Global Journal of Advanced Research
on Classical and Modern Geometries
ISSN: 2284-5569, Vol.8, (2019), Issue 1, pp.18-25

SUBMANIFOLDS OF EXPONENTIAL FAMILIES

MAHESH T. V. AND K.S. SUBRAHAMANIAN MOOSATH

ABSTRACT. Exponential family with ± 1 - connection plays an important role in information geometry. Amari proved that a submanifold M of an exponential family S is exponential if and only if M is a ∇^1- autoparallel submanifold. We show that if all ∇^1- auto parallel proper submanifolds of a $\pm 1-$ flat statistical manifold S are exponential then S is an exponential family. Also shown that the submanifold of a parameterized model S which is an exponential family is a ∇^1 - autoparallel submanifold.

Keywords: statistical manifold, exponential family, autoparallel submanifold. 2010 MSC: 53A15

1. Introduction

Information geometry emerged from the geometric study of a statistical model of probability distributions. The information geometric tools are widely applied to various fields such as statistics, information theory, stochastic processes, neural networks, statistical physics, neuroscience etc.[3][7]. The importance of the differential geometric approach to the field of statistics was first noticed by C R Rao [6]. On a statistical model of probability distributions he introduced a Riemannian metric defined by the Fisher information known as the Fisher information metric. Another milestone in this area is the work of Amari [1][2][5]. He introduced the α - geometric structures on a statistical manifold consisting of Fisher information metric and the $\pm\alpha$ - connections. Harsha and Moosath [4] introduced more generalized geometric structures called the (F, G) geometry on a statistical manifold which is a generalization of $\alpha-$ geometry. There are many attempts to understand the geometry of the statistical manifold and also to develop a differential geometric framework for the estimation theory.
In this paper we shall study the geometry of exponential family. An exponential family is an important statistical model which is attracted by many of the researchers from Physics, Mathematics and Statistics. The exponential family contains as special cases most of the standard discrete and continuous distributions that we use for practical modelling, such as the normal, Poisson, Binomial, exponential, Gamma, multivariate normal, etc. Distributions in the exponential family have been used in classical statistics for decades. We discuss the dually flat structure of the finite dimensional exponential family with respect to the $\pm 1-$ connections defined by Amari. Then we prove a condition for a $\pm 1-$ flat statistical manifold to be an exponential family. Also show that submanifold of a statistical manifold which is an exponential family is a ∇^1 - autoparallel submanifold.

2. Statistical Manifold

Consider the sample space $\mathcal{X} \subseteq \mathbb{R}^n$. A probability measure on $\mathcal{X}$ can be represented in terms of density function with respect to Lebesgue measure.

Key words and phrases. statistical manifold, exponential family, autoparallel submanifold.
AMS 2010 Mathematics Subject Classification:53A15.

Definition 2.1. Consider a family $\mathcal{S}$ of probability distributions on $\mathcal{X}$. Suppose each element of $\mathcal{S}$ can be parametrized using n real-valued variables $(\theta^1, ..., \theta^n)$ so that

$$\mathcal{S} = \{p_\theta = p(x;\theta) \ / \ \theta = (\theta^1, ..., \theta^n) \in \mathbb{E}\} \tag{2.1}$$

where $\mathbb{E}$ is a subset of $\mathbb{R}^n$ and the mapping $\theta \mapsto p_\theta$ is injective. We call such family $\mathcal{S}$ an n-dimensional **statistical model** or a **parametric model** or simply a **model** on θ.

Let us now state certain regularity conditions which are required for our geometric theory.
Regularity conditions

(1) We assume that $\mathbb{E}$ is an open subset of $\mathbb{R}^n$ and for each $x \in \mathcal{X}$, the function $\theta \mapsto p(x;\theta)$ is of class c^∞.
(2) Let $\ell(x;\theta) = \log p(x;\theta)$. For every fixed θ, n functions in x, $\{\partial_i \ell(x;\theta); i = 1, ..., n\}$ are linearly independent, where $\partial_i = \frac{\partial}{\partial \theta_i}$.
(3) The order of integration and differentiation may be freely rearranged.
(4) The moments of $\partial_i \ell(x;\theta)$ exists upto necessary orders.
(5) For a probability distribution p on Ω, let the support of p be defined as, $\text{supp}(p) := \{x \mid p(x) > 0\}$. The case when $\text{supp}(p_\theta)$ varies with θ poses rather significant difficulties for analysis. Hence we assume that $\text{supp}(p_\theta)$ is constant with respect to θ. Then we can redefine $\mathcal{X}$ to be $\text{supp}(p_\theta)$. This is equivalent to assuming that $p(x;\theta) > 0$ holds for all $\theta \in \mathbb{E}$ and all $x \in \mathcal{X}$. This means that the model $\mathcal{S}$ is a subset of

$$\mathcal{P}(\mathcal{X}) := \{p : \mathcal{X} \longrightarrow \mathbb{R} \mid p(x) > 0 \ (\forall \ x \in \mathcal{X}); \int_{\mathcal{X}} p(x)dx = 1\} \tag{2.2}$$

Definition 2.2. For a model $\mathcal{S} = \{p_\theta \ / \ \theta \in \mathbb{E}\}$, the mapping $\varphi : \mathcal{S} \longrightarrow \mathbb{R}^n$ defined by $\varphi(p_\theta) = \theta$ allows us to consider $\varphi = (\theta^i)$ as a coordinate system for $\mathcal{S}$. Suppose we have a c^∞ diffeomorphism $\psi : \mathbb{E} \longrightarrow \psi(\mathbb{E})$, where $\psi(\mathbb{E})$ is an open subset of $\mathbb{R}^n$. Then if we use $\rho = \psi(\theta)$ instead of θ as our parameter, we obtain $\mathcal{S} = \{p_{\psi^{-1}(\rho)} \mid \rho \in \psi(\mathbb{E})\}$. This expresses the same family of probability distributions $\mathcal{S} = \{p_\theta\}$. If we consider parametrizations which are c^∞ diffeomorphic to each other to be equivalent, then we may consider $\mathcal{S}$ as a c^∞ differentiable manifold and we call it as a **statistical manifold**.

For the statistical manifold $\mathcal{S} = \{p(x;\theta)\}$, define $\ell(x;\theta) = \log p(x;\theta)$ and consider the partial derivatives $\partial_i \ell; i = 1, ..., n$. By our assumption, $\partial_i \ell; i = 1, ..., n$ are linearly independent functions in x. We can construct the following n-dimensional vector space spanned by n functions $\partial_i \ell; i = 1, ..., n$ in x as,

$$T_\theta^1(\mathcal{S}) = \{A(x) \ / \ A(x) = \sum_{i=1}^{n} A^i \partial_i \ell\}. \tag{2.3}$$

Define expectation with respect to the distribution $p(x;\theta)$ as

$$E_\theta(f) = \int f(x) p(x;\theta) dx. \tag{2.4}$$

Note that $E_\theta[\partial_i \ell_{x;\theta}] = 0$ since $p(x;\theta)$ satisfies

$$\int p(x;\theta) dx = 1. \tag{2.5}$$

Hence for any random variable $A(x) \in T_\theta^1(\mathcal{S})$, we have $E_\theta[A(x)] = 0$.
This expectation induces an inner product on $\mathcal{S}$ in a natural way.

$$< A(x), B(x) >_\theta = \ E_\theta[A(x)B(x)]; \ \ for A(x), B(x) \in T_\theta^1(\mathcal{S})$$

Especially the inner product of the basis vectors ∂_i and ∂_j is

$$g_{ij}(\theta) = <\partial_i, \partial_j>_\theta = E_\theta[\partial_i \ell(x;\theta)\partial_j \ell(x;\theta)] \quad (2.6)$$
$$= -E[\partial_i \partial_j \ell(x;\theta)] \quad (2.7)$$
$$= \int \partial_i \ell(x;\theta)\partial_j \ell(x;\theta) p(x;\theta)dx. \quad (2.8)$$

It is clear that the matrix $G(\theta) = (g_{ij}(\theta))$ is symmetric (i.e $g_{ij} = g_{ji}$). For any n-dimensional vector $c = [c^1, ..., c^n]^t$

$$c^t G(\theta)c = \int \{\sum_{i=1}^{n} c^i \partial_i \ell(x;\theta)\}^2 p(x;\theta)dx \geqslant 0 \quad (2.9)$$

since $\{\partial_1 \ell(x;\theta), ..., \partial_n \ell(x;\theta)\}$ are linearly independent, G is positive definite. Hence $g =<,>$ defined in (2.8) is a Riemannian metric on the statistical manifold $\mathcal{S}$, called the **Fisher information metric**.

Example 2.3. Normal distribution
$\mathcal{X} = \mathbb{R}$, $n = 2$, $\theta = (\mu, \sigma)$, $E = \{(\mu, \sigma) \ / \ -\infty < \mu < \infty, 0 < \sigma < \infty\}$

$$\mathcal{S} = N(\mu, \sigma) = \{p(x;\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}\}. \quad (2.10)$$

This is a 2-dimensional manifold which can be identified with the upper half plane. The log-likelihood function is given by

$$\ell(x,\theta) = -\frac{(x-\mu)^2}{2\sigma^2} - \log \sqrt{2\pi}\sigma$$

The tangent space $T^1_\theta \mathcal{S}$ is spanned by $\partial_1 = \frac{\partial}{\partial \mu}$ and $\partial_2 = \frac{\partial}{\partial \sigma}$.

$$\partial_1 = \frac{(x-\mu)}{\sigma^2}\ , \ \partial_2 = -\frac{(x-\mu)^2}{\sigma^3} - \frac{1}{\sigma}$$

Then the Fisher information matrix $G(\theta) = (g_{ij})$ is given by

$$\begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{bmatrix}$$

Definition 2.4. Let $\mathcal{S} = \{p(x;\theta) \ / \ \theta \in \mathbb{E}\}$ be an n-dimensional statistical manifold with the Fisher metric g. We can define n^3 functions Γ^1_{ijk} by

$$\Gamma^1_{ijk} = E_\theta[(\partial_i \partial_j \ell(x;\theta))(\partial_k \ell(x;\theta))] \quad (2.11)$$

Γ^1_{ijk} uniquely determine an affine connection ∇^1 on the statistical manifold $\mathcal{S}$ by

$$\Gamma^1_{ijk} =< \nabla^1_{\partial_i} \partial_j, \partial_k > \quad (2.12)$$

∇^1 is called the $1-$**connection** or the **exponential connection**.

Here $\ell(x;\theta)$ the logarithm of the density function $p(x;\theta)$ is used to define the fundamental geometric structures in a statistical model $S = \{p(x;\theta)\}$. Amari defined one parameter family of functions called the α - embedding indexed by $\alpha \in \mathbb{R}$.

Definition 2.5. Let $L_{(\alpha)}(p)$ be a one parameter family of functions defined by

$$L_{(\alpha)}(p) = \begin{cases} \frac{2}{1-\alpha} p^{\frac{1-\alpha}{2}} & \alpha \neq 1 \\ \log p & \alpha = 1 \end{cases} \quad (2.13)$$

and we call

$$\ell_{(\alpha)}(x;\theta) = L_{(\alpha)}(p(x;\theta)) \tag{2.14}$$

the $\alpha-$**representation** of the density function $p(x;\theta)$.

The $1-$representation $\ell_1(x;\theta)$ is the log-likelihood function $\ell(x;\theta)$ and the $(-1)-$ representation $\ell_{-1}(x;\theta)$ is the density function $p(x;\theta)$ itself.
Let $T^{\alpha}_{\theta}(\mathcal{S})$ be the vector space spanned by n linearly independent functions $\partial_i \ell_\alpha(x;\theta)$ in x for $i = 1, ..., n$.

$$T^{\alpha}_{\theta}(\mathcal{S}) = \{A(x) \ / \ A(x) = \sum_{i=1}^{n} A^i \partial_i \ell_\alpha(x;\theta)\}. \tag{2.15}$$

There is a natural isomorphism between these two vector spaces $T^1_\theta(\mathcal{S})$ and $T^\alpha_\theta(\mathcal{S})$ given by

$$\partial_i \ell_1(x;\theta) \in T^1_\theta(\mathcal{S}) \longleftrightarrow \partial_i \ell_\alpha(x;\theta) \in T^\alpha_\theta(\mathcal{S}) \tag{2.16}$$

The vector space $T^\alpha_\theta(\mathcal{S})$ is called the $\alpha-$**representation of the tangent space** $T^1_\theta(\mathcal{S})$. The $\alpha-$representation of a vector $A = \sum_{i=1}^{n} A^i \partial_i \ell \in T^1_\theta(\mathcal{S})$ is the random variable

$$A_\alpha(x) = \sum_{i=1}^{n} A^i \partial_i \ell_\alpha(x;\theta) \tag{2.17}$$

Let us define the $\alpha-$**expectation** of a random variable f with respect to the density $p(x;\theta)$ as

$$E^\alpha_\theta(f) = \int f(x) p(x;\theta)^\alpha dx. \tag{2.18}$$

Then an inner product can be defined naturally as

$$< A_\alpha(x), B_\alpha(x) >^\alpha_\theta = E^\alpha_\theta[A_\alpha(x) B_\alpha(x)]. \tag{2.19}$$

We have the relations

$$\partial_i \ell_\alpha(x;\theta) = p^{\frac{(1-\alpha)}{2}} \partial_i \ell(x;\theta) \tag{2.20}$$

$$(\partial_i \ell_\alpha)(\partial_j \ell_{-\alpha}) = p(x;\theta)\ \partial_i \ell\ \partial_j \ell. \tag{2.21}$$

Thus we have

$$\begin{aligned} < \partial_i, \partial_j >^\alpha_\theta &= \int \partial_i \ell_\alpha(x;\theta) \partial_j \ell_\alpha(x;\theta) p(x;\theta)^\alpha dx && (2.22)\\ &= \int \partial_i \ell\ \partial_j \ell\ p(x;\theta)\ dx && (2.23)\\ &= g_{ij}(\theta) && (2.24) \end{aligned}$$

and the inner product has the following dualistic expression for any α,

$$< A_\alpha(x), B_\alpha(x) >_\theta = \int A_\alpha(x,\theta) B_{-\alpha}(x;\theta) dx. \tag{2.25}$$

Then we say that the two vector spaces $T^\alpha_\theta(\mathcal{S})$ and $T^{-\alpha}_\theta(\mathcal{S})$ are **dually coupled**. That is the inner product of two vectors A and B is given by the integration of the product of their $\alpha-$ and $(-\alpha)-$representations.
We have,

$$\partial_i \partial_j \ell_\alpha = p^{\frac{(1-\alpha)}{2}} (\partial_i \partial_j \ell + \frac{1-\alpha}{2} \partial_i \ell \partial_j \ell). \tag{2.26}$$

Hence we can define n^3 functions Γ^α_{ijk} by

$$\Gamma^\alpha_{ijk} = \int \partial_i \partial_j \ell_\alpha(x;\theta) \partial_k \ell_{-\alpha}(x;\theta) dx. \tag{2.27}$$

These Γ^α_{ijk} uniquely determine connections ∇^α on the statistical manifold $\mathcal{S}$ by

$$\Gamma^\alpha_{ijk} = < \nabla^\alpha_{\partial_i} \partial_j, \partial_k > \tag{2.28}$$

which is called α**-connection**.
Thus the one parameter family of functions $L_\alpha(p)$ defines a family of connections ∇^α, $\alpha \in \mathbb{R}$ on the statistical manifold $\mathcal{S}$.

Lemma 2.6. *The $\alpha-$connection ∇^α and the $-(\alpha)-$connection $\nabla^{-\alpha}$ are dual with respect to the Fisher information metric. In particular, the $0-$ connection is the Levi-Civita connection or the metric connection.*

Proof. By the use of $\alpha-$representation, we have

$$\begin{aligned} A < B, C > &= A \int B_\alpha(x,\theta) C_{-\alpha}(x;\theta) dx && (2.29) \\ &= \int (AB_\alpha(x,\theta)) C_{-\alpha}(x;\theta) dx + \int B_\alpha(x,\theta)(AC_{-\alpha}(x;\theta)) dx \\ &= < \nabla^\alpha_A B, C > + < B, \nabla^{-\alpha}_A C > . && (2.30) \end{aligned}$$

□

3. The Exponential Family

The Exponential family is a practically convenient and widely used unified family of distributions on finite dimensional Euclidean spaces parametrized by a finite dimensional parameter vector. It contains as special cases most of the standard discrete and continuous distributions that we use for practical modelling, such as the normal, Poisson, Binomial, exponential, Gamma, multivariate normal, etc.

Definition 3.1. The standard form of a n-dimensional exponential family of distributions $\mathcal{S} = \{p(x;\theta) \ / \ \theta \in E \subseteq \mathbb{R}^n\}$ is

$$p(x;\theta) = \exp(\sum_{i=1}^{n} \theta^i x_i - \psi(\theta)) \quad \text{or} \quad \log(p(x;\theta)) = \sum_{i=1}^{n} \theta^i x_i - \psi(\theta) \qquad (3.1)$$

where $x = (x_1, ..., x_n)$ is a set of random variables, $\theta = (\theta^1, .., \theta^n)$ are the canonical parameters and $\psi(\theta)$ is determined from the normalization condition.

Now consider the exponential family $\mathcal{S} = \{p(x;\theta) \ / \ \theta \in E \subseteq \mathbb{R}^n\}$ where $p(x,\theta) = exp[\sum_{i=1}^{n} \theta^i x_i - \psi(\theta)]$. Now $\partial_i l(x;\theta) = x_i - \partial_i \psi(\theta)$, $\quad \partial_i \partial_j l(x;\theta) = -\partial_i \partial_j \psi(\theta)$ Then $\Gamma^1_{ijk} = \partial_i \partial_j \psi(\theta) E_\theta(\partial_k l_\theta) = 0$
Thus we have $\nabla^1_{\partial_i} \partial_j = 0$. Then we say that the exponential family is 1 - flat. By duality we get it is -1 - flat also. Thus the exponential family is a dully flat space with respect to the $\pm 1-$ connections defined by Amari. Thus

Theorem 3.2. *The exponential family is a dually flat space with respect to the ± 1 connections defined by Amari.*

A dually flat space is an important tool in the geometric study of statistical estimation. Now we have seen that the important statistical model the exponential family has a dually flat structure with respect to the $\alpha = \pm 1-$connections.

4. Characterization of exponential family

Amari has given necessary and sufficient condition for a submanifold of an exponential family to be exponential. In this section we prove that a parameterized family which is flat with respect to ± 1-connection is an exponential family if all the ∇^1- autoparallel submanifolds are exponential. Also we show that if submanifold of a statistical manifold is exponential then it is ∇^1- autoparallel.

Definition 4.1. Let M be a smooth n-dimensional manifold and $\tau(\mathbf{M})$ denote the set of all smooth vector fields on M. Let $X \in \tau(\mathbf{M})$ and ∇ be a connection on $\mathbf{M}$, then X is parallel on $\mathbf{M}$ with respect to ∇ iff $\nabla_X Y = 0; \quad \forall Y \in \tau(\mathbf{M})$.
$\mathbf{M}$ is called flat with respect to ∇ if $\nabla_{\partial_i}\partial_j = 0$ for all i, j.

Definition 4.2. Let $\mathbf{N}$ be a submanifold of $\mathbf{M}$ and ∇ be an affine connection on $\mathbf{M}$. $\mathbf{N}$ is said to be autoparallel with respect to ∇ if $\nabla_X Y \in \tau(\mathbf{N})$ for all $X, Y \in \tau(\mathbf{N})$.
1-dimensional autoparallel submanifolds are called geodesics.

Remark 4.3. A necessary and sufficient condition for $\mathbf{N}$ to be autoparallel is that $\nabla_{\partial a}\partial b \in \tau(\mathbf{N})$ holds for all a, b.

Definition 4.4. Let $\mathbf{N}$ be a submanifold of $\mathbf{M}$. Let $p \in \mathbf{N}$, then $\mathrm{T}_p\mathbf{N} \subset \mathrm{T}_p\mathbf{M}$. Now consider the projection map $\pi_p : \mathrm{T}_p\mathbf{N} \longrightarrow \mathrm{T}_p\mathbf{M}$ and $\pi_p(D) = D$, $\forall D \in \mathrm{T}_p\mathbf{N}$. Let ∇ be a connection on $\mathbb{M}$, then we define a connection ∇^π on $\mathbf{N}$ as

$$(\nabla^\pi_X Y)_p = \pi_p(\nabla_X Y)_p; \ \ \forall p \in \mathbf{N}.$$

Now define

$$H(X, Y) = \nabla_X Y - \nabla^\pi_X Y$$

is called the second fundamental form or embedding curvature.

Remark 4.5. For each $p \in \mathbf{M}$. Let $\{(\partial_a)_p; 1 \le a \le m\}$, $m = \dim(\mathbf{N})$, be a basis for $\mathrm{T}_p\mathbf{N}$ and let $\{(\partial_k)_p; m+1 \le k \le n\}$, $n = \dim(\mathbf{M})$, be a basis for $\mathrm{T}_p\mathbf{N}^\perp$. Then we define $m^2(n-m)$ functions $\{H_{abk}\}$ in the following way

$$H_{abk} = \langle H(\partial_a, \partial_b), \partial_k \rangle = \langle \nabla_{\partial_a}\partial_b, \partial_k \rangle.$$

It follows that $H = 0$ iff $H_{abk} = 0; \ \ \forall a, b, k$

Remark 4.6. $H(X, Y) = 0$ iff $\mathbf{N}$ is ∇- autoparallel submanifold of $\mathbf{M}$.

Remark 4.7. We know that the exponential family is dually flat with respect to the ± 1- connection. But a parametrized model which is flat with respect to ± 1- connection need not be an exponential family.

Example 4.8. Let q be a smooth probability density function on $\mathbf{R}$ and q^k be the k^{th} iid extension. Then for

$$Y = (y^1, y^2, y^3, ...y^k)^t \tag{4.1}$$

we have

$$q^k(Y) = q(y^1)q(y^2)q(y^3), ...q(y^k). \tag{4.2}$$

For a regular matrix $A \in \mathbb{R}^{k\times k}$ and a vector $\mu \in \mathbb{R}^k$, we define a probability density function on $\mathbb{R}^k$ by

$$p(A, \mu, x) = \frac{q^k(A^{-1}(x-\mu))}{\mid det(A) \mid}. \tag{4.3}$$

Now define a statistical model

$$S = \{p(A, \mu, x) \mid \mu \in \mathbb{R}^k\}. \tag{4.4}$$

Now consider

$$log(p(A, \mu, x)) = \sum_{i=1}^{k} log(q(A^{-1}(x-\mu))) - log(\mid det(A) \mid). \tag{4.5}$$

Then clearly $\frac{\partial log(p(A,\mu,x))}{\partial \mu_i}$ is constant. So from the definition of Amari's α- connection $\Gamma^\alpha_{ij,k} = 0$, it implies that S is α-flat for all α, but in general it is not necessarily to be an exponential family.

Amari [1] has proved that a submanifold M of an exponential family **S** is an exponential family if and only if **M** is autoparallel with respect to ∇^1 in **S**.
Now we prove a condition for a statistical manifold to be an exponential family.

Theorem 4.9. *Let $S = \{P(x,\theta) \mid \theta \in \Theta\}$ be a parametrized family which is flat with respect to ∇^1 and ∇^{-1}. If all ∇^1- autoparallel proper submanifolds of S are exponential family, then S is an exponential family.*

Proof. Let $S = \{P(x,\theta) \mid \theta \in \Theta\}$ be an n-dimensional statistical manifold with dually flat structure $(S, g, \nabla^1, \nabla^{-1})$, where g is the Fisher information metric. Let $\theta = [\theta^i]$ and $\eta = [\eta_j]$ be the coordinate system of S with respect to ∇^1 and ∇^{-1} respectively. Now subdivide the range of index $i = 1, 2, ...n$ into indexing sets $I = \{i = 1, 2, ...k\}$ and $II = \{i = k+1, k+2, ...n\}$. Let $M(C_{II})$ be the set of points whose coordinates $[\theta^i]$ in II are fixed to constant $C_{II} = (C^i_{II})$ for $i = k+1, k+2, ..n$. That is

$$M(C_{II}) = \{p \in S \mid \theta^{k+1} = C^{k+1}_{II}, \theta^{k+2} = C^{k+2}_{II}, ...\theta^n = C^n_{II}\} \tag{4.6}$$

where $C_{II} \in \mathbb{R}^{n-k}$, then clearly this is an affine space with respect to θ- coordinate system, which implies $M(C_{II})$ is a ∇^1- autoparallel submanifold of S. Also if $C_{II} \neq C'_{II}$ then $M(C_{II}) \cap M(C'_{II})$ = ϕ and $\bigcup_{C_{II}} M(C_{II}) = S$. Now by our assumption $M(C_{II})$ is an exponential family for all C_{II}. If $p(x,\theta) \in S$, then $p(x,\theta) \in M(C_{II})$ for some constant C_{II}, this implies that

$$p(x,\theta) = exp(\sum_{i=1}^{k} \theta^i x_i - \psi^\beta(\theta)) \tag{4.7}$$

where $\psi^\beta(\theta)$ defined on $\Theta^\beta = \{\theta \in \Theta \mid \theta^{k+1} = C^{k+1}_{II}, \theta^{k+2} = C^{k+2}_{II}, ...\theta^n = C^n_{II}\}$. Now define $\phi(\theta)$ = $\psi^\beta(\theta)$ if $\theta \in \Theta^\beta$. Then we can write

$$\begin{aligned} p(x,\theta) &= exp(\sum_{i=1}^{k} \theta^i x_i - \phi(\theta)) && (4.8)\\ &= exp(\sum_{i=1}^{k} \theta^i x_i + \sum_{i=k+1}^{n} C^i_{II} x_i - \sum_{i=1}^{k} C^i_{II} x_i - \phi(\theta)) && (4.9)\\ &= exp(\sum_{i=1}^{n} \theta^i x_i + F(x) - \phi(\theta)) && (4.10) \end{aligned}$$

where $F(x) = -\sum_{i=1}^{k} C^i_{II} x_i$ for $p(x,\theta) \in M(C_{II})$, then S is an exponential family. □

Theorem 4.10. *Let $S = \{p(x,\theta) \mid \theta \in \Theta\}$ be a statistical manifold with ∇^1 connection. Let M be a submanifold of S. If M is an exponential family then M is ∇^1 autoparallel submanifold of S.*

Proof. Let $S = \{p(x,\theta) \mid \theta \in \Theta\}$ and $M = \{q(x,u)\}$ be a submanifold of S, $[\theta^i]$ be the coordinates on S and $[u_a]$ be the coordinates on M. Suppose M is an exponential family, then

$$q(x,u) = p(x,\theta(u)) = \exp\{\sum_{a=1}^{n} u_a G^a(x) + D(x) - \phi(u)\} \tag{4.11}$$

we have, $\Gamma^1_{ab,k} = E_\xi[(\partial_a \partial_a \ell_\theta)\partial_k \ell_\theta]$, where $\ell_\theta = log(p(x,\theta))$. Then $\partial_a \partial_a \ell_\theta = -\frac{\partial^2 \phi}{\partial u_a \partial u_b}$. Therefore we have, $\Gamma^1_{ab,k} = 0$ which implies $\langle \nabla^1_{\partial_a} \partial_b, \partial_k \rangle = 0; \quad \forall k$. Hence $H_{abk} = 0$, which implies that M is a ∇^1- autoparallel submanifold of S. □

5. Acknowledgments

The first named author was supported by Doctoral Research Fellowship from Indian Institute of Space Science and Technology (IIST), Kerala, India.

REFERENCES

[1] Amari S, *Methods of information geometry*, *Oxford University Press* (2000).

[2] Amari S, *Information geometry and its applications*, *Springer* (2016).

[3] Ay N., Jost J., Le H.V., Schwachhöfer L., *Information Geometry*, *Springer* (2017).

[4] Harsha K.V. and Subrahamanian Moosath K.S., F-geometry and Amari's α-geometry on a Statistical Manifold. *Entropy*, 16(5): 2472-2487, 2014.

[5] M. K. Murry and W. J. Rice, *Differential Geometry and Statistics*, *Chapman & Hall* (1993).

[6] Rao C.R., Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37: 81-91, 1945.

[7] Crasmareanu M.; Hreţcanu C. E., Statistical structures on metric path spaces, Chin. Ann. Math., Ser. B 33, No. 6, 889-902 (2012).

DEPARTMENT OF MATHEMATICS, INDIAN INSTITUTE OF SPACE SCIENCE AND TECHNOLOGY, TRIVANDRUM, KERALA, INDIA

E-mail address: maheshtv32@gmail.com

E-mail address: smoosath@iist.ac.in