Otto's metric on location-scale models and warped Riemannian metric

K. L. Chitou, F. Djibril Moussa, A. Gbaguidi Amoussou, C. Ogouyandjou

Abstract. In this paper, we show that the Otto's metric on a location-scale model defined on a Riemannian manifold is a warped Riemannian metric. This has been done by assuming that the location-scale model is invariant under the action of some Lie group. The obtained result is applied to the von Mises-Fisher model and to the Riemannian Gaussian model.

M.S.C. 2010: 31C12, 76M60, 53B12.
Key words: Otto's metric; location-scale model; warped Riemannian metric; Riemannian manifold; Lie group.

1 Introduction

Let (M, g_M) and (N, g_N) be two Riemannian manifolds of positive dimensions. Given a positive smooth function f on M, the warped product $B = M \times_f N$, is by definition the manifold $M \times N$ equipped with the warped product Riemannian metric $g = g_M + f^2 g_N$(see [1]). Warped product plays important roles in many fields such as differential geometry, physics and information geometry. In information geometry, warped product is an efficient tool to find the expression of the Rao–Fisher information metric of location-scale models defined on high-dimensional Riemannian manifolds.

Information geometry began as the geometric study on a statistical model $\mathcal{P}$ which is a set of probability distributions. A fundamental idea of information geometry is to identify the statistical model $\mathcal{P}$ to its parameter space which is assumed to be a manifold. Therefore, a Riemannian structure on $\mathcal{P}$ can be defined on its parameter space and a natural one is given by the Rao–Fisher information [2]. Said et al. [10] showed 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.

In this paper, we are interested in the Otto's metric, which has been introduced in [9] to give a gradient flow on Wasserstein space. Lott [6] studied the Riemannian geometric structure of Wasserstein space and performed some valuable geometric calculations on that space using Otto's metric. Considering this metric on a set

Applied Sciences, Vol. 24, 2022, pp. 43-61.

of probability distributions, Ogouyandjou et al. [7] proposed a family of torsion-free α-connections that coincide with the Levi-Civita connection when $\alpha = 0$. The present contribution investigates the geometry of location-scale models equipped with the Riemannian Otto's metric. We show, under some assumptions, that the Otto's metric of any location-scale model defines a warped Riemannian metric.

The rest of the paper is organized as follows: Section 2 is devoted to some preliminary results. In Section 3, we give the main result on the geometry of location-scale model. In Section 4 and Section 5, we focus on the applications of the result to the von Mises-Fisher model and to the Riemannian Gaussian model.

2 Preliminaries

Let (M, R) be a Riemannian manifold and L denotes the Lie group of the isometries of M. We define an action of L on M by:

$$g \cdot x = g(x) \quad \forall g \in L\,, \forall x \in M.$$

Let $K_{\bar{x}}$ be the subgroup of L which consists of the elements k such that $k \cdot \bar{x} = \bar{x}$. For each $k \in K_{\bar{x}}$, the derivative $d_{\bar{x}}k$ of k at the point $\bar{x}$ is a linear map of the tangent space $T_{\bar{x}}M$ of M at $\bar{x}$. The map $k \mapsto d_{\bar{x}}k$ is a representation of $K_{\bar{x}}$ in $T_{\bar{x}}M$, called the isotropy representation.

The Riemannian manifold M is said to be a Riemannian symmetric space, if for each $\bar{x} \in M$ there exists an isometry $s_{\bar{x}} \in L$, whose effect is to fix $\bar{x}$ and to reverse the geodesic curves passing through $\bar{x}$. In addition, if the isotropy representation is irreducible, that is, if the isotropy representation has no invariant subspaces in $T_{\bar{x}}M$, except $\{0\}$ and $T_{\bar{x}}M$, M is called irreducible Riemannian symmetric space.

Let $\mathcal{M} = M \times (0, \infty)$, α and β be positive smooth functions defined on $(0, \infty)$. A warped Riemannian metric G on $\mathcal{M}$ can be defined as follows: for each $z = (\bar{x}, \sigma) \in \mathcal{M}$,

$$G_z\left(U, U\right) = (\alpha(\sigma) u_\sigma)^2 + \beta^2(\sigma) R_{\bar{x}}\left(u, u\right) \quad \forall U \in T_z\mathcal{M},$$

where $T_z\mathcal{M}$ denotes the tangent space of $\mathcal{M}$ at z and $U = u_\sigma \partial_\sigma + u$, with $u_\sigma \in \mathbb{R}$, $\partial_\sigma := \dfrac{\partial}{\partial_\sigma}$ and $u \in T_{\bar{x}}M$. For instance see [10].

Given a manifold M, a location-scale model on M is a family $\mathcal{P}$ of parametric probability distributions $p\left(\cdot \,|\, \bar{x},\, \sigma\right)$ on M ($\bar{x} \in M$ and $\sigma \in (0, \infty)$), with respect to the Riemannian density of M, such that the map $(\bar{x}, \sigma) \mapsto p\left(\cdot \,|\, \bar{x},\, \sigma\right)$ is injective:

$$\mathcal{P} = \{\, p\left(\cdot \,|z\right) : z = (\bar{x}, \sigma) \in \mathcal{M}\}.$$

Throughout the paper, we will make use of the following notations:

- $\ell^{(\cdot)}$ will denote the function given by $\ell^{(x)}(z) = \log\left(p\left(x \,|\, z\right)\right)$;
- $d_z\ell^{(\cdot)}$, the map $x \mapsto d_z\ell^{(x)}$ where $d_z\ell^{(x)}$ is the derivative of $\ell^{(x)}$ at the point z;
- $d_{\bar{x}}\ell^{(x)}$, the derivative of the map $(\bar{x}, \sigma) \mapsto \ell^{(x)}(\bar{x}, \sigma)$ with σ fixed;
- $d_z\ell^{(\cdot)}u$, for $u \in T_{\bar{x}}M$, the map $x \mapsto d_z\ell^{(x)}u$;

- $\mathcal{C}(M)$, the set of all smooth real-valued functions on M.

The expectation $\mathbb{E}_z$ of a function f on M with respect to the probability density function $p(\cdot|z)$ is defined by:

$$\mathbb{E}_z(f) = \int_M f(x)p(x|z)\,\mathrm{dvol}_M.$$

In [7], the Otto's metric I is defined on the model $\mathcal{P}$ as follows: for each $z = (\bar{x}, \sigma) \in \mathcal{M}$,

$$I_z\,(U, U) = \mathbb{E}_z R\left(\mathrm{grad}_x\, d_z\ell^{(\cdot)}\,U,\, \mathrm{grad}_x\, d_z\ell^{(\cdot)}\,U\right) \quad \forall U \in T_z\mathcal{M}, \tag{2.1}$$

where grad_x denotes the Riemannian gradient with respect to x.

3 Main results

In this section, we show that Otto's metric on a location-scale model defined on the Riemannian manifold M is a warped Riemannian metric. To do this, we assume that M is an irreducible symmetric space under the transitive action of the Lie group L and that the location-scale model $\mathcal{P}$ is invariant under the action of L. The invariance condition of the model $\mathcal{P}$ under the action of L is:

$$p\,(g \cdot x \,|\, g \cdot \bar{x},\, \sigma) = p\,(x \,|\, \bar{x},\, \sigma)\ \forall g \in L, \tag{3.1}$$

where $g \cdot x$ denotes the action of $g \in L$ on $x \in M$. This identity is equivalent to the following relation:

$$\mathbb{E}_{g\cdot z}\, f = \mathbb{E}_z\, f \circ g, \quad \forall f \in \mathcal{C}(M),\, \forall g \in L, \tag{3.2}$$

where the function $f \circ g$ is given by $f \circ g(x) = f(g \cdot x)$, (see [10]).

Proposition 3.1. *[10] , If Condition* (3.2) *holds, then*

$$\partial_\sigma\, \ell^{(\cdot)}(z) \circ g = \partial_\sigma\, \ell^{(\cdot)}\,(g^{-1} \cdot z) \text{ and } d_{\bar{x}}\, \ell^{(\cdot)} \circ g = d_{g^{-1}\cdot\bar{x}}\ell^{(\cdot)} \circ d_{\bar{x}}g^{-1}.$$

If $g = s_{\bar{x}}$, then

$$\partial_\sigma\ell^{(\cdot)}\,(z) \circ s_{\bar{x}} = \partial_\sigma\ell^{(\cdot)}\,(z) \text{ and } d_{\bar{x}}\,\ell^{(\cdot)} \circ s_{\bar{x}} = -d_{\bar{x}}\,\ell^{(\cdot)}$$

Using this proposition, we prove the following result.

Proposition 3.2. *If Condition* (3.1) *holds, then for all $u \in T_{\bar{x}}M$, we have:*

$$\left(grad_x\, \partial_\sigma\ell^{(\cdot)}\,(z)\right) \circ g = dg \cdot grad_x\, \partial_\sigma\ell^{(\cdot)}\,(g^{-1} \cdot z), \tag{3.3}$$

$$\left(grad_x\, d_{\bar{x}}\ell^{(\cdot)}u\right) \circ g = dg \cdot grad_x\left(d_{g^{-1}\cdot\bar{x}}\ell^{(\cdot)}(d_{\bar{x}}g^{-1}\,u)\right), \tag{3.4}$$

$$\left(grad_x\, \partial_\sigma\ell^{(\cdot)}\,(z)\right) \circ s_{\bar{x}} = ds_{\bar{x}} \cdot grad_x\, \partial_\sigma\ell^{(\cdot)}\,(z), \tag{3.5}$$

and

$$\left(grad_x\, d_{\bar{x}}\ell^{(\cdot)}u\right) \circ s_{\bar{x}} = -ds_{\bar{x}} \cdot grad_x\left(d_{\bar{x}}\ell^{(\cdot)}u\right). \tag{3.6}$$

Proof. Set $u \in T_{\bar{x}}M$. First, note that for all $x \in M$: $\big(\mathrm{grad}_x\, \partial_\sigma \ell^{(\cdot)}\,(z)\big) \circ g(x) \in T_{g\cdot x}M$ and $\big(\mathrm{grad}_x\, d\ell(z)u\big) \circ g(x) \in T_{g\cdot x}M$. For any $v \in T_{g\cdot x}M$, we have:

$$\begin{aligned} R_{g\cdot x}\left(\big(\mathrm{grad}_x\, \partial_\sigma \ell^{(\cdot)}\,(z)\big) \circ g(x),\, v\right) &= R_{g\cdot x}\left(\mathrm{grad}_x\, \partial_\sigma \ell^{(\cdot)}\,(z)(g\cdot x),\, v\right), \\ &= \left(d_{g\cdot x} \partial_\sigma \ell^{(\cdot)}\,(z)\right) v, \end{aligned}$$

by the definition of Riemannian gradient. From the chain rule,

$$d_x\left(\partial_\sigma \ell^{(\cdot)}(z) \circ g\right) = d_{g\cdot x} \partial_\sigma\, \ell^{(\cdot)}(z) \circ d_x g,$$

and composing at right-hand side of the equality by $(d_x g)^{-1}$, we get:

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

Hence

$$\begin{aligned} R_{g\cdot x}\left(\big(\mathrm{grad}_x\, \partial_\sigma \ell^{(\cdot)}\,(z)\big) \circ g(x),\, v\right) &= d_x(\partial_\sigma \ell^{(\cdot)}(z) \circ g) \cdot \big((d_x g)^{-1} v\big), \\ &= d_x \partial_\sigma \ell^{(\cdot)}(g^{-1}\cdot z) \cdot \big((d_x g)^{-1} v\big) \quad \text{by Proposition 3.1}, \\ &= R_x\left(\mathrm{grad}_x\, \partial_\sigma \ell^{(x)}(g^{-1}\cdot z),\, (d_x g)^{-1} v\right), \\ &= R_{g\cdot x}\left(dg \cdot \mathrm{grad}_x\, \partial_\sigma \ell^{(\cdot)}(g^{-1}\cdot z)(x),\, v\right), \end{aligned}$$

since g is an isometry of M. Thus,

$$\begin{aligned} \big(\mathrm{grad}_x\, \partial_\sigma \ell^{(\cdot)}\,(z)\big) \circ g(x) &= d_x g \cdot \mathrm{grad}_x\, \partial_\sigma \ell^{(\cdot)}(g^{-1}\cdot z)(x), \\ &= \big(dg \cdot \mathrm{grad}_x\, \partial_\sigma \ell^{(\cdot)}(g^{-1}\cdot z)\big)(x), \end{aligned}$$

and (3.3) is proved.

Next, for all v in $T_{g\cdot x}M$,

$$\begin{aligned} R_{g\cdot x}\left(\big(\mathrm{grad}_x\, d_{\bar{x}} \ell^{(\cdot)} u\big) \circ g(x),\, v\right) &= d_x(d_{\bar{x}} \ell^{(\cdot)} u \circ g) \cdot [(d_x g)^{-1} \cdot v], \\ &= R_x\left(\mathrm{grad}_x\big(d_{\bar{x}} \ell^{(\cdot)} u \circ g\big)(x),\, (d_x g)^{-1} \cdot v\right), \\ &= R_{g\cdot x}\left(d_x g \cdot \mathrm{grad}_x\big(d_{\bar{x}} \ell^{(\cdot)} u \circ g\big)(x),\, v\right), \\ &= R_{g\cdot x}\left(dg \cdot \mathrm{grad}_x\big(d_{\bar{x}} \ell^{(\cdot)} \circ g\, u\big)(x),\, v\right). \end{aligned}$$

Then, from Proposition 3.1, we have for any $u \in T_{\bar{x}}M$:

$$\begin{aligned} \big(\mathrm{grad}_x\, d_{\bar{x}} \ell^{(\cdot)} u\big) \circ g &= dg \cdot \mathrm{grad}_x\big(\, d_{\bar{x}} \ell^{(\cdot)} \circ g\, u\big), \\ &= dg \cdot \mathrm{grad}_x\left(d_{g^{-1}\cdot \bar{x}} \ell^{(\cdot)} \cdot \big(d_{\bar{x}} g^{-1}\, u\big)\right), \end{aligned}$$

which is (3.4).

Let's now prove (3.5) and (3.6). Taking $g = s_{\bar{x}}$ in (3.3) and (3.4) give:

$$\begin{aligned} \big(\mathrm{grad}_x\, \partial_\sigma \ell^{(\cdot)}(z)\big) \circ s_{\bar{x}} &= ds_{\bar{x}} \cdot \mathrm{grad}_x\, \partial_\sigma \ell^{(\cdot)}(s_{\bar{x}}^{-1}\cdot z), \\ &= ds_{\bar{x}} \cdot \mathrm{grad}_x\, \partial_\sigma \ell^{(\cdot)}(z), \end{aligned}$$

and

$$\left(\mathrm{grad}_x\, d_{\bar{x}}\ell^{(\cdot)}u\right) \circ s_{\bar{x}} = ds_{\bar{x}} \cdot \mathrm{grad}_x\Big(d_{s_{\bar{x}}^{-1}\cdot\bar{x}}\ell^{(\cdot)} \cdot \left(d_{\bar{x}}(s_{\bar{x}})^{-1}\,u\right)\Big),$$
$$= -ds_{\bar{x}} \cdot \mathrm{grad}_x\left(d_{\bar{x}}\ell^{(\cdot)}\,u\right),$$

where $s_{\bar{x}}^{-1} \cdot z = z$ and $d_{\bar{x}}s_{\bar{x}}^{-1} = (d_{\bar{x}}s_{\bar{x}})^{-1} = -I$. Then,

$$\left(\mathrm{grad}_x\,\partial_\sigma\ell^{(\cdot)}(z)\right) \circ s_{\bar{x}} = ds_{\bar{x}} \cdot \mathrm{grad}_x\,\partial_\sigma\ell^{(\cdot)}(z)$$

and

$$\left[\mathrm{grad}_x\left(d_{\bar{x}}\ell^{(\cdot)}u\right)\right] \circ s_{\bar{x}} = -ds_{\bar{x}} \cdot \mathrm{grad}_x\,\left(d_{\bar{x}}\ell\,u\right).$$

□

Theorem 3.3 (Main Theorem)**.** *Let (M,R) be an irreducible symmetric space under the transitive action of the Lie group L of its isometries and $\mathcal{P}$ a location-scale model invariant under the action of L. The Otto's metric I of the model $\mathcal{P}$ is a warped Riemannian metric given by:*

$$I_z\,(U,U) = (\alpha(\sigma)u_\sigma)^2 + \beta^2(\sigma)R_{\bar{x}}\,(u,u) \quad \forall U \in T_z\mathcal{M}, \tag{3.7}$$

with

$$\alpha^2(\sigma) = \mathbb{E}_z\,R\left(grad_x\,\partial_\sigma\ell^{(\cdot)}(z), grad_x\,\partial_\sigma\ell^{(\cdot)}(z)\right), \tag{3.8}$$

and

$$\beta^2(\sigma) = \frac{tr(I_z|_{T_{\bar{x}}M})}{dim\ M}, \tag{3.9}$$

where $I_z|_{T_{\bar{x}}M}$ denotes the restriction of I_z to T_zM and $\partial_\sigma := \frac{\partial}{\partial_\sigma}$.

Proof. Since I is bilinear and symmetric,

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

with $U = u_\sigma\,\partial_\sigma + u \in T_z\mathcal{M}$ with $u_\sigma \in \mathbb{R}$ and $u \in T_{\bar{x}}M$. We shall prove the following relations,

$$I_z(\partial_\sigma,\partial_\sigma) = \alpha^2(\sigma), \tag{3.10}$$
$$I_z(\partial_\sigma,u) = 0, \tag{3.11}$$
$$I_z(u,u) = \beta^2(\sigma)\,R_{\bar{x}}(u,u), \tag{3.12}$$

where $\alpha^2(\sigma)$ and $\beta^2(\sigma)$ are given by (3.8) and (3.9).
We shall first prove (3.10). From (2.1),

$$I_z(\partial_\sigma,\partial_\sigma) = \mathbb{E}_zR\left(\mathrm{grad}_x\,d_z\ell^{(\cdot)}\,\partial_\sigma,\,\mathrm{grad}_x\,d_z\ell^{(\cdot)}\,\partial_\sigma\right),$$
$$= \mathbb{E}_zR\left(\mathrm{grad}_x\,\partial_\sigma\ell^{(\cdot)}(z),\,\mathrm{grad}_x\,\partial_\sigma\ell^{(\cdot)}(z)\right). \tag{3.13}$$

We shall show that the right-hand side of (3.13) does not depend on z. Let f denote the function defined by $f(x) = R_x\left(\mathrm{grad}_x\, \partial_\sigma \ell^{(x)}(z),\, \mathrm{grad}_x\, \partial_\sigma \ell^{(x)}(z)\right)$ and set $\alpha^2(z) = \mathbb{E}_z f$. For all $g \in L$, we have from (3.2):

$$\alpha^2(g\cdot z) = \mathbb{E}_{g\cdot z} f = \mathbb{E}_z f \circ g. \tag{3.14}$$

For all $x \in M$,

$$\begin{aligned} f\circ g\,(x) &= R_{g\cdot x}\left(\mathrm{grad}_x\, \partial_\sigma \ell^{(g\cdot x)}(g\cdot z), \mathrm{grad}_x\, \partial_\sigma \ell^{(g\cdot x)}(g\cdot z)\right),\\ &= R_{g\cdot x}\left(d_x g\cdot \mathrm{grad}_x\, \partial_\sigma \ell^{(x)}(z),\, d_x g\cdot \mathrm{grad}_x\, \partial_\sigma \ell^{(x)}(z)\right),\ \text{by Proposition 3.2}\\ &= R_x\left(\mathrm{grad}_x\, \partial_\sigma \ell^{(x)}(z),\, \mathrm{grad}_x\, \partial_\sigma \ell^{(x)}(z)\right),\ \text{since } g \text{ is an isometry of } M\\ &= f(x). \end{aligned} \tag{3.15}$$

Substituting (3.15) in (3.14) yields:

$$\alpha^2(g\cdot z) = \alpha^2(z).$$

Since L acts transitively on M, for all $\bar{x}_1$ and $\bar{x}_2$ in M, we have

$$\alpha^2(z_1) = \alpha^2(z_2),$$

where $z_1 = (\bar{x}_1, \sigma)$ and $z_2 = (\bar{x}_2, \sigma)$. Then, the right-hand side of (3.13) does not depend on $\bar{x}$ so that we get (3.10):

$$I_z(\partial_\sigma, \partial_\sigma) = \mathbb{E}_z R\left(\mathrm{grad}_x \partial_\sigma \ell^{(\cdot)}(z),\, \mathrm{grad}_x \partial_\sigma \ell^{(\cdot)}(z)\right) = \alpha^2(\sigma).$$

Now let's prove (3.11). We have:

$$I_z(\partial_\sigma,\, u) = \mathbb{E}_z R\left(\mathrm{grad}_x\big(d_z \ell^{(\cdot)}\, u\big),\, \mathrm{grad}_x\, \partial_\sigma \ell^{(\cdot)}(z)\right),$$

Let h denote the function under the expectation. Applying (3.2) with $g = s_{\bar{x}}$ yields:

$$I_z(\partial_\sigma,\, u) = \mathbb{E}_z h = \mathbb{E}_{s_{\bar{x}}\cdot z}\, h = \mathbb{E}_z\,(h\circ s_{\bar{x}}), \tag{3.16}$$

since $s_{\bar{x}}\cdot z = z$. For all $x \in M$,

$$\begin{aligned} (h\circ s_{\bar{x}})(x) &= R_{s_{\bar{x}}\cdot x}\ \left(\mathrm{grad}_x\big(d_z \ell^{(s_{\bar{x}}\cdot x)} u\big),\, \mathrm{grad}_x\, \partial_\sigma \ell^{(s_{\bar{x}}\cdot x)}(z)\right),\\ &= R_{s_{\bar{x}}\cdot x}\left(-d_x s_{\bar{x}}\cdot \mathrm{grad}_x\big(d_z \ell^{(x)}\, u\big),\, d_x s_{\bar{x}}\cdot \mathrm{grad}_x\, \partial_\sigma \ell^{(x)}(z)\right)\\ &= -R_x\left(\mathrm{grad}_x\big(d_z \ell^{(x)}\, u\big),\, \mathrm{grad}_x\, \partial_\sigma \ell^{(x)}(z)\right)\\ &= -h(x). \end{aligned} \tag{3.17}$$

where the second equality follows from Proposition 3.2 and where the third equality follows from the fact that $s_{\bar{x}}$ is an isometry of M. Replacing (3.17) in (3.16), we get

$$I_z(\partial_\sigma,\, u) = \mathbb{E}_z\, h = 0.$$

(3.12) is a consequence of Schur's lemma. As $I_z|_{T_{\bar{x}}M}$ and $R_{\bar{x}}$ are two bilinears forms on $T_{\bar{x}}M$ with $R_{\bar{x}}$ definite positive, then for all $k \in K_{\bar{x}}$:

$$R_{\bar{x}}(u,\, u) = R_{\bar{x}}(d_{\bar{x}}k\, u,\, d_{\bar{x}}k\, u) \quad \text{and} \quad I_z\,(u,\, u) = I_z\,(d_{\bar{x}}k\, u,\, d_{\bar{x}}k\, u) \tag{3.18}$$

The first equality holds because k is an isometry of (M, R). To obtain the last equality, first recall that

$$I_z\,(u,\,u) = \mathbb{E}_z\, R\Big(\mathrm{grad}_x\big(d_z\ell^{(\cdot)}\,u\big),\,\mathrm{grad}_x\big(d_z\ell^{(\cdot)}\,u\big)\Big).$$

Denoting by j the function under the expectation and applying (3.2) yield

$$I_z\,(u,\,u) = \mathbb{E}_z\, j = \mathbb{E}_{k^{-1}\cdot z}\, j = \mathbb{E}_z\, j\circ k^{-1} \tag{3.19}$$

since $k^{-1}\cdot z = z$ for $k \in K_{\bar{x}}$. For all $x \in M$, by Proposition 3.2, we have:

$$\begin{aligned} j\circ k^{-1}(x) &= R_{k^{-1}\cdot x}\,\Big(\mathrm{grad}_x\big(d_z\ell^{(k^{-1}\cdot x)}u\big),\,\mathrm{grad}_x\big(d_z\ell^{(k^{-1}\cdot x)}u\big)\Big)\,,\\ &= R_{k^{-1}\cdot x}\,\Big(d_xk^{-1}\mathrm{grad}_x\big(d_{\bar{x}}\ell^{(x)}(d_{\bar{x}}k\,u)\big),\,d_xk^{-1}\mathrm{grad}_x\big(d_{\bar{x}}\ell^{(x)}(d_{\bar{x}}k\,u)\big)\Big),\\ &= R_x\,\Big(\mathrm{grad}_x\big(d_{\bar{x}}\ell^{(x)}(d_{\bar{x}}k\,u)\big),\,\mathrm{grad}_x\big(d_{\bar{x}}\ell^{(x)}(d_{\bar{x}}k\,u)\big)\Big), \end{aligned} \tag{3.20}$$

because k is an isometry of M and $k\cdot z = z$. Then, we get (3.18) by subtituting (3.20) in (3.19). (3.18) means that $I_z|_{T_{\bar{x}}M}$ and $R_{\bar{x}}$ are invariant by the isotropy representation. Hence, applying the Schur's lemma, one can find some positive multiplicative factor β^2, such that:

$$I_z\,(u,\,u) = \beta^2\,R_{\bar{x}}\,(u,\,u). \tag{3.21}$$

We have to show that β^2 is given by (3.9). Taking the trace in (3.21),

$$\mathrm{tr}(I_z|_{T_{\bar{x}}M}) = \beta^2\,\mathrm{tr}\,R_{\bar{x}} = \beta^2\dim M.$$

Then

$$\beta^2 = \beta^2\,(z) = \frac{\mathrm{tr}(I_z|_{T_{\bar{x}}M})}{\dim M}.$$

Let's now show that β^2 is not depend on $\bar{x}$. Taking $\bar{x}_1$ and $\bar{x}_2$ in M, we have to show that

$$\beta^2\,(z_1) = \beta^2\,(z_2), \tag{3.22}$$

where $z_1 = (\bar{x}_1,\,\sigma)$ and $z_2 = (\bar{x}_2,\,\sigma)$. We have:

$$\mathrm{tr}(I_{z_2}|_{T_{\bar{x}_2}M}) = \sum_{i=1}^{n} I_{z_2}\,(e_i, e_i), \tag{3.23}$$

where $(e_1,\,e_2,\,\cdots,\,e_n)$ is an orthonormal basis of $T_{\bar{x}_2}M$. As the action of L on M is transitive, there exists $g \in L$ such that:

$$\bar{x}_2 = g\cdot\bar{x}_1.$$

Using this, we get:

$$I_{z_2}\,(e_i, e_i) = I_{g\cdot z_1}\,(e_i, e_i) = \mathbb{E}_{g\cdot z_1}\,t = \mathbb{E}_{z_1}\,t\circ g, \tag{3.24}$$

where $t = R\left(\mathrm{grad}_x\, d_{\bar{x}_2}\ell^{(\cdot)}\, e_i,\, \mathrm{grad}_x\, d_{\bar{x}_2}\ell^{(\cdot)}\, e_i\right)$. For all $x \in M$,

$$\begin{aligned} t \circ g\,(x) &= R_{g\cdot x}\left(\mathrm{grad}_x\, d_{\bar{x}_2}\ell^{(g\cdot x)}\, e_i,\, \mathrm{grad}_x\, d_{\bar{x}_2}\ell^{(g\cdot x)}\, e_i\right), \\ &= R_{g\cdot x}\left(d_x g\ \mathrm{grad}_x\big(d_{\bar{x}_1}\ell^{(x)}(d_{\bar{x}_2}g^{-1}\, e_i)\big), d_x g\ \mathrm{grad}_x\big(d_{\bar{x}_1}\ell^{(x)}(d_{\bar{x}_2}g^{-1}\, e_i)\big)\right), \\ &= R_x\left(\mathrm{grad}_x\big(d_{\bar{x}_1}\ell^{(x)}(d_{\bar{x}_2}g^{-1}\, e_i)\big),\ \mathrm{grad}_x\big(d_{\bar{x}_1}\ell^{(x)}(d_{\bar{x}_2}g^{-1}\, e_i)\big)\right). \end{aligned}$$

where the second equality holds by Proposition 3.2 and (3).

Replacing the last equality in (3.24) yields

$$\begin{aligned} I_{z_2}\,(e_i, e_i) &= \mathbb{E}_{z_1}\, R\left(\mathrm{grad}_x\big(d_{\bar{x}_1}\ell^{(\cdot)}(d_{\bar{x}_2}g^{-1}\, e_i)\big),\, \mathrm{grad}_x\big(d_{\bar{x}_1}\ell^{(\cdot)}(d_{\bar{x}_2}g^{-1}\, e_i)\big)\right) \\ &= I_{z_1}\,\big(d_{\bar{x}_2}g^{-1}\, e_i,\, d_{\bar{x}_2}g^{-1}\, e_i\big). \end{aligned}$$

Then (3.23) becomes

$$\mathrm{tr}(I_{z_2}|_{T_{\bar{x}_2}M}) = \sum_{i=1}^{n} I_{z_1}\,(d_{\bar{x}_2}g^{-1}\, e_i,\, d_{\bar{x}_2}g^{-1}\, e_i). \tag{3.25}$$

To complete the proof, note that $d_{\bar{x}_2}g^{-1}\, e_i$ is in $T_{\bar{x}_1}M$ since $d_{\bar{x}_2}g^{-1}$ maps an element of $T_{\bar{x}_2}M$ to an element of $T_{g^{-1}\cdot\bar{x}_2}M = T_{\bar{x}_1}M$. Therefore, for all $1 \leq i,\, j \leq n$, we get:

$$\begin{aligned} R_{\bar{x}_1}\,(d_{\bar{x}_2}g^{-1}\, e_i,\, d_{\bar{x}_2}g^{-1}\, e_j) &= R_{g^{-1}\cdot\bar{x}_2}\,(d_{\bar{x}_2}g^{-1}\, e_i,\, d_{\bar{x}_2}g^{-1}\, e_j) \\ &= R_{\bar{x}_2}\,(e_i,\, e_j). \end{aligned} \tag{3.26}$$

Since the family $(e_1,\, e_2,\, \cdots,\, e_n)$ is an orthonormal basis of $T_{\bar{x}_2}M$, the equality (3.26) shows that $\big(d_{\bar{x}_2}g^{-1}\, e_1,\, d_{\bar{x}_2}g^{-1}\, e_2,\, \cdots,\, d_{\bar{x}_2}g^{-1}\, e_n\big)$ is an orthonormal basis of $T_{\bar{x}_1}M$. Then, (3.25) becomes

$$\mathrm{tr}(I_{z_2}|_{T_{\bar{x}_2}M}) = \sum_{i=1}^{n} I_{z_1}\,(d_{\bar{x}_2}g^{-1}\, e_i,\, d_{\bar{x}_2}g^{-1}\, e_i) = \mathrm{tr}(I_{z_1}|_{T_{\bar{x}_1}M}).$$

From this, we obtain (3.22). □

4 Application to the von Mises-Fisher model

In this section, we give an application of Theorem 3.3 to the von Mises-Fisher model.

Definition 4.1. The von Mises-Fisher model is a location-scale model $\mathcal{P}$ defined on $\mathbb{S}^{n-1}$, the unit sphere of $\mathbb{R}^n$ $(n \geq 3)$ given by: for all $p\,(\cdot|\bar{x}, \eta) \in \mathcal{P}$

$$p\,(x|\bar{x},\, \eta) = \exp\left[\eta\, D\,(x,\, \bar{x}) - \psi\,(\eta)\right] \quad \forall x \in \mathbb{S}^{n-1},\, \eta \in (0, \infty), \tag{4.1}$$

where

$$D(x, \bar{x}) = \langle x, \bar{x}\rangle \text{ and } \psi(\eta) = \nu \log(2\pi) + \log\left(\eta^{1-\nu} I_{\nu-1}(\eta)\right), \tag{4.2}$$

with $\langle\cdot,\cdot\rangle$ denoting the Euclidean product of $\mathbb{R}^n$ and $I_{\nu-1}$ denoting the modified Bessel function of order $\nu - 1$, with $\nu = n/2$.

Remark 4.2. The parameter space of the model is not of the form $M \times (0, \infty)$. So, Theorem 3.3 can't be directly applied to the von Mises-Fisher model. To remedy this, one identifies the parameter space of the model to $\mathbb{R}^n$ (see [10]). In the following proposition, we call warped Riemannian metric I on $\mathbb{R}^n \backslash \{0\}$, the pull-forward of the warped Riemannian metric on $\mathbb{S}^{n-1} \times (0, \infty)$ given by (3.7) via the diffeomorphism $(\bar{x}, \eta) \mapsto \eta\, \bar{x}$.

$$I_z\,(U, U) = (\alpha\,(\eta)\, u_\eta)^2 + \beta^2\,(\eta)\, R_{\bar{x}}(u, u) \quad U \in T_z\mathbb{R}^n,$$

where $z = \eta\, \bar{x}$ and $U = u_\eta\, \bar{x} + \eta\, u$, with $\eta \in (0,\, \infty)$, $\bar{x} \in \mathbb{S}^{n-1}$, $u_\eta \in \mathbb{R}$ and $u \in T_{\bar{x}}\mathbb{S}^{n-1}$.

Proposition 4.1. *The Otto's metric I of the von Mises Fisher model is a warped Riemannian metric on the parameter space $\mathbb{R}^n$, given by*

$$I_z = \begin{cases} \alpha^2\,(\eta)\, d\eta^2 + \beta^2\,(\eta)\, R_{\bar{x}} & \text{if } z \neq 0 \\ \dfrac{n-1}{n} || \cdot || & \text{if } z = 0 \end{cases}$$

where

$$\alpha^2\,(\eta) = \frac{n-1}{n}\Big[1 - \frac{I_{\nu+1}\,(\eta)}{I_{\nu-1}\,(\eta)}\Big], \quad \beta^2\,(\eta) = \frac{\eta^2}{n}\Big(n - 1 + \frac{I_{\nu+1}\,(\eta)}{I_{\nu-1}\,(\eta)}\Big), \tag{4.3}$$

and $|| \cdot ||$ denotes the Euclidean norm.

Proof. The model $\mathcal{P}$ is invariant under the action of the Lie group L. Indeed, for all $g \in L$:

$$\begin{aligned} p\,(g \cdot x | g \cdot \bar{x},\, \eta) &= \exp\,[\,\eta\,\langle g \cdot x,\, g \cdot \bar{x}\rangle - \psi\,(\eta)], \\ &= \exp\,[\,\eta\,\langle x,\, \bar{x}\rangle - \psi\,(\eta)], \\ &= p\,(x | \,\bar{x},\, \eta). \end{aligned}$$

In addition, $\mathbb{S}^{n-1}$ is an irreducible symmetric space [3](Table II, p. 354). Then, applying Theorem 3.3 to the model $\mathcal{P}$ and transferring the metric obtained on $\mathbb{R}^n \backslash \{0\}$ we get:

$$I_z\,(U,\, U) = (\alpha\,(\eta)\, u_\eta)^2 + \beta^2\,(\eta)\, R_{\bar{x}}(u, u) \quad U \in T_z\mathbb{R}^n,$$

with $U = u_\eta\, \bar{x} + \eta\, u$ with $u_\eta \in \mathbb{R}$ and $u \in T_{\bar{x}}\mathbb{S}^{n-1}$, and where $\alpha^2\,(\eta)$ and $\beta^2\,(\eta)$ are given by

$$\alpha^2(\eta) = \mathbb{E}_z\, R\left(\operatorname{grad}_x \partial_\eta \ell^{(\cdot)}(z), \operatorname{grad}_x \partial_\eta \ell^{(\cdot)}(z)\right), \tag{4.4}$$

and

$$\beta^2(\eta) = \frac{\operatorname{tr}(I_z|_{T_{\bar{x}}\mathbb{S}^{n-1}})}{\dim \mathbb{S}^{n-1}}. \tag{4.5}$$

Note that the probability distributions related to the points $(\bar{x}, 0)$ are not considered in $\mathcal{P}$. Let's show that $\alpha^2\,(\eta)$ is given by (4.3). From (4.1), we have:

$$\partial_\eta \ell^{(x)}(z) = \partial_\eta\Big(\eta\, D\,(x,\, \bar{x}) - \psi\,(\eta)\Big) = D\,(x,\, \bar{x}) - \psi'\,(\eta).$$

Replacing this equality in (4.4), we get:

$$\alpha^2(\eta) = \mathbb{E}_z\, R\left(\operatorname{grad}_x D(\cdot,\, \bar{x}), \operatorname{grad}_x D(\cdot,\, \bar{x})\right). \tag{4.6}$$

For all $x \in \mathbb{S}^{n-1}$,

$$R_x\left(\operatorname{grad}_x D(x,\, \bar{x}), \operatorname{grad}_x D(x,\, \bar{x})\right) = ||\operatorname{grad}_x \langle x,\, \bar{x}\rangle||^2.$$

Here, note that the metric R considered on $M = \mathbb{S}^{n-1}$ is the scalar Euclidean product and D is given by (4.2). We have:

$$\operatorname{grad}_x \langle x,\, \bar{x}\rangle = \tan \overline{\operatorname{grad}_x} \langle x,\, \bar{x}\rangle,$$

where tan and $\overline{\operatorname{grad}_x}$ denote respectively the orthogonal projection on $T_{\bar{x}}\mathbb{S}^{n-1}$ and the gradient on $\mathbb{R}^n$. Then,

$$\operatorname{grad}_x \langle x,\, \bar{x}\rangle = \tan \bar{x} = \bar{x} - \langle x,\, \bar{x}\rangle\, \bar{x},$$

by the decomposition $T_{\bar{x}}\mathbb{R}^n = T_{\bar{x}}\mathbb{S}^{n-1} + \left(T_{\bar{x}}\mathbb{S}^{n-1}\right)^{\perp}$ (see [8]). Consequently, by Pythagora's Theorem:

$$R_x\left(\operatorname{grad}_x D(x,\, \bar{x}), \operatorname{grad}_x D(x,\, \bar{x})\right) = ||\bar{x} - \langle x,\, \bar{x}\rangle||^2 = 1 - \langle x,\, \bar{x}\rangle^2.$$

Substituting this in (4.6), we get:

$$\alpha^2(\eta) = 1 - \mathbb{E}_z \langle x,\, \bar{x}\rangle^2.$$

One can show that

$$\mathbb{E}_z \langle x,\, \bar{x}\rangle^2 = \frac{1}{n} + \frac{n-1}{n}\frac{I_{\nu+1}(\eta)}{I_{\nu-1}(\eta)}, \tag{4.7}$$

see for instance Appendix B of [10]. Then

$$\alpha^2(\eta) = \frac{n-1}{n}\Big[1 - \frac{I_{\nu+1}(\eta)}{I_{\nu-1}(\eta)}\Big].$$

Let's show that $\beta^2(\eta)$ is also given by (4.3). For this purpose, let $(e_i)_{1\le i\le n-1}$ be an orthonormal basis of $T_{\bar{x}}\mathbb{S}^{n-1}$. Then

$$\begin{aligned}
\operatorname{tr}(I_z|_{T_{\bar{x}}\mathbb{S}^{n-1}}) &= \sum_{i=1}^{n-1} I_z(e_i,\, e_i),\\
&= \sum_{i=1}^{n-1} \mathbb{E}_z\, R\left(\operatorname{grad}_x d_{\bar{x}}\ell^{(\cdot)}e_i,\, \operatorname{grad}_x d_{\bar{x}}\ell^{(\cdot)}e_i\right),\ \text{by (2.1)}\\
&= \mathbb{E}_z \sum_{i=1}^{n-1} R\left(\operatorname{grad}_x d_{\bar{x}}\ell^{(\cdot)}e_i,\, \operatorname{grad}_x d_{\bar{x}}\ell^{(\cdot)}e_i\right),
\end{aligned} \tag{4.8}$$

by (2.1) and by the linearity of the expectation. We have:

$$d_{\bar{x}}\ell^{(x)}e_i = R_{\bar{x}}\left(\operatorname{grad}_{\bar{x}} \ell^{(x)}(z),\, e_i\right) = \langle \operatorname{grad}_{\bar{x}} \ell^{(x)}(z), e_i\rangle, \tag{4.9}$$

where $\operatorname{grad}_{\bar{x}}$ is the gradient with respect to $\bar{x}$. By (4.1):

$$\operatorname{grad}_{\bar{x}} \ell^{(x)}(z) = \eta \operatorname{grad}_{\bar{x}} \langle x, \bar{x}\rangle = \eta\,(x - \langle x, \bar{x}\rangle \bar{x}). \tag{4.10}$$

Combining (4.9) with (4.10) we have for all $x \in \mathbb{S}^{n-1}$,

$$\begin{aligned} R_x\left(\operatorname{grad}_x d_{\bar{x}}\ell^{(x)}e_i, \operatorname{grad}_x d_{\bar{x}}\ell^{(x)}e_i\right) &= ||\operatorname{grad}_x\left(\eta\left\langle x - \langle x, \bar{x}\rangle \bar{x}, e_i\right\rangle\right)||^2, \\ &= \eta^2 ||\operatorname{grad}_x\left(\left\langle x - \langle x, \bar{x}\rangle \bar{x}, e_i\right\rangle\right)||^2, \\ &= \eta^2\left(1 - \left\langle x - \left\langle x, \bar{x}\right\rangle \bar{x}, \left\langle x - \langle x, \bar{x}\rangle \bar{x}, e_i\right\rangle e_i\right\rangle\right). \end{aligned}$$

Indeed

$$\begin{aligned} \operatorname{grad}_x \left\langle x - \langle x, \bar{x}\rangle \bar{x}, e_i\right\rangle &= \operatorname{grad}_x \langle x, e_i\rangle - \langle \bar{x}, e_i\rangle \operatorname{grad}_x \langle x, \bar{x}\rangle, \\ &= e_i - \langle x, e_i\rangle\, x, \end{aligned}$$

because $\langle \bar{x}, e_i\rangle = 0$. Then, taking the square of the norm and using Pythagora's Theorem, we get:

$$\begin{aligned} ||\operatorname{grad}_x \left\langle x - \langle x, \bar{x}\rangle \bar{x}, e_i\right\rangle||^2 &= 1 - \langle e_i, x\rangle^2, \\ &= 1 - \left\langle x - \langle x, \bar{x}\rangle \bar{x}, e_i\right\rangle^2 \quad \text{since } \langle \bar{x}, e_i\rangle = 0, \\ &= 1 - \left\langle x - \langle x, \bar{x}\rangle \bar{x}, \left\langle x - \langle x, \bar{x}\rangle \bar{x}, e_i\right\rangle e_i\right\rangle. \end{aligned}$$

Thus, replacing the expression of $R_x\left(\operatorname{grad}_x d_{\bar{x}}\ell^{(x)}e_i, \operatorname{grad}_x d_{\bar{x}}\ell^{(x)}e_i\right)$ in (4.8), we get:

$$\begin{aligned} \operatorname{tr}(I_z|_{T_{\bar{x}}\mathbb{S}^{n-1}}) &= \mathbb{E}_z\eta^2\left(n - 1 - \left\langle x - \left\langle x, \bar{x}\right\rangle \bar{x}, \sum_{i=1}^{n-1}\left\langle x - \langle x, \bar{x}\rangle \bar{x}, e_i\right\rangle e_i\right\rangle\right), \\ &= \mathbb{E}_z\eta^2\left(n - 1 - \left\langle x - \langle x, \bar{x}\rangle \bar{x}, \langle x - \langle x, \bar{x}\rangle \bar{x}\right\rangle\right), \\ &= \mathbb{E}_z\eta^2\left(n - 2 - \langle x, \bar{x}\rangle^2\right), \\ &= \eta^2(n - 2 + \mathbb{E}_z \langle x, \bar{x}\rangle^2), \quad \text{by (4.7)} \\ &= \eta^2\Big(n - 2 + \frac{1}{n} + \frac{n-1}{n}\frac{I_{\nu+1}(\eta)}{I_{\nu-1}(\eta)}\Big), \\ &= \eta^2\Big(\frac{(n-1)^2}{n} + \frac{n-1}{n}\frac{I_{\nu+1}(\eta)}{I_{\nu-1}(\eta)}\Big). \end{aligned}$$

Then,

$$\operatorname{tr}(I_z|_{T_{\bar{x}}\mathbb{S}^{n-1}}) = \eta^2\Big(\frac{(n-1)^2}{n} + \frac{n-1}{n}\frac{I_{\nu+1}(\eta)}{I_{\nu-1}(\eta)}\Big).$$

Replacing this in (4.5) yields

$$\beta^2(\eta) = \frac{\eta^2}{n}\Big(n - 1 + \frac{I_{\nu+1}(\eta)}{I_{\nu-1}(\eta)}\Big).$$

Let's now prove that I_z is well defined on the parameter space $\mathbb{R}^n$ of the model. Using the power series development of modified Bessel functions [11], one can show that

$$\frac{I_{\nu+1}(\eta)}{I_{\nu-1}(\eta)} = \frac{4}{n(n+2)}\left(\frac{1}{2}\eta\right)^2 + O(\eta^4).$$

From this equality, we infer

$$\alpha^2(\eta) = \frac{n-1}{n}\left(1 - \frac{4}{n(n+2)}\left(\frac{1}{2}\eta\right)^2\right) + O(\eta^4),$$

and

$$\beta^2(\eta) = \frac{\eta^2(n-1)}{n} + O(\eta^4).$$

Hence

$$\lim_{\eta\to 0}\alpha^2(\eta) = \frac{n-1}{n} \text{ and } \lim_{\eta\to 0}\beta^2(\eta) = 0. \tag{4.11}$$

Let $U_0 \in T_0\mathbb{R}^n$. Then there exists a vector field V on $\mathbb{R}^n$ such that $V(0) = U_0$ (see [8]). For any point $z \neq 0$, $V(z) \in T_z\mathbb{R}^n$ can be written as

$$V(z) = u_\eta\,\bar{x} + \eta\,u,$$

with $u_\eta \in \mathbb{R}$, $u \in T_{\bar{x}}\mathbb{S}^{n-1}$ and $z = \eta\,\bar{x}$ so that by Pythagora's Theorem

$$||V(z)||^2 = u_\eta^2 + \eta^2\,||u||^2.$$

Then,

$$\lim_{\eta\to 0} u_\eta^2 = \lim_{z\to 0}||V(z)||^2 = ||V(0)||^2 = ||U_0||^2. \tag{4.12}$$

Recall that

$$I_z(V(z), V(z)) = \alpha^2(\eta)\,u_\eta^2 + \beta^2(\eta)R_{\bar{x}}(u, u).$$

Note that the point $z = 0$ corresponds to $\eta = 0$. Now, using equations (4.11) and (4.12), we get:

$$\lim_{z\to 0} I_z(V(z), V(z)) = \frac{n-1}{n}||U_0||^2.$$

Let's then define a scalar product I_0 on $T_0\mathbb{R}^n$ by:

$$I_0(U_0, U_0) = \frac{n-1}{n}||U_0||^2 \quad \forall U_0 \in T_0\mathbb{R}^n.$$

To complete the proof, we have to show that for all vector field X and Y on $\mathbb{R}^n$, the map $\phi : \mathbb{R}^n \longrightarrow \mathbb{R}$ defined by:

$$\phi(z) = \begin{cases} \alpha^2(\eta)\,u_\eta\,v_\eta + \beta^2(\eta)\,\langle u, v\rangle \text{ if } z \neq 0 \\ \frac{n-1}{n}\langle X(0), Y(0)\rangle \text{ if } z = 0 \end{cases}$$

where $z = \eta\,\bar{x}$, $X(z) = u_\eta\,\bar{x} + \eta\,u$ and $Y(z) = v_\eta\,\bar{x} + \eta\,v$, is differentiable at the point $z = 0$. Let denote respectively by $\bar{x}_k$ and $z_k (k \in [\![1;n]\!])$ the components of $\bar{x}$ and z in $\mathbb{R}^n$. Using the chain rule [4], one can show that for any $z = \eta\,\bar{x}$:

$$\frac{\partial\phi(z)}{\partial z_k} = \frac{1}{\eta}\frac{\partial\phi(z)}{\partial\bar{x}_k} = \frac{\beta^2(\eta)}{\eta}\frac{\partial\langle u, v\rangle}{\partial\bar{x}_k} \quad \forall k \in [\![1;n]\!].$$

Then, since $\lim_{\eta\to 0}\frac{\beta^2(\eta)}{\eta} = 0$, we have $\lim_{z\to 0}\frac{\partial\phi(z)}{\partial z_k} = 0$.
On the other hand, we have

$$\frac{\partial\phi(0)}{\partial z_k} = \lim_{t\to 0}\frac{\phi(0,\cdots,t,\cdots,0) - \phi(0,\cdots,0)}{t} = 0.$$

To see this, we use L'Hopital's rule, by showing that

$$\lim_{t\to 0}\frac{d}{dt}\phi(0,\cdots,t,\cdots,0) = \lim_{t\to 0}\frac{\partial\phi}{\partial z_k}(0,\cdots,t,\cdots,0) = \lim_{z\to 0}\frac{\partial\phi(z)}{\partial z_k} = 0,$$

by the chain rule. Hence, ϕ is differentiable at the point $z = 0$. □

5 Application to the Riemannian Gaussian model

In this section, we find the expression of the Otto's metric I of the Riemannian Gaussian model as we did for the von Mises-Fisher model. We shall assume that M is simply connected.

Definition 5.1. The Riemannian Gaussian model is a location-scale model $\mathcal{P}$ defined on a Riemannian symmetric space of non-positive sectional curvature M by: for all $p(\cdot|\bar{x},\eta) \in \mathcal{P}$,

$$p(x|\bar{x},\,\eta) = \exp\left[\eta(\sigma)\; d^2(x,\,\bar{x}) - \psi\;(\eta(\sigma))\right]\ \forall x \in M, \tag{5.1}$$

where d is the Riemannian distance, $\psi(\eta(\sigma))$ normalizes $p(x|\bar{x},\,\eta)$ and $\eta(\sigma) = -\dfrac{1}{2\,\sigma^2}$.

In most cases of interest, the Riemannian Gaussian model is defined on a symmetric space and Theorem 3.3 can't be applied to a model defined on a such space.

Proposition 5.1. *Let (M,R) be an irreducible symmetric space such that the action of L on M is transitive. Let $\mathcal{P}$ be the Riemannian Gaussian model defined on M. Then, the Otto's metric of the model $\mathcal{P}$ is a warped Riemannian metric given by*

$$I_z(U,U) = (\alpha(\eta)\,u_\eta)^2 + \beta^2(\eta)\,R_{\bar{x}}(u,u) \quad \forall U \in T_z\mathcal{M},$$

with

$$\alpha^2(\eta) = 4\,\psi'\;(\eta)\,, \quad \beta^2(\eta) = 4\,\frac{\eta^2}{n}\sum_{l,k}\mathbb{E}_z R^3_{kl}, \tag{5.2}$$

where R_{kl} are components of R with respect to the normal coordinate system at x.

Proof. Let us first check the assumptions of Theorem 3.3. For $g \in L$ and $p(\cdot|\bar{x}, \eta) \in \mathcal{P}$, we have

$$\begin{aligned} p\,(g \cdot x \,|\, g \cdot z) &= \exp\left[\eta\, d^2\,(g \cdot x,\, g \cdot \bar{x}) - \psi\,(\eta)\right], \\ &= \exp\left[\eta\, d^2\,(x,\, \bar{x}) - \psi\,(\eta)\right], \end{aligned}$$

where $z = (\bar{x},\, \eta)$ and $g \cdot z = (g \cdot \bar{x},\, \eta)$. Then, the model is invariant under the action of L on M. So, applying Theorem 3.3 to the model, the Otto's metric turns to a warped Riemannian metric which is given by

$$I_z\,(U, U) = (\alpha\,(\eta)\, u_\eta)^2 + \beta^2\,(\eta)\, R_{\bar{x}}(u, u) \quad \forall U \in T_z\mathcal{M},$$

where

$$\alpha^2(\eta) = \mathbb{E}_z R\left(\operatorname{grad}_x \partial_\eta \ell^{(\cdot)}(z), \operatorname{grad}_x \partial_\eta \ell^{(\cdot)}(z)\right),$$

and

$$\beta^2(\eta) = \frac{\operatorname{tr}(I_z|_{T_{\bar{x}}M})}{\dim M}.$$

Let's show that $\alpha^2\,(\eta)$ and $\beta^2\,(\eta)$ are given by (5.2). By (5.1)

$$\operatorname{grad}_x \partial_\eta \ell^{(x)}(z) = \operatorname{grad}_x d^2\,(x,\, \bar{x}) = -2\ \exp_x^{-1}(\bar{x})\,.$$

The latter equality can be found in [5]. By replacing this in (5), we get:

$$\alpha^2(\eta) = \mathbb{E}_z\, R\,\left(-2\ \exp_x^{-1}(\bar{x})\,, -2\ \exp_x^{-1}(\bar{x})\right) = 4\,\mathbb{E}_z\, d^2\,(x,\, \bar{x})\,,$$

since $R_x\left(\exp_x^{-1}(\bar{x})\,, \exp_x^{-1}(\bar{x})\right) = d^2\,(x,\, \bar{x})$ (see [5]). Then, we show that

$$\mathbb{E}_z\, d^2\,(x,\, \bar{x}) = \psi'\,(\eta)\,.$$

Indeed, we have

$$\int_M \exp\left[\eta\, d^2\,(x,\, \bar{x}) - \psi\,(\eta)\right] \operatorname{dvol}_M(x) = 1.$$

From this, we get by differentiation

$$\int_M \left(d^2\,(x,\, \bar{x}) - \psi'\,(\eta)\right) p\,(x|z) \operatorname{dvol}_M(x) = 0. \tag{5.3}$$

A calculation of the left-hand side of (5.3) gives $\psi'\,(\eta) = \mathbb{E}_z\, d^2\,(x,\, \bar{x})$. Then, $\alpha^2\,(\eta) = 4\,\psi'\,(\eta)$. In addition, by the definition of Otto's metric

$$\begin{aligned} \operatorname{tr}\ (I_z|_{T_{\bar{x}}M}) &= \sum_{i=1}^n I_z\,(e_i,\, e_i)\,, \\ &= \sum_{i=1}^n \mathbb{E}_z\, R\left(\operatorname{grad}_x d_{\bar{x}}\ell^{(\cdot)} e_i,\, \operatorname{grad}_x d_{\bar{x}}\ell^{(\cdot)} e_i\right), \\ &= \sum_{i=1}^n \mathbb{E}_z\, R\left(\operatorname{grad}_x R_{\bar{x}}\left(\operatorname{grad}_{\bar{x}} \ell^{(\cdot)}(z), e_i\right),\, \operatorname{grad}_x R_{\bar{x}}\left(\operatorname{grad}_{\bar{x}} \ell^{(\cdot)}(z), e_i\right)\right), \end{aligned}$$

where $(e_1, e_2, \cdots, e_n)$ is a frame on M at $\bar{x}$. Let's note first that

$$\begin{aligned} \operatorname{grad}_x R_{\bar{x}} \left(\operatorname{grad}_x \ell^{(x)}(z), e_i\right) &= \eta \operatorname{grad}_x R_{\bar{x}} \left(\operatorname{grad}_{\bar{x}} d^2(x, \bar{x}), e_i\right), \\ &= -2\,\eta \operatorname{grad}_x R_{\bar{x}} \left(\exp_{\bar{x}}^{-1}(x), e_i\right). \end{aligned}$$

Let's introduce the normal coordinate system $h^1, \cdots, h^n$ of $\bar{x}$ relative to the basis $e_1, \cdots, e_n$ (see [8]). Then

$$\begin{aligned} \operatorname{grad}_x R_x \left(\operatorname{grad}_x \ell^{(x)}(z), e_i\right) &= -2\,\eta \operatorname{grad}_x h^i(x), \\ &= -2\,\eta \sum_{k,l} R_{kl}(x) \frac{\partial h^i(x)}{\partial h^k} \frac{\partial}{\partial h_l}\Big|_x, \\ &= -2\,\eta \sum_{l} R_{il}(x) \frac{\partial}{\partial h_l}\Big|_x. \end{aligned}$$

Replacing this in the expression of $\operatorname{tr}\left(I_z|_{T_{\bar{x}}M}\right)$ and performing a short calculation yield:

$$\begin{aligned} \operatorname{tr}\left(I_z|_{T_{\bar{x}}M}\right) &= 4\,\eta^2 \mathbb{E}_z \sum_{i,l,k} R_{il} R_{ik} R_{kl}, \\ &= 4\,\eta^2 \sum_{l,k} \mathbb{E}_z R_{kl}^3. \end{aligned}$$

Hence, $\beta^2(\eta)$ is given by:

$$\beta^2(\eta) = 4\,\frac{\eta^2}{n} \sum_{l,k} \mathbb{E}_z R_{kl}^3.$$

□

Now, let's assume that M is not irreducible. Since M is simply connected, we use the De Rham decomposition [3] (Proposition 5.5, p. 310) to write M as the product of irreducible symmetric spaces

$$M = M_1 \times \cdots \times M_r,$$

so that the metric R of M and the squared Riemannian distance d^2 can be expressed as

$$R_{\bar{x}}(u, u) = \sum_{i=1}^{n} R_{\bar{x}_i}(u_i, u_i), \tag{5.4}$$

and

$$d^2(\bar{x}, x) = \sum_{i=1}^{n} d_i^2(\bar{x}_i, x_i), \tag{5.5}$$

where $u = u_1 + u_2 + \cdots + u_n$, $x = (x_1, \cdots, x_n)$ and $\bar{x} = (\bar{x}_1, \cdots, \bar{x}_n)$ with $x_i \in M_i$, $\bar{x}_i \in M_i$ and $u_i \in T_{\bar{x}_i} M_i$. In this case, the Otto's metric of the Riemannian Gaussian model is given by the following proposition.

Proposition 5.2. *Let M be a simply connected Riemannian symmetric space of the compact type or the noncompact type and $\mathcal{P}$ a Gaussian Riemannian model defined on M. We assume that the action of L on M is transitive. The Otto's metric I of the model $\mathcal{P}$ is a multiply-warped Riemannian metric which is given by:*

$$I_z\,(U,U) = (\alpha\,(\eta)\,u_\eta)^2 + \sum_{i=1}^{n} \beta_i^2\,(\eta)\,R_{\bar{x}_i}(u_i,u_i) \quad \forall U \in T_z\mathcal{M},$$

where $U = u_\eta\,\partial_\eta + u_1 + u_2 + \cdots + u_n$, with $u_\eta \in \mathbb{R}$, $u_i \in T_{\bar{x}_i}M_i$ and where the functions α and β are given by:

$$\alpha^2\,(\eta) = 4\,\psi'\,(\eta) \;\; \text{and}\;\; \beta_i^2\,(\eta) = 4\,\frac{\eta^2}{n}\sum_{l,k}\mathbb{E}_{z_i}R_{kl}^3,$$

with $z_i = (\bar{x}_i,\eta) \in M_i \times (0,\infty)$.

To prove the proposition, we use the following lemma.

Lemma 5.3. *Let $\mathcal{P}$ be a location-scale model defined on a product manifold $M = M_1 \times M_2 \times \cdots \times M_n$ such that for any probability density function $p\,(\cdot|\bar{x},\eta) \in \mathcal{P}$:*

$$p\,(x|z) = \prod_{i=1}^{n} p_i\,(x_i|z_i) \quad \forall x = (x_1,\cdots,x_n) \in M_1 \times \cdots \times M_n, \tag{5.6}$$

where $z = (\bar{x},\eta)$, $z_i = (\bar{x}_i,\eta) \in M_i \times (0,\infty)$ and where $p_i\,(\cdot|z_i)$ is a probability density function with respect to the Riemannian density of M_i. Then,

$$d_z\ell^{(x)}\,U = u_\eta \sum_{k=1}^{n} \partial_\eta \ell_k^{(x_k)}(z_k) + d_{\bar{x}_i}\ell_i^{(x_i)}u_i \quad \forall U = u_\eta\,\partial_\eta + u_i \in T_{z_i}\Big(M_i \times (0,\infty)\Big),$$

where ℓ and ℓ_i are log-likehood functions given respectively by $\ell^{(x)}\,(z) = \log p\,(x|z)$ and $\ell_i^{(x_i)}\,(z_i) = \log p_i\,(x_i|z_i)$.

Proof. (of the lemma) From (5.6), we have:

$$\ell^{(x)}(z) = \sum_{k=1}^{n} \ell_k^{(x_k)}(z_k).$$

Taking the derivative at the point z in this equality and applying it to $U = u_\eta\,\partial_\eta + u_i$, we get:

$$\begin{aligned} d_z\ell^{(x)}\,U &= \sum_{k=1}^{n} d_{z_k}\ell_k^{(x_k)}\,U, \\ &= \sum_{k=1}^{n} d_{z_k}\ell_k^{(x_k)}\,(u_\eta\,\partial_\eta + u_i)\,, \\ &= u_\eta \sum_{k=1}^{n} \partial_\eta \ell_k^{(x_k)}(z_k) + \sum_{k=1}^{n} d_{z_k}\ell_k^{(x_k)}u_i. \end{aligned}$$

Then,

$$d_z \ell^{(x)}\, U = u_\eta \sum_{k=1}^{n} \partial_\eta \ell_k^{(x_k)}(z_k) + d_{\bar{x}_i} \ell_i^{(x_i)} u_i.$$

□

Let's now prove the proposition.

Proof. We have:

$$p\,(x|z) = \exp\, \left[\eta\, d^2\,(x,\, \bar{x}) - \psi\,(\eta)\right] \text{ for } x \in M.$$

where $z = (\bar{x}, \eta) \in \mathcal{M}$. Since M is a simply connected Riemannian symmetric space, we have from the decomposition of De Rham's Theorem [3]

$$M = M_1 \times \cdots \times M_r.$$

Then, we can write $d^2\,(x,\, \bar{x})$ in (5.11) as in (5.5) so that

$$\begin{aligned} p\,(x|z) &= \exp\left[\eta \sum_{i=1}^{n} d_i^2\,(\bar{x}_i, x_i) - \psi\,(\eta)\right], \\ &= \exp\left[\eta \sum_{i=1}^{n} d_i^2\,(\bar{x}_i, x_i) - \sum_{i=1}^{n} \psi_i\,(\eta)\right], \\ &= \prod_{i=1}^{n} \exp\, \left[\eta\, d_i^2\,(\bar{x}_i, x_i) - \psi_i\;(\eta)\right], \end{aligned}$$

where $\psi_i\,(\eta) = \mathbb{E}_z\left(d_i^2\,(\bar{x}_i, x_i)\right)$ and the second equality holds by the following equality:

$$\psi\,(\eta) = \mathbb{E}_z\left(d^2\,(\bar{x}, x)\right) = \mathbb{E}_z\left(\sum_{i=1}^{n} d_i^2\,(\bar{x}_i, x_i)\right) = \sum_{i=1}^{n} \psi_i(\eta).$$

Then,

$$p\,(x|z) = \prod_{i=1}^{n} p_i\,(x_i|z_i)\,.$$

where $z_i = (\bar{x}_i, \eta)$ and $p_i\,(\cdot|z_i)$ is the Riemannian Gaussian density with respect to the Riemannian density of M_i. Applying Lemma 5.3 to $\mathcal{P}$, we get:

$$d_z \ell^{(x)}\, U = u_\eta \sum_{k=1}^{n} \partial_\eta \ell_k^{(x_k)}(z_k) + d_{\bar{x}_i} \ell^{(x_i)} u_i \text{ for } U = u_\eta\, \partial_\eta + u_i. \tag{5.7}$$

Using the bilinearity and the symmetry of I_z, we have:

$$I_z\,(U, U) = u_\eta^2 I_z\,(\partial_\eta, \partial_\eta) + 2u_\eta \sum_{i=1}^{n} I_z\,(\partial_\eta, u_i) + \sum_{i=1}^{n}\sum_{j=1}^{n} I_z\,(u_i, u_j)\,, \tag{5.8}$$

where $U = u_\eta \partial_\eta + u_1 + u_2 + \cdots + u_n$, with $u_\eta \in \mathbb{R}$, $u_i \in T_{\bar{x}_i} M$. For $i \in [\![1;n]\!]$,

$$\begin{aligned}
I_z(\partial_\eta, u_i) &= \mathbb{E}_z R\left(\mathrm{grad}_x\left(d_z \ell^{(\cdot)} \partial_\eta\right), \mathrm{grad}_x\left(d_z \ell^{(\cdot)} u_i\right)\right), \\
&= \mathbb{E}_z R\left(\mathrm{grad}_x \sum_{k=1}^{n} \partial_\eta \ell_k^{(\cdot)}(z_k), \mathrm{grad}_{x_i} d_{\bar{x}_i} \ell_i^{(\cdot)} u_i\right) \quad \text{by (5.7)}, \\
&= \mathbb{E}_{z_i} R\left(\mathrm{grad}_{x_i} \partial_\eta \ell_i^{(\cdot)}(z_i), \mathrm{grad}_{x_i} d_{\bar{x}_i} \ell_i^{(\cdot)} u_i\right) \quad \text{by (5.4)}, \\
&= I_{z_i}(\partial_\eta, u_i).
\end{aligned}$$

From the identity of polarisation, we have:

$$I_{z_i}(\partial_\eta, u_i) = \frac{1}{4}\left[I_{z_i}(u_i + \partial_\eta, u_i + \partial_\eta) - I_{z_i}(u_i - \partial_\eta, u_i - \partial_\eta)\right].$$

Then

$$I_z(\partial_\eta, u_i) = \frac{1}{4}\left[I_{z_i}(u_i + \partial_\eta, u_i + \partial_\eta) - I_{z_i}(u_i - \partial_\eta, u_i - \partial_\eta)\right]. \tag{5.9}$$

Since each symmetric Riemannian space M_i is irreducible, Theorem 3.3 can be applied to the Riemannian Gaussian model $\{p_i(\cdot|z_i)\}$

$$I_{z_i}(u_\eta \partial_\eta + u_i, u_\eta \partial_\eta + u_i) = (\alpha_i(\eta) u_\eta)^2 + \beta_i^2(\eta) R_{\bar{x}_i}(u_i, u_i), \tag{5.10}$$

where $u_i \in T_{\bar{x}_i} M_i$, $\alpha_i^2(\eta) = 4\,\psi_i(\eta)$ and $\beta_i^2(\eta) = 4\,\dfrac{\eta^2}{n} \sum_{l,k} \mathbb{E}_{z_i} R_{kl}^3$. Using this equality, (5.9) becomes

$$I_z(\partial_\eta, u_i) = \alpha_i^2(\eta) + \beta_i^2(\eta) R_{\bar{x}_i}(u_i, u_i) - \left[\alpha_i^2(\eta) + \beta_i^2(\eta) R_{\bar{x}_i}(u_i, u_i)\right].$$

Then,

$$I_z(\partial_\eta, u_i) = 0. \tag{5.11}$$

For $(i,j) \in ([\![1;n]\!])^2$ such that $i \neq j$, we have:

$$\begin{aligned}
I_z(u_i, u_j) &= \mathbb{E}_z R\left(\mathrm{grad}_x d_z \ell^{(\cdot)} u_i, \mathrm{grad}_x d_z \ell^{(\cdot)} u_j\right), \\
&= \mathbb{E}_z R\left(\mathrm{grad}_{x_i} d_{z_i} \ell_i^{(\cdot)} u_i, \mathrm{grad}_{x_j} d_{z_j} \ell_j^{(\cdot)} u_j\right) \text{ by (5.4)}, \\
&= 0 \text{ by (2.1)}.
\end{aligned}$$

Thus,

$$I_z(u_i, u_j) = 0. \tag{5.12}$$

Using (5.11) and (5.12), (5.8) becomes

$$\begin{aligned}
I_z(U, U) &= u_\eta^2 I_z(\partial_\eta, \partial_\eta) + \sum_{i=1}^{n} I_{z_i}(u_i, u_i), \\
&= (\alpha(\eta) u_\eta)^2 + \sum_{i=1}^{n} I_{z_i}(u_i, u_i).
\end{aligned}$$

Then, by (5.10):

$$I_z\left(U,U\right)=\left(\alpha\left(\eta\right)u_\eta\right)^2+\sum_{i=1}^{n}\beta_i^2\left(\eta\right)R_{\bar{x}_i}\left(u_i,u_i\right).$$

$\square$

Acknowledgements. This work was supported by CEA-SMIA.

References

[1] R. L. Bishop, B. O'Neill, *Manifolds of negative curvature*, Transactions of the American Mathematical Society, 145 (1969), 1-49.

[2] F. Gianfelici, V. Battistelli, *Methods of Information geometry (Amari, s. and Nagaoka, h.;2000)[book review]*, IEEE Transactions on Information Theory 55 (6) (2009), 2905-2906.

[3] S. Helgason, *Differential Geometry and Symmetric Spaces*, American Mathematical Society 2001.

[4] M. L. John, *Introduction to smooth manifolds* Springer 2012.

[5] J. Jost, J. Jost, *Riemannian geometry and geometric analysis*, Springer 2008.

[6] J. Lott, *Some geometric calculations on Wasserstein space*, Communications in Mathematical Physics, 227 (2008) 423-437.

[7] C. Ogouyandjou, N. Wadagni, *Wasserstein Riemannian geometry on statistical manifold*, International Electronic Journal of Geometry 13 (2) (2020), 144-151.

[8] B. O'Neill, *Semi-Riemannian geometry with application to relativity*, Academic Press 1983.

[9] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*. Comm. Partial Differential Equations 26 (2001), 101-174.

[10] S. Said, L. Bombrun, Y. Berthoumieu, *Warped Riemannian metrics for location-scale models*, Geometric Structures of Information, Springer (2019), 251-296.

[11] G. N. Watson, *A treatise on the theory of Bessel functions*, Cambridge University Press 1995.

Authors' addresses:

Kabirou Lawalé Chitou, Amour Gbaguidi Amoussou and Carlos Ogouyandjou
Institut de Mathématiques et de Sciences Physiques(IMSP),
Université d'Abomey-Calavi(UAC), Bénin,
E-mail: kabirou.chitou@imsp-uac.org
amour.gbaguidi@imsp-uac.org
ogouyandjou@imsp-uac.org

Freedath Djibril Moussa
Faculté des Sciences et Techniques (FAST),
Université d'Abomey-Calavi(UAC), Bénin,
E-mail: freedath.djibril@imsp-uac.org