HAL open science

Curvature effects on the empirical mean in Riemannian and affine Manifolds: a non-asymptotic high concentration expansion in the small-sample regime

Xavier Pennec

To cite this version:

Xavier Pennec. Curvature effects on the empirical mean in Riemannian and affine Manifolds: a non-asymptotic high concentration expansion in the small-sample regime. 2019. hal-02157952

HAL Id: hal-02157952

https://inria.hal.science/hal-02157952v1

Preprint submitted on 17 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Curvature effects on the empirical mean in Riemannian and affine Manifolds: a non-asymptotic high concentration expansion in the small-sample regime

Xavier Pennec*

June 17, 2019

Abstract

The asymptotic concentration of the Fréchet mean of IID random variables on a Riemannian manifold was established with a central limit theorem by Bhattacharya & Patrangenaru (BP-CLT) [6]. This asymptotic result shows that the Fréchet mean behaves almost as the usual Euclidean case for sufficiently concentrated distributions. However, the asymptotic covariance matrix of the empirical mean is modified by the expected Hessian of the squared distance. This Hessian matrix was explicitly computed in [5] for constant curvature spaces in order to relate it to the sectional curvature. Although explicit, the formula remains quite difficult to interpret, and the intuitive effect of the curvature on the asymptotic convergence remains unclear. Moreover, we are most often interested in the mean of a finite sample of small size in practice. In this work, we aim at understanding the effect of the manifold curvature in this small sample regime. Last but not least, one would like computable and interpretable approximations that can be extended from the empirical Fréchet mean in Riemannian manifolds to the empirical exponential barycenters in affine connection manifolds.

For distributions that are highly concentrated around their mean, and for any finite number of samples, we establish explicit Taylor expansions on the first and second moment of the empirical mean thanks to a new Taylor expansion of the Riemannian log-map in affine connection spaces. This shows that the empirical mean has a bias in $1/n$ proportional to the gradient of the curvature tensor contracted twice with the covariance matrix, and a modulation of the convergence rate of the covariance matrix proportional to the covariance-curvature tensor. We show that our non-asymptotic high concentration expansion is consistent with the asymptotic expansion of the BP-CLT. Experiments on constant curvature spaces demonstrate that both expansions are very accurate in their domain of validity. Moreover, the modulation of the convergence rate of the empirical mean's covariance matrix is explicitly encoded using a scalar multiplicative factor that gives an intuitive vision of the impact of the curvature: the variance of the empirical mean decreases faster than in the Euclidean case in negatively curved space forms, with an infinite speed for an infinite negative curvature. This suggests potential links with the stickiness of the Fréchet mean described in stratified spaces. On the contrary, the variance of the empirical mean decreases more slowly than in the Euclidean case in positive curvature space forms, with divergence when we approach the limits of the Karcher & Kendall concentration conditions with a uniform distribution on the equator of the sphere, for which the Fréchet mean is not a single point any more.

*Universtié Côte d'Azur and Inria, Sophia Antipolis, France. Email: xavier.pennec@inria.fr

1 Introduction and overview of the results

The asymptotic concentration of the Fréchet mean of IID random variables on a Riemannian manifold was established in a central limit theorem by Bhattacharya & Patrangenaru (BP-CLT) [6]. This asymptotic result showed that the Fréchet mean behaves almost as the usual Euclidean case for sufficiently concentrated distributions once we have taken into account the curvature effects in the Hessian of the variance. These results were later generalized in [26] to establish weak laws of large numbers and central limit theorems of Lindeberg type for empirical Fréchet means of independent but non-identically distributed random variables. The essential quantity modifying the asymptotic covariance matrix of the empirical mean is the Hessian of the squared variance (the expectation of the Hessian of the squared distance). This Hessian matrix was explicitly computed in [5] for constant curvature spaces in order to relate the asymptotic dispersion of the Fréchet mean to the sectional curvature. Although explicit, the formula remains quite difficult to interpret, and the intuitive effect of the curvature on the asymptotic convergence remains cryptic. Moreover, we most often have in practice a finite sample of small size and it is interesting to understand if the manifold curvature has addition effects on the estimation of the mean in this small sample regime. Last but not least, we would like to obtain results that can be extended from the Fréchet mean in Riemannian manifolds to exponential barycenters in affine connection manifolds.

We investigate in this work the moments of the distribution of the empirical mean of a fixed number of samples in Riemannian and affine connection manifolds. We derive interpretable approximations for sufficiently concentrated distributions showing how the moments of the empirical mean deviate from the usual Euclidean case in the non-asymptotic case. We also study the speed of convergence towards their asymptotic value. In order to obtain a high concentration expansion of the minimum of the expected Riemannian squared distance, we may base our computations on a Taylor expansion of the Riemannian metric in a normal coordinate system at a fixed point. Such an extension is well known for the first orders, but going to orders higher than 4 is computationally much more involved and requires computational algebraic methods such as the ones developed in [8]. This leads to quite involved formulas with contractions over many covariant and contravariant indices[1]. In this paper, we base our Taylor expansions on a coordinate free expansion of the composition of two exponential maps developed by Gavrilov [19, 20]. This formula only involves the curvature and torsion tensors and their covariant derivatives at the development point. Moreover, the proposed extension relies only on the connection and not on the Riemannian metric, which makes it also suitable for the more general case of affine connection spaces with distributions sufficiently concentrated around their exponential barycenter.

Assumptions and objectives Let $X_n = \sum_{i=1}^n \delta_{x_i}$ be an IID n-sample drawn from a probability distribution μ on a Riemannian manifold $\mathcal{M}$. We want to investigate how the moments of the empirical Fréchet mean $\bar{x}_n$ of the empirical distribution X_n compares to the ones of the population Fréchet mean $\bar{x}$ of μ. This formulation obviously assumes that the Fréchet mean is unique. Thus, we assume that the distribution μ belongs to a sufficiently small strongly convex neighborhood $\mathcal{U}$ of diameter ε in $\mathcal{M}$. In Riemannian manifolds, this can be quantified by the Kendall and Karcher concentration (KKC) condition. This allows us to lift the distribution and all the computations to the convex subset $V_x = \log_x(\mathcal{U}) \subset T_x\mathcal{M}$ of the tangent space at any point $x \in \mathcal{U}$. In a convex affine connection space, the notion of mean that makes sense in convex subsets is based on exponential barycenters. One can show uniqueness in Arnaudon & Li convexity (ALC) conditions for convex affine manifold with p-convex geometry. This notion

[1]The results presented in this paper were originally developed in 2015 using Brewin formulas of this type. However, the difficulty to make sense of pages of computations involving summations over many indices led us to delay the publication until we found a simpler and more intuitive coordinate free formulation.

involves an auxiliary metric with which we can compute the diameter $\varepsilon = \mathrm{diam}(\mathcal{U})$ of the convex set $\mathcal{U}$.

Note that any distribution μ with a support in $\mathcal{U}$, including the empirical distribution $X_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$, has a unique mean. Moreover, the k-th order moments $\mathfrak{M}_k\mu(x) = \int_{\mathcal{M}} \log_x(z)^{\otimes k}\mu(dz)$ is of order k in $\varepsilon = \mathrm{diam}(\mathcal{U})$ since $\|\log_x(z)\|_x \leq \varepsilon$ for any x and $z \in V_x$. For our expansions, we also need to control the order of curvature terms that will appear. The Riemannian curvature tensor $R(u,v)w$ being multilinear in its variables, it is a term of order 3 in ε at each point $x \in \mathcal{M}$ for vectors $u, v, w \in V_x = \log_x(\mathcal{U}) \subset B(x,\varepsilon) \subset T_x\mathcal{M}$. Likewise, the covariant derivative $\nabla_t R(u,v)w$ is of order 4 for vectors in the same subset.

A distribution μ satisfying the KKC or the ALC conditions has a unique mean in $\mathcal{U}$ implicitly defined with the exponential barycenter formulation: $\mathfrak{M}_1\mu(z) = \int_{\mathcal{M}} \log_z(y)\,\mu(dy) = 0$. The goal is to find a high order approximation of the solution of the barycentric equation $\mathfrak{M}_1\mu(z) = 0$ in the neighborhood of a point $x \in \mathcal{U}$. For that purpose, we parametrize the points of $\mathcal{U}$ by $x_v = \exp_x(v)$ with $v \in V_x = \log_x(\mathcal{U}) \subset T_x\mathcal{M}$ and we look for a Taylor expansion of $\mathfrak{M}_1\mu(x_v) = 0$ with respect to v. However, the tangent mean $\mathfrak{M}_1\mu(x_v) = \int_{\mathcal{U}} \log_{x_v}(y)\,\mu(dy)$ is a vector field on $\mathcal{U}$, i.e. a mapping from $\mathcal{U} \subset \mathcal{M}$ to $T_{x_v}\mathcal{M}$. In order to have a unique image space, we may take a chart, which would imply working in a specific coordinate system, or more interestingly parallel transport each tangent vector from x_v to x: we obtain this way the recentered tangent mean map (Definition 5):

$$\mathfrak{N}_x^\mu(v) = \Pi^x_{\exp_x(v)}\mathfrak{M}_1\mu(\exp_x(v)) \in T_x\mathcal{M}.$$

The recentered tangent mean map is a mapping of vector spaces from $V_x \subset T_x\mathcal{M}$ to $T_x\mathcal{M}$ whose zeros parametrize the exponential barycenters of μ. In order to localize them, we aim at computing a series expansion of the recentered tangent mean map with respect to v.

Taylor expansions in affine connection spaces For that purpose, we develop methods for Taylor expansions in manifolds in Section 3. Based on the coordinate free expansion of the composition of two exponential maps of [19, 20] (the double exponential, theorem 2), we derive a series expansion up to order 5 of the logarithm of a fixed point $x_w = \exp_x(w)$ at a point $x_v = \exp_x(v)$, parallel transported back to x (Theorem 3):

$$\begin{aligned} l_x(v,w) =& \Pi^x_{x_v}\log_{x_v}(\exp_x(w)) = w - v + \frac{1}{6}R(w,v)(v-2w) \\ &+ \frac{1}{24}(\nabla_v R)(w,v)(2v-3w) + \frac{1}{24}(\nabla_w R)(w,v)(v-2w) + O(5). \end{aligned}$$

This neighboring log expansion is non-metric and valid for general affine connection manifolds. In the Riemannian case, we can take the square norm of that vector to get an expansion of the square Riemannian distance between two points that are away from the base point x. This expansion is the one needed to write an expansion of the variance around the point x.

However, since the minimum of the variance is in particular a critical point and thus a zero of the recentered tangent mean map $\mathfrak{N}_x^\mu(v)$, it is more convenient to compute directly the polynomial expansion of this vector space mapping. This is the focus of Section 4. Thanks to the previous work, it is relatively straightforward to show that

$$\begin{aligned} \mathfrak{N}_x^\mu(v) = \mathfrak{M}_1 - v + \frac{1}{6}R(\mathfrak{M}_1,v)v - \frac{1}{3}R(\bullet,v)_\bullet \vdots \mathfrak{M}_2 + \frac{1}{12}(\nabla_v R)(\mathfrak{M}_1,v)v \\ + \frac{1}{24}(\nabla_\bullet R)(\bullet,v)v \vdots \mathfrak{M}_2 - \frac{1}{8}(\nabla_v R)(\bullet,v)_\bullet \vdots \mathfrak{M}_2 - \frac{1}{12}(\nabla_\bullet R)(\bullet,v)_\bullet \vdots \mathfrak{M}_3 + O(\varepsilon^5) \end{aligned}$$

where the notation $R(\bullet,v)_\bullet\vdots\mathfrak{M}_2$ denotes the contraction of the tensor moment with the curvature tensor R along the axes specified by the bullets.

Solving for the value of $v = \log_x(\bar{x})$ that zeros out this expression leads to the polynomial expansion of the field $\log_x(\bar{x})$ pointing from the points x to the mean $\bar{x}$ (theorem 5):

$$\begin{aligned}\log_x(\bar{x}) =& \mathfrak{M}_1 - \frac{1}{3} R(\bullet, \mathfrak{M}_1)_\bullet \,\vdots\, \mathfrak{M}_2 - \frac{1}{24} \nabla_\bullet R(\bullet, \mathfrak{M}_1) \mathfrak{M}_1 \,\vdots\, \mathfrak{M}_2 \\ & - \frac{1}{8} \nabla_{\mathfrak{M}_1} R(\bullet, \mathfrak{M}_1)_\bullet \,\vdots\, \mathfrak{M}_2 - \frac{1}{12} \nabla_\bullet R(\bullet, \mathfrak{M}_1)_\bullet \,\vdots\, \mathfrak{M}_3 + O(\epsilon^5).\end{aligned}$$

Non asymptotic high concentration expansion of the moments of the empirical mean Equipped with this expansion, we analyze in Section 5 the expected first moment $\mathbf{E}\left[\, \log_x(\bar{x}_n) \,\right]$ and the expected second moment $\mathbf{E}\left[\, \log_x(\bar{x}_n) \otimes \log_x(\bar{x}_n) \,\right]$ of the empirical Fréchet mean $\bar{x}_n$ of n IID samples at the population mean $\bar{x}$. Using the previous expansion gives us a formula involving the tensor product of empirical moments of a sample. Taking the expectation of empirical moments is simple: it gives the moment of the underlying distribution. Taking the expectation of tensor products of empirical moments is more complex since it generates some tensor products of moments of mixed orders. We establish in Theorem 7 that the expected log of the empirical mean at the population mean is:

$$\mathbf{E}\left[\, \log_{\bar{x}}(\bar{x}_n) \,\right] = \tfrac{1}{6n} \left(1 - \tfrac{1}{n}\right) \mathfrak{M}_2 \,\overset{\circ}{\circ}\, \nabla_\bullet R(\bullet, \circ)_\circ \,\vdots\, \mathfrak{M}_2 + O\left(\epsilon^5\right).$$

This non asymptotic high concentration expansion of the first moment of the empirical mean on manifolds exhibits an unexpected bias in $1/n$ proportional to the (covariant) gradient of the curvature tensor contracted twice with the covariance matrix. This bias appearing in the small sample size regime was apparently completely unnoticed before.

For the second moment, Theorem 7 states that the covariance of the empirical mean is:

$$\mathbf{E}\left[\, \log_{\bar{x}}(\bar{x}_n) \otimes \log_{\bar{x}}(\bar{x}_n) \,\right] = \tfrac{1}{n} \left(\mathfrak{M}_2 - \tfrac{1}{3}\left(1 - \tfrac{1}{n}\right) \mathfrak{M}_2 \,\overset{\circ}{\circ}\, (\circ \otimes R(\bullet, \circ)_\bullet + R(\bullet, \circ)_\bullet \otimes \circ) \,\vdots\, \mathfrak{M}_2\right) + O\left(\epsilon^5\right).$$

This non asymptotic high concentration expansion of the covariance matrix clearly shows a modulation of the convergence rate in $1/n$ proportional to the covariance-curvature tensor. Experiments on constant curvature spaces demonstrate that the expansions are very accurate for variances that are smaller than the curvature, and allows to draw an intuitive understanding of the impact of the curvature on the statistical estimation.

Link with the BP-CLT We compare in Section 6 our non-asymptotic expansion to the asymptotic expansion on Riemannian manifolds [6, Theorem 2.2 p. 1231]. Rephrased with our notations, the BP-CLT states that, in KKC conditions, the empirical Fréchet mean $\bar{x}_n$ is a consistent estimator of the population Fréchet mean $\bar{x}$ and the random variable $\sqrt{n} \log_{\bar{x}}(\bar{x}_n) \in T_{\bar{x}}\mathcal{M}$ converges in law to a normal distribution of mean 0 and covariance $4\bar{H}^{(-1)}\, \mathfrak{M}_2\, \bar{H}^{(-1)}$, where the matrix $\bar{H}$ is the expectation of the Riemannian Hessian of the squared distance $\operatorname{dist}(., y)^2$ (Theorem 8). In this formula, we see that the expected Hessian is controlling the speed of convergence to the mean. Thus, the breaking of the BP-CLT in the case of zero eigenvalues of the expected Hessian may also be interpreted as an absence of convergence.

Establishing the high concentration expansion of $\bar{H}$ shows that both expansions are asymptotically consistent. However, our new non-asymptotic high concentration expansion has additional correction terms for the small data regime while the BP-CLT includes implicit correction terms in the expected Hessian of the squared distance for less concentrated distributions. One of the main interests of our new expansion is also to give an intuitive and intelligible interpretation of the expected Hessian of the squared distance in terms of the curvature and its derivatives.

Modulation of the convergence rate in space forms In order to better visualize the influence of the curvature on the empirical mean, we investigate in Section 7 the case of isotropic

distributions in constant curvature spaces, also called space forms. In this case, both the asymptotic and the high concentration expansions can be reduced to a scalar equation where the rate of convergence of the variance of the empirical mean with respect to the number of samples is modulated by a scalar factor $\alpha = \mathrm{Var}(\bar{x}_n)n/\sigma^2$ indicating how much the variance of the empirical mean deviates from the Euclidean case: a modulation factor $\alpha > 1$ indicates that the convergence is slower than in the Euclidean case, while a value $\alpha < 1$ indicates a faster convergence.

The setup can be further simplified by considering a uniform distribution on a Riemannian hypersphere of radius θ around the population mean $\bar{x}$. For a large number of samples in a manifold of large dimension, we obtain an archetypal modulation factor $\alpha = \frac{\tan^2(\sqrt{\kappa\theta^2})}{\kappa\theta^2}$ on the convergence of the variance of the empirical mean. We see that the variable controlling the modulation is actually $\kappa\theta^2$, the product of the sectional curvature with the variance. For a positive variance-curvature, the modulation of the rate of convergence is larger than one (the convergence is slower) and goes to infinity when $\kappa\theta^2$ approaches $\pi^2/4$. This corresponds exactly to the Kendall & Karcher concentration conditions under which all the results of this paper are restricted. This was expected since a uniform distributions on a Riemannian hypersphere of radius $\pi/2$ on a sphere fails to have a unique mean: the distribution of the empirical Fréchet mean converges to a mixture of Diracs rather than concentrating on a point as usual. For negative curvature, the modulation factor is below 1, meaning that the convergence is accelerated, and actually goes to zero (an infinite acceleration) for an infinitely negative variance-curvature. Such a phenomenon has been observed in specific cases for other types of means in negatively curves manifolds but was apparently not recognized so far as a general phenomenon of least-squares in manifolds. We conjecture that it is related to the phenomenon of stickiness of the Fréchet mean in stratified spaces.

These theoretical predictions are illustrated by experiments on the 2- and 3-sphere and on the hyperbolic space of dimension 3 which demonstrate that the formulas that we have obtained are very accurate in their own domains.

2 Means on Riemannian and affine connection manifolds

2.1 Riemannian manifolds

We consider a differential manifold $\mathcal{M}$ provided with a smooth scalar product $\langle\, . \mid . \,\rangle_x$ on each tangent space $T_x\mathcal{M}$ at point x of $\mathcal{M}$, called the Riemannian metric. In a chart, the metric is fully specified by the dot product of the coordinate vector fields: $g_{ij}(x) = \langle\, \partial_i \mid \partial_j \,\rangle$. The Riemannian distance between any two points on $\mathcal{M}$ is the infimum of the length of the curves joining these points. Geodesics are defined as the critical points of the energy functional. Geodesics are parametrized by arc-length in addition to optimizing the length functional. We assume in this paper that the manifold is geodesically complete, i.e. that 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. As an important consequence, the Hopf-Rinow-De Rham theorem states that there always exists at least one minimizing geodesic between any two points of the manifold (i.e. whose length is the distance between the two points).

From the theory of second order differential equations, we know that there exists one and only one geodesic $\gamma_{(x,v)}(t)$ starting from the point x with the tangent vector $v \in T_x\mathcal{M}$. The exponential map at point x maps each tangent vector $v \in T_x\mathcal{M}$ to the point of the manifold that is reached after a unit time by the geodesic: $\exp_x(v) = \gamma_{(x,v)}(1)$. The exponential map is locally one-to-one around 0: we denote by $\log_x(y)$ its inverse. To shorten formulas, we sometimes use the notation $\overrightarrow{xy}$ instead of $\log_x(y)$.

The cut time $t_{cut}(x, v)$ is the maximal time for which the normal geodesic starting at x with unit tangent vector $v \in T_x\mathcal{M}$ is length minimizing. By homogeneity, the cut-time can be

extended to the tangent bundle $T\mathcal{M}^*$: $t_{cut}(x,v) = t_{cut}(x, v/\|v\|_x)\|v\|_x$, except for null tangent vectors. The tangent cut-locus is defined as the set of vectors of $T_x\mathcal{M}^* = T_x\mathcal{M} \setminus \{0\}$ where the distance ceases to be minimizing: $C(x) = \{t_{cut}(x,v)v | v \in T_x\mathcal{M}^*\}$. The cut-locus is the image of the tangent cut-locus by the exponential map: $\mathrm{Cut}(x) = \exp_x(C(x))$. The distance to a point x is C^2 except at the cut locus where it is only continuous. The tangent cut locus delimits a star shaped domain around 0 in each tangent space whose interior is called the injectivity domain of the exponential map:

$$\mathrm{Inj}(x) = \{tv \mid 0 \leq t < t_{cut}(x,v), v \in T_x\mathcal{M}^*\}.$$

The injection radius is the infimum of the cut values of the various geodesics emanating from x (i.e. the radius of the largest open geodesic ball included in the injectivity domain, and thus on which $\exp_x$ is a diffeomorphism):

$$\mathrm{inj}(x) = \min\{t_{cut}(x, v/\|v\|_x) \mid v \in T_x\mathcal{M}^*\}.$$

There exists multiple notions of convexity in manifolds [4]. We use here the following: an open subset $U \subset \mathcal{M}$ is (strongly) convex if for any points $p, q \in U$, there exists a unique minimal geodesic γ joining p and q which belongs entirely to the subset: $\gamma \subset U$.

2.2 Convex affine manifolds

An affine manifold is a differential manifold $\mathcal{M}$ endowed with an affine connection ∇. Important classes of affine manifolds are: Riemannian manifolds with their Levi-Civita connection, Lie groups with their canonical symmetric space structure, and more generally affine symmetric spaces. The connection allows us to define geodesics as auto-parallel curves, or zero acceleration curves. These geodesics have an affine parametrization that measures relative distances along each geodesic, but there is no reference length to compare relative distances along different geodesics. Convexity may be defined as in Riemannian manifolds: an open set $\mathcal{U} \subset \mathcal{M}$ is convex if for every pair of points $x, y \in \mathcal{U}$ there exists a unique ∇-geodesic joining x and y lying entirely in $\mathcal{U}$ and that depends smoothly on its endpoints. This allows us to define the logarithm in a unique way within this subset. In that case, $(\mathcal{U}, \nabla)$ is called a convex (sub)-manifold. Whitehead Theorem tells us that there exists a convex neighborhood at each point of an affine manifold.

2.3 Curvature of a connection

In order to derive Taylor expansions in manifolds, we need to specify the notations for torsion and curvature operators and their coordinates in charts. The torsion tensor $T(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y] = -T(Y,X)$ measures how the skew-symmetric part of the connection differ from the Lie derivative $\mathcal{L}_X Y = [X,Y]$. The connection is torsion free if the torsion tensor vanishes identically. Two connections have the same geodesics if they have the same symmetric part $(\nabla_X Y + \nabla_Y X)/2$. i.e. if they only differ by torsion. Because means and barycenters in manifolds only involve geodesics and not parallel transport, we can restrict our attention to torsion free (also called symmetric) connections. This notably simplifies the expression of the Taylor expansions. In the sequel, all connections are assumed to be symmetric.

The Riemann curvature tensor in affine connection manifolds The curvature of an affine manifold is described by the $(1,3)$ curvature tensor $R : T\mathcal{M} \times T\mathcal{M} \times T\mathcal{M} \to TM$. It is defined from the covariant derivative by evaluation on vector fields u, v, w:

$$R(u,v)w = \nabla_u \nabla_v w - \nabla_v \nabla_u w - \nabla_{[u,v]} w. \tag{1}$$

The curvature can be interpreted as the difference between parallel transporting the vector w along an infinitesimal parallelogram with sides given by u and v. Two different sign conventions exist for the curvature tensor: the above definition is the one used in a number of reference

books in physics and mathematics [31, 30, 34]. Other authors use a minus sign to simplify some of the tensor notations [35, 32, 14, 4]. There exists moreover different conventions for the order of the tensors subscripts. We use here the convention $[R(u,v)w]^a = R^a_{bcd}u^c v^d w^b$ so that the tensor can be written using the Christoffel symbols Γ^a_{bc}:

$$R^a_{bcd} = dx^a(R(\partial_c, \partial_d)\partial_b) = \partial_c \Gamma^a_{db} - \partial_d \Gamma^a_{cb} + \Gamma^a_{ce}\Gamma^e_{db} - \Gamma^a_{de}\Gamma^e_{cb}, \tag{2}$$

where $\partial_i = \partial/\partial x_i$ are coordinate vector fields.

The curvature tensor has several symmetries that we will use to simplify expressions: it is skew-symmetric in the first two variables, and we we have the First and second Bianchi identities.

$$\begin{aligned} R(u,v) = -R(v,u) \quad &\text{or} \quad R^a_{bcd} = -R^a_{bdc}; \\ R(u,v)w + R(v,w)u + R(w,u)v = 0 \quad &\text{or} \quad R^a_{[bcd]} = R^a_{bcd} + R^a_{cdb} + R^a_{dbc} = 0; \\ (\nabla_u R)(v,w) + (\nabla_v R)(w,u) + (\nabla_w R)(u,v) = 0 \quad &\text{or} \quad R^a_{b[cd;e]} = \nabla_e R^a_{bcd} + \nabla_c R^a_{bde} + \nabla_d R^a_{bec} = 0. \end{aligned}$$

Additional symmetries in a Riemannian manifold In a Riemannian manifold, we can lower the first coordinate with the metric to obtain the $(0,4)$ version of the Riemannian curvature tensor:

$$R(u,v,w,z) = \langle\, R(w,z)v \mid u \,\rangle \quad \text{or in coordinates} \quad R_{abcd} = g_{ae}R^e_{bcd}. \tag{3}$$

This tensor inherits the above symmetries. Additionally, it is symmetric in the first and last two variables, which implies being skew-symmetric in the last two variables:

$$\begin{aligned} \langle\, R(u,v)w \mid z \,\rangle = -\langle\, R(u,v)z \mid w \,\rangle \quad &\text{or} \quad R_{abcd} = R_{cdab}; \\ \langle\, R(u,v)w \mid z \,\rangle = \langle\, R(w,z)u \mid w \,\rangle \quad &\text{or} \quad R_{abcd} = -R_{abdc}. \end{aligned}$$

This means in particular that its contraction with any symmetric tensor in the first two or last two indices is zero.

Note that in an affine manifold with an auxiliary metric g we can also lower the first index of the curvature tensor, but this metric should be kept in all computations as it is not the identity in a normal coordinate system and its covariant derivative is not zero as with the Levi-Civita connection.

Sectional curvature In Riemannian manifolds, the sectional curvature $\kappa(u,v)(x)$ measures the Gaussian curvature (the product of the principal curvatures) in the 2-planes of $T_x\mathcal{M}$ generated by the vectors u and v. It can be expressed from the curvature tensor by:

$$\kappa(u,v)(x) = \frac{\langle\, R(u,v)v \mid u \,\rangle_x}{\|u\|_x^2\|v\|_x^2 - \langle\, u \mid v \,\rangle_x^2}\,. \tag{4}$$

The upper bound of the sectional curvature plays a very important role for convexity in Riemannian manifolds, as we will see below.

2.4 Moments of a probability measure

Definition 1 *Let $\mu \in Prob(\mathcal{M})$ be a probability measure on a manifold $\mathcal{M}$. The k-order moment of μ is the $(k,0)$ tensor:*

$$\mathfrak{M}_k\mu(x) = \int_{\mathcal{M}} \underbrace{\log_x(y) \otimes \log_x(y) \ldots \otimes \log_x(y)}_{k \text{ times}}\, \mu(dy) = \int_{\mathcal{M}} \log_x(y)^{\otimes k}\, \mu(dy) \tag{5}$$

For a general distribution in a Riemannian manifold, the tensor field is not defined (and non-smooth) at points where the cut locus has a non-zero mass. If the density μ is uniformly bounded

by the Riemannian measure (i.e. if $\mu(dx) = \mu(x)d\mathcal{M}(x)$), then this integral defines a smooth $(k, 0)$ tensor field over $\mathcal{M}$ since the cut locus of the point x has null measure. In an affine manifold, the support of μ needs to be limited to an convex neighborhood so that the logarithm is well defined and all moments are smooth within this neighborhood. The 0-th order moment $\mathfrak{M}_0\mu = \int_{\mathcal{M}} d\mu(x) = 1$ is unit by definition. The first order moment $\mathfrak{M}_1\mu(x) = \int_{\mathcal{M}} \overrightarrow{xx_i}\, d\mu(x)$ is a vector field on the manifold $\mathcal{M}$.

In the following, we only consider distributions that have support in a convex affine manifold $(\mathcal{M}, \nabla)$ that is fixed once for all. Such a convex manifold is diffeomorphic to an open set of $\mathbb{R}^d$. Cartan-Hadamard manifolds (complete, simply connected manifolds with sectional curvature less than or equal to 0) are classical convex complete Riemannian manifolds. Examples of incomplete convex manifolds are given by small geodesic balls in Riemannian manifolds or small balls centered at the origin in an exponential chart in an affine manifold [2].

2.5 Means on Riemannian manifolds

The classical mean value of random numeric values with distribution μ is defined through an integral $\bar{x} = \int x\, \mu(dx)$. Because μ is normalized, this definition can be rewritten as an implicit barycentric equation: $\int (x - \bar{x})\, \mu(dx)$. With this formulation, it is obvious that this notion is affine and not metric. In the context of probability and statistics, Maurice Fréchet was the first to unify several type of typical statistical values, in particular the mean and the median, and to generalize them to abstract spaces. In a preparatory work, he first investigated [17] different ways to compute mean values of random triangles, independently of their position and orientation in space, with experiments to confront the theoretical results to real world data. In this respect, he may be considered as pioneering the statistical study of shapes. In a second study, motivated by the study of random curves, he first introduced a mean value and a law of large numbers defined by a generalization of the integral to normed vector (Wiener or Banach) space. Finally, Fréchet considered in [18, p.233] a family of central values (including the mean and the median) and its generalization to random elements in abstract metric spaces:

Definition 2 (Fréchet mean in a metric space [18]) *The p-mean (typical position of order p according to Fréchet) of a distribution (a random element) μ in an abstract metric space $\mathcal{M}$ is set of minima of the p-standard deviation :*

$$p\text{-}Mean(\mu) = \left\{ \underset{y \in \mathcal{M}}{\arg\min} \int_{\mathcal{M}} \mathrm{dist}(x, y)^p\, \mu(dx) \right\}. \tag{6}$$

The case $p = 2$ corresponds in vector spaces to the arithmetic mean, the case $p = 1$ to the median ("equiprobable value" in Fréchet's words).

The first key contribution of Fréchet was to consider many different types of typical elements, including of course the mean but also the median. Fréchet considered mainly the case $p \geq 1$, but he observed that many of the properties could be also generalized to $0 < p < 1$. His second revolutionary idea was to considered a set of mean elements rather than one unique mean. This idea was later developed by Ziezold [37] with a strong law of large numbers for sets of random elements in separable finite quasi-metric spaces. These two innovations justify the name of Fréchet mean that is used in geometric statistics.

For smooth differential geometric spaces like Riemannian manifolds, and restricting to the classical 2-mean, Berger [4, p.235] reported that "the existence of a unique center of mass in the large for manifolds with non-positive curvature was proven and used by Élie Cartan back in the 1920's." In order to find the fixed point of a group of isometries, Cartan indeed showed in [11] that the sum of the square distances from one point to a finite number of points has a unique minimum in simply connected Riemannian manifolds with non-positive curvature[2] (now

[2]Note III on normal spaces with negative or null Riemannian curvature, p. 267.

called a Hadamard space). This result was extended in [12] to closed subgroups of isometries. It is obvious in this text that Cartan is only using the uniqueness of the minimum of the sum of square distances as a tool in the specific case of negative curvature Riemannian manifolds and not as a general definition of the mean on manifolds as usually thought in probability or statistics.

For similar purposes, Grove and Karcher extended this idea in 1973 to positive curvature manifolds for distributions with sufficiently small support, typically within convex balls so that the mean exists and is unique [21]. The notion was coined *the Riemannian center of mass.* In this publication and in successive ones, Karcher and colleagues determined Jacobi field estimates to find optimal conditions for the convexity of the ball that support this definition. The Riemannian barycenter is commonly refered to as describes in [24] but the most complete description of the related properties is certainly found in [9], where a notion of barycenter in affine connection manifolds is also worked out. A good historical note on the history of the Riemannian barycenter is given in [1] and by Karcher himself in [25].

In all these works, the Riemannian center of mass is unique by definition. Considering a set-valued barycenter on an affine connection manifold was the contribution of Emery and Mokobodzki:

Definition 3 (Exponential barycenters in an affine connection manifold [16]) *We consider a probability measure μ with support in a convex local neighborhood $\mathcal{U}$ of an affine connection manifold $(\mathcal{M}, \nabla)$. Exponential barycenters of the probability measure μ are the points*

$$Bary_{\mathcal{U}}(\mu) = \left\{ x \in \mathcal{U} \mid \mathfrak{M}_1(x) = \int_{\mathcal{M}} \log_x(y)\, \mu(dy) = 0 \right\}.$$

The non-uniqueness of the expectation of a random variable considerably extended the usability of this definition, in particular in positive curvature manifolds. We notice that the notion remains purely affine, provided that the distribution has a support on a convex neighborhood in which the logarithm can be defined uniquely. In the Riemannian case, [16] showed that exponential barycenters are the critical points of the variance. Thus exponential barycenters contain in particular the minimum of the variance $\sigma^2(x) = \int_{\mathcal{M}} \mathrm{dist}^2(x, y)\, \mu(dy)$ (sometimes called the Fréchet functional), except if the variance is not differentiable at this point.

The optimal conditions for the existence and uniqueness of the minimum of the variance have been studied in [24, 9, 27, 28, 29]. The result has been extended to Fréchet p-means defined as the minima of the p-variance in [1, 36]. The Hessian of the squared distance to a point plays a key role in these conditions. Karcher looked for a ball where this Hessian is positive definite everywhere, while Kendall relaxed this condition to the locus of the minimum. We will see in Section 7.3 that the inverse of the expected Hessian also controls the speed of convergence to the mean, so that the non-uniqueness may also be interpreted as an absence of convergence.

Theorem 1 (Karcher & Kendall Concentration (KKC) Conditions) *Let $\mathcal{M}$ be a geodesically complete Riemannian manifold with injection radius $inj(x)$ Let μ is a probability distribution on $\mathcal{M}$ whose support is contained in a closed regular geodesic ball $\bar{B}(x, r)$ of radius $r < \frac{1}{2} inj(x)$. We assume moreover that the upper bound $\kappa = \sup_{y \in B(x,r), u \neq v \in T_y\mathcal{M}} \kappa(u, v)(y)$ of the sectional curvatures in the ball satisfies $\kappa < \pi^2/(2r)^2$. This second condition is always verified on spaces of negative curvature and it specifies a maximal radius $r^* = \frac{\pi}{2\sqrt{\kappa}}$ when there is positive sectional curvature. These concentration assumptions ensure that the variance has a unique global minimum that belongs to the ball $\bar{B}(x, r)$.*

In order to differentiate the different notions of means in Riemannian manifolds, it is usual in geometric statistics to name Fréchet mean the set of global minima of the variance, Karcher mean the set of local minima, and exponential barycenters the set of critical points satisfying the implicit equation $\mathfrak{M}_1(x) = 0$. It is clear that all these definition boil down to the same

unique point within the ball $B(x,r)$ in KKC conditions for the classical 2-mean, although some local minima and critical points may exist outside this ball. Throughout this paper, we assume that the support of all the distributions that we consider on Riemannian manifolds are included in a regular geodesic ball $B(x_0,r)$ of diameter $2r < \varepsilon$ that satisfies the KKC conditions. As a consequence, one can conclude that there is a unique exponential barycenter $\bar{x}$ included in the ball, necessarily the Fréchet mean [24, 27]. The maximal diameter ε of the support of the distribution is used in this paper as the scale variable to control Taylor expansions.

2.6 Means in convex affine connection manifolds

To define the mean of a probability distribution μ in a convex manifold $(\mathcal{M},\nabla)$, we cannot rely on the Fréchet mean since there is no distance. However, the notion of exponential barycenter still makes sense because the affine logarithm is well defined [16]: they are the zeros of the first moment field $\mathfrak{M}_1(\mu) = \int_{\mathcal{M}} \log_x(y)\, \mu(dy)$ for $x \in M$. This definition, studied in [9, 16], is close to the Riemannian center of mass but uses the logarithm of the affine connection in a convex domain instead of the Riemannian logarithm in the injectivity domain.

Distributions with compact support in a convex affine manifold have at least one exponential barycenter. Moreover, exponential barycenter are stable by affine diffeomorphisms (connection preserving maps, which thus preserve the geodesics and the normal convex neighborhoods) [9]. However, the classical notion of convexity is not sufficient to ensure the uniqueness of the mean. For that purpose, stronger convexity conditions have been proposed.

Definition 4 (Arnaudon & Li convexity (ALC) conditions [2]) *Let $(\mathcal{M},\nabla)$ be an affine manifold. A separating function on $\mathcal{M}$ is a convex function $\rho : \mathcal{M}\times\mathcal{M} \to \mathbb{R}^+$ vanishing exactly on the diagonal of the product manifold (considered as an affine manifold with the direct product connection). Here, convex means that the restriction of $\rho(\gamma(t))$ to any geodesic $\gamma(t)$ of $\mathcal{M}\times\mathcal{M}$ is a convex function from $\mathbb{R}$ to $\mathbb{R}^+$. A manifold which carries a smooth separating function ρ such that*

$$c\ \mathrm{dist}(x,y)^p \le \rho(x,y) \le C\ \mathrm{dist}(x,y)^p,$$

for some constants $0 < c < C$, some positive integer $p \ge 2$ and some auxiliary Riemannian distance function dist *is called a manifold with p-convex geometry. On a manifold with p-convex geometry, every compactly supported probability measure has a unique exponential barycenter.*

Whitehead theorem tells us in essence that any point in an affine connection manifold has a convex neighborhood with 2-convex geometry. In Riemannian manifolds, geodesic balls in KKC conditions have 2-convex geometry. These conditions will be sufficient for our goal of studying the behavior of the empirical means with high concentration. We should note that the notion is too strong in general since there are examples of manifolds where one can prove uniqueness although they do not have p-convex geometry for any p. This has motivated the definition of CSLCG (convex, with semi-local convex geometry) manifolds in [2] but we will not need it in this paper.

Throughout this paper, we only consider distributions that satisfy the ALC condition, i.e. compact distributions in a convex affine manifold $(\mathcal{M},\nabla)$ with p-convex geometry. With this extra structure, we have an auxiliary Riemannian metric on $\mathcal{M}$ whose norm in $T_x\mathcal{M}$ is denoted $\|.\|_x$ and whose Riemannian distance is denoted dist. Thanks to this distance, we can compute the diameter ε of the support of the distribution, which is then used as in Riemannian manifolds as the typical scale variable to control Taylor expansions.

3 Taylor expansions in Riemannian and affine manifolds

In order to analyze the variance of a random element in a Riemannian manifold, we may compute the Taylor expansion of the squared Riemannian distance functions in a local coordinate system.

This requires to have an expression for the geodesics. In a normal coordinate system centered at x, a geodesic starting at x with tangent vector v is a straight line: $\log_x(\exp_x(tv)) = tv$. However, the geodesic starting at $x_v = \exp_x(v)$ and ending at $x_w = \exp_x(w)$ deviates from the straight line from v to w in our chart because of curvature. Geodesic may be approximated by a polynomial expansion in the variables v and w that solves the geodesic equation.

The expansion of the Riemannian logarithm can be based on the Taylor expansion of the Riemannian metric in a normal coordinate system. Such an extension is well known for the first orders, but going to orders higher than 4 is computationally much more involved and requires computational algebraic methods such as the ones developed by Brewin [8]. Part of the results presented in this paper were originally developed in 2015 using this method. However, the intelligibility of the formulas involving lots of terms and indices was difficult.

In this paper, we base our Taylor expansions on a more geometric formulation based on a coordinate free expansion of the composition of two exponential maps developed by Gavrilov [19, 20]. This formula only involves the curvature and torsion tensors and their covariant derivatives at the development point. Moreover, the proposed extension relies only on the connection and not on the Riemannian metric, which makes it immediately suitable for the general affine connection case.

3.1 Gavrilov's expansions of the double exponential in affine manifolds

In the spirit of the Baker-Campbell-Hausdorff (BCH) formula for Lie groups, Gavrilov [19, 20] developed a coordinate free expansion of the composition of two exponential maps. The double exponential $\exp_x(v, u) = \exp_{\exp_x(v)}(\Pi_x^{\exp_x(v)} u)$ corresponds to a first geodesic shooting from the point x along the vector v, followed by a second geodesic shooting from $y = \exp_x(v)$ along the parallel transport $\Pi_x^y u$ of the vector u along the first geodesic (Fig. 1, left). This expansion holds in general affine connection manifolds, and has a surprisingly simple coordinate free formulation. We formulate it here in the torsion free case.

Theorem 2 (The double exponential expansion [19, 20]) *In a torsion-free affine connection manifold, the log of the double exponential $h_x(v, u) = \log_x(\exp_x(v, u))$ has the following series expansion in v and w at order 5:*

$$\begin{aligned} h_x(v,u) = {} & v + u + \frac{1}{6}R(u,v)v + \frac{1}{3}R(u,v)u + \frac{1}{24}(\nabla_v R)(u,v)(2v+5u) \\ & + \frac{1}{24}(\nabla_u R)(u,v)(v+2u) + O(5), \end{aligned} \tag{7}$$

where $O(5)$ represents polynomial terms of order 5 or more in u and v.

In the right-hand side, the tensor values are taken at x. By extending the vectors u and v of $T_x\mathcal{M}$ to vector fields in the neighborhood of x using parallel transport, we see that the series is valid point-wise at any point in that neighborhood.

The fundamental idea of this expansion is the use of parallel transport to identify tangent spaces rather than the differential of the exponential as is implicitly done in Taylor expansions in normal coordinate systems [8]. This trick drastically simplifies the expressions with coordinate free expressions. We notice that h_x is a mapping of vector spaces from $T_x\mathcal{M} \times T_x\mathcal{M}$ to $T_x\mathcal{M}$.

3.2 Moving the foot-point: the neighboring log expansion

The double exponential formula may be used to evaluate how the log of a fixed point $y = \exp_x(w)$ is modified when x is moved along the geodesic $x_v = \exp_x(v)$: we define the neighboring log as $l_x(v, w) = \Pi_{x_v}^x \log_{x_v}(\exp_x(w))$ (Fig. 1, right). This amounts to say that the vector $u = l_x(v, w)$ is solution of $\exp_{x_v}(\Pi_x^{x_v} u) = \exp_x(w)$. We recognize the double exponential $\exp_x(v, u)$ on the left-hand side. Thus, determining $u = l_x(v, w)$ amounts to solve $h_x(v, u) = \log_x(\exp_x(v, u)) = w$.

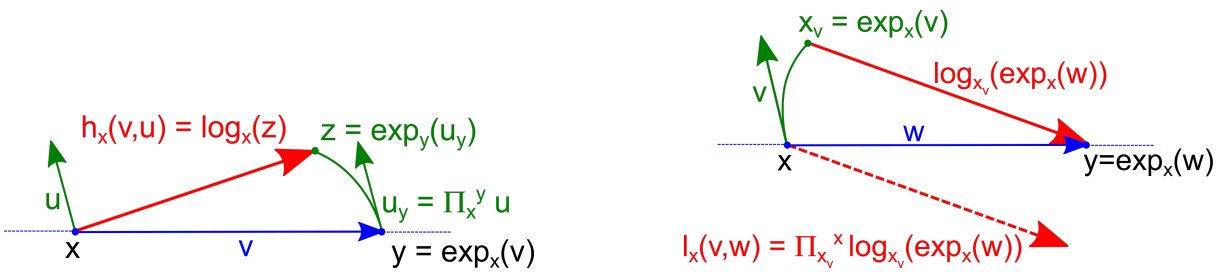

Figure 1: **Left:** the log of the composition of two exponentials (BCH-type formula) in a normal coordinate system at x. **Right:** The neighboring log of a fixed point $\exp_x(w)$ when the foot-point x is moved along the geodesic $x_v = \exp_x(v)$, in a normal coordinate system at x.

The first order solution of $h_x(v,u) = w$ is obviously $u = w - v + O(2)$. To find the second order, we assume that $u = w - v + u_2 + O(3)$ where u_2 is an unknown polynomial expression of order 2 in v and w. Plugging this value in Eq.7 gives $u_2 = O(3)$ since the curvature terms are of order 3. Assuming a third order term u_3, we now get:

$$h_x(v, w - v + u_3 + O(4)) = w + u_3 - \frac{1}{6}R(w,v)v + \frac{1}{3}R(w,v)w + O(4).$$

Thus we find $u_3 = \frac{1}{6}R(w,v)(v-2w)$. Finally, assuming a forth-order term u_4 in $u = w - v + \frac{1}{6}R(w,v)(v-2w) + u_4 + O(5)$ gives:

$$h_x(v,u) = w + u_4 + \frac{1}{24}(\nabla_v R)(w,v)(5w-3v) + \frac{1}{24}(\nabla_{(w-v)}R)(w,v)(2w-v) + O(5),$$

so that $h_x(v,u) = w$ implies $u_4 = \frac{1}{24}(\nabla_v R)(w,v)(2v-3w) + \frac{1}{24}(\nabla_w R)(w,v)(v-2w)$.

Theorem 3 (The neighboring log expansion) *In a torsion-free affine connection manifold, the neighboring log $l_x(v,w) = \Pi^x_{x_v}\log_{x_v}(\exp_x(w))$ has the following series expansion in v and w at order 5:*

$$\begin{aligned} l_x(v,w) =& w - v + \frac{1}{6}R(w,v)(v-2w) \\ &+ \frac{1}{24}(\nabla_v R)(w,v)(2v-3w) + \frac{1}{24}(\nabla_w R)(w,v)(v-2w) + O(5). \end{aligned} \tag{8}$$

3.3 Expansion of the Riemannian distance

With the neighboring log expansion, we can now come back to the expansion of the squared Riemannian distance between two points $x_v = \exp_x(v)$ and $x_w = \exp_x(w)$ that are close to x:

$$\begin{aligned} \mathrm{dist}(x_v, x_w)^2 &= \|\log_{x_v}(x_w)\|^2_{x_v} = \|\Pi^x_{x_v}\log_{x_v}(x_w)\|^2_x = \|l_x(v,w)\|^2_x \\ &= \|w-v\|^2_x + \frac{1}{3}\langle R(w,v)(v-2w) \mid w-v\rangle_x \\ &\quad + \frac{1}{12}\langle (\nabla_v R)(w,v)(2v-3w) \mid w-v\rangle_x \\ &\quad + \frac{1}{12}\langle (\nabla_w R)(w,v)(v-2w) \mid w-v\rangle_x + O(6). \end{aligned}$$

We now exploit the skew symmetry $\langle R(u,v)w \mid z\rangle = -\langle R(u,v)z \mid w\rangle$ to simplify the terms of order 4: $\langle R(w,v)(v-2w) \mid w-v\rangle = \langle R(w,v)v \mid w\rangle + 2\langle R(w,v)w \mid v\rangle = \langle R(w,v)w \mid v\rangle$. The derived skew-symmetry for the covariant derivative $\langle (\nabla_v R)(w,v)u \mid z\rangle = -\langle (\nabla_v R)(w,v)z \mid u\rangle$ allows us to simplify the terms of order 5:

$$\langle (\nabla_v R)(w,v)(2v-3w) \mid w-v\rangle_x = 2\langle \nabla_v R(w,v)v \mid w\rangle_x + 3\langle \nabla_v R(w,v)w \mid v\rangle_x = \langle \nabla_v R(w,v)w \mid v\rangle_x .$$

$\langle\, (\nabla_w R)(w,v)(v-2w) \mid w-v \,\rangle_x = \langle\, \nabla_w R(w,v)v \mid w \,\rangle_x + 2\,\langle\, \nabla_w R(w,v)w \mid v \,\rangle_x = \langle\, \nabla_w R(w,v)w \mid v \,\rangle_x$.

Finally, we obtain the Taylor expansion of the squared geodesic distance in a Riemannian manifold between two points $x_v = \exp_x(v)$ and $x_w = \exp_x(w)$ that are close to x:

$$\mathrm{dist}^2(x_v, x_w) = \|w-v\|_x^2 + \frac{1}{3}\,\langle R(w,v)w\,,\, v\rangle_x + \frac{1}{12}\,\left\langle \nabla_{(v+w)} R(w,v)w\,,\, v\right\rangle_x + O(6). \tag{9}$$

In coordinates, using the orthonormal property of a normal coordinate system, this reads: $\mathrm{dist}^2(x_v, x_w) = (w-v)^a(w-v)_a + \frac{1}{3}R_{abcd}v^a w^b w^c v^d + \frac{1}{12}\nabla_e R_{abcd} v^a w^b w^c v^d (w+v)^e + O(6)$.

4 Expansion of the first moment of the mean

Let μ be a probability distribution with support in a convex neighborhood $\mathcal{U}$ of diameter less than ε respecting the KKC conditions in a Riemannian manifold or the ALC conditions in a convex affine manifold. In such conditions, all the definitions of the mean boils down to the unique zero of the first moment: $\mathfrak{M}_1\mu(x) = \int_{\mathcal{U}} \log_x(y)\,\mu(dy) = 0$. The goal is this section is to find a polynomial approximation of the solution of this equation in the neighborhood of a point $x \in \mathcal{U}$.

For that purpose, we parametrize the points of $\mathcal{U}$ by $x_v = \exp_x(v)$ with $v \in V_x = \log_x(\mathcal{U}) \subset T_x\mathcal{M}$. Since the tangent mean $\mathfrak{M}_1\mu(x_v)$ is a vector field on $\mathcal{U}$, its value belongs to a different tangent space $T_{x_v}\mathcal{M}$ for each value of $v \in V_x$. Thus, establishing a Taylor expansion requires to express them all in a single chart. A natural choice is a normal coordinate system at x, but the resulting expansion will still depend on that specific coordinate system. In order to have a unique image space, we may also parallel transport each tangent vector from x_v to x to obtain the following mapping of vector space.

Definition 5 (Recentered tangent mean map) *Let μ be a probability distribution with support in a convex open subset $\mathcal{U}$ of an affine manifold $\mathcal{M}$. Then, for any vector $v \in V_x = \log_x(\mathcal{U}) \subset T_x\mathcal{M}$ for which $\mathfrak{M}_1\mu(\exp_x(v))$ is finite, we define the recentered tangent mean map as:*

$$\mathfrak{N}_x^\mu(v) = \Pi^x_{\exp_x(v)}\mathfrak{M}_1\mu(\exp_x(v)) \in T_x\mathcal{M}. \tag{10}$$

This is a mapping of vector spaces from $V_x \subset T_x\mathcal{M}$ to $T_x\mathcal{M}$ whose zeros parametrize the exponential barycenters of μ.

The mapping can be extended to the whole injection domain $\mathrm{Inj}(x)$ in the Riemannian case. The zeros of the recentered tangent map parametrize the exponential barycenters of μ because parallel transport is an isometry of tangent spaces in Riemannian manifolds: $\mathfrak{N}_x^\mu(v) = 0$ if and only if $\mathfrak{M}_1\mu(\exp_x(v)) = 0$. This property also holds for affine manifolds since parallel transport is an isomorphism of tangent spaces in that case. In KKC or ALC conditions, we know that the zero is unique.

Since we only deal with one unique probability distribution μ in the sequel, we simplify the notations below by dropping the dependency on μ in $\mathfrak{M}_k$ and $\mathfrak{N}_x$.

4.1 Taylor expansion of the recentered mean map

Using the neighboring log formula (Eq. 8), we can compute the series expansion of the recentered mean map. By linearity of the integral, we have:

$$\begin{aligned}\mathfrak{N}_x(v) &= \int_{\mathcal{M}} \Pi^x_{\exp_{x_v}} \log_{x_v}(y)\,\mu(dy) = \int_{\mathcal{M}} l_x(v, \log_x(y))\,\mu(dy)\\ &= \int_{\mathcal{M}} (\log_x(y) - v)\,\mu(dy) + \frac{1}{6}\int_{\mathcal{M}} R(\log_x(y), v)(v - 2\log_x(y))\,\mu(dy)\\ &\quad + \int_{\mathcal{M}} \frac{1}{24}(\nabla_v R)(\log_x(y), v)(2v - 3\log_x(y))\,\mu(dy)\\ &\quad + \int_{\mathcal{M}} \frac{1}{24}(\nabla_{\log_x(y)} R)(\log_x(y), v)(v - 2\log_x(y))\,\mu(dy) + \int_{\mathcal{M}} O(5)\,\mu(dy).\end{aligned}$$

In the integral of the first terms of the series, we recognize the first moment and the normalization (zeroth moment) of the distribution: $\int_{\mathcal{M}}(\log_x(y) - v)\,\mu(dy) = \mathfrak{M}_1 - v$. For the second order term, we take the curvature tensor out of the integral thanks to its mutlilinearity, and we see the first and second moment appearing:

$$\int_{\mathcal{M}} R(\log_x(y), v)(v - 2\log_x(y))\,\mu(dy) = R(\mathfrak{M}_1, v)v - 2R(\bullet, v)_\bullet \,\vdots\, \mathfrak{M}_2.$$

In this formula, the notation $R(\bullet, v)_\bullet \vdots \mathfrak{M}_2$ denotes the contraction of the $(2,0)$ symmetric moment tensor with the $(1,3)$ curvature tensor R along the axes specified by the bullets. In coordinates, this writes $[R(\bullet, v)_\bullet \,\vdots\, \mathfrak{M}_2]^a = R^a_{bcd}\, v^d\, \mathfrak{M}_2^{bc}$. In further computation, we will also use contractions like $\nabla_\bullet R(\bullet, v)_\bullet \,\vdots\, \mathfrak{M}_3$ or $\mathfrak{M}_2 \,\overset{\circ}{\vdots}\, \nabla_\bullet R(\bullet, \circ)_\circ \,\vdots\, \mathfrak{M}_2$. For the 4th order terms, we get:

$$\int_{\mathcal{M}} (\nabla_v R)(\log_x(y), v)(2v - 3\log_x(y))\,\mu(dy) = 2(\nabla_v R)(\mathfrak{M}_1, v)v - 3(\nabla_v R)(\bullet, v)_\bullet \,\vdots\, \mathfrak{M}_2,$$

$$\int_{\mathcal{M}} (\nabla_{\log_x(y)} R)(\log_x(y), v)(v - 2\log_x(y))\,\mu(dy) = (\nabla_\bullet R)(\bullet, v)v \,\vdots\, \mathfrak{M}_2 - 2(\nabla_\bullet R)(\bullet, v)_\bullet \,\vdots\, \mathfrak{M}_3.$$

Last but not least, $O(5)$ is a homogeneous polynomial of order at least 5 in v and $\log_x(y)$. Both vectors have a norm less than the diameter ε of the distribution, so that $\int_{\mathcal{M}} O(5)\,\mu(dy) = O(\varepsilon^5)$. We finally obtain the following expansion.

Theorem 4 (The recentered tangent mean map expansion) *Let μ be a probability distribution with support in a convex open subset $\mathcal{U}$ of diameter ε of an affine manifold $\mathcal{M}$ satisfying the KKC or ALC conditions. The expansion of the recentered tangent mean map around a point $x \in \mathcal{U}$ with respect to the vector $v \in V_x = \log_x(\mathcal{U}) \subset T_x\mathcal{M}$ is:*

$$\begin{aligned}\mathfrak{N}_x(v) = \mathfrak{M}_1 - v &+ \frac{1}{6}R(\mathfrak{M}_1, v)v - \frac{1}{3}R(\bullet, v)_\bullet \,\vdots\, \mathfrak{M}_2\\ &+ \frac{1}{12}(\nabla_v R)(\mathfrak{M}_1, v)v - \frac{1}{8}(\nabla_v R)(\bullet, v)_\bullet \,\vdots\, \mathfrak{M}_2\\ &+ \frac{1}{24}(\nabla_\bullet R)(\bullet, v)v \,\vdots\, \mathfrak{M}_2 - \frac{1}{12}(\nabla_\bullet R)(\bullet, v)_\bullet \,\vdots\, \mathfrak{M}_3 + O(\epsilon^5)\end{aligned} \tag{11}$$

4.2 Solving for the zero of the recentered mean map

Now that we have the expansion of the recentered mean map, the last step is to solve for the value of v that zeros it out. The second order approximation of $\mathfrak{N}_x(v) = 0$ is obviously $v = \mathfrak{M}_1 + O(\varepsilon^3)$ since the curvature terms are of order three. Now, to identify the third order term v_3, we plug the value $\mathfrak{M}_1 + v_3 + O(\varepsilon^4)$ in Eq.(11):

$$\mathfrak{N}_x(\mathfrak{M}_1 + v_3 + O(\varepsilon^4)) = -v_3 + \frac{1}{6}R(\mathfrak{M}_1, \mathfrak{M}_1)\mathfrak{M}_1 - \frac{1}{3}R(\bullet, \mathfrak{M}_1)_\bullet \,\vdots\, \mathfrak{M}_2 + O(\varepsilon^4).$$

The first term cancels out thanks to the skew-symmetry of the curvature tensor, so that we find that $v = \mathfrak{M}_1 - \frac{1}{3}R(\bullet, \mathfrak{M}_1)_\bullet \vdots \mathfrak{M}_2 + O(\varepsilon^4)$ is required to cancel out the recentered mean map. The fourth order give:

$$\mathfrak{N}_x\left(\mathfrak{M}_1 - \frac{1}{3}R(\bullet, \mathfrak{M}_1)_\bullet \vdots \mathfrak{M}_2 + v_4 + O(\varepsilon^5)\right) = -v_4 + \frac{1}{12}(\nabla_{\mathfrak{M}_1}R)(\mathfrak{M}_1, \mathfrak{M}_1)\mathfrak{M}_1$$
$$- \frac{1}{8}(\nabla_{\mathfrak{M}_1}R)(\bullet, \mathfrak{M}_1)_\bullet \vdots \mathfrak{M}_2 + \frac{1}{24}(\nabla_\bullet R)(\bullet, \mathfrak{M}_1)\mathfrak{M}_1 \vdots \mathfrak{M}_2 - \frac{1}{12}(\nabla_\bullet R)(\bullet, \mathfrak{M}_1)_\bullet \vdots \mathfrak{M}_3 + O(\epsilon^5).$$

Removing the vanishing term due to the skew-symmetry of the curvature tensor, we find:

$$v_4 = -\frac{1}{8}(\nabla_{\mathfrak{M}_1}R)(\bullet, \mathfrak{M}_1)_\bullet \vdots \mathfrak{M}_2 + \frac{1}{24}(\nabla_\bullet R)(\bullet, \mathfrak{M}_1)\mathfrak{M}_1 \vdots \mathfrak{M}_2 - \frac{1}{12}(\nabla_\bullet R)(\bullet, \mathfrak{M}_1)_\bullet \vdots \mathfrak{M}_3 + O(\epsilon^5).$$

Theorem 5 *Let μ be a probability distribution with support in a convex open subset $\mathcal{U}$ of diameter ε of an affine manifold $\mathcal{M}$ satisfying the KKC or ALC conditions. The vector field that points from a point of $\mathcal{U}$ to the mean $\bar{x}$ characterized by $\mathfrak{M}_1(\bar{x}) = 0$ has the following Taylor expansion:*

$$\begin{aligned}\log_x(\bar{x}) = \mathfrak{M}_1 - \frac{1}{3}R(\bullet, \mathfrak{M}_1)_\bullet \vdots \mathfrak{M}_2 + \frac{1}{24}\nabla_\bullet R(\bullet, \mathfrak{M}_1)\mathfrak{M}_1 \vdots \mathfrak{M}_2 \\ - \frac{1}{8}\nabla_{\mathfrak{M}_1}R(\bullet, \mathfrak{M}_1)_\bullet \vdots \mathfrak{M}_2 - \frac{1}{12}\nabla_\bullet R(\bullet, \mathfrak{M}_1)_\bullet \vdots \mathfrak{M}_3 + O(\epsilon^5).\end{aligned} \tag{12}$$

In coordinates, we have $[R(u,v)w]^a = R^a_{bcd}w^b u^c v^d$ and $[\nabla_u R(v,w)z]^a = \nabla_b R^a_{cde} z^c v^d w^e = R^a_{cde;b} z^c v^d w^e$, so that this formula writes:

$$\begin{aligned}\log_x(\bar{x})^a = \mathfrak{M}_1^a - \frac{1}{3}R^a_{bcd}\mathfrak{M}_2^{bc}\mathfrak{M}_1^d + \frac{1}{24}\nabla_e R^a_{bcd}\mathfrak{M}_2^{ce}\mathfrak{M}_1^b\mathfrak{M}_1^d \\ - \frac{1}{8}\nabla_e R^a_{bcd}\mathfrak{M}_1^e\mathfrak{M}_2^{bc}\mathfrak{M}_1^d - \frac{1}{12}\nabla_e R^a_{bcd}\mathfrak{M}_1^d\mathfrak{M}_3^{bce} + O(\epsilon^5),\end{aligned}$$

The order 3 of this expression was earlier derived by [13][Theorem 3.2] with an opposite sign convention for the curvature (see appendix A for the details of the notations equivalences). In this work, we add the order 4, which will turn out to be crucial for establishing the bias on the empirical Fréchet mean, and the method to get higher orders.

5 Non-asymptotic high concentration expansion of the empirical mean

Let μ be a distribution on $\mathcal{M}$ satisfying the KKC or ALC conditions, with mean $\bar{x}$. An IID n-sample $X_n = \{x_1, \dots x_n\} \in \mathcal{M}^n$ may be identified with the empirical distribution $X_n \simeq \frac{1}{n}\sum_i \delta_{x_i}$. To study the law of its empirical mean $\bar{x}_n$, we focus in this paper on its expected moments. To simplify the notations, we denote in this section $\mathfrak{M}_k = \mathfrak{M}_k\mu$ the moments of the underlying distribution μ and $\mathfrak{X}_k^n = \mathfrak{M}_k X_n$ the k-th moments of the n-sample X_n.

Since X_n is a distribution, we can use directly Theorem 5 to find that the location of the empirical Fréchet mean $\bar{x}_n$ with respect to a point x:

$$\begin{aligned}\log_x(\bar{x}_n) = \mathfrak{X}_1^n - \frac{1}{3}R(\bullet, \mathfrak{X}_1^n)_\bullet \vdots \mathfrak{X}_2^n + \frac{1}{24}\nabla_\bullet R(\bullet, \mathfrak{X}_1^n)\mathfrak{X}_1^n \vdots \mathfrak{X}_2^n \\ - \frac{1}{8}\nabla_{\mathfrak{X}_1^n}R(\bullet, \mathfrak{X}_1^n)_\bullet \vdots \mathfrak{X}_2^n - \frac{1}{12}\nabla_\bullet R(\bullet, \mathfrak{X}_1^n)_\bullet \vdots \mathfrak{X}_3^n + O(\epsilon^5).\end{aligned} \tag{13}$$

This formula makes use of the empirical moment $\mathfrak{X}_k^n(x) = \frac{1}{n}\sum_{i=1}^n \overrightarrow{xx_i} \otimes \dots \otimes \overrightarrow{xx_i}$ of the n-sample X_n. The expectation of this tensor when the x_i's are IID with law μ is simply

$$\mathbf{E}\left[\,\mathfrak{X}_k^n(x)\,\right] = \frac{1}{n}\sum_{i=1}^n \mathbf{E}\left[\,\overrightarrow{xx_i} \otimes \dots \otimes \overrightarrow{xx_i}\,\right] = \frac{1}{n}\sum_{i=1}^n \int_{\mathcal{M}} \overrightarrow{xx_i} \otimes \dots \otimes \overrightarrow{xx_i}\,\mu(dx_i) = \mathfrak{M}_k(x).$$

However, the expectation of the tensor product of moments is more complex than the simple contraction of their expectation. The reason is that the random variables x_i and x_j appearing in two empirical moments are independent if $i \neq j$ but not if $i = j$, in which case a higher order moments appears.

5.1 Expectation of tensor product of empirical moments

In this section, we keep the sums over the variables i, j, k explicit to stress that these are the indices of the data points and not covariant indices of vectors. For the product of the first and second moments appearing above, we have:

$$\begin{aligned}
\mathbf{E}\left[\,(\mathfrak{X}_1^n)\otimes(\mathfrak{X}_2^n)\,\right] &= \frac{1}{n^2}\sum_{i,j}\mathbf{E}\left[\,\overrightarrow{xx_i}\otimes\overrightarrow{xx_j}\otimes\overrightarrow{xx_j}\,\right] \\
&= \frac{1}{n^2}\sum_{i,j\neq i}\mathbf{E}\left[\,\overrightarrow{xx_i}\otimes\overrightarrow{xx_j}\otimes\overrightarrow{xx_j}\,\right] + \frac{1}{n^2}\sum_{i}\mathbf{E}\left[\,\overrightarrow{xx_i}\otimes\overrightarrow{xx_i}\otimes\overrightarrow{xx_i}\,\right] \\
&= \frac{n-1}{n}\mathfrak{M}_1\otimes\mathfrak{M}_2 + \frac{1}{n}\mathfrak{M}_3
\end{aligned}$$

This computation can be generalized to the expectation of the product of two moments of any order:

$$\mathbf{E}\left[\,\mathfrak{X}_p^n\otimes\mathfrak{X}_q^n\,\right] = \frac{n-1}{n}\mathfrak{M}_p\otimes\mathfrak{M}_q + \frac{1}{n}\mathfrak{M}_{p+q}$$

For three moments, we have:

$$\begin{aligned}
\mathbf{E}\left[\,\mathfrak{X}_p^n\otimes\mathfrak{X}_q^n\otimes\mathfrak{X}_r^n\,\right] =& \frac{1}{n^3}\sum_{i,j,k}\mathbf{E}\left[\,(\overrightarrow{xx_i}^{\otimes p})(\overrightarrow{xx_j}^{\otimes q})(\overrightarrow{xx_k}^{\otimes r})\,\right] \\
=& \frac{(n-1)(n-2)}{n^2}\mathfrak{M}_p\otimes\mathfrak{M}_q\otimes\mathfrak{M}_r + \frac{1}{n^2}\mathfrak{M}_{p+q+r} \\
&+ \frac{(n-1)}{n^2}(\mathfrak{M}_{p+q}\otimes\mathfrak{M}_r + (\bullet\otimes\mathfrak{M}_q\otimes\bullet)\,\vdots\,\mathfrak{M}_{p+r} + \mathfrak{M}_p\otimes\mathfrak{M}_{q+r}).
\end{aligned}$$

5.2 First moment of the empirical mean

Now we can come back to the location of the empirical Fréchet mean $\bar{x}_n$. To obtain its first moment, we take the expectation of $\log_x(\bar{x}_n)$ with respect to the joint law $\mu^{\otimes n}$ of the sample. Using Eq.13, this gives:

$$\begin{aligned}
\mathbf{E}\left[\,\log_x(\bar{x}_n)\,\right] =& \mathbf{E}\left[\,\mathfrak{X}_1^n\,\right] - \frac{1}{3}\mathbf{E}\left[\,R(\bullet,\mathfrak{X}_1^n)_\bullet\,\vdots\,\mathfrak{X}_2^n\,\right] + \frac{1}{24}\mathbf{E}\left[\,\nabla_\bullet R(\bullet,\mathfrak{X}_1^n)\mathfrak{X}_1^n\,\vdots\,\mathfrak{X}_2^n\,\right] \\
&- \frac{1}{8}\mathbf{E}\left[\,\nabla_{\mathfrak{X}_1^n}R(\bullet,\mathfrak{X}_1^n)_\bullet\,\vdots\,\mathfrak{X}_2^n\,\right] - \frac{1}{12}\mathbf{E}\left[\,\nabla_\bullet R(\bullet,\mathfrak{X}_1^n)_\bullet\,\vdots\,\mathfrak{X}_3^n\,\right] + O(\epsilon^5).
\end{aligned}$$

Then we expand the expectation of the tensor product of empirical moments using the rules of Section 5.1:

$$\begin{aligned}
\mathbf{E}\left[\,\log_x(\bar{x}_n)\,\right] =& \mathfrak{M}_1 - \frac{n-1}{3n}R(\bullet,\mathfrak{M}_1)_\bullet\,\vdots\,\mathfrak{M}_2 - \frac{1}{3n}R(\bullet,\bullet)_\bullet\,\vdots\,\mathfrak{M}_3 \\
&+ \frac{1}{24n^2}\nabla_\bullet R(\bullet,\bullet)_\bullet\,\vdots\,\mathfrak{M}_4 + \frac{(n-1)(n-2)}{24n^2}\nabla_\bullet R(\bullet,\mathfrak{M}_1)\mathfrak{M}_1\,\vdots\,\mathfrak{M}_2 \\
&+ \frac{(n-1)}{24n^2}\left(\nabla_\bullet R(\bullet,\circ)_\circ\,\vdots\,\mathfrak{M}_2\,\overset{\circ}{\vdots}\,\mathfrak{M}_2 + \nabla_\bullet R(\bullet,\mathfrak{M}_1)_\bullet\,\vdots\,\mathfrak{M}_3 + \nabla_\bullet R(\bullet,\bullet)\mathfrak{M}_1\,\vdots\,\mathfrak{M}_3\right) \\
&- \frac{1}{8n^2}\nabla_\bullet R(\bullet,\bullet)_\bullet\,\vdots\,\mathfrak{M}_4 - \frac{(n-1)(n-2)}{8n^2}\nabla_{\mathfrak{M}_1}R(\bullet,\mathfrak{M}_1)_\bullet\,\vdots\,\mathfrak{M}_2 \\
&- \frac{(n-1)}{8n^2}\left(\nabla_\circ R(\bullet,\circ)_\bullet\,\vdots\,\mathfrak{M}_2\,\overset{\circ}{\vdots}\,\mathfrak{M}_2 + \nabla_\bullet R(\bullet,\mathfrak{M}_1)_\bullet\,\vdots\,\mathfrak{M}_3 + \nabla_{\mathfrak{M}_1}R(\bullet,\bullet)_\bullet\,\vdots\,\mathfrak{M}_3\right) \\
&- \frac{(n-1)}{12n}\nabla_\bullet R(\bullet,\mathfrak{M}_1)_\bullet\,\vdots\,\mathfrak{M}_3 - \frac{1}{12n}\nabla_\bullet R(\bullet,\bullet)_\bullet\,\vdots\,\mathfrak{M}_4 + O(\epsilon^5).
\end{aligned}$$

We now remove the terms that vanish because they are the contraction of a symmetric tensor $\mathfrak{M}_k$ with skew symmetric indices of the curvature tensor and we factor similar terms (sometimes changing sign thanks to the skew-symmetry again). We finally find:

Theorem 6 (First moment of the empirical mean) *Let μ be a probability distribution satisfying the KKC or ALC conditions with support of diameter less than ε and mean $\bar{x}$. The first moment of the empirical Fréchet mean $\bar{x}_n$ of an IID n-sample is:*

$$\begin{aligned}\mathbf{E}\left[\log_x(\bar{x}_n)\right] = \mathfrak{M}_1 &- \frac{n-1}{3n} R(\bullet, \mathfrak{M}_1)_\bullet \,\vdots\, \mathfrak{M}_2 \\ &+ \frac{(n-1)(n-2)}{24n^2}\left(\nabla_\bullet R(\bullet, \mathfrak{M}_1)\mathfrak{M}_1 \,\vdots\, \mathfrak{M}_2 - 3\nabla_{\mathfrak{M}_1} R(\bullet, \mathfrak{M}_1)_\bullet \,\vdots\, \mathfrak{M}_2\right) \\ &+ \frac{(n-1)}{12n^2}\left(2\nabla_\circ R(\circ, \bullet)_\bullet \,\vdots\, \mathfrak{M}_2 \,\overset{\circ}{\otimes}\, \mathfrak{M}_2 - (1+n)\,\nabla_\bullet R(\bullet, \mathfrak{M}_1)_\bullet \,\vdots\, \mathfrak{M}_3\right) + O(\epsilon^5).\end{aligned}$$

At the mean $\bar{x}$ of the distribution μ, the first moment $\mathfrak{M}_1(\bar{x}) = 0$ vanishes, so that we end up with the estimation of how the empirical mean $\bar{x}_n$ deviates from the mean $\bar{x}$ of the underlying distribution μ in expectation (i.e. the bias):

$$\mathbf{E}\left[\log_{\bar{x}}(\bar{x}_n)\right] = \frac{(n-1)}{6n^2}\nabla_\bullet R(\bullet, \circ)_\circ \,\vdots\, \mathfrak{M}_2 \,\overset{\circ}{\otimes}\, \mathfrak{M}_2 + O(\epsilon^5). \tag{14}$$

5.3 Second moment of the empirical mean

Let $\Sigma_x^n = \log_x(\bar{x}_n) \otimes \log_x(\bar{x}_n)$ be the tensor product of the empirical mean with itself. From Eq. 13, we can write it as:

$$\Sigma_x^n = \mathfrak{X}_1^n \otimes \mathfrak{X}_1^n - \frac{1}{3}(R(\bullet, \mathfrak{X}_1^n)_\bullet \,\vdots\, \mathfrak{X}_2^n) \otimes \mathfrak{X}_1^n - \frac{1}{3}\mathfrak{X}_1^n \otimes (R(\bullet, \mathfrak{X}_1^n)_\bullet \,\vdots\, \mathfrak{X}_2^n) + O(\epsilon^5).$$

Taking the expectation to obtaining the covariance field, we have to expand the expression of the tensor product of empirical moments:

$$\mathbf{E}\left[\mathfrak{X}_1^n \otimes \mathfrak{X}_1^n\right] = \frac{1}{n}\mathfrak{M}_2 + \frac{n-1}{n}\mathfrak{M}_1 \otimes \mathfrak{M}_1;$$

$$\begin{aligned}\mathbf{E}\left[\mathfrak{X}_1^n \otimes (R(\bullet, \mathfrak{X}_1^n)_\bullet \,\vdots\, \mathfrak{X}_2^n)\right] = &\frac{n-1}{n^2}\Big(\mathfrak{M}_2 \,\overset{\circ}{\otimes}\, \circ \otimes R(\bullet, \circ)_\bullet \,\vdots\, \mathfrak{M}_2 + \bullet \otimes R(\bullet, \mathfrak{M}_1)_\bullet \,\vdots\, \mathfrak{M}_3\Big) \\ &+ \frac{(n-1)(n-2)}{n^2}\mathfrak{M}_1 \otimes R(\bullet, \mathfrak{M}_1)_\bullet \,\vdots\, \mathfrak{M}_2.\end{aligned}$$

Recombining the terms, we get:

$$\begin{aligned}\mathbf{E}\left[\Sigma_x^n\right] = &\mathfrak{M}_2 + \frac{n-1}{n}\mathfrak{M}_1 \otimes \mathfrak{M}_1 - \frac{(n-1)}{3n^2}\mathfrak{M}_2 \,\overset{\circ}{\otimes}\, (\circ \otimes R(\bullet, \circ)_\bullet + R(\bullet, \circ)_\bullet \otimes \circ) \,\vdots\, \mathfrak{M}_2 \\ &- \frac{(n-1)(n-2)}{3n^2}\Big(R(\bullet, \mathfrak{M}_1)_\bullet \,\vdots\, \mathfrak{M}_2 \otimes \mathfrak{M}_1 + \mathfrak{M}_1 \otimes R(\bullet, \mathfrak{M}_1)_\bullet \,\vdots\, \mathfrak{M}_2\Big) \\ &- \frac{n-1}{3n^2}\Big(R(\bullet, \mathfrak{M}_1)_\bullet \,\vdots\, \mathfrak{M}_3 \otimes \bullet + \bullet \otimes R(\bullet, \mathfrak{M}_1)_\bullet \,\vdots\, \mathfrak{M}_3\Big) + O(\epsilon^5)\end{aligned}$$

At the mean $\bar{x}$ of the distribution μ, the first moment $\mathfrak{M}_1(\bar{x}) = 0$ vanishes, so that we end up with the following estimation of the covariance matrix of the empirical mean:

$$\mathbf{E}\left[\log_{\bar{x}}(\bar{x}_n) \otimes \log_{\bar{x}}(\bar{x}_n)\right] = \frac{1}{n}\mathfrak{M}_2 - \frac{(n-1)}{3n^2}\mathfrak{M}_2 \,\overset{\circ}{\otimes}\, (\circ \otimes R(\bullet, \circ)_\bullet + R(\bullet, \circ)_\bullet \otimes \circ) \,\vdots\, \mathfrak{M}_2 + O(\epsilon^5). \tag{15}$$

In coordinates, this write perhaps more simply:

$$\mathbf{E}\left[\log_{\bar{x}}(\bar{x}_n) \otimes \log_{\bar{x}}(\bar{x}_n)\right]^{ab} = \frac{1}{n}\mathfrak{M}_2^{ab} - \frac{(n-1)}{3n^2}\mathfrak{M}_2^{cd}(\mathfrak{M}_2^{ae}R^b_{cde} + R^a_{cde}\mathfrak{M}_2^{be}) + O(\epsilon^5).$$

Theorem 7 (First moments of the empirical mean) *Let μ be a probability distribution satisfying the KKC or ALC conditions with support of diameter less than ε and mean $\bar{x}$. We denote $\mathfrak{M}_k = \mathfrak{M}_k\mu(\bar{x})$ its k-order moment at the mean $\bar{x}$. By definition, we have $\mathfrak{M}_0 = 1$ and $\mathfrak{M}_1 = 0$. Let $X_n = \{x_1, \dots x_n\} \in \mathcal{M}^n$ be an IID n-sample of this distribution. The empirical mean $\bar{x}_n$ of this sample is also unique. Its expected first moment at $\bar{x}$, the bias $Bias(\bar{x}_n) = \mathbf{E}\left[\log_{\bar{x}}(\bar{x}_n)\right]$, is*

$$Bias(\bar{x}_n) = \tfrac{1}{6n}\left(1 - \tfrac{1}{n}\right)\mathfrak{M}_2 \,\overset{\circ}{\circ}\, \nabla_\bullet R(\bullet,\circ)_\circ \,\overset{\bullet}{\bullet}\, \mathfrak{M}_2 + O\left(\epsilon^5\right). \tag{16}$$

Its expected second moment at $\bar{x}$, the covariance $Cov(\bar{x}_n) = \mathbf{E}\left[\log_{\bar{x}}(\bar{x}_n) \otimes \log_{\bar{x}}(\bar{x}_n)\right]$, is:

$$Cov(\bar{x}_n) = \tfrac{1}{n}\left(\mathfrak{M}_2 - \tfrac{1}{3}\left(1 - \tfrac{1}{n}\right)\mathfrak{M}_2 \,\overset{\circ}{\circ}\, (\circ \otimes R(\bullet,\circ)_\bullet + R(\bullet,\circ)_\bullet \otimes \circ) \,\overset{\bullet}{\bullet}\, \mathfrak{M}_2\right) + O\left(\epsilon^5\right). \tag{17}$$

In coordinates, this writes

$$\begin{aligned} Bias(\bar{x}_n)^a &= \tfrac{1}{6n}\left(1 - \tfrac{1}{n}\right)\nabla_b R^a_{cde}\mathfrak{M}_2^{ce}\mathfrak{M}_2^{bd} + O(\epsilon^5) \\ Cov(\bar{x}_n)^{ab} &= \tfrac{1}{n}\big(\mathfrak{M}_2^{ab} - \tfrac{1}{3}\left(1 - \tfrac{1}{n}\right)\mathfrak{M}_2^{cd}(\mathfrak{M}_2^{ae}R^b_{cde} + R^a_{cde}\mathfrak{M}_2^{be})\big) + O(\epsilon^5). \end{aligned}$$

Thus, in general manifolds, there is a bias on the empirical mean of order $1/n$ in the number of samples and of order 4 in ε. Asymptotically, this bias disappear with the number of samples. However, keeping a fixed number of samples, this bias increases and could possibly blow up (outside the domain of validity of our assumptions) when we approach a singularity with an unbounded curvature, in which case the gradient of the curvature has to become large. Of course, the KKC or ALC conditions on the support of the distribution only holds at a sufficient distance of such a singularity for a fixed covariance, which limits the conclusion that we may draw from this trend. Interestingly, such a bias is not visible in symmetric spaces since the curvature tensor is covariantly constant in these manifolds.

The covariance matrix of the empirical mean also has a curvature correction term modulating (accelerating or decelerating) the convergence. In order to better understand the effect of curvature, we study in the next section how this relates to the Bhattacharya-Patrangenaru central limit theorem in Riemannian manifolds, and in Section 7 how the formulas simplify for constant curvature spaces.

6 Asymptotic covariance of the empirical Riemannian mean

The Bhattacharya-Patrangenaru central limit theorem (BP-CLT) for sample Fréchet means in manifolds [6, Theorem 2.1 p. 1230] is a general CLT valid for non-Riemannian twice differentiable distances. We consider here the case of the intrinsic Fréchet mean on a Riemannian manifold [6, Theorem 2.2 p. 1231].

6.1 The Bhattacharya-Patrangenaru CLT

We first align our notations to the ones of [6]. Recall that the support of μ is included in $\mathcal{U} \subset B(x,r)$ with $r < r^*$ by the KKC condition. Because we use a normal coordinate system, a point $y \in \mathcal{U}$ is parametrized by $v = \log_x(y) \in V_x = \log_x(\mathcal{U})$, so that our v is the θ for BP and the BP function $\phi(y)$ is $\log_x(y)$ for us. With our notation $x_v = \phi^{(-1)}(v) = \exp_x(v)$, the pullback of the Riemannian distance to the chart writes: $\left(\rho^\phi(u,v)\right)^2 = \operatorname{dist}(x_u, x_v)^2 = \|l_x(u,v)\|_x^2$. The BP function $\Psi(u;v) = \left(\frac{\partial}{\partial v}\operatorname{dist}(x_u,x_v)^2\right)^\top$ is the Euclidean gradient of this pulled back squared distance with respect to v. The BP matrix $D_v\Psi(u;v)$ is second order derivative of the pulled back squared distance $v \to \operatorname{dist}(x_u,x_v)^2$ in our chart. In the original BP-CLT theorem, the covariance of the normal law is $\Lambda^{(-1)}\,\Sigma\,\Lambda^{(-1)}$, where Σ is the covariance of $\Psi(u;v)$ and the matrix $\Lambda = \mathbf{E}\left[D_v\Psi(\log_x(x_i);v)\right]$ is the expected value of the second order derivative under the law μ of the sample x_i.

In a normal coordinate system centered at $\bar{x}$, the Christoffel symbols and their derivatives vanish at $\bar{x}$ so that the standard differential corresponds to the Riemannian gradient and the standard second order derivative corresponds to the Riemannian Hessian. Let $d_y^2(x) = \mathrm{dist}^2(y,x)$. For $y \notin \mathrm{Cut}(x)$, we have $\nabla d_y^2(x) = -2\log_x(y)$ and the Hessian is $H_x(y) = \nabla^2 d_y^2(x) = -2D_x \log_x(y)$ (see for instance [33, appendix A]). Thus, we have $\Psi(u;0) = \nabla d_{x_v}^2(x_u)|_{v=0} = -2\log_{\bar{x}}(x_u) = -2u$ and its covariance is simply $\Sigma = 4\int_{\mathcal{M}} \log_{\bar{x}}(y)\log_{\bar{x}}(y)^\top \,\mu(dy) = 4\mathfrak{M}_2$. For the second order derivatives, we have $D_v\Psi(\log_{\bar{x}}(y);v)|_{v=0} = H_{\bar{x}}(y)$ and its expected value for y following the law μ is $\Lambda = \bar{H} = \int_{\mathcal{M}} H_{\bar{x}}(y)\,\mu(dy)$. These alignments of notations lead to the following formulation of the BP-CLT for intrinsic means.

Theorem 8 (BP-CLT for intrinsic sample means [6]) *Let μ is a probability distribution on a Riemannian manifold $(\mathcal{M},g)$ of Fréchet mean $\bar{x}$ and covariance matrix $\mathfrak{M}_2 = \mathfrak{M}_2\mu(\bar{x})$ whose support is included in a regular geodesic ball $B(x_0,r)$ satisfying the KKC conditions.*

Then (a) the empirical Fréchet mean $\bar{x}_n$ of a sample $X_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ is a consistent estimator of the population Fréchet mean $\bar{x}$ of μ and (b) the random variable $\sqrt{n}\log_{\bar{x}}(\bar{x}_n) \in T_{\bar{x}}\mathcal{M}$ converges in law to a normal distribution of mean 0 and covariance $4\bar{H}^{(-1)}\,\mathfrak{M}_2\,\bar{H}^{(-1)}$, where the matrix $\bar{H}$ is the expectation of the Riemannian Hessian of the squared distance $\mathrm{dist}(.,y)^2$ *according to the distribution μ.*

In order to compare our non-asymptotic high concentration expansion (Theorem 7) to the BP-CLT, we first observe that the consistency of the empirical Fréchet mean of the BP-CLT states that the limit of the bias in Eq.(16) should vanish when n goes to infinity. This is indeed the case since $\mathrm{Bias}(\bar{x}_n) = \mathbf{E}\left[\log_{\bar{x}}(\bar{x}_n)\right]$ is of order $1/n$. To compare the covariance, we essentially need to compute the Hessian of the squared Riemannian distance $H_x(y)$, integrate it to get its expectation $\bar{H}$, and show that $4\bar{H}^{(-1)}\mathfrak{M}_2\bar{H}^{(-1)}$ has the same Taylor expansion as in Eq.(17).

6.2 Hessian of the squared Riemannian distance

The Taylor expansion of $H_x(x_u)$ has been established in Eq.(1) of [33] using Brewin Taylor expansion[3]:

$$\frac{1}{2}[H_x(x_u)]_b^a = -[D_x\log_x(x_u)]_b^a = \delta_b^a + \frac{1}{3}R_{cdb}^a u^c u^d + \frac{1}{12}\nabla_c R_{deb}^a u^c u^d u^e + O(4).$$

We verify below that we obtain the same formulation using the Gavrilov's Taylor expansions used in the current paper. The Taylor expansion of the squared distance in a normal coordinate system at x is given by Eq.(9):

$$\mathrm{dist}^2(x_v,x_u) = (u-v)^a(u-v)_a + \frac{1}{3}R_{abcd}v^a u^b u^c v^d + \frac{1}{12}\nabla_e R_{abcd}v^a u^b u^c v^d (u+v)^e + O(6).$$

Taking the derivative with respect to the coordinate α of v and raising the index α, we get the gradient:

$$\Psi(u,v) = -2(u-v)^\alpha + \frac{2}{3}R_{bcd}^\alpha u^b u^c v^d + \frac{1}{6}\nabla_e R_{bcd}^\alpha u^b u^c v^d (u+v)^e + \frac{1}{12}\nabla^\alpha R_{ebcd}v^e u^b u^c v^d + O(5).$$

Taking the derivative with respect to the coordinate β of v, we have:

$$D_\beta\Psi(u,v) = 2\delta_\beta^\alpha + \frac{2}{3}R_{bc\beta}^\alpha u^b u^c + \frac{1}{6}\nabla_e R_{bc\beta}^\alpha u^b u^c (u+v)^e + \frac{1}{6}\nabla_\beta R_{ecd}^\alpha u^e u^c v^d + \frac{1}{6}\nabla^\alpha R_{ebc\beta}v^e u^b u^c + O(4)$$

The value at $v=0$ is $\frac{1}{2}[H_x(x_u)]_\beta^\alpha = \delta_\beta^\alpha + \frac{1}{3}R_{bc\beta}^\alpha u^b u^c + \frac{1}{12}\nabla_e R_{bc\beta}^\alpha u^b u^c u^e + O(4)$, in accordance with the previous expansion of [33]. The expectation of this Hessian matrix when x_u has probability μ is:

$$\frac{1}{2}\bar{H}_b^a = \delta_b^a + \frac{1}{3}R_{dcb}^a\mathfrak{M}_2^{dc} + \frac{1}{12}\nabla_e R_{dcb}^a\mathfrak{M}_3^{dce} + O(\varepsilon^4)$$

[3] We indicate here the order $O(4)$ corresponding to the notations of the current paper since the third order of Brewin expansions has a different meaning which includes a conformal factor accounting for curvature.

6.3 Covariance matrix of the BP-CLT

The inverse of the matrix $\bar{H}$ can be determined by identification in $[\bar{H}^{(-1)}]^a_b[\bar{H}]^b_c = \delta^a_c$:

$$[\bar{H}^{(-1)}]^a_b = \frac{1}{2}\left(\delta^a_b - \frac{1}{3}R^a_{dcb}\mathfrak{M}_2^{dc} - \frac{1}{12}\nabla_e R^a_{dcb}\mathfrak{M}_3^{dce}\right) + O(\varepsilon^4).$$

Finally, we get

$$4[\bar{H}^{(-1)}\mathfrak{M}_2\bar{H}^{(-1)}]^{ab} = 4[\bar{H}^{(-1)}]^a_c[\mathfrak{M}_2]^{cd}[\bar{H}^{(-1)}]^b_d = \mathfrak{M}_2^{ab} - \frac{1}{3}\mathfrak{M}_2^{ef}\left(R^a_{efc}\mathfrak{M}_2^{cb} + \mathfrak{M}_2^{ad}R^b_{efd}\right) + O(\varepsilon^5),$$

which is in accordance with the term in $1/n$ of Eq.17 of Theorem 7.

7 Modulation of the rate of convergence in space forms

In order to verify experimentally our predicted bias and rate of convergence, it is better to have explicit formulas for geodesics, which are known only for symmetric spaces. In these spaces, the curvature is covariantly constant, so that the empirical mean $\bar{x}_n$ of an IID n-sample has no measurable bias even for small sample sizes up to order 5. Thus, the only visible impact of the curvature is on the covariance of the empirical mean. To simplify further the setup and to minimize the number of parameters, we focus more particularly on constant curvature spaces, which include the Euclidean space (sectional curvature $\kappa = 0$), the sphere of radius R (positive sectional curvature $\kappa = 1/R^2$) and the hyperbolic space of negative sectional curvature $\kappa < 0$, which can be viewed as a pseudo-sphere of radius $R = -1/\sqrt{|\kappa|}$ in the Minkowski space. Because the space is symmetric, the expected

7.1 Asymptotic BP-CLT for isotropic distributions

The Hessian of the squared distance in constant curvature spaces has been computed in closed form in [5] using Jacobi fields. The interested reader may also find in [33] a more pedestrian approach using the embedding of the sphere (resp. the hyperbolic space) in the Euclidean space (resp. the Minkowski space) that obtains similar formulas for $\kappa = \pm 1$. With the notations $\log_x(y) = \theta u$ where $\theta = \|\log_x(y)\|_x$ is the distance from x to y and $u = \log_x(y)/\theta$ is the unit vector of $T_x\mathcal{M}$ pointing from x to y, the Hessian of the squared distance is $\frac{1}{2}H_x(y) = uu^\top + f_\kappa(\theta)(\operatorname{Id} - uu^\top)$ with

$$f_\kappa(\theta) = \begin{cases} \sqrt{|\kappa|}\theta\coth(\sqrt{|\kappa|}\theta) & \text{if } \kappa < 0, \\ 1 & \text{if } \kappa = 0, \\ \sqrt{\kappa}\theta\cot(\sqrt{\kappa}\theta) & \text{if } \kappa > 0. \end{cases}.$$

We can unify the notations for all curvatures by observing that $f_\kappa(\theta) = h(\kappa\theta^2)$ where $h(t) = \sqrt{t}\cot(\sqrt{t})$. This function is analytic at 0: its Taylor expansion is $h(t) = 1 - t/3 + O(t^2)$. Thus, the formulation is valid with positive, null and negative curvature:

$$\frac{1}{2}H_x(y) = uu^\top + h(\kappa\theta^2)(\operatorname{Id} - uu^\top) \qquad \text{with} \qquad h(t) = \sqrt{t}\cot(\sqrt{t}) \tag{18}$$

When the point y follows an isotropic distribution at x (i.e. circularly symmetric in the tangent space $T_x\mathcal{M}$), the distribution of u is uniform on the unit sphere of dimension $d-1$, where d is the dimension of the manifold. It is also independent of the distribution $dP(\theta)$ on the distance. This means that $\mathbf{E}\left[\, uu^\top \,\right] = \frac{1}{d}\operatorname{Id}$. Thus, we are left with the simple expected Hessian $\bar{H} = 2\gamma\operatorname{Id}$ with

$$\gamma = \frac{1}{d} + \left(1 - \frac{1}{d}\right)\bar{h} \qquad \text{where} \qquad \bar{h} = \mathbf{E}\left[\, h(\kappa\operatorname{dist}(\bar{x}, .)^2)\,\right] = \int_{\mathcal{M}} h(\kappa(\operatorname{dist}(\bar{x}, y)^2)\,\mu(dy). \tag{19}$$

The BP-CLT tells us that the covariance matrix of the sample mean is:

$$\mathrm{Cov}(\log_{\bar{x}}(\bar{x}_n)) = \gamma^{-2}\left(\frac{1}{n}\mathfrak{M}_2\right) + O\left(\frac{1}{n^2}\right). \tag{20}$$

Because the distribution is isotropic, the covariance matrix $\mathfrak{M}_2$ at the mean point x is diagonal, as well as the resulting covariance on the empirical mean. Thus, the formula boils down to a scalar equation $\mathrm{Var}(\log_{\bar{x}}(\bar{x}_n)) = \gamma^{-2}\frac{\sigma^2}{n} + O\left(\frac{1}{n^2}\right)$ using the variance $\sigma^2 = \mathrm{Tr}(\mathfrak{M}_2)$ of the original distribution. In this formula, we see that the factor $\alpha = \gamma^{-2}$ plays the role of a modulation factor indicating how the rate of convergence of the covariance (or variance) differs from the Euclidean case.

7.2 Non-asymptotic high concentration expansion

In constant curvature spaces, the Riemannian curvature tensor is given by $R_{abcd} = \kappa\,(g_{ac}g_{db} - g_{ad}g_{cb})$. Equivalently the $(1,3)$ curvature tensor is $R^a_{bcd} = g^{ae}R_{ebcd} = \kappa\,(\delta^a_c g_{db} - \delta^a_d g_{cb})$. Thus, in a normal coordinate system where $g_{ab} = \delta_{ab}$:

$$\begin{aligned}\mathfrak{M}_2^{cd}(\mathfrak{M}_2^{ae}R^b_{cde} + R^a_{cde}\mathfrak{M}_2^{be}) =&\kappa\,\mathfrak{M}_2^{cd}\mathfrak{M}_2^{be}(\delta^a_d g_{ec} - \delta^a_e g_{dc}) + \kappa\,\mathfrak{M}_2^{cd}(\delta^b_d g_{ec} - \delta^b_e g_{dc})\mathfrak{M}_2^{ae}\\ =&\kappa\,(\mathfrak{M}_2^{ca}g_{ec}\mathfrak{M}_2^{be} - \mathfrak{M}_2^{cd}g_{dc}\mathfrak{M}_2^{ba} + \mathfrak{M}_2^{cb}g_{ec}\mathfrak{M}_2^{ae} - \mathfrak{M}_2^{cd}g_{dc}\mathfrak{M}_2^{ab})\\ =&2\kappa([\mathfrak{M}_2\mathfrak{M}_2]^{ab} - \mathrm{Tr}(\mathfrak{M}_2)\mathfrak{M}_2^{ab})\end{aligned}$$

Thus, Eq.(17) of Theorem 16 states that:

$$\mathrm{Cov}(\bar{x}_n) = \tfrac{1}{n}\mathfrak{M}_2\big(\,\mathrm{Id} - \tfrac{\kappa}{3}\left(1-\tfrac{1}{n}\right)(\mathfrak{M}_2 - \mathrm{Tr}(\mathfrak{M}_2)\,\mathrm{Id})\big) + O\big(\epsilon^5\big)$$

With an isotropic distribution of variance $\sigma^2 = \mathrm{Tr}(\mathfrak{M}_2)$ (i.e. $\mathfrak{M}_2 = \sigma^2/d\,\mathrm{Id}$), the covariance matrix of the empirical mean is diagonal and this expression boils down to:

$$\mathrm{Var}(\bar{x}_n) = \tfrac{\sigma^2}{n}\big(1 + \tfrac{2\kappa\sigma^2}{3}\left(1-\tfrac{1}{d}\right)\left(1-\tfrac{1}{n}\right)\big) + O\big(\epsilon^5\big).$$

We can summarize the results in the following theorem.

Theorem 9 (Modulation of the convergence speed by the curvature in space forms) *We consider an isotropic distribution of mean $\bar{x}$ and variance σ^2 on a space of constant sectional curvature κ, whose support satisfies the KKC conditions. Because the space is symmetric, the bias of the empirical Fréchet mean $\bar{x}_n$ of an IID n-sample vanishes for small sample sizes at order 5:* $\mathbf{E}\,[\,\log_{\bar{x}}(\bar{x}_n)\,] = 0 + O\big(\epsilon^5\big)$.

The non-asymptotic variance of the empirical Fréchet mean of an IID n-sample with sufficiently small variance $\sigma^2 < \varepsilon$ is:

$$Var(\bar{x}_n) = \alpha\frac{\sigma^2}{n} \qquad \textit{with} \qquad \alpha = \big(1 + \tfrac{2}{3}\kappa\sigma^2\left(1-\tfrac{1}{d}\right)\left(1-\tfrac{1}{n}\right)\big) + O\big(\epsilon^5\big). \tag{21}$$

The asymptotic variance of the empirical Fréchet mean of an IID n-sample is:

$$Var(\bar{x}_n) = \alpha\frac{\sigma^2}{n} \qquad \textit{with} \qquad \alpha = \left(\tfrac{1}{d} + \left(1-\tfrac{1}{d}\right)\overline{h}\right)^{-2} + O\left(\tfrac{1}{n^2}\right), \tag{22}$$

where

$$\overline{h} = \mathbf{E}\left[\,h(\kappa\,\mathrm{dist}(\bar{x},.)^2)\,\right] \qquad \textit{with} \qquad h(t) = \sqrt{t}\coth(\sqrt{t}). \tag{23}$$

In both expansions, the modulation factor $\alpha = Var(\bar{x}_n)\frac{n}{\sigma^2}$ indicates how much the variance of the empirical Fréchet mean deviates from the Euclidean case, i.e. how much the convergence speed is modulated by the curvature of the space: a modulation factor $\alpha > 1$ indicates that the convergence is slower than in the Euclidean case, while $\alpha < 1$ indicates a faster convergence.

A first observation is that there is no modulation for $n = 1$ in the non-asymptotic high concentration expansion. Indeed, the mean of one sample is the sample itself, so that the variance of the empirical mean is the one of the sample. Note that this feature is not shared by the large sample CLT approximation.

The second observation is that there is no modulation either in dimension 1 for both expansions, which is expected since there is no intrinsic curvature in that case. For higher dimensions, positive sectional curvature induces an increase of the dispersion of the empirical mean which slows down the convergence of the law of large number with respect to the Euclidean case. A negative sectional curvature accelerates the convergence of the law of large number with respect to the Euclidean case, as we will see more clearly below. At first sight, this modulation of the rate of convergence could seem to be related to the smeary means of [15]. However, the Taylor expansion that we get is still in $1/n$. Moreover, there is no mechanism in our expansion to create a term in $1/n^\alpha$ with $\alpha < 1$ in the series. The two phenomena are thus different.

7.3 Archetypal modulation factor

In order to give a more intuitive idea of the impact of the curvature, we consider a uniform distribution on the Riemannian hypersphere of radius θ around the point x. Such a distribution maximizes the variance among isotropic distributions with support in a closed geodesic ball, and its singular distribution on θ allows us to compute the integral of the Hessian of the distance in closed form. Moreover, the covariance matrix is $\mathfrak{M}_2 = \theta^2/d\,\mathrm{Id}$ so that the variance is θ^2. The modulation factor for a large number n of samples is then:

$$\alpha = \left(\tfrac{1}{d} + \left(1 - \tfrac{1}{d}\right) h(\kappa\theta^2)\right)^{-2} + O\left(\tfrac{1}{n}\right) = h(\kappa\theta^2)^{-2} + O(\tfrac{1}{d}) + O(\tfrac{1}{n}).$$

The archetypal values of this modulation factor are obtained for positive or negative curvature with a large number of samples in a manifold of large dimension d:

$$\tilde{\alpha} = h(\kappa\theta^2)^{-2} = \frac{\tan^2(\sqrt{\kappa\theta^2})}{\kappa\theta^2}. \tag{24}$$

This archetypal modulation factor is displayed in Figure (2). The Taylor expansion of the modulation factor for a small variance shows that the modulation factor is smooth at 0 and corresponds exactly to our non-asymptotic high concentration prediction:

$$\alpha = 1 + \tfrac{2}{3}\left(1 - \tfrac{1}{d}\right)\kappa\theta^2 + O(\theta^4, \tfrac{1}{n}).$$

It is important to notice that the variable controlling the modulation is actually $\kappa\theta^2$, i.e. the product of the sectional curvature with the variance.

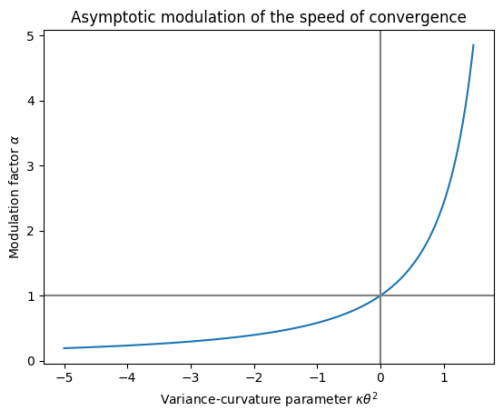

Figure 2: Archetypal modulation $\tilde{\alpha} = h(\kappa\theta^2)^{-2}$ of the speed of convergence as a function of the curvature-variance parameter for a large dimension d and a large sample n.

For a positive variance-curvature, the modulation of the rate of convergence is above one (slower convergence than in Euclidean spaces) and actually goes to infinity when $\kappa\theta^2$ approaches $\pi^2/4$. Interestingly, this corresponds to the Kendall & Karcher concentration conditions under which all this paper is restricted since we should have $\theta < \frac{\pi}{2\sqrt{\kappa}}$ (Eq. 1). Actually, as we will see below with experiments, there are examples on the sphere with $\theta = \frac{\pi}{2\sqrt{\kappa}}$ (namely a uniform distribution on the equator) where the distribution of the empirical Fréchet mean converges to a mixture of Diracs rather than concentrating on a point as usual.

For negative curvature, our formula indicates a modulation factor below 1, meaning that the variance of the estimated mean decreases faster than in Euclidean spaces. Such a phenomenon

was observed in specific cases of other types of means in negatively curved manifolds [3, 22]. However, it was apparently not recognized so far as a general phenomenon of least-squares in manifolds. For an infinite negative variance-curvature, the modulation factor actually goes to zero. This effect suggests that we could see here the beginning of one phenomenon related to the stickiness of the Fréchet mean described in stratified spaces [23]. In very specific stratification cases such as corners of positive or negative curvature, one can indeed interpret the singularity with infinite curvature as the limit of a smooth manifold whose curvature is concentrating at point. In such a process, not only the curvature is becoming (positively or negatively) infinite, but the gradient is also becoming very large, which can attract or repulse the empirical Fréchet mean from the singularity. Of course, the formula that we present here is only an approximation, so that we cannot conclude anything on this basis for large variances. However, the modulation of the convergence speed combined with the bias appear to be significant elements departing from the Euclidean situation that might partly explain the sticky mean in the range of very concentrated data on smooth affine connection or Riemannian manifolds.

7.4 Experiments on the 2-dimensional sphere

In order to illustrate very practically the effect of curvature on the convergence of the empirical mean estimation, we can design a very simple experiment on the sphere $\mathcal{S}_2$ embedded in $\mathbb{R}^3$. We consider a uniform distribution on the hypersphere of radius θ around the north pole (for $\theta \in [0, \pi]$). In the embedding 3D space, this corresponds to the horizontal small circle of radius $\sin\theta$ centered at the point $(0, 0, \cos\theta)$. Such a distribution has two advantages: it is very symmetric, which considerably simplifies the computations, and it maximizes the variance among isotropic distributions with a prescribed support. Moreover, continuous or discrete distribution which entirely lie in one hemisphere (excluding the equator) are know to have a unique Fréchet mean located in that hemisphere [10], which is in our case at the north or the south pole.

In order to see that with simple arguments, we compute the variance at any point on the sphere. By symmetry of our distribution, it only depends on the latitude of the point at which we compute it. Here, we count the latitude ϕ from 0 at the north pole to π at the south pole. Since the Riemannian distance is the angle between points on the sphere, the variance is:

$$\mathrm{Var}(\theta, \phi) = \int_{-\pi}^{\pi} \arccos^2 \left(\sin\theta \sin\phi \cos\alpha + \cos\theta \cos\phi\right) \frac{d\alpha}{2\pi}.$$

There is generally no closed-form expression for this integral. We display in Fig.3 its value as a function of the latitude for different values of θ. We clearly see that the the north pole is the unique population Fréchet mean for $\theta \in [0, \pi/2[$, with a variance $\mathrm{Var}(e_3) = \theta^2$ and a covariance matrix $\mathrm{Cov}(e_3) = \sigma^2\,\mathrm{Id}$ with $\sigma^2 = \theta^2/2$.

For low values of θ, the variance is monotone and the south pole is a global maximum. For a value of θ larger than about 1.2 radian, a maximum appears in the south hemisphere and the south pole becomes a local minimum. For a uniform distribution on the equator ($\theta = \pi/2$), the north and south poles become global minima with variance $\pi^2/4$, while the equator is maximizing the variance with the value $\pi^2/3$. For $\theta \in]\pi/2; \pi]$, the roles of the north and south pole are exchanged, so that we restrict in the sequel to $0 \leq \theta \leq \pi/2$.

We now draw an empirical distribution of n points uniformly sampled on the horizontal small circle of radius $\sin\theta$. With the classical Gauss-Newton gradient descent algorithm $\bar{x}^{t+1} = \exp_{\bar{x}^t}(\frac{1}{n}\sum_i \log_{\bar{x}^t}(x_i))$, we compute the empirical Fréchet mean of this sample. In order to avoid the convergence problems due to potentially null gradients at maximima or saddle points, we initialize the algorithm with the north pole, which is the unique Fréchet mean when $\theta < \pi/2$. Averaging the square distance of the result to the north pole for a large number N of repeated random sampling allows us to compute a stochastic integral of the variance $\mathrm{Var}(n, \theta) = \mathrm{Var}(\bar{x}_n)$. Finally, we compute the normalized modulation factor $\alpha(n, \theta) = n\mathrm{Var}(n, \theta)/\theta^2$ which indicates how the rate of convergence differs from the Euclidean case. Results are plotted in

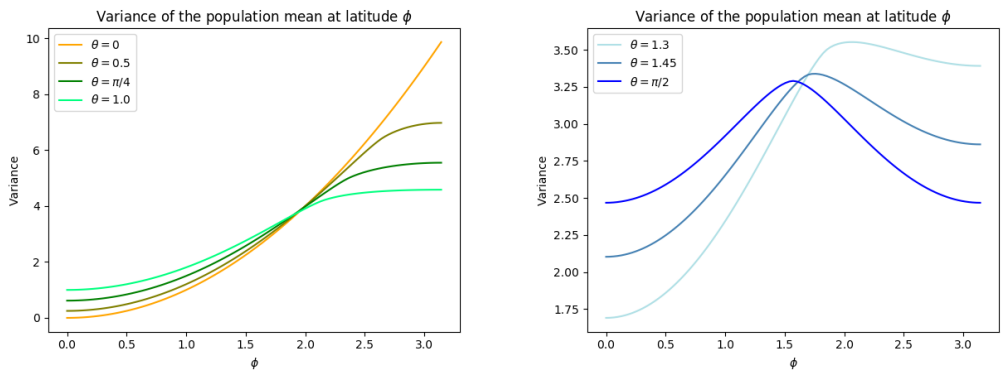

Figure 3: Variance of the distribution on the circle of radius $\sin\theta$ at a point of latitude ϕ (counted from 0 at the north pole to π at the south pole) on the sphere $\mathcal{S}_2$.

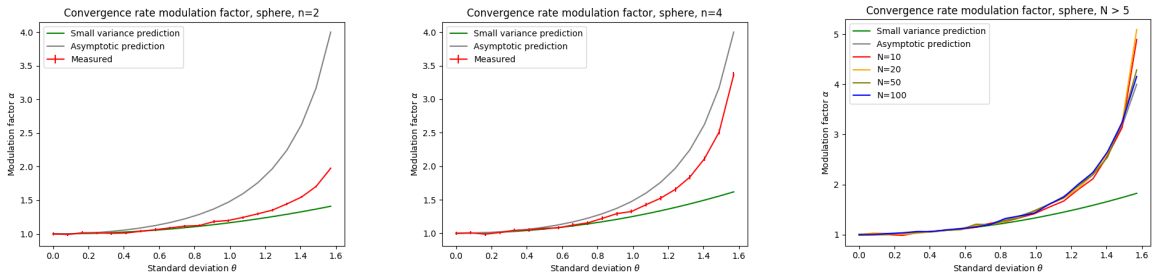

Figure 4: Predicted (green curve) versus measured modulation factor $\alpha(n,\theta) = n\mathrm{Var}(n,\theta)/\theta^2$ on the speed of convergence of the empirical Fréchet mean to the population Fréchet mean on the sphere $\mathcal{S}_2$, computed with $N = 5000$ drawings to approximate the expectation for a number of samples $n = 2$, $n = 4$ and $n = 10$ to $n = 100$.

Figure 4 for several values of n and compared to the value predicted by our non-asymptotic high concentration formula Eq.(21) (green curve) and by the asymptotic CLT formula Eq.(22) (grey curve). We see that the non-asymptotic high concentration expansion closely predicts the normalized modulation factor for $\theta < 0.8rad$, whatever the number of samples. Above this value, the neglected terms in $O(\theta^5)$ take the lead and increase to a maximum which depends on how close we are to $\theta = \pi/2$. In this formula, the influence of the number n of sample points is visible for a very small value, but disappears for 5 to 10 points both in the predicted and measured modulation factors (right of Fig.4). For a small number of samples, the asymptotic CLT formula is significantly overestimating the modulation factor for a large range of values of θ, while is become a very good predictor on almost all the range of values for $n \geq 10$.

Very similar results are obtained for a uniform distribution on the hypersphere of radius θ on the sphere $\mathcal{S}_3$ around the north pole $(0,0,0,1)$ as shown in Fig. 5 and in higher dimensions. The more we increase the dimension, the more the asymptotic BP-CLT is overestimating the modulation factor for a small number of samples. However, the asymptotic approximation remains quite good for a large number of samples even close to $\theta = \pi/2$.

7.5 Experiments on the hyperbolic space

A similar experiment can be made on the hyperbolic space $\mathbb{H}_3$. We consider the positive sheet $(t > 0)$ of the hyperboloid of equation $x^2 + y^2 + z^2 - t^2 = -1$ embedded in $\mathbb{R}^4$. Using the pseudo-metric $\|(x,y,z,t)\|_*^2 = x^2 + y^2 + z^2 - t^2$, the hyperbolic space $\mathbb{H}_3$ can be seen as the pseudo-sphere of radius -1 in the Minkowski space $\mathbb{R}^{3,1}$. With these conventions, geodesics are the trace of 2-planes passing through the origin and the Riemannian distance is the arc-length

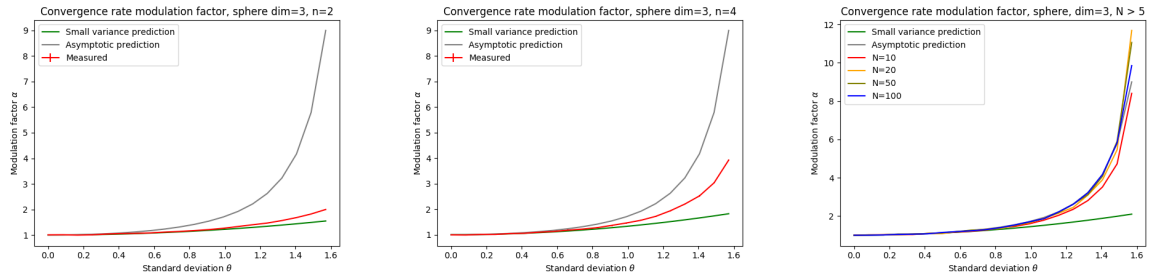

Figure 5: Predicted (green curve) versus measured modulation factor $\alpha(n,\theta) = n\mathrm{Var}(n,\theta)/\theta^2$ on the speed of convergence of the empirical Fréchet mean to the population Fréchet mean on the sphere $\mathcal{S}_3$, computed with $N = 5000$ drawings to approximate the expectation for a number of samples $n = 2$, $n = 4$ and $n = 10$ to $n = 100$.

$d(x,y) = \mathrm{arccosh}(-\langle x \mid y \rangle_*)$.

As in the previous case, we consider a uniform distribution on the hypersphere of radius $\theta \in \mathbb{R}^+$ centered at $(0,0,0,1)$ in $\mathbb{H}_3$. In the embedding space $\mathbb{R}^4$, this hypersphere in $\mathbb{H}_3$ is the intersection of the hyperplane $t = \cosh\theta$ with the hyperboloid. It is thus simply a sphere of radius $\sinh\theta$ in this hyperplane. A uniform distribution on the hypersphere is thus easily obtained from a uniform distribution on the 3-sphere. Because the hyperbolic space $\mathbb{H}_3$ has negative curvature and is Hadamard, there exists a unique Fréchet mean which is the center of the hypersphere $\bar{x} = (0,0,0,1)$. However, the negative curvature may also cause the classical Gauss-Newton gradient descent algorithm to diverge when the variance of the distribution is large (see [7] for an example on SPD matrices with the affine-invariant metric). The algorithm may simply be modified with an adaptive Levenberg-Marquardt time-step $\tau < 1$ in the formula $\bar{x}^{t+1} = \exp_{\bar{x}^t}(\tau \frac{1}{n} \sum_i \log_{\bar{x}^t}(x_i))$.

We draw an empirical distribution of n points uniformly sampled on the hypersphere of radius θ. We compute the square geodesic distance between the pole $\bar{x} = (0,0,0,1)$ and the obtained empirical Fréchet mean $\bar{x}_n$of this sample. Averaging this value for a large number N of repeated random sampling allows us to compute a stochastic integral of the variance at the Fréchet mean $\mathrm{Var}(n,\theta) = \mathrm{Var}(\bar{x}_n)$. Finally, we compute the normalized modulation factor $\alpha(n,\theta) = n\mathrm{Var}(n,\theta)/\theta^2$ which indicates how the rate of convergence differs from the Euclidean case.

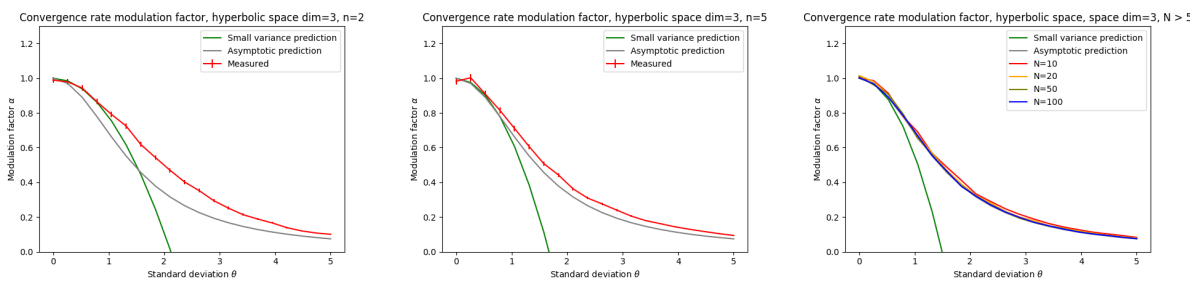

Figure 6: Predicted (green curve) versus measured modulation factor $\alpha(n,\theta) = n\mathrm{Var}(n,\theta)/\theta^2$ on the speed of convergence of the empirical Fréchet mean to the population Fréchet mean on the hyperbolic space $\mathbb{H}_3$, computed with $N = 5000$ drawings to approximate the expectation. Top raw: curve along $\theta = [0;5]$ for a number of samples $n = 2$, $n = 5$ and $n = 10$ to $n = 100$.

Results are displayed in Figure 6 for several values of θ and n (red curve) and compared to the value predicted by our non-asymptotic high concentration formula Eq.(21) (green curve) and by the asymptotic CLT formula Eq.(22) (grey curve). We see that the non-asymptotic high

concentration expansion closely predicts the normalized modulation factor for $\theta < 1$, whatever the number of samples. Above this value, the terms in $O(\theta^5)$ neglected in the Taylor expansion take the lead and the formula gets useless. The asymptotic CLT formula is significantly underestimating the modulation factor for the whole range of values of θ for a small number of samples. It becomes a very good predictor on all the range of values for $n \geq 10$. In the presented experiments, we limited the radius θ to 6 for visualization purposes. However, we have observed a modulating factor as low as $\alpha = 0.01$ for $\theta = 15$ (i.e. an acceleration of the convergence by two orders of magnitude with respect to the Euclidean case) for two points. It is also interesting to notice that the number n of points in the sample has a very limited influence on the modulation factor for $n \geq 10$: the main factor is visibly the variance θ^2 of the sample.

8 Discussion

We have derived in this paper a new type of approximation of the moments of the empirical mean $\bar{x}_n$ of an IID n-sample of a sufficiently concentrated distribution on a Riemannian and affine connection manifolds. This high concentration expansion of the moments of the empirical mean shows a bias in $1/n$ which is directly proportional to the gradient of the curvature tensor contracted twice with the covariance matrix. This unexpected bias, apparently never described before, is important in the small sample regime. The high concentration expansion of the covariance matrix of the empirical mean has also a curvature correction term modulating (accelerating or decelerating) the convergence with respect to the Euclidean case. The main variable controlling the modulation is the covariance-curvature tensor $R(\bullet, \circ)\bullet : \mathfrak{M}_2$ (the matrix $A^i_j = R^i_{klj}\mathfrak{M}_2^{kl}$ in coordinates).

Our new high-concentration expansion is valid in the small sample regime. Thus, it is a useful complement to the Bhattacharya-Patrangenaru central-limit theorem, which gives an asymptotic expansion for a large number of samples. With our notations, the BP-CLT states that the empirical mean $\bar{x}_n$ converges in expectation to the true mean $\bar{x}$ with a rate of at least $1/\sqrt{n}$ and with covariance matrix (at the true mean) $\mathrm{Cov}(\bar{x}_n) = \frac{1}{n} 4 \bar{H}^{(-1)} \mathfrak{M}_2 \bar{H}^{(-1)}$, where $\bar{H}$ is the expected Hessian of the Riemannian square distance (according to the distribution). The high concentration expansion of this expected Hessian $\bar{H} = 2\,\mathrm{Id} - \frac{2}{3} R(\bullet, \circ)\bullet : \mathfrak{M}_2 + O(\varepsilon^3)$ is controlled once again by the covariance-curvature tensor, as already observed in [33], so that both expansions are consistent in the high-concentration asymptotic regime.

In constant curvature spaces, we showed that both the asymptotic BP-CLT and the non-asymptotic high-concentration expansion predict a deviation of the decrease of the covariance of the empirical mean with respect to the Euclidean case. This modulation of the convergence speed can be encoded with a single multiplicative factor α which indicates that the variance of the Fréchet mean decreases faster in negatively curved space forms than in the Euclidean case, while it decreases more slowly in positive curvature space forms. The archetypal modulation factor goes to zero for an infinite negative curvature. This suggests that we could see here the beginning of the stickiness of the Fréchet mean described in stratified spaces. On the contrary, the modulation factor goes to infinity when we approach the limits of the Karcher & Kendall concentration conditions with a uniform distribution on the equator of the sphere, for which the Fréchet mean is not a single point anymore. Although an explicit formula has previously been established for the BP-CLT in space forms, it seems that the interpretation of the influence of the curvature of empirical means got unnoticed so far.

Acknowledgements. This project has received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation program (grant G-Statistics agreement No 786854). Sincere thanks Yann Thanwerdas and Nicolas Guigui for their proofreading of the manuscript.

A Equivalence of notations with Darling's paper

In the process of establishing the foundations of Kalman filtering on manifolds [13], Darling derived a coordinate free approximation of the exponential barycenter estimation in a manifold. The work of this research report was apparently unpublished in a journal, which is very unfortunate as it got completely unnoticed.

His Exponential Barycenter Formula Eq.(26) (Section 3.2, p.12) actually corresponds to the order 3 of our formula Eq.(12). However, the 4th order necessary to see that there is a bias on the empirical mean was not derived. Rephrased in the notations of this, his formula reads

$$\log_x(\bar{x}) = \mathfrak{M}_1 - \frac{1}{3} R_{ijk} \mathfrak{M}_1^i \mathfrak{M}_2^{jk} + O(\varepsilon^4).$$

However, the convention for the sign of the Riemannian curvature tensor is the opposite of ours. Indeed, Darling defines

$$R(u,v)w = D\Gamma(v)(w \otimes u) - D\Gamma(u)(w \otimes v) + \Gamma(\Gamma(w \otimes u) \otimes v) - \Gamma(\Gamma(w \otimes v) \otimes u).$$

Taking $w = \partial_b$, $u = \partial_c$, $v = \partial_d$, we obtain

$$R(\partial_c, \partial_d)\partial_b = \partial_d \Gamma_{bc} - \partial_c \Gamma_{bd} + \Gamma_{ed}\Gamma^e_{bc} - \Gamma_{ec}\Gamma^e_{bd}.$$

Given that the connection is torsion-free, we have $\Gamma_{ij} = \Gamma_{ji}$, and we see that this formula is the opposite convention of the one used in this paper in Eq.2.

The index of the Riemannian curvature tensor also differ from ours. Here $R_{ijk} = R(\partial_i, \partial_j)\partial_k$, so that $R_{ijk}\mathfrak{M}_1^i \mathfrak{M}_2^{jk} = R(\mathfrak{M}_1, \bullet)_\bullet \vdots \mathfrak{M}_2$. Taking the opposite due to the curvature convention correctly give the term $R(\bullet, \mathfrak{M}_1)_\bullet \vdots \mathfrak{M}_2$ of our formula Eq.(12).

References

[1] Bijan Afsari. Riemannian L^p center of mass: Existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139(02):655–673, February 2011.

[2] Marc Arnaudon and Xue-Mei Li. Barycenters of measures transported by stochastic flows. *The Annals of Probability*, 33(4):1509–1543, July 2005.

[3] Bojan Basrak. Limit theorems for the inductive mean on metric trees. *Journal of Applied Probability*, 47(4):1136–1149, December 2010.

[4] Marcel Berger. *A Panoramic View of Riemannian Geometry.* Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[5] Abhishek Bhattacharya and Rabi Bhattacharya. Statistics on Riemannian manifolds: asymptotic distribution and curvature. *Proceedings of the American Mathematical Society*, 136(8):2959–2967, 2008.

[6] Rabi Bhattacharya and Vic Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds-II. *The Annals of Statistics*, 33(3):1225–1259, June 2005.

[7] Dario A. Bini and Bruno Iannazzo. Computing the Karcher mean of symmetric positive definite matrices. *Linear Algebra and its Applications*, 438(4):1700–1710, February 2013.

[8] Leo Brewin. Riemann normal coordinate expansions using Cadabra. *Classical and Quantum Gravity*, 26(17):175017, September 2009.

[9] Peter Buser and Hermann Karcher. *Gromov's almost flat manifolds.* Number 81 in Astérisque. Société mathématique de France, 1981.

[10] Samuel R. Buss and Jay P. Fillmore. Spherical Averages and Applications to Spherical Splines and Interpolation. *ACM Trans. Graph.*, 20(2):95–126, April 2001.

[11] Élie Cartan. *Leçons sur la géométrie des espaces de Riemann.* Gauthier-Villars, Paris, France, 1928.

[12] Élie Cartan. Groupes simples clos et ouverts et géométrie riemannienne. *Journal de Mathématiques Pures et Appliquées*, 9e série(tome 8):1–34, 1929.

[13] Richard William Ramsay Darling. Geometrically Intrinsic Nonlinear Recursive Filters II: Foundations. *Electronic Journal of Probability*, 5:1–18, 2000. arXiv: math/9809029.

[14] Manfredo Perdigao do Carmo. *Riemannian geometry.* Mathematics. Theory & Applications. Birkhäuser, Boston, 1992.

[15] Benjamin Eltzner and Stephan F. Huckemann. A Smeary Central Limit Theorem for Manifolds with Application to High Dimensional Spheres. *arXiv:1801.06581 [math, stat]*, January 2018. arXiv: 1801.06581.

[16] M. Emery and G. Mokobodzki. Sur le barycentre d'une probabilité dans une variété. In *Séminaire de Probabilités XXV*, volume 1485, pages 220–233. Springer Berlin Heidelberg, Berlin, Heidelberg, 1991.

[17] Maurice Fréchet. Valeurs moyennes attachées a un triangle aléatoire. *La revue scientifique*, Fascicule 10:475–482, December 1943.

[18] Maurice Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut Henri Poincaré*, 10:215–310, 1948.

[19] A V Gavrilov. Algebraic Properties of Covariant Derivative and Composition of Exponential Maps. *Math. Tr.*, 9(1):3–20, 2006. English version in Siberian Advances in Mathematics, 2006, 16:3, 54-70.

[20] A. V. Gavrilov. The double exponential map and covariant derivation. *Siberian Mathematical Journal*, 48(1):56–61, January 2007.

[21] Karsten Grove and Hermann Karcher. How to conjugate C1-close group actions. *Mathematische Zeitschrift*, 132(1):11–20, March 1973.

[22] Susan Holmes. Personnal communication.

[23] Thomas Hotz, Stephan Huckemann, Huiling Le, J. S. Marron, Jonathan C. Mattingly, Ezra Miller, James Nolen, Megan Owen, Vic Patrangenaru, and Sean Skwerer. Sticky central limit theorems on open books. *The Annals of Applied Probability*, 23(6):2238–2258, December 2013.

[24] H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30(5):509–541, September 1977.

[25] Hermann Karcher. Riemannian Center of Mass and so called Karcher mean. *arXiv:1407.2087 [math]*, July 2014.

[26] Wilfrid S. Kendall and Huiling Le. Limit theorems for empirical Fréchet means of independent and non-identically distributed manifold-valued random variables. *Brazilian Journal of Probability and Statistics*, 25(3):323–352, November 2011.

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

[28] Huiling Le. Locating Fréchet Means with Application to Shape Spaces. *Advances in Applied Probability*, 33(2):324–338, 2001.

[29] Huiling Le. Estimation of Riemannian barycenters. *LMS J. Comput. Math.*, 7:193–200, 2004.

[30] John M. Lee. *Riemannian manifolds: an introduction to curvature*. Number 176 in Graduate texts in mathematics. Springer, New York, 1997.

[31] Charles W. Misner, Kip S. Thorne, and John Archibald Wheeler. *Gravitation*. W. H. Freeman, San Francisco, 1973.

[32] Barrett O'Neill. *Semi-Riemannian geometry: with applications to relativity*. Number 103 in Pure and applied mathematics. Academic Press, New York, 1983.

[33] Xavier Pennec. Barycentric subspace analysis on manifolds. *Annals of Statistics*, 46(6A):2711–2746, 2018.

[34] M. M Postnikov. *Geometry VI: Riemannian geometry*. Springer, Berlin, 2010.

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

[36] Le Yang. Riemannian median and its estimation. *LMS Journal of Computation and Mathematics*, 13:461–479, December 2010.

[37] Herbert Ziezold. On Expected Figures and a Strong Law of Large Numbers for Random Elements in Quasi-Metric Spaces. In J. Kožešnik, editor, *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the 1974 European Meeting of Statisticians*, number 7A, pages 591–602. Springer Netherlands, 1977.