Advanced Studies in Pure Mathematics 57, 2010
Probabilistic Approach to Geometry
pp. 303–321

Statistical manifolds and affine differential geometry

Hiroshi Matsuzoe

Abstract.

In this paper, we give a summary of geometry of statistical manifolds, and discuss relations between information geometry and affine differential geometry. Dually flat spaces and canonical divergence functions are important objects in information geometry. We show that such objects can be generalized in the framework of affine differential geometry.

In addition, we give a brief summary of geometry of statistical manifolds admitting torsion, which is regarded as a quantum version of statistical manifolds. We discuss relations between statistical manifolds admitting torsion and geometry of affine distributions.

§1. Introduction

A statistical manifold (M, ∇, h) is a (semi-)Riemannian manifold (M, h) with a torsion-free affine connection ∇ which satisfies some compatible condition. (See Definition 2.4 in Section 2.) For a statistical manifold, a pair of mutually dual affine connections can be defined naturally. These geometric structures have been studied in differential geometry. However, statistical manifolds and dual affine connections were refound in statistics in 1970's to construct geometric theory for statistical inferences (cf. [2] and [4]). Nowadays, this geometric method is called information geometry, and is applied various fields of mathematical sciences. For example, EM algorithms [3], boosting algorithms [17], algorithms of belief propagations [7], etc., are elucidated clearly by information geometry.

Received January 20, 2009.
Revised February 28, 2009.
2000 *Mathematics Subject Classification.* 53A15, 53A30, 53B05, 62B10, 81Q70.
Key words and phrases. Statistical manifold, affine differential geometry, information geometry, semi-Weyl manifold, affine distribution.
This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Young Scientists (B), No. 19740033.

From the viewpoint of differential geometry, statistical manifold structures arise in affine hypersurface theory. In fact, the notion of statistical manifold was originally introduced by S. L. Lauritzen [12], and it was reformulated by Kurose from the viewpoint of affine differential geometry [9]. We then find that statistical manifold structures and important objects of information geometry can be generalized in the framework of affine differential geometry.

In this paper, we summarize geometry of statistical manifolds and discuss relations between information geometry and affine differential geometry. In particular, the Legendre transformations and canonical divergences are important object in information geometry. We show that these objects are generalized by conormal (or dual map) transformations and geometric divergences in affine differential geometry.

In the later part of this paper, we consider a quantum version of statistical manifolds. It is known that an affine connection on a set of quantum states has the non-zero torsion tensor. Hence Kurose introduced the notion of statistical manifolds admitting torsion to discuss geometric structures on quantum state spaces [11]. As statistical manifolds have close relations to geometry of affine immersions, statistical manifolds admitting torsion have relations to geometry of affine distributions. We then give brief summaries of these geometry.

§2. Dual connections and Statistical manifolds

We assume that all the objects are smooth throughout this paper. In addition, we may assume that a manifold is simply connected since we discuss local geometric properties on a manifold.

At first, we introduce the notion of dual affine connections.

Let (M, h) be a semi-Riemannian manifold, and ∇ an affine connection on M. We define the *dual* (or *conjugate*) *connection* ∇^* of ∇ with respect to h by

$$Xh(Y,Z) = h(\nabla_X Y, Z) + h(Y, \nabla^*_X Z), \tag{1}$$

where X, Y and Z are arbitrary vector fields on M. We say that ∇ and ∇^* are *mutually dual* with respect to h.

Denote by R the curvature tensor of ∇, and by T the torsion of ∇, that is,

$$\begin{aligned} R(X,Y)Z &= \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z, \\ T(X,Y) &= \nabla_X Y - \nabla_Y X - [X,Y]. \end{aligned}$$

From straightforward calculations, we have the following proposition.

Proposition 2.1. *Suppose that affine connections* ∇ *and* ∇^* *are mutually dual with respect to* h. *Then the following formulae hold:*

(1) $(\nabla^*)^* = \nabla$.
(2) *Set* $\nabla^{(0)} := (\nabla + \nabla^*)/2$. *Then* $\nabla^{(0)} h = 0$.
(3) $h(R(X,Y)Z,V) = -h(Z, R^*(X,Y)V)$.

We remark that that $\nabla^{(0)}$ may not be the Levi–Civita connection of h. The connection $\nabla^{(0)}$ has a torsion in general.

Let us consider when dual affine connections are torsion-free.

Proposition 2.2 ([16]). *Consider the following conditions:*

(1) ∇ *is torsion-free.*
(2) ∇^* *is torsion-free.*
(3) $C := \nabla h$ *is totally symmetric.*
(4) $\nabla^{(0)} = (\nabla + \nabla^*)/2$ *is the Levi–Civita connection with respect to* h.

Then any two of the above conditions imply the rest of them.

This proposition implies that torsion-free mutually dual affine connections induce a symmetric $(0,3)$-tensor field C. We call the tensor field C a *cubic form* on M, and we say that the connection ∇ is *compatible* with h. Conversely, for a given totally symmetric $(0,3)$-tensor field, we can define mutually dual affine connections.

Proposition 2.3 ([12] and [16]). *Suppose that* (M,h) *is a semi-Riemannian manifold,* C *is a totally symmetric* $(0,3)$*-tensor field on* M, *and* $\nabla^{(0)}$ *is the Levi–Civita connection with respect to* h. *Set* ∇ *and* ∇^* *by*

$$
\begin{aligned}
h(\nabla_X Y, Z) &:= h(\nabla^{(0)}_X Y, Z) - \frac{1}{2}C(X,Y,Z), \\
h(\nabla^*_X Y, Z) &:= h(\nabla^{(0)}_X Y, Z) + \frac{1}{2}C(X,Y,Z).
\end{aligned}
$$

Then ∇ *and* ∇^* *are mutually dual torsion-free affine connections compatible with* h.

Let us give the definition of statistical manifold.

Definition 2.4. *Let* (M,h) *be a semi-Riemannian manifold, and* ∇ *a torsion-free affine connection on* M. *We call the triplet* (M,∇,h) *a* statistical manifold *if* ∇h *is totally symmetric, that is,* ∇ *is compatible with* h.

For a given statistical manifold (M,∇,h), denote by ∇^* the dual connection of ∇ with respect to h. In this case, the triplet (M,∇^*,h)

is also a statistical manifold, and we call (M, ∇^*, h) the *dual statistical manifold* of (M, ∇, h).

The notion of statistical manifold was originally introduced by S. L. Lauritzen [12]. He called a triplet (M, h, C) a statistical manifold. On the other hand, our definition follows Kurose's formulation [9]. From Propositions 2.2 and 2.3, these definitions are essentially equivalent.

We remark that a parametric statistical model has a statistical manifold structure (cf.[4] and [12]). Let (Ω, β) be a measurable space. Suppose that $\mathcal{M}$ is a parametric statistical model on Ω, that is, $\mathcal{M}$ is a set of probability distributions on (Ω, β) parametrized by $\zeta = (\zeta^1, \dots, \zeta^n) \in U \subset \boldsymbol{R}^n$:

$$\mathcal{M} = \left\{ p(x, \zeta) \,\middle|\, p(x, \zeta) > 0, \int_\Omega p(x, \zeta) dx = 1 \right\}.$$

Under suitable conditions (see Chapter 2.1 in [4]), $\mathcal{M}$ is regarded as a manifold with a local coordinate system $(\zeta^1, \dots, \zeta^n)$.

For simplicity, set $l(x; \zeta) := \log p(x; \zeta)$. We define a symmetric matrix (g_{ij}) $(i, j = 1, 2, \dots, n)$ by

$$g_{ij}(\zeta) := \int_\Omega \left(\frac{\partial l}{\partial \zeta^i}(x; \zeta) \right) \left(\frac{\partial l}{\partial \zeta^j}(x; \zeta) \right) p(x; \zeta) dx.$$

If $\{g_{ij}(\zeta)\}$ is positive definite, it determines a Riemannian metric on $\mathcal{M}$. We call g the *Fisher metric* on $\mathcal{M}$.

For an arbitrary constant $\alpha \in \boldsymbol{R}$, we can define a torsion-free affine connection on $\mathcal{M}$ by

$$\begin{aligned} \Gamma^{(\alpha)}_{ij,k}(\zeta) \;\; &:= \;\; \int_\Omega \left\{ \frac{\partial^2 l}{\partial \zeta^i \partial \zeta^j}(x; \zeta) + \frac{1-\alpha}{2} \left(\frac{\partial l}{\partial \zeta^i}(x; \zeta) \right) \left(\frac{\partial l}{\partial \zeta^j}(x; \zeta) \right) \right\} \\ &\qquad\qquad \times \left(\frac{\partial l}{\partial \zeta^k}(x; \zeta) \right) p(x; \zeta) dx, \\ g\left(\nabla^{(\alpha)}_{\frac{\partial}{\partial \zeta^i}} \frac{\partial}{\partial \zeta^j}, \frac{\partial}{\partial \zeta^k} \right) &= \Gamma^{(\alpha)}_{ij,k}(\zeta). \end{aligned}$$

The affine connection $\nabla^{(\alpha)}$ is called the *α-connection* on $\mathcal{M}$. For an arbitrary constant α, $\nabla^{(\alpha)}$ and $\nabla^{(-\alpha)}$ are mutually dual with respect to g. In addition, $\nabla^{(\alpha)} g$ and $\nabla^{(-\alpha)} g$ are totally symmetric. Hence the triplet $(\mathcal{M}, \nabla^{(\alpha)}, g)$ is a statistical manifold, and $(\mathcal{M}, \nabla^{(-\alpha)}, g)$ is its dual statistical manifold.

§3. Flatness and conformal flatness on statistical manifolds

In this section, we summarize flatness and several conformal flatness of statistical manifolds.

3.1. Dually flat spaces

Let (M, ∇, h) be a statistical manifold. For simplicity, we assume that M is simply connected and M has a global coordinate system.

The connection ∇ (or a statistical manifold (M, ∇, h)) is said to be *flat* if ∇ is torsion-free and curvature-free everywhere on M. If a statistical manifold (M, ∇, h) is flat, from Proposition 2.1, the dual statistical manifold (M, ∇^*, h) is also flat. In this case, we call the tetrad (M, h, ∇, ∇^*) a *dually flat space.*

For a dually flat space (M, h, ∇, ∇^*), since ∇ is flat and M has a global coordinate chart, there exists a coordinate system $\{\theta^i\}$ such that all the connection coefficients $\{\Gamma^{\nabla}_{ij}{}^{k}\}$ $(i, j, k = 1, \ldots, n)$ vanish everywhere on M. We call such a coordinate system $\{\theta^i\}$ an *affine coordinate system* of ∇. In addition, there exists a ∇^*-affine coordinate system $\{\eta_i\}$ such that

$$h\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\eta_j}\right) = \delta_i^j \quad (i, j = 1, \ldots, n).$$

The coordinate system $\{\eta_i\}$ is called the *dual coordinate system* of $\{\theta^i\}$ with respect to h.

Proposition 3.1. *Let (M, h, ∇, ∇^*) be a dually flat space. Suppose that M has a globally defined ∇-affine coordinate system $\{\theta^i\}$. Denote by $\{\eta_i\}$ the dual coordinate system of $\{\theta^i\}$. Then there exist functions ψ and ϕ on M such that*

$$(2) \quad \frac{\partial\psi}{\partial\theta^i} = \eta_i, \quad \frac{\partial\phi}{\partial\eta_i} = \theta^i, \quad \psi(p) + \phi(p) - \sum_{i=1}^{n} \theta^i(p)\eta_i(p) = 0 \ (p \in M).$$

In addition, the second derivatives induce the original semi-Riemannian metric:

$$(3) \qquad h_{ij} = \frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}, \quad h^{ij} = \frac{\partial^2\phi}{\partial\eta_i\partial\eta_j},$$

where h^{ij} is the (i, j)-component of the inverse matrix of (h_{ij}).

For proof, see Chapter 3.3 in [4].

Formulae (2) are the *Legendre transformation* on (M, h, ∇, ∇^*). Formulae (3) imply that h is a *Hessian metric* on M, since h is given as a

Hessian of the function ψ (or ϕ). Hence (∇, h) and (∇^*, h) are called *Hessian structures.* (See [23].) For these reasons, ψ and ϕ are important objects on (M, h, ∇, ∇^*). We call ψ the *θ-potential function*, and ϕ the *η-potential function.*

From the functions ψ and ϕ, we define an asymmetric distance-like function on $M \times M$ by

$$\rho(p, q) := \psi(p) + \phi(q) - \sum_{i=1}^{n} \theta^i(p)\eta_i(q) \quad (p, q \in M). \tag{4}$$

We call the function ρ the *(∇-canonical) divergence* of (M, h, ∇, ∇^*).

Further arguments of dually flat spaces and divergence functions, see Chapter 3 in [4]. Canonical divergences are an important object in information geometry. See [4] and [17], for example. We remark that these structures are related to Hessian structures [23] and α-Hessian structures [26].

3.2. Conformal equivalence relations

In this section, we summarize several conformal equivalence relations on manifolds. For more details, see [8], [10] and [14].

3.2.1. *Conformal equivalence:* Let g and $\tilde{g}$ be Riemannian metrics on M. We say that g and $\tilde{g}$ are *conformally equivalent* if there exists a function ϕ on M such that

$$\tilde{g}(X, Y) = e^{2\phi} g(X, Y).$$

We say that a Riemannian metric g is *flat* if the Levi–Civita connection of g is flat. We say that g is *conformally flat* if g is conformally equivalent to a flat Riemannian metric in a neighbourhood of an arbitrary point of M.

If Riemannian metrics g and $\tilde{g}$ are conformally equivalent, then their Levi–Civita connections ∇^g and $\nabla^{\tilde{g}}$ satisfy

$$\nabla^{\tilde{g}}_X Y = \nabla^g_X Y - g(X, Y)\mathrm{grad}_g \phi + d\phi(Y)X + d\phi(X)Y,$$

where $\mathrm{grad}_g \phi$ is the gradient vector field of ϕ with respect to g.

3.2.2. *Projective equivalence:* Let M be an n-dimensional manifold. We say that two affine connections ∇ and $\tilde{\nabla}$ on M are *projectively equivalent* if there exists a 1-form τ on M such that

$$\tilde{\nabla}_X Y = \nabla_X Y + \tau(Y)X + \tau(X)Y. \tag{5}$$

We say that an affine connection ∇ is *projectively flat* if ∇ is projectively equivalent to some flat affine connection in a neighbourhood of an arbitrary point of M.

3.2.3. *Dual-projective equivalence:* Let (M, h) be a semi-Riemannian manifold. We say that two affine connections ∇ and $\tilde{\nabla}$ on M are *dual-projectively equivalent* [8] if there exists a 1-form τ on M such that

$$\tilde{\nabla}_X Y = \nabla_X Y - h(X, Y)\tau^{\#}, \tag{6}$$

where $\tau^{\#}$ is the metrical dual vector field defined by $h(\tau^{\#}, X) := \tau(X)$.

We say that an affine connection ∇ is *dual-projectively flat* if ∇ is dual-projectively equivalent to some flat affine connection in a neighbourhood of an arbitrary point of M.

3.2.4. *Conformal-projective equivalence:* As in the duality of affine connections, a duality of projective (or conformal) equivalence relations arise naturally on statistical manifolds. (See [10], [14] and [22], for example.) Let us consider such relations.

Suppose that (M, ∇, h) and $(M, \tilde{\nabla}, \tilde{h})$ are statistical manifolds. We say that (M, ∇, h) and $(M, \tilde{\nabla}, \tilde{h})$ are *conformally-projectively equivalent* (or *generalized conformally equivalent*) [13] if there exist two functions ϕ and ψ on M such that

$$\begin{aligned} \tilde{h}(X, Y) &= e^{\psi+\phi} h(X, Y), \\ \tilde{\nabla}_X Y &= \nabla_X Y - h(X, Y)\mathrm{grad}_h \psi + d\phi(Y)X + d\phi(X)Y. \end{aligned}$$

In a special case, for an arbitrary constant $\alpha \in \boldsymbol{R}$, we say that (M, ∇, h) and $(M, \tilde{\nabla}, \tilde{h})$ are *α-conformally equivalent* [9] if there exists a function ψ on M such that

$$\begin{aligned} \tilde{h}(X, Y) &= e^{\psi} h(X, Y), \\ \tilde{\nabla}_X Y &= \nabla_X Y - \frac{1+\alpha}{2} h(X, Y)\mathrm{grad}_h \psi \\ &\quad + \frac{1-\alpha}{2}\left(d\psi(Y)X + d\psi(X)Y\right). \end{aligned}$$

We say that a statistical manifold (M, ∇, h) is *conformally- projectively flat* (or *α-conformally flat*) if (M, ∇, h) is conformally-projectively equivalent (or α-conformally equivalent) to a flat statistical manifold in a neighbourhood of an arbitrary point of M, respectively.

Suppose that (M, ∇, h) and $(M, \tilde{\nabla}, \tilde{h})$ are conformally-projectively equivalent. From Equations (5) or (6), if ψ is constant, then ∇ and $\tilde{\nabla}$ are projectively equivalent. In this case, (M, ∇, h) and $(M, \tilde{\nabla}, \tilde{h})$ are (-1)-conformally equivalent. If ϕ is constant, then ∇ and $\tilde{\nabla}$ are dual-projectively equivalent. In this case, (M, ∇, h) and $(M, \tilde{\nabla}, \tilde{h})$ are 1-conformally equivalent.

It is well-known that projective flatness or conformal flatness are characterized by Weyl's projective curvature tensor or conformal curvature tensor, respectively. Similarly, dual-projective flatness is characterized by Ivanov's dual-projective curvature tensor [8], and conformal-projective flatness is characterized by Kurose's conformal-projective curvature tensor [10]. These curvature tensors are summarized in [14].

§4. Generalized dual connections and semi-Weyl manifolds

Let (M, h) be a semi-Riemannian manifold, ∇ an affine connection on M, and τ a 1-form on M. We define the *generalized dual* (or *generalized conjugate*) *connection* $\overline{\nabla}^*$ of ∇ by τ with respect to h by

$$Xh(Y,Z) = h(\nabla_X Y, Z) + h(Y, \overline{\nabla}^*_X Z) - \tau(X)h(Y,Z). \tag{7}$$

The notion of generalized dual connection was originally introduced by A. P. Norden to study Weyl geometry [20]. (See also [18]. Nomizu called $\overline{\nabla}^*$ the generalized conjugate connection.) Generalized dual connections are invariant under gauge transformations. Let ϕ be a function, and τ a 1-form on M. Set

$$(\bar{h}, \bar{\tau}) := (e^{\phi}h, \tau - d\phi).$$

The pair $(\bar{h}, \bar{\tau})$ is called a *gauge transformation* of (h, τ). We then obtain

$$\begin{aligned} &Xh(Y,Z) = h(\nabla_X Y, Z) + h(Y, \overline{\nabla}^*_X Z) - \tau(X)h(Y,Z) \\ \iff\ &X\bar{h}(Y,Z) = \bar{h}(\nabla_X Y, Z) + \bar{h}(Y, \overline{\nabla}^*_X Z) - \bar{\tau}(X)\bar{h}(Y,Z). \end{aligned}$$

For generalized dual connections, as in the case of dual connections, we obtain the following proposition.

Proposition 4.1 ([16]). *Consider the following conditions:*

(1) ∇ *is torsion-free.*
(2) $\overline{\nabla}^*$ *is torsion-free.*
(3) $\bar{C}(X,Y,Z) := (\nabla_X h)(Y,Z) + \tau(X)h(Y,Z)$ *is totally symmetric.*
(4) $\overline{\nabla}^{(0)} = (\nabla + \overline{\nabla}^*)/2$ *is a Weyl connection, that is,* $\overline{\nabla}^{(0)}$ *is torsion-free and*

$$\bar{C}^{(0)}(X,Y,Z) := (\overline{\nabla}^{(0)}_X h)(Y,Z) + \tau(X)h(Y,Z) = 0.$$

Then any two of the above conditions imply the rest of them.

Let (M, h) be a semi-Riemannian manifold, ∇ a torsion-free affine connection on M, and τ a 1-form on M. We say that the tetrad

(M, ∇, h, τ) is a *semi-Weyl manifold* if $\nabla h + \tau \otimes h$ is totally symmetric, that is,

$$(\nabla_X h)(Y,Z) + \tau(X)h(Y,Z) = (\nabla_Y h)(X,Z) + \tau(Y)h(X,Z).$$

For a given semi-Weyl manifold (M, ∇, h, τ), it is easy to check the following properties.

(1) If $(\nabla_X h)(Y,Z) + \tau(X) = 0$, then (M, ∇, h) is a Weyl manifold.

(2) If $\tau = 0$, then (M, ∇, h) is a statistical manifold.

As we will see next section, semi-Weyl manifolds arise naturally in affine differential geometry. See also [15].

§5. Affine immersions

In this section, we summarize the definitions and basic results of affine differential geometry, then elucidate relations between affine differential geometry and information geometry. For more details about affine differential geometry, see [19] and [24].

Let M be an n-dimensional manifold ($n \geq 2$), f an immersion from M to $\boldsymbol{R}^{n+1}$, and ξ a vector field along f. We say that the pair $\{f, \xi\}$ is an *affine immersion* if, for an arbitrary point p in M, the tangent space is decomposed as

$$T_{f(p)}\boldsymbol{R}^{n+1} = f_*(T_pM) \oplus \boldsymbol{R}\{\xi_p\}.$$

We call ξ a *transversal vector field.*

Let D be the standard flat affine connection on $\boldsymbol{R}^{n+1}$. From the decomposition of the tangent space, the covariant derivatives are decomposed as follows:

$$D_X f_* Y = f_*(\nabla_X Y) + h(X,Y)\xi, \tag{8}$$

$$D_X \xi = -f_*(SX) + \tau(X)\xi. \tag{9}$$

We call ∇ a *induced connection*, h an *affine fundamental form*, S an *affine shape operator*, and τ a *transversal connection form.*

If the affine fundamental form h is nondegenerate everywhere on M, the immersion f is called *nondegenerate.* It is independent of the choice of transversal vector fields. If $\tau = 0$, the affine immersion $\{f, \xi\}$ is called *equiaffine.*

Since the affine connection D is flat, we obtain the fundamental structural equations for affine immersions.

The Gauss equation:

$$R(X,Y)Z = h(Y,Z)SX - h(X,Z)SY. \tag{10}$$

The Codazzi equations:

$$(11)\quad \begin{aligned}(\nabla_X h)(Y,Z)+\tau(X)h(Y,Z) &= (\nabla_Y h)(X,Z)+\tau(Y)h(X,Z),\\ (\nabla_X S)(Y)-\tau(X)SY &= (\nabla_Y S)(X)-\tau(Y)SX.\end{aligned}$$

The Ricci equation:

$$h(X,SY)-h(Y,SX)=(\nabla_X\tau)(Y)-(\nabla_Y\tau)(X).$$

From Codazzi equation for h, we directly obtain the following proposition.

Proposition 5.1. *Let $\{f,\xi\}$ be an affine immersion. If f is nondegenerate, then induced objects (M,∇,h,τ) is a semi-Weyl manifold. If $\{f,\xi\}$ is nondegenerate and equiaffine, then (M,∇,h) is a statistical manifold.*

For a non-zero function ϕ on M, and a vector field Z, we change a transversal vector field ξ to

$$\bar{\xi}=\frac{1}{\phi}\{\xi+f_*(Z)\}.$$

Then the induced objects change as follows:

$$\begin{aligned}\bar{h}(X,Y) &= \phi h(X,Y),\\ \overline{\nabla}_X Y &= \nabla_X Y-h(X,Y)Z,\\ \bar{\tau}(X) &= \tau(X)+h(X,Z)-d\log|\phi|(X),\\ \bar{S}X &= \frac{1}{\phi}\left\{SX-\nabla_X Z+\bar{\tau}(X)Z+d\log|\phi|(X)Z\right\}.\end{aligned}$$

Suppose that $\{f,\xi\}$ is a nondegenerate equiaffine affine immersion. If $\bar{\xi}$ is a parallel constant vector field, $\bar{\tau}=0,\bar{S}=0$ and hence $\overline{\nabla}$ is a flat affine connection. In this case, Z should be the gradient vector field of $\log|\phi|$ with respect to h. Hence we conclude

Proposition 5.2. *Let $\{f,\xi\}$ be a nondegenerate equiaffine affine immersion. Suppose that (M,∇,h) is the induced statistical manifold. Then (M,∇,h) is 1-conformally flat.*

If M is simply connected, we can give the fundamental theorem for affine immersions.

Theorem 5.3 ([5], [8] and [9]). *Let (M,∇,h,τ) be a simply connected semi-Weyl manifold. If ∇ is dual-projectively flat, then there exists a nondegenerate affine immersion $\{f,\xi\}$ such that it induces the original semi-Weyl manifold (M,∇,h,τ).*

If (M, ∇, h) is a simply connected 1-conformally flat statistical manifold, then there exists a nondegenerate equiaffine immersion $\{f, \xi\}$ such that it induces (M, ∇, h).

Let $\boldsymbol{R}_{n+1}$ be the dual vector space of $\boldsymbol{R}^{n+1}$, and $\langle\ ,\ \rangle$ the pairing of $\boldsymbol{R}_{n+1}$ and $\boldsymbol{R}^{n+1}$. Suppose that $\{f, \xi\}$ is an affine immersion from M to $\boldsymbol{R}^{n+1}$. We can define the *conormal map* $v : M \to \boldsymbol{R}_{n+1}$ of $\{f, \xi\}$ by

$$\langle v(p), \xi_p \rangle = 1, \quad \langle v(p), f_* X_p \rangle = 0. \tag{12}$$

Differentiating above equations, we have

$$\langle v_* X_p, \xi_p \rangle = -\tau(X), \quad \langle v_* X_p, f_* Y_p \rangle = -h(X, Y).$$

If h is nondegenerate, v is an immersion from M to $\boldsymbol{R}_{n+1}$, and v is transversal to v itself. Hence $\{v, -v\}$ is an affine immersion from M to $\boldsymbol{R}_{n+1}$. The induced connection of $\{v, -v\}$ is given by

$$D_X v_* Y \;=\; v_*(\hat{\nabla}^*_X Y) - h^*(X, Y)v.$$

Then the connection $\hat{\nabla}^*$ satisfies

$$Xh(Y, Z) = h(\nabla_X Y, Z) + h(Y, \hat{\nabla}^*_X Z) + \tau(Z)h(X, Y).$$

We remark that the connection $\hat{\nabla}^*$ is called the *semi-conjugate connection* of ∇ with respect to h by τ [8].

If $\tau = 0$, then ∇ and $\hat{\nabla}$ are mutually dual affine connections. Hence we conclude

Proposition 5.4. *Let $\{f, \xi\}$ be a nondegenerate equiaffine immersion from M to $\boldsymbol{R}^{n+1}$, and v the conormal map of $\{f, \xi\}$. Suppose that ∇ and $\hat{\nabla}^*$ are induced connections $\{f, \xi\}$ and $\{v, -v\}$, respectively. Then induced connections ∇ and $\hat{\nabla}^*$ are mutually dual with respect to the affine fundamental form h.*

Let us construct divergence functions on M. Let $\{f, \xi\}$ be a nondegenerate equiaffine immersion and v the conormal map of $\{f, \xi\}$. We define a function ρ on $M \times M$ by

$$\rho^G(p, q) = \langle v(q), f(p) - f(q) \rangle.$$

We call ρ^G a *geometric divergence* on M [9].

The geometric divergence is an example of contrast functions [6], Then we can induce a statistical manifold (M, ∇, h) from ρ^G.

Proposition 5.5. *Let (M, h, ∇, ∇^*) be a simply connected dually flat space. Then the canonical divergence and the geometric divergence on (M, ∇, h) coincide.*

Proof. Suppose that $\{\theta^i\}$ is a ∇-affine coordinate system, and $\{\eta_i\}$ is its dual coordinate system. Denote by ψ the θ-potential function, and by ϕ the η-potential function. Then the canonical divergence $\rho(p,q)$ is given by

$$\rho(p,q) = \psi(p) + \phi(q) - \sum_{i=1}^{n} \theta^i(p)\eta_i(q), \quad (p, q \in M).$$

On the other hand, the statistical manifold (M, ∇, h) is realized in $\boldsymbol{R}^{n+1}$ as a graph immersion $\{f, \xi\}$:

$$f : \begin{pmatrix} \theta^1 \\ \vdots \\ \theta^n \end{pmatrix} \mapsto \begin{pmatrix} \theta^1 \\ \vdots \\ \theta^n \\ \psi \end{pmatrix}, \quad \xi = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

In fact, Equation (8) is given by

$$D_{\frac{\partial}{\partial\theta^i}} f_* \left(\frac{\partial}{\partial\theta^j} \right) = \frac{\partial^2\psi}{\partial\theta^i \partial\theta^j} \xi.$$

Denote by $\{\Gamma_{ij}^k\}$ the connection coefficients of ∇. Then $\Gamma_{ij}^k \equiv 0$ $(i, j, k = 1, \ldots, n)$, and hence $\{\theta^i\}$ is a ∇-affine coordinate system. From Equation (3), the affine fundamental form coincides with h. In this case, from the definitions of conormal map (12) and the Legendre transformation (2), the conormal map is given by

$$v(q) = (-\eta_1(q), \ldots, -\eta_n(q), 1).$$

From the Legendre transformation (2), the geometric divergence ρ^G is given by

$$\begin{aligned} \rho^G(p,q) &= \langle v(q), f(p) - f(q) \rangle \\ &= \psi(p) + \sum_{i=1}^{n} v_i(q)\theta^i(p) - \psi(q) - \sum_{i=1}^{n} v_i(q)\theta^i(q) \\ &= \psi(p) + \phi(q) + \sum_{i=1}^{n} v_i(q)\theta^i(p) \\ &= \psi(p) + \phi(q) - \sum_{i=1}^{n} \eta_i(q)\theta^i(p). \end{aligned}$$

Therefore, the geometric divergence coincides with the canonical divergence. Q.E.D.

In this section, we mainly discussed dually flatness and 1-conformal equivalence of statistical manifolds. For α-conformal equivalence, see [25] for example. We remark that Ohara recently showed a relation between centroaffine immersions and the Tsallis statistics [21], which is related to geometry of statistical manifolds of constant curvature. We also remark that quite another formulation has given in [1], which is also useful in information geometry.

§6. Centroaffine immersions of codimension two

Let M be an n-dimensional manifold ($n \geq 3$), f an immersion from M to $\boldsymbol{R}^{n+2}$, and ξ a vector field along f. We identify f with the position vector of $\boldsymbol{R}^{n+2}$.

We say that the pair $\{f, \xi\}$ is a *centroaffine immersion of codimension two* from M to $\boldsymbol{R}^{n+2}$ if, for an arbitrary point p in M, the tangent space $T_{f(p)}\boldsymbol{R}^{n+2}$ is decomposed as

$$T_{f(p)}\boldsymbol{R}^{n+2} = f_*(T_pM) \oplus \boldsymbol{R}\{\xi_p\} \oplus \boldsymbol{R}\{f(p)\}.$$

We call ξ a *transverse vector field.*

The induced objects can be defined as follows:

$$\begin{aligned} D_X f_* Y &= f_* \nabla_X Y + h(X,Y)\xi + k(X,Y)f, \\ D_X \xi &= -f_* SX + \tau(X)\xi + \mu(X)f. \end{aligned}$$

We call ∇ a *induced connection*, h an *affine fundamental form*, S an *affine shape operator*, and τ a *transversal connection form.*

If the affine fundamental form h is nondegenerate everywhere on M, the immersion f is called *nondegenerate.* If $\tau = 0$, the centroaffine immersion $\{f, \xi\}$ is called *equiaffine.*

As in affine immersions (of codimension one), the fundamental structural equations hold. Hence a nondegenerate centroaffine immersion of codimension two $\{f, \xi\}$ induces a semi-Weyl manifold. If $\{f, \xi\}$ is nondegenerate and equiaffine, then $\{f, \xi\}$ induces a statistical manifold.

Conversely, the following theorem has been obtained in [14].

Theorem 6.1 ([14]). *Suppose that (M, ∇, h) is a simply connected conformally-projectively flat statistical manifold, then there exists a centroaffine immersion of codimension two $\{f, \xi\}$ which realizes (M, ∇, h) in $\boldsymbol{R}^{n+2}$.*

Let $\{f,\xi\}$ be a non-degenerate equiaffine centroaffine immersion of codimension two. Denote by $\boldsymbol{R}_{n+2}$ the dual space of $\boldsymbol{R}^{n+2}$, and by $\langle\ ,\ \rangle$ the pairing of $\boldsymbol{R}_{n+2}$ and $\boldsymbol{R}^{n+2}$. Then we can define a centroaffine immersion of codimension two $\{v,w\} : M \to \boldsymbol{R}_{n+2}$ by

$$\langle v(p), \xi_p\rangle = 1, \qquad \langle w(p), \xi_p\rangle = 0,$$
$$\langle v(p), f(p)\rangle = 0, \qquad \langle w(p), f(p)\rangle = 1, \tag{13}$$
$$\langle v(p), f_*X_p\rangle = 0, \qquad \langle w(p), f_*X_p\rangle = 0. \tag{14}$$

We call $\{v,w\}$ the *dual map* of $\{f,\xi\}$, and the correspondence $\{f,\xi\} \mapsto \{v,w\}$ the *dual map transformation.*

We define a function ρ on $M \times M$ by

$$\rho^G(p,q) = \langle v(q), f(p)\rangle.$$

We call ρ^G a *geometric divergence* on M. The geometric divergence is a contrast function on M, hence ρ^G induces a statistical manifold (M,∇,h). Conversely, for a conformally-projectively flat statistical manifold (M,∇,h), the geometric divergence is determined uniquely. It is independent of the realization of (M,∇,h) into $\boldsymbol{R}^{n+2}$.

We remark that the dual map transformation is a generalization of the Legendre transformation. In fact, suppose that (M,h,∇,∇^*) is a dually flat space. Then (M,h,∇) is realized in $\boldsymbol{R}^{n+2}$ by

$$f : \begin{pmatrix} \theta^1 \\ \vdots \\ \theta^n \end{pmatrix} \mapsto \begin{pmatrix} \theta^1 \\ \vdots \\ \theta^n \\ \psi \\ 1 \end{pmatrix}, \qquad \xi = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \end{pmatrix}.$$

The dual map is given by

$$\begin{aligned} v(q) &= (-\eta_1(q), \ldots, -\eta_n(q), 1, \phi(q)), \\ w &= (0, \ldots, 0, 0, 1). \end{aligned}$$

The differentials of f and v are

$$\begin{aligned} \frac{\partial f}{\partial \theta^i} &= \left(0, \ldots, 1, \ldots, 0, \frac{\partial \psi}{\partial \theta^i}, 0\right)^T, \\ \frac{\partial v}{\partial \eta_i} &= \left(0, \ldots, 1, \ldots, 0, 0, \frac{\partial \phi}{\partial \eta_i}\right). \end{aligned}$$

Differentiating the left side of Equation (13), using Equation (14), we have

$$(15) \qquad \langle v_* X_p, f(p) \rangle = 0.$$

Equations (14) and (15) imply

$$\frac{\partial \psi}{\partial \theta^i} = \eta_i, \quad \frac{\partial \phi}{\partial \eta_i} = \theta^i.$$

Substituting $f(p)$ and $v(p)$ into (13), we have

$$-\sum_{i=1}^{n} \eta_i(p)\theta^i(p) + \psi(p) + \phi(p) = 0.$$

Therefore, the definition of the dual map coincides with the Legendre transformation if (M, h, ∇, ∇^*) is dually flat. The following table shows the correspondence.

$$\begin{aligned}
\langle v(p), f_* X_p \rangle = 0 \quad &\Longleftrightarrow \quad \frac{\partial \psi}{\partial \theta^i} = \eta_i, \\
\langle v_* X_p, f(p) \rangle = 0 \quad &\Longleftrightarrow \quad \frac{\partial \phi}{\partial \eta_i} = \theta^i, \\
\langle v(p), f(p) \rangle = 0 \quad &\Longleftrightarrow \quad \psi(p) + \phi(p) - \sum_{i=1}^{n} \eta_i(p)\theta^i(p) = 0.
\end{aligned}$$

Similar calculations as in affine hypersurface theory, the geometric divergence ρ^G on a dually flat space coincides with the canonical divergence ρ.

§7. Statistical manifolds admitting torsion

In this section, we consider a quantum version of statistical manifold.

Let (M, h) be a semi-Riemannian manifold, and ∇ an affine connection on M. Denote by T^∇ the torsion tensor of ∇. We say that the triplet (M, ∇, h) is a *statistical manifold admitting torsion* if the following formula holds [11]:

$$(\nabla_X h)(Y, Z) - (\nabla_Y h)(X, Z) = -h(T^\nabla(X, Y), Z).$$

It is known that a set of quantum states has a non-zero torsion tensor in general. (See Chapter 7.3 in [4].) Hence the notion of statistical manifolds admitting torsion can be regarded as a quantum version of the notion of statistical manifolds. We will give an example in Section 8.

From a straightforward calculation, we can check the dual connection ∇^* of ∇ is torsion-free. Conversely, we obtain the following proposition.

Proposition 7.1. *Let (M,h) be a semi-Riemannian manifold, ∇^* a torsion-free affine connection on M, and ∇ the dual connection of ∇^* with respect to h. Then (M,∇,h) is a statistical manifold admitting torsion.*

For a semi-Weyl manifold, we naturally obtain a statistical manifold admitting torsion.

Proposition 7.2. *Let (M,∇,h,τ) be a semi-Weyl manifold. Denote by $\overline{\nabla}^*$ the generalized dual connection of ∇ by τ with respect to h. Set $\nabla'_X Y = \nabla_X Y - \tau(X)Y$. Then $\overline{\nabla}^*$ is the dual connection of ∇' with respect to h, and (M,∇',h) is a statistical manifold admitting torsion.*

Proof. From the definition of dual connection (1) and generalized dual connection (7), we have

$$\begin{aligned} Xh(Y,Z) &= h(\nabla_X Y, Z) + h(Y, \overline{\nabla}^*_X Z) - \tau(X)h(Y,Z) \\ &= h(\nabla'_X Y, Z) + h(Y, \overline{\nabla}^*_X Z). \end{aligned}$$

The torsion tensor is given by $T^{\nabla'}(X,Y) = \tau(Y)X - \tau(X)Y$. Q.E.D.

We say that (M,∇',h) is the statistical manifold admitting torsion *associated* to a semi-Weyl manifold (M,∇,h,ω).

§8. Affine distributions

Let M be an n-dimensional manifold, ω a $\boldsymbol{R}^{n+1}$-valued 1-form, and ξ a $\boldsymbol{R}^{n+1}$-valued function on M. We say that the pair $\{\omega,\xi\}$ is an *affine distribution* if, for an arbitrary point p in M, the following conditions hold:

$$\begin{aligned} &\boldsymbol{R}^{n+1} = \text{Image } \omega_p \oplus \boldsymbol{R}\{\xi_x\}, \\ &\text{Image } (d\omega)_p \subset \text{Image } \omega_p. \end{aligned} \tag{16}$$

If $\{f,\xi\}$ is an affine immersion from M to $\boldsymbol{R}^{n+1}$, then $\{df,\xi\}$ is an affine distribution.

The induced objects can be defined by

$$\begin{aligned} X\omega(Y) &= \omega(\nabla_X Y) + h(X,Y)\xi, \\ X\xi &= -\omega(SX) + \tau(X)\xi. \end{aligned}$$

We call ∇ an *induced connection*, h an *affine fundamental form*. From Equation (16) in the definition of affine distribution, h is a symmetric $(0,2)$-tensor field.

If h is nondegenerate everywhere on M, ω is called *nondegenerate.* If $\tau = 0$, the affine distribution $\{\omega, \xi\}$ is called *equiaffine.* A necessary and sufficient condition for an affine distribution $\{\omega, \xi\}$ to be equiaffine is

$$\text{Image } (d\xi)_p \subset \text{Image } \omega_p.$$

As in affine immersions, the fundamental structural equations hold for affine distributions. Then we have the following proposition.

Proposition 8.1. *Let $\{\omega, \xi\}$ be a nondegenerate equiaffine affine distribution. Denote by ∇ the induced connection, and by h the affine fundamental form. Then (M, ∇, h) is a statistical manifold admitting torsion.*

Example 8.2 ([11]). *Let* $\mathrm{Herm}(d)$ *be the set of all Hermitian matrices of degree d. Denote by $\mathcal{S}$ the set of quantum states defined by*

$$\mathcal{S} = \{P \in \mathrm{Herm}(d) \mid P > 0,\ \mathrm{Trace}P = 1\}.$$

For an arbitrary point P in $\mathcal{S}$, we identify the tangent space $T_P\mathcal{S}$ with the set of traceless Hermitian matrices $\mathcal{A}_0$:

$$\mathcal{A}_0 = \{X \in \mathrm{Herm}(d) \mid \mathrm{Trace}X = 0\}.$$

For each vector $X \in \mathcal{A}_0$, we denote the corresponding vector field as $\widetilde{X}$.

For an arbitrary point $P \in \mathcal{S}$ and a vector $X \in \mathcal{A}_0$, we define $\omega_P(\widetilde{X}) \in \mathrm{Herm}(d)$ *and ξ by*

$$\begin{aligned} X &= \frac{1}{2}(P\omega_P(\widetilde{X}) + \omega_P(\widetilde{X})P), \\ \xi &= -I_d. \end{aligned} \tag{17}$$

An immersion $\omega_P(\widetilde{X})$ in (17) is called a symmetric logarithmic derivative.

Then $\{\omega, \xi\}$ is an equiaffine affine distribution from $\mathcal{S}$ to $\mathrm{Herm}(d)$. *The affine fundamental form and the induced connection for $\{\omega, \xi\}$ can be written*

$$\begin{aligned} h_P(\widetilde{X}, \widetilde{Y}) &= \frac{1}{2}\mathrm{Trace}\left(P(\omega_P(\widetilde{X})\omega_P(\widetilde{Y}) + \omega_P(\widetilde{Y})\omega_P(\widetilde{X}))\right), \\ \left(\nabla_{\widetilde{X}}\widetilde{Y}\right)_P &= \widetilde{\left(h_P(\widetilde{X}, \widetilde{Y})P - \frac{1}{2}(X\omega_P(\widetilde{Y}) + \omega_P(\widetilde{Y})X)\right)}. \end{aligned}$$

The affine fundamental form h coincides with the SLD Fisher metric *in quantum information theory. We can check that ∇ is curvature-free but it is not torsion-free in general.*

Lastly, let us define geometric pre-divergence function.

Let $\{\omega, \xi\}$ be an equiaffine affine distribution to $\boldsymbol{R}^{n+1}$, and $\boldsymbol{R}_{n+1}$ the dual vector space of $\boldsymbol{R}^{n+1}$. For an arbitrary points $p \in M$, we define a map $v : M \to \boldsymbol{R}_{n+1}$ by

$$\langle v(p), \xi_p \rangle = 1, \quad \langle v(p), \omega_p(X) \rangle = 0.$$

We call v the *conormal map* of $\{\omega, \xi\}$.

For an equiaffine affine distribution $\{\omega, \xi\}$, and its conormal map v, we define a function ρ on $TM \times M$ by

$$\rho(X, q) = \langle v(q), \omega(X) \rangle.$$

We call ρ a *geometric pre-divergence* on M.

Some important theorems for divergences, e.g. the projection theorem [4], hold for a geometric pre-divergence. See [11].

References

[1] N. Abe, Affine immersions and conjugate connections, Tensor, **55** (1994), 276–280.

[2] S. Amari, Differential-Geometrical Methods in Statistics, Lecture Notes in Statist., **28**, Springer-Verlag, 1985.

[3] S. Amari. Information geometry of the EM and em algorithms for neural networks, Neural Networks, **8** (1995), 1379–1408.

[4] S. Amari and H. Nagaoka, Method of Information Geometry, Amer. Math. Soc., Providence, Oxford Univ. Press, Oxford, 2000.

[5] F. Dillen, K. Nomizu and L. Vranken, Conjugate connections and Radon's theorem in affine differential geometry, Monatsh. Math., **109** (1990), 221–235.

[6] S. Eguchi, Geometry of minimum contrast, Hiroshima Math. J., **22** (1992), 631–647.

[7] S. Ikeda, T. Tanaka and S. Amari, Information geometry of turbo and low-density parity-check codes, IEEE Trans. Inform. Theory, **50** (2004), 1097–1114.

[8] S. Ivanov, On dual-projectively flat affine connections, J. Geom., **53** (1995), 89–99.

[9] T. Kurose, On the divergences of 1-conformally flat statistical manifolds, Tôhoku Math. J., **46** (1994), 427–433.

[10] T. Kurose, Conformal-projective geometry of statistical manifolds, Interdiscip. Inform. Sci., **8** (2002), 89–100.

[11] T. Kurose, Statistical manifolds admitting torsion, Geometry and Something 2007, (in Japanese).

[12] S. L. Lauritzen, Statistical manifolds, In: Differential Geometry in Statistical Inferences, IMS Lecture Notes Monogr. Ser., **10**, Inst. Math. Statist., Hayward California, 1987, pp. 96–163.

[13] H. Matsuzoe, On realization of conformally-projectively flat statistical manifolds and the divergences, Hokkaido Math. J., **27** (1998), 409–421.

[14] H. Matsuzoe, Geometry of contrast functions and conformal geometry, Hiroshima Math. J., **29** (1999), 175–191.

[15] H. Matsuzoe, Geometry of semi-Weyl manifolds and Weyl manifolds, Kyushu J. Math., **55** (2001), 107–117.

[16] H. Matsuzoe, Geometry of statistical manifolds and its generalization, In: Proceedings of the 8th International Workshop on Complex Structures and Vector Fields, World Scientific, 2007, pp. 244–251.

[17] N. Murata, T. Takenouchi, T. Kanamori and S. Eguchi, Information geometry of U-Boost and Bregman divergence, Neural Computation, **16** (2004), 1437–1481.

[18] K. Nomizu, Affine connections and their use, In: Geometry and Topology of Submanifolds VII, (ed. F. Dillen), World Scientific, 1995, pp. 197–203.

[19] K. Nomizu and T. Sasaki, Affine Differential Geometry—Geometry of Affine Immersions—, Cambridge Univ. Press, Cambridge, 1994.

[20] A. P. Norden, Affinely Connected Spaces (in Russian), Moskva-Leningrad, 1950.

[21] A. Ohara, Geometry of distributions associated with Tsallis statistics and properties of relative entropy minimization, Phys. Lett. A, **370** (2007), 184–193.

[22] I. Okamoto, S. Amari and K. Takeuchi, Asymptotic theory of sequential estimation: differential geometrical approach, Ann. Statist., **19** (1991), 961–981.

[23] H. Shima, The Geometry of Hessian Structures, World Scientific, 2007.

[24] U. Simon, A. Schwenk-Schellschmidt and H. Viesel, Introduction to the Affine Differential Geometry of Hypersurfaces, Lecture Notes Sci. Univ. Tokyo, 1991.

[25] K. Uohashi, On α-conformal equivalence of statistical submanifolds, J. Geom., **75** (2002), 179–184.

[26] J. Zhang and H. Matsuzoe, Dualistic differential geometry associated with a convex function, In: Advances in Applied Mathematics and Global Optimization (in Honor of Gilbert Strang), (eds. D. Gao and H. Sherali), Springer-Verlag, 2009, pp. 437–464.

Department of Computer Science and Engineering
Graduate School of Engineering
Nagoya Institute of Technology
Gokiso-cho, Showa-ku
Nagoya, 466-8555
Japan

E-mail address: `matsuzoe@nitech.ac.jp`