INTERNATIONAL ELECTRONIC JOURNAL OF GEOMETRY
VOLUME 13 NO. 2 PAGE 144–151 (2020)
DOI: HTTPS://DOI.ORG/10.36890/IEJG.689702

Wasserstein Riemannian Geometry on Statistical Manifold

Carlos Ogouyandjou* **and Nestor Wadagni**

(Communicated by Murat Tosun)

ABSTRACT

In this paper, we study some geometric properties of statistical manifold equipped with the Riemannian Otto metric which is related to the $\mathcal{L}^2$-Wasserstein distance of optimal mass transport. We construct some α-connections on such manifold and we prove that the proposed connections are torsion-free and coincide with the Levi-Civita connection when $\alpha = 0$. In addition, the exponentialy families and the mixture families are shown to be respectively (1)-flat and (-1)-flat.

Keywords: *Statistical manifold; Riemannian metric; Otto metric; α-connections; Wasserstein Riemannian space; flatness.*
AMS Subject Classification (2020): *15B48 ; 53C23; 53C25 ; 60D05.*

1. Introduction

Information geometry started as the investigation of the differential geometric stucture of some set of probability distributions which constitutes a statistical manifold. Since the seminal work of Rao [11] where Fisher information geometry is viewed as a Riemannian metric on a space of probability distributions, it became obvious that as differentiable manifold, a space of probability distributions can be equipped with a multitude of Riemannian metrics that are not necessarily the Fisher metric. Considering the Riemannian structure obtained by the Fisher information on a statistical manifold, Amari [2] defines a one-parameter family of affine connections called α-connections. Hence α-connections have become key tools in information geometry and have been widely investigated by several authors such as Gbaguidi et al. [7] who constructed a family of α-connections on a Hilbert bundle of generalized statistical manifold.

In this paper we are interested in statistical manifold equipped with the Wasserstein metric which is related to optimal transport. Kantorovich and Rubinstein [8] stated that the Wasserstein metric can be taken as a reasonable distance on spaces of random variables or of probability distributions. However, explicit calculations based on that metric seems to be somewhat difficult to perform. Lott [9] showed that the Riemannian Otto metric related to Wasserstein metric makes the calculations on Wasserstein space easier. We make use of the Otto metric to investigate the Wasserstein Riemannian geometry on statistical manifold.

Let $\mathcal{M}$ be a set of probability densities endowed with the Otto Riemannian metric. We construct on $\mathcal{M}$ a family $\nabla^{(\alpha)}$ of torsion-free α-connections that is exactly the Levi-Civita connection on $\mathcal{M}$ when $\alpha = 0$. We also find out that the exponential families and the mixture families are respectively (1)-flat and (-1)-flat. The rest of the paper is organized as follows: we recall some preliminaries on α-connections in section 2, and we present useful results on Otto metric and Wasserstein metric in section 3. Finally, the main results are given in section 4.

2. Preliminary remarks on α-connections

For some integer $d \geq 1$, let $\mathcal{X}$ be a non-empty subset of $\mathbb{R}^d$ and $\mathcal{M}$ be a family of probability distributions on $\mathcal{X}$. Each element of $\mathcal{M}$, can be identified with $\theta = (\theta_1, \cdots, \theta_n) \in \Theta$ a subset of $\mathbb{R}^n$ and the mapping $\theta \mapsto p_\theta$ is

Received : *15-February-2020,* ***Accepted :*** *16-August-2020*
*** Corresponding author***

injective. $\mathcal{M}$ is a C^∞ differentiable manifold.

Example 2.1. $\mathcal{X} = \mathbb{R}$, $n = 2$, $\theta = (\mu, \sigma)$, $\Theta = \{(\mu, \sigma) : \mu \in \mathbb{R}, \sigma \in \mathbb{R}_+^*\}$

$$p(x, \theta) = \frac{1}{\sigma\sqrt{2\pi}} exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

Put $\ell(.;\theta) = \log p(., \theta)$. $\frac{\partial \ell(.;\theta)}{\partial \theta^i}$ for $i = 1, \cdots, n$ are the scores functions.
The tangent space $\top_\theta(\mathcal{M})$ can be identifed with $\tilde{\top}_\theta(\mathcal{M})$, the vector space spanned by $\frac{\partial \ell(x;\theta)}{\partial \theta^i}$, and endowed with the inner product $\langle \tilde{X}, \tilde{Y} \rangle_\theta = \mathbb{E}_\theta[\tilde{X}\tilde{Y}]$. The mapping

$$\sum_i a_i \frac{\partial}{\partial \theta^i} \mapsto \sum_i a_i \frac{\partial \ell(x;\theta)}{\partial \theta^i}$$

defines an isometry between $\top_\theta \mathcal{M}$ and $\tilde{\top}_\theta(\mathcal{M})$, (see[12]).

Definition 2.1. The Fisher information metric

The Fisher information matrix of $\mathcal{M}$ at θ is the $n \times n$ matrix $G(\theta) = (\tilde{g}_{ij}(\theta))$ defined by :

$$\tilde{g}_{ij}(\theta) := \mathbb{E}_\theta[\partial_i \ell(X, \theta) \partial_j \ell(X, \theta)] = \int_{\mathcal{X}} \partial_i \ell(x, \theta) \partial_j \ell(x, \theta) p(x; \theta) dx$$

where $\partial_i := \frac{\partial}{\partial \theta_i}$ and $\ell(x, \theta) = \log p(x; \theta)$. In particular, when $n = 1$, we call this the Fisher information.
The inner product of the natural basis of the coordinate system $(\theta_1, \cdots, \theta_n)$

$$\langle \partial_i, \partial_j \rangle = \tilde{g}_{ij}$$

uniquely determines a Riemannian metric $\tilde{g} = \langle \cdot, \cdot \rangle$ such that for all $\theta \in \Theta$, and for all $X, Y \in \top_\theta \mathcal{M}$; $\tilde{g}_\theta(X, Y) = \langle X, Y \rangle_\theta = \mathbb{E}_\theta[(X\ell)(Y\ell)]$. $\tilde{g}$ is called Fisher metric or alternatively, the information metric.

Definition 2.2. An affine connection ∇ on a differentiable manifold $\mathcal{M}$ is a mapping

$$\nabla : \mathcal{X}(\mathcal{M}) \times \mathcal{X}(\mathcal{M}) \to \mathcal{X}(\mathcal{M})$$

which is denoted by $(X, Y) \to \nabla_X Y$ and which satisfies the following properties:

- $\nabla_{fX+gY} Z = f \nabla_X Z + g \nabla_Y Z$
- $\nabla_X (Y + Z) = \nabla_X Y + \nabla_X Z$
- $\nabla_X (fY) = f \nabla_X Y + X(f) Y$ in which $X, Y, Z \in \mathcal{X}(\mathcal{M})$ and $f, g \in C^\infty(\mathcal{M})$.

Theorem 2.1. *[6] Given a Riemannian manifold $(\mathcal{M}, g)$, there exists a unique affine connection ∇ on $\mathcal{M}$ satisfing the conditions:*

- *∇ is symmetric.*
- *∇ is compatible with the Riemannian metric g.*

This affine connection is the Levi-Civita connection on the manifold $(\mathcal{M}, g)$.

In a coordinate system (U, θ), the function $\overset{\circ}{\Gamma}{}^k_{ij}$ defined on U by $\nabla_{\partial_i} \partial_j = \sum_k \Gamma^k_{ij} \partial_k$ are called the Christoffel symbol of the the Levi-Civita connection and we have

$$\overset{\circ}{\Gamma}{}^k_{ij} = \frac{1}{2}\left(\frac{\partial g_{jm}}{\partial \theta^i} + \frac{\partial g_{mi}}{\partial \theta^j} - \frac{\partial g_{ij}}{\partial \theta^m}\right) g^{mk}. \tag{2.1}$$

Amari[2] considers the function $\Gamma^{(\alpha)}_{ij,k}$ which maps each point θ to the following value:

$$\left(\Gamma^{(\alpha)}_{ij,k}\right)_\theta := \mathbb{E}_\theta\left[\left(\partial_i \partial_j \ell(X, \theta) + \frac{1-\alpha}{2} \partial_i \ell(X, \theta) \partial_j \ell(X, \theta)\right)(\partial_k \ell(X, \theta))\right]$$

where α is some arbitrary real number. The α-connection $\nabla^{(\alpha)}$,which is an affine connection, is defined by

$$\langle \nabla^{(\alpha)}_{\partial_i} \partial_j, \partial_k \rangle = \Gamma^{(\alpha)}_{ij,k},$$

where $g = \langle \cdot, \cdot \rangle$ is the Fisher metric and $\nabla^{(\alpha)}_{\partial_i} \partial_j$ is the α covariant derivative of ∂_j in the direction of ∂_i.
Next, we recall some important results on the Otto metric which is a Riemannian metric on the Wasserstein space.

3. Otto metric

3.1. Wasserstein metric

Let $(\mathcal{X}, \mu)$ and $(\mathcal{Y}, \nu)$ be two probability spaces. A coupling of (μ, ν) is a random vector (X, Y) such that the law of X is μ and the law of Y is ν. By abuse of language, the law of (X, Y) is also called a coupling of (μ, ν). We denote by $\Pi(\mu, \nu)$ the set of coupling of (μ, ν).

Definition 3.1. Let $\mathcal{X}$ be a subset of $\mathbb{R}^n$, $n \in \mathbb{N}^*$ and let $p \in [1; \infty[$. For any two probability measures μ, ν on $\mathcal{X}$, the Wasserstein distance of order p between μ and ν is defined by:

$$W_p(\mu, \nu) = \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}} \|x - y\|^p d\pi(x, y) \right)^{1/p}. \tag{3.1}$$

Definition 3.2. Let $P(\mathcal{X})$ be the set of probability measures on $\mathcal{X}$. The Wasserstein space of order p, $p \in [1, \infty[$ is defined as

$$P_p(\mathcal{X}) = \left\{ \mu \in P(\mathcal{X}); \int_{\mathcal{X}} \|x\|^p d\mu(x) < +\infty \right\}. \tag{3.2}$$

W_p defines a (finite) distance on $P_p(\mathcal{X})$. For more details on Wasserstein space see [13].

3.2. Otto metric

We consider an n-dimensional regular statistical manifold $\mathcal{M} = \{p(\cdot; \theta); \theta = (\theta_1, \cdots, \theta_n) \in \Theta\}$ where Θ is an open subset of $\mathbb{R}^n$ and the mapping $\theta \mapsto p_\theta$ is injective.

Motivated by the study of a class of partial differential equation, in [10], Otto considered an inner product defined on smooth functions of the θ-fiber, $\top_\theta \mathcal{M}$ of the tangent bundle, as

$$(u, v) \mapsto \int_{\mathcal{X}} \partial_x u(x) \cdot \partial_x v(x) p(x; \theta) dx \tag{3.3}$$

where ∂_x is the gradient function. Then, this inner produit defines on $\mathcal{M}$ a Riemannian metric so called Otto metric g, with coordinate functions:

$$g_{ij} = \int_{\mathcal{X}} \partial_x \frac{\partial \ell(x; \theta)}{\partial \theta^i} \cdot \partial_x \frac{\partial \ell(x; \theta)}{\partial \theta^j} p(x; \theta) dx = \mathbb{E}\left(\partial_{\theta_i} \ell \cdot \partial_{\theta_j} \ell \right). \tag{3.4}$$

The following theorem states that the Riemannian Otto metric is related to the wasserstein metric.

Theorem 3.1. *[9] Let $P^\infty(\chi) = \left\{ f : f \in C^\infty(\chi), f > 0, \int_\chi f(x) d(x) = 1 \right\}$. If $c : [0, 1] \to P^\infty(\mathcal{M})$ is a smooth immersed curve then its length $L(c)$ in the Wasserstein space $P_2(\chi)$ satisfies*

$$L(c) = \int_0^1 \langle c'(t), c'(t) \rangle^{\frac{1}{2}} dt$$

where

$$L(c) = \sup_{j \in \mathbb{N}} \sup_{0 = t_0 \le t_1 \le \cdots \le t_J = 1} \sum_{j=1}^{J} W_2(c(t_{j-1}), c(t_j)).$$

Proposition 3.1. *Let $\mathcal{M} = \{p(\cdot; \theta); \theta = (\theta_1, \cdots, \theta_n) \in \Theta\}$ be a statistical manifold endowed with the Otto metric g (3.4). For all $i, j, k \in \{1, \cdots, n\}$*

$$\begin{aligned} \frac{\partial g_{jk}}{\partial \theta^i} + \frac{\partial g_{ki}}{\partial \theta^j} - \frac{\partial g_{ij}}{\partial \theta^k} &= \mathbb{E}_\theta \left[2\partial_x \partial_{ij} \ell \partial_x \partial_k \ell \right. \\ &\quad + \partial_x \partial_j \ell \partial_x \partial_k \ell \partial_i \ell + \partial_x \partial_k \ell \partial_x \partial_i \ell \partial_j \ell \\ &\quad \left. - \partial_x \partial_j \ell \partial_x \partial_i \ell \partial_k \ell \right]. \end{aligned} \tag{3.5}$$

Proof. Taking the partial derivative of g_{ij} in Equation (3.4) with respect to θ_i, θ_j, θ_k yields the result. □

4. α-connection related to Otto metric

4.1. Construction of our α-connection

In the remainder, we consider an n-dimensional statistical manifold $\mathcal{M} = \{p(\cdot;\theta); \theta = (\theta_1, \cdots, \theta_n) \in \Theta\}$ where Θ is a open subset of $\mathbb{R}^n$ and the mapping $\theta \mapsto p_\theta$ is injective. We endowed $\mathcal{M}$ with the Riemannian Otto metric g which is related to Wasserstein distance. For any $\alpha \in \mathbb{R}$, $i, j, k \in \{1, \cdots, n\}$, we introduce the function $\Gamma^{(\alpha)}_{ij,k}$ which maps each point θ to the following value:

$$\Gamma^{(\alpha)}_{ij,k} = \mathbb{E}_\theta \left[\partial_x \partial_{ij} \ell \partial_x \partial_k \ell\right] + \frac{1-\alpha}{2} \top^{\alpha}_{ij,k} \tag{4.1}$$

where $\top^{\alpha}_{ij,k}$ is a tensor defined by :

$$\top^{\alpha}_{ij,k} = \mathbb{E}_\theta \left[\partial_x \partial_j \ell \partial_x \partial_k \ell \partial_i \ell\right] + \mathbb{E}_\theta \left[\partial_x \partial_k \ell \partial_x \partial_i \ell \partial_j \ell\right] - (1+\alpha)\mathbb{E}_\theta \left[\partial_x \partial_j \ell \partial_x \partial_i \ell \partial_k \ell\right]. \tag{4.2}$$

Let ϕ be the parameter of dimension n of some parametrization of $\mathcal{M}$, alternative to that indicated by θ. Coordinates of ϕ will be denoted by $\phi = (\phi_1, \cdots, \phi_n)$, and we write ∂_{ϕ_u} for $\frac{\partial}{\partial \phi_u}$ and $\theta_{i/u} = \frac{\partial \theta_i}{\partial \phi_u}$.

Lemma 4.1. *For any change of coordinate system $\top^{\alpha}_{ij,k}$ satisfies the equation $\top^{\alpha}_{uv,w} = \top^{\alpha}_{ij,k} \theta_{i/u} \theta_{j/v} \phi_{w/k}$.*

Proof. Using (4.2), we have

$$\begin{aligned} \top^{\alpha}_{ij,k} \theta_{i/u} \theta_{j/v} \phi_{w/k} &= \mathbb{E}_\theta \left[\partial_x \partial_j \ell \partial_x \partial_k \ell \partial_u \ell\right] \theta_{j/v} \phi_{w/k} \\ &\quad + \mathbb{E}_\theta \left[\partial_x \partial_k \ell \partial_x \partial_i \ell \partial_v \ell\right] \theta_{i/u} \phi_{w/k} \\ &\quad - (1+\alpha)\mathbb{E}_\theta \left[\partial_x \partial_j \ell \partial_x \partial_i \ell \partial_w \ell\right] \theta_{i/u} \theta_{j/v} \\ &=: I_1 + I_2 - (1+\alpha) I_3. \end{aligned} \tag{4.3}$$

We have

$$\begin{aligned} I_1 &= \mathbb{E}_\theta \left[\partial_x \partial_j \ell \partial_x \partial_k \ell \partial_u \ell\right] \theta_{j/v} \phi_{w/k} \\ &= \mathbb{E}_\theta \left[\partial_x \partial_v \ell \partial_x \partial_w \ell \partial_u \ell\right] \\ &\quad + \mathbb{E}_\theta \left[\partial_x \partial_v \ell \partial_u \ell \partial_w \ell\right] \left(\partial_x \theta_{j/v}\right) \theta_{j/v} \\ &\quad + \mathbb{E}_\theta \left[\partial_x \partial_w \ell \partial_u \ell \partial_v \ell\right] \left(\partial_x \phi_{w/k}\right) \phi_{w/k} \\ &\quad + \mathbb{E}_\theta \left[\partial_v \ell \partial_u \ell \partial_w \ell\right] \left(\partial_x \theta_{j/v}\right) \left(\partial_x \phi_{w/k}\right) \theta_{j/v} \phi_{w/k} \\ &= \mathbb{E}_\theta \left[\partial_x \partial_v \ell \partial_x \partial_w \ell \partial_u \ell\right] \end{aligned} \tag{4.4}$$

because $\partial_x \theta_{i/u} = \partial_x \theta_{j/v} = \partial_x \phi_{w/k} = 0$. Similarly, we deduce

$$\begin{aligned} I_2 &= \mathbb{E}_\theta \left[\partial_x \partial_w \ell \partial_x \partial_u \ell \partial_v \ell\right] \\ I_3 &= \mathbb{E}_\theta \left[\partial_x \partial_v \ell \partial_x \partial_u \ell \partial_w \ell\right]. \end{aligned}$$

Then $\top^{\alpha}_{uv,w} = \top^{\alpha}_{ij,k} \theta_{i/u} \theta_{j/v} \phi_{w/k}$. □

The following result based on the transformation law (see [1]) gives a characterization of affine connections on a Riemannian manifold.

Lemma 4.2. *On Riemannian manifold $(\mathcal{M}, g)$:*

(a) all affine connection ∇ with connection symbol Γ^k_{ij} (i.e.$\nabla_{\partial_i} \partial_j = \Gamma^k_{ij} \partial_k$) are of the form

$$\Gamma^k_{ij} = \overset{\circ}{\Gamma}{}^k_{ij} + S^k_{ij} \tag{4.5}$$

where $S_{ij,k}$ satisfied

$$S^w_{uv} = S^k_{ij} \theta_{i/u} \theta_{j/v} \phi_{w/k}$$

(b) any set of smooth function $\Gamma_{ij,k}$ on $\mathcal{M}$ which satisfies the law (4.5) constitutes the connection symbols of an affine connection on $\mathcal{M}$.

Proof. Let's first prove (a). Let ∇ an affine connection on M. Then

$$\begin{aligned}
\Gamma_{uv}^{w}\partial_w = \nabla_{\frac{\partial}{\partial\phi^u}}\frac{\partial}{\partial\phi^v} &= \nabla_{\frac{\partial\theta^i}{\partial\phi^u}\frac{\partial}{\partial\theta^i}}\left(\frac{\partial\theta^j}{\partial\phi^v}\frac{\partial}{\partial\theta^j}\right)\\
&= \frac{\partial\theta^i}{\partial\phi^u}\nabla_{\frac{\partial}{\partial\theta^i}}\left(\frac{\partial\theta^j}{\partial\phi^v}\frac{\partial}{\partial\theta^j}\right)\\
&= \frac{\partial\theta^i}{\partial\phi^u}\left[\frac{\partial\theta^j}{\partial\phi^v}\nabla_{\frac{\partial}{\partial\theta^i}}\frac{\partial}{\partial\theta^j}+\frac{\partial}{\partial\theta^i}\left(\frac{\partial\theta^j}{\partial\phi^v}\right)\frac{\partial}{\partial\theta^j}\right]\\
&= \frac{\partial\theta^i}{\partial\phi^u}\left[\frac{\partial\theta^j}{\partial\phi^v}\Gamma_{ij}^{k}\frac{\partial}{\partial\theta^k}+\frac{\partial}{\partial\theta^i}\left(\frac{\partial\theta^j}{\partial\phi^v}\right)\frac{\partial}{\partial\theta^j}\right]\\
&= \Gamma_{ij}^{k}\frac{\partial\theta^i}{\partial\phi^u}\frac{\partial\theta^j}{\partial\phi^v}\frac{\partial\phi^w}{\partial\theta^k}\frac{\partial}{\partial\phi^w}+\frac{\partial^2\theta^j}{\partial\phi^u\partial\phi^v}\frac{\partial}{\partial\theta^j}\\
&= \Gamma_{ij}^{k}\frac{\partial\theta^i}{\partial\phi^u}\frac{\partial\theta^j}{\partial\phi^v}\frac{\partial\phi^w}{\partial\theta^k}\frac{\partial}{\partial\phi^w}+\frac{\partial^2\theta^k}{\partial\phi^u\partial\phi^v}\frac{\partial\phi^w}{\partial\theta^k}\frac{\partial}{\partial\phi^w}\\
&= \left[\Gamma_{ij}^{k}\frac{\partial\theta^i}{\partial\phi^u}\frac{\partial\theta^j}{\partial\phi^v}\frac{\partial\phi^w}{\partial\theta^k}+\frac{\partial^2\theta^w}{\partial\phi^u\partial\phi^v}\frac{\partial\phi^w}{\partial\theta^k}\right]\frac{\partial}{\partial\phi^w}.
\end{aligned}$$

Then

$$\Gamma_{uv}^{w}(\phi) = \Gamma_{ij}^{k}(\theta)\frac{\partial\theta^i}{\partial\phi^u}\frac{\partial\theta^j}{\partial\phi^v}\frac{\partial\phi^w}{\partial\theta^k}+\frac{\partial^2\theta^w}{\partial\phi^u\partial\phi^v}\frac{\partial\phi^w}{\partial\theta^k}.$$

It is well known that the Christofell symbol satisfies the transformation law: $\overset{\circ}{\Gamma}{}_{uv}^{w}(\phi) = \overset{\circ}{\Gamma}{}_{ij}^{k}(\theta)\frac{\partial\theta^i}{\partial\phi^u}\frac{\partial\theta^j}{\partial\phi^v}\frac{\partial\phi^w}{\partial\theta^k}+\frac{\partial^2\theta^w}{\partial\phi^u\partial\phi^v}$ (see [3]). One has

$$\Gamma_{uv}^{w}(\phi)-\overset{\circ}{\Gamma}{}_{uv}^{w}(\phi) = \left(\Gamma_{ij}^{k}(\theta)-\overset{\circ}{\Gamma}{}_{ij}^{k}(\theta)\right)\frac{\partial\theta^i}{\partial\phi^u}\frac{\partial\theta^j}{\partial\phi^v}\frac{\partial\phi^w}{\partial\theta^k}.$$

The last equation shows that

$$\Gamma_{ij}^{k} = \overset{\circ}{\Gamma}{}_{ij}^{k}+S_{ij}^{k}$$

where S_{ij}^{k} satisfies

$$S_{uv}^{w} = S_{ij}^{k}\theta_{i/u}\theta_{j/v}\phi_{w/k}$$

To the proof of (b), one shows that the following map:

$$\begin{array}{rrc}
\nabla: & \mathcal{X}(\mathcal{M})\times\mathcal{X}(\mathcal{M})\to & \mathcal{X}(\mathcal{M})\\
 & (X=x^i\partial_i, Y=y^j\partial_j)\mapsto & x^iy^j\Gamma_{ij}^{k}\partial_k+x^i\partial_i(y^j)\partial_j
\end{array}$$

is an affine connection on $\mathcal{M}$. □

Theorem 4.1. *Let $\mathcal{M}=\{p(\cdot;\theta);\theta=(\theta_1,\cdots,\theta_n)\in\Theta\}$ be a statistical manifold endowed with the Otto metric g (3.4). There exists an affine connection $\nabla^{(\alpha)}:\mathcal{X}(\mathcal{M})\times\mathcal{X}(\mathcal{M})\to\mathcal{X}(\mathcal{M})$ defined by :*

$$g\left(\nabla^{(\alpha)}_{\partial_i}\partial_j,\partial_j\right) = \Gamma^{(\alpha)}_{ij,k}. \tag{4.6}$$

Proof. Set $\Gamma^{(\alpha),k}_{i,j} = \Gamma^{(\alpha)}_{ij,m}g^{mk}$. By using Lemma 4.2, Lemma 4.1 and Proposition 3.1

$$\begin{array}{rrc}
\nabla^{(\alpha)}: & \mathcal{X}(\mathcal{M})\times\mathcal{X}(\mathcal{M})\to & \mathcal{X}(\mathcal{M})\\
 & (X=x^i\partial_i, Y=y^j\partial_j)\mapsto & x^iy^j\Gamma_{ij}^{(\alpha),k}\partial_k+x^i\partial_i(y^j)\partial_j
\end{array} \tag{4.7}$$

is an affine connection on $\mathcal{M}$. The proof is completed. □

Now, we prove that this α-connection is torsion-free and for $\alpha=0$ this connection is the Levi-Civita connection.

Theorem 4.2. 1. *$\nabla^{(\alpha)}$ is a torsion-free affine connection.*

2. *The* 0*-connection is the Levi-Civita connection with respect to the Otto metric.*

Proof. 1. We have

$$\begin{aligned}\nabla^{(\alpha)}_{\partial_i}\partial_j - \nabla^{(\alpha)}_{\partial_j}\partial_i &= \Gamma^{(\alpha),k}_{ij}\partial_k - \Gamma^{(\alpha),k}_{ji}\partial_k \\ &= \Gamma^{(\alpha),k}_{ij}\partial_k - \Gamma^{(\alpha),k}_{ij}\partial_k \\ &= 0\end{aligned}$$

where $\Gamma^{(\alpha),k}_{ij} = \Gamma^{(\alpha)}_{ij,m} g^{mk}$.

2. Taking the partial derivative of g_{ij} in Equation (3.4) with respect to θ_k, we obtain $\partial_k g_{ij} = \Gamma^{(0)}_{ij,k} + \Gamma^{(0)}_{kj,i}$. □

4.2. Flatness of exponential and mixture families

Let's introduce now the notion of exponential family. In general, if an n-dimensional model $\mathcal{M} = \{p(\cdot;\theta), \theta \in \Theta\}$ can be expressed in terms of functions $\{C, F_1, \cdots, F_n\}$ on $\mathcal{X}$ and a function ψ on Θ such that

$$p(x;\theta) = \exp\left[C(x) + \sum_{i=1}^{n} \theta^i F_i(x) - \psi(\theta)\right], \tag{4.8}$$

then we say that $\mathcal{M}$ is an exponential family, and that the $[\theta^i]$ are its natural or its canonical parameters. Next, let's consider the case where an n-dimensional model $\mathcal{M}$ can be expressed in terms of functions $\{C, F_1, \cdots, F_n\}$ on $\mathcal{X}$ as

$$p(x;\theta) = C(x) + \sum_{i=1}^{n} \theta^i F_i(x), \tag{4.9}$$

then we say that $\mathcal{M}$ is a mixture family, and that the $[\theta^i]$ are its mixture parameters.
The following theorem gives the flatness result of exponential family.

Theorem 4.3. *An exponential family*

$$\mathcal{M} = \left\{p(\cdot;\theta) = \exp\left(C(\cdot) + \sum_{i=1}^{n} \theta^i F_i(\cdot) - \psi(\theta)\right), \theta \in \Theta\right\}$$

equipped with Otto metric is (1)*-flat.*

Proof. Let $p(\cdot;\theta) \in \mathcal{M}$ with $\mathcal{M}$ the exponential family. We have $p(x;\theta) = \exp\left\{C(x) + \sum_{i=1}^{n} \theta^i F_i(x) - \psi(\theta)\right\}$. One has

$$\ell(x) = C(x) + \sum_{i=1}^{n} \theta^i F_i(x) - \psi(\theta)$$

Then

$$\partial_i \ell(x) = F_i(x) - \partial_i \psi(\theta); \partial_{ij}\ell(x) = -\partial_{ij}\psi(\theta); \partial_x \partial_{ij}\ell(x) = 0.$$

Thus

$$\mathit{\Gamma}^{(1)}_{ij,k} \quad = \quad -\partial_x \partial_{ij}\psi(\theta) \cdot \mathbb{E}_\theta\left[\partial_x \partial_k l\right] = 0.$$

This completes the proof. □

Similarly, we state the flatness for a mixture family.

Theorem 4.4. *A mixture familly*

$$\mathcal{M} = \left\{p(\cdot;\theta) = C(\cdot) + \sum_{i=1}^{n} \theta^i F_i(\cdot) - \psi(\theta), \theta \in \Theta\right\}$$

equiped with Otto metric is (-1)*-flat.*

Proof. Let $p(\cdot;\theta) \in \mathcal{M}$ with $\mathcal{M}$ the mixture family. We have $p(x;\theta) = C(x) + \sum_{i=1}^{n} \theta^i F_i(x)$. One has

$$\ell(x) = \log p(x;\theta).$$

Then

$$\partial_i \ell(x) = \frac{F_j(x)}{p(x;\theta)}; \partial_{ij}\ell(x) = -\frac{F_i(x)F_j(x)}{p^2(x;\theta)},$$

$$\begin{aligned} \partial_x \partial_{ij}\ell &= -\partial_x[\partial_i \ell \partial_j \ell] \\ &= -\partial_x \partial_i \ell \partial_j \ell - \partial_i \ell \partial_x \partial_j \ell. \end{aligned}$$

Thus

$$\partial_x \partial_{ij}\ell(x) \cdot \partial_x \partial_k \ell(x) = -\partial_x \partial_i \ell \cdot \partial_x \partial_k \ell(x) \partial_j \ell - \partial_x \partial_j \ell \partial_x \partial_k \ell(x) \partial_i \ell(x). \tag{4.10}$$

Using the previous equations and the definition of $\Gamma_{ij,k}^{(\alpha)}$ (4.1) we have

$$\Gamma_{ij,k}^{(-1)} = 0.$$

We conclude that the mixture family is (-1)-flat. □

Acknowledgments

The authors would like to express their profound gratitude to the anonymous referees for their fruitful remarks and to the "Centre d'Excellence Africain en Sciences Mathématiques et Applications" (CEA-SMA) at the "Institut de Mathématiques et de Sciences Physiques" (IMSP) for their financial support.

References

[1] Amari, S-I., Barndorff-Nielsen O. E., Kass R. E., Lauritzen,S. L., Rao C. R.: Differential geometry in statistical inference. Lecture notes-monograph series Shanti S., Series Editor vol. 10. (1987).
[2] Amari, S., Nagaoko, H.: Methods of information geometry. American Mathematical Soc. (2007).
[3] De Souza, E.: Tensor Calculus for Engineers and Physicists. Springer (2016).
[4] De Giorgi, E.: New problems on minimizing movements, Boundary Value Problems for PDE and Applications, pp. 81-98, Masson, Paris (1993).
[5] Ambrosio, L., Gigli, N., Savaré, G.: Gradient flows in metric space and in the space of probability measures, Lectures in Mathematics ETH, second ed., Birkhäuser Verlag, Besel (2008).
[6] Do Carmo, M.: Riemannian Geometry, Birkhauser Inc., Boston (1992).
[7] Gbaguidi Amoussou, A., Djibril Moussa, F., Ogouyandjou, C., Diop, M. A.: New connections on the fiber-bundle of generalized statistical manifolds. In: Proceedings of the International Conference on Differential Geometry- Dynamical Systems (BSG DGDS), 30 Aug- 2 Sept 2018, Mangalia-ROMANIA. BSG Proceedings. **26**, 23-32 (2019).
[8] Kantorovich, L.: *On the translocation of masses*. C. R. Acad. Sci. URSS (N.S) **37**, 199-201 (1942).
[9] Lott, J.: *Some geometric calculation on Wasserstein space*. Commun. Math. Phys. **277**, 423-437 (2008).
[10] Olkin, I., Pukelsheim, F.: *The distance between two random vectors with given dispersion matrices*. Linear Algebra Appl. **48**, 257–263 (1982). https://doi.org/10.1016/0024-3795(82)90112-4.
[11] Rao, C.R.: *Information and accuracy attainable in the estimation of statistical parameter*. Bull. Calcutta. Math. Soc. **37**, 81-91 (1945).
[12] De Souza, D., Vigelis R., Cavalcante C.: *Geometry induced by a generalization of Rényi divergence*. Entropy, **18(11)**, 407 (2016).
[13] Villani, C.: Topics in Optimal Transportation. Graduate studies in Mathematics 58, Providence, RI: Ameri. Math. Soc. (2003).

Affiliations

CARLOS OGOUYANDJOU
ADDRESS: Institut de Mathématiques et de Sciences Physiques(IMSP) Université d'Abomey-Calavi(UAC), Bénin.
E-MAIL: ogouyandjou@imsp-uac.org
ORCID ID: 0000-0001-7115-7724

NESTOR WADAGNI
ADDRESS: Institut de Mathématiques et de Sciences Physiques(IMSP) Université d'Abomey-Calavi(UAC), Bénin.
E-MAIL: nestor.wadagni@imsp-uac.org
ORCID ID: 0000-0001-8757-176X