HIROSHIMA MATH. J.
23 (1993), 327–332

Any statistical manifold has a contrast function — On the C^3-functions taking the minimum at the diagonal of the product manifold

Dedicated to Professor Masahisa Adachi on his 60th birthday

Takao MATUMOTO

(Received January 16, 1992)

§1. Introduction

A statistical manifold is defined by Lauritzen [5] and nothing but a Riemannian manifold (M, g) with a symmetric covariant tensor T of order 3. The symmetric tensor defines a pair of torsion free dual connections and vice versa; The latter geometry has been studied by Amari and others in connection with statistical inferences (cf. [1] and [2]). Moreover, Eguchi [3] has proved that a contrast function also gives all the data of the statistical manifold if it exists; We will find in this paper a contrast function which induces a given statistical manifold.

DEFINITION. A contrast function of a differentiable manifold M is a real-valued smooth function ρ on $M \times M$ such that $\rho(x, y) \geq 0$ with equality if and only if $x = y$ and

$$X_x X_x \rho(x, y)_{|x=y} > 0$$

for any smooth vector field X on M which is non-zero at x.

We use the following notation for a function on M defined by the value of the partial derivative in $M \times M$ with respect to the smooth vector fields $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ on M as in [3]:

$$\rho(X_1 \ldots X_n | Y_1 \ldots Y_m)(z) = (X_1)_x \ldots (X_n)_x (Y_1)_y \ldots (Y_m)_y \rho(x, y)_{|x=z, y=z}.$$

In particular, the partial derivative $Z\rho(X_1 \ldots X_n | Y_1 \ldots Y_m)$ on M is equal to $\rho(ZX_1 \ldots X_n | Y_1 \ldots Y_m) + \rho(X_1 \ldots X_n | ZY_1 \ldots Y_m)$.

According to Eguchi [3] the Riemannian metric tensor g is defined by

$$g(X, Y) = -\rho(X | Y).$$

Since the contrast function takes the minimum at the diagonal manifold

$\Delta_M = \{(x, y) \in M \times M | x = y\}$, we have $\rho(X|\cdot) = 0$ for any X. Moreover applying X to $\rho(Y|\cdot) = 0$, we get $\rho(XY|\cdot) = -\rho(X|Y)$. Thus $g(X, Y) = \rho(XY|\cdot) = \rho(\cdot|XY)$ and g is positive definite by the definition of our contract function. The symmetry also follows from $g(X, Y) - g(Y, X) = \rho([X, Y]|\cdot) = 0$.

The trilinear form T is defined by

$$T(X, Y, Z) = -\rho(XY|Z) + \rho(Z|XY).$$

We see that $T(fX, Y, Z) = fT(X, Y, Z)$, $T(X, Y, Z) - T(X, Y, Z) = \rho([X, Y]|Z) - \rho(Z|[X, Y]) = 0$ and $T(X, Z, Y) - T(X, Y, Z) = \rho(XY|Z) + \rho(Y|XZ) - \rho(XZ|Y) - \rho(Z|XY) = X\{\rho(Y|Z) - \rho(Z|Y)\} = 0$. Therefore T is a symmetric covariant tensor of order 3.

THEOREM 1. *For any statistical manifold (M, g, T) we can find a contrast function ρ of M which gives g and T by the above formulas.*

We will make several remarks concering Theorem 1.

For the proof of Theorem 1 we use only the fact that the first derivatives of ρ should vanish; The free parameters for the values of the partial derivatives of order up to 3 on Δ_M are just given by $n(n+1)/2$ parameters g_{ij} and $n(n+1)(n+2)/6$ parameters T_{ijk}.

Due to [1], if one of the torsion free dual connections is flat (i.e., of zero curvature), then the other dual connection is also flat; Moreover if M is simply connected in addition, we have a canonical contrast function called divergence which has the form

$$\rho(x, y) = \phi(x) + \psi(x^*(y)) - x \cdot x^*(y)$$

where $x = (x_1, \dots, x_n)$ and $x^* = (x^1, \dots, x^n)$ are the affine coordinates with respect to the flat connections and $x \cdot x^*$ denotes the Euclidean inner product $\sum_i x_i x^i$; We see that the affine coordinates x and x^* of the same point are related by the Legendre transformation $x \cdot x^* = \phi(x) + \psi(x^*)$ and satisfy $g_{ij} = g(\partial_i, \partial_j) = \partial_i \partial_j \phi$, $g^{ij} = g(\partial^i, \partial^j) = \partial^i \partial^j \psi$ and $g(\partial^i, \partial_j) = \delta^i_j$. Even if M is not simply connected, the Riemannian metric g is given by $g(\partial_i, \partial_j) = \partial_i \partial_j \phi$ for a local coordinate; So such a manifold is called Hessian manifold [7] as well as 1-flat statistical manifold (cf. [5]).

If N is a regular submanifold of M for a statistical manifold (M, g, T), then N inherits the structure of statistical manifold by restricting the tensors g and T; The trosion free dual connections are given by the orthogonal projection; The restriction of a contrast function gives also a contrast function of $(N, g|N, T|N)$. So, the embedding problem of a given statistical manifold into a 1-flat statistical manifold will be an interesting problem to seek more canonical contrast functions.

If ρ is symmetric, then the tensor T is zero and the induced statistical manifold is a Riemannian manifold; A contrast function for a Riemannian manifold is given by the square of the Riemannian distance between two points x and y; For the smoothness see [4; Th. 4.3.6] for example; What Eguchi [3] proved implies that the square of the Riemannian distance for the submanifold and the distance measured in an ambient manifold coincide up to the third order. In case of Finsler spaces the square of the distance may not give a smooth function on the product manifold; The normal coordinate may not be smooth at the origin in Finsler geometry (cf. [6; §3.6]).

Eguchi [3] proved also that the curvature tensors are related to some of the fourth order property of the contrast functions; What are the free parameters in this case will be another problem.

§2. Proof of Theorem 1

By the condition that the function ρ on $M \times M$ takes the minimum at the diagonal we have already get $\rho(X|\cdot) = \rho(\cdot|X) = 0$ and then

$$\rho(XY|\cdot) = \rho(YX|\cdot) = \rho(\cdot|XY) = \rho(\cdot|YX) = -\rho(X|Y) = -\rho(Y|X)$$

for any smooth vector fields X and Y on M. The latter function is denoted by $g(X, Y)$.

LEMMA 2.1. *If* $[X, Y] = [Y, Z] = [Z, X] = 0$, *we have*

(1) $2\rho(XYZ|\cdot) = Xg(Y, Z) + Yg(Z, X) + Zg(X, Y) + T(X, Y, Z)$,
(2) $2\rho(\cdot|XYZ) = Xg(Y, Z) + Yg(Z, X) + Zg(X, Y) - T(X, Y, Z)$,
(3) $2\rho(XY|Z) = -Xg(Y, Z) - Yg(Z, X) + Zg(X, Y) - T(X, Y, Z)$ and
(4) $2\rho(X|YZ) = Xg(Y, Z) - Yg(Z, X) - Zg(X, Y) + T(X, Y, Z)$.

Proof. The first equality comes from

$$\begin{aligned} &Xg(Y, Z) + Yg(Z, X) + Zg(X, Y) \\ &\quad = X\rho(YZ|\cdot) - Y\rho(X|Z) + Z\rho(XY|\cdot) = 2\rho(XYZ|\cdot) - T(Y, Z, X). \end{aligned}$$

The right-hand side of the third one is equal to

$$\begin{aligned} &X\rho(Y|Z) + Y\rho(Z|X) - Z\rho(Y|X) - T(X, Y, Z) \\ &\quad = \rho(XY|Z) + \rho(Z|XY) - T(X, Y, Z) = 2\rho(XY|Z). \end{aligned}$$

The others are proved in the same way. q.e.d

Proof of Theorem 1. We use a local coordinate $x_1, \ldots, x_n$ of M of the first factor and write $y_1, \dots, y_n$ for the same coordinate for the second factor; Also we write $z_1, \dots, z_n$ for the same coordinate for $\Delta_M = M$. The corresponding vector fields on $M \times M$ will be denoted by $\partial_1, \dots, \partial_n$ and $\partial_1^*, \dots, \partial_n^*$; Also the corresponding vector fields on $\Delta_M = M$ by $\delta_1, \dots, \delta_n$. The Riemannian metric $g_{ij}(z) = g(\delta_i, \delta_j)(z)$ on M determines the partial derivatives $\partial_i \partial_j \rho = -\partial_i \partial_j^* \rho = -\partial_i^* \partial_j \rho = \partial_i^* \partial_j^* \rho$ on Δ_M by $\partial_i \partial_j \rho(x, y)_{|x=y=z} = g_{ij}(z)$. Now Lemma 2.1 (1) means that after g is defind on M, $T(\delta_i, \delta_j, \delta_k)(z)$ determines $\partial_i \partial_j \partial_k \rho(x, y)_{|x=y=z}$ and vice versa. Moreover the values of the other partial derivatives of order 3 on Δ_M are determined uniquely and compatibly by the other formulas in Lemma 2.1. So, we can define ρ locally by

$$\begin{aligned}\rho(x, y) = (1/2) \sum_{i=1}^n \sum_{j=1}^n g_{ij}(z)_{|z=y}(x_i - y_i)(x_j - y_j) \\ + (1/12) \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \{\delta_i g_{jk} + \delta_j g_{ki} + \delta_k g_{ij} \\ + T(\delta_i, \delta_j, \delta_k)\}(z)_{|z=y}(x_i - y_i)(x_j - y_j)(x_k - y_k).\end{aligned}$$

Take a smooth partition of unity $\{\varepsilon_\alpha(z)\}$ subordinate to the covering of $M = \Delta_M$ by the coordinate neighborhoods $\{U_\alpha\}$. We take the local coordinate $\bar{z}_1, \dots, \bar{z}_n$ of U_α and define a local contrast function ρ_α on a product neighborhood of $U_\alpha \subset \Delta_M$ by applying the above formula to the corresponding local coordinate $\bar{x}_1, \dots, \bar{x}_n, \bar{y}_1, \dots, \bar{y}_n$. In other words we write with the Einstein convention

$$\begin{aligned}\rho_\alpha(x, y) &= (1/2)\bar{g}_{ab}(\bar{z})_{|\bar{z}=\bar{y}}(\bar{x}_a - \bar{y}_a)(\bar{x}_b - \bar{y}_b) \\ &\quad + (1/12)\{\bar{\delta}_a \bar{g}_{bc} + \bar{\delta}_b \bar{g}_{ca} + \bar{\delta}_c \bar{g}_{ab} \\ &\qquad + T(\bar{\delta}_a, \bar{\delta}_b, \bar{\delta}_c)\}(\bar{z})_{|\bar{z}=\bar{y}}(\bar{x}_a - \bar{y}_a)(\bar{x}_b - \bar{y}_b)(\bar{x}_c - \bar{y}_c) \\ &= (1/2)\delta_a^i \delta_b^j g_{ij}(z)_{|z=y}(\bar{x}_a - \bar{y}_a)(\bar{x}_b - \bar{y}_b) \\ &\quad + (1/12)\{\delta_a^i \delta_i(\delta_b^j \delta_c^k g_{jk}) + \delta_b^j \delta_j(\delta_c^k \delta_a^i g_{ki}) + \delta_c^k \delta_k(\delta_a^i \delta_b^j g_{ij}) \\ &\qquad + \delta_a^i \delta_b^j \delta_c^k T(\delta_i, \delta_j, \delta_k)\}(z)_{|z=y}(\bar{x}_a - \bar{y}_a)(\bar{x}_b - \bar{y}_b)(\bar{x}_c - \bar{y}_c),\end{aligned}$$

where $\bar{\delta}_a = \partial/\partial\bar{z}_a$, $\bar{g}_{ab} = g(\bar{\delta}_a, \bar{\delta}_b)$ and $\delta_a^i = \partial z_i/\partial\bar{z}_a$. Using also the notation $\partial_i^a = \partial\bar{x}_a/\partial x_i$ and $\delta_i^a = \partial\bar{z}_a/\partial z_i$, we have

$$\begin{aligned}&\partial_i \partial_j \bar{g}_{ab}(\bar{z})_{|\bar{z}=\bar{y}}(\bar{x}_b - \bar{y}_b) \\ &= \partial_a^{i'} \delta_b^{j'} g_{i'j'}(z)_{|z=y}\{\partial_i^a \partial_j^b + \partial_j^a \partial_i^b + (\partial_i \partial_j^a)(\bar{x}_b - \bar{y}_b) + (\bar{x}_a - \bar{y}_a)(\partial_i \partial_j^b)\}.\end{aligned}$$

So, we get $\partial_i \partial_j \rho_\alpha(x, y)_{|x=y=z} = (1/2)\{g_{ij}(z) + g_{ji}(z)\} = g_{ij}(z)$.

Moreover, we have

$$\partial_i\partial_j\partial_k\rho_\alpha(x, y)_{|x=y=z} - (1/2)\{\delta_i g_{jk} + \delta_j g_{ki} + \delta_k g_{ij} + T(\delta_i, \delta_j, \delta_k)\}(z)$$
$$= (1/2)\delta_a^{i'}\delta_b^{j'}g_{i'j'}(z)\{\partial_k(\partial_i^a\partial_j^b) + \partial_k(\partial_j^a\partial_i^b) + (\partial_i\partial_j^a)\partial_k^b + \partial_k^a(\partial_i\partial_j^b)\}_{|x=z}$$
$$+ (1/12)\{\delta_{i'}(\delta_b^{j'}\delta_c^{k'})\delta_a^{i'}g_{j'k'} + \delta_{j'}(\delta_c^{k'}\delta_a^{i'})\delta_b^{j'}g_{k'i'} + \delta_{k'}(\delta_a^{i'}\delta_b^{j'})\delta_c^{k'}g_{i'j'}\}$$
$$(\partial_i^a\partial_j^b\partial_k^c + \partial_i^a\partial_k^b\partial_j^c + \partial_j^a\partial_i^b\partial_k^c + \partial_j^a\partial_k^b\partial_i^c + \partial_k^a\partial_j^b\partial_i^c + \partial_k^a\partial_i^b\partial_j^c)_{|x=z} = 0$$

because $\partial_{k|x=z} = \delta_k$, $\partial^a_{i|x=z} = \delta_i^a$, $\delta_k(\phi\psi) = (\delta_k\phi)\psi + \phi(\delta_k\psi)$ and $\delta_k(\delta_a^l\delta_i^a) = 0$ for any k, a, i, l, ϕ and ψ; In fact, we use the fact that the value of the first term $(1/2)\delta_a^{i'}\delta_b^{j'}g_{i'j'}(z)\{\partial_k(\partial_i^a\partial_j^b) + \partial_k(\partial_j^a\partial_i^b) + (\partial_i\partial_j^a)\partial_k^b + \partial_k^a(\partial_i\partial_j^b)\}_{|x=z}$ is not changed by any permutation of the set$\{i, j, k\}$; By multiplying by 12 we get the 36 terms obtained by applying the permutation of $\{i, j, k\}$ to the triple of $g_{lj}\delta_a^l(\delta_k\delta_i^a) + g_{jl}\delta_b^l(\delta_k\delta_i^b)$ and the 36 terms obtained from $g_{lj}(\delta_k\delta_a^l)\delta_i^a + g_{lj}(\delta_k\delta_b^l)\delta_i^b + g_{lj}(\delta_k\delta_c^l)\delta_i^c + g_{jl}(\delta_k\delta_a^l)\delta_i^a + g_{jl}(\delta_k\delta_b^l)\delta_i^b + g_{jl}(\delta_k\delta_c^l)\delta_i^c$.

Now without using the Einstein convention we redefine

$$\rho(x, y) = \sum_\alpha \varepsilon_\alpha(z)_{|z=y}\rho_\alpha(x, y)$$

near Δ_M. Then, in the original local coordinate

$$\partial_i\partial_j(\sum_\alpha \varepsilon_\alpha(y)\rho_\alpha(x, y))_{|x=y=z} = \sum_\alpha \varepsilon_\alpha(y)\partial_i\partial_j\rho_\alpha(x, y)_{|x=y=z} = g_{ij}(z)$$

and

$$\partial_i\partial_j\partial_k(\sum_\alpha \varepsilon_\alpha(y)\rho_\alpha(x, y))_{|x=y=z} = \sum_\alpha \varepsilon_\alpha(y)\partial_i\partial_j\partial_k\rho_\alpha(x, y)_{|x=y=z}$$
$$= (1/2)\{\delta_i g_{jk} + \delta_j g_{ki} + \delta_k g_{ij} + T(\delta_i, \delta_j, \delta_k)\}(z).$$

So, by extending this ρ which is defined only on a neighborhood of Δ_M smoothly on the whole $M \times M$ we get a desired constrast function of the given statistical manifold (M, g, T). q.e.d

Acknowledgement: I would like to express my deepest thanks to Dr. Eguchi who kindly taught the contents of [3] and after my having written this paper sent a copy of Nagoka's note which gave ρ in a local coordinate neighborhood concerning the problem 5 of [1; p.101]. Nagoka's formula for a local ρ is essentially the same as ours but not prepared enough to get a global constrast function ρ.

References

[1] S. -I. Amari, Differential-geometrical Methods in Statistics, Lecture Notes in Statistics **28**, Springer-Verlag, Berlin, 1985.

[2] S. -I. Amari, O. E. Barndorff-Nielsen, R. E. Kass, S. L. Lauritzen and C. R. Rao, Differential Geometry in Statistical Inferences, IMS Lecture Notes-Monograph Series, **10**, Institute of Mathematical Statistics, Hayward, 1987.

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

[4] S. Kobayashi and K. Nomizu, Foundation of Differential Geometry, Interscience Publisheres, New York, 1963.

[5] S. L. Lauritzen, Statistical manifolds, in Differential Geometry in Statistical Inferences, [2] above, 1987, 163–216.

[6] H. Rund, The Differential Geometry of Finsler Spaces, Grund. Math. Wiss., Springer-Verlag, Berlin, 1959.

[7] H. Shima, Hessian manifolds and convexity, in Manifolds and Lie Groups, Progress in Mathematics vol. **14**, Birkhäuser, Boston, 1981, 385–392.

Department of Mathematics
Faculty of Science
Hiroshima University
Higashi-Hiroshima 724, *Japan*