INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Probabilities and Statistics on Riemannian Manifolds: A Geometric approach

Xavier Pennec

N° 5093

January 2004

THÈME 3

ISSN 0249-6399 ISRN INRIA/RR--5093--FR+ENG

INRIA
SOPHIA ANTIPOLIS

Probabilities and Statistics on Riemannian Manifolds: A Geometric approach

Xavier Pennec

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet Epidaure

Rapport de recherche n° 5093 — January 2004 — 49 pages

Abstract: Measurements of geometric primitives are often noisy in real applications and we need to use statistics either to reduce the uncertainty (estimation), to compare measurements, or to test hypotheses. Unfortunately, geometric primitives often belong to manifolds that are not vector spaces. In previous works [Pennec, 1996, Pennec and Ayache, 1998], we used invariance requirements to develop some basic probability tools on transformation groups and homogeneous manifolds that avoids paradoxes.

In this paper, we consider the Riemannian metric as the basic structure for the manifold. 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 test. We provide a simple (but highly non trivial) characterization of Karcher means and an original gradient descent algorithm to efficiently compute them. The notion of Normal law we propose is based on the the minimization of the information 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.

To come back to more practical cases, we then reconsider the case of connected Lie groups and homogeneous manifolds. In our Riemannian context, we investigate the use of invariance principles to choose the metric: we show that it can provide the stability of our statistical definitions w.r.t. geometric operations (composition, inversion and action of transformations). However, an invariant metric does not always exists for homogeneous manifolds, nor does a left and right invariant metric for non-compact Lie groups. In this case, we cannot guaranty the full consistency of geometric and statistical operations. Thus, future work will have to concentrate on constraints weaker than invariance.

Key-words: Riemannian geometry, statistics, probabilties, invariance.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : +33 4 92 38 77 77 — Télécopie : +33 4 92 38 77 65

Probabilités et statistiques sur des variétés riemanniennes : une approche géométrique

Résumé : Les mesures de primitives géométriques sont souvent bruitées dans les applications réelles et nous devons utiliser les statistiques pour en réduire l'incertitude (estimation), pour les comparer ou pour tester des hypothèses. Malheureusement, ces primitives géométriques appartiennent souvent à des variétés qui ne sont pas des espaces vectoriels. Dans des travaux précédents [Pennec, 1996, Pennec and Ayache, 1998], nous nous sommes appuyés sur l'invariance pour développer des outils statistiques de base qui évitent les paradoxes dans des groupes de transformation et des variétés homogènes.

Dans ce rapport, nous considérons la métrique riemannienne comme la structure déterminant la variété. En se basant sur cette métrique, nous développons les notions de valeur moyenne et de matrice de covariance d'un élément aléatoire, de loi normale, de distance de Mahalanobis et de test du χ^2. Nous présentons une caractérisation simple (mais hautement non triviale !) de la moyenne de Karcher ainsi qu'un algorithme de descente de gradient efficace pour l'obtenir. La notion de loi normale que nous proposons repose sur la minimisation de l'information de la distribution connaissant la moyenne et la covariance. La famille de densités qui en résulte va de la distribution uniforme (mesure ou densité pour le cas compacte) à la distribution ponctuelle (Dirac). Nous fournissons de plus des approximations simple (ainsi que leurs limites) pour de faibles variances qui montrent que l'on peut travailler efficacement avec ces définitions.

Pour en revenir à des cas plus pratiques, nous reconsidérons les groupes de Lie et les variétés homogènes. Dans le contexte riemannien, les principes d'invariance s'expriment sur la métrique. Nous montrons que cala apporte la stabilité de nos définitions statistiques par rapport aux opérations géométriques (composition, inversion et action de transformations). Cependant, une métrique invariante n'existe pas toujours pour une variété homogène, tout comme une métrique bi-invariante (à droite et à gauche) n'existe en général pas pour un groupe de Lie non compact. Dans ce cas, nous ne pouvons pas garantir la consistance de toutes les opérations statistiques et géométriques. Les travaux futurs devront donc se concentrer sur des contraintes plus faibles que l'invariance.

Mots-clés : Géométrie riemannienne, statistiques, probabilités, invariance.

Contents

1 Introduction

To represent the results of a random experiment, one usually construct a probabilized space $(\Omega, \mathcal{A}, \Pr)$ where Ω is the space of elementary events (the possible outcomes of the experiment), $\mathcal{A}$ is a tribe on Ω, usually the Borelian one[1], and Pr is a measure of probability (a normalized σ-additive function from $\mathcal{A}$ to $\mathbb{R}_+$).

Although this probabilized space contains all the information about the random experiment, we are often only interested in some measurements depending of the outcome of the experiment. The mathematical way to formalize that is to investigate *random variables* or *observables* which are maps[2] $\boldsymbol{x} = x(\omega)$ from Ω to $\mathbb{R}$. This formalism allows to "forget" the original probabilized space $(\Omega, \mathcal{A}, \Pr)$ and work directly in $\mathbb{R}$ by associating to each random variable $\boldsymbol{x}$ or $\boldsymbol{y}$ different probabilities.

Often, working directly with probabilities on events is difficult and one can restrict to random variables that have a *probability density function*, i.e. a normalized function $p_{\boldsymbol{x}}$ from $\mathbb{R}$ to $\mathbb{R}_+$ such that for all interval $]a; b[$:

$$\Pr(\boldsymbol{x} \in]a, b[) = \int_a^b p_{\boldsymbol{x}}(y).dy$$

However, from a computational point of view, the pdf is 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* $\sigma^2_{\boldsymbol{x}} = \mathbf{E}\left[\, (\boldsymbol{x} - \overline{x})^2 \,\right]$.

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* which is a map $\boldsymbol{x} = x(\omega)$ from Ω to $\mathbb{R}^n$. One can easily generalize the probability density function to random vector (using open sets or "paves" instead of intervals). Also, as the expectation is a linear operator, it is easily generalized to vectorial or matricial functions in order to define the mean value and the covariance matrix of a random vector:

$$\overline{x} = \mathbf{E}\left[\, \boldsymbol{x} \,\right] = (\mathbf{E}\left[\, \boldsymbol{x}_1 \,\right], \dots, \mathbf{E}\left[\, \boldsymbol{x}_n \,\right])^{\mathrm{T}} = (\overline{x}_1, \dots, \overline{x}_n)^{\mathrm{T}}$$

$$\Sigma_{\boldsymbol{x}\boldsymbol{x}} = \mathbf{E}\left[\, (\boldsymbol{x} - \overline{x}).(\boldsymbol{x} - \overline{x})^{\mathrm{T}} \,\right] = \int (y - \overline{x}).(y - \overline{x})^{\mathrm{T}}.p_{\boldsymbol{x}}(y).dy$$

However, for many statistical problems, one have to assume a probability distribution. The Gaussian distribution is especially well adapted, as it is completely determined by the mean and the covariance. It is moreover the distribution that minimizes the information knowing only these moments. Then, one can use standard statistical tests such as the χ^2 test and use a probabilistic distance between distributions such as the Mahalanobis distance. For more details about all these notions, one can refer to [Neveu, 1990, Papoulis, 1991, Pelat, 1992].

[1] A tribe of Ω is a family of parts of Ω containing $\varnothing$ which is stable by complementation, countable union and intersection. The Borelian tribe of a topological space is the tribe generated by the class of open sets of this space.

[2] The random variables should however map an element of the tribe $\mathcal{A}$ to an element of the considered tribe of $\mathbb{R}$. On usually use the Borelian tribes on each space and restrict the random variables to functions that respect these structures (Borelian functions).

The problem we investigate in this article is to generalize this framework to measurements in Riemannian manifolds instead of measurements in a vector space. We call them *random primitives*. Examples of manifolds we routinely use are 3D rotations and 3D rigid transformations as transformation groups and 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), oriented or directed points. We have already shown in [Pennec and Thirion, 1997, Pennec and Ayache, 1998] 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.

We review in the remainder of this section some basic notions of differential and Riemannian geometry that will be used afterward. This synthesis was inspired from [Spivak, 1979, chap. 9], [Klingenberg, 1982] and [Carmo, 1992], and the reader can refer to these books to find more details. In section 2, we develop the general theory for (connected and geodesically complete) Riemannian manifolds. In section 3 we show how to apply and simplify this framework in the case of transformation (Lie) groups, by using the left invariant Riemannian metric. In section 4, we focus on homogeneous manifolds and emphasize the very simple results obtained using the invariant metric if it exists.

1.1 Riemannian metric, distance and geodesics

In the geometric framework, one specify 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 ∂/∂_{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

$$\langle \, v \mid w \, \rangle_x = v^{\mathrm{T}} \cdot G(x) \cdot w$$

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:

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

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 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:

$$\mathrm{dist}(\mathrm{x}, \mathrm{y}) = \min_{\gamma} \mathcal{L}(\gamma) \qquad \text{with} \qquad \gamma(0) = \mathrm{x} \quad \text{and} \quad \gamma(1) = \mathrm{y} \tag{2}$$

The curves realizing this minimum for any two points of the manifold are called geodesics[3]. Let $[g^{ij}] = [g_{ij}]^{(-1)}$ be the inverse of the metric matrix (in a given coordinate system x) and Γ^i_{jk} the Christoffel symbols (using Einstein summation convention[4]):

$$\Gamma^i_{jk} = \frac{1}{2} g^{im} \left(\partial_k g_{mj} + \partial_j g_{mk} - \partial_m g_{jk} \right) \tag{3}$$

The calculus of variations shows the geodesics are the curves satisfying the following second order differential system (in the chart $x = (x_1, \dots x_n)$):

$$\ddot{\gamma}_i + \Gamma^i_{jk} \dot{\gamma}^j \dot{\gamma}^k = 0 \tag{4}$$

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 2), 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.

1.2 Exponential map and cut locus

From the second order differential equations theory, 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:

$$\exp_{\mathrm{x}} \quad : \quad \begin{array}{ccc} T_{\mathrm{x}}\mathcal{M} & \longrightarrow & \mathcal{M} \\ \partial_v & \longmapsto & \exp_{\mathrm{x}}(\partial_v) = \gamma_{(\mathrm{x},\partial_v)}(1) \end{array} \tag{5}$$

This function realizes a local diffeomorphism from a sufficiently small neighborhood $\mathcal{V}$ of $0 \in T_{\mathrm{x}}\mathcal{M}$ into a neighborhood $\mathcal{U}_{\mathrm{x}}$ 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:

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

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 has 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

[3] 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.

[4] Einstein summation convention is the implicit sum upon each index that appear up and down in the formula.

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. Moreover, the boundary of the domain $\mathcal{C}(\mathrm{x}) = \partial\mathcal{D}(\mathrm{x})$ is continuous and contains a dense set of points where several minimizing geodesics meet.

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 but we have to take care of the multiplicity.

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 $\underline{\mathrm{x}}$.

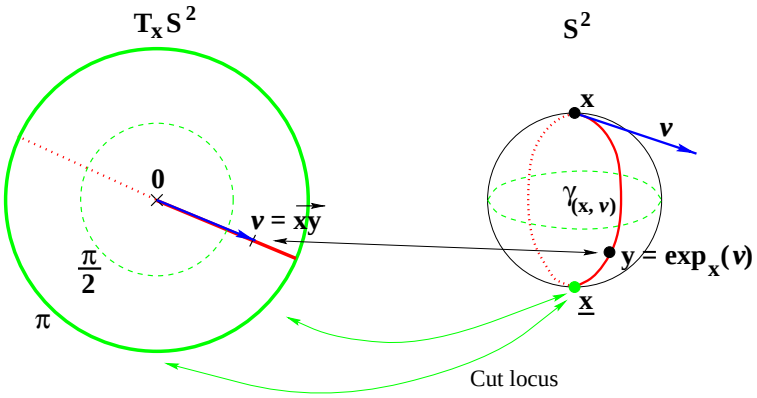


Figure 1: Exponential chart and cut locus for the sphere $\mathcal{S}_2$ and the projective space $\mathcal{P}_2$

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.

1.3 Riemannian measure or volume form

In a vector space with coordinate frame $\mathcal{A} = (a_1, \dots a_n)$, the representation of the canonical 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 the canonical 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:

$$d\mathcal{M}(\mathrm{x}) = \sqrt{|G(x)|}dx \tag{6}$$

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:

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

1.4 Gradient, Hessian and Taylor expansion of a real function

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

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

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 $\langle\, \omega \mid v \,\rangle_x = \omega^{\mathrm{T}} \cdot G(x) \cdot v$. Thus, the expression of the gradient in a chart is:

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

This definition corresponds to the classical gradient in $\mathbb{R}^n$ even in the case of a non orthonormal basis.

Hessian 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:

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

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

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

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: $J_{f_{\mathrm{x}}} = \mathrm{grad}\, f^{\mathrm{T}}$, and the Christoffel symbols vanished so that the matrix of second derivatives $H_{f_{\mathrm{x}}}$ corresponds to the Hessian $\mathrm{Hess}\, f$. Thus, The Taylor expansion can be written in any coordinate system:

$$f\left(\exp_{\mathrm{x}}(v)\right) = f(\mathrm{x}) + \mathrm{grad}\, f(v) + \frac{1}{2}\mathrm{Hess}\, f(v, v) + O(\|v\|^3) \tag{7}$$

1.5 Curvature and regular geodesic balls

Riemannian curvature Gauss showed that the curvature of a surface can be expressed from its metric. Riemann used this property to generalize the notion of curvature to manifolds in the following way. Let $\mathcal{V} \subset T_{\mathrm{x}}\mathcal{M}$ be a 2-dimensional vector subset of the tangent space at x. The space spanned by the geodesics of $\mathcal{M}$ starting at x with a tangent vector in $\mathcal{V}$ is a sub-manifold with the induced metric. Thus, it is a surface and we can determine its Gaussian curvature: we call *sectional curvature* or *Riemannian curvature* this intrinsic quantity $\kappa(\mathrm{x}, \mathcal{V})$

It is clear that the curvature is always null in a vector space. In simple cases, the "surface" of the manifold (viewed as embedded in $\mathbb{R}^p$) is always on the same side of the tangent space at any point. The curvature is in this case positive. This is in particular the case of the sphere or of $\mathcal{S}O_n$. For rigid transformations of $\mathbb{R}^n$, we add the translations (thus a vector space of null curvature), and the manifold has only a *non negative curvature*. There are also examples of manifolds with negative curvature, such as the hyperbolic paraboloid $z = x^2 - y^2$ (the saddle surface).

Ricci curvature The Ricci curvature tensor is a linear operator of the tangent space (i.e. a matrix). It is obtained a a partial trace of the full Riemannian curvature tensor, and can be though of as a Laplacian of the metric: this is the property we will be interested in to make a Taylor expansion of the metric.

Let $u \in T_{\mathrm{x}}\mathcal{M}$ be a unit vector of the tangent plane, then $\langle\, Ric(u) \mid u \,\rangle$ is the sum of the sectional curvatures of all the planes spanned by the vector u and one of the $n-1$ vector from an orthonormal frame containing u. In dimensions 2 and 3 the Ricci curvature matrix specifies all sectional curvatures of the manifold, but in higher dimensions Ricci curvature contains less information than the full curvature tensor. Using standard tensor notations and Einstein summation convention, the Riemannian curvature tensor and its contraction into the Ricci tensor are given by:

$$R^{\gamma}_{ijk} = \partial_k \Gamma^{\gamma}_{ij} + \Gamma^{\sigma}_{ij}\Gamma^{\gamma}_{\sigma k} - \partial_j \Gamma^{\gamma}_{ik} - \Gamma^{\sigma}_{ik}\Gamma^{\gamma}_{\sigma j} \qquad \text{and} \qquad \mathrm{Ric}_{ij} = R^{\sigma}_{ij\sigma}$$

Regular geodesic balls The ball $\mathcal{B}(\mathrm{x}, r)$ is the set of points $\mathrm{y} \in \mathcal{M}$ such that the distance to the center x is less than the radius r. This ball is said to be *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. Let κ be the maximum of the Riemannian curvature in this geodesic ball. The ball is said to be *regular* if its radius verifies $2.r.\sqrt{\kappa} < \pi$.

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 Hadamard theorem, such a manifold is diffeomorphic to $\mathbb{R}^n$ if it is connected).

2 Probabilities and statistics on a Riemannian Manifold

In this section, we do not consider measurements that are real variables or vectors depending on the outcome of a random experiment, but rather measurements of elements of a manifold. We call such a measurement a *random primitive*, reserving the names of *random transformation* and *random feature* for the particular case of transformation groups and homogeneous manifolds.

Definition 1 (Random primitive)
Let $(\Omega, \mathcal{B}, \Pr)$ *be a probabilized space with the Borelian tribe. A random primitive in the Riemannian manifold* $\mathcal{M}$ *is a Borelian 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 $\mathcal{M}$.

2.1 Probability density function

Definition 2 *Let* $\mathcal{A}$ *be the Borelian tribe of* $\mathcal{M}$ *(the tribe generated by the class of open sets). The random primitive* $\mathbf{x}$ *has a probability density function* $p_{\mathbf{x}}$ *(real, positive and integrable function) if:*

$$\forall \mathcal{X} \in \mathcal{A}, \qquad \Pr(\mathbf{x} \in \mathcal{X}) = \int_{\mathcal{X}} p(\mathrm{y}).d\mathcal{M}(\mathrm{y}) \qquad \textit{and} \qquad \Pr(\mathcal{M}) = \int_{\mathcal{M}} p(\mathrm{y}).d\mathcal{M}(\mathrm{y}) = 1 \tag{8}$$

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

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

where $\mathcal{V}(\mathcal{X})$ is the "volume" of the set $\mathcal{X}$.

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 [Poincaré, 1912, Kendall and Moran, 1963, Pennec and Ayache, 1998] 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 ? We will address this question in the next sections for transformation groups and homogeneous manifolds and show that an invariant metric is a good geometric choice.

Expression of the density in a chart Let $\mathbf{x}$ be a random primitive 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 of equation (6), the two pdfs are related by

$$\rho_{\boldsymbol{x}}(y) = p_{\mathbf{x}}(y).\sqrt{|G(y)|} \tag{9}$$

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.

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

$$\mathbf{E}\left[\,\varphi(\mathbf{x})\,\right] = \mathbf{E}_{\mathbf{x}}\left[\,\varphi\,\right] = \int_{\mathcal{M}} \varphi(\mathrm{y}).p_{\mathbf{x}}(\mathrm{y}).d\mathcal{M}(\mathrm{y}) \tag{10}$$

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.

2.2 Expectation and Mean value

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

Fréchet expectation or mean value Let $\boldsymbol{x}$ be a random vector of $\mathbb{R}^n$. [Fréchet, 1944, Fréchet, 1948] observed that the variance $\sigma^2_{\boldsymbol{x}}(y) = \mathbf{E}\left[\ \mathrm{dist}(\boldsymbol{x}, y)^2\ \right]$ is minimized for the mean vector $\overline{x} = \mathbf{E}\left[\,\boldsymbol{x}\,\right]$. 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 primitive
Let $\mathbf{x}$ *be a random primitive of pdf* $p_{\mathbf{x}}$. *The variance* $\sigma^2_{\mathbf{x}}(\mathrm{y})$ *is the expectation of the squared distance between the random primitive and the fixed primitive* y*:*

$$\sigma^2_{\mathbf{x}}(\mathrm{y}) = \mathbf{E}\left[\ \mathrm{dist}(\mathrm{y}, \mathbf{x})^2\ \right] = \int_{\mathcal{M}} \mathrm{dist}(\mathrm{y}, \mathrm{z})^2.p_{\mathbf{x}}(\mathrm{z}).d\mathcal{M}(\mathrm{z}) \tag{11}$$

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

$$\mathbb{E}[\,\mathbf{x}\,] = \arg\min_{\mathrm{y}\in\mathcal{M}} \left(\mathbf{E}\left[\ \mathrm{dist}(\mathrm{y}, \mathbf{x})^2\ \right]\right) \tag{12}$$

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

In the same way, one define the *empirical or discrete mean primitive* of a set of measures $\mathrm{x}_1, \ldots \mathrm{x}_n$ with the discrete version:

$$\mathbb{E}[\ \{\mathrm{x}_i\}\ ] = \arg\min_{\mathrm{y}\in\mathcal{M}} \left(\mathbf{E}\left[\ \{\ \mathrm{dist}(\mathrm{y}, \mathrm{x}_i)^2\}\ \right]\right) = \arg\min_{\mathrm{y}\in\mathcal{M}} \left(\frac{1}{n}\sum_i \mathrm{dist}(\mathrm{y}, \mathrm{x}_i)^2\right) \tag{13}$$

If there exists a least a mean primitive $\bar{\mathrm{x}}$, one call *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

$$\sigma_{\mathbf{x},\alpha}(\mathrm{y}) = \left(\mathbf{E}\left[\ \mathrm{dist}(\mathrm{y}, \mathbf{x})^\alpha\ \right]\right)^{1/\alpha} = \left(\int_{\mathcal{M}} \mathrm{dist}(\mathrm{y}, \mathrm{z})^\alpha.p_{\mathbf{x}}(\mathrm{z}).d\mathcal{M}(\mathrm{z})\right)^{1/\alpha} \tag{14}$$

If this function is bounded on $\mathcal{M}$, one call *central primitive at order* α every primitive $\bar{\mathrm{x}}_\alpha$ minimizing it. For instance, the *modes* are obtained for $\alpha = 0$. Exactly like in a vector space, they are the primitives where the density is maximal on the manifold (which is generally not a maximum for the density on the charts). The *median primitive* 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.

Existence and uniqueness: Karcher expectation As our mean primitive 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 is 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, 1977] proposed to consider the local minima of the variance $\sigma^2_{\mathbf{x}}(\mathrm{y})$ (equation 11) instead of the global ones. As global minima are local minima, the Fréchet expected primitives are a subset of the Karcher expected primitives. However, the use of local minima allows to characterize the Karcher means using only local derivatives of order two.

Using this extended definition, [Karcher, 1977] and [Kendall, 1990] 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:

Theorem 1 (Existence and uniqueness of the Karcher mean)
Let $\mathbf{x}$ *be a random primitive of pdf* $p_{\mathbf{x}}$.

- **[Kendall, 1990]** *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 Karcher mean* $\mathbf{x}$ *on this ball.*
- **[Karcher, 1977]** *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 Karcher mean.*

These conditions are relatively restrictive but ensure a correct behavior of our mean for localized distributions.

Other possible definitions of the mean primitives The Karcher mean is perfectly adapted for our purpose, thanks to the good properties it has for optimization (see 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, 1949] uses 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 $\mathbf{E}\left[\, \boldsymbol{x} \,\right]$. Thus, in a metric space, the *mean according to Doss* is defined as the set of primitives $\bar{\mathrm{x}} \in \mathcal{M}$ verifying:

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

Herer shows in [Herer, 1986, Herer, 1988] 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 in [Emery and Mokobodzki, 1991] and [Arnaudon, 1994, Arnaudon, 1995]. 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 primitive $\mathbf{x}$ with density $p_{\mathbf{x}}$ is the set $\mathbb{B}(\mathbf{x})$ of primitives $\mathrm{y} \in \mathcal{M}$ 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 primitive for which the support 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*[5] $\mathcal{U}$, Emery shows that the *exponential barycenters*, defined as the critical points of the variance $\sigma^2_{\mathbf{x}}(\mathrm{y})$ are are subset of the convex barycenter $\mathbb{B}(\mathbf{x})$. Local and global minima being particular critical points, the exponential barycenters include the Karcher means that include themselves the Fréchet means.

[Picard, 1994] realizes 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 primitives, all these values are close.

2.3 Characterizing a Karcher mean

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. Fortunately, we were able to generalize in appendix A a differentiability proof of Pr Maillot [Maillot, 1997] 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 function

$$\sigma^2(\mathrm{x}) = \int_{\mathcal{M}} \mathrm{dist}(\mathrm{x},\mathrm{z})^2 . dP(\mathrm{z})$$

is differentiable at any point $\mathrm{y} \in \mathcal{M}$ where the variance is finite and the cut locus $C(\mathrm{y})$ has a null probability measure:

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

At such a point, it has the following gradient:

$$(\mathrm{grad}\,\sigma^2)(\mathrm{y}) = -2. \int_{\mathcal{M}/C(\mathrm{y})} \overrightarrow{\mathrm{yz}} . dP(\mathrm{z}) \tag{15}$$

[5] 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.

Using the random primitive formalism, this result can be rewritten more simply

$$\text{grad}\ \left(\sigma^2_{\mathbf{x}}(\mathrm{y})\right) = -2.\mathbf{E}\left[\ \overrightarrow{\mathrm{y}\mathbf{x}}\ \right] \tag{16}$$

Now, we know that the variance is continuous but not 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 $(\text{grad}\,\sigma^2)(y) = 0$ if this is defined or $P(C(y)) > 0$.

Corollary 1 (Characterization of Karcher means for manifolds with a cut locus) *Assume that the random primitive* $\mathbf{x}$ *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* $\bar{\mathrm{x}}$ *to be a Karcher mean is:*

$$\bar{\mathrm{x}} \in \mathbb{E}[\ \mathbf{x}\ ] \quad \Longrightarrow \quad \left| \begin{array}{l} \mathbf{E}\left[\ \overrightarrow{\bar{\mathrm{x}}\mathbf{x}}\ \right] = 0 \quad \textit{if} \quad \bar{\mathrm{x}} \notin \mathcal{A} \\ \textit{or} \\ \bar{\mathrm{x}} \in \mathcal{A} \end{array} \right. \tag{17}$$

For discrete or empirical means, the characterization is the same but we can write explicitly the set $\mathcal{A} = \cup_i C(\mathrm{x}_i)$ *and the expectation* $\mathbf{E}\left[\ \{\overrightarrow{\bar{\mathrm{x}}\mathrm{x}_\mathrm{i}}\}\ \right] = \sum_i \overrightarrow{\bar{\mathrm{x}}\mathrm{x}_\mathrm{i}}$.

If the manifold does not have a cut locus, we have no differentiation problem. One can even differentiate one step further to obtain the Hessian matrix:

$$\text{Hess}\ \left(\sigma^2_{\mathbf{x}}(\mathrm{y})\right) = -2.\ \mathrm{Id}$$

In this case, there is one and only one minimum of the variance, necessarily the Fréchet mean.

Corollary 2 (Characterization of the Fréchet mean for manifolds without a cut locus) *Assume that the random primitive* $\mathbf{x}$ *has a finite variance. Then, there is one and only one Fréchet (or Karcher) mean characterized by*

$$\mathbb{E}[\ \mathbf{x}\ ] = \{\bar{\mathrm{x}}\} \quad \Longleftrightarrow \quad \mathbf{E}\left[\ \overrightarrow{\bar{\mathrm{x}}\mathbf{x}}\ \right] = 0 \tag{18}$$

For discrete or empirical means, the characterization is the similar

$$\mathbb{E}[\ \{\mathrm{x}_i\}\ ] = \{\bar{\mathrm{x}}\} \quad \Longleftrightarrow \quad \mathbf{E}\left[\ \{\overrightarrow{\bar{\mathrm{x}}\mathrm{x}_\mathrm{i}}\}\ \right] = \sum_i \overrightarrow{\bar{\mathrm{x}}\mathrm{x}_\mathrm{i}} = 0 \tag{19}$$

Basically, the characterization is the same as in Euclidean spaces, except that the non-zero probability cut loci induce some discontinuities in the first derivative of the variance. This corresponds to something like a Dirac on the second order derivative, which is the main reason why it is much more difficult to compute the Hessian matrix of the variance on a manifold with a cut locus.

Example on the circle The easiest example is probably a symmetric distribution on the circle. Let $p = \cos(\theta)^2/\pi$ be the probability density function of our random primitive $\boldsymbol{\theta}$ 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 $\text{dist}(\alpha,\theta) = |\alpha - \theta|$ within this domain.

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

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

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 $H(\pm\pi) = +2$.

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 calculus. 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 calculus.

2.4 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 primitive $\mathbf{x}$ 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 (7), 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 concave and has thus a minimum characterized by a null gradient. Let $H_f(v)$ denote the linear form verifying $\langle\, H_f(v) \mid w \,\rangle = \operatorname{Hess} f(v,w)$ 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 $\text{grad}\, f = -2.\mathbf{E}\left[\, \overrightarrow{\mathrm{y}\mathbf{x}} \,\right]$. Neglecting the "cut locus term" in the Hessian matrix gives us a perfect positive definite matrix $\text{Hess}\, f \simeq 2.\,\text{Id}$. Thus, the gradient descent algorithm is

$$\mathrm{y}_{t+1} = \exp_{\mathrm{y}_t}\left(\mathbf{E}\left[\, \overrightarrow{\mathrm{y_t}\mathbf{x}} \,\right]\right) \tag{20}$$

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:

$$\mathrm{y}_{t+1} = \exp_{\mathrm{y}_t}\left(\sum_i \overrightarrow{\mathrm{y_t x_i}}\right) \tag{21}$$

We note that in the case of a vector space, these two formula simplify to $y_{t+1} = \mathbf{E}\left[\, \boldsymbol{x} \,\right]$ 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 measurements $\{\mathrm{x}_i\}$, on can choose at random one of the measurements as the starting point. Another solution is to map to each point x_i is 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 measurements) but the method can be randomized efficiently [Huber, 1981, Rousseeuw and Leroy, 1987].

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

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

2.5 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 $\overrightarrow{y\boldsymbol{x}} = \boldsymbol{x} - y$:

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

This definition is easily 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 5 (Covariance)
Let $\mathbf{x}$ *be a random primitive 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 note* $\Sigma_{\mathbf{xx}}$ *and we call covariance the expression:*

$$\Sigma_{\mathbf{xx}} = \text{Cov}_{\bar{\mathrm{x}}}(\mathbf{x}) = \mathbf{E}\left[\, \overrightarrow{\bar{\mathrm{x}}\mathbf{x}}.\overrightarrow{\bar{\mathrm{x}}\mathbf{x}}^{T} \,\right] = \int_{\mathcal{D}(\bar{\mathrm{x}})} (\overrightarrow{\bar{\mathrm{x}}\mathrm{x}}).(\overrightarrow{\bar{\mathrm{x}}\mathrm{x}})^{T}.p_{\mathbf{x}}(\mathrm{x}).d\mathcal{M}(\mathrm{x}) \tag{22}$$

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:

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

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

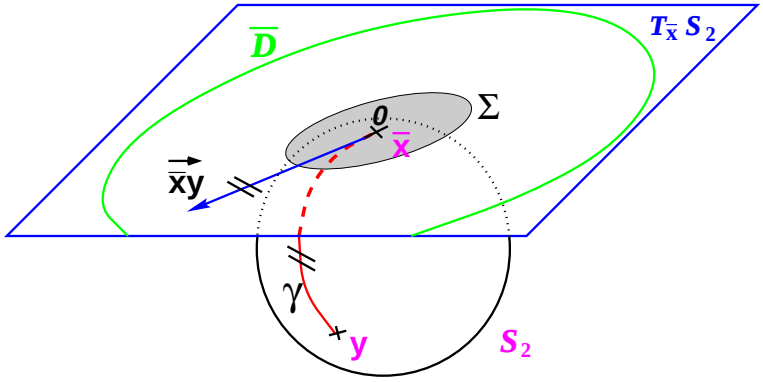


Figure 2: The covariance is defined in the tangent plane at $\mathcal{S}_2$ at the mean point as the classical covariance matrix of the random vector "deviation from the mean" $\Sigma_{\mathbf{xx}} = \mathbf{E}\left[\ \overrightarrow{\bar{\mathrm{x}}\mathbf{x}}\ .\ \overrightarrow{\bar{\mathrm{x}}\mathbf{x}}^{\mathrm{T}}\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

In fact, as soon as we have found a (or the) mean value, 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.

2.6 Several random primitives

Now assume that we have several simultaneous measures of the same random experiment. We quickly develop in this section the case of two random primitives $\mathbf{x}$ and $\mathbf{y}$ in two manifolds $\mathcal{M}$ and $\mathcal{N}$. The generalization to any (finite) number of random primitives is straightforward.

Let $\mathbf{z} = (\mathbf{x}, \mathbf{y})$ be the corresponding random primitive in the product manifold $\mathcal{M} \times \mathcal{N}$. The measure and metric and this product manifold is simply the product of the measures and metrics. Thus the pdf is $p_{\mathbf{z}}(\mathrm{z}) = p_{(\mathbf{x},\mathbf{y})}(\mathrm{x},\mathrm{y})$ with respect to the measure $d\mathcal{M}.d\mathcal{N}$.

As in the vector case, the *marginal* densities are obtained by partial integration:

$$p_{\mathbf{x}}(\mathrm{x}) = \int_{\mathcal{N}} p_{(\mathbf{x},\mathbf{y})}(\mathrm{x},\mathrm{y}).d\mathcal{N}(\mathrm{y}) \qquad \text{et} \qquad p_{\mathbf{y}}(\mathrm{y}) = \int_{\mathcal{M}} p_{(\mathbf{x},\mathbf{y})}(\mathrm{x},\mathrm{y}).d\mathcal{M}(\mathrm{x})$$

The mean value of the joint measure is

$$\bar{\mathrm{z}} \in \mathbb{E}[\ \mathbf{z}\ ] = \mathbb{E}[\ \mathbf{x}\ ] \times \mathbb{E}[\ \mathbf{y}\ ] \iff \bar{\mathrm{x}} \in \mathbb{E}[\ \mathbf{x}\ ] \quad \text{and} \quad \bar{\mathrm{y}} \in \mathbb{E}[\ \mathbf{y}\ ]$$

Assuming that there is only one mean, the covariance is given by:

$$\Sigma_{\mathbf{zz}} = \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix}$$

where $\Sigma_{\mathbf{xy}} = \Sigma_{\mathbf{yx}}^{\mathrm{T}}$ is the following cross covariance

$$\Sigma_{\mathbf{xy}} = \mathbf{E}\left[\overrightarrow{\bar{\mathrm{x}}\mathbf{x}}.\overrightarrow{\bar{\mathrm{y}}\mathbf{y}}^{\mathrm{T}} \right]$$

Independent primitives *The pdf of the joint measure $p_{\mathbf{z}}(\mathrm{z})$ can by factored in*

$$p_{(\mathbf{x},\mathbf{y})}(\mathrm{x},\mathrm{y}) = p_{\mathbf{x}}(\mathrm{x}).p_{\mathbf{y}}(\mathrm{y})$$

if and only if the random primitives $\mathbf{x}$ *and* $\mathbf{y}$ *are independent.*

The independence assumption will often be made in statistics. In this case the cross covariance can be factored in $\Sigma_{\mathbf{xy}} = \mathbf{E}\left[\overrightarrow{\bar{\mathrm{x}}\mathbf{x}} \right].\mathbf{E}\left[\overrightarrow{\bar{\mathrm{y}}\mathbf{y}}^{\mathrm{T}} \right]$ but these two integrals are null by the mean value characterization. Thus, we have : $\Sigma_{\mathbf{xy}} = 0$.

2.7 A generalization of the Normal distribution

We now present an approach based on the the information minimization to generalize the Normal distribution to a manifold. This may not be the best generalization but it allows us to approximate such a distribution by the usual Gaussian in the tangent space of the mean value in the case of small covariances. In this section the symbols log and exp denote the standard logarithmic and exponential functions in $\mathbb{R}$.

Information and uniform law As we can integrate a real valued function, the extension of the *entropy* $\mathbf{H}[\,\mathbf{x}\,]$ (or its opposite, the *information* $\mathbf{I}[\,\mathbf{x}\,]$) of a random primitive is straightforward:

$$\mathbf{I}[\,\mathbf{x}\,] = -\mathbf{H}[\,\mathbf{x}\,] = \mathbf{E}[\,\log(p_{\mathbf{x}}(\mathbf{x}))\,] = \int_{\mathcal{M}} \log(p_{\mathbf{x}}(\mathrm{x})).p_{\mathbf{x}}(\mathrm{x}).d\mathcal{M}(\mathrm{x}) \tag{23}$$

This definition is coherent since the pdf $p_{\mathcal{U}}$ that minimize the information when we only know that the measure is in a compact set $\mathcal{U}$ is the uniform density in this set:

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

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

$$\mathbf{I}[\,\mathbf{x} \mid \bar{\mathrm{x}} \in \mathbb{E}[\,\mathrm{x}\,]\ ,\ \Sigma_{\mathbf{xx}} = \Sigma\,]$$

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

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 like if we were in the open domain $\mathcal{D}(\bar{\mathrm{x}}) \in \mathbb{R}^n$.

Using a scalar Lagrange multiplier α for the normalization constraint, a vector multiplier β for the mean value and a symmetric matrix Γ for the covariance, the Lagrangian is:

$$\Lambda(\rho) = \int_{\mathcal{D}(\bar{\mathrm{x}})} \left(\rho.\log(\rho) + \alpha.\rho + \beta^{\mathrm{T}}.y.\rho + \frac{y^{\mathrm{T}}.\Gamma.y}{2}.\rho\right).d\mathcal{M}_{\bar{\mathrm{x}}}(y)$$

It is stationary at a critical point:

$$\frac{\partial\Lambda(\rho)}{\partial\rho} = 0 = \log(\rho) + 1 + \alpha + \beta^{\mathrm{T}}.y + \frac{y^{\mathrm{T}}.\Gamma.y}{2}$$

Thus we find $\rho(y) = k.\exp\left(-\beta^{\mathrm{T}}.y - \frac{y^{\mathrm{T}}.\Gamma.y}{2}\right)$ with $k = \exp(-1-\alpha)$. 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. Reporting in the constraints gives the following equations.

Theorem 3 (Normal law)
We call Normal law on the manifold $\mathcal{M}$ the pdf minimizing the information with a fixed 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 Normal law of mean $\bar{\mathrm{x}}$ (the mean value) and concentration matrix Γ is given by:

$$N_{(\bar{\mathrm{x}},\Gamma)}(\mathrm{y}) = k.\exp\left(-\frac{\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}^{T}.\Gamma.\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}}{2}\right) \tag{24}$$

where the normalization constant is

$$k^{(-1)} = \int_{\mathcal{M}} \exp\left(-\frac{\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}^{T}.\Gamma.\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}}{2}\right).d\mathcal{M}(\mathrm{y}) \tag{25}$$

The covariance and concentration are related by:

$$\Sigma = k.\int_{\mathcal{M}} \overrightarrow{\bar{\mathrm{x}}\mathrm{y}}.\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}^{T}.\exp\left(-\frac{\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}^{T}.\Gamma.\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}}{2}\right).d\mathcal{M}(\mathrm{y}) \tag{26}$$

From the concentration matrix, we can compute the covariance of the random primitive, at least numerically, but the reverse is more difficult. As $y^{\mathrm{T}}.\Gamma.y = \mathrm{Tr}(\Gamma.y.y^{\mathrm{T}})$, the information of this pdf simplifies to:

$$\mathbf{I}\left[\,N_{(\bar{\mathrm{x}},\Gamma)}\,\right] = \log(k) - \frac{1}{2}.\int_{\mathcal{M}} y^{\mathrm{T}}.\Gamma.y.N_{(\bar{\mathrm{x}},\Gamma)}.d\mathcal{M}_{\bar{\mathrm{x}}}(y) = \log(k) - \frac{1}{2}.\mathrm{Tr}\left(\Gamma.\Sigma\right)$$

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:

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

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.

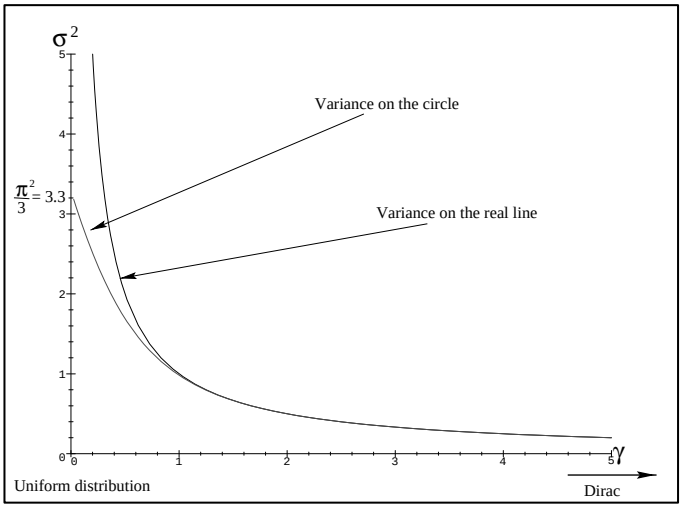


Figure 3: Variance σ^2 with respect to the concentration parameter γ on the circle of radius 1 and the real line. This variance tends toward $\sigma_0^2 = \pi^2/3$ for the uniform distribution on the circle ($\gamma = 0$) whereas it tends to infinity for the uniform measure on $\mathbb{R}$. For a strong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

As the circle is compact, the variance cannot becomes infinite as in the real case. If the concentration parameter γ goes to zero (which corresponds to $\sigma^2 \to +\infty$ in the real case), a Taylor

expansion gives $\sigma^2 = a^2/3 + O(\gamma)$. Thus, the maximal variance on the circle is

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

As expected, the Normal density of concentration 0 is the uniform density. On the other hand, if γ goes to infinity, the variance goes to zero and the density tends to a Dirac (see figure 3).

Approximation for a small Σ 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 [Chavel, 1993, 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):

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

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

Theorem 4 (Approximate normal density)
In a complete Riemannian manifold, the normal density is $N(\mathrm{y}) = k.\exp(-\overrightarrow{\mathrm{xy}}^T.\Gamma.\overrightarrow{\mathrm{xy}}/2)$. *Let* r *be the injection radius at the mean point. The normalization constant and the concentration matrices are approximated by the following expressions for a covariance matrix* Σ *of small variance* $\sigma^2 = \mathrm{Tr}(\Sigma)$*:*

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

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* $\forall k \in \mathbb{R}^+,\ \lim_{0+} x^{-k}.\varepsilon(x) = 0$

Discussion The information minimization approach to generalize the normal distribution to Riemannian manifolds is interesting since we obtain a whole family of densities going from the Dirac to the uniform distribution (or the uniform measure if the manifold in 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 statistical purposes.

It would be interesting to see if there exists similar families of densities that are continuous and differentiable at the cut locus. A starting point could be to extend some of the distributions used in directional statistics (statistics on the sphere) as in [Bingham, 1974, Jupp and Mardia, 1989, Kent, 1992, Mardia, 1995].

2.8 Mahalanobis distance and χ^2 law

The problem we are now tackling is to determine if a measure $\hat{\mathrm{y}}$ was drawn from a random primitive $\mathbf{x}$. From a statistical point of view, the pdf of the measurement process is often too rich an information to be estimated or handled and, in practice, one often characterize this pdf by its moments, and more particularly by the mean and the covariance[6]. We denote it by

$$\mathbf{x} \sim (\bar{\mathrm{x}}, \Sigma_{\mathbf{xx}})$$

Based on these characteristics only, we want to decide if the measure $\hat{\mathrm{y}}$ is compatible with this measurement process.

In the vectorial case and assuming a Gaussian distribution, the χ^2 test is well adapted to do that. This test measures the probability of the Mahalanobis distance $\chi^2 = (\hat{x} - \bar{x})^{\mathrm{T}} \cdot \Sigma^{(-1)}_{\boldsymbol{xx}} \cdot (\hat{x} - \bar{x})$ assuming that $\hat{x}$ is drawn from $\boldsymbol{x}$. If the probability is too small (i.e. χ^2 is too large), the hypothesis is rejected.

This 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 primitive and not only the normal one.

Definition 6 (Mahalanobis distance)
We call Mahalanobis distance between a random primitive $\mathbf{x} \sim (\bar{\mathrm{x}}, \Sigma_{\mathbf{xx}})$ *and a (deterministic) point* y *on the manifold the value*

$$\mu^2_{\mathbf{x}}(\mathrm{y}) = \overrightarrow{\bar{\mathrm{x}}\mathrm{y}}^{T}.\Sigma^{(-1)}_{\mathbf{xx}}.\overrightarrow{\bar{\mathrm{x}}\mathrm{y}} \tag{27}$$

In fact, the Mahalanobis distance measures the distance between y and the mean value $\bar{\mathrm{x}}$ according to the "metric" $\Sigma^{(-1)}_{\mathbf{xx}}$.

Property 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:

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

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

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

Theorem 5 *The expected Mahalanobis distance of a random primitive with itself is independent of the distribution and does only depend on the dimension of the manifold:*

$$\mathbf{E}\left[\,\mu^2_{\mathbf{x}}(\mathbf{x})\,\right] = n \tag{28}$$

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

[6] If we have to assume a density for this random primitive, the least informative one is the normal density that can be approximated by theorem 4.

A generalized χ^2 law Assuming that the random primitive $\mathbf{x} \sim (\bar{\mathrm{x}}, \Sigma_{\mathbf{xx}})$ 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 primitive (at least up to the order $O(\sigma^3)$):

Theorem 6 (Approximate χ^2 law)
With the same hypotheses as for the approximate normal law, the χ^2 probability is

$$\Pr\{\chi^2 \leq \alpha^2\} = (2\pi)^{-\frac{n}{2}} \int_{\|x\| \leq \alpha} \exp\left(-\frac{\|x\|^2}{2}\right) .dx + O(\sigma^3) + \varepsilon\left(\frac{\sigma}{r}\right) \tag{29}$$

while the density is

$$p_{\chi^2}(u) = \frac{1 + O(\sigma^3) + \varepsilon\left(\frac{\sigma}{r}\right)}{2.\mathbf{\Gamma}\left(\frac{n}{2}\right)} \left(\frac{u}{2}\right)^{\frac{n}{2}-1} \exp\left(-\frac{u}{2}\right) \tag{30}$$

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 [Press et al., 1991]).

In practice, one often use this law to test if a measurement $\hat{\mathrm{x}}$ has been drawn from a random primitive $\mathbf{x}$: 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 $\mu^2_{\mathbf{x}}(\hat{\mathrm{x}}) \leq \alpha^2$.

3 The case of connected Lie groups

A Lie group being a differentiable manifold, we are just left with the choice of a Riemannian metric to apply our statistical framework. Firstly, we detail the two canonical Riemannian metrics (left- and right-invariant metrics). Then, choosing the left invariant metric, we show that all the exponential could be identified to the exponential chart at the identity (called the principal chart). From a computer science point of view, this means that we only have only one chart to deal with in order to cover the whole group (with special cases of the points of the cut locus). In Section 3.3, we investigate the propagation of the pdfs of random transformations for the group operations: left and right translation by a deterministic transformation, inversion, composition of two random transformations. Interestingly, this last pdf is based on a generalisation of the convolution product used for the addition of two random variables in $\mathbb{R}^n$. Section 3.4 details the simplifications of our gradient descent algorithm to obtain the mean using the group operations and the properties of the left invariant metric, while Section 3.5 investigates the propagation of the Karcher mean values under the group operations: if the left translation of a random transformation (by a deterministic transformation) is stable, we need to require the left and right invariance of the metric in order to ensure the stability of the mean values with respect to the right translation and the inversion. An example with 2D rigid transformations, where the metric is only left invariant, shows that the empirical mean value of the inverse is generally not the inverse of the mean value. Last but not least, we show in section 3.6 that the Normal distribution is stable by left translation.

3.1 Left and Right Invariant Metrics

We have two canonical way to compare the tangent spaces at different points of the manifold: the left translation $L_{\mathrm{g}}(\mathrm{f}) = \mathrm{g} \circ \mathrm{f}$ and the right translation $R_{\mathrm{g}}(\mathrm{f}) = \mathrm{f} \circ \mathrm{g}$. These applications can obviously

be differentiated to provide linear operators from $T_{\mathrm{f}}\mathcal{G}$ to $T_{(\mathrm{g}\circ\mathrm{f})}\mathcal{G}$ and $T_{(\mathrm{f}\circ\mathrm{g})}\mathcal{G}$. Let us denote by J_L and J_R the expression of the differentials (in a given chart) at the identity:

$$J_L(f) = \left.\frac{\partial(\mathrm{f}\circ\mathrm{e})}{\partial \mathrm{e}}\right|_{\mathrm{e}=\mathrm{Id}} \qquad \text{and} \qquad J_R(f) = \left.\frac{\partial(\mathrm{e}\circ\mathrm{f})}{\partial \mathrm{e}}\right|_{\mathrm{e}=\mathrm{Id}}$$

Let v be a tangent vector at the identity. It corresponds to a tangent vector of $T_{\mathrm{f}}\mathcal{M}$ $v_l = J_L(\mathrm{f}).v$ by left translation and $v_r = J_R(\mathrm{f}).v$ by right translation.

Let now $\langle\, .\mid . \,\rangle$ be any dot product on the tangent space at the identity, characterised by a symmetric matrix Q in our chart. We can translate this dot product in any tangent space by left or right translation. If v and W are tangent vectors at f, we define the left-invariant metric by:

$$\langle\, v \mid w \,\rangle_{\mathrm{f}} = \left\langle\, J_L(f)^{(-1)}.v \;\middle|\; J_L(\mathrm{f})^{(-1)}.w \,\right\rangle = v^{\mathrm{T}}.J_L(\mathrm{f})^{(-\mathrm{T})}.Q.J_L(\mathrm{f})^{(-1)}.w,$$

which means that the expression of the metric is in our chart: $G_L(\mathrm{f}) = J_L(\mathrm{f})^{(-\mathrm{T})}.Q.J_L(\mathrm{f})^{(-1)}$. The right-invariant metric is defined similarly with $G_R(\mathrm{f}) = J_R(\mathrm{f})^{(-\mathrm{T})}.Q.J_R(\mathrm{f})^{(-1)}$.

We note that the volume element associated to the these invariant metrics are the Haar measures of the group ($|J| = |\det(J)|$) [Pennec and Ayache, 1998]:

$$d_L\mathcal{G}(f) = \sqrt{|G_L(f)|}.df = \lambda.\frac{df}{|J_L(f)|} \qquad \text{and} \qquad d_R\mathcal{G}(f) = \sqrt{|G_R(f)|}.df = \lambda.\frac{df}{|J_R(f)|}$$

Since the metric is left (resp. right) invariant, the geodesics are globally conserved by left (resp. right) translation and one can determine only the geodesics starting from the identity. One should be careful that left and right invariant metrics are generally different. However, if the group is compact (for instance rotations), there exists a left and right invariant metric [Spivak, 1979, Carmo, 1992] and the geodesics for this metric (starting from the identity) are the one parameter sub-groups, i.e. the curves verifying $\forall (s,t) \in \mathbb{R}^2,\ \gamma(s+t) = \gamma(s)\circ\gamma(t) = \gamma(t)\circ\gamma(s)$. This equation is often easier to solve that the standard second order differential equations system, but it is not valid for only locally compact groups (such as rigid transformations for instance).

3.2 Principal chart

From now on, we consider that the group is geodesically complete and that geodesics are determined for the left-invariant metric. The same developments could of course be done (with different results) with the right invariant one.

We call *principal chart* the exponential chart at the identity and we denote by $\vec{\mathrm{f}}$ the representation of f in this chart. To simplify the notations, we will write simply exp and log the exponential map and its inverse function at the identity. As the metric matrix Q is symmetric and positive, it can be written $Q = A^{\mathrm{T}}.A$. Thus, changing the coordinate system into $\overrightarrow{\mathrm{f}}' = A.\overrightarrow{\mathrm{f}}$ allows to obtain a principal chart where $Q = \mathrm{Id}$. We will consider $Q = \mathrm{Id}$ in the sequel.

Let $\gamma_{(\mathrm{Id},\vec{\mathrm{f}})}$ be a geodesic starting at the identity with tangent vector $\vec{\mathrm{f}}$. We have $\gamma(0) = \mathrm{Id}$ and $\gamma(1) = \mathrm{f}$. Translating this geodesic by g, we obtain a new geodesic $\delta = \mathrm{g}\circ\gamma_{(\mathrm{Id},\vec{\mathrm{f}})}$ starting at g with tangent vector $J_L(\vec{\mathrm{g}}).\overrightarrow{\mathrm{f}}$ and arriving at time one at $\mathrm{g}\circ\mathrm{f}$. Thus, by definition of the exponential map, we get $\exp_{\mathrm{g}}(J_L(\vec{\mathrm{g}}).\overrightarrow{\mathrm{f}}) = \mathrm{g}\circ\mathrm{f}$. Taking $v = J_L(\vec{\mathrm{g}}).\vec{\mathrm{f}}$, we obtain the following formula for the exponential map at any point:

$$\exp_{\mathrm{g}}(v) = \mathrm{g}\circ\exp(J_L(\vec{\mathrm{g}})^{(-1)}.v) \tag{31}$$

Taking now $v = J_L(\vec{g}).(\vec{g}^{(-1)} \circ \vec{f})$, we obtain the expression $\overrightarrow{gf}$ of the f in the exponential chart at g (expressed in the basis induced by the principal chart):

$$\overrightarrow{gf} = \log_{\vec{g}}(f) = J_L(\vec{g}).(\vec{g}^{(-1)} \circ \vec{f}) \tag{32}$$

The main interest of all these developments is that, from a computer science point of view, we can use only one chart (the principal chart) and obtain all the others by the left translation.

The distance between two transformations f and g is the length of the minimising geodesic joining the two points. By definition, such a geodesics starts at g with tangent vector $\overrightarrow{gf}$:

$$\mathrm{dist}(f,g)^2 = \|\overrightarrow{gf}\|^2_{\vec{g}} = \overrightarrow{gf}^{\mathrm{T}}.Q(\vec{g}).\overrightarrow{gf} = (\vec{g}^{(-1)} \circ \vec{f})^{\mathrm{T}}.(\vec{g}^{(-1)} \circ \vec{f}) = \|\vec{g}^{(-1)} \circ \vec{f}\|^2_{\mathrm{Id}} = \mathrm{dist}(g^{(-1)} \circ f,\ \mathrm{Id})^2 \tag{33}$$

3.3 Propagation of the pdfs for some simple group operations

Since we chose the left invariant metric on our Lie group, the probability density functions are defined using the left Haar measure:

$$\Pr(\mathbf{f} \in \mathcal{Y}) = \int_{\mathcal{Y}} p(y).d_L\mathcal{G}(y)$$

As already mentioned, the left and right Haar measures are generally different. Their difference is quantified by the *module* $\Delta\mathcal{G}(g)$, defined by the relation: $d_L\mathcal{G}(g) = \Delta\mathcal{G}(g).d_R\mathcal{G}(g)$. With our notations, we have $\Delta\mathcal{G}(g) = |J_R(g)|/|J_L(g)|$. The group is said unimodular if $\Delta\mathcal{G}(g)$ is constant at each point of the group (i.e. equal to 1 up to a normalisation factor). A compact group is always unimodular but only locally compact group may have different left and right Haar measures (they may have in this case different left and right invariant metrics).

Theorem 7 (Translation of a random transformation)
Let $\mathbf{f}$ *be a random transformation of pdf* $p_{\mathbf{f}}$*, and* f_o *a deterministic transformation. The pdf of* $\mathbf{g}_l = f_o \circ \mathbf{f}$ *(left translation of* $\mathbf{f}$*) is:*

$$p_{(f_o \circ \mathbf{f})}(g) = p_{\mathbf{f}}(f_o^{(-1)} \circ g) \tag{34}$$

whereas the pdf of the right translation is:

$$p_{(\mathbf{f} \circ f_o)}(g) = \frac{\Delta\mathcal{G}(g \circ f_o^{(-1)})}{\Delta\mathcal{G}(g)}.p_{\mathbf{f}}(g \circ f_o^{(-1)}) \tag{35}$$

Proof: The probability of $\mathbf{g}_l = f_o \circ \mathbf{f}$ to be in a set $\mathcal{Y} \subset \mathcal{G}$ is:

$$P(f_o \circ \mathbf{f} \in \mathcal{Y}) = P(\mathbf{f} \in f_o^{(-1)} \circ \mathcal{Y}) = \int_{f_o^{(-1)} \circ \mathcal{Y}} p_{\mathbf{f}}(h).d_L\mathcal{G}(h) \int_{\mathcal{Y}} p_{\mathbf{f}}(f_o^{(-1)} \circ g).d_L\mathcal{G}(f_o^{(-1)} \circ g)$$

where the last equality is obtained by the change of variable $h = f_o^{(-1)} \circ g$. The first result is obtained using the left invariance of the measure.
Now, for the right translation, the probability of $\mathbf{g}_r = \mathbf{f} \circ f_o$ to be in a set $\mathcal{Y} \subset \mathcal{G}$ is:

$$\Pr((\mathbf{f} \circ f_o) \in \mathcal{Y}) = \Pr(\mathbf{f} \in (\mathcal{Y} \circ f_o^{(-1)})) = \int_{(\mathcal{Y} \circ f_o^{(-1)})} p_{\mathbf{f}}(h).d_L\mathcal{G}(h)$$

With the change of variable $h = g \circ f_o^{(-1)}$, we have

$$\Pr((\mathbf{f} \circ f_o) \in \mathcal{Y}) = \int_{\mathcal{Y}} p_{\mathbf{f}}(g \circ f_o^{(-1)}).d_L\mathcal{G}(g \circ f_o^{(-1)}) = \int_{\mathcal{Y}} \Delta\mathcal{G}(g \circ f_o^{(-1)}).p_{\mathbf{f}}(g \circ f_o^{(-1)}).d_R\mathcal{G}(g \circ f_o^{(-1)})$$

Since $d_R\mathcal{G}$ is the right invariant measure, we can simplify and come back to the left invariant measure:

$$\Pr((\mathbf{f}\circ \mathrm{f}_o)\in\mathcal{Y}) = \int_{\mathcal{Y}} \Delta\mathcal{G}(\mathrm{g}\circ\mathrm{f}_o^{(-1)}).p_{\mathbf{f}}(\mathrm{g}\circ\mathrm{f}_o^{(-1)}).d_R\mathcal{G}(\mathrm{g}) = \int_{\mathcal{Y}} \frac{\Delta\mathcal{G}(\mathrm{g}\circ\mathrm{f}_o^{(-1)})}{\Delta\mathcal{G}(\mathrm{g})}.p_{\mathbf{f}}(\mathrm{g}\circ\mathrm{f}_o^{(-1)}).d_L\mathcal{G}(\mathrm{g})$$

This gives the second result. ■

Theorem 8 Composition of random transformations
Let $\mathbf{f}_1$ *and* $\mathbf{f}_2$ *be two random transformations of pdf* $p_{\mathbf{f}_1}$ *and* $p_{\mathbf{f}_2}$*. The pdf of their composition is:*

$$p_{(\mathbf{f}_1\circ\mathbf{f}_2)}(\mathrm{f}) = \int_{\mathcal{G}} p_{\mathbf{f}_1}(\mathrm{g}).p_{\mathbf{f}_2}(\mathrm{g}^{(-1)}\circ\mathrm{f}).d_L\mathcal{G}(\mathrm{g}) \tag{36}$$

Proof: Let $\mathcal{F}\subset\mathcal{G}$ be a set of transformation. The set of couples $(\mathrm{g}_1,\mathrm{g}_2)$ such that $\mathrm{g}_1\circ\mathrm{g}_2\in\mathcal{F}$ is $\mathcal{A} = \left\{(\mathrm{g}_1,\mathrm{g}_2)\ /\ \mathrm{g}_1\in\mathcal{G}\ ,\ \mathrm{g}_2\in\mathrm{g}_1^{(-1)}\circ\mathcal{F}\right\}$. Thus, the probability of $\mathbf{f}=\mathbf{f}_1\circ\mathbf{f}_2$ to be in $\mathcal{F}$ is:

$$\begin{aligned}
\Pr(\mathbf{f}_1\circ\mathbf{f}_2\in\mathcal{F}) &= \int_{\mathcal{G}}\left(\int_{\mathrm{g}_1^{(-1)}\circ\mathcal{F}} p_{\mathbf{f}_2}(\mathrm{g}_2).d_L\mathcal{G}(\mathrm{g}_2)\right) p_{\mathbf{f}_1}(\mathrm{g}_1).d_L\mathcal{G}(\mathrm{g}_1)\\
&= \int_{\mathcal{G}}\left(\int_{\mathcal{F}} p_{\mathbf{f}_2}(\mathrm{g}_1^{(-1)}\circ\mathrm{g}_2).d_L\mathcal{G}(\mathrm{g}_2)\right) p_{\mathbf{f}_1}(\mathrm{g}_1).d_L\mathcal{G}(\mathrm{g}_1)\\
&= \int_{\mathcal{F}}\left(\int_{\mathcal{G}} p_{\mathbf{f}_2}(\mathrm{g}_1^{(-1)}\circ\mathrm{g}_2).p_{\mathbf{f}_1}(\mathrm{g}_1).d_L\mathcal{G}(\mathrm{g}_1)\right) d_L\mathcal{G}(\mathrm{g}_2)
\end{aligned}$$

■

Analogy with the convolution product Equation (36) is remarkably similar to the classical convolution product obtain for the addition of two random vectors in $\mathbb{R}^n$:

$$p_{(\boldsymbol{x}+\boldsymbol{y})}(z) = p_{\boldsymbol{x}}\circledast p_{\boldsymbol{y}}(z) = \int_{\mathbb{R}^n} p_{\boldsymbol{x}}(t).p_{\boldsymbol{y}}(z-t).dt$$

This formula can be found as a particular case of equation (36) for the group of translations of $\mathbb{R}^n$ for which we have $x^{(-1)}=-x$, $x\circ y = x+y$ and $d_L\mathcal{G}(x)=dx$.

Theorem 9 (Inversion of a random transformation
Let $\mathbf{f}$ *be a random transformation of density* $p_{\mathbf{f}}$*. The density of the inverse transformation is:*

$$p_{\mathbf{f}^{(-1)}}(\mathrm{g}) = \Delta\mathcal{G}(\mathrm{g}^{(-1)}).p_{\mathbf{f}}(\mathrm{g}^{(-1)}) \tag{37}$$

Proof: The first step is to determine the relation between $d_L\mathcal{G}(\mathrm{g}^{(-1)})$ and $d_L\mathcal{G}(\mathrm{g})$. We have:

$$\begin{aligned}
d_L\mathcal{G}(\mathrm{g}^{(-1)}) &= \left|J_L(\mathrm{g}^{(-1)})\right|^{(-1)}.d(\mathrm{g}^{(-1)} = \left|J_L(\mathrm{g}^{(-1)})\right|^{(-1)}\cdot\left|\frac{\partial \mathrm{g}^{(-1)}}{\partial \mathrm{g}}\right|.d\mathrm{g} = \left|\frac{\partial \mathrm{g}}{\partial \mathrm{g}^{(-1)}}.\frac{\partial \mathrm{g}^{(-1)}\circ\mathrm{e}}{\partial \mathrm{e}}\right|_{\mathrm{e}=\mathrm{Id}}\Bigg|^{(-1)}.d\mathrm{g}\\
&= \left|\frac{\partial \mathrm{e}^{(-1)}\circ\mathrm{g}}{\partial \mathrm{e}}\right|_{\mathrm{e}=\mathrm{Id}}\Bigg|^{(-1)}.d\mathrm{g} = |J_R(\mathrm{g})|^{(-1)}\cdot\left|\frac{\partial \mathrm{e}^{(-1)}}{\partial \mathrm{e}}\right|_{\mathrm{e}=\mathrm{Id}}\Bigg|^{(-1)}.d\mathrm{g} = \frac{|J_L(\mathrm{g})|}{|J_R(\mathrm{g})|}.d_L\mathcal{G}(\mathrm{g})
\end{aligned}$$

the last equality being due to $\partial \mathrm{e}^{(-1)}/\partial \mathrm{e} = -\mathrm{Id}$ at $\mathrm{e}=\mathrm{Id}$. Thus, we obtain:
$d_L\mathcal{G}(\mathrm{g}^{(-1)}) = \Delta\mathcal{G}(\mathrm{g}^{(-1)}).d_L\mathcal{G}(\mathrm{g})$. Now, if $\mathcal{F}$ is a set of transformations, we have:

$$\begin{aligned}
\Pr(\mathbf{f}^{(-1)}\in\mathcal{F}) &= \Pr(\mathbf{f}\in\mathcal{F}^{(-1)}) = \int_{\mathcal{F}^{(-1)}} p_{\mathbf{f}}(\mathrm{g}).d_L\mathcal{G}(\mathrm{g})\\
&= \int_{\mathcal{F}} p_{\mathbf{f}}(\mathrm{g}^{(-1)}).d_L\mathcal{G}(\mathrm{g}^{(-1)}) = \int_{\mathcal{F}} \Delta\mathcal{G}(\mathrm{g}^{(-1)})p_{\mathbf{f}}(\mathrm{g}^{(-1)}).d_L\mathcal{G}(\mathrm{g})
\end{aligned}$$

which gives the proposed density. ■

3.4 Obtaining the Karcher mean values

The Karcher means are minimising the variance function defined with the left invariant distance as defined in the previous section. There are few simplifications due to the left invariant structure: thanks to formula (32), we can write:

$$\mathbf{E}\left[\overrightarrow{\mathrm{gf}}\right] = J_L(\vec{\mathrm{g}}).\int_{\mathcal{G}/C(\mathrm{g})} (\vec{\mathrm{g}}^{(-1)} \circ \vec{\mathrm{f}}).dP(\mathrm{f})$$

Thus, assuming that the conditions of theorem (2) (i.e. that the variance is finite and the cut locus has a null measure), our gradient descent algorithm to obtain the mean can be entirely written in the principal chart (we focus here on the practical case: the empirical mean value):

$$\vec{\mathrm{g}}_{t+1} = \vec{\mathrm{g}}_t \circ \left(\frac{1}{n}\sum_i \vec{\mathrm{g}}_t^{(-1)} \circ \vec{\mathrm{f}}_i\right) \tag{38}$$

which simply corresponds to the following algorithm: (left) translate all measured values to the origin using the inverse of the current transformation estimation; compute the barycenter of the transformations in the principal chart; use that barycenter as the (right) increment for the current value.

Example with 3D rigid rotations We have shown in [Pennec, 1996] that the principal chart is the rotation vector (the unit axis of rotation multiplied by the angle of rotation) with an norm (angle) less that π. Thus, the mean rotation vector satisfies[7]:

$$\mathbf{E}\left[\,\bar{r}^{(-1)} \circ \boldsymbol{r}\,\right] = 0$$

If the mean rotation $\bar{r}$ is sufficiently centered in the principal chart (i.e. if the angle of rotation is small), and if the distribution is sufficiently peaked, then a Taylor expansion shows that the classical expectation is an approximation of the Karcher one:

$$\bar{r}^{(-1)} \circ r = r - \bar{r} + O(\|\bar{r}\|^2) + O(\|\bar{r} - r\|^2)$$

However, as soon as the Karcher mean value goes farther from the origin, the difference with the classical expectation is more and more visible (see figure 4).

3.5 Properties of the Karcher expectation for the group operations

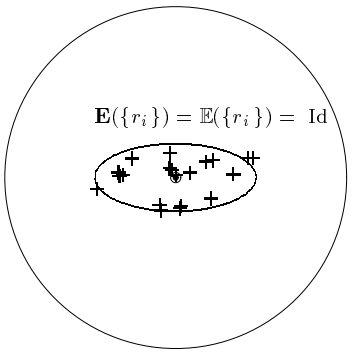
Let us now turn to the properties of the mean transformations for the group operations.

Theorem 10 (Left translation of a random transformation)

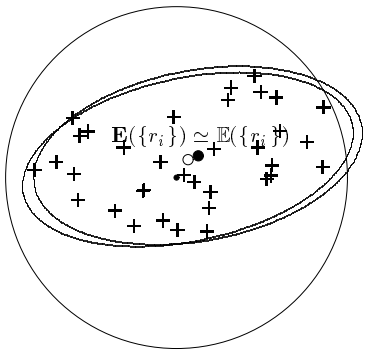
$$\mathbb{E}[\,\mathrm{g} \circ \mathbf{f}\,] = \mathrm{g} \circ \mathbb{E}[\,\mathbf{f}\,] \qquad \textit{and} \qquad \mathbb{E}[\,\{\mathrm{g} \circ \mathrm{f}_i\}\,] = \mathrm{g} \circ \mathbb{E}[\,\{f_i\}\,] \tag{39}$$

This result also holds for the sets of central primitives of any order.

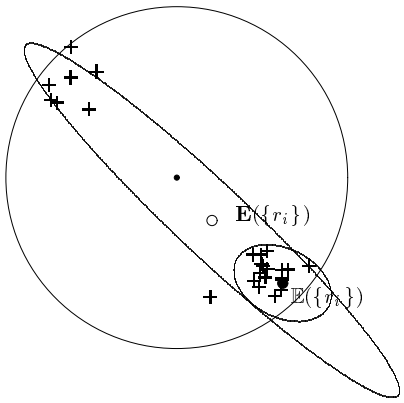
[7] the formulas to compute the composition, inversion and all the necessary derivatives are provided in [Pennec, 1996].



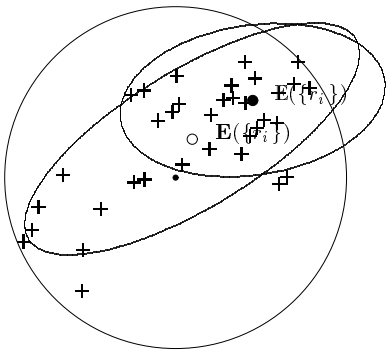
(a) Standard and Fréchet expectations are identical at the origin.



(b) When the Fréchet expectation is near the origin, the standard expectation can be considered as a first order approximation.



(c) Near the Domain boundary, standard and Fréchet expectations greatly differ. We note that with this relatively small noise, it would be possible to detect the boundary effect.



(d) With a greater amount of noise, it is no longer possible to guess the correct clustering to avoid the boundary effect.

Figure 4: Behaviour of expected rotations and their covariance matrices: projection of the measured rotation vectors, their standard and Fréchet expectation and the corresponding uncertainty ellipsoid at $\chi^2 = 15$ onto the (r_x, r_y) plane. The circle represent the domain boundary for the rotation vector $(\theta = \|r\| = \pi)$. Remember that when we cross this boundary on one side at point $r = \theta.n$, we reenter the domain at the symmetric point $r' = -\theta.n$. In fact, the write way to visualise the covariance matrices would be to left-translate all measurements with respect to the "mean" point, as in figure (4(a)) so that the representation corresponds to the exponential chart at this point.

Proof: Thanks to the (left) invariant distance, we have:
$$\sigma^2_{\mathrm{g}\circ\mathbf{f}}(\mathrm{h}) = \mathbf{E}\left[\ \mathrm{dist}(\mathrm{g}\circ\mathbf{f},\mathrm{h})^2\ \right] = \mathbf{E}\left[\ \mathrm{dist}(\mathbf{f},\mathrm{g}^{(-1)}\circ\mathrm{h})^2\ \right] = \sigma^2_{\mathbf{f}}(\mathrm{g}^{(-1)}\circ\mathrm{h})$$
Thus, $\bar{\mathrm{f}}$ minimises $\sigma^2_{\mathbf{f}}(\mathrm{f})$ if and only if $\bar{\mathrm{h}} = \mathrm{g}\circ\bar{\mathrm{f}}$ minimises $\sigma^2_{\mathrm{g}\circ\mathbf{f}}(\mathrm{h})$. The same argument holds for other types of central features (since only the invariance of the distance was used) and for the empirical mean. ■

Theorem 11 (Right translation and inversion of a random transformation)
If the distance is **left and right** *invariant, we have:*

$$\mathbb{E}[\ \mathbf{f}\circ\mathrm{g}\ ] = \mathbb{E}[\ \mathbf{f}\ ]\circ\mathrm{g} \qquad \textit{and} \qquad \mathbb{E}[\ \{\mathrm{f}_i\}\circ\mathrm{g}\ ] = \mathbb{E}[\ \{f_i\}\ ]\circ\mathrm{g} \tag{40}$$

$$\mathbb{E}\left[\ \mathbf{f}^{(-1)}\ \right] = \mathbb{E}[\ \mathbf{f}\ ]^{(-1)} \qquad \textit{and} \qquad \mathbb{E}\left[\ \{\mathrm{f}_i\}^{(-1)}\ \right] = \mathbb{E}[\ \{f_i\}\ ]^{(-1)} \tag{41}$$

These results also holds for the sets of central primitives of any order.

Proof: The first result is obtained as in theorem (10) but using the right invariance. For the inversion, let us first notice that
$$\mathrm{dist}(\mathrm{g}^{(-1)},\mathrm{f}) = \mathrm{dist}(\mathrm{Id},\mathrm{g}\circ\mathrm{f}) = \mathrm{dist}(\mathrm{f}^{(-1)},\mathrm{g})$$
where the first equality is obtained using the left invariance of teh distance and the second one using the right invariance. Thus, we have
$$\sigma^2_{\mathbf{f}^{(-1)}}(\mathrm{g}) = \mathbf{E}\left[\ \mathrm{dist}(\mathbf{f}^{(-1)},\mathrm{g})^2\ \right] = \mathbf{E}\left[\ \mathrm{dist}(\mathbf{f},\mathrm{g}^{(-1)})^2\ \right] = \sigma^2_{\mathbf{f}}(\mathrm{g}^{(-1)})$$
The result is then obtained as before. The same argument holds for other types of central features (since only the invariance of the distance was used) and for the empirical mean. ■

Since a compact group has a left and right invariant distance, the stability of the set of Karcher means by the group operations is ensured. It may seem weird that this is generally not the case for a non compact group. We present below an example on 2D rigid transformations where the stability is observed for the rotation part (a compact group) but not for the translation part which is responsible for the non compacity.

Example with 2D rigid transformations A 2D rigid transformation f is characterised by an angle of rotation $\theta \in [-\pi;\pi[$ and a translation vector $t \in \mathbb{R}^2$. The composition of two transformations is given by:

$$f_1 \circ f_2 = \left| \begin{array}{l} (\theta_1+\theta_2)\ [2\pi] \\ R(\theta_1).t_2 + t_1 \end{array} \right. \qquad \text{with} \qquad R(\theta) = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

The modulus on the angle mean that we keep the value between $-\pi$ and $+\pi$.

Excluding the rotations of $\pm\pi$, this representation is the principal chart for the left invariant distance. Indeed, we have:

$$J_L(f) = \left.\frac{\partial(f\circ e)}{\partial e}\right|_{e=\mathrm{Id}} = \begin{bmatrix} 1 & 0 \\ 0 & R(\theta) \end{bmatrix}$$

Choosing an identity metric at the identity, the representation of the left invariant metric is $G_L(f) = J_L(f)^{(-\mathrm{T})}.J_L(f)^{(-1)} = \mathrm{Id}$. Thus, the Christoffel symbols vanish and the geodesics starting from 0 (the identity) are strait lines.

Now, let $f_1 = (\pi/4\,;\, -\sqrt{2}/2\,;\, \sqrt{2}/2)$, $f_2 = (0\,;\, \sqrt{2}\,;\, 0)$ and $f_3 = (-\pi/4\,;\, -\sqrt{2}/2\,;\, -\sqrt{2}/2)$ be three transformations in our principal chart. One easily verify that their barycenter is the origin. Thus, $\bar{f} = (0\,;\, 0\,;\, 0) = \mathrm{Id}$ is a Karcher mean. In this case, this is the only one:

$$\mathbb{E}[\, f_1, f_2, f_3\,] = \{(0\,;\, 0\,;\, 0)\}$$

The inverse of these transformations is easy to compute: $f_1^{(-1)} = (-\pi/4; 0; -1)$, $f_2^{(-1)} = (0; -\sqrt{2}; 0)$ and $f_3^{(-1)} = (+\pi/4\,;\, 0\,;\, 1)$. This time, the barycenter $\overline{f^{(-1)}} = (0\,;\, -\sqrt{2}/3\,;\, 0)$ is null for the rotation part, but not for the translation: the mean value of the inverse transformations is not the identity

$$\mathbb{E}[\, f_1^{(-1)}, f_2^{(-1)}, f_3^{(-1)}\,] = \left\{(0\,;\, -\sqrt{2}/3\,;\, 0)\right\}$$

3.6 A word on the propagation of the Normal law parameters

Theorem 12 (Left translation of a Normal random transformation)
If $\mathbf{f}$ *is a random transformation following the Normal law* $N_{(\bar{\mathrm{f}},\Gamma_{\mathbf{f}})}$ *(in the conditions of theorem 3), then* $\mathrm{g}\circ\mathbf{f}$ *is a random transformation following the Normal law* $N_{(\mathrm{g}\circ\bar{\mathrm{f}},\Gamma_{\mathrm{g}\circ\mathbf{f}})}$ *with*

$$\Gamma_{\mathrm{g}\circ\mathbf{f}} = J^{(-T)}.\Gamma_{\mathbf{f}}.J^{(-1)} \qquad \textit{where} \qquad J = \left.\frac{\partial(\vec{\mathrm{g}}\circ\vec{\mathrm{f}})}{\partial\vec{\mathrm{f}}}\right|_{\vec{\mathrm{f}}=\bar{\mathrm{f}}}$$

Moreover, the covariances matrices are related by $\Sigma_{\mathrm{g}\circ\mathbf{f}} = J.\Sigma_{\mathbf{f}}.J^T$.

Proof: According to eq. (34), the random transformation $\mathrm{g}\circ\mathbf{f}$ has the pdf:

$$p(\mathrm{h}) = N_{(\bar{\mathrm{f}},\Gamma_{\mathbf{f}})}(\mathrm{g}^{(-1)}\circ\mathrm{h}) = k(\bar{\mathrm{f}},\Gamma_{\mathbf{f}}).\exp\left(-\frac{1}{2}.\left(\vec{\mathrm{f}}^{(-1)}\circ(\vec{\mathrm{g}}^{(-1)}\circ\mathrm{h})\right)^{\mathrm{T}}.J_L(\bar{\mathrm{f}})^{\mathrm{T}}.\Gamma_{\mathbf{f}}.J_L(\bar{\mathrm{f}}).\left(\vec{\mathrm{f}}^{(-1)}\circ(\vec{\mathrm{g}}^{(-1)}\circ\mathrm{h})\right)\right)$$

But we have

$$J_L(\mathrm{g}\circ\bar{\mathrm{f}}) = \left.\frac{\partial(\vec{\mathrm{g}}\circ\vec{\mathrm{f}})\circ\vec{\mathrm{e}}}{\partial\vec{\mathrm{e}}}\right|_{\vec{\mathrm{e}}=\mathrm{Id}} = \left.\frac{\partial(\vec{\mathrm{g}}\circ\vec{\mathrm{f}})}{\partial\vec{\mathrm{f}}}\right|_{\vec{\mathrm{f}}=\bar{\mathrm{f}}}.\left.\frac{\vec{\mathrm{f}}\circ\vec{\mathrm{e}}}{\partial\vec{\mathrm{e}}}\right|_{\vec{\mathrm{e}}=\mathrm{Id}} = J.J_L(\bar{\mathrm{f}})$$

with the definition of J given in the theorem. Thus, we have $J_L(\bar{\mathrm{f}}) = J^{(-1)}.J_L(\mathrm{g}\circ\bar{\mathrm{f}})$ and

$$p(\mathrm{h}) = k(\bar{\mathrm{f}},\Gamma_{\mathbf{f}}).\exp\left(-\frac{1}{2}.\left((\vec{\mathrm{g}}\circ\vec{\mathrm{f}})^{(-1)}\circ h\right)^{\mathrm{T}}.J_L(\mathrm{g}\circ\bar{\mathrm{f}})^{\mathrm{T}}.\Gamma_{\mathrm{g}\circ\mathbf{f}}.J_L(\mathrm{g}\circ\bar{\mathrm{f}}).\left((\vec{\mathrm{g}}\circ\vec{\mathrm{f}})^{(-1)}\circ h\right)\right)$$

Now, since the measure is left invariant, the change of variable $\mathrm{h}\to\mathrm{g}^{(-1)}\circ\mathrm{h}$ in the definition of the normalisation coefficient gives (eqn. 25):

$$k(\bar{\mathrm{f}},\Gamma_{\mathbf{f}})^{(-1)} = \int_G \exp\left(-\frac{1}{2}.\left(\vec{\mathrm{f}}^{(-1)}\circ\mathrm{h}\right)^{\mathrm{T}}.J_L(\bar{\mathrm{f}})^{\mathrm{T}}.\Gamma_{\mathbf{f}}.J_L(\bar{\mathrm{f}}).\left(\vec{\mathrm{f}}^{(-1)}\circ\mathrm{h}\right)\right).d_L\mathcal{G}(\mathrm{h}) = k(\mathrm{g}\circ\bar{\mathrm{f}},\Gamma_{\mathrm{g}\circ\mathbf{f}})$$

Thus, the pdf p is the Normal law of mean $(\mathrm{g}\circ\bar{\mathrm{f}})$ and concentration matrix $\Gamma_{\mathrm{g}\circ\mathbf{f}}$. The relation between the covariance matrices relies on the fact that

$$\overrightarrow{(\mathrm{g}\circ\bar{\mathrm{f}})\mathrm{h}} = J_L(\vec{\mathrm{g}}\circ\vec{\bar{\mathrm{f}}}).\left(\left(\vec{\mathrm{g}}\circ\vec{\bar{\mathrm{f}}}\right)^{(-1)}\circ\vec{\mathrm{h}}\right) = J.J_L(\vec{\bar{\mathrm{f}}}).\left(\vec{\bar{\mathrm{f}}}^{(-1)}\circ\left(\vec{\mathrm{g}}^{(-1)}\circ\vec{\mathrm{h}}\right)\right) = J.\overrightarrow{\bar{\mathrm{f}}(\mathrm{g}^{(-1)}\circ\mathrm{h})}$$

Then, the change of variable $\mathrm{h}\to\mathrm{g}^{(-1)}\circ\mathrm{h}$ in the definition of the covariance matrix of the random transformation $\mathrm{g}\circ\mathbf{f}$ (eqn. 26) gives the result. ■

One could obtain similar results for the right composition and the inversion in the case of a left and right invariant distance. The proof for the right translation would be obtained by the right equivalent of the construction we have done with the left invariant distance. The proof for the inversion is more tricky, but the properties of the left and right invariant metric should be sufficient.

4 The case of connected homogeneous manifolds

A manifold on which acts a transformation group is said to be *transitive* or *homogeneous* with respect to this group if, for any two points of the manifold, there exists a transformation of the group that transform one point to the other. Let o be a particular point of the manifold that we call *the origin* and let us denote by $\mathrm{g} \star \mathrm{x}$ the action of a transformation g on a point x of the manifold. The manifold is homogeneous if $\mathcal{G} \star \mathrm{o} = \{\mathrm{g} \star \mathrm{o} \ / \ \mathrm{g} \in \mathcal{G}\} = \mathcal{M}$.

4.1 Invariant Riemannian metric

In the case of an homogeneous manifold, we can identify the elements $\mathrm{x} \in \mathcal{M}$ with subsets of of the group $\mathcal{G}$. Let $\mathcal{H}$ be the *isotropy subgroup* of the origin, i.e. the set of transformations that leave it unchanged:

$$\mathcal{H} = \{\mathrm{h} \in \mathcal{G} \ / \ \mathrm{h} \star \mathrm{o} = \mathrm{o}\}$$

Let now $\mathrm{f_x}$ be a transformation that bring the origin into the point $\mathrm{x} = \mathrm{f_x} \star \mathrm{o}$. The set of all transformation that bring the origin into the point x is the left translation of $\mathcal{H}$ by $\mathrm{f_x}$:

$$\mathcal{F}_\mathrm{x} = \{\mathrm{g} \in \mathcal{G} \ / \ \mathrm{g} \star \mathrm{o} = \mathrm{x}\} = \mathrm{f_x} \circ \mathcal{H}$$

These sets are in fact the elements of the quotient space $\mathcal{G}/\mathcal{H}$ are are thus called the *coset* $\mathcal{F}_\mathrm{x}$ of x.

To compare the tangent spaces $T_\mathrm{o}\mathcal{M}$ and $T_\mathrm{x}\mathcal{M}$ at the origine and at point x, we now have a whole set of transformations from $\mathcal{F}_\mathrm{x}$. Let us chose a *placement function* $\mathrm{f_x}$, i.e. a representant of each coset. In a local coordinate system, the vector $v \in T\mathcal{M}$ is transformed by the placement function into

$$v_x = J(\mathrm{f}_x).v \qquad \text{with} \qquad J(\mathrm{f}_x) = \left.\frac{\partial(\mathrm{f}_x \star e)}{\partial e}\right|_{e=o}$$

This transportation can by used to define the dot product from the tangent space at the origin on each tangent space $T_\mathrm{x}\mathcal{M}$:

$$\langle\, y_1 \mid y_2 \,\rangle_x = \left\langle\, J(f_x)^{(-1)}.y_1 \;\middle|\; J(f_x)^{(-1)}.y_2 \,\right\rangle \qquad \text{avec} \qquad J(f_x) = \left.\frac{\partial(f_x \star e)}{\partial e}\right|_{e=\mathrm{o}}$$

Thus, the expression of the metric in a local chart is

$$G(x) = J(\mathrm{f}_x)^{(-\mathrm{T})}.Q.J(\mathrm{f}_x)^{(-1)} \tag{42}$$

A necessary condition to obtain an invariant metric is obviously the stability of the metric with respect to the choice of the placement function. From the definition of the cosets, this reduces to the invariance of the metric Q at the origin by the action of the isotropy group $\mathcal{H}$:

$$\forall \mathrm{h} \in \mathcal{H} \qquad\qquad J(\mathrm{h})^{\mathrm{T}}.Q.J(\mathrm{h}) = Q \tag{43}$$

The dot product $\langle\, . \mid . \,\rangle_x$ is in this case a continuous function of x: we have obtained an invariant metric (the condition above is sufficient). Note that there does not always exists an invariant metric: think for instance to points under the affine group.

Assuming that there exists an invariant metric, the associated infinitesimal volume element is given by:

$$d\mathcal{M}(\mathrm{x}) = \sqrt{|G(x)|}.dx = \lambda.\frac{dx}{|J(\mathrm{f}_x)|} \qquad \text{with} \qquad \lambda = \sqrt{|Q|} \qquad \text{and} \qquad \mathrm{f}_x \in \mathcal{F}_x$$

This measure was the invariant measure determined in [Pennec and Ayache, 1998]. However, we have a stronger constraint on the existance: an invariant distance implies an invariant measure, but the converse is false.

4.2 Principal chart

In the sequel, we assume that there exists and invariant distance and that the manifold is geodesically complete for this metric. We call principal chart the exponential chart at the origin with $Q = \mathrm{Id}$. The action of a transformation is define in this chart by $\mathrm{f} \star \vec{\mathrm{x}} \equiv \log(\mathrm{f} \star \mathrm{x}) = \log(\mathrm{f} \star \exp(\vec{\mathrm{x}}))$.

Since geodesics are globally invariant by the action of transformation, we can easily express the exponential chart at any other point. If $\overrightarrow{\mathrm{xy}}$ is a tangent vector of $T_{\mathrm{x}}\mathcal{M}$ express in the basis induced by the principal chart, we have:

$$\exp_{\vec{\mathrm{x}}}(\overrightarrow{\mathrm{xy}}) = \mathrm{f}_{\vec{\mathrm{x}}} \star \exp(J(f_{\vec{\mathrm{x}}})^{(-1)}.\vec{\mathrm{y}})$$

The inverse function gives (in the star-shaped domain delimited by the cut-locus):

$$\overrightarrow{\mathrm{xy}} = \log_{\vec{\mathrm{x}}}(\mathrm{y}) = J(\mathrm{f}_{\vec{\mathrm{x}}}).(\mathrm{f}_{\vec{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{y}}) \tag{44}$$

The functions are independent of the chosen placement function. Indeed, let f'_{x} be another placement function. Then, by definition of the cosets, for each point x there exists a transformation $\mathrm{h}_{\mathrm{x}} \in \mathcal{H}$ such that $\mathrm{f}'_{\mathrm{x}} = \mathrm{f}_{\mathrm{x}} \circ \mathrm{h}_{\mathrm{x}}$. Now, the geodesics starting from the origin are strait lines in the exponential chart and are globally invariant: they are transformed into strait lines in the exponential chart by the isotropy group. This means that the action of the isotropy group is linear in the principal chart: we can write $\mathrm{h} \star \vec{\mathrm{x}} = J(\mathrm{h}).\vec{\mathrm{x}}$. Thus, $\mathrm{f}'^{(-1)}_{\vec{\mathrm{x}}} \star \vec{\mathrm{y}} = \mathrm{h}^{(-1)}_{\vec{\mathrm{x}}} \star (\mathrm{f}_{\mathrm{x}} \star \vec{\mathrm{y}}) = J(\mathrm{h}_{\vec{\mathrm{x}}})^{(-1)}.(\mathrm{f}_{\mathrm{x}} \star \vec{\mathrm{y}})$. Since $J(\mathrm{f}'_{\vec{\mathrm{x}}}) = J(\mathrm{f}_{\vec{\mathrm{x}}}).J(\mathrm{h}_{\vec{\mathrm{x}}})$, we get the sought equality $J(\mathrm{f}_{\vec{\mathrm{x}}}).(\mathrm{f}_{\vec{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{y}}) = J(\mathrm{f}'_{\vec{\mathrm{x}}}).(\mathrm{f}'^{(-1)}_{\vec{\mathrm{x}}} \star \vec{\mathrm{y}})$ The invariance of the exponential is similar.

The distance between two points x and y is by definition the length of the minimizing geodesic joining them, and by definition of the exponential chart the length of $\overrightarrow{\mathrm{xy}}$:

$$\mathrm{dist}(\mathrm{x},\mathrm{y})^2 = \| \log_{\vec{\mathrm{x}}}(\mathrm{y}) \|^2_{\vec{\mathrm{x}}} = \overrightarrow{\mathrm{xy}}^{\mathrm{T}}.Q(\vec{\mathrm{x}}).\overrightarrow{\mathrm{xy}} = (\vec{\mathrm{f}}_{\vec{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{y}})^{\mathrm{T}}.(\vec{\mathrm{f}}_{\vec{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{y}}) \tag{45}$$

4.3 Propagation of pdfs

Theorem 13 (Action of a deterministic transformation on a random feature)
Let $\mathbf{x}$ *be a random feature of pdf* $p_{\mathbf{x}}$ *and* $\mathrm{f} \in \mathcal{G}$ *a deterministic transformation. Then, the pdf of the random feature* $\mathrm{f} \star \mathbf{x}$ *is*

$$p_{(\mathrm{f}\star\mathbf{x})}(\mathrm{y}) = p_{\mathbf{x}}(\mathrm{f}^{(-1)} \star \mathrm{y}) \tag{46}$$

Proof: The probability of $\mathbf{y} = \mathrm{f} \star \mathbf{x}$ to be in a set $\mathcal{Y}$ is

$$P(\mathrm{f} \star \mathbf{x} \in \mathcal{Y}) = P(\mathbf{x} \in \mathrm{f}^{(-1)} \star \mathcal{Y}) = \int_{(\mathrm{f}^{(-1)} \star \mathcal{Y})} p_{\mathbf{x}}(\mathrm{z}).d\mathcal{M}(\mathrm{z}) = \int_{\mathcal{Y}} p_{\mathbf{x}}(\mathrm{f}^{(-1)} \star \mathrm{y}).d\mathcal{M}(\mathrm{y})$$

The last equality is obtained using the change of variable $\mathrm{z} = \mathrm{f}^{(-1)} \star \mathrm{y}$ and the invariance of the measure. ■

Theorem 14 (Action of a random transformation on the origin)
Let $\mathbf{f}$ *be a random transformation of pdf* $p_{\mathbf{f}}$*. Its action on the origin* o *determines a random feature* $\mathbf{y} = \mathbf{f} \star \mathrm{o}$ *that has the pdf*

$$p_{(\mathbf{f}\star\mathrm{o})}(\mathrm{y}) = \int_{\mathcal{H}} p_{\mathbf{f}}(\mathrm{f}_{\mathrm{y}} \circ \mathrm{h}).d\mathcal{H}(\mathrm{h}) \tag{47}$$

Proof: The probability of $\mathbf{x} = \mathbf{f} \star \mathrm{o}$ to be in a set $\mathcal{X} \in \mathcal{M}$ is:

$$\Pr(\mathbf{f} \star \mathrm{o} \in \mathcal{X}) = \int_{\mathcal{F}_\mathcal{X}} p_\mathbf{f}(\mathrm{g}).d\mathcal{G}(\mathrm{g}) \qquad \text{avec} \qquad \mathcal{F}_\mathcal{X} = \{\mathrm{g} \in \mathcal{G} \ / \ \mathrm{g} \star \mathrm{o} \in \mathcal{X}\}$$

Using a placement function, the set of transformations $\mathcal{F}_\mathcal{X}$ is $\mathcal{F}_\mathcal{X} = \{\mathrm{f_x} \circ \mathrm{h} \ / \ \mathrm{x} \in \mathcal{X} \ , \ \mathrm{h} \in \mathcal{H}\}$. Thus:

$$\Pr(\mathbf{x} \in \mathcal{X}) = \int_{\mathcal{X},\mathcal{H}} p_\mathbf{f}(\mathrm{f_x} \circ \mathrm{h}).d\mathcal{M}(\mathrm{x}).d\mathcal{H}(\mathrm{h}) = \int_\mathcal{X} \left(\int_\mathcal{H} p_\mathbf{f}(\mathrm{f_x} \circ \mathrm{h}).d\mathcal{H}(\mathrm{h}) \right) .d\mathcal{M}(\mathrm{x})$$

■

Theorem 15 (Action of a random transformation of a deterministic feature)
Let $\mathbf{f}$ *be a random transformation of pdf* $p_\mathbf{f}$*. Its action on the deterministic feature* x *determines a random feature* $\mathbf{y} = \mathbf{f} \star \mathrm{x}$ *that has the pdf:*

$$p_{(\mathbf{f} \star x)}(y) = \int_\mathcal{H} p_{(\mathbf{f} \circ f_x)}(f_y \circ h).d\mathcal{H}(h) = \int_\mathcal{H} \frac{\Delta\mathcal{G}(f_y \circ h \circ f_x^{(-1)})}{\Delta\mathcal{G}(f_y \circ h)} . p_\mathbf{f}(f_y \circ h \circ f_x^{(-1)}).d\mathcal{H}(h) \tag{48}$$

Proof: Choosing $\mathrm{f_x} \in \mathcal{F}_\mathrm{x}$ and writing $\mathbf{y} = \mathbf{f} \star x = (\mathbf{f} \circ \mathrm{f_x}) \star \mathrm{o}$, we apply equation 35 to obtain the pdf of $\mathbf{f} \circ \mathrm{f_x}$ and then theorem 47. ■

Theorem 16 (Action of a random transformation on a random feature)
Let $\mathbf{f}$ *be a random transformation of pdf* $p_\mathbf{f}$ *and* $\mathbf{x}$ *be a random feature of pdf* $p_\mathbf{x}$*. The pdf of the random feature* $\mathbf{y} = \mathbf{f} \star \mathbf{x}$ *is:*

$$p_{\mathbf{f} \star \mathbf{x}}(\mathrm{y}) = \int_\mathcal{G} p_\mathbf{f}(\mathrm{g}).p_\mathbf{x}(\mathrm{g}^{(-1)} \star \mathrm{y}).d_L\mathcal{G}(\mathrm{g}) \tag{49}$$

Once again, this is very similar to a convolution product.

Proof: Let $\mathcal{X} \subset \mathcal{M}$ be a set of features. The set $\mathcal{A}$ of couples (f, x) such that $\mathrm{f} \star \mathrm{x} \in \mathcal{X}$ can be writen:

$$\mathcal{A} = \left\{(\mathrm{f}, \mathrm{x}) \ / \ \mathrm{f} \in \mathcal{G} \ , \ \mathrm{x} \in \mathrm{f}^{(-1)} \star \mathcal{X}\right\}$$

Thus, the probability of $\mathbf{y} = \mathrm{f} \star \mathrm{x}$ to be in $\mathcal{X}$ is

$$\begin{aligned}
\Pr(\mathbf{f} \star \mathbf{x} \in \mathcal{X}) &= \int_\mathcal{G} \left(\int_{\mathrm{g}^{(-1)} \star \mathcal{X}} p_\mathbf{x}(\mathrm{y}).d\mathcal{M}(\mathrm{y}) \right) .p_\mathbf{f}(\mathrm{g}).d_L\mathcal{G}(\mathrm{g}) \\
&= \int_\mathcal{G} \left(\int_\mathcal{X} p_\mathbf{x}(\mathrm{g}^{(-1)} \star \mathrm{z}).d\mathcal{M}(\mathrm{z}) \right) p_\mathbf{f}(\mathrm{g}).d_L\mathcal{G}(\mathrm{g}) \\
&= \int_\mathcal{X} \left(\int_\mathcal{G} p_\mathbf{x}(\mathrm{g}^{(-1)} \star \mathrm{z}).p_\mathbf{f}(\mathrm{g}).d_L\mathcal{G}(\mathrm{g}) \right) .d\mathcal{M}(\mathrm{z})
\end{aligned}$$

■

4.4 Obtaining the Karcher mean values

As for the Lie groups, we have few simplifications due to the left invariant structure in the Karcher means characterization: thanks to formula (44), we can write:

$$\mathbf{E}\left[\ \overrightarrow{\mathrm{y}\mathbf{x}}\ \right] = J(\mathrm{f}_{\vec{\mathrm{y}}}). \int_{\mathcal{M}/C(\mathrm{y})} (\mathrm{f}_{\vec{\mathrm{y}}}^{(-1)} \star \vec{\mathrm{z}}).dP(\vec{\mathrm{z}})$$

Thus, assuming that the conditions of theorem (2) (i.e. that the variance is finite and the cut locus has a null measure), our gradient descent algorithm to obtain the mean can be entirely written in the principal chart (we focus here on the practical case: the empirical mean value):

$$\bar{\mathrm{x}}_{t+1} = f_{\bar{\mathrm{x}}_t} \star \left(\frac{1}{n} \sum_i \left(f^{(-1)}_{\bar{\mathrm{x}}_t} \star \vec{\mathrm{x}}_i \right) \right) \tag{50}$$

which simply corresponds to the following algorithm: translate all measured values to the origin using the inverse of the placement function at the current estimation; compute the barycenter of the points in the principal chart; use that barycenter as the increment for the current value.

4.5 Stability of the Fréchet expectation and Normal law

Let us now turn to the stability of the Fréchet expectation with respect to the action of transformations.

Theorem 17 (Action of a deterministic transformation on a random feature)

$$\mathbb{E}[\, \mathrm{g} \star \mathbf{x} \,] = \mathrm{g} \star \mathbb{E}[\, \mathbf{x} \,] \qquad et \qquad \mathbb{E}[\, \{\mathrm{g} \star \mathrm{x}_i\} \,] = \mathrm{g} \star \mathbb{E}[\, \{x_i\} \,] \tag{51}$$

This resul also holds for the set of central features of any order.

Proof: Let $\mathbf{z} = \mathrm{g} \star \mathbf{x}$ be the random feature obtained by the action of the deterministic transformation g on the random feature $\mathbf{x}$. We have:

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

Thanks to the invariance of the distance, we get:

$$\sigma^2_{\mathbf{z}}(\mathrm{y}) = \int_{\mathcal{M}} \mathrm{dist}(\mathrm{g}^{(-1)} \star \mathrm{y}, \mathrm{x})^2 . p_{\mathbf{x}}(\mathrm{x}) . d\mathcal{M}(\mathrm{x}) = \sigma^2_{\mathbf{x}}(\mathrm{g}^{(-1)} \star \mathrm{y})$$

Thus, the feature $\bar{\mathrm{x}}$ minimizes $\sigma^2_{\mathbf{x}}(\mathrm{x})$ if and only if $\bar{\mathrm{z}} = \mathrm{g} \star \bar{\mathrm{x}}$ minimizes $\sigma^2_{\mathbf{z}}(\mathrm{z})$, which may be rewritten $\mathbb{E}[\, \mathbf{z} \,] = \mathrm{g} \star \mathbb{E}[\, \mathbf{x} \,]$. Moreover, variances are equal: $\sigma_{\mathbf{z}} = \sigma_{\mathbf{x}}$.
The same demonstration holds for the stability of the central feature of any order, as well as for the empirical mean feature. ■

However, if the transformation turns out to be probabilistic, then we have no simple expression for $\mathbb{E}[\, \mathbf{f} \star \mathbf{x} \,]$.

Theorem 18 (Action of a deterministic transformation on a normal random feature) *Let* $\mathbf{x}$ *be a random Normal feature of law* $N_{(\bar{\mathrm{x}}, \Gamma_{\mathbf{x}})}$ *(in the conditions of theorem 3), and* f *be a deterministic transformation. Then,* $\mathrm{f} \star \mathbf{x}$ *is a random Normal feature following the Normal law* $N_{(\mathrm{f} \star \bar{\mathrm{x}}, \Gamma_{\mathrm{f} \star \mathbf{x}})}$ *with*

$$\Gamma_{\mathrm{f} \star \mathbf{x}} = J^{(-T)} . \Gamma_{\mathbf{x}} . J^{(-1)} \qquad where \qquad J = \left. \frac{\partial (\mathrm{f} \star \vec{\mathrm{x}})}{\partial \vec{\mathrm{x}}} \right|_{\vec{\mathrm{x}} = \bar{\mathrm{x}}}$$

Moreover, the covariances matrices are related by $\Sigma_{\mathrm{f} \star \mathbf{x}} = J . \Sigma_{\mathbf{x}} . J^T$.

The demonstration of this theorem is essentially the same as the one for Lie groups (Section 3.6).

5 Discussion

On a (geodesically complete) Riemannian manifold, it is easy to define probability density functions associated to random features, 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 feature). Thus, the definition of a mean value for a random feature is much more complex than for the vectorial case and it requires a distance-based variational formulation: the Fréchet or Karcher expected features basically minimize globally (or locally) the variance. As the mean is now defined through a minimization procedure, its existence and uniqueness are not ensured any more (except under some specific conditions). 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 Karcher values of order 0) in the vectorial case. 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 of the mean value, and a re-centering step of the chart at the point of the manifold that corresponds to the 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 feature 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 feature that basically weights the distance between the deterministic feature and the mean feature using the inverse of the covariance matrix. Interestingly, the expected Mahalanobis distance of a random primitive 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 chose a variational approach to generalize the Normal law: we define it as the distribution that minimizes the information knowing the mean and the covariance. Neglecting the cut-locus constraints, we show that is 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 feature, 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 work, all definitions are derived from the Riemannian metric of the manifold. More generally, we could conceive other definitions of the mean value using the notion of connector introduced in [Picard, 1994]. 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.

When we deal with geometric problems, we often have transformation groups (Lie groups in ou case) and manifolds that are subject to the action of such a transformation group. The question that arise is how to chose an adapted Riemannian metric on these structures? In the case of a Lie group, we have two canonical Riemannian structures given by the left and right invariant metrics. Choosing one of these metrics lead to important simplifications in the propagation of the pdf of random transformations. One may notice the remarkable similarity between the pdf of composition of two random transformations and the convolution product used to compute the pdf of the sum of two random vectors. However, the propagation formulas are not symmetric, and a Riemannian metric that is only left invariant only guarantees the stability of the mean, covariance (and Normal parameters) under the action of a deterministic transformation on the left. The stability by the action on both side and by inversion is only guaranteed if we have a left and right invariant metric. Unfortunately, such a bi-invariant metric does not exists in general for non-compact Lie groups. Is there another way to find a metric that leads to symmetric properties for non-compact Lie groups (or simply for locally compact such as rigid transformation)?

The second kind of geometrical objects we investigated is homogeneous manifolds, i.e. manifolds that have no invariant w.r.t. the action of a transformation group. The basic idea is to ensure the stability of our statistical operations w.r.t. the group action by choosing an appropriate (i.e. invariant) Riemannian metric. We show that a necessary condition is the invariance of the metric at any given point by the isotropy group of this point. Such a metric does not always exist (e.g. for points under the action of the linear or affine group). If it exists, then we can write quite simply the propagation of the pdf of a random feature subject to the action of a random or deterministic transformation, and show that the mean, covariance and Normal parameters are stable under the action of a deterministic transformation.

A third kind of important geometrical objects we have not investigated yet are the so-called "shape spaces", i.e. the quotient of a manifold by a transformation group. The difficult part here is that we have to remove the assumption of geodesical completeness and deal with manifolds with boundaries. Indeed, the shape manifolds often present some singularities [Small, 1996]: for instance, the shape of pairs of points (subject to rigid transformations) is the distance between them, which is positive but can reach the singular point zero. On a geometric aspect, the question arise whether we can find compatible metrics on the feature, transformation and shape spaces, i.e. such that the metric on the feature space be the direct product of the metrics on the shape and transformation spaces.

In conclusion, we believe that we could derive most of the interesting statistical operations on a finite-dimensional manifold from a Riemannian metric (or perhaps only a connector following [Picard, 1994]), even if there the theory still has to be generalized to manifolds with boundaries. In the case of geometric objects (Lie groups or homogeneous manifolds), requiring the invariance of the metric ensures the stability of our operations w.r.t. the structural operations (composition or action, inversion), but such a metric does not always exists. Would there be some weaker conditions than invariance that could still guaranty the stability of our operations? At that point, it seems to us that the choice of the Riemannian metric is the key point to investigate. Another interesting but difficult research direction would concern the generalization of our statistical framework to infinite-dimensional groups and transformations, such as manifolds of curves or surfaces, and groups of diffeomorphisms. For this purpose, the work of Younes at al [Younes, 1998, Miller and Younes, 2001], may provide a thorough basis.

References

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

[Arnaudon, 1995] Arnaudon, M. (1995). Barycentres convexes et approximations des martingales continues dans les variétés. In J. Azema, P.A. Meyer, M. Y., editor, *Séminaire de probabilités XXIX*, volume 1613 of *Lect. Notes in Math.*, pages 70–85. Springer-Verlag.

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

[Carmo, 1992] Carmo, M. d. (1992). *Riemannian Geometry*. Mathematics. Birkhäuser, Boston, Basel, Berlin.

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

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

[Emery and Mokobodzki, 1991] Emery, M. and Mokobodzki, G. (1991). Sur le barycentre d'une probabilité dans une variété. In J. Azema, P.A. Meyer, M. Y., editor, *Séminaire de probabilités XXV*, volume 1485 of *Lect. Notes in Math.*, pages 220–233. Springer-Verlag.

[Fréchet, 1944] Fréchet, M. (1944). 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.

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

[Herer, 1986] Herer, W. (1986). 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.

[Herer, 1988] Herer, W. (1988). 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.

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

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

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

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

[Kendall, 1990] Kendall, W. (1990). Probability, convexity, and harmonic maps with small image i: uniqueness and fine existence. *Proc. London Math. Soc.*, 61(2):371–406.

[Kent, 1992] Kent, J. (1992). *The art of Statistical Science*, chapter 10 : New Directions in Shape Analysis, pages 115–127. John Wiley & Sons. K.V. Mardia, ed.

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

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

[Mardia, 1995] Mardia, K. (1995). Directional statistics and shape analysis. Research Report STAT95/24, University of Leeds, UK.

[Miller and Younes, 2001] Miller, M. and Younes, L. (2001). Group actions, homeomorphisms, and matching: A general framework. *International Journal of Computer Vision*, 41(1/2):61–84.

[Neveu, 1990] Neveu, J. (1990). *Introduction aux probabilités.* Ecole Polytechnique. Cours de l'Ecole Polytechnique.

[Papoulis, 1991] Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill, Inc.

[Pelat, 1992] Pelat, D. (1992). *Bruits et Signaux.* Cours de l'Ecole doctorale d'astrophysique d'Ile de France.

[Pennec, 1996] Pennec, X. (1996). *L'Incertitude dans les Problèmes de Reconnaissance et de Recalage – Applications en Imagerie Médicale et Biologie Moléculaire.* PhD thesis, Ecole Polytechnique, Palaiseau (France).

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

[Pennec and Thirion, 1997] Pennec, X. and Thirion, J.-P. (1997). A framework for uncertainty and validation of 3D registration methods based on points and frames. *Int. Journal of Computer Vision*, 25(3):203–229.

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

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

[Press et al., 1991] Press, W., Flannery, B., Teukolsky, S., and Vetterling, W. (1991). *Numerical Recipices in C.* Cambridge Univ. Press.

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

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

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

[Younes, 1998] Younes, L. (1998). Computable elastic distances between shapes. *SIAM Journal on Applied Mathematics*, 58(2):565–586.

A Gradient of the variance

This proof is a generalization of a differentiability proof for the uniform distribution on compact manifolds by Pr Maillot [Maillot, 1997]. One of the main difficulty was to remove the compactness hypothesis.

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:

$$P(C(\mathrm{y})) = \int_{C(\mathrm{y})} dP(\mathrm{z}) = 0 \qquad \text{and} \qquad \sigma^2(\mathrm{y}) = \int_{\mathcal{M}} \mathrm{dist}(\mathrm{y},\mathrm{z})^2 . dP(\mathrm{z}) < \infty \tag{52}$$

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:

$$\forall v \in T_{\mathrm{y}}\mathcal{M},\ \|v\| < \varepsilon \quad \left|\sigma^2(\exp_{\mathrm{y}}(v)) - \sigma^2(\mathrm{y}) - \left\langle (\mathrm{grad}\, \sigma^2)(\mathrm{y}) \mid v \right\rangle\right| \leq \eta.\|v\| \tag{53}$$

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

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

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

Lemma 1 A basis of neighbourhoods of $C(\mathrm{y})$
$W_\varepsilon = \bigcup_{\mathrm{x} \in \mathcal{B}(\mathrm{y}, \varepsilon)} C(\mathrm{y})$ *is a continuous series of included and decreasing open sets all containing* $C(\mathrm{y})$ *and converging toward 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 $\underline{U} = \rho^{(-1)}(+\infty)$ is open. Let $\pi : U \to \mathcal{M}$ be the canonical projection along the fiber $\pi((\mathrm{x}, v)) = \mathrm{x}$ (it is obviously continuous) and 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 cuting 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, this set is open.

Now, we are left the the proof that $\pi(U_\varepsilon) = W_\varepsilon$ is open. Assume that W_ε is not open. Then, $\overline{W_\varepsilon} = \mathcal{M} \backslash W_\varepsilon$ is not closed: there exists a series of points $\mathrm{z}_i \in \overline{W_\varepsilon}$ converging toward $\mathrm{z} \in W_\varepsilon$. By definition of this set, there exists $u = (\mathrm{z}, v) \in U_\varepsilon$ such that $\pi(u) = \mathrm{z}$. Conversely, there is not unit tangent vector at $z_i \notin W_\varepsilon$ with that property: $U_{\mathrm{z}_i} \cap U_\varepsilon = \varnothing$. Consider the parallel transportation v_i of v from z to z_i. The series $u_i = (\mathrm{z}_i, v) \in \overline{U_\varepsilon}$ is obviously converging toward $u = (\mathrm{z}, v) \in U_\varepsilon$. Since U_ε is open, $\overline{U_\varepsilon}$ is closed and the series converges within this set, which is in contradiction with the previous statement.

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

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

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

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: $\int_{C(\mathrm{y})} \mathrm{dist}(\mathrm{y},\mathrm{z}).dP(\mathrm{z}) = 0$. Thus, there exists an open set $W \in \mathcal{M}$ containing $C(\mathrm{y})$ and having an arbitrarily small measures for the identity and the distance function (say less than α). Since W_ε is a basis of neighborhoods of $C(\mathrm{y})$, there exists ε sufficiently small such that $W_\varepsilon \in W$.

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 belong 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. This allows us to simplify the notations. The density is

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

The covariance and concentration are related by:

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

Since the concentration matrix Γ is symmetric and positive definite, we can diagonalize it in the form $\Gamma = A^{\mathrm{T}}.\Lambda^2.A$ with A beeing ortogonal and Λ beeing the diagonal matrix of positive square roots of eigenvalues. With the change of variable $z = \Lambda.A.y$, we can rewrite the above integrals as:

$$k^{(-1)} = \det(\Lambda)^{(-1)}.\int_{\mathcal{D}'} \exp\left(-\frac{\|z\|^2}{2}\right).\sqrt{\det(G(A^{\mathrm{T}}.\Lambda^{(-1)}.z))}.dz$$

$$J = k^{(-1)}.\det(\Lambda).\Lambda.A.\Sigma.A^{\mathrm{T}}.\Lambda = \int_{\mathcal{D}'} z.z^{\mathrm{T}}.\exp\left(-\frac{\|z\|^2}{2}\right).\sqrt{\det(G(A^{\mathrm{T}}.\Lambda^{(-1)}.z))}.dz$$

where the new definition domain is $\mathcal{D}' = \Lambda.A.\mathcal{D}$.

From [Chavel, 1993, section 2.8, corollary 2.3, p.84], we know the Taylor expansion of the metric around the origin in a normal coordinate system:

$$\det(g_{ij}(\exp v)) = 1 - \mathrm{Ric}\,(v,v)/3 + O(\|v\|^3)$$

Thus:

$$d\mathcal{M}(y) = \sqrt{\det(G(y))} = 1 - \frac{y^{\mathrm{T}}.\mathrm{Ric}\,.y}{6} + O(\|y\|^3) \tag{54}$$

where Ric is the expression of the Ricci tensor of scalar curvatures in the exponential chart. Let $T = \Lambda^{(-1)}.A.\mathrm{Ric}\,.A^{\mathrm{T}}.\Lambda^{(-1)}$. Since $\|A^{\mathrm{T}}.\Lambda^{(-1)}.z\|^2 = \|\Lambda^{(-1)}.z\|^2 = \sum_i z_i^2/\lambda_i^2 \leq \|z\|^2/\lambda_{min}^2$, we obtain:

$$\sqrt{\det(G(A^{\mathrm{T}}.\Lambda^{(-1)}.z))} = 1 - \frac{z^{\mathrm{T}}.T.z}{6} + O\left(\frac{\|z\|^3}{\lambda_{min}^3}\right)$$

B.1 Manifolds of non positive curvature at the mean point

At such a point, the cut locus is void and the definition domain of the exponential chart is $\mathcal{D} = \mathbb{R}^n$. This greatly simplifies the integrals. We have:

$$\det(\Lambda).k^{(-1)} = \int_{\mathbb{R}^n} \exp\left(-\frac{\|z\|^2}{2}\right).\left\{1 - \frac{1}{6}z^{\mathrm{T}}.T.z + O\left(\frac{\|z\|^3}{\lambda_{min}^3}\right)\right\}.dz$$

The first term is

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

The third term in simple: as the integral converges to a constant, we obtain:

$$\int_{\mathbb{R}^n} O\left(\frac{\|z\|^3}{\lambda_{min}^3}\right).\exp\left(-\frac{\|z\|^2}{2}\right).dz = O\left(\lambda_{min}^{-3}\right)$$

The second term is

$$\begin{aligned}\int_{\mathbb{R}^n} z^{\mathrm{T}}.T.z.\exp\left(-\tfrac{\|z\|^2}{2}\right).dz &= \textstyle\sum_{i,j} T_{ij}\int z_i.z_j\exp\left(-\tfrac{\|z\|^2}{2}\right).dz\\ &= \textstyle\sum_i T_{ii}\left(\int z_i^2.\exp\left(-\tfrac{z_i^2}{2}\right).dz_i\right).\left(\prod_{j\neq i}\int\exp\left(-\tfrac{z_j^2}{2}\right).dz_j\right)\\ &= \textstyle\sum_i T_{ii}.\left(\sqrt{2\pi}\right).\left((2.\pi)^{(n-1)/2}\right)\\ &= (2\pi)^{n/2}.\mathrm{Tr}(T)\\ &= (2\pi)^{n/2}.\mathrm{Tr}(\Gamma^{(-1)}.\mathrm{Ric}\,)\end{aligned}$$

since $\mathrm{Tr}(A.B) = \mathrm{Tr}(B.A)$. Thus we obtain:

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

As $\mathrm{Tr}(\Gamma^{(-1)}.\mathrm{Ric}\,)$ is a term in λ_{min}^{-2}, we have:

$$k = \frac{\sqrt{\det(\Gamma)}}{(2\pi)^{n/2}}\left(1 + \frac{\mathrm{Tr}(\Gamma^{(-1)}.\mathrm{Ric}\,)}{6} + O(\lambda_{min}^{-3})\right) \tag{55}$$

Now, for the covariance matrix, we have:

$$J = \int_{\mathbb{R}^n} z.z^{\mathrm{T}}.\exp\left(-\frac{\|z\|^2}{2}\right).\left(1 - \frac{1}{6}\mathrm{Tr}(T.z.z^{\mathrm{T}}) + O\left(\frac{\|z\|^3}{\lambda_{min}^3}\right)\right).dz$$

We consider first the off diagonal elements J_{ij} (with $i \neq j$): the first term of the integral vanished because we integrate antisymmetric functions over $\mathbb{R}$: $\int z_i.z_j.\exp(\|z\|^2/2).dz = 0$. The integral converges to a constant for the third term and the result is thus $O(\lambda_{min}^{-3})$. Since $\mathrm{Tr}(T.z.z^{\mathrm{T}}) = \sum_{k,l} T_{kl}.z_k.z_l$, we are left with:

$$J_{ij} = -\frac{1}{6}\sum_{k,l} T_{kl}.\int z_k.z_l.z_i.z_j.\exp\left(-\frac{\|z\|^2}{2}\right).dz$$

For $i \neq k$ and $j \neq l$ or $i \neq l$ and $j \neq k$, 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:

$$J_{ij} = -\frac{1}{6}(T_{ij} + T_{ji}).\int z_i^2.z_j^2.\exp\left(-\frac{\|z\|^2}{2}\right).dz = -\frac{T_{ij}}{3}(2\pi)^{n/2}$$

Now, we consider the diagonal elements J_{ii}. The first term is

$$\begin{aligned}\int_{\mathbb{R}^n} z_i^2.\exp(-\|z\|^2/2).dz &= \left(\textstyle\int_{\mathbb{R}} z_i^2.\exp(-z_i^2/2).dz_i\right).\prod_{j\neq i}\left(\int_{\mathbb{R}}\exp(-z_j^2/2).dz_j\right)\\ &= (2\pi)^{1/2}.(2\pi)^{(n-1)/2} = (2\pi)^{n/2}\end{aligned}$$

The second term is

$$\alpha_i = -\frac{1}{6}\int_{\mathbb{R}^n} z_i^2.z^{\mathrm{T}}.T.z.\exp(-\|z\|^2/2).dz = -\frac{1}{6}\sum_{k,l} T_{kl}\int_{\mathbb{R}^n} z_i^2.z_k.z_l.\exp(-\|z\|^2/2).dz$$

For $k \neq l$, we integrate antisymmetric functions: the result is zero. Thus, we are left with:

$$\begin{aligned}
\alpha_i &= -\tfrac{1}{6}\textstyle\sum_{k\neq i} T_{kk}\int_{\mathbb{R}^n} z_i^2.z_k^2.\exp(-\|z\|^2/2).dz - \tfrac{1}{6}T_{ii}\int_{\mathbb{R}^n} z_i^4.\exp(-\|z\|^2/2).dz \\
&= -\tfrac{1}{6}\textstyle\sum_{k\neq i} T_{kk}\left(\int_{\mathbb{R}} z_i^2.\exp(-z_i^2/2).dz_i\right).\left(\int_{\mathbb{R}} z_k^2.\exp(-z_k^2/2).dz_k\right).\left(\int_{\mathbb{R}}\exp(-z_j^2/2).dz_j\right)^{(n-2)} \\
&\quad -\tfrac{1}{6}T_{ii}\left(\int_{\mathbb{R}} z_i^4.\exp(-z_i^2/2).dz_i\right).\left(\int_{\mathbb{R}}\exp(-z_j^2/2).dz_j\right)^{(n-1)} \\
&= -\tfrac{1}{6}\textstyle\sum_{k\neq i} T_{kk}.(2\pi)^{n/2} - \tfrac{1}{6}T_{ii}.3.(2\pi)^{n/2} \\
&= -\tfrac{1}{6}\left(\mathrm{Tr}(T) + 2.T_{ii}\right) = -\tfrac{1}{6}\left(\mathrm{Tr}(\Gamma^{(\text{-}1)}.\mathrm{Ric}\,) + 2.T_{ii}\right)
\end{aligned}$$

For the third term, the integral converges to a constant and the result is $O(\lambda_{min}^{-3})$. Then, the diagonal element are:

$$J_{ii} = (2\pi)^{n/2}.\left(1 - \frac{\mathrm{Tr}(\Gamma^{(\text{-}1)}.\mathrm{Ric}\,)}{6} - \frac{T_{ii}}{3} + O(\lambda_{min}^{-3})\right)$$

Combining with off diagonal terms, we get:

$$J = (2\pi)^{n/2}.\left\{\mathrm{Id}.\left(1 - \frac{\mathrm{Tr}(\Gamma^{(\text{-}1)}.\mathrm{Ric}\,)}{6}\right) - \frac{T}{3} + O(\lambda_{min}^{-3})\right\}$$

As $J = k^{(\text{-}1)}.\det(\Lambda).\Lambda.A.\Sigma.A^{\mathrm{T}}.\Lambda$, we have (using equation 55 for the expansion of k):

$$\begin{aligned}
\Lambda.A.\Sigma.A^{\mathrm{T}}.\Lambda &= (k/\det(\Lambda)).J \\
&= \left\{1 + \tfrac{1}{6}\mathrm{Tr}(\Gamma^{(\text{-}1)}.\mathrm{Ric}\,) + O(\lambda_{min}^3)\right\}.\left\{\mathrm{Id}.\left(1 - \tfrac{1}{6}\mathrm{Tr}(\Gamma^{(\text{-}1)}.\mathrm{Ric}\,)\right) - \tfrac{1}{3}T + O(\lambda_{min}^{-3})\right\} \\
&= \mathrm{Id} - \tfrac{1}{3}T + O(\lambda_{min}^{-3})
\end{aligned}$$

We obtain thus the following relation between the covariance and the concentration matrices:

$$\Sigma = \Gamma^{(\text{-}1)} - \frac{1}{3}\Gamma^{(\text{-}1)}.\mathrm{Ric}\,.\Gamma^{(\text{-}1)} + O(\lambda_{min}^{-5}) \tag{56}$$

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, $O(\lambda_{min}^{-5}) = O(\sigma_{max}^5)$ 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 $\sigma_{max}^2 < \sum_i \sigma_i^2 = \sigma^2$, we have $O(\sigma_{max}) = O(\sigma)$ and the Taylor expansion of $\Gamma^{(\text{-}1)}$ is:

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

Then, we have

$$\Gamma.\Sigma = \mathrm{Id} - \frac{1}{3}\mathrm{Ric}\,.\Gamma^{(\text{-}1)} + O(\lambda_{min}^{-3}) = \mathrm{Id} - \frac{1}{3}\mathrm{Ric}\,.\Sigma + O(\sigma^3)$$

This means that:

$$\Gamma = \Sigma^{(\text{-}1)} - \frac{1}{3}\mathrm{Ric}\, + O(\sigma) \tag{57}$$

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

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

For the determinant, we observe that

$$\frac{\partial \det(A(\alpha))}{\partial \alpha} = \det(A(\alpha)).\mathrm{Tr}\left(\frac{\partial A(\alpha)}{\partial \alpha}.A(\alpha)^{(-1)}\right)$$

Thus, we have the following Taylor expansion (with $A(\alpha) = \mathrm{Id} + \alpha.B$):

$$\begin{aligned}\det(\mathrm{Id} + \alpha.B) &= \det(\mathrm{Id}) + \alpha.\left.\frac{\partial \det(\mathrm{Id}+\alpha.B)}{\partial \alpha}\right|_{\alpha=0} + O(\alpha^2)\\ &= 1 + \alpha.\det(\mathrm{Id}).\mathrm{Tr}\,(B.(\mathrm{Id} + B.0)^{(-1)}) + O(\alpha^2)\\ &= 1 + \mathrm{Tr}(\alpha.B) + O(\alpha^2)\end{aligned}$$

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

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

and thus

$$\sqrt{\det(\Gamma)} = \det(\Sigma)^{-1/2}.\left(1 - \frac{1}{6}.\mathrm{Tr}(\Sigma.\mathrm{Ric}\,) + O(\sigma^3)\right)$$

Reporting this expression in equation 55, 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:

$$k = \frac{1 + O(\sigma^3)}{\sqrt{(2\pi)^n.\det(\Sigma)}} \tag{58}$$

Summary of the approximate normal density: In a manifold of non positive curvature, 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. We are going to bound each term of all the integrals 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 injective 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 $z = \Lambda.A.y$, the ball is transformed into the set $\mathcal{B}_z = \{z / \|\Lambda^{(-1)}.z\| < r\}$. Since $\|\Lambda^{(-1)}\|^2 \leq \lambda_{min}^{-2}.\|z\|^2$, we can use the smaller ball $\mathcal{B} = \mathcal{B}(\lambda_{min}.r)$ for the integrations in polar coordinates. For the integrals that are separable along axes, it will be simpler to use the maximal cube $\mathcal{C} = \{z / -\lambda_{min}.r/\sqrt{n} < z_i < \lambda_{min}.r/\sqrt{n}\}$ included in $\mathcal{B}$.

In the following, we bound the different integrals used in the last section and summarized 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 $\mathcal{C}$:

$$\begin{aligned}\int_{\mathcal{D}'} \exp\left(-\frac{\|z\|^2}{2}\right).dz &\geq \prod_i \left(\int_{-\lambda_{min}.r/\sqrt{n}}^{\lambda_{min}.r/\sqrt{n}} \exp\left(-\frac{z_i^2}{2}\right).dz_i\right) \\ &\geq (2\pi)^{n/2}.\left(1 - \operatorname{erfc}\left(\frac{\lambda_{min}.r}{\sqrt{2n}}\right)\right)^n\end{aligned}$$

Here, $\operatorname{erfc}(x) = 1 - \frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2).dt$ 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:

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

We can summarize the value of the first integral by:

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

Now, for the following integral, we have

$$\begin{aligned}\int_{-\lambda_{min}.r/\sqrt{n}}^{\lambda_{min}.r/\sqrt{n}} z_i^2.\exp\left(-\frac{z_i^2}{2}\right).dz_i &= \sqrt{2\pi} - \sqrt{2\pi}.\operatorname{erfc}\left(\frac{\lambda_{min}.r}{\sqrt{2n}}\right) - \frac{2\sqrt{n}}{\lambda_{min}.r}\exp\left(-\frac{\lambda_{min}^2.r^2}{2}\right) \\ &= \sqrt{2\pi} + \varepsilon(\lambda_{min}^{-1}.r^{-1})\end{aligned}$$

We obtain thus

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

In fact, it is not hard to show 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 $\varepsilon(\lambda_{min}^{-1}.r^{-1})$. Thus, we can directly generalize the previous results by replacing $O(\sigma^k)$ with $O(\sigma^k) + \varepsilon\left(\frac{\sigma}{r}\right)$.

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 $\sigma^2 = \mathrm{Tr}(\Sigma)$:

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

C Approximated generalized χ^2 law

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

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

Let $\mathcal{B}_\Sigma(\alpha)$ be the "elliptic ball" of covariance Σ and radius α: $\mathcal{B}_\Sigma(\alpha) = \{y \;/\; y^{\mathrm{T}}.\Sigma^{(\text{-}1)}.y \leq \alpha^2\}$. Assuming that there is not cut locus, this probability can be written in a normal coordinate system:

$$\Pr\{\chi^2 \leq \alpha^2\} = \int_{\chi^2 \leq \alpha^2} N(\mathrm{y}).d\mathcal{M}y = \int_{\mathcal{B}_\Sigma(\alpha)} k.\exp(-y^{\mathrm{T}}.\Gamma.y/2).\sqrt{G(y)}.dy$$

Since $\Gamma = \Sigma^{(\text{-}1)} - \frac{1}{3}\mathrm{Ric} + O(\sigma)$ and $\sqrt{\det(G(y))} = 1 - \frac{y^{\mathrm{T}}.\mathrm{Ric}.y}{6} + O(\|y\|^3)$ we have:

$$\Pr\{\chi^2 \leq \alpha^2\} = k\int_{\mathcal{B}_\Sigma(\alpha)} \exp\left(-\frac{y^{\mathrm{T}}\left(\Sigma^{(\text{-}1)} - \frac{1}{3}\mathrm{Ric} + O(\sigma)\right)y}{2}\right)\left(1 - \frac{y^{\mathrm{T}}\mathrm{Ric}\,y}{6} + O(\|y\|^3)\right)dy$$

Let Θ be a positive square root of Σ such that $\Sigma = \Theta.\Theta^{\mathrm{T}}$. Which the change of variable $y = \Theta.x$, we have $dy = \sqrt{\det(\Sigma)}.dx$ and the probability is now:

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

where $S = \frac{1}{3}\Theta^{\mathrm{T}}\mathrm{Ric}\,\Theta$. As this is a term of the order $O(\sigma^2)$, we have the following Taylor expansion:

$$\exp\left(-\frac{x^{\mathrm{T}}\left(\mathrm{Id} - S + O(\sigma^3)\right)x}{2}\right) = \exp\left(-\frac{\|x\|^2}{2}\right)\left(1 + \frac{x^{\mathrm{T}}.S.x}{2} + O(\sigma^3)\right)$$

Reporting in the probability, we find that:

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

Now, we have:

$$\int_{\|x\| \leq \alpha} \exp\left(-\frac{\|x\|^2}{2}\right)\left(1 + \|x\|^2\right).O(\sigma^3).dx \leq O(\sigma^3)\int_{\mathbb{R}^n}(1 + \|x\|^2)\exp\left(-\frac{\|x\|^2}{2}\right)dx = O(\sigma^3)$$

Thus, using equation 58 for the value of k, we find the usual χ^2 probability, which is independent of the covariance Σ, up the the order σ^3:

$$\Pr\{\chi^2 \leq \alpha^2\} = (2\pi)^{-\frac{n}{2}} \int_{\|x\| \leq \alpha} \exp\left(-\frac{\|x\|^2}{2}\right).dx + O(\sigma^3) \tag{59}$$

This integral can be computed in polar coordinates in R^n: if $r = \|x\|$ is the radius and n is the corresponding unit vector ($x = r.n$), we have $dx = r^{n-1}.dr.dn$ and thus:

$$\begin{aligned}\int_{\|x\| \leq \alpha} \exp\left(-\frac{\|x\|^2}{2}\right).dx &= \left(\int_{\mathcal{S}_{n-1}} dn\right).\left(\int_0^\alpha r^{n-1}.\exp(-r^2/2).dr\right) \\ &= \left(\frac{2\pi^{n/2}}{\mathbf{\Gamma}(n/2)}\right).\left(\frac{1}{2}\int_0^{\alpha^2} t^{n/2-1}.\exp\left(-\frac{t}{2}\right).dt\right)\end{aligned}$$

with the change of variable $t = r^2$. In this formula, $\mathbf{\Gamma}$ is the Gamma function which can be recursively computed from $\mathbf{\Gamma}(x+1) = x.\mathbf{\Gamma}(x)$ with $\mathbf{\Gamma}(1) = 1$ and $\mathbf{\Gamma}(\frac{1}{2}) = \sqrt{\pi}$. For half integer values, the result is:

$$\mathbf{\Gamma}(\frac{n}{2}) = (k-1)! \quad \text{if} \quad n = 2.k \qquad \text{and} \qquad \mathbf{\Gamma}(\frac{n}{2}) = \sqrt{\pi}.\prod_{i=0}^{k-1}(i+\frac{1}{2}) \quad \text{if} \quad n = 2.k+1$$

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

$$p_{\chi^2}(u) = \frac{1 + O(\sigma^3)}{2.\mathbf{\Gamma}\left(\frac{n}{2}\right)}\left(\frac{u}{2}\right)^{\frac{n}{2}-1}\exp\left(-\frac{u}{2}\right) \tag{60}$$

If there is a cut locus, we have to replace $O(\sigma^i)$ by $O(\sigma^i) + \varepsilon\left(\frac{\sigma}{r}\right)$ 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 $\mathcal{B}(\lambda_{min}.r) \subset \mathcal{D}' \subset \mathbb{R}^n$, we can enclose the researched domain into: $\mathcal{B}(\min(\lambda_{min}.r, \alpha) \subset \mathcal{B}(\alpha) \cap \mathcal{D}' \subset \mathcal{B}(\alpha)$. For $\alpha \leq \lambda_{min}.r$, there is not problem but for $\alpha > \lambda_{min}.r$, we have:

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

and we have already seen that this integral is $1 + \varepsilon\left(\frac{\sigma}{r}\right)$. As $\alpha > \lambda_{min}.r$, 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 $O(\sigma^3) + \varepsilon\left(\frac{\sigma}{r}\right)$.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399