

Article

Induced Dualistic Geometry of Finitely Parametrized Probability Densities on Manifolds

Simon Fong^{1,†,*} and Peter Tiño^{1,†}¹ School of Computer Science, The University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

* Correspondence: r.s.fong@cs.bham.ac.uk; Tel.: +44-07-482-458-248

† These authors contributed equally to this work.

‡ Preliminary work of this paper was presented in TGS1 2017 held in CIRM, Marseille from 28th August 2017 to 1st September 2017. Poster presented was titled: Induced Dualistic Structure and Probability Densities on Riemannian Manifolds

Abstract: This paper aims to describe the geometrical structure and explicit expressions of family of finitely parametrized probability densities over smooth manifold M . The geometry of family of probability densities on M are inherited from probability densities on Euclidean spaces $\{U_\alpha\}$ via bundle morphisms, induced by an orientation-preserving diffeomorphisms $\rho_\alpha : U_\alpha \rightarrow M$. Current literature inherits densities on M from tangent spaces via Riemannian exponential map $\exp : T_x M \rightarrow M$; densities on M are defined locally on region where the exponential map is a diffeomorphism. We generalize this approach with an arbitrary orientation-preserving bundle morphism; we show that the dualistic geometry of family of densities on U_α can be inherited to family of densities on M . Furthermore, we provide explicit expressions for parametrized probability densities on $\rho_\alpha(U_\alpha) \subset M$. Finally, using the component densities on $\rho_\alpha(U_\alpha)$, we construct parametrized mixture densities on totally bounded subsets of M . We provide a description of inherited mixture product dualistic geometry of the family of mixture densities.

Keywords: Probability densities on manifold, geometric statistics, Hessian manifold

1. Introduction

Statistics and data analysis on manifolds is of interest in fields such as image processing, shape analysis, machine learning, and natural computation. For instance, the notion of probability densities on manifold of motions [1] arises naturally; yet a description of geometrical properties and the explicit expressions of families of probability densities over a such general manifold is lacking.

Striving for explicit expressions of probability density functions over manifolds goes back to the field of directional statistics on spheres [2], and recently [3] on the space of symmetric positive definite matrices. Both studies employ the notion of Riemannian exponential map described by Pennec in [4].

However, finding the exact expression of Riemannian exponential map for general Riemannian manifolds can in practice be computationally expensive, as it involves solving the geodesic equation, which is a second-order differential equation. Therefore, in this work we aim to extend and generalize the approach by [4] in two ways:

1. Inherit the dualistic geometry of family of parametrized probability densities over Euclidean spaces to probability densities over manifolds via orientation-preserving bundle morphism. This generalizes the Riemannian exponential map
2. Construct and study the product dualistic geometry of families of parametrized mixture densities on totally bounded subsets of M . This generalizes the extent of region where family of finitely parametrized probability densities and their corresponding geometrical structure can be described on M .

The discussion of this paper revolves around the following bundle morphism induced by a local orientation-preserving diffeomorphism $\rho : U \subset \mathbb{R}^n \rightarrow M$.

$$\begin{array}{ccc}
 S \subset \text{Prob}(U) & \xrightarrow{\rho^{-1*}} & \tilde{S} \subset \text{Prob}(M) \\
 \downarrow & & \downarrow \\
 U \subset \mathbb{R}^n & \xleftrightarrow[\rho]{\rho^{-1}} & \rho(U) \subset M
 \end{array}$$

In this work we will discuss how the geometry of S , a family of probability densities over Euclidean space $U \subset \mathbb{R}^n$, is inherited to \tilde{S} via pulled-back bundle morphism defined by ρ^{-1} .

The paper is organized as follows: In the second section we discuss and establish the domain of discourse. We provide a brief summary of the natural identification of probability densities over manifolds both as functions and as volume forms described in the existing literature. This allows us to use density and volume form interchangeably in further discussion for simplicity. This describes the vertical parts of the bundle morphism shown above.

In the third section we extend the notion of naturality of Riemannian structure to dualistic structure [5] of Hessian-Riemannian manifolds. We study how dualistic structure can be inherited from one manifold to another via a diffeomorphism. We show how the pulled-back dualistic structure can be determined explicitly via the pulled-back local coordinates. As it is well known that statistical manifolds can be endowed with dualistic geometry. This will be used throughout the rest of the paper to naturally inherit family of probability densities over Euclidean spaces to probability densities over manifolds. This describes the top horizontal part of the bundle morphism shown above.

In the fourth section we consider the entire bundle morphism and present a construction of probability densities on manifolds that generalizes the current approach [4] as discussed above. We first show how family of parametrized probability densities \tilde{S} on M is inherited from open subset $U \subset \mathbb{R}^n$ via the bundle morphism, while preserving the dualistic geometrical structure of S on U . This allows us to derive an exact expression of probability density functions on M as a pulled-back densities. Moreover, the induced family of probability densities \tilde{S} on M inherits the dualistic geometry of S on U , which in turn allows us to inherit useful properties such as metric and divergence. This local construction is then extended to construct parametrized mixture densities over totally bounded subsets of M . We show that, under rather mild natural conditions, the family of mixture densities is a product manifold with the corresponding dualistic product geometry (generated by locally induced family of probability densities).

Finally, in section five we provide a simple example of inducing family of mixture Gaussian density functions over the unit 2-sphere \mathbb{S}^2 .

2. Setting

Let M be connected n -dimensional smooth Riemannian manifold with finite Riemann volume.

Let $\text{Vol}(M)$ denote the line bundle of smooth densities over M , and $\mathcal{E}(\text{Vol}(M))$ denote the smooth sections of $\text{Vol}(M)$. Let $\text{Dens}_+(M)$ denote the subspace of positive densities over M (reader is kindly referred to relevant publications [6,7] for details on volume bundle):

$$\text{Dens}_+(M) = \{\mu \in \mathcal{E}(\text{Vol}(M)) \mid \mu(x) > 0 \forall x \in M\} \subset \text{Vol}(M) \quad .$$

We consider the subspace of positive densities that integrate to 1 over M : $\text{Prob}(M) = \{\mu \in \text{Dens}_+(M) \mid \int_M \mu = 1\} \subset \text{Dens}_+(M)$.

Let $\mu_0 := dV_g$ denote the Riemannian volume form, be the reference measure on M . The space of probability density function over M denoted by $P(M) \subset L^1(M, \mu_0) \subset C^\infty(M, \mathbb{R}_+)$ can be associated naturally to $\text{Prob}(M) \subset \text{Dens}_+(M)$ volume forms in via Hodge star operator [8], or more notably in recent work, a diffeomorphism [7,9] given by the map:

$$\Phi : \text{Dens}_+(M) \rightarrow C^\infty(M, \mathbb{R}_+)$$

$$\mu \mapsto \sqrt{\frac{\mu}{\mu_0}} = \sqrt{\text{Jac}_{\mu_0}\mu} \quad .$$

For the rest of the paper we will associate the space of probability densities as a function space and space of densities naturally with respect to the Riemannian volume form dV_g .

In this work we will also restrict our attention to finitely parametrized families of probability distributions. Family of finitely parametrized probability distributions, equipped with Fisher information metric, has the structure of a Riemannian manifold also known as a statistical manifold [5].

That is, the statistical manifold of the family of probability distributions will be finite dimensional for the rest of the paper.

3. Preliminary: Induced dualistic structure

Consider a statistical manifold S with its corresponding metric g together with a pair of g -conjugate connections (∇, ∇^*) . The triplet (g, ∇, ∇^*) , known as the dualistic structure [5], is fundamental to the study of the intrinsic geometry of statistical manifolds [10]. The triplet (g, ∇, ∇^*) satisfies the following:

$$X\langle Y, Z \rangle_g = \langle \nabla_X Y, Z \rangle_g + \langle Y, \nabla_Y Z \rangle_g \quad , \quad \forall X, Y, Z \in \mathcal{E}(TS) \quad .$$

In this section we show how dualistic structure can be inherited to an arbitrary smooth manifold from a given manifold with dualistic structure naturally via a diffeomorphism.

We discuss two different ways of pulling back (dually flat) dualistic structures given a diffeomorphism from one manifold to another. We first show that general dualistic structures can be pulled back directly via diffeomorphism. We then show when the manifolds are dually flat, the induced dualistic structure can be computed implicitly via the pulled-back coordinates and metric.

Whilst the first method arises more naturally in a theoretical setting, the second provides a more computable way to describe an such an induced dualistic structure that is equivalent to the first when the manifolds are dually flat.

It is worth noting that, the intrinsic geometry of a statistical manifolds (S, g, ∇, ∇^*) can be alternatively be associated to a pair (g, T) , where T denotes Amari-Chenstov tensor $T(X, Y, Z) = \langle \nabla_X^* Y - \nabla_X Y, Z \rangle_g$.

In a recent paper, Ay et al. [11] showed how (g, T) can be inherited via sufficient statistics. In our case since diffeomorphisms are injective therefore is a sufficient statistics, our discussion is a special case of what they showed.

In this section we provide an alternative proof for the case when S is finite dimensional, and when the map between statistical manifolds is a diffeomorphism. In particular we show the relation between induced dualistic structure, Hessian structure, and local coordinate systems on finite dimensional statistical manifolds.

3.1. Naturality of dualistic structure

Suppose S is a finite dimensional manifold equipped with torsion-free dually flat dualistic structure (g, ∇, ∇^*) , then we can induce via a diffeomorphism a (dually flat) dualistic structure onto another manifold \tilde{S} .

Theorem 1. *If $\varphi : \tilde{S} \rightarrow S$ is a diffeomorphism between smooth manifolds, and S is equipped with torsion-free dualistic structure (g, ∇, ∇^*) , then \tilde{S} is a Riemannian manifold with induced dualistic structure $(\varphi^*g, \varphi^*\nabla, \varphi^*\nabla^*)$.*

Proof. Let \tilde{S} be a smooth n-dimensional manifold, and let S be a smooth n-dimensional manifold with torsion-free dualistic structure (g, ∇, ∇^*) then the following condition is satisfied:

$$X\langle Y, Z \rangle_g = \langle \nabla_X Y, Z \rangle_g + \langle Y, \nabla_Y^* Z \rangle_g \quad , \quad \forall X, Y, Z \in \mathcal{E}(TS) \quad .$$

Let $\varphi : \tilde{S} \rightarrow S$ be a diffeomorphism. Then the pullback of g along φ given by $\tilde{g} = \varphi^*g$ defines a Riemannian metric on \tilde{S} .

Consider pullback of ∇ via φ is given by:

$$\begin{aligned} \varphi^*\nabla : \mathcal{E}(T\tilde{S}) \times \mathcal{E}(T\tilde{S}) &\rightarrow \mathcal{E}(T\tilde{S}) \\ (\varphi^*\nabla)(\tilde{X}, \tilde{Y}) &= \varphi_*^{-1}\nabla(\varphi_*\tilde{X}, \varphi_*\tilde{Y}) = \varphi_*^{-1}\nabla_{\varphi_*\tilde{X}}\varphi_*\tilde{Y} \quad , \end{aligned}$$

where $\mathcal{E}(T\tilde{S})$ denote the set of smooth sections of tangent bundle over \tilde{S} , φ_* is the push-forward of φ , and φ^* is the pullback of φ . Since pullback of torsion-free connection by diffeomorphism is a torsion-free connection. Therefore $\varphi^*\nabla$ and $\varphi^*\nabla^*$ are torsion-free connections on the tangent bundle over \tilde{S} : $\pi_{\sim} : T\tilde{S} \rightarrow \tilde{S}$.

Therefore it remains to show that $(\tilde{\nabla}, \tilde{\nabla}^*) := (\varphi^*\nabla, \varphi^*\nabla^*)$ is \tilde{g} -conjugate pair of connections on \tilde{S} . i.e. they satisfy the following equation:

$$\tilde{X}\langle \tilde{Y}, \tilde{Z} \rangle_{\tilde{g}} = \langle \tilde{\nabla}_{\tilde{X}}\tilde{Y}, \tilde{Z} \rangle_{\tilde{g}} + \langle \tilde{Y}, \tilde{\nabla}_{\tilde{X}}^*\tilde{Z} \rangle_{\tilde{g}} \quad , \quad \forall \tilde{X}, \tilde{Y}, \tilde{Z} \in \mathcal{E}(T\tilde{S}) \quad .$$

Let $\tilde{X}, \tilde{Y}, \tilde{Z} \in \mathcal{E}(T\tilde{S})$, and let $\tilde{p} \in \tilde{S}$ be arbitrary.

$$\begin{aligned} \langle \tilde{\nabla}_{\tilde{X}}\tilde{Y}, \tilde{Z} \rangle_{\tilde{g}}(\tilde{p}) &= \langle (\varphi^*\nabla)_{\tilde{X}}\tilde{Y}, \tilde{Z} \rangle_{\tilde{g}}(\tilde{p}) = \langle \varphi_*^{-1}\nabla_{\varphi_*\tilde{X}}\varphi_*\tilde{Y}_{\tilde{p}}, \tilde{Z}_{\tilde{p}} \rangle_{\tilde{g}} \\ &= \langle \varphi_*^{-1} \left(\nabla_{\varphi_*\tilde{X}}\varphi_*\tilde{Y} \right)_{\varphi(\tilde{p})}, \varphi_*^{-1} \left(\varphi_*\tilde{Z} \right)_{\varphi(\tilde{p})} \rangle_{\tilde{g}} \\ &= \left(\varphi_*^{-1}\tilde{g} \right)_{\varphi(\tilde{p})} \left(\nabla_{\varphi_*\tilde{X}}\varphi_*\tilde{Y}_{\varphi(\tilde{p})}, \varphi_*\tilde{Z}_{\varphi(\tilde{p})} \right) \\ &= \langle \nabla_{\varphi_*\tilde{X}}\varphi_*\tilde{Y}_{\varphi(\tilde{p})}, \varphi_*\tilde{Z}_{\varphi(\tilde{p})} \rangle_g = \langle \nabla_{\varphi_*\tilde{X}}\varphi_*\tilde{Y}, \varphi_*\tilde{Z} \rangle_g(\varphi(\tilde{p})) \quad . \end{aligned}$$

Similarly, we have:

$$\langle \tilde{Y}, \tilde{\nabla}_{\tilde{X}}^*\tilde{Z} \rangle_{\tilde{g}}(\tilde{p}) = \langle \varphi_*\tilde{Y}, \nabla_{\varphi_*\tilde{X}}\varphi_*\tilde{Z} \rangle_g \quad .(\varphi(\tilde{p}))$$

Since (∇, ∇^*) is g -conjugate pair of connection on S , we have for each $\tilde{p} \in \tilde{S}$:

$$\langle \nabla_{\varphi_*\tilde{X}}\varphi_*\tilde{Y}_{\varphi(\tilde{p})}, \varphi_*\tilde{Z}_{\varphi(\tilde{p})} \rangle_g + \langle \varphi_*\tilde{Y}_{\varphi(\tilde{p})}, \nabla_{\varphi_*\tilde{X}}\varphi_*\tilde{Z}_{\varphi(\tilde{p})} \rangle_g = \varphi_*\tilde{X} \left(\langle \varphi_*\tilde{Y}, \varphi_*\tilde{Z} \rangle_g \right) (\varphi(\tilde{p})) \quad .$$

Hence:

$$\begin{aligned}
\varphi_* \tilde{X} (\langle \varphi_* \tilde{Y}, \varphi_* \tilde{Z} \rangle_g) (\varphi(\tilde{p})) &= \varphi_* \tilde{X} (\langle \varphi_* \tilde{Y}, \varphi_* \tilde{Z} \rangle_g \circ \varphi \circ \varphi^{-1}) (\varphi(\tilde{p})) \\
&= \varphi_* \tilde{X} (\langle (\varphi^* g)(\tilde{Y}, \tilde{Z}) \circ \varphi^{-1} \rangle) (\varphi(\tilde{p})) \\
&= \tilde{X} (\langle \tilde{Y}, \tilde{Z} \rangle_{\tilde{g}} \circ \varphi^{-1} \circ \varphi) \circ \varphi^{-1} (\varphi(\tilde{p})) \\
&= \tilde{X} \langle \tilde{Y}, \tilde{Z} \rangle_{\tilde{g}} (\tilde{p}) \quad .
\end{aligned}$$

Combining the above results we get:

$$\langle \tilde{\nabla}_{\tilde{X}} \tilde{Y}, \tilde{Z} \rangle_{\tilde{g}} (\tilde{p}) + \langle \tilde{Y}, \tilde{\nabla}_{\tilde{X}}^* \tilde{Z} \rangle_{\tilde{g}} (\tilde{p}) = \tilde{X} \langle \tilde{Y}, \tilde{Z} \rangle_{\tilde{g}} (\tilde{p}) \quad , \quad \forall \tilde{X}, \tilde{Y}, \tilde{Z} \in \mathcal{E}(T\tilde{S}) \quad .$$

Therefore \tilde{S} can be equipped with the induced torsion-free dualistic structure $(\tilde{g}, \tilde{\nabla}, \tilde{\nabla}^*) = (\varphi^* g, \varphi^* \nabla, \varphi^* \nabla^*)$ as desired. \square

Remark 1. This means the diffeomorphism $\varphi : \tilde{S} \rightarrow \tilde{S}$ is a local isometry.

This also implies if (∇, g) satisfies Codazzi's equation [12]:

$$Xg(Y, Z) - g(\nabla_X Y, Z) - g(Y, \nabla_X Z) = (\nabla_X g)(Y, Z) = (\nabla_Z g)(Y, X) \quad ,$$

then $(\tilde{\nabla}, \tilde{g}) := (\varphi^* \nabla, \varphi^* g)$ also satisfies Codazzi's equation. The derivation is as follows (cf. proof of Prop 4.2 [12]):

$$(\tilde{\nabla}_{\tilde{X}} \tilde{g})(\tilde{Y}, \tilde{Z}) = \tilde{X} \tilde{g}(\tilde{Y}, \tilde{Z}) - \tilde{g}(\nabla_{\tilde{X}} \tilde{Y}, \tilde{Z}) - \tilde{g}(\tilde{Y}, \nabla_{\tilde{X}} \tilde{Z}) \quad .$$

By duality of (∇, ∇^*) , the above equals to:

$$= \tilde{g}(\nabla_{\tilde{X}} \tilde{Y}, \tilde{Z}) + \tilde{g}(\tilde{Y}, \tilde{\nabla}_{\tilde{X}}^* \tilde{Z}) - \tilde{g}(\nabla_{\tilde{X}} \tilde{Y}, \tilde{Z}) - \tilde{g}(\tilde{Y}, \nabla_{\tilde{X}} \tilde{Z}) = \tilde{g}(\tilde{Y}, \tilde{\nabla}_{\tilde{X}}^* \tilde{Z}) - \tilde{g}(\tilde{Y}, \nabla_{\tilde{X}} \tilde{Z}) \quad .$$

By a similar argument, we also have:

$$(\tilde{\nabla}_{\tilde{Z}} \tilde{g})(\tilde{Y}, \tilde{X}) = \tilde{g}(\tilde{Y}, \tilde{\nabla}_{\tilde{Z}}^* \tilde{X}) - \tilde{g}(\tilde{Y}, \nabla_{\tilde{Z}} \tilde{X}) \quad .$$

Hence the following is satisfied:

$$\begin{aligned}
&(\tilde{\nabla}_{\tilde{X}} \tilde{g})(\tilde{Y}, \tilde{Z}) - (\tilde{\nabla}_{\tilde{Z}} \tilde{g})(\tilde{Y}, \tilde{X}) \\
&= \langle \tilde{Y}, \tilde{T}^*(\tilde{X}, \tilde{Z}) - \tilde{T}(\tilde{X}, \tilde{Z}) \rangle_{\tilde{g}} \quad ,
\end{aligned}$$

where \tilde{T}, \tilde{T}^* denote the torsion of $\tilde{\nabla}, \tilde{\nabla}^*$ respectively. Since $\tilde{\nabla}$ and $\tilde{\nabla}^*$ are torsion free, the expression is identically zero on all $\tilde{p} \in \tilde{S}$. Therefore $(\tilde{\nabla}, \tilde{g})$ satisfies Codazzi's equation. By Proposition 2.1 in [13], \tilde{g} is a Hessian metric with respect to $\tilde{\nabla}$, meaning there exists a potential function $\tilde{\psi}$ on \tilde{S} such that $\tilde{g} = \tilde{\nabla} \tilde{\nabla} \tilde{\psi}$ is a Hessian metric, and $(\tilde{S}, \tilde{g}, \tilde{\psi})$ is a Hessian manifold.

We can also determine the pull-back curvature on \tilde{S} with dualistic structure $(\varphi^* g, \varphi^* \nabla, \varphi^* \nabla^*)$ by the following immediate corollary:

Corollary 1. Let $\varphi : (\tilde{S}, \tilde{g}) \rightarrow (S, g)$ be a local isometric diffeomorphism. Suppose S has g -conjugate connections (∇, ∇^*) , and let $(\tilde{\nabla}, \tilde{\nabla}^*) := (\varphi^* \nabla, \varphi^* \nabla^*)$ be the induced \tilde{g} -conjugate connections on \tilde{S} , then $\varphi^* R = \tilde{R}$ and $\varphi^* R^* = \tilde{R}^*$. In particular if S is dually flat, then so is \tilde{S} .

Proof. Let $\varphi : (\tilde{S}, \tilde{g}, \tilde{\nabla}, \tilde{\nabla}^*) \rightarrow (S, g, \nabla, \nabla^*)$ be a local isometry, where $(\tilde{g}, \tilde{\nabla}, \tilde{\nabla}^*) = (\varphi^* g, \varphi^* \nabla, \varphi^* \nabla^*)$, then by the proof of theorem 1:

$$\tilde{\nabla}_{\tilde{X}} \tilde{Y} = (\varphi^* \nabla)_{\tilde{X}} \tilde{Y} = \varphi_*^{-1} (\nabla_{\varphi_* \tilde{X}} \varphi_* \tilde{Y}), \quad \forall \tilde{X}, \tilde{Y} \in \mathcal{E}(T\tilde{S}) .$$

Hence we have:

$$\begin{aligned} \varphi_* (\tilde{\nabla}_{\tilde{X}} \tilde{\nabla}_{\tilde{Y}} \tilde{Z}) &= \nabla_{\varphi_* \tilde{X}} \nabla_{\varphi_* \tilde{Y}} \varphi_* \tilde{Z} , \\ \varphi_* (\tilde{\nabla}_{[\tilde{X}, \tilde{Y}]} \tilde{Z}) &= \nabla_{\varphi_* [\tilde{X}, \tilde{Y}]} \varphi_* \tilde{Z} = \tilde{\nabla}_{[\varphi_* \tilde{X}, \varphi_* \tilde{Y}]} \varphi_* \tilde{Z} . \end{aligned} \quad (1)$$

Equation 1 is obtained by:

$$\begin{aligned} \varphi_* (\tilde{\nabla}_{\tilde{X}} \tilde{\nabla}_{\tilde{Y}} \tilde{Z}) &= \varphi_* \tilde{\nabla} (\tilde{X}, \tilde{\nabla}_{\tilde{Y}} \tilde{Z}) \\ &= \nabla (\varphi_* \tilde{X}, \varphi_* \tilde{\nabla}_{\tilde{Y}} \tilde{Z}) \\ &= \nabla (\varphi_* \tilde{X}, \nabla (\varphi_* \tilde{Y}, \varphi_* \tilde{Z})) \\ &= \nabla_{\varphi_* \tilde{X}} \nabla_{\varphi_* \tilde{Y}} \varphi_* \tilde{Z} . \end{aligned}$$

By definition the Riemannian curvature tensor on S is given by: $X, Y, Z \in TS$:

$$\langle R(X, Y)Z, W \rangle = \langle \nabla_X \nabla_Y Z, W \rangle - \langle \nabla_Y \nabla_X Z, W \rangle - \langle \nabla_{[X, Y]} Z, W \rangle .$$

Let $\tilde{X}, \tilde{Y}, \tilde{Z}, \tilde{W} \in T\tilde{S}$. The pullback curvature tensor $\varphi^* R$ is thus given by:

$$\begin{aligned} \langle (\varphi^* R)(\tilde{X}, \tilde{Y})\tilde{Z}, \tilde{W} \rangle &= \langle \tilde{\nabla}_{\varphi_* \tilde{X}} \tilde{\nabla}_{\varphi_* \tilde{Y}} \varphi_* \tilde{Z}, \tilde{W} \rangle \\ &\quad - \langle \tilde{\nabla}_{\varphi_* \tilde{Y}} \tilde{\nabla}_{\varphi_* \tilde{X}} \varphi_* \tilde{Z}, \tilde{W} \rangle - \langle \tilde{\nabla}_{[\varphi_* \tilde{X}, \varphi_* \tilde{Y}]} \varphi_* \tilde{Z}, \tilde{W} \rangle \\ &= \varphi_* (\nabla_X \nabla_Y Z) - \varphi_* (\nabla_Y \nabla_X Z) - \varphi_* (\nabla_{[X, Y]} Z) \\ &= \varphi_* (\langle R(X, Y)Z, W \rangle) . \end{aligned}$$

By symmetry, $\varphi^* R^*$ satisfies:

$$\langle (\varphi^* R^*)(\tilde{X}, \tilde{Y})\tilde{Z}, \tilde{W} \rangle = \varphi_* (\langle R^*(X, Y)Z, W \rangle) .$$

If S is dually flat, meaning $R = 0 = R^*$, we then have the equality:

$$R^* \equiv 0 \Leftrightarrow R \equiv 0 \Leftrightarrow \varphi^* R \equiv 0 \Leftrightarrow \varphi^* R^* \equiv 0 .$$

□

This result has been known for Levi-Civita connection on Riemannian manifolds. Here we generalize it slightly to pair of g -conjugate dual connections.

3.2. Computing induced dualistic structure

Finally we discuss how pulled-back dually flat dualistic structure $(\varphi^*g, \varphi^*\nabla, \varphi^*\nabla^*)$ can be determined explicitly via the Hessian dualistic structure generated by the pulled-back metric, local coordinates, and the corresponding induced potential function.

Definition 1. Given a smooth manifold M , a *divergence* [5] D or *contrast function* [14] on M is a smooth function $D : M \times M \rightarrow \mathbb{R}_+$ satisfying the following:

1. $D(p; q) \geq 0$, and
2. $D(p, q) = 0$ iff $p = q$.

A dualistic structure $(g^D, \nabla^D, \nabla^{D*})$ on S can be determined by divergence function D via the following equations [5,14,15] for each point $p \in S$:

$$\begin{aligned} g_{ij}^D \Big|_p &= g_p^D(\partial_i, \partial_j) := -\partial_i^1 \partial_j^2 D[p; q] \Big|_{q=p} \\ \Gamma_{ijk}^D \Big|_p &= \langle \nabla_{\partial_i}^D \partial_j, \partial_k \rangle \Big|_p := -\partial_i^1 \partial_j^1 \partial_k^2 D[p; q] \Big|_{q=p} \quad , \end{aligned}$$

where (θ_i) denote local coordinates on M with corresponding local coordinate frame (∂_i) about p . Let ∂_i^ℓ denote the i^{th} partial derivative on the ℓ^{th} argument of D . By an abuse of notation, we may write [5]:

$$\begin{aligned} D[\partial_i; \partial_j] &:= -\partial_i^1 \partial_j^2 D[p; q] \Big|_{q=p} \quad , \quad \text{and} \\ -D[\partial_i; \partial_j; \partial_k] &:= -\partial_i^1 \partial_j^1 \partial_k^2 D[p; q] \Big|_{q=p} \quad . \end{aligned} \quad (2)$$

Remark 2. Conversely, given a torsion-free dualistic structure and a local coordinate system, there exists a divergence that induces the dualistic structure [16]. We will refer to the divergence \tilde{D} on \tilde{S} corresponding to the pulled-back dualistic structure $(\varphi^*g, \varphi^*\nabla, \varphi^*\nabla^*)$ (not necessarily dually flat) as the *induced divergence* on \tilde{S} .

For the rest of the section we will assume both S and \tilde{S} are dually flat, and we show how the pulled-back dually flat dualistic structure can be determined explicitly.

Let $(\partial_i := \frac{\partial}{\partial \theta_i})$ denote the local coordinate frame for TS corresponding to local ∇ -affine coordinates (θ_i) , and let $\mathcal{D} \subset T\tilde{S}$ denote the tangent subspace spanned by vector fields $(\tilde{\partial}_i := \varphi_*^{-1} \partial_i)$. Since $[\varphi_*^{-1} \partial_i, \varphi_*^{-1} \partial_j] = \varphi_*^{-1} [\partial_i, \partial_j] = 0$, the vector fields $(\tilde{\partial}_i := \varphi_*^{-1} \partial_i)$ commute, hence the space \mathcal{D} spanned by $(\tilde{\partial}_i)$ is involutive. Therefore by theorem of Frobenius, \mathcal{D} is completely integrable, hence there exists $(\varphi^*\nabla)$ -affine coordinates $(\tilde{\theta}_i)$ on \tilde{S} with corresponding local frame $(\tilde{\partial}_i)$.

Remark 3. The pulled-back coordinates and the corresponding local coordinate frame described in this fashion does not depend on the dual flatness of S and \tilde{S} . Of course, when S and \tilde{S} are not dually flat, the coordinate systems may no longer be $\nabla, \tilde{\nabla}$ -affine.

Let $(\partial^i := \frac{\partial}{\partial \eta_i})$ denote local coordinate frame of TS corresponding to local ∇^* -affine coordinates (η_i) . We can define local $(\varphi^*\nabla^*)$ -affine coordinates $(\tilde{\eta}_i)$ on \tilde{S} with corresponding local coordinate frame $(\tilde{\partial}^i := \varphi_*^{-1} \partial^i)$.

It is immediate that $(\tilde{\theta}_i), (\tilde{\eta}_i)$ are φ^*g -dual coordinates:

$$\begin{aligned}
(\tilde{\partial}_i, \tilde{\partial}^j)_{\tilde{g}} &= (\varphi^*g) \left(\varphi_*^{-1}\partial_i, \varphi_*^{-1}\partial^j \right) \\
&= g \left(\varphi_*\varphi_*^{-1}\partial_i, \varphi_*\varphi_*^{-1}\partial^j \right) \\
&= g \left(\partial_i, \partial^j \right) = \delta_i^j \quad .
\end{aligned}$$

Moreover, if we consider smooth pulled back coordinates $(\bar{\theta}_i := \theta_i \circ \varphi)$ and the correspond local coordinate frame $(\bar{\partial}_i) := \left(\frac{\partial}{\partial \bar{\theta}_i} \right)$ on $T\tilde{S}$. Then

$$\begin{aligned}
\tilde{\partial}_i(\bar{\theta}_j) &= \varphi_*^{-1}\partial_i(\theta_j \circ \varphi) \\
&= \partial_i(\theta_j \circ \varphi \circ \varphi^{-1}) = \partial_i\theta_j = \delta_i^j \quad .
\end{aligned}$$

This implies for each i , there exists a constant c_i such that $\tilde{\theta}_i = \bar{\theta}_i + c_i$, this implies $\tilde{\partial}_i = \bar{\partial}_i$ for all i . Since \tilde{g} is a Hessian metric with respect to $\tilde{\nabla}$, there exists a potential function $\tilde{\psi}$ such that $\tilde{g} = \tilde{\nabla}\tilde{\nabla}\tilde{\psi}$ $\tilde{g}_{ij} = \tilde{\partial}_i\tilde{\partial}_j\tilde{\psi} = \bar{\partial}_i\bar{\partial}_j\tilde{\psi}$. The corresponding \tilde{g} -dual local coordinate system of \tilde{S} with respect to $(\bar{\theta}_i)$, denoted by $(\bar{\eta}_i)$ can be defined by $(\bar{\eta}_i) = \bar{\partial}_i\tilde{\psi}$ with correspond local coordinate frame $(\bar{\partial}^i)$ of $T\tilde{S}$ [5].

Now let's consider the Hessian Riemannian structure of dually flat manifolds [13] induced by the following divergence function:

$$\begin{aligned}
\bar{D} : \tilde{S} \times \tilde{S} &\rightarrow \mathbb{R}_+ \\
(\tilde{p}, \tilde{q}) &\mapsto \bar{D}(\tilde{p}, \tilde{q}) = \tilde{\psi}(\tilde{p}) + \tilde{\psi}^\dagger(\tilde{q}) - \langle \bar{\theta}(\tilde{p}), \bar{\eta}(\tilde{q}) \rangle \quad ,
\end{aligned}$$

where $\tilde{\psi}^\dagger$ is a smooth function on \tilde{S} representing the Legendre-Fréchet transformation of $\tilde{\psi}$ with respect to the pair of \tilde{g} -dual local coordinates $(\bar{\theta}_i), (\bar{\eta}_i)$ on \tilde{S} . Let \bar{g} denote the Hessian metric generated by \bar{D} : $\bar{g}_{ij}|_{\tilde{p}} := g_{ij}^{\bar{D}}|_{\tilde{p}}$. By definition $\bar{g}_{ij}|_{\tilde{p}} = -\bar{\partial}_i\bar{\partial}_j\bar{D} = g_{ij}^{\bar{D}}|_{\tilde{p}}$.

Let $(\nabla^{\bar{D}}, \nabla^{\bar{D}^*})$ denote pair of \tilde{g} -dual connections defined by \bar{D} . We now show $(\tilde{\nabla}, \tilde{\nabla}^*) = (\bar{\nabla}, \bar{\nabla}^*)$

Let $\tilde{X}, \tilde{Y}, \tilde{Z} \in \mathcal{E}(T\tilde{S})$ and $\tilde{p} \in \tilde{S}$, the following is satisfied by construction:

$$\begin{aligned}
\tilde{X}\langle \tilde{Y}, \tilde{Z} \rangle_{\tilde{g}} &= \langle \tilde{\nabla}_{\tilde{X}}\tilde{Y}, \tilde{Z} \rangle_{\tilde{g}} + \langle \tilde{Y}, \tilde{\nabla}_{\tilde{X}}^*\tilde{Z} \rangle_{\tilde{g}} \\
\tilde{X}\langle \tilde{Y}, \tilde{Z} \rangle_{\bar{g}} &= \langle \bar{\nabla}_{\tilde{X}}\tilde{Y}, \tilde{Z} \rangle_{\bar{g}} + \langle \tilde{Y}, \bar{\nabla}_{\tilde{X}}^*\tilde{Z} \rangle_{\bar{g}} \quad .
\end{aligned}$$

Since $\tilde{g} = \bar{g}$, the two equations are equal, hence if we let for $p \in \tilde{S}$ let $(\bar{\partial}_i = \tilde{\partial}_i)$ denote the local frame of $T_p\tilde{S}$, and let $\tilde{X} = \tilde{\partial}_i = \bar{\partial}_i$, $\tilde{Y} = \tilde{\partial}_j = \bar{\partial}_j$, and $\tilde{Z} = \tilde{\partial}_k = \bar{\partial}_k$ then:

$$\begin{aligned}
\langle \tilde{\nabla}_{\tilde{X}}\tilde{Y}, \tilde{Z} \rangle_{\tilde{g}} + \langle \tilde{Y}, \tilde{\nabla}_{\tilde{X}}^*\tilde{Z} \rangle_{\tilde{g}} &= \langle \bar{\nabla}_{\tilde{X}}\tilde{Y}, \tilde{Z} \rangle_{\bar{g}} + \langle \tilde{Y}, \bar{\nabla}_{\tilde{X}}^*\tilde{Z} \rangle_{\bar{g}} \\
&= \langle \tilde{\nabla}_{\tilde{\partial}_i}\tilde{\partial}_j, \tilde{\partial}_k \rangle_{\tilde{g}} + \langle \tilde{\partial}_j, \tilde{\nabla}_{\tilde{\partial}_i}^*\tilde{\partial}_k \rangle_{\tilde{g}} = \langle \bar{\nabla}_{\bar{\partial}_i}\bar{\partial}_j, \bar{\partial}_k \rangle_{\bar{g}} + \langle \bar{\partial}_j, \bar{\nabla}_{\bar{\partial}_i}^*\bar{\partial}_k \rangle_{\bar{g}} \\
&= 0 + \langle \tilde{\partial}_j, \tilde{\nabla}_{\tilde{\partial}_i}^*\tilde{\partial}_k \rangle_{\tilde{g}} = 0 + \langle \bar{\partial}_j, \bar{\nabla}_{\bar{\partial}_i}^*\bar{\partial}_k \rangle_{\bar{g}}
\end{aligned}$$

Since $\tilde{\partial}_i = \bar{\partial}_i$ for all i , and $\tilde{g} = \bar{g}$, we have for all $\tilde{p} \in \tilde{S}$:

$$\langle \tilde{\partial}_j, \tilde{\nabla}_{\tilde{\partial}_i}^* \tilde{\partial}_k \rangle_{\tilde{g}} = \langle \tilde{\partial}_j, \overline{\nabla}_{\tilde{\partial}_i}^* \tilde{\partial}_k \rangle_{\tilde{g}} \quad , \quad \forall i, j, k \quad .$$

Therefore $\tilde{\nabla}^* = \overline{\nabla}^*$, and hence by symmetry $(\tilde{\nabla}, \tilde{\nabla}^*) = (\overline{\nabla}, \overline{\nabla}^*)$.

Furthermore, we can determine the explicit expression of the Christoffel symbols of the induced connection $\overline{\nabla}^* = \tilde{\nabla}^* = \varphi^* \nabla^*$ at $\tilde{p} \in \tilde{S}$ in pulled-back coordinates $\tilde{\theta}_i = \theta_i \circ \varphi$ as follows. Let $p \in S$ be an arbitrary, and let $\tilde{p} := \tilde{\varphi}(p) = \varphi^{-1}(p) \in \tilde{S}$, then:

$$\begin{aligned} \tilde{\Gamma}_{ijk}^* \Big|_{\tilde{p}} &= \tilde{\partial}_i \tilde{\partial}_j \tilde{\partial}_k \tilde{\psi}(\tilde{p}) = \tilde{\partial}_i \tilde{g}_{jk}(\tilde{p}) \\ &= \tilde{\partial}_i \tilde{g}_{jk}(\tilde{p}) = \left(\varphi_*^{-1} \partial_i \right) \tilde{g}_{jk}(\tilde{p}) \\ &= \partial_i \left(g_{jk} \circ \varphi \circ \varphi^{-1} \right) (\varphi(\tilde{p})) = \Gamma_{ijk}^* \Big|_p \quad . \end{aligned}$$

4. Probability densities on manifold

In this section we construct family of probability densities over smooth manifolds via orientation-preserving bundle morphism. This extend and generalize the construction of probability distributions on geodesically complete Riemannian manifolds via Riemannian exponential map described in previous literature [4], summarized as follows:

Given an arbitrary point x on Riemannian manifold M , the geodesic $\gamma_v : [0, I_v) \subset \mathbb{R} \rightarrow M$ with initial point p and initial velocity $v \in T_x M$ is uniquely determined on an interval $[0, I_v)$. This allows us to define the Riemannian exponential map at a given point x , which maps the tangent space $T_x M$ to M by tracing along the geodesic γ_v starting at x , determined by initial velocity $v \in T_x M$ for time 1. To be precise, we have the following definition [17]:

Definition 2. Given a point $x \in M$, consider the subset O_x of $T_x M$ given by:

$$O_p := \{v \in T_x M \mid \gamma_v \text{ is defined on } [0, I_v), \text{ and } I_v > 1\} \quad ,$$

then the **exponential map** at x is the map:

$$\begin{aligned} \exp_x : O_x &\rightarrow M \\ \exp_x(v) &:= \gamma_v(1) \quad . \end{aligned}$$

For each $x \in M$, there exists a local neighbourhood W_x in $T_x M$ where the Riemannian exponential map is a diffeomorphism. The the image of Riemannian exponential map can be viewed as a generalization of "going from p with direction v in the shortest path for time 1" within this local neighbourhood. It is also well known that if M is geodesically complete, then the Riemannian exponential map is defined on the entire tangent space $T_x M$ for any $x \in M$.

In previous literature [4], probability densities and the corresponding statistical properties on geodesically complete Riemannian manifolds are constructed by inheriting probability densities on the tangent space via the Riemannian exponential map.

In particular, let $W_x \subset T_x M$ denote the region where the exponential map is diffeomorphism. For each $y \in \exp_x(W_x)$ there is a unique $v \in W_x$ such that $\exp_x(v) = y$.

Given a probability density function p on $T_x M$, a probability density function \tilde{p} on M whose function value on $y = \exp_x(v) \in \exp_x(W_x)$ can be constructed by:

$$p(v) = \tilde{p}(\exp_x(v)) \quad ,$$

or equivalently, since \exp_x is a diffeomorphism,

$$\tilde{p}(y) = p(\log_x(y)) \quad , \quad (3)$$

where \log_x is the inverse of the Riemannian exponential map on W_x .

However, finding the exact expression for Riemannian exponential map for general Riemannian manifolds maybe computationally expensive in practice, as it involves solving the geodesic equation, which is a second order differential equation. Therefore in this work we are aim to find the explicit expression of parametrized probability distributions on manifold with a more general map.

We extend and generalize the above construction in two ways:

We first discuss family of probability densities on M inherited locally from open subsets of \mathbb{R}^n via an orientation-preserving bundle diffeomorphism. We discuss how the pulled-back family of probability densities on M inherit the geometrical properties from the family of probability densities on \mathbb{R}^n , and show that it generalities the above construction with the Riemannian exponential map.

We then extend this to construct probability distributions on M supported beyond the region where the map ρ is a diffeomorphism. In particular, we show that parametrized family of probability densities on inherited in this fashion can be extended to any totally bounded subset V of M . We first consider an orientation-preserving open cover of V , where each element of the open cover is equipped with a local family of inherited probability distributions. A family of parametrized mixture densities \mathcal{L}_V on the entire V can thus be constructed by gluing the locally inherited densities. Finally we discuss the geometrical properties of elements of \mathcal{L}_V , and show that it is a product manifold of locally inherited family under two conditions.

We shall begin our discussion by first considering the case when M is Riemannian manifold. In the last subsection we generalize it to arbitrary smooth topological manifold.

4.1. Local densities on M via local bundle morphism

Let M be a smooth topological manifold, let U be an open subset of \mathbb{R}^n . Suppose there exists orientation-preserving diffeomorphism $\rho : U \subset \mathbb{R}^n \rightarrow M$. Here we consider locally inherited family of probability densities on M via orientation-preserving diffeomorphism ρ . In particular we construct local parametrized families of probability densities over $\rho(U) \subset M$ as a subspace of the density bundle $\text{Vol}(M)$ via the bundle morphism induced by ρ .

Let $\hat{S} := \{p_\theta | \theta \in \Xi \subset \mathbb{R}^m\} \subset \mathcal{P}(U)$ denote a finitely parametrized family of probability density functions over $U \subset \mathbb{R}^n$. We further assume elements of \hat{S} are mutually absolute continuous, then \hat{S} has the structure of a statistical manifold [5].

Let μ_0 be an arbitrary reference measure on U , and let $S := \{v_\theta | \theta \in \Xi\}$ denote the set of volume densities over U naturally associated to \hat{S} with respect to μ_0 (as discussed in section 2).

Locally defined family of probability densities over $\rho(U) \subset M$ can thus be constructed via the pullback bundle morphism defined by orientation-preserving diffeomorphism¹ $\rho^{-1} : M \rightarrow U$:

¹ Since ρ is an orientation-preserving diffeomorphism, then so is ρ^{-1}

$$\begin{array}{ccc}
 S \subset \text{Prob}(U) & \xrightarrow{\rho^{-1*}} & \tilde{S} \subset \text{Prob}(M) \\
 \downarrow & & \downarrow \\
 U & \xrightleftharpoons[\rho]{\rho^{-1}} & \rho(U) \subset M
 \end{array}$$

where $\tilde{S} \subset \text{Prob}(M)$ is a family of probability densities over M given by:

$$\tilde{S} := \rho^{-1*} S = \left\{ \tilde{\nu}_\theta = \rho^{-1*} \nu_\theta \right\} .$$

More precisely, let $x \in V := \rho(U) \subseteq M$, and let $X_1, \dots, X_n \in T_x M$ be arbitrary vectors. Given a density $\nu_\theta \in S \subset \text{Prob}(U)$, the pulled-back density $\tilde{\nu}_\theta$ on $V \subset M$ is given by:

$$\tilde{\nu}_\theta := \rho_\alpha^{-1*} \nu_\theta(X_1, \dots, X_n) \Big|_y = \nu_\theta(\rho_\alpha^{-1} X_1, \dots, \rho_\alpha^{-1} X_n) \Big|_{\rho_\alpha^{-1}(y)} , \quad \forall y \in V_\alpha \subseteq M. \quad (4)$$

Since ρ is an orientation-preserving diffeomorphism, so is ρ^{-1} , hence we have the following equality:

$$1 = \int_U \nu_\theta = \int_{V:=\rho(U)} \rho^{-1*} \nu_\theta .$$

Suppose ν_θ has probability density function p_θ with respect to the reference measure μ_0 on U , i.e. $\nu_\theta = p_\theta d\mu_0$, then in local coordinates (x^1, \dots, x^n) of M , the above integral has the following form:

$$1 = \int_U p_\theta d\mu_0 = \int_V (p_\theta \circ \rho^{-1}) (\det D\rho^{-1}) dx^1 \wedge \dots \wedge dx^n .$$

Next we show the diagram commutes: since ρ is a local diffeomorphism, for each $v \in U$, $\exists! y \in \rho(U) \subset M$ such that $y = \rho(v)$, and (v, p_θ) is a section in the line bundle $\pi_U : \text{Prob}(U) \rightarrow U$. We have the following equalities:

$$\begin{aligned}
 \rho \circ \pi_U(v, p_\theta) &= \rho(v) = y , \\
 \pi_M \circ \rho^{-1*}(v, p_\theta) &= \pi_M(\rho^{-1}(v), \rho^{-1*} p_\theta) = \rho^{-1}(v) = \rho(v) = y .
 \end{aligned}$$

Finally, suppose S has dualistic structure given by (g, ∇, ∇^*) . Since ρ^{-1} is a diffeomorphism, so is ρ^{-1*} . Therefore by the discussion in Section 3, \tilde{S} has inherited dualistic structure $(\varphi^* g, \varphi^* \nabla, \varphi^* \nabla^*)$, where $\varphi = \rho^{-1*}$. In particular, the induced family of probability distributions inherits the geometrical structure via the bundle morphism as well.

Remark 4. Note that both the local coordinate map $p_\theta \in S \mapsto \theta \in \mathbb{R}^\ell$ and $\varphi := \rho^{-1*}$ are diffeomorphisms. For the rest of the paper we will, without loss of generality, assume \tilde{S} to be parametrized by (θ_i) instead of the pulled-back local coordinates $(\theta_i \circ \rho^{-1*})$ unless specified otherwise (described in Section 3.2 and Remark 3) if the map ρ^{-1*} is clear.

Example 1. Suppose $(S, g, \nabla^{(\alpha)}, \nabla^{(-\alpha)}) \subset \text{Prob}(U)$ is an α -affine statistical manifold for some $\alpha \in \mathbb{R}$, with Fisher metric g , the associated g -dual α -connections $(\nabla^{(\alpha)}, \nabla^{(-\alpha)})$, and the corresponding α -divergence D_α on S [5]. Since $\rho : U \rightarrow M$ is a (local) diffeomorphism, it is injective, hence a sufficient statistic for S [5].

By the invariance of Fisher metric and α -connection under sufficient statistic, the induced family \tilde{S} is also an α -affine statistical manifold.

Furthermore, due to the monotonicity of α -divergence (as a special case of f -divergence), the induced divergence \tilde{D}_α (see remark 2) on \tilde{S} can be computed by:

$$\tilde{D}_\alpha(\rho^{-1*} p, \rho^{-1*} q) = D_\alpha(p, q) \quad , \quad \text{for } p, q \in S \quad .$$

4.1.1. Special case: Riemannian exponential map

We now illustrate the framework outlined above on a special case of locally inherited family of probability densities discussed in [4] via Riemann exponential map:

Example 2. Let M be a complete Riemannian manifold. For each $x \in M$ [17]:

$$\begin{aligned} \text{seg}(x) &:= \{v \in T_x M \mid \exp_x(tv) \text{ is a minimizing unit speed curve, } t \in [0, 1]\} \quad , \\ \text{seg}^0(x) &:= \{sv \mid s \in [0, 1], v \in \text{seg}(x)\} \quad , \\ \text{cut}(x) &:= \text{seg}(x) \setminus \text{seg}^0(x) \quad , \end{aligned}$$

where $\text{cut}(x)$ is known as **cut locus** in current literature.

Consider the open set $U_x := T_x M \setminus \text{cut}(x) \subset T_x M$, then $\exp_x : U_x \rightarrow M$ is a local diffeomorphism.

Since each tangent $T_x M$ is a topological vector space, it can be considered naturally as a metric space with the metric topology induced by the Riemannian metric. Since finite dimensional topological vector spaces of the same dimension $n := \dim(M)$ are unique up to isomorphism, $T_x M$ is isomorphic to \mathbb{R}^n . Moreover, since Euclidean metric and Riemannian metric are equivalent on finite dimensional topological vector spaces, the respective induced metric topologies are also equivalent. This means probability density functions over $T_x M$ can be considered naturally as density functions over \mathbb{R}^n .

Let S_x denote a finitely parametrized family of probability densities over U_x . Since \exp_x is a diffeomorphism in U_x , we can construct parametrized family of probability distributions on $\exp_x(U_x)$ by:

$$\begin{array}{ccc} S_x \subset \text{Prob}(U_x) & \xrightarrow{\log_x^*} & \text{Prob}(M) \\ \downarrow & & \downarrow \\ U_x \subset T_x M & \xleftrightarrow[\exp_x]{\log_x} & M \end{array}$$

where $\log_x := \exp_x^{-1}$ denotes the Riemannian log function. $\tilde{p} \in \tilde{S}_x = \log_x^*$ are given by $\tilde{p}(y) = \log_x^* p(y) = p(\exp_x^{-1}(y)) = p(\log_x(y))$. This coincides with equation 3.

Since \exp_x is an orientation preserving diffeomorphism on $U_x = M \setminus \text{cut}(x)$, this reduces to a special case of the construction via orientation-preserving bundle morphism discussed above.

It is worth nothing that for general Riemannian manifolds, this approach maybe quite limiting since U_x maybe a small region in the tangent space.

Throughout the rest of the paper, we will be using the Riemannian exponential map as an example to illustrate our approach. It is however worth noting that our construction applies to all orientation-preserving diffeomorphism, not just the Riemannian exponential map.

4.2. Mixture densities on totally bounded subsets of M

In this section we discuss probability distributions on M supported beyond the region where the map ρ is a diffeomorphism. In particular, we discuss how parametrized family of mixture probability densities with locally inherited dualistic geometry can be defined on totally bounded subsets of M . We begin with the notion of orientation-preserving open cover on M .

4.2.1. Orientation-preserving open cover

Definition 3. Let M be an n -dimensional smooth manifold, an **orientation-preserving open cover** of M is a set of pairs $\{(\rho_\alpha, U_\alpha)\}$, where $U_\alpha \subset \mathbb{R}^n$, $\rho_\alpha : U_\alpha \rightarrow M$ are orientation preserving diffeomorphisms, and $\{\rho_\alpha(U_\alpha)\}$ is an open cover of M .

Remark 5. There always exists an orientation-preserving open cover for orientable smooth manifold M ; we may simply consider the smooth atlas $\mathcal{A} := \{(\rho_\alpha, U_\alpha)\}$ of M . It is worth noting that we only require $\rho_\alpha : U_\alpha \rightarrow M$ to be orientation-preserving locally on each U_α . Therefore, orientation-preserving open cover exists even when M is inorientable.

We now provide two examples of orientation-preserving open cover using Riemannian exponential map:

Example 3. Given a complete, Riemannian manifold M , and a point $x \in M$, the **injectivity radius** at x is the real number [17]:

$$\text{inj}(x) := \max_{r \in \mathbb{R}} \{ \exp_x : B(0, r) \subset T_x M \rightarrow M \text{ is a diffeomorphism} \} \quad .$$

For $x \in M$, let $B_x := B(0, \text{inj}(x)) \subset T_x M \cong \mathbb{R}^n$ denote the ball of injectivity radius centred at 0 i.e. the largest metric ball in $T_x M$ such that \exp_x is a diffeomorphism, then $\{\exp_x(B_x)\}$ is an open cover of M .

In B_x , the pushforward of the Riemannian exponential map denoted by $(\exp_x)_* : B_x \subset T_x M \rightarrow M$ is the identity map, a linear isomorphism. Therefore \exp_x is an orientation preserving local diffeomorphism on the ball of injectivity radius [17]. Hence $\{(\rho_x, U_x)\} := \{(\exp_x, B_x)\}$ is an orientation-preserving open cover of M .

Example 4. Alternatively we can consider another orientation-preserving open cover extended from the one defined above:

$$\{(\exp_x, T_x M \setminus \text{cut}(x))\} = \{(\exp_x, \text{seg}^0(x))\}$$

is also an orientation-preserving open cover of M .

4.2.2. Refinement of orientation-preserving open cover:

Given an orientation-preserving open cover over Riemannian manifold M , there exists a refinement of open cover by metric balls in M .

This will be used to construct a finite orientation-preserving open cover of totally bounded subsets of M .

Lemma 1. Let M be a n -dimensional Riemannian manifold, and orientation-preserving open cover of $\{(\rho_\alpha, U_\alpha)\}$ of M . There always exist refinement $\{(\rho_\alpha, W_\alpha)\}$ of $\{U_\alpha\}$ satisfying:

1. $\{\rho_\alpha(W_\alpha)\}$ covers M , and
2. $\rho_\alpha(W_\alpha)$ are metric balls in M .

Proof. Given (orientation-preserving) open cover $\{(\rho_\alpha, U_\alpha)\}$ of M . For each α , $\forall x_\alpha \in \rho_\alpha(U_\alpha)$ there exists normal neighbourhood $N_{x_\alpha} \subset \rho_\alpha(U_\alpha)$.

Since N_{x_α} is open for all $x_\alpha \in \rho_\alpha(U_\alpha)$, there exists $\epsilon_{x_\alpha} > 0$ such that the metric ball centred at x_α denoted by B_{x_α} satisfies: $B_{x_\alpha} := B(x_\alpha, \epsilon_{x_\alpha}) \subset N_{x_\alpha} \subset \rho_\alpha(U_\alpha) \subset M$.

In other words, B_{x_α} is the metric ball centred at x_α in normal coordinates under norm given by radial distance function.

Since ρ_α is a diffeomorphism for all α , this implies $\rho_\alpha^{-1}(B_{x_\alpha}) =: W_{x_\alpha} \subset U_\alpha$ for each $x_\alpha \in \rho_\alpha(U_\alpha)$. Hence $\{(\rho_\alpha, W_{x_\alpha})\}$ is the desired refinement of $\{(\rho_\alpha, U_\alpha)\}$.

□

Observe that the proof did not use the fact that ρ_α is orientation-preserving. It suffices to consider open cover $\{(\rho_\alpha, U_\alpha)\}$ of M such that ρ_α are just diffeomorphisms with the following simple result:

Lemma 2. *Let $f : M \rightarrow N$ be a local diffeomorphism between manifolds M, N . Then there exists local orientation-preserving diffeomorphism $\tilde{f} : M \rightarrow N$.*

Proof. Since $f : M \rightarrow N$ is a local diffeomorphism, the pushforward $f_* : T_p M \rightarrow T_{f(p)} N$ is a linear isomorphism for all $p \in M$. Since f_* is a linear isomorphism, the determinant of the matrix Df is non-zero: $\det Df \neq 0$. If f is orientation-preserving, then there's nothing left to prove. Hence we will now assume f is orientation reversing, in other words: $\det Df < 0$.

Let (x^1, \dots, x^n) denote local coordinates in M , let \bar{f} denote coordinate representation of f , then we can write:

$$\bar{f}(x^1(p), \dots, x^n(p)) = (\bar{f}_1(x), \dots, \bar{f}_n(x)) \quad ,$$

where $x := (x^1(p), \dots, x^n(p))$.

Choose $a \in [1, \dots, n]$, and let $\tilde{f} : M \rightarrow N$ denote the diffeomorphism from M to N defined by the following coordinate representation:

$$\tilde{f}(x^1(p), \dots, x^n(p)) = (\bar{f}_1(x), \dots, \bar{f}_{a+1}(x), \bar{f}_a(x), \dots, \bar{f}_n(x)) \quad .$$

In other words, we define \tilde{f} by swap the a^{th} and $a + 1^{\text{st}}$ coordinates of f . The matrix representation of \tilde{f}_* in standard coordinates is thus given by:

$$D\tilde{f} = I' \cdot Df \quad ,$$

where I' is the matrix given by:

$$I' = \begin{bmatrix} I_{a-1} & & & & & \\ & 0 & 1 & & & \\ & 1 & 0 & & & \\ & & & & & \\ & & & & & I_{n-(a+1)} \end{bmatrix} \quad ,$$

where I_k is the identity matrix of the dimension k , the sub-matrix $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is located at the $(a, a)^{\text{th}}$ to the $(a + 1, a + 1)^{\text{st}}$ position of I' , and the rest of the entries are all zero. Since $\det Df < 0$, this means $\det D\tilde{f} = \det I' \cdot \det Df = -1 \cdot \det Df > 0$. □

For the rest of the discussion we will consider the orientation-preserving open cover by metric balls $\{(\rho_\alpha, W_{x_\alpha})\}$ of Riemannian manifold M .

Example 5. *Consider once again orientation-preserving open cover of Example 3 given by $\{(\rho_\alpha, U_\alpha)\} = \{(\exp_x, T_x M \setminus \text{cut}(x))\} = \{(\exp_x, \text{seg}^o(x))\}_{x \in M}$.*

Metric balls in normal neighbourhoods are geodesic balls, i.e. for all $B(x_\alpha, \epsilon_{x_\alpha})$, there exists δ_{x_α} such that $B(x_\alpha, \epsilon_{x_\alpha}) = \exp_{x_\alpha}(B(\vec{0}, \delta_{x_\alpha}))$

An orientation-preserving open cover by metric balls can be given by the refinement: $\{(\rho_\alpha, W_{x_\alpha})\} = \left\{ \left(\exp_{x_\alpha}, B(\vec{0}, \delta_{x_\alpha}) \right) \right\}_{x_\alpha \in M}$.

4.2.3. Mixture densities on totally bounded subsets of M

Now we are ready to construct parametrized family of probability densities on totally bounded subsets of Riemannian manifold M . Let $V \subset M$ be totally bounded subset of M , and let $\{(\rho_\alpha, U_\alpha)\}$ be an orientation-preserving open cover of M . Let $\{(\rho_{x_\alpha}, W_{x_\alpha})\}$ denote a refinement of $\{(\rho_\alpha, U_\alpha)\}$ by open metric balls in M discussed in Lemma 1.

Since $\{\rho_\alpha(W_{x_\alpha}) = B(x_\alpha, \epsilon_{x_\alpha})\}_{x_\alpha \in M}$ is an open cover of M by metric balls, it is an open cover of $V \subset M$ as well. Moreover, since V is totally bounded, there exists a finite subcover $\{B(x_\alpha, \epsilon_{x_\alpha})\}_{\alpha=1}^\Lambda$ of V .

For simplicity, by an abuse of notation, we will denote the finite subcover by $\{(\rho_\alpha, W_\alpha)\}_{\alpha=1}^\Lambda := \{(\rho_{x_\alpha}, W_{x_\alpha})\}_{\alpha=1}^\Lambda$.

For each $\alpha \in [1, \dots, \Lambda]$, let $S_\alpha := \{v_\alpha := \nu_{\theta^\alpha} | \theta^\alpha \in \Xi_\alpha \subset \mathbb{R}^{m_\alpha}\} \subset \text{Prob}(W_\alpha)$ denote finitely parametrized volume form over W_α parametrized by $\theta^\alpha \in \Xi_\alpha$. Note that in general we allow the parametric families S_α to have different parametrizations and dimensions m_α . Furthermore, consider for each α the induced family of local probability densities $\tilde{S}_\alpha := \rho_\alpha^{-1*} S_\alpha = \{\tilde{v}_\alpha = \rho_\alpha^{-1*} v_\alpha\} \subset \text{Prob}(M) \subset \text{Vol}(M)$ on $\rho_\alpha(W_\alpha) \subset M$, we can then define parametrized mixture densities over $V \subset M$ by patching together the locally induced ones:

$$\tilde{v} := \sum_{\alpha=1}^\Lambda \varphi_\alpha \cdot \tilde{v}_\alpha \quad , \quad (5)$$

where $\varphi_\alpha \in (0, 1)$ for all $\alpha \in [1, \dots, \Lambda]$, and $\sum_{\alpha=1}^\Lambda \varphi_\alpha = 1$.

Let S_0 denote the simplex of mixture coefficients treated as a family of discrete distributions:

$$\begin{aligned} S_0 &:= \left\{ \{\varphi_\alpha\}_{\alpha=1}^\Lambda \mid \varphi_\alpha \in (0, 1), \sum_{\alpha=1}^\Lambda \varphi_\alpha = 1 \right\} \subset (0, 1)^{\Lambda-1} \\ &= \left\{ \{\varphi_\alpha = P(A = \alpha)\}_{\alpha=1}^\Lambda \right\} \quad . \end{aligned}$$

Then S_0 is an exponential family, hence a dually flat manifold with local parametrization $(\theta_j^0)_{j=1}^{\Lambda-1} \mapsto \{\varphi_\alpha\}_{\alpha=1}^\Lambda$, given by [5]:

$$\theta_j^0 = \log \frac{\varphi_j}{1 - \sum_{\alpha=1}^{\Lambda-1} \varphi_\alpha} \quad , \quad j \in [1, \dots, \Lambda - 1] \quad .$$

Let $\Xi_0 := \left\{ (\theta_j^0)_{j=1}^{\Lambda-1} \right\}$ denote the set of parameters of S_0 . We denote the set of mixture densities by \mathcal{L}_V :

$$\mathcal{L}_V := \left\{ \tilde{v} = \sum_{\alpha=1}^\Lambda \varphi_\alpha \cdot \tilde{v}_\alpha \mid \{\varphi_\alpha\}_{\alpha=1}^\Lambda \in S_0, \tilde{v}_\alpha \in \tilde{S}_\alpha \right\} \quad .$$

In local coordinates (x^1, \dots, x^n) of M , the mixture volume form \tilde{v}_α can be expressed in the following form:

$$\tilde{\nu} = \tilde{p}(x, \zeta) dx^1 \wedge \cdots \wedge dx^n = \sum_{\alpha} \varphi_{\alpha} \cdot \tilde{p}_{\alpha}(x, \theta^{\alpha}) dx^1 \wedge \cdots \wedge dx^n \quad , \quad (6)$$

where $\tilde{p}_{\alpha}(x, \theta^{\alpha})$ are parametrized probability density functions naturally associated to $\tilde{\nu}_{\alpha}$ that are parametrized by sets of parameters $(\theta^{\alpha}) \in \Xi_{\alpha}$ (see remark 4). In local coordinates (x^1, \dots, x^n) of M , $\tilde{p}_{\alpha}(x, \theta^{\alpha})$ is defined implicitly by $\tilde{p}_{\alpha}(x, \theta^{\alpha}) dx^1 \wedge \cdots \wedge dx^n := \tilde{\nu}_{\alpha}$. The parameters of the mixture distribution $\tilde{p}(x, \zeta)$ are collected in $(\zeta_i)_{i=1}^d := (\theta^0, \theta^1, \dots, \theta^{\Lambda}) \in \Xi_0 \times \Xi_1 \times \cdots \times \Xi_{\Lambda}$, where $d := \dim(\Xi_0 \times \Xi_1 \times \cdots \times \Xi_{\Lambda})$.

4.2.4. Remark: Exhaustion by compact set and the extent of extended support

It is worth noting that since every smooth topological manifold is σ -locally compact, M admits a compact exhaustion:

Definition 4. An *exhaustion by compact sets* is an increasing sequence of compact subsets K_j of M such that $\phi \neq K_1 \subset \text{int}(K_2) \subset K_2 \subseteq \cdots$ and:

$$\lim_{j \rightarrow \infty} K_j = \bigcup_{i=1}^{\infty} K_i = M \quad .$$

For each $x \in M$, for any totally bounded subset $V \subsetneq M$, there exists $N \in \mathbb{N}$ such that for all $n > N$:

$$\overline{K_n} (\supset K_n) \supset V \quad .$$

Even though at first glance totally bound-ness might be quite restrictive, this shows that it allows us to approximate the manifold sufficiently well.

4.3. Geometrical structure of \mathcal{L}_V

Consider a totally bounded subset $V \subset M$, orientation-preserving finite open cover $\{(\rho_{\alpha}, W_{\alpha})\}_{\alpha=1}^{\Lambda}$ of V by metric balls, and dually flat families of densities S_{α} over W_{α} . We will show that the family of mixture distributions \mathcal{L}_V is, under two conditions, a dually flat product Riemannian manifold, hence naturally inheriting the local geometry of families of component densities S_{α} established in the beginning of section 3.

For each α , let $(S_{\alpha}, g_{\alpha}, \nabla^{\alpha}, \nabla^{\alpha*}) \subset \text{Prob}(W_{\alpha})$ be a family of probability densities over W_{α} that has the structure of a dually flat statistical manifold. Let $\tilde{S}_{\alpha} := \rho_{\alpha}^{-1*} S_{\alpha}$ denote the pulled-back family of probability densities on $\rho_{\alpha}(W_{\alpha}) \subset M$ with corresponding pulled-back dualistic structures $(\tilde{g}_{\alpha}, \tilde{\nabla}^{\alpha}, \tilde{\nabla}^{\alpha*})$ discussed in section 3. Let $S_0 := \left\{ \{\varphi_{\alpha}\}_{\alpha=1}^{\Lambda} \mid \varphi_{\alpha} \in (0, 1), \sum_{\alpha=1}^{\Lambda} \varphi_{\alpha} = 1 \right\} \subset (0, 1)^{\Lambda-1}$ denote the simplex of mixture coefficients with corresponding dualistic structure $(g_0, \nabla^0, \nabla^{0*})$.

4.3.1. \mathcal{L}_V as a smooth manifold

For simplicity, let $\tilde{p}_{\alpha}(x) := \tilde{p}(x, \theta^{\alpha})$ denote the probability density function corresponding to $\tilde{\nu}_{\alpha} \in \tilde{S}_{\alpha}$ for all α . We first show that \mathcal{L}_V is indeed a smooth manifold under the following two natural conditions:

(C1). **Family of mixture component distributions have different proper support:** Let $K_{\alpha} := \{x \in M \mid \tilde{p}_{\alpha}(x) > 0, \forall \tilde{p}_{\alpha} \in \tilde{S}_{\alpha}\}$ denote the proper support of probability densities $\tilde{p}_{\alpha} \in \tilde{S}_{\alpha}$ for each $\alpha \in \{1, \dots, \Lambda\}$. We assume $K_{\alpha} \setminus K_{\beta} \neq \emptyset$ for $\beta \neq \alpha$.

(C2). **No functional dependency between mixture component densities** : We construct mixture densities in \mathcal{L}_V as unconstrained mixtures, meaning there are no functional dependency between mixture component. In other words, changing parameters $\theta^\beta \in \Xi_\beta$ of mixture component $\tilde{p}_\beta \in \tilde{\mathcal{S}}_\beta$ has no influence on $\tilde{p}_\alpha \in \tilde{\mathcal{S}}_\alpha$ for $\beta \neq \alpha$ and vice versa. We write this condition as follows: For each $\tilde{p}_\alpha \in \tilde{\mathcal{S}}_\alpha$, $\frac{\partial \tilde{p}_\alpha}{\partial \theta_k^\beta} = 0$, $\forall \theta_k^\beta \in \Xi^\beta \subset \mathbb{R}^{m_\beta}, \forall \beta \neq \alpha$.

Remark 6. 1. The first condition C1 can always be satisfied simply by choosing a suitable open cover of V .
2. The second condition C2 is automatically fulfilled for unconstrained mixture models. One can imagine introducing functional dependencies among mixture component distributions, but this is not the case considered here. We make the assumption that: if we alter one distribution $\tilde{p}_\alpha \in \tilde{\mathcal{S}}_\alpha$, it does not affect distributions in $\tilde{\mathcal{S}}_\beta$ for $\beta \neq \alpha$.

We now discuss the implications of conditions C1 and C2 in further detail:

The first condition C1 implies: The component distributions $\tilde{p}_\alpha \in \tilde{\mathcal{S}}_\alpha$ are linearly independent functions, and the map $(\theta^0) \in \Xi_0 \mapsto (\varphi_\alpha)_{\alpha=1}^\Lambda \mapsto \sum_{\alpha=1}^\Lambda \varphi_\alpha \tilde{p}_\alpha(x, \theta^\alpha)$ is injective.

The second condition C2 implies:

Consider two distributions $\tilde{p}, \tilde{q} \in \mathcal{L}_V$ sharing the same mixture coefficients $\{\varphi_\alpha\}_{\alpha=1}^\Lambda$, i.e. $\tilde{p}(x) = \sum_{\alpha=1}^\Lambda \varphi_\alpha \tilde{p}_\alpha(x)$ and $\tilde{q} = \sum_{\alpha=1}^\Lambda \varphi_\alpha \tilde{q}_\alpha(x)$.

If $\tilde{p}(x) = \tilde{q}(x)$ for all $x \in M$:

$$\begin{aligned} \frac{\partial \tilde{p}(x)}{\partial \theta_i^\alpha} &= \frac{\partial \tilde{q}(x)}{\partial \theta_i^\alpha} \quad , \quad \forall \alpha, \forall i \\ \frac{\partial^\ell \tilde{p}(x)}{(\partial \theta_i^\alpha)^\ell} &= \frac{\partial^\ell \tilde{q}(x)}{(\partial \theta_i^\alpha)^\ell} \quad , \quad \forall \ell \in \mathbb{N}_+ \quad , \end{aligned}$$

hence by condition C2, for each $\alpha \in [1, \dots, \Lambda]$, there exists a constant c_α such that

$$\varphi_\alpha \tilde{p}_\alpha(x) = \varphi_\alpha \tilde{q}_\alpha(x) + c_\alpha, \quad \forall x \in M \quad .$$

Since \tilde{p}_α and \tilde{q}_α are probability densities:

$$\varphi_\alpha \cdot \underbrace{\int_M \tilde{p}_\alpha}_{=1} = \varphi_\alpha \cdot \underbrace{\int_M \tilde{q}_\alpha}_{=1} + \int_M c_\alpha \quad .$$

This means $\int_M c_\alpha = c_\alpha \cdot \int_M 1 = 0$. Since M has finite (non-zero) Riemannian volume: $\int_M 1 > 0$, this implies $c_\alpha = 0$ and hence $\tilde{p}_\alpha = \tilde{q}_\alpha$ for all α .

Hence by injectivity of the parametrization mapping $\theta^\alpha \mapsto \tilde{p}_\alpha$ of $\tilde{\mathcal{S}}_\alpha$ for $\alpha \in [1, \dots, \Lambda]$, the parametrization of mixture component parameters $(\theta^1, \dots, \theta^\Lambda) \mapsto (\tilde{p}_1, \dots, \tilde{p}_\Lambda)$ is injective as well.

Therefore the parametrization map $(\theta^0, \theta^1, \dots, \theta^\Lambda) \in \Xi_0 \times \Xi_1 \times \dots \times \Xi_\Lambda \mapsto \tilde{p}(x) = \sum_{\alpha=1}^\Lambda \varphi_\alpha \tilde{p}_\alpha(x)$ onto \mathcal{L}_V is an isomorphism.

By condition C2, the parameters are also independent in the sense that for $\beta \neq \alpha$:

$$0 = \frac{\partial \tilde{p}_\alpha}{\partial \theta_k^\beta} = \frac{\partial \tilde{p}_\alpha}{\partial \theta_i^\alpha} \cdot \frac{\partial \theta_i^\alpha}{\partial \theta_k^\beta} \quad \forall \tilde{p}_\alpha \in \tilde{\mathcal{S}}_\alpha \Leftrightarrow \frac{\partial \theta_i^\alpha}{\partial \theta_k^\beta} = 0 \quad .$$

Moreover, recall the pulled-back local coordinates maps $(\bar{\theta}_i^\alpha := \theta_i^\alpha \circ (\rho^{-1})^* = \theta_i^\alpha \circ \rho^*)$ on \tilde{S}_α , then:

$$\frac{\partial \theta_i^\alpha \circ \rho^*}{\partial \theta_k^\beta \circ \rho^*} = (\rho^{-1})^* \frac{\partial}{\partial \theta_k^\beta} (\theta_i^\alpha \circ \rho^*) = \frac{\partial}{\partial \theta_k^\beta} (\theta_i^\alpha \circ \rho^* \circ \rho^{-1}) = 0 \quad .$$

In other words, pullback by ρ does not introduce additional functional dependencies among parameters.

Mixture densities $\tilde{p} \in \mathcal{L}_V$ can thus be identified naturally by the map $\tilde{p} = \sum_{\alpha=1}^{\Lambda} \varphi_\alpha \tilde{p}_\alpha \mapsto (\varphi_1, \dots, \varphi_\Lambda, \tilde{p}_1, \dots, \tilde{p}_\Lambda)$, where the image represents the mixture coefficients and the mixture component distributions. Since parametrizations $\theta^\alpha \mapsto \tilde{p}_\alpha$ are smooth with smooth inverse, \mathcal{L}_V is a smooth manifold with coordinates $\mathcal{L}_V \ni \tilde{p} := \sum_{\alpha=1}^{\Lambda} \varphi_\alpha \tilde{p}_\alpha(x, \theta^\alpha) \mapsto (\theta^0, \theta^1, \dots, \theta^\Lambda) \in \Xi_0 \times \Xi_1 \times \dots \times \Xi_\Lambda$.

4.3.2. Torsion-free dualistic structure on \mathcal{L}_V

Now consider the following function on $\mathcal{L}_V \times \mathcal{L}_V$:

$$\begin{aligned} D : \mathcal{L}_V \times \mathcal{L}_V &\rightarrow \mathbb{R} \\ D(\tilde{p}, \tilde{q}) &= D\left(\sum_{\alpha=1}^{\Lambda} \varphi_\alpha p_\alpha, \sum_{\alpha=1}^{\Lambda} \varphi'_\alpha q_\alpha\right) := D_{KL}(\{\varphi_\alpha\}, \{\varphi'_\alpha\}) + \sum_{\alpha=1}^{\Lambda} \bar{D}_\alpha(\tilde{p}_\alpha, \tilde{q}_\alpha) \\ &= \sum_{\alpha=1}^{\Lambda} \varphi_\alpha \log\left(\frac{\varphi_\alpha}{\varphi'_\alpha}\right) + \sum_{\alpha=1}^{\Lambda} \bar{D}_\alpha(\tilde{p}_\alpha, \tilde{q}_\alpha) \quad . \end{aligned} \quad (7)$$

where D_{KL} is the Kullback-Leibler divergence (relative entropy) on mixture coefficients $S_0 = \{\{\varphi_\alpha = P(A = \alpha)\}_{\alpha=1}^{\Lambda}\}$ as family of discrete distributions, and \bar{D}_α is the induced divergence on smooth manifolds \tilde{S}_α described by Remark 2 in Section 3.

It is immediate by definition that D satisfies the conditions of a divergence:

1. Non-negativity: Since \bar{D}_α 's and D_{KL} are both non-negative, so is D :

$$D = \underbrace{D_{KL}}_{\geq 0} + \sum_{\alpha=1}^{\Lambda} \underbrace{\bar{D}_\alpha}_{\geq 0}$$

2. Identity: Since \bar{D}_α 's and D_{KL} are divergences, the following is satisfied:

$$\begin{aligned} D = 0 &\Leftrightarrow \left\{ \begin{array}{l} D_{KL}(\varphi) = 0 \\ \bar{D}_\alpha(\tilde{p}_\alpha, \tilde{q}_\alpha) = 0 \quad \forall \alpha \end{array} \right\} \\ &\Leftrightarrow \left\{ \begin{array}{l} \varphi_\alpha = \varphi'_\alpha \\ \tilde{p}_\alpha = \tilde{q}_\alpha \end{array} \right\} \quad \forall \alpha \\ &\Leftrightarrow \tilde{p} = \tilde{q} \end{aligned}$$

Let $\theta^\alpha \in \Xi^\alpha \subset \mathbb{R}^{m_\alpha}$ be coordinates of S_α , and one again let $(\xi_i)_{i=1}^d := (\theta^0, \tilde{\theta}^1, \dots, \tilde{\theta}^\Lambda)$ denote the coordinates of \mathcal{L}_V , where $d := \dim(\Xi_0 \times \Xi_1 \times \dots \times \Xi_\Lambda)$ and $(\tilde{\theta}_j^\alpha := \theta_j^\alpha \circ \rho^*)$ denote the local pulled-back coordinates of \tilde{S}_α discussed in section 3.2 (see also Remark 3 and 4). Let the corresponding coordinate frame be denoted by $(\bar{\partial}_i := \frac{\partial}{\partial \xi_i})$. We can then construct a dualistic structure on \mathcal{L}_V [5,15]

via divergence function D denoted by $(g^D, \nabla^D, \nabla^{D*})$ with the following equations as discussed in Section 3

$$\begin{aligned} g_{ij}^D(\bar{p}) &:= g^D(\bar{\partial}_i, \bar{\partial}_j) \Big|_{\bar{p}} = -D[\bar{\partial}_i; \bar{\partial}_j] \Big|_{\bar{p}} \\ &= -D_{KL}[\bar{\partial}_i; \bar{\partial}_j] \Big|_{\bar{p}} + \sum_{\alpha=1}^{\Lambda} -\bar{D}_\alpha[\bar{\partial}_i; \bar{\partial}_j] \Big|_{\bar{p}} \\ &= g^{DKL}(\bar{\partial}_i, \bar{\partial}_j) \Big|_{\bar{p}} + \sum_{\alpha=1}^{\Lambda} g^{\bar{D}_\alpha}(\bar{\partial}_i, \bar{\partial}_j) \Big|_{\bar{p}} . \end{aligned} \quad (8)$$

Moreover, if $\bar{\partial}_i = \frac{\partial}{\partial \theta_i^0}$ then:

$$\begin{aligned} g^D(\bar{\partial}_i, \bar{\partial}_j) \Big|_{\bar{p}} &= g^{DKL}(\bar{\partial}_i, \bar{\partial}_j) \Big|_{\bar{p}} + 0 \\ &= g^{DKL}(\bar{\partial}_i, \bar{\partial}_j) \Big|_{\{\varphi_\alpha\}_{\alpha=1}^{\Lambda}} , \end{aligned} \quad (9)$$

where $\{\varphi_\alpha\}_{\alpha=1}^{\Lambda} \in S_0$. If $\bar{\partial}_i = \partial_i^\beta = \frac{\partial}{\partial \theta_i^\beta}$, where (θ_i^β) denote parametrization of \tilde{S}_β for some $\beta \in \{1, \dots, \Lambda\}$, then by condition C2:

$$\begin{aligned} g^D(\bar{\partial}_i, \bar{\partial}_j) \Big|_{\bar{p}} &= 0 + \sum_{\alpha=1}^{\Lambda} g^{\bar{D}_\alpha}(\partial_i^\beta, \partial_j^\beta) \Big|_{\bar{p}_\alpha} \\ &= 0 + \sum_{\alpha=\beta} g^{\bar{D}_\alpha}(\partial_i^\beta, \partial_j^\beta) \Big|_{\bar{p}_\alpha} \\ &= g^{\bar{D}_\beta}(\partial_i^\beta, \partial_j^\beta) \Big|_{\bar{p}_\beta} = g^\beta(\partial_i^\beta, \partial_j^\beta) \Big|_{p_\beta} , \end{aligned} \quad (10)$$

where g^β is the metric on S_β , and $p_\beta := (\rho^{-1*})^{-1} \bar{p}_\beta = \rho^* \bar{p}_\beta$. The last equality is due to the analysis towards the end of Section 3. Otherwise, let $\bar{\partial}_i = \frac{\partial}{\partial \theta_i^0}$, and $\bar{\partial}_j = \partial_j^\beta = \frac{\partial}{\partial \theta_j^\beta}$ for $i \neq j$, then

$$\begin{aligned} g^D(\bar{\partial}_i, \bar{\partial}_j) \Big|_{\bar{p}} &= g^{DKL}(\bar{\partial}_i, 0) \Big|_{\{\varphi_\alpha\}_{\alpha=1}^{\Lambda}} + g^{\bar{D}_\beta}(0, \partial_j^\beta) \Big|_{\bar{p}_\beta} \\ &= 0 + 0 = 0 . \end{aligned}$$

The Christoffel symbols of the connection ∇^D is given by:

$$\begin{aligned} \Gamma_{ijk}^D \Big|_{\bar{p}} &:= \langle \nabla_{\bar{\partial}_i}^D \bar{\partial}_j, \bar{\partial}_k \rangle_{g^D} \Big|_{\bar{p}} = -D[\bar{\partial}_i \bar{\partial}_j; \bar{\partial}_k] \Big|_{\bar{p}} \\ &= -D_{KL}[\bar{\partial}_i \bar{\partial}_j; \bar{\partial}_k] \Big|_{\bar{p}} + \sum_{\alpha=1}^{\Lambda} -\bar{D}_\alpha[\bar{\partial}_i \bar{\partial}_j; \bar{\partial}_k] \Big|_{\bar{p}} \\ &= \langle \nabla_{\bar{\partial}_i}^{DKL} \bar{\partial}_j, \bar{\partial}_k \rangle_{g^D} \Big|_{\bar{p}} + \sum_{\alpha=1}^{\Lambda} \langle \nabla_{\bar{\partial}_i}^{\bar{D}_\alpha} \bar{\partial}_j, \bar{\partial}_k \rangle_{g^D} \Big|_{\bar{p}} , \end{aligned} \quad (11)$$

where we recall $-D[\partial_i \partial_j; \partial_k] := -\partial_i^1 \partial_j^1 \partial_k^2 D[p; q] \Big|_{q=p}$ by equation (2) in Section 3. By a similar argument as above in equation 9 and 10, we obtain the following:

$$\langle \nabla_{\bar{\partial}_i}^D \bar{\partial}_j, \bar{\partial}_k \rangle_{g^D} \Big|_{\bar{p}} = \begin{cases} \langle \nabla_{\bar{\partial}_i}^{D_{KL}} \bar{\partial}_j, \bar{\partial}_k \rangle_{g^{D_{KL}}} \Big|_{\{\varphi_\alpha\}_{\alpha=1}^\Lambda} & \text{if } \bar{\partial}_\ell = \frac{\partial}{\partial \theta_\ell^0}, \text{ for } \ell = i, j, k, \\ \langle \nabla_{\bar{\partial}_i}^{\bar{D}^\beta} \bar{\partial}_j, \bar{\partial}_k \rangle_{g^{D_\beta}} \Big|_{p_\beta} & \text{if } \bar{\partial}_\ell = \partial_\ell^\beta = \frac{\partial}{\partial \theta_\ell^\beta}, \text{ for } \ell = i, j, k, \\ 0 & \text{Otherwise.} \end{cases} \quad (12)$$

This implies that for $X_0, Y_0 \in \mathcal{E}(TS_0)$, $X_\alpha, Y_\alpha \in \mathcal{E}(T\tilde{S}_\alpha)$ for $\alpha \in [1, \dots, \Lambda]$, we have the following:

$$\begin{aligned} \nabla_{(Y_0 + \sum_{\alpha=1}^\Lambda Y_\alpha)}^D \left(X_0 + \sum_{\alpha=1}^\Lambda X_\alpha \right) &= \nabla^D \left(X_0 + \sum_{\alpha=1}^\Lambda X_\alpha, Y_0 + \sum_{\alpha=1}^\Lambda Y_\alpha \right) \\ &= \nabla_{Y_0}^{D_{KL}} X_0 + \sum_{\alpha=1}^\Lambda \nabla_{Y_\alpha}^{\bar{D}^\alpha} X_\alpha. \end{aligned} \quad (13)$$

By symmetry and the fact that $g^D = g^{D^*}$ [5], we also have the following:

$$\langle \nabla_{\bar{\partial}_i}^{D^*} \bar{\partial}_j, \bar{\partial}_k \rangle_{g^D} \Big|_{\bar{p}} = \langle \nabla_{\bar{\partial}_i}^{D_{KL}} \bar{\partial}_j, \bar{\partial}_k \rangle_{g^D} \Big|_{\bar{p}} + \sum_{\alpha=1}^\Lambda \langle \nabla_{\bar{\partial}_i}^{\bar{D}^\alpha} \bar{\partial}_j, \bar{\partial}_k \rangle_{g^D} \Big|_{\bar{p}}, \quad (14)$$

and we also obtain the following result analogous to equation (12):

$$\langle \nabla_{\bar{\partial}_i}^{D^*} \bar{\partial}_j, \bar{\partial}_k \rangle_{g^D} \Big|_{\bar{p}} = \begin{cases} \langle \nabla_{\bar{\partial}_i}^{D_{KL}^*} \bar{\partial}_j, \bar{\partial}_k \rangle_{g^{D_{KL}^*}} \Big|_{\{\varphi_\alpha\}_{\alpha=1}^\Lambda} & \text{if } \bar{\partial}_\ell = \frac{\partial}{\partial \theta_\ell^0}, \text{ for } \ell = i, j, k \\ \langle \nabla_{\bar{\partial}_i}^{\bar{D}^{\beta^*}} \bar{\partial}_j, \bar{\partial}_k \rangle_{g^{\bar{D}_\beta^*}} \Big|_{p_\beta} & \text{if } \bar{\partial}_\ell = \partial_\ell^\beta = \frac{\partial}{\partial \theta_\ell^\beta}, \text{ for } \ell = i, j, k, \\ 0 & \text{Otherwise.} \end{cases} \quad (15)$$

Finally, for $X_0, Y_0 \in \mathcal{E}(TS_0)$, $X_\alpha, Y_\alpha \in \mathcal{E}(T\tilde{S}_\alpha)$ for $\alpha \in [1, \dots, \Lambda]$, we would have the following:

$$\begin{aligned} \nabla_{(Y_0 + \sum_{\alpha=1}^\Lambda Y_\alpha)}^{D^*} \left(X_0 + \sum_{\alpha=1}^\Lambda X_\alpha \right) &= \nabla^{D^*} \left(X_0 + \sum_{\alpha=1}^\Lambda X_\alpha, Y_0 + \sum_{\alpha=1}^\Lambda Y_\alpha \right) \\ &= \nabla_{Y_0}^{D_{KL}^*} X_0 + \sum_{\alpha=1}^\Lambda \nabla_{Y_\alpha}^{\bar{D}_\alpha^*} X_\alpha. \end{aligned} \quad (16)$$

Hence by equations (8, 13, 16), the dualistic structure $\{g^D, \nabla^D, \nabla^{D^*}\}$ naturally decomposes into the part of mixture coefficients and mixture components. We abbreviate equations (8, 13, 16) to the following compact notation:

$$\left\{ g^{D_{KL}} \oplus \bigoplus_{\alpha=1}^\Lambda g^{\bar{D}_\alpha}, \nabla^{D_{KL}} \oplus \bigoplus_{\alpha=1}^\Lambda \nabla^{\bar{D}_\alpha}, \nabla^{D_{KL}^*} \oplus \bigoplus_{\alpha=1}^\Lambda \nabla^{\bar{D}_\alpha^*} \right\}. \quad (17)$$

4.4. Dualistic structure of \mathcal{L}_V

To show that \mathcal{L}_V is indeed a product Riemannian manifold, we recall some properties of product Riemannian manifolds [18, 19].

Remark 7. Note that since \mathcal{L}_V consists of a finite mixture of probability distributions, to show that $\mathcal{L}_V = S_0 \times \tilde{S}_1 \times \cdots \times \tilde{S}_\Lambda$, it suffices to consider the dualistic structure of the product of two manifolds.

Given two Riemannian manifolds $(M, g_1), (N, g_2)$, the product Riemannian metric on $M \times N$ is given by [18]:

$$g(X, Y) = g_1(PX, PY) + g_2(QX, QY) \quad ,$$

where P, Q are projections from $T(M \times N)$ to TM, TN respectively. Suppose ∇^1 and ∇^2 are connections of M, N respectively, then the product connection is given by [19]:

$$\nabla_{Y_1+Y_2}X_1 + X_2 = \nabla_{Y_1}^1 X_1 + \nabla_{Y_2}^2 X_2 \quad , \quad (18)$$

where $X_1, Y_1 \in TM$ and $X_2, Y_2 \in TN$. We once again abbreviate the product connection to a more compact notation for simplicity: $\nabla = \nabla^1 \oplus \nabla^2$. Since the Lie bracket of $M \times N$ is :

$$[X_1 + X_2, Y_1 + Y_2]_{M \times N} = [X_1, Y_1]_M + [X_2, Y_2]_N \quad ,$$

and the curvature tensor is given by:

$$R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X, Y]} Z \quad .$$

The curvature tensor on the $M \times N$ is thus:

$$R(X_1 + X_2, Y_1 + Y_2, Z_1 + Z_2, W_1 + W_2) = R_1(X_1, Y_1, Z_1, W_1) + R_2(X_2, Y_2, Z_2, W_2) \quad , \quad (19)$$

where R_1, R_2 denote the curvature tensor of M, N respectively. Hence if M and N are flat, so is $M \times N$.

Therefore to show that the Riemannian structure derived from the divergence D in equation (7) (given by equation (17)) coincides with the product Riemannian structure discussed above, it suffices to show the following result.

Theorem 2. Let $(M, g_1, \nabla^1, \nabla^{1*}), (N, g_2, \nabla^2, \nabla^{2*})$ be two smooth manifolds with their corresponding dualistic structure, and consider their product $(M \times N, g, \nabla)$ with the product metric g and product connection $\nabla := \nabla^1 \oplus \nabla^2$, then the connection $\nabla^{1*} \oplus \nabla^{2*}$ is g -dual to ∇ . In particular $(\nabla^1 \oplus \nabla^2)^* = \nabla^{1*} \oplus \nabla^{2*}$. Furthermore, if M and N are dully flat, then so is $M \times N$.

Proof. Let $(M, g_1, \nabla^1, \nabla^{1*}), (N, g_2, \nabla^2, \nabla^{2*})$ be two smooth manifolds with their corresponding dualistic structure. Let $\nabla = \nabla^1 \oplus \nabla^2$ denote the product connection on $M \times N$ given by equation (18) in compact notation. ² Let $\langle \cdot, \cdot \rangle := \langle \cdot, \cdot \rangle_g$, where $g = g_1 \oplus g_2$ denote the product Riemannian metric on $M \times N$.

Let $(x_1, x_2) \in M \times N$ be arbitrary, and let $\{X, Y, Z\} = \{X_1 + X_2, Y_1 + Y_2, Z_1 + Z_2\} \in \mathcal{E}(T(M \times N))$, then we have:

² Please refer to comments following equation (18)

$$X\langle Y, Z \rangle = (X_1 + X_2)\langle Y_1 + Y_2, Z_1 + Z_2 \rangle \quad (20)$$

$$\begin{aligned} &= X_1 (\langle Y_1, Z_1 \rangle + \langle Y_1, Z_2 \rangle + \langle Y_2, Z_1 \rangle + \langle Y_2, Z_2 \rangle) \\ &\quad + X_2 (\langle Y_1, Z_1 \rangle + \langle Y_1, Z_2 \rangle + \langle Y_2, Z_1 \rangle + \langle Y_2, Z_2 \rangle) \quad , \end{aligned} \quad (21)$$

where $X_1, Y_1, Z_1 \in \mathcal{E}(TM)$, $X_2, Y_2, Z_2 \in \mathcal{E}(TN)$ follows from the natural identification $T_{(x_1, x_2)}(M \times N) \cong T_{x_1}M \times T_{x_2}N$.

Let $\pi_{*1} : T(M \times N) \rightarrow M$ and $\pi_{*2} : T(M \times N) \rightarrow N$ denote the natural projection map, then:

$$\begin{aligned} \langle Y_1, Z_2 \rangle &= \langle Y_1, \pi_{*1} Z_2 \rangle_{g_1} + \langle \pi_{*2} Y_1, Z_2 \rangle_{g_2} \\ &= \langle Y_1, 0 \rangle_{g_1} + \langle 0, Z_2 \rangle_{g_2} = 0 \quad , \end{aligned} \quad (22)$$

and similarly $\langle Y_2, Z_1 \rangle = 0$ as well.

Hence equation (20) equals to:

$$\begin{aligned} &X_1 (\langle Y_1, Z_1 \rangle + \langle Y_2, Z_2 \rangle) + X_2 (\langle Y_1, Z_1 \rangle + \langle Y_2, Z_2 \rangle) \\ &= \nabla_{X_1}^1 \langle Y_1, Z_1 \rangle + \nabla_{X_1}^1 \langle Y_2, Z_2 \rangle \\ &\quad + \nabla_{X_2}^2 \langle Y_1, Z_1 \rangle + \nabla_{X_2}^2 \langle Y_2, Z_2 \rangle \\ &= \nabla_{X_1}^1 \langle Y_1, Z_1 \rangle + 0 + 0 + \nabla_{X_2}^2 \langle Y_2, Z_2 \rangle \quad . \end{aligned} \quad (23)$$

The last equality is due to the fact that $\langle Y_1, Z_1 \rangle$ is a function on M , hence $\nabla_{X_2}^2 \langle Y_1, Z_1 \rangle = 0$, and by symmetry $\nabla_{X_1}^1 \langle Y_2, Z_2 \rangle = 0$ as well. By a similar argument, we also have the following:

$$\begin{aligned} \langle \nabla_X Y, Z \rangle &= \langle \nabla_{X_1}^1 Y_1, Z_1 + Z_2 \rangle + \langle \nabla_{X_2}^2 Y_2, Z_1 + Z_2 \rangle \\ &= \langle \nabla_{X_1}^1 Y_1, Z_1 \rangle + \langle \nabla_{X_1}^1 Y_1, Z_2 \rangle \\ &\quad + \langle \nabla_{X_2}^2 Y_2, Z_1 \rangle + \langle \nabla_{X_2}^2 Y_2, Z_2 \rangle \\ &= \langle \nabla_{X_1}^1 Y_1, Z_1 \rangle + \langle \nabla_{X_2}^2 Y_2, Z_2 \rangle \quad . \end{aligned} \quad (24)$$

The first equality is due to equation (18), the assumption that ∇ is a product connection. The last equality is due to a similar reason as equation (22), where $\langle \nabla_{X_1}^1 Y_1, Z_2 \rangle = \langle \nabla_{X_2}^2 Y_2, Z_1 \rangle = 0$, since $\pi_{*i} Z_j = 0$, and $\pi_{*i} \nabla_{X_j}^j Y_j = 0$ for $i \neq j$.

Subtracting equation (24) from equation (23), we have the equality:

$$\begin{aligned} X\langle Y, Z \rangle - \langle \nabla_X Y, Z \rangle &= \nabla_{X_1}^1 \langle Y_1, Z_1 \rangle + 0 + 0 + \nabla_{X_2}^2 \langle Y_2, Z_2 \rangle \\ &\quad - \left(\langle \nabla_{X_1}^1 Y_1, Z_1 \rangle + \langle \nabla_{X_2}^2 Y_2, Z_2 \rangle \right) \\ &= \langle \nabla_{X_1}^{1*} Z_1, Y_1 \rangle + \langle \nabla_{X_1}^{2*} Z_1, Y_1 \rangle \quad , \end{aligned}$$

where ∇^{1*}, ∇^{2*} denote the g_1, g_2 -dual connection to ∇^1, ∇^2 on M, N respectively. The unique [12] g -dual connection to ∇ of $M \times N$, denoted by ∇^* , is thus given by the following:

$$\nabla_{Y_1+Y_2}^* X_1 + X_2 = \nabla_{Y_1}^{1*} X_1 + \nabla_{Y_2}^{2*} X_2 \quad .$$

Furthermore, since the curvature of $M \times N$ satisfies product curvature tensor described in equation (19), if $(M, g_1, \nabla^1, \nabla^{1*}), (N, g_2, \nabla^2, \nabla^{2*})$ are both dually flat, then so is their product $(M \times N, g_1 + g_2, \nabla^1 \oplus \nabla^2, \nabla^{1*} \oplus \nabla^{2*})$.

□

Remark 8. By the theorem above, $\mathcal{L}_V = S_0 \times \tilde{S}_1 \times \cdots \times \tilde{S}_\Lambda$ is therefore a product manifold with product dualistic structure:

$$\left\{ g_0 \oplus \bigoplus_{\alpha=1}^{\Lambda} \tilde{g}_\alpha, \nabla^0 \oplus \bigoplus_{\alpha=1}^{\Lambda} \tilde{\nabla}^\alpha, \nabla^{0*} \oplus \bigoplus_{\alpha=1}^{\Lambda} \tilde{\nabla}^{\alpha*} \right\} = \left\{ g^{D_{KL}} \oplus \bigoplus_{\alpha=1}^{\Lambda} g^{\bar{D}_\alpha}, \nabla^{D_{KL}} \oplus \bigoplus_{\alpha=1}^{\Lambda} \nabla^{\bar{D}_\alpha}, \nabla^{D_{KL}^*} \oplus \bigoplus_{\alpha=1}^{\Lambda} \nabla^{\bar{D}_\alpha^*} \right\},$$

where the equality follows from equation (17) and the discussion towards the end of Section 3.2.³

Since mixture coefficients S_0 correspond to the family of multinomial distributions, which in turn is a member of the exponential family, it is dually flat. Therefore by the above result, if the mixture component families S_1, \dots, S_Λ on orientation-preserving open cover $\{W_\alpha\}_{\alpha=1}^\Lambda$ are all dually flat, then so is $\tilde{S}_1, \dots, \tilde{S}_\Lambda$. Therefore by applying induction on the above result, $\mathcal{L}_V = S_0 \times \tilde{S}_1 \times \cdots \times \tilde{S}_\Lambda$ is also dually flat.

Recall from equation (7) that a divergence function D on \mathcal{L}_V can be defined by the following:

$$\begin{aligned} D : \mathcal{L}_V \times \mathcal{L}_V &\rightarrow \mathbb{R} \\ D(\tilde{p}, \tilde{q}) &= D\left(\sum_{\alpha=1}^{\Lambda} \varphi_\alpha p_\alpha, \sum_{\alpha=1}^{\Lambda} \varphi'_\alpha q_\alpha\right) := D_{KL}(\{\varphi_\alpha\}, \{\varphi'_\alpha\}) + \sum_{\alpha=1}^{\Lambda} \bar{D}_\alpha(\tilde{p}_\alpha, \tilde{q}_\alpha) \\ &= \sum_{\alpha=1}^{\Lambda} \varphi_\alpha \log\left(\frac{\varphi_\alpha}{\varphi'_\alpha}\right) + \sum_{\alpha=1}^{\Lambda} \bar{D}_\alpha(\tilde{p}_\alpha, \tilde{q}_\alpha). \end{aligned}$$

Consider $(S_0, g_0, \nabla^0, \nabla^{0*}), (\tilde{S}_1, \tilde{g}_1, \tilde{\nabla}^1, \tilde{\nabla}^{1*}), \dots, (\tilde{S}_\Lambda, \tilde{g}_\Lambda, \tilde{\nabla}^\Lambda, \tilde{\nabla}^{\Lambda*})$ as dually flat manifolds with their corresponding dualistic structures.⁴

We will show that: if $S_0, S_1, \dots, S_\Lambda$ are all dually flat, the divergence function D defined in equation (7) is in fact the canonical divergence of $\mathcal{L}_V = S_0 \times \tilde{S}_1 \times \cdots \times \tilde{S}_\Lambda$, with respect to product dualistic structure

$$\left\{ g_0 \oplus \bigoplus_{\alpha=1}^{\Lambda} \tilde{g}_\alpha, \nabla^0 \oplus \bigoplus_{\alpha=1}^{\Lambda} \tilde{\nabla}^\alpha, \nabla^{0*} \oplus \bigoplus_{\alpha=1}^{\Lambda} \tilde{\nabla}^{\alpha*} \right\}.$$

Let (θ^0, η_0) denote the local g_0 -dual coordinates of S_0 , and let $(\tilde{\theta}^\alpha)$ denote the local $\tilde{\nabla}^\alpha$ -affine coordinates of \tilde{S}_α for $\alpha \in [1, \dots, \Lambda]$.

By the discussion towards the end of section 3, let $\tilde{\psi}_\alpha$ denote the pulled-back potential function on \tilde{S}_α defined by local coordinates $(\tilde{\theta}^\alpha)$ and pulled-back metric \tilde{g}_α on \tilde{S}_α . The \tilde{g}_α -dual local coordinates to $(\tilde{\theta}^\alpha)$ can be defined via induced potential function $\tilde{\psi}_\alpha$ by: $(\tilde{\eta}_\alpha^i := \frac{\partial}{\partial \tilde{\theta}_\alpha^i} \tilde{\psi}_\alpha)$.

Since S_0 and \tilde{S}_α are all dually flat for $\alpha \in [1, \dots, \Lambda]$, we can write the divergences D_{KL} and \bar{D}_α of S_0 and \tilde{S}_α in the canonical form [5] respectively as follows:

³ For detailed discussion of the dualistic structures on $S_0, \tilde{S}_1, \dots, \tilde{S}_\Lambda$ please refer to the beginning of Section 4.3.

⁴ Please refer to the beginning of this section (4.3) for detailed discussion of the induced dualistic structure on \tilde{S}_α for $\alpha \in [1, \dots, \Lambda]$.

$$D_{KL}(\varphi, \varphi') := \psi_0(\varphi) + \psi_0^\dagger(\varphi') - \langle \theta^0(\varphi), \eta_0(\varphi') \rangle, \\ \bar{D}_\alpha(\tilde{p}_\alpha, \tilde{q}_\alpha) := \tilde{\psi}_\alpha(\tilde{p}_\alpha) + \tilde{\psi}_\alpha^\dagger(\tilde{q}_\alpha) - \langle \tilde{\theta}^\alpha(\tilde{p}_\alpha), \tilde{\eta}_\alpha(\tilde{q}_\alpha) \rangle.$$

The functions ψ_0^\dagger and $\tilde{\psi}_\alpha^\dagger$ denote the Legendre-Fenchel transformation of ψ_0 and $\tilde{\psi}_\alpha$, and given by the following equations, respectively:

$$\psi_0^\dagger(\varphi') := \sup_{\varphi \in \mathcal{S}_0} \left\{ \langle \theta^0(\varphi), \eta_0(\varphi') \rangle - \psi_0(\varphi) \right\}, \\ \tilde{\psi}_\alpha^\dagger(\tilde{q}_\alpha) := \sup_{\tilde{p}_\alpha \in \mathcal{S}_\alpha} \left\{ \langle \tilde{\theta}^\alpha(\tilde{p}_\alpha), \tilde{\eta}_\alpha(\tilde{q}_\alpha) \rangle - \tilde{\psi}_\alpha(\tilde{p}_\alpha) \right\}.$$

The divergence D on \mathcal{L}_V from equation (7) can then be expressed as:

$$D(\tilde{p}, \tilde{q}) = D\left(\sum_{\alpha=1}^{\Lambda} \varphi_\alpha p_\alpha, \sum_{\alpha=1}^{\Lambda} \varphi'_\alpha q_\alpha\right) = D_{KL}(\{\varphi_\alpha\}, \{\varphi'_\alpha\}) + \sum_{\alpha=1}^{\Lambda} \bar{D}_\alpha(\tilde{p}_\alpha, \tilde{q}_\alpha) \\ = \psi_0(\varphi) + \psi_0^\dagger(\varphi') - \langle \theta(\varphi), \eta(\varphi') \rangle \\ + \sum_{\alpha=1}^{\Lambda} \tilde{\psi}_\alpha(\tilde{p}_\alpha) + \tilde{\psi}_\alpha^\dagger(\tilde{q}_\alpha) - \langle \tilde{\theta}^\alpha(\tilde{p}_\alpha), \tilde{\eta}_\alpha(\tilde{q}_\alpha) \rangle \\ = \left(\psi_0(\varphi) + \sum_{\alpha} \tilde{\psi}_\alpha(\tilde{p}_\alpha) \right) + \left(\psi_0^\dagger(\varphi') + \sum_{\alpha=1}^{\Lambda} \tilde{\psi}_\alpha^\dagger(\tilde{q}_\alpha) \right) \\ - \left(\langle \theta(\varphi), \eta(\varphi') \rangle + \sum_{\alpha=1}^{\Lambda} \langle \tilde{\theta}^\alpha(\tilde{p}_\alpha), \tilde{\eta}_\alpha(\tilde{q}_\alpha) \rangle \right).$$

First and second part is convex due to linearity of derivative, the independence of parameters given by condition C2, and the Hessians of potential functions $\psi_0, \tilde{\psi}_1, \dots, \tilde{\psi}_\Lambda$ are positive semi-definite. The third part is a sum of inner products, which is again an inner product on the product parameter space in $\Xi_0 \times \Xi_1 \times \dots \times \Xi_\Lambda$.

Recall that the parameters of the mixture distribution $\tilde{p}(x, \zeta) \in \mathcal{L}_V$ are collected in $(\zeta_i)_{i=1}^d := (\theta^0, \tilde{\theta}^1, \dots, \tilde{\theta}^\Lambda) \in \Xi_0 \times \Xi_1 \times \dots \times \Xi_\Lambda$, where $d := \dim(\Xi_0 \times \Xi_1 \times \dots \times \Xi_\Lambda)$.⁵

By linearity of derivative and condition C2, the g^D -coordinate dual to $(\zeta_i)_{i=1}^d$ is given by $(\eta_0, \tilde{\eta}_1, \dots, \tilde{\eta}_\Lambda) := \left(\frac{\partial}{\partial \zeta_i} \left(\psi_0(\varphi) + \sum_{\alpha=1}^{\Lambda} \tilde{\psi}_\alpha(\tilde{p}_\alpha) \right) \right)$. Furthermore, the dual potential of $\left(\psi_0(\varphi) + \sum_{\alpha=1}^{\Lambda} \tilde{\psi}_\alpha(\tilde{p}_\alpha) \right)$ is given by the following Legendre-Fenchel transformation:

⁵ Recall that coordinates θ^0 of mixture coefficients are not pulled-back.

$$\begin{aligned}
& \left(\psi_0(\varphi) + \sum_{\alpha} \tilde{\psi}_{\alpha}(\tilde{p}_{\alpha}) \right)^{\dagger} = \\
& \sup_{\varphi \in S_0, \tilde{p}_{\alpha} \in \tilde{S}_{\alpha}} \left\{ \left(\langle \theta^0(\varphi), \eta_0(\varphi') \rangle + \sum_{\alpha=1}^{\Lambda} \langle \tilde{\theta}^{\alpha}(\tilde{p}_{\alpha}), \tilde{\eta}_{\alpha}(\tilde{q}_{\alpha}) \rangle \right) - \left(\psi_0(\varphi) + \sum_{\alpha} \tilde{\psi}_{\alpha}(\tilde{p}_{\alpha}) \right) \right\} \\
& = \sup_{\varphi \in S_0, \tilde{p}_{\alpha} \in \tilde{S}_{\alpha}} \left\{ \left(\langle \theta^0(\varphi), \eta_0(\varphi') \rangle - \psi_0(\varphi) \right) + \sum_{\alpha=1}^{\Lambda} \left(\langle \tilde{\theta}^{\alpha}(\tilde{p}_{\alpha}), \tilde{\eta}_{\alpha}(\tilde{q}_{\alpha}) \rangle - \tilde{\psi}_{\alpha}(\tilde{p}_{\alpha}) \right) \right\} \\
& = \sup_{\varphi \in S_0} \left\{ \langle \theta^0(\varphi), \eta_0(\varphi') \rangle - \psi_0(\varphi) \right\} + \sum_{\alpha=1}^{\Lambda} \sup_{\tilde{p}_{\alpha} \in \tilde{S}_{\alpha}} \left\{ \langle \tilde{\theta}^{\alpha}(\tilde{p}_{\alpha}), \tilde{\eta}_{\alpha}(\tilde{q}_{\alpha}) \rangle - \tilde{\psi}_{\alpha}(\tilde{p}_{\alpha}) \right\} \\
& = \psi_0^{\dagger}(\varphi') + \sum_{\alpha=1}^{\Lambda} \tilde{\psi}_{\alpha}^{\dagger}(\tilde{q}_{\alpha}) \quad .
\end{aligned}$$

The third equality follows from the functional independence of $\varphi \in S_0$ and \tilde{p}_{α} 's in \tilde{S}_{α} . Hence the Legendre-Fenchel transform of the first component of D is exactly the second component of D , therefore D is the canonical divergence of $(\mathcal{L}_V, g_0 \oplus \bigoplus_{\alpha=1}^{\Lambda} \tilde{g}_{\alpha}, \nabla^0 \oplus \bigoplus_{\alpha=1}^{\Lambda} \tilde{\nabla}^{\alpha}, \nabla^{0*} \oplus \bigoplus_{\alpha=1}^{\Lambda} \tilde{\nabla}^{\alpha*})$.

Finally we discuss generalizations of the constructions discussed in this section.

4.5. Inheriting densities of unbounded support

Whilst the previous discussion allows us to inherit families of distributions supported in open subsets of Euclidean spaces \mathbb{R}^n , it can be extended to inherit families of distributions with unbounded support over \mathbb{R}^n .

One way is to inherit family of distributions with unbounded support by applying the orientation-preserving bundle morphism construction twice.

Let $x \in M$ be arbitrary. The first bundle morphism is constructed via a diffeomorphism between \mathbb{R}^n (linearly isomorphic to tangent spaces $T_x M$ of M) to star-shaped open neighbourhoods V_x about the origin of tangent spaces $T_x M$ of M .

Using the bundle morphism construction discussed in the beginning of this section, this allows us to construct a family of distributions supported V_x .

The second bundle morphism is constructed via local diffeomorphism $\exp_x : V_x \subset T_x M \rightarrow M$ from a (star-shaped) open subset $V_x \subset T_x M \cong \mathbb{R}^n$ to $\exp_x(V_x) \subset M$ as discussed in the beginning of this section. We discuss this more formally as follows:

Let M be a Riemannian manifold and let $x \in M$ be arbitrary. Consider a metric ball $B(x, \epsilon_x) \subset M$.

Metric balls $B(x, \epsilon_x)$ in M are geodesically convex, hence $V_x := \exp_x^{-1}(B(x, \epsilon_x))$ is open and star-shaped in $T_x M \cong \mathbb{R}^n$ about $\vec{0} \in T_x M$.

By Gonnord and Tosel [20], there exists a diffeomorphism $f_x : V_x \rightarrow \mathbb{R}^n$, and by lemma 2, we can assume f_x to be orientation-preserving.

Let S be a family of probability densities supported in \mathbb{R}^n . We can induce a family of probability densities \tilde{S} on M by the following composition of orientation-preserving bundle morphisms.

⁶ One could, for example obtain it via an element of an orientation-preserving open cover of M by metric balls: $\{\rho_{\alpha}(W_{x_{\alpha}}) = B(x_{\alpha}, \epsilon_{x_{\alpha}})\}_{x_{\alpha} \in M}$ discussed in lemma 1

$$\begin{array}{ccccc}
 S \subset \text{Prob}(\mathbb{R}^n) & \xrightarrow{f_x^*} & \text{Prob}(V_x) & \xrightarrow{\exp_x^{-1*}} & \tilde{S} \subset \text{Prob}(M) \\
 \downarrow & & \downarrow & & \downarrow \\
 \mathbb{R}^n & \xleftarrow{f_x} & V_x \subset T_x M & \xleftrightarrow[\exp_x]{\exp_x^{-1}} & \exp_x(V_x) = (B(x, \epsilon_x)) \subset M
 \end{array}$$

Alternatively, we may also consider the family of truncated distributions $\bar{S} \subset \text{Prob}(V_x)$ supported in V_x obtained from $S \subset \text{Prob}(\mathbb{R}^n)$ by:

$$\bar{S} := \left\{ \frac{p}{\int_{V_x} p} \mid p \in S \right\} \subset \text{Prob}(V_x) \quad .$$

4.6. Densities on general M

In the previous discussion, we assumed M to be a smooth Riemannian manifold. In this subsection we discuss briefly how the approach can be extended easily to the case where M is a smooth topological manifold with no pre-defined Riemannian structure. Indeed, a Riemannian structure is only required for lemma 1 to construct an open cover by metric balls of totally bounded set V of M . Since there is no additional conditions on the Riemannian structure, it can be arbitrarily defined.

Since smooth topological manifolds are paracompact, there always exist a Riemannian metric. Here we discuss a specific metric inherited via a given orientation-preserving open cover.

Let $\{(\rho_\alpha, U_\alpha)\}$ denote once again an orientation-preserving open cover of M .

Let h_α be a positive definite 2-form on U_α . Then $(\rho_\alpha^{-1*} h_\alpha) : T_{x_\alpha} M \times T_{x_\alpha} M \rightarrow \mathbb{R}$ is the pulled-back metric on each tangent space for $x_\alpha \in \rho_\alpha(U_\alpha)$. A Riemannian metric on M can thus be constructed via partition of unity subordinate to regular refinement of $\{U_\alpha\}$.

5. Example

Finally we illustrate our approach with the following example on unit 2-sphere \mathbb{S}^2 .

5.1. Local bivariate Gaussian distribution on \mathbb{S}^2

Let \mathbb{S}^n denote the unit n -sphere in \mathbb{R}^{n+1} :

$$\mathbb{S}^n := \{x \in \mathbb{R}^{n+1} \mid \|x\| = 1\} \quad .$$

For simplicity, we will let $n = 2$ for the rest of the discussion. Consider the "northern" hemisphere denoted by $V_1 = \{(x, y, z) \in \mathbb{S}^2 \mid z \geq 0\}$.

In the existing literature, discussed in Example 2, family of probability densities on V_1 can be inherited via the following bundle morphism:

$$\begin{array}{ccc}
 S_{U_1} \subset \text{Prob}(U_1) & \xrightarrow{\log_N^*} & \text{Prob}(V_1) \\
 \downarrow & & \downarrow \\
 U_1 \subset T_N \mathbb{S}^2 & \xleftrightarrow[\exp_N]{\log_N} & V_1 = \exp_N(U_1) \subset \mathbb{S}^2
 \end{array}$$

where $N = (0, 0, 1) \in V_1$ denote the north pole on the sphere \mathbb{S}^2 , $\exp_N : T_N \mathbb{S}^2 \rightarrow \mathbb{S}^2$ the Riemannian exponential map centred at N , and $U_1 \subset T_N \mathbb{S}^2$ a neighbourhood about $\vec{0} \in T_N \mathbb{S}^2$ locally diffeomorphic to V_1 under \exp_N . Family of probability volume densities on V_1 is thus given by the following map:

$$p \in S_{U_1} \subset \text{Prob}(U_1) \mapsto \log_N^* p \in \text{Prob}(V_1) \quad .$$

In practice this approach may be computationally expensive, as finding the inverse of Riemannian exponential map involves solving a second order differential equation: geodesic equation. In the 2-sphere it is given by the following: Let $x \in \mathbb{S}^2$ and $v \in T_x \mathbb{S}^2$ be arbitrary. Geodesic $\gamma(t)$ on \mathbb{S}^2 with $\gamma(0) = x, \gamma'(0) = v$ is given by:

$$\begin{aligned} \gamma : [a, b] &\rightarrow \mathbb{S}^2 \\ t &\mapsto \cos(t\|v\|)x + \sin(t\|v\|) \frac{v}{\|v\|} . \end{aligned}$$

In this work, our approach allows us to consider the alternative maps that also preserve the geometrical structure of S_{U_1} . In particular we may consider any arbitrary diffeomorphism (by lemma 1), and in this case we consider the restricted spherical coordinates on V_1 from $\{1\} \times U_1$, where $U_1 := (0, \frac{\pi}{2}) \times (0, 2 \cdot \pi)$, is given by:

$$\begin{aligned} \rho : \{1\} \times U_1 &\rightarrow V_1 \subset \mathbb{S}^2 \subset \mathbb{R}^3 \\ (1, \varphi, \theta) &\rightarrow (1 \cdot \sin \varphi \cos \theta, 1 \cdot \sin \varphi \sin \theta, 1 \cdot \cos \varphi) . \end{aligned} \quad (25)$$

This map is the restriction of 3-dimensional spherical coordinate parametrization on the unit ball, which is an orientation-preserving diffeomorphism on $[0, 1] \times U_1$:

$$\begin{aligned} \tilde{\rho} : [0, 1] \times U_1 &\rightarrow \mathbb{R}^3 \\ (r, \varphi, \theta) &\rightarrow (r \cdot \sin \varphi \cos \theta, r \cdot \sin \varphi \sin \theta, r \cdot \cos \varphi) . \end{aligned}$$

Since $\tilde{\rho}$ is an orientation-preserving on $[0, 1] \times U_1$, so is $\rho(1, \varphi, \theta) = \tilde{\rho}(r, \varphi, \theta)|_{r=1}$ on $\{1\} \times U_1$ [8]. Moreover, ρ is a bijective immersion⁷ hence a diffeomorphism on $\{1\} \times U_1$ with inverse $\rho^{-1}(x, y, z) = \tilde{\rho}^{-1}(x, y, z)|_{z=\sqrt{1-x^2-y^2}}$, where $\tilde{\rho}^{-1}$ denote the inverse of $\tilde{\rho}$ given by the following:

$$\begin{aligned} \tilde{\rho}^{-1} : \mathbb{R}^3 &\rightarrow [0, 1] \times U_1 \\ (x, y, z) &\mapsto (r, \varphi, \theta) = \left(\sqrt{x^2 + y^2 + z^2}, \tan^{-1} \left(\frac{y}{x} \right), \cos^{-1} \left(\frac{z}{\sqrt{x^2 + y^2 + z^2}} \right) \right) . \end{aligned}$$

The matrix $D\tilde{\rho}^{-1}$ is given by:

$$\begin{bmatrix} \frac{x}{\sqrt{x^2+y^2+z^2}} & \frac{y}{\sqrt{x^2+y^2+z^2}} & \frac{z}{\sqrt{x^2+y^2+z^2}} \\ \frac{zx}{(x^2+y^2+z^2)^{3/2}} \frac{1}{\sqrt{-\frac{z^2}{x^2+y^2+z^2}+1}} & \frac{zy}{(x^2+y^2+z^2)^{3/2}} \frac{1}{\sqrt{-\frac{z^2}{x^2+y^2+z^2}+1}} & -1 \left(\frac{1}{\sqrt{x^2+y^2+z^2}} - \frac{z^2}{(x^2+y^2+z^2)^{3/2}} \right) \frac{1}{\sqrt{-\frac{z^2}{x^2+y^2+z^2}+1}} \\ -\frac{y}{x^2} \left(\frac{y^2}{x^2} + 1 \right)^{-1} & \frac{1}{x} \left(\frac{y^2}{x^2} + 1 \right)^{-1} & 0 \end{bmatrix}$$

⁷ Since rank of ρ_* is $2 = \dim(U_1)$.

and the corresponding determinant is given by the following, which is always positive for $x, y, z \in \mathbb{R}$:

$$\frac{1}{x^2 + y^2 + z^2} \frac{1}{\sqrt{\frac{x^2 + y^2}{x^2 + y^2 + z^2}}} .$$

Therefore $\tilde{\rho}^{-1}$ is orientation-preserving, and so is ρ^{-1} .⁸ Since $\sqrt{x^2 + y^2 + z^2} = 1$ is a constant in $\mathbb{S}^2 \hookrightarrow \mathbb{R}^3$, the determinant of $D\rho^{-1}$ becomes:

$$\frac{1}{\sqrt{x^2 + y^2}} ,$$

which is always positive on \mathbb{S}^2 .

In this example we will inherit truncated bivariate Gaussian distributions on the rectangle $U_1 = (0, \frac{\pi}{2}) \times (0, 2 \cdot \pi) \cong \{1\} \times U_1$:

$$p(\varphi, \theta) := \frac{1}{A_p} \frac{1}{c_p} e^{((\varphi, \theta) - \mu_{U_1})^\top \Sigma_{U_1} ((\varphi, \theta) - \mu_{U_1})} ,$$

where $A_p := \int_{U_1} p(\varphi, \theta) d\varphi d\theta$, and c_p is the standard scaling factor of bivariate normal distributions given by $2\pi\sqrt{|\Sigma_{U_1}|}$. Let $\bar{S}_{U_1} := \{p(\varphi, \theta)\}$ denote the family of truncated Gaussian distributions supported in U_1 .

Induced distributions \tilde{S}_{V_1} on $V_1 \subset \mathbb{S}^2$ via pullback bundle morphism induced by ρ^{-1} is given by the following commuting diagram (c.f. Section 4.1):

$$\begin{array}{ccc} \bar{S}_{U_1} \subset \text{Prob}(U_1) & \xrightarrow{\rho^{-1*}} & \tilde{S}_{V_1} \subset \text{Prob}(\mathbb{S}^2) \\ \downarrow & & \downarrow \\ U_1 & \xleftrightarrow[\rho]{\rho^{-1}} & V_1 = \rho(U_1) \subset \mathbb{S}^2 \end{array}$$

The induced probability density $\tilde{p} \in \tilde{S}_{V_1} := \rho^{-1*} \bar{S}_{U_1}$ on $\rho(U_1) \subset \mathbb{S}^2$ is given by (see equation (4)):

$$\tilde{p} := \rho^{-1*} p = p \circ \rho^{-1} : \mathbb{S}^2 \rightarrow \mathbb{R} , \quad \text{for } p \in \bar{S}_{U_1} .$$

The close form expression of $\tilde{p} \in \tilde{S}_{V_1}$ is thus given by the following expression:

$$\begin{aligned} \tilde{p}(x, y, z) &:= \frac{1}{A_p} \frac{1}{c_p} e^{(\rho^{-1}(x, y, z) - \mu_{U_1})^\top \Sigma_{U_1} (\rho^{-1}(x, y, z) - \mu_{U_1})} \cdot \left| \det D\rho^{-1} \right|_{r=1} \\ &= \frac{1}{A_p} \frac{1}{c_p} e^{(\rho^{-1}(x, y, z) - \mu_{U_1})^\top \Sigma_{U_1} (\rho^{-1}(x, y, z) - \mu_{U_1})} \cdot \frac{1}{\sqrt{x^2 + y^2}} , \end{aligned}$$

where $r := x^2 + y^2 + z^2 = 1$ in \mathbb{S}^2 . Notice since $z = \sqrt{1 - x^2 - y^2}$ on \mathbb{S}^2 , the change of measure only depends on x, y .

Furthermore, since \bar{S}_{U_1} is an exponential family, which is an α -affine manifold with $\alpha = 1$ and ρ is a (local) diffeomorphism hence a sufficient statistics for \tilde{S}_{V_1} , by example 1, \tilde{S}_{V_1} is an 1-affine manifold

⁸ By a slight modification to the proof of Lemma 13.20 of [8].

as well. The canonical divergence on \bar{S}_{U_1} is given by the Kullback-Leibler divergence D_{KL} , hence the induced divergence \bar{D}_1 on \bar{S}_{V_1} can thus be computed by:

$$\bar{D}_1(\rho^{-1*}p, \rho^{-1*}q) = D_{KL}(p, q) \quad , \quad \text{for } p, q \in \bar{S}_{U_1} \quad . \quad (26)$$

As an illustration of the construction of pulled-back dualistic structure and induced divergence discussed in Remark 2 Section 3.

Let $V_2 = \{(x, y, z) \in \mathbb{S}^2 | y \geq 0\}$ denote another closed subset of \mathbb{S}^2 and let $V = V_1 \cup V_2$ denote a totally bounded subset of \mathbb{S}^2 .

Analogous to the previous construction, we once again consider restricted spherical coordinate in equation (25): $\rho : \{1\} \times U_2 \rightarrow V_2$ on V_2 from $\{1\} \times U_2$, where $U_2 := (0, \pi) \times (0, \pi)$. ρ is an orientation-preserving diffeomorphism on $\{1\} \times U_2$. Let \bar{S}_{U_2} denote family of truncated Gaussian distributions supported in U_2 , and let $\bar{S}_{V_2} := \rho^{-1*}\bar{S}_{U_2}$.

Let $S_0 := \{\{\varphi_1, \varphi_2\} | \varphi_1, \varphi_2 \in [0, 1], \varphi_1 + \varphi_2 = 1\}$ denote the set of mixture coefficients. Consider the set of mixture densities \mathcal{L}_V given by:

$$\mathcal{L}_V := \{\varphi_1 \cdot \tilde{p}_1 + \varphi_2 \cdot \tilde{p}_2 | \{\varphi_1, \varphi_2\} \in S_0, \tilde{p}_1 \in \bar{S}_{V_1}, \tilde{p}_2 \in \bar{S}_{V_2}\} \quad .$$

Since $\bar{S}_{V_1}, \bar{S}_{V_2}$ are both dually flat, by the previous section, the canonical divergence D_V of dually flat manifold $\mathcal{L}_V = S_0 \times \bar{S}_{V_1} \times \bar{S}_{V_2}$ can thus be evaluated by equation (7):

$$D_V(\tilde{p}, \tilde{q}) = \sum_{\alpha=1}^2 \varphi_\alpha \log \left(\frac{\varphi_\alpha}{\varphi'_\alpha} \right) + \sum_{\alpha=1}^2 \bar{D}_\alpha(\tilde{p}_\alpha, \tilde{q}_\alpha) \quad ,$$

where $\tilde{p} := \sum_{\alpha=1}^2 \varphi_\alpha \cdot \tilde{p}_\alpha, \tilde{q} := \sum_{\alpha=1}^2 \varphi_\alpha \cdot \tilde{q}_\alpha \in \mathcal{L}_V$, and \bar{D}_α is the induced divergence on \bar{S}_{V_α} for $\alpha = 1, 2$ respectively given by equation (26).

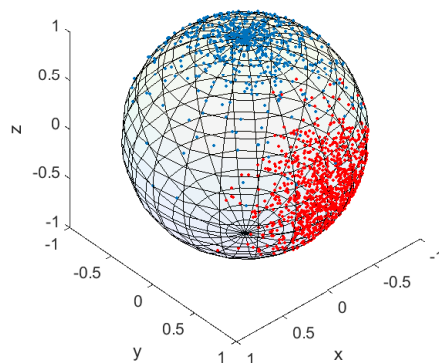


Figure 1. A sample from a mixture of induced Gaussian distribution on $V = V_1 \cup V_2$

$$\varphi_1 = \frac{1}{2}; \quad \mu_{U_1} = (0, 0); \quad \Sigma_{U_1} = \begin{bmatrix} \pi & 0 \\ 0 & \frac{\pi}{10} \end{bmatrix}$$

$$\varphi_2 = \frac{1}{2}; \quad \mu_{U_2} = \left(\frac{\pi}{2}, \frac{\pi}{2} \right); \quad \Sigma_{U_2} = \begin{bmatrix} \pi/32 & 0 \\ 0 & \frac{\pi}{32} \end{bmatrix}$$

Acknowledgments: The authors would like to thank Joshua Knowles for the valuable input and discussions on the poster presented in TGSI 2017.

Author Contributions: Both authors contributed equally to the design of the research. The research was carried out by both authors, and the manuscript was subsequently prepared together. Both authors read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Elgammal, A.; Lee, C.S. The role of manifold learning in human motion analysis. In *Human Motion*; Springer, 2008; pp. 25–56.
2. Mardia, K.V.; Jupp, P.E. *Directional statistics*; Vol. 494, John Wiley & Sons, 2009.
3. Said, S.; Bombrun, L.; Berthoumieu, Y.; Manton, J.H. Riemannian Gaussian distributions on the space of symmetric positive definite matrices. *IEEE Transactions on Information Theory* **2017**, *63*, 2153–2170.
4. Pennec, X. Probabilities and statistics on Riemannian manifolds: A geometric approach. PhD thesis, INRIA, 2004.
5. Amari, S.; Nagaoka, H. *Methods of Information Geometry*; Vol. 191, *Translations of Mathematical monographs*, Oxford University Press, 2000.
6. Lang, S. *Fundamentals of Differential Geometry*; Vol. 191, Springer Science & Business Media, 2012.
7. Bruveris, M.; Michor, P.W. Geometry of the Fisher-Rao metric on the space of smooth densities on a compact manifold. *arXiv preprint arXiv:1607.04550* **2016**.
8. Lee, J.M. *Smooth manifolds*; Springer, 2003.
9. Khesin, B.; Lenells, J.; Misiołek, G.; Preston, S. Geometry of diffeomorphism groups, complete integrability and geometric statistics. *Geometric and Functional Analysis* **2013**, *23*, 334–366.
10. Ay, N.; Jost, J.; Vân Lê, H.; Schwachhöfer, L. *Information Geometry*; Vol. 64, Springer, 2017.
11. Ay, N.; Jost, J.; Vân Lê, H.; Schwachhöfer, L. Information Geometry and Sufficient Statistics. *Probability Theory and Related Fields* **2015**, *162*, 327–364.
12. Nomizu, K.; Sasaki, T. *Affine differential geometry: geometry of affine immersions*; Cambridge University Press, 1994.
13. Shima, H. *The geometry of Hessian structures*; World Scientific, 2007.
14. Calin, O.; Udrişte, C. *Geometric modeling in probability and statistics*; Springer, 2014.
15. Eguchi, S.; others. Geometry of minimum contrast. *Hiroshima Mathematical Journal* **1992**, *22*, 631–647.
16. Matumoto, T. Any statistical manifold has a contrast function—on the C^3 -functions taking the minimum at the diagonal of the product manifold. *Hiroshima Math. J.* **1993**, *23*, 327–332.
17. Petersen, P. *Riemannian geometry*; Vol. 171, Springer, 2006.
18. Sakai, T. *Riemannian Geometry*; Vol. 149, American Mathematical Soc., 1996.
19. Do Carmo, M.P.; Flaherty Francis, J. *Riemannian Geometry*; Vol. 115, Birkhäuser Boston, 1992.
20. Gonnord, S.; Tosel, N. *Calcul différentiel*; Ellipses, 1998.