CONSTRUCTION OF GEOMETRIC DIVERGENCE ON q-EXPONENTIAL FAMILY

Hiroshi Matsuzoe

Abstract: *A divergence function is a skew-symmetric distance like function on a manifold. In the geometric theory of statistical inference, such a divergence function is useful. In complex systems, Tsallis anomalous statistics is developing rapidly. A q-exponential family is an important statistical model in Tsallis statistics. For this q-exponential family, a divergence function is constructed from the viewpoint of affine differential geometry.*

Keywords: *information geometry, affine differential geometry, Tsallis statistics, divergence, q-exponential family*

ACM Classification Keywords: *G.0 General*

MSC: *53A15, 62B10, 53A30*

Introduction

A divergence function is a skew-symmetric squared distance like function on a manifold. In information geometry, which is a geometric theory of statistical inference, divergence functions play important roles. The Kullback-Leibler divergence is a typical example of such divergence functions [Amari and Nagaoka, 2000].

Recently, theory of complex systems has been developing rapidly. In complex systems, Tsallis statistics is one of anomalous statistics based on generalized entropies [Tsallis, 2009]. A q-exponential family is an important statistical model in Tsallis statistics, which includes various long tail probability distributions.

In this paper, we construct a divergence function on a q-exponential family from the viewpoint of affine differential geometry. For this purpose, we study generalized conformal geometry and affine hypersurface theory.

In information geometry, affine differential geometry is more useful than Riemannian geometry, and it is known that affine differential geometry generalizes geometry of distance. See [Matsuzoe, 2009], for example.

Statistical manifolds and generalized conformal geometry

First, let us recall geometry of statistical manifolds. Further details about statistical manifolds and generalized conformal equivalence relations, see [Matsuzoe, 2009; Matsuzoe, 2010], for example.

Definition 1 ([Kurose, 1994]). Let (M, h) be a semi-Riemannian manifold and let ∇ be a torsion-free affine connection on M. We say that the triplet (M, ∇, h) is a *statistical manifold* if the covariant derivative ∇h is a totally symmetric, that is, the following equation folds:

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

where X, Y and Z are arbitrary vector fields on M. The symmetric $(0, 3)$-tensor field $C := \nabla h$ is called the *cubic form* of (M, ∇, h). (In information geometry, the triplet (M, h, C) is also called a statistical manifold.)

For a statistical manifold (M, ∇, h), we can define another torsion-free affine connection by

$$Xh(Y, Z) = h(\nabla_X Y, Z) + h(Y, \nabla^*_X Z).$$

We call ∇^* the *dual connection* of ∇ with respect to h. In this case, the triplet (M, ∇^*, h) is a statistical manifold, which is called the *dual statistical manifold* of (M, ∇, h).

Here let us recall that a statistical model is a statistical manifold [Amari and Nagaoka, 2000]. Let $(\Omega, \mathcal{F})$ be a measurable space, and let S be a parametric statistical model on Ω. That is, S is a set of all probability densities on Ω parametrized by $\theta = (\theta^1, \dots, \theta^n) \in \Theta \subset \mathbf{R}$ such that

$$S = \left\{ p(x;\theta) \,\middle|\, p(x;\theta) > 0, \ \int_\Omega p(x;\theta)dx = 1 \right\},$$

where x is a random variable on Ω. We assume that S is a manifold with a local coordinate system $(\theta^1, \dots, \theta^n)$. For simplicity, set $\partial_i := \partial/\partial\theta^i$. We define a symmetric matrix (g^F_{ij}) $(i, j = 1, 2, \dots, n)$ by

$$g^F_{ij}(\theta) := \int_\Omega \left(\partial_i \log p(x;\theta)\right)\left(\partial_j \log p(x;\theta)\right) p(x;\theta)dx.$$

If $(g^F_{ij}(\theta))$ is positive definite, it determines a Riemannian metric on M. We call g^F the *Fisher metric* on M.

For an arbitrary constant $\alpha \in \mathbf{R}$, we can define a torsion-free affine connection $\nabla^{(\alpha)}$ on S by

$$\Gamma^{(\alpha)}_{ij,k}(\theta) := \int_\Omega \left\{ \partial_i\partial_j \log p(x;\theta) + \frac{1-\alpha}{2}\left(\partial_i \log p(x;\theta)\right)\left(\partial_j \log p(x;\theta)\right) \right\} \left(\partial_k \log p(x;\theta)\right) p(x;\theta)dx,$$

where $\Gamma^{(\alpha)}_{ij,k}(\theta)$ is the Christoffel symbol of the first kind, i.e., $g^F(\nabla^{(\alpha)}_{\partial_i}\partial_j, \partial_k) = \Gamma^{(\alpha)}_{ij,k}(\theta)$. The affine connection $\nabla^{(\alpha)}$ is called the α-*connection* on M. For an arbitrary constant $\alpha \in \mathbf{R}$, $\nabla^{(\alpha)}$ and $\nabla^{(-\alpha)}$ are mutually dual with respect to g^F, and the triplets $(M, \nabla^{(\alpha)}, g^F)$ and $(M, \nabla^{(-\alpha)}, g^F)$ are dual statistical manifolds. We remark that $(M, \nabla^{(\alpha)}, g^F)$ is invariant under the choice of the reference measure on Ω. Hence we call $(M, \nabla^{(\alpha)}, g^F)$ an *invariant statistical manifold*.

A statistical model S_e is an *exponential family* if $S_e = \left\{ p(x;\theta) \,\middle|\, p(x;\theta) = \exp\left[\sum_{i=1}^n F_i(x)\theta^i - \psi(\theta)\right] \right\}$, where $F_1(x), \dots, F_n(x)$ are random variables on Ω, and ψ is a convex function on Θ. For an exponential family, the Fisher metric g^F and the α-connection $\nabla^{(\alpha)}$ are given by

$$g^F_{ij}(\theta) = \partial_i\partial_j\psi(\theta), \quad \Gamma^{(\alpha)}_{ij,k} = \frac{1-\alpha}{2} C^F_{ijk}(\theta) = \frac{1-\alpha}{2}\partial_i\partial_j\partial_k\psi(\theta).$$

Since $\Gamma^{(1)}_{ij,k} \equiv 0$, the 1-connection $\nabla^{(1)}$ is flat and $\{\theta^1, \dots, \theta^n\}$ is an affine coordinate system on S_e. For this reason, the 1-connection $\nabla^{(1)}$ on an exponential family S_e is called an *exponential connection*.

Next, we consider the 1-conformal equivalence relation on statistical manifolds.

Definition 2 ([Kurose, 1994]). We say that statistical manifolds (M, ∇, h) and $(M, \bar{\nabla}, \bar{h})$ are 1-*conformally equivalent* if there exists a function λ such that

$$\begin{aligned} \bar{h}(X,Y) &= e^\lambda h(X,Y), \\ \bar{\nabla}_X Y &= \nabla_X Y - h(X,Y)\mathrm{grad}_h\lambda, \end{aligned}$$

where $\mathrm{grad}_h\lambda$ is the gradient vector field of λ with respect to h. A statistical manifold (M, ∇, h) is said to be 1-*conformally flat* if (M, ∇, h) is locally 1-conformally equivalent to some flat statistical manifold.

Geometry of q-exponential family

In this section, we consider geometry of q-exponential families. First, let us recall the definitions of q-exponential functions and q-exponential families

For a fixed positive number q, the q-*exponential function* and the q-*logarithm function* are defined by

$$\exp_q x := \begin{cases} (1 + (1-q)x)^{\frac{1}{1-q}}, & q \neq 1, \quad (1 + (1-q)x > 0), \\ \exp x, & q = 1, \end{cases}$$

$$\log_q x := \begin{cases} \dfrac{x^{1-q} - 1}{1-q}, & q \neq 1, \ (x > 0), \\ \log x, & q = 1, \ (x > 0), \end{cases}$$

respectively. When $q \to 1$, the q-exponential function recovers the standard exponential, and the q-logarithm function recovers the standard logarithm. A statistical model S_q is said to be a *q-exponential family* if $S_q = \left\{p(x,\theta) \,\middle|\, p(x;\theta) = \exp_q\left[\sum_{i=1}^n \theta^i F_i(x) - \psi(\theta)\right], \theta \in \Theta \subset \mathbf{R}^n\right\}$. We remark that a q-exponential family is obtained from the maximization principle of the generalized entropy in Tsallis statistics [Tsallis, 2009].

In the same manner as an exponential family, we can define geometric objects on S_q. We define the *q-Fisher metric* g^q, the *q-cubic form* C^q, and the *q-exponential connection* $\nabla^{q(e)}$ by

$$g^q_{ij} := \frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}, \quad C^q_{ijk} := \frac{\partial^3\psi}{\partial\theta^i\partial\theta^j\partial\theta^k}, \quad g^q(\nabla^{q(e)}_X Y, Z) := g^q(\nabla^{q(LC)}_X Y, Z) - \frac{1}{2}C^q(X,Y,Z),$$

respectively, where $\nabla^{q(LC)}$ is the Levi-Civita connection with respect to g^q. In this case, $(S_q, \nabla^{q(e)}, g^q)$ is a flat statistical manifold. Denote by $\nabla^{(2q-1)}$ the $(2q-1)$-connection on S_q, i.e., $\alpha = 2q-1$. The triplet $(S_q, \nabla^{(2q-1)}, g^F)$ is an invariant statistical manifold. The following lemma is given in [Matsuzoe and Ohara, 2011].

Lemma 1. *For a q-exponential family S_q, consider a flat statistical manifold $(S_q, \nabla^{q(e)}, g^q)$, and an invariant statistical manifold $(S_q, \nabla^{(2q-1)}, g^F)$. Then these two statistical manifolds are 1-conformally equivalent, that is,*

$$\begin{aligned} g^q(X,Y) &= \frac{q}{Z_q} g^F(X,Y), \\ \nabla^{q(e)}_X Y &= \nabla^{(2q-1)}_X Y - g^F(X,Y)\mathrm{grad}_h\left(\log\frac{q}{Z_q}\right), \end{aligned}$$

where $Z_q(\theta) = \int_\Omega p(x;\theta)^q dx$. In particular, $(S_q, \nabla^{(2q-1)}, g^F)$ is 1-conformally flat.

The geometric divergence on q-exponential family

In this section, we consider realizations of q-exponential families into affine space, and constructions of geometric divergences. For more details about affine differential geometry, see [Nomizu and Sasaki, 1994].

Let M be an n-dimensional manifold, and let f be an immersion from M to $\mathbf{R}^{n+1}$. Denote by ξ a *transversal vector field*, that is, the tangent space is decomposed as $T_{f(p)}\mathbf{R}^{n+1} = f_*(T_pM) \oplus \mathrm{Span}\{\xi_p\}, ({}^\forall p \in M)$. We call the pair $\{f, \xi\}$ an *affine immersion* from M to $\mathbf{R}^{n+1}$.

Denote by D the standard flat affine connection. Then we have the following decompositions:

$$\begin{aligned} D_X f_* Y &= f_*(\nabla_X Y) + h(X,Y)\xi, & (1)\\ D_X \xi &= -f_*(SX) + \tau(X). & (2) \end{aligned}$$

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, the immersion f is said *nondegenerate*. If $\tau = 0$, the affine immersion $\{f, \xi\}$ is said *equiaffine*. It is known that the induced objects (M, ∇, h) becomes a 1-conformally flat statistical manifold if the affine immersion is nondegenerate and equiaffine [Kurose, 1994].

The induced objects depend on the choice of transversal vector field. For a function ϕ on M, and a vector field Z on M, if we change a transversal vector field ξ to $\bar{\xi} = e^\phi \xi + f_*(Z)$,Then the induced objects change as follows (See Chapter 2 in [Nomizu and Sasaki, 1994]):

$$\begin{aligned} \bar{h}(X,Y) &= e^{-\phi}h(X,Y), & (3)\\ \overline{\nabla}_X Y &= \nabla_X Y - e^{-\phi}h(X,Y)Z, & (4)\\ \bar{\tau}(X) &= \tau(X) + e^{-\phi}h(X,Z) - d\phi(X). & (5) \end{aligned}$$

Theorem 1. *For a q-exponential family $S_q = \left\{p(x,\theta) \,\middle|\, p(x;\theta) = \exp_q\left[\sum_{i=1}^n \theta^i F_i(x) - \psi(\theta)\right]\right\}$, set*

$$\begin{aligned} f_q(p(\theta)) &= \{\theta^1, \ldots, \theta^n, \psi(\theta)\}^T, \\ \xi_q &= \{0, \ldots, 0, 1\}^T, \\ \xi_q^F &= \{e^{-\phi}\xi_q + f_* \mathrm{grad}_h \phi\}, \end{aligned}$$

where $\phi = \log(Z_q/q)$ *and* $Z_q(\theta) = \int_\Omega p(x;\theta)^q dx$. *Then the pair* $\{f_q, \xi_q\}$ *is an affine immersion which realizes the flat statistical manifold* $(S_q, \nabla^{q(e)}, g^q)$. *The pair* $\{f_q, \xi_q^F\}$ *is an affine immersion which realizes the invariant statistical manifold* $(S_q, \nabla^{2q-1)}, g^F)$.

Proof. From the definitions of q-Fisher metric, q-exponential connection and Equations (1)-(2), the affine immersion $\{f_q, \xi_q\}$ realizes $(S_q, \nabla^{q(e)}, g^q)$ in $\mathbf{R}^{n+1}$. From Lemma 1 and Equations (3)-(5), the affine immersion $\{f_q, \xi_q^F\}$ realizes $(S_q, \nabla^{2q-1)}, g^F)$ in $\mathbf{R}^{n+1}$. □

Finally, we define the geometric divergence on q-exponential family. For a nondegenerate equiaffine immersion $\{f, \xi\}$, the *conormal map* $\nu : M \to \mathbf{R}_{n+1}$ of $\{f, \xi\}$ by $\nu_p(f_*X) = 0$ and $\nu_p(\xi(p)) = 1$. Then we define a skew-symmetric function ρ on $M \times M$ by

$$\rho(p, q) = \nu_p(f(q) - f(p)).$$

The function ρ is called the *geometric divergence* [Kurose, 1994; Matsuzoe, 2010]. In fact, ρ induces the given statistical manifold structure, that is, the geometric divergence ρ_q for $\{f_q, \xi_q\}$ induces $(S_q, \nabla^{q(e)}, g^q)$, and ρ_F for $\{f_q, \xi_q^F\}$ induces $(S_q, \nabla^{2q-1)}, g^F)$.

Conclusion

A q-exponential family is an important statistical model in Tsallis statistics. In this paper, we give a hypersurface affine immersion of q-exponential family. As a consequence, we obtain a divergence function on a q-exponential family, which is an important distance like function in information geometry.

Acknowledgments

This research was partially supported by JSPS, Grant-in-Aid for Young Scientists (B), No. 23740047.

Bibliography

[Amari and Nagaoka, 2000] S. Amari and H. Nagaoka, Methods of information geometry, Amer. Math. Soc., Providence, Oxford University Press, Oxford, 2000.

[Kurose, 1994] T. Kurose, On the divergences of 1-conformally flat statistical manifolds, Tôhoku Math. J., 46, 427-433, 1996.

[Matsuzoe, 2009] H. Matsuzoe, Computational Geometry from the Viewpoint of Affine Differential Geometry, Lecture Notes in Computer Science, 5416, 103-123, 2009.

[Matsuzoe, 2010] H. Matsuzoe, Statistical manifolds and affine differential geometry, Adv. Stud. Pure Math., 57, 303-321, 2010.

[Matsuzoe and Ohara, 2011] H. Matsuzoe and A. Ohara, Geometry for q-exponential families, Recent Progress in Differential Geometry and Its Related Fields: Proceedings of the 2nd International Colloquium on Differential Geometry and its Related Fields, World Sci. Publ., 55-71, 2011.

[Nomizu and Sasaki, 1994] K. Nomizu and T. Sasaki, Affine differential geometry – Geometry of Affine Immersions –, Cambridge University Press, 1994.

[Tsallis, 2009] C. Tsallis, Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World, Springer, 2009.

Authors' Information

Hiroshi Matsuzoe *- Associate Professor, Ph.D., Department of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology, Nagoya 466-8555, Japan; e-mail:* *matsuzoe@nitech.ac.jp*