Warped Riemannian metrics for location-scale models

Salem Said, Lionel Bombrun, Yannick Berthoumieu

Abstract The present contribution shows that warped Riemannian metrics, a class of Riemannian metrics which play a prominent role in Riemannian geometry, are also of fundamental importance in information geometry. Precisely, the starting point is a new theorem, which states that the Rao-Fisher information metric of any location-scale model, defined on a Riemannian manifold, is a warped Riemannian metric, whenever this model is invariant under the action of some Lie group. This theorem is a valuable tool in finding the expression of the Rao-Fisher information metric of location-scale models defined on high-dimensional Riemannian manifolds. Indeed, a warped Riemannian metric is fully determined by only two functions of a single variable, irrespective of the dimension of the underlying Riemannian manifold. Starting from this theorem, several original results are obtained. The expression of the Rao-Fisher information metric of the Riemannian Gaussian model is provided, for the first time in the literature. A generalised definition of the Mahalanobis distance is introduced, which is applicable to any location-scale model defined on a Riemannian manifold. The solution of the geodesic equation, as well as an explicit construction of Riemannian Brownian motion, are obtained, for any Rao-Fisher information metric defined in terms of warped Riemannian metrics. Finally, using a mixture of analytical and numerical computations, it is shown that the parameter space of the von Mises-Fisher model of n-dimensional directional data, when equipped with its Rao-Fisher information metric, becomes a Hadamard manifold, a simply-connected complete Riemannian manifold of negative sectional curvature, for $n = 2, \ldots, 8$. Hopefully, in upcoming work, this will be proved for any value of n.

Salem Said
Laboratoire IMS, CNRS, Université de Bordeaux, e-mail: `salem.said@u-bordeaux.fr`

Lionel Bombrun
Laboratoire IMS, Bordeaux Sciences Agro, e-mail: `lionel.bombrun@u-bordeaux.fr`

Yannick Berthoumieu
Laboratoire IMS, Bordeaux INP,e-mail: `yannick.berthoumieu@u-bordeaux.fr`

Notation: hard to differentiate $\tilde{M}$, M ?

1 Introduction

[11, 40]

Warped Riemannian metrics are a class of Riemannian metrics which arise throughout Riemannian geometry [11][40]. For example, the Riemannian metrics of surfaces of revolution, and of spaces of constant curvature (when restricted to polar coordinate charts), are warped Riemannian metrics. Closely similar to warped Riemannian metrics, warped semi-Riemannian metrics are very important in theoretical physics. Indeed, many gravitational models are given by warped semi-Riemannian metrics [38]. The present contribution shows that warped metrics, in addition to their well-known role in geometry and physics, play a fundamental role in information geometry, and have a strong potential for applications in statistical inference and statistical learning.

A unified definition of warped Riemannian metrics was first formulated in [11]. Here, only a special case of this definition is required. Precisely, let M be a complete Riemannian manifold, with length element ds^2_M, and consider the product manifold $\mathcal{M} = M \times (0,\infty)$, equipped with the length element $ds^2_{\mathcal{M}}$

$$ds^2_{\mathcal{M}}(z) = dr^2 + \beta^2(r)\, ds^2_M(x) \qquad \text{for } z = (x,r) \in \mathcal{M} \tag{1a}$$

where $\beta^2(r)$ is a strictly positive function. Then, the length element $ds^2_{\mathcal{M}}$ defines a warped Riemannian metric on $\mathcal{M}$. In Riemannian geometry, the coordinate r is a distance function, measuring the distance to some point or hypersurface [40]. In physics, r is replaced by the time t, and dr^2 is replaced by $-dt^2$ in formula (1a) (this is the meaning of "semi-Riemannian") [38]. In any case, the coordinate x can be thought of as a spatial coordinate which determines a position in M.

The intuition behind the present contribution is that warped Riemannian metrics are natural candidates for Riemannian metrics on location-scale models. Indeed, if $\mathcal{P}$ is a location-scale model on M, with location parameter $\bar{x} \in M$ and scale parameter $\sigma > 0$, then the parameter space of $\mathcal{P}$ is exactly $\mathcal{M} = M \times (0,\infty)$ with its points $z = (\bar{x},\sigma)$. Thus, a warped Riemannian metric on $\mathcal{M}$ can be defined using (1a), after introducing a new scale parameter $r = r(\sigma)$ and setting $x = \bar{x}$.

As it turns out, this intuition is far from arbitrary. The main new result in the present contribution, Theorem 1 of Section 3, states that the Rao-Fisher information metric of any location-scale model is a warped Riemannian metric, whenever this model is invariant under the action of some Lie group. Roughly, Theorem 1 states that if M is a Riemannian symmetric space under the transitive action of a Lie group of isometries G, and if each probability density $p(x|\,\bar{x},\sigma)$, belonging to the model $\mathcal{P}$, verifies the invariance condition

$$p(\,g\cdot x|\,g\cdot\bar{x},\sigma) = p(x|\bar{x},\sigma) \quad \text{for all } g \in G \tag{1b}$$

where $g \cdot x$ denotes the action of $g \in G$ on $x \in M$, then the Rao-Fisher information metric of the model $\mathcal{P}$ is a warped Riemannian metric.

A technical requirement for Theorem 1 is that the Riemannian symmetric space M should be irreducible. The meaning of this requirement, and the fact that it can be relaxed in certain cases, are discussed in Remarks 4 and 5 of Section 3. The proof of Theorem 1 is given in Appendix A.

A fundamental idea of information geometry is that the parameter space of a statistical model $\mathcal{P}$ should be considered as a Riemannian manifold [2][15]. [2,15] According to [15][7], the unique way of doing so is by turning Fisher's information matrix into a Riemannian metric, the Rao-Fisher information metric. In this connection, Theorem 1 shows that, when the statistical model $\mathcal{P}$ is a location-scale model which is invariant under the action of a Lie group, information geometry inevitably leads to the study of warped Riemannian metrics.

In addition to stating and proving Theorem 1, the present contribution aims to explore its implications, with regard to the Riemannian geometry of location-scale models, and to lay the foundation for its applications in statistical inference and statistical learning.

To begin, Section 4 applies Theorem 1 to two location-scale models, the von Mises-Fisher model of directional data [33][16], and the Riemannian Gaussian model of data in spaces of covariance matrices [14, 41, 42]. This leads to the analytic expression of the Rao-Fisher information metric of each one of these two models. Precisely, the Rao-Fisher information metric of the von Mises-Fisher model is given in Proposition 2, and that of the Riemannian Gaussian model is given in Proposition 3. The result of Proposition 2 is essentially already contained in [33], (see Page 199), but Proposition 3 is new in the literature.

Finding the analytic expression of the Rao-Fisher information metric, or equivalently of Fisher's information matrix, of a location-scale model $\mathcal{P}$ defined on a high-dimensional non-trivial manifold M, is a very difficult task when attempted by direct calculation. Propositions 2 and 3 show that this task is greatly simplified by Theorem 1. Precisely, if the dimension of M is d, then the dimension of the parameter space $\mathcal{M} = M \times (0,\infty)$ is $d+1$. Therefore, *a priori*, the expression of the Rao-Fisher information metric involves $(d+1)(d+2)/2$ functions of both parameters $\bar{x}$ and σ of the model $\mathcal{P}$. Instead of so many functions of both $\bar{x}$ and σ, Theorem 1 reduces the expression of the Rao-Fisher information metric to only two functions of σ alone. In the notation of (1a), these two functions are $\alpha(\sigma) = dr/d\sigma$ and $\beta(\sigma) = \beta(r(\sigma))$.

Section 5 builds on Theorem 1 to introduce a general definition of the Mahalanobis distance, applicable to any location-scale model $\mathcal{P}$ defined on a manifold M. Precisely, assume that the model $\mathcal{P}$ verifies the conditions of Theorem 1, so its Rao-Fisher information metric is a warped Riemannian metric. Then, the *generalised Mahalanobis distance* is defined as the Riemannian distance on M which is induced by the restriction of the Rao-Fisher information metric to M. The expression of the generalised Mahalanobis dis-

tance is given in Propositions 4 and 5. It was recently applied to visual content classification in [10].

The generalised Mahalanobis distance includes the classical Mahalanobis distance as a special case. Precisely, assume $\mathcal{P}$ is the isotropic normal model defined on $M = \mathbb{R}^d$, so each density $p(x|\bar{x},\sigma)$ is a d-variate normal density with mean $\bar{x}$ and covariance matrix σ^2 times the identity. Then, $\mathcal{P}$ verifies the invariance condition (1b) under the action of the group G of translations in $\mathbb{R}^d$. Therefore, by Theorem 1, its Rao-Fisher information metric is a warped Riemannian metric. This metric is already known in the literature, in terms of the length element [8][9]

$$ds^2_{\mathcal{M}}(z) = \frac{2d}{\sigma^2}\,d\sigma^2 + \frac{1}{\sigma^2}\,\|d\bar{x}\|^2 \qquad \text{for } z = (\bar{x},\sigma) \in \mathcal{M} \tag{1c}$$

where $\|d\bar{x}\|^2$ denotes the Euclidean length element on $\mathbb{R}^d$. Now, the restriction of this Rao-Fisher information metric to $M = \mathbb{R}^d$ is given by the second term in (1c), which is clearly the Euclidean length element divided by σ^2, and corresponds to the classical Mahalanobis distance [36] — note that (1c) is brought into the form (1a) by letting $r(\sigma) = \sqrt{2d}\log(\sigma)$ and $\beta(r) = \exp(-r/\sqrt{2d})$. trop petit.

Section 6 illustrates the results of Sections 4 and 5, by applying them to the special case of the Riemannian Gaussian model defined on $M = \mathcal{P}_n$, the space of $n \times n$ real covariance matrices. In particular, it gives directly applicable expressions of the Rao-Fisher information metric, and of the generalised Mahalanobis distance, corresponding to this model. As it turns out, the generalised Mahalanobis distance defines a whole new family of affine-invariant distances on $\mathcal{P}_n$, in addition to the usual affine-invariant distance, which was introduced to the information science community in [39].

Section 7 provides the solution of the geodesic equation of any of the Rao-Fisher information metrics arising from Theorem 1. The main result is Proposition 6, which states that the solution of this equation, for given initial conditions, reduces to the solution of a one-dimensional second-order differential equation. This implies that geodesics with given initial conditions can be constructed at a reasonable numerical cost, which opens the possibility, with regard to future work, of a practical implementation of Riemannian line-search optimisation algorithms, which find the extrema of cost functions by searching for them along geodesics [1][32].

Section 8 is a follow up to Section 7, focusing on the construction of Riemannian Brownian motion, instead of the solution of the geodesic equation. The main result is Proposition 7, which states that the construction of Riemannian Brownian motion reduces to the solution of a one-dimensional stochastic differential equation. This implies that Brownian paths can be constructed at a reasonable computational cost, which opens the possibility of a practical implementation of Riemannian stochastic search algorithms, which have the ability to avoid local minima and saddle points [5].

Section 9 is motivated by the special case of the isotropic normal model, with its Rao-Fisher information metric given by (1c). It is well-known that,

after a trivial change of coordinates, the length element (1c) coincides with the length element of the Poincaré half-space model of hyperbolic geometry [8][9]. This means that the parameter space of the isotropic normal model, when equipped with its Rao-Fisher information metric, becomes a space of constant negative curvature, and in particular a Hadamard manifold, a simply-connected complete Riemannian manifold of negative sectional curvature [40][13]. One cannot but wonder whether other location-scale models also give rise to Hadamard manifolds in this way. This is investigated using Propositions 2 and 6, for the case of the von Mises-Fisher model. A mixture of analytical and numerical computations leads to a surprising new observation: the parameter space of the von Mises-Fisher model of n-dimensional directional data, when equipped with its Rao-Fisher information metric, becomes a Hadamard manifold, for $n = 2,\ldots,8$. Ongoing research warrants the conjecture that this is true for any value of n, but this is yet to be proved.

Theorem 1, the main new result in the present contribution, has many potential applications, which will be developed in future work. Indeed, this theorem can provide the expression of the Rao-Fisher information metric, or equivalently of Fisher's information matrix, of a location-scale model, even if this model is defined on a high-dimensional non-trivial manifold. By doing so, it unlocks access to the many applications which require this expression, both in statistical inference and in statistical learning. In statistical inference, the expression of the Rao-Fisher information metric allows the computation of the Cramér-Rao lower bound, and the construction of asymptotic chi-squared statistics [30][48]. In statistical learning, it allows the practical implementation of the natural gradient algorithm, which has the advantages of full reparameterisation invariance, of asymptotic efficiency of its stochastic version, and of the linear rate of convergence of its deterministic version [3, 12, 35].

A first step, towards developing the applications of Theorem 1, was recently taken in [49]. Using the expression of the Rao-Fisher information metric of the Riemannian Gaussian model, this derived and successfully implemented the natural gradient algorithm, for the problem of on-line learning of an unknown probability density, on the space $\mathcal{P}_n$ of $n \times n$ real covariance matrices. Future work will focus on extending this algorithm to more problems of statistical learning, including problems of on-line classification and regression in the space $\mathcal{P}_n$, and in other spaces of covariance matrices. In addition to its use in deriving the natural gradient algorithm for problems of statistical learning, the expression of the Rao-Fisher information metric of the Riemannian Gaussian model can be used in deriving so-called natural evolutionary strategies, for black-box optimisation in spaces of covariance matrices. These would generalize currently existing natural evolutionary strategies, mostly restricted to black-box optimisation in Euclidean space [47][37].

The following Section 2 provides background on warped Riemannian metrics. While the present introduction expressed Riemannian metrics using length elements, Section 2 will use scalar products on the tangent space.

It was felt that the use of length elements, while somewhat less rigorous, is more intuitive, and therefore better suited for the introduction.

2 Background on warped Riemannian metrics

Assume M is a complete Riemannian manifold with Riemannian metric Q, and consider the manifold $\mathcal{M} = M \times (0,\infty)$. A warped Riemannian metric I on $\mathcal{M}$ is given in the following way [11, 38, 40]. Let α and β be positive functions, defined on $(0,\infty)$. Then, for $z = (x,\sigma) \in \mathcal{M}$, let the scalar product I_z on the tangent space $T_z\mathcal{M}$ be defined by

$$I_z(U,U) = (\alpha(\sigma)\,u_\sigma)^2 + \beta^2(\sigma)\,Q_x(u,u) \qquad U \in T_z\mathcal{M} \tag{2a}$$

where $U = u_\sigma\,\partial_\sigma + u$ with $u_\sigma \in \mathbb{R}$ and $u \in T_xM$. The functions α and β are part of the definition of the warped metric I. Once these functions are fixed, it is possible to introduce a change of coordinates $r = r(\sigma)$ which eliminates α from (2a). Precisely, if $dr/d\sigma = \alpha(\sigma)$ then

$$I_z(U,U) = u_r^2 + \beta^2(r)\,Q_x(u,u) \tag{2b}$$

where $U = u_r\,\partial_r + u$ and $\beta(r) = \beta(\sigma(r))$.

The coordinate r will be called *vertical distance*. This is not a standard terminology, but is suggested as part of the following geometric picture. For $z = (x,\sigma) \in \mathcal{M}$, think of x as a horizontal coordinate, and of σ as a vertical coordinate. Accordingly, the points $z_0 = (x,\sigma_0)$ and $z_1 = (x,\sigma_1)$ lie on the same vertical line. It can be shown from (2b) that the Riemannian distance between z_0 and z_1 is

$$d(z_0,z_1) = r(\sigma_1) - r(\sigma_0) \qquad \text{where } \sigma_0 < \sigma_1 \tag{3a}$$

Precisely, $d(z_0,z_1)$ is the Riemannian distance induced by the warped Riemannian metric I.

The vertical distance r can be used to express a necessary and sufficient condition for completeness of the manifold $\mathcal{M}$, equipped with the warped Riemannian metric I. Namely, $\mathcal{M}$ is a complete Riemannian manifold, if and only if

$$\lim_{\sigma\to\infty} r(\sigma) - r(\sigma_0) = \infty \quad \text{and} \quad \lim_{\sigma\to 0} r(\sigma_1) - r(\sigma) = \infty \tag{3b}$$

where σ_0 and σ_1 are arbitrary. This condition is a special case of Lemma 7.2 in [11].

Let $K^{\mathcal{M}}$ and K^M denote the sectional curvatures of $\mathcal{M}$ and M, respectively. The relation between these two is given by the curvature equations of Riemannian geometry [40][17]. These are, rendu .

and β^{-1}

$$\text{Gauss equation}: K^{\mathcal{M}}_z(u,v) = \frac{1}{\beta^2}K^{M}_x(u,v) - \left(\frac{\dot\beta}{\beta}\right)^2 \tag{4a}$$

$$\text{Jacobi equation}: K^{\mathcal{M}}_z(u,\partial_r) = -\frac{\ddot\beta}{\beta} \tag{4b}$$

small

for $u, v \in T_xM$. Here, the notations $K^{\mathcal{M}}_z$ and K^M_x mean that $K^{\mathcal{M}}$ is computed at z, and K^M is computed at x, where $z = (x,\sigma)$. Equations (4) are a special case of Lemma 7.4 in [11].

Note, as a corollary of these equations, that $\mathcal{M}$ has negative sectional curvature $K^{\mathcal{M}} < 0$, if M has negative sectional curvature $K^M < 0$ and β is a strictly convex function of r.

Remark 1 : equations (2) contain an abuse of notation. Namely, u denotes a tangent vector to M at x, and a tangent vector to $\mathcal{M}$ at z, at the same time. In the mathematical literature (for example, in [11][38]), one writes $d\pi_z(U)$ instead of u, using the derivative $d\pi$ of the projection mapping $\pi(z) = x$, and this eliminates any ambiguity. In the present contribution, a deliberate choice is made to use a lighter, though not entirely correct, notation. ■

Remark 2 : consider the proof of equations (3). For (3a), let $\gamma(t)$ and $c(t)$ be curves connecting $z_o = (x,\sigma_o)$ and $z_1 = (x,\sigma_1)$. Assume these are parameterised as follows,

$$\gamma(t) : \begin{Bmatrix} x(t) = x \quad \text{(constant)} \\ r(t) = r(\sigma_o) + t\,(r(\sigma_1) - r(\sigma_o)) \end{Bmatrix} \qquad c(t) : \begin{Bmatrix} x(t) \\ r(t) \end{Bmatrix}$$

seems like $\gamma(t) = c(t)$. $c(t)$ is arbitrary right?

where $t \in [0,1]$. If $L(\gamma)$ and $L(c)$ denote the lengths of these curves, then from (2b),

$$L(c) = \int_0^1 \left((\dot r)^2 + \beta^2(r)Q(\dot x,\dot x)\right)^{1/2} dt \geq \int_0^1 \dot r\, dt = r(\sigma_1) - r(\sigma_o) = L(\gamma)$$

where the dot denotes differentiation with respect to t, and the inequality is strict unless $\gamma = c$. This shows that $\gamma(t)$ is the unique length-minimising geodesic connecting z_o and z_1. Thus, $d(z_o, z_1) = L(\gamma)$, and this gives (3a). For (3b), note that Lemma 7.2 in [11] states that $\mathcal{M}$ is complete, if and only if $(0,\infty)$ is complete when equipped with the distance

$$d_{(0,\infty)}(\sigma_o,\sigma_1) = |r(\sigma_1) - r(\sigma_o)|$$

However, this is equivalent to (3b). An alternative way of understanding (3b) is provided in Remark 14 of Section 7. Precisely, let a geodesic of the form $\gamma(t)$ be called a vertical geodesic. The first condition in (3b) states that a vertical geodesic cannot reach the value $\sigma = \infty$ within a finite time, and the second condition in (3b) states that a vertical geodesic cannot reach the value $\sigma = 0$ within a finite time. ■

Remark 3 : for each $\sigma \in (0,\infty)$, the manifold M can be embedded into the

manifold $\mathcal{M}$, in the form of the hypersurface $M_\sigma = M \times \{\sigma\}$. Through this embedding, the warped Riemannian metric I of $\mathcal{M}$ induces a Riemannian metric Q^σ on M. By definition, this metric Q^σ is obtained by the restriction of I to the tangent vectors of M_σ [40][17]. It follows from (2) that

$$Q^\sigma_x(u,u) = \beta^2(\sigma)\,Q_x(u,u) \tag{5}$$

The induced metric Q^σ will be called an extrinsic metric on M, since it comes from the ambient space $\mathcal{M}$. By (5), the extrinsic metric Q^σ is equal to a scaled version of the Riemannian metric Q of M, with scaling factor $\beta(\sigma)$. ■

3 Connection with location-scale models

This section establishes the connection between warped Riemannian metrics and location-scale models. The main result is Theorem 1, which states that the Rao-Fisher information metric of any location-scale model is a warped Riemannian metric, whenever this model is invariant under the action of some Lie group.

To state this theorem, assume M is an irreducible Riemannian symmetric space, with invariant Riemannian metric Q, under the transitive action of a Lie group of isometries G [22]. Consider a location-scale model $\mathcal{P}$ defined on M,

$$\mathcal{P} = \{\,p(x|\bar x,\sigma)\,;\, \bar x \in M\,,\, \sigma \in (0,\infty)\} \tag{6}$$

To each point $z = (\bar x,\sigma)$ in the parameter space $\mathcal{M} = M \times (0,\infty)$, this model associates a probability density $p(x|\bar x,\sigma)$ on M, which has a location parameter $\bar x$ and a scale parameter σ. Precisely, $p(x|\bar x,\sigma)$ is a probability density with respect to the invariant Riemannian volume element of M.

The condition that the model $\mathcal{P}$ is invariant under the action of the Lie group G means that,

$$p(\,g\cdot x|\,g\cdot\bar x\,,\sigma) = p(x|\bar x,\sigma) \quad \text{for all } g \in G \tag{7}$$

where $g \cdot x$ denotes the action of $g \in G$ on $x \in M$.

The Rao-Fisher information metric of the location-scale model $\mathcal{P}$ is a Riemannian metric I on the parameter space $\mathcal{M}$ of this model [2]. It is defined as follows, for $z = (\bar x,\sigma) \in \mathcal{M}$ and $U \in T_z\mathcal{M}$,

$$I_z(U,U) = \mathbb{E}_z\left[\left(d\ell(z)\,U\right)^2\right] \tag{8}$$

$d\ell(z) \cdot U$? Notation

where $\mathbb{E}_z$ denotes expectation with respect to the probability density $p(x|z) = p(x|\bar x,\sigma)$, and $\ell(z)$ is the log-likelihood function, given by $\ell(z)(x) = \log p(x|z)$.

In the following statement, $\nabla_{\bar x}\,\ell(z)$ denotes the Riemannian gradient of $\ell(z)$, taken with respect to $\bar x \in M$, while the value of σ is fixed.

Theorem 1. *if condition (7) is verified, then the Rao-Fisher information metric I of (8) is a warped Riemannian metric given by (2a), where*

$$\alpha^2(\sigma) = \mathbb{E}_z\,(\partial_\sigma \ell(z))^2 \qquad \beta^2(\sigma) = \mathbb{E}_z\,Q\,(\nabla_{\bar x}\,\ell(z)\,,\nabla_{\bar x}\,\ell(z)\,)/\dim M \tag{9}$$

The expectations appearing in (9) do not depend on $\bar x$, so $\alpha(\sigma)$ and $\beta(\sigma)$ are well-defined functions of σ.

Remark 4 : recall the definition of an irreducible Riemannian symmetric space [22]. A Riemannian manifold M, whose group of isometries is denoted G, is called a Riemannian symmetric space, if for each $\bar x \in M$ there exists an isometry $s_{\bar x} \in G$, whose effect is to fix $\bar x$ and reverse the geodesic curves passing through $\bar x$. Further, M is called irreducible if it verifies the following condition. Let $K_{\bar x}$ be the subgroup of G which consists of those elements k such that $k \cdot \bar x = \bar x$. For each $k \in K_{\bar x}$, its derivative $dk_{\bar x}$ is a linear mapping of $T_{\bar x}M$. The mapping $k \mapsto dk_{\bar x}$ is a representation of $K_{\bar x}$ in $T_{\bar x}M$, called the isotropy representation, and M is called an irreducible Riemannian symmetric space if the isotropy representation is irreducible. That is, if the isotropy representation has no invariant subspaces in $T_{\bar x}M$, except $\{0\}$ and $T_{\bar x}M$. Irreducible Riemannian symmetric spaces are classified in [22] (Table I, Page 346 and Table II, Page 354). They include spaces of constant curvature, such as spheres and hyperbolic spaces, as well as spaces of positive definite matrices which have determinant equal to 1, and whose entries are real or complex numbers, or quaternions. ■

Remark 5 : it is sometimes possible to apply Theorem 1, even when the Riemannian symmetric space M is not irreducible. For example, in 4.2, Theorem 1 will be used to find the expression of the Rao-Fisher information metric of the Riemannian Gaussian model [14,41,42]. For this model, when M is not irreducible, the Rao-Fisher information metric turns out to be a so-called multiply-warped Riemannian metric, rather than a warped Riemannian metric. The concrete case of $M = \mathcal{P}_n$, the space of $n \times n$ real covariance matrices, is detailed in Section 6. ■

Proof of Theorem 1 : recall the expression $U = u_\sigma \partial_\sigma + u$ with $u_\sigma \in \mathbb{R}$ and $u \in T_{\bar x}M$. Since the Rao-Fisher information metric I is bilinear and symmetric,

$$I_z(U,U) = I_z(\partial_\sigma,\partial_\sigma)\,u^2_\sigma + 2I_z(\partial_\sigma,u)\,u_\sigma + I_z(u,u)$$

It is possible to show the following,

$$I_z(\partial_\sigma,\partial_\sigma) = \quad \alpha^2(\sigma) \tag{10a}$$

$$I_z(\partial_\sigma,u) = \quad 0 \tag{10b}$$

$$I_z(u,u) = \beta^2(\sigma)\,Q_{\bar x}(u,u) \tag{10c}$$

where $\alpha^2(\sigma)$ and $\beta^2(\sigma)$ are given by (9).

Proof of (10a) : this is immediate from (8). Indeed,

$$I_z(\partial_\sigma,\partial_\sigma) = \mathbb{E}_z\left[\left(d\ell(z)\,\partial_\sigma\right)^2\right] = \mathbb{E}_z\left[\left(\partial_\sigma \ell(z)\right)^2\right]$$

Proof of (10b) : this is carried out in Appendix A, using the fact that M is a Riemannian symmetric space.

Proof of (10c) : this is carried out in Appendix A, using the fact that M is irreducible, by an application of Schur's lemma from the theory of group representations [27].

The fact that the expectations appearing in (9) do not depend on $\bar x$ is also proved in Appendix A. Throughout the proof of the theorem, the following identity is used, which is equivalent to condition (7). For any real-valued function f on M,

$$\mathbb{E}_{g\cdot z}\,f = \mathbb{E}_z\,(f \circ g) \tag{11}$$

Here, $g \cdot z = (g \cdot \bar x, \sigma)$, and $f \circ g$ is the function $(f \circ g)(x) = f(g \cdot x)$, for $g \in G$ and $z = (\bar x,\sigma)$. ■

4 Examples : von Mises-Fisher and Riemannian Gaussian

This section applies Theorem 1 to finding the expression of the Rao-Fisher information metric of two location-scale models. These are the von Mises-Fisher model, which is widely used in the study of directional data [33][16], and the Riemannian Gaussian model, recently introduced in the study of data with values in spaces of covariance matrices [14,41,42].

The application of Theorem 1 to these two models is encapsulated in the following Proposition 1. Precisely, both of these models are of a common exponential form, which can be described as follows. Let M be an irreducible Riemannian symmetric space, as in Section 3. In the notation of (6), consider a location-scale model $\mathcal{P}$ defined on M, by

$$p(x|\bar x,\sigma) = \exp\left[\,\eta(\sigma)\,D(x,\bar x) - \psi(\eta(\sigma))\right] \tag{12a}$$

where $\eta(\sigma)$ is a certain parameter, to be called the natural parameter, and where $D : M \times M \to \mathbb{R}$ verifies the condition,

$$D(\,g\cdot x\,,\,g\cdot\bar x) = D(x\,,\bar x) \quad \text{for all } g \in G \tag{12b}$$

There is no need to assume that the function D is positive.

Proposition 1. *if the model $\mathcal{P}$ is given by equations (12), then the Rao-Fisher information metric I of this model is a warped Riemannian metric,*

$$I_z(U,U) = \psi''(\eta)\,u^2_\eta + \beta^2(\eta)\,Q_{\bar x}(u,u) \tag{13a}$$

where $U = u_\eta\, \partial_\eta + u$ with $u_\eta \in \mathbb{R}$ and $u \in T_x M$, and where

$$\beta^2(\eta) = \eta^2\, \mathbb{E}_z\, Q\left(\nabla_z\, D, \nabla_z\, D\right)/\dim M \tag{13b}$$

$\nabla_{\bar{x}} D$ not clear.

Proof: for a model $\mathcal{P}$ defined by (12a), condition (12b) is equivalent to condition (7). Therefore, by application of Theorem 1, it follows that I is a warped Riemannian metric, of the form (2a),

$$I_z(U,U) = (\alpha(\sigma)\, u_\sigma)^2 + \beta^2(\sigma)\, Q_x(u,u) \tag{14a}$$

where $\alpha^2(\sigma)$ and $\beta^2(\sigma)$ are given by (9). Consider the first term in (14a). By the change of coordinates formula [29], $u_\sigma = \sigma'(\eta)\, u_\eta$, where the prime denotes differentiation with respect to η. It follows that

$$(\alpha(\sigma)\, u_\sigma)^2 = \alpha^2(\sigma)\, (\sigma'(\eta))^2\, u_\eta^2 \tag{14b}$$

However, by (9),

$$\alpha^2(\sigma)\,(\sigma'(\eta))^2 = \mathbb{E}_z\left(\partial_\sigma \ell(z)\, \sigma'(\eta)\right)^2 = \mathbb{E}_z\left(\partial_\eta \ell(z)\right)^2 \tag{14c}$$

Here, the log-likelihood $\ell(z)$ is found from (12a),

$$\ell(z)(x) = \eta(\sigma)\, D(x,\bar{x}) - \psi(\eta(\sigma)) \tag{14d}$$

Therefore, the last expression in (14c) is

$$\mathbb{E}_z\left(\partial_\eta \ell(z)\right)^2 = -\mathbb{E}_z\, \partial^2_\eta \ell(z) = \psi''(\eta) \tag{14e}$$

where the first equality is the same as in [2], (see Page 28). Now, (14b) and (14c) imply

$$(\alpha(\sigma)\, u_\sigma)^2 = \psi''(\eta)\, u_\eta^2 \tag{14f}$$

Replacing this in (14a), and writing $\beta(\eta) = \beta(\sigma(\eta))$, gives

$$I_z(U,U) = \psi''(\eta)\, u_\eta^2 + \beta^2(\eta)\, Q_x(u,u) \tag{14g}$$

which is the same as (13a). To prove the proposition, it remains to show that $\beta^2(\eta)$ is given by (13b). To do so, note that it follows from (14d),

$$\nabla_z\, \ell(z) = \nabla_z\left[\eta(\sigma)\, D(x,\bar{x}) - \psi(\eta(\sigma))\right] = \eta(\sigma)\, \nabla_z\, D(x,\bar{x})$$

(x)

Replacing this in (9) gives,

$$\beta^2(\eta) = \mathbb{E}_z\, Q\left(\nabla_z\, \ell(z), \nabla_z\, \ell(z)\right)/\dim M = \eta^2\, \mathbb{E}_z\, Q\left(\nabla_z\, D, \nabla_z\, D\right)/\dim M$$

and this is the same as (13b). ■

$\mathbb{R}^n$ mal écrit

4.1 The von Mises-Fisher model

The von Mises-Fisher model is a mainstay of directional statistics [33][16]. In the notation of (12), this model corresponds to $M = S^{n-1}$, the unit sphere in $\mathbb{R}^n$, and to $G = O(n)$, the group of $n \times n$ real orthogonal matrices, which acts on $\mathbb{R}^n$ by rotations. Then, the expressions appearing in (12a) are

$$D(x,\bar{x}) = \langle x, \bar{x}\rangle \qquad \psi(\eta) = \nu \log(2\pi) + \log\left(\eta^{1-\nu} I_{\nu-1}(\eta)\right) \tag{15}$$

for $\eta \in [0,\infty)$. Here, $\langle x,\bar{x}\rangle$ denotes the Euclidean scalar product in $\mathbb{R}^n$, so that condition (12b) is clearly verified, and $I_{\nu-1}$ denotes the modified Bessel function of order $\nu - 1$, where $\nu = n/2$. The natural parameter η and the scale parameter σ should be considered identical, in the sense that $\eta(\sigma) = \sigma$, as long as $\sigma \in (0,\infty)$. However, η takes on the additional value $\eta = 0$, which requires a special treatment.

Remark 6: the parameter space of the von Mises-Fisher model will be identified with the space $\mathbb{R}^n$. This is done by mapping each couple $(\bar{x},\eta)$ to the point $z = \eta\,\bar{x}$ in $\mathbb{R}^n$. This mapping defines a diffeomorphism from the set of couples $(\bar{x},\eta)$ where $\eta \in (0,\infty)$, to the open subset $\mathbb{R}^n - \{0\} \subset \mathbb{R}^n$. On the other hand, it maps all couples $(\bar{x},\eta = 0)$, to the same point $z = 0 \in \mathbb{R}^n$. Note that each couple $(\bar{x},\eta)$ where $\eta \in (0,\infty)$ defines a distinct von Mises-Fisher distribution, which is a unimodal distribution with its mode at $\bar{x}$. On the other hand, all couples $(\bar{x},\eta = 0)$ define the same von Mises-Fisher distribution, which is the uniform distribution on S^{n-1}. Therefore, it is correct to map all of these couples to the same point $z = 0$. ■

Proposition 1 will only provide the Rao-Fisher information metric of the von Mises-Fisher model on the subset $\mathbb{R}^n - \{0\}$ of the parameter space $\mathbb{R}^n$. Therefore, it is necessary to verify that this metric has a well-defined limit at the point $z = 0$. This is carried out in Proposition 2 below. In the statement of this proposition, a tangent vector $U \in T_z\mathbb{R}^n$, at a point $z \in \mathbb{R}^n - \{0\}$, is written in the form

$$U = u_\eta\, \bar{x} + \eta\, u \tag{16a}$$

identification rapide. pourquoi pas ∂_η?

where $z = \eta\,\bar{x}$, and where $u_\eta \in \mathbb{R}$ and $u \in T_x S^{n-1}$. Here, u_η and u are unique, for a given U. Precisely,

$$u_\eta = \langle U, \bar{x}\rangle \qquad u = \frac{1}{\eta}\left[U - \langle U, \bar{x}\rangle\, \bar{x}\right] \tag{16b}$$

as follows since $\bar{x}$ and u are orthogonal.

Proposition 2. *the Rao-Fisher information metric I of the von Mises-Fisher model is a well-defined Riemannian metric on the parameter space $\mathbb{R}^n$. On $\mathbb{R}^n - \{0\}$, it is a warped Riemannian metric of the form (13a), where*

faille

$$\psi''(\eta) = \tfrac{1}{n} + \tfrac{n-1}{n}\, \frac{I_{\nu+1}(\eta)}{I_{\nu-1}(\eta)} - \frac{I^2_\nu(\eta)}{I^2_{\nu-1}(\eta)} \tag{17a}$$

$$\beta^2(\eta) = \tfrac{\eta^2}{n}\left(1 + \frac{I_{\nu+1}(\eta)}{I_{\nu-1}(\eta)}\right) \tag{17b}$$

and it extends smoothly to the value,

$$I_0(U,U) = \frac{1}{n}\,\|U\|^2 \tag{17c}$$

at the point $z = 0$. Here, $\|\cdot\|$ denotes the Euclidean norm.

Proof: the Rao-Fisher information metric I on $\mathbb{R}^n - \{0\}$ is given by Proposition 1. This proposition applies because $M = S^{n-1}$ is an irreducible Riemannian symmetric space [22] (Table II, Page 354). Accordingly, for any point $z \in \mathbb{R}^n - \{0\}$, the metric I_z is given by (13a). Formulae (17) are proved as follows.

Proof of (17a): this is carried out in Appendix B, using the derivative and recurrence relations of modified Bessel functions [46].

Proof of (17b): this follows from (13b) and (15). By (15),

$$\nabla_x\, D(x,\bar{x}) = x - \langle x,\bar{x}\rangle\, \bar{x}$$

which is just the orthogonal projection of x onto the tangent space $T_x S^{n-1}$. Replacing in (13b) gives

$$\beta^2(\eta) = \frac{\eta^2}{n-1}\, \mathbb{E}_z\, \|\nabla_x\, D\|^2 = \frac{\eta^2}{n-1}\, \mathbb{E}_z\left(1 - \langle x,\bar{x}\rangle^2\right) \tag{18a}$$

Here, in the first equality, $n - 1$ appears because $\dim S^{n-1} = n - 1$. The second equality follows by Pythagoras' theorem,

$$\|x - \langle x,\bar{x}\rangle\, \bar{x}\|^2 = \|x\|^2 - \|\langle x,\bar{x}\rangle\, \bar{x}\|^2 = 1 - \langle x,\bar{x}\rangle^2$$

since x and $\bar{x}$ belong to the unit sphere S^{n-1}. Formula (17b) is derived from (18a) in Appendix B, using the derivative and recurrence relations of modified Bessel functions [46].

Proof of (17c): for any point $z \in \mathbb{R}^n - \{0\}$, the metric I_z is given by (13a). This reads

$$I_z(U,U) = \psi''(\eta)\, u_\eta^2 + \beta^2(\eta)\, \|u\|^2 \tag{18b}$$

Consider the limit of this expression at the point $z = 0$. In (17a) and (17b), this corresponds to the limit at $\eta = 0$. This can be evaluated using the power series development of modified Bessel functions [46]. When replaced in (17a) and (17b), this gives the following developments,

faille

$$\psi''(\eta) = \tfrac{1}{n} - \tfrac{12}{n^2(n+2)}\left(\tfrac{\eta}{2}\right)^2 + O\left(\eta^4\right)$$

$$\beta^2(\eta) = \tfrac{4}{n}\left(\tfrac{\eta}{2}\right)^2 + O\left(\eta^4\right)$$

which immediately imply that

$$\lim_{\eta\to 0}\, \psi''(\eta) = \tfrac{1}{n} \tag{18c}$$

$$\lim_{\eta\to 0}\, \beta^2(\eta) = 0 \tag{18d}$$

Replacing (18c) and (18d) in (18b) gives,

$$\lim_{z\to 0}\, I_z(U,U) = \frac{1}{n}\, u_\eta^2 \tag{18e}$$

Note that, from (16a),

$$\|U\|^2 = u_\eta^2 + \eta^2\, \|u\|^2$$

by Pythagoras' theorem, since $\bar{x}$ and u are orthogonal. At the point $z = 0$, one has $\eta = 0$, so that

$$\|U\|^2_{z=0} = u_\eta^2$$

This shows that (18e) is the same as

$$\lim_{z\to 0}\, I_z(U,U) = \frac{1}{n}\, \|U\|^2 \tag{18f}$$

This limit does not depend on the path along which z tends to $z = 0$. Therefore, I_z extends smoothly to I_0, which is given by (17c), at the point $z = 0$. This shows that I is a well-defined Riemannian metric throughout the parameter space $\mathbb{R}^n$. ■

4.2 The Riemannian Gaussian model

The Riemannian Gaussian model was recently introduced as a means of describing unimodal populations of covariance matrices [14,41,42]. This model can be defined on any Riemannian symmetric space of non-positive sectional curvature. Let M be such a symmetric space and denote G its group of isometries. Then, the expressions appearing in (12a) are

$$D(x,\bar{x}) = d^2(x,\bar{x}) \qquad \eta(\sigma) = -\frac{1}{2\sigma^2} \tag{19}$$

where $d(x,\bar{x})$ denotes the Riemannian distance in M, and condition (12b) is verified since each isometry $g \in G$ preserves this Riemannian distance. The function $\psi(\eta)$ is a strictly convex function of $\eta \in (-\infty, 0)$, which can be expressed by means of a multiple integral [42], (see Proposition 1 in this reference). Precisely, $\psi(\eta)$ is the cumulant generating function of the squared Riemannian distance $d^2(x,\bar{x})$.

Proposition 1 cannot be applied directly to the Riemannian Gaussian model (19). This is because, in most cases of interest, the Riemannian symmetric space M is not irreducible. In such cases, before applying Proposition 1, it is necessary to introduce the De Rham decomposition theorem [40][22].

Remark 7 : assume the Riemannian symmetric space M is moreover simply-connected. Then, the De Rham decomposition theorem implies that M is a Riemannian product of irreducible Riemannian symmetric spaces [22] (Proposition 5.5, Page 310). Precisely, $M = M_1 \times \ldots \times M_r$ where each M_q is an irreducible Riemannian symmetric space, and the Riemannian metric and distance of M can be expressed as follows,

$$Q(u,u) = \sum_{q=1}^{r} Q(u_q, u_q) \tag{20a}$$

Q_q ? or explain $Q(u_q, u_q) = Q(0,\ldots,u_q,\ldots,0)$

$$d^2(x,y) = \sum_{q=1}^{r} d^2(x_q, y_q) \tag{20b}$$

d_q ?

where $x, y \in M$ are written $x = (x_1, \ldots, x_r)$ and $y = (y_1, \ldots, y_r)$ with $x_q, y_q \in M_q$, and where $u \in T_xM$ is written $u = u_1 + \ldots + u_r$ with $u_q \in T_{x_q}M_q$. Since M has non-positive sectional curvature, each M_q is either a Euclidean space, or a so-called space of non-compact type, having negative sectional curvature [22]. A concrete example of the De Rham decomposition is treated in Section 6, where $M = \mathcal{P}_n$ is the space of $n \times n$ real covariance matrices. ■

The following Proposition 3 gives the Rao-Fisher information metric of the Riemannian Gaussian model. Since this model is, in general, defined on a Riemannian symmetric space M which is not irreducible, the Rao-Fisher information metric turns out to be a multiply-warped Riemannian metric, rather than a warped Riemannian metric.

Remark 8 : in the notation of (20), a multiply-warped Riemannian metric I is a Riemannian metric defined on $\mathcal{M} = M \times (0,\infty)$, in the following way [18][45]

$$I_z(U,U) = (\alpha(\sigma)\,u_\sigma)^2 + \sum_{q=1}^{r} \beta_q^2(\sigma)\, Q_x(u_q, u_q) \tag{21}$$

for $z = (\bar{x},\sigma)$ and $U \in T_z\mathcal{M}$, where $U = u_\sigma \partial_\sigma + u$ with $u_\sigma \in \mathbb{R}$ and $u \in T_{\bar{x}}M$. Here, the functions α and β_q are positive functions defined on $(0,\infty)$. Clearly, when $r = 1$, definition (21) reduces to definition (2a) of a warped Riemannian metric. Therefore, warped Riemannian metrics are a special case of multiply-warped Riemannian metrics. ■

Proposition 3. *the Rao-Fisher information metric of the Riemannian Gaussian model is a multiply-warped Riemannian metric. In terms of $\bar{x} \in M$ and $\eta = -1/2\sigma^2$, this metric has the following expression*

$$I_z(U,U) = \psi''(\eta)\,u_\eta^2 + \sum_{q=1}^{r} \left(4\eta^2\psi_q'(\eta)/\dim M_q\right) Q_x(u_q,u_q) \tag{22}$$

where $U = u_\eta\partial_\eta + u$, and where $\psi_q(\eta)$ is the cumulant generating function of the squared Riemannian distance $d^2(x_q,\bar{x}_q)$.

Proof: assume first that the Riemannian symmetric space M is irreducible, so that Proposition 1 applies directly, and the Rao-Fisher information metric is given by (13a),

$$I_z(U,U) = \psi''(\eta)\,u_\eta^2 + \beta^2(\eta)\,Q_x(u,u) \tag{23a}$$

To obtain $\beta^2(\eta)$, replace into (13b) the fact that

$$\nabla_x D(x,\bar{x}) = -2\exp_x^{-1}(\bar{x}) \qquad Q(\nabla_x D, \nabla_x D) = 4\,d^2(x,\bar{x})$$

where exp denotes the Riemannian exponential mapping, corresponding to the Riemannian metric Q of M [13], (see Page 407). It then follows from (13b) that,

$$\beta^2(\eta) = 4\eta^2\,\mathbb{E}_z\, d^2(x,\bar{x})/\dim M = 4\eta^2\psi'(\eta)/\dim M \tag{23b}$$

where the second equality holds since $\psi(\eta)$ is the cumulant generating function of $d^2(x,\bar{x})$. From (23a) and (23b),

correct terminology ?

$$I_z(U,U) = \psi''(\eta)\,u_\eta^2 + \left(4\eta^2\psi'(\eta)/\dim M\right) Q_x(u,u) \tag{23c}$$

which is the same as (22) with $r = 1$. This proves the proposition in the special case where M is irreducible. For the general case where M is not irreducible, write $U = u_\eta\partial_\eta + u$, with $u = u_1 + \ldots + u_r$ as in Remark 7. It is possible to prove that,

$$q \neq p\;:\; I_z(u_p,u_q) = 0 \tag{24a}$$

$$u = u_p\;:\; I_z(U,U) = \psi''(\eta)\,u_\eta^2 + \left(4\eta^2\psi_p'(\eta)/\dim M_p\right) Q_x(u_p,u_p) \tag{24b}$$

Then, since the Rao-Fisher information metric I is bilinear and symmetric, (22) follows immediately, and the proposition is proved in the general case.

Proof of identities (24) : this is carried out using the following properties (25). Note first that the probability density function of the Riemannian Gaussian model is given by (12a) and (19),

$$p(x|\bar{x},\sigma) = \exp\left[\eta(\sigma)\,d^2(x,\bar{x}) - \psi(\eta(\sigma))\right] \tag{25a}$$

By substituting (20b) in this expression, it is seen that

$$p(x|\bar{x},\sigma) = \prod_{q=1}^{r} \exp\left[\eta(\sigma)\,d^2(x_q,\bar{x}_q) - \psi_q(\eta(\sigma))\right] = \prod_{q=1}^{r} p(x_q|\bar{x}_q,\sigma) \tag{25b}$$

where $\psi_q(\eta)$ is the cumulant generating function of $d^2(x_q,\bar{x}_q)$, as stated after (22). The last equality shows that $(x_q\,; q = 1,\ldots,r)$ are independent, and that each x_q has a Riemannian Gaussian density on the irreducible Riemannian symmetric space M_q, with parameters $z_q = (\bar{x}_q,\sigma)$. Now, identities (24) can be obtained from definition (8) of the Rao-Fisher information metric. To apply this definition, note from (25b), that the log-likelihood function $\ell(z)$ can be written,

$$\ell(z)(x) = \log p(x|z) = \sum_{q=1}^{r} \ell(z_q)(x_q) \quad \text{where} \quad \ell(z_q)(x_q) = \log p(x_q|z_q) \tag{25c}$$

Proof of (24a) : recall the polarisation identity, from elementary linear algebra [28], (see Page 29),

$$I_z(u_p,u_q) = \frac{1}{4}\,I_z(u_p+u_q, u_p+u_q) - \frac{1}{4}\,I_z(u_p-u_q, u_p-u_q)$$

By replacing (8) into this identity, it can be seen that,

$$I_z(u_p,u_q) = \mathbb{E}_z\left(\,(d\ell(z)\,u_p)\,(d\ell(z)\,u_q)\,\right) \tag{26}$$

petit abus de notation.

Using (25c), it is then possible to write

$$I_z(u_p,u_q) = \mathbb{E}_z\left(\,(d\ell(z_p)\,u_p)\,(d\ell(z_q)\,u_q)\,\right) = \mathbb{E}_{z_p}\left(d\ell(z_p)\,u_p\right)\,\mathbb{E}_{z_q}\left(d\ell(z_q)\,u_q\right)$$

Here, the first equality follows from (26), since $u_p \in T_{x_p}M_p$ and $u_q \in T_{x_q}M_q$, and the second equality holds since x_p and x_q are independent. Now, each one of the two expectations appearing on the right-hand side is equal to zero, since the expectation of the derivative of the log-likelihood must be zero [2], (see Page 28). This shows that (24a) holds. ■

Proof of (24b) : the condition $u = u_p$ implies $U = u_\eta\partial_\eta + u_p$. Replacing this in (8), it follows using (25c),

$$I_z(U,U) = \mathbb{E}_z\left(\sum_{q=1}^{r} d\ell(z_q)\,U\right)^2 = \mathbb{E}_z\left(\sum_{q=1}^{r} u_\eta\,\partial_\eta\ell(z_q) + d\ell(z_p)\,u_p\right)^2 \tag{27a}$$

where the second equality holds since $u_p \in T_{x_p}M_p$. Since the x_q are independent, it is clear from (25c) that the $\ell(z_q)$ are independent. Accordingly, by expanding the right-hand side of (27a),

$$I_z(U,U) = \sum_{q\neq p} u_\eta^2\,\mathbb{E}_{z_q}\left(\partial_\eta\ell(z_q)\right)^2 + \mathbb{E}_{z_p}\left(\,d\ell(z_p)\,U\right)^2 \tag{27b}$$

→ étape 27a → 27b

Applying (14c) from the proof of Proposition 1 to each term in the sum over $q \neq p$, it follows that

$$I_z(U,U) = \sum_{q\neq p} \psi_q''(\eta)\,u_\eta^2 + \mathbb{E}_{z_p}\left(\,d\ell(z_p)\,U\right)^2 \tag{27c}$$

By (8), the expectation appearing in the second term is given by the Rao-Fisher information metric of the Riemannian Gaussian model on the irreducible Riemannian symmetric space M_p. This can be replaced from (23c), so that

$$\begin{aligned} I_z(U,U) &= \textstyle\sum_{q\neq p} \psi_q''(\eta)\,u_\eta^2 + \psi_p''(\eta)\,u_\eta^2 + \left(4\eta^2\psi_p'(\eta)/\dim M_p\right) Q_x(u_p,u_p) \\ &= \textstyle\sum_q \psi_q''(\eta)\,u_\eta^2 \qquad\quad + \left(4\eta^2\psi_p'(\eta)/\dim M_p\right) Q_x(u_p,u_p) \end{aligned}$$

This immediately yields (24b), upon noting from (25a) and (25b) that $\psi(\eta) = \Sigma_q\,\psi_q(\eta)$. ■

Now, since identities (24) have been proved, (22) follows from the fact that the Rao-Fisher information metric I is bilinear and symmetric. ■

5 The generalised Mahalanobis distance

This section builds on Remark 3, made at the end of Section 2, in order to generalise the definition of the classical Mahalanobis distance, to the context of a location-scale model $\mathcal{P}$ defined on a Riemannian symmetric space M.

To begin, assume that, as in Theorem 1, the Riemannian symmetric space M is irreducible and the location-scale model $\mathcal{P}$ verifies condition (7). Then, according to Theorem 1, the Rao-Fisher information metric I of the model $\mathcal{P}$ is a warped Riemannian metric on the parameter space $\mathcal{M}$.

Recall from Remark 3 that this warped Riemannian metric I induces an extrinsic Riemannian metric Q^σ on M, for each $\sigma \in (0,\infty)$. The *generalised Mahalanobis distance* is defined to be the Riemannian distance on M which is induced by the extrinsic Riemannian metric Q^σ. The generalised Mahalanobis distance between $\bar{x}$ and $\bar{y}$ in M is denoted $d(\bar{x},\bar{y}|\sigma)$. It is given by the following proposition.

Proposition 4. *the generalised Mahalanobis distance $d(\bar{x},\bar{y}|\sigma)$ between $\bar{x}$ and $\bar{y}$ in M is given by*

$$d(\bar{x},\bar{y}|\sigma) = \beta(\sigma)\,d(\bar{x},\bar{y}) \tag{28}$$

where the function $\beta(\sigma)$ is given by (9), and where $d(\bar{x},\bar{y})$ denotes the Riemannian distance in M.

je ne comprends pas cette partie. juste un scaling ? de $d(\bar{x},\bar{y})$?

Uniquement intéressant dans le cas non irréductible ?
↳ car les $\beta_q^2(\sigma)$ ~~dép~~ dépendent du même paramètre σ ?

Remark 9 : the generalised Mahalanobis distance (28) reduces to the classical Mahalanobis distance, when $\mathcal{P}$ is the isotropic normal model on $M = \mathbb{R}^d$. In this case, the Rao-Fisher metric I is given by (1c) in the introduction, so that $\beta(\sigma) = 1/\sigma$. Replacing this in (28) yields

$$d(\bar{x},\bar{y}\,|\sigma) = \frac{1}{\sigma}\,\|\bar{x}-\bar{y}\| \tag{29}$$

where $\|\bar{x}-\bar{y}\|$ is the Euclidean distance in $M = \mathbb{R}^d$. Now, (29) is the classical Mahalanobis distance [36]. ■

Proof of Proposition 4 : the extrinsic metric Q^σ is induced by the warped Riemannian metric I, which is given by (9). Therefore, it follows from (5) that

$$Q^\sigma_x(u,u) = \beta^2(\sigma)\,Q_x(u,u) \tag{30a}$$

where the function $\beta(\sigma)$ is given by (9). To find the generalised Mahalanobis distance between $\bar{x}$ and $\bar{y}$ in M, let $c(t)$ be a curve in M with $c(0) = \bar{x}$ and $c(1) = \bar{y}$. Denote by $L(c|\sigma)$ and by $L(c)$ the length of this curve, as measured by the Riemannian metrics Q^σ and Q, respectively. Then,

$$L(c|\sigma) = \int_0^1 \left(Q^\sigma(\dot{c},\dot{c})\right)^{1/2} dt = \beta(\sigma)\int_0^1 \left(Q(\dot{c},\dot{c})\right)^{1/2} dt = \beta(\sigma)\,L(c) \tag{30b}$$

where the second equality follows from (30a). To obtain (28), it is enough to write

$$d(\bar{x},\bar{y}\,|\sigma) = \inf_c L(c|\sigma) = \beta(\sigma)\inf_c L(c) = \beta(\sigma)\,d(\bar{x},\bar{y}) \tag{30c}$$

where the infimum is over all curves $c(t)$ as above, and the second equality follows from (30b). ■

Expression (28) of the generalised Mahalanobis distance is valid only under the assumption that the Riemannian symmetric space M is irreducible. This assumption does not hold, when the model $\mathcal{P}$ is the Riemannian Gaussian model studied in 4.2. For this model, an alternative expression of the generalised Mahalanobis distance is given in Proposition 5 below.

As in 4.2, let $\mathcal{P}$ be the Riemannian Gaussian model on a Riemannian symmetric space M, where M is simply-connected and has non-positive sectional curvature. Proposition 3 states that the Rao-Fisher information metric I of the model $\mathcal{P}$ is a multiply-warped Riemannian metric on the parameter space $\mathcal{M}$. For each $\sigma \in (0,\infty)$, this multiply-warped Riemannian metric I induces an extrinsic Riemannian metric Q^σ on M. Precisely, Q^σ can be obtained from (22) of Proposition 3,

$$Q^\sigma_x(u,u) = \sum_{q=1}^{r} \beta^2_q(\sigma)\,Q_x(u_q,u_q) \qquad \beta^2_q(\sigma) = 4\eta^2\psi'_q(\eta)/\dim M_q \tag{31}$$

The generalised Mahalanobis distance $d(\bar{x},\bar{y}\,|\sigma)$ is the Riemannian distance between $\bar{x}$ and $\bar{y}$ in M, induced by this extrinsic Riemannian metric Q^σ.

Proposition 5. *when $\mathcal{P}$ is the Riemannian Gaussian model, the generalised Mahalanobis distance $d(\bar{x},\bar{y}\,|\sigma)$ between $\bar{x}$ and $\bar{y}$ in M is given by*

$$d^2(\bar{x},\bar{y}\,|\sigma) = \sum_{q=1}^{r} \beta^2_q(\sigma)\,d^2(\bar{x}_q,\bar{y}_q) \tag{32}$$

where the notation is that of (20b).

Proof: the proof hinges on the fact that the extrinsic Riemannian metric Q^σ of (31) is an invariant Riemannian metric on M. In other words, if G is the group of isometries of M, then

$$Q^\sigma_{g\cdot x}(dg_x\,u, dg_x\,u) = Q^\sigma_x(u,u) \qquad \text{for all } g \in G \tag{33a}$$

where dg_x is the derivative of the isometry g at the point $\bar{x}$. The proof of (33a) is not detailed here. It follows since the Riemannian metric Q is also an invariant Riemannian metric on M, so that Q also verifies (33a), and since Q^σ is related to Q by (31). A general result in [22] (Corollary 4.3, Page 182), states that all invariant Riemannian metrics on M have the same geodesics. In particular, the metrics Q^σ and Q have the same geodesics, and therefore the same Riemannian exponential mapping exp. To find the generalised Mahalanobis distance between $\bar{x}$ and $\bar{y}$ in M, let $u = \exp^{-1}_{\bar{x}}(\bar{y})$, and note that

$$d^2(\bar{x},\bar{y}\,|\sigma) = Q^\sigma_x(u,u) = \sum_{q=1}^{r} \beta^2_q(\sigma)\,Q_x(u_q,u_q) \tag{33b}$$

where the second equality follows from (31). Now, to prove (32) it is enough to prove that

$$Q_x(u_q,u_q) = d^2(\bar{x}_q,\bar{y}_q) \tag{33c}$$

Indeed, (32) is then obtained by replacing (33c) into (33b). The proof of (33c) follows by writing, as in (33b),

$$d^2(\bar{x},\bar{y}) = Q_x(u,u) = \sum_{q=1}^{r} Q_x(u_q,u_q) = \sum_{q=1}^{r} d^2(\bar{x}_q,\bar{y}_q) \tag{33d}$$

where the second equality follows from (20a), and the third equality follows from (20b). Since (33d) is an identity which holds for arbitrary $\bar{x} = (\bar{x}_1,\ldots,\bar{x}_r)$ and $\bar{y} = (\bar{y}_1,\ldots,\bar{y}_r)$, it follows that (33c) must hold true, as required. ■

Remark 10 : the generalised Mahalanobis distance, whether given by (28) or by (32), is an invariant Riemannian distance on M,

$$d(g\cdot\bar{x}, g\cdot\bar{y}\,|\sigma) = d(\bar{x},\bar{y}\,|\sigma) \qquad \text{for all } g \in G \tag{34}$$

This follows since the Riemannian distance in M is invariant : $d(g\cdot\bar{x}, g\cdot\bar{y}) = d(\bar{x},\bar{y})$. ■

6 A concrete example for the Riemannian Gaussian model

The aim of this section is to illustrate the geometric concepts involved in Propositions 3 and 5, by applying these concepts to the concrete example of the Riemannian Gaussian model defined on $M = \mathcal{P}_n$, the space of $n\times n$ real covariance matrices.

The space $\mathcal{P}_n$ is a Riemannian symmetric space, which is simply-connected and has non-positive sectional curvature [22][44]. It is usually equipped with its affine-invariant Riemannian metric [44][6],

$$Q_x(u,u) = \operatorname{tr}[\bar{x}^{-1}u]^2 \qquad \bar{x} \in \mathcal{P}_n,\ u \in T_x\mathcal{P}_n \tag{35a}$$

This metric is invariant under the action of the group of isometries $G = GL(n,\mathbb{R})$ on $\mathcal{P}_n$, which is given by affine transformations,

$$g\cdot\bar{x} = g\,\bar{x}\,g^t \tag{35b}$$

where ^{t} denotes the transpose. Moreover, this metric induces a Riemannian distance on $\mathcal{P}_n$, which is given by,

$$d^2(\bar{x},\bar{y}) = \operatorname{tr}\left[\log\left(\bar{x}^{-1/2}\,\bar{y}\,\bar{x}^{-1/2}\right)\right]^2 \tag{35c}$$

This distance is also invariant under the action of the group $GL(n,\mathbb{R})$ on $\mathcal{P}_n$. In other words : $d(g\cdot\bar{x}, g\cdot\bar{y}) = d(\bar{x},\bar{y})$.

The Riemannian Gaussian model on $\mathcal{P}_n$ is given by the probability density function [14][41]

$$p(x|\bar{x},\sigma) = Z^{-1}(\sigma)\exp\left[-\frac{d^2(x,\bar{x})}{2\sigma^2}\right] \tag{36a}$$

which is a probability density function with respect to the invariant volume element associated to the Riemannian metric (35a). The normalising factor $Z(\sigma)$ can be expressed as a multiple integral [41], (see Proposition 4 in this reference),

$$Z(\sigma) = \mathrm{C}_n \int_{\mathbb{R}^n} e^{-\|r\|^2/2\sigma^2} \prod_{i<j} \sinh\left(|r_i - r_j|/2\right)\, dr_1 \ldots dr_n \tag{36b}$$

where C_n is a numerical constant which only depends on n, and the integration variable is denoted $r = (r_1,\ldots,r_n) \in \mathbb{R}^n$. If $\eta(\sigma) = -1/2\sigma^2$, then $\psi(\eta) = \log Z(\sigma)$ is a strictly convex function of $\eta \in (-\infty,0)$.

With (35) and (36) in mind, consider the application of Proposition 3 to the Riemannian Gaussian model on $\mathcal{P}_n$. This will lead to the expression of the Rao-Fisher information metric I of this model.

De Rham decomposition of $\mathcal{P}_n$: recall first that Proposition 3 uses the De Rham decomposition, introduced in Remark 7. For the Riemannian symmetric space $\mathcal{P}_n$, the De Rham decomposition states that $\mathcal{P}_n$ is a Riemannian product of irreducible Riemannian symmetric spaces $\mathcal{P}_n = \mathbb{R}\times S\mathcal{P}_n$, where $S\mathcal{P}_n$ is the set of $\bar{s} \in \mathcal{P}_n$ such that $\det(\bar{s}) = 1$. The identification of $\mathcal{P}_n$ with $\mathbb{R}\times S\mathcal{P}_n$ is obtained by identifying each $\bar{x} \in \mathcal{P}_n$ with a couple $(\bar{\tau},\bar{s})$, where $\bar{\tau} \in \mathbb{R}$ and $\bar{s} \in S\mathcal{P}_n$ are given by

R and SP_n such that $\mathbb{R}\times SP_n$.

$$\bar{\tau} = \log\det(\bar{x}) \qquad \bar{s} = e^{-\tau/n}\,\bar{x} \tag{37a}$$

Note that the spaces $\mathbb{R}$ and $S\mathcal{P}_n$ are indeed irreducible Riemannian symmetric spaces. This is clear for $\mathbb{R}$, which is one-dimensional and cannot be decomposed into a product of lower-dimensional spaces. The fact that $S\mathcal{P}_n$ is irreducible can be found in [22], (Table II, Page 354). It will be convenient to write $\bar{x} = (\bar{x}_1,\bar{x}_2)$ where $\bar{x}_1 = \bar{\tau}$ and $\bar{x}_2 = \bar{s}$. If $u \in T_x\mathcal{P}_n$, then $u = u_1 + u_2$,

$$u_1 = \frac{1}{n}\operatorname{tr}(\bar{x}^{-1}u)\,\bar{x} \qquad u_2 = u - \frac{1}{n}\operatorname{tr}(\bar{x}^{-1}u)\,\bar{x} \tag{37b}$$

Here, $u_1 \in T_{\bar{x}_1}\mathbb{R}$, where $T_{\bar{x}_1}\mathbb{R} \subset T_x\mathcal{P}_n$ is the one-dimensional subspace consisting of symmetric matrices v of the form $v = t\,\bar{x}$ with t any real number. On the other hand, $u_2 \in T_{\bar{x}_2}S\mathcal{P}_n$, where $T_{\bar{x}_2}S\mathcal{P}_n \subset T_x\mathcal{P}_n$ is the subspace consisting of symmetric matrices v which satisfy $\operatorname{tr}(\bar{x}^{-1}v) = 0$. Using (37a) and (37b), (20a) and (20b) of Remark 7 can be written down,

$$Q_x(u,u) = Q_x(u_1,u_1) + Q_x(u_2,u_2) \tag{37c}$$

$$d^2(\bar{x},\bar{y}) = \frac{1}{n}\,|\bar{x}_1 - \bar{y}_1|^2 + d^2(\bar{x}_2,\bar{y}_2) \tag{37d}$$

where Q_x is the affine-invariant metric (35a) and $d(\bar{x},\bar{y})$ or $d(\bar{x}_2,\bar{y}_2)$ is the Riemannian distance (35c). The proof of formulae (37c) and (37d) is a direct calculation, and is not detailed here.

The Rao-Fisher metric I : according to (22) of Proposition 3, the Rao-Fisher information metric I of the Riemannian Gaussian model on $\mathcal{P}_n$ is given by,

$$I_z(U,U) = \psi''(\eta)\,u_\eta^2 + \frac{4\eta^2\psi'_1(\eta)}{\dim\mathbb{R}}\,Q_x(u_1,u_1) + \frac{4\eta^2\psi'_2(\eta)}{\dim S\mathcal{P}_n}\,Q_x(u_2,u_2) \tag{38a}$$

for $z = (\bar{x},\sigma)$ in the parameter space $\mathcal{M} = \mathcal{P}_n\times(0,\infty)$, and for $U = u_\eta\,\partial_\eta + u$ where $u = u_1 + u_2$ is given by (37b). Indeed, (38a) results from (22), by putting

$r = 2$, as well as $M_1 = \mathbb{R}$ and $M_2 = SP_n$. The functions appearing in (38a) are $\psi(\eta) = \log Z(\sigma)$ with $Z(\sigma)$ given by (36b), and, as shown in Remark 11 below.

$$\psi_1(\eta) = \frac{1}{2}\log(2\pi n) - \frac{1}{2}\log(-2\eta) \qquad \psi_2(\eta) = \psi(\eta) - \psi_1(\eta) \tag{38b}$$

Moreover, $\dim \mathbb{R} = 1$ and $\dim SP_n = \dim \mathcal{P}_n - 1 = n(n+1)/2 - 1$. Replacing into (38a) gives,

$$I_z(U,U) = \psi''(\eta)\, u_\eta^2 - 2\eta\, Q_z(u_1, u_1) + \frac{8\eta^2\psi_2'(\eta)}{n^2+n-2} Q_z(u_2,u_2) \tag{38c}$$

This expression of the Rao-Fisher information metric of the Riemannian Gaussian model on $\mathcal{P}_n$ can be computed directly from (35a), (37b) and (38b), once the function $\psi(\eta)$ is known. This function $\psi(\eta)$ has been tabulated for values of n up to $n = 50$, using a Monte Carlo method which was developed specifically for the evaluation of (36b) [50].

Remark 11 : assume x follows the Riemannian Gaussian probability density (36a) on $\mathcal{P}_n$. If $x = (x_1, x_2)$ where $x_1 \in \mathbb{R}$ and $x_2 \in SP_n$, then the densities of x_1 and x_2 can be found by replacing (37d) into (36a). Precisely, this gives

$$p(x|\bar{x},\sigma) \propto \exp\left[-\frac{|x_1 - \bar{x}_1|^2}{2n\sigma^2}\right] \times \exp\left[-\frac{d^2(x_2,\bar{x}_2)}{2\sigma^2}\right]$$

It follows from this decomposition that x_1 and x_2 are independent, and that x_1 follows a univariate normal distribution of mean $\bar{x}_1$ and of variance $n\sigma^2$. In particular, the moment generating function $\psi_1(\eta)$ of the squared distance $|x_1 - \bar{x}_1|^2$ has the expression stated in (38b). ■

The generalised Mahalanobis distance on $\mathcal{P}_n$: applying Proposition 5 will yield the expression of the generalised Mahalanobis distance on $\mathcal{P}_n$. The Rao-Fisher information metric I as given by (38c) induces an extrinsic Riemannian metric Q^σ on $\mathcal{P}_n$, for each $\sigma \in (0,\infty)$,

$$Q^\sigma_z(u,u) = -2\eta\, Q_z(u_1,u_1) + \frac{8\eta^2\psi_2'(\eta)}{n^2+n-2} Q_z(u_2,u_2) \qquad \eta = -\frac{1}{2\sigma^2} \tag{39a}$$

The generalised Mahalanobis distance on $\mathcal{P}_n$ is the Riemannian distance induced on $\mathcal{P}_n$ by the extrinsic Riemannian metric Q^σ. If the generalised Mahalanobis distance between $\bar{x}$ and $\bar{y}$ in $\mathcal{P}_n$ is denoted $d(\bar{x},\bar{y}\,|\sigma)$, then (32) of Proposition 5, along with (39a), imply

petit

$$d^2(\bar{x},\bar{y}\,|\sigma) = \frac{|\bar{x}_1 - \bar{y}_1|^2}{n\sigma^2} + \frac{4\psi_2'\left(-\frac{1}{2\sigma^2}\right)}{(n^2+n-2)\sigma^4} d^2(\bar{x}_2,\bar{y}_2) \tag{39b}$$

$\to$ 2 ? Erreur ?

$Q^\sigma_{\bar{y}} = \sum_q \beta_q^2(\sigma) Q_{\bar{x}}(v_q, v_q)$

$\Leftrightarrow d^2(\bar{x},\bar{y}\,|\sigma) = \sum \beta_q^2\, d_q^2(\bar{x}_q, \bar{y}_q)$

This distance can be computed directly from (35c), (37a) and (38b), once the function $\psi(\eta)$ has been tabulated using the Monte Carlo method of [50], or computed in any other way.

Affine invariance of the generalised Mahalanobis distance : the affine-invariant Riemannian metric Q of (35a) is well-known to the information science community, having been introduced in [39]. Besides the metric Q, a whole new family of affine-invariant Riemannian metrics Q^σ is provided by (39a). Indeed, to say that Q is affine-invariant means that it is invariant under affine transformations (35b). In other words

$$Q_{g\cdot z}(dg_z\, u, dg_z\, u) = Q_z(u,u) \qquad \text{for all } g \in GL(n,\mathbb{R}) \tag{40a}$$

where dg_z denotes the derivative of the affine transformation (35b) at the point $\bar{x} \in \mathcal{P}_n$. On the other hand, it is shown in Remark 12 below that each one of the metrics Q^σ also verifies (40a), so that

$$Q^\sigma_{g\cdot z}(dg_z\, u, dg_z\, u) = Q^\sigma_z(u,u) \qquad \text{for all } g \in GL(n,\mathbb{R}) \tag{40b}$$

This means that each one of the metrics Q^σ is an affine-invariant Riemannian metric, as claimed. Furthermore, the fact that the metric Q^σ is invariant under affine transformations implies that the generalised Mahalanobis distance (39b) is also invariant under these transformations,

$$d(g\cdot\bar{x}, g\cdot\bar{y}\,|\sigma) = d(\bar{x},\bar{y}\,|\sigma) \qquad \text{for all } g \in GL(n,\mathbb{R}) \tag{40c}$$

This is because the generalised Mahalanobis distance (39b) is the Riemannian distance induced on $\mathcal{P}_n$ by Q^σ.

Remark 12 : to prove that (40b) holds for each one of the metrics Q^σ, write (39a) in the form

$$Q^\sigma_z(u,u) = \beta_1^2(\sigma)\, Q_z(u_1,u_1) + \beta_2^2(\sigma)\, Q_z(u_2,u_2) \tag{41a}$$

Let $g \in GL(n,\mathbb{R})$ and let $\bar{y} = g\cdot\bar{x}$ and $v = dg_z\, u$, so that $v \in T_y\mathcal{P}_n$. Using (41a), the left-hand side of (40b) takes on the form

$$Q^\sigma_y(v,v) = \beta_1^2(\sigma)\, Q_y(v_1,v_1) + \beta_2^2(\sigma)\, Q_y(v_2,v_2) \tag{41b}$$

Note from (35b) that

$$\bar{y} = g\cdot\bar{x} = g\,\bar{x}\,g^t \qquad v = dg_z\, u = g\,u\,g^t \tag{41c}$$

Therefore, it follows from (37b) that v_1 is given by

$$v_1 = \frac{1}{n}\operatorname{tr}[\bar{y}^{-1}v]\,\bar{y} = \frac{1}{n}\operatorname{tr}[\bar{x}^{-1}u]\,\bar{y}$$

where the second equality follows by a direct calculation. From (41c), it is now seen that

$$v_1 = \frac{1}{n}\operatorname{tr}[\bar{x}^{-1}u]\, g\,\bar{x}\,g^t = g\,u_1\,g^t = dg_z\, u_1 \tag{41d}$$

where the second equality follows from (37b). Since $v = v_1 + v_2$, this implies that v_2 is given by

$$v_2 = v - v_1 = g\,u\,g^t - g\,u_1\,g^t = g\,(u - u_1)\,g^t = g\,u_2\,g^t = dg_z\, u_2 \tag{41e}$$

where the second equality follows from (41c) and (41d). Using (40a), it follows from (41d) and (41e) that

$$Q_y(v_1,v_1) = Q_z(u_1,u_1) \qquad Q_y(v_2,v_2) = Q_z(u_2,u_2) \tag{41f}$$

Finally, replacing (41f) into (41b), it is found that

$$Q^\sigma_y(v,v) = \beta_1^2(\sigma)\, Q_z(u_1,u_1) + \beta_2^2(\sigma)\, Q_z(u_2,u_2) = Q^\sigma_z(u,u)$$

where the second equality follows from (41a). Recalling the definitions of $\bar{y}$ and v, it is clear that the proof of (40b) is now complete. ■

7 The solution of the geodesic equation

The present section provides the solution of the geodesic equation of a multiply-warped Riemannian metric. The main result is the following Proposition 6. This proposition shows that the solution of the geodesic equation of a multiply-warped Riemannian metric, for given initial conditions, reduces to the solution of a one-dimensional second-order differential equation. As stated in Remark 8, warped Riemannian metrics are a special case of multiply-warped Riemannian metrics. Therefore, Proposition 6 also applies to the solution of the geodesic equation of a warped Riemannian metric. This special case of warped Riemannian metrics was treated separately in [38].

Let I be a multiply-warped Riemannian metric defined on $\mathcal{M} = M \times (0,\infty)$, in the notation of (21),

$$I_z(U,U) = (\alpha(\sigma)\, u_\sigma)^2 + \sum_{q=1}^{r} \beta_q^2(\sigma)\, Q_z(u_q,u_q) \tag{42}$$

for $z = (\bar{x},\sigma)$ and $U \in T_z\mathcal{M}$, with $u = u_\sigma\partial_\sigma + u$ and $u = u_1 + \ldots + u_r$. As in (2b) of Section 2, introduce the vertical distance coordinate r, which is defined by $dr/d\sigma = \alpha(\sigma)$.

Proposition 6. *let $\gamma(t)$ be a geodesic of the multiply-warped Riemannian metric I, with initial conditions $\gamma(0) = z$ and $\dot\gamma(0) = U$, and let $\gamma(t) = (\bar{x}(t),\sigma(t))$ and $r(t) = r(\sigma(t))$. Then, $r(t)$ verifies the second-order differential equation*

$$\ddot{r} = -\frac{1}{2}\frac{d}{dr}V(r) \qquad V(r) = \sum_{q=1}^{r} \frac{\beta_q^2(r(0))}{\beta_q^2(r)} I_z(u_q,u_q) \tag{43a}$$

and $\bar{x}(t)$ is given by

$$\bar{x}(t) = \exp_z\left[\sum_{q=1}^{r}\left(\int_0^t \frac{\beta_q^2(r(0))}{\beta_q^2(r(s))}ds\right) u_q\right] \tag{43b}$$

where exp *denotes the Riemannian exponential mapping of the metric Q on M.*

Proof : the proof is given in Appendix C. It is a generalisation of the proof dealing with the special case of warped Riemannian metrics, which can be found in [38] (Proposition 38, Page 208). ■

Proposition 6 shows that the main difficulty, involved in computing a geodesic $\gamma(t)$ of the multiply-warped Riemannian metric I, lies in the solution of the second order differential equation (43a). Indeed, once this equation is solved, computing $\gamma(t)$ essentially reduces to an application of exp, which is the Riemannian exponential mapping of the metric Q on M. In the context of the present contribution, Q is an invariant metric on M, where M is a Riemannian symmetric space. Therefore, exp has a straightforward expression [22] (Theorem 3.3, Page 173). In particular, for the examples treated in Section 4, the expression of exp is well-known in the literature. For the von Mises-Fisher model, this expression is elementary, since geodesics on a sphere in Euclidean space are the great circles on this sphere. For the Riemannian Gaussian model, when this model is defined on the space $M = \mathcal{P}_n$ of $n \times n$ real covariance matrices, the expression of exp is widely used in the literature, as found in [39].

Remark 13 : the differential equation (43a) is the equation of motion of a one-dimensional conservative mechanical system. As such, its solution can be carried out by quadrature [21], (see Page 11). Precisely, the solution reduces to computing the integral

$$t = \pm \int_{r(0)}^{r(t)} \frac{dr}{\sqrt{E - V(r)}} \tag{44a}$$

where the total energy E is a conserved quantity, in the sense that $\dot{E} = 0$, and it can be shown that $E = I_z(U,U)$. Recalling that $dr/d\sigma = \alpha(\sigma)$, this integral can be written

$$t = \pm \int_{\sigma(0)}^{\sigma(t)} \frac{\alpha(\sigma)}{\sqrt{E - V(\sigma)}}\, d\sigma \qquad V(\sigma) = V(r(\sigma)) \tag{44b}$$

Here, if t is interpreted as time, then the integral on the right-hand side gives the time necessary to go from $\sigma(0)$ to $\sigma(t)$. In particular, replacing $\sigma(t)$ by ∞ and by 0 gives the two quantities

$$t_{\infty} = \int_{\sigma(0)}^{\infty} \frac{\alpha(\sigma)}{\sqrt{E - V(\sigma)}}\, d\sigma \qquad t_{0} = \int_{0}^{\sigma(0)} \frac{\alpha(\sigma)}{\sqrt{E - V(\sigma)}}\, d\sigma \tag{44c}$$

where t_∞ is the time necessary for $\sigma(t)$ to reach the value $\sigma = \infty$, and t_0 is the time necessary for $\sigma(t)$ to reach the value $\sigma = 0$. Since $\mathcal{M} = M \times (0,\infty)$, these two values $\sigma = \infty$ and $\sigma = 0$ are excluded from $\mathcal{M}$. Therefore, the geodesic $\gamma(t)$ cannot be extended beyond the time $t = \min(t_\infty, t_0)$, as it would then escape from $\mathcal{M}$. ■

Remark 14: a vertical geodesic is a geodesic $\gamma(t)$ for which $\dot\gamma(0) = U$ with $U = u_\sigma \partial_\sigma$. This means that all the u_x are zero. In (43a), this implies that $V(r) = 0$, so that $\ddot r = 0$ and $r(t)$ is an affine function of t. In (43b), this implies that $\bar x(t) = \bar x$ is constant. In Remark 2 of Section 2, it was shown that a vertical geodesic is a unique length-minimising geodesic. For a vertical geodesic, (44c) reads

$$t_{\infty} = \frac{1}{\sqrt{E}} \int_{\sigma(0)}^{\infty} \alpha(\sigma)\, d\sigma \qquad t_{0} = \frac{1}{\sqrt{E}} \int_{0}^{\sigma(0)} \alpha(\sigma)\, d\sigma \tag{45a}$$

These formulae provide another way of understanding conditions (3b) from Section 2. Precisely, since $dr/d\sigma = \alpha(\sigma)$, it is clear that t_∞ and t_0 are given by

$$t_{\infty} = \frac{1}{\sqrt{E}} \lim_{\sigma\to\infty} r(\sigma) - r(\sigma(0)) \qquad t_{0} = \frac{1}{\sqrt{E}} \lim_{\sigma\to 0} r(\sigma(0)) - r(\sigma) \tag{45b}$$

Thus, the first condition in (3b) is equivalent to $t_\infty = \infty$, which means that $\sigma(t)$ cannot reach the value $\sigma = \infty$ within a finite time, and the second condition in (3b) is equivalent to $t_0 = \infty$, which means that $\sigma(t)$ cannot reach the value $\sigma = 0$ within a finite time. ■

8 The construction of Riemannian Brownian motion

The present section provides an explicit construction of the Riemannian Brownian motion associated to any multiply-warped Riemannian metric. The main result is the following Proposition 7, which shows that this construction reduces to the solution of a one-dimensional stochastic differential equation. This is in analogy with Proposition 6 of the previous section.

Let I be a multiply-warped Riemannian metric on $\mathcal{M} = M \times (0,\infty)$, given by (42). Recall that a Riemannian Brownian motion associated to I is a diffusion process z in $\mathcal{M}$ which satisfies the identity, see [24][23],

$$df(z(t)) = \frac{1}{2}\Delta_{\mathcal{M}} f(z(t)) + dm^f(t) \tag{46a}$$

for each smooth function f on $\mathcal{M}$, where $\Delta_{\mathcal{M}}$ is the Laplace-Beltrami operator of I, and where m^f is a local martingale with respect to the augmented natural filtration of z. In other words, z is a Riemannian Brownian motion associated to I, if z solves the martingale problem associated to $(1/2)\Delta_{\mathcal{M}}$. The Laplace-Beltrami operator $\Delta_{\mathcal{M}}$ is given by the following formula, which will be proved in Appendix D,

show dependence on r

$$\Delta_{\mathcal{M}} f = \frac{1}{G}\,\partial_r\left(G\,\partial_r f\right) + \sum_{q=1}^{r} \frac{1}{\beta_q^2(r)}\,\Delta_{M_q} f \tag{46b}$$

where $G = \prod_{q=1}^{r} \beta_q^{\dim M_q}$. Equation (46a) defines a Riemannian Brownian motion process z, associated to the multiply-warped Riemannian metric I, through its relationship to the Laplace-Beltrami operator $\Delta_{\mathcal{M}}$. On the other hand, the following Proposition 7 shows how such a Riemannian Brownian motion process may be constructed explicitly.

Proposition 7. *Let z be a stochastic process with values in $\mathcal{M}$, and write $z(t) = (\bar x(t), \sigma(t))$ where $\bar x(t) = (\bar x_1(t), \ldots, \bar x_r(t))$ and $\bar x_q(t) \in M_q$. Let r and $(\theta_1, \ldots, \theta_r)$ be independent diffusion processes, where r is a one-dimensional diffusion process, which satisfies the stochastic differential equation*

$$dr(t) = \frac{1}{2}\frac{\partial_r G}{G}(r(t))dt + dw(t) \tag{47a}$$

with w a standard Brownian motion, and where θ_q is a Riemannian Brownian motion in M_q. If each $\bar x_q$ is a time-changed version of θ_q,

$$\bar x_q(t) = (\theta_q \circ \tau_q)(t) \qquad \tau_q(t) = \int_0^t \frac{ds}{\beta_q^2(r(s))} \tag{47b}$$

and if $\sigma(t) = \sigma(r(t))$, then the process z is a Riemannian Brownian motion associated to the multiply-warped Riemannian metric I given by (42).

Proof: the proof of this proposition is given in Appendix D. ■

Remark 15: the above Proposition 7 can be used for numerical simulation of a Riemannian Brownian motion z associated to a multiply-warped Riemannian metric I. According to Proposition 7, to simulate z, it is enough to simulate the solution r of the one-dimensional stochastic differential equation (47a), and then to independently simulate each θ_q as a Riemannian Brownian motion in the Riemannian symmetric space M_q. To simulate the solution r of (47a), it is enough to use any of the numerical schemes described in [26].

On the other hand, Riemannian Brownian motion in a symmetric space can be simulated using Lie group stochastic exponentials [4, 20, 31]. A detailed description of these methods falls outside the present scope, and is reserved for future work. ■

9 Surprising observation: Hadamard manifolds

In Section 2, the completeness and curvature of a warped Riemannian metric I were characterised by Formulae (3b) and (4), respectively. Here, based on 4.1, these formulae will be applied to the case where I is the Rao-Fisher information metric of the von Mises-Fisher model defined on S^{n-1}. Precisely, this application is carried out using a mixture of analytical and numerical computations, for the von Mises-Fisher model defined on S^{n-1} where $n = 2, \ldots, 8$. The result is a surprising observation: the parameter space of the von Mises-Fisher model, when equipped with the Rao-Fisher information metric I, becomes a Hadamard manifold, a simply-connected complete Riemannian manifold of negative sectional curvature [40][13]. Since this observation is true for several values of n, it gives rise to a family of Hadamard manifolds. Part of this claim can be proved for any value of $n = 2, \ldots$, as in the following proposition.

Proposition 8. *for any value of $n = 2, \ldots$, the parameter space of the von Mises-Fisher model defined on S^{n-1} is a simply-connected manifold, which moreover becomes a complete Riemannian manifold when equipped with the Rao-Fisher information metric I.*

Proof: recall from Remark 6 that the parameter space of the von Mises-Fisher model defined on S^{n-1} is identified with $\mathbb{R}^n$. Of course, $\mathbb{R}^n$ is a simply-connected manifold [43]. Thus, to prove the proposition, it remains to prove that the parameter space $\mathbb{R}^n$ becomes a complete Riemannian manifold when equipped with the Rao-Fisher information metric I. This will be done by proving that all geodesics of the metric I which pass through the point $z = 0$ in $\mathbb{R}^n$ can be extended indefinitely. Then, a corollary of the Hopf-Rinow theorem [13] (Corollary I.7.2, Page 29) implies the required completeness of the parameter space $\mathbb{R}^n$.

First, note that the geodesics of the metric I which pass through the point $z = 0$ are exactly the radial straight lines in $\mathbb{R}^n$. Indeed, according to Remark 6, if $\gamma(t)$ is a geodesic of I where $\gamma(t) = (\bar x(t), \eta(t))$, then $\gamma(t)$ is identified with the curve $z(t) = \eta(t)\,\bar x(t)$ in $\mathbb{R}^n$. Moreover, by Proposition 2, the restriction of I to $\mathbb{R}^n - \{0\}$ is a warped Riemannian metric of the general form (13a). Then, by Remark 14, the vertical geodesics $\gamma(t)$ of this warped Riemannian metric are parameterised by $r(t) =$ affine function of t and $\bar x(t) = \bar x =$ constant, where r is the vertical distance coordinate. Therefore, each vertical geodesic $\gamma(t)$ can be parameterised by $\eta(t) = \eta(r(t))$ and $\bar x(t) = \bar x =$ constant. This is

identified with the curve $z(t) = \eta(t)\,\bar x$, which is a radial straight line in $\mathbb{R}^n$, in the direction $\bar x$. It remains to note that the geodesics of the metric I are just the geodesics of its restriction to $\mathbb{R}^n - \{0\}$, extended by continuity whenever they reach the point $z = 0$.

Let $z(t) = \eta(t)\,\bar x$ describe a geodesic of the metric I, as just explained. To say that this geodesic can be extended indefinitely is equivalent to saying that $\eta(t)$ cannot reach the value $\eta = \infty$ within a finite time. For the von Mises-Fisher model, $\eta(\sigma) = \dot\sigma$ as long as $\sigma \in (0,\infty)$. Therefore, according to Remark 14, by evaluating the two conditions (3b), it is possible to know whether $\eta(t)$ can reach the two values $\eta = \infty$ and $\eta = 0$ within a finite time. These conditions now read

$$\lim_{\eta\to\infty} r(\eta) - r(\eta_0) \stackrel{?}{=} \infty \quad\text{and}\quad \lim_{\eta\to 0} r(\eta_1) - r(\eta) \stackrel{?}{=} \infty$$

where η_0 and η_1 are arbitrary. By (13a), $r(\eta)$ is defined by $dr/d\eta = (\psi''(\eta))^{1/2}$. Therefore, the two conditions in (3b) are identical to

$$\int_{\eta_0}^{\infty} (\psi''(\eta))^{1/2}\, d\eta \stackrel{?}{=} \infty \quad\text{and}\quad \int_0^{\eta_1} (\psi''(\eta))^{1/2}\, d\eta \stackrel{?}{=} \infty \tag{48a}$$

where $\psi''(\eta)$ is given by (17a). For the first integral, recall the asymptotic expansion of modified Bessel functions at $\eta = \infty$ [46], (Section 7.23, Page 203. This formula appears with the wrong sign for the second term in parentheses, in [33]), ok, no pb

$$I_\nu(\eta) = \frac{e^\eta}{\sqrt{2\pi\eta}}\left(1 - \frac{4\nu^2 - 1}{8\eta} + \frac{(4\nu^2-1)(4\nu^2-3^2)}{2!(8\eta)^2}\right) + O(\eta^{-3})$$

Using this asymptotic expansion, it follows by performing some direct calculations, and recalling that $\nu = n/2$,

Erreur indexes

$$\frac{I_{\nu+1}(\eta)}{I_{\nu+1}(\eta)} = 1 - \frac{n}{\eta} + \frac{n(n-1)}{2\eta^2} + O(\eta^{-3})$$

$$\frac{I_\nu^2(\eta)}{I_{\nu-1}^2(\eta)} = 1 - \frac{n-1}{\eta} + \frac{(n-1)(n-2)}{2\eta^2} + O(\eta^{-3})$$

Replacing these expressions into (17a) immediately gives

$$\psi''(\eta) = \frac{n-1}{2\eta^2} + O(\eta^{-3}) \tag{48b}$$

Since $n > 1$, this implies that the first integral in (48a) is divergent, as required. The second integral in (48a) is actually convergent. Indeed, $\psi''(\eta)$ is a continuous function in the neighborhood of $\eta = 0$, as seen in the proof of Proposition 2, for the limit (18c). Thus, the first condition in (3b) is verified, while the second condition is not verified. This means that $\eta(t)$ cannot reach the value $\eta = \infty$ within a finite time, but that it can reach the value $\eta = 0$

within a finite time. The first of these two statements shows that the geodesic described by $z(t)$ can indeed be extended indefinitely. Now, any geodesic of the metric I which passes through the point $z = 0$ is described by some $z(t)$ of this form. ■

The idea behind the proof of the completeness part of Proposition 8 can be summarised as follows. The restriction of the Rao-Fisher information metric I to $\mathbb{R}^n - \{0\}$ is a warped Riemannian metric. Thus, as stated in Section 2, $\mathbb{R}^n - \{0\}$ will be a complete Riemannian manifold, when equipped with this warped Riemannian metric, if and only if the two conditions in (3b) are verified. Once these conditions are evaluated, it turns out the first one is verified, but the second one is not. Thus, $\mathbb{R}^n - \{0\}$ is not a complete Riemannian manifold. However, this is only due to the fact that the point $z = 0$ is excluded. Once this point is included, the parameter space $\mathbb{R}^n$ is obtained, and this is a complete Riemannian manifold, when equipped with the Rao-Fisher information metric I. Precisely, a vertical geodesic in $\mathbb{R}^n - \{0\}$ can reach the point $z = 0$ within a finite time, but then all it does is pass through this point, and immediately return to $\mathbb{R}^n - \{0\}$. However, this vertical geodesic cannot escape to infinity within a finite time. With regard to the simple connectedness part of Proposition 8, excluding the point $z = 0$ has no influence if $n > 2$, since $\mathbb{R}^n - \{0\}$ is still simply-connected in this case. However, for $n = 2$, $\mathbb{R}^n - \{0\}$ is not simply-connected [43].

Proposition 8 established that the parameter space $\mathbb{R}^n$ of the von Mises-Fisher model is a simply-connected complete Riemannian manifold, for any value of n. To show that this parameter space is a Hadamard manifold, it remains to show that it has negative sectional curvature. This is done using numerical computation, for $n = 2, \ldots, 8$.

Precisely, let K_z denote the sectional curvature of $\mathbb{R}^n$ at a point $z \in \mathbb{R}^n$, with respect to the Rao-Fisher information metric I. Since, according to Proposition 2, I is a warped Riemannian metric, the sectional curvature K_z can be computed from formulae (4). To evaluate formula (4a), the Gauss equation, it is enough to note that for the von Mises-Fisher model, K^M is a constant equal to $+1$. Indeed, K^M is the sectional curvature of the unit sphere S^{n-1}. Then, formula (4a) reads

$$K_z(u,v) = \frac{1}{\beta^2} - \left(\frac{\partial_r \beta}{\beta}\right)^2 \quad u,v \in T_{\bar{z}} S^{n-1} \tag{49a}$$

where $z = \eta \bar{z}$. On the other hand, formula (4b), the radial curvature equation, can be copied directly,

$$K_z(u,\partial_r) = -\frac{\partial_r^2 \beta}{\beta} \tag{49b}$$

Here, $\beta(\eta)$ is given by (17b) of Proposition 2, and r is the vertical distance coordinate defined by $dr/d\eta = (\psi''(\eta))^{1/2}$. In each one of formulae (49), the

right-hand side is a real-valued function of η, independent of the vectors u, v. This will be denoted in the following way

$$K_z(u,v) = K^s(\eta) \qquad K_z(u,\partial_r) = K^r(\eta) \tag{50}$$

Precisely, $K^s(\eta)$ is the sectional curvature of any section (u,v) tangent to the surface of a sphere centred at $z = 0$ and with Euclidean radius η. On the other hand, $K^r(\eta)$ is the sectional curvature of a radial section (u, ∂_r). In the following, $K^s(\eta)$ will be called the surface curvature, and $K^r(\eta)$ will be called the radial curvature.

Formulae (49) were computed numerically for $n = 2, \ldots, 8$. It was systematically found that the sectional curvatures $K^s(\eta)$ and $K^r(\eta)$ are negative for all values of η. From these numerical results, it can be concluded with certainty that the sectional curvature of $\mathbb{R}^n$, with respect to the Rao-Fisher information metric I, is negative, when n ranges from $n = 2$ to $n = 8$. Figures 1-3 below give a graphic representation for $n = 3, 4, 5$. The sectional curvatures $K^s(\eta)$ and $K^r(\eta)$ behave in the same way, for all considered values of n. Precisely, they are equal to zero at $\eta = 0$, and decrease to a limiting negative value, as η becomes large. This limiting value, denoted $K^s(\infty)$ and $K^r(\infty)$, for $K^s(\eta)$ and $K^r(\eta)$, respectively, is given in the following table

	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 8$
$K^s(\infty)$	-0.50	-0.25	-0.16	-0.12	- 0.10	-0.08	-0.07
$K^r(\infty)$	-0.50	-0.25	-0.16	-0.12	-0.10	-0.08	-0.07

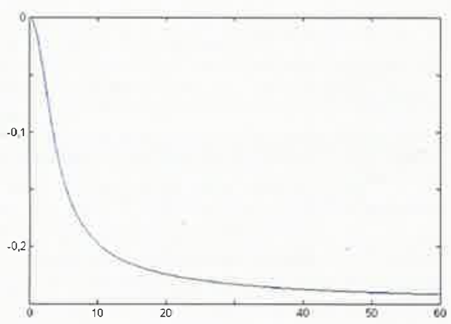
Table 1: Limiting value of the surface and radial curvatures

Remarkably, it appears from this table that $K^s(\infty)$ and $K^r(\infty)$ are close to each other, having the same first two digits after the decimal point. However, it is currently not clear why, or whether it is because $K^s(\infty)$ and $K^r(\infty)$ are equal.

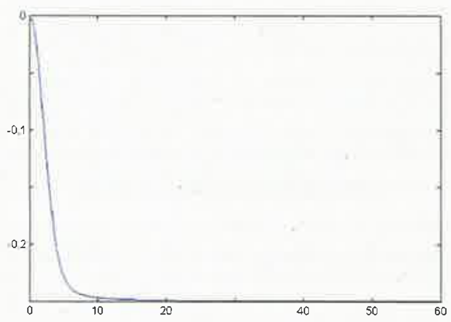
Remark 16 : based on Proposition 8, and on the numerical results reported here, it has been found that the parameter space $\mathbb{R}^n$ of the von Mises-Fisher model becomes a Hadamard manifold, when equipped with the Rao-Fisher information metric I, for $n = 2, \ldots, 8$. Indeed, Proposition 8 shows that this parameter space is a simply-connected complete Riemannian manifold, for any value of $n = 2, \ldots$, while the numerical results of Figures 1-3 and Table 1 show that it has negative sectional curvature. Hopefully, future work will soon remove the need to appeal to numerical results, by giving a mathematical proof of the fact that the sectional curvature of the parameter space $\mathbb{R}^n$ is negative for any value of $n = 2, \ldots$, without restriction. ■

Remark 17 : preliminary results from ongoing research clearly indicate that the Riemannian Gaussian model, which was studied in 4.2, when defined on

$M = H^{n-1}$, the $(n-1)$-dimensional hyperbolic space, has similar properties to the von Mises-Fisher model, with regard to the sectional curvature of its parameter space. Indeed, numerical computations show the sectional curvature of this parameter space $\mathcal{M} = H^{n-1} \times (0, \infty)$, equipped with the Rao-Fisher information metric I, is negative for $n = 3, 4, 5$. These numerical



Graph of the surface curvature $K^s(\eta)$

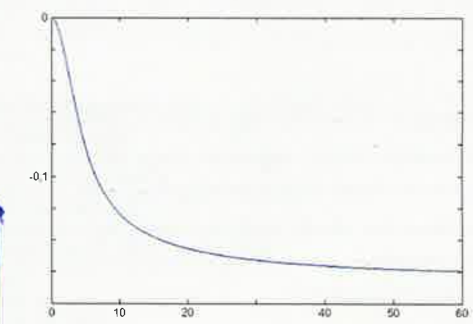


Graph of the radial curvature $K^r(\eta)$

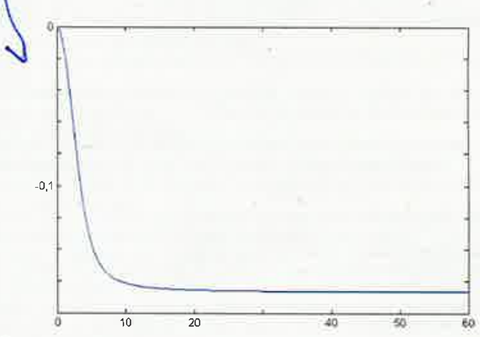
Fig. 1: $(n = 3)$ Sectional curvature of the parameter space of the von Mises-Fisher model

K on the same plot

on same plot



Graph of the surface curvature $K^s(\eta)$

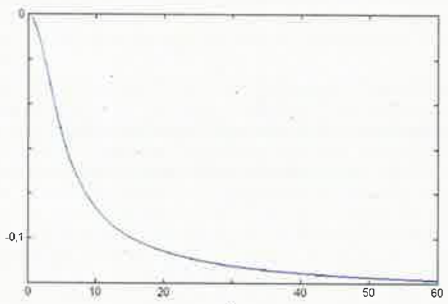


Graph of the radial curvature $K^r(\eta)$

Fig. 2: $(n = 4)$ Sectional curvature of the parameter space of the von Mises-Fisher model

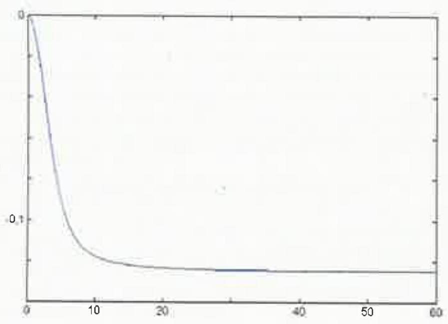
computations were carried out using formulae (4), for the sectional curvature of a warped Riemannian metric. This is justified because the hyperbolic space H^{n-1} is an irreducible Riemannian symmetric space, since it is a space of constant negative curvature, so the Rao-Fisher information metric (22) is a warped Riemannian metric. ■

Limiting values $K^s(\infty)$, $K^r(\infty)$ for Riemannian Gaussian model on hyperbolic space? as in Tab. 1. Are they identical?



Graph of the surface curvature $K^s(\eta)$

Idem.



Graph of the radial curvature $K^r(\eta)$

Fig. 3: ($n = 5$) Sectional curvature of the parameter space of the von Mises-Fisher model

References

1. Absil, P.A., Mahony, R., Sepulchre, R.: Optimization algorithms on matrix manifolds. Princeton University Press, Princeton, NJ (2008)
2. Amari, S., Nagaoka, H.: Methods of information geometry. American Mathematical Society (2000)
3. Amari, S.I.: Natural gradient works efficiently in learning. Neural Comput. **10**(2), 251–276 (1998)
4. Arnaudon, M.: Semi-martingales dans les espaces homogènes. Annales de l'I.H.P. **29**(3), 269–288 (1993)
5. Arnaudon, M., Miclo, L.: A stochastic algorithm finding generalized means on compact manifolds. Stochastic. Process. Appl. **124**(10), 3463 – 3479 (2014)
6. Atkinson, C., Mitchell, A.: Rao's distance measure. Sankhya Ser. A **43**, 345–365 (1981)
7. Ay, N., Jost, J., Lê, H.V., Schwachhöfer, L.: Information geometry and sufficient statistics. Probab. Theory Relat. Fields **162**(1), 327–364 (2015)
8. Bensadon, J.: Black-box optimization using geodesics in statistical manifolds. Entropy **17**(1), 304–345 (2015)
9. Bensadon, J.: Applications of information theory to machine learning. Ph.D. thesis, Université Paris-Saclay (2016). URL `https://tel.archives-ouvertes.fr/tel-01297163`
10. Berthoumieu, Y., Bombrun, L., Said, S.: Classification approach based on the product of Riemannian manifolds from Gaussian parameterization space. In: International Conference on Image Processing (ICIP) (2017)
11. Bishop, R.L., O'NEILL, B.: Manifolds of negative curvature. Trans. Amer. Math. Soc. **145**, 1–49 (1969)
12. Bottou, L., Curtis, F.E., Nocedal, J.: Optimization methods for large-scale machine learning (2017). URL `http://arxiv.org/abs/1606.04838v2`
13. Chavel, I.: Riemannian geometry, a modern introduction. Cambridge University Press, Cambridge, UK (2006)
14. Cheng, G., Vemuri, B.C.: A novel dynamic system in the space of SPD matrices with applications to appearance tracking. SIAM J. Imaging Sci. **6**(1), 592–615 (2013)
15. Chentsov, N.N.: Statistical decision rules and optimal inference. American Mathematical Society, Providence, RI (1982)
16. Chikuse, Y.: Statistics on special manifolds, (Lecture Notes in Statistics, 174). Springer Science+Business Media, LLC (2003)
17. Do Carmo, M.P.: Riemannian geometry (second edition). Birkhauser, Boston (1992)
18. Dobarro, F., Unal, B.: Curvature of multiply warped products. J. Geom. Phys. **55**, 75–106 (2005)
19. Emery, M.: Stochastic Calculus in Manifolds. Springer-Verlag, Berlin Heidelberg (1980)
20. Estrade, A.: Exponentielle stochastique et intégrale multiplicative discontinues. Annales de l'I.H.P. **28**(1), 107–129 (1992)
21. Gallavotti, G.: The elements of mechanics. Springer Science+Business Media, LLC, New York (1983)
22. Helgason, S.: Differential geometry and symmetric spaces. Academic Press, New York and London (1962)
23. Hsu, E.P.: Stochastic analysis on manifolds. American Mathematical Society, Providence, RI (2002)
24. Ikeda, N., Watanabe, S.: Stochastic differential equations and diffusion processes. North-Holland Publishing Company, Amsterdam (1981)

25. Kallenberg, O.: Foundations of modern probability (second edition). Springer-Verlag, New York (2001)
26. Kloeden, P.E., Platen, E.: Numerical Solution of Stochastic Differential Equations. Springer, New York (2011)
27. Knapp, A.W.: Lie groups beyond an introduction (second edition). Birkhauser, Boston (2002)
28. Lang, S.: Introduction to linear algebra (second edition). Springer-Verlag, New York (1986)
29. Lee, J.M.: Introduction to smooth manifolds (second edition). Springer Science+Business Media, New York (2012)
30. Lehmann, E.L., Romano, J.P.: Testing statistical hypotheses (third edition). Springer Science+Business Media, Inc., New York (2005)
31. Liao, E.: Lévy processes on Lie groups. Cambridge University Press, Cambridge, UK (2004)
32. Luenbeger, D.G.: Linear and nonlinear programming (second edition). Addison-Wesley publishing company, Reading, MA (1973)
33. Mardia, K.V., Jupp, P.E.: Directional statistics. John Wiley & Sons, Ltd., Chichester (2002)
34. Marsden, J.E., Ratiu, T.: Manifolds, tensor analysis, and applications. Springer-Verlag Publishing Company Inc., New York (2001)
35. Martens, J.: New insights and perspectives on the natural gradient method (2017). URL `http://arxiv.org/abs/1412.1193v8`
36. McLachlan, G.J.: Discriminant analysis and statistical pattern recognition. John Wiley & Sons, Inc., Hoboken (2004)
37. Ollivier, Y., Arnold, L., Auger, A., Hansen, N.: Information geometric optimization algorithms: a unifying picture via invariance principles. J. Mach. Learn. Res. **18**(18), 1–65 (2017)
38. O'Neill, B.: Semi-Riemannian geometry with applications to relativity. Academic Press, San Diego, California (1983)
39. Pennec, X., Fillard, P., Ayache, N.: A Riemannian framework for tensor computing. Int. J. Comput. Vision **66**(1), 41–66 (2006)
40. Petersen, P.: Riemannian geometry (second edition). Springer Science+Business Media, LLC, New York (2006)
41. Said, S., Bombrun, L., Berthoumieu, Y., Manton, J.H.: Riemannian Gaussian distributions on the space of symmetric positive definite matrices. IEEE. Trans. Inf. Theory **63**(4), 2153–2170 (2017)
42. Said, S., Hajri, H., Bombrun, L., Vemuri, B.C.: Gaussian distributions on Riemannian symmetric spaces: statistical learning with structured covariance matrices. IEEE. Trans. Inf. Theory **64**(2), 752–772 (2018)
43. Spanier, E.H.: Algebraic topology. Springer-Verlag New York, Inc., New York (1966)
44. Terras, A.: Harmonic analysis on symmetric spaces and applications, Vol. II. Springer-Verlag, New York (1988)
45. Unal, B.: Doubly warped products. Differential Geom. Appl. **15**, 253–263 (2001)
46. Watson, G.N.: A treatise on the theory of Bessel functions. Cambridge University Press, Cambridge, UK (1922)
47. Wierstra, D., Schaul, T., Glasmachers, T., Sun, Y., Peters, J., Schmidhuber, J.: Natural evolution strategies. J. Mach. Learn. Res. **15**(1), 949–980 (2014)
48. Young, G.A., Smith, R.L.: Essentials of statistical inference. Cambridge University Press, Cambridge, UK (2005)
49. Zanini, P., Said, S., Berthoumieu, Y., Congedo, M., Jutten, C.: Riemannian online algorithm for estimating mixture model parameters. In: Geometric Science of Information (GSI) (2017)

50. Zanini, P., Said, S., Congedo, M., Berthoumieu, Y., Jutten, C.: Parameter estimates of Riemannian Gaussian distributions in the manifold of covariance matrices. In: Sensor Array and Multichannel Signal Processing Workshop (SAM) (2016)

Appendix A – Proof of Theorem 1

In order to complete the proof of Theorem 1, the following proposition is needed. The notation is that of (11).

Proposition 9. *assume condition (7) holds. Then,*

$$\partial_\sigma \ell(z) \circ g = \partial_\sigma \ell(g^{-1}\cdot z) \qquad \nabla_{\bar x}\, \ell(z) \circ g = dg_x \nabla_{\bar x}\, \ell(g^{-1}\cdot z) \tag{51a}$$

In particular, if $g = s_x$,

$$\partial_\sigma \ell(z) \circ s_x = \partial_\sigma \ell(z) \qquad \nabla_{\bar x}\, \ell(z) \circ s_x = -\nabla_{\bar x}\, \ell(z) \tag{51b}$$

$\bar x$ not defined at this point neither its action on $\mathcal{M}$: i.e. a $(0,\infty)$

Proof: note that (51b) follows from (51a), by the definition of the geodesic-reversing isometry s_x [22]. Indeed, $s_x \cdot \bar x = \bar x$ so $s_x^{-1}\cdot z = z$. Moreover, $ds_x = -\mathrm{Id}$, as a linear mapping of T_xM, where Id denotes the identity. To prove (51a), note that

$$(\partial_\sigma \ell(z) \circ g)\,(x) = \partial_\sigma \log p(g\cdot x|z) = \partial_\sigma \log p(x|g^{-1}\cdot z) \tag{52a}$$

where the second equality follows from condition (7). However,

$$\partial_\sigma \log p(x|g^{-1}\cdot z) = \partial_\sigma \ell(g^{-1}\cdot z)(x) \tag{52b}$$

Replacing (52b) in (52a) gives,

$$(\partial_\sigma \ell(z) \circ g)\,(x) = \partial_\sigma \ell(g^{-1}\cdot z)(x)$$

which is the first part of (51a). For the second part, a similar reasoning can be applied. Precisely, using condition (7), it follows,

$$(d\ell(z) \circ g)\,(x) = d\log p(g\cdot x|z) = d\log p(x|g^{-1}\cdot z) = d\ell^{(g)}(z)(x) \tag{53a}$$

where $d\ell(z)$ denotes the derivative of $\ell(z)$ with respect to $\bar x$, and $\ell^{(g)}(z) = \ell(g^{-1}\cdot z)$, so (53a) implies that,

$$d\ell(z) \circ g = d\ell^{(g)}(z) \tag{53b}$$

By the chain rule [29], for $u \in T_xM$,

$$d\ell^{(g)}(z)|_x\, u = d\ell(g^{-1}\cdot z)\, dg_x^{-1}\, u$$

notation.

Replacing in (53b),

$$d\ell(z) \circ g|_x = d\ell(g^{-1}\cdot z)\, dg_x^{-1} \tag{53c}$$

The second part of (51a) can now be obtained as follows. By the definition of the Riemannian gradient [40],

$$Q\,(\nabla_{\bar x}\, \ell(z) \circ g\,, u) = d\ell(z) \circ g|_x\, u = d\ell(g^{-1}\cdot z)\, dg_x^{-1}\, u \tag{54a}$$

where the second equality follows from (53c). However,

$$d\ell(g^{-1}\cdot z)\, dg_x^{-1}\, u = Q\,(\nabla_{\bar x}\, \ell(g^{-1}\cdot z), dg_x^{-1}\, u)$$

Since g is an isometry of M, its derivative dg_x preserves the Riemannian metric Q. Therefore,

$$Q\,(\nabla_{\bar x}\, \ell(g^{-1}\cdot z), dg_x^{-1}\, u) = Q\,(dg_x \nabla_{\bar x}\, \ell(g^{-1}\cdot z),\, u) \tag{54b}$$

Replacing (54b) in (54a) gives,

$$Q\,(\nabla_{\bar x}\, \ell(z) \circ g\,, u) = Q\,(dg_x \nabla_{\bar x}\, \ell(g^{-1}\cdot z),\, u)$$

To finish the proof, it is enough to note that the vector u is arbitrary. ■

Proof of (10b): recall the polarisation identity, from elementary linear algebra [28], (see Page 29),

$$I_z(\partial_\sigma, u) = \frac{1}{4}\, I_z(\partial_\sigma + u, \partial_\sigma + u) - \frac{1}{4}\, I_z(\partial_\sigma - u, \partial_\sigma - u)$$

by replacing (8) into this identity, it can be seen that,

$$I_z(\partial_\sigma, u) = \mathbb{E}_z\,(\partial_\sigma \ell(z)\, d\ell(z)\, u)$$

Then, by recalling the definition of the Riemannian gradient [40],

$$I_z(\partial_\sigma, u) = \mathbb{E}_z\,(\,\partial_\sigma \ell(z)\, Q\,(\nabla_{\bar x}\, \ell(z), u)\,) \tag{55a}$$

Denote the function under the expectation by f, and apply (11) with $g = s_x$. Then,

$$\mathbb{E}_z\, f = \mathbb{E}_{s_x\cdot z}\, f = \mathbb{E}_z\,(f \circ s_x) \tag{55b}$$

since $s_x \cdot z = z$. Note that (10b) amounts to saying that $\mathbb{E}_z\, f = 0$. To prove this, note that

$$f \circ s_x = (\partial_\sigma \ell(z) \circ s_x)\, Q\,(\nabla_{\bar x} \ell(z) \circ s_x\,, u) = -\partial_\sigma \ell(z) Q\,(\nabla_{\bar x} \ell(z), u) = -f \tag{55c}$$

where the second equality follows from (51b). Replacing in (55b) shows that $\mathbb{E}_z\, f = 0$. ■

Proof of (10c): the idea is to apply Schur's lemma to $I_z(u,u)$, considered as a symmetric bilinear form on T_xM. First, it is shown that this symmetric bilinear form is invariant under the isotropy representation. That is,

$$I_z(u,u) = I_z\,(dk_x\, u\,, dk_x\, u) \qquad \text{for all } k \in K_x \tag{56a}$$

This is done using (11). Note from (8),

$$I_z(u,u) = \mathbb{E}_z\left[Q\,(\nabla_{\bar x}\, \ell(z), u)\right]^2 \tag{56b}$$

Denote the function under the expectation by f. By (11),

$$\mathbb{E}_z\, f = \mathbb{E}_{k^{-1}\cdot z}\, f = \mathbb{E}_z\,(f \circ k^{-1}) \tag{56c}$$

since $k^{-1}\cdot z = z$ for $k \in K_x$. To find $f \circ k^{-1}$, note that,

$$Q\,(\nabla_{\bar x}\, \ell(z) \circ k^{-1}, u) = Q\,(dk_x^{-1} \nabla_{\bar x}\, \ell(z), u) = Q\,(\nabla_{\bar x}\, \ell(z), dk_x\, u)$$

where the first equality follows from (51a) and the fact that $k\cdot\bar x = \bar x$, and the second equality from the fact that dk_x preserves the Riemannian metric Q. Now, by (56b) and (56c),

$$I_z(u,u) = \mathbb{E}_z\, f = \mathbb{E}_z\,(f \circ k^{-1}) = \mathbb{E}_z\,(Q\,(\nabla_{\bar x}\ell(z), dk_x\, u))^2 = I_z\,(dk_x\, u\,, dk_x\, u)$$

and this proves (56a). Recall Schur's lemma, ([27], Page 240). Applied to (56a), this lemma implies that there exists some multiplicative factor β^2, such that

$$I_z(u,u) = \beta^2\, Q_x(u,u) \tag{57a}$$

It remains to show that β^2 is given by (9). Taking the trace of (57a),

$$\operatorname{tr} I_z = \beta^2 \operatorname{tr} Q_x = \beta^2 \dim M \tag{57b}$$

If $e_1, \ldots, e_d$ is an orthonormal basis of T_xM, then by (56b),

$$\operatorname{tr} I_z = \mathbb{E}_z \sum_{i=1}^{d} (Q\,(\nabla_{\bar x}\, \ell(z), e_i))^2 = \mathbb{E}_z\, Q\,(\nabla_{\bar x}\ell(z), \nabla_{\bar x}\ell(z)) \tag{57c}$$

Thus, (57b) and (57c) show that β^2 is given by (9). ■

To complete the proof of Theorem 1, it remains to show that the expectations appearing in (9) do not depend on $\bar x$.

For the first expectation, giving $\alpha^2(\sigma)$, note that,

$$\mathbb{E}_{g\cdot z}\,(\partial_\sigma \ell(g\cdot z))^2 = \mathbb{E}_z\,(\partial_\sigma \ell(g\cdot z) \circ g)^2 = \mathbb{E}_z\,(\partial_\sigma \ell(z))^2 \tag{58}$$

where the first equality follows from (11) and the second equality follows from (51a). Thus, this expectation has the same value, whether computed at $g\cdot z = (g\cdot\bar x, \sigma)$, or at $z = (\bar x, \sigma)$. Therefore, it does not depend on $\bar x$, since the action of G on M is transitive.

For the second expectation, giving $\beta^2(\sigma)$, note that by (11),

$$\mathbb{E}_{g\cdot z}\, Q\,(\nabla_{\bar x}\, \ell(g\cdot z)\,, \nabla_{\bar x}\, \ell(g\cdot z)\,) = \mathbb{E}_z\, Q\,(\nabla_{\bar x}\, \ell(g\cdot z) \circ g\,, \nabla_{\bar x}\, \ell(g\cdot z) \circ g\,) \tag{59a}$$

On the other hand, by (51a),

detail: applied to representations $k \mapsto dk_{\bar x}$ of $K_{\bar x}$.

↳ use of irreducibility of $K_{\bar x}$

ok... car

$I : T_x \mapsto T_x^*$

invariant iff $I \in \mathrm{Hom}_G(T_x, T_x^*)$

$$\nabla_{\bar x}\, \ell(g\cdot z) \circ g = dg_x \nabla_{\bar x}\, \ell(z)$$

Moreover, since dg_x preserves the Riemannian metric Q,

$$\begin{aligned} Q(\nabla_{\bar x}\, \ell(g\cdot z) \circ g, \nabla_{\bar x}\, \ell(g\cdot z) \circ g) &= Q(dg_x \nabla_{\bar x}\, \ell(z)\,, dg_x \nabla_{\bar x}\, \ell(z)) \\ &= Q(\nabla_{\bar x}\, \ell(z)\,, \nabla_{\bar x}\, \ell(z)) \end{aligned}$$

Replacing in (59a) gives

$$\mathbb{E}_{g\cdot z}\, Q\,(\nabla_{\bar x}\, \ell(g\cdot z)\,, \nabla_{\bar x}\, \ell(g\cdot z)\,) = \mathbb{E}_z\, Q(\nabla_{\bar x}\, \ell(z)\,, \nabla_{\bar x}\, \ell(z)) \tag{59b}$$

so this expectation has the same value, at $g\cdot z$ and at z. By the same argument made after (58), it does not depend on $\bar x$. ■

Proof of (11): let dv denote the invariant Riemannian volume element of M, and note that,

$$\mathbb{E}_{g\cdot z}\, f = \int_M f(x)\, p(x|g\cdot z)\, dv(x) = \int_M f(x)\, p(g^{-1}\cdot x|z)\, dv(x) \tag{60a}$$

where the second equality follows from (7). Introduce the variable $y = g^{-1}\cdot x$. Since the volume element dv is invariant,

$$\int_M f(x)\, p(g^{-1}\cdot x|z)\, dv(x) = \int_M f(g\cdot y)\, p(y|z)\, dv(y) \tag{60b}$$

The last integral is the same as $\mathbb{E}_z\,(f \circ g)$. Therefore, (11) follows from (60a) and (60b). ■

Appendix B – Proof of Proposition 2

It remains to prove (17a) and (17b). To do so, introduce the following notation, using (15),

$$t = \langle x, \bar x\rangle \qquad Z(\eta) = e^{\psi(\eta)} = (2\pi)^{\nu}\, \eta^{1-\nu}\, I_{\nu-1}(\eta) \tag{61a}$$

Then, $Z(\eta)$ is the moment generating function of t, so

$$\mathbb{E}_z(t) = \frac{Z'(\eta)}{Z(\eta)} \quad \text{and} \quad \mathbb{E}_z\,(t^2) = \frac{Z''(\eta)}{Z(\eta)} \tag{61b}$$

where the prime denotes differentiation with respect to η. Recall the derivative and recurrence relations of modified Bessel functions [46],

$$(\eta^{-a} I_a(\eta))' = \eta^{-a} I_{a+1}(\eta) \qquad I_{a-1}(\eta) - I_{a+1}(\eta) = \frac{2a}{\eta}\, I_a(\eta) \tag{62}$$

where a is any complex number. By applying these relations to (61b), it is possible to show, through a direct calculation,

$$\mathbb{E}_z(t) = \frac{I_\nu(\eta)}{I_{\nu-1}(\eta)} \tag{63a}$$

$$\mathbb{E}_z(t^2) = \frac{1}{n} + \frac{n-1}{n}\frac{I_{\nu+1}(\eta)}{I_{\nu-1}(\eta)} \tag{63b}$$

taille

Formulae (63) will provide the proof of (17a) and (17b).

Proof of (17a): since $\psi(\eta)$ is the cumulant generating function of t,

$$\psi''(\eta) = \mathrm{Var}_z(t) = \mathbb{E}_z(t^2) - \mathbb{E}_z(t)^2 \tag{64a}$$

where Var denotes the variance. Now, (17a) follows immediately by replacing from (63) into the right-hand side. ■

Proof of (17b): recall from (18a),

$$\beta^2(\eta) = \frac{\eta^2}{n-1}\mathbb{E}_z(1-t^2) \tag{64b}$$

However, from (63b),

$$\mathbb{E}_z(1-t^2) = \frac{n-1}{n}\left(1 + \frac{I_{\nu+1}(\eta)}{I_{\nu-1}(\eta)}\right) \tag{64c}$$

Now, (17b) follows by replacing (64c) into (64b). ■

Proof of (63a): using the derivative relation of modified Bessel functions, which is the first relation in (62), with $a = \nu - 1$, it follows that

$$Z'(\eta) = (2\pi)^\nu \eta^{1-\nu} I_\nu(\eta) \tag{65a}$$

Now, Formula (63a) follows by replacing this into (61b) and using (61a). ■

Proof of (63b): write (65a) in the form

$$Z'(\eta) = (2\pi)^\nu \eta\,(\eta^{-\nu} I_\nu(\eta))$$

By the product rule

$$Z''(\eta) = (2\pi)^\nu \eta^{-\nu} I_\nu(\eta) + (2\pi)^\nu \eta\,(\eta^{-\nu} I_\nu(\eta))'$$

The derivative in the second term can be evaluated from the derivative relation of modified Bessel functions, with $a = \nu$. Then,

$$Z''(\eta) = (2\pi)^\nu \eta^{-\nu} I_\nu(\eta) + (2\pi)^\nu \eta^{1-\nu} I_{\nu+1}(\eta)$$

Rearrange this formula as

$$Z''(\eta) = (2\pi)^\nu \eta^{1-\nu}\left(\eta^{-1} I_\nu(\eta) + I_{\nu+1}(\eta)\right)$$

By the recurrence relation of modified Bessel functions, which is the second relation in (62), with $a = \nu$, it then follows

$$Z''(\eta) = (2\pi)^\nu \eta^{1-\nu}\left(\frac{1}{2\nu} I_{\nu-1}(\eta) - \frac{1}{2\nu} I_{\nu+1}(\eta) + I_{\nu+1}(\eta)\right)$$

Recalling that $2\nu = n$, this can be written,

$$Z''(\eta) = (2\pi)^\nu \eta^{1-\nu}\left(\frac{1}{n} I_{\nu-1}(\eta) + \frac{n-1}{n} I_{\nu+1}(\eta)\right) \tag{65b}$$

Now, Formula (63b) follows by replacing this into (61b) and using (61a). ■

Appendix C – Proof of Proposition 6

The setting and notations are the same as in Section 7, except for the fact that $\bar{x}$ is written as x, without the bar, in order to avoid notations such as $\dot{\bar{x}}$ or $\ddot{\bar{x}}$. This being said, let $\tilde{\nabla}$ and ∇ denote the Levi-Civita connections of the Riemannian metrics I and Q, respectively. Thus, $\tilde{\nabla}$ is a connection on the tangent bundle of the manifold $\mathcal{M}$, and ∇ is a connection on the tangent bundle of the manifold M [40][13]. Introduce the shape operator $S : T_xM \to T_xM$, which is given as in [40],

Pourquoi pas des $T_x\mathcal{M}$?

$$S(u) = \tilde{\nabla}_u \partial_r \qquad u \in T_xM \tag{66}$$

for any $x \in M$. The following identities can be found in [40] (Section 2.4, Page 41),

$$\tilde{\nabla}_{\partial_r}\partial_r = 0 \tag{67a}$$

$$\tilde{\nabla}_{\partial_r} X = S(X) \tag{67b}$$

$$\tilde{\nabla}_X Y = \nabla_X Y - I(S(X), Y)\,\partial_r \tag{67c}$$

for any vector fields X and Y on M. Using these identities, it is possible to write the geodesic equation of the Riemannian metric I, in terms of the shape operator S. This is given in the following proposition.

Proposition 10. *let $\gamma(t)$ be a curve in $\mathcal{M}$, with $\gamma(t) = (x(t), \sigma(t))$ and let $r(t) = r(\sigma(t))$. The curve $\gamma(t)$ is a geodesic of the Riemannian metric I if and only if it satisfies the geodesic equation*

$$\ddot{r} = I(S(\dot{x}), \dot{x}) \tag{68a}$$

$$\ddot{x} = -2\dot{r}S(\dot{x}) \tag{68b}$$

where $\ddot{x} = \nabla_{\dot{x}}\dot{x}$ is the acceleration of the curve $x(t)$ in M.

The shape operator S moreover admits a simple expression, which can be derived from expression (42) of the Riemannian metric I, using the fact that $\tilde{\nabla}$ is a metric connection [13] (Theorem I.5.1, Page 16).

Proposition 11. *in the notation of (42), the shape operator S is given by*

$$S(u) = \sum_{q=1}^{r} \frac{\partial_r \beta_q(r)}{\beta_q(r)} u_q \tag{69}$$

In other words, the decomposition $u = u_1 + \ldots + u_r$ provides a block-diagonalisation of S, where each block is a multiple of identity.

Combining Propositions 10 and 11, the geodesic equation (68) takes on a new form. Precisely, replacing (69) into (68) gives the following equations

$$\ddot{r} = \sum_{q=1}^{r} \beta_q(r)\partial_r\beta_q(r)\,Q(\dot{x}_q, \dot{x}_q) \tag{70a}$$

$$\ddot{x}_q = -2\dot{r}\,\frac{\partial_r \beta_q(r)}{\beta_q(r)}\,\dot{x}_q \tag{70b}$$

where $\dot{x} = \dot{x}_1 + \ldots + \dot{x}_r$ and $\ddot{x} = \ddot{x}_1 + \ldots + \ddot{x}_r$. The proof of Proposition 6 can be obtained directly from equations (70), using the following conservation laws.

Proposition 12. *each one of the following quantities C_q is a conserved quantity,*

$$C_q = \beta_q^4(r)\,Q(\dot{x}_q, \dot{x}_q) \qquad \text{for } q = 1, \ldots, r \tag{71}$$

In other words, C_q remains constant when evaluated along any geodesic $\gamma(t)$ of the Riemannian metric I.

For now, assume that Propositions 10 through 12 are true. To prove Proposition 6, note the following.

Proof of (43a): it is enough to show that the right-hand side of (43a) is the same as the right-hand side of (70a). To do so, note from (43a) and (71) that

$$V(r) = \sum_{q=1}^{r} \frac{\beta_q^2(r(0))}{\beta_q^2(r)} I_z(u_q, u_q) = \sum_{q=1}^{r} \frac{C_q}{\beta_q^2(r)} \tag{72a}$$

Indeed, since $\dot{x}_q(0) = u_q$ and since C_q is a conserved quantity

$$\beta_q^2(r(0))\,I_z(u_q, u_q) = \beta_q^4(r)\,Q(\dot{x}_q, \dot{x}_q)\big|_{t=0} = C_q$$

Now, replacing the derivative of (72a) into the right-hand side of (43a) directly leads to the right-hand side of (70a). ■

Proof of (43b): recall from Remark 7 that M is the Riemannian product of the M_q. Therefore, the Riemannian exponential mapping of M is also the product of the Riemannian exponential mappings of the M_q. Precisely, (43b) is equivalent to

$$x_q(t) = \exp_{x_q(0)}\left[\left(\int_0^t \frac{\beta_q^2(r(0))}{\beta_q^2(r(s))}ds\right) u_q\right] \qquad \text{for } q = 1, \ldots, r \tag{72b}$$

This means that the curve $x_q(t)$ in M_q is a reparameterised geodesic $(\delta_q \circ F)(t)$ where $\delta_q(t)$ is the geodesic given by $\delta_q(t) = \exp(t\,u_q)$ and $F(t)$ is the integral inside the parentheses in (72b). To prove (43b), it is sufficient to prove that (72b) solves equation (70b). Using the chain rule, (72b) implies that

$$\ddot{x}_q = \dot{F}^2\left(\ddot{\delta}_q \circ F\right) + F''\left(\dot{\delta}_q \circ F\right) = \dot{F}^2\left(\ddot{\delta}_q \circ F\right) + \frac{F''}{F'}\dot{x}_q = \frac{F''}{F'}\dot{x}_q \tag{72c}$$

where the third equality follows because δ_q is a geodesic, and therefore its acceleration $\ddot{\delta}_q$ is zero. By replacing the definition of the function $F(t)$, it is seen that (72c) is the same as (70b). It follows that (72b) solves (70b), as required. ■

Proof of Proposition 10: recall the geodesic equation is $\tilde{\nabla}_{\dot{\gamma}}\dot{\gamma} = 0$, which means that the velocity $\dot{\gamma}(t)$ is self-parallel [40][13]. Here, the velocity $\dot{\gamma}(t)$ is given by $\dot{\gamma}(t) = \dot{r}\,\partial_r + \dot{x}$. Accordingly, the left-hand side of the geodesic equation is

$$\tilde{\nabla}_{\dot{\gamma}}\dot{\gamma} = \tilde{\nabla}_{\dot{\gamma}}\dot{r}\,\partial_r + \tilde{\nabla}_{\dot{\gamma}}\dot{x} = \ddot{r}\,\partial_r + \dot{r}\,\tilde{\nabla}_{\dot{\gamma}}\partial_r + \tilde{\nabla}_{\dot{\gamma}}\dot{x} \tag{73a}$$

where the second equality follows by the product rule for the covariant derivative [40][13]. The second and third terms on the right-hand side of (73a) can be written in terms of the shape operator S. Precisely, for the second term,

$$\tilde{\nabla}_{\dot{\gamma}}\partial_r = \dot{r}\,\tilde{\nabla}_{\partial_r}\partial_r + \tilde{\nabla}_{\dot{x}}\partial_r = S(\dot{x}) \tag{73b}$$

where the second equality follows from (66) and (67a). Moreover, for the third term,

$$\tilde{\nabla}_{\dot{\gamma}}\dot{x} = \dot{r}\,\tilde{\nabla}_{\partial_r}\dot{x} + \tilde{\nabla}_{\dot{x}}\dot{x} = \dot{r}\,S(\dot{x}) + \ddot{x} - I(S(\dot{x}), \dot{x})\,\partial_r \tag{73c}$$

where the second equality follows from (67b) and (67c). Replacing (73b) and (73c) into (73a), the left-hand side of the geodesic equation becomes

$$\tilde{\nabla}_{\dot{\gamma}}\dot{\gamma} = (\ddot{r} - I(S(\dot{x}), \dot{x}))\,\partial_r + (\ddot{x} + 2\dot{r}S(\dot{x}))$$

Setting this equal to zero immediately gives equations (68). ■

Proof of Proposition 11 : recall the shape operator S is symmetric, since it is essentially the Riemannian Hessian of r [40] (Section 2.4, Page 41). Therefore, it is enough to evaluate $I(S(u), u)$ for $u \in T_x M$. Let X by a vector field on M, with $X(x) = u$. Then,

$$I(S(u), u) = I(S(X), X) = I\left(\tilde{\nabla}_{\partial_r} X, X\right) \tag{74a}$$

where the second equality follows from (67b). Using the fact that $\tilde{\nabla}$ is a metric connection [13] (Theorem I.5.1, Page 16), the right-hand side can be written as

$$I\left(\tilde{\nabla}_{\partial_r} X, X\right) = \frac{1}{2}\partial_r I(X, X) = \frac{1}{2}\partial_r \sum_{q=1}^{r} \beta_q^2(r)\, Q_q(u_q, u_q) \tag{74b}$$

where the second equality follows from (42). It remains to note that

$$\frac{1}{2}\partial_r \beta_q^2(r)\, Q_q(u_q, u_q) = \frac{\partial_r \beta_q(r)}{\beta_q(r)} I_q(u_q, u_q)$$

Accordingly, (74a) and (74b) imply

$$I(S(u), u) = I\left(\sum_{q=1}^{r} \frac{\partial_r \beta_q(r)}{\beta_q(r)} u_q, u_q\right) = I\left(\sum_{q=1}^{r} \frac{\partial_r \beta_q(r)}{\beta_q(r)} u_q, u\right) \tag{74c}$$

and (69) follows from the fact that S is symmetric. ■

Proof of Proposition 12 : to say that C_q is a conserved quantity means that $\dot{C}_q = 0$. From (71),

$$\dot{C}_q = 4\dot{r}\, \beta_q^3(r)\, \partial_r \beta_q(r)\, Q(\dot{x}_q, \dot{x}_q) + \beta_q^4(r) \frac{d}{dt} Q(\dot{x}_q, \dot{x}_q) \tag{75a}$$

The last derivative can be expressed as

$$\frac{d}{dt} Q(\dot{x}_q, \dot{x}_q) = 2\,Q(\ddot{x}_q, \dot{x}_q) = -4\dot{r}\, \frac{\partial_r \beta_q(r)}{\beta_q(r)} Q(\dot{x}_q, \dot{x}_q) \tag{75b}$$

where the second equality follows from (70b). By replacing (75b) into (75a), it follows immediately that $\dot{C}_q = 0$. ■

Appendix D – Proof of Proposition 7

The proof of Formula (46b), for the Laplace-Beltrami operator $\Delta_{\mathcal{M}}$, will introduce some useful notation.

Proof of Formula (46b) : for any smooth function f on $\mathcal{M}$, it follows from (42) that the Riemannian gradient $\tilde{\nabla} f$ of f, with respect to the multiply-warped Riemannian metric I, is given by

$$\tilde{\nabla} f = (\partial_r f)\, \partial_r + \sum_{q=1}^{r} \beta_q^{-2}(r)\, \nabla_{\bar{x}_q} f \tag{76a}$$

where $\nabla_{\bar{x}_q} f$ is the Riemannian gradient of f with respect to $\bar{x}_q \in M_q$, computed with all other arguments of f being fixed. Expression (76a) can be verified by checking that

$$df_{\!z} U = I(\tilde{\nabla} f, U)$$

for any tangent vector U to $\mathcal{M}$. This follows directly from (42) and (76a). Now, by definition of the Laplace-Beltrami operator [34], (see Page 443),

$$\Delta_{\mathcal{M}} f = \operatorname{div} \tilde{\nabla} f \tag{76b}$$

where the divergence $\operatorname{div} V$ of a vector field V on $\mathcal{M}$ is found from

$$\mathcal{L}_V \operatorname{vol} = (\operatorname{div} V) \operatorname{vol} \tag{76c}$$

with the notation $\mathcal{L}$ for the Lie derivative, and vol for the Riemannian volume element of the metric I. This last formula can be applied along with the following expression of vol, which follows from (42),

$$\operatorname{vol} = G(r)\, dr \bigwedge_q \operatorname{vol}_q(\bar{x}_q) \tag{76d}$$

where the function $G(r)$ was defined after (46b), and where $\wedge$ denotes the exterior product, and vol_q is the Riemannian volume of M_q. From (76c) and (76d), applying the product formula of the Lie derivative [34] (Theorem 7.4.8, Page 414),

$$\operatorname{div} V = \frac{1}{G}\partial_r (G V_r) + \sum_{q=1}^{r} \operatorname{div}_{M_q} V_q \tag{76e}$$

where $V = V_r \partial_r + \sum_q V_q$ with each V_q tangent to M_q, and where $\operatorname{div}_{M_q} V_q$ denotes the divergence of V_q with respect to $\bar{x}_q \in M_q$. Formula (46b) follows directly from (76a), (76b) and (76e). Indeed, for the vector field $V = \tilde{\nabla} f$,

$$V_r = \partial_r f \qquad V_q = \beta_q^{-2}(r)\, \nabla_{\bar{x}_q} f$$

as can be seen from (76a). ■

For the proof of Proposition 7, assume the process z is a Riemannian Brownian motion associated to I. Write $z(t) = (\bar{x}(t), \sigma(t))$ where $\bar{x}(t) = (\bar{x}_1(t), \ldots, \bar{x}_r(t))$ and each $\bar{x}_q(t)$ belongs to M_q. The proof consists in showing that the joint distribution of the processes $r(t) = r(\sigma(t))$ and $\bar{x}_q(t)$ is the same as described in Proposition 7. This is done through the following steps.

Step 1 – $r(t)$ verifies (47a) : recall that, for any smooth function f on $\mathcal{M}$, the process z verifies (46a). If $f = r$, then by (46a) and (46b)

$$dr(t) = \frac{1}{2}\Delta_{\mathcal{M}} r(t) dt + dm^r(t) = \frac{1}{2}\frac{\partial_r G}{G}(r(t)) dt + dm^r(t) \tag{77}$$

To prove that $r(t)$ verifies (47a), it is enough to prove that $dm^r(t) = dw(t)$ where w is a standard Brownian motion. Note that $dr^2(t)$ can be computed in two ways. The first way, by Itô's formula and (77)

$$dr^2 = r\Delta_{\mathcal{M}} r(t) dt + 2r(t) dm^r(t) + d[m^r](t)$$

where $[m^r](t)$ is the quadratic variation process of the local martingale $m^r(t)$ [25] (Theorem 17.16, Page 339). The second way, by (46a) with $f = r^2$,

$$\begin{aligned} dr^2 &= \left(I(\tilde{\nabla} r, \tilde{\nabla} r) + r\Delta_{\mathcal{M}} r(t)\right) dt + dm^{r^2}(t) dt \\ &= (1 + r\Delta_{\mathcal{M}} r(t))\, dt + dm^{r^2}(t) dt \end{aligned}$$

where the second line follows from (42) because $\tilde{\nabla} r = \partial_r$. By equating these two expressions of $dr^2(t)$, it follows that $dt - d[m^r](t)$ is the differential of a continuous local martingale of finite variation, and therefore identically zero [25] (Proposition 17.2, Page 330). In other words, $d[m^r](t) = dt$ and Lévy's characterisation implies that $dm^r(t) = dw(t)$, where w is a standard Brownian motion [25] (Theorem 18.3, Page 352). ■

Step 2 – $\bar{x}_q(t)$ verifies (47b) : if f is a smooth function on $\mathcal{M}$, such that $f(z) = f(\bar{x}_q)$, then by (46a) and (46b)

$$df(\bar{x}_q(t)) = \frac{1}{2}\beta_q^{-2}(r(t))\, \Delta_{M_q} f(\bar{x}_q(t))\, dt + dm^f(t) \tag{78}$$

Define $l_q(t)$ to be the inverse of the time change process $\tau_q(t)$ defined in (47b), and let $\theta_q(t) = (\bar{x}_q \circ l_q)(t)$. By applying the time change $l_q(t)$ to (78)

$$df(\theta_q(t)) = \frac{1}{2}\Delta_{M_q} f(\theta_q(t))\, dt + d(m^f \circ l_q)(t)$$

where the first term on the right-hand side is obtained by replacing from the definition of $\tau_q(t)$ in (47b). Recall that a time change of a local martingale is a local martingale [25] (Theorem 17.24, Page 344). Therefore, $m^f \circ l_q$ is a local martingale, so θ_q solves the martingale problem associated to $(1/2)\Delta_{M_q}$. This means that θ_q is a Riemannian Brownian motion in M_q. ■

Step 3 – $r(t)$ and $\theta_q(t)$ are independent : it is required to prove that the processes $r(t)$ and $\theta_1(t), \ldots, \theta_r(t)$ are jointly independent. A detailed proof is given only of the fact that $r(t)$ and $\theta_q(t)$ are independent, for any fixed q. The complete proof is obtained by repeating similar arguments.

The following proof is modeled on [23] (Example 3.3.3, Page 84). Let $(\xi^i(t)\,; i = 1, \ldots, \dim M_q)$ be the stochastic anti-development of $\bar{x}_q(t)$. Precisely [19] (Definition 8.23, Page 119)

$$d\xi^i(t) = Q\left(e^i, d\bar{x}_q(t)\right) = \beta_q^{-2}(r(t))\, I\left(e^i, dz(t)\right) \tag{79a}$$

where the e^i form a parallel orthonormal moving frame above the stochastic process $\bar{x}_q(t)$ in M_q. On the other hand, it is possible to write, using Itô's formula [19] (Proposition 7.34, Page 109)

$$dr(t) = I\left(\partial_r, dz(t)\right) + \frac{1}{2}\Delta_{\mathcal{M}} r(t)\, dt \tag{79b}$$

Let $[\xi^i, r]$ denote the quadratic covariation of ξ^i and r. From [19] (Proposition 5.18, Page 63), since z is a Riemannian Brownian motion, it follows from (79a) and (79b) that

$$d[\xi^i, r](t) = \beta_q^{-2}(r(t))\, I(e^i, \partial_r)(z(t))\, dt = 0 \tag{79c}$$

as e^i and ∂_r are orthogonal with respect to I. This means that $(\xi^i(t))$ and $r(t)$ have zero quadratic covariation, and therefore $(\xi^i(t))$ and $w(t)$ have zero quadratic covariation. It follows as in [23] (Lemma 3.3.4, Page 85), that $r(t)$ and $\theta_q(t)$ are independent. ■