Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements

Xavier Pennec

This paper will appear in the *Journal of Mathematical Imaging and Vision* (Springer).
The original publication will be available at www.springerlink.com.

Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements

Xavier Pennec
INRIA - Epidaure / Asclepios project-team
2004 Route des Lucioles, BP 93
F-06902 Sophia Antipolis Cedex, France
Email: Xavier.Pennec@sophia.inria.fr

Abstract

In medical image analysis and high level computer vision, there is an intensive use of geometric features like orientations, lines, and geometric transformations ranging from simple ones (orientations, lines, rigid body or affine transformations, etc.) to very complex ones like curves, surfaces, or general diffeomorphic transformations. The measurement of such geometric primitives is generally noisy in real applications and we need to use statistics either to reduce the uncertainty (estimation), to compare observations, or to test hypotheses. Unfortunately, even simple geometric primitives often belong to manifolds that are not vector spaces. In previous works [1, 2], we investigated invariance requirements to build some statistical tools on transformation groups and homogeneous manifolds that avoids paradoxes. In this paper, we consider finite dimensional manifolds with a Riemannian metric as the basic structure. Based on this metric, we develop the notions of mean value and covariance matrix of a random element, normal law, Mahalanobis distance and χ^2 law. We provide a new proof of the characterization of Riemannian centers of mass and an original gradient descent algorithm to efficiently compute them. The notion of Normal law we propose is based on the maximization of the entropy knowing the mean and covariance of the distribution. The resulting family of pdfs spans the whole range from uniform (on compact manifolds) to the point mass distribution. Moreover, we were able to provide tractable approximations (with their limits) for small variances which show that we can effectively implement and work with these definitions.

1 Introduction

To represent the results of a random experiment, one theoretically consider a probability measure on the space of all events. Although this probabilized space contains all the information about the random experiment, one often have access only to some measurements depending of the outcome of the experiment. The mathematical way to formalize this is to investigate *random variables* or *observables* which are maps from the probabilized space into $\mathbb{R}$. One usually further simplify by restricting to random variables that have a *probability density function* (pdf).

However, from a computational point of view, the pdf is still too informative and we have to restrict the measurements to a few numeric characteristics of a random variable. Thus, one usually approximate a unimodal pdf by a central value and a dispersion value around it. The most used central value is the *mean value* or *expectation* of the random variable: $\overline{x} = \mathbf{E}\left[\, \boldsymbol{x} \,\right] = \int \boldsymbol{x} \, \mathrm{dPr} = \int_{\mathbb{R}} y \, p_{\boldsymbol{x}}(y) \, dy$. The corresponding dispersion value is the *variance* $\sigma_{\boldsymbol{x}}^2 = \mathbf{E}\left[\, (\boldsymbol{x} - \overline{x})^2 \,\right]$.

In real problems, we can have several simultaneous measurements of the same random experiment. If we arrange these n random variables $\boldsymbol{x}_i$ into a vector $\boldsymbol{x} = (\boldsymbol{x}_1 \dots \boldsymbol{x}_n)$, we obtain a *random vector*. As the expectation is a linear operator, it is easily generalized to vector or matrix functions in order to define the mean value and the covariance matrix of a random vector: $\Sigma_{\boldsymbol{x}\boldsymbol{x}} = \mathbf{E}\left[\, (\boldsymbol{x} - \overline{x}) \, (\boldsymbol{x} - \overline{x})^{\mathrm{T}} \,\right] = \int (y - \overline{x}) \, (y - \overline{x})^{\mathrm{T}} \, p_{\boldsymbol{x}}(y) \, dy$. If one has to assume a probability distribution, the Gaussian distribution is usually well adapted, as it is completely determined by the mean and the covariance. It is moreover the entropy maximizing distribution knowing only these moments. Then, one can use a statistical distance between distributions such as the Mahalanobis distance and the associated statistical tests.

The problem we investigate in this article is to generalize this framework to measurements in finite dimensional Riemannian manifolds instead of measurements in a vector space. Examples of manifolds we routinely use in medical imaging applications are 3D rotations, 3D rigid transformations, frames (a 3D point and an orthonormal trihedron), semi- or non-oriented frames (where 2 (resp. 3) of the trihedron unit vectors are given up to their sign) [3, 4], oriented or directed points [5, 6], positive definite symmetric matrices coming from diffusion tensor imaging [7, 8, 9, 10, 11] or from variability measurements [12]. We have already shown in [13, 2] that this is not an easy problem and that some paradoxes can arise. In particular, we cannot generalize the expectation to give a mean value since it would be an integral with value in the manifold: a new definition of mean is needed, which implies revisiting an important part of the theory of statistics.

Statistical analysis on manifolds is a relatively new domain at the confluent of several mathematical and application domains. Its goal is to study statistically geometric object living in differential manifolds. It is linked to the theory of statistical manifolds [14, 15], which aims at providing a Riemannian structure to the the space of parameters of statistical distribution. However, the targeted geometrical objects are usually different. Directional statistics [16, 17, 18, 19] 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. what remains invariant under the action of a given group of transformation (usually rigid body ones or similarities). The statistics on these shape spaces [20, 21, 22, 23] raised the need for intrinsic tools. However, the link between the tools developed in these works, 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, [24] 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 [25, 26, 27, 28, 29, 30]. On the applied mathematics and computer science side, people get interested in computing and optimizing in specific manifolds, like rotations and rigid body transformations [4, 31, 32, 33, 34], Stiefel and Grassmann manifolds [35], etc.

Over the last years, several groups attempted to federate some of the above approaches in a general statistical framework, with different objectives in mind. For instance, [36] and [15] aimed at characterizing the performances of statistical parametric estimators, like the bias and the mean square error. [36] considered extrinsic statistics, based on the Euclidean distance of the embedding space, while [15] considered the intrinsic Riemannian distance, and refined the Cramer-Rao lower bound using bounds on the sectional curvature of the manifold. In [37, 38], 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. Here, in view of computer vision and medical image analysis applications, our concern is quite different: we aim 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 are interested in small sample sizes rather than large one, and we prefer to obtain approximations rather than bounds on the quality of the estimation. Thus, one of our special interest is 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 all cases, 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 of the theory, we rely in this paper only on intrinsic properties of the Riemannian manifold, thus excluding methods based on the embedding of the manifold in an ambient Euclidean space.

We review in Section 2 some basic notions of differential and Riemannian geometry that will be needed afterward. This synthesis was inspired from [39, chap. 9], [40, 41, 42], and the reader can refer to these books to find more details. In the remaining of the paper, we consider that our Riemannian manifold is connected and geodesically complete. We first detail in Section 3 the measure induced by the Riemannian metric on the manifold, which allows to define probability density functions, in particular the uniform one. Then, we turn in Section 4 to the expectation of a random point. We provide a quite comprehensive survey of the definitions that have been proposed. Among them, we focus on the Karcher and Fréchet means, defined as the set of points minimizing locally or globally the variance (the expected Riemannian distance). We provide a new proof of the barycentric characterization theorem and an original Gauss-Newton gradient descent algorithm to practically compute the mean. Once the mean value is determined, one can easily define the covariance matrix of a random point (and possibly higher order moments) using the exponential map at the mean point (Section 4). To generalize the Gaussian distribution, we propose in Section 6 a new family of distributions based on a maximum entropy approach. Under some reasonable hypotheses, we show that is amounts to take a truncated Gaussian distribution in the exponential map at the mean point. We illustrate the properties of this pdf family on the circle, and provide computationally tractable approximations for concentrated distributions. Last but not least, we investigate in Section 7 the generalization of the Mahalanobis distance and the χ^2 law. A careful analysis shows that, with our definition of the generalized Gaussian, the χ^2 law remains independent of the variance and of the manifold curvature, up to the order 3. This demonstrate that the whole framework is computationally sound and particularly consistent.

2 Differential geometry background

2.1 Riemannian metric, distance and geodesics

In the geometric framework, one specifies the structure of a manifold $\mathcal{M}$ by a *Riemannian metric*. This is a continuous collection of dot products $\langle \, . \mid . \, \rangle_{\mathrm{x}}$ on the tangent space $T_{\mathrm{x}}\mathcal{M}$ at each point x of the manifold. A local coordinate system $x = (x^1, \dots x^n)$ induces a basis $\frac{\partial}{\partial x} = (\partial_1, \dots \partial_n)$ of the tangent spaces (∂_i is a shorter notation for $\partial/\partial x^i$). Thus, we can express the metric in this basis by a symmetric positive definite matrix $G(x) = [g_{ij}(x)]$ where each element is given by the dot product of the tangent vector to the coordinate curves: $g_{ij}(x) = \langle \, \partial_i \mid \partial_j \, \rangle$. This matrix is called the *local representation of the Riemannian metric* in the chart x and the dot products of two vectors v and w in $T_x\mathcal{M}$ is now $\langle \, v \mid w \, \rangle_x = v^{\mathrm{T}} \, G(x) \, w$. The matrix $G(x)$ is called the *local representation of the Riemannian metric* in the chart x.

If we consider a curve $\gamma(t)$ on the manifold, we can compute at each point its instantaneous speed vector $\dot{\gamma}(t)$ and its norm, the instantaneous speed. To compute the length of the curve, we can proceed as usual by integrating this value along the curve:

$$\mathcal{L}_a^b(\gamma) = \int_a^b \|\dot{\gamma}(t)\| \, dt = \int_a^b \left(\langle \, \dot{\gamma}(t) \mid \dot{\gamma}(t) \, \rangle_{\gamma(t)} \right)^{\frac{1}{2}} dt$$

The Riemannian metric is the intrinsic way of measuring length on a manifold. The extrinsic way is to consider the manifold as embedded in a larger vector space E (think for instance to the sphere $\mathcal{S}_2$ in $\mathbb{R}^3$) and compute the length of a curve in $\mathcal{M}$ as for any curve in E. In this case, the corresponding Riemannian metric is the restriction of the dot product of E onto the tangent space at each point of the manifold. By Whitney's theorem, there always exists such an embedding for a large enough vector space E ($\dim(E) \leq 2\dim(\mathcal{M}) + 1$).

To obtain a distance between two points of a connected Riemannian manifold, we simply have to take the minimum length among the smooth curves joining these points:

$$\mathrm{dist}(\mathrm{x}, \mathrm{y}) = \min_{\gamma} \mathcal{L}(\gamma) \qquad \text{with} \qquad \gamma(0) = \mathrm{x} \quad \text{and} \quad \gamma(1) = \mathrm{y} \tag{1}$$

The curves realizing this minimum for any two points of the manifold are called geodesics[1]. Let $[g^{ij}] = [g_{ij}]^{(-1)}$ be the inverse of the metric matrix (in a given coordinate system x) and $\Gamma_{jk}^i = \frac{1}{2} g^{im} \left(\partial_k g_{mj} + \partial_j g_{mk} - \partial_m g_{jk} \right)$ the Christoffel symbols (using Einstein summation convention that implicit sum upon each index that appear up and down in the formula). The calculus of variations shows the geodesics are the curves satisfying the following second order differential system (in the chart $x = (x_1, \dots x_n)$):

$$\ddot{\gamma}^i + \Gamma_{jk}^i \dot{\gamma}^j \dot{\gamma}^k = 0$$

The manifold is said to be *geodesically complete* if the definition domain of all geodesics can be extended to $\mathbb{R}$. This means that the manifold has no boundary nor any singular point that we can reach in a finite time (for instance, $\mathbb{R}^n - \{0\}$ with the usual metric is not geodesically complete, but $\mathbb{R}^n$ or $\mathcal{S}_n$ are). As an important consequence, the Hopf-Rinow-De Rham theorem state that such a manifold is complete for the induced distance (equation 1), and that there always exist at least one minimizing geodesic between any two points of the manifold (i.e. which length is the distance between the two points). *From now on, we will assume that the manifold is geodesically complete.*

[1] In facts, geodesics are defined as the critical points of the energy functional $\mathcal{E}(\gamma) = \frac{1}{2} \int_a^b \|\partial_\gamma\|^2 \, dt$. It turns out that they also optimize the length functional but they are moreover parameterized proportionally to arc-length.

2.2 Exponential map and cut locus

From the theory of second order differential equations, we know that there exists one and only one geodesic $\gamma_{(\mathrm{x},\partial_v)}$ going through the point $\mathrm{x} \in \mathcal{M}$ at $t = 0$ with tangent vector $\partial_v \in T_\mathrm{x}\mathcal{M}$. This geodesic is theoretically defined in a sufficiently small interval around zero but since we the manifold is geodesically complete, its definition domain can be extended to $\mathbb{R}$. Thus, the point $\gamma_{(\mathrm{x},\partial_v)}(t)$ is defined for all vector $\partial_v \in T_\mathrm{x}\mathcal{M}$ and all parameter t. The *exponential map* maps each vector ∂_v to the point of the manifold reached in a unit time:

$$\exp_\mathrm{x} \quad : \quad \begin{array}{ccc} T_\mathrm{x}\mathcal{M} & \longrightarrow & \mathcal{M} \\ \partial_v & \longmapsto & \exp_\mathrm{x}(\partial_v) = \gamma_{(\mathrm{x},\partial_v)}(1) \end{array}$$

This function realizes a local diffeomorphism from a sufficiently small neighborhood of 0 in $T_\mathrm{x}\mathcal{M}$ into a neighborhood of the point $\mathrm{x} \in \mathcal{M}$. We denote by $\log_x = \exp_\mathrm{x}^{(-1)}$ the inverse map or simply $\overrightarrow{\mathrm{xy}} = \log_\mathrm{x}(\mathrm{y})$. In this chart, the geodesics going through x are the represented by the lines going through the origin: $\log_\mathrm{x} \gamma_{(\mathrm{x},\overrightarrow{\mathrm{xy}})}(t) = t\,\overrightarrow{\mathrm{xy}}$. Moreover, the distance with respect to the development point x is preserved:

$$\mathrm{dist}(\mathrm{x},\mathrm{y}) = \|\overrightarrow{\mathrm{xy}}\| = (\langle\, \overrightarrow{\mathrm{xy}} \mid \overrightarrow{\mathrm{xy}} \,\rangle_\mathrm{x})^{1/2}$$

Thus, the *exponential chart at* x can be seen as the development of the manifold in the tangent space at a given point along the geodesics. This is also called a *normal coordinate system* if it is provided with an orthonormal basis. At the origin of such a chart, the metric reduces to the identity matrix and the Christoffel symbols vanish.

Now, it is natural to search for the maximal domain where the exponential map is a diffeomorphism. If we follow a geodesic $\gamma_{(\mathrm{x},\partial_v)}(t) = \exp_\mathrm{x}(t\,\partial_v)$ from $t = 0$ to infinity, it is either always minimizing all along or it is minimizing up to a time $t_0 < \infty$ and not any more after (thanks to the geodesic completeness). In this last case, the point $\mathrm{z} = \gamma_{(\mathrm{x},\partial_v)}(t_0)$ is called a *cut point* and the corresponding tangent vector $t_0\,\partial_v$ a *tangential cut point*. The set of of all cut points of all geodesics starting from x is the *cut locus* $C(\mathrm{x}) \in \mathcal{M}$ and the set of corresponding vectors the *tangential cut locus* $\mathcal{C}(\mathrm{x}) \in T_\mathrm{x}\mathcal{M}$. Thus, we have $C(\mathrm{x}) = \exp_\mathrm{x}(\mathcal{C}(\mathrm{x}))$, and the maximal definition domain for the exponential chart is the domain $\mathcal{D}(\mathrm{x})$ containing 0 and delimited by the tangential cut locus.

It is easy to see that this domain is connected and star-shaped with respect to the origin of $T_\mathrm{x}\mathcal{M}$. Its image by the exponential map covers all the manifold except the cut locus and the segment $[0, \overrightarrow{\mathrm{xy}}]$ is transformed into the unique minimizing geodesic from x to y. Hence, the exponential chart at x is a chart centered at x with a connected and star-shaped definition domain that covers all the manifold except the cut locus $C(\mathrm{x})$:

$$\begin{array}{ccc} \mathcal{D}(\mathrm{x}) \in \mathbb{R}^n & \longleftrightarrow & \mathcal{M} - C(\mathrm{x}) \\ \overrightarrow{\mathrm{xy}} = \log_\mathrm{x}(\mathrm{y}) & \longleftrightarrow & \mathrm{y} = \exp_\mathrm{x}(\overrightarrow{\mathrm{xy}}) \end{array}$$

From a computational point of view, it is often interesting to extend this representation to include the tangential cut locus. However, we have to take care of the multiple representations: points in the cut locus where several minimizing geodesics meet are represented by several points on the tangential cut locus as the geodesics are starting with different tangent vectors (e.g. rotation of π around the axis $\pm n$ for 3D rotations, antipodal point on the sphere). This multiplicity problem cannot be avoided as the set of such points is dense in the cut locus.

The size of this definition domain is quantified by the *injectivity radius* $\mathrm{i}(\mathcal{M}, x) = \mathrm{dist}(\mathrm{x}, \mathcal{C}(\mathrm{x}))$, which is the maximal radius of centered balls in $T_\mathrm{x}\mathcal{M}$ on which the exponential map is one-to-one. The injectivity radius of the manifold $\mathrm{i}(\mathcal{M})$ is the infimum of the injectivity over the manifold. It may be zero, in which case the manifold somehow tends towards a singularity (think e.g. to the surface $z = 1/\sqrt{x^2 + y^2}$ as a sub-manifold of $\mathbb{R}^3$).

Example: On the sphere $\mathcal{S}_n$ (center 0 and radius 1) with the canonical Riemannian metric (induced by the ambient Euclidean space $\mathbb{R}^{n+1}$), the geodesics are the great circles and the cut locus of a points x is its antipodal point $\underline{\mathrm{x}} = -\mathrm{x}$. The exponential chart is obtained by rolling the sphere onto its tangent space so that the great circles going through x become lines. The maximal definition domain is thus the open ball $\mathcal{D} = \mathcal{B}_n(\pi)$. On its boundary $\partial\mathcal{D} = \mathcal{C} = \mathcal{S}_{n-1}(\pi)$, all the points represent $\underline{\mathrm{x}}$.

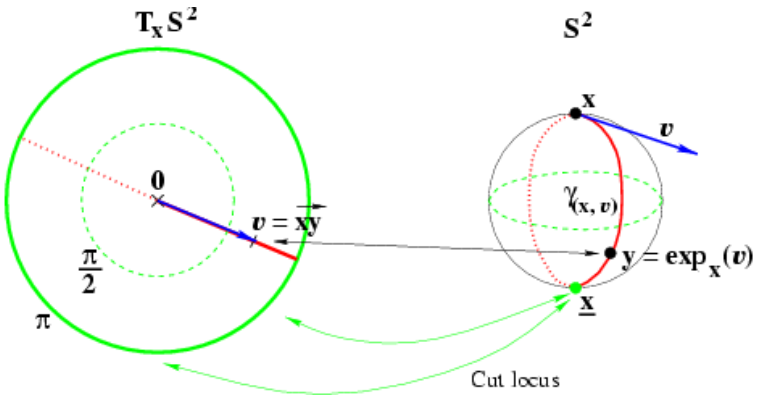

Figure 1: Exponential chart and cut locus for the sphere $\mathcal{S}_2$ and the projective space $\mathcal{P}_2$

For the real projective space $\mathcal{P}_n$ (obtained by identification of antipodal points of the sphere $\mathcal{S}_n$), the geodesics are still the great circles, but the cut locus of the point $\{\mathrm{x}, -\mathrm{x}\}$ is now the equator of the two points where antipodal points are still identified (thus the cut locus is $\mathcal{P}_{n-1}$). The definition domain of the exponential chart is the open ball $\mathcal{D} = \mathcal{B}_n(\frac{\pi}{2})$, and the tangential cut locus is the sphere $\partial\mathcal{D} = \mathcal{S}_{n-1}(\frac{\pi}{2})$ where antipodal points are identified.

2.3 Taylor expansion of a real function

Let f be a smooth function from $\mathcal{M}$ to $\mathbb{R}$ (an observable). Its Gradient $\operatorname{grad} f$ at point x is the linear form on $T_{\mathrm{x}}\mathcal{M}$ corresponding to the directional derivatives ∂_v:

$$\forall\ v \in T_{\mathrm{x}}\mathcal{M} \qquad \operatorname{grad} f(v) = \partial_v f$$

Thanks to the dot product, we can identify the linear form $d\omega$ in the tangent space $T_{\mathrm{x}}\mathcal{M}$ with the vector ω if $d\omega(v) = \langle\, \omega \mid v \,\rangle_{\mathrm{x}}$ for all vector $v \in T_{\mathrm{x}}\mathcal{M}$. All these notions can be extended to the whole manifold using vector fields: in a local chart x, we have $\partial_v f = \frac{\partial f(x)}{\partial x}\, v$ and $\langle\, \omega \mid v \,\rangle_x = \omega^{\mathrm{T}}\, G(x)\, v$. Thus, the expression of the gradient in a chart is:

$$\operatorname{grad} f = G^{(-1)}(x)\ \frac{\partial f}{\partial x}^{\mathrm{T}} = g^{ij}\, \partial_j f$$

This definition corresponds to the classical gradient in $\mathbb{R}^n$ even in the case of a non orthonormal basis. The second covariant derivative (the *Hessian*) is a little bit more complex and makes use of the connection ∇. We just give here its expression in a local coordinate system:

$$\operatorname{Hess} f = \nabla df = (\partial_{ij} f - \Gamma_{ij}^k\, \partial_k f)\, dx^i\, dx^j$$

Let now f_{x} be the expression of f in a normal coordinate system at x. Its Taylor expansion around the origin is:

$$f_{\mathrm{x}}(v) = f_{\mathrm{x}}(0) + J_{f_{\mathrm{x}}}\, v + \frac{1}{2} v^{\mathrm{T}}\, H_{f_{\mathrm{x}}}\, v + O(\|v\|^3)$$

where $J_{f_{\mathrm{x}}} = [\partial_i f]$ and $H_{f_{\mathrm{x}}} = [\partial_{ij} f]$. Since we are in a normal coordinate system, we have $f_{\mathrm{x}}(v) = f(\exp_{\mathrm{x}}(v))$. Moreover, the metric at the origin reduces to the identity: $J_{f_{\mathrm{x}}} = \operatorname{grad} f^{\mathrm{T}}$, and the Christoffel symbols vanish so that the matrix of second derivatives $H_{f_{\mathrm{x}}}$ corresponds to the Hessian $\operatorname{Hess} f$. Thus, The Taylor expansion can be written in any coordinate system:

$$f\left(\exp_{\mathrm{x}}(v)\right) = f(\mathrm{x}) + \operatorname{grad} f(v) + \frac{1}{2}\operatorname{Hess} f(v,v) + O(\|v\|^3) \tag{2}$$

3 Random points on a Riemannian Manifold

In this paper, we are interested in measurements of elements of a Riemannian manifold that depend on the outcome of a random experiment. Particular cases are given by *random transformation* and *random feature* for the particular case of transformation groups and homogeneous manifolds.

Definition 1 (Random point on a Riemannian Manifold)
Let $(\Omega, \mathcal{B}(\Omega), \Pr)$ *be a probability space,* $\mathcal{B}(\Omega)$ *being the Borel* σ*-algebra of* Ω *(i.e. the smallest* σ*-algebra containing all the open subsets of* Ω*) and* $\Pr$ *a measure on that* σ*-algebra such that* $\Pr(\Omega) = 1$*. A random point in the Riemannian manifold* $\mathcal{M}$ *is a Borel measurable function* $\mathbf{x} = \mathrm{x}(\omega)$ *from* Ω *to* $\mathcal{M}$*.*

As in the real or vectorial case, we can now make abstraction of the original space Ω and directly work with the induced probability measure on $\mathcal{M}$.

3.1 Riemannian measure or volume form

In a vector space with basis $\mathcal{A} = (a_1, \dots a_n)$, the local representation of the metric is given by $G = A^{\mathrm{T}} A$ where $A = [a_1, \dots a_n]$ is the matrix of coordinates change from $\mathcal{A}$ to an orthonormal basis. Similarly, the measure (or the infinitesimal volume element) is given by the volume of the parallelepipedon spanned by the basis vectors: $d\mathcal{V} = |\det(A)|\, dx = \sqrt{|\det(G)|}\, dx$. Assuming now a Riemannian manifold $\mathcal{M}$, we can see that the Riemannian metric $G(x)$ induces an infinitesimal volume element on each tangent space, and thus a measure on the manifold:

$$d\mathcal{M}(\mathrm{x}) = \sqrt{|G(x)|}\, dx$$

One can show that the cut locus has a null measure. This means that we can integrate indifferently in $\mathcal{M}$ or in any exponential chart. If f is an integrable function of the manifold and $f_{\mathrm{x}}(\overrightarrow{\mathrm{xy}}) = f(\exp_{\mathrm{x}}(\overrightarrow{\mathrm{xy}}))$ is its image in the exponential chart at x, we have:

$$\int_{\mathcal{M}} f(\mathrm{x})\, d\mathcal{M} = \int_{\mathcal{D}(\mathrm{x})} f_{\mathrm{x}}(\vec{\mathrm{z}})\, \sqrt{G_{\vec{\mathrm{x}}}(\vec{\mathrm{z}})}\, d\vec{\mathrm{z}}$$

3.2 Probability density function

Definition 2 *Let* $\mathcal{B}(\mathcal{M})$ *be the Borel* σ*-algebra of* $\mathcal{M}$*. The random point* $\mathbf{x}$ *has a probability density function* $p_{\mathbf{x}}$ *(real, positive and integrable function) if:*

$$\forall \mathcal{X} \in \mathcal{B}(\mathcal{M}), \qquad \Pr(\mathbf{x} \in \mathcal{X}) = \int_{\mathcal{X}} p(\mathrm{y})\, d\mathcal{M}(\mathrm{y}) \qquad \textit{and} \qquad \Pr(\mathcal{M}) = \int_{\mathcal{M}} p(\mathrm{y})\, d\mathcal{M}(\mathrm{y}) = 1$$

A simple example of a pdf is the *uniform pdf* in a bounded set $\mathcal{X}$:

$$p_{\mathcal{X}}(\mathrm{y}) = \frac{1}{\int_{\mathcal{X}} d\mathcal{M}} \mathbf{1}_{\mathcal{X}}(\mathrm{y}) = \frac{\mathbf{1}_{\mathcal{X}}(\mathrm{y})}{\mathcal{V}ol(\mathcal{X})}$$

One must be careful that this pdf is uniform with respect to the measure $d\mathcal{M}$ and is not uniform for another measure on the manifold. This problem is the basis of the Bertrand paradox for geometrical probabilities [43, 44, 2] and raise the problem of the measure to choose on the manifold. In our case, the measure is induced by the Riemannian metric but the problem is only lifted: which Riemannian metric do we have to choose ? For transformation groups and homogeneous manifolds, we showed in [1] that an invariant metric is a good geometric choice, even if such a metric does not always exist for homogeneous manifolds or if it leads in general to a partial consistency only between the geometric and statistical operations in non compact transformation groups [45].

3.3 Expression of the density in a chart

Let $\mathbf{x}$ be a random point of pdf $p_{\mathbf{x}}$. If $x = \pi(\mathrm{x})$ is a chart of the manifold defined almost everywhere, we obtain a random vector $\boldsymbol{x} = \pi(\mathbf{x})$ which pdf $\rho_{\boldsymbol{x}}$ is defined with respect to the Lebesgue measure dx in $\mathbb{R}^n$ instead of $d\mathcal{M}$ in $\mathcal{M}$. Using the expression of the Riemannian measure, the two pdfs are related by

$$\rho_{\boldsymbol{x}}(y) = p_{\mathbf{x}}(y) \sqrt{|G(y)|}$$

We shall note that the density $\rho_{\boldsymbol{x}}$ depends on the chart used whereas the pdf $p_{\mathbf{x}}$ is intrinsic to the manifold.

3.4 Expectation of an observable

Let $\varphi(\mathrm{x})$ be a Borelian real valued function defined on $\mathcal{M}$ and $\mathbf{x}$ a random point of pdf $p_{\mathbf{x}}$. Then, $\varphi(\mathbf{x})$ is a real random variable and we can compute its expectation:

$$\mathbf{E}\,[\,\varphi(\mathbf{x})\,] = \mathbf{E}_{\mathbf{x}}\,[\,\varphi\,] = \int_{\mathcal{M}} \varphi(\mathrm{y})\, p_{\mathbf{x}}(\mathrm{y})\, d\mathcal{M}(\mathrm{y})$$

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 have no way to generalize this integral in $\mathbb{R}$ into an integral with value in the manifold.

4 Expectation or Mean of a Random point

We focus in this section to the notion of central value of a distribution. We will preferably use the denomination *mean value or mean point* than *expected point* to stress the difference between this notion and the expectation of a real function.

4.1 Fréchet expectation or mean value

Let $\boldsymbol{x}$ be a random vector of $\mathbb{R}^n$. Fréchet observed in [46, 47] that the variance $\sigma^2_{\boldsymbol{x}}(y) = \mathbf{E}\left[\ \mathrm{dist}(\boldsymbol{x}, y)^2\ \right]$ is minimized for the mean vector $\overline{x} = \mathbf{E}\,[\,\boldsymbol{x}\,]$. The major point for the generalization is that the expectation of a real valued function is well defined for our connected and geodesically complete Riemannian manifold $\mathcal{M}$.

Definition 3 Variance of a random point
Let $\mathbf{x}$ *be a random point of pdf* $p_{\mathbf{x}}$*. The variance* $\sigma^2_{\mathbf{x}}(\mathrm{y})$ *is the expectation of the squared distance between the random point and the fixed point* y*:*

$$\sigma^2_{\mathbf{x}}(\mathrm{y}) = \mathbf{E}\left[\ \mathrm{dist}(\mathrm{y},\mathbf{x})^2\ \right] = \int_{\mathcal{M}} \mathrm{dist}(\mathrm{y},\mathrm{z})^2\, p_{\mathbf{x}}(\mathrm{z})\, d\mathcal{M}(\mathrm{z}) \tag{3}$$

Definition 4 Fréchet expectation of a random point
Let $\mathbf{x}$ *be a random point. If the variance* $\sigma^2_{\mathbf{x}}(\mathrm{y})$ *is finite for all point* $\mathrm{y} \in \mathcal{M}$ *(which is in particular true for a density with a compact support), every point* $\bar{\mathrm{x}}$ *minimizing this variance is called an expected or mean point. Thus, the set of the mean points is:*

$$\mathbb{E}\left[\,\mathbf{x}\,\right] = \arg\min_{\mathrm{y}\in\mathcal{M}} \left(\mathbf{E}\left[\ \mathrm{dist}(\mathrm{y},\mathbf{x})^2\ \right]\right)$$

If there exists a least one mean point $\bar{\mathrm{x}}$*, we call variance the minimal value* $\sigma^2_{\mathbf{x}} = \sigma^2_{\mathbf{x}}(\bar{\mathrm{x}})$ *and standard deviation its square-root.*

Similarly, one defines the *empirical or discrete mean point* of a set of measures $\mathrm{x}_1, \ldots \mathrm{x}_n$:

$$\mathbb{E}\left[\,\{\mathrm{x}_i\}\,\right] = \arg\min_{\mathrm{y}\in\mathcal{M}} \left(\mathbf{E}\left[\,\{\,\mathrm{dist}(\mathrm{y},\mathrm{x}_i)^2\}\,\right]\right) = \arg\min_{\mathrm{y}\in\mathcal{M}} \left(\frac{1}{n}\sum_i \mathrm{dist}(\mathrm{y},\mathrm{x}_i)^2\right)$$

If there exists a least a mean point $\bar{\mathrm{x}}$, one calls *empirical variance* the minimal value $s^2 = \frac{1}{n}\sum_i \mathrm{dist}(\bar{\mathrm{x}},\mathrm{x}_i)^2$ and *empirical standard deviation* or RMS (for Root Mean Square) its square-root.

Following the same principle, one can define other types of central values. The *mean deviation at order* α is

$$\sigma_{\mathbf{x},\alpha}(\mathrm{y}) = \left(\mathbf{E}\left[\ \mathrm{dist}(\mathrm{y},\mathbf{x})^\alpha\ \right]\right)^{1/\alpha} = \left(\int_{\mathcal{M}} \mathrm{dist}(\mathrm{y},\mathrm{z})^\alpha\, p_{\mathbf{x}}(\mathrm{z})\, d\mathcal{M}(\mathrm{z})\right)^{1/\alpha}$$

If this function is bounded on $\mathcal{M}$, one call *central point at order* α every point $\bar{\mathrm{x}}_\alpha$ minimizing it. For instance, the *modes* are obtained for $\alpha = 0$. Exactly like in a vector space, they are the points where the density is maximal on the manifold (which is generally not a maximum for the density on the charts). The *median point* is obtained for $\alpha = 1$. For $\alpha \to \infty$, we obtain the "barycenter" of the distribution support (which has to be compact).

The definition of these central values can be extended to the discrete case easily, except perhaps for the modes and for $\alpha \to \infty$. We note that the Fréchet expectation is defined for all metric space and not only for Riemannian manifolds.

4.2 Existence and uniqueness: Riemannian center of mass

As our mean point is the result of a minimization, its existence is not ensured (the global minimum could be unreachable) and anyway the result is a set and no longer a single element. This has to be compared with some central values in vector spaces, for instance the modes. However, the Fréchet expectation does not define all the modes even in vector spaces: one only keep the modes of maximal intensity.

To get rid of this constraint, Karcher [25] proposed to consider the local minima of the variance $\sigma^2_{\mathbf{x}}(\mathrm{y})$ (equation 3) instead of the global ones. We call these new set of means *Riemannian centers*

of mass. As global minima are local minima, the Fréchet expected points are a subset of the Riemannian centers of mass. However, the use of local minima allows to characterize the Riemannian centers of mass using only local derivatives of order two.

Using this extended definition, Karcher [25] and Kendall [48] established conditions on the manifold and the distribution to ensure the existence and uniqueness of the mean. We just recall here the results without the proofs.

Definition 5 (Regular geodesic balls) *The ball* $\mathcal{B}(\mathrm{x}, r) = \{\mathrm{y} \in \mathcal{M} / \ \mathrm{dist}(\mathrm{x}, \mathrm{y}) < r\}$ *is said* geodesic *if it does not meet the cut locus of its center. This means that there exists a unique minimizing geodesic from the center to any point of a geodesic ball. The ball is said* regular *if its radius verifies* $2r\sqrt{\kappa} < \pi$*, where* κ *is the maximum of the Riemannian curvature in this ball.*

For instance, on the sphere $\mathcal{S}_2$ with radius one, the curvature is constant and equal to 1. A geodesic ball is regular if $r < \pi/2$. Such a ball can almost cover an hemisphere, but not the equator. In a Riemannian manifold with non positive curvature, a regular geodesic ball can cover the whole manifold (according to the Cartan-Hadamard theorem, such a manifold is diffeomorphic to $\mathbb{R}^n$ if it is simply connected and complete).

Theorem 1 (Existence and uniqueness of the Riemannian center of mass)
Let $\mathbf{x}$ *be a random point of pdf* $p_{\mathbf{x}}$*.*

- **Kendall [48]** *If the support of* $p_{\mathbf{x}}$ *is included in a regular geodesic ball* $\mathcal{B}(\mathrm{y}, r)$*, then there exists one and only one Riemannian center of mass* $\mathbf{x}$ *on this ball.*
- **Karcher [25]** *If the support of* $p_{\mathbf{x}}$ *is included in a geodesic ball* $\mathcal{B}(\mathrm{y}, r)$ *and if the ball of double radius* $\mathcal{B}(\mathrm{y}, 2\,r)$ *is still geodesic and regular, then the variance* $\sigma^2_{\mathbf{x}}(\mathrm{z})$ *is a convex function of* z *and has only one critical point on* $\mathcal{B}(\mathrm{y}, r)$*, necessarily the Riemannian center of mass.*

These conditions ensure a correct behavior of the mean for sufficiently localized distributions. However, they are quite restrictive as they only address pdfs with a compact support. Kendall's existence and uniqueness theorem was extended by [30] to distributions with non-compact support in manifolds with Ψ-convexity. This notion, already introduced by Kendall in his original proof, is here extended to the whole manifold. Unfortunately, this type of argument can only be used for a restricted class of manifolds as a non-compact connected and geodesically complete Ψ-convex manifold is diffeomorphic to $\mathbb{R}^m$. It remains that this extension of the theorem applies to the important class of Hadamard manifolds (i.e. simply connected, complete and with non-positive sectional curvature), whose curvature is bounded from below.

4.3 Other possible definitions of the mean points

The Riemannian center of mass is perfectly adapted for our purpose, thanks to its good properties for optimization (see Section 4.6 below). However, there are other works proposing different ways to generalize the notion of mean value or barycenter of a distribution in a manifold. We review them for the sake of completeness and for their mathematical interest, but they do not seem to be practically applicable.

Doss [49] used another property of the expectation as the starting point for the generalization: if $\boldsymbol{x}$ is a real random variable, the only real number $\bar{x}$ verifying:

$$\forall y \in \mathbb{R} \qquad |y - \bar{x}| \leq \mathbf{E}\left[\, |\boldsymbol{x} - \bar{x}| \,\right]$$

is the mean value $\mathbf{E}[\,\boldsymbol{x}\,]$. Thus, in a metric space, the *mean according to Doss* is defined as the set of points $\bar{\mathrm{x}} \in \mathcal{M}$ verifying:

$$\forall \mathrm{y} \in \mathcal{M} \qquad \mathrm{dist}(\mathrm{y}, \bar{\mathrm{x}}) \leq \mathbf{E}[\ \mathrm{dist}(\mathbf{x}, \bar{\mathrm{x}})\ ]$$

Herer shows in [50, 51] that this definition includes the classical expectation in a Banach space (with possibly other points) and develop on this basis a conditional expectation.

A similar definition that uses convex functions on the manifold instead of metric properties proposed by Emery [27] and Arnaudon [52, 28]. A function from $\mathcal{M}$ to $\mathbb{R}$ is convex if its restriction to all geodesic is convex (considered as a function from $\mathbb{R}$ to $\mathbb{R}$). The *convex barycenter* of a random point $\mathbf{x}$ with density $p_{\mathbf{x}}$ is the set $\mathbb{B}(\mathbf{x})$ of points $\mathrm{y} \in \mathcal{M}$ such that $\alpha(\mathrm{y}) \leq \mathbf{E}[\,\alpha(\mathbf{x})\,]$ holds for every real bounded and convex function α on a neighborhood of the support of $p_{\mathbf{x}}$.

This definition seems to be of little interest in our case since for compact manifolds, such as the sphere or $\mathcal{S}O_3$ (the manifold of 3D rotations), the geodesics are closed and the only convex functions on the manifold are the constant ones. Thus, every random point for which the support of the distribution is the whole manifold has the whole manifold as convex barycenter.

However, in the case where the support of the distribution is included in a *strongly convex open set*[2] $\mathcal{U}$, Emery [27] showed that the *exponential barycenters*, defined as the critical points of the variance $\sigma^2_{\mathbf{x}}(\mathrm{y})$ are subset of the convex barycenter $\mathbb{B}(\mathbf{x})$. Local and global minima being particular critical points, the exponential barycenters include the Riemannian centers of mass that include themselves the Fréchet means.

Picard [29] 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 $O(\sigma_{\mathbf{x}})$. Thus, for sufficiently centered random points, all these values are close.

4.4 Characterizing a Riemannian center of mass

To characterize a local minimum of a twice differentiable function, we just have to require a null gradient and a negative definite Hessian matrix. The problem with the variance function $\sigma^2(\mathrm{y})$ is that the integration domain (namely $\mathcal{M} \backslash C(\mathrm{y})$) depends on the derivation point y. Thus we cannot just use the Lebesgue theorem to differentiate under the sum, unless there is no cut locus, or the distribution has a sufficiently small compact support, which is the property used by Karcher Kendall and Emery for the previous existence, uniqueness and equivalence theorems. We were able to generalize in appendix A a differentiability proof of Pr Maillot [53] originally designed for the uniform distribution on compact manifolds. The theorem we obtain is the following:

Theorem 2 Gradient of the variance function
Let P be a probability on the Riemannian manifold $\mathcal{M}$. The variance $\sigma^2(\mathrm{y}) = \int_{\mathcal{M}} \mathrm{dist}(\mathrm{y},\mathrm{z})^2\, dP(\mathrm{z})$ is differentiable at any point $\mathrm{y} \in \mathcal{M}$ where it is finite and where the cut locus $C(\mathrm{y})$ has a null probability measure:

$$P(C(\mathrm{y})) = \int_{C(\mathrm{y})} dP(\mathrm{z}) = 0 \qquad \textit{and} \qquad \sigma^2(\mathrm{y}) = \int_{\mathcal{M}} \mathrm{dist}(\mathrm{y},\mathrm{z})^2\, dP(\mathrm{z}) < \infty$$

[2] Here, strongly convex means that for every two points of $\mathcal{U}$ there is a unique minimizing geodesic joining them that depend in a C^∞ of the two points.

At such a point, it has the following gradient:

$$(\operatorname{grad}\sigma^2)(\mathrm{y}) = -2\int_{\mathcal{M}/C(\mathrm{y})} \overrightarrow{\mathrm{yz}}\; dP(\mathrm{z}) = -2\,\mathbf{E}\left[\,\overrightarrow{\mathrm{yx}}\,\right]$$

Now, we know that the variance is continuous but may not be differentiable at the points where the cut locus has a non-zero probability measure. At these points, the variance can have an extremum (think for instance to $\|x\|$ in vector spaces). Thus, the extrema of σ^2 are characterized by $(\operatorname{grad}\sigma^2)(y) = 0$ if this is defined or $P(C(y)) > 0$.

Corollary 1 (Characterization of Riemannian centers of mass)
Assume that the random point $\mathbf{x}$ *has a finite variance everywhere and let* $\mathcal{A}$ *be the set of points where the cut locus has a non-zero probability measure. A necessary condition for* $\bar{\mathrm{x}}$ *to be a Riemannian center of mass is* $\mathbf{E}\left[\,\overrightarrow{\bar{\mathrm{x}}\mathbf{x}}\,\right] = 0$ *if* $\bar{\mathrm{x}} \notin \mathcal{A}$*, or* $\bar{\mathrm{x}} \in \mathcal{A}$*. For discrete or empirical means, the characterization is the same but we can write explicitly the set* $\mathcal{A} = \cup_i C(\mathrm{x}_i)$*.*

If the manifold does not have a cut locus (for instance in Hadamard manifolds), we have no differentiation problem. One could think of going one step further and computing the Hessian matrix. Indeed, we have in the vector case: $\operatorname{Hess}\left(\sigma^2_{\boldsymbol{x}}(y)\right) = -2\,\mathrm{Id}$ everywhere, which proves that any extremum of the variance is a minimum. In Riemannian manifolds, one has to be more careful because the Hessian is modified by the curvature of the manifold. One solution is to compute the Hessian matrix using the theory of Jacobi fields, and then take estimates of its eigenvalues based on bounds on the curvature. This is essentially the idea exploited in [25] to show the uniqueness of the mean is small enough geodesic balls, and by [54] to exhibit an example of a manifold without cut-locus that is strongly convex (i.e. there is one and only one minimizing geodesic joining any two points), but that support finite mass measures that have non-unique Riemannian centers of mass. Thus, the absence of a cut locus is not enough: one should also have some constraints on the curvature of the manifold. In order to remain simple, we stick in this paper to the existence and uniqueness theorem provided by [30] for simply connected and complete manifolds whose curvature is non-positive (i.e. Hadamard) and bounded from below.

Corollary 2 (Characterization of the Fréchet mean for Hadamard manifolds with a curvature bounded from below).
Assume that the random point $\mathbf{x}$ *has a finite variance. Then, there is one and only one Riemannian center of mass characterized by* $\mathbf{E}\left[\,\overrightarrow{\bar{\mathrm{x}}\mathbf{x}}\,\right] = 0$*. For discrete or empirical means, the characterization is similar.*

Results similar to Theorem 2 and the above corollaries have been derived independently. [15] defined the mean values in manifolds as the exponential barycenters. To relate them with the Riemannian centers of mass, they determined the gradient of the variance. However, they only investigate the relation between the two notions when the probability is dominated by the Riemannian measure, which excludes explicitly point-mass distributions. In [37, 38], the gradient of the variance is also determined and the existence of the mean is established for simply connected Riemannian manifolds with non-positive curvature.

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 (we assumed that the manifold was complete and simply connected). 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 Hessian matrix of the variance on these manifolds.

4.5 Example on the circle

The easiest example of this difficulty is probably a symmetric distribution on the circle. Let $p = \cos(\theta)^2/\pi$ be the probability density function of our random point $\boldsymbol{\theta}$ on the circle. For a circle with the canonical metric, the exponential chart centered at α is $\overrightarrow{\alpha\theta} = \theta - \alpha$ for $\theta \in]\alpha - \pi; \alpha + \pi[$, the distance being obviously $\operatorname{dist}(\alpha, \theta) = |\alpha - \theta|$ within this domain.

Let us first compute the mean points by computing explicitly the variance and its derivatives. The variance is:

$$\sigma^2(\alpha) = \int_{\alpha-\pi}^{\alpha+\pi} \operatorname{dist}(\alpha,\theta)^2 \, p(\theta) \, d\theta = \int_{-\pi}^{\pi} \gamma^2 \frac{\cos(\gamma+\alpha)^2}{\pi} d\gamma = \frac{\pi^2}{3} - \frac{1}{2} + \cos(\alpha)^2.$$

Its derivative is rather easy to compute: $\operatorname{grad}\sigma^2(\alpha) = -2\cos(\alpha)\sin(\alpha)$, and the second order derivative is $H(\alpha) = 4\sin(\alpha)^2 - 2$. Solving for $\operatorname{grad}\sigma^2(\alpha) = 0$, we get four critical points:

- $\alpha = 0$ and $\alpha = \pm\pi$ with $H(0) = H(\pm\pi) = -2$,
- $\alpha = \pm\pi/2$ with $H(\pm\pi/2) = +2$.

Thus, there are two relative (and here absolute) minima: $\mathbb{E}[\,\boldsymbol{\theta}\,] = \{0\,,\,\pm\pi\}$.

Let us use now the general framework developed on Riemannian manifolds. According to theorem 2, the gradient of the variance is

$$\operatorname{grad}\sigma^2(\alpha) = -2\mathbb{E}\left[\,\overrightarrow{\alpha\boldsymbol{\theta}}\,\right] = -2\int_{\alpha-\pi}^{\alpha+\pi} \overrightarrow{\alpha\theta}\, d\theta = -2\int_{\alpha-\pi}^{\alpha+\pi} (\theta-\alpha)\,\frac{\cos(\theta)^2}{\pi}\, d\theta = -2\cos(\alpha)\sin(\alpha),$$

which is in accordance with our previous computations. Now, differentiating once again under the sum, we get:

$$\int_{\alpha-\pi}^{\alpha+\pi} \frac{\partial^2 \operatorname{dist}(\alpha,\theta)}{\partial\alpha^2}\, p(\theta)\, d\theta = -2\int_{\alpha-\pi}^{\alpha+\pi} \frac{\partial\overrightarrow{\alpha\theta}}{\partial\alpha}\, p(\theta)\, d\theta = 2\int_{\alpha-\pi}^{\alpha+\pi} p(\theta)\, d\theta = 2,$$

which is clearly different from our direct calculation. One way to see the problem is the following: the vector field $\overrightarrow{\alpha\theta}$ is continuous and differentiable on the circle except at the cut locus of α (i.e. at $\theta = \alpha \pm \pi$) where it has a jump of 2π. Thus, the second order derivative of the squared distance should be $-2(-1 + 2\pi\delta_{(\alpha\pm\pi)}(\theta))$, where δ is the Dirac distribution, and the integral becomes:

$$H(\alpha) = -2\int_{\alpha-\pi}^{\alpha+\pi} \left(-1 + 2\pi\delta_{(\alpha\pm\pi)}(\theta)\right)\, p(\theta)\, d\theta = 2 - 4\pi\, p(\alpha \pm \pi) = 2 - 4\cos(\theta)^2$$

which is this time in accordance with the direct calculation.

4.6 A gradient descent algorithm to obtain the mean

Gradient descent is a usual technique to compute a minimum. Moreover, as we have a canonical way to go from the tangent space to the manifold thanks to the exponential map, this iterative

algorithm seems to be perfectly adapted. In this section, we assume that the conditions of theorem (2) are fulfilled.

Let y be an estimation of the mean of the random point $\mathbf{x}$ and $f(\mathrm{y}) = \sigma^2_{\mathbf{x}}(\mathrm{y})$ the variance. A practical gradient descent algorithm is to minimize the second order approximation of the cost function at the current point. According to the Taylor expansion of equation (2), the second order approximation of f and y is:

$$f(\exp_{\mathrm{y}}(v)) = f(\mathrm{y}) + \operatorname{grad} f(v) + \frac{1}{2}\operatorname{Hess} f(v, v)$$

This is a function of the vector $v \in T_{\mathrm{y}}\mathcal{M}$. Assuming that $\operatorname{Hess} f$ is positive definite, this function is convex and has thus a minimum characterized by a null gradient. Let $H_f(v)$ denote the linear form verifying $H_f(v)(w) = \operatorname{Hess} f(v, w)$ for all w and $H_f^{(-1)}$ denote the inverse map. The minimum is characterized by

$$\operatorname{grad}_v f_{\mathrm{y}} = 0 = \operatorname{grad} f + H_f(v) \qquad \Leftrightarrow \qquad v = -H_f^{(-1)}(\operatorname{grad} f)$$

We saw in the previous section that $\operatorname{grad} f = -2\,\mathbf{E}\left[\,\overrightarrow{\mathrm{y}\mathbf{x}}\,\right]$. Neglecting the "cut locus term" in the Hessian matrix gives us a perfect positive definite matrix $\operatorname{Hess} f \simeq 2\ \mathrm{Id}$. Thus, the gradient descent algorithm is

$$\mathrm{y}_{t+1} = \exp_{\mathrm{y}_t}\left(\mathbf{E}\left[\,\overrightarrow{\mathrm{y_t}\mathbf{x}}\,\right]\right)$$

This gradient descent algorithm can be seen as the discretization of the ordinary differential equation $\dot{\mathrm{y}}(t) = \mathbf{E}\left[\,\overrightarrow{\mathrm{y(t)}\mathbf{x}}\,\right]$. Other discretization scheme are possible [55], sometimes with convergence theorems [56].

In the case of the discrete or empirical mean, which is much more interesting from a statistical point of view, we have exactly the same algorithm, but with the empirical expectation:

$$\mathrm{y}_{t+1} = \exp_{\mathrm{y}_t}\left(\frac{1}{n}\sum_{i=1}^{n}\overrightarrow{\mathrm{y_t}\mathrm{x_i}}\right)$$

We note that in the case of a vector space, these two formula simplify to $y_{t+1} = \mathbf{E}[\,\boldsymbol{x}\,]$ and $y_{t+1} = \frac{1}{n}\sum_i x_i$, which are the definition of the mean value and the barycenter. Moreover, the algorithm converges in a single step.

An important point for this algorithm is to determine a good starting point. In the case on a set of observations $\{\mathrm{x}_i\}$, one can choose at random one of the observations as the starting point. Another solution is to map to each point x_i its mean distance with respect to other points (or the median distance to be robust) and choose as the starting point the minimizing point. From a computer science point of view, the complexity is k^2 (where k is the number of observations) but the method can be randomized efficiently [57, 58].

To verify the uniqueness of the solution, we can repeat the algorithm from several starting points (for instance all the observations x_i). If we know the Riemannian curvature of the manifold (for instance if it is constant or if there is an upper bound κ), we can use theorem (1, Section 4.2). We just have to verify that the maximum distance between the observations and the mean value we have found is sufficiently small so that all observations fits into a regular geodesic ball of radius:

$$r = \max_i \operatorname{dist}(\bar{\mathrm{x}}, \mathrm{x}_i) < \frac{\pi}{2\sqrt{\kappa}}$$

5 Covariance matrix

With the mean value, we have a dispersion value: the variance. To go one step further, we observe that the covariance matrix of a random vector $\boldsymbol{x}$ with respect to a point y is the *directional* dispersion of the "difference" vector $\overrightarrow{y\boldsymbol{x}} = \boldsymbol{x} - y$:

$$\mathrm{Cov}_{\boldsymbol{x}}(y) = \mathbf{E}\left[\, \overrightarrow{y\boldsymbol{x}}\, \overrightarrow{y\boldsymbol{x}}^{\mathrm{T}} \,\right] = \int_{\mathbb{R}^n} (\overrightarrow{yx})\, (\overrightarrow{yx})^{\mathrm{T}}\, p_{\boldsymbol{x}}(x)\, dx$$

This definition is readily extendible to a complete Riemannian manifold using the random vector $\overrightarrow{\mathrm{y}\mathbf{x}}$ in $T_{\mathrm{y}}\mathcal{M}$ and the Riemannian measure. In fact, we are usually interested in the covariance relative to the mean value:

Definition 6 (Covariance)
Let $\mathbf{x}$ *be a random point and* $\bar{\mathrm{x}} \in \mathbb{E}[\,\mathbf{x}\,]$ *a mean value that we assume to be unique to simplify the notations (otherwise we have to keep a reference to the mean value). We define the covariance* $\Sigma_{\mathbf{xx}}$ *by the expression:*

$$\Sigma_{\mathbf{xx}} = \mathrm{Cov}_{\mathbf{x}}(\bar{\mathrm{x}}) = \mathbf{E}\left[\, \overrightarrow{\bar{\mathrm{x}}\mathbf{x}}\, \overrightarrow{\bar{\mathrm{x}}\mathbf{x}}^{T} \,\right] = \int_{\mathcal{D}(\bar{\mathrm{x}})} (\overrightarrow{\bar{\mathrm{x}}\mathrm{x}})\, (\overrightarrow{\bar{\mathrm{x}}\mathrm{x}})^{T}\, p_{\mathbf{x}}(\mathrm{x})\, d\mathcal{M}(\mathrm{x})$$

The empirical covariance is defined in the same way using the discrete version of the expectation operator.

We observe that the covariance depends on the basis used for the exponential chart if we see it as a matrix, but it does not depend on it if we consider it as a bilinear form over the tangent plane.

The covariance is related to the variance just as in the vector case:

$$\mathrm{Tr}(\Sigma_{\mathbf{xx}}) = \mathbf{E}\left[\, \mathrm{Tr}(\overrightarrow{\bar{\mathrm{x}}\mathbf{x}}\, \overrightarrow{\bar{\mathrm{x}}\mathbf{x}}^{\mathrm{T}}) \,\right] = \mathbf{E}\left[\, \mathrm{dist}(\bar{\mathrm{x}}, \mathbf{x})^2 \,\right] = \sigma_{\mathbf{x}}^2$$

This formula is still valid relatively to any fixed point: $\mathrm{Tr}(\mathrm{Cov}_{\mathbf{x}}(\mathrm{y})) = \sigma_{\mathbf{x}}^2(\mathrm{y})$.

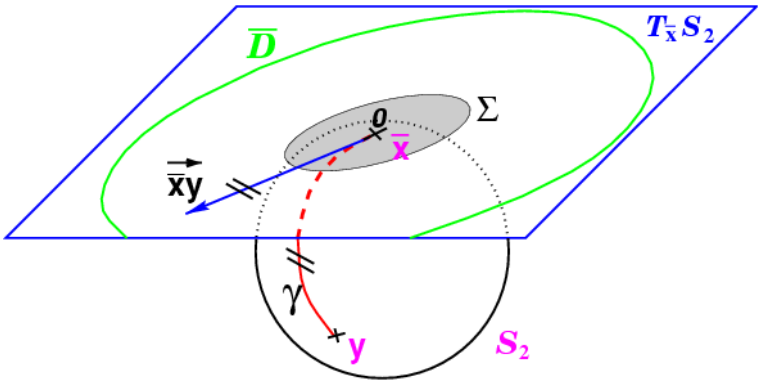

Figure 2: The covariance is defined in the tangent plane at $\mathcal{S}_2$ at the mean point as the classical covariance matrix of the random vector "deviation from the mean" $\Sigma_{\mathbf{xx}} = \mathbf{E}\left[\, \overrightarrow{\bar{\mathrm{x}}\mathbf{x}}\, \overrightarrow{\bar{\mathrm{x}}\mathbf{x}}^{\mathrm{T}} \,\right]$.

In fact, as soon as we have found a (or the) mean value and that the probability of its cut locus is null, everything appears to be similar to the case of a centered random vector by developing

the manifold onto the tangent space at the mean value. Indeed, $\overrightarrow{\bar{\mathrm{x}}\mathbf{x}}$ is a random vector of pdf $\rho_{\boldsymbol{x}}(y) = p_{\mathbf{x}}(y)\,\sqrt{|G(y)|}$ with respect to the Lebesgue measure in the connected and star-shaped domain $\mathcal{D}(\bar{\mathrm{x}}) \subset T_{\bar{\mathrm{x}}}\mathcal{M}$. We know that its expectation is $\mathbf{E}\left[\,\overrightarrow{\bar{\mathrm{x}}\mathbf{x}}\,\right] = 0$ and its covariance matrix is defined as usual. Thus, we could define higher order moments of the distribution by tensors on this tangent space, just as we have done for the covariance.

6 An information-based generalization of the Normal distribution

In this section, we turn to the generalization of the Gaussian distribution to Riemannian manifolds. Several generalizations have already be proposed so far. In the stochastic calculus community, one usually consider the heat kernel $p(x, y, t)$, which is the transition density of the Brownian motion [24, 59, 60]. This is the smallest positive fundamental solution to the heat equation $\frac{\partial f}{\partial t} - \Delta f = 0$, where Δ 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 λ_i with associated eigenfunctions f_i solutions of $\Delta f = \lambda f$). The practical problem is that the computation of this spectrum is impossible but in very few cases [42].

To obtain tractable formulas, several distributions have been proposed in directional statistics [16, 17, 18, 19, 61], 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. [61] for the circular and spherical case). It is easy to see that the wrapped Gaussian distribution tends towards the mass distribution if the variance goes to zero. In the circular case, one can also show that is tends toward the uniform distribution for a large variance. [15] extended this definition 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.

Thus, instead of keeping a Gaussian pdf in some tangent space, we propose in this paper a new variational approach based on global properties, consistent with the previous definitions of the mean and covariance. The property that we take for granted is the maximization of the entropy knowing the mean and the covariance matrix. For many aspects, this may not be the best generalization of the Gaussian. However, we demonstrate that it provides a family ranging from the point-mass distribution to the uniform measure (for compact manifolds) and that we can provide computationally tractable approximations for any manifold in case of small variances. In this section the symbols log and exp denote the standard logarithmic and exponential functions in $\mathbb{R}$.

6.1 Entropy and uniform law

As we can integrate a real valued function, the extension of the *entropy* $\mathbf{H}[\,\mathbf{x}\,]$ of a random point is straightforward:

$$\mathbf{H}[\,\mathbf{x}\,] = \mathbf{E}[\,-\log(p_{\mathbf{x}}(\mathbf{x}))\,] = -\int_{\mathcal{M}} \log(p_{\mathbf{x}}(\mathrm{x}))\, p_{\mathbf{x}}(\mathrm{x})\, d\mathcal{M}(\mathrm{x})$$

This definition is coherent with the measure inherited from our Riemannian metric since the pdf $p_{\mathcal{U}}$ that maximizes the entropy when we only know that the measure is in a compact set $\mathcal{U}$ is the uniform density in this set:

$$p_{\mathcal{U}}(\mathrm{x}) = \mathbf{1}_{\mathcal{U}}(\mathrm{x}) \Big/ \int_{\mathcal{U}} d\mathcal{M}(\mathrm{y})$$

6.2 Maximum entropy characterization

Now assume that we only know the mean (that we suppose to be unique) and the covariance of a random point: we denote it by $\mathbf{x} \sim (\bar{\mathrm{x}}, \Sigma)$. If we need to assume a pdf for that random point, it seems reasonable to choose the one which is the least informative, while fitting the mean and the covariance. The pdf is maximizing in this case the conditional entropy

$$\mathbf{H}\left[\, \mathbf{x} \,|\, \bar{\mathrm{x}} \in \mathbb{E}\left[\, \mathrm{x} \,\right] \,,\, \Sigma_{\mathbf{xx}} = \Sigma \,\right]$$

In standard multivariate statistics, this maximum entropy principle is one characterization of the Gaussian distributions [62, p. 409]. In the Riemannian case, we can express all the constraints directly in the exponential chart at the mean value. Let $\rho(y) = p(\exp_{\bar{\mathrm{x}}}(y))$ be the density in the chart with respect to the induced Riemannian measure $d\mathcal{M}_{\bar{\mathrm{x}}}(y) = \sqrt{|G_{\bar{\mathrm{x}}}(y)|}\, dy$ (we use here the Riemannian measure instead of the Lebesgue one to simplify equations below). The constraints are:

- the normalization: $\mathbf{E}\left[\, \mathbf{1}_{\mathcal{M}} \,\right] = \displaystyle\int_{\mathcal{D}(\bar{\mathrm{x}})} \rho(y)\, d\mathcal{M}_{\bar{\mathrm{x}}}(y) = 1$
- a nul mean value: $\mathbf{E}\left[\, \overrightarrow{\bar{\mathrm{x}}\mathbf{x}} \,\right] = \displaystyle\int_{\mathcal{D}(\bar{\mathrm{x}})} y\, \rho(y)\, d\mathcal{M}_{\bar{\mathrm{x}}}(y) = 0,$
- and a fixed covariance Σ : $\mathbf{E}\left[\, \overrightarrow{\bar{\mathrm{x}}\mathbf{x}}\, \overrightarrow{\bar{\mathrm{x}}\mathbf{x}}^{\mathrm{T}} \,\right] = \displaystyle\int_{\mathcal{D}(\bar{\mathrm{x}})} y\, y^{\mathrm{T}}\, \rho(y)\, d\mathcal{M}_{\bar{\mathrm{x}}}(y) = \Sigma$

To simplify the optimization, we won't consider any continuity or differentiability constraint on the cut locus $C(\bar{\mathrm{x}})$ (which would mean constraints on the border of the domain). This means that we can do the optimization in the exponential chart at the mean point as if we were in the open domain $\mathcal{D}(\bar{\mathrm{x}}) \in \mathbb{R}^n$.

Using the convexity of the function $-x\log(x)$, one can show that the maximum entropy is attained by distributions of density $\rho(y) = k\, \exp\left(-\beta^{\mathrm{T}}\, y - \frac{y^{\mathrm{T}}\, \Gamma\, y}{2}\right)$, if there exists constants k, β and Γ such that our constrains are fulfilled [62, Theorem 13.2.1, p. 409]. Assuming that the definition domain $\mathcal{D}(\bar{\mathrm{x}})$ is symmetric with respect to the origin, we find that $\beta = 0$ ensures a null mean. Substituting in the constraints gives the following equations.

Theorem 3 (Normal law)
We call Normal law on the manifold $\mathcal{M}$ the maximum entropy distribution knowing the mean value and covariance. Assuming no continuity nor differentiability constraint on the cut locus $C(\bar{\mathrm{x}})$ and a symmetric domain $\mathcal{D}(\bar{\mathrm{x}})$, the pdf of the Normal law of mean $\bar{\mathrm{x}}$ (the mean value) and concentration matrix Γ is given by:

$$N_{(\bar{\mathrm{x}},\Gamma)}(\mathrm{y}) = k\, \exp\left(-\frac{\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}^{\,T}\, \Gamma\, \overrightarrow{\bar{\mathrm{x}}\mathrm{y}}}{2}\right)$$

where the normalization constant and the covariance are related to the concentration matrix by:

$$k^{(-1)} = \int_{\mathcal{M}} \exp\left(-\frac{\overrightarrow{\bar{x}y}^T \, \Gamma \, \overrightarrow{\bar{x}y}}{2}\right) d\mathcal{M}(y) \qquad \textit{and} \qquad \Sigma = k \int_{\mathcal{M}} \overrightarrow{\bar{x}y} \, \overrightarrow{\bar{x}y}^T \exp\left(-\frac{\overrightarrow{\bar{x}y}^T \, \Gamma \, \overrightarrow{\bar{x}y}}{2}\right) d\mathcal{M}(y)$$

From the concentration matrix, we can compute the covariance of the random point, at least numerically, but the reverse is more difficult.

6.3 The vectorial case

The integrals can be entirely computed, and we find $k^{(-1)} = \frac{(2\pi)^{\frac{n}{2}}}{\sqrt{|\Gamma|}}$ and $\Gamma = \Sigma^{(-1)}$. The Normal density is thus the usual Gaussian density:

$$N_{(\bar{x},\Gamma)}(x) = k \, \exp\left(-\frac{(x-\bar{x})^{\mathrm{T}} \, \Gamma \, (x-\bar{x})}{2}\right) = \frac{1}{(2\,\pi)^{n/2}\sqrt{|\Sigma|}} \, \exp\left(-\frac{(x-\bar{x})^{\mathrm{T}} \, \Sigma^{(-1)} \, (x-\bar{x})}{2}\right)$$

6.4 Example on a simple manifold: the circle

The exponential chart for the circle of radius 1 with the canonical metric is the angle $\theta \in \mathcal{D} =]-\pi;\pi[$ with respect to the development point, and the measure is simply $d\theta$. For a circle of radius r, the exponential chart becomes $x = r\,\theta$. The domain is $\mathcal{D} =]-a;a[$ (with $a = \pi\,r$) and the measure is $dx = r\,d\theta$. Thus, the normalization factor of the Normal density is:

$$k^{(-1)} = \int_{-a}^{a} \exp\left(-\frac{\gamma\,x^2}{2}\right) dx = \sqrt{\frac{2\pi}{\gamma}} \, \mathrm{erf}\left(\sqrt{\frac{\gamma}{2}}\,a\right)$$

where erf is the error function $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2)\,dt$. The density is the truncated Gaussian $N_{(0,\gamma)}(x) = k \, \exp\left(-\frac{\gamma\,x^2}{2}\right)$. It is continuous but not differentiable on the cut locus $\pi \equiv -\pi$. The truncation introduces a bias in the relation between the variance and the concentration parameter:

$$\sigma^2 = \int_{-a}^{a} x^2 \, k \, \exp\left(-\frac{\gamma\,x^2}{2}\right) dx = \frac{1}{\gamma}\left(1 - 2\,a\,k\,\exp\left(-\frac{\gamma\,a^2}{2}\right)\right)$$

It is interesting to have a look on limit properties: if the circle radius goes to infinity, the circle becomes the real line and we obtain the usual Gaussian with the relation $\sigma^2 = 1/\gamma$, as expected. Now, let us consider a circle with a fixed radius. As anticipated, the variance goes to zero and the density tends toward a point mass distribution (see figure 3) if the concentration γ goes to infinity. On the other hand, the variance cannot becomes infinite (as in the real case) when the concentration parameter γ goes to zero because the circle is compact: a Taylor expansion gives $\sigma^2 = a^2/3 + O(\gamma)$. Thus, the maximal variance on the circle is

$$\sigma_0^2 = \lim_{\gamma \to 0} \sigma^2 = \frac{a^2}{3} \qquad \text{with the density} \qquad N_{(0,0)}(x) = \frac{1}{2\,a}$$

Interestingly, the Normal density of concentration 0 is the uniform density. In fact, this result can be generalized to all compact manifolds

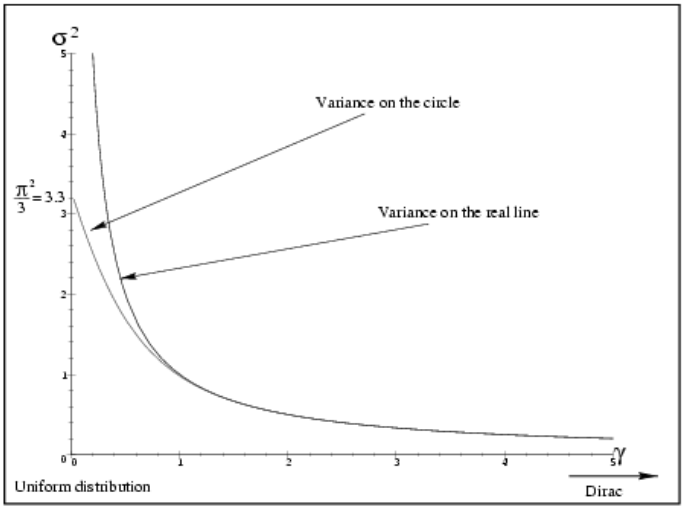

Figure 3: Variance σ^2 with respect to the concentration parameter γ on the circle of radius 1 and the real line. This variance tends toward $\sigma_0^2 = \pi^2/3$ for the uniform distribution on the circle ($\gamma = 0$) whereas it tends to infinity for the uniform measure on $\mathbb{R}$. For a strong concentration ($\gamma > 1$), the variance on the circle can be accurately approximated by $\sigma^2 \simeq 1/\gamma$, as in the real case.

6.5 Small concentration matrix on compact manifolds: uniform distribution

Let n be the dimension of the manifold. Using bounds on the eigenvalues of Γ, we have: $\mathrm{Tr}(\Gamma)\|\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}\|^2/n \leq \overrightarrow{\bar{\mathrm{x}}\mathrm{y}}^{\mathrm{T}}\,\Gamma\,\overrightarrow{\bar{\mathrm{x}}\mathrm{y}} \leq \mathrm{Tr}(\Gamma)\|\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}\|^2$. This means that we can bound the exponential by:

$$1 + \sum_{k=1}^{+\infty} \frac{(-1)^k\,\mathrm{Tr}(\Gamma)^k\,\|\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}\|^{2k}}{k!\,2^k} \leq \exp(-\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}^{\mathrm{T}}\,\Gamma\,\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}/2) \leq 1 + \sum_{k=1}^{+\infty} \frac{(-1)^k\,\mathrm{Tr}(\Gamma)^k\,\|\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}\|^{2k}}{k!\,2^k\,n^k}$$

As all the moments $\int_{\mathcal{M}} \|\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}\|^{2k}\,d\mathcal{M}$ are finite since the manifold is compact, one can conclude that

$$k^{(-1)} = \int_{\mathcal{M}} \exp(-\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}^{\mathrm{T}}\,\Gamma\,\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}/2)\,d\mathcal{M} = Vol(\mathcal{M}) + O(\mathrm{Tr}(\Gamma))$$

It follows immediately that the limit of the normal distribution is the uniform one, and that the limit covariance is finite.

Theorem 4 *Let $\mathcal{M}$ be a compact manifold. The limit of the normal distribution for small concentration matrices is the uniform density $N(\mathrm{y}) = 1/Vol(\mathcal{M}) + O(\mathrm{Tr}(\Gamma))$. Moreover, the covariance matrix tends towards a finite value:*

$$\Sigma = \frac{1}{Vol(\mathcal{M})} \int_M \overrightarrow{\bar{\mathrm{x}}\mathrm{y}}\,\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}^{T}\,d\mathcal{M} + O(\mathrm{Tr}(\Gamma)) < +\infty$$

6.6 Approximation for a large concentration matrix

If the pdf is sufficiently concentrated (a high concentration matrix Γ or a small covariance matrix Σ), then we can use a Taylor expansion of the metric in a normal coordinate system around the mean value to approximate the previous integrals and obtain a Taylor expansions of the normalization factor and the concentration matrix with respect to the covariance matrix.

The Taylor expansion of the metric is given by [63, p84]. We easily deduce the Taylor expansion of the measure around the origin (Ric is the Ricci (or scalar) curvature matrix in the considered normal coordinate system):

$$d\mathcal{M}(y) = \sqrt{\det(G(y))}\, dy = \left(1 - \frac{y^{\mathrm{T}}\, \mathrm{Ric}\; y}{6} + O(\|y\|^3)\right) dy$$

Substituting this Taylor expansion in the integrals and manipulating the formulas (see appendix B) leads to the following theorem.

Theorem 5 (Approximate normal density)
In a complete Riemannian manifold, the normal density is $N(\mathrm{y}) = k\, \exp(-\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}^T\, \Gamma\, \overrightarrow{\bar{\mathrm{x}}\mathrm{y}}/2)$. *Let* $r = i(\mathcal{M}, \bar{\mathrm{x}})$ *be the injectivity radius at the mean point (by convention* $r = +\infty$ *if there is no cut-locus). Assuming a finite variance for any concentration matrix* Γ*, we have the following approximations of the normalization constant and concentration matrix for a covariance matrix* Σ *of small variance* $\sigma^2 = \mathrm{Tr}(\Sigma)$*:*

$$k = \frac{1 + O(\sigma^3) + \varepsilon\left(\frac{\sigma}{r}\right)}{\sqrt{(2\pi)^n\, \det(\Sigma)}} \qquad \textit{and} \qquad \Gamma = \Sigma^{(\text{-}1)} - \frac{1}{3}\mathrm{Ric} + O(\sigma) + \varepsilon\left(\frac{\sigma}{r}\right)$$

Here, $\varepsilon(x)$ *is a function that is a* $O(x^k)$ *for any positive* k*, with the convention that* $\varepsilon\left(\frac{\sigma}{+\infty}\right) = \varepsilon(0) = 0$*. More precisely, this is a function such that* $\forall k \in \mathbb{R}^+,\ \lim_{0+} x^{-k}\, \varepsilon(x) = 0$

6.7 Discussion

The maximum entropy approach to generalize the normal distribution to Riemannian manifolds is interesting since we obtain a whole family of densities going from the Dirac measure to the uniform distribution (or the uniform measure if the manifold is only locally compact). Unfortunately, this distribution is generally not differentiable at the cut locus, and often even not continuous.

However, if the relation between the parameters and the moments of the distribution are not as simple as in the vector case (but can we expect something simpler in the general case of Riemannian manifolds?), the approximation for small covariances turns out to be rather simple. Thus, this approximate distribution can be handled quite easily for computational purposes. It is likely that similar approximations hold for wrapped Gaussian, but this remains to establish.

7 Mahalanobis distance and χ^2 law

The problem we are now tackling is to determine if an observation $\hat{\mathrm{x}}$ could reasonably have come from a given probability distribution $\mathbf{x}$ on $\mathcal{M}$ or if it should be considered as an outlier. From a computational point of view, the pdf of the measurement process is often too rich an information to be estimated or handled. In practice, one often characterizes this pdf by its moments, and more particularly by the mean and the covariance. We denote it by $\mathbf{x} \sim (\bar{\mathrm{x}}, \Sigma_{\mathbf{xx}})$. Based on these characteristics only, we have to decide if the observation $\hat{\mathrm{x}}$ is compatible with this measurement process.

In the vector case, a well adapted tool is the Mahalanobis distance $\mu^2 = (\hat{x} - \bar{x})^{\mathrm{T}}\, \Sigma^{(\text{-}1)}_{\boldsymbol{xx}}\, (\hat{x} - \bar{x})$, which measures the distance between the observation $\hat{\mathrm{x}}$ and the mean value $\bar{\mathrm{x}}$ according to the "metric" $\Sigma^{(\text{-}1)}_{\mathbf{xx}}$. To test if the observation $\hat{\mathrm{x}}$ is an outlier, the principle of the Mahalanobis D^2 test is to compute the tail probabilities of the resulting distribution (the so called p-value), i.e. the risk

of error when saying that the observation does not come from the distribution. The distribution of $\mathbf{x}$ is usually assumed to be Gaussian, as this distribution minimizes the added information (i.e. minimizes the entropy) when we only know the mean and the covariance. In that case, we know that the Mahalanobis distance should be χ^2_n distributed if the observation is correct (n is the dimension of the vector space). If the probability of the current distance is too small (i.e. μ^2 is too large), the observation $\hat{x}$ can safely be considered as an outlier.

The definition of the Mahalanobis distance can be easily generalized to complete Riemannian manifolds with our tools. We note that it is well defined for any distribution of the random point and not only the normal one.

Definition 7 (Mahalanobis distance)
We call Mahalanobis distance between a random point $\mathbf{x} \sim (\bar{\mathrm{x}}, \Sigma_{\mathbf{xx}})$ *and a (deterministic) point* y *on the manifold the value*

$$\mu^2_{\mathbf{x}}(\mathrm{y}) = \overrightarrow{\bar{\mathrm{x}}\mathrm{y}}^{\,T}\, \Sigma^{(\text{-}1)}_{\mathbf{xx}}\, \overrightarrow{\bar{\mathrm{x}}\mathrm{y}}.$$

7.1 Properties

Since $\mu^2_{\mathbf{x}}$ is a function from $\mathcal{M}$ to $\mathbb{R}$, $\mu^2_{\mathbf{x}}(\mathbf{y})$ is a real random variable. The expectation of this random variable is well defined and turns out to be quite simple:

$$\begin{aligned}\mathbf{E}\left[\, \mu^2_{\mathbf{x}}(\mathbf{y}) \,\right] &= \int_{\mathcal{M}} \mu^2_{\mathbf{x}}(\mathrm{z})\, p_{\mathbf{y}}(\mathrm{z})\, d\mathcal{M}(\mathrm{z}) = \int_{\mathcal{M}} \overrightarrow{\bar{\mathrm{x}}\mathrm{z}}^{\mathrm{T}}\, \Sigma^{(\text{-}1)}_{\mathbf{xx}}\, \overrightarrow{\bar{\mathrm{x}}\mathrm{z}}\, p_{\mathbf{y}}(\mathrm{z})\, d\mathcal{M}(\mathrm{z}) \\ &= \mathrm{Tr}\left(\Sigma^{(\text{-}1)}_{\mathbf{xx}} \int_{\mathcal{M}} \overrightarrow{\bar{\mathrm{x}}\mathrm{z}}\, \overrightarrow{\bar{\mathrm{x}}\mathrm{z}}^{\mathrm{T}}\, p_{\mathbf{y}}(\mathrm{z})\, d\mathcal{M}(\mathrm{z}) \right) = \mathrm{Tr}\left(\Sigma^{(\text{-}1)}_{\mathbf{xx}}\, \mathrm{Cov}_{\boldsymbol{y}}(\bar{\mathrm{x}}) \right)\end{aligned}$$

The expectation of the Mahalanobis distance of a random point with itself is even simpler:

$$\mathbf{E}\left[\, \mu^2_{\mathbf{x}}(\mathbf{x}) \,\right] = \mathrm{Tr}(\Sigma^{(\text{-}1)}_{\mathbf{xx}}\, \Sigma_{\mathbf{xx}}) = \mathrm{Tr}(\,\mathrm{Id}_n) = n$$

Theorem 6 *The expected Mahalanobis distance of a random point with itself is independent of the distribution and does only depend on the dimension of the manifold:* $\mathbf{E}\left[\, \mu^2_{\mathbf{x}}(\mathbf{x}) \,\right] = n$.

This identity can be used to verify with a posteriori observations that the covariance matrix has been correctly estimated. It can be compared with the expectation of the "normalized" squared distance, which is by definition: $\mathbf{E}\left[\, \mathrm{dist}(\mathbf{x}, \bar{\mathrm{x}})^2/\sigma^2_{\mathbf{x}} \,\right] = 1$.

7.2 A generalized χ^2 law

Assuming that the random point $\mathbf{x} \sim (\bar{\mathrm{x}}, \Sigma_{\mathbf{xx}})$ is normal, we can go one step further and compute the probability that $\chi^2 = \mu^2_{\mathbf{x}} < \alpha^2$ (see appendix C). This generalization of the χ^2 law turns out to be still independent of the mean value and the covariance matrix of the random point (at least up to the order $O(\sigma^3)$):

Theorem 7 (Approximate χ^2 law)
With the same hypotheses as for the approximate normal law, the χ^2 *probability is*

$$\Pr\{\chi^2 \leq \alpha^2\} = (2\pi)^{-\frac{n}{2}} \int_{\|x\| \leq \alpha} \exp\left(-\frac{\|x\|^2}{2} \right)\, dx + O(\sigma^3) + \varepsilon\left(\frac{\sigma}{r}\right)$$

while the density is

$$p_{\chi^2}(u) = \frac{1}{2\,\boldsymbol{\Gamma}\left(\frac{n}{2}\right)} \left(\frac{u}{2}\right)^{\frac{n}{2}-1} \exp\left(-\frac{u}{2}\right) + O(\sigma^3) + \varepsilon\left(\frac{\sigma}{r}\right)$$

The χ^2 probability can be computed using the incomplete gamma function $\Pr\{\chi^2 \leq \alpha^2\} = P\left(\frac{n}{2}, \frac{\alpha^2}{2}\right)$ (see for instance [64]).

In practice, one often use this law to test if an observation $\hat{\mathrm{x}}$ has been drawn from a random point $\mathbf{x}$ that we assume to be Gaussian: if the hypothesis is true, the value $\mu^2_{\mathbf{x}}(\hat{\mathrm{x}})$ will be less than α^2 with a probability $\gamma = \Pr\{\chi^2 \leq \alpha^2\}$. Thus, one choose a confidence level γ (for instance 95% or 99%), then we find the value $\alpha(\gamma)$ such that $\gamma = \Pr\{\chi^2 \leq \alpha^2\}$ and accept the hypothesis if $\mu^2_{\mathbf{x}}(\hat{\mathrm{x}}) \leq \alpha^2$.

8 Discussion

On a (geodesically complete) Riemannian manifold, it is easy to define probability density functions associated to random points, thanks to the availability of a metric. However, as soon as the expectation is concerned, we may only define the expectation of an observable (a real or vectorial function of the random point). Thus, the definition of a mean value for a random point is much more complex than for the vectorial case and it requires a distance-based variational formulation: the Riemannian center of mass basically minimize locally the variance. As the mean is now defined through a minimization procedure, its existence and uniqueness are not ensured any more (except for distributions with a sufficiently small compact support). In practice, one mean value almost always exists, and it is unique as soon as the distribution is sufficiently peaked. The properties of the mean are very similar to those of the modes (that can be defined as central values of order 0) in the vectorial case. We present here a new proof of the barycentric characterization theorem that is valid for distribution with any kind of support. The main difference with the vector case is that we have to ensure a null probability measure of the cut locus. To compute the mean value, we designed an original Gauss-Newton gradient descent algorithm that essentially alternates the computation of the barycenter in the exponential chart centered at the current estimation, and a re-centering step of the chart at the newly computed barycenter.

To define higher moments of the distribution, we used the exponential chart at the mean point (which may be seen as the development of the manifold onto its tangent space at this point along the geodesics): the random point is thus represented as a random vector with null mean in a star-shaped and symmetric domain. With this representation, there is no more problem to define the covariance matrix and potentially higher order moments. Based on this covariance matrix, we defined a Mahalanobis distance between a random and a deterministic point that basically weights the distance between the deterministic point and the mean point using the inverse of the covariance matrix. Interestingly, the expected Mahalanobis distance of a random point with itself is independent of the distribution and is equal to the dimension of the manifold, as in the vectorial case.

Like for the mean, we choose a variational approach to generalize the Normal law: we define it as the maximum entropy distribution knowing the mean and the covariance. Neglecting the cut-locus constraints, we show that this amounts to consider a truncated Gaussian distribution on the exponential chart centered at the mean point. However, the relation between the concentration matrix (the "metric" used in the exponential of the pdf) and the covariance matrix is slightly more complex that the simple inversion of the vectorial case, as it has to be corrected for the curvature of the manifold.

Last but not least, using the Mahalanobis distance of a Normally distributed random point, we can generalize the χ^2 law: we were able to show that is has the same density as in the vectorial case up to an order 3 in σ. This opens the way to the generalization of many other statistical tests,

as we may expect similarly simple approximations for sufficiently centered distributions.

In this paper, we focused on purpose on the theoretical formulation of the statistical framework in geodesically complete Riemannian spaces, eluding applications examples. The interested reader will find practical applications in computer vision to compute the mean rotation [32, 33] or for the generalization of matching algorithms to arbitrary geometric features [65]. In medical image analysis, selected applications cover the validation of the rigid registration accuracy [4, 66, 67], shape statistics [7] and more recently tensor computing, either for processing and analyzing diffusion tensor images [10, 8, 9, 11], or to model the brain variability [12]. One can even find applications in rock mechanics with the analysis of fracture geometry [68].

In the theory presented here, all definitions are derived from the Riemannian metric of the manifold. A natural question is how to chose this metric? Invariance properties requirements provide a partial answer for connected Lie groups and homogeneous manifolds [1, 45]. However, an invariant metric does not always exist on an homogeneous manifolds. Likewise, left and right invariant metric and generally different in non-compact Lie groups, so that we only have a partial consistency between the geometric and statistical operations. Another way to chose the metric could be to estimate it from empirical data.

More generally, we could design other definitions of the mean value using the notion of connector introduced in [29]. This connector formalizes a relationship between the manifold and its tangent space at one point, exactly in the way we used the exponential map of the Riemannian metric. Thus, we could generalize easily the higher order moments and the other statistical operations we defined by replacing everywhere the exponential map with the connector. One important point is that these connectors can model extrinsic distances (like the Euclidean distance on unit vectors), and could lead to very efficient approximations of the mean value for sufficiently peaked distributions. For instance, the "barycenter / re-centering" algorithm we designed will most probably converge toward a first order approximation of the Riemannian mean if we use any chart that is consistent to the exponential chart at the first order (e.g. Euler's angle on re-centered 3D rotations). We believe that this research track may become one of the most productive from the practical point of view.

Acknowledgments

The main part of this work was done at MIT, Artificial Intelligence Lab in 1997. If all the ideas presented in this paper were in place at that time [69], the formula for gradient of the variance (appendix A) remained a conjecture. The author would like to thank specially Pr Maillot for providing a first proof of the differentiability of the variance for the uniform distribution on compact manifolds [53]. Its generalization to the non compact case with arbitrary distributions considerably delayed the submission of this paper. In the meantime, other (and simpler) proofs were proposed [15, 38]. The author would also like to thank two anonymous referees for very valuable advises and for the detection of a number of technical errors.

A Gradient of the variance

This proof is a generalization of a differentiability proof for the *uniform* distribution on *compact* manifolds by Pr Maillot [53]. The main difficulties were to remove the compactness hypothesis, and to generalize to arbitrary distributions.

Hypotheses

Let P be a probability on the Riemannian manifold $\mathcal{M}$. We assume that the cut locus $C(\mathrm{y})$ of the derivation point $\mathrm{y} \in \mathcal{M}$ has a null measure with respect to this probability (as it has with the Riemannian measure) and that the variance is finite at that point:

$$P(C(\mathrm{y})) = \int_{C(\mathrm{y})} dP(\mathrm{z}) = 0 \qquad \text{and} \qquad \sigma^2(\mathrm{y}) = \int_{\mathcal{M}} \mathrm{dist}(\mathrm{y},\mathrm{z})^2 \, dP(\mathrm{z}) < \infty$$

Goal and problem

Let now $\vec{\mathrm{g}}(\mathrm{y}) = \int_{\mathcal{M}\setminus C(\mathrm{y})} \overrightarrow{\mathrm{yz}} \, dP(\mathrm{z})$. As $\|\overrightarrow{\mathrm{yz}}\| = d(\mathrm{z},\mathrm{y}) \leq 1 + d(\mathrm{z},\mathrm{y})^2$, and using the null probability of the cut locus, we have:

$$\|\vec{\mathrm{g}}(\mathrm{y})\| \leq \int_{\mathcal{M}\setminus C(\mathrm{y})} \|\overrightarrow{\mathrm{yz}}\| \, dP(\mathrm{z}) \leq \int_{\mathcal{M}\setminus C(\mathrm{y})} \left(1 + d(\mathrm{z},\mathrm{y})^2\right) \, dP(\mathrm{z}) = 1 + \sigma^2(\mathrm{y}) < \infty$$

Thus $\vec{\mathrm{g}}(\mathrm{y})$ is well defined everywhere. Let $h(\mathrm{y}) = d(\mathrm{z},\mathrm{y})^2 = \|\overrightarrow{\mathrm{yz}}\|^2$. For a fixed $\mathrm{z} \notin C(\mathrm{y})$ (which is equivalent to $\mathrm{y} \notin C(\mathrm{z})$), we have $(\mathrm{grad}\, h)(\mathrm{y}) = -2\overrightarrow{\mathrm{yz}}$. Thus the proposition:

$$(\mathrm{grad}\, \sigma^2)(\mathrm{y}) = -2 \int_{\mathcal{M}/C(\mathrm{y})} \overrightarrow{\mathrm{yz}} \, dP(\mathrm{z})$$

corresponds to a derivation under the sum, but the usual conditions of the Lebesgue theorem are not fulfilled: the zero measure set $C(\mathrm{y})$ varies with y. Thus, we have to come back to the original definition of the gradient.

Definition of the gradient

Let $\gamma(t)$ be a curve with $\gamma(0) = \mathrm{y}$ and $\dot{\gamma}(0) = w$. By definition, the function $\sigma^2 : \mathcal{M} \to R$ is derivable if there exists a vector $(\mathrm{grad}\, \sigma^2)(\mathrm{y}) \in T_{\mathrm{y}}\mathcal{M}$ (the gradient) such that:

$$\forall w \in T_{\mathrm{y}}\mathcal{M} \quad \left\langle \, (\mathrm{grad}\, \sigma^2)(\mathrm{y}) \mid w \, \right\rangle = \partial_w \sigma^2(y) = \lim_{t \to 0} \frac{\sigma^2(\gamma(t)) - \sigma^2(\mathrm{y})}{t}$$

Since tangent vectors are defined as equivalent classes, we can choose the geodesic curve $\gamma(t) = \exp_{\mathrm{y}}(t\, w)$. Using $v = t\, w$, the above condition can then be rewritten:

$$\forall v \in T_{\mathrm{y}}\mathcal{M} \quad \lim_{\|v\| \to 0} \frac{\sigma^2(\exp_{\mathrm{y}}(v)) - \sigma^2(\mathrm{y}) - \left\langle \, (\mathrm{grad}\, \sigma^2)(\mathrm{y}) \mid v \, \right\rangle}{\|v\|} = 0$$

which can be rephrased as: for all $\eta \in \mathbb{R}_+$, there exists ε sufficiently small such that:

$$\forall v \in T_{\mathrm{y}}\mathcal{M},\ \|v\| < \varepsilon \quad \left| \sigma^2(\exp_{\mathrm{y}}(v)) - \sigma^2(\mathrm{y}) - \left\langle \, (\mathrm{grad}\, \sigma^2)(\mathrm{y}) \mid v \, \right\rangle \right| \leq \eta \, \|v\|$$

General idea

Let $\Delta(\mathrm{z}, v)$ be the integrated function (for $\mathrm{z} \notin C(\mathrm{y})$):

$$\Delta(\mathrm{z}, v) = \ \mathrm{dist}(\exp_{\mathrm{y}}(v), \mathrm{z})^2 - \ \mathrm{dist}(\mathrm{y},\mathrm{z})^2 + 2 \left\langle \, \overrightarrow{\mathrm{yz}} \mid v \, \right\rangle$$

and $H(v) = \int_{\mathcal{M}\setminus C(\mathrm{y})} \Delta(\mathrm{z}, v)\, dP(\mathrm{z})$ be the function to bound:

$$\begin{aligned} H(v) &= \sigma^2(\exp_{\mathrm{y}}(v)) - \sigma^2(\mathrm{y}) - \left\langle\, (\operatorname{grad}\sigma^2)(\mathrm{y}) \mid v \,\right\rangle \\ &= \int_{\mathcal{M}} \operatorname{dist}(\exp_{\mathrm{y}}(v), \mathrm{z})^2\, dP(\mathrm{z}) - \int_{\mathcal{M}} \operatorname{dist}(\mathrm{y}, \mathrm{z})^2\, dP(\mathrm{z}) + 2\int_{\mathcal{M}\setminus C(\mathrm{y})} \left\langle\, \overrightarrow{\mathrm{yz}} \mid v \,\right\rangle\, dP(\mathrm{z}) \end{aligned}$$

The idea is to split this integral in two in order to bound Δ on a small neighborhood W around the cut locus of y and to use the standard Lebesgue theorem to bound the integral of Δ on $\mathcal{M}\setminus W$.

Lemma 1 *$W_\varepsilon = \bigcup_{\mathrm{x}\in\mathcal{B}(\mathrm{y},\varepsilon)} C(\mathrm{x})$ is a continuous series of included and decreasing open sets all containing $C(\mathrm{y})$ and converging towards it.*

Let us first reformulate the definition of W_ε:

$$\mathrm{z} \in W_\varepsilon \quad\Leftrightarrow\quad \exists \mathrm{x} \in \mathcal{B}(\mathrm{y},\varepsilon) \,/\, z \in C(\mathrm{x}) \quad\Leftrightarrow\quad \exists \mathrm{x} \in \mathcal{B}(\mathrm{y},\varepsilon) \,/\, x \in C(\mathrm{z}) \quad\Leftrightarrow\quad C(\mathrm{z}) \cap \mathcal{B}(\mathrm{y},\varepsilon) \neq \varnothing$$

Going to the limit, we have: $\mathrm{z} \in W_0 = \lim_{\varepsilon\to 0} W_\varepsilon \Leftrightarrow C(\mathrm{z}) \cap \{\mathrm{y}\} \neq \varnothing \Leftrightarrow \mathrm{z} \in C(\mathrm{y})$. Thus, we have a continuous series of included and decreasing sets all containing $C(\mathrm{y})$ and converging toward it. Now, let us prove that W_ε is an open set.

Let $U = \{u = (\mathrm{x}, v) \in \mathcal{M} \times T_{\mathrm{x}}\mathcal{M}\,;\, \|v\| = 1\}$ be the unit tangent bundle of $\mathcal{M}$ and $\rho : U \to \bar{R}_+ = \mathbb{R}_+ \cup \{+\infty\}$ be the cutting abscissa of the geodesic starting at x with the tangent vector v. Let now $\underline{U} = \rho^{(-1)}(\mathbb{R}_+)$ be the subset of the unit tangent bundle where $\rho(u) < +\infty$. The function ρ being continuous on U, the subspace $\underline{U} = \rho^{(-1)}([0, +\infty[)$ is open. Let $\pi : U \to \mathcal{M}$ be the canonical projection along the fiber $\pi((\mathrm{x}, v)) = \mathrm{x}$. This is a continuous and open map. Let us denote by $U_{\mathrm{x}} = \pi^{(-1)}(\{\mathrm{x}\})$ the unit tangent bundle at point x and $\underline{U}_{\mathrm{x}} = \pi^{(-1)}(\{\mathrm{x}\}) \cap \underline{U}$ its subset that lead to a cutting point of x.

Consider $u = (\mathrm{x}, v) \in \underline{U}_{\mathrm{x}}$ and the geodesic $\exp_{\mathrm{x}}(t\,\rho(u)\,v)$ for $t \in [0; 1]$: it is starting from x with tangent vector v and arriving at $\mathrm{z} = \exp_{\mathrm{x}}(\rho(u)\,v)$ with the same tangent vector by parallel transportation. Reverting the time course $(t \to 1-t)$, we have a geodesic starting at z with tangent vector $-v$ and arriving at x with the same tangent vector. By definition of the cutting function, we have $\rho(u) = \rho(u')$ with $u' = (\mathrm{z}, -v)$. Thus, the function $q(u) = (\exp_{\mathrm{x}}(\rho(u)\,v), -v)$ is a continuous bijection from $\underline{U}$ into itself with $q^{(-1)} = q$.

Let $\mathrm{z} \in W_\varepsilon$. By definition of W_ε, z is in the cut locus of a point $\mathrm{x} \in \mathcal{B}(\mathrm{y},\varepsilon)$: there exists a unit tangent vector v at that point such that $z = \pi(q(x, v))$. Conversely, the cut locus of z intersects $\mathcal{B}(\mathrm{y},\varepsilon)$: there exists a unit tangent vector v at z such that $\mathrm{x} = \pi(q(z, v)) \in \mathcal{B}(\mathrm{y},\varepsilon)$. Thus, we can rewrite $W_\varepsilon = \pi(U_\varepsilon)$ where $U_\varepsilon = q^{(-1)}\left(\pi^{(-1)}\left(\mathcal{B}(\mathrm{y},\varepsilon)\right) \cap \underline{U}\right)$. The functions π and q being continuous, U_ε is open. Finally, π being an open map, we conclude that W_ε is open.

Lemma 2 *Let $\mathrm{y} \in \mathcal{M}$ and $\alpha > 0$. Then there exists an open neighborhood W_ε of $C(\mathrm{y})$ such that*

(i) For all $\mathrm{x} \in \mathcal{B}(\mathrm{y},\varepsilon)$, $C(\mathrm{x}) \subset W_\varepsilon$,

(ii) $P(W_\varepsilon) = \int_{W_\varepsilon} dP(\mathrm{z}) < \alpha$

(iii) $\int_{W_\varepsilon} \operatorname{dist}(\mathrm{y}, \mathrm{z})\, dP(\mathrm{z}) < \alpha$

By hypothesis, the cut locus $C(\mathrm{y})$ has a null measure for the measures $dP(\mathrm{z})$. The distance being a measurable function, its measure is null on the cut locus: $\int_{C(\mathrm{y})} \operatorname{dist}(\mathrm{y}, \mathrm{z})\, dP(\mathrm{z}) = 0$. Thus, the functions $\int_{W_\varepsilon} dP(\mathrm{z})$ and $\int_{W_\varepsilon} \operatorname{dist}(\mathrm{y}, \mathrm{z})\, dP(\mathrm{z})$ are converging toward zero as ε goes to zero. By continuity, we can make both terms smaller than any positive α by choosing ε sufficiently small.

Bounding Δ on W_ε

Let W_ε, ε and α verifying the conditions of lemma 2 and $\mathrm{x}, \mathrm{x}' \in \mathcal{B}(\mathrm{y}, \varepsilon)$. We have $\mathrm{dist}(\mathrm{z}, \mathrm{x}) \leq \mathrm{dist}(\mathrm{z}, \mathrm{x}') + \mathrm{dist}(\mathrm{x}', \mathrm{x})$. Thus:

$$\mathrm{dist}(\mathrm{z}, \mathrm{x})^2 - \mathrm{dist}(\mathrm{z}, \mathrm{x}')^2 \leq \mathrm{dist}(\mathrm{x}, \mathrm{x}')\big(\mathrm{dist}(\mathrm{x}, \mathrm{x}') + 2\,\mathrm{dist}(\mathrm{z}, \mathrm{x}')\big)$$

Using the symmetry of x and x′ and the inequalities $\mathrm{dist}(\mathrm{x}, \mathrm{x}') \leq 2\varepsilon$ and $\mathrm{dist}(\mathrm{z}, \mathrm{x}') \leq \mathrm{dist}(\mathrm{z}, \mathrm{y}) + \varepsilon$, we have: $\big|\,\mathrm{dist}(\mathrm{z}, \mathrm{x})^2 - \mathrm{dist}(\mathrm{z}, \mathrm{x}')^2\big| \leq 2\,\mathrm{dist}(\mathrm{x}, \mathrm{x}')\big(2\varepsilon + \mathrm{dist}(\mathrm{z}, \mathrm{y})\big)$. Applying this bound to $\mathrm{x} = \exp_{\mathrm{y}}(v)$ and $\mathrm{x}' = \mathrm{y}$, we obtain:

$$\big|\,\mathrm{dist}(\mathrm{z}, \exp_{\mathrm{y}}(v))^2 - \mathrm{dist}(\mathrm{z}, \mathrm{y})^2\big| \leq 2\big(2\varepsilon + \mathrm{dist}(\mathrm{z}, \mathrm{y})\big)\,\|v\|$$

Now, the last term of $\Delta(\mathrm{z}, v)$ is easily bounded by: $\langle\, \overrightarrow{\mathrm{yz}} \mid v \,\rangle \leq \mathrm{dist}(\mathrm{y}, \mathrm{z})\,\|v\|$. Thus, we have: $|\Delta(\mathrm{z}, v)| \leq 4\big(\varepsilon + \mathrm{dist}(\mathrm{z}, \mathrm{y})\big)\,\|v\|$ and its integral over W_ε is bounded by:

$$\int_{W_\varepsilon} \Delta(\mathrm{z}, v)\, dP(\mathrm{z}) \leq 4\|v\| \int_{W_\varepsilon} \big(\varepsilon + \mathrm{dist}(\mathrm{z}, \mathrm{y})\big)\, dP(\mathrm{z}) < 8\alpha\|v\|$$

Bounding Δ on $\mathcal{M}\backslash W_\varepsilon$:

Let $\mathrm{x} = \exp_{\mathrm{y}}(v) \in \mathcal{B}(\mathrm{y}, \varepsilon)$. We know from lemma 2 that the cut locus $C(\mathrm{x})$ of such a point belongs to W_ε. Thus, the integration domain $\mathcal{M}\backslash W_\varepsilon$ is now independent of y and we can use the usual Lebesgue theorem to differentiate under the sum:

$$\mathrm{grad}\left(\int_{\mathcal{M}\backslash W_\varepsilon} \mathrm{dist}(\mathrm{y}, \mathrm{z})^2\, dP(\mathrm{z})\right) = -2\int_{\mathcal{M}\backslash W_\varepsilon} \overrightarrow{\mathrm{yz}}\, dP(\mathrm{z})$$

By definition, this means that for $\|v\|$ small enough, we have:

$$\int_{\mathcal{M}\backslash W_\varepsilon} \Delta(\mathrm{z}, v)\, dP(\mathrm{z}) < \alpha\|v\|$$

Conclusion

Thus, for $\|v\|$ small enough, we have $\int_{\mathcal{M}} \Delta(\mathrm{z}, v)\, dP(\mathrm{z}) < 9\alpha\|v\|$, which means that the variance has a derivative at the point y:

$$(\mathrm{grad}\,\sigma^2)(\mathrm{y}) = -2\int_{\mathcal{M}/C(\mathrm{y})} \overrightarrow{\mathrm{yz}}\, dP(\mathrm{z})$$

B Approximation of the generalized Normal density

In this section, we only work with a normal coordinate system at the mean value of the considered normal law. The generalized Normal density is: $N(y) = k \exp\left(-y^{\mathrm{T}}\,\Gamma\, y/2\right)$, where the normalization constant, the covariance and concentration are related by:

$$k^{(-1)} = \int_{\mathcal{M}} \exp\left(-\frac{y^{\mathrm{T}}\,\Gamma\, y}{2}\right) d\mathcal{M}(y) \qquad \text{and} \qquad \Sigma = k\int_{\mathcal{M}} y\, y^{\mathrm{T}} \exp\left(-\frac{y^{\mathrm{T}}\,\Gamma\, y}{2}\right) d\mathcal{M}(y)$$

We assume here that there two integrals converge toward a finite value for all concentration matrices Γ. In these expressions, the parameter is the concentration matrix Γ. The goal of this appendix is

to invert these relations in order to obtain a Taylor expansion of the concentration matrix and the normalization coefficient k with respect to the variance. This can be realized thanks to the Taylor expansion of the Riemannian metric around the origin in a normal coordinate system [63, section 2.8, corollary 2.3, p.84]: $\det(g_{ij}(\exp v)) = 1 - \mathrm{Ric}\,(v,v)/3 + O(\|v\|^3)$, where Ric is the expression of the Ricci tensor of scalar curvatures in the exponential chart. Thus [42, p.144]:

$$d\mathcal{M}(y) = \sqrt{\det(G(y))}\,dy = \left(1 - \frac{y^{\mathrm{T}}\,\mathrm{Ric}\;y}{6} + R_{d\mathcal{M}}(y)\right)dy \qquad \text{with} \qquad \lim_{y\to 0} \frac{R_{d\mathcal{M}}(y)}{\|y\|}^3 = 0$$

Let us investigate the normalization coefficient first. We have:

$$k^{(\text{-}1)} = \int_{\mathcal{D}} \exp\left(-\frac{y^{\mathrm{T}}\,\Gamma\,y}{2}\right)\,\left(1 - \frac{y^{\mathrm{T}}\,\mathrm{Ric}\;y}{6} + R_{d\mathcal{M}}(y)\right)dy$$

where $\mathcal{D}$ is the definition domain of the exponential chart. Since the concentration matrix Γ is a symmetric positive definite matrix, we have a unique symmetric positive definite square root $\Gamma^{1/2}$, which is obtained by changing the eigenvalues of Γ into their square root in a diagonalization. Using the change of variable $z = \Gamma^{1/2}\,y$ in the first two terms and the matrix $T = \Gamma^{-\frac{1}{2}}\mathrm{Ric}\,\Gamma^{-\frac{1}{2}}$ to simplify the notations, we have:

$$k^{(\text{-}1)} = \det(\Gamma)^{-1/2}\int_{\mathcal{D}'} \exp\left(-\frac{\|z\|^2}{2}\right)\left(1 - \frac{z^{\mathrm{T}}\,T\,z}{6}\right)\,dy + \int_{\mathcal{D}} \exp\left(-\frac{y^{\mathrm{T}}\,\Gamma\,y}{2}\right)\,R_{d\mathcal{M}}(y)\,dy$$

where the new definition domain is $\mathcal{D}' = \Gamma^{1/2}\mathcal{D}$. Likewise, we have for the covariance:

$$k^{(\text{-}1)}\Sigma = \frac{1}{\sqrt{\det(\Gamma)}}\int_{\mathcal{D}'} \Gamma^{-\frac{1}{2}}(zz^{\mathrm{T}})\Gamma^{-\frac{1}{2}} \exp\left(-\frac{\|z\|^2}{2}\right)\left(1 - \frac{z^{\mathrm{T}}Tz}{6}\right)dz + \int_{\mathcal{D}} yy^{\mathrm{T}} \exp\left(-\frac{y^{\mathrm{T}}\Gamma y}{2}\right)R_{d\mathcal{M}}(y)dy$$

B.1 Manifolds with an infinite injectivity radius at the mean point

At such a point, the cut locus is void and the definition domain of the exponential chart is $\mathcal{D} = \mathcal{D}' = \mathbb{R}^n$. This simplifies the integrals so that we can compute the first two terms explicitly.

The normalization coefficient: The first term is

$$\int_{\mathbb{R}^n} \exp\left(-\frac{\|z\|^2}{2}\right)\,dz = \prod_i \int_{\mathbb{R}} \exp\left(-\frac{z_i^2}{2}\right)\,dz_i = (2\pi)^{n/2}$$

By linearity, we also easily compute the second term:

$$\int_{\mathbb{R}^n} z^{\mathrm{T}}\,T\,z\,\exp\left(-\frac{\|z\|^2}{2}\right)\,dz = \mathrm{Tr}\left(T\int_{\mathbb{R}^n} z\,z^{\mathrm{T}}\,\exp\left(-\frac{\|z\|^2}{2}\right)\,dz\right) = (2\pi)^{n/2}\,\mathrm{Tr}(T)$$

Thus, we end up with:

$$k^{(\text{-}1)} = \frac{(2\pi)^{n/2}}{\sqrt{\det(\Gamma)}}\left(1 - \frac{\mathrm{Tr}(\Gamma^{(\text{-}1)}\,\mathrm{Ric}\,)}{6}\right) + \int_{\mathbb{R}^n} \exp\left(-\frac{y^{\mathrm{T}}\,\Gamma\,y}{2}\right)\,R_{d\mathcal{M}}(y)\,dy$$

Unfortunately, the last term is not so simple to simplify as we only know that the remainder $R_{d\mathcal{M}}(y)$ behaves as a $O(\|y\|^3)$ for small values of y. The idea is to split the right integral into one part around zero, say for $\|y\| < \alpha$, where we know how to bound the remainder, and to show

that $\exp(-\|y\|^2)$ dominate the Remainder elsewhere. Let γ_m be the smallest eigenvalue of Γ. As $y^{\mathrm{T}}\Gamma y \geq \gamma_m \|y\|^2$, we have:

$$\int_{\mathbb{R}^n} \exp\left(-\frac{y^{\mathrm{T}}\,\Gamma\,y}{2}\right)\, R_{d\mathcal{M}}(y)\,dy \leq \int_{\mathbb{R}^n} \exp\left(-\gamma_m \frac{\|y\|^2}{2}\right)\, R_{d\mathcal{M}}(y)\,dy$$

For any $\eta > 0$, there exists a constant $\alpha > 0$ such that $\|y\| < \alpha$ implies that $|R_{d\mathcal{M}}(y)| < \eta\,\|y\|^3$. Thus:

$$\begin{aligned}\int_{\|y\|<\alpha} \exp\left(-\gamma_m \frac{\|y\|^2}{2}\right)\, R_{d\mathcal{M}}(y)\,dy &< \eta \int_{\|y\|<\alpha} \exp\left(-\gamma_m \frac{\|y\|^2}{2}\right)\, \|y\|^3\,dy \\ &< \eta \int_{\mathbb{R}^n} \exp\left(-\gamma_m \frac{\|y\|^2}{2}\right)\, \|y\|^3\,dy = \frac{2\,\eta\,Vol(\mathcal{S}_{n-1})}{\gamma_m^2}\end{aligned}$$

Thus, we know that this part of the integral behaves as a $O\left(\gamma_m^{-2}\right)$. For the other part, as exp is a monotonous function, one has: $\exp(-\gamma\,\|y\|^2) < \exp(-\|y\|^2)/\gamma^2$ for $\|y\|^2 > 2\,\log(\gamma)/(\gamma-1)$. Moreover, as the limit of $\log(\gamma)/(\gamma-1)$ is zero at $\gamma = +\infty$, it can be made smaller than $\alpha/2$ provided that γ is large enough. Thus, we have that (for γ_m large enough):

$$\begin{aligned}\left|\int_{\|y\|^2>\alpha} \exp\left(-\frac{\gamma_m\,\|y\|^2}{2}\right)\, R_{d\mathcal{M}}(y)\,dy\right| &< \frac{1}{\gamma_m^2}\int_{\|y\|^2>\alpha} \exp\left(-\frac{\|y\|^2}{2}\right)\, |R_{d\mathcal{M}}(y)|\,dy \\ &< \frac{1}{\gamma_m^2}\int_{\mathbb{R}^n} \exp\left(-\frac{\|y\|^2}{2}\right)\, |R_{d\mathcal{M}}(y)|\,dy\end{aligned}$$

The last integral is a constant which is finite, since we assumed that $k^{(\text{-}1)}$ was finite for all Γ. Thus, this part of the integral behaves also as a $O\left(\gamma_m^{-2}\right)$ when γ_m goes to infinity. We have obtained that:

$$\int_{\mathbb{R}^n} \exp(-y^{\mathrm{T}}\,\Gamma\,y/2)\, R_{d\mathcal{M}}(y)\,dy = O\left(\gamma_m^{-2}\right) \tag{4}$$

so that

$$k^{(\text{-}1)} = \frac{(2\pi)^{n/2}}{\sqrt{\det(\Gamma)}}\left(1 - \frac{\mathrm{Tr}(\Gamma^{(\text{-}1)}\,\mathrm{Ric}\,)}{6}\right) + O\left(\gamma_m^{-2}\right)$$

As $\mathrm{Tr}(\Gamma^{(\text{-}1)}\,\mathrm{Ric}\,)$ is a term in γ_m^{-1} and $\sqrt{\det(\Gamma)}$ a term in $\gamma_m^{1/2}$, we finally have:

$$k = \frac{\sqrt{\det(\Gamma)}}{(2\pi)^{n/2}}\left(1 + \frac{\mathrm{Tr}(\Gamma^{(\text{-}1)}\,\mathrm{Ric}\,)}{6} + O\left(\gamma_m^{-3/2}\right)\right) \tag{5}$$

The covariance matrix: The principle is the same as for the normalization coefficient, with slightly more complex integrals. We basically have to compute

$$k^{(\text{-}1)}\,\Sigma = \det(\Gamma)^{-1/2}\,\Gamma^{-1/2}\,\left(I - \frac{1}{6}J\right)\,\Gamma^{-1/2} + \int_{\mathcal{D}} y\,y^{\mathrm{T}}\,\exp\left(-\frac{y^{\mathrm{T}}\,\Gamma\,y}{2}\right)\,R_{d\mathcal{M}}(y)\,dy$$

$$\text{with}\qquad I = \int_{\mathbb{R}^n} z\,z^{\mathrm{T}}\,\exp\left(-\frac{\|z\|^2}{2}\right)\,dz \qquad \text{and} \qquad J = \int_{\mathbb{R}^n} \mathrm{Tr}(T\,z\,z^{\mathrm{T}})\,z\,z^{\mathrm{T}}\,\exp\left(-\frac{\|z\|^2}{2}\right)\,dz$$

These are classical calculations in multivariate statistics: for the first integral, the off diagonal elements I_{ij} (with $i \neq j$) vanish because we integrate antisymmetric functions over $\mathbb{R}$:

$$\int_{\mathbb{R}^n} z_i\,z_j\,\exp(\|z\|^2/2)\,dz = 0 \qquad \text{for} \qquad i \neq j$$

The diagonal elements are:

$$I_{ii} = \int_{\mathbb{R}^n} z_i^2 \, \exp(-\|z\|^2/2) \, dz = \left(\int_{\mathbb{R}} z_i^2 \, \exp(-z_i^2/2) \, dz_i \right) \prod_{j \neq i} \left(\int_{\mathbb{R}} \exp(-z_j^2/2) \, dz_j \right) = (2\pi)^{n/2}$$

so that we have $I = (2\pi)^{n/2}$ Id. Since $\mathrm{Tr}(T\, z\, z^{\mathrm{T}}) = \sum_{k,l} T_{kl}\, z_k\, z_l$, the elements of the second integral are:

$$J_{ij} = \sum_{k,l} T_{kl} \int z_k \, z_l \, z_i \, z_j \, \exp\left(-\frac{\|z\|^2}{2} \right) \, dz$$

Let us investigate first off diagonal elements ($i \neq j$): for $k \neq i$ and $l \neq j$ or $l \neq i$ and $k \neq j$, we integrate an antisymmetric function and the result is zero. Since the Ricci curvature matrix is symmetric, the matrix T is also symmetric and the sum is reduced to a single term:

$$J_{ij} = (T_{ij} + T_{ji}) \int z_i^2 \, z_j^2 \, \exp\left(-\frac{\|z\|^2}{2} \right) \, dz = 2\, T_{ij} (2\pi)^{n/2} \quad \text{for} \quad i \neq j$$

The diagonal terms are: $J_{ii} = \sum_{k,l} T_{kl} \int_{\mathbb{R}^n} z_i^2 \, z_k \, z_l \, \exp(-\|z\|^2/2) \, dz$. For $k \neq l$, we integrate antisymmetric functions: the result is zero. Thus, we are left with:

$$\begin{array}{rcl} J_{ii} & = & \sum_{k \neq i} T_{kk} \int_{\mathbb{R}^n} z_i^2 \, z_k^2 \, \exp(-\|z\|^2/2) \, dz + T_{ii} \int_{\mathbb{R}^n} z_i^4 \, \exp(-\|z\|^2/2) \, dz \\ & = & \sum_{k \neq i} T_{kk} \, (2\pi)^{n/2} + T_{ii} \, 3 \, (2\pi)^{n/2} = (2\pi)^{n/2} \, (\, \mathrm{Tr}(T) + 2 \, T_{ii} \,) \end{array}$$

Combining with off diagonal terms, we get: $J = (2\pi)^{n/2} \, (\mathrm{Tr}(T) \;\, \mathrm{Id} + 2\, T)$, so that:

$$k^{(-1)} \, \Sigma = \frac{(2\pi)^{n/2}}{\sqrt{\det(\Gamma)}} \, \Gamma^{-1/2} \, \left(\mathrm{Id} - \frac{1}{6} (\mathrm{Tr}(T) \;\, \mathrm{Id} + 2\, T) \right) \, \Gamma^{-1/2} + \int_{\mathcal{D}} y \, y^{\mathrm{T}} \, \exp\left(-\frac{y^{\mathrm{T}} \, \Gamma \, y}{2} \right) \, R_{d\mathcal{M}}(y) \, dy$$

Like for the normalization constant, we now have to bound the integral of the remainder. The principle is the same. Let γ_m be the smallest eigenvalue of Γ. As $y^{\mathrm{T}} \Gamma y \geq \gamma_m \|y\|^2$, we have:

$$\left| \int_{\mathbb{R}^n} y \, y^{\mathrm{T}} \, \exp\left(-\frac{y^{\mathrm{T}} \, \Gamma \, y}{2} \right) \, R_{d\mathcal{M}}(y) \, dy \right| \leq \int_{\mathbb{R}^n} \|y\|^2 \, \exp\left(-\gamma_m \frac{\|y\|^2}{2} \right) \, |R_{d\mathcal{M}}(y)| \, dy$$

For any $\eta > 0$, we can find a constant $\alpha > 0$ such that $\|y\| < \alpha$ implies that $|R_{d\mathcal{M}}(y)| < \eta \, \|y\|^3$, i.e:

$$\begin{array}{rcl} \displaystyle\int_{\|y\|<\alpha} \|y\|^2 \, \exp\left(-\gamma_m \frac{\|y\|^2}{2} \right) \, |R_{d\mathcal{M}}(y)| \, dy & < & \eta \displaystyle\int_{\|y\|<\alpha} \|y\|^2 \, \exp\left(-\gamma_m \frac{\|y\|^2}{2} \right) \, \|y\|^3 \, dy \\ & < & \eta \displaystyle\int_{\mathbb{R}^n} \exp\left(-\gamma_m \frac{\|y\|^2}{2} \right) \, \|y\|^5 \, dy = \frac{8 \, \eta \, Vol(\mathcal{S}_{n-1})}{\gamma_m^3} \end{array}$$

Thus, we know that this part of the integral behaves as a $O\left(\gamma_m^{-3}\right)$. For the other part, as exp is a monotonous function, one has: $\exp(-\gamma \, \|y\|^2) < \exp(-\|y\|^2)/\gamma^3$ for $\|y\|^2 > 3 \, \log(\gamma)/(\gamma - 1)$. Moreover, as the limit of $\log(\gamma)/(\gamma - 1)$ is zero at $\gamma = +\infty$, it can be made smaller than $\alpha/3$ provided that γ is large enough. Thus, we have that (for γ_m large enough):

$$\begin{array}{rcl} \displaystyle\int_{\|y\|^2>\alpha} \|y\|^2 \, \exp\left(-\frac{\gamma_m \, \|y\|^2}{2} \right) \, |R_{d\mathcal{M}}(y)| \, dy & < & \displaystyle\frac{1}{\gamma_m^3} \int_{\|y\|^2>\alpha} \|y\|^2 \, \exp\left(-\frac{\|y\|^2}{2} \right) \, |R_{d\mathcal{M}}(y)| \, dy \\ & < & \displaystyle\frac{1}{\gamma_m^3} \int_{\mathbb{R}^n} \|y\|^2 \, \exp\left(-\frac{\|y\|^2}{2} \right) \, |R_{d\mathcal{M}}(y)| \, dy \end{array}$$

The last integral is a constant which is finite, since we assumed that Σ was finite for all Γ. Thus, this part of the integral behaves also as a $O\left(\gamma_m^{-3}\right)$ when γ_m goes to infinity. As $1/\sqrt{\det(\Gamma)}$ a term in $\gamma_m^{-1/2}$, we finally have:

$$\Sigma = k\frac{(2\pi)^n}{\sqrt{\det(\Gamma)}}\left(\Gamma^{(\text{-}1)} - \frac{\mathrm{Tr}(\Gamma^{(\text{-}1)}\,\mathrm{Ric}\,)}{6}\Gamma^{(\text{-}1)} - \frac{\Gamma^{(\text{-}1)}\,\mathrm{Ric}\;\Gamma^{(\text{-}1)}}{3} + O\left(\gamma_m^{-5/2}\right)\right)$$

Simplifying this expression using the previous Taylor expansion of k (Eq. 5), we obtain thus the following relation between the covariance and the concentration matrices:

$$\Sigma = \Gamma^{(\text{-}1)} - \frac{1}{3}\Gamma^{(\text{-}1)}\,\mathrm{Ric}\;\Gamma^{(\text{-}1)} + O\left(\gamma_m^{-5/2}\right) \tag{6}$$

Inverting the variables of the Taylor expansions This relation can be inverted to obtain the Taylor expansion of the concentration matrix with respect to the covariance. First, we shall note that from the above equation, $O(\gamma_m^{-5/2}) = O(\sigma_{max}^5)$ where σ_{max} is the square root of the largest eigenvalue of Σ. However, a global variable such as the variance $\sigma^2 = \mathrm{Tr}(\Sigma)$ is more appropriate. Since $\sigma_{max}^2 \leq \sum_i \sigma_i^2 = \sigma^2 \leq n\,\sigma_{max}^2$, we have $O(\sigma_{max}) = O(\sigma)$. Then, one easily verifies that the Taylor expansion of $\Gamma^{(\text{-}1)}$ is: $\Gamma^{(\text{-}1)} = \Sigma + \Sigma\,\mathrm{Ric}\;\Sigma/3 + O(\sigma^5)$, and finally:

$$\Gamma = \Sigma^{(\text{-}1)} - \frac{1}{3}\mathrm{Ric} + O(\sigma) \tag{7}$$

To express k with respect to Σ instead of Γ, we have to compute $\mathrm{Tr}(\Gamma^{(\text{-}1)}\,\mathrm{Ric}\,)$ and $\sqrt{\det(\Gamma)}$.

$$\mathrm{Tr}(\Gamma^{(\text{-}1)}\,\mathrm{Ric}\,) = \mathrm{Tr}(\Sigma\,\mathrm{Ric} + \frac{1}{3}\Sigma\,\mathrm{Ric}\;\Sigma\,\mathrm{Ric} + O(\sigma^5)) = \mathrm{Tr}(\Sigma\,\mathrm{Ric}\,) + O(\sigma^3)$$

For the determinant, one verifies that if A is a differential map from $\mathbb{R}$ to the matrix group, one has $\det(A)' = (\det(A))\,\mathrm{Tr}(A'\,A^{(\text{-}1)})$, so that $\det(\,\mathrm{Id} + \eta\,B) = 1 + \eta\,\mathrm{Tr}(B) + O(\eta^2)$. Since $\Gamma\,\Sigma = \,\mathrm{Id} - \frac{1}{3}\mathrm{Ric}\;\Sigma + O(\sigma^3)$ and Σ is a term in $O(\sigma^2)$, we have

$$\det(\Gamma)\det(\Sigma) = \det\left(\,\mathrm{Id} - \frac{1}{3}\mathrm{Ric}\;\Sigma + O(\sigma^3)\right) = 1 - \frac{1}{3}\,\mathrm{Tr}(\Sigma\,\mathrm{Ric}\,) + O(\sigma^3)$$

and thus

$$\sqrt{\det(\Gamma)} = \frac{1}{\sqrt{\det(\Sigma)}}\left(1 - \frac{1}{6}\,\mathrm{Tr}(\Sigma\,\mathrm{Ric}\,) + O(\sigma^3)\right)$$

Substituting this expression in equation 5, we obtain:

$$k = \frac{1}{\sqrt{(2\pi)^n\;\det(\Sigma)}}\left(1 - \frac{1}{6}\,\mathrm{Tr}(\Sigma\,\mathrm{Ric}\,) + O(\sigma^3)\right)\left(1 + \frac{1}{6}\,\mathrm{Tr}(\Sigma\,\mathrm{Ric}\,) + O(\sigma^3)\right)$$

Which simplifies in:

$$k = \frac{1 + O(\sigma^3)}{\sqrt{(2\pi)^n\;\det(\Sigma)}} \tag{8}$$

Summary of the approximate normal density: In a manifold without a cut locus at the mean point, the normal density in a normal coordinate system at the mean value is:

$$N(y) = k\, \exp\left(-\frac{y^{\mathrm{T}}\, \Gamma\, y}{2}\right)$$

The normalization constant and the concentration matrices are approximated by the following expressions for a covariance matrix Σ of small variance $\sigma^2 = \mathrm{Tr}(\Sigma)$:

$$k = \frac{1 + O(\sigma^3)}{\sqrt{(2\pi)^n\, \det(\Sigma)}} \qquad \text{and} \qquad \Gamma = \Sigma^{(-1)} - \frac{1}{3}\mathrm{Ric} + O(\sigma)$$

B.2 Manifolds with a cut locus at the mean point

Now, we have to integrate all the integrals over the definition domain $\mathcal{D}$ of the exponential chart at the mean point, which is only a subset of $\mathbb{R}^n$. For the integrals involving the remainder of the Taylor expansion of the metric, there is no problem as a bound on $\mathcal{D}$ is dominated by the bound on $\mathbb{R}^n$ we already computed. For the other integrals, the idea is to bound each term from above and below to show that the same results still holds with just a slight modification of the Taylor expansion bounding term. Let r be the (finite) injectivity radius at the mean point. The open ball $\mathcal{B}(r)$ is the greatest geodesic ball included in the definition domain: $\mathcal{B}(r) \subset \mathcal{D} \subset \mathbb{R}^n$. The idea is to bound the integrals of positive terms by integrating over these three sets.

In fact, with the change in variable $z = \Gamma^{1/2}\, y$, the ball is transformed into the ellipsoid $\Gamma^{1/2}\, \mathcal{B}(r) = \{z \mid z^{\mathrm{T}}\Gamma^{(-1)}z < r^2\}$. As $z^{\mathrm{T}}\Gamma^{(-1)}z \leq \gamma_m^{(-1)}\|z\|^2$, we may use smaller ball $\mathcal{B}(\sqrt{\gamma_m}\, r)$ for the integrations in polar coordinates. For the integrals that are separable along axes, it will be simpler to use the maximal cube $\mathcal{C} = \{z / -\sqrt{\gamma_m}\, r/\sqrt{n} < z_i < \sqrt{\gamma_m}\, r/\sqrt{n}\}$. As a summary, we have the inclusion chain: $\mathcal{C} \subset \mathcal{B}(\sqrt{\gamma_m}\, r) \subset \Gamma^{1/2}\, \mathcal{B}(r) \subset \mathcal{D}' = \Gamma^{1/2}\, \mathcal{D} \subset \mathbb{R}^n$.

In the following, we bound the different integrals used in the last section and summarize afterward the modifications it implies in the results of the last section. The first integral is bounded above by the previous value:

$$\int_{\mathcal{D}'} \exp\left(-\frac{\|z\|^2}{2}\right)\, dz \leq \int_{\mathbb{R}^n} \exp\left(-\frac{\|z\|^2}{2}\right)\, dz = (2\pi)^{n/2}$$

and below by an integration over $\mathcal{C}$:

$$\int_{\mathcal{D}'} \exp\left(-\frac{\|z\|^2}{2}\right)\, dz \geq \prod_i \left(\int_{-\sqrt{\gamma_m}\, r/\sqrt{n}}^{\sqrt{\gamma_m}\, r/\sqrt{n}} \exp\left(-\frac{z_i^2}{2}\right)\, dz_i\right) \geq (2\pi)^{n/2}\, \left(1 - \mathrm{erfc}\, \left(\frac{\sqrt{\gamma_m}\, r}{\sqrt{2n}}\right)\right)^n$$

Here, $\mathrm{erfc}\,(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2)\, dt$ is the complement of the error function. An interesting property is that this function (like the exponential) tends toward zero faster than any fraction $1/x^k$ as x goes to infinity. Putting things the other way, like for Taylor expansions, we will denote by $\varepsilon(x)$ a function that is an $O(x^k)$ for every positive k:

$$\lim_{0+} \frac{\varepsilon(x)}{x^k} = \lim_{0+} \frac{\mathrm{erfc}\,(1/x)}{x^k} = \lim_{0+} \frac{\exp(-1/x)}{x^k} = 0$$

We can summarize the value of the first integral by:

$$\int_{\mathcal{D}'} \exp\left(-\frac{\|z\|^2}{2}\right)\, dz = (2\pi)^{n/2} + \varepsilon\left({\gamma_m}^{-1/2}\, r^{-1}\right)$$

Now, for the second integral, we have

$$\begin{aligned}\int_{-\sqrt{\gamma_m}\,r/\sqrt{n}}^{\sqrt{\gamma_m}\,r/\sqrt{n}} z_i^2\,\exp\left(-\frac{z_i^2}{2}\right)\,dz_i &= \sqrt{2\pi}-\sqrt{2\pi}\,\mathrm{erfc}\left(\tfrac{\sqrt{\gamma_m}\,r}{\sqrt{2n}}\right)-\tfrac{2\sqrt{n}}{\sqrt{\gamma_m}\,r}\exp\left(-\tfrac{\gamma_m\,r^2}{2}\right)\\ &= \sqrt{2\pi}+\varepsilon\left({\gamma_m}^{-1/2}\,r^{-1}\right)\end{aligned}$$

We obtain thus

$$\int_{\mathcal{D}'} z_i^2\,\exp\left(-\frac{\|z\|^2}{2}\right)\,dz = (2\pi)^{n/2}+\varepsilon({\gamma_m}^{-1/2}\,r^{-1})$$

In fact, it is easy to see that every integral we computed over $\mathbb{R}^n$ in the previous paragraph has the same value over $\mathcal{D}'$ plus a term whose absolute value is of the order of $\varepsilon(\gamma_m^{-1/2}\,r^{-1})=\varepsilon(\sigma/r)$. Thus, we can directly generalize the previous results by replacing $O(\sigma^k)$ with $O(\sigma^k)+\varepsilon\,(\sigma/r)$.

Summary of the approximate normal density: In a manifold with injectivity radius r at the mean point, the normal density in a normal coordinate system at this mean value is:

$$N(y)=k\,\exp\left(-\frac{y^{\mathrm{T}}\,\Gamma\,y}{2}\right)$$

The normalization constant and the concentration matrices are approximated by the following expressions for a covariance matrix Σ of small variance $\sigma^2=\mathrm{Tr}(\Sigma)$:

$$k=\frac{1+O(\sigma^3)+\varepsilon\,(\sigma/r)}{\sqrt{(2\pi)^n\,\det(\Sigma)}}\qquad\text{and}\qquad\Gamma=\Sigma^{(\text{-}1)}-\frac{1}{3}\mathrm{Ric}\,+O(\sigma)+\varepsilon\left(\frac{\sigma}{r}\right)$$

C Approximated generalized χ^2 law

Assuming that the random point $\mathbf{x}\sim(\bar{\mathrm{x}},\Sigma)$ follows a normal law, we want to compute the probability that

$$\chi^2=\mu_{\mathbf{x}}^2=\overrightarrow{\bar{\mathrm{x}}\mathbf{x}}^{\mathrm{T}}\,\Sigma^{(\text{-}1)}\,\overrightarrow{\bar{\mathrm{x}}\mathbf{x}}\le\alpha^2$$

Let $\mathcal{B}_\Sigma(\alpha)$ be the "elliptic ball" of covariance Σ and radius α: $\mathcal{B}_\Sigma(\alpha)=\{y\,/\,y^{\mathrm{T}}\,\Sigma^{(\text{-}1)}\,y\le\alpha^2\}$.

C.1 Manifold without a cut locus at the mean point

Assuming that there is not cut locus, this probability can be written in a normal coordinate system:

$$\Pr\{\chi^2\le\alpha^2\}=\int_{\chi^2\le\alpha^2}N(\mathrm{y})\,d\mathcal{M}y=\int_{\mathcal{B}_\Sigma(\alpha)}k\,\exp\left(-\frac{y^{\mathrm{T}}\,\Gamma\,y}{2}\right)\,\sqrt{G(y)}\,dy$$

Since $\sqrt{\det(G(y))}=1-\frac{1}{6}y^{\mathrm{T}}\,\mathrm{Ric}\;y+R_{d\mathcal{M}}(y)$, we have:

$$\Pr\{\chi^2\le\alpha^2\}=k\int_{\mathcal{B}_\Sigma(\alpha)}\exp\left(-\frac{y^{\mathrm{T}}\Gamma y}{2}\right)\left(1-\frac{y^{\mathrm{T}}\mathrm{Ric}\,y}{6}\right)dy+k\int_{\mathcal{B}_\Sigma(\alpha)}\exp\left(-\frac{y^{\mathrm{T}}\,\Gamma\,y}{2}\right)\,R_{d\mathcal{M}}(y)\,dy$$

The last integral, involving the remainder of the metric, is obviously bounded by the integral over the whole tangent space $k\,\int_{\mathbb{R}^n}\exp(-y^{\mathrm{T}}\,\Gamma\,y/2)\,|R_{d\mathcal{M}}(y)|\,dy$ which we have shown to be a $O(\sigma^3)$ in

appendix B. Let $\Sigma^{1/2}$ be the positive symmetric square root of Σ. With the change of variable $y = \Sigma^{1/2}\, x$, this probability becomes:

$$\Pr\{\chi^2 \leq \alpha^2\} = \sqrt{2}\, k\, \sqrt{\det(\Sigma)} \int_{\|x\| \leq \alpha} \exp\left(-\frac{1}{2} x^{\mathrm{T}}\, \Sigma^{\frac{1}{2}}\, \Gamma\, \Sigma^{\frac{1}{2}}\, x\right) \left(1 - \frac{1}{6} x^{\mathrm{T}}\, \Sigma^{\frac{1}{2}}\, \mathrm{Ric}\, \Sigma^{\frac{1}{2}}\, x\right) dx + O(\sigma^3)$$

Using $S = \frac{1}{3}\Sigma^{1/2}\, \mathrm{Ric}\, \Sigma^{1/2}$ and the fact that $k\, \sqrt{\det(\Sigma)} = (2\,\pi)^{-n/2}\, (1 + O(\sigma^3))$ (from (Eq. 8)), we end up with

$$\Pr\{\chi^2 \leq \alpha^2\} = \frac{1 + O(\sigma^3)}{\sqrt{(2\,\pi)^n}} \int_{\|x\| \leq \alpha} \exp\left(-\frac{1}{2} x^{\mathrm{T}}\, \Sigma^{\frac{1}{2}}\, \Gamma\, \Sigma^{\frac{1}{2}}\, x\right) \left(1 - \frac{1}{2} x^{\mathrm{T}}\, S\, x\right) dx + O(\sigma^3)$$

The goal is now to show that the above integrated term is basically $\exp(-\|x\|^2/2)$ plus a remainder which integrates into a $O(\sigma^3)$. In this process, we will have to compute the integrals:

$$I_k(\alpha) = \int_{\|x\| < \alpha} \exp\left(-\frac{\|x\|^2}{2}\right)\, \|x\|^{2k}\, dx$$

By changing into polar coordinates ($r = \|x\|$ is the radius and u is the corresponding unit vector so that $x = r\, u$), we have $dx = r^{n-1}\, dr\, du$ and thus:

$$\begin{aligned} I_k(\alpha) &= \left(\int_{\mathcal{S}_{n-1}} du\right) \left(\int_0^{\alpha} r^{2k+n-1}\, \exp(-r^2/2)\, dr\right) \\ &= \left(\frac{\pi^{n/2}}{\mathbf{\Gamma}(n/2)}\right) \left(\int_0^{\alpha^2} t^{k-1+\frac{n}{2}}\, \exp\left(-\frac{t}{2}\right)\, dt\right) \end{aligned}$$

with the change of variable $t = r^2$. In this formula, $\mathbf{\Gamma} = \int_0^{+\infty} t^{x-1}\, \exp(-t) dt$ is the Gamma function which can be recursively computed from $\mathbf{\Gamma}(x+1) = x\, \mathbf{\Gamma}(x)$ with $\mathbf{\Gamma}(1) = 1$ and $\mathbf{\Gamma}(1/2) = \sqrt{\pi}$. The value $I_0(\alpha)$ is in fact the standard cumulated probability function of the χ^2 law (up to a normalization factor). For $\alpha = +\infty$, the remaining integral can be computed using the Gamma function:

$$I_k(+\infty) = \frac{(2\pi)^{\frac{n}{2}}\, 2^k\, \mathbf{\Gamma}(k + \frac{n}{2})}{\mathbf{\Gamma}(\frac{n}{2})}$$

Series expansion of the exponential From Eq. 7, we have $\Sigma^{1/2}\, \Gamma\, \Sigma^{1/2} = \mathrm{Id} - S - R_\Gamma$, where the remainder R_Γ behaves as a $O(\sigma^3)$. Thus, we have

$$\exp\left(-\frac{x^{\mathrm{T}}\, \Sigma^{1/2}\, \Gamma\, \Sigma^{1/2}\, x}{2}\right) = \exp\left(-\frac{1}{2}\|x\|^2\right) \left(1 + \frac{x^{\mathrm{T}}\, S\, x}{2} + R_{\exp}\right)$$

where the remainder is given by the convergent series

$$R_{\exp}(x) = \frac{1}{2} x^{\mathrm{T}}\, R_\Gamma\, x + \sum_{k=2}^{+\infty} \frac{(x^{\mathrm{T}}\, (S + R_\Gamma)\, x)^k}{2^k\, k!}.$$

Substituting in the probability, we find that:

$$\Pr\{\chi^2 \leq \alpha^2\} = \frac{1 + O(\sigma^3)}{(2\,\pi)^{n/2}} \int_{\|x\| \leq \alpha} \exp\left(-\frac{\|x\|^2}{2}\right) (1 + R_{\exp}(x))\, dx + O(\sigma^3)$$

Integral of the Remainder $R_{\exp}$ $S = \frac{1}{3}\Sigma^{1/2}\,\mathrm{Ric}\,\Sigma^{1/2}$ is a term of the order $O(\sigma^2)$ and the remainder R_Γ behaves as a $O(\sigma^3)$. Thus, we can find positive constants C_1 and C_2 such that, for σ sufficiently small, we have: $x^{\mathrm{T}}\,R_\Gamma\,x \leq C_1\,\sigma^3\|x\|^2$ and $x^{\mathrm{T}}\,(S+R_\Gamma)\,x \leq C_2\,\sigma^2\,\|x\|^2$. This means that we can bound the integral of the remainder by:

$$\int_{\|x\|<\alpha} \exp\left(-\frac{\|x\|^2}{2}\right) R_{\exp}(x)\,dx \leq C_1\,\sigma^3\,I_1(+\infty) + \sum_{k=2}^{+\infty} \frac{C_2^k\,\sigma^{2k}}{2^k\,k!} I_k(+\infty) =\leq C_1'\,\sigma^3 + \frac{(2\,\pi)^{\frac{n}{2}}}{\mathbf{\Gamma}\left(\frac{n}{2}\right))} \sum_{k=2}^{+\infty} a_k$$

where $a_k = C_2^k\,\sigma^{2k}\,\mathbf{\Gamma}(k+n/2)/k!$. To show that the series converges, let us investigate the ratio of successive terms:

$$\frac{a_{k+1}}{a_k} = \frac{C_2\,\sigma^2}{k+1}\,\frac{\mathbf{\Gamma}(k+1+n/2)}{\mathbf{\Gamma}(k+n/2)} = C_2\,\sigma^2\,\frac{k+n/2}{k+1}$$

The limit for $k=+\infty$ is obviously $C_2\,\sigma^2$, which can be made smaller than 1 for σ sufficiently small. Thus, by d'Alembert's ratio test, the series converges. Moreover, the smaller term of the series is a σ^4 (for k=2), which shows that the integral of the remainder is a finally dominated by the first term, a $O(\sigma^3)$.

Conclusion Finally, the researched probability is

$$\Pr\{\chi^2 \leq \alpha^2\} = (2\pi)^{-\frac{n}{2}} \int_{\|x\|\leq\alpha} \exp\left(-\frac{\|x\|^2}{2}\right) dx + O(\sigma^3) = \frac{1}{2\,\mathbf{\Gamma}(\frac{n}{2})} \int_0^{\alpha^2} \left(\frac{t}{2}\right)^{n/2} \exp(-t/2)dt + O(\sigma^3)$$

Thus, the probability density function of a χ^2 is:

$$p_{\chi^2}(u) = \frac{1}{2\,\mathbf{\Gamma}\left(\frac{n}{2}\right)} \left(\frac{u}{2}\right)^{\frac{n}{2}-1} \exp\left(-\frac{u}{2}\right) + O(\sigma^3)$$

C.2 Manifold with a cut locus at the mean point

If there is a cut locus, we have to replace $O(\sigma^k)$ by $O(\sigma^k) + \varepsilon\,(\sigma/r)$ and we should only integrate on $\mathcal{B}_\Sigma(\alpha) \cap \mathcal{D}$. After the change of coordinates, we should integrate on $\mathcal{B}(\alpha) \cap \mathcal{D}'$. Thus, we have:

$$\Pr\{\chi^2 \leq \alpha^2\} = (2\pi)^{-\frac{n}{2}} \int_{\mathcal{B}(\alpha)\cap\mathcal{D}'} \exp\left(-\frac{\|x\|^2}{2}\right)\,dx + O(\sigma^3) + \varepsilon\left(\frac{\sigma}{r}\right)$$

As we have $\mathcal{B}(\sqrt{\gamma_m}r) \subset \mathcal{D}' \subset \mathbb{R}^n$, we can enclose the researched domain into: $\mathcal{B}\left(\min(\sqrt{\gamma_m}\,r, \alpha)\right) \subset \mathcal{B}(\alpha) \cap \mathcal{D}' \subset \mathcal{B}(\alpha)$. For $\alpha \leq \sqrt{\gamma_m}\,r$, there is not problem but for $\alpha > \sqrt{\gamma_m}\,r$, we have:

$$\Pr\{\chi^2 \leq \alpha^2\} \geq (2\pi)^{-\frac{n}{2}} \int_{\mathcal{B}(\sqrt{\gamma_m}\,r)} \exp\left(-\frac{\|x\|^2}{2}\right)\,dx + O(\sigma^3) + \varepsilon\left(\frac{\sigma}{r}\right)$$

and we have already seen that this integral is $1 + \varepsilon\left(\frac{\sigma}{r}\right)$. As $\alpha > \sqrt{\gamma_m}\,r$, the same integral is itself of the same order, and we obtain in all cases the same result as with no cut locus where $O(\sigma^3)$ is replaced by $O(\sigma^3) + \varepsilon\,(\sigma/r)$.

References

[1] Xavier Pennec. *L'incertitude dans les problèmes de reconnaissance et de recalage – Applications en imagerie médicale et biologie moléculaire.* Thèse de sciences (PhD thesis), Ecole Polytechnique, Palaiseau (France), December 1996.

[2] X. Pennec and N. Ayache. Uniform distribution, distance and expectation problems for geometric features processing. *Journal of Mathematical Imaging and Vision*, 9(1):49–67, July 1998.

[3] X. Pennec, N. Ayache, and J.-P. Thirion. Landmark-based registration using features identified through differential geometry. In I. Bankman, editor, *Handbook of Medical Imaging*, chapter 31, pages 499–513. Academic Press, September 2000.

[4] X. Pennec, C.R.G. Guttmann, and J.-P. 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.

[5] S. Granger, X. Pennec, and A. Roche. Rigid point-surface registration using an EM variant of ICP for computer guided oral implantology. In W.J. Niessen and M.A. Viergever, editors, *4th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI'01)*, volume 2208 of *LNCS*, pages 752–761, Utrecht, The Netherlands, October 2001.

[6] S. Granger and X. Pennec. Statistiques exactes et approchées sur les normales aléatoires. Research report RR-4533, INRIA, 2002.

[7] P.T. Fletcher, S. Joshi, C. Lu, and S Pizer. Gaussian distributions on Lie groups and their application to statistical shape analysis. In *Poc of Information Processing in Medical Imaging (IPMI'2003)*, pages 450–462, 2003.

[8] P.T. Fletcher and S.C. Joshi. Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In *Proc. of CVAMIA and MMBIA Workshops, Prague, Czech Republic, May 15, 2004*, LNCS 3117, pages 87–98. Springer, 2004.

[9] Ch. Lenglet, M. Rousson, R. Deriche, and O. Faugeras. Statistics on multivariate normal distributions: A geometric approach and its application to diffusion tensor MRI. Research Report 5242, INRIA, 2004.

[10] P. Batchelor, M. Moakher, D. Atkinson, F. Calamante, and A. Connelly. A rigorous framework for diffusion tensor calculus. *Mag. Res. in Med.*, 53:221–225, 2005.

[11] Xavier Pennec, Pierre Fillard, and Nicholas Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 65(1), October 2005.

[12] Pierre Fillard, Vincent Arsigny, Xavier Pennec, Paul 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.

[13] X. Pennec and J.-P. 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.

[14] Shun-ichi Amari. *Differential-geometric methods in Statistics*, volume 28 of *Lecture Notes in Statistics*. Springer, 2nd corr. print edition, 1990.

[15] J.M. Oller and J.M. Corcuera. Intrinsic analysis of statistical estimation. *Annals of Statistics*, 23(5):1562–1581, 1995.

[16] C. Bingham. An antipodally symmetric distribution on the sphere. *The Annals of Statistics*, 2(6):1201–1225, 1974.

[17] P.E. Jupp and K.V. Mardia. A unified view of the theory of directional statistics, 1975-1988. *Int. Statistical Review*, 57(3):261–294, 1989.

[18] 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.

[19] K.V. Mardia. Directional statistics and shape analysis. *Journal of applied Statistics*, 26(949-957), 1999.

[20] D.G. Kendall. A survey of the statistical theory of shape (with discussion). *Statist. Sci.*, 4:87–120, 1989.

[21] 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.

[22] H. Le and D.G. Kendall. The Riemannian structure of euclidean shape space: a novel environment for statistics. *Ann. Statist.*, 21:1225–1271, 1993.

[23] C.G. Small. *The Statistical Theory of Shapes*. Springer series in statistics. Springer, 1996.

[24] U. Grenander. *Probabilities on Algebraic Structures*. Whiley, 1963.

[25] H. Karcher. Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math*, 30:509–541, 1977.

[26] W.S. Kendall. Convexity and the hemisphere. *Journ. London Math. Soc.*, 43(2):567–576, 1991.

[27] 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.

[28] 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.

[29] J. Picard. Barycentres et martingales sur une variété. *Annales de l'institut Poincaré - Probabilités et Statistiques*, 30(4):647–702, 1994.

[30] R.W.R. Darling. Martingales on non-compact manifolds: maximal inequalities and prescribed limits. *Ann. Inst. H. Poincarré Proba. Statistics*, 32(4):431–454, 1996.

[31] U. Grenander, M.I. Miller, and A. Srivastava. Hilbert-schmidt lower bounds for estimators on matrix Lie groups for atr. *IEEE Trans. on PAMI*, 20(8):790–802, 1998.

[32] X. Pennec. Computing the mean of geometric features - application to the mean rotation. Research Report RR-3371, INRIA, March 1998.

[33] C. Gramkow. On averaging rotations. *Int. Jour. Computer Vision*, 42(1-2):7–16, April/May 2001.

[34] M. Moakher. Means and averaging in the group of rotations. *SIAM J. of Matrix Anal. Appl.*, 24(1):1–16, 2002.

[35] 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.

[36] H. Hendricks. A Cramer-Rao type lower bound for estimators with values in a manifold. *Journal of Multivariate Analysis*, 38:245–261, 1991.

[37] R. Bhattacharya and V. Patrangenaru. Nonparametric estimation of location and dispersion on Riemannian manifolds. *Journal of Statistical Planning and Inference*, 108:23–36, 2002.

[38] 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.

[39] M. Spivak. *Differential Geometry*, volume 1. Publish or Perish, Inc., 2nd edition, 1979.

[40] W. Klingenberg. *Riemannian Geometry*. Walter de Gruyter, Berlin, New York, 1982.

[41] M. do Carmo. *Riemannian Geometry*. Mathematics. Birkhäuser, Boston, Basel, Berlin, 1992.

[42] S. Gallot, D. Hulin, and J. Lafontaine. *Riemannian Geometry*. Springer Verlag, 2nd edition edition, 1993.

[43] H. Poincaré. *Calcul des probabilités*. 2nd edition, Paris, 1912.

[44] M.G. Kendall and P.A.P. Moran. *Geometrical probability*. Number 10 in Griffin's statistical monographs and courses. Charles Griffin & Co. Ltd., 1963.

[45] Xavier Pennec. Probabilities and Statistics on Riemannian Manifolds: A Geometric approach. Research Report 5093, INRIA, January 2004.

[46] 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.

[47] M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré*, 10:215–310, 1948.

[48] 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.

[49] S. Doss. Sur la moyenne d'un élément aléatoire dans un espace distancié. *Bull. Sc. Math.*, 73:48–72, 1949.

[50] W. Herer. Espérance mathématique au sens de Doss d'une variable aléatoire à valeur dans un espace métrique. *C. R. Acad. Sc. Paris, Série I*, t.302(3):131–134, 1986.

[51] W. Herer. Espérance mathématique d'une variable aléatoire à valeur dans un espace métrique à courbure négative. *C. R. Acad. Sc. Paris, Série I*, t.306:681–684, 1988.

[52] M. Arnaudon. Espérances conditionnelles et C-martingales dans les variétés. In M. Yor J. Azema, P.A. Meyer, editor, *Séminaire de probabilités XXVIII*, volume 1583 of *Lect. Notes in Math.*, pages 300–311. Springer-Verlag, 1994.

[53] H. Maillot. Différentielle de la variance et centrage de la plaque de coupure sur une variété riemannienne compacte. Communication personnelle, 1997.

[54] W.S. Kendall. The propeller: a counterexample to a conjectured criterion for the existence of certain harmonic functions. *Journal of the London Mathematical Society*, 46:364–374, 1992.

[55] 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.

[56] 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.

[57] P. Huber. *Robust Statistics*. John Wiley, New York, 1981.

[58] P.J. Rousseeuw and A.M. Leroy. *Robust Regression and Outliers Detection*. Wiley series in prob. and math. stat. J. Wiley and Sons, 1987.

[59] M. Emery. *Stochastic Calculus in Manifolds*. Springer, Berlin, 1989.

[60] Alexander Grigor'yan. Heat kernels on weighted manifolds and applications. *Cont. Math*, 2005. To appear. http://www.ma.ic.ac.uk/ grigor/pubs.htm.

[61] K.V. Mardia and P.E. Jupp. *Directional statistics*. Whiley, Chichester, 2000.

[62] A.M. Kagan, Y.V. Linnik, and C.R. Rao. *Characterization problems in Mathematical Statistics*. Whiley-Interscience, New-York, 1973.

[63] I. Chavel. *Riemannian geometry - A modern introduction*, volume 108 of *Cambridge tracts in mathematics*. Cambridge university press, 1993.

[64] W.H. Press, B.P. Flannery, S.A Teukolsky, and W.T. Vetterling. *Numerical Recipices in C*. Cambridge Univ. Press, 1991.

[65] X. Pennec. Toward a generic framework for recognition based on uncertain geometric features. *Videre: Journal of Computer Vision Research*, 1(2):58–87, 1998.

[66] A. Roche, X. Pennec, G. Malandain, and N. Ayache. Rigid registration of 3D ultrasound with mr images: a new approach combining intensity and gradient information. *IEEE Transactions on Medical Imaging*, 20(10):1038–1049, October 2001.

[67] S. Nicolau, X. Pennec, L. Soler, and N. Ayache. Evaluation of a new 3D/2D registration criterion for liver radio-frequencies guided by augmented reality. In N. Ayache and H. Delingette, editors, *International Symposium on Surgery Simulation and Soft Tissue Modeling (IS4TM'03)*, volume 2673 of *Lecture Notes in Computer Science*, pages 270–283, Juan-les-Pins, France, 2003. INRIA Sophia Antipolis, Springer-Verlag.

[68] V. Rasouli. *Application of Riemannian multivariate statistics to the analysis of rock fracture surface roughness.* PhD thesis, University of London, 2002.

[69] X. 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.