D(x_j)$, over all combinations of extrema (x_i, x_j) [Gyulassy 2008]. The Morse-Smale complex includes regions (i.e., sub-manifolds of Ω) of dimensions 0 through D, where D is the dimension of the domain Ω, i.e. $D = 2$ or $D = 3$ for our purposes. The *Morse-Smale (MS) complex of* I is a partition of the domain Ω into regions defined by the set of integral lines that share common starting and ending points. The interior of each region is monotonic with respect to the intensity I: a region contains no critical points and has a single local minimum and maximum on its boundary, see for example Figure 6.10(ii) where the maximum $I_{\max}$ and minimum $I_{\min}$ are shown on the boundary of the grey region. The MS complex can also be seen as a graph on the brain image domain Ω whose nodes are the critical points of the brain image intensity.

The *persistence of a critical point* x_i *of* I is the amount of change in intensity I required to remove this critical point:

$$p(x_i) = |I(x_i) - I(n(x_i))| \tag{6.8}$$

where $n(x_i)$ is the critical point closest to x_i *in intensity*, among the critical points connected to x_i by an integral line [Gerber 2012]. The persistence of x_i is a measure of its significance as a critical point, i.e. importance of the topological feature. Figure 6.9 illustrates the definition of persistence on a 1D example. The function represented has 4 critic points: two minima and two maxima. The figure shows how they pair, as well as their persistence. On the x-axis, colors show the regions of the corresponding 1D Morse-Smale complex.

Beside this usual definition of persistence of a critical point, we define here the *persistence of a region of the Morse-Smale complex* as the amount of change in intensity required to remove this region from the MS complex, and more precisely:

$$p(\text{region}) = |I_{\max} - I_{\min}| \tag{6.9}$$

where $I_{\max}$ and $I_{\min}$ are respectively the maximum and the minimum in intensity of this region. In contrast to the definition of the persistence of a critical point, we do not rely on the saddle points, but only on the extrema i.e. the minima and maxima.

Hierarchy of Morse-Smale complexes The notion of persistence of a region enables the definition of a *hierarchy of MS complexes* of one brain image I [Gyulassy 2008, Gerber 2012]. One uses the ordering given by persistence to successively remove topological features from the image I. One starts with the MS complex of the brain image I defined above and one recursively removes the critical points with minimal persistence. This leads to a nested series of successively simplified Morse-Smale complexes. At each level, some of the MS regions are merged into a single region. Ultimately the Morse-Smale complex consists of only one region which is the entire domain Ω.

The persistence introduces a notion of scale at which the Morse-Smale complex of I is considered. One keeps only the nodes whose persistence is above the threshold. Figure 6.9 shows that the one dimensional domain is partitioned differently if one takes a threshold p below p_1 or between p_1 and p_2. We say that a Morse-Smale complex is represented at a given persistence level. At the scale of the persistence threshold p, the intensity is considered monotonic on each region of the MS.

We note that this MS hierarchy is different from a Gaussian scale space (GSS) hierarchy of images [Reininghaus 2011]. The latter takes critical points across smoothing scales and not across persistence levels.

6.3.2 Computing Morse-Smale complexes of (brain) images in practice

The previous definitions are relevant to (continuous) Morse-Smale theory and apply strictly for a continuous intensity function I. Nevertheless, the MS complex, introduced in terms of ascending and descending manifolds, can be computed for a discrete image I, i.e. for the brain images of our database.

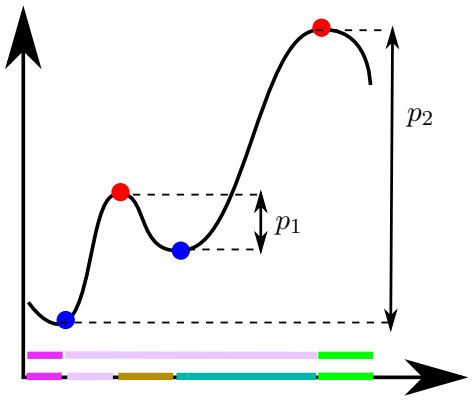


Figure 6.9: Persistences $p_1 < p_2$ of two pairs min-max. A threshold $p_1 < p < p_2$ divides the domain into 3 regions (pink, violet, green), while $p < p_1$ divides it in 5 regions (pink, violet, brown, turquoise, green).

Computing the Morse-Smale of a brain image We compute the Morse-Smale complex of a brain image, which will later be the brain *template* image. Our input are $\{x_i, I_i\}$, i.e. the intensity values $\{I_i\}$ on a grid $\{x_i\}_i$ of Ω. We compute the integral lines of the intensity gradient, which we then gather to get the regions of the Morse-Smale complex. For each element of the grid x_i, following the gradient ∇I leads to computing the integral line going through x_i and in particular its starting and ending points [Gerber 2012]. The domain Ω can be approximated via a k nearest-neighbor graph and one computes the integral lines by considering the connectivity of the graph. Then, elements x_i's with same starting and ending points belong to the same Morse-Smale region. This gives the partition of the domain Ω and therefore the Morse-Smale complex. We remark that x_i's necessarily belong to a 3-dimensional (for a tridimensional image) component of the Morse-Smale complex because the 0-, 1- and 2-dimensional components have measure zero.

Figure 6.10 shows the Morse-Smale complex of the 2D slice of a 3D brain image for level of persistence of $p = 0.1$. The image's 2D domain is divided in different regions, represented by the different colors. The quadrant shows part of the underlying Morse-Smale graph. The red dot represents a maximum in intensity, and the blue dot a minimum of intensity. They are nodes of the underlying graph on the domain Ω.

Morse-Smale (MS) graph and labeled Morse-Smale (MS) graph There are two ways of representing the Morse-Smale graph corresponding to the computed Morse-Smale complex. Both will be useful for analyzing the template's asymptotic bias. One can consider the graph as the set of nodes and edges, without any intensity information at the nodes. We simply call this graph the *Morse-Smale (MS) graph*: this is the graph illustrated on Figure 6.10(i). Alternatively, one can label the nodes

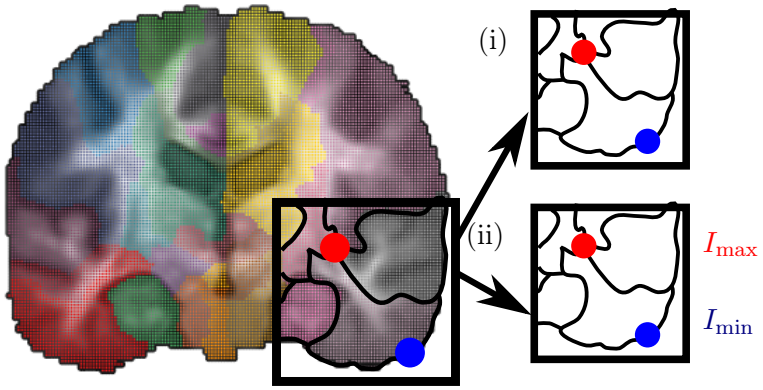


Figure 6.10: Computational representation of the geometry: Regions on a 2D domain, induced by a MSC with threshold $p = 0.1$: (i) The MS graph represents the isotropy group's class of the image, (ii) the labeled MS graph represents the image's orbit under the diffeomorphisms.

with the intensity information. We call this graph the *labeled Morse-Smale (MS) graph*: this is the graph illustrated on Figure 6.10(ii).

6.3.3 Template's computation and Morse-Smale complexes

We have defined the Morse-Smale complex and show how to compute it in practice. Now we show how the MS complex of an image can represent its geometry and in particular isotropy group. We first link the image's isotropy group with the image's gradient and then the image's gradient with the Morse-Smale graph, through two Lemmas.

Lie algebra of the isotropy group and intensity gradient of the brain template We show a lemma that links the Lie algebra $\mathfrak{g}_I$ of the brain template isotropy group to the gradient of its intensity.

Lemma 6.3.1 *Let $\mathfrak{g}_I$ be the Lie algebra of the isotropy group G_I of image I. The infinitesimal transformations of G_I, i.e. the elements of $\mathfrak{g}_I$, are vector fields on Ω, satisfying:*

$$\forall x \in \Omega, \quad \nabla I(x)^T . v(x) = 0 \tag{6.10}$$

Proof The proof is given in Appendix C.

The vector fields in the Lie algebra $\mathfrak{g}_I$ of the isotropy group of I are necessarily perpendicular to the intensity gradient at any point x of the image's domain Ω. To understand the intuition behind this lemma, consider a uniform image, i.e. with a constant intensity. In this case, there is no restrictions a priori on the vector fields of the Lie algebra of the image's isotropy group. Thus, the isotropy group is as large

as it can be. We get that the isotropy group of an image with constant intensity is the whole group of diffeomorphisms.

We note that this lemma does not give a characterization of the vector fields in $\mathfrak{g}_I$. It gives the inclusion: $\mathfrak{g}_I \subset \{v | \forall x \in \Omega, \nabla I(x).v(x) = 0\}$. Thus, it allows to control the complexity of the isotropy group's Lie algebra.

Intensity gradient and MS graph We now present a lemma showing that the MS graph can be used to computationally represent the isotropy group of an image.

Lemma 6.3.2 *Take two images I_1 and I_2. Assume that their MS graphs are the same, regardless of the nodes' positions. Then, there exists a diffeomorphism ψ and a function κ such that: $\nabla\psi(x).\nabla I_1 \circ \psi(x) = \kappa(x)\nabla I_2(x), \forall x \in \Omega$.*

Proof The proof is given in Appendix C.

In others words: if two images I_1, I_2 have the same MS graph, then I_1 can be diffeomorphically deformed so that its intensity gradient is parallel at every point to the intensity gradient of I_2.

From Lemma 6.3.1, the sets $\{v_1 | \forall x \in \Omega, \nabla I_1(x).v_1(x) = 0\}$ and $\{v_2 | \forall x \in \Omega, \nabla I_2(x).v_2(x) = 0\}$ control the isotropy groups of I_1 and I_2. If I_1 and I_2 have same MS graph, any vector field in the first set can be diffeomorphically deformed to get a vector field in the second set, and conversely. As a consequence, the MS graph represents the image's isotropy group. From Section 6.2, we know that the isotropy group controls, in turn, the orbit's type of the image i.e. to which stratum the image belongs.

Furthermore, we note that we could have considered the labeled MS graph of the image, i.e. the MS graph with intensities at the nodes, see Figure 6.10(ii). The labeled MS graph controls the orbit of the image: images in the same orbit have the same topology but also same intensities.

6.4 Topology quantifies and controls the template's asymptotic bias

This Section gathers the elements of Sections 6.2 and 6.3 to quantify the asymptotic bias in the brain template computation. Section 6.2 shows how the *geometry is responsible for the asymptotic bias* in the template estimation, see in particular Equation A.15. Section 6.3 shows how *Morse-Smale complexes provide a computational representation of this geometry*, see in particular Subsection 6.3.3.

This Section uses Morse-Smale complexes to quantify, and then control, the asymptotic bias of brain template estimation in neuroimaging. More precisely, we will see how the notion of persistence defined in Section 6.3 relates to the estimation of the asymptotic bias B_∞. This shows that topology can be used as a tool in estimation theory.

6.4.1 Quantify the template inconsistency

Section 6.2 shows through Equation A.15 that the asymptotic bias B_∞ of the brain template estimator $\hat{T}$ depends on the ratio $\left(\frac{\sigma}{d}\right)^2$, where d is the distance of the template to an image with larger isotropy group, and σ is the standard deviation of the Gaussian noise in the generative model, see again Figure 6.7. We express d and σ using the Morse-Smale complexes of Section 6.3.

Understand and estimate the geometric parameter d The distance d is the distance to a brain image with larger isotropy group, measured in sum of squared differences of intensities, see Figure 6.7. How can we measure this distance d locally on the template's image? From Section 6.2, we know that the isotropy group becomes larger when the image is "more symmetric". From Section 6.3, we know that the isotropy group becomes larger when the image topology becomes simpler. Thus, the distance d is a distance in intensity from the template image to a similar image with simpler topology.

We want to express this distance locally on the template image. Modifying the intensity locally on the template image modifies the image itself and may simplify its topology. For example, modifying the intensity locally in a region of the image can suppress a min-max pair and the image becomes "more symmetric". Thus we describe the distance d locally on the image by the amount of intensity needed to be changed in this region, so that the topology is simplified.

We quantify the local intensity needed to simplify the template image's topology using the Morse-Smale complex representation of Section 6.3. Let be given the Morse-Smale complex of the template image. The intensity needed to simplify the image's topology is, by definition, the intensity needed to simplify the Morse-Smale graph. We consider the partition of the image's domain Ω induced by the Morse-Smale complex. For each region of the partition, the intensity needed to simplify the topology can be represented by the amount of intensity needed to remove the min-max pair of the region:

$$\hat{d}(\text{region}) = \hat{T}_{\max} - \hat{T}_{\min} = p_{\hat{T}}(\text{region}) \tag{6.11}$$

This quantifies the importance of the region as a representative of the brain anatomy: if the intensity difference between the region's min and max is low, then one can assume that this min-max pair has been created by chance because of the noise on the images. We see that the notion of persistence defined in Section 6.3 estimates the first geometric parameter d.

Understand and estimate the geometric parameter σ Now we turn to the second geometric parameter that causes the asymptotic bias: the standard deviation of the noise σ, see again Equation A.15 and Figure 6.7. The noise σ is a parameter of the generative model that we assume has produced the observed images of the subjects brain anatomies. The parameter σ is unknown but it can be estimated from the observed images. Since we want to compute the asymptotic bias locally,

we are interested in estimating the parameter σ locally, and for example on a region of the Morse-Smale complex of the template image. We estimate it as the average of the variability in intensity of the registered images in the region:

$$\hat{\sigma}(\text{region}) \simeq \frac{1}{\#x} \sum_{x \in \text{region}} \hat{\sigma}(x), \tag{6.12}$$

where $\hat{\sigma}(x)$ is the variability in intensity of the registered images at the voxel x, and serves as an estimate of the noise at this voxel. This quantifies the amount of noise in this region. The larger the noise, the more chances for the template to show min-max pairs that appeared by chance.

Compute the asymptotic bias using the persistence of the whitened brain template The local estimates of the geometric parameters d and σ enable us to estimate the asymptotic bias locally on a brain region:

$$\hat{B}_\infty(\text{region}) = \left(\frac{\hat{d}(\text{region})}{\hat{\sigma}(\text{region})} \right)^{-2}, \tag{6.13}$$

We emphasize here that $\hat{B}_\infty$ is an *estimate* of the asymptotic bias B_∞ (of the brain template estimation), and not an exact computation.

We link the estimate $\hat{B}_\infty$ to the definition of persistence in the Morse-Smale complex framework. First, we define the *whitened brain template estimate* $\hat{t}$ of $\hat{T}$ as:

$$\forall x \in \Omega, \quad \hat{t}(x) = \frac{\hat{T}(x)}{\hat{\sigma}(x)}. \tag{6.14}$$

In other words, we divide the brain template intensity of each voxel x by the estimation of the standard deviation of the noise at this voxel $\hat{\sigma}(x)$. This whitens the noise all over the brain template.

We assume that the critical points of $\hat{t}$ are close to the critical points of $\hat{T}$ and consider the Morse-Smale complex of the whitened template. We further assume that: $\hat{\sigma}_{\text{region}} \simeq \hat{\sigma}(\max) \simeq \hat{\sigma}(\min)$, where $\hat{\sigma}(\max), \hat{\sigma}(\min)$ are the variabilities at the respective min and max of the Morse-Smale complex. We can write:

$$\hat{B}_\infty(\text{region}) \simeq \left(\frac{\hat{T}_{\max}}{\hat{\sigma}(\max)} - \frac{\hat{T}_{\min}}{\hat{\sigma}(\min)} \right)^{-2} = \big(\hat{t}(\max) - \hat{t}(\min)\big)^{-2} = p_{\hat{t}}(\text{region})^{-2} \tag{6.15}$$

where we recognize the persistence $p_{\hat{t}}(\text{region})$ of the corresponding region of the whitened template $\hat{t}$. This links the estimation of the asymptotic bias to the persistence of the whitened template's Morse-Smale complex. This shows how a topological property of the image in fact represents a statistical property of this image as the estimate of the brain template.

Hierarchy of the whitened template The persistence of the whitened template quantifies locally the asymptotic bias, i.e. how far the brain template is from the unique brain anatomy of the generative model. Is there a statistical interpretation of the hierarchies of Morse-Smale complexes, introduced in Section 6.3? Let us consider another Morse-Smale of the whitened template's hierarchy, i.e. a Morse-Smale computed at a given persistence threshold $p_{\text{threshold}}$. There is an asymptotic bias threshold that corresponds, which we can write: $p_{\text{threshold}}^{-1/2}$. The regions kept in the new Morse-Smale are those having a persistence higher than the persistence threshold $p_{\text{threshold}}$, i.e. those having an asymptotic bias lower than the asymptotic bias threshold $p_{\text{threshold}}^{-1/2}$.

Therefore, if we can impose the topology of the brain template to match the new Morse-Smale of threshold $p_{\text{threshold}}^{-1/2}$, we control its asymptotic bias. This means that we preserve only the min-max pairs shown on the Morse-Smale graph chosen. It eliminates the min-max pairs that have been created by chance, because the noise on the images was at a similar level than the intensity signal on these regions. The next subsection explains how to impose the topology of a given Morse-Smale on the template's image.

6.4.2 Controlling the template's asymptotic bias by constraining its topology

We are given the template's image and we want to force its asymptotic bias to be below a threshold, so that it is closer to estimating the anatomy of the database, i.e. the anatomy shared by the subject brains. The development above suggests to compute the Morse-Smale complex with a persistence threshold corresponding to the desired bias threshold. Then, enforcing template's topology to match the Morse-Smale complex will control its asymptotic bias. This enforcement procedure is called "Topological denoising".

6.4.2.1 Topological denoising

Topological denoising is a procedure for smoothing an image, like our template image, while preserving topological features [Jacobson 2012, Günther 2014]. The input of the procedure is the intensity function defining the template $\hat{T} : \Omega \to \mathbb{R}$ and a MS complex with intensity values at its nodes. Enforcing the template's topology to match the MS complex means that we compute $\hat{T}' : \Omega \to \mathbb{R}$ which is a smoothed version of original template estimate $\hat{T}$ containing only the intensity min-max pairs specified by the MS complex chosen. $\hat{T}'$ should be otherwise as close as possible to the original template estimate $\hat{T}$ in terms of intensity. The values and positions of the MS extrema are preserved, while all other extrema are removed from the brain template estimate. Such procedure provides control over the topology of the brain image $\hat{T}$.

Formally, the original Topological denoising problem is written as the minimiza-

tion [Jacobson 2012]:

$$\begin{aligned}
&\text{argmin}_{T'} \sum_{x_i \in \Omega} ||\hat{T}(x_i) - T'(x_i)||^2 + \int_\Omega ||\Delta T'||^2 \\
&\quad \text{s.t. } T'(x_i) = \hat{T}(x_i) \text{ for } x_i \text{ a node of the MS complex} \\
&\qquad T'(x_j) > T'(x_i) \text{ for } x_j \text{ neighbor of } x_i \text{ and } x_i \text{ minimum} \\
&\qquad T'(x_j) < T'(x_i) \text{ for } x_j \text{ neighbor of } x_i \text{ and } x_i \text{ maximum} \\
&\qquad T'(x_i) > \min_{\text{neighbor } x_j} T'(x_j) \text{ for } x_i \text{ not an extremum} \\
&\qquad T'(x_i) < \max_{\text{neighbor } x_j} T'(x_j) \text{ for } x_i \text{ not an extremum}
\end{aligned}$$

The non linear inequality constraints make this optimization problem hard to solve. The solution suggested by [Jacobson 2012] is to compute a *representative function* u that verifies the last four inequality constraints. Given this function u, the topological denoising problem becomes:

$$\begin{aligned}
&\underset{T'}{\text{argmin}} \sum_{x_i \in \Omega} ||\hat{T}(x_i) - T'(x_i)||^2 + \int_\Omega ||\Delta T'||^2 \\
&\quad \text{s.t. } T'(x_i) = \hat{T}(x_i) \text{ for } x_i \text{ a node of the MS complex} \\
&\qquad (T'(x_i) - T'(x_j))(u(x_i) - u(x_j)) > 0, \text{ for } (x_i, x_j) \text{ a pair of neighbors}
\end{aligned}$$

where the last constraint means that the direction of T' shall be aligned with the direction of u. This alternative optimization problem is easily solved [Jacobson 2012].

The representative function u can be computed by solving the Dirichlet problem:

$$\begin{aligned}
&\underset{u}{\text{argmin}} \int_\Omega ||\nabla u||^2 \\
&\quad \text{s.t. } u(x_i) = 0 \text{ for } x_i \text{ a minimum} \\
&\qquad u(x_i) = 1 \text{ for } x_i \text{ a maximum}
\end{aligned}$$

Minimizers of the Dirichlet energy are harmonic functions, and their properties guarantee that x_i and x_j are minima and maxima and that u contains no other extrema inside the MS regions. We refer to [Jacobson 2012] for further details.

Figure 6.11 shows examples of topological denoising. The topology to be enforced is represented by the red and blue dots, which are nodes of the MS complex: red for intensity maxima and blue for intensity minima. On the left example, the circle motifs that were inducing undesirable minima and maxima are removed. On the right example, two of the initial four maxima in the center of the image are removed too. Only the topology dictated by the input Morse-Smale complex is preserved.

6.4.2.2 Integrating the topological denoising in the template computation pipeline

The original template's computation is performed with the algorithm of [Joshi 2004] and use the LCC Log-daemons for the registrations [Lorenzi 2013]. We adapt it by

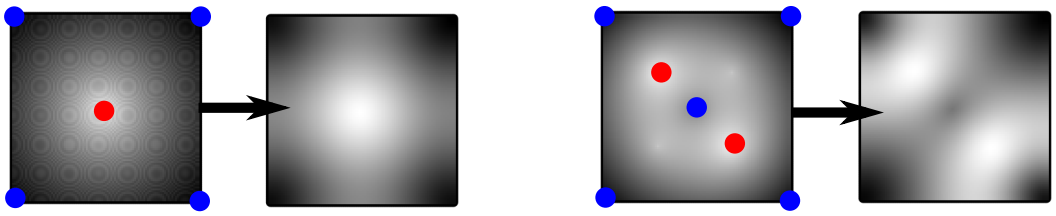


Figure 6.11: Topological denoising on two toy examples. We impose topological constraints on the initial images, on the left in both cases: minima in blue and maxima in red. The arrows denote the action of the topological denoising and point to the output image.

adding a Topological denoising step, in order to control the template's asymptotic bias.

Algorithm 3 shows the adapted procedure. One initiates with the template being one of the subject images: $\hat{T}_1 = I_1$. At each iteration k of the template's computation, one registers the subject images to the current template $\hat{T}_k$ and performs the average of the registered images' intensities to get a first version of the updated template $\hat{T}_{k+1}$. So far, this matches the usual template estimation procedure. Our adaptation is what follows. The MS complex of the updated template $\hat{T}_{k+1}$ is computed, using the R package msr [Gerber 2012]. Then, the updated template $\hat{T}_{k+1}$ is smoothed using Topological denoising, see Figure 6.13. These steps are iterated until convergence.

Algorithm 3 Controlled brain template estimation

Input: Images $\{I_i\}_{i=1}^n$, noise variance σ^2, persistence threshold $p_{\text{threshold}}$
Initialization:
$\hat{T}_1 = I_1$ (one of the subjects images)
$k = 1$
Repeat:
Non-linearly register the images to $\hat{T}_k$, i.e. compute ϕ_k^i: $J_k^i \simeq I^i \circ \phi_k^i$
Compute the mean deformation: $\bar{\phi}_k$
Register subject image: $L_k^i = I^i \circ \phi_k^i \circ \bar{\phi}_k^{-1}$
Compute the mean intensity image for template iteration: $\hat{T}_{k+1} = \frac{1}{n}\sum_{i=1}^n L_k^i$
Compute the MS complex of $\hat{T}_{k+1}$ at persistence level p
Topological denoising of T_{k+1} using the MS complex
$k \leftarrow k + 1$
until convergence: $||\hat{T}_k - \hat{T}_{k+1}|| < \varepsilon$
Output: $\hat{T}_k$

The main parameter controlling this adapted procedure is the asymptotic bias threshold, i.e. the persistence threshold $p_{\text{threshold}}$ for the MS complex computation. The next section discusses the choice of this parameter $p_{\text{threshold}}$. Varying the

threshold $p_{\text{threshold}}$ leads to the construction of a hierarchy of templates. The other parameter is σ, which is the noise on the subject images. Either one knows it from the experimental design, or one estimates it with the variability of the registered subject images, as we did in Section 6.3.

6.5 Experimental results

This section presents experimental results on the quantification of the template's asymptotic bias and the adapted algorithm that bounds this bias. We use the Open Access Series of Imaging Studies (OASIS) database consisting of 136 T1 weighted MR images of brains [Marcus 2007].

6.5.1 Quantification of the template inconsistency

We quantify the asymptotic bias locally on the brain template computed from the OASIS database with the usual procedure. This shows how faithfully the computed template represents human brain anatomy for the neuroimaging studies.

First, we produce maps showing the local asymptotic bias directly with a color code superimposed on the original tridimensional template image, see Figure 6.12. We call these maps the *asymptotic bias maps*. A green color indicates a low asymptotic bias for the region and a red color indicates a high asymptotic bias on the region.

The scale for the color code corresponds to a logarithmic scale, and more precisely to SNR_{dB}, where:

$$\text{SNR}_{\text{dB}} = 10 \log_{10} \left[\left(\frac{d}{\sigma} \right)^2 \right] \tag{6.16}$$

The scale is thus in dB, as the decibel is the logarithmic unit that expresses the ratio of two values of a physical quantity, which is the squared intensity in our case. This unit emphasizes that the quantification of the asymptotic bias depends on a signal-noise ratio (SNR). Indeed, one can consider that the signal is d, which is the template's intensities representing the brain anatomies and the "noise" is σ, the intersubject variability after registration. The larger is the SNR, the lower is the asymptotic bias on the brain template.

We compute several maps, see (c)-(d)-(e) on Figure 6.12 for the same brain template. The difference between the maps is the Morse-Smale complex's persistence threshold used to compute the asymptotic bias. The threshold is increased from left to right on Figure 6.12 (c)-(d)-(e). Increasing the threshold makes more and more regions appear and these are more and more biased: they become colored in orange-red.

The asymptotic bias maps have the following interpretation with respect to neuroimaging. The maps show regions, in orange-red, where the template's brain structures are small with respect to the subjects' variability in the database. In these orange-red regions, it is not reasonable to have a sharply defined template,

because the structures may have appeared by chance, by registration of noise between the different subjects. In other words, the maps reveal brain regions where the assumption of a unique anatomy in the subject population may break down.

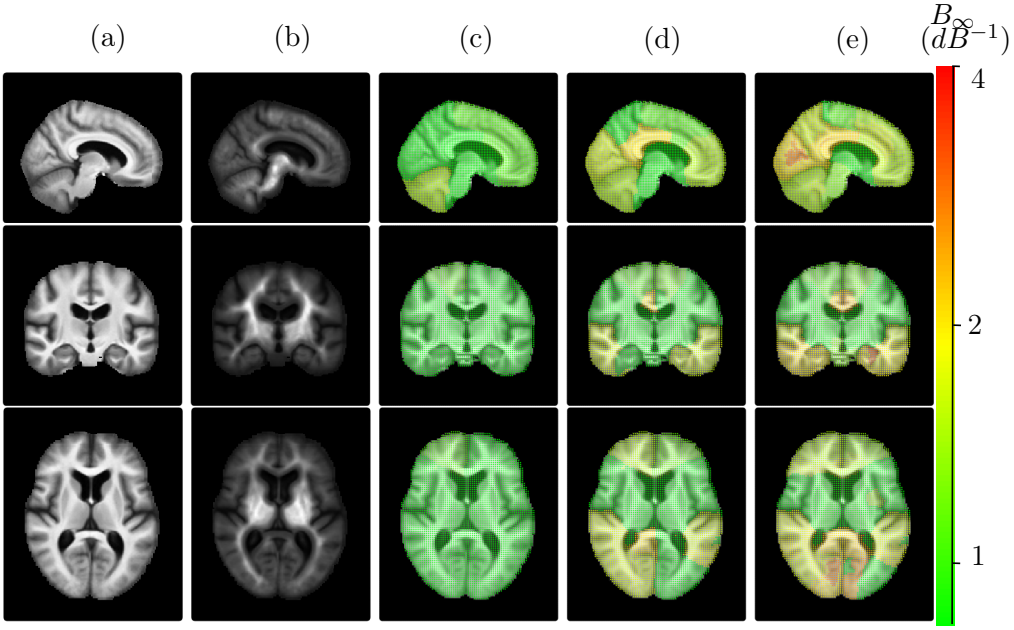


Figure 6.12: Investigation of the template's consistency as an estimator of a unique anatomy. (a) Template. (b) Template whitened by the intersubject variability. (c) Region-wise inconsistency for a threshold= 1.3, (d) for threshold = 2, (e) for threshold = 4 (dimensionless).

6.5.2 Topological denoising for a consistent template

Choice of the persistence threshold Each map of Figure 6.12 (c)-(d)-(e) represents the asymptotic bias of the brain template we would obtain if we were constraining the image to the topology of the corresponding Morse-Smale. The persistence threshold gives a way to investigate the trade-off between asymptotic unbiasedness and sharpness of the template. On the one hand, a complex topology - i.e. a low persistence threshold - implies an important asymptotic bias on the template, which may not represent faithfully the brain anatomy shared by the subjects in the OASIS database. On the other hand, a topology that is too simple - i.e. a high persistence threshold - has no chance of representing a brain anatomy at all. If we want to look at small brain structures, we have to allow for some precision in the topology.

Therefore, which topology shall we choose in this trade-off of asymptotic unbiasedness versus sharpness? If the local intensity of the computed template is below the noise, there is no hope to compute a consistent template. As in the 1D example of [Allassonnière 2007], if the noise is of the same order of magnitude as the signal,

the template may estimate the noise instead of the signal. Thus it makes sense to choose an inconsistency threshold between -1 and 0 dB, that expresses the limit situation where signal (intensity on the brain image) and noise are of the same order of magnitude.

Applying topological denoising to control the brain template's bias We apply the methodology of Section 6.4.2 to enforce the asymptotic bias to be below a threshold, using Topological denoising. Enforcing the unbiasedness in the procedure enables us to build the template of Figure 6.13. As a proof of concept, we have run it on the subject coronal slices of the OASIS database. Following the development above, we bound the asymptotic bias by setting the SNR threshold to -0.8 dB. We observe that the brain regions that were the more biased - i.e. in orange-red in Figure 6.12 are now blurred. Thus, Topological denoising decides where the sharply defined brain template makes sense as a representative of the shared brain anatomy, and blurs it where it does not.

One could be interested in a template, that would be sharp *and* unbiased. In this case, one could consider dropping the assumption of a unique anatomy and consider multiple templates, i.e. use a mixture model. Further work is needed to investigate the construction of a stratified template, which would add a new stratification every time a region's asymptotic bias crosses the threshold $B_\infty \sim 1dB$.

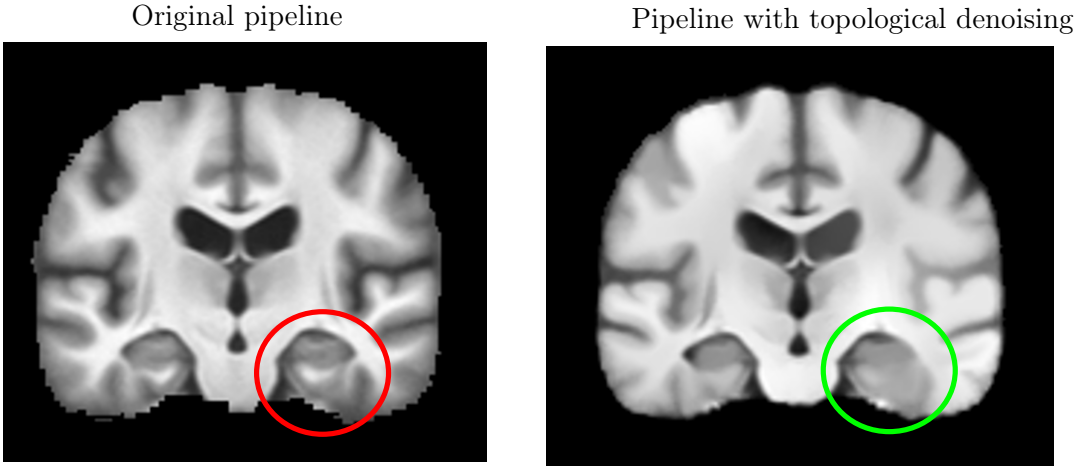


Figure 6.13: Results of Topological Denoising integration in the pipeline for the template's estimation. Left: Template from pipeline without topological denoising. Right: Template with the topologically constrained pipeline. Inconsistent regions from Figure 6.12 are now blurred.

6.6 Conclusion and perspectives

We have investigated the algorithm of brain template computation under the assumption of a unique anatomy. Such computations have been used in the medical imaging literature for at the last fifteen years. We have presented a methodology relying on images topology that quantifies the asymptotic bias of the brain template, as an estimate of the anatomy shared by the subjects of the database under study. This is, to our knowledge, the first attempt to assess the uncertainty on this statistical procedure.

Our methodology builds a bridge between the diffeomorphic registration framework of Medical Imaging and Morse-Smale theory in Topology. This link is an interesting application of topology in itself and has been developed in Section 6.3. Furthermore, we have shown how the MS framework identifies biased regions in the brain template in Section 6.4. In these regions, a sharp template might not be desirable. We can control the template's asymptotic bias by adding a Topological Denoising step in the template algorithm, creating a trade-off between sharpness and unbiasedness. Our methodology is illustrated on a real database of 136 brain images in Section 6.5. In particular, we have shown how the topological denoising blurs the regions that were the most biased. Our method allows us to control the template asymptotic bias, at the price of dropping its sharpness.

It would be very interesting to be able to keep both the unbiasedness and the sharpness of the brain template. In fact, the template being biased can be seen as an indication that the assumption of a unique anatomy within the population may be relaxed. One could think about estimating a mixture of several templates or stratified templates. Each of the templates would represent only a subgroup of the brain population. This subgroup would have a lower variability. Therefore, the parameter σ will be decreasing and the biasedness too. This will allow for templates that are sharper and still unbiased.

Chapter 7

Geometric Statistics on manifolds with Lie group actions: geometric correction of the bias induced by registration of anatomical shapes

---

Chapter 5 uses tools of Geometric Statistics on quotient spaces to analyze the algorithm of template shape computation. Chapter 6 shows the impact of the results on biased template in neuroimaging. In this Chapter, we generalize the use of Geometric Statistics to the analysis of other algorithms on shape data. More precisely, we focus on statistical analyses of shapes that use registration as a pre-processing step. We show that a bias also occurs in these analyses. We offer a theoretical geometric method to correct it, by enforcing a modification of the quotient space geometry.

In order for this Chapter to be self-contained, Section 7.2 is very similar to the first section of Chapter 5. Thus Section 7.2 can be skipped in the reading of this thesis..

7.1 Introduction

Consider a dataset of n observations X_i, $i = 1, .., n$ in a finite dimensional manifold M. For example, each X_i can represent a 3D configuration of an object, described by k landmarks on its surface: $X_i \in (\mathbb{R}^3)^k$. Unsupervised learning seeks to discover the underlying structure of the X_i's, and to answer questions as: what is the probability distribution function of the 3D configurations of these objects? Both clustering or submanifold learning in M are techniques unveiling properties of such a probabilistic structure.

This paper considers a related problem. What if we are not interested in the 3D configurations of the objects, but rather in the shapes of these configurations? In this case, the goal is to discover the probabilistic structure of some *equivalence classes* of the data: $[X_i]$'s. The shape of a set of landmarks, of a curve or of a surface can be defined as follows: the remainder after we have filtered out the position, the orientation [Kendall 1984] and the parameterization. Objects - landmarks, curves, surfaces, etc.- are the observations X_i's in a given manifold M. Their shapes are their equivalence classes $[X_i]$'s under a group of transformations, like rotations, translations, scalings, etc., or a group of reparameterizations. If one wants to do

statistics on shapes, one does statistics on equivalence classes. We consider the specific case where the equivalence classes come from the isometric action of a Lie group on the manifold M.

Applications show various situations where one needs to perform statistics on equivalence classes $[X_i]$'s (the shapes) while observing only X_i's (the objects). Paleontologists consider human skulls described by the 3D coordinates of landmarks detected on the skull's surface. They seek to correlate the skull *shapes* with biogeographic data [Zalawadia 2010]. Biologists observe proteins, also described by the 3D coordinates of their atomes. They study how their *shapes* are related to their function. Statistics on misfolded proteins are used to understand diseases, like Parkinson's disease [Li 2008]. Orthopaedic surgeons analyze *bones' shapes* for surgical pre-planning [Darmanté 2014].

The intuitive procedure to perform statistics on the equivalence classes (the shapes) goes as follows. First, one takes the equivalence class of each observation: $X_i \to [X_i]$, $i = 1, .., n$. One goes from the object's parameterization to its shape's parameterization $X_i \to [X_i]$ by "eliminating" the position, orientation or parameterization of the object. In other words, one projects the observations $X_i \in M$ in a quotient shape space Q: $[X_i] \in Q$. We call π this projection from M to Q: $X_i \xrightarrow{\pi} [X_i]$. In the literature, M is sometimes called the top space. Second, one conducts the statistical analysis on the $[X_i]$'s, for example subspace learning or clustering.

This two step procedure seems straightforward and is commonly used in the literature [Kendall 1989, Bookstein 1986]. But one may wonder if its first step, i.e. the projection π from M to the quotient shape space Q: $X_i \xrightarrow{\pi} [X_i], i = 1..n$, affects the consistency of the statistical analysis. This is a legitimate question as any inconsistency would make the conclusions of the study less accurate.

Our paper analyzes the procedure's consistency, under a generative model of the X_i's where the data are subject to a centered noise in the space of objects M. The generative model includes: (i) the *probability distribution p_Q of some shapes Y_i's in* Q and (ii) the mechanism generating the objects X_i with probability distribution p_M in M. An important characteristic of (ii) is the noise model on the objects, which is taken to be in the space M and not in the quotient shape space Q. This means that the measurement error is made on the recording of the objects' configurations in 3D rather than on their shapes, in our previous example.

The goal is to learn the distribution of the shapes p_Q. The two step procedure ultimately performs learning on the $[X_i]$'s, so it tries to unveil $\pi(p_M)$. We demonstrate that the procedure is biased as a learning of p_Q, in the sense that: $\pi(p_M) \neq p_Q$.

We give a geometric interpretation. $\pi(p_M) \neq p_Q$ comes from the fact that the equivalence classes $[X_i]$'s are submanifolds in M, that have an external curvature. Thus the projection $\pi : M \to Q$, which represents the registration pre-processing step, is a non-orthogonal projection. All in all: *the geometry of the equivalence classes - called orbits in the case of the Lie group action - creates a bias in the estimation of the distribution.* We leverage this geometric understanding to suggest a method that corrects for this bias. The correction performs a conformal transfor-

mation of the Riemannian metric of M. As a consequence, the equivalence classes $[X_i]$'s loose their curvature in M and the projection $\pi : M \to Q$ becomes orthogonal.

Related work on the shape space's geometry as a quotient space

Statistics on shape spaces have been studied since the 1980's. Kendall first investigated the geometry of shape spaces of k labeled landmarks in $\mathbb{R}^m$ [Kendall 1989]. A shape is represented as the equivalence class of the landmarks in $(\mathbb{R}^m)^k$ under the action of translations, rotations and scalings. [Huckemann 2010] study statistics on general shapes, that are equivalence classes of objects in a manifold M under the action of a Lie group G. This unifies the theory for shapes of landmarks, of curves and of surfaces.

Note that there is a distinction between "form" and "shape" is the literature. "Form" relates to the quotient of the object by rotations and translations only. "Shape" denotes the quotient of the object by rotations, translations, and scalings. Kendall shape spaces refer to "shape": the scalings are quotiented by constraining the size of the landmarks' set to be 1.

The models of [Kendall 1989] include the variability of the *shapes*, as opposed to the noise on the *objects*, which is a pervasive effect of imperfect measuring instruments. They do not consider that the data are observed in the space of landmarks $(\mathbb{R}^m)^k$ i.e. in the ambient manifold M and projected in the shape space Σ_m^k i.e. the quotient space Q. The question of the bias of the procedure is not raised.

Procrustean analysis is another method for performing statistics on landmarks shape spaces [Dryden 1998]. Kendall analyses perform an explicit projection of the data in the shape space: they explicitly compute their coordinates in Σ_m^k. In contrast, Procrustean analyses have an implicit approach. They keep the coordinates in $(\mathbb{R}^m)^k$ and stay in the ambient space: the projection of the data in the shape space is algorithmic and performed by "aligning" or "registering" them. For example, orthogonal Procrustes analysis "aligns" the sets of landmarks by rotating each set to minimize the Euclidean distance to the other sets. Procrustean analysis considers the fact that the data are observed in the space $(\mathbb{R}^m)^k$ but does not consider the geometry of the shape space.

The issue of estimating the mean shape, which is the simplest form of learning on the shape space, has been studied in the literature. Computing the mean shape is one method for summarizing shape data. The mean shape is an average of *equivalence classes of the data*: one computes the mean after projection of the data in the shape space. Recent studies show that the statistical properties of the mean shape estimation are modified when one takes into account this noise [Miolane 2015a]. Consistency of the mean shape estimation has been discussed with and without the noise assumption. The mean shape was shown consistent for shapes of landmarks in 2D and 3D in [Lele 1993, Le 1998] and for shapes of signals in [Kurtek 2011], both under the simplifying assumption of no noise on the objects. [Allassonnière 2007] provided examples of inconsistency of the mean shape for noise on the objects, in examples with discretized signals and 2D images [Allassonnière 2015b]. [Bigot 2011]

computed lower and upper bounds on the inconsistency of the mean shape for discretized curves in 2D. But this is restricted to the mean shape estimation and no geometric intuition nor correction strategy was given.

Recently, the inconsistency of the mean shape has been studied in the abstract setting of Lie group actions [Miolane 2016]. The authors provide a geometric interpretation of the bias of this procedure. Their result applies to the literature of shape statistics and to signals, curves and surfaces. For example, formulated in the Procrustean terminology it gives: the Generalized Procrustes Analysis (GPA) estimator of mean "form" is asymptotically biased.

[Miolane 2016] identify the geometric variable that controls it and propose a correction. However, it corrects for the bias in the mean shape estimate and in the whole probability distribution function of shapes p_Q.

Contributions and outline

We make three contributions. First, we show that the projection $\pi : M \to Q$ defined by $X_i \xrightarrow{\pi} [X_i], i = 1..n,$, and representing the registration pre-processing step, biases the statistical analysis of shapes. This is done by computing $\pi(p_M)$ and showing that $\pi(p_M) \neq p_Q$. We offer an interpretation of the bias through the geometry of the shape space. Second, we leverage our understanding to suggest a geometric correction. These results are shown in the following mathematical framework: we assume a proper, effective and isometric action of a finite dimensional Lie group G - defining the registration process - on a finite dimensional Riemannian manifold M.

The paper has three Sections. Section 1 introduces the geometric framework of quotient spaces and details the geometry behind the projection $\pi : X_i \xrightarrow{\pi} [X_i], i = 1..n,$. Section 2 shows that a statistical analysis of shapes starting with a registration step is biased and gives a geometric interpretation: the bias comes from the fact that the projection π is non-orthogonal. Section 3 describes the procedure to correct the bias: it amounts at changing the metric of the space M in order to force the projection π to be orthogonal.

7.2 Geometrization of unsupervised learning algorithms on shape data

7.2.1 Two running examples

We introduce two simple examples of shape spaces as in [Miolane 2016]. We will refer to them constantly to provide intuition.

First, we consider two landmarks in the plane $\mathbb{R}^2$ as in Figure 7.1 (a). The landmarks are parameterized each with 2 coordinates. For simplicity we consider that one landmark is fixed at the origin on $\mathbb{R}^2$. Thus the landmarks are now parameterized by the 2 coordinates of the second landmark only, e.g. in polar coordinates (r, θ). We are interested in the shape of the 2 landmarks, i.e. in their distance which is simply r.

Second, we consider two landmarks on the sphere S^2 as in Figure 7.1 (b). One of the landmark is fixed at the origin of S^2. The landmarks are now parameterized by the 2 coordinates of the second landmark only, i.e. $(\theta.\phi)$. The shape of the two landmarks is the angle between them and is simply θ.

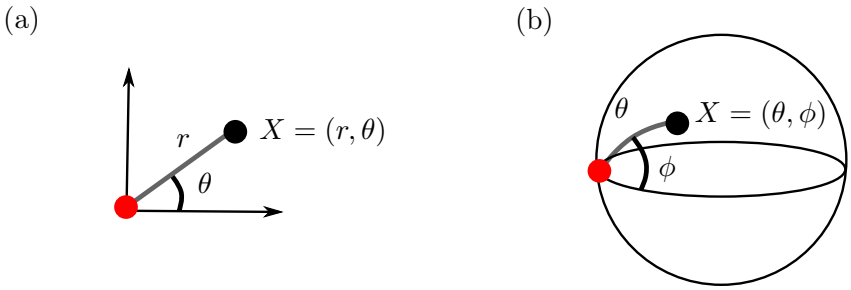


Figure 7.1: Two landmarks, one in red and one in black, on the plane $\mathbb{R}^2$ (a) and on the sphere S^2 (b). The landmark in red is fixed at the origin of the coordinates. The system of landmarks is entirely represented by the coordinates X of the landmark in black.

7.2.2 Elements of Differential Geometry of shapes

The framework of Differential Geometry of shapes that serves this study has been introduced in [Miolane 2016]. We only summarize the most important aspects here.

The data are objects X_i, $i = 1...n$, that are either sets of landmarks, curves, images, etc. We consider that each object X_i is a point in a finite dimensional Riemannian manifold M, which may be called the top space in other studies [Allassonnière 2015a]. We have $M = \mathbb{R}^2$ in the plane example: a flat manifold of dimension 2. We have $M = S^2$ in the sphere example: a manifold of constant positive curvature and of dimension 2.

Shapes are orbits (equivalence classes) of a Lie group action By definition, the objects' shapes are their equivalence classes $[X_i]$, $i = 1...n$, under the isometric action of some finite dimensional Lie group G: G is a group of continuous transformations that models what does not change the shape. A *Lie group action of* $\mathcal{G}$ on $\mathcal{M}$ is a differentiable map $\rho : \mathcal{G} \times \mathcal{M} \ni (g, X) \mapsto g \cdot X \in \mathcal{M}$, such that $g_1 \cdot (g_2 \cdot X) = (g_1 \cdot g_2) \cdot X$ and $Id \cdot X = X$, where Id is the identity of the group. The action of G on M will be written with "$\cdot$".

The equivalence classes $[X_i]$'s have a precise definition in the framework of Lie group actions. The equivalence class of X_i is the orbit of X_i under the action of G, i.e. the submanifold O_{X_i} of all objects in M that have the same shape as X. In other words: $[X_i] = O_{X_i}$ and both notations and appellations may be found in the literature. The curvature of the orbit as a submanifold of M is the key point of the results in Section 7.4.

In our examples, the rotations are the transformations that leave the shape of the systems invariant. Let us take g a rotation. The action of g on the landmark X is illustrated by a blue arrow in Figures 7.2 (a) for the plane and (d) for the sphere. We observe that the action does not change the shape of the systems: the distance between the two landmarks is preserved in (a), the angle between the two landmarks is preserved in (d). The equivalence class of X_i is also called its orbit and written O_{X_i}. The equivalence class/orbit of X is illustrated with the blue dotted circle in Figure 7.2 (a) for the plane example and in Figure 7.2 (b) for the sphere example.

Shape space is a quotient space The *shape space* is by definition the space of orbits, i.e. the space of equivalence classes. This is a quotient space denoted $Q = M/G$. One orbit in M, i.e. one circle in Figure 7.2 (b) or (e), corresponds to a point in Q. The shape space is $Q = \mathbb{R}_+$ in the plane example. This is the space of all possible distances between the two landmarks, see Figure 7.2 (c). The shape space is $Q = [0, \pi]$ in the sphere example. This is the space of all possible angles between the two landmarks, see Figure 7.2 (f).

The distance in M induces a quasi-distance d_Q in the *shape space* Q:

$$d_Q(O_{X_1}, O_{X_2}) = \inf_{g \in G} d_M(g \cdot X_1, X_2). \tag{7.1}$$

In the plane example, the distance between two shapes is the difference in distances between the landmarks. One can compute it by first aligning the landmarks, say on the first axis of $\mathbb{R}^2$. Then, one uses the distance in $\mathbb{R}^2$.

Isotropy group of an object The *isotropy group of* X_i is the subgroup of transformations of G that leave X_i invariant. For the plane example, every $X_i \neq (0, 0)$ has isotropy group the identity and $(0, 0)$ has isotropy group the whole group of 2D rotations. Objects on the same orbit, i.e. objects that have the same shape, have conjugate isotropy groups. The *orbit type* of an orbit is the corresponding conjugation class of the objects isotropy groups. The manifold M (resp. the quotient space) are stratified into orbit types (resp. projection of the orbit types) [Miolane 2016]: this means that both M and Q can be seen as a collection of submanifolds of various dimensions, called strata, glued together in a certain way.

7.2.3 Geometric statistics

Probability distribution on the shape space and on the object space Random points and random probability distributions on manifolds were defined for instance in [Pennec 2006]. In particular, for manifolds, the generalization of the uniform measure is the Riemannian measure defined on M by $dM(X) = \sqrt{\det(g_{AB}(X))}dX$.

In our case, we will integrate over three manifolds: M, Q and O_X for different orbits. Moreover, for a given orbit O_X, we have the intrinsic measure dO^{int}, which does not consider the embedding of O_X in M, or the induced or extrinsic measure

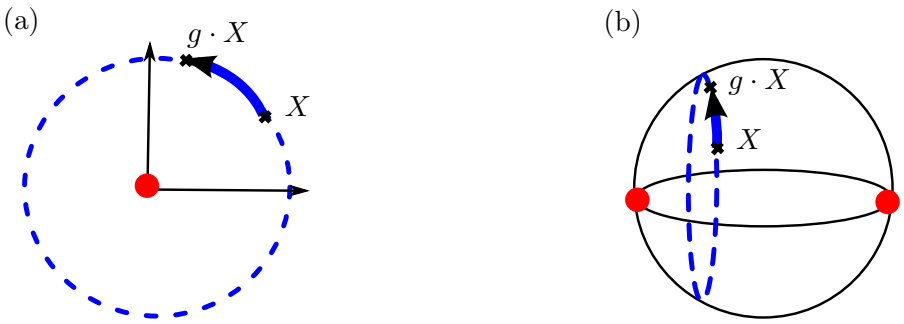


Figure 7.2: (a) Action of rotation $g \in SO(2)$ on point $X \in \mathbb{R}^2$, and orbit of X in blue dotted line; (b) Action of rotation $g \in SO(2)$ on point $X \in S^2$, and orbit of X in blue dotted line.

dO^{ext} which considers the embedding of O_X in M as it comes from the pull-back on O_X of the Riemannian metric on M. On Q, we have the induced measure. Because the action is isometric, the measures on M, Q and the extrinsic measure on O are related as follows:

$$dM(X) = \sqrt{\det(g^M(X))}dX = dQ(x).d\mathcal{O}^{\text{ext}}(X).$$

where we emphasize that the extrinsic measure is used on O here.

The ambient space M is a manifold, the orbit $\mathcal{O}_X$ is a submanifold of $\mathcal{M}$. The quotient $\mathcal{Q}$ is a stratified manifold. To perform statistics on these spaces, we need reference measures. We use the Riemannian measure on M, the induced measure on O, and the induced measure on Q. These measures enable to define probability distribution functions on manifolds, and for example the notion of a Riemannian Gaussian on Q, O and M [Pennec 2006].

We have focused on an intuitive introduction of the concepts. We refer to [Postnikov 2001, Alekseevsky 2003, Huckemann 2010] for mathematical details. From now on, the mathematical setting is the following: we assume a proper, effective and isometric action of a finite dimensional Lie group G on a finite dimensional Riemannian manifold M.

7.3 Geometrization of generative models of shape data and learning procedures

7.3.1 Generative models

We recall that the data are the X_i, $i = 1..n$, which are sets of landmarks, curves, images, etc. We interpret the data X_i's as random realizations of the generative model:

$$X_i = \text{Exp}(g_i \cdot Y_i, \varepsilon_i) \qquad i = 1...n \tag{7.2}$$

where Exp(p,u) denotes the Riemannian exponential of u at point p. The shape $Y_i \in M/G \subset M$ is given position/parameterization $g_i \in G$ and observed with noise

$\varepsilon_i \in T_{g_i \cdot Y_i} M$. The Y_i, g_i, ε_i are realizations of random variables. Their individual laws are presented in the next paragraphs. They define the nature of the generative model and the expression of the probability distribution of the X_i's in M, which we denote p_M. Drawing the shape Y_i, its position g_i and the noise ε_i lead to the following three step interpretation of the generative model of Equation 7.2. These steps are also illustrated with shiny applications available online at `https://nmiolane.shinyapps.io/shinyPlane` and `https://nmiolane.shinyapps.io/shinySphere`.

Step 1: Generate the shape $Y_i \in M/G$ We assume that there is an probability density of shapes p_Q in $Q = M/G$, with respect to the measure on Q induced by the Riemannian measure of M. The Y_i's are assumed to be i.i.d. samples drawn from this distribution p_Q. For example, p_Q can be a Gaussian as illustrated in Figure 7.3 on the shape spaces for the plane and sphere examples. The variability of shapes described by p_Q is the variability that is meaningful for the statistical study, whether we are analyzing shapes of skulls, proteins, bones, neural spike trains, handwritten digits or brains.

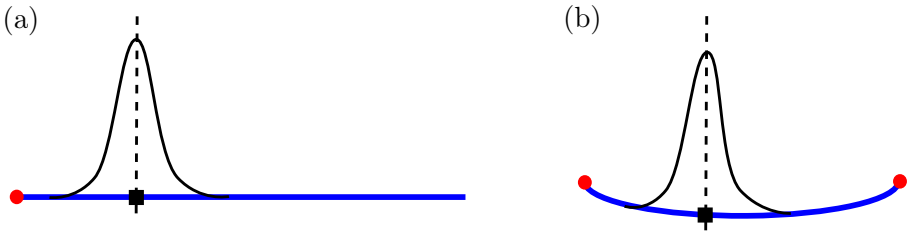


Figure 7.3: Step 1 of the Generative model illustrated for the Rrunning examples. In both cases, the probability distribution function p_Q on shapes is illustrated with the black curve. Its expectation is illustrated by the green square. (a): Representative case $\mathbb{R}^2$, the distribution of shapes p_Q is a distribution on $r \in \mathbb{R}_+$. (b): Representative case S^2, the distribution of shapes p_Q is a distribution on $\theta \in [0, \pi]$.

Step 2: Generate its position/parameterization $g_i \in G$ We cannot observe shapes in $Q = M/G$. We rather observe objects in M, that are shapes posed or parameterized in a certain way. There are two classes of generative models, depending on the assumption we make on the poses: functional and structural models. This terminology comes from the literature of Error-in-Variables and will be used in Section 4.

In the *functional model*, the positions on the orbit are parameters which we want to estimate. In this model, the number of parameters increases with the number of data. In the *structural model*, the positions on the orbits are random variables generated by a probability distribution, which we denote p_O. We assume that there is a probability distribution on the positions/parameterizations of G, or equivalently a probability distribution p_O on principal orbits with respect to their

intrinsic measure. In order to draw a position on the orbit, we draw an element g_i from the probability distribution on the Lie group and make g_i act on the Y_i. We assume that the distribution does not depend on the shape Y_i that has been drawn. The g_i's are assumed to be i.i.d. from this distribution. For example, p_O can be a Gaussian as illustrated in Figure 7.4 on the shape spaces for the plane and sphere examples.

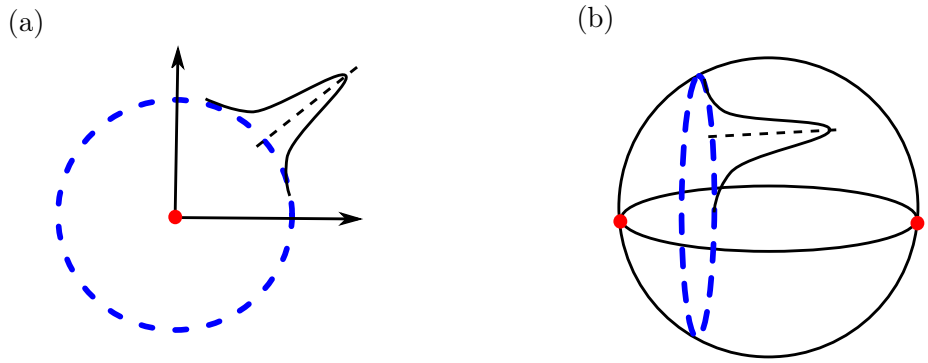


Figure 7.4: Step 2 of the Generative model illustrated for the running examples. In both cases, the probability distribution function on poses is illustrated with the black curve. (a) the distribution on orbits is a distribution in angles: each point is posed on its orbits through a distribution on $\phi_1, ...\phi_{m-1}$. (b) the distribution on orbits is a distribution is angles: each point is posed on its orbits through a distribution on $\phi_1 ...\phi_{m-1}$.

Step 3: Generate the noise $\varepsilon_i \in T_{g_i \cdot Y_i} M$ The observed X_i's are the outcomes of noisy measurements. We assume that there is a probability distribution function on $T_{g_i \cdot Y_i} M$ representing the noise. We further assume that this is a Gaussian centered at $g_i \cdot Y_i$, the origin of the tangent space $T_{g_i \cdot Y_i} M$, and with standard deviation σ, see Figures 7.5. This assumes that the noise model is the same at $g_j \cdot Y_j$ and $g_i \cdot Y_i$, i.e. that the noise is isotropic and stationary. The parameter σ representing its standard deviation will be extremely important in the developments of Section, as we will compute Taylor expansions around $\sigma = 0$, i.e. for low measurement error. All in all, these three steps produce observations X_i with a given probability distribution p_M that depends on the laws of Y, g and ε.

7.3.2 Learning the distribution of shapes

Our goal is to unveil the structure of *shape* data Y_i's, from the observations of noisy objects X_i's. Geometrically, the goal is to learn the underlying structure of the shapes Y_i, represented by the probability distribution p_Q in the quotient space Q, from the noisy observations X_i drawn from p_M on the manifold. We assume the structural model for the distribution of the g_i's. We do not introduce a new procedure to learn p_Q, but we rather analyze the intuitive procedure usually performed

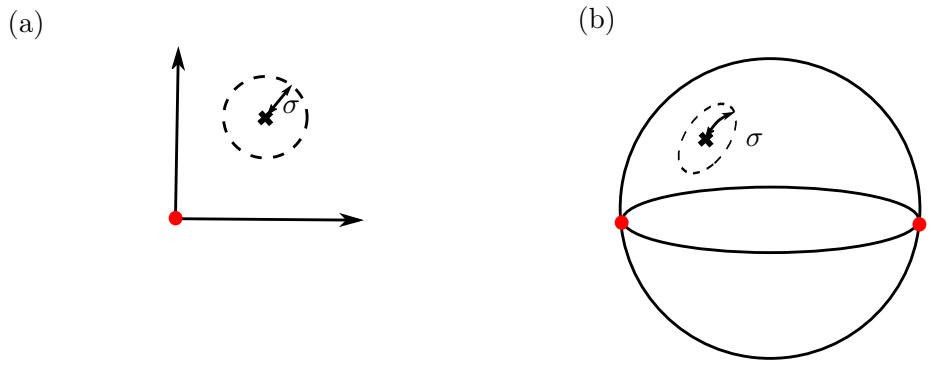


Figure 7.5: Step 3 of the Generative model. In both cases, the probability distribution function on the ambient space is multivariate, one isolevel is illustrated with the dotted curve. (a) Noise in the ambient space $\mathbb{R}^2$. (b) Noise in the ambient space S^2.

in the literature, which usually takes two steps [Kurtek 2011][Kendall 1989]: one registration step (like a pre-processing step) and the learning step.

First step: Registration step In a first step, one registers (i.e. aligns, i.e. puts in optimal position) the objects. By doing so, one filters out the position/parameterization component, i.e. one projects the objects in the shape space Q: $X_i \xrightarrow{\pi} [X_i], i = 1..n,$.

In practice, the projection is performed either explicitly, or implicitly by solving an optimization problem. *The explicit approach* computes the "shape coordinates" of each object. This is easily done for shapes like landmarks shapes, using Kendall or Bookstein coordinates. *The implicit approach registers* the objects. One takes representatives of each object that are aligned in a certain sense: two objects X_1 and X_2 are aligned if the X_1 is in optimal position with respect to X_2, i.e. we have chosen an optimal g to act on X_1 so that it is as close as possible to X_2. These representatives are objects and still belong to the manifold M. However, they belong to a subspace of M that is isomorphic to the shape space Q. As a consequence, on can perform statistics on them as if we were in the shape space. This step is illustrated on Figure 7.6.

Second step: Learning step In a second step, one proceeds with unsupervised learning techniques, for example subspace learning and clustering on the $[X_i]$'s. The standard situation would have been to perform the learning on the Y_i's. Here, one uses the $[X_i]$'s as proxies.

We could consider the mean shape estimation which is the simplest case of subspace learning, as it is a 0-dimensional subspace learning. We could also consider K-means on shapes which is a simple case of clustering algorithm.

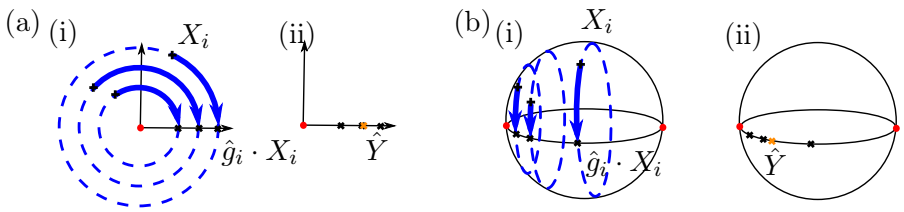


Figure 7.6: The 2 steps of Unsupervised learning on shapes. First, the observations are registered: the data follow their orbits until they are aligned. Then, unsupervised learning as for example the mean shape computation or K-means on shapes, is performed on the projected objects.

7.4 Bias on the usual procedure

This section shows that registration, as a pre-processing step for statistical shape analysis, biases the studies when the data are generated with the model of Equation 7.2. We recall that the registration is geometrically interpreted as the projection π of the data in the quotient shape space: $\pi : X_i \xrightarrow{\pi} [X_i], i = 1..n,$. We thus show that $\pi(p_M) \neq p_Q$: in other words, the statistical analysis is not even performed on the probability distribution function that we want to learn. In this sense, we say that unsupervised learning on shapes are biased. We show this result for a Riemannian Gaussian noise of variance σ^2 in M, truncated at 3σ.

7.4.1 Definitions of curvature

The geometric origin of this bias is the curvature of the orbits. Curvature can be defined in various ways, both intrinsic and extrinsic. We recall them here as they appear in the developments later. First, the extrinsic definitions describe the curvature of an orbit O, as embedded in a manifold M. The *second fundamental form* h of a submanifold O of M is defined on $T_XO \times T_XO$ by $h(v, w) = (\nabla_v w)^\perp \in N_XO$, where $(\nabla_v w)^\perp$ denotes the orthogonal projection of covariant derivative $\nabla_v w$ onto the normal bundle. The second fundamental form h is a $(2, 1)$-tensor which can be considered as a matrix in 3D. Thus, its trace will be a vector. The *mean curvature vector* H *of* O is defined as: $H = Tr(h)$. Both h and H are measures of extrinsic curvature of O in M. For example, an hypersphere of radius R in $\mathbb{R}^m$ has mean curvature vector of magnitude: $||H|| = \frac{m-1}{R}$.

Second, taking the intrinsic point of view on the orbit, one can consider the Riemannian curvature tensor of the orbit R^O_{ijkl} and the associated Ricci curvature R^O_{ij} and scalar curvature R^O [Postnikov 2001]. For example, an hypersphere of radius R in $\mathbb{R}^m$ has intrinsic scalar curvature equal to 1.

7.4.2 Elementary bias induced by the registration step

We first consider that the distribution of shapes that we want to learn, p_Q, is a Dirac: $p_Q = \delta_Y$. This means that the data X_i's are all generated from the same

shape Y, but possibly with different positions and with measurement error. The registration step induces a bias in the estimation of p_Q, i.e. in the estimation of the unique shape Y, because of the extrinsic curvature of the orbit of Y [Miolane 2016]. More precisely, the distribution $\pi(p_M)$ is not centered around the unique shape Y.

Theorem 7.4.1 (Miolane et al. (2016) [Miolane 2016]) *The probability distribution function on the shapes induced by the generative model is:*

$$\pi(p_M)(x) = b(x) = \frac{1}{c_q \sigma^q} \exp\left(-\frac{x^T x}{2\sigma^2}\right) \cdot \left(c_p^0 + c_p \sum_{ij}\left(\delta_{ij} - \frac{1}{6}R^O_{ij} - \sum_a \frac{1}{2} h^a_{ij}(x) x_a\right)\sigma^2 + \mathcal{O}(\sigma^4)\right).$$

Here x is the Riemannian logarithm $x = Log_Y[X]$ of the shape $[X]$ at the template shape Y, h is the second fundamental form of the orbit of x, R is the Ricci curvature of M, c_q, c_p, c_p^0 are constants independent of x and σ.

We call $\pi(p_M)(x) = b(x)$ the *elementary bias*, because it corresponds to the elementary effect of the registration step for the further statistical analysis. It shows that the isotropic Gaussian noise on the objects is integrated into a skewed distribution on the quotient shape space Q, because of the registration step. In other words, this elementary bias depends on the noise model. The exponential in the expression of $\pi(p_M)$ in Theorem 7.4.1 belongs to a Gaussian distribution centered at $x = 0$. This is a Riemannian Gaussian centered at the template shape Y, because x are coordinates at the tangent space of Q at Y. However the whole distribution $\pi(p_M)$ differs from the Gaussian because of the x-dependent term in the right parenthesis. This dependency in x comes from the external curvature h^a_{ij} of the orbit of Y. Therefore, the curvature of the orbits induces a skew of the distribution $\pi(p_M)$, as observed for the examples in Figure 7.7 taken from [Miolane 2016].

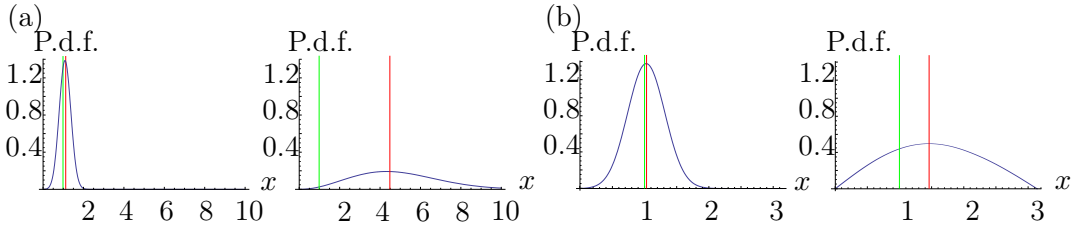


Figure 7.7: (a) Induced distributions on the distance r between two landmarks in $\mathbb{R}^3$ for real distance $y = 1$ (in green) and noise level $\sigma = 0.3$ and $\sigma = 3$. (b) Induced distributions on the angle x between the two landmarks on S^3, for real angle $y = 1$ and noise levels $\sigma = 0.3$ and $\sigma = 3$. In both cases the mean shape estimate $\hat{y}$ is shown in red.

This skew biases the further statistical analysis. In particular, the usual procedure for estimating Y is to compute the mean shape, i.e. the expectation of $\pi(p_M)$.

As $\pi(p_M)$ is skewed, the estimator is different from Y, the unique shape that it is designed to estimate [Miolane 2016]. This is shown on Figure 7.7: the shape Y is shown in green and the estimator is shown in red. We emphasize that the bias is asymptotic, as we are already in the limit of a sample size equal to infinity.

7.4.3 Bias induced by the registration step

The previous subsection shows the bias that is induced by the registration step, when the distribution of shapes p_Q is reduced to a unique shape Y. Here generalize this result to any distribution of shapes p_Q. Intuitively, a general distribution of shapes p_Q can be seen as the continuous sum of Diracs. The elementary bias of the previous subsection thus impacts each of these Diracs. The induced distribution $\pi(p_M)$, on which the statistical learning, is like a convolution of the elementary bias and the original distribution of shapes p_Q.

Theorem 7.4.2 *The projection π from the manifold M to the quotient shape space: $X_i \xrightarrow{\pi} [X_i], i = 1..n$, implies that the statistical analysis is performed on the distribution $\pi(F)$:*

$$\pi(p_M)(x) = \mathbb{E}_{p_Q}\left[b(x)\right] \tag{7.3}$$

where p_Q is the distribution we want to learn and f_Y is the bias function coming from Theorem 7.4.1. We could see this as a generalization of the convolution $\pi(p_M)(x) = p_Q * b(x)$.

The projection $X_i \xrightarrow{\pi} [X_i], i = 1..n$, transforms the probability distribution p_Q we seek to learn into $p_Q * b$. This transformation is skewed, which induces a bias in the statistical moments of $p_Q * b$, with respect to the statistical moments of p_Q that one could be interested in. From the very start, the learning procedure tries to learn a biased structure. In this sense, the registration step through the projection $\pi : M \to Q$ biases the following statistical analysis or learning algorithm.

The systematic bias could impact the conclusions from learning algorithms on shapes. Correcting the bias may be preferable in order to ensure accurate conclusions from a statistical analysis or to unveil the expected structure in the shape data.

7.5 Correcting the usual procedure

We have highlighted the variables responsible for the bias: this is the external curvature of the orbits at the scale of the noise. This is induces an elementary bias function that is skewed: the registration step registers the Gaussian isotropic noise into a skewed noise on the quotient shape space. If we can correct this skew, we can correct the elementary bias and therefore unbias the further statistical analysis on shapes.

The intuition we have gained from the study enables us to suggest a geometric correction of the systematic bias. We perform a transformation of the orbits, through a conformal change of the Riemannian metric of M. The orbits loose their external

curvature. The projection $\pi : M \to Q$, corresponding to the registration step, becomes orthogonal and the skew disappears on the induced distribution $\pi(p_M)$. We exhibit our geometric correction in two flavors in Theorems 7.5.2 and Theorem 7.5.3, equivalently offering to transform the metric of the ambient space M in Theorem 7.5.2 or the distribution p_M of the data X_i's in Theorem 7.5.3.

7.5.1 Correction of the bias by changing the metric M

We consider g_M the Riemannian metric on M and ϕ a smooth real-valued function on M. Then:

$$g'_M = \phi \cdot g_M = e^{2\varphi} g_M = e^{2\varphi} g_M \tag{7.4}$$

is also a Riemannian metric on M. We say that g'_M is conformal to g_M. We show that there exists a conformal change of metric, i.e. a function ϕ, that can make the orbits loose their external curvature. The notation $\cdot$ is purely schematic and does not represent the action of a Lie group.

Lemma 7.5.1 *There exists a smooth G-invariant $\phi : M \mapsto \mathbb{R}$ that makes M a product space almost everywhere. Equivalently, taking X_{ref} on a principal orbit:*

$$\forall X \in M^*, e^{2\phi(X)} g^{\mathcal{O},ext}(X) = g^{\mathcal{O},ext}(X_{ref}) \quad \text{and} \quad \phi(g \cdot X) = \phi(X) \quad \text{(G-invariance)} \tag{7.5}$$

where $g^{\mathcal{O},ext}$ is the extrinsic Riemannian metric on the orbit, i.e. induced by the metric of M.

This can be seen as follows. As the principal orbit type is dense, there are principal orbits almost everywhere. Thus we remove the singularities of the space. These principal orbits are diffeomorphic, from the orbit-stabilizer theorem. The function ϕ takes a principal orbit, sets it as a reference, and then maps all principal orbits to it. Note that ϕ is not smooth at the singularities, thus of the set of measure 0: $M - M^*$.

Now we rely on this conformal transformation of the metric to modify the usual statistical analysis of shapes, by adding two geometric corrections.

Theorem 7.5.2 (Correction of the systematic bias - Metric version) *Two successive transformations of the space correct the systematic bias in the procedure of statistical analysis on shapes:*

$$\begin{aligned} &(1) \quad g^M(X) \to g'^M(X) = e^{2\phi(X)} g^M(X) \quad \textit{(before the registration step)} \\ &(2) \quad g^Q(X) \to g'^Q(X) = e^{-2\phi(X)} g^Q(X) \quad \textit{(after the registration step)} \end{aligned}$$

where ϕ of the conformal change of metric is given by Lemma 7.5.1. When taking the elementary shape distribution p_Q, i.e. the Dirac at Y describing a unique shape in Q, the probability distribution of shapes $\pi(p_M)$ in Q writes:

$$\pi(p_M) = \frac{1}{\mathfrak{M}_0(Y)} \exp\left(-\frac{1}{2} d_Q(Y, X)\right)$$

Proof The proof is given in Appendix B.

The distribution $\pi(p_M)$, which is the corrected version of the elementary bias of the previous section, is not skewed. As a consequence, its expectation gives a unbiased estimator of the unique anatomy Y in the case of a Dirac. For a general distribution of shapes p_Q, it does not induce a global shift of the projected distribution $\pi(p_M)$. In this sense, the geometric transformation has corrected the bias. We also note that this correction only focuses on adapting the registration step and does not depend on the statistical analysis that one wants to perform on the shapes. In particular, it corrects the bias for the mean shape estimate and K-means on shapes.

Figure 7.8: Illustration for the successive transformations of metric in $\mathbb{R}^2/SO(2)$ and $S^2/SO(2)$. The reference orbit chosen is shown in light blue. In both cases, singularities are put at infinity and the orbit have null external curvature in the new ambient space.

This change of metric is illustrated in Figure 7.8 for the two running examples. More precisely, the change of metric amounts to replace the extrinsic metric on the orbit part by the metric of one of the orbit, which plays the role of an intrinsic metric. It puts the external curvatures of all orbits to 0 and send the singularities of Q to the infinity, so Q is now a manifold. Moreover, π' is now a Riemannian submersion from the whole M. The change of metric makes M^* a principal fiber bundle over Q^* with fiber $G/G_{\min}$, where $G_{\min}$ is the minimal isotropy group.

7.5.2 Correction of the bias by transforming the distribution of the data

Now we show another method for correcting for the bias, by transforming the distribution of the data X_i's or, more precisely, by redistributing the X_i's. We consider p_M the distribution of the data X_i's on M, as given by the generative model, and ϕ a smooth real-valued function on M. Then the following is also a distribution on M:

$$p'_M(Z) = \phi \cdot p_M = \frac{e^{m\phi(X)}}{C'_M(\sigma)} \exp\left(-\frac{{d'}^2_M(Z,Y)}{2\sigma^2}\right) \tag{7.6}$$

where m is the dimension of M and the new geodesic distance d'_M is computed as:

$${d'}_M(Z,Y)^2 = \int g'_{AB}(\gamma(t))\frac{d\gamma(t)^A}{dt}\frac{d\gamma(t)^B}{dt}dt = \int e^{2\phi(\gamma(t)))}g_{AB}(\gamma(t))\frac{d\gamma(t)^A}{dt}\frac{d\gamma(t)^B}{dt}dt.$$

This amounts to changing the distances between data in the Riemannian Gaussian, and to weight by a factor which would correspond to the change in measure. Again, the notation $\cdot$ is purely schematic and does not represent the action of a Lie group. Taking the function ϕ of Lemma 7.5.1, we rely on this redistribution of the data to modify the usual statistical analysis of shapes as follows.

Theorem 7.5.3 (Correction of the systematic bias - Probability distribution version) *Two successive transformations of the probability distribution function correct the systematic bias:*

$$\begin{aligned} (1) \quad & p_M(X) \to p'_M(X) = \phi * p_M \quad \textit{(before registration)} \\ (2) \quad & f(X) \to f'(X) = (-\phi) * f \quad \textit{(after registration)} \end{aligned}$$

where ϕ is given by Lemma 7.5.1. As a result, the probability distribution of shapes in Q, with the initial induced metric, writes:

$$f(X) = \frac{1}{\mathfrak{M}_0(Y)} \exp\left(-\frac{1}{2} d_Q(Y, X)\right)$$

Proof Like for Theorem 7.5.2, the proof is given in Appendix B.

This transformation of the probability distribution is also illustrated in Figure 7.8 for the two running examples. It amounts to re-balance the distribution of objects with respect to the orbit they belong to, by applying a transformation depending on ϕ to each data, before and after registration. It corrects for the unbalanced distribution we had in Figure 6.7.

7.5.3 The two corrections are equivalent and boil down to computing ϕ

We consider again the probability distribution function p_M of the data X_i's:

$$p_M(Z) = \frac{1}{C_M(\sigma)} \exp\left(-\frac{d_M^2(Z, Y)}{2\sigma^2}\right) \tag{7.7}$$

The distribution p_M depends on the Riemannian metric g_M of M, through the exponential with the Riemannian geodesic distance d_M. As a consequence, varying the Riemannian metric g_M gives rise to a family of distributions on M.

We consider the family of distributions associated to conformal changes of metric g_M. Formally, taking a conformal change $g'^M(X) = e^{2\phi(X)} g^M(X)$ where ϕ is a smooth function $\phi : M \to \mathbb{R}$, the corresponding probability density transformation writes:

$$p'_M(Z) \equiv \phi * p_M(Z) = \frac{e^{m\phi(X)}}{C'_M(\sigma)} \exp\left(-\frac{d'^2_M(Z, Y)}{2\sigma^2}\right) \tag{7.8}$$

where m is the dimension of M and the new geodesic distance d'_M is as in the previous section.

This shows the equivalence between our two corrections, i.e. between the geometry and the data distribution: the probability distribution p'_M on the manifold with metric g_M is equivalent to the probability distribution p_M on the new ambient manifold with metric g'_M. This is not surprising given the interpretation provided by Theoretical Physics. In General Relativity, the geometry of space-time is equivalent to the distribution of masses in space-time. Likewise in the correction we suggest, we may transform the metric of the ambient space or equivalently, transform the probability distribution function of the data.

All in all, the correction boils down to the functional equations 7.5 in ϕ. If we have expressions of the metric g_M, we can perform the correction analytically. However, in real world examples, we may not have these expressions. This correction needs to be performed in a non-parametric way. In this case, our "correction" is closely related to metric learning methods, known to "enhance" the algorithms. Here, we give it a geometric foundation to seek such a new metric on the quotient space.

7.6 Conclusion

We have investigated the consequence of the projection step in statistical analysis of shapes or unsupervised learning on shapes. We have shown that the curved geometry of the orbits - the equivalence classes - implies that the unsupervised learning tries to learn a biased probability distribution function. We have computed the induced probability distribution on the shape space and show how it differs from the probability distribution one tries to learn. We have leveraged this geometric understanding to propose a correction of the bias. By changing the metric on the shape space, we flatten the orbits and thus eliminate the source of bias.

Chapter 8

Geometric Statistics on fiber bundles: processing of 3D medical images

In the previous Chapters, we have defined and used Geometric Statistics to analyze organ shapes. Now we show how the Geometric Statistics can be applied to medical imaging processing as well.

This chapter has been published in the Proceedings of the workshop "Medical Computer Vision" of the MICCAI conference, under the title: "A Survey of Mathematical Structures for Extending 2D Neurogeometry to 3D Image Processing".

8.1 Introduction

Machine learning algorithms using big data are often "black-boxes". Thus, they can be hard to interpret. There is still a need of constructive models, so that the big data framework can be fed by new structures. The visual cortex offers inspiration for new methods in (medical) computer vision. From the biological model of human vision, one builds a geometric model of the visual cortex. The geometric model is in turn implemented for computer vision purposes. This is the field of *(2D)-Neurogeometry*.

8.1.0.1 Biological Intuition behind Neurogeometry

The geometric model of the visual cortex's is built as follows. From the biological point of view, neurons of the primary visual cortex $V1$ are local detectors called "point processors" [Koenderink 1987]. They are retinotopically connected to small domains of the retina, called their "receptive field" [Jones 1987]. Mathematically, this structure is an isomorphic map from the 2D retina to the 2D cortical layer. It means that each neuron is associated to a position in our retina $(x, y) \in \mathbb{R}^2$, or equivalently in our visual field.

Then, the neuron acts as a filter on the optical signal of the retina's photoreceptors. Its transfer function is called its "receptive profile". The so-called "simple neurons" of $V1$ have a highly anisotropic profile [Jones 1987]. They are sensitive to the orientation $\theta \in S^1$ of the optical signal, in terms of the intensity gradient. A simple neuron is thus represented by the corresponding position $(x, y) \in \mathbb{R}^2$ of the retina and by the preferred orientation $\theta \in S^1$ of its filter [Petitot 2013].

Interestingly, Hubel and Wiesel have shown that neurons detecting all orientations at the same position (x, y) form an anatomical structure, called an "orientation hypercolumn" [Hubel 1977]. This discovery led to the Nobel Prize in 1981. It means that *the fiber bundle* $\mathbb{R}^2 \times S^1$ *is neurally implemented in the brain.*

Ultimately, one models the neuronal activity propagation in $\mathbb{R}^2 \times S^2$. The horizontal cortico-cortical connections of $V1$ are represented by a horizontal distribution in sub-Riemannian geometry [Petitot 2013]. The propagation of the cortical activity is then a propagation along sub-Riemannian geodesics [Citti 2006].

8.1.0.2 Implementations of Neurogeometry for Computer Vision

One finds implementations of 2D-Neurogeometry in computer vision. For example, a sub-Riemannian diffusion process leads to algorithms for image completion or in-painting [Boscain 2014]. Fitting a sub-Riemannian geodesic enables contour completion [Citti 2006]. Furthermore, a sub-Riemannian smoothing can smooth the image while preserving crossings [Duits 2010a]. But the framework lacks general applications in *medical* computer vision, although some exist [Duits 2011]. One reason is that Neurogeometry is essentially 2D, as the retina is 2D. And the generalization of 2D-Neurogeometry to 3D-Neurogeometry is conceptually subtle. There is a need of a theoretical survey summarizing the mathematical structures in the 3D framework. The purpose of this paper is to fill this gap.

8.1.0.3 Contribution and Outline of the Paper

This paper aims to be a guide for understanding and generalizing 2D-Neurogeometry to 3D-Neurogeometry. It is a *theoretical toolbox of 3D-Neurogeometry* for: 1) conceiving new algorithms in medical computer vision; and 2) interpreting existing algorithms. In Section 1, we recall briefly some concepts of Differential Geometry. In Section 2, we describe 2D-Neurogeometry and its applications, as an introduction to the 3D case. In Section 3, we describe 3D-Neurogeometry and its possible applications.

8.2 Requirements of Differential Geometry

The following is summarized in Table 8.1 at the end of the section. We assume that the reader is familiar with the following concepts of Differential Geometry: manifolds, (principal) fiber bundles, (pseudo-) Riemannian manifolds [Postnikov 2001], Lie groups, Lie algebra, bi-invariant (pseudo-)metrics [Bourbaki 1989][Miolane 2015b], Lie group action on a manifold, homogeneous manifolds [Alekseevsky 2003], sub-Riemannian manifolds [Bellaıche 1996]. Some are illustrated in Fig. 8.1.

The aforementioned structures are present simultaneously in the computational framework of Neurogeometry. They arise with their set of related curves, as shown in Table 8.1. Depending on the application for image processing, one is interested

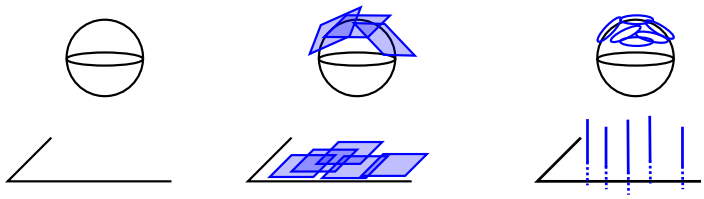


Figure 8.1: Left: S^2 and $\mathbb{R}^2$ are manifolds. Center: tangent bundles of S^2 and $\mathbb{R}^2$. Right-top: principal bundle with base S^2 and structure group S^1. Right-bottom: principal bundle with base $\mathbb{R}^2$ and structure group $\mathbb{R}$. In all cases, the fibers are drawn in blue.

in computing one curve or another. Thus, on shall understand their differences and relations. Some curves are illustrated in Fig. 8.2.

Figure 8.2: From Left to Right. S^2 and $\mathbb{R}^2$ with action of $SO(2)$. The orbits are in blue and coincide with the curves created by the action of 1-parameter subgroups of $SO(2)$ (as $SO(2)$ is 1-dimensional). Riemannian geodesics on S^2 and $\mathbb{R}^2$ for standard induced metric from $\mathbb{R}^3$ on S^2 and the Euclidean metric on $\mathbb{R}^2$.

8.3 The Example of 2D-Neurogeometry

This section serves as an introduction of the 3D case. An image processing pipeline using 2D-Neurogeometry usually follows three steps: 1. Lift (L), 2. Processing (P) and 3. Projection (P) (LPP-framework, see Fig. 8.3). These steps can be iterated [Citti 2006][Sanguinetti 2008] or not [Boscain 2012][Duits 2010a]. Biologically, the lift represents the activation of the neurons in V1. The processing of the lifted image represents the propagation of the neuronal activity in V1. The projection corresponds to our visual interpretation of the information given by the visual cortex after neuronal propagation.

First, we survey the mathematical structures (subsections 2.1, 2.2). We summarize them in Table 8.2. Then we present the LPP-frame of standard algorithms in 2D-Neurogeometry (subsection 2.3). Application-oriented readers can start with subsection 2.3, then go to subsections 2.1, 2.2.

8.3.1 Structures on the Lifted Space $SE(2) = \mathbb{R}^2 \times SO(2) = \mathbb{R}^2 \times S^1$

8.3.1.1 Group Actions

The law of $SE(2)$ is, for all $(t_1, R_1), (t_2, R_2) \in SE(2)$:

$$(t_1, R_1) * (t_2, R_2) = (R_1.t_2 + t_1, R_1.R_2)$$

Table 8.1: Curves related to the different structures of Differential Geometry in Neurogeometry. **For spaces:** M is a manifold, (P, M) a fiber bundle of base M, (M, G) is M endowed with a G action, (P, M, G) is a principal bundle of base M and structure group G, G is a Lie group. **For verticality/horizontality:** V^F, V^O, V and V^Δ: vertical in the sense of fibers, orbits, orbits=fibers (same notion for principal bundle), Δ-distribution. Same notations using H for horizontal. **For metric structures:** g^R is a (pseudo)-Riemannian metric, g^{SR} is a sub-Riemannian metric. **For curves:** γ denotes a notion of geodesics. We have γ^G, γ^G_e, γ^R, γ^{SR} for group geodesic, 1-parameter subgroup, Riemannian geodesics and sub-Riemannian geodesic.

	On M	On (P, M)	On (M, G)	On (P, M, G)	On G
No metric		• V^F-curves • H^F-curves	• V^O-curves ex: γ^G-action • H^O-curves	• V-curves ex: γ^G-action • H-curves	• γ^G ex: γ^G_e
Metric g^R	• γ^R	• V^F-γ^R • H^F-γ^R	• V^O-γ^R • H^O-γ^R	• V-γ^R • H-γ^R	IF g^R bi-inv.: $\gamma^G = \gamma^R$
SR-metric g^{SR}	• V^Δ-curves • H^Δ-curves ex: γ^{SR}	IF $\Delta \perp$ fibers: $V^\Delta = V^F$ $H^\Delta = H^F$	IF $\Delta \perp$ orbits: $V^\Delta = V^O$ $H^\Delta = H^O$	IF $\Delta \perp$ fibers: $V^\Delta = V$ $H^\Delta = H$	IF G Carnot: $\exists \quad g^{SR}$

In this law, we read the group actions on $SE(2)$ and their general properties. $SE(2)$ acts on itself through the left and right translations (freely and transitively). As a Lie subgroup, $SO(2)$ acts on $SE(2)$ on the left and right (freely but not transitively). Note that the right $SO(2)$-action is trivial on the $\mathbb{R}^2$ part. Moreover, the right $SO(2)$-action makes $SE(2)$ a principal bundle of base $\mathbb{R}^2$ and structure group $SO(2)$.

8.3.1.2 A Sub-Riemannian Metric and two Riemannian Metrics

To introduce the sub-Riemannian metric, one first defines its horizontal distribution Δ. In 2D-Neurogeometry one takes the moving frame (X_1, X_2, X_3) on $\mathbb{R}^2 \times S^1$:

$$\begin{cases} X_1 = \cos\theta.\partial_x + \sin\theta.\partial_y, \\ X_2 = \partial_\theta, \\ X_3 = -\sin\theta.\partial_x + \cos\theta.\partial_y \end{cases}$$

to define $\Delta = (X_1, X_2)$. The sub-Riemannian metric g^{SR} is defined as the Euclidean metric on Δ. In the standard basis ∂_x, ∂_y, ∂_z, its inverse writes:

$$g^{SR}(x, y, \theta)^{ij} = \begin{pmatrix} \cos^2\theta & \sin\theta\cos\theta & 0 \\ \sin\theta\cos\theta & \sin^2\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

In practice, the sub-Riemannian metric is usually approximated the Riemannian metric g_ε^R whose inverse is [Citti 2006]:

$$g_\varepsilon^R(x,y,\theta)^{ij} = \begin{pmatrix} \cos^2\theta + \varepsilon^2 \sin^2\theta & (1-\varepsilon^2)\sin\theta\cos\theta & 0 \\ (1-\varepsilon^2)\sin\theta\cos\theta & \sin^2\theta + \varepsilon^2\cos^2\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

In addition, one defines a left-invariant metric g_μ^R as:

$$g_\mu^R(0,0,0)_{ij} = \begin{pmatrix} \mu & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

on the Lie algebra $\mathfrak{se}(2)$. Then, one propagates it on $SE(2)$ through left translations. g_μ^R is $(SE(2))$-left-invariant by construction. But g_μ^R is not $(SE(2))$-right-invariant as there is no bi-invariant metric on $SE(2)$ [Miolane 2015b]. g_μ^R is invariant by the $SO(2)$-left and $SO(2)$-right actions.

8.3.1.3 A Survey of Curves

From Table 8.1 and the aforementioned structures, we survey the curves on $SE(2)$. We have group geodesics of $SE(2)$, Riemannian geodesics of g_μ^R, sub-Riemannian geodesics of g^{SR} and their Riemannian approximation through g_ε^R. Group and Riemannian geodesics differ as g_μ^R is not bi-invariant.

w.r.t. the right $SO(2)$-action, some are vertical or horizontal (taken With respect to. g_μ^R). Examples of vertical group geodesics, vertical Riemannian geodesics for g_μ^R and vertical sub-Riemannian geodesics for g^{SR} are orbits of the $SO(2)$-action. Examples of horizontal group geodesics, horizontal Riemannian geodesics for g_μ^R are straight lines between two translations. Examples of horizontal sub-Riemannian geodesics for g^{SR} are integral curves of X_1.

With respect to. Δ, some are Δ-vertical or Δ-horizontal. There is no Δ-vertical group geodesics, and no Δ-vertical Riemannian geodesic. Example of Δ-horizontal group geodesics and Δ-horizontal Riemannian geodesic for g_μ^R are orbits of the right $SO(2)$-action. Sub-Riemannian geodesics are always Δ-horizontal.

8.3.2 Structures on the Image Domain $\mathbb{R}^2$

Projecting $\mathbb{R}^2 \times S^1$ along the fibers S^1 gives $\mathbb{R}^2$. Equivalently, we can quotient $SE(2)$ by the $SO(2)$-right. The residual left $SE(2)$-action on $\mathbb{R}^2$ is: $(t,R) \circ x = R.x + t$. We read the related left $SO(2)$-action. Regarding the metric structures, g_μ^R was right $SO(2)$-invariant. Thus the projection is a Riemannian submersion for g_μ^R. It induces a Riemannian metric on $\mathbb{R}^2$ which is the Euclidean metric. Projecting the horizontal Riemannian geodesics gives linear curves in $\mathbb{R}^2$. The projection of the sub-Riemannian geodesics gives the elastica curves [Citti 2006], which can be linear or curvilinear.

Table 8.2: Structures of 2D-Neurogeometry. Use Table 8.1 to get the related curves.

	Actions	Metrics
$SE(2) = \mathbb{R}^2 \times S^1$	• left, right translations of $SE(2)$ • left, right actions of $SO(2)$	• g^R_μ • g^{SR} • g^{SR}_ε
$\mathbb{R}^2$	• left action of $SE(2)$ • left action of $SO(2)$	• Euclidean metric (projection of g^R_μ)

8.3.3 The Three Steps: LPP Framework for 2D

8.3.3.1 First Step: Lift (L)

The image domain $D \subset \mathbb{R}^2$ is lifted to $\tilde{D} \in \mathbb{R}^2 \times S^1$ [Citti 2006] (positions and orientations taken with directions). The lift is implemented by detecting the direction of the intensity gradient:

$$\frac{\nabla I}{||\nabla I||} = (-\sin\theta, \cos\theta)$$

at each point $(x, y) \in D$. Then D is mapped to a surface $\tilde{D}$: $(x, y) \mapsto (x, y, \theta(x, y))$ in $\mathbb{R}^2 \times S^1$. At the end of this step, the intensity is a function of $\tilde{D}$.

Alternatively, one can lift to the projective tangent bundle $PT\mathbb{R}^2 = SE(2)/\mathbb{Z}_2$ (positions and orientations taken without directions) [Boscain 2012]. Whether one should use $SE(2)$ of $PT\mathbb{R}^2$ is discussed here [Boscain 2012] (rem. 4, 13).

8.3.3.2 Second Step: Processing (P)

First, the processing can be the evolution of partial differential equations (PDEs) with sub-Riemannian operators. For example, the sub-Riemannian diffusion is defined with the sub-Riemannian Laplacian $\Delta_{SR} = X_1^2 + X_2^2$. Depending on the goal of the processing, one adds drift (also called convection) to the PDE: there is drift for completion purposes [Boscain 2012] and for enhancement [Duits 2011]. Equivalently, one can formulate this step as an oriented random walk. One writes the corresponding Kolmogorov equations.

Some PDEs are computed with the lifted intensity $I(x, y, \theta)$. In-painting methods provide examples: one "paints" directly in the lifted space [Boscain 2012]. Others compute with the activity function: $u(x, y, \theta) = u(x, y, \theta)\delta_\Sigma$ where $u(x, y, \theta) = |X_3(\theta).\nabla I(x, y)|$. In-painting methods provide also examples of this approach [Sanguinetti 2008]. The corrupted image has a hole in $\tilde{D}$. The activity propagation amounts to "fill the hole" by a minimal surface. Then, one "paints" the surface by linking the isolevel sets with sub-Riemannian geodesics.

Then, the processing can be curve fitting. Which curve do we fit? One can fit a sub-Riemannian geodesic, as in the second example of in-painting above

[Sanguinetti 2008]. Another example is contour completion [Citti 2006]. One can also fit a Riemannian geodesic or a group geodesic for enhancement of 1-dimensional structures. A comparison of the two suggests that one should prefer the group geodesic [Duits 2011][Duits 2010b](called the "exponential curve" here).

8.3.3.3 Third Step: Projection (P)

The processed lifted image on $\mathbb{R}^2 \times S^1$ is projected to a "standard" image defined on $\mathbb{R}^2$. The projection can be done in two different ways. First, one can use the "verticality" along fibers of the bundle $\mathbb{R}^2 \times S^1$. In this case, one projects along the fiber S^1, choosing a θ that maximizes a likelihood criterion [Boscain 2014]. Second, one can use the "Δ-verticality" of sub-Riemannian geometry. In this case, one projects along the normal of the horizontal distribution Δ through a concentration scheme [Sanguinetti 2008]. This allows for several maxima at each point, i.e. crossings on the image.

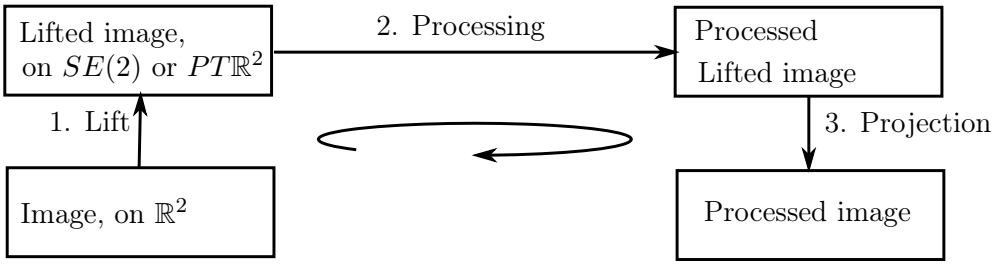


Figure 8.3: The 3 steps of image processing in 2D-Neurogeometry: LPP framework.

8.4 A Theoretical Toolbox for 3D-Neurogeometry

As implemented by now [Duits 2010a] [Duits 2011], image processing pipelines using 3D-neurogeometry also follows the three same steps: 1. Lift (L), 2. Processing (P) and 3. Projection (P) (see Fig. 8.4). This steps could be iterated or not. The difference with the 2D-neurogeometry however is the Processing. In this step, there is an additional level of structure in 3D-Neurogeometry w.r.t. the 2D case.

As in the 2D-case, we first survey the mathematical structures and summarize them in Table 8.5 (subsections 3.1,3.2, 3.3). Then we present the LPP-frame of a 3D-Neurogeometry (subsection 3.4). Application-oriented readers can read subsection 3.4 first, and then go to subsections 3.1, 3.2, 3.3.

Table 8.3: Properties of group actions on $SE(3)$. "Isotropy" means the isotropy groups. Actions on the $\mathbb{R}^3$-part of $SE(3) = \mathbb{R}^3 \times SO(3)$ are the main distinction between Left and Right. "Fundamental" denotes the fundamental representation on $\mathbb{R}^3$, and "Trivial" the trivial representation on $\mathbb{R}^3$.

		Free	**Transitive**	**Orbits**	**Isotropy**	**Quotient**	**On $\mathbb{R}^3$-part**
$SE(3)$**-actions**	**Left**	yes	yes	$SE(3)$	$\{e\}$	$\{[e]\}$	fundamental
	Right	yes	yes	$SE(3)$	$\{e\}$	$\{[e]\}$	trivial
$SO(3)$**-actions**	**Left**	yes	no	$\sim SO(3)$	$\{e\}$	$\mathbb{R}^3$	fundamental
	Right	yes	no	$\sim SO(3)$	$\{e\}$	$\mathbb{R}^3$	trivial
$SO(2)$**-actions**	**Left**	yes	no	$\sim SO(2)$	$\{e\}$	$\mathbb{R}^3 \times S^2$	fundamental
	Right	yes	no	$\sim SO(2)$	$\{e\}$	$\mathbb{R}^3 \times S^2$	trivial

8.4.1 Structures on the Lie Group $SE(3) = \mathbb{R}^3 \times SO(3)$

8.4.1.1 Group Actions

The law of $SE(3)$ is, for all $(t_1, R_1), (t_2, R_2) \in \mathbb{R}^3 \times SO(3)$:

$$(t_1, R_1) * (t_2, R_2) = (R_1.t_2 + t_1, R_1.R_2).$$

We read the group actions on $SE(3)$ and their properties (see Table 8.3).

As a Lie group, $SE(3)$ acts on itself through left and right translations. As subgroups of $SE(3)$, $SO(3)$ and $SO(2)$ also act on $SE(3)$, on the left and right. The right $SO(3)$-action makes $SE(3)$ a trivial principal bundle over $\mathbb{R}^3$ with structure group $SO(3)$. The right $SO(2)$-action on $SE(3)$ makes the $SE(3)$ a principal bundle over $\mathbb{R}^3 \times S^2$ with structure group $SO(2)$.

8.4.1.2 A Left-Invariant Metric and a Bi-Invariant Pseudo-Metric

As in the 2D case, one defines the left-invariant metric g^R_μ on $SE(3)$. g^R_μ is left-invariant by construction. But g^R_μ is not right-invariant.

g^R_μ is invariant by the left and right $SO(3)$-actions. The left $SO(3)$-invariance comes from the left invariance of g^R_μ. The right $SO(3)$-invariance is shown considering the right action on the parts $\mathbb{R}^3$ and $SO(3)$ separately, as g^R_μ is diagonal. Consequently, g^R_μ is also invariant by left and right $SO(2)$-actions.

As opposed as the $2D$ case, there exist bi-invariant pseudo-metrics on $SE(3)$. We refer to [Miolane 2015b] for their explicit construction. A possible choice is:

$$g^{BI}(0, \mathbb{I}_3)_{ij} = \begin{pmatrix} 0 & \mathbb{I}_3 \\ \mathbb{I}_3 & 0 \end{pmatrix}$$

known as the Klein form. Here $\mathbb{I}_3$ is the 3D identity matrix.

Table 8.4: Induced group actions on $\mathbb{R}^3 \times S^2$. Note that there are no more right actions, as $SO(2)$ is not a normal group of $SO(3)$ nor $SE(3)$. "Isotropy" means the "isotropy groups". "Fundamental" denotes the fundamental representation on $\mathbb{R}^3$.

		Free	**Trans.**	**Orbits**	**Isotropy**	**Quotient**	**On $\mathbb{R}^3$-part**
$SE(3)$**-action**	**Left**	no	yes	$\mathbb{R}^3 \times S^2$	$SO(2)$	$\{[e]\}$	fundamental
$SO(3)$**-action**	**Left**	no	no	$\sim SO(3)/SO(2)$	$SO(2)$	$\mathbb{R}^3$	fundamental

8.4.1.3 A Survey of Curves on $SE(3)$

From Table 8.1 and the aforementioned structures, we survey the curves on $SE(3)$. We have the group geodesics of $SE(3)$, the Riemannian geodesics of g^R_μ an the pseudo-Riemannian geodesics of g^{BI}. The group geodesics coincide with the pseudo-Riemannian ones, but differ from the Riemannian ones [Postnikov 2001] (see Table 8.1).

With respect to. a $SO(3)$- or $SO(2)$-action, some of these curves are vertical, some are horizontal (taken w.r.t. g^R_μ). Examples of vertical group geodesics are the orbits of the $SO(2)$-action or the action of the group geodesics of $SO(3)$. Examples of horizontal group geodesics are those generated by an element of the Lie algebra of the translations.

8.4.2 Structures on the Lifted Space $\mathbb{R}^3 \times S^2$ and on $\mathbb{R}^3$

We go from $SE(3)$ to $\mathbb{R}^3 \times S^2$, by quotienting the right $SO(2)$-action. The quotient is implemented by choosing an origin in $\mathbb{R}^3 \times S^2$, usually $(0, a)$. An element $(x, n) \in \mathbb{R}^3 \times S^2$ is represented as the result of the action of the corresponding (x, R) on $(0, a)$, where R is precisely the rotation bringing a onto n.

8.4.2.1 Induced Group Actions

The induced action of $SE(3)$ on $\mathbb{R}^3 \times S^2$ writes, for all $(t, R) \in SE(3)$ and $(x, n) \in \mathbb{R}^3 \times S^2$:

$$(t, R) * (x, n) = (R.x + t, R.n)$$

We read the group actions on $\mathbb{R}^3 \times S^2$ and their properties (see Table 8.4).

The $SE(3)$-action is transitive on $\mathbb{R}^3 \times S^2$. It makes $\mathbb{R}^3 \times S^2$ a homogeneous space. As the isotropy group is $SO(2)$ everywhere, the orbit-stabilizer theorem gives: $\mathbb{R}^3 \times S^2 = SE(3)/SO(2)$. Moreover, it provides the justification of the choice of an origin $(0, a)$ in computer vision algorithms. All points are equivalent in a homogeneous space. Computations do not depend on the choice of origin.

8.4.2.2 Induced Riemannian and Pseudo-Riemannian Metrics

g^R_μ was invariant by the right $SO(2)$-action. Thus, the projection onto $\mathbb{R}^3 \times S^2$ is a Riemannian submersion for g^R_μ. It induces a Riemannian metric on $\mathbb{R}^3 \times S^2$, still denoted g^R_μ. g^R_μ is still $SE(3)$- and $SO(3)$- invariant.

Similarly, g^{BI} was invariant by the right $SO(2)$-action. It induces a Riemannian pseudo-metric on $\mathbb{R}^3 \times S^2$, which is still $SE(3)$- and $SO(3)$- invariant.

8.4.2.3 A Sub-Riemannian Metric

As in 2D, one defines a sub-Riemannian metric g^{SR} on $\mathbb{R}^3 \times S^2$ by first defining Δ. We take $(X_1, X_2, X_3, X_4, X_5)$ on $\mathbb{R}^3 \times S^2$ as:

$$\begin{cases} X_1 = \cos\theta\cos\phi.\partial_x + \cos\theta\sin\phi.\partial_y - \sin\theta.\partial_z, \\ X_2 = -\sin\phi.\partial_x + \cos\phi.\partial_y, \\ X_3 = \partial_\theta, \\ X_4 = \partial_\phi, \\ X_5 = \sin\theta\cos\phi.\partial_x + \sin\theta\sin\phi.\partial_y + \cos\theta.\partial_z \end{cases}$$

and $\Delta = \text{Span}\{X_1, X_2, X_3, X_4\}$. g^{SR} is defined as the Euclidean metric on Δ. As in the 2D-case, it would be approximated by a Riemannian metric in practice.

8.4.2.4 A Survey of Curves

From Table 8.1 and the aforementioned structures, we survey the curves on $SE(3)$. However, we have a new class of curves in 3D-Neurogeometry w.r.t. 2D-Neurogeometry: the curves of the lifted space $\mathbb{R}^3 \times S^2$ that are projection of curves of $SE(3)$, as the projection of the group geodesics.

In the following, "verticality" and "horizontality" are taken w.r.t. the right $SO(2)$-action. Projecting horizontal (g^R_μ) Riemannian geodesics gives *generalized Riemannian geodesics*. Projecting horizontal (for g^{BI}) pseudo-Riemannian geodesics gives *generalized pseudo-Riemannian geodesics*. More precisely, a smooth horizontal curve in $SE(3)$ is a (pseudo-) Riemannian geodesics if and only if it is a (pseudo-) Riemannian geodesics in $\mathbb{R}^3 \times S^2$. The projection of vertical curves are points. The projection of a curve that is vertical at one point has a "cusp".

Ultimately, we have the curves that are Δ-horizontal in the sense of the sub-Riemannian geometry. Among them, we have sub-Riemannian geodesics.

8.4.3 Structures on the Image Domain $\mathbb{R}^3$

The previous structures are projected to $\mathbb{R}^3$, using the projection of the trivial bundle $\mathbb{R}^3 \times S^2$ on the first component. In particular, projecting the previous curves give curves in $\mathbb{R}^3$. We have: the projection of the sub-Riemannian geodesics (an equivalent of 2D elastica curves), the double-projection of the group geodesics (equivalently the double-projection of the pseudo-Riemannian curves for g^{BI}), the double-projection of the Riemannian geodesics for g^R_μ.

Table 8.5: Structures of 3D-Neurogeometry. Use Table 8.1 to get the related curves.

	Actions	Metrics
$SE(3)$	• left, right translations of $SE(3)$ • left, right actions of $SO(3)$ • left, right actions of $SO(2)$	• g^R_μ • g^{BI}
$\mathbb{R}^3 \times S^2$	• left action of $SE(3)$ • left action of $SO(3)$	• projection of g^R_μ • projection of g^{BI} • g^{SR} • g^{SR}_ε
$\mathbb{R}^3$	• left action of $SE(3)$ • left action of $SO(3)$	• Euclidean metric (double-projection of g^R_μ)

8.4.4 The Three Steps: LPP Framework for 3D

8.4.4.1 First Step: Lift (L)

As in 2D, one lifts the medical image defined on $D \subset \mathbb{R}^3$ to an image defined on $\tilde{D} \subset \mathbb{R}^3 \times S^2$, using the gradient direction at each $(x, y, z) \in D$:

$$\frac{\nabla I}{||\nabla I||} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$$

8.4.4.2 Second Step: Processing (P)

First, as in 2D, the processing could be performed on $\mathbb{R}^3 \times S^2$ without taking into account the $SE(3)$ structure. One would only consider the sub-Riemannian structure on the lifted space $\mathbb{R}^3 \times S^2$. In doing so, one could define sub-Riemannian partial differential equations as in 2D-Neurogeometry, using the X_i as differential operators. For in-painting purposes, the 2D work of [Boscain 2014][Sanguinetti 2008] provides intuition. Similarly, one could add drift (or convection) depending on the application.

Then, in contrast to 2D, the processing *can* be performed on $SE(3)$. This is done by embedding $\mathbb{R}^3 \times S^2$ in $SE(3)$ as the quotient of $SE(3)$ by a $SO(2)$-action. Then, performing $SO(2)$-invariant computations on $SE(3)$ is equivalent to performing computations on $\mathbb{R}^3 \times S^2$. The advantage is that one has more structures, e.g. more curves for curve fitting (compare subsections 2.2 and 3.4).

This is the first main distinction between the 2D and the 3D case. In 2D-Neurogeometry, we have one (trivial) quotient of $\mathbb{R}^2 \times S^2$. In contrast in 3D-Neurogeometry, one has two successive quotients of $SE(3) = \mathbb{R}^3 \times SO(3)$.

The second distinction is the existence of bi-invariant pseudo-metrics g^{BI} in the 3D-case, but not in the 2D-case [Miolane 2015b]. As such, g^{BI} could represent a new powerful tool of 3D-Neurogeometry. We note that in medical computer vision, the bi-invariant pseudo-metric g^{BI} is rarely used as opposed to algorithms in robotics [Zefran 1995]. Considering its bi-invariance property, it would be interesting to consider it for the computations. For example, g^{BI} characterizes the group geodesics of $SE(3)$: this could simplify computations. g^{BI} could replace the use of g^R_μ as an auxiliary metric, suppressing the need of a choice of μ.

8.4.4.3 Third Step: Projection (P)

The projection of the lifted image to an image defined on $\mathbb{R}^3$ could be defined in two different ways, exactly as in the 2D-case.

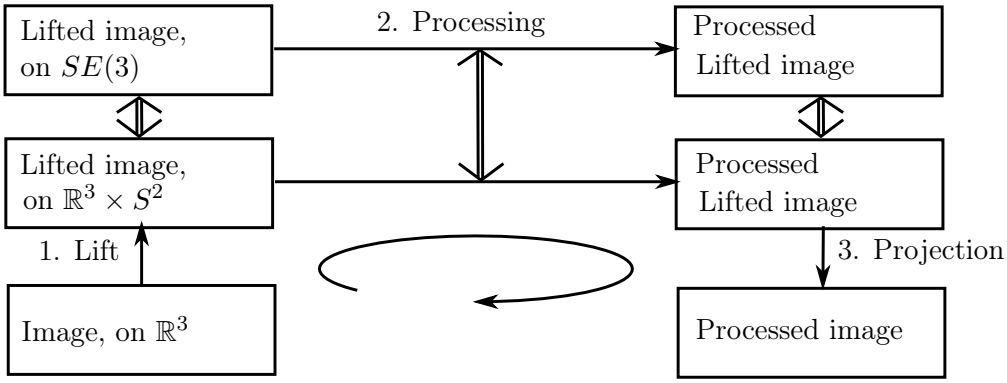


Figure 8.4: The 3 steps of image processing for 3D-Neurogeometry: LPP framework.

8.5 Conclusion

This paper is a *theoretical toolbox* for creating new algorithms in 3D medical computer vision. We have described the mathematical structures arising in the generalization of (2D-)Neurogeometry to 3D images.

Chapter 9

Conclusion

This Ph.D thesis developed Geometric Statistics for Computational Anatomy. Computational anatomy seeks to uncover the normal and pathological variability of the human anatomy, through the statistical analysis of organ shapes in medical images. Organ shapes are complex data that typically live in manifolds with additional geometric structures, like Lie groups or quotient spaces. Geometric Statistics generalizes the theory of statistics to such data spaces. Geometric Statistics is a fusion of Differential Geometry and Statistics that provides a successful framework for the computational analysis of the human anatomy.

9.1 Summary of the contributions

We summarize here the contributions of this Ph.D thesis with respect to the definition and use of Geometric Statistics for Computational Anatomy.

Theoretical developments on Geometric Statistics Geometric Statistics is a mathematical theory in construction. We have first tackled its definition of the structure of finite dimensional Lie groups. Chapter 4 shows that a consistent theory of statistics, that works for all Lie groups, needs to be defined without (pseudo-)Riemannian metric. This is because a bi-invariant (pseudo-)Riemannian metric fails to exist for some Lie groups. Geometric Statistics on Lie groups may be instead defined with a bi-invariant affine connection, like the Cartan-Schouten connection, that exists on all Lie groups.

We also considered the case where one still wants to try Statistics on (pseudo-)Riemannian manifolds on a given Lie group. We have thus provided an algorithm that constructs a bi-invariant (pseudo-)metric on the group given as input, in case of existence. In other words, our algorithm tells if Riemannian statistics are applicable for the Lie group of the study. If they are, the algorithm outputs a Riemannian metric that respects the Lie group structure, i.e. that is bi-invariant.

The above contributions are examples where the geometry is given - the Lie group structure - but we need to define the statistics. Our next contribution comes from the converse situation: the statistical principles are given, but we seek to construct the appropriate geometric structures. In this respect, Chapter 8 has set geometric foundations to use the sub-Riemannian Statistics, already applied in 2D image processing, for 3D image processing. The geometric structures involved in the 3D case are indeed significantly different, and more complicated, than the structures involved in the 2D case. In particular, we have seen that a new quotient space

appears in the 3D case. We have thus provided a toolbox of Differential Geometry needed to develop Geometric Statistics for the processing of 3D medical images.

Geometric Statistics for the analysis of statistics on shapes Then, we have shown that Geometric Statistics is an efficient formal setting to analyze statistical properties of algorithms on shape data. We can use Differential Calculus to derive the statistical properties of procedures employed in Computational Anatomy. Chapter 5 has focused on the algorithm of template shape computation, which has been used for 15 years in the field. We have evaluated its consistency as an estimator of the unique anatomy, which is a parameter of the generative model. We have shown that the algorithm is asymptotically biased.

Then, we have extended our analysis to algorithms that use registration as a pre-processing step for statistics on shapes. Chapter 7 has applied Geometric Statistics to emphasize that registration corresponds to a non-orthogonal projection in a quotient space. As a consequence, there is bias in such statistical analysis of shapes, just as there was bias in the computation of the template anatomy. Geometric Statistics thus enables to analyze a wide class of algorithms.

Geometric Statistics for the improvement of algorithms on shape data Geometric Statistics facilitates the analysis of estimators on manifolds with additional properties. As presented in Chapter 2, this analysis shall be used to improve the properties of the estimators. Geometric Statistics helps us to understand the geometric origin of pathological statistical properties of some algorithms in Computational Anatomy. As a consequence, we have leveraged this intuition to correct the algorithms.

First, we have proposed methods to correct for the bias in template shape estimation. These methods are inspired by estimation theory on vector spaces, reviewed in Chapter 2. Chapter 5 adapts bootstrap techniques to perform bias correction in the quotient space. The bootstrap methods proposed are very interesting because they are resampling methods and as such, they work on any type of (shape) data: landmarks, curves, surfaces, etc.

Second, we have suggested a theoretical approach to correct for the general bias occurring in every algorithm of statistical shape analysis that uses registration as a pre-processing step. This bias comes from the non-linearity of the projection in the quotient shape space. Our method, presented in Chapter 7 offers to modify the ambient Riemannian metric through a conformal transformation is order to make this projection linear. We deform the data space to overcome the bias induced by the geometry. This correction is therefore very general and applicable to a wide class of algorithms.

Applications to anatomical data Our last contributions focus on the applications of Geometric Statistics to simulated and real datasets in Computational Anatomy. We have validated our results on toy examples of landmarks in $\mathbb{R}^m$ and

the hypersphere S^m to illustrate the apparition of the bias in the template shape estimation. Then, we have moved to real datasets and shown the bias on shapes of the optical nerve on monkeys in Chapter 5. We have further illustrated this bias on brain images in Chapter 6, where we have developed a topological method to model the quotient by the action of diffeomorphisms. This has enabled us to show where the bias is the most important locally on the brain template.

We have quantified the bias on simulated and real datasets but we have also shown the usefulness of our methods to correct it. The two bootstrap methods have been validated on simulated datasets. Then we have used them to quantify and correct the bias on template shapes of landmarks describing optical nerves of monkeys in Chapter 5. For neuroimaging, we have introduced a new procedure relying on topological constraints in Chapter 6, in order to control the template's inconsistency by forcing it to stay below a threshold. These contributions show the impact of Geometric Statistics for Computational Anatomy.

9.2 Perspectives

This Ph.D thesis opens the door to other developments of Geometric Statistics, and to its applications in Computational Anatomy as well as in other fields. We present here some of its perspectives.

Extend Geometric Statistics to other manifolds and statistical procedures Geometric Statistics relies on Differential Geometry and Statistics: the field is therefore immensely rich as we can vary both the properties of the data space and the statistical procedures. For example, we can consider the sample mean on Lie groups defined with affine connection and study its asymptotic properties in this context. We can also define a sub-Riemannian structure on Lie groups, define a notion of mean and study its properties. Beyond Lie groups, we can explore notions of means on fiber bundles, principal bundles etc.

While focusing on one geometric space, we can investigated statistical moments beyond the mean, like the variance for example. Eventually, statistical procedures like dimension reduction techniques or regression can be generalized to Lie groups, sub-Riemannian spaces, bundles, etc. We have shown that bootstrap techniques could be adapted to improve estimators on quotient spaces. We can also adapt them to improve future estimators on other manifolds with geometric properties, like affine connection spaces. The theoretical aspect of Geometric Statistics thus promises the development of various extensions.

Use Geometric Statistics to analyze and improve other algorithms We have investigated the asymptotic bias of the template estimation algorithm of Computational Anatomy. But we could also analyze other statistical properties of the same algorithm, and for example start with its bias in the low sample setting (small

data). Again, the analysis of the bias in the low sample setting would lead to the development of methods designed to improve the estimator.

Then, we can investigate other algorithms of medical imaging or machine learning through the lens of Differential Geometry. For example, general algorithms including trees or forests could get an interesting formulation in terms of stratified spaces. Such algorithms could be reformulated using Geometric Statistics so that their variance and bias could be studied.

Use intuition of Geometric Statistics to create new algorithms Just as Statistics and Machine learning can be considered distinct, one could consider going from Geometric Statistics to "Geometric Machine Learning". Instead of dealing with input from Euclidean spaces, one would deal with input from spaces with geometric properties. This idea has already been exploited with the introduction of kernel in usual learning algorithms. Kernels enable to run the usual machine learning algorithm on spaces with another metric structure than the Euclidean one. Nevertheless, "Geometric Machine Learning" would go further by enabling to take into account other geometric structures of the space, non only the one related to its inner product, like algebraic structures.

Furthermore, we could build on the mathematical toolbox of Chapter 8 to develop some Geometric Statistics applied to Computer Vision. Sub-Riemannian geometry indeed allows to envision various applications regarding tridimensional image processing, like in-painting or dis-occlusion, anisotropic smoothing etc.

Geometric Statistics and Heterogeneous data analysis The ultimate goal of Computational Anatomy within Computational Medicine is to describe the observed anatomical variability with a biological and biophysical model. Therefore, we could merge Geometric Statistics for anatomical data with other input data. We could integrate anatomical, functional and physiological or clinical parameters. Such an heterogeneous statistical theory holds promises with respect to the clinical applications.

9.3 Epilogue

We have striven to express a tendency of the medical imaging community to converge towards a theory of statistics adapted to the data spaces of Computational Anatomy. With this Ph.D thesis, we have proposed *Geometric Statistics* as a general framework to encompass the efforts made in this direction.

We have provided clear results on the current research boundaries of Geometric Statistics as a mathematical theory, as well as practical applications of the powerful tools it provides. We believe that this double exposition will inspire bright minds in order to keep on building this theory, with the aim to provide the medical community with novel and efficient statistical tools. Ultimately, we are certain that Geometric

Statistics will also find important developments in other fields and look forward to its expansion.

Bibliography

[Afsari 2011] Bijan Afsari. *Riemannian L^p center of mass: existence, uniqueness, and convexity.* Proc. Am. Math. Soc., vol. 139, no. 2, pages 655–673, 2011. (Cited on page 42.)

[Alekseevsky 2003] Dmitri Alekseevsky, Andreas Kriegl, Mark Losik and Peter W. Michor. *The Riemannian Geometry of Orbit Spaces. The Metric, Geodesics, and Integrable Systems.* Publ. Math. Debrecen, vol. 62, 2003. (Cited on pages 76, 106, 110, 138 and 150.)

[Allassonnière 2007] S. Allassonnière, Y. Amit and A. Trouvé. *Towards a Coherent Statistical Framework for Dense Deformable Template Estimation.* Journal of the Royal Statistical Society, vol. 69, no. 1, pages 3–29, 2007. (Cited on pages 73, 78, 80, 83, 102, 104, 105, 115, 129 and 134.)

[Allassonnière 2015a] Stéphanie Allassonnière, Loïc Devilliers and Xavier Pennec. *Estimating the Template in the Total Space with the Fréchet Mean on Quotient Spaces may have a Bias.* In Proceedings of the fifth international workshop on Mathematical Foundations of Computational Anatomy (MFCA'15), pages 131–142, 2015. (Cited on pages 78, 79, 83, 102, 114, 116 and 136.)

[Allassonnière 2015b] Stéphanie Allassonnière and Estelle Kuhn. *Convergent stochastic Expectation Maximization algorithm with efficient sampling in high dimension. Application to deformable template model estimation.* Computational Statistics & Data Analysis, vol. 91, pages 4 – 19, 2015. (Cited on pages 71 and 134.)

[Allassonnière 2016] Stéphanie Allassonnière, Loïc Devilliers and Xavier Pennec. *Fréchet means top and quotient space may not be consistent. (Personal communication).* 2016. (Cited on pages 73, 78 and 80.)

[Amari 1990] Shun-ichi Amari. Differential-geometrical methods in statistics. Lecture notes in statistics. Springer-Verlag, Berlin, Heidelberg, 1990. (Cited on page 8.)

[Amit 1991] Yali Amit, Ulf Grenander and Mauro Piccioni. *Structural Image Restoration Through Deformable Templates.* Journal of the American Statistical Association, vol. 86, no. 414, pages 376–387, 1991. (Cited on page 5.)

[Apostolova 2007] Liana G. Apostolova and Paul M. Thompson. *Brain mapping as a tool to study neurodegeneration.* Neurotherapeutics, vol. 4, no. 3, pages 387–400, 2007. (Cited on page 4.)

[Ashburner 1998] John Ashburner, Chloe Hutton, Richard Frackowiak, Ingrid Johnsrude, Cathy Price and Karl Friston. *Identifying global anatomical differences: Deformation-based morphometry.* Human Brain Mapping, vol. 6, no. 5-6, pages 348–357, 1998. (Cited on page 104.)

[Ashburner 2001] John Ashburner and Karl Friston. *Why voxel-based morphometry should be used.* Neuroimage, vol. 14, pages 1238–1243, 2001. (Cited on page 104.)

[Ashburner 2003] J. Ashburner and K.J. Friston. *Rigid body registration.* In R.S.J. Frackowiak, K.J. Friston, C. Frith, R. Dolan, K.J. Friston, C.J. Price, S. Zeki, J. Ashburner and W.D. Penny, editeurs, Human Brain Function. Academic Press, 2nd édition, 2003. (Cited on page 5.)

[Astrakhantsev 1978] V. V. Astrakhantsev. *Decomposability of metrizable Lie algebras.* Functional Analysis and Its Applications, vol. 12, no. 3, pages 210–212, July 1978. (Cited on pages 43, 44, 46, 51 and 53.)

[Baloch 2009] Sajjad Baloch and Christos Davatzikos. *Morphological appearance manifolds in computational anatomy: Groupwise registration and morphological analysis.* NeuroImage, vol. 45, no. 1, Supplement 1, pages S73 – S85, 2009. (Cited on page 104.)

[Baron 2001] J.C. Baron, G. ChÃľtelat, B. Desgranges, G. Perchey, B. Landeau, V. de la Sayette and F. Eustache. *In Vivo Mapping of Gray Matter Loss with Voxel-Based Morphometry in Mild Alzheimer's Disease.* NeuroImage, vol. 14, no. 2, pages 298 – 309, 2001. (Cited on page 4.)

[Bartels 1972] R. H. Bartels and G. W. Stewart. *Solution of the Matrix Equation AX + XB = C [F4].* Commun. ACM, vol. 15, no. 9, pages 820–826, September 1972. (Cited on page 48.)

[Beg 2005] M. Faisal Beg, Michael I. Miller, Alain Trouvé and Laurent Younes. *Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms.* International Journal of Computer Vision, vol. 61, no. 2, pages 139–157, 2005. (Cited on page 5.)

[Belavin 1998] A. Belavin and V. Drinfeld. *Triangle Equations and Simple Lie Algebras.* Mathematical physics reviews Vol. 4, 1998. (Cited on pages 44 and 56.)

[Bellaıche 1996] André Bellaıche and Jean-Jacques Risler. *Sub-Riemannian geometry.* Progress in Mathematics, vol. 144, 1996. (Cited on page 150.)

[Benayadi 2014] Said Benayadi and Alberto Elduque. *Classification of quadratic Lie algebras of low dimension.* Journal of Mathematical Physics, vol. 55, no. 8, page 081703, 2014. (Cited on pages 44 and 46.)

[Berman 2000] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov and Philip E. Bourne. *The Protein Data Bank*. Nucleic Acids Res, vol. 28, pages 235–242, 2000. (Cited on page 99.)

[Bhattacharya 2003] Rabi Bhattacharya and Vic Patrangenaru. *Large sample theory of intrinsic and extrinsic sample means on manifolds*. Ann. Statist., vol. 31, no. 1, pages 1–29, 02 2003. (Cited on page 7.)

[Bhattacharya 2005] Rabi Bhattacharya and Vic Patrangenaru. *Large sample theory of intrinsic and extrinsic sample means on manifoldsâĂŤII*. Ann. Statist., vol. 33, no. 3, pages 1225–1259, 06 2005. (Cited on page 7.)

[Bhattacharya 2012] A. Bhattacharya and R. Bhattacharya. Nonparametric inference on manifolds: With applications to shape spaces. Institute of Mathematical Statistics Monographs. Cambridge University Press, 2012. (Cited on page 6.)

[Bigot 2010] Jeremie Bigot and Sebastien Gadat. *A deconvolution approach to estimation of a common shape in a shifted curves model*. Ann. Statist., vol. 38, no. 4, pages 2422–2464, 2010. (Cited on pages 78, 80, 83 and 116.)

[Bigot 2011] Jeremie Bigot and Benjamin Charlier. *On the consistency of Fréchet means in deformable models for curve and image analysis*. Electronic Journal of Statistics, no. 5, pages 1054–1089, 2011. (Cited on pages 73, 78, 80, 105 and 134.)

[Billera 2001] Louis J. Billera, Susan P. Holmes and Karen Vogtmann. *Geometry of the Space of Phylogenetic Trees*. Advances in Applied Mathematics, vol. 27, no. 4, pages 733 – 767, 2001. (Cited on page 7.)

[Boisvert 2008] Jonathan Boisvert, Farida Cheriet, Xavier Pennec, Hubert Labelle and Nicholas Ayache. *Articulated Spine Models for 3D Reconstruction from Partial Radiographic Data*. IEEE Transactions on Bio-Medical Engineering, vol. 55, no. 11, pages 2565–2574, November 2008. (Cited on page 5.)

[Bookstein 1986] Fred L. Bookstein. *Size and Shape Spaces for Landmark Data in Two Dimensions*. Statist. Sci., vol. 1, no. 2, pages 181–222, 05 1986. (Cited on page 133.)

[Bookstein 1997] F. L. Bookstein. Morphometric Tools for Landmark Data. Cambridge University Press., June 1997. (Cited on page 6.)

[Bookstein 2001] F. Bookstein. *Voxel-based morphometry should not be used with imperfectly registered images*. Neuroimage, vol. 14 (6), pages 1454–1462, 2001. (Cited on page 104.)

[Bordemann 1997] M. Bordemann. *Nondegenerate invariant bilinear forms on nonassociative algebras.* Acta Mathematica Universitatis Comenianae. New Series, vol. 66, no. 2, pages 151–201, 1997. (Cited on page 44.)

[Boscain 2012] Ugo V. Boscain, Jean Duplaix, Jean-Paul Gauthier and Francesco Rossi. *Anthropomorphic Image Reconstruction via Hypoelliptic Diffusion.* SIAM J. Control and Optimization, vol. 50, no. 3, pages 1309–1336, 2012. (Cited on pages 117, 151 and 154.)

[Boscain 2014] U. Boscain, R. A. Chertovskih, J. P. Gauthier and A. O. Remizov. *Hypoelliptic Diffusion and Human Vision: A Semidiscrete New Twist.* SIAM Journal on Imaging Sciences, Society for Industrial and Applied Mathematics, vol. 7, no. 2, pages pp.669–695, 2014. (Cited on pages 6, 150, 155 and 159.)

[Bourbaki 1989] N. Bourbaki. Lie groups and lie algebras: Chapters 1-3. Elements of mathematics. Hermann, 1989. (Cited on pages 45, 48 and 150.)

[Breiman 1996] Leo Breiman. *Bagging Predictors.* Machine Learning, vol. 24, no. 2, pages 123–140, 1996. (Cited on page 26.)

[Brennan 1989] R G Brennan and B W Matthews. *The helix-turn-helix DNA binding motif.* Journal of Biological Chemistry, vol. 264, no. 4, pages 1903–6, 1989. (Cited on page 100.)

[Brewin 2009] Leo Brewin. *Riemann normal coordinate expansions using Cadabra.* Classical and Quantum Gravity, vol. 26, no. 17, page 175017, 2009. (Cited on pages 185, 192 and 202.)

[Brignell 2010] Christopher J. Brignell, Ian L. Dryden, S. Antonio Gattone, Bert Park, Stuart Leask, William J. Browne and Sean Flynn. *Surface shape analysis with an application to brain surface asymmetry in schizophrenia.* Biostatistics, vol. 11, no. 4, pages 609–630, 2010. (Cited on page 4.)

[Brooksbank 2008] Peter A. Brooksbank and Eugene M. Luks. *Testing isomorphism of modules.* Journal of Algebra, vol. 320, no. 11, pages 4020 – 4029, 2008. Computational Algebra. (Cited on page 61.)

[Brown 1992] Lisa Gottesfeld Brown. *A Survey of Image Registration Techniques.* ACM Comput. Surv., vol. 24, no. 4, pages 325–376, December 1992. (Cited on page 5.)

[Campoamor-Stursberg 2008] R. Campoamor-Stursberg. *Quasi-Classical Lie Algebras and their Contractions.* International Journal of Theoretical Physics, vol. 47, no. 2, pages 583–598, 2008. (Cited on pages 44 and 46.)

[Cartan 1933] Elie Cartan. Sur la structure des groupes de transformations finis et continus. 2. Edit. Paris: Vuibert. 157 S. (1933)., 1933. (Cited on pages 43 and 48.)

[Ceballos 2012] M. Ceballos, J. Núñez and Á. F. Tenorio. *Algorithmic Method to Obtain Abelian Subalgebras and Ideals in Lie Algebras.* Int. J. Comput. Math., vol. 89, no. 10, pages 1388–1411, 07 2012. (Cited on page 60.)

[Chasseing 2016] Sénat. Commission des affaires sociales. Chasseing. *Rapport d'information fait au nom de la commission des affaires sociales sur l'enquete de la Cour des comptes sur l'adaptation aux besoins des moyens materiels et humains consacres a l'imagerie medicale.* Les Rapports du Sénat, 2016. (Cited on pages 2 and 3.)

[Chou 2008] Yi-Yu Chou, Natasha Leporé, Greig I de Zubicaray, Owen T Carmichael, James T Becker, Arthur W Toga and Paul M Thompson. *Automated ventricular mapping with multi-atlas fluid image alignment reveals genetic effects in Alzheimer's disease.* Neuroimage, vol. 40, no. 2, pages 615–630, 2008. (Cited on page 4.)

[Chung 2009] Moo K. Chung, Vikas Singh, Peter T. Kim, Kim M. Dalton and Richard J. Davidson. Topological characterization of signal in brain images using min-max diagrams, pages 158–166. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. (Cited on page 105.)

[Citti 2006] Giovanna Citti and Alessandro Sarti. *A cortical based model of perceptual completion in the roto-translation space.* Journal of Mathematical Imaging and Vision, vol. 24, no. 3, pages 307–326, 2006. (Cited on pages 6, 150, 151, 153, 154 and 155.)

[Cohen 1997] A. M. Cohen, De Graaf W. A. and L. Rónyai. *Computations in finite-dimensional Lie algebras.* Discrete Mathematics and Theoretical Computer Science, vol. 1, no. 1, pages 129–138, 1997. (Cited on pages 44, 46 and 59.)

[Corcuera 1998] José Manuel Corcuera and Wilfrid S. Kendall. *Riemannian Barycentres and Geodesic Convexity.* Math. Proc. Cambridge Philos. Soc, vol. 127, pages 253–269, 1998. (Cited on pages 7 and 42.)

[Darmanté 2014] Hugo Darmanté, Benoit Bugnas, Regis Bernard De Dompsure, Laurent Barresi, Nina Miolane, Xavier Pennec, Fernand de Peretti and Nicolas Bronsard. *Analyse biométrique de lánneau pelvien en 3 dimensions âĂŞ à propos de 100 scanners.* Revue de Chirurgie Orthopédique et Traumatologique, vol. 100, no. 7, Supplement, pages S241 –, 2014. (Cited on pages 5, 71 and 133.)

[Davis 2010] Brad C. Davis, P. Thomas Fletcher, Elizabeth Bullitt and Sarang Joshi. *Population Shape Regression from Random Design Data.* International Journal of Computer Vision, vol. 90, no. 2, pages 255–266, 2010. (Cited on page 7.)

[De Graaf 2000] W.A. De Graaf. Lie algebras: Theory and algorithms. North-Holland mathematical library. Elsevier, 2000. (Cited on pages 44, 46 and 48.)

[de La Rosa 2006] José Ismael de La Rosa and Gilles Fleury. *Bootstrap Methods for a Measurement Estimation Problem.* IEEE Transactions on Instrumentation and Measurement, Institute of Electrical and Electronics Engineers, vol. 55(3), pages 820–827, 2006. (Cited on page 25.)

[Delorme 2001] Patrick Delorme. *Classification des triples de Manin pour les algebres de Lie reductives complexes: Avec un appendice de Guillaume Macey.* Journal of Algebra, vol. 246, no. 1, pages 97 – 174, 2001. (Cited on page 44.)

[Dempster 1977] A. P. Dempster, N. M. Laird and D. B. Rubin. *Maximum likelihood from incomplete data via the EM algorithm.* Journal of the Royal Society, Series B, vol. 39, no. 1, pages 1–38, 1977. (Cited on page 112.)

[Despotovic 2015] Ivana Despotovic, Bart Goossens and Wilfried Philips. *MRI Segmentation of the Human Brain: Challenges, Methods, and Applications.* J Computational and Mathematical Methods in Medicine, 2015. (Cited on page 104.)

[Drinfeld 1987] V. G. Drinfeld. Quantum groups. Amer. Math. Soc., Providence, RI, 1987. (Cited on pages 44 and 56.)

[Dryden 1992] I. L. Dryden and K. V. Mardia. *Size and shape analysis of landmark data.* Biometrika, vol. 79, no. 1, pages 57–68, 1992. (Cited on page 6.)

[Dryden 1998] I.L. Dryden and K.V. Mardia. Statistical shape analysis. John Wiley & Sons, New York, 1998. (Cited on pages 6, 72 and 134.)

[Du 2015] Jiejun Du, Ian L. Dryden and Xianzheng Huang. *Size and Shape Analysis of Error-Prone Shape Data.* Journal of the American Statistical Association, vol. 110, no. 509, pages 368–379, 2015. (Cited on page 73.)

[Duits 2010a] Remco Duits and EM Franken. *Left-invariant parabolic evolutions on SE (2) and contour enhancement via invertible orientation scores. Part I: Linear left-invariant diffusion equations on SE (2).* Quart. Appl. Math., vol. 68, pages 255–292, 2010. (Cited on pages 6, 150, 151 and 155.)

[Duits 2010b] Remco Duits and EM Franken. *Left-invariant parabolic evolutions on SE (2) and contour enhancement via invertible orientation scores. Part II : non linear left invariant diffusion equations on invertible orientation scores.* Quart. Appl. Math. 68, pages 293–331, 2010. (Cited on page 155.)

[Duits 2011] Remco Duits and Erik Franken. *Left-Invariant Diffusions on the Space of Positions and Orientations and their Application to Crossing-Preserving Smoothing of HARDI images.* International Journal of Computer Vision, vol. 92, no. 3, pages 231–264, 2011. (Cited on pages 150, 154 and 155.)

[Duong 2011] Minh-Thanh Duong. *A new invariant of quadratic lie algebras and quadratic lie superalgebras.* Theses, Université de Bourgogne, 2011. (Cited on pages 44, 48 and 63.)

[Dupuis 1998] Paul Dupuis and Ulf Grenander. *Variational problems on flows of diffeomorphisms for image matching.* Q. Appl. Math., vol. LVI, no. 3, pages 587–600, 1998. (Cited on page 5.)

[Durrleman 2010] Stanley Durrleman. *Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution. (Modèles statistiques de courants pour mesurer la variabilité anatomique de courbes, de surfaces et de leur évolution).* PhD thesis, University of Nice Sophia Antipolis, France, 2010. (Cited on page 5.)

[Efron 1979] B. Efron. *Bootstrap Methods: Another Look at the Jackknife.* The Annals of Statistics, vol. 7, no. 1, pages 1–26, 1979. (Cited on page 90.)

[Efron 1987] Bradley Efron. *Better Bootstrap Confidence Intervals.* Journal of the American Statistical Association, vol. 82, no. 397, pages 171–185, 1987. (Cited on page 26.)

[Efron 1993] B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*, 1993. (Cited on pages 20, 23 and 24.)

[Ehlers 2012] Jürgen Ehlers, Felix A. E. Pirani and Alfred Schild. *Republication of: The geometry of free fall and light propagation.* General Relativity and Gravitation, vol. 44, no. 6, pages 1587–1609, 2012. (Cited on page 92.)

[Elewa 2012] Ashraf M. T. Elewa. Morphometrics for nonmorphometricians. Springer, 2012. (Cited on page 71.)

[Émery 1991] M. Émery and G. Mokobodzki. *Sur le barycentre d'une probabilité dans une variété.* Séminaire de probabilités de Strasbourg, vol. 25, pages 220–233, 1991. (Cited on pages 7, 42 and 81.)

[Eriksson 2001] S. H. Eriksson, F. J. Rugg-Gunn, M. R. Symms, G. J. Barker and J. S. Duncan. *Diffusion tensor imaging in patients with epilepsy and malformations of cortical development.* Brain, vol. 124, no. 3, pages 617–626, 2001. (Cited on page 4.)

[Evans 2012] A.C. Evans, A.L. Janke, D.L. Collins and S. Baillet. *Brain templates and atlases.* Neuroimage, vol. 62(2), pages 911–922, 2012. (Cited on pages 72, 103 and 104.)

[Favre 1987] G. Favre and L.J Santharoubane. *Symmetric, invariant, non-degenerate bilinear form on a Lie algebra.* Journal of Algebra, vol. 105, no. 2, pages 451 – 464, 1987. (Cited on pages 44 and 46.)

[Fletcher 2004] P Thomas Fletcher, Conglin Lu, Stephen M Pizer and Sarang Joshi. *Principal geodesic analysis for the study of nonlinear statistics of shape.* IEEE transactions on medical imaging, vol. 23, no. 8, pages 995–1005, 2004. (Cited on page 7.)

[Fréchet 1944] M. Fréchet. L'intégrale abstraite d'une fonction abstraite d'une variable abstraite et son application a la moyenne d'un élément aléatoire de nature quelconque. La Revue Scientifique, 1944. (Cited on page 42.)

[Fréchet 1948] M. Fréchet. *Les éléments aléatoires de nature quelconque dans un espace distancié*. Annales de l'institut Henri Poincaré, vol. 10, no. 4, pages 215–310, 1948. (Cited on pages 6 and 42.)

[Frobenius 1878] F.G. Frobenius. Ueber lineare substitutionen und bilineare formen. Reimer, 1878. (Cited on page 61.)

[Gerber 2010] Samuel Gerber, Tolga Tasdizen, P. Thomas Fletcher, Sarang C. Joshi and Ross T. Whitaker. *Manifold modeling for brain population analysis.* Medical Image Analysis, vol. 14, no. 5, pages 643–653, 2010. (Cited on page 4.)

[Gerber 2012] Samuel Gerber and Kristin Potter. *Data Analysis with the Morse-Smale Complex: The msr Package for R.* Journal of Statistical Software, vol. 050, no. i02, 2012. (Cited on pages 109, 119, 120 and 127.)

[GlobalData 2011] GlobalData. *US cloud computing market for medical imaging.* Strategic Report, 2011. (Cited on page 4.)

[Goodall 1991] Colin Goodall. *Procrustes Methods in the Statistical Analysis of Shape.* Journal of the Royal Statistical Society. Series B (Methodological), vol. 53, no. 2, pages 285–339, 1991. (Cited on pages 6 and 72.)

[Goodall 1993] Colin R. Goodall and Kanti V. Mardia. *Multivariate Aspects of Shape Theory.* Ann. Statist., vol. 21, no. 2, pages 848–866, 06 1993. (Cited on page 6.)

[Gower 2004] John C. Gower and Garmt B. Dijksterhuis. Procrustes problems, volume 30 of *Oxford Statistical Science Series*. Oxford University Press, Oxford, UK, January 2004. (Cited on pages 6 and 72.)

[Grenander 1993] U. Grenander. General pattern theory: a mathematical study of regular structures. Oxford mathematical monographs. Clarendon, 1993. (Cited on page 4.)

[Grunwald 1906] Josef Grunwald. *Uber duale Zahlen und ihre Anwendung in der Geometrie.* Monatshefte fuer Mathematik und Physik, vol. 17, no. 1, pages 81–136, 1906. (Cited on page 57.)

[Guimond 1998] A. Guimond, J. Meunier and J.-P. Thirion. *Automatic Computation of Average Brain Models.* In Proc. of First Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI'98), volume 1496 of *LNCS*, pages 631–640, Cambridge, USA, October 1998. Springer. (Cited on page 103.)

[Günther 2014] D. Günther, A. Jacobson, J. Reininghaus, H.-P. Seidel, O. Sorkine-Hornung and T. Weinkauf. *Fast and Memory-Efficient Topological Denoising of 2D and 3D Scalar Fields.* IEEE Transactions on Visualization and Computer Graphics (Proc. IEEE VIS), vol. 20, no. 12, pages 2585–2594, December 2014. (Cited on page 125.)

[Guts 1984] A. K. Guts and A. V. Levichev. *On the Foundations of Relativity Theory.* Doklady Akademii Nauk SSSR, vol. 277, no. 6, pages pp. 1299–1303., 1984. (Cited on page 43.)

[Gyulassy 2008] A. Gyulassy, P.-T. Bremer, B. Hamann and V. Pascucci. *A Practical Approach to Morse-Smale Complex Computation: Scalability and Generality.* IEEE Transactions on Visualization and Computer Graphics, vol. 14, no. 6, pages 1619–1626, 2008. (Cited on pages 118 and 119.)

[Hadj-Hamou 2016] Mehdi Hadj-Hamou, Marco Lorenzi, Nicholas Ayache and Xavier Pennec. *Longitudinal Analysis of Image Time Series with Diffeomorphic Deformations: A Computational Framework Based on Stationary Velocity Fields.* Frontiers in Neuroscience, June 2016. (Cited on pages 5 and 103.)

[Han 2003] Xiao Han. *Topology preserving level set method for geometric deformable models.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 6, pages 755–768, 2003. (Cited on page 106.)

[Hendriks 1991] Harrie Hendriks. *A Cramer-Rao type lower bound for estimators with values in a manifold.* Journal of Multivariate Analysis, vol. 38, no. 2, pages 245 – 261, 1991. (Cited on page 6.)

[Hilgert 1985] Joachim Hilgert and KarlH. Hofmann. *Lorentzian cones in real Lie algebras.* Monatshefte fuer Mathematik, vol. 100, no. 3, pages 183–210, 1985. (Cited on page 44.)

[Hofmann 1986] Karl H. Hofmann and Verena S. Keith. *Invariant quadratic forms on finite dimensional lie algebras.* Bulletin of the Australian Mathematical Society, vol. 33, pages 21–36, 2 1986. (Cited on page 44.)

[Hubel 1977] David Hubel, CM Willmer and JC Rutter. *Orientation columns in macaque monkey visual cortex demonstrated by the 2-deoxyglucose autoradiographic technique.* Nature, vol. 269, page 22, 1977. (Cited on page 150.)

[Huckemann 2010] S Huckemann, T Hotz and A Munk. *Intrinsic shape analysis: Geodesic principal component analysis for Riemannian manifolds modulo Lie group actions.* Statistica Sinica, vol. 20, pages 1–100, 2010. (Cited on pages 6, 7, 32, 73, 75, 76, 106, 134 and 138.)

[Hughes 1996] Bruce Hughes. *Geometric topology of stratified spaces.* Electron. Res. Announc. Am. Math. Soc., vol. 02, no. 2, pages 73–81, 1996. (Cited on page 111.)

[Huiling 2004] Le Huiling. *Estimation of Riemannian Barycentres.* LMS Journal of Computation and Mathematics, vol. 7, pages 193–200, 1 2004. (Cited on page 42.)

[I. L. Dryden 1991] K. V. Mardia I. L. Dryden. *General Shape Distributions in a Plane.* Advances in Applied Probability, vol. 23, no. 2, pages 259–276, 1991. (Cited on page 6.)

[Jacobson 2012] A. Jacobson, T. Weinkauf and O. Sorkine. *Smooth Shape-Aware Functions with Controlled Extrema.* Computer Graphics Forum (Proc. SGP), vol. 31, no. 5, pages 1577–1586, July 2012. (Cited on pages 125 and 126.)

[Jones 1987] Judson P Jones and Larry A Palmer. *An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex.* Journal of neurophysiology, vol. 58, no. 6, pages 1233–1258, 1987. (Cited on page 149.)

[Joshi 2000] Sarang C. Joshi and Michael I. Miller. *Landmark matching via large deformation diffeomorphisms.* IEEE Trans. Image Processing, vol. 9, no. 8, pages 1357–1370, 2000. (Cited on page 5.)

[Joshi 2004] S. Joshi, Brad Davis, B Matthieu Jomier and Guido Gerig B. *Unbiased diffeomorphic atlas construction for computational anatomy.* Neuroimage, vol. 23, pages 151–160, 2004. (Cited on pages 103, 113 and 126.)

[Joshi 2006] Shantanu Joshi, David Kaziska, Anuj Srivastava and Washington Mio. *Riemannian Structures on Shape Spaces: A Framework for Statistical Inferences.* In Statistics and Analysis of Shapes, pages 313–333. Birkhaeuser Boston, 2006. (Cited on pages 73 and 90.)

[Jupp 1989] P. E. Jupp and K. V. Mardia. *A Unified View of the Theory of Directional Statistics, 1975-1988.* International Statistical Review / Revue Internationale de Statistique, vol. 57, no. 3, pages 261–294, 1989. (Cited on page 6.)

[Karcher 1977] H. Karcher. *Riemannian center of mass and mollifier smoothing.* Communications on Pure and Applied Mathematics, vol. 30, no. 5, pages 509–541, 1977. (Cited on pages 6 and 42.)

[Karger 1985] Adolf Karger and 1905 Novák Josef. Space kinematics and lie groups. New York : Gordon and Breach Science Publishers, 1985. Translation of: Prostorová kinematika a Liehovy grupy. (Cited on page 69.)

[Kath 2004] Ines Kath and Martin Olbrich. *Metric Lie algebras with maximal isotropic centre.* Mathematische Zeitschrift, vol. 246, pages 23–53, 2004. (Cited on pages 44 and 46.)

[Kath 2006] Ines Kath and Martin Olbrich. *Metric Lie algebras and quadratic extensions.* Transformation Groups, vol. 11, pages 87–131, 2006. (Cited on pages 44 and 46.)

[Keith 1984] V.S. Keith. *On Invariant Bilinear Forms on Finite-dimensional Lie Algebras.* Phd thesis, Tulane University, 1984. (Cited on pages 43, 44, 52 and 54.)

[Kendall 1977] David G Kendall. *The diffusion of shape.* Advances in applied probability, vol. 9, no. 3, pages 428–430, 1977. (Cited on page 72.)

[Kendall 1984] David G. Kendall. *Shape Manifolds, Procrustean Metrics, and Complex Projective Spaces.* Bulletin of the London Mathematical Society, vol. 16, no. 2, pages 81–121, 1984. (Cited on pages 6, 71 and 132.)

[Kendall 1989] David G. Kendall. *A Survey of the Statistical Theory of Shape.* Statistical Science, vol. 4, no. 2, pages pp. 87–99, 1989. (Cited on pages 133, 134 and 141.)

[Kendall 1990] Wilfrid S. Kendall. *Probability, Convexity, and Harmonic Maps with Small Image I: Uniqueness and Fine Existence.* Proceedings of the London Mathematical Society, vol. s3-61, no. 2, pages 371–406, 1990. (Cited on page 42.)

[Kenwright 2012] Ben Kenwright. *A Beginners Guide to Dual-Quaternions: What They Are, How They Work, and How to Use Them for 3D Character Hierarchies.* The 20th International Conference on Computer Graphics, Visualization and Computer Vision, pages 1–13, 2012. WSCG 2012 Communication Proceedings. (Cited on page 70.)

[Khanal 2016] Bishesh Khanal, Marco Lorenzi, Nicholas Ayache and Xavier Pennec. *A biophysical model of brain deformation to simulate and analyze longitudinal {MRIs} of patients with Alzheimer's disease.* NeuroImage, vol. 134, pages 35 – 52, 2016. (Cited on page 5.)

[Kitagawa 1977] G. Kitagawa. *An algorithm for solving the matrix equation $X = FXF\ T + S$.* International Journal of Control, vol. 25, no. 5, pages 745–753, 1977. (Cited on page 48.)

[Koenderink 1987] J.J. Koenderink and A.J. van Doorn. *Representation of local geometry in the visual system.* Biological Cybernetics, vol. 55, no. 6, pages 367–375, 1987. (Cited on page 149.)

[Kubicki 2007] Marek Kubicki, Robert McCarley, Carl-Fredrik Westin, Hae-Jeong Park, Stephan Maier, Ron Kikinis, Ferenc Jolesz and Martha Shenton. *A Review of Diffusion Tensor Imaging Studies in Schizophrenia.* Journal of Psychiatric Research, vol. 41, no. 1-2, pages 15–30, 01 2007. (Cited on page 4.)

[Kurtek 2011] Sebastian A. Kurtek, Anuj Srivastava and Wei Wu. *Signal Estimation Under Random Time-Warpings and Nonlinear Signal Alignment.* In J. Shawe-taylor, R.s. Zemel, P. Bartlett, F.c.n. Pereira and K.q. Weinberger, editeurs, Advances in Neural Information Processing Systems 24, pages 675–683, 2011. (Cited on pages 71, 73, 78, 79, 80, 83, 105, 134 and 141.)

[Le 1993] Huiling Le and David G. Kendall. *The Riemannian Structure of Euclidean Shape Spaces: A Novel Environment for Statistics.* The Annals of Statistics, vol. 21, no. 3, pages 1225–1271, 09 1993. (Cited on pages 6 and 72.)

[Le 1998] Huiling Le. *On the Consistency of Procrustean Mean Shapes.* Advances in Applied Probability, vol. 30, no. 1, pages 53–63, March 1998. (Cited on pages 72 and 134.)

[Lee 2007] John A. Lee and Michel Verleysen. Nonlinear dimensionality reduction. Springer Publishing Company, Incorporated, 1st édition, 2007. (Cited on page 8.)

[Lehmann 1998] E.L. Lehmann and G. Casella. Theory of Point Estimation. Springer Verlag, 1998. (Cited on page 16.)

[Lele 1993] S. Lele. *Euclidean distance matrix analysis (EDMA): estimation of mean form and mean form difference.* Mathematical Geology, vol. 25, pages 573–602, 1993. (Cited on pages 72, 73 and 134.)

[Levi 1905] E. Levi. Sulla struttura dei gruppi finiti e continui, volume 40. Atti della Reale Accademia delle Scienze di Torino, 1905. (Cited on page 59.)

[Li 2008] Jia-Yi Li, Elisabet Englund, Janice Holton, Denis Soulet, Peter Hagell, Andrew Lees, Tammaryn Lashley, Niall Quinn, Stig Rehncrona, Anders Bjorklund, Hakan Widner, Tamas Revesz, Olle Lindvall and Patrik Brundin. *Lewy bodies in grafted neurons in subjects with Parkinson's disease suggest host-to-graft disease propagation.* Nature, vol. 14, 2008. (Cited on pages 71 and 133.)

[Lifshits 1995] M. A. Lifshits. Infinite-dimensional gaussian distributions, pages 68–83. Springer Netherlands, Dordrecht, 1995. (Cited on page 111.)

[Lorenzi 2010] Marco Lorenzi, Nicholas Ayache, G. Frisoni and Xavier Pennec. *4D registration of serial brain MR's images: a robust measure of changes applied*

to Alzheimer's disease. In Miccai Workshop on Spatio-Temporal Image Analysis for Longitudinal and Time-Series Image Data, Beijing, China, September 2010. Best Oral Presentation award. (Cited on page 5.)

[Lorenzi 2011] Marco Lorenzi, Nicholas Ayache, Giovanni B. Frisoni and Xavier Pennec. *Mapping the effects of $A\beta_{1-42}$ levels on the longitudinal changes in healthy aging: hierarchical modeling based on stationary velocity fields*. In Proceedings of Medical Image Computing and Computer Assisted Intervention (MICCAI), volume 6892 of *LNCS*, pages 663–670. Springer, 2011. (Cited on pages 4 and 71.)

[Lorenzi 2013] Marco Lorenzi, Nicholas Ayache, Giovanni B. Frisoni and Xavier Pennec. *LCC-Demons: a robust and accurate symmetric diffeomorphic registration algorithm*. NeuroImage, vol. 81, no. 1, pages 470–483, 2013. (Cited on page 126.)

[Lytchak 2010] Alexander Lytchak and Gudlaugur Thorbergsson. *Curvature explosion in quotients and applications*. J. Differential Geom., vol. 85, no. 1, pages 117–140, 2010. (Cited on pages 88 and 115.)

[MacDonald 2000] David MacDonald, Noor Kabani, David Avis and Alan C Evans. *Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI*. NeuroImage, vol. 12, no. 3, pages 340–356, 2000. (Cited on page 106.)

[Mangin 1995] Jean-François Mangin, Vincent Frouin, Isabelle Bloch, Jean Régis and Jaime López-Krahe. *From 3D magnetic resonance images to structural representations of the cortex topography using topology preserving deformations*. Journal of Mathematical Imaging and Vision, vol. 5, no. 4, pages 297–318, 1995. (Cited on page 106.)

[Marco 2013] Lorenzi Marco and Xavier Pennec. Parallel transport with pole ladder: Application to deformations of time series of images, pages 68–75. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. (Cited on page 92.)

[Marcus 2007] DS Marcus, TH Wang, J Parker, JG Csernansky, JC Morris and RL. Buckner. *Open Access Series of Imaging Studies (OASIS): Cross-Sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults*. Journal of Cognitive Neuroscience., vol. 19, pages 1498–1507, 2007. (Cited on page 128.)

[Mardia 1999] K. V. Mardia. *Directional statistics and shape analysis*. Journal of Applied Statistics, vol. 26, no. 8, pages 949–957, 1999. (Cited on page 6.)

[Matheron 1975] G. Georges Matheron. Random sets and integral geometry. Wiley series in probability and mathematical statistics. Wiley, New York, London, 1975. Includes index. (Cited on page 8.)

[Medina 1982] A. Medina. *Groupes de Lie munis de pseudo-métriques de Riemann bi-invariantes.* Sémin. géométrie différentielle 1981-1982, Montepellier 1982, Exp. No.6, 37 p. (1982)., 1982. (Cited on pages 44, 51, 52, 53 and 67.)

[Medina 1985a] A. Medina and P. Revoy. *Algèbres de Lie et produit scalaire invariant.* Annales scientifiques de l'École Normale Supérieure, vol. 18, no. 3, pages 553–561, 1985. (Cited on pages 44, 48, 51, 52, 53, 54 and 56.)

[Medina 1985b] Alberto Medina and Philippe Revoy. *Les groupes oscillateurs et leurs reseaux.* manuscripta mathematica, vol. 52, no. 1-3, pages 81–95, 1985. (Cited on page 43.)

[Michor 2013] Peter W. Michor and David Mumford. *A zoo of diffeomorphism groups on Rn.* Annals of Global Analysis and Geometry, vol. 44, no. 4, pages 529–540, 2013. (Cited on page 107.)

[Miller 1997] M. Miller, A. Banerjee, G. Christensen, S. Joshi, N. Khaneja, U. Grenander and L. Matejic. *Statistical Methods in Computational Anatomy.* Stat Methods Med Res, vol. 6, no. 3, pages 267–299, September 1997. (Cited on page 4.)

[Milnor 1976] John Milnor. *Curvatures of left invariant metrics on lie groups.* Advances in Mathematics, vol. 21, no. 3, pages 293 – 329, 1976. (Cited on page 46.)

[Miolane 2015a] Nina Miolane and Xavier Pennec. *Biased estimators on quotient spaces.* Proceedings of the 2nd international of Geometric Science of Information (GSI'2015), 2015. (Cited on pages 73, 90, 105, 114, 115, 116 and 134.)

[Miolane 2015b] Nina Miolane and Xavier Pennec. *Computing Bi-Invariant Pseudo-Metrics on Lie Groups for Consistent Statistics.* Entropy, vol. 17, no. 4, pages 1850–1881, April 2015. (Cited on pages 150, 153, 156 and 160.)

[Miolane 2016] Nina Miolane, Susan Holmes and Xavier Pennec. Template shape estimation: correcting an asymptotic bias. working paper or preprint, July 2016. (Cited on pages 105, 107, 115, 116, 117, 135, 136, 137, 143, 144 and 205.)

[Motsak 2006] O. Motsak. *Computation of the central elements and centralizers of sets of elements in non-commutative polynomial algebras.* Thèse de sciences (phd thesis), Technische Universität Kaiserslautern, 2006. (Cited on page 60.)

[Nakahara 2003] Mikio Nakahara. Geometry, topology, and physics. Graduate student series in physics. Institute of Physics Publishing, Bristol, Philadelphia, 2003. (Cited on pages 29, 30, 31 and 32.)

[Pacard 2005] Frank Pacard. *Constant mean curvature hypersurfaces in Riemannian manifolds.* Riv. Mat. Univ. Parma , vol. 7, no. 4, pages 141–162, 2005. (Cited on page 115.)

[Patrangenaru 2015] V. Patrangenaru and L. Ellingson. Nonparametric statistics on manifolds and their applications to object data analysis. Taylor & Francis, 2015. (Cited on pages 6 and 97.)

[Pennec 1998a] X Pennec and N Ayache. *A geometric algorithm to find small but highly similar 3D substructures in proteins.* Bioinformatics, vol. 14, no. 6, pages 516–522, 1998. (Cited on page 101.)

[Pennec 1998b] Xavier Pennec. *Toward a generic framework for recognition based on uncertain geometric features.* Videre: Journal of Computer Vision Research, vol. 1, no. 2, pages 58–87, 1998. (Cited on page 100.)

[Pennec 2006] Xavier Pennec. *Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements.* Journal of Mathematical Imaging and Vision, vol. 25, no. 1, pages 127–154, 2006. (Cited on pages 7, 29, 78, 79, 81, 111, 113, 137 and 138.)

[Pennec 2012] Xavier Pennec and Vincent Arsigny. *Exponential Barycenters of the Canonical Cartan Connection and Invariant Means on Lie Groups.* Matrix Information Geometry, pages 123–168, 2012. (Cited on pages 7, 41, 43 and 64.)

[Pennec 2015] Xavier Pennec. *Barycentric Subspaces and Affine Spans in Manifolds.* In Geometric Science of Information GSI'2015, Second International Conference, volume 9389 of *Lecture Notes in Compuer Science*, pages 12–21, Palaiseau, France, October 2015. (Cited on page 7.)

[Petitot 2013] Jean Petitot. *Neurogeometry of neural functional architectures.* Chaos, Solitons & Fractals, vol. 50, no. 0, pages 75 – 92, 2013. Functionality and Dynamics in Biological Systems. (Cited on pages 6, 149 and 150.)

[Postnikov 2001] M. Postnikov. Riemannian geometry. Encyclopaedia of Mathem. Sciences. Springer, 2001. (Cited on pages 7, 29, 36, 37, 45, 46, 47, 55, 76, 84, 92, 106, 138, 142, 150, 157 and 185.)

[Rand 1986] D.W. Rand. *PASCAL programs for identification of Lie algebras: Part 1. Radical - a program to calculate the radical and nil radical of parameter-free and parameter-dependent lie algebras.* Computer Physics Communications, vol. 41, no. 1, pages 105 – 125, 1986. (Cited on pages 44 and 46.)

[Rand 1988] D. Rand, P. Winternitz and H. Zassenhaus. *On the identification of a Lie algebra given by its structure constants. I. Direct decompositions, levi decompositions, and nilradicals.* Linear Algebra and its Applications, vol. 109, no. 0, pages 197 – 246, 1988. (Cited on pages 44, 46 and 59.)

[Reininghaus 2011] Jan Reininghaus, Natallia Kotava, David Günther, Jens Kasten, Hans Hagen and Ingrid Hotz. *A Scale Space Based Persistence Measure for Critical Points in 2D Scalar Fields.* IEEE Trans. Vis. Comput. Graph., vol. 17, no. 12, pages 2045–2052, 2011. (Cited on page 119.)

[Roentgen 1896] W. C. Roentgen. *On a New Kind of Rays.* Science, vol. 3, no. 59, pages 227–231, 1896. (Cited on page 2.)

[Ronyai 1997] L. Ronyai, G. Ivanyos, A. Kuronya and W. A. de Graaf. *Computing Levi Decompositions in Lie algebras.* Applicable Algebra in Engineering, Communication and Computing, vol. 8, no. 4, pages 291–303, 1997. (Cited on page 44.)

[RStudio, Inc 2013] RStudio, Inc. *Easy web applications in R.*, 2013. (Cited on page 85.)

[Sanguinetti 2008] Gonzalo Sanguinetti, Giovanna Citti and Alessandro Sarti. *Image Completion Using a Diffusion Driven Mean Curvature Flowing A Sub-Riemannian Space.* In International on Computer Vision Theory and Applications, 3rd. Proceedings. VISAPP 2008. Funchal, Portugal., volume 2, pages 46–53. INSTICC - Institute for Systems and Technologies of Control and Communication, jan 2008. (Cited on pages 151, 154, 155 and 159.)

[Santalo 2004] Luis Antonio Santalo and Mark Kac. Integral geometry and geometric probability. Cambridge mathematical library. Cambridge University Press, Cambridge, 2004. 1re edition London : Addison-Wesley, 1976. (Cited on page 8.)

[Schur 1905] Issai Schur. Neue Begründung der Theorie der Gruppencharaktere. Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin, 1905. (Cited on page 60.)

[Singh 2013] Nikhil Singh, Jacob Hinkle, Sarang Joshi and P. Thomas Fletcher. A hierarchical geodesic model for diffeomorphic longitudinal shape analysis, pages 560–571. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. (Cited on page 5.)

[Small 1989] Christopher G. Small. *[A Survey of the Statistical Theory of Shape]: Comment.* Statist. Sci., vol. 4, no. 2, pages 105–108, 05 1989. (Cited on page 6.)

[Sommer 2014] S. Sommer, F. Lauze and M. Nielsen. *Optimization over geodesics for exact principal geodesic analysis.* Advances in Computational Mathematics, vol. 40, no. 2, pages 283–313, 2014. (Cited on page 7.)

[Sternberg 1964] Shlomo Sternberg. Lectures on differential geometry. Prentice-Hall mathematics series. Englewood Cliffs N.J. Prentice-Hall, 1964. Autre tirage : 1965. (Cited on pages 43 and 47.)

[Stigler 1981] Stephen M. Stigler. *Gauss and the Invention of Least Squares.* Ann. Statist., vol. 9, no. 3, pages 465–474, 05 1981. (Cited on page 15.)

[Sudlow 2015] Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, Bette Liu, Paul Matthews, Giok Ong, Jill Pell, Alan Silman, Alan Young, Tim Sprosen, Tim Peakman and Rory Collins. *UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age.* PLOS Medicine, 2015. (Cited on page 3.)

[Thompson 1992] D.A.W. Thompson and J.T. Bonner. On growth and form. Canto. Cambridge University Press, 1992. (Cited on pages 5 and 40.)

[Tickle 1998] Ian J. Tickle, Roman A. Laskowski and David S. Moss. *Error Estimates of Protein Structure Coordinates and Deviations from Standard Geometry by Full-Matrix Refinement of γB- and βB2-Crystallin.* Acta Crystallographica Section D, vol. 54, no. 2, pages 243–252, Mar 1998. (Cited on page 99.)

[Trouvé 1998] Alain Trouvé. *Diffeomorphisms Groups and Pattern Matching in Image Analysis.* International Journal of Computer Vision, vol. 28, no. 3, pages 213–221, 1998. (Cited on page 5.)

[Tsou 1957] S.-T. Tsou and A. G. Walker. *XIX. Metrisable Lie Groups and Algebras.* Proceedings of the Royal Society of Edinburgh. Section A. Mathematical and Physical Sciences, vol. 64, pages 290–304, 1 1957. (Cited on pages 43 and 46.)

[Tsou 1962] S-T. Tsou. *XI. On the Construction of Metrisable Lie Algebras.* Proceedings of the Royal Society of Edinburgh. Section A. Mathematical and Physical Sciences, vol. 66, pages 116–127, 1 1962. (Cited on pages 43 and 46.)

[Vaillant 2005] Marc Vaillant and Joan Glaunès. Surface matching via currents, pages 381–392. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005. (Cited on page 5.)

[Yang 2010] Le Yang. *Riemannian median and its estimation.* LMS Journal of Computation and Mathematics, vol. 13, pages 461–479, 12 2010. (Cited on page 42.)

[Younes 2012] L. Younes. Shapes and diffeomorphisms. Applied Mathematical Sciences. Springer London, Limited, 2012. (Cited on pages 101 and 107.)

[Zalawadia 2010] DA Zalawadia, DJ Vadgama, DS Ruparelia and DS Patel. *Morphometric Study Of Pterion In Dry Skull Of Gujarat Region.* NJIRM, vol. 1(4), pages 25–29, 2010. (Cited on page 133.)

[Zefran 1995] Milos Zefran, Vijay Kumar and Christopher Croke. *On the Generation of Smooth Three-Dimensional Rigid Body Motions.* IEEE Transactions on Robotics and Automation, 1995. (Cited on page 160.)

Appendix A

Proofs of the theorems of Chapter 5

We present the detailed proofs of the theorems. The precise statements of the theorems are given again in each section of this supplementary materials.

A.1 Notations

We denote Y the template shape and $\hat{Y}$ its estimate. X is a point in the manifold M. We consider that X belongs to a principal orbit and we recall that the set of principal orbits is dense in M. We write $Z = \pi(X)$ the projection of X in the shape space Q. We denote m the dimension of M, p the dimension of the principal orbits and q the dimension of the quotient space. Figure A.1 shows the elements Y, X, $O_Z = O_X$, $Z = \pi(X)$.

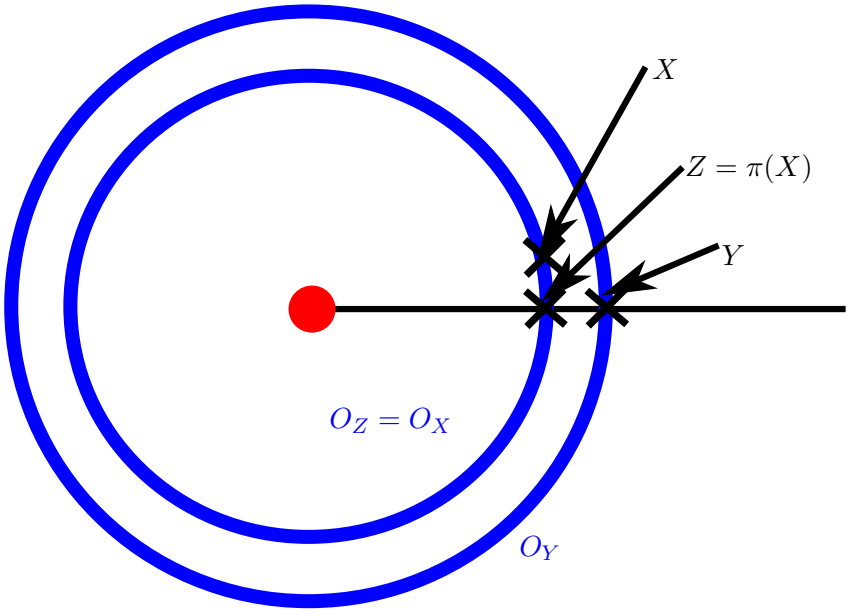


Figure A.1: Summary of the notations used in the proofs. Y is the template shape, X is a point in M, belonging to a principal orbit O_X. $Z = \pi(X)$ is the projection of X in the shape space.

Normal Coordinate Systems We often use Normal Coordinate Systems (NCS) to express the coordinates of tangent vectors. We refer to [Postnikov 2001] for theoretical developments about the NCS and to [Brewin 2009] for Taylor expansions of differential geometric tensors in a NCS.

For example, we may consider a NCS centered at the point Y, with respect to the Riemannian metric of M. This NCS is valid on an open neighborhood of Y, that is start-shaped domain around Y. Moreover, we assume that M is geodesically complete : thus, this domain is equal to the whole manifold M, with the exception of the cut locus which is of null measure [Postnikov 2001]. The Riemannian logarithm of a point X in the NCS at Y is denoted: $\overrightarrow{YX}$.

Asymptotic behavior for $\sigma \to 0$ We denote: (i) $\Theta(\sigma^k)$ a function that is proportional to σ^k, (ii) $\mathcal{O}(\sigma^k)$ a function equivalent to σ^k for $\sigma \to 0$ and (iii) $\varepsilon(\sigma)$ a function that is exponentially decreasing for $\sigma \to 0$.

A.2 Preliminaries

A.2.1 A first computation

First, we show a technical result that will be used throughout the proofs. We show that the following integral on $T_Y M$, the tangent space of M at Y:

$$\int_{\mathcal{C}_{B_r}} d_M(Y,X)^k \exp\left(-\frac{d_M^2(Y,X)}{2\sigma^2}\right) d\overrightarrow{YX}$$

is a function $\sigma \to \varepsilon(\sigma)$, i.e. decreases exponentially when $\sigma \to 0$. In the integral above, the notation $\mathcal{C}_{B_r}$ denotes the complement in M of the geodesic ball B_r of center Y and of radius r.

We split the coordinates in $T_Y M$ into (ρ, u) (polar coordinates):

$$\begin{aligned} || \int_{\mathcal{C}_{B_r}} d_M(Y,X)^k \exp\left(-\frac{d_M^2(Y,X)}{2\sigma^2}\right) dX|| &\leq K \int_{u\in S^m} \int_r^{\rho(u)} \rho^k \exp\left(-\frac{\rho^2}{2\sigma^2}\right) \rho^{m-1} d\rho dS^m \\ &\leq K \int_{u\in S^m} \int_r^{\rho(u)} \rho^{m+k-1} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) d\rho dS^m \end{aligned}$$

where $\rho(u)$ is the distance to the cutlocus in the direction u.
The positive integral is upper bounded by the same integral defined on the larger domain:

$$|| \int_{\mathcal{C}_{B_r}} d_M(Y,X)^k \exp\left(-\frac{d_M^2(Y,X)}{2\sigma^2}\right) dX|| \leq K \int_{u\in S^m} \int_r^{+\infty} \rho^{m+k-1} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) d\rho dS^m$$

Integrating the volume of the unit hypersphere:

$$|| \int_{\mathcal{C}_{B_r}} d_M(Y,X)^k \exp\left(-\frac{d_M^2(Y,X)}{2\sigma^2}\right) dX|| \leq K \int_r^{+\infty} \rho^{m+k-1} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) d\rho$$

where the right-hand-side is dominated by $\exp(-\frac{r^2}{2\sigma^2})$ by the dominated convergence theorem. Therefore:

$$\int_{\mathcal{C}_{B_r}} d_M(Y,X)^k \exp\left(-\frac{d_M^2(Y,X)}{2\sigma^2}\right) dX = \varepsilon(\sigma) \tag{A.1}$$

i.e. decreases exponentially when $\sigma \to 0$.

A.2.2 Truncated Gaussian moments in a Euclidean space

We give preliminary computations of Gaussian moments in a m-dimensional vector space $\mathbb{R}^m$, using the curved notation $\mathcal{M}$. We refer to the *(unnormalized) moment of order k of the m-dimensional Gaussian of covariance $\sigma^2 A$* as:

$$\mathcal{M}^{i_1 \dots i_k}(\sigma^2 A) = \int_{\mathbb{R}^m} X^{i_1} \dots X^{i_k} \exp\left(-\frac{X^T A^{-1} X}{2\sigma^2}\right) dX$$

and to the *truncated (unnormalized) moment at radius r of the m-dimensional Gaussian of covariance $\sigma^2 A$* as:

$$\mathcal{M}_r^{i_1 \dots i_k}(\sigma^2 A) = \int_{B_r} X^{i_1} \dots X^{i_k} \exp\left(-\frac{X^T A^{-1} X}{2\sigma^2}\right) dX$$

where the integration domain is now the m-dimensional ball B_r.

A.2.2.1 First, we recall the expressions of some unnormalized Gaussian moments.

The order 0, 2 and 4 are:

$$\boxed{\begin{aligned} \mathcal{M}^0(\sigma^2.A) &= \sigma^m \sqrt{(2\pi)^m} \sqrt{\det(A)} \\ \mathcal{M}^{ab}(\sigma^2.A) &= \sigma^{m+2} \sqrt{(2\pi)^m} \sqrt{\det(A)}.A^{ab} \\ \mathcal{M}^{abcd}(\sigma^2.A) &= \sigma^{m+4} \sqrt{(2\pi)^m} \sqrt{\det(A)}.\left(A^{ab}A^{cd} + A^{ac}A^{bd} + A^{ad}A^{bc}\right) \\ \mathcal{M}^{i_1 \dots i_k}(\sigma^2.A) &= \Theta\left(\sigma^{m+k}\right)^{i_1 \dots i_k} \quad \text{if } k \text{ even} \\ &= 0 \quad \text{if } k \text{ odd} \end{aligned}} \tag{A.2}$$

where we recall that $\Theta(\sigma^{m+k})$ denotes the proportionality to σ^{m+k}.

A.2.2.2 Second, we turn to the (unnormalized) truncated moments.

They write, with respect to the total moments:

$$\mathcal{M}_r^{i_1...i_k}(\sigma^2 A) = \mathcal{M}^{i_1...i_k}(\sigma^2 A) - \int_{\mathcal{C}_{B_r}} X^{i_1}...X^{i_k} \exp\left(-\frac{X^T A^{-1} X}{2\sigma^2}\right) dX$$

The second term of the sum is negligible when $\sigma \to 0$. To see this, we put an upper bound on its norm through triangular inequality:

$$||\int_{\mathcal{C}_{B_r}} X^{i_1}...X^{i_k} \exp\left(-\frac{X^T A^{-1} X}{2\sigma^2}\right) dX|| \leq \int_{\mathcal{C}_{B_r}} ||X^{i_1}||...||X^{i_k}|| \exp\left(-\frac{X^T A^{-1} X}{2\sigma^2}\right) dX,$$

using triangular inequality on their coordinates:

$$||\int_{\mathcal{C}_{B_r}} X^{i_1}...X^{i_k} \exp\left(-\frac{X^T A^{-1} X}{2\sigma^2}\right) dX|| \leq \int_{\mathcal{C}_{B_r}} ||X||^k \exp\left(-\frac{X^T A^{-1} X}{2\sigma^2}\right) dX,$$

and performing the change of variables $X' = A^{-1/2}X$, i.e. taking the matrix square root of the positive definite matrix A^{-1}:

$$\begin{aligned}||\int_{\mathcal{C}_{B_r}} X^{i_1}...X^{i_k} \exp\left(-\frac{X^T A^{-1} X}{2\sigma^2}\right) dX|| &\leq \int_{C_{B_{|||A^{1/2}|||r}}} |||A|||^{k/2} ||X'||^k \exp\left(-\frac{X'^T X'}{2\sigma^2}\right) \det\left(A^{1/2}\right) dX, \\ &= |||A|||^{k/2} \det\left(A^{1/2}\right) \int_{\mathcal{C}_{B_{|||A^{1/2}|||r}}} ||X'||^k \exp\left(-\frac{||X'||^2}{2\sigma^2}\right) dX.\end{aligned}$$

By the computations in Subsection A.2.1, we have:

$$||\int_{\mathcal{C}_{B_r}} X^{i_1}...X^{i_k} \exp\left(-\frac{X^T A^{-1} X}{2\sigma^2}\right) dX|| = \varepsilon(\sigma)$$

Therefore, the truncated moments are equivalent to the (full) moments for $\sigma \to 0$:

$$\boxed{\mathcal{M}_r^{i_1...i_k}(\sigma^2 A) = \mathcal{M}^{i_1...i_k}(\sigma^2 A) + \varepsilon(\sigma)} \tag{A.3}$$

A.2.3 Isotropic Gaussian Moments on a Riemannian manifolds

We turn to computations of Gaussian moments in a m-dimensional Riemannian manifold M, using the notation $\mathfrak{M}$. We refer to the *(unnormalized) moment of order k of the m-dimensional isotropic Gaussian of covariance* $\sigma^2\mathbb{I}$ as:

$$\mathfrak{M}^{i_1...i_k}\left(\sigma^2\mathbb{I}\right) = \int_M \overrightarrow{YX}^{i_1}...\overrightarrow{YX}^{i_k} \exp\left(-\frac{d_M^2(X,Y)}{2\sigma^2}\right) dM(X)$$

and to the *truncated (unnormalized) moment at radius r of the m-dimensional isotropic Gaussian of covariance* $\sigma^2\mathbb{I}$ as:

$$\mathfrak{M}_r^{i_1...i_k}\left(\sigma^2\mathbb{I}\right) = \int_{B_r} \overrightarrow{YX}^{i_1}...\overrightarrow{YX}^{i_k} \exp\left(-\frac{d_M^2(X,Y)}{2\sigma^2}\right) d\overrightarrow{YX}$$

where the integration domain is now the m-dimensional geodesic ball B_r of radius r and centered at Y.

A.2.3.1 First, we consider the truncated moments:

$$\mathfrak{M}_r^{i_1\dots i_k}\left(\sigma^2\mathbb{I}\right) = \int_{B_r} \overrightarrow{YX}^{i_1}\dots\overrightarrow{YX}^{i_k} \exp\left(-\frac{d_M^2(X,Y)}{2\sigma^2}\right) dM(X)$$

In a NCS at Y : $d_M^2(X,Y) = \overrightarrow{YX}^T\overrightarrow{YX}$:

$$\mathfrak{M}_r^{i_1\dots i_k}\left(\sigma^2\mathbb{I}\right) = \int_{B_r} \overrightarrow{YX}^{i_1}\dots\overrightarrow{YX}^{i_k} \exp\left(-\frac{\overrightarrow{YX}^T\overrightarrow{YX}}{2\sigma^2}\right) dM(X)$$

On the small ball B_r, $dM(X) = d\overrightarrow{YX} + \frac{1}{6}R_{ab}(Y)\overrightarrow{YX}^a\overrightarrow{YX}^b d\overrightarrow{YX} + \mathcal{O}(||\overrightarrow{YX}||^3)d\overrightarrow{YX}$:

$$\begin{aligned}\mathfrak{M}_r^{i_1\dots i_k}\left(\sigma^2\mathbb{I}\right) = \int_{B_r} \overrightarrow{YX}^{i_1}\dots\overrightarrow{YX}^{i_k} \exp\left(-\frac{\overrightarrow{YX}^T\overrightarrow{YX}}{2\sigma^2}\right) d\overrightarrow{YX}\\ + \frac{1}{6}\sum_{ab} R_{ab}(Y)\int_{B_r} \overrightarrow{YX}^a\overrightarrow{YX}^b.\overrightarrow{YX}^{i_1}\dots\overrightarrow{YX}^{i_k} \exp\left(-\frac{\overrightarrow{YX}^T\overrightarrow{YX}}{2\sigma^2}\right) d\overrightarrow{YX}\\ + \int_{B_r} O(||\overrightarrow{YX}||^3)\overrightarrow{YX}^{i_1}\dots\overrightarrow{YX}^{i_k} \exp\left(-\frac{\overrightarrow{YX}^T\overrightarrow{YX}}{2\sigma^2}\right) d\overrightarrow{YX}\end{aligned}$$

Now we recognize the (un-normalized) moments of a truncated isotropic Gaussian a B_r in the vector space $T_YM \simeq \mathbb{R}^m$. We replace them by the expressions given in the previous subsection:

$$\mathfrak{M}_r^{i_1\dots i_k}\left(\sigma^2\mathbb{I}\right) = \left(\mathcal{M}^{i_1\dots i_k}(\sigma^2\mathbb{I}) + \varepsilon(\sigma)\right) + \frac{1}{6}\sum_{ab} R_{ab}(Y)\left(\mathcal{M}^{abi_1\dots i_k}(\sigma^2\mathbb{I}) + \varepsilon(\sigma)\right) + \mathcal{O}(\sigma^{m+k+4})$$

The sum of two functions that decrease exponentially for $\sigma \to 0$ is a function that decreases exponentially for $\sigma \to 0$. So that:

$$\mathfrak{M}_r^{i_1\dots i_k}\left(\sigma^2\mathbb{I}\right) = \mathcal{M}^{i_1\dots i_k}(\sigma^2\mathbb{I}) + \frac{1}{6}\sum_{ab} R_{ab}(Y)\mathcal{M}^{abi_1\dots i_k}(\sigma^2\mathbb{I}) + \mathcal{O}(\sigma^{m+k+4}) + \varepsilon(\sigma)$$

We take the σ's out:

$$\boxed{\mathfrak{M}_r^{i_1\dots i_k}\left(\sigma^2\mathbb{I}\right) = \sigma^{k+m}\mathcal{M}^{i_1\dots i_k}(\mathbb{I}) + \frac{\sigma^{k+m+2}}{6}\sum_{ab} R_{ab}(Y)\mathcal{M}^{abi_1\dots i_k}(\mathbb{I}) + \mathcal{O}(\sigma^{m+k+4}) + \varepsilon(\sigma)} \tag{A.4}$$

A.2.3.2 Second, we consider the (full) moments.

They write, with respect to the non-truncated moments:

$$\mathfrak{M}^{i_1\dots i_k}\left(\sigma^2\mathbb{I}\right) = \mathfrak{M}_r^{i_1\dots i_k}\left(\sigma^2\mathbb{I}\right) + \int_{\mathcal{C}_{B_r}} X^{i_1}\dots X^{i_k} \exp\left(-\frac{d_M^2(Y,X)}{2\sigma^2}\right) dM(X)$$

The second term of the sum is negligible when $\sigma \to 0$. To see this, we put an upper bound on its norm through triangular inequality:

$$||\int_{\mathcal{C}_{B_r}} X^{i_1}\dots X^{i_k} \exp\left(-\frac{d_M^2(Y,X)}{2\sigma^2}\right) dM(X)|| \leq \int_{\mathcal{C}_{B_r}} ||X^{i_1}||\dots||X^{i_k}|| \exp\left(-\frac{d_M^2(Y,X)}{2\sigma^2}\right) dM(X)$$

then using triangular inequality on the coordinates:

$$||\int_{\mathcal{C}_{B_r}} X^{i_1}\dots X^{i_k} \exp\left(-\frac{d_M^2(Y,X)}{2\sigma^2}\right) dM(X)|| \leq \int_{\mathcal{C}_{B_r}} d_M(Y,X)^k \exp\left(-\frac{d_M^2(Y,X)}{2\sigma^2}\right) dM(X).$$

We assume that the Ricci curvature of M is bounded from below. Therefore, the measure dM has an upper bound with respect to the Lebesgue measure on the tangent space, which we write K: $dM(X) \leq KdX$:

$$|| \int_{\mathcal{C}_{B_r}} X^{i_1}...X^{i_k} \exp\left(-\frac{d_M^2(Y,X)}{2\sigma^2}\right) dM(X)|| \leq K \int_{\mathcal{C}_{B_r}} d_M(Y,X)^k \exp\left(-\frac{d_M^2(Y,X)}{2\sigma^2}\right) dX$$

By the computations in Subsection A.2.1, this inequality together with $\sigma \to 0$ shows that:

$$\int_{\mathcal{C}_{B_r}} X^{i_1}...X^{i_k} \exp\left(-\frac{d_M^2(Y,X)}{2\sigma^2}\right) dM(X) = \varepsilon(\sigma) \tag{A.5}$$

i.e. is a function that decreases exponentially when $\sigma \to 0$.

Therefore, the (unnormalized) moment of order k of the m-dimensional isotropic Gaussian of covariance $\sigma^2\mathbb{I}$ writes:

$$\boxed{\mathfrak{M}^{i_1...i_k}\left(\sigma^2\mathbb{I}\right) = \sigma^{k+m}\mathcal{M}^{i_1...i_k}(\mathbb{I}) + \frac{\sigma^{k+m+2}}{6}\sum_{ab} R_{ab}(Y)\mathcal{M}^{abi_1...i_k}(\mathbb{I}) + \mathcal{O}(\sigma^{m+k+4}) + \varepsilon(\sigma).} \tag{A.6}$$

A.3 Proof of Theorem A.3.1: Induced probability density on shapes

In this section we prove the Theorem 1 of our paper. We recall Theorem 1 below.

Theorem A.3.1 *The data X_i's are generated in the finite-dimensional Riemannian manifold M following the model: $X_i = Exp(g_i \cdot Y, \varepsilon_i), i = 1...n$, described in the paper. In this model: (i) the action of the finite dimensional Lie group G on M, denoted $\cdot$, is isometric, (ii) the parameter Y is the template shape in the shape space Q, (iii) ε_i is the noise and follows a (generalization to manifolds of a) Gaussian of variance σ^2, see Section 1 of the paper.*

Then, the probability distribution function f on the shapes of the X_i's, $i = 1...n$, in the asymptotic regime on an infinite number of data $n \to +\infty$, has the following Taylor expansion around the noise level $\sigma = 0$:

$$f(Z) = \frac{1}{(\sqrt{2\pi}\sigma)^q} \exp\left(-\frac{d_M^2(Y,Z)}{2\sigma^2}\right)\left(F_0(Z) + \sigma^2 F_2(Z) + \mathcal{O}(\sigma^4) + \varepsilon(\sigma)\right)$$

where (i) Z denotes a point in the shape space Q, (ii) F_0 and F_2 are functions of Z involving the derivatives of the Riemannian tensor at Z and the derivatives of the graph G describing the orbit O_Z at Z, and (iii) ε is a function of σ that decreases exponentially for $\sigma \to 0$.

We consider $Z \in B_r$ i.e. in the geodesic ball of center Y and radius Y. We use a NCS at Z. The notations are summarized on Figure A.2. The reader can refer to this Figure along the proof.

The generative model implies the following Riemannian Gaussian distribution on the objects:

$$f(Z) = \frac{1}{C_M(\sigma)} \exp\left(-\frac{d_M^2(X,Y)}{2\sigma^2}\right), \tag{A.7}$$

where $C_M(\sigma)$ is the integration constant:

$$C_M(\sigma) = \int_M \exp\left(-\frac{d_M^2(X,Y)}{2\sigma^2}\right) dM(X) \tag{A.8}$$

We compute the induced probability distribution f on shapes by integrating the distribution on the orbit of X out of $f(Z)$:

$$f(Z) = \frac{1}{C_M(\sigma)} \int_{O_Z} \exp\left(-\frac{d_M^2(Y,X)}{2\sigma^2}\right) dO_Z(X).$$

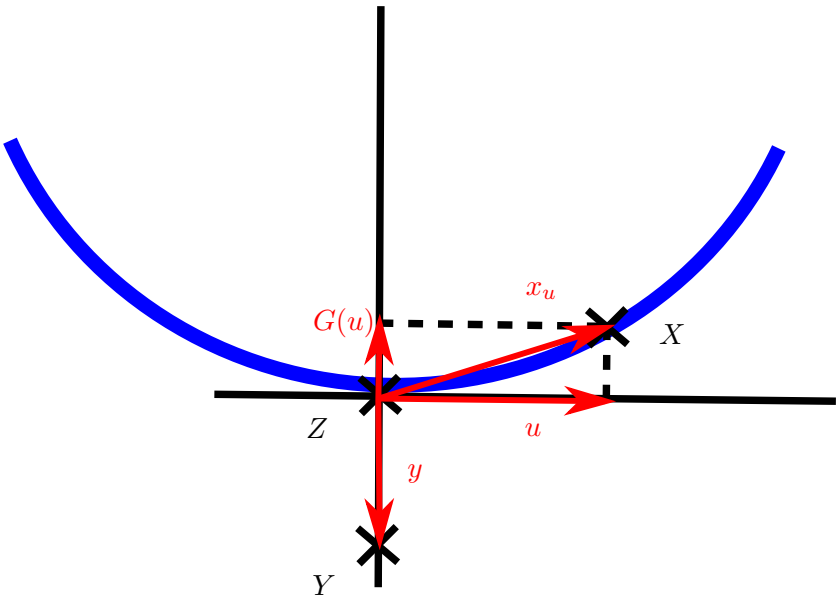


Figure A.2: Summary of the notations used in the proof of Theorem 1. The vectors of the tangent space $T_Z M$ are in red.

We start by dividing the integral:

$$f(Z) = \frac{1}{C_M(\sigma)}.\left(\int_{O_Z \cap B_r} \exp\left(-\frac{d_M^2(Y,X)}{2\sigma^2}\right) dO_Z(X) + \int_{O_Z \cap \mathcal{C}_{B_r}} \exp\left(-\frac{d_M^2(Y,X)}{2\sigma^2}\right) dO_Z(X)\right)$$

In the parenthesis, we call the first integral $I_Z(\sigma)$ and the second term is an $\eta(\sigma)$:

$$f(Z) = C_M(\sigma)^{-1}.(I_Z(\sigma) + \eta(\sigma))$$

We compute the Taylor expansions of $C_M(\sigma)^{-1}$ and $I_Z(\sigma)$ for $\sigma \to 0$ and show that $\eta(\sigma) = \varepsilon(\sigma)$.

A.3.1 Taylor expansion of $C_M(\sigma)^{-1}$

The integration constant $C_M(\sigma)$ is the moment of order 0 of the m-dimensional isotropic Gaussian of variance $\sigma^2\mathbb{I}$ in the m-dimensional manifold M. From Subsection A.2.3:

$$\begin{aligned} C_M(\sigma) &= (\sqrt{2\pi}\sigma)^m + \frac{\sigma^{m+2}}{6}\sum_{ab} R_{ab}(Y)(\sqrt{2\pi})^m \delta_{ab} + \mathcal{O}(\sigma^{m+3}) + \varepsilon(\sigma) \\ &= (\sqrt{2\pi}\sigma)^m + \frac{\sigma^{m+2}}{6} R(Y)(\sqrt{2\pi})^m + \mathcal{O}(\sigma^{m+3}) + \varepsilon(\sigma) \end{aligned}$$

where $\varepsilon(\sigma)$ is exponentially decreasing wrt σ and $R(Y)$ is the scalar curvature of M at Y. Its inverse:

$$\begin{aligned} C_M(\sigma)^{-1} &= \left((\sqrt{2\pi}\sigma)^m + \frac{(\sqrt{2\pi})^m}{6} R(Y)\sigma^{m+2} + \mathcal{O}(\sigma^{m+3}) + \varepsilon(\sigma)\right)^{-1} \\ &= (\sqrt{2\pi}\sigma)^{-m}.\left(1 + \frac{\sigma^2}{6}R(Y) + O(\sigma^3) + \varepsilon(\sigma)\right)^{-1} \\ &= (\sqrt{2\pi}\sigma)^{-m}.\left(1 - \frac{\sigma^2}{6}R(Y) + O(\sigma^3) + \varepsilon(\sigma)\right) \end{aligned}$$

A.3.2 Taylor expansion of $I_Z(\sigma)$

Now we compute the integral $I_Z(\sigma)$ involved in the formula of $f(Z)$:

$$I_Z(\sigma) = \int_{O_Z \cap B_r} \exp\left(-\frac{d_M^2(Y,X)}{2\sigma^2}\right) dO(X)$$

We denote: $B_{r_Z} = B_{r_Z}$. We first perform the computations for any σ. We recall that r is fixed, and small enough in a sense made precise later.

A.3.2.1 Computing $d_M^2(Y,X)$

The first component of $I_Z(\sigma)$ is $d_M^2(Y,X)$. We express the Taylor expansion of d_M^2 for $X \in O_Z$ close to Z.

First, we parameterize the point X. The points X, $Z = \pi(X)$ and the orbit $O_Z \subset M$ are illustrated on Figure A.1: the orbit $O_Z = O_X$ is the blue circle in $M = \mathbb{R}^2$ going through X and Z. The orbit O_Z can be seen in the tangent space T_ZM through the Logarithm map at Z. For r small enough, i.e. for X close enough to Z, we can locally represent O_Z in T_ZM as the graph of a smooth function G from T_ZO to N_ZO, using the vector $u \in T_ZO_Z$ around $u = 0$:

$$I : T_ZO_Z \mapsto T_ZM = T_ZO_Z \oplus N_ZO_Z$$
$$u \mapsto x_u = (u, G(u))$$

The local graph $u \to G(u)$ is illustrated on Figure A.1 and has the following Taylor expansion around $u = 0$ in the NCS at Z:

$$G(u)^a = \frac{1}{2}h^a_{bc}(Z)u^bu^c + G_3(Z)^a_{bcd}u^bu^cu^d + G_4(Z)^a_{bcde}u^bu^cu^du^e + O(||u||^5)$$

The 0-th and 1-th order derivatives of $u \to G(u)$ are zero because the graph goes through Z and is tangent at T_ZO_Z. The second order derivative is by definition the second fundamental form $h(Z)$ introduced in Section 2 of the paper: $h(Z)$ represents the best quadratic approximation of the graph G. The third and fourth orders $G_3(Z)$ and $G_4(Z)$ are further refinements on the shape of the graph G around Z.

Second, we compute the Taylor expansion of $d_M^2(Y,X)$ with respect to u:

$$\boxed{d_M^2(Y,X) = d_M^2(Y,Z) + u^aL_a(Z) + u^au^bM_{ab}(Z) + u^au^bu^cP_{abc}(Z) + u^au^bu^cu^dS_{abcd}(Z) + \mathcal{O}(||u||^5)}$$

and our goal is to compute the different tensors.

Y and X are represented by their Riemannian Logarithms at Z: $y = \mathrm{Log}_Z Y$, and $x_u = \mathrm{Log}_Z X$. We also recall: $u^Ty = 0$. Using these, we express the squared distance $d_M^2(Y,X)$ in the NCS at Z. We use the formula p.23 in

[Brewin 2009] with the notations $\Delta x \leftarrow y - x_u$ and $x \leftarrow x_u$:

$$
\begin{aligned}
d_M^2 &= d_M^2(\mathrm{Exp}_Z(x(u)), \mathrm{Exp}_Z y) \\
&= \delta_{ab}(y-x_u)^a(y-x_u)^b \\
&- \frac{60}{180} R_{cadb} x_u^c x_u^d (y-x_u)^a (y-x_u)^b \\
&- \frac{15}{180} x_u^d x_u^e \nabla_a R_{dbec} (y-x_u)^a (y-x_u)^b (y-x_u)^c \\
&- \frac{3}{540} x_u^e x_u^f (44 R_{gaeb} R_{gcfd} + 3\nabla_{ab} R_{ecfd}) (y-x_u)^a (y-x_u)^b (y-x_u)^c (y-x_u)^d \\
&+ \frac{1}{54} x_u^f x_u^g R_{hafb} \nabla_c R_{hdge} (y-x_u)^a (y-x_u)^b (y-x_u)^c (y-x_u)^d (y-x_u)^e \\
&- \frac{30}{180} x_u^c x_u^d x_u^e \nabla_c R_{daeb} \\
&+ \frac{1}{180} x_u^d x_u^e x_u^f (8 R_{gdea} R_{gbfc} - 9\nabla_{da} R_{ebfc}) (y-x_u)^a (y-x_u)^b (y-x_u)^c \\
&- \frac{5}{540} x_u^e x_u^f x_u^g (8 R_{haeb} \nabla_c R_{hfgd} + 9 R_{haeb} \nabla_h R_{fcgd} \\
&\qquad + 20 R_{haeb} \nabla_f R_{hcgd} - 6 R_{hefa} \nabla_b R_{hcgd}) (y-x_u)^a (y-x_u)^b (y-x_u)^c (y-x_u)^d \\
&+ \frac{1}{180} x_u^c x_u^d x_u^e x_u^f (8 R_{gcda} R_{gefb} - 9\nabla_{cd} R_{eafb}) (y-x_u)^a (y-x_u)^b \\
&+ \frac{1}{180} x_u^d x_u^e x_u^f x_u^g (4 R_{hadb} \nabla_e R_{hfgc} + 4 R_{hdea} \nabla_b R_{hfgc} + 4 R_{hdea} \nabla_f R_{hbgc} \\
&\qquad + 3\nabla_{dea} R_{fbgc}) (y-x_u)^a (y-x_u)^b (y-x_u)^c \\
&+ \mathcal{O}(||x_u||^5)
\end{aligned}
$$

We express this in orders of x_u:

$$
\begin{aligned}
d_M^2 = {} & \delta_{ab} y^a y^b \\
& - 2\delta_{ab} y^a x_u^b \\
& - \frac{1}{3} R_{cadb} x_u^c x_u^d y^a y^d \\
& + \frac{1}{12} x_u^d x_u^e \nabla_a R_{dbec} y^a y^b y^c \\
& - \frac{1}{180} x_u^e x_u^f (44 R_{gaeb} R_{gcfd} + 3 \nabla_{ab} R_{ecfd}) y^a y^b y^c y^d \\
& - \frac{1}{54} x_u^f x_u^g R_{hafb} \nabla_c R_{hdge} y^a y^b y^c y^d y^e \\
& + \frac{2}{3} x_u^c x_u^d R_{cadb} x_u^a y^b \\
& - \frac{1}{12} x_u^d x_u^e \nabla_a R_{dbec} (x_u^a y^b y^c + 2 y^a y^b x_u^c) \\
& - \frac{1}{180} x_u^e x_u^f (44 R_{gaeb} R_{gcfd} + 3 \nabla_{ab} R_{ecfd}) (2 x_u^a y^b y^c y^d + 2 y^a y^b x_u^c y^d) \\
& + \frac{1}{54} x_u^f x_u^g R_{hafb} \nabla_c R_{hdge} (4 x_u^a y^b y^c y^d y^e + y^a y^b x_u^c y^d y^e) \\
& - \frac{1}{3} x_u^c x_u^d x_u^e \nabla_c R_{daeb} y^a y^b \\
& \frac{1}{180} x_u^d x_u^e x_u^f (8 R_{gdea} R_{gbfc} - 9 \nabla_{da} R_{ebfc}) y^a y^b y^c \\
& - \frac{1}{180} x_u^e x_u^f x_u^g (8 R_{haeb} \nabla_c R_{hfgd} + 9 R_{haeb} \nabla_h R_{fcgd} + 20 R_{haeb} \nabla_f R_{hcgd} - 6 R_{hefa} \nabla_b R_{hcgd}) y^a y^b y^c y^d \\
& - \frac{1}{3} R_{cadb} x_u^c x_u^d x_u^a x_u^b \\
& - \frac{1}{12} x_u^d x_u^e \nabla_a R_{dbec} (2 x_u^a x_u^b y^c + y^a x_u^b x_u^c) \\
& - \frac{1}{180} x_u^e x_u^f (44 R_{gaeb} R_{gcfd} + 3 \nabla_{ab} R_{ecfd}) (2 x_u^a x_u^b y^c y^d + 2 y^a x_u^b x_u^c y^d + 2 y^a y^b x_u^c x_u^d + 4 x_u^a y^b x_u^c y^d) \\
& + \frac{1}{54} x_u^f x_u^g R_{hafb} \nabla_c R_{hdge} (4 x_u^a x_u^b y^c y^d y^e + 4 y^a x_u^b x_u^c y^d y^e) \\
& + \frac{1}{180} x_u^d x_u^f x_u^g (8 R_{gdea} R_{gbfc} - 9 \nabla_{da} R_{ebfc}) (x_u^a y^b y^c + 2 y^a x_u^b x_u^c) \\
& - \frac{1}{180} x_u^e x_u^f x_u^g (8 R_{haeb} \nabla_c R_{hfgd} + 9 R_{haeb} \nabla_h R_{fcgd} + 20 R_{haeb} \nabla_f R_{hcgd} - 6 R_{hefa} \nabla_b R_{hcgd}) (x_u^a y^b y^c y^d + y^a y^b x_u^c y^d) \\
& + \frac{1}{180} x_u^c x_u^d x_u^e x_u^f (8 R_{gdea} R_{gbfc} - 9 \nabla_{da} R_{ebfc}) y^a y^b \\
& + \frac{1}{180} (4 R_{hadb} \nabla_e R_{hfgc} + 4 R_{hdea} \nabla_b R_{hfgc} + 4 R_{hdea} \nabla_f R_{hbgc} - 3 \nabla_{dea} R_{fbgc}) y^a y^b y^c \\
& + \mathcal{O}(||x_u||^5)
\end{aligned}
$$

We replace x_u^a by $x_u^a = u^a + G(u)^a$ where the expression of $G(u)$ is a $\mathcal{O}(||u||^2)$, so that:

$$
\begin{aligned}
d_M^2 = {} & d_M^2(Y,Z) - 2\delta_{ab}y^a u^b \\
& - 2\delta_{ab}y^a G(u)^b - \frac{1}{3}R_{cadb}u^c u^d y^a y^d + \frac{1}{12}u^d u^e \nabla_a R_{dbec} y^a y^b y^c - \frac{1}{180}u^e u^f (44R_{gaeb}R_{gcfd} + 3\nabla_{ab}R_{ecfd}) y^a y^b y^c y^d \\
& - \frac{1}{54}u^f u^g R_{hafb}\nabla_c R_{hdge} y^a y^b y^c y^d y^e \\
& - \frac{2}{3}R_{cadb}u^c G(u)^d y^a y^d + \frac{2}{12}u^d G(u)^e \nabla_a R_{dbec} y^a y^b y^c - \frac{2}{180}u^e G(u)^f (44R_{gaeb}R_{gcfd} + 3\nabla_{ab}R_{ecfd}) y^a y^b y^c y^d \\
& - \frac{2}{54}u^f G(u)^g R_{hafb}\nabla_c R_{hdge} y^a y^b y^c y^d y^e + \frac{2}{3}u^c u^d R_{cadb}u^a y^b - \frac{1}{12}u^d u^e \nabla_a R_{dbec}(u^a y^b y^c + 2y^a y^b u^c) \\
& - \frac{1}{180}u^e u^f (44R_{gaeb}R_{gcfd} + 3\nabla_{ab}R_{ecfd})(2u^a y^b y^c y^d + 2y^a y^b u^c y^d) \\
& + \frac{1}{54}u^f u^g R_{hafb}\nabla_c R_{hdge}(4u^a y^b y^c y^d y^e + y^a y^b u^c y^d y^e) \\
& - \frac{1}{3}u^c u^d u^e \nabla_c R_{daeb} y^a y^b + \frac{1}{180}u^d u^e u^f (8R_{gdea}R_{gbfc} - 9\nabla_{da}R_{ebfc}) y^a y^b y^c \\
& - \frac{1}{180}u^e u^f u^g (8R_{haeb}\nabla_c R_{hfgd} + 9R_{haeb}\nabla_h R_{fcgd} + 20R_{haeb}\nabla_f R_{hcgd} - 6R_{hefa}\nabla_b R_{hcgd}) y^a y^b y^c y^d \\
& - \frac{1}{3}R_{cadb}G(u)^c G(u)^d y^a y^d + \frac{1}{12}G(u)^d G(u)^e \nabla_a R_{dbec} y^a y^b y^c \\
& - \frac{1}{180}G(u)^e G(u)^f (44R_{gaeb}R_{gcfd} + 3\nabla_{ab}R_{ecfd}) y^a y^b y^c y^d \\
& - \frac{1}{54}G(u)^f G(u)^g R_{hafb}\nabla_c R_{hdge} y^a y^b y^c y^d y^e + \frac{2}{3}u^c u^d R_{cadb}G(u)^a y^b \\
& - \frac{1}{12}u^d u^e \nabla_a R_{dbec}(G(u)^a y^b y^c + 2y^a y^b G(u)^c) \\
& - \frac{1}{180}u^e u^f (44R_{gaeb}R_{gcfd} + 3\nabla_{ab}R_{ecfd})(2G(u)^a y^b y^c y^d + 2y^a y^b G(u)^c y^d) \\
& + \frac{1}{54}u^f u^g R_{hafb}\nabla_c R_{hdge}(4G(u)^a y^b y^c y^d y^e + y^a y^b G(u)^c y^d y^e) \\
& - \frac{1}{3}u^c u^d G(u)^e \nabla_c R_{daeb} y^a y^b + \frac{1}{180}u^d u^e G(u)^f (8R_{gdea}R_{gbfc} - 9\nabla_{da}R_{ebfc}) y^a y^b y^c \\
& - \frac{1}{180}u^e u^f G(u)^g (8R_{haeb}\nabla_c R_{hfgd} + 9R_{haeb}\nabla_h R_{fcgd} + 20R_{haeb}\nabla_f R_{hcgd} - 6R_{hefa}\nabla_b R_{hcgd}) y^a y^b y^c y^d \\
& - \frac{1}{3}R_{cadb}u^c u^d u^a u^b \\
& - \frac{1}{12}u^d u^e \nabla_a R_{dbec}(2u^a u^b y^c + y^a u^b u^c) \\
& - \frac{1}{180}u^e u^f (44R_{gaeb}R_{gcfd} + 3\nabla_{ab}R_{ecfd})(2u^a u^b y^c y^d + 2y^a u^b u^c y^d + 2y^a y^b u^c u^d + 4u^a y^b u^c y^d) \\
& + \frac{1}{54}u^f u^g R_{hafb}\nabla_c R_{hdge}(4u^a u^b y^c y^d y^e + 4y^a u^b u^c y^d y^e) \\
& + \frac{1}{180}u^d u^f u^g (8R_{gdea}R_{gbfc} - 9\nabla_{da}R_{ebfc})(u^a y^b y^c + 2y^a u^b u^c) \\
& - \frac{1}{180}u^e u^f u^g (8R_{haeb}\nabla_c R_{hfgd} + 9R_{haeb}\nabla_h R_{fcgd} + 20R_{haeb}\nabla_f R_{hcgd} - 6R_{hefa}\nabla_b R_{hcgd})(u^a y^b y^c y^d + y^a y^b u^c y^d) \\
& + \frac{1}{180}u^c u^d u^e u^f (8R_{gdea}R_{gbfc} - 9\nabla_{da}R_{ebfc}) y^a y^b \\
& + \frac{1}{180}(4R_{hadb}\nabla_e R_{hfgc} + 4R_{hdea}\nabla_b R_{hfgc} + 4R_{hdea}\nabla_f R_{hbgc} - 3\nabla_{dea}R_{fbgc}) y^a y^b y^c \\
& + \mathcal{O}(||u||^5)
\end{aligned}
$$

We replace $G(u)$ by its Taylor expansion. Identifying the tensors gives:

$$\boxed{\begin{aligned} L_a(Z) &= 0 \\ M_{ab}(Z) &= -y^c h_{ab}(Z)^c - \frac{1}{3}R_{acbd}(Z)y^c y^d + \frac{1}{12}\nabla_d R_{aebc}(Z)y^d y^e y^c \\ &\quad - \frac{1}{180}(44R_{geaf}(Z)R_{gcbd}(Z) + 3\nabla_{ef}R_{acbd}(Z))y^e y^f y^c y^d \\ &\quad - \frac{1}{54}R_{hfag}(Z)\nabla_c R_{hdbe}(Z)y^f y^g y^c y^d y^e \end{aligned}} \tag{A.9}$$

and $P_{abc}(Z)$ and $S_{abcd}(Z)$ are tensors mixing the derivatives of the graph G and the derivatives of the Riemannian curvature R of M at Z.

A.3.2.2 Computing $dO_Z(X)$

The second component of $I_Z(\sigma)$ is the measure of the orbit $dO_Z(X)$. We seek the Taylor expansion of the measure:

$$\boxed{dO_Z(X) = \left(1 + T_c(Z)u^c - N_{cb}(Z)u^b u^c + \mathcal{O}(||u||^3)\right) du} \tag{A.10}$$

and our goal is, again, to express the tensors $T_c(Z)$ and $N_{cb}(Z)$.

The measure $dO_Z(X)$ is the restriction of the measure $dM(X)$:

$$dO_Z(X) = dM(X)|_{T_X O_Z}$$

and we know that for X close enough to Z: $dM(X) = dx_u - \frac{1}{6}\mathrm{Ric}_{ab}(Z)x_u^a x_u^b dx_u$. Thus:

$$dO_Z(X) = \left(1 - \frac{1}{6}\mathrm{Ric}_{ab}(Z)x_u^a x_u^b + \mathcal{O}(||x_u||^3)\right) dx_u$$

We replace x_u^a by $x_u^a = u^a + \frac{1}{2}h_{bc}^a(Z)u^b u^c + O(||u||^4)$

$$dO_Z(X) = \left(1 - \frac{1}{6}\mathrm{Ric}_{ab}(Z)u^a u^b + \mathcal{O}(||u||^3)\right) dx_u$$

We express dx_u with respect to du:

$$\begin{aligned} dx_u &= \det\left(\frac{dx_u^a}{du^b}\right) du \\ &= \det\left(\frac{d\left(u^a + \frac{1}{2}h_{bc}^a(Z)u^b u^c + O(||u||^4)\right)}{du^b}\right) du \\ &= \det\left(\delta_b^a + \frac{1}{2}h_{cb}^a(Z)u^c + \frac{1}{2}h_{bc}^a(Z)u^c + \mathcal{O}(||u||^3)\right) du \\ &= \det\left(\delta_b^a + h_{cb}^a(Z)u^c + \mathcal{O}(||u||^3)\right) du \end{aligned}$$

Developing the determinant:

$$dx_u = \left(1 + h_{ca}^a(Z)u^c + \frac{1}{2}\left((h_{ca}^a(Z)u^c)^2 - h_{cb}^a(Z)u^c h_{da}^b(Z)u^d\right) + \mathcal{O}(||u||^3)\right) du$$

Plugging in $dO_Z(X)$:

$$\begin{aligned} dO_Z(X) &= \left(1 - \frac{1}{6}\mathrm{Ric}_{ab}(Z)u^a u^b + \mathcal{O}(||u||^3)\right)(1 + h_{ca}^a(Z)u^c) \\ &\quad + \left(1 - \frac{1}{6}\mathrm{Ric}_{ab}(Z)u^a u^b + \mathcal{O}(||u||^3)\right)\left(\frac{1}{2}\left((h_{ca}^a(Z)u^c)^2 - h_{cb}^a(Z)u^c h_{da}^b(Z)u^d\right)\right) \\ &\quad + \left(1 - \frac{1}{6}\mathrm{Ric}_{ab}(Z)u^a u^b + \mathcal{O}(||u||^3)\right)\left(\mathcal{O}(||u||^3)\right) du \end{aligned}$$

We develop by keeping up to the quadratic terms only:

$$dO_Z(X) = \left(1 - \frac{1}{6}\mathrm{Ric}_{ab}(Z)u^a u^b + h^a_{ca}(Z)u^c + \frac{1}{2}\left((h^a_{ca}(Z)u^c)^2 - h^a_{cb}(Z)u^c h^b_{da}(Z)u^d\right) + \mathcal{O}(||u||^3)\right) du$$

We reorganize the terms, relabeling the mute labels:

$$dO_Z(X) = \left(1 + h^a_{ca}(Z)u^c - \frac{1}{6}\mathrm{Ric}_{cb}(Z)u^c u^b + \frac{1}{2}\left(h^a_{ca}(Z)u^c . h^a_{ca}(Z)u^b - h^a_{cd}(Z)u^c h^d_{ba}(Z)u^b\right) + \mathcal{O}(||u||^3)\right) du$$

Now we can factorize the quadratic terms:

$$dO_Z(X) = \left(1 + h^a_{ca}(Z)u^c - u^b u^c\left(\frac{1}{6}\mathrm{Ric}_{cb}(Z) + \frac{1}{2}h^a_{ca}(Z)h^a_{ca}(Z) - \frac{1}{2}h^a_{cd}(Z)h^d_{ba}(Z)\right) + \mathcal{O}(||u||^3)\right) du$$

So that we find the expressions of the tensors:

$$\boxed{\begin{aligned} T_c(Z) &= h^a_{ca}(Z) \\ N_{cb}(Z) &= \frac{1}{6}\mathrm{Ric}_{cb}(Z) + \frac{1}{2}h^a_{ca}(Z)h^a_{ca}(Z) - \frac{1}{2}h^a_{cd}(Z)h^d_{ba}(Z) \end{aligned}} \tag{A.11}$$

A.3.2.3 Gathering to compute $I_Z(\sigma)$

We plug the expressions of $d^2_M(Y,X)$ and $dO_Z(X)$, computed in the previous subsections, in the expression of $I_Z(\sigma)$:

$$I_Z(\sigma) = \int_{B_{r_Z}} \exp\left(-\frac{d^2_M(Y,X)}{2\sigma^2}\right) dO_Z(X)$$

Plugging the squared distance $d^2_M(Y,X)$ first:

$$I_Z(\sigma) = \int_{B_{r_Z}} \exp\left(-\frac{d^2_M(Y,Z) + u^a u^b M_{ab}(Z) + u^a u^b u^c P_{abc}(Z) + u^a u^b u^c u^d S_{abcd}(Z) + \mathcal{O}(||u||^5)}{2\sigma^2}\right) dO_Z(X)$$

We split the exponential and extract the part of the exponential that does not depend on u:

$$I_Z(\sigma) = \exp\left(-\frac{d^2_M(Y,Z)}{2\sigma^2}\right)\int_{B_{r_Z}} \exp\left(-\frac{u^a u^b M_{ab}(Z)}{2\sigma^2}\right)\exp\left(-\frac{u^a u^b u^c P_{abc}(Z) + u^a u^b u^c u^d S_{abcd}(Z) + \mathcal{O}(||u||^5)}{2\sigma^2}\right) dO_Z(X$$

We perform the Taylor expansion of the term with $\mathcal{O}(||u||^4)$, recalling that at this point, σ can still be anything.

$$I_Z(\sigma) = \exp\left(-\frac{d^2_M(Y,Z)}{2\sigma^2}\right)\int_{B_{r_Z}} \exp\left(-\frac{u^a u^b M_{ab}(Z)}{2\sigma^2}\right)\left(1 - \frac{u^a u^b u^c P_{abc}(Z) + u^a u^b u^c u^d S_{abcd}(Z) + \mathcal{O}(||u||^5)}{2\sigma^2}\right) dO_Z(X)$$

Now we plug the $dO_Z(X)$:

$$I_Z(\sigma) = \exp\left(-\frac{d^2_M(Y,Z)}{2\sigma^2}\right)\int_{B_{r_Z}} \exp\left(-\frac{u^a u^b M_{ab}(Z)}{2\sigma^2}\right)\left(1 - \frac{u^a u^b u^c P_{abc}(Z) + u^a u^b u^c u^d S_{abcd}(Z)}{2\sigma^2} + \frac{\mathcal{O}(||u||^5)}{2\sigma^2}\right).$$
$$\left(1 + T_c u^c(Z) - N_{cb}(Z)u^b u^c + \mathcal{O}(||u||^3)\right) du$$

We develop the product of the parenthesis on the right:

$$I_Z(\sigma) = \exp\left(-\frac{d^2_M(Y,Z)}{2\sigma^2}\right)\int_{B_{r_Z}} \exp\left(-\frac{u^a u^b M_{ab}(Z)}{2\sigma^2}\right)\left(1 + T_c u^c(Z) - N_{cb}(Z)u^b u^c + \frac{u^a u^b u^c P_{abc}(Z)}{2\sigma^2}\right) du$$
$$+ \exp\left(-\frac{d^2_M(Y,Z)}{2\sigma^2}\right)\int_{B_{r_Z}} \exp\left(-\frac{u^a u^b M_{ab}(Z)}{2\sigma^2}\right)\left(\mathcal{O}(||u||^4) - \frac{u^a u^b u^c u^d S_{abcd}(Z)}{2\sigma^2} + \frac{\mathcal{O}(||u||^5)}{2\sigma^2}\right) du$$

By skew symmetry, the terms in $T_c u^c(Z)$ and $u^a u^b u^c P_{abc}(Z)$ integrate to 0. Moreover, $\mathcal{O}(||u||^3)$, $\mathcal{O}(||u||^5)$ become $\mathcal{O}(||u||^4)$, $\mathcal{O}(||u||^6)$:

$$I_Z(\sigma) = \exp\left(-\frac{d_M^2(Y,Z)}{2\sigma^2}\right) \int_{B_{r_Z}} \exp\left(-\frac{u^a u^b M_{ab}(Z)}{2\sigma^2}\right) \left(1 - N_{cb}(Z)u^b u^c + \mathcal{O}(||u||^4) - \frac{u^a u^b u^c u^d S_{abcd}(Z)}{2\sigma^2} + \frac{\mathcal{O}(||u||^6)}{2\sigma^2}\right) du$$

We recognize the unnormalized truncated Gaussian moments of in $\mathbb{R}^p$, where p is the dimension of the orbit O_Z, see Subsection A.2.2:

$$I_Z(\sigma) = \exp\left(-\frac{d_M^2(Y,Z)}{2\sigma^2}\right) . \left(\mathcal{M}_{r_Z}(\sigma^2 M^{-1}) - N_{cb}(Z)\mathcal{M}_{r_Z}^{bc}(\sigma^2 M^{-1}) + \mathcal{O}(\sigma^{p+4}) - \frac{S_{abcd}(Z)}{2\sigma^2}\mathcal{M}_{r_Z}^{abcd}(\sigma^2 M^{-1}) + \frac{\mathcal{O}(\sigma^{p+6})}{2\sigma^2}\right)$$

We express them in terms of the unnormalized Gaussian moments in $\mathbb{R}^p$, see Subsection A.2.2:

$$I_Z(\sigma) = \exp\left(-\frac{d_M^2(Y,Z)}{2\sigma^2}\right) . . \left(\sigma^p \mathcal{M}(M^{-1}) + N_{cb}(Z)\sigma^{p+2}\mathcal{M}(M^{-1})^{bc} - \frac{S_{abcd}}{2\sigma^2}\sigma^{p+4}\mathcal{M}(M^{-1})^{abcd} + \mathcal{O}(\sigma^{p+4}) + \frac{\mathcal{O}(\sigma^{p+6})}{\sigma^2} + \varepsilon(\sigma)\right)$$

We simplify the σ's:

$$I_Z(\sigma) = \exp\left(-\frac{d_M^2(Y,Z)}{2\sigma^2}\right) . . \left(\sigma^p \mathcal{M}(M^{-1})^0 + \sigma^{p+2}\left(N_{cb}\mathcal{M}(M^{-1})^{bc} - \frac{S_{abcd}(Z)}{2}\mathcal{M}(M^{-1})^{abcd}\right) + \mathcal{O}(\sigma^{p+4}) + \varepsilon(\sigma)\right)$$

For convenience in the later computation, we define the notations:

$$I_Z(\sigma) = \exp\left(-\frac{d_M^2(Y,Z)}{2\sigma^2}\right)\left(\sigma^p m_0(Z) + \sigma^{p+2} m_2(Z) + \mathcal{O}(\sigma^4) + \varepsilon(\sigma)\right)$$

where:

$$\boxed{\begin{aligned} m_0(Z) &= \mathcal{M}(M)^0 = \sqrt{(2\pi)^p}\sqrt{\det\left((M_{ab}(Z)^{-1}\right)} \\ m_2(Z) &= N_{cb}(Z)\mathcal{M}(M^{-1}(Z))^{bc} - \frac{1}{2}S_{abcd}(Z)\mathcal{M}(M^{-1}(Z))^{abcd} \end{aligned}} \tag{A.12}$$

where M_{ab} and S_{abcd} are given in the previous subsections.

A.3.3 Upper bound on $\eta(\sigma)$

We proceed with an adaptation of the method in Subsection A.2.1:

$$\eta(\sigma) = \int_{O_Z \cap \mathcal{C}_{B_r}} \exp\left(-\frac{d_M^2(Y,X)}{2\sigma^2}\right) dO(X)$$

Assuming that the Ricci curvature of the orbit is bounded by below, the measure of the orbit is bounded by above, by a constant that we write K_O:

$$||\eta(\sigma)|| \leq K_O \int_{O_Z \cap \mathcal{C}_{B_r}} \exp\left(-\frac{d_M^2(Y,X)}{2\sigma^2}\right) du$$

We integrate on the orbit by filiating it with hyperspheres of radii ρ:

$$\begin{aligned}||\eta(\sigma)|| &\leq K_O \int_r^{+\infty} \int_{S_\rho \cap O_Z} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) d(S_\rho \cap O_Z) d\rho \\ &= K_O \int_r^{+\infty} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) \mathrm{Vol}(S_\rho \cap O_Z) d\rho\end{aligned}$$

Now the volume of $S_\rho \cap O_Z$ is polynomial in ρ. We denote P this polynomial:

$$||\eta(\sigma)|| \leq K_O \int_r^{+\infty} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) P(\rho) d\rho$$

By dominated convergence theorem, the right-hand-side is dominated by $\exp(-\frac{r^2}{2\sigma^2})$. Thus:

$$\eta(\sigma) = \varepsilon(\sigma) \tag{A.13}$$

i.e. $\sigma \to \eta(\sigma)$ is exponentially decreasing when $\sigma \to 0$.

A.3.4 Final result: Taylor expansion of $f(Z)$

Replacing the terms:

$$f(Z) = \frac{\left(1 - \frac{\sigma^2}{6} R(Y) + O(\sigma^3) + \varepsilon(\sigma)\right)}{(\sqrt{2\pi}\sigma)^m} \left(\exp\left(-\frac{d_M^2(Y,Z)}{2\sigma^2}\right) \left(\sigma^p m_0(Z) + \sigma^{p+2} m_2(Z) + \mathcal{O}(\sigma^{p+4}) + \varepsilon(\sigma)\right) + \varepsilon(\sigma)\right)$$

We put the ε inside the main parenthesis:

$$\begin{aligned}f(Z) = &\frac{\left(1 - \frac{\sigma^2}{6} R(Y) + O(\sigma^3) + \varepsilon(\sigma)\right)}{(\sqrt{2\pi}\sigma)^m} \cdot \\ &\cdot \exp\left(-\frac{d_M^2(Y,Z)}{2\sigma^2}\right) \left(\sigma^p m_0(Z) + \sigma^{p+2} m_2(Z) + \mathcal{O}(\sigma^{p+4}) + \varepsilon(\sigma) + \exp\left(+\frac{d_M^2(Y,Z)}{2\sigma^2}\right) \varepsilon(\sigma)\right)\end{aligned}$$

We recall that $Z \in B_r$ so that:

$$f(Z) = \frac{\left(1 - \frac{\sigma^2}{6} R(Y) + O(\sigma^3)\right)}{(\sqrt{2\pi}\sigma)^m} \cdot \exp\left(-\frac{d_M^2(Y,Z)}{2\sigma^2}\right) \left(\sigma^p m_0(Z) + \sigma^{p+2} m_2(Z) + \mathcal{O}(\sigma^{p+4}) + \varepsilon(\sigma)\right)$$

We define:

$$f_Q(Z) = \frac{\exp\left(-\frac{d_M^2(Y,Z)}{2\sigma^2}\right)}{(\sqrt{2\pi}\sigma)^q}$$

and put it in the front, remembering that $m = p + q$:

$$f(Z) = f_Q(Z) \frac{\left(1 - \frac{\sigma^2}{6} R(Y) + O(\sigma^3)\right)}{(\sqrt{2\pi}\sigma)^p} \cdot \left(\sigma^p m_0(Z) + \sigma^{p+2} m_2(Z) + \mathcal{O}(\sigma^{p+4}) + \varepsilon(\sigma)\right)$$

We divide by σ^p:

$$f(Z) = f_Q(Z) \left(1 - \frac{\sigma^2}{6} R(Y) + O(\sigma^3)\right) \cdot \left(\frac{m_0(Z)}{(\sqrt{2\pi})^p} + \sigma^2 \frac{m_2(Z)}{(\sqrt{2\pi})^p} + \mathcal{O}(\sigma^4) + \varepsilon(\sigma)\right)$$

We develop everything except the Gaussian in front:

$$f(Z) = f_Q(Z) \left(\frac{m_0(Z)}{(\sqrt{2\pi})^p} - \frac{\sigma^2}{6} \frac{m_0(Z)}{(\sqrt{2\pi})^p} R(Z) + \sigma^2 \frac{m_2(Z)}{(\sqrt{2\pi})^p} + \mathcal{O}(\sigma^4) + \varepsilon(\sigma)\right)$$

We write this:

$$f(Z) = f_Q(Z)\left(F_0(Z) + \sigma^2 F_2(Z) + \mathcal{O}(\sigma^4) + \varepsilon(\sigma)\right)$$

where:

$$\begin{aligned} F_0(Z) &= \frac{m_0(Z)}{(\sqrt{2\pi})^p} \\ &= \frac{\sqrt{(2\pi)^p}\sqrt{\det\left((M_{ab}(Z)^{-1}\right)}}{(\sqrt{2\pi})^p} \\ &= \sqrt{\det\left((M_{ab}(Z)^{-1}\right)} \end{aligned}$$

And:

$$\begin{aligned} F_2(Z) &= -\frac{1}{6}\frac{m_0(Z)}{(\sqrt{2\pi})^p}R(Z) + \frac{m_2(Z)}{(\sqrt{2\pi})^p} \\ &= -\frac{1}{6}\sqrt{\det\left((M_{ab}(Z)^{-1}\right)}R(Z) + \frac{m_2(Z)}{(\sqrt{2\pi})^p} \end{aligned}$$

So that:

$$\boxed{\begin{aligned} F_0(Z) &= \sqrt{\det\left((M_{ab}(Z)^{-1}\right)} \\ F_2(Z) &= -\frac{1}{6}\sqrt{\det\left((M_{ab}(Z)^{-1}\right)}R(Z) + \frac{m_2(Z)}{(\sqrt{2\pi})^p} \end{aligned}} \tag{A.14}$$

and we refer to the previous subsections for the formula of $M_{ab}(Z)$ and $m_2(Z)$.

A.4 Proof of Theorem A.4.1: Bias on the template shape

Now we prove the second theorem given in the paper "Template shape estimation".

Theorem A.4.1 *The data X_i's are generated with the model described in the paper "Template shape estimation", where the template shape Y is a parameter and under the assumptions of Theorem 1. The template shape Y is estimated with $\hat{Y}$, which is computed by the usual procedure described the paper.*

In the regime of an infinite number of data $n \to +\infty$, the asymptotic bias of the template's shape estimator $\hat{Y}$, with respect to the parameter Y, has the following Taylor expansion around the noise level $\sigma = 0$:

$$Bias(\hat{Y}, Y) = -\frac{\sigma^2}{2}H(Y) + \mathcal{O}(\sigma^4) + \varepsilon(\sigma) \tag{A.15}$$

where (i) H is the mean curvature vector of the template shape's orbit which represents the external curvature of the orbit in M, and (ii) ε is a function of σ that decreases exponentially for $\sigma \to 0$.

We compute the bias $\mathrm{Bias}(Y, \hat{Y})$ of $\hat{Y}$ as an estimator of Y. In the following, we take a NCS at Y. In particular, the vector $\overrightarrow{YZ} = \mathrm{Log}_Y Z$ has coordinates written z.

The expectation of the distribution f of shapes in Q is $\hat{Y}$ by definition. The point $\hat{Y}$, expressed in a NCS at the template Y, gives $\mathrm{Bias}(Y, \hat{Y})$, a tangent vector at $T_Y M$ that indicates how much one has to shoot to reach the estimator $\hat{Y}$:

$$\mathrm{Bias}(Y, \hat{Y}) = \mathrm{Log}_Y \hat{Y} = \int_Q \overrightarrow{YZ} f(Z) dQ(Z). \tag{A.16}$$

First, we take a ball of small radius r in Q and fix r. We split the integral:

$$\mathrm{Bias}(Y, \hat{Y}) = \int_{B_r^Q} \overrightarrow{YZ} f(Z) dQ(Z) + \int_{\mathcal{C}_{B_r^Q}} \overrightarrow{YZ} f(Z) dQ(Z)$$

By the result of the preliminaries, adapted to Q, the right part is a function $\sigma \to \varepsilon(\sigma)$ that is exponentially decreasing for $\sigma \to 0$.

$$\mathrm{Bias}(Y, \hat{Y}) = \int_{B_r^Q} \overrightarrow{YZ} f(Z) dQ(Z) + \varepsilon(\sigma)$$

A.4.1 Using the result of Theorem A.3.1

We plug the expression of the density f using Theorem A.3.1:

$$\begin{aligned}\text{Bias}(Y,\hat{Y}) &= \int_{B_r^Q} \overrightarrow{YZ} f_Q(Z)\left(F_0(Z)+\sigma^2 F_2(Z)+\mathcal{O}(\sigma^4)+\varepsilon(\sigma)\right) dQ(Z)+\varepsilon(\sigma)\\ &= \int_{B_r^Q} \overrightarrow{YZ} f_Q(Z)\left(F_0(Z)+\sigma^2 F_2(Z)+\mathcal{O}(\sigma^4)+\varepsilon(\sigma)\right) dQ(Z)+\varepsilon(\sigma)\end{aligned}$$

Computing the a-coordinate of the bias:

$$\text{Bias}(Y,\hat{Y})^a = \int_{B_r^Q} z^a f_Q(Z)\left(F_0(Z)+\sigma^2 F_2(Z)+\mathcal{O}(\sigma^4)+\varepsilon(\sigma)\right) dQ(Z)+\varepsilon(\sigma)$$

We are on a ball of small radius r around Y. We write the Taylor expansions of the F's terms around $z=0$:

$$\begin{aligned}F_0(Z) &= F_{00}(Y)+F_{01d}(Y)z^d+\mathcal{O}(||z||^2)\\ F_2(Z) &= F_{20}(Y)+F_{21d}(Y)z^d+\mathcal{O}(||z||^2)\end{aligned}$$

We replace these Taylor expansions in the expression of the bias:

$$\begin{aligned}\text{Bias}(Y,\hat{Y})^a = &\int_{B_r^Q} z^a f_Q(Z)\left(F_{00}(Y)+F_{01d}(Y)z^d+\mathcal{O}(||z||^2)\right) dQ(Z)\\ &+\int_{B_r^Q} z^a f_Q(Z)\left(\sigma^2 F_{20}(Y)+\sigma^2 F_{21d}(Y)z^d+\sigma^2\mathcal{O}(||z||^2)+\mathcal{O}(\sigma^4)+\varepsilon(\sigma)\right) dQ(Z)+\varepsilon(\sigma)\end{aligned}$$

Reorganizing:

$$\begin{aligned}\text{Bias}(Y,\hat{Y}) = &\int_{B_r^Q} z^a f_Q(Z)\left(F_{00}(Y)+F_{01d}(Y)z^d+\sigma^2 F_{20}(Y)+\sigma^2 F_{21d}(Y)z^d\right) dQ(Z)\\ &+\int_{B_r^Q} \overrightarrow{YZ} f_Q(Z)\left(\mathcal{O}(||z||^2)+\sigma^2\mathcal{O}(||z||^2)+\mathcal{O}(\sigma^4)+\varepsilon(\sigma)\right) dQ(Z)\\ &+\varepsilon(\sigma)\end{aligned}$$

By the dominated convergence theorem, we can put the inside $\varepsilon(\sigma)$ outside the parenthesis.

$$\begin{aligned}\text{Bias}(Y,\hat{Y}) = &\int_{B_r^Q} z^a f_Q(Z)\left(F_{00}(Y)+F_{01d}(Y)z^d+\sigma^2 F_{20}(Y)+\sigma^2 F_{21d}(Y)z^d\right) dQ(Z)\\ &+\int_{B_r^Q} \overrightarrow{YZ} f_Q(Z)\left(\mathcal{O}(||z||^2)+\sigma^2\mathcal{O}(||z||^2)+\mathcal{O}(\sigma^4)\right) dQ(Z)\\ &+\varepsilon(\sigma)\end{aligned}$$

We develop the measure on dQ: $dQ(Z) = (1-\frac{1}{6}\text{Ric}(Y)_{bc}z^b z^c+\mathcal{O}(||z||^3))dz$:

$$\begin{aligned}\text{Bias}(Y,\hat{Y})^a = &\int_{B_r^Q} f_Q(Z)\left(F_{00}(Y)z^a+F_{01d}(Y)z^a z^d\right)(1-\frac{1}{6}\text{Ric}(Y)_{bc}z^b z^c+\mathcal{O}(||z||^3))dz\\ &+\int_{B_r^Q} f_Q(Z)\left(\sigma^2 F_{20}(Y)z^a+\sigma^2 F_{21d}(Y).z^a z^d\right)(1-\frac{1}{6}\text{Ric}(Y)_{bc}z^b z^c+\mathcal{O}(||z||^3))dz\\ &+\int_{B_r^Q} f_Q(Z)z^a\left(\mathcal{O}(||z||^2)+\sigma^2\mathcal{O}(||z||^2)+\mathcal{O}(\sigma^4)\right)(1-\frac{1}{6}\text{Ric}(Y)_{bc}z^b z^c+\mathcal{O}(||z||^3))dz\\ &+\varepsilon(\sigma)\end{aligned}$$

We develop:

$$\begin{aligned}\mathrm{Bias}(Y,\hat{Y})^a = &\int_{B_r^Q} f_Q(Z)\left(F_{00}(Y)z^a + F_{01d}(Y)z^a z^d + \sigma^2 F_{20}(Y)z^a + \sigma^2 F_{21d}(Y).z^a z^d\right)dz\\ &+\int_{B_r^Q} f_Q(Z)\left(F_{00}(Y)z^a + F_{01d}(Y)z^a z^d\right)(-\frac{1}{6}\mathrm{Ric}(Y)_{bc}z^b z^c)dz\\ &+\int_{B_r^Q} f_Q(Z)\left(\sigma^2 F_{20}(Y)z^a + \sigma^2 F_{21d}(Y).z^a z^d\right)(-\frac{1}{6}\mathrm{Ric}(Y)_{bc}z^b z^c)dz\\ &+\int_{B_r^Q} f_Q(Z)\left(F_{00}(Y)z^a + F_{01d}(Y)z^a z^d + \sigma^2 F_{20}(Y)z^a + \sigma^2 F_{21d}(Y).z^a z^d\right)\mathcal{O}(||z||^3)dz\\ &+\int_{B_r^Q} f_Q(Z)z^a\left(\mathcal{O}(||z||^2) + \sigma^2\mathcal{O}(||z||^2) + \mathcal{O}(\sigma^4)\right)dz\\ &+\int_{B_r^Q} f_Q(Z)z^a\left(\mathcal{O}(||z||^2) + \sigma^2\mathcal{O}(||z||^2) + \mathcal{O}(\sigma^4)\right)(-\frac{1}{6}\mathrm{Ric}(Y)_{bc}z^b z^c)dz\\ &+\int_{B_r^Q} f_Q(Z)z^a\left(\mathcal{O}(||z||^2) + \sigma^2\mathcal{O}(||z||^2) + \mathcal{O}(\sigma^4)\right)\mathcal{O}(||z||^3)dz\\ &+\varepsilon(\sigma)\end{aligned}$$

We eliminate the odd terms that give 0 by skewsymmetry:

$$\begin{aligned}\mathrm{Bias}(Y,\hat{Y})^a = &\int_{B_r^Q} f_Q(Z)\left(+F_{01d}(Y)z^a z^d + \sigma^2 F_{21d}(Y).z^a z^d\right)dz\\ &+\int_{B_r^Q} f_Q(Z)\left(F_{01d}(Y)z^a z^d + \sigma^2 F_{21d}(Y).z^a z^d\right)(-\frac{1}{6}\mathrm{Ric}(Y)_{bc}z^b z^c)dz\\ &+\int_{B_r^Q} f_Q(Z)\left(F_{00}(Y)z^a + \sigma^2 F_{20}(Y)z^a\right)\mathcal{O}(||z||^3)dz\\ &+\int_{B_r^Q} f_Q(Z)z^a\left(\mathcal{O}(\sigma^4)\right)dz\\ &+\int_{B_r^Q} f_Q(Z)z^a\left(\mathcal{O}(\sigma^4)\right)(-\frac{1}{6}\mathrm{Ric}(Y)_{bc}z^b z^c)dz\\ &+\int_{B_r^Q} f_Q(Z)z^a\left(\mathcal{O}(||z||^2) + \sigma^2\mathcal{O}(||z||^2) + \mathcal{O}(\sigma^4)\right)\mathcal{O}(||z||^3)dz\\ &+\varepsilon(\sigma)\end{aligned}$$

We delete the terms that will give more than $\mathcal{O}(\sigma^2)$ by integration (these are normalized moments) and put them in $\mathcal{O}(\sigma^4)$ which is the next order since there is no$\mathcal{O}(\sigma^3)$:

$$\mathrm{Bias}(Y,\hat{Y})^a = \int_{B_r^Q} f_Q(Z)F_{01d}(Y)z^a z^d dz + \int_{B_r^Q} f_Q(Z)z^a\left(\mathcal{O}(\sigma^4)\right)dz + \mathcal{O}(\sigma^4) + \varepsilon(\sigma)$$

We gather the terms in $\mathcal{O}(\sigma^4)$:

$$\mathrm{Bias}(Y,\hat{Y})^a = \int_{B_r^Q} f_Q(Z)F_{01d}(Y)z^a z^d dz + \mathcal{O}(\sigma^4) + \varepsilon(\sigma)$$

We recognize the *normalized* truncated moment of order 2 in the q-dimensional Riemannian manifold Q, see Subsection A.2.3:

$$\mathrm{Bias}(Y,\hat{Y})^a = F_{01d}(Y)\mathfrak{M}_r^{Qad}(\sigma^2\mathbb{I}) + \mathcal{O}(\sigma^4) + \varepsilon(\sigma)$$

We express it with respect to the non-truncated *normalized* moment in $\mathbb{R}^q$, see Subsection A.2.3:

$$\mathrm{Bias}(Y,\hat{Y})^a = F_{01d}(Y)\left(\sigma^2\mathcal{M}^{Qad}(\mathbb{I}) + \mathcal{O}(\sigma^4) - \varepsilon(\sigma)\right) + \mathcal{O}(\sigma^4) + \varepsilon(\sigma)$$

We gather the ε's and the $\mathcal{O}(\sigma^4)$:

$$\text{Bias}(Y,\hat{Y})^a = F_{01d}(Y)\sigma^2\mathcal{M}_r^{Qad}(\mathbb{I}) + \mathcal{O}(\sigma^4) + \varepsilon(\sigma)$$

We replace the *normalized* 2nd order moment by its expression which is simply δ^{ad}, see Subsection A.2.2:

$$\begin{aligned}\text{Bias}(Y,\hat{Y})^a &= F_{01d}(Y)\sigma^2\delta^{ad} + \mathcal{O}(\sigma^4) + \varepsilon'(\sigma)\\ &= F_{01}^a(Y)\sigma^2 + \mathcal{O}(\sigma^4) + \varepsilon(\sigma)\end{aligned}$$

A.4.2 Computation of $F_{01}^a(Y)$: Taylor expansion of $F_0(Z)$ in the coordinate z

The term $F_{01}^a(Y)$ is the first order coefficient in the Taylor expansion of $F_0(Z)$ around Y in the coordinate z. Thus, we compute this Taylor expansion.

We first compute the Taylor expression of $F_0(Z)$ using the coordinate $y = \overrightarrow{ZY} = \text{Log}_Z Y$ in the NCS at Z. The expression of $F_0(Z)$ in the previous section gives:

$$F_0(Z) = \Theta_{Z,0}(1) = \sqrt{(2\pi)^p}\sqrt{\det\left((M_{ab}(Z)^{-1}\right)}$$

The previous subsections give:

$$M_{ab}(Z) = -y_c h_{ab}^c(Z) + d_{ab}(Z)$$

and we replace $d_{ab}(Z)$ by the formula p.23 in [Brewin 2009] but keeping only the first order:

$$M_{ab}(Z) = \delta_{ab} - y_c h_{ab}^c(Y) + \mathcal{O}(||y||^2)$$

So that:

$$M_{ab}(Z)^{-1} = \delta_{ab} + y_c h_{ab}^c(Y) + \mathcal{O}(||y||^2)$$

We plug this in $F_0(Z)$:

$$\begin{aligned}F_0(Z) &= \sqrt{\det\left(\delta_{ab} + y_c h_{ab}^c(Y) + \mathcal{O}(||y||^2)\right)}\\ &= \sqrt{1 + y_c \text{Trace}(h_{ab}^c(Y)) + \mathcal{O}(||y||^2)}\\ &= \sqrt{1 + y_c H^c(Y) + \mathcal{O}(||y||^2)}\\ &= \left(1 + \frac{1}{2}y_c H^c(Y) + \mathcal{O}(||y||^2)\right)\end{aligned}$$

where the trace of the second fundamental form is the external curvature vector $H(Y)$ by definition.

We convert this Taylor expansion in y, the coordinate of Y in a NCS at Z, into a Taylor expansion in z, the coordinate of Z in a NCS at Y. To express y with respect to z, we consider the geodesic $\gamma_{ZY}(t)$ from Z to Y and the geodesic $\gamma_{YZ}(t)$ from Y to Z. When parameterized by the arclength s, they are related as follows:

$$\begin{aligned}\gamma_{ZY}(s) &= \text{Exp}_Z\left(s\overrightarrow{ZY}\right)\\ &= \gamma_{YZ}(1-s)\\ &= \text{Exp}_Y\left((1-s)\overrightarrow{YZ}\right)\end{aligned}$$

Differentiating this relation gives:

$$D\text{Exp}_Z|_{s\overrightarrow{ZY}}.\overrightarrow{ZY} = -D\text{Exp}_Y|_{(1-s)\overrightarrow{YZ}}.\overrightarrow{YZ}$$

Taking the relation at $s = 0$:

$$D\text{Exp}_Z|_0.\overrightarrow{ZY} = -D\text{Exp}_Y|_{\overrightarrow{YZ}}.\overrightarrow{YZ}$$

where $D\text{Exp}_Z|_0 = Id$, so that:

$$y = -D\text{Exp}_Y|_z.z$$

We use the definition of the NCS at Y. When $\text{Exp}_Y(u) = U$, then U has coordinates u in the NCS at Y. So that: $D\text{Exp}_Y|_u.u = u$. This gives, with $u = z$:

$$y = -z$$

and we get the Taylor expansion of $F_0(Z)$ expressed in the coordinate z:

$$F_0(Z) = \left(1 - \frac{1}{2}z_c H^c(Y) + \mathcal{O}(||z||^2)\right)$$

And we identify the term $F_{01}^a(Y)$ needed:

$$F_{01}^a(Y) = -\frac{1}{2}H^a(Y)$$

A.4.3 Final result: Taylor expansion of the bias

Replacing $F_{01}^a(Y)$ by its value computed above:

$$\text{Bias}(Y, \hat{Y})^a = -\frac{1}{2}.H^a(Y)\sigma^2 + \mathcal{O}(\sigma^4) + \varepsilon(\sigma)$$

Appendix B

Proofs of the lemmas of Chapter 6

B.1 Proof of lemma 6.3.1.

Take $I \in L_2(\Omega)$ and an element $\phi \in G_I$: $I \circ \phi^{(-1)} = I$. We write the diffeomorphism ϕ as the Lie group exponential of the vector field v of its Lie algebra:

$$\phi = \exp(tv) = Id + tv + \mathcal{O}(t^2)$$

Then, the equation above leads to:

$$\begin{aligned} I(x) &= I(x - tv(x) + \mathcal{O}(t^2)) \\ I(x) &= I(x) - DI(x).(tv(x)) + \mathcal{O}(t^2) \\ 0 &= DI(x).(tv(x)) + \mathcal{O}(t^2) \end{aligned}$$

The identification of the coefficients in this Taylor expansion leads to:

$$\nabla I(x).v(x) = 0$$

A vector field of the Lie algebra of the isotropy group of the image I is perpendicular the image's gradient at any point x of the image's domain Ω.

B.2 Proof of lemma 6.3.2.

The images I_1 and I_2 have the same MS graph. The graph of I_1, taken with the nodes and edges positions on Ω, can be diffeomorphically deformed on the graph of I_2. We take ψ_1 a diffeomorphism that realizes the graphs' matching. I_1 is transformed to $I_1 \circ \psi_1^{(-1)}$.

$I_1 \circ \psi_1^{(-1)}$ and I_2 share the same MS graph, taken with the nodes and edges' positions on Ω. We consider one cell of this graph. We consider the integral lines of their respective gradients $\nabla\psi_1.\nabla I_1 \circ \psi_1$ and ∇I_2 on the cell. Both define a "parallel" partition of the cell. As a consequence, the set of integral lines of the gradient of $I_1 \circ \psi_1^{(-1)}$ can be mapped diffeomorphically to the set of integral lines of the gradient of I_2. We take ψ_2 a diffeormorphism that realizes the matching of the integral lines. We write $\psi = \psi_2 \circ \psi_1$. At the end of this step, I_1 is transformed to $I_1 \circ \psi^{(-1)}$.

The gradients of $I_1 \circ \psi^{(-1)}$ and I_2 share the same integral lines. These integral lines have same direction too. Thus the gradients are parallel at any point $x \in \Omega$: there exists $\kappa \in F(\Omega, \mathbb{R}_+)$ such that:

$$\begin{aligned} \nabla \left(I_1 \circ \psi^{(-1)} \right)(x) &= \kappa(x) \nabla I_2(x), \quad \forall x \in \Omega \\ \nabla\psi.\nabla I_1 \circ \psi &= \kappa(x) \nabla I_2(x), \quad \forall x \in \Omega \end{aligned}$$

Appendix C

Proofs of the theorems of Chapter 7

C.1 Proof of Theorem 7.4.2

Proof We take a normal coordinate system centered at the template Y. We have the decomposition $X = (X_O, X_H) \in T_Y O \oplus H$, where O is the orbit of Y.

As in [Miolane 2016], we can consider that the distribution on the orbits is a Dirac without changing the result of our computations. As we will project back, it does not matter where the data sits on the orbit.

The generative model implies the following Riemannian normal distribution on the objects:

$$F(X) = \int_{Y \in Q} \frac{p_Q(Y)}{C_M(\sigma)} \exp\left(-\frac{d_M^2(X,Y)}{2\sigma^2}\right) dQ(Y) \quad \text{with} \quad C_M(\sigma) = \int_M \exp\left(-\frac{d_M^2(X,Y)}{2\sigma^2}\right) dM(X) \tag{C.1}$$

The distance $d_M(X,Y)$ expressed in the normal coordinate system at Y is simply $d_M(X,Y)^2 = X^T X$.

To use the result of [Miolane 2016], we truncate the Riemannian Gaussian:

$$F(X) = \int_{Y \in Q} \frac{p_Q(Y)}{S^m(\sqrt{2\pi})^m \sigma^m} \exp\left(-\frac{d_M^2(X,Y)}{2\sigma^2}\right) dQ(Y), \tag{C.2}$$

where S is the normalization coefficient coming of the univariate truncated Gaussian at $\sigma/2$. B_M, B_O and B_Q refer to geodesic balls of radius $\sigma/2$ in their respective spaces. We denote $C_M = S^m(\sqrt{2\pi})^m \sigma^m$ and remark that $C_M = C_H C_O = S^q(\sqrt{2\pi})^q \sigma^q . S^p(\sqrt{2\pi})^p \sigma^p$ where q is the dimension of the quotient space and p the dimension of the principal orbits.

Our goal is to compute the projected distribution on the shape space. This is done by integrating the distribution on the orbit of X out of $F(X)$:

$$\pi(F)(x) = \frac{1}{C_M} \int_{B_O} \int_{Y \in Q} p_Q(Y) \exp\left(-\frac{d_M^2(Y,X)^2}{2\sigma^2}\right) dO(X_O) dQ(Y).$$

Plugging this into the expression of the projected distribution:

$$\pi(F)(x) = \int_O \int_Q F_Y(X) p_Q(Y) dQ(Y) dO(X)$$

The integrals $\int_Q$ and $\int_O$ can be inverted. First, $F(X)p_Q(t)$ is a continuous function with a compact support induced by the truncated Gaussian. Thus, $F(X)p_Q(t)$ is dominated by a constant function which is its maximum on the compact. Fubini's theorem on integral inversion gives:

$$\begin{aligned}\pi(F)(x) &= \int_Q p_Q(Y) \int_O F_Y(X) dO(X) dQ(Y) \\ &= \int_Q p_Q(Y) f_Y(x) dQ(Y) \\ &= \mathbb{E}_{p_Q}\left[f_Y(x)\right]\end{aligned}$$

C.2 Proofs of Theorems 7.5.2 and 7.5.3.

We perform the change of metric:

$$g'^M(X) = e^{2\phi(X)} g^M(X) = \begin{pmatrix} e^{2\phi(X)}.g^Q(X) & 0 \\ 0 & g^{\mathcal{O},\text{int}}(X) \end{pmatrix} \tag{C.3}$$

The metric g' on M is also G-invariant. Thus we are under the same assumptions that in Theorem 1 and we can follow the computations of its proof. We will show that this new metric is chosen exactly in order to cancel the bias. Intuitively, this change of metric makes the manifold a product space almost everywhere. Also, it becomes a principal bundle. Therefore, there are no crossed terms and what is in the orbit does not affect what happens in the quotient space.

First, we show that under this new metric, there is no external curvature of the principal orbits. Recall that the external curvature is defined as:

$$h(u,v) = (\nabla_u(v))^\perp = \Gamma^a_{ij} u_i v_j \tag{C.4}$$

where the second equality comes from the fact that orbit and quotient are perpendicular for a G-invariant metric.

We show that the Christoffel symbol is 0.

$$\begin{aligned}\Gamma^a_{ij} &= \frac{1}{2} g'^{aC} \left(\frac{\partial g'_{Ci}}{\partial x_j} + \frac{\partial g_{Cj}}{\partial x_i} - \frac{\partial g_{ij}}{\partial x_C} \right) \\ &= \frac{1}{2} g'^{ab} \left(-\frac{\partial g'_{ij}}{\partial x_b} \right) \qquad \text{because g' is diagonal per blocs} \\ &= 0\end{aligned}$$

where the last line comes from the fact that g'_{ij} on the quotient coordinate anymore.

Therefore the external curvature is null. As it represents the second derivative of the local graph of the orbit, and this is true everywhere on the orbit, we conclude that this graph is locally null.

Because $G = 0$, we have without Taylor expansion nor approximation:

$$\begin{aligned} d_M(Y,Z)^2 &= \sigma^2(x_a x^a + z_i z^i) \\ C_M(\sigma) &= \sigma^m \int_M \exp\left(-\frac{d_M(Y,Z)^2}{2\sigma^2}\right) dM'(Z) \\ &= \sigma^m . \mathfrak{M}'^0(Y) \end{aligned}$$

Then, we add the weights $w(X)$, and thus, following the computations of the proof of Theorem 1, we see that the integral defining the systematic bias can be split:

$$\begin{aligned} \text{Bias}_\infty^a &= \frac{\sigma^{m+1}}{C_M(\sigma)} \int_M x^a w(X) \exp\left(-\frac{d_M(Y,Z)^2}{2\sigma^2}\right) dM'(Z) \\ &= \frac{\sigma^{m+1}}{\sigma^m . \mathfrak{M}'^0(Y)} \int_Q x^a w(X) \exp\left(-\frac{x_a x^a}{2}\right) dQ'(X) \int_O \exp\left(-\frac{z_i z^i}{2}\right) dO(Z) \\ &= \frac{\sigma}{\mathfrak{M}'^0(Y)} \int_Q x^a \exp\left(-\frac{x_a x^a}{2}\right) dQ(X) \int_O \exp\left(-\frac{z_i z^i}{2}\right) dO(Z) \\ &= 0 \end{aligned}$$

where the last equality comes from symmetry considerations on the first integral.