

# Statistical Computing on Manifolds: From Riemannian geometry to Computational Anatomy

Xavier Pennec<sup>1</sup>

INRIA Sophia Antipolis - Asclepios Team, 2004 Rte des Lucioles BP 93  
06902 Sophia Antipolis Cedex, France

**Abstract.** Computational anatomy is an emerging discipline that aims at analyzing and modeling the individual anatomy of organs and their biological variability across a population. The goal is not only to model the normal variations among a population, but also discover morphological differences between normal and pathological populations, and possibly to detect, model and classify the pathologies from structural abnormalities. Applications are very important both in neuroscience, to minimize the influence of the anatomical variability in functional group analysis, and in medical imaging, to better drive the adaptation of generic models of the anatomy (atlas) into patient-specific data (personalization).

However, understanding and modeling the shape of organs is made difficult by the absence of physical models for comparing different subjects, the complexity of shapes, and the high number of degrees of freedom implied. Moreover, the geometric nature of the anatomical features usually extracted raises the need for statistics and computational methods on objects that do not belong to standard Euclidean spaces. We investigate in this chapter the Riemannian metric as a basis for developing generic algorithms to compute on manifolds. We show that few computational tools derived from this structure can be used in practice as the atoms to build more complex generic algorithms such as mean computation, Mahalanobis distance, interpolation, filtering and anisotropic diffusion on fields of geometric features. This computational framework is illustrated with the joint estimation and anisotropic smoothing of diffusion tensor images and with the modeling of the brain variability from sulcal lines.

## 1 Introduction

### 1.1 Computational Anatomy

Anatomy is the science that studies the structure and the relationship in space of different organs and tissues in living systems. Since the 1980ies, an ever growing number of imaging modalities allows observing both the anatomy and the function *in vivo* and *in situ* at many spatial scales (from cells to the whole body) and at multiple time scales: milliseconds (e.g. beating heart), years (growth or aging), or even ages (evolution of species). Moreover, the non-invasive aspect

allows repeating the observations on multiple subjects. This has a strong impact on the goals of the anatomy which are changing from the description of a *representative individual* to the description of the structure and organization of organs at the *population level*. This led in the last 10 to 20 years to the gradual evolution of *descriptive atlases* into interactive and *generative models*, allowing the simulation of new observations. Typical examples are given for the brain by the MNI 305 [45] and ICBM 152 [97] templates that are the basis of the Brain Web MRI simulation engine [30]. In the orthopedic domain, one may cite the "bone morphing" method [55, 125] that allows to simulate the shape of bones.

The combination of these new observation means and of the computerized methods is at the heart of computational anatomy, an emerging discipline at the interface of geometry, statistics and image analysis which aims at developing algorithms to model and analyze the biological shape of tissues and organs. The goal is to estimate representative organ anatomies across diseases, populations, species or ages, to model the organ development across time (growth or aging), to establish their variability, and to correlate this variability information with other functional, genetic or structural information (e.g. fiber bundles extracted from diffusion tensor images). From an applicative point of view, a first objective is to understand and to model how life is functioning at the population level, for instance by classifying pathologies from structural deviations (taxonomy) and by integrating individual measures at the population level (spatial normalization) to relate anatomy and function. A second application objective is to provide better quantitative and objective measures to detect, understand and correct dysfunctions at the individual level in order to help therapy planning (before), control (during) and follow-up (after).

The method is generally to map some generic (atlas-based) knowledge to patients-specific data through atlas-patient registration. In the case of observations of the same subject, many geometrical and physically based registration methods were proposed to faithfully model and recover the deformations. However, in the case of different subjects, the absence of physical models relating the anatomies leads to a reliance on statistics to learn the geometrical relationship from many observations. This is usually done by identifying anatomically representative geometric features (points, tensors, curves, surfaces, volume transformations), and then modeling their statistical distribution across the population, for instance via a mean shape and covariance structure analysis after a group-wise matching. In the case of the brain, one can rely on a hierarchy of structural models:

- Anatomical or functional landmarks like the AC and PC points [133, 22];
- Curves like crest lines [132] or sulcal lines [93, 88, 50];
- Surfaces like the cortical surface or sulcal ribbons [135, 3, 141];
- images seen as 3D functions, which lead to voxel-based morphometry (VBM) [11];
- Rigid, multi-affine or diffeomorphic transformations [137, 101, 5], leading to Tensor-based morphometry (TBM).

To exemplify the methodology of computational anatomy, we will detail in Section 6.2 a method to estimate the variability of the cortex shape which relies on sulcal lines manually delineated within MRI brain images of many subjects. A first problem is to define what is the mean and the variance of a set of lines. In this case, the mean line is computed by optimization, and we chose a very simple variability model based on the covariance matrix of the corresponding points in each subject at each point of the mean line. In order to use this model in other applications (for instance to better guide the deformation of a brain template to a specific patient image), a second problem is to extrapolate the covariance information from the lines to the whole brain surface and volume. Indeed, positive definite symmetric matrices only constitute a convex cone in the vector space of symmetric matrices. Thus, convex operations like the mean are stable, but more general algorithms involving partial differential equations (PDEs) or gradient descent inevitably lead to negative eigenvalues which are not physically acceptable. Designing well behaved algorithms to work on covariance matrices also turns out to be crucial for the second application that will be described in Section 6.1: the processing of diffusion tensor images (DTI), a new type of MRI modality that reveals in vivo the anatomical architecture of the brain connections.

Actually, these examples are typical of the difficulty of computing statistics on geometric features. The underlying reason is that these features most often belong to curved manifolds rather than to Euclidean spaces. Thus, one cannot simply use the classical linear statistics and one needs to develop a more general theory on which consistent algorithms could be designed.

## 1.2 Statistical analysis on manifolds

Statistical computing on simple manifolds like the 3D sphere or a flat torus (for instance an image with opposite boundary points identified) might seem easy as we can see the geometrical properties (e.g. invariance by rotation or translation) and imagine tricks to alleviate the different problems. For instance, the average of points on a sphere is located inside the sphere and not on its surface, but unless the distribution is perfectly symmetric, one can always project the mean point on the sphere surface. However, when it comes to slightly more complex manifolds like the space of positive definite matrices or the space of rigid-body motions (rotations and translations), without even thinking to infinite dimensional manifolds like spaces of curves, surfaces or diffeomorphisms, computational tricks are much more difficult to find and have to be determined on a case by case basis.

Statistical analysis on manifolds is a relatively new domain at the confluence of several mathematical and application domains. Its goal is to statistically study geometric objects living in differential manifolds. Directional statistics [21, 74, 82, 95] provide a first approach to statistics on manifold. As the manifolds considered here are spheres and projective spaces, the tools developed were mostly *extrinsic*, i.e. relying on the embedding of the manifold in the ambient Euclidean space. More complex objects are obtained when we consider the “shape” of a set

of  $k$  points, i.e. the part that remains invariant under the action of a given group of transformations (usually rigid body ones or similarities). Statistics on shape spaces [78, 36, 86, 130] raised the need for intrinsic tools. However, the link between the tools developed, the metric used and the space structure was not always very clear.

Another mathematical approach was provided by the study of stochastic processes on Lie groups. For instance, [64] derived central limit theorems on different families of groups and semi-groups with specific algebraic properties. Since then, several authors in the area of stochastic differential geometry and stochastic calculus on manifolds proposed results related to mean values [75, 81, 44, 4, 122, 31].

In the area of numerical methods for dynamic systems and partial differential equations, quite a few interesting numerical methods were developed to preserve the geometric properties of the flow of a differential equation such as symplectic integrator for Hamiltonian systems, symmetric integrators for reversible systems and optimization methods on Lie groups and manifolds [67, 69]. In particular, several Newton iteration schemes relying on different structures were proposed to optimize a function on a matrix Lie group or on Riemannian manifolds [110, 92, 33]. From the applied mathematics and computer science point of view, people get interested in computing and optimizing on specific manifolds, like rotations and rigid body transformations [118, 65, 112, 62, 104], Stiefel and Grassmann manifolds [42].

Over the last years, several groups attempted to federate some of the above approaches in a general statistical and computing framework, with different objectives in mind. For instance, the aim of the theory of statistical manifolds [2, 109] is to provide a Riemannian structure to the space of parameters of statistical distribution. This evolved into the more general theory of information geometry [72, 76]. Seen from the point of view of statistics on manifolds rather than manifolds of statistical parameters, a few authors characterized the performances of some statistical parametric estimators in manifolds like the bias and the mean square error. For instance, [70] considered extrinsic statistics, based on the Euclidean distance of the embedding space, while [109] considered the intrinsic Riemannian distance, and refined the Cramer-Rao lower bound using bounds on the sectional curvature of the manifold. In [17–19], the authors focused on the asymptotic consistency properties of the extrinsic and intrinsic means and variances for large sample sizes, and were able to propose a central limit theorem for flat manifolds.

In view of computer vision and medical image analysis applications, our concern in [113, 114] was quite different: we aimed at developing computational tools that can consistently deal with geometric features, or that provide at least good approximations. As we often have few measurements, we were interested in small sample sizes rather than large one, and we preferred to obtain approximations rather than bounds on the quality of the estimation. Thus, a special interest was to develop Taylor expansions with respect to the variance, in order to evaluate the quality of the computations with respect to the curvature of the manifold.

In [109, 17–19] as well as in our work, the chosen framework is the one of geodesically complete Riemannian manifolds, which appears to be powerful enough to support an interesting theory. To ensure a maximal consistency, we chose to rely only on intrinsic properties of the Riemannian manifold, thus excluding methods based on the embedding of the manifold in an ambient Euclidean space.

### 1.3 Chapter Organization

We summarize in Section 2 the mathematical bases that are needed to deal with finite dimensional manifolds. Then, we show in Section 3 that a consistent set of statistical tools, including mean and covariance matrix analysis, can be developed based on the choice of a Riemannian metric. This algorithmic framework to compute on manifolds is then extended in Section 4 to process fields of geometric features (manifold-valued image). In particular, we show that one can perform interpolation, filtering, isotropic and anisotropic regularization and restoration of missing data (extrapolation or in-painting) on manifold valued images by using generalized weighted means and partial differential equations (PDEs). Finally, the methodology is exemplified in Section 6 with two example applications: the joint estimation and smoothing of diffusion tensor fields from diffusion weighted images, and the modeling of the variability of the brain from a data-set of precisely delineated sulcal lines, where covariance matrices are used to describe the anatomical variability of points in the brain.

## 2 A Riemannian computing framework

The goal of this section is to establish the mathematical bases that will allow to build a simple but consistent statistical computing framework on manifolds. We describe a few computational tools (namely the Riemannian Exp and Log maps) derived from a chosen Riemannian metric on a given manifold. The implementation of these atomic tools will then constitute the basis to build more complex generic algorithms in Section 3.4. The interested reader may refer to [34] for a more complete but still affordable presentation of Riemannian geometry and to [131, chap. 9] and [84, 59] for more details.

### 2.1 The Riemannian structure

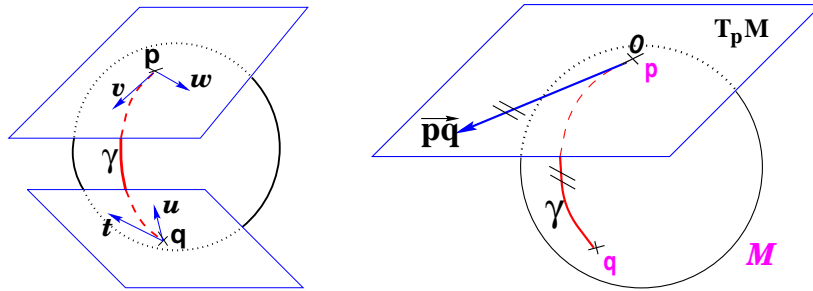
In the geometric framework, one has to separate the topological and differential properties of the manifold from the geometric and metric ones. The first ones determine the local structure of a manifold  $\mathcal{M}$  by specifying neighboring points and tangent vectors, which allows us to differentiate smooth functions on the manifold. This also allows us to define continuous paths on the manifold and to classify them by the number of loops they are doing around "holes" in the manifold. However, within each of these homotopy classes, there is no tool to choose something like the "straightest path". To obtain such a notion, we need to add

a geometric structure, called a connection, which allows to compare neighboring tangent spaces. Indeed, differentiating a path on a manifold gives tangent vectors belonging at each point to a different tangent vector space. In order to compute the second order derivative (the acceleration of the path), we need a way to map the tangent space at a point to the tangent space at any neighboring point. This is the goal of a connection  $\nabla_X Y$ , which specifies how the vector field  $Y(p)$  is derived in the direction of the vector field  $X(p)$ . Such a connection operator also describes how a vector is transported from a tangent space to another along a given curve and specifies the local parallel transport. However, there is usually no global parallelism. As a matter of facts, transporting the same vector along two different curves arriving at the same point might lead to different ending vectors: this is easily seen on the sphere where traveling from north pole to the equator, then along the equator for 90 degrees and back to North pole turns any tangent vector by 90 degrees. This defect of global parallelism is the sign of curvature. By looking for curves that remains locally parallel to themselves (i.e. such that  $\nabla_{\dot{\gamma}}\dot{\gamma} = 0$ ), one defines the equivalent of "straight lines" in the manifold: geodesics. One should notice that there exists many different choices of connections on a given manifold which lead to different geodesics.

Geodesics by themselves do not quantify how far away from each other two points are. For that purpose, we need an additional structure: a distance. By restricting to distances which are compatible with the differential structure, we enter into the realm of Riemannian geometry. A *Riemannian metric* is defined by a continuous collection of scalar products  $\langle \cdot | \cdot \rangle_p$  (or equivalently norms  $\|\cdot\|_p$ ) on each tangent space  $T_p\mathcal{M}$  at point  $p$  of the manifold. Thus, if we consider a curve on the manifold, we can compute at each point its instantaneous speed vector (this operation only involves the differential structure) and its norm to obtain the instantaneous speed (the Riemannian metric is needed for this operation). To compute the length of the curve, this value is integrated as usual along the curve. The distance between two points of a connected Riemannian manifold is the minimum length among the curves joining these points. The curves realizing this minimum are called metric geodesics. The fundamental theorem of Riemannian geometry states that on any Riemannian manifold there is a unique (torsion-free) connection which is compatible with the metric, called the Levi-Civita (or metric) connection. For that choice of connection, shortest path are geodesics ("straight lines"). In the following, we only consider the Levi-Civita connection. Moreover, we assume that the manifold is geodesically complete, i.e. that all geodesics can be indefinitely extended. This means that the manifold has neither boundary nor any singular point that we can reach in a finite time. As an important consequence, the Hopf-Rinow-De Rham theorem states that there always exists at least one minimizing geodesic between any two points of the manifold (i.e. whose length is the distance between the two points).

## 2.2 Exponential charts

The calculus of variations shows that geodesics are the solutions of a system of second order differential equations depending on the connection (thus on the



**Fig. 1. Left:** The tangent planes at points  $p$  and  $q$  of the sphere  $\mathcal{S}_2$  are different: the vectors  $v$  and  $w$  of  $T_p\mathcal{M}$  cannot be compared to the vectors  $t$  and  $u$  of  $T_q\mathcal{M}$ . Thus, it is natural to define the scalar product on each tangent plane. **Right:** The geodesics starting at  $p$  are straight lines in the exponential map and the distance along them is conserved.

metric)<sup>1</sup>. Let  $p$  be a point of the manifold that we consider as a local reference and  $\mathbf{v}$  a vector of the tangent space  $T_p\mathcal{M}$  at that point. From the theory of second order differential equations, we know that there exists one and only one geodesic  $\gamma_{(p,\mathbf{v})}(t)$  starting from that point with this tangent vector. This allows to wrap the tangent space onto the manifold, or equivalently to develop the manifold in the tangent space along the geodesics (think of rolling a sphere along its tangent plane at a given point), by mapping to each vector  $\mathbf{v} \in T_p\mathcal{M}$  the point  $q$  of the manifold that is reached after a unit time by the geodesic  $\gamma_{(p,\mathbf{v})}(t)$  starting at  $p$  with tangent vector  $\vec{v}$ . This mapping  $\text{Exp}_p(\mathbf{v}) = \gamma_{(p,\mathbf{v})}(1)$  is called the *exponential map* at point  $p$ . Straight lines going through  $0$  in the tangent space are transformed into geodesics going through point  $p$  on the manifold and distances along these lines are conserved (Fig. 1).

The exponential map is defined in the whole tangent space  $T_p\mathcal{M}$  (since the manifold is geodesically complete) but it is generally one-to-one only locally around  $0$  in the tangent space (i.e. around  $p$  in the manifold). In the sequel, we denote by  $\vec{pq} = \text{Log}_p(q)$  the inverse of the exponential map: this is the smallest vector (in norm) such that  $q = \text{Exp}_p(\vec{pq})$ . If we look for the maximal definition domain, we find out that it is an open and star-shaped domain which boundary is called the *tangential cut-locus*  $C_p$ . The image of  $C_p$  by the exponential map is the cut locus  $\mathcal{C}_p$  of point  $p$ . This is (the closure of) the set of points where several minimizing geodesics starting from  $p$  meet. On the sphere  $\mathcal{S}_2(1)$  for instance, the cut locus of a point  $p$  is its antipodal point and the tangential cut locus is the circle of radius  $\pi$ .

The exponential and log maps within this domain realizes a chart (a local parameterization of the manifold) called *the exponential chart at point p*. It

<sup>1</sup> The Christoffel symbols  $\Gamma_{ab}^c$  determine the connection in a local coordinate system through  $\nabla_{\partial_a} \partial_b = \sum_c \Gamma_{ab}^c \partial_c$ . The Levi-Civita connection is determined from the metric tensor  $g_{ab}(p) = \langle \partial_a | \partial_b \rangle_p$  and its inverse  $g^{cd}$  by  $\Gamma_{ab}^c = \frac{1}{2} \sum_d g^{cd} (\partial_a g_{db} + \partial_b g_{da} - \partial_c g_{ab})$ .

covers all the manifold except the cut locus of the reference point  $p$ , which has a null measure. In this chart, geodesics starting from  $p$  are straight lines, and the distance from the reference point are conserved. This chart is somehow the “most linear” chart of the manifold with respect to the reference point  $p$ . The set of all the exponential charts at each point of the manifold realize an atlas which allows working very easily on the manifold, as we will see in the following.

### 2.3 Practical implementation

The exponential and logarithmic maps (from now on Exp and Log maps) are obviously different for each manifold and for each metric. Thus they have to be determined and implemented on a case by case basis. Examples of closed form expressions for rotations and rigid body transformations can be found for the left invariant metric in [120], and for covariance matrices (positive definite symmetric matrices, so called tensors in medical image analysis) in [117, 8] and Section 5. It has to be noticed that the equation of the geodesics are only needed for the sake of computational efficiency: geodesics are curves minimizing the distance but also the Riemannian energy (the integral of the squared speed) between two points. Thus computing  $\vec{pq} = \log_p(q)$  may be posed as an optimal control problem [77, 1], and computing  $\text{Exp}_p(v)$  as a numerical integration problem (see e.g. [69, 67]). This opens the way to statistical computing in more complex spaces than the one we considered up to now, like curves [100, 83, 147], surfaces, and diffeomorphic transformations. For instance, the large deformation diffeomorphic metric mapping (LDDMM) method proposed for inter-subject image registration in computational anatomy [15, 101, 102, 73] finds the geodesic in the joint intensity and deformation space by minimizing the Riemannian length of the deformation for a given right-invariant metric on a diffeomorphism group. Through the so called EPDiff equation (Euler-Poincaré equation for diffeomorphisms), this optimization framework has been recently rephrased in an exponential/logarithm framework similar to the one developed here [103]. Despite the infinite number of dimensions, simple statistics like the mean and the principal component analysis of a (finite) set of samples may still be computed [140, 40]. Exponential charts constitute very powerful atomic functions in terms of implementation on which we will be able to express practically all the geometric operations: the implementation of  $\text{Log}_p$  and  $\text{Exp}_q$  is the basis of programming on Riemannian manifolds, as we will see in the following.

In a Euclidean space, the exponential charts are nothing but one orthonormal coordinates system translated at each point: we have in this case  $\vec{pq} = \text{Log}_p(q) = q - p$  and  $\text{Exp}_p(v) = p + v$ . This example is more than a simple coincidence. In fact, most of the usual operations using additions and subtractions may be reinterpreted in a Riemannian framework using the notion of *bi-point*, an antecedent of vector introduced during the 19th Century. Indeed, vectors are defined as equivalent classes of bi-points in a Euclidean space. This is possible because we have a canonical way (the translation) to compare what happens at two different points. In a Riemannian manifold, we can still compare things locally (by parallel transportation), but not any more globally. This means that



each “vector” has to remember at which point of the manifold it is attached, which comes back to a bi-point.

A second way to see the vector  $\vec{pq}$  is as a vector of the tangent space at point  $p$ . Such a vector may be identified to a point on the manifold using the geodesic starting at  $p$  with tangent vector  $\vec{pq}$ , i.e. using the exponential map  $q = \text{Exp}_p(\vec{pq})$ . Conversely, the logarithmic map may be used to map almost any bi-point  $(p, q)$  into a vector  $\vec{pq} = \text{Log}_p(q)$  of  $T_p\mathcal{M}$ . This reinterpretation of addition and subtraction using logarithmic and exponential maps is very powerful to generalize algorithms working on vector spaces to algorithms on Riemannian manifolds, as illustrated in Table 1 and the in following sections.

	Euclidean space	Riemannian manifold
Subtraction	$\vec{pq} = q - p$	$\vec{pq} = \text{Log}_p(q)$
Addition	$p = q + v$	$q = \text{Exp}_p(v)$
Distance	$\text{dist}(p, q) = \ q - p\ $	$\text{dist}(p, q) = \ \vec{pq}\ _p$
Mean value (implicit)	$\sum_i (p_i - \bar{p}) = 0$	$\sum_i \vec{pp}_i = 0$
Gradient descent	$p_{t+\varepsilon} = p_t - \varepsilon \nabla C(p_t)$	$p_{t+\varepsilon} = \text{Exp}_{p_t}(-\varepsilon \nabla C(p_t))$
Geodesic interpolation	$p(t) = p_0 + t \vec{p_0p_1}$	$p(t) = \text{Exp}_{p_0}(t \vec{p_0p_1})$

**Table 1.** Re-interpretation of standard operations in a Riemannian manifold.

### 3 Simple statistics on Riemannian manifolds

The Riemannian metric induces an infinitesimal volume element on each tangent space, and thus a reference measure  $d\mathcal{M}(p)$  on the manifold that can be used to measure random elements on the manifold (generalization of random variables). Without entering into the details of measure theory, such an element is characterized by its probability measure  $dP(p)$ . Its probability density function (pdf) is the function  $\rho$  such that  $dP(p) = \rho(p)d\mathcal{M}(p)$ , if it exists. The induced measure  $d\mathcal{M}$  actually represents the notion of *uniformity* according to the chosen Riemannian metric. This automatic derivation of the uniform measure from the metric gives a rather elegant solution to the Bertrand paradox for geometric probabilities [123, 79]. This paradox proposes three equally acceptable ways to compute the probability that the length of a "random chord" on a circle is greater than the side of an inscribed equilateral triangle, which lead to a probability of 1/2, 1/3 and 1/4. All methods are correct but actually rely on different uniform measures. The canonical definition of the uniform measure by the Riemannian metric prevents such a paradox to appear in our Riemannian setting.

With the probability measure  $dP$  of a random element, we can integrate any function  $f(p)$  from the manifold to any vector space, thus defining the expected

value of this function  $\mathbf{E}[f] = \int_{\mathcal{M}} f(p).dP(p)$ . This notion of expectation corresponds to the one we defined on real random variables and vectors. However, we cannot directly extend it to define the mean value of the distribution since we generally cannot integrate manifold-valued functions. Thus, one cannot define the mean or expected “value” of a random manifold element that way.

### 3.1 First statistical moment: the mean

As one cannot define the mean or expected “value” of a random manifold element using a weighted sum or an integral as usual, several alternative definitions based on properties on the usual mean were proposed (see [122] and [114, Sec. 4.3] for a review). The most interesting ones for general geodesically complete Riemannian manifolds were proposed by Fréchet, Karcher and Emery.

One solution is to rely on a distance-based variational formulation: the Fréchet [57, 58] (resp. Karcher [75]) expected features minimize globally (resp. locally) the variance:

$$\sigma^2(q) = \int_{\mathcal{M}} \text{dist}(p, q)^2 dP(p) = \frac{1}{n} \sum_{i=1}^n \text{dist}(p_i, q)^2,$$

written respectively in the continuous and discrete forms. One can generalize the variance to a dispersion at order  $\alpha$  by changing the  $L_2$  with an  $\alpha$ -norm:  $\sigma_\alpha(p) = (\int \text{dist}(p, q)^\alpha dP(p))^{1/\alpha}$ . The minimizers are called the central Karcher values at order  $\alpha$ . For instance, the median is obtained for  $\alpha = 1$  and the modes for  $\alpha = 0$ , exactly like in the vector case. It is worth noticing that the median and the modes are not unique in general in a vector space, and that even the mean may not exist (e.g. for heavy tailed distribution). In Riemannian manifolds, the existence and uniqueness of all central Karcher values is generally not ensured as they are obtained through a minimization procedure. However, [75] and [80] were able to establish existence and uniqueness theorems for distributions with a compact and small enough support. These theorems were then extended in [31] to distributions with non-compact support in a very specific class of manifolds that includes the Hadamard manifolds<sup>2</sup> whose curvature is bounded from below. This does not include rigid body transformations, but this includes the manifold of tensors. For a finite number of discrete samples at a finite distance of each other (which is the practical case in statistics) a mean value always exists (the variance is finite everywhere in a complete space so there exists a minimizer). and it is unique as soon as the distribution is sufficiently peaked.

Emery [44] proposed to use the *exponential barycenters*, i.e. the points at which the mean is null in the local exponential chart :  $\int_M \overrightarrow{xy} dP(y) = 0$ . If the support of the distribution is included in a *strongly convex open set*<sup>3</sup>, he showed that the exponential barycenters were the critical points of the variance. They

<sup>2</sup> Simply connected and complete manifolds with non-positive sectional curvature

<sup>3</sup> Here, strongly convex means that for every two points there is a unique minimizing geodesic joining them that depend in a  $C^\infty$  of the two points.

are thus a superset of the Riemannian centers of mass that include themselves the Fréchet means. Showing that these two notions continue to be essentially equivalent for distributions with a larger support is more difficult in the presence of a cut locus. Indeed, the distance is continuous but not differentiable at cut locus points where several minimizing geodesic meets. For instance, the distance from a point of the sphere to its antipodal point is maximal although the directional derivatives of the distance at this points are non zero in all directions. Thus, although the variance is continuous, it might not be differentiable everywhere. We showed in [113, 114] that it is actually differentiable at all points where the variance is finite and where the cut locus has a null mass for the considered probability measure<sup>4</sup>. In that case, its gradient is:

$$\nabla\sigma^2(q) = -2 \int \vec{qp} dP(p) = \frac{-2}{n} \sum_{i=1}^n \vec{qp}_i$$

respectively in the continuous (probabilistic) and discrete (statistical) formulations.

When we have a positive mass on the cut-locus, the right hand side of this equation is obviously not defined: the variance is continuous but can have a sharp extremum (most probably a maximum).

Thus, the extrema of the Riemannian variance are exponential barycenters or points with  $P(C(y)) > 0$ : apart from the specific problems with masses on the cut-locus, we have the implicit characterization of Karcher mean points as exponential barycenters which was presented in Table 1. Similar results have been derived independently in [109], where it is assumed that the probability is dominated by the Riemannian measure (which explicitly excludes point-mass distributions and the case  $P(C(y)) > 0$ ), and in [17, 18] for simply connected Riemannian manifolds with non-positive curvature. Our proof extends this result to any kind of manifold. Basically, the characterization of the Riemannian center of mass is the same as in Euclidean spaces if the curvature of manifold is non-positive (and bounded from below), in which case there is no cut-locus. If the sectional curvature becomes positive, a cut locus may appear, and a non-zero probability on this cut-locus induces some discontinuities in the first derivative of the variance. This corresponds to something like a Dirac measure on the second order derivative, which is an additional difficulty to compute the exact Hessian matrix of the variance on these manifolds. In practice, the gradient is well defined for discrete samples as soon as there is no sample lying exactly on the cut-locus of the current test point. Of course, perturbing the point position solves the problem (locally), but this might turn out to be a problem for designing certain certified computational algorithmic procedure if the same point is not perturbed exactly the same at different times.

---

<sup>4</sup> Notice that this is always the case when the random element has a density with respect to the Riemannian measure, but this does unfortunately not include the discrete (statistical) formulation where the probability measure is the sum of point masses at sample locations.

Picard [122] realized a good synthesis of most of these notions of mean value and show that the definition of a “barycenter” (i.e. a mean value) is linked to a connector, which determines itself a connection, and thus possibly a metric. An interesting property brought by this formulation is that the distance between two barycenters (with different definitions) is of the order of  $O(\sigma_x)$ . Thus, for sufficiently concentrated random points, all these values are close.

### 3.2 A Newton algorithm to compute the mean

To effectively compute the mean value, we proposed in [111, 112] a Gauss-Newton gradient descent algorithm on rotations and rigid-body motions. This algorithm was readily extended to general Riemannian manifolds in [113, 114] by approximating the variance using a Taylor expansion in a normal coordinate system: for a vector field  $v \in T_q\mathcal{M}$ , we have

$$\sigma^2(\text{Exp}_q(v)) = \sigma^2(q) + \left\langle \overrightarrow{\nabla\sigma^2}(q) \mid v \right\rangle_q + \frac{1}{2}\text{Hess}\sigma^2(v, v) + O(\|v\|_q^2)$$

The gradient of the variance being a vector field, the second order derivative (the Hessian) is obtained using the connection. However, we know that the gradient is not continuous at the cut locus. To circumscribe this problem, one can split the integral into one part that does not take into account the cut locus, which gives us a perfect positive definite matrix (2 times the identity), and one part that account for the cut locus, which can be expressed using integrals of Jacobi fields [75]. For a toy example on the circle, see also [114]. Deliberately neglecting this second term gives us a perfectly concave “second order approximation” with the following Gauss-Newton iterative scheme:

$$\bar{p}^{t+1} = \text{Exp}_{\bar{p}^t} \left( \frac{1}{n} \sum_{i=1}^n \overrightarrow{\bar{p}^t p_i} \right).$$

This algorithm essentially alternates the computation of the barycenter in the exponential chart centered at the current estimation of the mean value, and a re-centering step of the chart at the point of the manifold that corresponds to the computed barycenter (geodesic marching step). In practice, we found that this algorithm was very efficient and typically converges in 5 to 10 iterations to the numerical accuracy of the machine for rotations, rigid body transformations and positive definite symmetric matrices. Notice that it converges toward the real mean in a single step in a vector space. One can actually show that the convergence of this type of Newton iteration is locally quadratic around non degenerated critical points [110, 92, 33].

### 3.3 Covariance matrix and Principal Geodesic Analysis

Once the mean point is determined, using the exponential chart at the mean point is particularly interesting as the random feature is represented by a random

vector with null mean in a star-shaped domain. However, one important difference with the Euclidean case is that the reference measure is not the Lebesgue one but the pull-back of the Riemannian measure  $d\mathcal{M}$  by the Exponential map at the mean point. With this representation, there is no difficulty to define the covariance matrix (respectively continuous and discrete forms):

$$\Sigma = \int \vec{p}q \cdot \vec{p}q^T dP(q) = \frac{1}{n} \sum_{i=1}^n \vec{p}q_i \cdot \vec{p}q_i^T$$

and potentially higher order moments. This covariance matrix can then be used to define the Mahalanobis distance between a random and a deterministic feature that basically weights the distance between the deterministic feature and the mean feature using the inverse of the covariance matrix:  $\mu_{(\vec{p}, \Sigma)}(q) = \vec{p}q^T \Sigma^{-1} \vec{p}q$ . Interestingly, the expected Mahalanobis distance of a random element with itself is independent of the distribution and is equal to the dimension of the manifold, as in the vector case. This statistical distance can be used as a basis to generalize some statistical tests such as the Mahalanobis  $D^2$  test [114].

To analyze the results of a set of measurements in a Euclidean space, one often performs a principal component analysis (PCA). A generalization to Riemannian manifolds called Principal Geodesic Analysis (PGA) was proposed in [54] to analyze shapes based on the medial axis representations (M-reps). The basic idea is to find a low dimensional sub-manifold generated by some geodesic subspaces that best explain the measurements (i.e. such that the squared Riemannian distance from the measurements to that sub-manifold is minimized). Another point of view is to assume that the measurements are generated by a low dimensional Gaussian model. Estimating the model parameters amounts to a covariance analysis in order to find the  $k$ -dimensional subspace that best explains the variance. In a Euclidean space, these two definitions correspond thanks to Pythagoras's theorem. However, in the Riemannian setting, geodesic subspaces are generally not orthogonal due to the curvature. Thus, the two notions differ: while the Riemannian covariance analysis can easily be performed in the tangent space of the mean, finding Riemannian sub-manifolds turns out to become an almost intractable problem. As a matter of fact, the solution retained by [54] was finally to rely on the covariance analysis.

When the distribution is unimodal and sufficiently peaked, we believe that covariance analysis is anyway much better suited. However, for many problems, the goal is rather to find a sub-manifold on which measurements are more or less uniformly distributed. This is the case for instance for features sampled on a surface or points sampled along a trajectory (time sequences). While the one dimensional case can be tackled by regression [32], the problem for higher dimensional sub-manifolds remains quite open. Some solutions may come from manifold embedding techniques as exemplified for instance in [24].

### 3.4 Gaussian and $\chi^2$ law

Several generalizations of the Gaussian distribution to Riemannian manifolds have already been proposed so far. In the stochastic calculus community, one usu-

ally consider the heat kernel  $\rho(p, q, t)$ , which is the transition density of the Brownian motion [64, 43, 66]. This is the smallest positive fundamental solution to the heat equation  $\frac{\partial f}{\partial t} - \Delta f = 0$ , where  $\Delta$  is the Laplace-Beltrami operator (i.e. the standard Laplacian with corrections for the Riemannian metric). On compact manifolds, an explicit basis of the heat kernel is given by the spectrum of the manifold-Laplacian (eigenvalues  $\lambda_i$  with associated eigenfunctions  $f_i$  solutions of  $\Delta f = \lambda f$ ). However, the explicit computation of this spectrum is impossible but in very few cases [59].

To obtain tractable formulas, several alternative distributions have been proposed in directional statistics [21, 74, 82, 95, 96], in particular the wrapped Gaussian distributions. The basic idea is to take the image by the exponential of a Gaussian distribution on the tangent space centered at the mean value (see e.g. [96] for the circular and spherical case). It is easy to see that the wrapped Gaussian distribution tends toward the mass distribution if the variance goes to zero. In the circular case, one can also show that it tends toward the uniform distribution for a large variance. This definition was extended in [109] by considering non-centered Gaussian distributions on the tangent spaces of the manifold in order to tackle the asymptotic properties of estimators. In this case, the mean value is generally not any more simply linked to the Gaussian parameters. In view of a computational theory, the main problem is that the pdf of the wrapped distributions can only be expressed if there is a particularly simple geometrical shape of the cut-locus. For instance, considering an anisotropic covariance on the  $n$ -dimensional sphere leads to very complex calculations.

Instead of keeping a Gaussian pdf in some tangent space, we propose in [114, 113, 111] a new variational approach which is consistent with the previous definitions of the mean and covariance. The property that we took as axiom is that the Gaussian distribution maximizes the entropy among all distributions when we know the mean and the covariance matrix. In the Riemannian setting, we defined the intrinsic entropy as the expectation of the logarithm of the intrinsic pdf:

$$\mathbf{H}[\rho] = - \int_{\mathcal{M}} \log(\rho(p)) \rho(p) d\mathcal{M}(p) = - \int_{\mathcal{M}} \log(\rho(p)) dP(p)$$

Our definition of the entropy is consistent with the measure inherited from the Riemannian metric since the pdf that maximizes the entropy when we only know that the result is in a compact set  $\mathcal{U}$  is the uniform density in this set:  $p_{\mathcal{U}}(p) = \mathbb{1}_{\mathcal{U}}(p) / \int_{\mathcal{U}} d\mathcal{M}(p)$ .

The intrinsic pdf maximizing this entropy knowing the mean  $\bar{x}$  and the covariance matrix  $\Sigma$  is a Gaussian distribution on the exponential chart centered at the mean point and truncated at the cut locus (if there is one)<sup>5</sup> [114]:

$$N_{(\bar{p}, \Gamma)}(q) = k \exp\left(-\frac{1}{2} \vec{p}\vec{q}^T \Gamma \vec{p}\vec{q}\right).$$

<sup>5</sup> The definition domain of the exponential map at the mean point has to be symmetric to obtain this result. This is the case in particular for symmetric spaces, i.e. Riemannian spaces which metric are invariant under some symmetry.

However, the relation between the concentration matrix (the “metric”  $\Gamma$  used in the exponential of the probability density function) and the covariance matrix  $\Sigma$  is more complex than the simple inversion of the vectorial case, as it has to be corrected for the curvature of the manifold. Using a Taylor expansion of the Riemannian measure, one can obtain computationally tractable approximations for any manifold in case of small variances: Let  $r = i(\mathcal{M}, \bar{x})$  be the injectivity radius at the mean point, i.e. the shortest distance to the cut-locus (by convention  $r = +\infty$  if there is no cut-locus). Assuming a finite variance for any concentration matrix  $\Gamma$ , we have the following Taylor expansions:

$$k = \frac{1 + O(\sigma^3) + \epsilon(\sigma/r)}{\sqrt{(2\pi)^n \det(\Sigma)}} \quad \text{and} \quad \Gamma = \Sigma^{(-1)} - \frac{1}{3} \text{Ric} + O(\sigma) + \epsilon(\sigma/r)$$

Here,  $\epsilon(x)$  is a function that is a  $O(x^k)$  for any positive  $k$  (more precisely, this is a function such that  $\forall k \in \mathbb{R}^+$ ,  $\lim_{0+} x^{-k} \epsilon(x) = 0$ ).

This family of distributions ranges from the point-mass distribution (for  $\Gamma = \infty$ ) to the uniform measure (i.e. uniform density for compact manifolds) for a null concentration matrix. For some theoretical reasons (including the non-differentiability of the pdf at the cut locus), this is probably not be the best generalization of the Gaussian. However, from a practical point of view, it provides effective and computationally tractable approximations for any manifold in case of small variances that we were not able to obtain from the other definitions.

Based on this generalized Gaussian, we investigated in [114, 113, 111] a generalization of the  $\chi^2$  law to manifolds by considering the Mahalanobis distance of a Gaussian random feature. In the same conditions as for the Gaussian, one can show that it has the same density as in the vectorial case up to an order 3 in  $\sigma$ . This opens the way to the generalization of many other statistical tests, as we may expect similarly simple approximations for sufficiently centered distributions.

### 3.5 A Link between Extrinsic and Robust Statistics

From a practical point of view, many of the efficient methods proposed to work on geometric data in real applications actually use tools which rely on an extrinsic distance in a carefully chosen embedding space rather than on the intrinsic Riemannian distance. The goal of this section is to investigate some bridges between extrinsic and intrinsic approaches.

The main tool that we use here is the the notion of statistical robustness, as defined in [71, 126] i.e. the property of an estimator to be insensitive to small departures from the statistical model assumptions. In particular, outliers (events that are not modeled) most often lie in the tail of the distribution and robust statistics aim at reducing their influence. What is interesting is that we can often see an extrinsic distance on a manifold as a robust version of the Riemannian distance, as we shall see below on specific examples.

Let us considered more specifically M-estimators [126] of the distance  $d_\phi(p, q) = \phi(\text{dist}(p, q))$  with  $\phi(0) = 0$  and  $\phi'$  decreasing monotonically from  $\phi'(0) = 1$

while remaining non negatives. These conditions ensure that  $d_\phi$  remains a distance which is equivalent to the Riemannian one for small distances, while giving less weight to points that are far away by tempering their distance. Thus, outliers that lie in the tail of the distribution have much less influence on the results of the estimation, hence the robustness. One can show [63] that using such a  $\phi$ -function amounts to replace in the computation of the mean the tangent vector  $\vec{pq}$  by the vector:

$$\psi(\vec{pq}) = \frac{\phi(\|\vec{pq}\|_p)}{\|\vec{pq}\|_p} \vec{pq} = \frac{\phi(\text{dist}(p, q))}{\text{dist}(p, q)} \vec{pq} = \vec{pq} + O(\|\vec{pq}\|^2)$$

This mapping constitute a connector in the sense of [122] (a smooth mapping that replaces the Euclidean difference  $q - p$ )<sup>6</sup>, exactly in the way we used the logarithmic map of the Riemannian metric. Thus, we could think of defining mean values, higher order moments and other statistical operations by replacing everywhere the Riemannian logarithmic and exponential map with their  $\phi$ -equivalent.

For instance, one can verify that  $\|\psi(\vec{pq})\|_p = d_\phi(p, q)$ . This show that the  $\phi$ -variance of a random point  $\sigma_\phi^2(p) = \mathbf{E} [ d_\phi^2(\mathbf{q}, p) ] = \int_{\mathcal{M}} \|\psi(\vec{pq})\|_p^2 dP(q)$  is properly defined. Likewise, one can define the  $\phi$ -covariance  $\Sigma_\phi(p) = \mathbf{E} [ \psi(\vec{pq}) \cdot \psi(\vec{pq})^t ]$ , which trace is still equal to the  $\phi$ -variance. This  $\phi$ -variance can be differentiated at the points where the cut-locus has a null probability measure (because the  $\phi$ -distance is dominated by the Riemannian distance), and we obtain:

$$\nabla \sigma_\phi^2(p) = -2 \int_{\mathcal{M}} \phi'(\|\vec{pq}\|_p) \psi(\vec{pq}) dP(q).$$

This formula is interesting as it shows the divergence of the different notions of mean: the  $\phi$ -center of mass is a weighted barycenter both in the Riemannian and in the  $\phi$  exponential charts, but it is generally different from the (unweighted)  $\phi$ -exponential barycenter. The different notions of means are not any more subsets of each other: although the extrinsic mean is a robust estimator of the mean, the consistency of the distance minimization and the exponential-based algorithms is broken. From a numerical point of view, using an efficient and robust estimator might be an interesting feature, but we need to control the quality of this estimation to establish the domain in which the estimations are numerically consistent. Let us illustrate this with unit vectors and 3D rotations.

Euclidean metric induces on the sphere  $\mathcal{S}_{n-1}$  a rotationally invariant Riemannian metric for which geodesics are great circles, and the distance between two unit vectors  $u$  and  $v$  is the angle  $\theta = d(u, v) = \arccos(u^t \cdot v)$ . The Euclidean metric  $d_E(u, v) = \|u - v\|$  can be considered as a  $\phi$  estimator with  $\phi(\theta) = 2 \sin(\theta/2)$ . With the help of a Lagrange multiplier, one easily computes that the extrinsic

<sup>6</sup> Formally, a connector is a smooth mapping from  $\mathcal{M} \times \mathcal{M}$  to  $T\mathcal{M}$  that maps a bi-point  $(p, q)$  in the manifold to a vector in the tangent space  $T_p\mathcal{M}$ . The mapping should zero at  $q = p$  and its differential at that point should be the identity. This ensures that it is locally consistent with the Riemannian Log map.



Euclidean mean is the renormalized Euclidean mean  $\bar{u} = \int udP(u) / \|\int udP(u)\|$ , which is thus a robust estimator of the Riemannian mean.

For directions, a quite used encoding is the tensor  $u.u^t$ , which may be seen as an immersion of the projective space  $\mathcal{P}_{n-1}$  into the vector space of  $n \times n$  matrices ( $\mathbb{R}^{n^2}$ ). With this embedding, the squared extrinsic Euclidean distance (renormalized to be consistent with the previous ones) is  $d^2(u, v) = \frac{1}{2}\|u.u^t - v.v^t\|^2 = 1 - (u^t v)^2 = \sin^2(\theta)$ . This is also a robust distance with  $\phi(\theta) = \sin(\theta)$  (for  $\theta \leq \pi$ ). In the tensor space, the encoding of a random direction is the random tensor  $T_u = \mathbf{E}[u.u^t]$ . One should notice that the mean direction is represented by the tensor  $\bar{u}.\bar{u}^t$  which is closest to  $T_u$  in the Euclidean sense: this is the eigenvector(s) of  $T_u$  corresponding to the largest eigenvalue. One can also show that the  $\phi$ -covariance of the direction is given directly by the restriction of the tensor to the hyperplane orthogonal to the first eigenvector [63]. Thus, the random tensor encodes not only for a robust estimation of the Riemannian mean but also for (an approximation) of the second order moments.

Simulations were run on a large number of cases to measure the relative accuracy of the vector and tensor estimations with respect to the Riemannian mean. Up to a variance of 20 degrees, the three methods have a similar accuracy and results are almost not distinguishable. Between 20 and 40 degrees of variance, the tensor estimation becomes different from the two others while keeping a comparable global accuracy. After 40 degrees, the accuracy of the tensor mean highly degrades; the vector mean becomes different from the Riemannian means while keeping for a while a similar accuracy.

A very similar analysis can be done with 3D rotations: one can also model two well known extrinsic methods to compute the mean as  $\phi$ -connectors. The first method is to represent rotations using unit quaternions, and to compute the renormalized Euclidean mean on the sphere of unit quaternions. As rotation quaternions are defined up to their signs, one theoretically needs to iterate this process and to re-orient the unit quaternions at each step in the hemisphere chosen to represent the mean in order to converge. This method amounts to consider the  $\phi$ -distance  $d_{quat}(\theta) = 4 \sin(\theta/4)$ . The second method is to average the rotation matrices directly in the  $3 \times 3$  matrix space. Then, the mean is “renormalized” by looking for the rotation matrix which is closest to this result in the Euclidean matrix distance (Froebenius) sense. This can be easily realized using a SVD decomposition on the mean matrix. This method amounts to consider the  $\phi$ -distance  $d_{mat}(\theta) = 2 \sin(\theta/2)$ . Simulation experiments were performed for the two extrinsic methods by [41] in a registration context, and later on for the mean with the three methods by [62]. Like for unit directions/orientations, estimation results were similar up to 40 degrees of variance in the input rotations.

These experiments showed that efficient extrinsic approximations can be designed and used in practice, at the cost of potential inconsistencies between several notions of the mean that might be used in different algorithms (next Section will develop many algorithms based on the mean). However, the intrinsic Riemannian theory may be used as a central tool to compare different

extrinsic metrics, to establish the quality of the resulting approximations and to control the limits of their validity.

## 4 Computing with Manifold-valued images

The previous section showed how to derive from the atomic Exp and Log maps many important statistical notions, like the mean, covariance and Principal Component Analysis (PCA). We now turn to the generalization of some image processing algorithms like interpolation, diffusion and restoration of missing data (extrapolation) to manifold-valued images. We show that most interpolation and filtering methods can be reformulated using weighted means. The linear and non-linear diffusion schemes can be adapted to Manifolds through PDEs, provided that we take into account the variations of the metric. For details, we refer the reader to [117].

### 4.1 Interpolation and filtering as weighted means

One of the important operations in geometric data processing is to interpolate values between known measurements. In 3D image processing, (tri-) linear interpolation is often used thanks to its very low computational load and comparatively much better results than nearest neighbor interpolation. Other popular methods include the cubic and, more generally, spline interpolations [134, 99]. The standard way to interpolate on a regular lattice is to make a linear combination of samples  $f_k$  at integer (lattice) coordinates  $k \in \mathbb{Z}^d$ :  $f(x) = \sum_k w(x-k) f_k$ . A typical example is the sinus cardinal interpolation where the convolution kernel has an infinite support. With the nearest-neighbor, linear (or tri-linear in 3D), and higher order spline interpolations, the kernel is piecewise polynomial, and has a compact support [134, 99]. With normalized weights, this interpolation can be seen as a weighted mean. Thus, it can be generalized in the manifold framework as an optimization problem: the interpolated value  $p(x)$  on our feature manifold is the point that minimizes  $C(p(x)) = \sum_{i=1}^n w_i(x) \text{dist}^2(p_i, p(x))$ . This can easily be solved using the iterative Gauss-Newton scheme proposed for the Karcher mean. The linear interpolation is interesting and can be written explicitly since it is a simple geodesic walking scheme:  $p(t) = \text{Exp}_{p_0}(t\overrightarrow{p_0p_1}) = \text{Exp}_{p_1}((1-t)\overrightarrow{p_1p_0})$ .

Many other operators can be rephrased as weighted means. For instance approximations and convolutions like Gaussian filtering can be viewed as the average of the neighboring values weighted by a (Gaussian) function of their spatial distance. For instance,  $\hat{F}(x) = \int_{\mathbb{R}^n} K(u) F(x+u) du$  is the minimizer of  $C(\hat{F}) = \int_{\mathbb{R}^n} K(u) \text{dist}^2(F(x+u), \hat{F}(x)) du$ . In this formulation the kernel can be a discrete measure, for instance if samples are defined on the points of a grid. In a Riemannian manifold, this minimization problem is still valid, but instead of a closed-form solution, we have once again a Gauss-Newton iterative gradient descent algorithm to reach the filtered value:

$$\hat{p}^{t+1}(x) = \int_{\mathbb{R}^n} K(u) \text{Log}_{\hat{p}^t(x)}(p(x+u)) du.$$

We can also use anisotropic and non-stationary kernels  $K(x, u)$ . For instance, it can be modulated by the norm of the derivative of the field in the direction  $u$ . We should notice that for a manifold-value field  $p(x)$ , the directional derivatives  $\partial_u p(x)$  is a tangent vector of  $T_{p(x)}\mathcal{M}$  which can be practically approximated using finite “differences” in the exponential chart:  $\partial_u p(x) \simeq \text{Log}_{p(x)}(p(x+u)) + O(\|u\|^2)$ . However, to measure the norm of this vector, we have to use the Riemannian metric at that point:  $\|\partial_u p\|_p$ .

## 4.2 Harmonic diffusion

An alternative to kernel filtering is to consider a regularization criterion that penalizes the spatial variations of the field. A measure of variation is the spatial gradient (the linear form that maps to any spatial direction  $u$  the directional derivative  $\partial_u p(x)$ ), which can be robustly computed as the matrix that best approximates the directional derivatives in the neighborhood (e.g. 6, 18 or 26 connectivity in 3D). The simplest criterion based on the gradient is the Harmonic energy

$$\text{Reg}(p) = \frac{1}{2} \int_{\Omega} \|\nabla p(x)\|_{p(x)}^2 dx = \frac{1}{2} \sum_{i=1}^d \int_{\Omega} \|\partial_{x_i} p(x)\|_{p(x)}^2 dx.$$

The Euler-Lagrange equation of this Harmonic regularization criterion with Neumann boundary conditions is as usual  $\nabla \text{Reg}(p)(x) = -\Delta p(x)$ . However, the Laplace-Beltrami operator on the manifold  $\Delta p(x)$  is the sum of the usual flat Euclidean second order directional derivatives  $\partial_{x_i}^2 p(x)$  in a locally orthogonal system and an additional term due to the curvature of the manifold that distorts the ortho-normality of this coordinate system in the neighborhood. To practically compute this operator, we proposed in [117] an efficient and general scheme based on the observation that the Christoffel symbols and their derivatives along the geodesics vanish at the origin of the exponential chart. This means that the correction for the curvature is in fact already included: by computing the standard Laplacian *in that specific map*, one gets the directional Laplace-Beltrami operator for free:  $\Delta_u p = \text{Log}_{p(x)}(p(x+u)) + \text{Log}_{p(x)}(p(x-u)) + O(\|u\|^4)$ . Averaging over all the directions in a spatial neighborhood  $\mathcal{V}$  finally gives a robust and efficient estimation scheme:

$$\Delta p(x) \propto \sum_{u \in \mathcal{V}} \frac{1}{\|u\|^2} \text{Log}_{p(x)}(p(x+u))$$

A very simple scheme to perform Harmonic diffusion is to use a first order geodesic gradient descent. At each iteration and at each point  $x$ , one walks a little bit along the geodesic which start at the current point with the opposite of the gradient of the regularization criterion  $p^{t+1}(x) = \text{Exp}_{p^t(x)}(-\varepsilon \Delta p^t(x))$ .

## 4.3 Anisotropic diffusion

In order to filter within homogeneous regions but not across their boundaries, an idea is to penalize the smoothing in the directions where the derivatives are

important [121, 61]. This can be realized directly in the discrete implementation of the Laplacian by weighting the directional Laplacian with a decreasing function of the norm  $\|\partial_u p\|_p$  of the gradient in that direction. For instance, we used  $\Delta_u p = \sum_u c(\|\partial_u p\|_p) \Delta_u p$  with  $c(x) = \exp(-x^2/\kappa^2)$  in [117]. As the convergence of this scheme is not guaranteed (anisotropic regularization “forces” may not derive from a well-posed energy), the problem may be reformulated as the optimization of a  $\phi$ -function of the Riemannian norm of the spatial gradient (a kind of robust M-estimator):  $Reg_\phi(p) = \frac{1}{2} \int_\Omega \phi(\|\nabla p(x)\|_{p(x)}) dx$ . By choosing an adequate  $\phi$ -function, one can give to the regularization an isotropic or anisotropic behavior [12]. The main difference with a classical Euclidean calculation is that we have to take the curvature into account by using the Laplace-Beltrami operator, and by measuring the length of directional derivatives using the Riemannian metric at the right point [47]. Using  $\Psi(x) = \phi'(x)/x$ , we get:

$$\nabla Reg_\phi(p) = -\Psi(\|\nabla p\|_p) \Delta p - \sum_{i=1}^d \partial_{x_i} \Psi(\|\nabla p\|_p) \partial_{x_i} p.$$

#### 4.4 Diffusion-based interpolation and extrapolation

The pure diffusion reduces the noise in the data but also the amount of information. Moreover, the total diffusion time that controls the amount of smoothing is difficult to estimate. At an infinite diffusion time, the field will be completely homogeneous. Thus, it is more interesting to consider the data as noisy observations and the regularization as a prior on the spatial regularity of the field. Usually, one assumes a Gaussian noise independent at each position, which leads to a least-squares criterion through a maximum likelihood approach. For a dense data field  $q(x)$ , the similarity criterion that is added to the regularization criterion is simply  $Sim(p) = \int_\Omega \text{dist}^2(p(x), q(x)) dx$ . The only difference here is that it uses the Riemannian distance. It simply adds a linear (geodesic) spring  $\nabla_p \text{dist}^2(p, q) = -2 \vec{pq}$  to the global gradient to prevent the regularization from pulling to far away from the original data.

For sparse measures, using directly the maximum likelihood on the observed data leads to deal with Dirac (mass) distributions in the derivatives, which is a problem for the numerical implementation. One solution is to consider the Dirac distribution as the limit of the Gaussian function  $G_\sigma$  when  $\sigma$  goes to zero, which leads to the regularized derivative [117]:

$$\nabla Sim(x) = -2 \sum_{i=1}^n G_\sigma(x - x_i) \overrightarrow{p(x)p_i}.$$

## 5 The Example of Covariance matrices

Positive definite symmetric matrices, called tensors in medical image analysis, are used for instance to encode the covariance matrix of the Brownian motion (diffusion) of water in Diffusion Tensor Imaging (DTI) [13, 87], to encode the joint variability at different places (Green function) in shape analysis (see [51, 50, 52]),

and in image analysis to guide the segmentation, grouping and motion analysis [98, 144, 23, 145]. They are also appearing in many other application domains. For instance, they are a common tool in numerical analysis to locally drive the size of the adaptive meshes in order to optimize the cost of solving PDEs in 3D [106]. In the formation of echo-Doppler or radar images, Teoplitz Hermitian positive definite matrices uniquely characterized circular complex random processes with a null mean [107].

The main computational problem is that the tensor space is a manifold that is not a vector space with the usual additive structure. As the positive definiteness constraint delimits a convex half-cone in the vector space of symmetric matrices, convex operations (like the mean) are stable in this space but problems arise with more complex operations. For instance, there is inevitably a point in the image where the time step is not small enough when smoothing fields of tensors with gradient descents, and this results into negative eigenvalues. Even when a spectral decomposition is performed to smooth independently the rotation (eigenvectors basis trihedron) and eigenvalues [138, 27], there is a continuity problem around equal eigenvalues.

To answer that problem, it was proposed concurrently by several authors in the context of Diffusion Tensor Images (DTI) to endow the space of tensors with a Riemannian metric invariant by any change of the underlying space coordinates, i.e. invariant under the action of affine transformations of covariance matrices. This led to the distance  $\text{dist}^2(\Sigma, \Lambda) = \text{Tr}(\log(\Sigma^{-1/2}\Lambda\Sigma^{-1/2})^2)$  where  $\exp$  and  $\log$  stand for the matrix logarithm.

This metric leads to a very regular Hadamard manifold structure (a hyperbolic-like space without cut-locus) which simplifies the computations. Tensors with null and infinite eigenvalues are both at an infinite distance of any positive definite symmetric matrix: the cone of positive definite symmetric matrices is changed into a space of “constant” (homogeneous) non-scalar curvature without boundaries. Moreover, there is one and only one geodesic joining any two tensors, the mean of a set of tensors is uniquely defined, and we can even define globally consistent orthonormal coordinate systems of tangent spaces. Thus, the structure we obtain is very close to a vector space, except that the space is curved. The invariant metric has been independently proposed in [53] for the analysis of principal modes of sets of diffusion tensors; in [105] for its mathematical properties which were exploited in [14] for a new anisotropic DTI index; and in [117] where we were not interested by the metric per-se, but rather as the basis for building the complete computational framework on manifold-valued images that we presented in last section. By looking for a suitable metric on the space of Gaussian distributions for the segmentation of diffusion tensor images, [89] also end-up with the same metric. It is interesting to see that completely different approaches, relying on invariance requirements on the one hand, and relying on an information measure to evaluate the distance between distributions on the other hand, lead to the same metric on the tensor space. The metric has been also previously introduced in statistics as the Fisher information metric to model the geometry of the multivariate normal family [25, 129, 26] and is considered as

a well known result in other branches of mathematics [16]. In computer vision, it was rediscovered to deal with covariance matrices [56]. An implicit form was introduced in [69] for developing flows and dynamic systems on the space of symmetric matrices. The corresponding integrator (which corresponds to a geodesic walking with this Riemannian metric) was used for the anisotropic regularization of diffusion tensor images in [28] and [20].

### 5.1 The one-parameter family of affine-invariant metrics

One can question about the uniqueness of this type of Riemannian metric. We have shown in [116] that there is actually a one-parameter family of such affine-invariant Riemannian metrics on tensors that all share the same connection. This is the affine invariant connection on homogeneous spaces of [108] which is used in many theorems on symmetric spaces in many differential geometry textbooks [85, 68, 60].

The basic idea is to define a group action, here the linear group  $GL_n$ , and to provide the space  $Sym_n^+$  of positive definite symmetric matrices (tensors) with a invariant Riemannian metric with respect to it. Here, the group action  $A \star \Sigma = A \Sigma A^T$  corresponds to the standard action of the affine group on the covariance matrix of random variables in  $\mathbb{R}^n$ , hence the name of the metric. When the group is sufficiently large to transport one point onto any other, the manifold is said homogeneous and we can also use the group to transport the metric from one point (called the origin) to any other point. The constraint is that the metric at the origin should be invariant by transformations that leave the origin unchanged (the isotropy group of that point).

In the case of tensors, the identity is left unchanged by rotations, so that the metric at that point should be rotationally invariant. All such dot products on symmetric matrices are given (up to a constant global multiplicative factor) by:

$$\langle V | W \rangle_{\text{Id}} = \text{Tr}(V W) + \beta \text{Tr}(V) \text{Tr}(W) \quad \text{with } \beta > -1/n$$

where  $n$  is the dimension of the space (the inequality ensures the positiveness). This metric at the identity can then be transported at any point by the group action using the (symmetric or any other) square root  $\Sigma^{1/2}$  considered as a group element:

$$\begin{aligned} \langle V | W \rangle_{\Sigma} &= \left\langle \Sigma^{-1/2} V \Sigma^{-1/2} \mid \Sigma^{-1/2} W \Sigma^{-1/2} \right\rangle_{\text{Id}} \\ &= \text{Tr}(V \Sigma^{-1} W \Sigma^{-1}) + \beta \text{Tr}(V \Sigma^{-1}) \text{Tr}(W \Sigma^{-1}) \end{aligned}$$

The Riemannian distance is obtained by integration, or more easily by the norm of the initial tangent vector of the geodesic joining the two points:

$$\begin{aligned} \text{dist}^2(\Sigma, \Lambda) &= \|\log_{\Sigma}(\Lambda)\|_{\Sigma}^2 = \|\Sigma^{-1/2} \log_{\Sigma}(\Lambda) \Sigma^{-1/2}\|_{\text{Id}}^2 \\ &= \text{Tr} \left( \log(\Sigma^{-1/2} \Lambda \Sigma^{-1/2})^2 \right) + \beta \text{Tr} \left( \log(\Sigma^{-1/2} \Lambda \Sigma^{-1/2}) \right)^2 \end{aligned}$$

It is worth noticing that tensors with null eigenvalues are at an infinite distance of any regular tensor, as are tensors with infinite eigenvalues: the original cone of

positive definite symmetric matrices, a linear manifold with a flat but incomplete metric (there is a boundary at a finite distance) has been changed into a regular and complete (but curved) manifold with an infinite development in each of its  $n(n+1)/2$  directions.

For  $\beta = -1/(n+1)$ , we have the metric that [91] proposed by embedding the space of tensors of dimension  $n$  into the space of  $n+1$  square matrices using homogeneous coordinates (this allows them to seamlessly take into account an additional position that represent the mean of the Gaussian distribution), and by quotienting out  $n+1$  dimensional rotations. The same trick could be used to embed the space in higher dimensional spaces (square matrices of dimension  $n+p+1$ , in which case one would obtain the invariant metric with  $\beta = -1/(n+p+1)$ ). Interestingly,  $-1/\beta = n+1$  is the first authorized integer to obtain a proper metric!

In fact, one can show that all the metrics of this affine-invariant family have the same Levy-Civita connection  $\nabla_V W = \nabla_W V = -1/2(V\Sigma^{-1}W + W\Sigma^{-1}V)$  [129]. This means that they share the same geodesics and the Riemannian Exp and Log maps at each point:

$$\begin{aligned}\text{Exp}_\Sigma(W) &= \Sigma^{1/2} \exp\left(\Sigma^{-1/2}W\Sigma^{-1/2}\right) \Sigma^{1/2} \\ \text{Log}_\Sigma(A) &= \Sigma^{1/2} \log\left(\Sigma^{-1/2}A\Sigma^{-1/2}\right) \Sigma^{1/2}\end{aligned}$$

However, one should be careful that the orthonormal bases are different for each metric which means that distances along the geodesics are different.

From the connection, one can compute the curvature tensor of the manifold [129]  $R(X, Y, V, W) = 1/4\text{Tr}(Y\Sigma^{-1}X\Sigma^{-1}V\Sigma^{-1}W - X\Sigma^{-1}Y\Sigma^{-1}V\Sigma^{-1}W)$ . From this tensor, one gets the sectional curvature and see that it is non positive and bounded from below (by  $-1/2$ ). Thus, it is Hadamard manifold, i.e. a kind of hyperbolic space in which we have for instance the existence and uniqueness of the mean. There is also no cut-locus, which simplifies the computations.

## 5.2 Log-Euclidean metrics

By trying to put a Lie group structure on the space of tensors, Vincent Arsigny observed that the matrix exponential was a diffeomorphism from the space of symmetric matrices to the tensor space. This well-known fact in mathematics was apparently never used to transport all the operations defined in the vector space of symmetric matrices to the tensor space, thus providing the tensor space with a commutative Lie group structure and even with a vector space structure [9, 8]. For instance, the composition (the log-product) is defined by  $\Sigma_1 \diamond \Sigma_2 = \exp(\log(\Sigma_1) + \log(\Sigma_2))$ . The Euclidean metric on symmetric matrices is transformed into a bi-invariant Riemannian metric on the tensor manifold (i.e. a metric which is invariant by both left and right compositions in the Lie group). As geodesics are straight lines in the space of symmetric matrices, the expression of the Exp, Log and distance maps for the Log-Euclidean metric is

easily determined:

$$\begin{aligned}\text{Exp}_{\Sigma}(W) &= \exp(\log(\Sigma) + \partial_W \log(\Sigma)) \\ \text{Log}_{\Sigma}(A) &= D \exp(\log(\Sigma)) (\log(A) - \log(\Sigma)) \\ \text{dist}_{LE}^2(\Sigma_1, \Sigma_2) &= \text{Tr}((\log(\Sigma_1) - \log(\Sigma_2))^2)\end{aligned}$$

These formulas look more complex than for the affine invariant metric because they involve the differential of the matrix exponential and logarithm in order to transport tangent vectors from one space to another [119]. However, they are in fact nothing but the transport of the addition and subtraction through the exponential of symmetric matrices. In practice, the log-Euclidean framework consist in taking the logarithm of the tensor data, computing like usual in the Euclidean space of symmetric matrices, and coming back at the end to the tensor space using the exponential [9, 7].

From a theoretical point of view, geodesics through the identity are the same as for affine-invariant metrics, but this is not true any more in general at other points of the tensor manifold [8]. The affine-invariant and log-Euclidean means are also identical if the mean commutes with all the data. When they are not equal, one can show that the log-Euclidean mean is slightly more anisotropic [8]. A careful comparison of both metrics in practical applications [7, 9] showed that there was very few differences on the results (of the order of 1%) on real DTI images, but that the log-Euclidean computations were 4 to 10 times faster. Thus, for this specific type of application, the log-Euclidean framework seems to be best suited. For other types of applications, like adaptive re-meshing [106], the anisotropy of the tensors can be much larger, which may lead to larger differences. In any case, initializing the iterative optimizations of affine-invariant algorithms with the log-Euclidean result drastically speeds-up the convergence.

### 5.3 Other families of metrics

Other families of metrics were also proposed to work with positive definite symmetric matrices, especially in view of processing diffusion tensor images. For instance, [143] proposed to parameterize tensors by their Cholesky decomposition  $\Sigma = LL^T$  where  $L$  is upper triangular with positive diagonal entries. Taking the standard flat Euclidean metric on the (positive) upper diagonal matrices leads to straight line geodesics in that space which are then transported to the tensor space as for the log-Euclidean framework. This lead to tensor space structure which is obviously flat, but where the null eigenvalues are at a finite distance, like in the Euclidean case.

Other square roots might be used to define other metrics on tensors as  $\Sigma = (LR)(LR)^T$  is also a valid decomposition for any rotation  $R$ . For instance, the symmetric square root  $UA^{1/2}U^T$  lead to a well defined metric on tensors which has similar properties as the Cholesky metric above, yet having different geodesics. The fact that the rotation  $R$  can be freely chosen to compute the square root recently led Dryden to propose a new metric in [35] which basically measure the shortest distance between all the square roots  $L_1R_1$  of  $\Sigma_1$  and  $L_2R_2$



of  $\Sigma_2$ . The minimal distance is realized by the Procrustes match of the square roots:

$$\text{dist}(\Sigma_1, \Sigma_2) = \min_{R \in O(n)} \|L_2 - L_1 R\|$$

and the optimal rotation  $\hat{R} = UV^T$  is obtained thanks to the singular value decomposition of  $L_2^T L_1 = USV^T$ . This metric is in fact the standard Kendall metric on the reflection size-and-shape space of  $n + 1$  points in dimension  $n$  [35, 37, 130], which geometry is well known. For instance, the minimal geodesic joining  $\Sigma_1$  to  $\Sigma_2$  is given by

$$\Sigma(t) = \left( (1-t)L_1 + tL_2\hat{R} \right) \left( (1-t)L_1 + tL_2\hat{R} \right)^T$$

From the equation of the geodesics, one can derive the Riemannian exp and log map and proceed with the general computing framework. However, one must be careful that this space is not complete and has singularities when the matrix  $\Sigma$  has rank  $n - 2$ , i.e. when 2 eigenvalues are going to zero [86].

In [142], Wang and Vemuri proposed to use the square root of the J-divergence (the symmetrized Kullback Leibler divergence) as a "distance"<sup>7</sup> on the tensor space:

$$\text{dist}_J^2(\Sigma_1, \Sigma_2) = \text{Tr}(\Sigma_1 \Sigma_2^{(-1)} + \Sigma_2 \Sigma_1^{(-1)}) - 2n$$

This J-distance has interesting properties: it is affine invariant, and the Fréchet mean value of a set of tensors  $\Sigma_i$  has a closed form solution:

$$\bar{\Sigma} = B^{-1/2} \left( B^{1/2} A B^{1/2} \right)^{1/2} A^{-1/2}$$

with  $A = \sum_i \Sigma_i$  and  $B = \sum_i \Sigma_i^{(-1)}$ . However, this is not a Riemannian distance as a Taylor expansion

$$\text{dist}_J^2(\Sigma, \Sigma + \epsilon V) = \frac{\epsilon^2}{2} \text{Tr}(\Sigma^{(-1)} V \Sigma^{(-1)} V) + O(\epsilon^3)$$

indicates that the underlying infinitesimal dot product is the usual affine invariant metric  $\langle V | W \rangle = 1/2 \text{Tr} \Sigma^{(-1)} V \Sigma^{(-1)} W$ . In fact, this metric might well be an extrinsic metric, like the one we investigated in Section 3.5 for unit vectors and rotations, and it would be interesting to determine what is the embedding space.

## 6 Applications in Computational Anatomy

Now that we have seen the generic statistical computing framework on Riemannian manifolds and families of Riemannian metrics that can be designed on the tensor manifold, let us turn to two important applications in medical image analysis and computational anatomy: diffusion tensor imaging, which provides unique in vivo information about the structure of the white matter fibers in the brain, and the estimation of the variability of the cortex among subjects.

<sup>7</sup> Quotation marks indicate here that the triangular inequality might not be verified.

## 6.1 Diffusion tensor imaging

Diffusion tensor Imaging (DTI) is a unique tool to assess in vivo oriented structures within tissues via the directional measure of water diffusion. Fiber tracking is the application targeted by most researchers in order to investigate non invasively the anatomical-functional architecture of the brain. Most of the current applications are currently in neuroscience, with high signal-to-noise ratios (SNR) images on healthy subjects rather. DTI might also prove to be an interesting quantification tool for medical diagnosis [124, 127]. However, using such a modality in a clinical environment is difficult: data often have to be acquired quickly because the patient cannot stay in a static position for too long due to pathologies. As this prevents the acquisition of multiple images for averaging, this results in a limited number of encoding gradients and low SNR images. The estimation of the diffusion tensor field from diffusion weighted images (DWI) being noise-sensitive, clinical DTI is often not suitable for fiber tracking. Thus, one need to regularize the tensor field without blurring the transitions between distinct fiber tracts, which delimit anatomical and functional brain regions. Smoothing independently each DWI before estimating the tensor results in a smoother tensor field but it also blurs the transitions between homogeneous regions, as this information is not accessible by taking each DWI individually. Consequently, one would like to perform an anisotropic regularization of the tensor field itself.

Most of the methods developed so far actually estimate the tensor field in a first phase with a simple algebraic method (see below) and then spatially regularize some of the geometric features of the tensor field. We believe that a better idea is to consider a prior on the spatial regularity when estimating the tensor field itself so that the estimation could remain statistically optimal with respect to the DWI noise model and could keep the maximum amount of information from the original data. We designed in [48, 49] a maximum likelihood (ML) criterion for the estimation of tensors fields from DWI with the MRI specific Rician noise (the amplitude of a complex Gaussian signal), and extended it to a maximum a posteriori (MAP) criterion by considering a spatial prior on the tensor field regularity. This results into an algorithm that jointly (rather than sequentially) performs the estimation and the regularization of the tensor field.

The Stejskal-Tanner diffusion equation [13] relates the diffusion tensor  $D$  to each noise-free DWI:

$$S_i = S_0 \exp(-b g_i^T D g_i)$$

where  $S_i$  is the original DWI corresponding to the encoding gradient  $g_i$ ,  $S_0$  the base image with a null gradient, and  $b$  the diffusion factor. By taking the logarithm of this equation, one obtain a linear system. Solving that system in a least square (LS) sense leads to the minimization of a quadratic criterion, which is easily performed using algebraic methods (see e.g. [146]). Doing this implicitly assumes a log-Gaussian noise on the images, which is justify only for high SNRs. Very few works did consider non log-Gaussian noise because it requires optimization techniques on tensors which are very difficult to control

with the standard Euclidean framework. With the log-Euclidean framework, such an optimization is not difficult (one could also restate everything within the affine-invariant framework but calculations are slightly more complex). For instance, in the case of a Gaussian noise on the DWIs, the tensor  $D = \exp(L)$  is parameterized by its logarithm  $L$ , an unconstrained symmetric matrix. The criterion to optimize is  $Sim_G(L) = \sum (\hat{S}_i - S_i(\exp(L)))^2$ , and the gradient is

$$\nabla Sim_G(L) = 2b \sum (\hat{S}_i - S_i) \cdot \partial_L S_i \quad \text{with} \quad \partial_L S_i = S_i \partial_{g_i \cdot g_i^t} \exp(L)$$

For low SNRs, we have to take into account the real nature of the noise in MRI, which is Gaussian in the complex domain (the k-space). [143] proposed an estimation criterion on the complex DWI signal that is adapted to that noise, with a computationally grounded optimization framework based on the Cholesky decomposition. However, one usually only have access to the amplitude of the signal complex signal in clinical images: in that case, the noise is thus Rician. One can show that such a noise induces a signal-dependent bias of the order of  $\sigma^2/2S$  on the DWI signal [128]. The signal being systematically larger than what it ought to be, the tensors will be under-estimated. To take explicitly the nature of this noise into account, we should optimize the log-likelihood of the signal corrupted by a Rician noise. This leads to a more complex criterion than above, but its gradient is very similar to the Gaussian case above:  $\nabla Sim_R(L) = -1/\sigma^2 \sum (S_i - \alpha \hat{S}_i) \partial_L S_i$ , except that we have a correcting factor  $\alpha = I'_0/I_0(\hat{S}_i S_i/\sigma^2)$  depending on the signal and the noise variance ( $I_0$  and  $I'_0$  are computable Bessel functions). The noise variance can easily be estimated on the background of the image (outside the head) where there is no signal.

For the spatial regularity, we proposed in [48, 49] to use a Markovian prior  $p(\Sigma(x+dx)|\Sigma(x)) \propto \exp(-\|\nabla \Sigma(x)^T \cdot dx\|_{\Sigma(x)}/\lambda)$ , and to account for discontinuities using a redescending M-estimator (a so-called  $\phi$ -functional). In the log-Euclidean framework, the tensor field  $\Sigma(x)$  is parameterized by its logarithm  $L(x)$ , and the log of the prior is simply:  $Reg(L) = \int_{\Omega} \phi(\|\nabla L\|)$ . In our experiments, we use  $\phi(s) = 2\sqrt{1+s^2/\kappa^2} - 2$ . The  $\phi$ -function preserves the edges of the tensor field while smoothing homogeneous regions. To include this regularization as an a-priori into the ML optimization process, we simply need to compute its gradient  $\nabla Reg(L) = -\psi(\|\nabla L\|) \Delta L - \sum_i \partial_i (\psi(\|\nabla L\|)) \cdot \partial_i L$  where  $\psi(s) = \phi'(s)/s$ . Directional derivatives, gradient and Laplacian were estimated with a finite differences scheme like with scalar images (see [48, 49] for details).

Experiments on synthetic data with contours and a Rician noise showed that the gradient descent technique was correctly removing the negative eigenvalues that did appear in the standard (Euclidean log-Gaussian) estimation technique. ML and MAP (with regularization) methods with a Gaussian noise model were underestimating the volume of tensors even more than the standard log-Gaussian method (30% instead of 20%), while Rician ML and MAP methods were estimating it within 5%. More interestingly, the methods were tested on two clinical datasets of low and medium quality: a brain image with a very low SNR (Fig. 2), and an experimental acquisition of a tumor in the spinal chord, both with 7 gradient directions (Fig. 3). This last type of acquisition is currently actively

investigated in clinical research (e.g. [46]). It is difficult to perform because the position is uncomfortable due to the tumor and the coil cannot be perfectly adapted to the body as it is for the head. The images are consequently much noisier than for the brain.

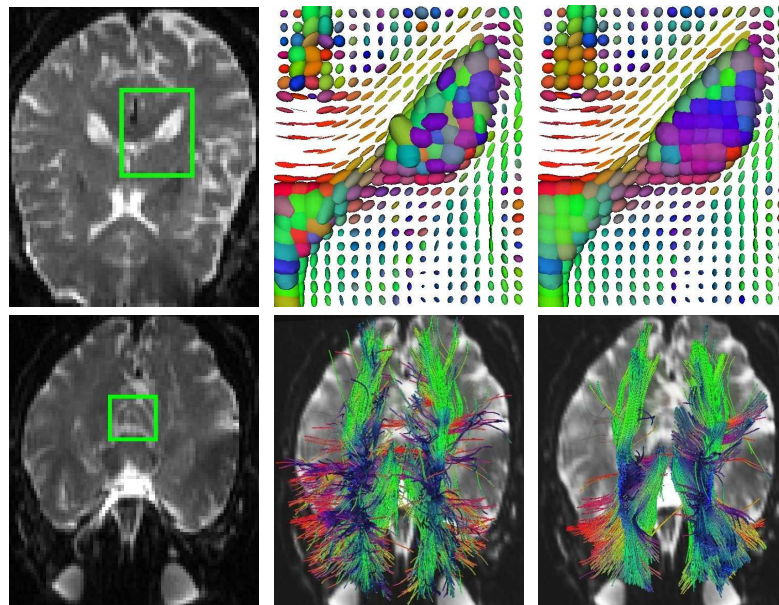
Like for synthetic data, using gradient descent techniques removed all the negative eigenvalues of the standard method. To evaluate the impact of the noise model on the tensor reconstruction in the brain, we computed the mean apparent diffusion coefficient (ADC), fractional anisotropy (FA) and volume of the diffusion tensors in the ventricles (high but anisotropic diffusion), and in the corpus callosum (lower diffusion with high anisotropy) [49]. Using the Rician noise model increase the tensor volume and the ADC by about 10% in isotropic regions and by 1 to 2% in anisotropic regions without modifying the FA. In the spinal chord, using the Rician noise model also lead to an increase of the tensors of about 30% in volume. This corresponds to the correction of the shrinking effect with Gaussian and Log-Gaussian noises. Adding some spatial regularization (MAP methods) systematically decreases the FA. However, this effect is much lower for anisotropic regions and minimized with the Rician noise model: 3% only in the corpus callosum (versus 11% with log-Gaussian), and 15% in the ventricles (versus 30% with log-Gaussian). Thus, it seems that these measurements are more reproducible with the MAP Rician reconstruction.

The tractography results in a much smoother and longer fibers with less dispersion for the MAP Rician model. The overall number of reconstructed fibers is also much larger. The smoothness of the tensor field indeed leads to more regular and longer fibers: tracts that were stopped due to the noise are now fully reconstructed. A careful quantitative evaluation and validation of the whole framework however remains to be done. In particular, it would be necessary to evaluate the reproducibility across acquisitions and scanners, for instance using repeated scans of the same subject, as well as evaluations of physical phantoms.

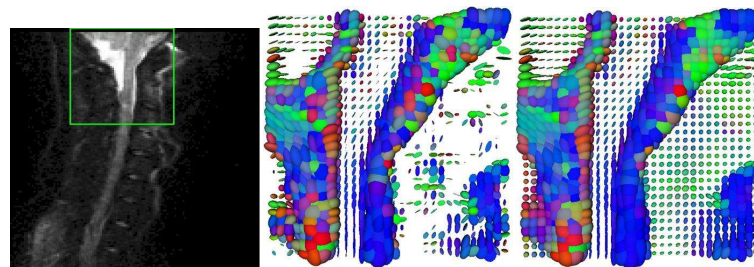
## 6.2 Learning Brain Variability from Sulcal Lines

A second interesting application is the statistical modeling of the brain variability in a given population of 3D images [51, 50]. In such a process, the identification of corresponding points among each individual anatomy (structural homologies) allows us to encode the brain variability by covariance matrices. The reason why we should not simplify these tensors into simpler scalar values is that there are evidences that structural variations are larger along certain preferred directions [136]. Thus the metrics on covariance matrices presented in Section 5 and the statical computing framework developed in earlier Sections are once again needed.

In the brain, a certain number of sulcal landmarks consistently appear in all normal individuals and allow a consistent subdivision of the cortex into major lobes and gyri [94]. Moreover, sulcal lines are low dimensional structures easily identified by neuroscientists. In the framework of the associated team program



**Fig. 2. Tensor field estimation of a brain (top row) and improvement of the fiber tracking (bottom row).** **Top Left:** A slice of the  $b_0$  image. **Top Middle:** The classic log-Gaussian estimation on the ROI. The color codes for the principal direction of tensors: **red:** left-right, **green:** anterior-posterior, **blue:** inferior-superior. Missing tensors in the splenium region are non-positive. **Top Right:** The MAP estimation of the same region. **Bottom row, Left:** ROI where the tracking is initiated. **Bottom row, middle:** The cortico-spinal tract reconstructed after a classic estimation. **Bottom row, Right:** Same tract reconstructed after our MAP estimation. Rendering is obtained using the MedINRIA software developed by P. Fillard and N. Toussaint.



**Fig. 3. Tensor field estimation of the spinal cord.** **Left:** A slice of the  $b_0$  image with the ROI squared in green. **Middle:** Classic **log-Gaussian ML** tensor estimation. There are many missing (non-positive) tensors around and in the spinal cord. **Right:** **Rician MAP** tensor estimation: tensors are all positive and the field is much more regular while preserving discontinuities. Original DWI are courtesy of D. Ducreux, MD. Rendering is obtained using the MedINRIA software (<http://www.inria.fr/sophia/asclepios/software/MedINRIA/>).

between Epidaure/Asclepios at INRIA and LONI at UCLA<sup>8</sup>, we use a data-set of sulcal lines manually delineated in 98 subjects by expert neuroanatomists according to a precise protocol<sup>9</sup>. We used the 72 sulcal curves that consistently appear in all normal subjects (abusively called *sulci* in the sequel).

To find the corresponding points between the 98 instances of each of the 72 sulci, we alternatively computed the matches that minimize the distance between the mean curve and the instances, and re-estimated the mean curve from the updated matches. As a result, we obtain for each point of each mean sulcal curve the set of corresponding anatomical positions in each subject. The number of tensor needed to represent the variability information along each sulcus was then adjusted by picking only a few tensors alors the mean line and linearly interpolating in-between them. The optimal subset of tensors is determined by optimizing the distance between interpolated and measured tensors along the line so that the error does not exceed a prescribed value. In this process, the distance and interpolation between covariance matrices was performed using the affine-invariant metric. Interestingly, selecting only 366 variability tensors was sufficient to encode the variability of the 72 sulci without a significant loss of accuracy. The result is a sparse field of tensors, that can naturally be extrapolated to the whole space using the framework described in Section 4.4 (Fig. 4). This dense map of tensors was shown to be in good agreement with previous published results: the highly specialized and lateralized areas such as the planum parietale and the temporo-parietal areas consistently shows the highest amount of variability. The lowest amount of variability is consistently found in phylogenetically older areas (e.g. orbitofrontal cortex) and primary cortices that myelinate earliest during development (e.g., primary somatosensory and auditory cortex). However, our variability map give more than the amount of variability since we can extract from the tensors the spatial directions where the variability is the greatest at every single anatomical position. We refer the reader to [51, 50] for a more detailed explanation of the method and for the neuroscience interpretation of these results.

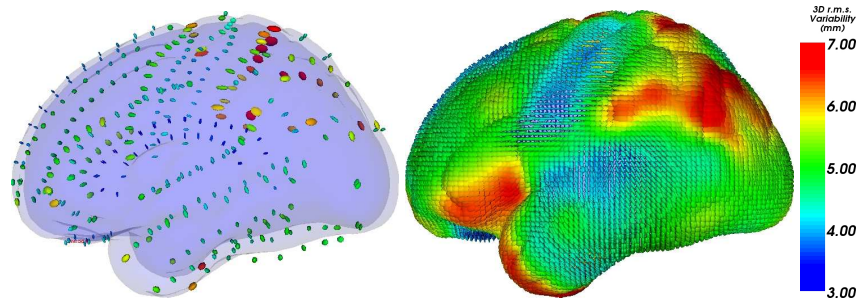
## 7 Challenges

We have shown in this chapter that the choice of a Riemannian metric and the implementation of a few tools derived from it, namely the Exp and Log maps, provides the bases for building a consistent algorithmic framework to compute on manifolds. In particular, we can compute consistent statistics, perform interpolation, filtering, isotropic and anisotropic regularization and restoration of missing data. Last but not least, powerful computational models of the anatomy could be built thanks to this Riemannian computing framework.

There are however many challenges opened both from the theoretical and application point of views. For instance, the Riemannian approach that we presented here is not perfectly consistent with the structure of Lie groups as soon

<sup>8</sup> <http://www-sop.inria.fr/epidaure/Collaborations/UCLA/atlas.html>

<sup>9</sup> [http://www.loni.ucla.edu/~khayashi/Public/medial\\_surface/](http://www.loni.ucla.edu/~khayashi/Public/medial_surface/)



**Fig. 4.** Variability tensor extrapolation. Left: The 366 tensors retained for our model. Right: Result of the extrapolation. Each point of this average brain shape contains a variability tensor.

as they are not compact nor Abelian, which is already the case for rigid body transformations. In that case, there is generally no left and right invariant metric, and most of the operations that we defined (e.g. the mean) with either the left or the right invariant metric are not consistent with inversion. To find an alternative to the Riemannian structure for Lie groups, we investigate with V. Arsigny the idea on relying on one-parameter subgroups instead of geodesics. Preliminary results indicate that this may provide an interesting structure [6, 5]. For instance, one can design bi-invariant means that are fully compatible with the group structure [10]. They are defined through fixed point equations which are very similar to the Riemannian ones. However, these equations do not derive from a well posed metric. It would be interesting to see what part of the statistical computing framework still holds if we replace the distance by a simple positive or negative energy. This probably amounts to considering the connection as the basic structure of the manifold instead of the Riemannian metric.

Another key problem is to extend our statistical computing framework to infinite dimensional manifolds such as surfaces and diffeomorphism groups. From a theoretical point of view, we know how to provide the diffeomorphism group with left or right invariant Riemannian metrics that are sufficiently smooth to compute the geodesics by optimization [15, 101, 102, 73]. Through the so called EPDiff equation (Euler-Poincaré equation for diffeomorphisms), this optimization framework has been recently rephrased in an exponential/logarithm framework similar to the one developed here [103]. Thus, the basic algorithmic tools are the same, except that optimizing each time to compute the exponential and the logarithm has a deep impact on the computational times. However, one difficulty is that the infinite number of dimensions forbids the use of many tools like the probability density functions! Thus, even if simple statistics like the mean and the principal component analysis of a finite set of samples may still be computed [140, 40], one should be very careful about ML-like statistical estimation in these spaces: there is always a finite number of data for an infinite number of parameters. In particular, there are infinitely many left- or right-invariant metrics on diffeomorphisms, and learning the optimal metric is an ill-posed problem.

Estimations need to be regularized with prior models or performed within finite dimensional families of metrics whose assumptions are suited for the problem at hand. An interesting track for that is to establish specific models of the Green's function based on the mixture of smoothly varying local and long-distance interaction convolution kernels. If we only consider the local kernel, the Riemannian elasticity [119, 115] could be an interesting family of metrics allowing to measure statistically the properties of the virtual underlying material. Moreover, it was recently shown that such a criterion was consistently separating monozygotic twins from others, which suggest that such deformation-based measures could be anatomically meaningful [90].

Last but not least, surfaces are an important source of anatomical structures in computational anatomy, and one need to design efficient methods and metrics to capture their statistical properties. It would also be useful to fuse the information coming from image deformations and from surfaces in a single framework. Courants (generalization of distributions) provide consistent mathematical tools for discrete and continuous surfaces [29]. A diffeomorphic registration algorithm of surfaces based on that notion was proposed for instance in [139]. The tools were then drastically improved in [40, 39, 38] to provide the basis of a computationally efficient statistical computing framework on curves and surfaces. We expect very interesting advances in this direction in the coming years.

From a computational anatomy standpoint, the huge number of degrees of freedom involved in the estimation of the anatomical variability will require to aggregate information coming from many different sources in order to improve the statistical power. As there is no gold standard, we should also be careful that many biases may be hidden in the results. Thus, methods to compare and fuse statistical information coming from many different anatomical features will need to be developed in order to confirm anatomical findings. For the brain variability, one could for instance add to the sulci other cortical landmarks like sulcal ribbons and gyri, the surface of internal structures like the ventricles, the hippocampus or the corpus callosum, or fiber pathways mapped from DTI. These sources of information are individually providing a partial and biased view of the whole variability. Thus, we expect to observe a good agreement in some areas, and complementary measures in other areas. This will most probably lead in a near future to new anatomical findings and more robust medical image analysis applications.

## References

1. S. Allasonnière, A. Trouvé, and L. Younes. Geodesic shooting and diffeomorphic matching via textured models. In A. Rangarajan et al., editor, *Proc. of EMMCVPR 2005*, LNCS 3757, page 365–381, 2005.
2. Shun-ichi Amari. *Differential-geometric methods in Statistics*, volume 28 of *Lecture Notes in Statistics*. Springer, 2nd corr. print edition, 1990.
3. A. Andrade, F. Kherif, J.-F. Mangin, K. Worsley, A.-L. Paradis, O. Simon, S. Dehaene, and J.-B. Poline. Detection of fMRI activation using cortical surface mapping. *Human Brain Mapping*, 12:79–93, 2001.



4. M. Arnaudon. Barycentres convexes et approximations des martingales continues dans les variétés. In M. Yor J. Azema, P.A. Meyer, editor, *Séminaire de probabilités XXIX*, volume 1613 of *Lect. Notes in Math.*, pages 70–85. Springer-Verlag, 1995.
5. Vincent Arsigny, Olivier Commowick, Xavier Pennec, and Nicholas Ayache. A log-Euclidean framework for statistics on diffeomorphisms. In *Proc. of the 9th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'06), Part I*, number 4190 in LNCS, pages 924–931, 2-4 October 2006. PMID: 17354979.
6. Vincent Arsigny, Olivier Commowick, Xavier Pennec, and Nicholas Ayache. A log-Euclidean polyaffine framework for locally rigid or affine registration. In J.P.W. Pluim, B. Likar, and F.A. Gerritsen, editors, *Proceedings of the Third International Workshop on Biomedical Image Registration (WBIR'06)*, volume 4057 of LNCS, pages 120–127, Utrecht, The Netherlands, 9 - 11 July 2006. Springer Verlag.
7. Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Fast and simple calculus on tensors in the log-Euclidean framework. In J. Duncan and G. Gerig, editors, *Proceedings of the 8th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention - MICCAI 2005, Part I*, volume 3749 of LNCS, pages 115–122, Palm Springs, CA, USA, October 26-29, 2005. Springer Verlag. PMID: 16685836.
8. Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 29(1):328–347, 2006.
9. Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine*, 56(2):411–421, August 2006. PMID: 16788917.
10. Vincent Arsigny, Xavier Pennec, and Nicholas Ayache. Bi-invariant means in lie groups. application to left-invariant polyaffine transformations. Research report rr-5885, INRIA Sophia-Antipolis, April 2006.
11. J. Ashburner and K. J. Friston. Voxel-based morphometry - the methods. *NeuroImage*, 11(6):805–821, 2000.
12. G. Aubert and P. Kornprobst. *Mathematical problems in image processing - Partial differential equations and the calculus of variations*, volume 147 of *Applied Mathematical Sciences*. Springer, 2001.
13. P.J. Basser, J. Mattiello, and D. Le Bihan. MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66:259–267, 1994.
14. P. Batchelor, M. Moakher, D. Atkinson, F. Calamante, and A. Connelly. A rigorous framework for diffusion tensor calculus. *Magnetic Resonance in Medicine*, 53:221–225, 2005.
15. M.F. Beg, M.I. Miller, A. Trouvé, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. Journal of Computer Vision*, 61(2):139–157, 2005.
16. R. Bhatia. On the exponential metric increasing property. *Linear Algebra and its Applications*, 375:211–220, 2003.
17. R. Bhattacharya and V. Patrangenaru. Nonparametric estimation of location and dispersion on Riemannian manifolds. *Journal of Statistical Planning and Inference*, 108:23–36, 2002.
18. R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds, I. *Annals of Statistics*, 31(1):1–29, 2003.

19. R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds, II. *Annals of Statistics*, 33(3):1225–1259, 2005.
20. G.N.J.C. Bierkens. Geometric methods in diffusion tensor regularization. Master's thesis, Technische Universiteit Eindhoven, Dept. of Math and Comp. Sci., 2004.
21. C. Bingham. An antipodally symmetric distribution on the sphere. *Annals of Statistics*, 2(6):1201–1225, 1974.
22. F.L. Bookstein. *The Measurement of Biological Shape and Shape Change*, volume 24 of *Lecture Notes in Biomathematics*. Springer-Verlag, 1978.
23. T. Brox, J. Weickert, B. Burgeth, and P. Mrázek. Nonlinear structure tensors. *Image and Vision Computing*, 24(1):41–55, 2006.
24. Anders Brun. *Manifolds in Image Science and Visualization*. PhD thesis, Linköping University, 2007. Linköping Studies in Science and Technology Dissertations No 1157.
25. J. Burbea and C.R. Rao. Entropy differential metric, distance and divergence measures in probability spaces: a unified approach. *Journal of Multivariate Analysis*, 12:575–596, 1982.
26. M. Calvo and J.M. Oller. An explicit solution of information geodesic equations for the multivariate normal model. *Statistics and Decisions*, 9:119–138, 1991.
27. C. Chéfd'hotel, D. Tschumperlé, R. Deriche, and O. Faugeras. Constrained flows of matrix-valued functions: Application to diffusion tensor regularization. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Proc. of ECCV 2002*, volume 2350 of *LNCS*, pages 251–265. Springer Verlag., 2002.
28. C. Chéfd'hotel, D. Tschumperlé, R. Deriche, and O. Faugeras. Regularizing flows for constrained matrix-valued images. *J. Math. Imaging and Vision*, 20(1-2):147–162, January - March 2004.
29. D. Cohen-Steiner and J.M. Morvan. Restricted delaunay triangulations and normal cycle. In *Proceedings of the nineteenth annual symposium on Computational geometry*, pages 312–321, 2003.
30. D.L. Collins, A.P. Zijdenbos, V. Kollokian, J.G. Sled, N.J. Kabani, C.J. Holmes, and A.C. Evans. Design and construction of a realistic digital brain phantom. *IEEE Transactions on Medical Imaging*, 17(3):463–468, June 1998.
31. R.W.R. Darling. Martingales on non-compact manifolds: maximal inequalities and prescribed limits. *Annales de l'institut Poincaré - Probabilités et Statistiques*, 32(4):431–454, 1996.
32. B. Davis, P.Th. Fletcher, E. Bullitt, and S. Joshi. Population shape regression from random design data. In *Proc. of ICCV'07*, 2007.
33. J.-P. Dedieu, G. Malajovich, and P. Priouret. Newton method on Riemannian manifolds: Covariant alpha-theory. *IMA Journal of Numerical Analysis*, 23:395–419, 2003.
34. M. do Carmo. *Riemannian Geometry*. Mathematics. Birkhäuser, Boston, Basel, Berlin, 1992.
35. I.L. Dryden, A. Koloydenko, and D. Zhou. Non-Euclidean statistics for covariance matrices with application to diffusion tensor imaging. *Submitted*, 2008.
36. I.L. Dryden and K.V. Mardia. Theoretical and distributional aspects of shape analysis. In *Probability Measures on Groups, X (Oberwolfach, 1990)*, pages 95–116, New York, 1991. Plenum.
37. I.L. Dryden and K.V. Mardia. *Statistical Shape Analysis*. John Wiley, Chichester, 1998.

38. Stanley Durrleman, Xavier Pennec, Alain Trouvé, and Nicholas Ayache. A forward model to build unbiased atlases from curves and surfaces. In X. Pennec and S. Joshi, editors, *Proc. of the International Workshop on the Mathematical Foundations of Computational Anatomy (MFCA-2008)*, September 2008.
39. Stanley Durrleman, Xavier Pennec, Alain Trouvé, and Nicholas Ayache. Sparse approximation of currents for statistics on curves and surfaces. In Dimitris Metaxas, Leon Axel, Gabor Szekely, and Gabor Fichtinger, editors, *Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI)*, LNCS, New-York, USA, September 2008. Springer. In press.
40. Stanley Durrleman, Xavier Pennec, Alain Trouvé, Paul Thompson, and Nicholas Ayache. Inferring brain variability from diffeomorphic deformations of currents: an integrative approach. *Medical Image Analysis*, 12/5(12):626–637, 2008.
41. R.B. Fisher D.W. Eggert, A. Lorusso. Estimating 3d rigid body transformations: A comparison of four major algorithms. *Machine Vision Applications, Special Issue on Performance Characteristics of Vision Algorithms*, 9(5/6):272–290, 1997.
42. A. Edelman, T. Arias, and S.T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal of Matrix Analysis and Applications*, 20(2):303–353, 1998.
43. M. Emery. *Stochastic Calculus in Manifolds*. Springer, Berlin, 1989.
44. M. Emery and G. Mokobodzki. Sur le barycentre d’une probabilité dans une variété. In M. Yor J. Azema, P.A. Meyer, editor, *Séminaire de probabilités XXV*, volume 1485 of *Lect. Notes in Math.*, pages 220–233. Springer-Verlag, 1991.
45. A. C. Evans, D. L. Collins, S. R. Mills, E. D. Brown, R. L. Kelly, and T. M. Peters. 3D statistical neuroanatomical models from 305 MRI volumes. In *Proc. IEEE-Nuclear Science Symposium and Medical Imaging Conference*, pages 1813–1817, 1993.
46. D. Facon, A. Ozanne, P. Fillard, J.-F. Lepeintre, C. Tournoux-Facon, and D. Ducreux. MR diffusion tensor imaging and fiber tracking in spinal cord compression. *American Journal of Neuroradiology (AJNR)*, 26:1587–1594, 2005.
47. Pierre Fillard, Vincent Arsigny, Nicholas Ayache, and Xavier Pennec. A Riemannian framework for the processing of tensor-valued images. In Ole Fogh Olsen, Luc Florak, and Arjan Kuijper, editors, *Deep Structure, Singularities, and Computer Vision (DSSCV)*, number 3753 in LNCS, pages 112–123. Springer Verlag, June 2005.
48. Pierre Fillard, Vincent Arsigny, Xavier Pennec, and Nicholas Ayache. Clinical DT-MRI estimation, smoothing and fiber tracking with log-Euclidean metrics. In *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI 2006)*, pages 786–789, Crystal Gateway Marriott, Arlington, Virginia, USA, April 2006.
49. Pierre Fillard, Vincent Arsigny, Xavier Pennec, and Nicholas Ayache. Clinical DT-MRI estimation, smoothing and fiber tracking with log-Euclidean metrics. *IEEE Transactions on Medical Imaging*, 26(11):1472–1482, November 2007. PMID: 18041263.
50. Pierre Fillard, Vincent Arsigny, Xavier Pennec, Kiralee M. Hayashi, Paul M. Thompson, and Nicholas Ayache. Measuring brain variability by extrapolating sparse tensor fields measured on sulcal lines. *Neuroimage*, 34(2):639–650, January 2007. Also as INRIA Research Report 5887, April 2006. PMID: 17113311.
51. Pierre Fillard, Vincent Arsigny, Xavier Pennec, Paul M. Thompson, and Nicholas Ayache. Extrapolation of sparse tensor fields: Application to the modeling of brain variability. In Gary Christensen and Milan Sonka, editors, *Proc. of Information*

- Processing in Medical Imaging 2005 (IPMI'05)*, volume 3565 of *LNCS*, pages 27–38, Glenwood springs, Colorado, USA, July 2005. Springer. PMID: 17354682.
52. Pierre Fillard, Xavier Pennec, Paul M. Thompson, and Nicholas Ayache. Evaluating brain anatomical correlations via canonical correlation analysis of sulcal lines. *NeuroImage*, 2008. Accepted for publication.
  53. P. Thomas Fletcher and Sarang C. Joshi. Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In *Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis, ECCV 2004 Workshops CVAMIA and MMBIA, Prague, Czech Republic, May 15, 2004*, volume 3117 of *LNCS*, pages 87–98. Springer, 2004.
  54. P.T. Fletcher, S. Joshi, C. Lu, and S Pizer. Gaussian distributions on Lie groups and their application to statistical shape analysis. In Chris Taylor and Alison Noble, editors, *Proc. of Information Processing in Medical Imaging (IPMI'2003)*, volume 2732 of *LNCS*, pages 450–462. Springer, 2003.
  55. M. Fleute and S. Lavallée. Building a complete surface model from sparse data using statistical shape models: Application to computer assisted knee surgery. In Springer, editor, *Proc. of Medical Image Computing and Computer-Assisted Intervention (MICCAI'98)*, volume 1496 of *LNCS*, pages 879–887, 1998.
  56. W. Förstner and B. Moonen. A metric for covariance matrices. In F. Krumm and V. S. Schwarz, editors, *Qua vadis geodesia...? Festschrift for Erik W. Grafarend on the occasion of his 60th birthday*, number 1999.6 in Tech. Report of the Dpt of Geodesy and Geoinformatics, pages 113–128. Stuttgart University, 1999.
  57. M. Fréchet. L'intégrale abstraite d'une fonction abstraite d'une variable abstraite et son application à la moyenne d'un élément aléatoire de nature quelconque. *Revue Scientifique*, pages 483–512, 1944.
  58. M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut Henri Poincaré*, 10:215–310, 1948.
  59. S. Gallot, D. Hulin, and J. Lafontaine. *Riemannian Geometry*. Springer Verlag, 2nd edition edition, 1993.
  60. R.V. Gamkrelidze, editor. *Geometry I*, volume 28 of *Encyclopaedia of Mathematical Sciences*. Springer Verlag, 1991.
  61. G. Gerig, R. Kikinis, O. Kübler, and F.A. Jolesz. Nonlinear anisotropic filtering of MRI data. *IEEE Transactions on Medical Imaging*, 11(2):221–232, June 1992.
  62. C. Gramkow. On averaging rotations. *International Journal of Computer Vision*, 42(1-2):7–16, April/May 2001.
  63. Sébastien Granger and Xavier Pennec. Statistiques exactes et approchées sur les normales aléatoires. Research report RR-4533, INRIA, 2002.
  64. U. Grenander. *Probabilities on Algebraic Structures*. Wiley, 1963.
  65. U. Grenander, M.I. Miller, and A. Srivastava. Hilbert-schmidt lower bounds for estimators on matrix Lie groups for ATR. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 20(8):790–802, 1998.
  66. Alexander Grigor'yan. Heat kernels on weighted manifolds and applications. In J. Jorgenson and L. Walling, editors, *The Ubiquitous Heat Kernel*, volume 398 of *Contemporary Mathematics*, pages 91–190. AMS, 2006.
  67. Ernst Hairer, Ch. Lubich, and Gerhard Wanner. *Geometric numerical integration : structure preserving algorithm for ordinary differential equations*, volume 31 of *Springer series in computational mathematics*. Springer, 2002.
  68. S. Helgason. *Differential Geometry, Lie groups, and Symmetric Spaces*. Academic Press, 1978.
  69. Uwe Helmke and J. B. Moore. *Optimization and Dynamical Systems*. Communication and Control Engineering Series. Springer, 1994.

70. H. Hendricks. A Cramer-Rao type lower bound for estimators with values in a manifold. *Journal of Multivariate Analysis*, 38:245–261, 1991.
71. P. Huber. *Robust Statistics*. John Wiley, New York, 1981.
72. Shun ichi Amari and Hiroshi Nagaoka. *Methods of Information Geometry*. Number 191 in Translations of Mathematical Monographs. American Mathematical Society, 2000.
73. Sarang C. Joshi and Michael I. Miller. Landmark matching via large deformation diffeomorphisms. *IEEE Trans. Image Processing*, 9(8):1357–1370, 2000.
74. P.E. Jupp and K.V. Mardia. A unified view of the theory of directional statistics, 1975-1988. *International Statistical Review*, 57(3):261–294, 1989.
75. H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications in Pure and Applied Mathematics*, 30:509–541, 1977.
76. Robert E. Kass and Paul W Vos. *Geometric Foundations of Asymptotic Inference*. Wiley series in Probability and Statistics. John Wiley & Sons, 1997.
77. C. Y. Kaya and J. L. Noakes. Geodesics and an optimal control algorithm. In *Proceedings of the 36th IEEE Conference on Decision and Control, San Diego, CA, U.S.A.*, pages 4918–4919, 1997.
78. D.G. Kendall. A survey of the statistical theory of shape (with discussion). *Statistical Science*, 4:87–120, 1989.
79. M.G. Kendall and P.A.P. Moran. *Geometrical probability*. Number 10 in Griffin's statistical monographs and courses. Charles Griffin & Co. Ltd., 1963.
80. W.S. Kendall. Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence. *Proc. London Math. Soc.*, 61(2):371–406, 1990.
81. W.S. Kendall. Convexity and the hemisphere. *Journal of the London Mathematical Society*, 43(2):567–576, 1991.
82. J.T. Kent. *The art of Statistical Science*, chapter 10 : New Directions in Shape Analysis, pages 115–127. John Wiley & Sons, 1992. K.V. Mardia, ed.
83. E. Klassen, A. Srivastava, W. Mio, and S. Joshi. Analysis of planar shapes using geodesic path on shape spaces. *IEEE Trans. on PAMI*, 26(3):372–383, 2004.
84. W. Klingenberg. *Riemannian Geometry*. Walter de Gruyter, Berlin, New York, 1982.
85. S. Kobayashi and K. Nomizu. *Foundations of differential geometry, vol. II*. Number 15 in Interscience tracts in pure and applied mathematics. John Wiley & Sons, 1969.
86. H. Le and D.G. Kendall. The Riemannian structure of Euclidean shape space: a novel environment for statistics. *Annals of Statistics*, 21:1225–1271, 1993.
87. D. Le Bihan, J.-F. Mangin, C. Poupon, C.A. Clark, S. Pappata, N. Molko, and H. Chabriat. Diffusion tensor imaging: Concepts and applications. *Journal Magnetic Resonance Imaging*, 13(4):534–546, 2001.
88. G. Le Goualher, E. Procyk, D. Collins, R. Venugopal, C. Barillot, and A. Evans. Automated extraction and variability analysis of sulcal neuroanatomy. *IEEE Transactions on Medical Imaging*, 18(3):206–217, 1999.
89. Ch. Lenglet, M. Rousson, R. Deriche, and O. Faugeras. Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor MRI processing. *Journal of Mathematical Imaging and Vision*, 25(3):423–444, October 2006.
90. Natasha Lepore, Caroline Brun, Yi-Yu Chou, Agatha D. Lee, Marina Barysheva, Xavier Pennec, Katie McMahon, Matthew Meredith, Greig I. de Zubicaray, Margaret J. Wright, Arthur W. Toga, and Paul M. Thompson. Best individual template selection from deformation tensor minimization. In *Proc. of the 2008 IEEE*

- Int. Symp. on Biomedical Imaging: From Nano to Macro (ISBI'08), Paris, France, May 14-17*, pages 460–463, 2008.
91. Miroslav Lovrić and Maung Min-Oo. Multivariate normal distributions parametrized as a Riemannian symmetric space. *Journal of Multivariate Analysis*, 74(1):36–48, 2000.
  92. R. Mahony and R. Manton. The geometry of the Newton method on non-compact Lie groups. *Journal of Global Optimization*, 23:309–327, 2002.
  93. J.-F. Mangin, D. Riviere, A. Cachia, E. Duchesnay, Y. Cointepas, D. Papadopoulos-Orfanos, D. L. Collins, A. C. Evans, and J. Régis. Object-based morphometry of the cerebral cortex. *IEEE Transactions on Medical Imaging*, 23(8):968–982, August 2004.
  94. J.-F. Mangin, D. Rivière, A. Cachia, E. Duchesnay, Y. Cointepas, D. Papadopoulos-Orfanos, P. Scifo, T. Ochiai, F. Brunelle, and J. Régis. A framework to study the cortical folding patterns. *NeuroImage*, 23(Supplement 1):S129–S138, 2004.
  95. K.V. Mardia. Directional statistics and shape analysis. *Journal of applied Statistics*, 26(949-957), 1999.
  96. K.V. Mardia and P.E. Jupp. *Directional statistics*. Wiley, Chichester, 2000.
  97. J. Mazziotta, A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, R. Woods, T Paus, G. Simpson, B. Pike, C. Holmes, L. Collins, P. Thompson, D. MacDonald, M. Iacoboni, T. Schormann, K. Amunts, N. Palomero-Gallagher, S. Geyer, L. Parsons, K. Narr, N. Kabani, G. Le Goualher, D. Boomsma, T. Cannon, R. Kawashima, and B. Mazoyer. A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (ICBM). *Philos Trans R Soc Lond B Biol Sci*, 356:1293–1322, 2001.
  98. G Medioni, M.-S. Lee, and Ch.-K. Tang. *A Computational Framework for Segmentation and Grouping*. Elsevier, 2000.
  99. E. Meijering. A chronology of interpolation: From ancient astronomy to modern signal and image processing. *Proceedings of the IEEE*, 90(3):319–342, March 2002.
  100. Peter W. Michor and David Mumford. Riemannian geometries on spaces of plane curves. *J. Eur. Math. Soc. (JEMS)*, 8:1–48, 2006. ESI Preprint 1425, arXiv:math.DG/0312384.
  101. M.I. Miller, A. Trouvé, and L. Younes. On the metrics and Euler-Lagrange equations of computational anatomy. *Annual Re-view of Biomedical Engineering*, pages 375–405, 2003.
  102. M.I. Miller and L. Younes. Group actions, homeomorphisms, and matching: A general framework. *International Journal of Computer Vision*, 41(1/2):61–84, 2001.
  103. Michael I. Miller, Alain Trouvé, and Laurent Younes. Geodesic shooting for computational anatomy. *Journal of Mathematical Imaging and Vision*, 2006.
  104. M. Moakher. Means and averaging in the group of rotations. *SIAM Journal of Matrix Analysis and Applications*, 24(1):1–16, 2002.
  105. M. Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal of Matrix Analysis and Applications*, 26(3):735–747, 2005.
  106. B. Mohammadi, H. Borouchaki, and P.L. George. Delaunay mesh generation governed by metric specifications. Part II: applications. *Finite Elements in Analysis and Design*, pages 85–109, 1997.
  107. Bill Moran, Sofia Suvorova, and Stephen Howard. Sensor management for radar: a tutorial. In *Advances in Sensing with Security Applications 17–30 July 2005, Il Ciocco, Italy*. NATO Advanced Study Institute, 2005.

108. K. Nomizu. Invariant affine connections on homogeneous spaces. *American J. of Math.*, 76:33–65, 1954.
109. J.M. Oller and J.M. Corcuera. Intrinsic analysis of statistical estimation. *Annals of Statistics*, 23(5):1562–1581, 1995.
110. B. Owren and B. Welfert. The Newton iteration on Lie groups. *BIT Numerical Mathematics*, 40(1):121–145, 2000.
111. Xavier Pennec. *L'incertitude dans les problèmes de reconnaissance et de recalage – Applications en imagerie médicale et biologie moléculaire*. Thèse de sciences (Ph.D thesis), Ecole Polytechnique, Palaiseau (France), December 1996.
112. Xavier Pennec. Computing the mean of geometric features - application to the mean rotation. Research Report RR-3371, INRIA, March 1998.
113. Xavier Pennec. Probabilities and statistics on Riemannian manifolds: Basic tools for geometric measurements. In A.E. Cetin, L. Akarun, A. Ertuzun, M.N. Gurcan, and Y. Yardimci, editors, *Proc. of Nonlinear Signal and Image Processing (NSIP'99)*, volume 1, pages 194–198, June 20-23, Antalya, Turkey, 1999. IEEE-EURASIP.
114. Xavier Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, July 2006. A preliminary appeared as INRIA RR-5093, January 2004.
115. Xavier Pennec. Left-invariant Riemannian elasticity: a distance on shape diffeomorphisms? In X. Pennec and S. Joshi, editors, *Proc. of the International Workshop on the Mathematical Foundations of Computational Anatomy (MFCA-2006)*, pages 1–13, 2006.
116. Xavier Pennec. *Statistical Computing on Manifolds for Computational Anatomy*. Habilitation à diriger des recherches, Université Nice Sophia-Antipolis, December 2006.
117. Xavier Pennec, Pierre Fillard, and Nicholas Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, January 2006. A preliminary version appeared as INRIA Research Report 5255, July 2004.
118. Xavier Pennec, Charles R.G. Guttmann, and Jean-Philippe Thirion. Feature-based registration of medical images: Estimation and validation of the pose accuracy. In *Proc. of First Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI'98)*, volume 1496 of LNCS, pages 1107–1114, Cambridge, USA, October 1998. Springer Verlag.
119. Xavier Pennec, Radu Stefanescu, Vincent Arsigny, Pierre Fillard, and Nicholas Ayache. Riemannian elasticity: A statistical regularization framework for non-linear registration. In J. Duncan and G. Gerig, editors, *Proceedings of the 8th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention - MICCAI 2005, Part II*, volume 3750 of LNCS, pages 943–950, Palm Springs, CA, USA, October 26-29, 2005. Springer Verlag. PMID: 16686051.
120. Xavier Pennec and Jean-Philippe Thirion. A framework for uncertainty and validation of 3D registration methods based on points and frames. *Int. Journal of Computer Vision*, 25(3):203–229, December 1997.
121. P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, 12(7):629–639, 1990.
122. J. Picard. Barycentres et martingales sur une variété. *Annales de l'institut Poincaré - Probabilités et Statistiques*, 30(4):647–702, 1994.
123. H. Poincaré. *Calcul des probabilités*. 2nd edition, Paris, 1912.

124. J. M. Provenzale, S. Mukundan, and D. P. Barboriak. Diffusion-weighted and perfusion MR imaging for brain tumor characterization and assessment of treatment response. *Radiology*, 239(3):632 – 649, 2006.
125. K.T. Rajamani, S.C. Joshi, and M.A Styner. Bone model morphing for enhanced surgical visualization. In IEEE, editor, *Proc of IEEE Symp. on Biomedical Imaging: Nano to Macro (ISBI) 2004*, volume 2, pages 1255–1258, April 2004.
126. P.J. Rousseeuw and A.M. Leroy. *Robust Regression and Outliers Detection*. Wiley series in prob. and math. stat. J. Wiley and Sons, 1987.
127. M. Rovaris, A. Gass, R. Bammer, S.J. Hickman, O. Ciccarelli, D.H. Miller, and M. Filippi. Diffusion MRI in multiple sclerosis. *Neurology*, 65:1526–1532, 2005.
128. J. Sijbers, A.J. den Dekker, P. Scheunders, and D. Van Dyck. Maximum likelihood estimation of Rician distribution parameters. *TMI*, 17(3), June 1998.
129. L.T. Skovgaard. A Riemannian geometry of the multivariate normal model. *Scand. J. Statistics*, 11:211–223, 1984.
130. C.G. Small. *The Statistical Theory of Shapes*. Springer series in statistics. Springer, 1996.
131. M. Spivak. *Differential Geometry*, volume 1. Publish or Perish, Inc., 2nd edition, 1979.
132. G. Subsol, J.-Ph. Thirion, and N. Ayache. A scheme for automatically building 3D morphometric anatomical atlases: application to a skull atlas. *Medical Image Analysis*, 2(1):37–60, 1998.
133. J. Talairach and P. Tournoux. *Co-Planar Stereotaxic Atlas of the Human Brain: 3-dimensional Proportional System : an Approach to Cerebral Imaging*. Thieme Medical Publishers, New York, 1988.
134. P. Thévenaz, T. Blu, and M. Unser. Interpolation revisited. *IEEE Transactions on Medical Imaging*, 19(7):739–758, July 2000.
135. P.M. Thompson, D. MacDonald, M.S. Mega, C.J. Holmes, A.C. Evans, and A.W. Toga. Detection and mapping of abnormal brain structure with a probabilistic atlas of cortical surfaces. *Journal of Computer Assisted Tomography*, 21(4):567–581, 1977.
136. P.M. Thompson, M.S. Mega, R.P. Woods, C.I. Zoumalan, C.J. Lindshield, R.E. Blanton, J. Moussai, C.J. Holmes, J.L. Cummings, and A.W. Toga. Cortical change in alzheimer’s disease detected with a disease-specific population-based brain atlas. *Cerebral Cortex*, 11(1):1–16, January 2001.
137. Alain Trouvé. Diffeomorphisms groups and pattern matching in image analysis. *International Journal of Computer Vision*, 28(3):213–221, 1998.
138. D. Tschumperlé. *PDE-Based Regularization of Multivalued Images and Applications*. PhD thesis, University of Nice-Sophia Antipolis, dec 2002.
139. M. Vaillant and J. Glaunes. Surface matching via currents. In *Proc. of IPMI’05*, pages 381–392, 2005.
140. M. Vaillant, M.I. Miller, L. Younes, and A. Trouvé. Statistics on diffeomorphisms via tangent space representations. *NeuroImage*, 23(Supp. 1):S161–S169, 2004.
141. M. Vaillant, A. Qiu, J. Glaunès, and M. Miller. Diffeomorphic metric surface mapping in subregion of the superior temporal gyrus. *NeuroImage*, 34(3):1149–1159, 2007.
142. Z. Wang and B.C. Vemuri. DTI segmentation using an information theoretic tensor dissimilarity measure. *IEEE Trans. on Medical Imaging*, 24(10):1267–1277, October 2005.
143. Z. Wang, B.C. Vemuri, Y. Chen, and T. Mareci. A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex DWI. *IEEE Trans. on Medical Imaging*, 2004.



144. J. Weickert and T. Brox. Diffusion and regularization of vector- and matrix-valued images. In M.Z. Nashed and O. Scherzer, editors, *Inverse Problems, Image Analysis, and Medical Imaging.*, volume 313 of *Contemporary Mathematics*, pages 251–268, Providence, 2002. AMS.
145. J. Weickert and H. Hagen, editors. *Visualization and Processing of Tensor Fields.* Mathematics and Visualization. Springer, 2006.
146. C.F. Westin, S.E. Maier, H. Mamata, A. Nabavi, F.A. Jolesz, and R. Kikinis. Processing and visualization for diffusion tensor MRI. *Medical Image Analysis*, 6(2):93–108, June 2002.
147. Laurent Younes. Computable elastic distances between shapes. *SIAM Journal on Applied Mathematics*, 58(2):565–586, 1998.