Tôhoku Math. J.
46 (1994), 427–433

ON THE DIVERGENCES OF 1-CONFORMALLY FLAT STATISTICAL MANIFOLDS

TAKASHI KUROSE

(Received June 7, 1993, revised October 8, 1993)

Abstract. We canonically define a square-distance-like function on each simply connected 1-conformally flat statistical manifold and prove that the generalized Pythagorean theorem holds if the statistical manifold has constant curvature.

Introduction. A manifold with a torsion-free affine connection ∇ and a pseudo-Riemannian metric h is called a *statistical manifold* if ∇h is symmetric. Recently, it has become recognized that the geometry of statistical manifolds is useful in statistics, since each family of probability distributions with sufficient regularity has the statistical manifold structure which is naturally determined by the family.

In 1982, through the geometric study of statistical inference, Nagaoka and Amari showed that each flat statistical manifold has a canonical square-distance-like function, which they called the *divergence* of the statistical manifold. Moreover, they proved that the divergence locally satisfies the Pythagorean theorem for geodesic right triangles (see [A1] and [A2]).

The purpose of this paper is to generalize their results to a wider class of statistical manifolds from the view point of affine geometry. We shall show the following: *Divergences can be canonically defined for any simply connected* 1-*conformally flat statistical manifolds. If a statistical manifold has constant curvature, the divergence satisfies the generalized Pythagorean theorem.*

Sections 1 and 2 are devoted to collecting preliminary facts on statistical manifolds and affine immersions, respectively. The main results are given in Section 3.

The author would like to thank the referee for his careful reading and kind comments.

1. Statistical manifolds. Let ∇ and h be a torsion-free affine connection and a pseudo-Riemannian metric on a manifold M, respectively. The triple (M, ∇, h) is called a *statistical manifold* if ∇h is symmetric.

When a statistical manifold (M, ∇, h) is given, we can define another torsion-free affine connection $\bar{\nabla}$ by

---

Partly supported by the Grants-in-Aid for Encouragement of Young Scientists, The Ministry of Education, Science and Culture, Japan.

1991 *Mathematics Subject Classification*. Primary 53A15.

$$Xh(Y, Z)=h(\nabla_X Y, Z)+h(Y, \bar{\nabla}_X Z),$$

where X, Y and Z are arbitrary vector fields on M. It is easy to see that $(M, \bar{\nabla}, h)$ is also a statistical manifold. We call $\bar{\nabla}$ the *dual connection* of ∇ with respect to h, and $(M, \bar{\nabla}, h)$ the *dual statistical manifold* of (M, ∇, h).

We say that (M, ∇, h) has *constant curvature* K if the curvature tensor R of ∇ satisfies

$$R(X, Y)Z=K\cdot\{h(Y, Z)X-h(X, Z)Y\}.$$

A statistical manifold with constant curvature 0 is said to be *flat*.

For a real number α, two statistical manifolds (M, ∇, h) and $(M, \tilde{\nabla}, \tilde{h})$ are said to be *α-conformally equivalent* if there exists a function φ on M such that

$$\tilde{h}(X, Y)=e^{\varphi}h(X, Y),$$

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

It is easily verified that α-conformal equivalence is an equivalence relation on the class of statistical manifolds. Two statistical manifolds (M, ∇, h) and $(M, \tilde{\nabla}, \tilde{h})$ are α-conformally equivalent if and only if their dual statistical manifolds $(M, \bar{\nabla}, h)$ and $(M, \bar{\tilde{\nabla}}, \tilde{h})$ are $(-\alpha)$-conformally equivalent, since

$$\begin{aligned}\tilde{h}(Y, \bar{\tilde{\nabla}}_X Z)&=X\tilde{h}(Y, Z)-\tilde{h}(\tilde{\nabla}_X Y, Z)\\ &=d\varphi(X)\tilde{h}(Y, Z)+\tilde{h}(\nabla_X Y, Z)+\tilde{h}(Y, \bar{\nabla}_X Z)-\Big\{\tilde{h}(\nabla_X Y, Z)\\ &\quad-\frac{1+\alpha}{2}d\varphi(Z)\tilde{h}(X, Y)+\frac{1-\alpha}{2}\{d\varphi(X)\tilde{h}(Y, Z)+d\varphi(Y)\tilde{h}(X, Z)\}\Big\}\\ &=\tilde{h}(Y, \bar{\nabla}_X Z)-\frac{1-\alpha}{2}d\varphi(Y)\tilde{h}(X, Z)+\frac{1+\alpha}{2}\{d\varphi(Z)\tilde{h}(X, Y)+d\varphi(X)\tilde{h}(Z, Y)\}.\end{aligned}$$

A statistical manifold (M, ∇, h) is said to be (∇-)*α-conformally flat* if (M, ∇, h) is α-conformally equivalent to a flat statistical manifold in a neighbourhood of an arbitrary point of M. By Proposition 9.1 in [NS], (M, ∇, h) is (-1)-conformally flat if and only if ∇ is a projectively flat connection with symmetric Ricci tensor. Hence the following proposition holds.

PROPOSITION 1. *A statistical manifold (M, ∇, h) is 1-conformally flat if and only if the dual connection $\bar{\nabla}$ is a projectively flat connection with symmetric Ricci tensor.*

The geometric meaning of ∇-α-conformal flatness ($\alpha\neq\pm1$) is not clear even if $\alpha=0$. The idea of α-conformal change was originally formulated by Okamoto, Amari and Takeuchi [OAT], through the geometric consideration of sequential estimation theory in statistical inference. In their study, however, it seems that h-α-conformal flatness, that is, conformal flatness of h in the usual sense is meaningful.

2. Affine immersions. In this section, we recall several definitions and preliminary facts on affine hypersurface theory. For more details, see [N] or [NP].

Let M be a manifold of dimension $n \geq 2$. A pair (x, ξ) is called an *affine immersion* of M into the $(n+1)$-dimensional affine space $\boldsymbol{R}^{n+1}$ if x is an immersion of M into $\boldsymbol{R}^{n+1}$ and ξ is a transversal vector field along x. For a given affine immersion (x, ξ) of M, the *induced connection* ∇ and the *second fundamental form* h are determined by

$$(x^*D)_X(x_* Y) = x_*(\nabla_X Y) + h(X, Y)\xi ,$$

where D is the standard affine connection of $\boldsymbol{R}^{n+1}$, and X and Y are arbitrary vector fields on M.

If h is non-degenerate everywhere on M, we say that (x, ξ) is *non-degenerate*. It is easy to show that the definition is independent of the choice of ξ. We say that (x, ξ) is *equiaffine* if $(x^*D)_X\xi$ is tangent to M for any vector field X on M, or equivalently, if ∇h is symmetric. Hence, (M, ∇, h) is a statistical manifold if and only if (x, ξ) is a non-degenerate equiaffine immersion. In this case, we say that the non-degenerate equiaffine immersion (x, ξ) realizes the statistical manifold (M, ∇, h) in $\boldsymbol{R}^{n+1}$. It is known that such an equiaffine immersion is uniquely determined up to affine transformations of $\boldsymbol{R}^{n+1}$.

Conversely, Dillen, Nomizu and Vrancken proved the following theorem.

THEOREM (cf. [DNV]). *A simply connected statistical manifold (M, ∇, h) can be realized in $\boldsymbol{R}^{n+1}$ if and only if $\bar{\nabla}$ is a projectively flat connection with symmetric Ricci tensor.*

By this theorem and Proposition 1 in Section 1, we have the following corollary.

COROLLARY. *A simply connected statistical manifold can be realized in $\boldsymbol{R}^{n+1}$ if and only if it is* 1*-conformally flat.*

When $\dim M \geq 3$, as in Riemannian geometry, (M, ∇, h) has constant curvature if both ∇ and $\bar{\nabla}$ are projectively flat connections (see [K3]). Therefore both a statistical manifold and its dual statistical manifold can be realized in $\boldsymbol{R}^{n+1}$ if and only if the statistical manifold has constant curvature. This fact was shown in [K1] in a direct way. The following result was also proved there.

THEOREM. *A simply connected statistical manifold with constant curvature K can be realized in $\boldsymbol{R}^{n+1}$ by an equiaffine immersion (x, ξ) such that $\xi + Kx$ is constant on M.*

The *affine shape operator* S of an equiaffine immersion (x, ξ) is the $(1, 1)$-tensor field on M defined by $x_*(S(X)) = -(x^*D)_X\xi$ for an arbitrary vector field X on M. The statistical manifold realized by (x, ξ) has constant curvature K if and only if $S = KI$.

Let (x, ξ) be a non-degenerate equiaffine immersion of M into $\boldsymbol{R}^{n+1}$. We denote by $\boldsymbol{R}_{n+1}$ the dual space of $\boldsymbol{R}^{n+1}$ and by $\langle \bar{u}, u \rangle$ the pairing of $\bar{u} \in \boldsymbol{R}_{n+1}$ and $u \in \boldsymbol{R}^{n+1}$. We define the *conormal map* $\bar{x}$ from M to $\boldsymbol{R}_{n+1}$ by

$$\langle \bar{x}(p), x_* X \rangle = 0 \qquad \text{for all } X \in T_p(M), \text{ and}$$
$$\langle \bar{x}(p), \xi(p) \rangle = 1$$

for $p \in M$. The conormal map has the following properties.

PROPOSITION 2. *For arbitrary vector fields X and Y on M, we have*

$$\langle \bar{x}_* X, \xi \rangle = 0 ,$$
$$\langle \bar{x}_* Y, x_* X \rangle = -h(Y, X) , \quad \textit{and}$$
$$(\bar{x}^* D)_X (\bar{x}_* Y) = \bar{x}_* (\bar{\nabla}_X Y) - h(Y, S(X)) \bar{x} .$$

For the proof and more information about conormal maps, see [NS], [NO] or [K2].

3. Divergences. Let (M, ∇, h) be a simply connected 1-conformally flat statistical manifold, and (x, ξ) a non-degenerate equiaffine immersion realizing (M, ∇, h) in $\boldsymbol{R}^{n+1}$. We define the *divergence* ρ of (M, ∇, h) by

$$\rho(p, q) = \langle \bar{x}(q) , \; x(p) - x(q) \rangle \qquad (p, q \in M) ,$$

where $\bar{x}$ is the conormal map of (x, ξ). The definition of ρ is independent of the choice of a realization of (M, ∇, h). The fucntion $\rho(p, \cdot)$ for fixed p is known as the affine distance function for (x, ξ) from the point $x(p)$.

Identifying the tangent space $M \times M$ at (p, q) with the direct sum $T_p(M) \oplus T_q(M)$, we use the following notation:

$$\rho[X_1 \cdots X_i \mid Y_1 \cdots Y_j](p) = (X_1, 0) \cdots (X_i, 0)(0, Y_1) \cdots (0, Y_j) \rho \mid_{(p, p)} ,$$

where $X_1, \cdots, X_i, Y_1, \cdots, Y_j$ $(i, j \geq 0)$ are arbitrary vector fields on M.

PROPOSITION 3. *The divergence ρ has the following properties*:
(1) *ρ is identically zero on the diagonal set of $M \times M$.*
(2) *For arbitrary vector fields X, Y and Z on M,*

$$\rho[X \mid \,] = 0 ,$$
$$\rho[X \mid Y] = -h(X, Y) , \quad \textit{and}$$
$$\rho[XY \mid Z] = -h(\nabla_X Y, Z) .$$

PROOF. We shall prove (2). By definition, we have

$$((X, 0)\rho)(p, q) = \langle \bar{x}(q), \; (x_* X)_p \rangle, \qquad ((X, 0)(0, Y)\rho)(p, q) = \langle (\bar{x}_* Y)_q , \; (x_* X)_p \rangle ,$$

and

$$((X, 0)(Y, 0)(0, Z)\rho)(p, q) = \langle(\bar{x}_* Z)_q,\ ((x^* D)_X(x_* Y))_p\rangle$$
$$= \langle(\bar{x}_* Z)_q,\ (x_*(\nabla_X Y))_p + (h(X, Y))(p)\xi(p)\rangle .$$

Setting $q=p$, we obtain the required equalities from the definition of the conormal map $\bar{x}$ and Proposition 2 in Section 2. □

In the geometric study of statistics, a function ρ on $M\times M$ with the properties in Proposition 3 is called a *normalized yoke* of the statistical manifold (M, ∇, h) (see, e.g., [BNJK]). A normalized yoke ρ is called a *contrast function* if it vanishes only on the diagonal set of $M\times M$ (see [E]). Recently, Matumoto [M] showed that each statistical manifold, not necessarily simply connected, has a contrast function. The divergence defined above is a contrast function when the realization is globally strictly convex.

REMARK. Let ρ be a normalized yoke of a statistical manifold (M, ∇, h). For a fuction φ on M, a function $\tilde{\rho}$ on $M\times M$ defined by

$$\tilde{\rho}(p, q) = e^{\varphi(q)}\rho(p, q) \qquad (p, q\in M)$$

is a normalized yoke of a statistical manifold $(M, \tilde{\nabla}, \tilde{h})$ which is 1-conformally equivalent to (M, ∇, h) as in Section 1. In particular, if ρ is the divergence of (M, ∇, h), the divergence of $(M, \tilde{\nabla}, \tilde{h})$ coincides with $\tilde{\rho}$.

The following is an easy consequence of Proposition 3.

COROLLARY. *For arbitrary vector fields X, Y and Z on M,*

$$\rho[\,|X] = 0 ,$$
$$\rho[XY|\,] = -\rho[X|Y] = \rho[\,|XY] , \quad \textit{and}$$
$$\rho[Y|XZ] = -h(Y, \bar{\nabla}_X Z) .$$

According to this corollary, the function $\bar{\rho}(p, q) = \rho(q, p)$ is a normalized yoke of the dual statistical manifold $(M, \bar{\nabla}, h)$.

LEMMA 4. *For three distinct points p, q and r of M, let P be a 2-plane of $\boldsymbol{R}^{n+1}$ through $x(p)$ and $x(q)$ parallel to $\xi(q)$, and $\bar{P}$ a 2-plane of $\boldsymbol{R}_{n+1}$ through $\bar{x}(q)$ and $\bar{x}(r)$ parallel to $\bar{x}(q)$. If P is vertical to $\bar{P}$, then*

$$\rho(p, r) = \rho(p, q) + \rho(q, r) - \rho(p, q)(1 - \langle\bar{x}(r), \xi(q)\rangle) .$$

PROOF. The assumption implies

$$\langle(\bar{x}(r) - \bar{x}(q)) \wedge \bar{x}(q) ,\ (x(p) - x(q)) \wedge \xi(q)\rangle = 0 .$$

Calculating directly the left-hand side of this equation, we obtain the desired result. □

LEMMA 5. *Suppose that a non-degenerate equiaffine immersion (x, ξ) realizes a statistical manifold with constant curvature K. Then ξ and the conormal map $\bar{x}$ of (x, ξ) satisfies*

$$1-\langle \bar{x}(r), \xi(q)\rangle = K\rho(q, r) \qquad \textit{for any points } q, r \in M .$$

PROOF. By assumption, $\xi + Kx$ is constant everywhere on M. Therefore

$$1-\langle \bar{x}(r), \xi(q)\rangle = \langle \bar{x}(r), \xi(r)-\xi(q)\rangle = K\langle \bar{x}(r), x(q)-x(r)\rangle = K\rho(q, r) .$$

□

REMARK. Conversely, if $1-\langle \bar{x}(r), \xi(q)\rangle = K\rho(q, r)$ holds for any points q, $r \in M$, then $-h(S(X), Y) = K\rho[X \mid Y] = -Kh(X, Y)$ for arbitrary vector fields X and Y. This implies that the statistical manifold realized by (x, ξ) has constant curvature K.

Our main result is the following theorem.

MAIN THEOREM. *Suppose that* (M, ∇, h) *is a simply connected statistical manifold with constant curvature* K. *Let* γ *be a* ∇*-geodesic joining two points* p *and* q *of* M, *and* $\bar{\gamma}$ *a* $\bar{\nabla}$*-geodesic joining* q *and a point* r *of* M. *If* $\dot{\gamma}$ *and* $\dot{\bar{\gamma}}$ *are mutually orthogonal at* q *with respect to* h, *then the divergence* ρ *satisfies*

$$\rho(p, r) = \rho(p, q) + \rho(q, r) - K\rho(p, q)\rho(q, r) .$$

PROOF. We may assume that $\gamma(0) = \bar{\gamma}(0) = q$. Let P be a 2-plane of $\boldsymbol{R}^{n+1}$ through $x(q)$ parallel to $x_*(\dot{\gamma}(0))$ and $\xi(q)$. In the theory of affine hypersurfaces, it is known that the image of the geodesic γ under x is contained in P. In a same manner, we observe that the curve $\bar{x} \circ \bar{\gamma}$ is contained in a 2-plane $\bar{P}$ of $\boldsymbol{R}_{n+1}$ through $\bar{x}(q)$ parallel to $\bar{x}_*(\dot{\bar{\gamma}}(0))$ and $\bar{x}(q)$. As a consequence, we know that $x(p) \in P$ and $\bar{x}(r) \in \bar{P}$.

Using Proposition 2 in Section 2, we can easily verify that

$$\langle \bar{x}_*(\dot{\bar{\gamma}}(0)) \wedge \bar{x}(q),\ x_*(\dot{\gamma}(0)) \wedge \xi(q)\rangle = 0 .$$

This implies that P is vertical to $\bar{P}$. Therefore the theorem follows from Lemma 4 and Lemma 5. □

In the case $K \neq 0$, the formula in the main theorem can be rewritten as

$$(1-K\rho(p, r)) = (1-K\rho(p, q))(1-K\rho(q, r)) .$$

When (M, ∇, h) is either the standard sphere ($K=1$), the Euclidean space ($K=0$) or the hyperbolic space ($K=-1$), our result reduces to the well-known Pythagorean theorem on each space: If a geodesic triangle Δpqr has a right angle at the vertex q,

$$\begin{cases} \cos|\widehat{pr}| = \cos|\widehat{pq}|\cdot\cos|\widehat{qr}| & \text{if} \quad K=1 , \\ |\overline{pr}|^2 = |\overline{pq}|^2 + |\overline{qr}|^2 & \text{if} \quad K=0 , \\ \cosh|\widehat{pr}| = \cosh|\widehat{pq}|\cdot\cosh|\widehat{qr}| & \text{if} \quad K=-1 . \end{cases}$$

REFERENCES

[A1] S.-I. AMARI, Differential Geometrical Methods in Statistics, Lecture Notes in Statist. 28, Springer-

Verlag, New York, 1985.

[A2] S.-I. AMARI, Differential geometrical theory of statistics, Differential Geometry in Statistical Inference, IMS Lecture Notes-Monograph Series, vol. 10, Inst. Math. Statist., Hayward, California, 1987.

[BNJK] O. E. BARNDORFF-NIELSEN, P. E. JUPP AND W. S. KENDALL, Stochastic calculus, statistical asymptotics, Taylor strings and phyla, Dept. of Theoretical Statistics, Institute of Mathematics, Univ. of Aarhus, Research Reports 236, 1991.

[DNV] F. DILLEN, K. NOMIZU AND L. VRANCKEN, Conjugate connections and Radon's theorem in affine differential geometry, Monatsh. Math. 109 (1990), 221–235.

[E] S. EGUCHI, Geometry of minimum contrast, Hiroshima Math. J. 22 (1992), 631–647.

[K1] T. KUROSE, Dual connections and affine geometry, Math. Z. 203 (1990), 115–121.

[K2] T. KUROSE, On the Minkowski problem in affine geometry, Results Math. 20 (1991), 643–649.

[K3] T. KUROSE, Dual connections and projective geometry, Hiroshima Math. J. 23 (1993), 327–332.

[M] T. MATUMOTO, Any statistical manifold has a contrast function—On the C^3-functions taking the minimum at the diagonal of the product manifold—, preprint.

[N] K. NOMIZU, Introduction to affine differential geometry, Part I, MPI preprint MPI 88-37, 1988.

[NO] K. NOMIZU AND B. OPOZDA, On normal and conormal maps for affine hypersurfaces, Tôhoku Math. J. 44 (1992), 425–431.

[NP] K. NOMIZU AND U. PINKALL, On the geometry of affine immersions, Math. Z. 195 (1987), 165–178.

[NS] K. NOMIZU AND U. SIMON, Notes on conjugate connections, in Geometry and Topology of Submanifolds, IV (F. Dillen and L. Verstraelen, eds.), World Scientific, 1992, pp. 152–173.

[OAT] I. OKAMOTO, S.-I. AMARI AND K. TAKEUCHI, Asymptotic theory of sequential estimation: differential geometrical approach, Ann. Statist. 19 (1991), 961–981.

DEPARTMENT OF APPLIED MATHEMATICS
FUKUOKA UNIVERSITY
FUKUOKA 814-80
JAPAN