Affine Immersions and Statistical Manifolds

Mahesh T V, K S Subrahamanian Moosath

Department of Mathematics
Indian Institute of Space Science And Technology
Valiyamalap(o), Thiruvananthapuram, kerala, India- 695547

Abstract

In this paper first we discuss about statistical manifolds and immersed hyper surfaces in $\mathbb{R}^{n+1}$. We show that every centro-affine hyper surface is a statistical manifold realized in $\mathbb{R}^{n+1}$. Then proved that dually flat statistical manifold can be immersed as a centro-affine hyper surface. Also observed that a Blaschke immersed manifold is a statistical manifold. Then in the case of the dually flat statistical manifold $(S, \nabla^e, \nabla^m)$ we show that if all $\nabla^e$- autoparallel submanifolds are exponential then $S$ is an exponential family. Also prove that submanifold of a statistical manifold with $\nabla^e$ connection is $\nabla^e$- autoparallel if it is exponential family.

Keywords: statistical manifold, affine immersion, exponential family

2010 MSC: 53A15

Introduction

Information geometry investigates the differential geometric structure of statistical manifolds. It has got a wide variety of applications in the areas of engineering and science. Parametrized family of probability measures on a sample space (finite or infinite) has got an affine space structure introduced by Amari and Chensov. They have also given a family of affine connections with dualistic property on the parametrized family of probability measures. So a family of parametrized probability density functions is a smooth manifold, called the statistical manifold, with the parameters as coordinates, Fisher information is the Riemannian metric and we have a family of affine connections with dualistic property. In the statistical manifold in which parameters are coordinate, the affine structure enables the nice representation of probability measures through the vector space of random variables. A statistical manifold is also viewed as a Riemannian manifold with a three tensor which is symmetric (Lauritzen). Kurose [3] studied the affine immersions of statistical manifolds and Dillen, Nomizu , Vranken [5] studied affine hyper surfaces which has relevance in simply connected statistical manifold theory.

In this paper we look at the affine geometric aspects of statistical manifolds and also of the $\nabla^e$ autoparallel submanifolds of exponential family. In section 1 we recall the basic results in affine differential geometry. In section (2)
we introduce the notion of statistical manifold realized in $\mathbb{R}^{n+1}$, there we show that every centro-affine hypersurface is a statistical manifold realized in $\mathbb{R}^{n+1}$. Dually flat statistical manifold is locally immersed as a centro-affine hypersurface, also we observed that for a Blaschke immersion $f : M \rightarrow \mathbb{R}^{n+1}$, $(M, \nabla, h)$ is a statistical manifold. In section 3 we discuss about exponential family. Family of probability measures is a subset of infinite dimensional affine subspace of positive measures up to scale. Parametrized models are surfaces in this affine space, among that the exponential family is a statistical manifold realized in $\mathbb{R}^{n+1}$. We show that if all $\nabla^c$- auto parallel submanifold of a parametrized model $S$ is exponential then $S$ is an exponential family. Also Shown that submanifold of a parametrized model $S$ which is an exponential family is a $\nabla^c$- auto parallel submanifold. Also observed that exponential family is locally immersed as a centro-affine hypersurface.

1. Affine Immersion

In this section we give a brief description about immersion of a manifold in $\mathbb{R}^{n+1}$ as a hyper surface.

Let $M$ be an $n$-dimensional smooth manifold, $\mathcal{F}(M)$ denote the set of all real valued smooth functions on $M$, $\tau(M)$ be the set of all smooth vector fields on $M$ and $\nabla$ be an affine connection on $M$. A vector field $X \in \tau(M)$ is called parallel on $M$ if $\nabla_X Y = 0$, $\forall Y \in \tau(M)$ and $M$ is called flat with respect to $\nabla$ if $\nabla_{\partial_i} \partial_j = 0$ for all $i, j$, where $\partial_i$ are the basic vector fields with respect to the local coordinates $(x_1, x_2, x_3...x_n)$. The torsion tensor field $T(X, Y)$ and curvature tensor field $R(X, Y)Z$ with respect to $\nabla$ are defined as

$$T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y]$$

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

The Ricci Tensor of type $(0, 2)$ is defined by $Ric(Y, Z) = \text{trace}\{X \rightarrow R(X, Y)Z\}$. A symmetric tensor field $g$ of type $(0, 2)$ is called a non-degenerate metric on $M$, if $g(X, Y) = 0, \forall Y \in \tau(M)$ then $X = 0$. It is also called pseudo-Riemannian metric. If $g$ is positive definite then we say $g$ is a Riemannian metric. An affine connection $\nabla$ is called a metric connection if $\nabla g = 0$, where $\nabla g$ denote the covariant derivative of the tensor field $g$. If $\nabla$ is metric then we have $X g(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)$ for every $X, Y, Z \in \tau(M)$.

**Definition.** Let $(M, \nabla)$ be $n$-manifold with an affine connection $\nabla$. An immersion $f : M \rightarrow \mathbb{R}^{n+1}$ is called an affine immersion if there is a vector field $\xi$ on $M$ such that the following holds

(a) $T_{f(x)}(\mathbb{R}^{n+1}) = f_*(T_x(M)) + \text{span}\{\xi_x\}$

(b) $D_X f_*(Y) = f_*(\nabla_X Y) + h(X, Y)\xi$

The vector field $\xi$ which satisfies (a) is called the transversal vector field. $h$ is a symmetric bilinear function on the tangent space $T_x(M)$, called the affine fundamental form and $D$ denote the standard connection on $\mathbb{R}^{n+1}$. 

2
Note that for an immersion $f : M \to \mathbb{R}^{n+1}$ having a transversal vector field $\xi$ on $M$ there is a torsion free induced connection $\nabla$ satisfying $D_X f_\ast (Y) = f_\ast (\nabla_X Y) + h(X,Y) \xi$. Also for all $X \in \tau(M)$, $\xi$ satisfies $D_X \xi = -f_\ast (S X) + \alpha(X) \xi$, where $S$ is a tensor of type $(1,1)$ called affine the shape operator and $\alpha$ is a 1-form called the transversal connection from.

**Definition.** Let $\omega$ be a fixed volume element on $\mathbb{R}^{n+1}$. For a hyper surface immersion $f : M \to \mathbb{R}^{n+1}$ with transversal vector field $\xi$ we define the induced volume element $\eta$ on $M$ as $\eta(X_1, X_2, \ldots X_n) = \omega(X_1, X_2, \ldots X_n, \xi)$.

Then we have $\nabla_X \eta = \alpha(X) \eta$ for every $X \in \tau(M)$.

**Theorem 1.1.** Let $f : M \to \mathbb{R}^{n+1}$ be an affine immersion with transversal vector field $\xi$, $\nabla$ be the the induced connection, $h$ be the affine fundamental form, $S$ be the affine shape operator and $\alpha$ be the the transversal connection then the following equations hold.

1. $R(X,Y)Z = h(Y,Z)SX - h(X,Z)SY$ (Gauss)
2. $(\nabla_X h)(Y, Z) + \alpha(X)h(Y,Z) = (\nabla_Y h)(X, Z) + \alpha(Y)h(X, Z)$ (Codazzi for $h$)
3. $(\nabla_X S)(Y) - \alpha(X)SY = (\nabla_Y S)(X) - \alpha(Y)SX$ (Codazzi for $S$)
4. $h(X, SY) - h(SX, Y) = d\alpha(X,Y)$ (Ricci)

**Definition.** An affine connection $\nabla$ is locally equiaffine if around each point $x$ in $M$ there is a parallel volume element, that is a non-vanishing $n$-form $\omega$ such that $\nabla \omega = 0$.

**Definition.** An affine connection $\nabla$ on $M$ is called equiaffine connection if $\nabla$ admits a parallel volume element $\omega$ on $M$, then we say $(\nabla, \omega)$ is an equiaffine structure.

**Definition.** Two torsion free locally equiaffine connections $\nabla$ and $\nabla'$ on a differentiable manifold $M$ are said to be projectively equivalent if there is a closed 1-form $\rho$ such that

$$\nabla'_X Y = \nabla_X Y + \rho(X)Y + \rho(Y)X, \quad \forall X, Y \in \tau(M).$$

$\nabla$ is called projectively flat if it is projectively equivalent to a flat connection.

**Definition.** Let $\mathbb{R}^{n+1}$ be the affine space with fixed point $o$. A hyper surface $M$ embedded in $\mathbb{R}^{n+1}$ is said to be centro-affine hyper surface if the position vector $\overrightarrow{o x}$ for each $x$ in $M$ is transversal to the tangent plane of $M$ at $x$.

Let $D$ be the standard affine connection on $\mathbb{R}^{n+1}$. Then $(D_X Y)_x = (\nabla^X_X)_x + h(X,Y)(-x), \quad \forall x \in M$, where $(\nabla^X_X)_x$ is tangent to $M$. $\nabla$ is a connection on $M$, called the induced connection and $h$ is called the affine fundamental form. Note that the curvature tensor $R$ and the Ricci tensor $Ric$ of the induced connection on centro affine hypersurface are given by $R(X,Y)Z = h(Y,Z)X - h(X,Z)Y$ and $Ric(Y,Z) = (n-1)h(Y,Z)$ respectively.

**Definition.** Let $f : M \to \mathbb{R}^{n+1}$ be a hyper surface immersion, a transversal vector field $\xi$ is called equiaffine if $D_X \xi$ is tangent to $M$ for each $X \in T_x(M)$. 

3
Definition. Let $f : M \rightarrow \mathbb{R}^{n+1}$ be a hyper surface immersion, a transversal vector field $\xi$ is called non-degenerate if $h$ is non-degenerate everywhere on $M$.

Example. Let $o$ be the point of the affine space $\mathbb{R}^{n+1}$ chosen as origin, $M$ be an $n$-dimensional manifold. Let $f : M \rightarrow \mathbb{R}^{n+1} - \{o\}$ be an immersion such that the position vector $\overrightarrow{of(x)}$ for $x \in M$ is always transversal to $f(M)$. Now take $\xi = -\overrightarrow{of(x)}$, with respect to this $\xi$, $f$ becomes an affine immersion.

We have $D_X \xi = -f_\ast X$ so that $\alpha = 0$ and $S = I$.

Proposition 1.2. Let $(M, \nabla)$ be $n$-manifold with projectively flat affine connection $\nabla$ with symmetric Ricci tensor. Then $(M, \nabla)$ can be locally immersed as a centro-affine hyper surface.

Definition. A transversal vector field is called Blaschke normal field if it satisfies

(a) $(\nabla, \eta)$ is an equiaffine structure, that is $\nabla \eta = 0$.

(b) $\eta$ coincides with the volume element $\omega_h$ of non degenerated metric $h$, where $\eta$ is the induced volume element.

An affine immersion with Blaschke normal field is called Blaschke immersion.

Definition. A Blaschke hyper surface $M$ is called an improper affine hyper shere if the shape operator $S$ is identically zero.

If $S = \lambda I$ where $\lambda$ is a non zero constant, then $M$ is called a proper affine hyper shere.

2. Statistical Manifold

The affine geometric aspects of statistical manifold are studied by Kurose [3, 4] and Uohashi,Ohara, Fujii [10]. Families of probability measures is a subset of the infinite dimensional affine space of the positive measures up to a scale. Parametrized models are surfaces in this affine space. In this section we look at various cases in which the statistical manifold is immersed as a hyper surface in $\mathbb{R}^{n+1}$. We show that every centro-affine hypersurface is a statistical manifold realized in $\mathbb{R}^{n+1}$, dually flat statistical manifold is locally immersed as a centro-affine hypersurface. Also if $f : M \rightarrow \mathbb{R}^{n+1}$ be a Blaschke immersion, then $(M, \nabla, h)$ is a statistical manifold.

Definition. Consider the family $S$ of probability distributions on a sample space $\chi$. Suppose each element of $S$ can be parametrized using $n$ real valued variables $\{\theta^1, \theta^2, \theta^3 ... \theta^n\}$ so that

$$S = \{ p_\theta = p(x, \theta) \mid \theta = (\theta^1, \theta^2, \theta^3 ... \theta^n) \in U\},$$

where $U \subseteq \mathbb{R}^n$ and $\theta \rightarrow p_\theta$ is injective. Such a family $S$ is called $n$ dimensional statistical model. There are some regularity conditions for statistical model which are important for the geometric study.
Regularity conditions

(a) \( U \) is an open subset of \( \mathbb{R}^n \) and for each \( x \in \chi \), the function \( \theta \rightarrow p(x; \theta) \) is of class \( C^\infty \).

(b) Let \( \ell(x; \theta) = \log p(x; \theta) \) and \( \partial_i = \frac{\partial}{\partial \theta_i} \). For every fixed \( \theta \), \( n \) functions in \( x \) \( \{ \partial_i \ell(x; \theta); i = 1, 2, ..n \} \) are linearly independent and are known as scores.

(c) The order of integration and differentiation may be freely interchange.

(d) Moment of scores exists upto necessary orders.

Definition. Let \( S = \{ p(x, \theta) | \theta \in U \subseteq \mathbb{R}^n \} \) be a statistical model, the mapping \( \phi : S \rightarrow \mathbb{R}^n \) defined by \( \phi(p_\theta) = \theta \) allow us to consider \( \phi = [\theta_i] \) as a coordinate system for \( S \). Suppose we have a \( C^\infty \) diffeomorphism \( \psi \) from \( U \) onto \( \psi(U) \subseteq \mathbb{R}^n \). If \( \rho = \psi(\theta) \) then, \( S = \{ p(x, \psi^{-1}(\rho)) \mid \rho \in \psi(U) \} \). By considering parametrization which are \( C^\infty \) diffeomorphic to each other to be equivalent we may consider \( S \) as a smooth manifold, called the statistical manifold.

Definition (Amri’s \( \alpha \)-connection). Let \( S = \{ p(x, \theta) | \theta \in U \subseteq \mathbb{R}^n \} \), be an \( n \)-dimensional statistical manifold \( g \) be the Fisher information metric. Let \( \alpha \in \mathbb{R} \). Define \( n^3 \) functions
\[
\Gamma^\alpha_{ij,k} = E_\theta[(\partial_i \partial_j \ell_\theta + \frac{1-\alpha}{2} \partial_i \ell_\theta \partial_j \ell_\theta) \partial_k \ell_\theta]
\]
Then we have an affine connection \( \nabla^\alpha \) on \( S \) defined by \( \langle \nabla^\alpha_{i,j} \theta, \partial_k \theta \rangle = \Gamma^\alpha_{ij,k} \), called Amri’s \( \alpha \)-connections.

Remark. \( \nabla^\alpha \) is flat iff \( \nabla^{-\alpha} \) is flat. If \( \alpha = 1 \) then \( \nabla^1 \) is called exponential connection or \( e \)-connection and we denote it by \( \nabla \) or \( \nabla^e \), \( \nabla^{-1} \) called \( m \)-connection is also denoted by \( \nabla^m \).

Definition (Lauritzen). A statistical manifold is a pseudo-Riemannian manifold \( (M, h) \) with a symmetric covariant tensor \( \nabla h \) of order 3.

Definition. Let \( M \) be an \( n \)-dimensional manifold and \( h \) be a non-degenerate metric and \( \nabla \) be an affine connection. We define a connection \( \nabla \) by \( X h(Y, Z) = h(\nabla_X Y, Z) + h(Y, \nabla_X Z) \).

This connection \( \nabla \) is called the conjugate connection. Note that the torsion tensor \( T \) and \( T \) of \( \nabla \) and \( \nabla \) are respectively satisfy
\[
(\nabla_X h)(Y, Z) + h(Y, T(X, Z)) = (\nabla_Y h)(X, Z) + h(Y, \nabla_T(X, Z))
\]
(1)

Then \( \nabla \) is torsion free if and only if \( (\nabla, h) \) satisfies the Codazzi’s equation
\[
(\nabla_X h)(Y, Z) = (\nabla_Z h)(Y, X), \quad \forall X, Y, Z \in \tau(M).
\]
(2)

Also the curvature tensor \( R \) and \( \nabla \) of \( \nabla \) and \( \nabla \) are related by
\[
h(R(X, Y)Z, U) = -h(Z, \nabla(R(X, Y)U),
\]
(3)
so \( R = 0 \) if and only if \( \nabla R = 0 \).
Definition. Let \( M \) be an \( n \)-dimensional smooth manifold with affine connection \( \nabla \). Then we have a connection in cotangent bundle \( T^*(M) \) called the dual connection denoted by \( \nabla^* \) defined as

\[
(\nabla^*_X \omega)(Y) = X(\omega(Y)) - \omega(\nabla_X Y)
\]  

(4)

where \( \omega \) is a 1-form.

Let \( M \) be a manifold with non-degenerate metric \( h \). Then we can identify the tangent bundle \( T(M) \) and cotangent bundle \( T^*(M) \) in the following way. For each \( x \in M \) we have a linear isomorphism \( \lambda_h : T(M) \rightarrow T^*(M) \) defined by

\[
\lambda_h(X)(Y) = h(X, Y) \quad \forall X, Y \in T_x(M)
\]  

(5)

Remark. Let \( M \) be a manifold with pseudo-Riemannian metric \( h \) and \( \nabla \) be an affine connection on \( M \) such that \( \nabla h \) is symmetric. The conjugate connection \( \nabla^* \) and dual connection \( \nabla^* \) in \( T^*(M) \) are correspond to each other by the isomorphism \( \lambda_h \). Thus for an \( n \) dimensional manifold \( (M, \nabla, h) \) with \( \nabla h \) symmetric we have a dual connection \( \nabla^* \) on \( T(M) \) so that we have the duality structure \( (M, \nabla, \nabla^*, h) \).

On a statistical manifold we always have a duality structure \( \nabla^* \). Also \( (M, \nabla) \) is flat if and only if \( (M, \nabla^*) \) is flat, in that case we call \( (M, h, \nabla, \nabla^*) \) is a dually flat space.

Definition. For a real number \( \alpha \) two statistical manifolds \( (\tilde{M}, \tilde{\nabla}, \tilde{h}) \) and \( (M, \nabla, h) \), are said to be \( (\alpha) \)-conformally equivalent if there exist a function \( \phi \) on \( M \) such that

\[
\tilde{h}(X, Y) = e^{\phi} h(X, Y),
\]

\[
\tilde{h}(\tilde{\nabla}_X Y, Z) = h(\nabla_X Y, Z) - \frac{1 + \alpha}{2} d\phi(Z) h(X, Y) + \frac{1 - \alpha}{2} \{ d\phi(X) \tilde{h}(Y, Z) + d\phi(Y) \tilde{h}(X, Z) \}.
\]

A statistical manifold is said to be \( (\alpha) \)-conformally flat if it is \( (\alpha) \)-conformally equivalent to a flat connection.

Note. In [3, 4] Kurose discussed about affine immersion of statistical manifolds into the affine space.

Let \( f : M \rightarrow \mathbb{R}^{n+1} \), be an affine immersion with transversal vector field \( \xi \) and induced connection \( \nabla \). From the definition of non-degenerate and equiaffine transversal vector fields, \( (M, \nabla, h) \) is a statistical manifold if and only if \( \xi \) is non-generate and equiaffine, in this case we say that \( (M, \nabla, h) \) realize in \( \mathbb{R}^{n+1} \) and by proposition(9.1) in [11] statistical manifold \( (M, \nabla, h) \) is \((-1)\)-conformally flat if and only if \( \nabla \) is a projectively flat connection with symmetric Ricci tensor. Also from [3] a simply connected statistical manifold \( (M, \nabla, h) \) can be realized in \( \mathbb{R}^{n+1} \) if and only if \( \nabla^* \) is projectively flat.

Proposition 2.1. Every centro-affine hypersurface is a statistical manifold realized in \( \mathbb{R}^{n+1} \).

Proof. Let \( \mathbb{R}^{n+1} \) be an affine space with fixed point \( o \), \( M \) be an \( n \)-dimensional manifold. Let \( f : M \rightarrow \mathbb{R}^{n+1} - \{ o \} \) be an immersion such that the position vector \( \overrightarrow{o \cdot f(x)} \) for \( x \in M \) is always transversal to \( f(M) \). Now take \( \xi = \overrightarrow{f(x) \cdot o} \).
with respect to this ξ, f becomes an affine immersion and we have \( D_X\xi = -f_*(X) \) (negative sign comes because of orientation), so the transversal connection α = 0 and the shape operator \( S = I \). Then from the Codazzi for \( h, \nabla h \) is symmetric. So \((M, \nabla, h)\) is a statistical manifold realized in \( \mathbb{R}^{n+1} \), where \( \nabla \) and \( h \) are the induced connection and affine fundamental form respectively. \( \square \)

**Proposition 2.2.** Let \((M, h, \nabla, \nabla^*)\) is a dually flat statistical manifold. Then both \((M, \nabla)\) and \((M, \nabla^*)\) can be locally immersed as a centro-affine hypersurface.

*Proof.* Since \((M, \nabla)\) is a flat manifold it is \((α)\)-conformally flat for all \( α \) by taking \( φ = 0 \) in the definition of \((α)\)-conformally equivalent. So in particular \((M, \nabla)\) is \((-1)\)-conformally flat. Then \( \nabla \) is projectively flat connection with symmetric Ricci tensor. Then from the proposition \((1.2)\) \((M, \nabla)\) is locally immersed as a centro-affine hypersurface. Similarly we can prove for \((M, \nabla^*)\). \( \square \)

**Proposition 2.3.** Let \( f : M \rightarrow \mathbb{R}^{n+1} \) be a Blaschke immersion. Then \((M, \nabla, h)\) is a statistical manifold, where \( \nabla \) is the induced connection and \( h \) is the induced affine fundamental form.

*Proof.* Let \( f : M \rightarrow \mathbb{R}^{n+1} \) be a Blaschke immersion. Then by the definition of Blaschke immersion \( \nabla \eta = 0 \). Since \( \nabla_X \eta = α(X)\eta \) for all vector field \( X \), we have \( α = 0 \). Then from the Codazzi for \( h, \nabla h \) is symmetric and it follows that \((M, \nabla, h)\) is a statistical manifold. \( \square \)

### 3. Exponential family

The set of probability measures is a subset of the affine space of positive measures up to scale. Parametrized family of probability measures can be considered as a surface in this affine space of positive measures up to scale. In particular the geometry of exponential family is the simplest geometry, the affine geometry. Exponential families are characterized as the finite dimensional affine subspaces of the affine space of measures up to scale.

In this section, based on the \(±\)-flat structure of exponential family, we explore certain geometric properties of exponential family. We show that if all \( \nabla^e \) auto parallel submanifold of a statistical manifold \( S \) are exponential family, then \( S \) is an exponential family. We prove that submanifold of a statistical manifold which is an exponential family is \( \nabla^e \)-auto parallel submanifold. Also observed that exponential family is locally immersed as a centro-affine hypersurface.

**Definition.** Let \( S \) and \( M \) be manifolds with \( M \subset S \). Let \([\theta_i] = (θ_1, θ_2, ... θ_n)\) and \([u_a] = (u_1, u_2, ... u_m)\) be coordinate system for \( S \) and \( M \) respectively, where \( n = \text{dim}(S) \) and \( m = \text{dim}(M) \). We call \( M \) a submanifold of \( S \) if the following conditions hold.

- The restriction \( \theta_i |_M \) to \( M \) is a \( C^∞ \) functions on \( M \), \( i = 1, 2, ... n \).
- Let \( B^i_a = \left( \frac{∂θ_i}{∂u_a} \right) \big|_p \) and \( B_a = [B^1_a, B^2_a ... B^n_a] \in \mathbb{R}^n \). Then for each \( p \in M \), \( \{B_1, B_2, ..., B_m\} \) are linearly independent. (so, \( m \leq n \)).
• For any open subset $W$ of $M$, there exists $U$ an open subset of $S$, such that $W = M \cap U$.

**Definition.** Let $M$ be a submanifold of $S$ and $\nabla$ be an affine connection on $S$. $M$ is said to be auto parallel with respect to $\nabla$ if $\nabla_X Y \in \tau(M)$ for all $X, Y \in \tau(M)$.

1-dimensional auto parallel submanifolds are called geodesics.

**Remark.** A necessary and sufficient condition for $M$ to be auto parallel is that $\nabla_{\partial_a} \partial_b \in \tau(M)$ holds for all $a, b$, where $\partial_a = \frac{\partial}{\partial u_a}$.

**Definition.** Let $M$ be a submanifold of $S$. Let $p \in M$, then $T_p M \subset T_p S$. Now consider the projection map $\pi_p : T_p M \rightarrow T_p S$ and $\pi_p(V) = V$, $\forall V \in T_p M$. Let $\nabla$ be a connection on $S$, then we define a connection $\nabla^\pi$ on $M$ as

$$(\nabla^\pi_X Y)_p = \pi_p(\nabla_X Y)_p, \forall p \in M, \forall X, Y \in \tau(M)$$

Define the second fundamental form or Embedding curvature as $H(X, Y) = \nabla_X Y - \nabla^\pi_X Y$

**Remark.** Let $S$ be an $n$-dimensional manifold and $M$ be a submanifold of dimension $m$. For each $p \in S$, let $\{(\partial_a)_p; 1 \leq a \leq m\}$, be the basis for $T_p M$ and let $\{(\partial_k)_p; m + 1 \leq k \leq n\}$, be the basis for $T_p M^\perp$. 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 $H_{abk} = 0$ for $1 \leq a, b \leq m$ and $m + 1 \leq k \leq n$. Also we have $H(X, Y) = 0$ iff $M$ is $\nabla$ auto parallel submanifold of $S$.

**Definition.** Let $\mathcal{X}$ be a sample space and $\mu$ be a measure on $\mathcal{X}$. The exponential family is the family of probability distributions $p(x, \theta)$ expressed in the following form

$$p(x, \theta) = \exp(\sum_{i=1}^{r} \theta_i x^i - K(\theta))d\mu,$$

where $\{x^1, x^2, .....x^r\}$ are random variables, $\theta = \{\theta_1, \theta_2, .....\theta_r\}$ are parameters and $K(\theta)$ is the normalizer.

**Example.** Let

$$p(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

be the normal family, then take

$$\theta_1 = \frac{-1}{2\sigma^2}, \quad \theta_2 = \frac{\mu}{\sigma^2}, \quad \text{and} \quad K(\theta) = \frac{\log(\frac{\sigma}{\theta_1})}{2} - \frac{(\theta_2)^2}{4\theta_1}$$

Then we can easily see $p(x, \mu, \sigma) = p(\theta_1, \theta_2) = \exp(x^2\theta_1 + x\theta_2 - K(\theta))$

**Note.** Exponential family is an affine space, it is also flat with respect to $\nabla^{\pm 1}$ connection.
A parametrized model which is flat with respect to ±1 connection need not be an exponential family.

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

$$Y = (y^1, y^2, y^3, \ldots y^k)^t$$

we have

$$q^k(Y) = q(y^1)q(y^2)q(y^3), \ldots q(y^k)$$

(6)

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

$$p(A, \mu, x) = \frac{q^k(A^{-1}(x - \mu))}{|\text{det}(A)|}$$

(7)

Now consider

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

(8)

Now define a statistical model

$$\log(p(A, \mu, x)) = \sum_{i=1}^{k} \log(q(A^{-1}(x - \mu))) - \log(|\text{det}(A)|)$$

(9)

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

**Theorem 3.1.** Let $\mathbf{S}$ be an exponential family and $\mathbf{M}$ be a submanifold of $\mathbf{S}$. Then $\mathbf{M}$ is exponential family if and only if $\mathbf{M}$ is auto parallel with respect to $\nabla^1$ in $\mathbf{S}$.

**Theorem 3.2.** Let $S = \{P(x, \theta) \mid \theta \in \Theta\}$, be an $n$-dimensional statistical manifold with dually flat structure $(S, \nabla^e, \nabla^m)$. If all $\nabla^e$ auto parallel 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 $(S, \nabla^e, \nabla^m)$, where $g$ is the fisher information metric. Let $\theta = [\theta^i]$ and $\eta = [\eta_j]$ be the coordinate system of $S$ with respect to $\nabla^e$ and $\nabla^m$ respectively. Now subdivide the range of index $i = 1, 2, \ldots n$ into indexing sets $I = \{i = 1, 2, \ldots k\}$ and $II = \{i = k + 1, k + 2, \ldots n\}$. Let $M(C_{11})$ be the set of points whose coordinates $[\theta^i]$ in $II$ are fixed to constant $C_{11} = (C_{11}^i)$ for $i = k + 1, k + 2, \ldots n$. That is

$$M(C_{11}) = \{p \in S \mid \theta^{k+1} = C_{11}^{k+1}, \theta^{k+2} = C_{11}^{k+2}, \ldots \theta^n = C_{11}^n\}$$

(10)

Where $C_{11} \in \mathbb{R}^{n-k}$, then clearly this is an affine space with respect to $\theta$- coordinate system, which implies $M(C_{11})$ is a $\nabla^e$- auto parallel submanifold of $S$. Also if $C_{11} \neq C_{11}'$ then $M(C_{11}) \cap M(C_{11}') = \phi$, and $\bigcup_{C_{11}} M(C_{11}) = S$.

Now by our assumption $M(C_{11})$ is an exponential family for all $C_{11}$. If $p(x, \theta) \in S$, then $p(x, \theta) \in M(C_{11})$ for some constant $C_{11}$, this implies that

$$p(x, \theta) = \exp(\sum_{i=1}^{k} \theta^i x_i - \psi^0(\theta))$$

(11)
where $\psi^\beta(\theta)$ defined on $\Theta^\beta = \{ \theta \in \Theta \mid \theta^{k+1} = C_{I1}^{k+1}, \theta^{k+2} = C_{I1}^{k+2}, \ldots \theta^n = C_{I1}^n \}$. Now define $\phi(\theta) = \psi^\beta(\theta)$ if $\theta \in \Theta^\beta$. Then we can write

$$p(x, \theta) = \exp \left( \sum_{i=1}^k \theta^i x_i - \phi(\theta) \right)$$

(13)

$$= \exp \left( \sum_{i=1}^k \theta^i x_i + \sum_{i=k+1}^n C_{I1}^i x_i - \sum_{i=1}^k C_{I1}^i x_i - \phi(\theta) \right)$$

(14)

$$= \exp \left( \sum_{i=1}^n \theta^i x_i + F(x) - \phi(\theta) \right)$$

(15)

where $F(x) = - \sum_{i=1}^k C_{I1}^i x_i$ for $p(x, \theta) \in M(C_{I1})$, then $S$ is an exponential family.

\[ \square \]

**Theorem 3.3.** Let $S = \{ p(x, \theta) \mid \theta \in \Theta \}$ be a statistical manifold with $\nabla^e$ connection. Let $M$ be a submanifold of $S$. If $M$ is an exponential family then $M$ is $\nabla^e$ auto parallel submanifold of $S$.

**Proof.** Let $S = \{ p(x, \theta) \mid \theta \in \Theta \}$ and $M = \{ q(x, u) \}$ be the 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

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

(16)

we have, $\Gamma^1_{ab,k} = E_\xi [\partial_a \partial_b \partial_\theta \partial_k \ell_\theta], \text{ where } \ell_\theta = \log(p(x, \theta))$. Then $\partial_a \partial_b \partial_\theta \ell_\theta = -\frac{\partial^2 \phi}{\partial u_a \partial u_b}$. Therefore we have,

$\Gamma^1_{ab,k} = 0 \text{ which implies } \langle \nabla^e_\partial_\theta, \partial_\theta \rangle = 0, \forall k$. Hence $H_{abk} = 0$, which implies that $M$ is a $\nabla^e$ auto parallel submanifold of $S$.

\[ \square \]

**Note.** From the proposition(2.2), exponential family with $\pm 1$ connection is locally immersed as a centro-affine hypersurface

**Acknowledgements**

The first author expresses his sincere gratitude to Indian institute of Space Science and technology, department of space, government of India for supporting his research work by providing fellowship and other necessary means.

**References**

**References**

[1] S.-i. Amari, Information geometry and its applications, Springer, 2016.

[2] S. L. Lauritzen, Statistical manifolds, JSTOR, 1984.
[3] T. Kurose, Dual connections and affine geometry, Mathematische Zeitschrift 203 (1) (1990) 115–121.

[4] T. Kurose, On the divergences of 1-conformally flat statistical manifolds, Tohoku Mathematical Journal, Second Series 46 (3) (1994) 427–433.

[5] F. Dillen, K. Nomizu, L. Vranken, Conjugate connections and radon’s theorem in affine differential geometry, Monatshefte für Mathematik 109 (3) (1990) 221–235.

[6] S.-i. Amari, H. Nagaoka, Method of information geometry. ams monograph (2000).

[7] M. Murray, J. Rice, Differential geometry and statistics. number 48 in monographs on statistics and applied probability (1993).

[8] K. Nomizu, T. Sasaki, Affine differential geometry: geometry of affine immersions, Cambridge University Press, 1994.

[9] A. Nihat, J. Jurgen, L. Hong, Van, S. Lorenz, Information geometry, Springer, 2017.

[10] K. Uohashi, A. Ohara, T. Fujii, 1-conformally flat statistical submanifolds.

[11] K. Nomizu, U. Simon, Notes on conjugate connections, in: Geometry and topology of submanifolds IV. Proc. Conf. Di. Geom. Vision, Leuven, World Scientific, 1992, pp. 152–173.