Information Geometry - Hessian Geometry

Robert A. Wolak, JU
joint work with Michel Nguiffo Boyom

Contents

Contents

Contents

Contents

General

Probability distribution on a set X is a non-negative real function function

$$p\colon X \to \mathbb{R}$$

1) if X discrete and countable $\Sigma_{x \in X} p(x) = 1$;

2) if $X = \mathbb{R}^n$ $\int_X p(x)dx = 1$.

p is a *probability density function.*

Let Λ be a domain in R^m. We consider families of probability distributions on a set $\mathcal{X}$ parametrized by $\lambda \in \Lambda$.

$$\mathcal{P} = \{p(x;\lambda) | \lambda \in \Lambda\}$$

(1) Λ is a domain in R^m,
(2) $p(x;\lambda)$ for a fixed x is a smooth function in λ,
(3) the operation of integration with respect to x and differentiation with respect to λ are commutative.

Λ is called an m-dimensional statistical model (parametric model).
Notation $\Lambda = \{p(x;\lambda)\} = \{p_\lambda(x)\}$, $p(x;\lambda) = p_\lambda(x)$

Examples

Example (Normal distribution)

$X = \mathbb{R}$, m=2

$$\Lambda = \{(\mu, \sigma) \colon -\infty < \mu < \infty, 0 < \sigma < \infty\}$$

$$p(x; \lambda) = \frac{1}{\sqrt{2\pi\sigma}} exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}$$

Example (Multivariate normal distribution)

$X = \mathbb{R}^k$, $m = k + \frac{k(k+1)}{2}$, $\lambda = (\mu, \Sigma)$

$$\Lambda = \{(\mu, \Sigma) \colon \mu \in \mathbb{R}^k, \quad \Sigma \in \mathbb{R}^{k^2} : positive \;\; definite\}$$

$$p(x; \lambda) = (2\pi)^{-k/2} (det\Sigma)^{-1/2} exp\{-\frac{1}{2}(x-\mu)^t \Sigma^{-1} (x-\mu)\}$$

Examples cont.

Example (Poisson distribution)

$X = \mathbb{N}$, $m = 1$, $\Lambda = (0, \infty)$

$$p(x; \lambda) = e^{-\lambda} \frac{\lambda^x}{x!}$$

Example ($P(X)$ for finite X)

$X = \{x_0, x_1, ..., x_n\}$, $\Lambda = \{(\lambda^1, ..., \lambda^n) \colon \lambda^i > 0, \;\; \Sigma_{i=1}^n \lambda^i < 1\}$

$$p(x; \lambda) = \begin{cases} \lambda^i & 1 \leq i \leq n \\ 1 - \Sigma_{i=1}^n \lambda^i & i = 0 \end{cases}$$

Fisher metric

Definition

Let $\mathcal{P} = \{p(x;\lambda) | \lambda \in \Lambda\}$ be a family of probability distributions on a set $\mathcal{X}$ parametrized by $\lambda \in \Lambda$.

We set $l_\lambda = l(x;\lambda) = logp(x;\lambda)$ and denote by E_λ the expectation with respect to $p_\lambda(x) = p(x;\lambda)$.

Then the matrix $g_F(\lambda) = [g_{ij}(\lambda)]$ defined by

$$g_{ij}(\lambda) = E_\lambda[\frac{\partial l_\lambda}{\partial \lambda^i}\frac{\partial l_\lambda}{\partial \lambda^j}] = \int_{\mathcal{X}} \frac{\partial l(x;\lambda)}{\partial \lambda^i}\frac{\partial l(x;\lambda)}{\partial \lambda^j} p(x;\lambda)dx$$

is called the Fisher information matrix tensor.

Fisher metric cont.

Simple calculations show that

$$g_{ij}(\lambda) = -E_\lambda[\frac{\partial^2 l_\lambda}{\partial\lambda^i\partial\lambda^j}].$$

The Fisher information matrix tensor $g_F(\lambda) = [g_{ij}(\lambda)]$ is positive semi-definite on Λ:

$$\Sigma_{i,j} g_{ij}(\lambda) c^i c^j = \int_{\mathcal{X}} \{\Sigma_i c^i \frac{\partial l(x;\lambda)}{\partial\lambda^i}\}^2 p(x;\lambda) dx \geq 0.$$

In information geometry the standard assumption has been:

(4) For a family of probability distributions $\mathcal{P} = \{p(x;\lambda) | \lambda \in \Lambda\}$ the Fisher information matrix tensor $g_F(\lambda) = [g_{ij}(\lambda)]$ is positive definite on Λ.

Remark *The general case seems to be difficult to study if not hopeless. Therefore to develop a meaningful more general theory in our paper* **[BW]** *we assume that the* **Fisher information matrix tensor is parallel with respect to some torsion-free connection** *on Λ. The condition permits us to construct a foliation, and under some reasonable assumptions it has a transverse Hessian structure.*

Connections

$$(\Gamma^{(\alpha)}_{ij,k})_\lambda = E_\lambda[(\partial_i\partial_j l_\lambda + \frac{1-\alpha}{2}\partial_i l_\lambda \partial_j l_\lambda)(\partial_k l_\lambda)]$$

where α is an arbitrary real number.

We define an affine connection $\nabla^{(\alpha)}$ on Λ by

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

$\nabla^{(\alpha)}$ is called the α**-connection.** $\nabla^{(\alpha)}$ is a symmetric connection.

$$\Gamma^{(\beta)}{}_{ij,k} = \Gamma^{(\alpha)}{}_{ij,k} + \frac{\alpha-\beta}{2} T_{ijk},$$

where T_{ijk} is a covariant symmetric tensor of degree 3 defined by

$$(T_{ijk})_\lambda = E_\lambda[\partial_i l_\lambda \partial_j l_\lambda \partial_k l_\lambda].$$

Moreover,

$$\nabla^{(\alpha)} = (1-\alpha)\nabla^{(0)} + \alpha\nabla^{(1)} = \frac{1+\alpha}{2}\nabla^{(1)} + \frac{1-\alpha}{2}\nabla^{(-1)}$$

Connections cont.

Theorem

The 0-connection is the Riemannian connection with respect to the Fisher metric.

Exponential family

If an *m*-dimensional model

$$S = \{p_\theta : \theta \in \Theta\}$$

can be expressed in terms of functions $\{C, F_1, ..., F_m\}$ on X and a function ψ on Θ:

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

then S is called an **exponential family** and $\{\theta^i\}$ are called natural or canonical parameters.
From the normality condition

$$\psi(\theta) = log \int exp[C(x) + \Sigma_{i=1}^{n} \theta^i F_i(x)] dx.$$

The parametrization $\theta \mapsto p_\theta$ is one-to-one iff the m+1 functions $\{F_1, ..., F_m, 1\}$ are linearly independent.
Always assumed!

Examples

Example (Normal distribution)

$$C(x) = 0, \quad F_1(x) = x, \quad F_2(x) = x^2, \quad \theta^1 = \frac{\mu}{\sigma^2} \quad \theta^2 = -\frac{1}{2\sigma^2}$$

$$\psi(\theta) = \frac{\mu^2}{2\sigma^2} + log(\sqrt{2\pi\sigma}) = -\frac{(\theta^1)^2}{4\theta^2} + \frac{1}{2} log(-\frac{\pi}{\theta^2}).$$

Example

Consider an m-dimensional model $S = \{p_\theta\}$ which can be expressed in terms of function $\{C, F_1, ..., F_m\}$ on X as

$$p(x;\theta) = C(x) + \Sigma_i \theta^i F_i(x).$$

S forms an affine subspace of $P(X)$.
S is called a **mixture family** with **mixture parameters** θ^i.

Examples

Example (Normal distribution)

$$C(x) = 0, \quad F_1(x) = x, \quad F_2(x) = x^2, \quad \theta^1 = \frac{\mu}{\sigma^2} \quad \theta^2 = -\frac{1}{2\sigma^2}$$

$$\psi(\theta) = \frac{\mu^2}{2\sigma^2} + log(\sqrt{2\pi\sigma}) = -\frac{(\theta^1)^2}{4\theta^2} + \frac{1}{2} log(-\frac{\pi}{\theta^2}).$$

Example

Consider an m-dimensional model $S = \{p_\theta\}$ which can be expressed in terms of function $\{C, F_1, ..., F_m\}$ on X as

$$p(x;\theta) = C(x) + \Sigma_i \theta^i F_i(x).$$

S forms an affine subspace of $P(X)$.
S is called a **mixture family** with **mixture parameters** θ^i.

Example (Multivariate normal distribution)

$$C(x) = 0, \quad F_i(x) = x_i, \quad F_{ij}(x) = x_i x_j \quad (i \leq j)$$

$$\theta^i = \Sigma_j (\Sigma^{-1})^{ij} \mu_j, \quad \theta^{ii} = (-1/2)(\Sigma^{-1})^{ii}, \quad \theta^{ij} = -(\Sigma^{-1})^{ij} \quad (i < j)$$

and

$$F_A(x) = x, \quad F_B(x) = xx^t, \quad \theta^A = \Sigma^{-1}\mu, \quad \theta^B = (-1/2)\Sigma^{-1},$$

We have

$$p(x;\theta) = exp[\Sigma_{1 \leq i \leq k} \theta^i F_i(x) + \Sigma_{1 \leq i \leq j \leq k} \theta^{ij} F_{ij}(x) - \psi(\theta)]$$

$$= exp[(\theta^A)^t F_A(x) + tr(\theta^B F_B(x)) - \psi(\theta)]$$

where $\psi(\theta) = ...$

Theorem

An exponential family (a mixture family, respectively) is $\nabla^{(1)}$-flat ($\nabla^{(-1)}$-flat, respectively) and its natural parameters (mixture parameters, respectively) form a $\nabla^{(1)}$-affine ($\nabla^{(-1)}$-affine, respectively) coordinate system.

Theorem

Let S be an exponential family (a mixture family, respectively) and M a submanifold of S. Them M is an exponential family (a mixture family, respectively) iff M is $\nabla^{(1)}$-autoparallel ($\nabla^{(-1)}$-autoparallel) in S.

Theorem

An exponential family (a mixture family, respectively) is $\nabla^{(1)}$-flat ($\nabla^{(-1)}$-flat, respectively) and its natural parameters (mixture parameters, respectively) form a $\nabla^{(1)}$-affine ($\nabla^{(-1)}$-affine, respectively) coordinate system.

Theorem

Let S be an exponential family (a mixture family, respectively) and M a submanifold of S. Them M is an exponential family (a mixture family, respectively) iff M is $\nabla^{(1)}$-autoparallel ($\nabla^{(-1)}$-autoparallel) in S.

Contents

Dual connections

When investigating the properties of the Fisher metric g and the α-connection $\nabla^{(\alpha)}$ it is important to consider them not individually, but rather as the triple $(g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$. The reason for this is that, through g, there exists a kind of duality between $\nabla^{(\alpha)}$ and $\nabla^{(-\alpha)}$ which is of fundamental significance. This notion of duality emerges not only when considering statistical models but also in many different problems related to information geometry.

(S, g) a Riemannian manifold, ∇ and ∇^* two connections.

Definition

If for any $X, Y, Z \in \mathcal{X}(S)$

$$Zg(X, Y) = g(\nabla_Z X, Y) + g(X, \nabla^*_Z Y)$$

holds then the connections ∇ and ∇^* are said to be dual (or conjugate). The triple (g, ∇, ∇^*) is called a dualistic structure on S.

In local coordinates we have

$$\partial_k g_{ij} = \Gamma_{ki,j} + \Gamma^*_{kj,i}$$

For given g and ∇ there exists a unique dual connection ∇^*

Moreover,

1. $(\nabla^*)^* = \nabla$,
2. $(\nabla + \nabla^*)/2$ is a metric connection,
3. if a connection ∇' has the same torsion as ∇^* and if $(\nabla + \nabla')/2$ is metric, then $\nabla' = \nabla^*$.

Theorem

For any statistical model, the (α)-connection and the $(-\alpha)$-connection are dual with respect to the Fisher metric.

Theorem

Let $h_\gamma : T_pS \to T_qS$ (resp. h_γ^* be the parallel transport along curve γ from p to q with respect to ∇ (resp. ∇^*), then

$$g(h_\gamma(X), h_\gamma^*(Y)) = g(X, Y)$$

for any vectors $X, Y \in T_pS$.

For any vector fields $X, Y, Z, W \in \mathcal{X}(X)$

$$g(R(X,Y)Z, W) = -g(R^*(X,Y)W, Z)$$

thus

$$R = 0 \quad iff \quad R^* = 0.$$

However, a similar property does not hold for the torsion tensors.

Let (g, ∇, ∇^*) be a dualistic structure on a manifold S. If the connections ∇ amd and ∇^* are both symmetric ($T = T^* = 0$), then the ∇-flatness and ∇^*-flatness are equivalent.

Since the α-connections are always symmetric, for any statistical model S and for any real number α S is α-flat iff S is $(-\alpha)$-flat.

We call (S, g, ∇, ∇^*) a dually flat space if both dual connections are flat.

Theorem

Let (S, g, ∇, ∇^*) be a dually flat space. If a submanifold M of S is autoparallel with respect to either ∇ or ∇^*, then M is a dually flat space with respect to the dualistic structure $(g_M, \nabla_M, \nabla^*_M)$ induced on M by (g, ∇, ∇^*).

$\hat{\xi}\colon X \to \mathbb{R}^m$ is called an **estimator.**

$\hat{\xi}$ is called an unbiased estimator if $E_\xi[\hat{\xi}(X)] = \xi$ for any ξ.

The mean squared error of an inbiased estimator $\hat{\xi}$ may be expressed as the variance-covariance matrix $V_\xi[\hat{\xi}] = [v_\xi^{ij}]$ where

$$v_\xi^{ij} = E_\xi[(\hat{\xi}^i(X) = \xi^i)(\hat{\xi}^j(X) = \xi^j)]$$

An unbiased estimator $\hat{\xi}$ achieving the equality $V_\xi[\hat{\xi}] = G(\xi)^{-1}$ for all ξ is called an **efficient estimator.**

Theorem

A necessary and sufficient condition for a coordinate system ξ of a model $S = \{p_\xi\}$ to have an efficient estimator is that S is an exponential family and ξ is (-1)-affine.

Contents

Hessian structures

Definition

A Riemannian metric g on a flat manifold (M, D) is called a Hessian metric if for any point x of M there exists a local function ϕ defined on an open nbhd of x such that

$$g = Dd\phi$$

If $(x^1, ..., x^m)$ is an affine coordinate system for D then

$$g_{ij} = \frac{\partial^2 \phi}{\partial x^i \partial x^j},$$

The pair (D, g) is called a **Hessian structure** on M; M is called **a Hessian manifold** - notation (M, D, g). A function ϕ is called a (local) potential of (D, g).

Definition

A Hessian structure (D,g) is said to be of Koszul type if there exists a closed 1-form ω such that $g = D\omega$.

Hessian structures cont.

Let ∇ be the Levi-Civita connection of the Riemannian metric g.
Let γ be the difference tensor

$$\gamma_X Y = \nabla_X Y - D_X Y$$

As ∇ and D are torsion-free $\gamma_X Y = \gamma_Y X$.

Proposition

Let (M,D) be a flat manifold and g a Riemannian metric on M. Then the following conditions are equivalent:

1. g is a Hessian metric,
2. $(D_X g)(Y, Z) = (D_Y g)(X, Z)$,
3. $\frac{\partial g_{ij}}{\partial x^k} = \frac{\partial g_{kj}}{\partial x^i}$,
4. $g(\gamma_X Y, Z) = g(Y, \gamma_X Z)$,
5. $\gamma_{ijk} = \gamma_{jik}$.

cont.

Let (M, D) be a flat manifold and $\pi : TM \to M$ its tangent bundle.
To an affine chart $(x^1, ..., x^m)$ we associate a complex chart on TM

$$z^j = \xi^j + i\xi^{m+j}$$

where $\xi^i = x^i\pi$ and $\xi^{m+i} = dx^i$ for $i = 1, ..., m$. J_D the associated complex structure on TM.

On TM we define the following Riemannian metric g^T

$$g^T = \Sigma g_{ij}\pi dz^i d\bar{z}^j$$

Proposition

Let (M, D) be a flat manifold and g a Riemannian metric on M. Then the following conditions are equivalent:
(1) g is a Hessian metric on (M, D),
(2) g^T is a Kählerian metric on (TM, J_D).

cont.

Theorem

Let (M,D,g) be a Hessian manifold and let ∇ be the Levi-Civita connection of g. Define a connection D' by

$$D' = 2\nabla - D.$$

Then
(1) D' is a flat connection,
(2) $Xg(Y,Z) = g(D_X Y, Z) + g(Y, D'_X Z)$,
(3) (D',g) is a Hessian structure.

Codazzi structures

Proposition

Let D be a torsion-free connection and let g be a Riemannian metric. Let D' be a new connection defined by

$$Xg(Y,Z) = g(D_X Y, Z) + g(Y, D'_X Z)$$

Then the following conditions are equivalent:
(1) the connection D' is torsion-free,
(2) The pair (D, g) satisfies the Codazzi equation

$$(D_X g)(Y,Z) = (D_Y g)(X,Z),$$

(3) Let ∇ be the Levi-Civita connection for g, let $\gamma_X Y = \nabla_X Y - D_X Y$. Then

$$g\gamma_X Y, Z) = g(Y, \gamma_X Z).$$

If the pair (D, g) satisfies the Codazzi equation, so does the pair (D', g) and

$$D' = 2\nabla - D \quad and \quad (D_X g)(Y,Z) = 2g(\gamma_X Y, Z).$$

Codazzi structures cont.

Definition

A pair (D, g) where D is a torsion-free connection and g a Riemannian metric on a manifold M is called a **Codazzi structure** if it satisfies the Codazzi

$$(D_X g)(Y, Z) = (D_Y g)(X, Z),$$

For a Codazzi structure (D, g) the connection D' defined by

$$Xg(Y, Z) = g(D_X Y, Z) + g(Y, D'_X Z)$$

is called the **dual connection** of D with respect to g, and the pair (D', g) the **dual Codazzi structure** of (D, g).

Definition

A Codazzi structure (D, g) is of constant curvature c if the curvature tensor R_D of D satisfies

$$R_D(X, Y)Z = c\{g(Y, Z)X - g(X, Z)Y\}.$$

Codazzi structures cont.

Definition

A pair (D, g) where D is a torsion-free connection and g a Riemannian metric on a manifold M is called a **Codazzi structure** if it satisfies the Codazzi

$$(D_X g)(Y, Z) = (D_Y g)(X, Z),$$

For a Codazzi structure (D, g) the connection D' defined by

$$Xg(Y, Z) = g(D_X Y, Z) + g(Y, D'_X Z)$$

is called the **dual connection** of D with respect to g, and the pair (D', g) the **dual Codazzi structure** of (D, g).

Definition

A Codazzi structure (D, g) is of constant curvature c if the curvature tensor R_D of D satisfies

$$R_D(X, Y)Z = c\{g(Y, Z)X - g(X, Z)Y\}.$$

Proposition

Let (D, g) be a Codazzi structure nad (D', g) its dual Codazzi structure. Then (1)

$$g(R_D(X, Y)Z, W) + g(Z, R_{D'}(X, Y)W) = 0;$$

(2) if (D, g) is a Codazzi struccture of constant curvature c, then (D', g) is also of constant curvature c.

Proposition

A Codazzi structure (D, g) is of constant curvature 0 iff (D, g) is a Hessian structure.

Proposition

Let (D, g) be a Codazzi structure of constant curvature. Then locally g is of the form

$$Dd\phi + \frac{\phi}{m-1} Ric_D$$

where Ric_D is the Ricci tensor of D and ϕ is a local function.

Proposition

Let (D, g) be a Codazzi structure nad (D', g) its dual Codazzi structure. Then (1)

$$g(R_D(X, Y)Z, W) + g(Z, R_{D'}(X, Y)W) = 0;$$

(2) if (D, g) is a Codazzi struccture of constant curvature c, then (D', g) is also of constant curvature c.

Proposition

A Codazzi structure (D, g) is of constant curvature 0 iff (D, g) is a Hessian structure.

Proposition

Let (D, g) be a Codazzi structure of constant curvature. Then locally g is of the form

$$Dd\phi + \frac{\phi}{m-1} Ric_D$$

where Ric_D is the Ricci tensor of D and ϕ is a local function.

Proposition

Let (D, g) be a Codazzi structure nad (D', g) its dual Codazzi structure. Then (1)

$$g(R_D(X, Y)Z, W) + g(Z, R_{D'}(X, Y)W) = 0;$$

(2) if (D, g) is a Codazzi struccture of constant curvature c, then (D', g) is also of constant curvature c.

Proposition

A Codazzi structure (D, g) is of constant curvature 0 iff (D, g) is a Hessian structure.

Proposition

Let (D, g) be a Codazzi structure of constant curvature. Then locally g is of the form

$$Dd\phi + \frac{\phi}{m-1} Ric_D$$

where Ric_D is the Ricci tensor of D and ϕ is a local function.

Example

Let $S(m)$ be the set of real symmetric matrices of degree m, and let $S(m)^+$ be the subset of $S(m)$ consisting of of positive-definite symmetric matrices. Put

$$p(x;\mu,\sigma) = (2\pi)^{-m/2}(det\sigma)^{-1/2}exp(\{-\frac{{}^t(x-\mu)\sigma^{-1}(x-\mu)}{2}\}),$$

where $\mu \in \mathbb{R}^m$ and $\sigma \in S(m)^+$. Then $\{p(x;\mu,\sigma)\colon (\mu,\sigma) \in \mathbb{R}^m \times S(m)^+\}$ is a family of probability distributions on $\mathbb{R}^m$ parametrized by (μ,σ) and called a family of **m-dimensional normal distributions**.

Let Ω be a domain in a finite dimensional real vector space V, and let ρ be an injective linear mapping from Ω into $S(m)^+$.

Proposition

Let $\{p(x;\mu,\omega)\colon (\mu,\omega) \in \mathbb{R}^m \times \Omega\}$ be a family of probability distributions induced by ρ. Then the family is an exponential family parametrized by $\theta = \rho(\omega)\mu \in \mathbb{R}^m$ and $\omega \in \Omega$. The Fisher information metric is a Hessian metric on $\mathbb{R}^m \times \Omega$ with potential function

$$\phi(\theta,\omega) = (1/2)\{{}^t\theta\rho(\omega)^{-1}\theta - log\ det\rho(\omega)\}.$$

References

1) Amari S-I and Nagaoka H.: Methods of Information Geometry, Translations of Mathematical monographs, AMS-OXFORD, vol 191
2) Shima, H.: The Geometry of Hessian Structures. World Scientific (2007)
3)**[BW]** Nguiffo Boyom M and W R.: Transversely Hessian foliations and information geometry, Int. J. Math. 27,11 (2016)

Conferences
1) Geometric Science of Information - GSI (Mines ParisTech, Paris 2014 and 2017)
2) Topological and Geometrical Structure of Information (CIRM 2017)

Foliations

Let $\mathcal{F}$ be a foliation on an m-manifold M. Then $\mathcal{F}$ is defined by a cocycle $\mathcal{U} = \{U_i, f_i, k_{ij}\}_{i \in I}$ modeled on a q-manifold N_0 $(0 < q < m)$ such that
(1) $\{U_i\}_{i \in I}$ is an open covering of M,
(2) $f_i : U_i \to N_0$ are submersions with connected fibres,
(3) $k_{ij} : N_0 \to N_0$ are local diffeomorphisms of N_0 with $f_i = k_{ij} f_j$ on $U_i \cap U_j$.

The connected components of the trace of any leaf of $\mathcal{F}$ on U_i are fibres of f_i, and the trace itself consists of at most a denombrable number of these fibres.

The open subsets $N_i = f_i(U_i) \subset N_0$ form a q-dimensional manifold $N_{\mathcal{U}} = \bigsqcup N_i$, which can be considered to be a complete transverse manifold of the foliation $\mathcal{F}$. The pseudogroup $\mathcal{H}_{\mathcal{U}}$ of local diffeomorphisms of N generated by k_{ij} is called the holonomy pseudogroup of the foliated manifold $(M, \mathcal{F})$ defined by the cocycle $\mathcal{U}$.

A foliation on a smooth manifold M understood as an involutive subbundle of TM, or equivalently, according to the Frobenius theorem as a partition of the manifold by submanifolds of the same dimension with some regularity condition, can be defined by many different cocycles.

There is a notion of equivalent cocycles, similar to the notion of equivalent atlases of a smooth manifold, and a foliation can be understood as an equivalence class of such cocycles. The equivalence class $\mathcal{H}$ of $\mathcal{H}_{\mathcal{U}}$, is called the **holonomy group** of $\mathcal{F}$, or of the foliated manifold $(M, \mathcal{F})$.

The vector bundle $N(M, \mathcal{F}) = TM/T\mathcal{F}$ is called the **normal bundle of the foliation** $\mathcal{F}$. Then the tangent bundle TM is isomorphic to the direct sum $T\mathcal{F} \oplus N(M, \mathcal{F})$. These isomorphisms are determined by the choice of a supplementary subbundle Q in TM to the tangent bundle to the foliation $T\mathcal{F}$.

The cocycle $\mathcal{U} = \{U_i, f_i, k_{ij}\}_{i \in I}$ modeled on a q-manifold N_0 induces on the normal bundle a cocycle $\mathcal{V} = \{V_i, \bar{f}_i, \bar{k}_{ij}\}_{i \in I}$ modeled on the $2q$-manifold TN_0, where $V_i = TU_i$, $\bar{f}_i$ is the mapping induced by df_i, and $\bar{k}_{ij} = dk_{ij}$.

The foliation $\mathcal{F}_N$ of the normal bundle is of codimension $2q$, its leaves project on leaves of $\mathcal{F}$. They are, in fact, coverings of these leaves.

In a similar way one can foliate any bundle obtained via a point-wise process from the normal bundle, e.g., the frame bundle of the normal bundle, the dual normal bundle, any tensor product of these bundles.

Geometric structures on foliated manifolds

In the case of a foliated manifold we can consider three types of geometrical structures related to the foliation:

transverse - defined on the transverse manifold, the associated holonomy pseudogroup consists of automorphisms of this geometrical structure;

foliated - only defined on the normal bundle, and when expressed in a local adapted chart, depending only on the transverse coordinates; a foliated structure projects to a transverse structure along submersions of the cocycle defining the foliation.;

associated - defined globally, on the tangent bundle but adapted to the spliting, and defining a foliated structure on the normal bundle.

Foliated and transverse structures are in one-to-one correspondence, an associated structure defines a foliated structure, but different associated structures can define the same foliated structure.