Differential Geometrical Aspects of Quantum State Estimation and Relative Entropy

Hiroshi Nagaoka

METR 94-14 September 1994

(communicated by S. Amari)

Differential Geometrical Aspects of Quantum State Estimation and Relative Entropy

Hiroshi Nagaoka*

Abstract

The space S of finite-dimensional positive density operators (quantum states) is studied in a differential geometrical viewpoint. We suppose that a generalized covariance for arbitrary two observables (Hermitian operators) is specified at each state in S, which includes the symmetrized inner product and the Bogoliubov inner product as special (but important) cases, and introduce a triplet structure $(g, \nabla^{(\mathrm{e})}, \nabla^{(\mathrm{m})})$ on S via the specified covariance, where g is a Riemannian metric and $\nabla^{(\mathrm{e})}$ and $\nabla^{(\mathrm{m})}$ are affine connections. The structure $(g, \nabla^{(\mathrm{e})}, \nabla^{(\mathrm{m})})$ is regarded as a quantum analogue of the triplet of Fisher metric, exponential connection and mixture connection on a space of probability densities introduced in the information geometry by S. Amari. Some aspects relating to the quantum state estimation and the relative entropy are treated in terms of the differential geometry, where the duality between $\nabla^{(\mathrm{e})}$ and $\nabla^{(\mathrm{m})}$ with respect to g plays an essential role.

1 Introduction

The work [12] of C. R. Rao (1945), which stated that a Riemannian metric is naturally defined on a manifold of probability densities by means of the Fisher information matrix, was undoubtedly one of the most important pioneering works in the history of differential geometrical methods in statistics. However, in order to see the real worth of this Riemannian metric (called the *Fisher metric*), we had to

*Graduate school of Information Systems, University of Electro–Communications, 1-5-1 Chofugaoka, Chofu-shi, Tokyo 182, Japan.

wait for the works of Chentsov [3] and Amari [2], who independently introduced one parameter family of affine connections $\{\alpha - \text{connection} \mid \alpha \in \boldsymbol{R}\}$ (, where we use the notation by Amari). In particular, the combination of the $(\alpha = 1)$-connection (also called the *exponential connection*, or *e-connection* for short), the $(\alpha = -1)$-connection (also called the *mixture connection*, or *m-connection* for short) and the Fisher metric has been shown to be a useful framework in studying many statistical estimation problems ([2]). This usefulness can be regarded as a consequence of a fundamental relation between (e,m)-geometry and the Cramér-Rao inequality.

Nagaoka and Amari [9] (see chap.3 of [2]) showed that the α-connection and the $(-\alpha)$-connection are mutually dual w.r.t. (with respect to) the Fisher metric in some sense, whereby a close relation between these connections and the α-divergence (a squared-distance like function) was elucidated. When $\alpha = \pm 1$, the α-divergence becomes the well known Kullback's information divergence (classical relative entropy), and thus the (e, m)-geometry can be regarded also as the geometry of the relative entropy, which is connected to the maximum entropy principle, the large deviation problems, etc.

In the quantum probability theory, statistical estimation problems on quantum states have been studied ([6] [13] [7]) based on some versions of Fisher information defined via symmetric logarithmic derivative (SLD), right logarithmic derivative (RLD), etc. We also have the quantum relative entropy

$$K(\rho, \sigma) = \mathrm{Tr}\, \rho(\log \rho - \log \sigma). \tag{1}$$

However, it seems difficult to find any essential relations between the quantum estimation theory and the quantum relative entropy. The purpose of the present paper is to develop a general differential-geometrical framework for finite quantum systems in which the world of SLD and the world of quantum relative entropy are both geometrized, whereby the gap between these two worlds is elucidated.

2 Riemannian metric and affine connection

Here we give a brief review on some notions in the general differential geometry (e.g., [8] [2]).

Let S be an n dimensional smooth manifold. We denote the tangent space of S at a point $\rho \in S$ by $T_\rho(S)$, and the totality of smooth vector fields on S by $\mathcal{X}(S)$. When an inner product g_ρ on $T_\rho(S)$ is specified for each point $\rho \in S$ and when the correspondence $\rho \mapsto g_\rho$ is smooth, we call g a Riemannian metric on S. For $\forall X, \forall Y \in \mathcal{X}(S)$, the mapping $g(X,Y) : \rho \mapsto g_\rho(X_\rho, Y_\rho)$ becomes a smooth function on S. When a coordinate system $(\xi^i) = (\xi^1, \ldots, \xi^n)$ of S is given, g is represented by the component functions $g_{ij} = g(\partial_i, \partial_j)$, where $\partial_i \stackrel{\mathrm{def}}{=} \partial/\partial\xi^i$.

The notion of affine connection can be introduced in several different ways, one of which is to represent an affine connection on S by a covariant derivative $\nabla : \mathcal{X}(S)^2 \to \mathcal{X}(S)$ $((X,Y) \mapsto \nabla_X Y)$. The coefficients of the connection ∇ w.r.t. (ξ^i) are defined by $\nabla_{\partial_i}\partial_j = \sum_k \Gamma_{ij}^k \partial_k$.

When a connection ∇ is given, the parallel displacement of a tangent vector $X = \sum_i X^i \partial_i$ along a curve $\xi^i = \xi^i(t)$ is defined by the equation $\dot{X}^k + \sum_{i,j} \Gamma_{ij}^k \dot{\xi}^i X^j = 0$, where the dot $\dot{}$ means d/dt. The condition that the parallel displacement does not depend on a curve but its end points is (locally) equivalent to $\mathcal{R} = 0$, where $\mathcal{R} : \mathcal{X}(S)^3 \to \mathcal{X}(S)$ is the *curvature tensor* defined by

$$\mathcal{R}(X,Y)Z = \nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z) - \nabla_{[X,Y]}Z, \tag{2}$$

where $[X,Y] \stackrel{\mathrm{def}}{=} XY - YX$.

A vector field $X \in \mathcal{X}(S)$ is said to be ∇-*parallel* on S when X is parallel on any curve in S, or equivalently when $\nabla_Y X = 0$ for $\forall Y \in \mathcal{X}(S)$. A coordinate system (ξ^i) is said to be ∇-*affine* when $\{\partial_i\,;\ i = 1, \ldots, n\}$ are all ∇-parallel, or equivalently when $\Gamma_{ij}^k = 0$ for $\forall i, \forall j, \forall k$. The connection ∇ is said to be *flat*, or S is said to be ∇-*flat*, when there exists a ∇-affine coordinate system. The flatness is (locally) equivalent to the condition that $\mathcal{R} = 0$ and $\mathcal{T} = 0$, where $\mathcal{T} : \mathcal{X}(S)^2 \to \mathcal{X}(S)$ is the *torsion tensor* defined by

$$\mathcal{T}(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y]. \tag{3}$$

Let M be a submanifold of S. When a metric g is given on S, a metric on M is naturally obtained by restricting g onto M. When a connection ∇ is given on S, on the other hand, the restriction of ∇ onto M does not yield a connection on M unless M is ∇-*autoparallel* in the sense that $\nabla_X Y \in \mathcal{X}(M)$ for $\forall X, \forall Y \in \mathcal{X}(M)$. However, if both g and ∇ are given on S, we can always define a connection on M by taking the orthogonal projection of $\nabla_X Y$ w.r.t. the metric g for $\forall X, \forall Y \in \mathcal{X}(S)$. The obtained connection is called the g-*projection* of ∇ onto M.

3 Geometrical structures on a manifold of quantum states

In this section we show that, when a generalized covariance is specified, natural analogues of the Fisher metric and the e,m-connections are defined on a manifold of quantum states.

Let $\mathcal{H}$ be a finite-dimensional Hilbert space, $\mathcal{A} = \{A \mid A = A^*\}$ be the totality of Hermitian operators on $\mathcal{H}$ and $S = \{\rho \mid \rho = \rho^* > 0, \mathrm{Tr}\, \rho = 1\}$ be the totality of positive density operators on $\mathcal{H}$. Since S is an open subset of $\mathcal{A}_1 \stackrel{\mathrm{def}}{=} \{A \mid A \in$

$\mathcal{A}, \mathrm{Tr}\ A = 1\}$, S is naturally regarded as a smooth manifold of dimension $n \stackrel{\mathrm{def}}{=} \dim \mathcal{A}_1 = (\dim \mathcal{H})^2 - 1$.

Let ι be the immersion $S \to \mathcal{A}_1 \subset \mathcal{A}$. Then a linear isomorphism $T_\rho(S) \to \mathcal{A}_0 \stackrel{\mathrm{def}}{=} \{A \mid A \in \mathcal{A}, \mathrm{Tr}\ A = 0\}$ is established by $D \mapsto \iota_*(D)$. We write $\iota_*(D) = D^{(\mathrm{m})}$ and call it the *m-representation* of a tangent vector D. Since $D^{(\mathrm{m})}$ is the derivative of $\iota : \rho \mapsto \rho$ by D, it may be written symbolically as $D^{(\mathrm{m})} = D\rho$. The isomorphism $D \mapsto D^{(\mathrm{m})}$ provides S with an affine connection, which we call the *(-1)-connection* or the *mixture connection*. This connection is represented by the covariant derivative $\nabla^{(\mathrm{m})} : \mathcal{X}(S) \times \mathcal{X}(S) \to \mathcal{X}(S)$ such that, for $\forall X, \forall Y \in \mathcal{X}(S)$

$$(\nabla_X^{(\mathrm{m})} Y)^{(\mathrm{m})} = X(Y^{(\mathrm{m})}),$$

where the RHS (right hand side) means the derivative by X of $Y^{(\mathrm{m})} : S \to \mathcal{A}_0$ ($\rho \mapsto Y_\rho^{(\mathrm{m})}$). For an arbitrary affine isomorphism $\omega : \mathcal{A}_1 \to \boldsymbol{R}^n$, the composition $\omega \circ \iota$ forms a $\nabla^{(\mathrm{m})}$-affine coordinate system of S, and hence $\nabla^{(\mathrm{m})}$ is flat. Such a coordinate system (ξ^i) is represented as $\xi^i(\rho) = \mathrm{Tr}\ (\rho F^i)$ by a set of operators $\{F^1, \ldots, F^n\}$ such that $\{1, F^1, \ldots, F^n\}$ forms a basis of $\mathcal{A}$, where 1 denotes the identity operator. The parallel displacement w.r.t. $\nabla^{(\mathrm{m})}$ of $D \in T_\rho(S)$ to $D' \in T_{\rho'}(S)$ is determined by $D^{(\mathrm{m})} = D'^{(\mathrm{m})}$.

Suppose that we are given a family $\{\langle \cdot, \cdot \rangle_\rho \mid \rho \in S\}$ of inner products on $\mathcal{A}$, where $\langle A, B \rangle_\rho \in \mathbf{R}$ is specified smoothly by ρ for $\forall A, \forall B \in \mathcal{A}$. We further assume that

$$\langle A, 1 \rangle_\rho = \langle A \rangle_\rho \stackrel{\mathrm{def}}{=} \mathrm{Tr}\ (\rho A) \tag{4}$$

for $\forall A \in \mathcal{A}$. Such a family $\{\langle \cdot, \cdot \rangle_\rho\}$ is called a *generalized covariance*. Two important examples are the *symmetrized inner product*

$$\langle A, B \rangle_\rho = \frac{1}{2} \mathrm{Tr}\ (\rho AB + \rho BA) \tag{5}$$

and the *Bogoliubov inner product* (also called the *Kubo-Mori inner product* or the *canonical correlation*)

$$\langle A, B \rangle_\rho = \int_0^1 \mathrm{Tr}\,(\rho^\lambda A \rho^{1-\lambda} B) d\lambda. \tag{6}$$

These examples are unified in the general form ([11])

$$\langle A, B \rangle_\rho = \int_0^1 \mathrm{Tr}\,(\rho^\lambda A \rho^{1-\lambda} B) \nu(d\lambda), \tag{7}$$

where ν is an arbitrary probability measure on $[0, 1]$ satisfying $\nu(d\lambda) = \nu(1 - d\lambda)$.

For $\forall \rho \in S$, we define the mapping $\Phi_\rho : \mathcal{A} \to \mathcal{A}$ by

$$\langle A, B \rangle_\rho = \mathrm{Tr}\,(A \Phi_\rho(B)), \quad \forall A, \forall B \in \mathcal{A}. \tag{8}$$

Since Φ_ρ is a linear isomorphism on $\mathcal{A}$, we can define the *e-representation* $D^{(\mathrm{e})} \in \mathcal{A}$ of given $D \in T_\rho(S)$ by

$$D^{(\mathrm{m})} = D\rho = \Phi_\rho(D^{(\mathrm{e})}). \tag{9}$$

Note that, for $\forall A \in \mathcal{A}$, the derivative of the function $\langle A \rangle : \rho \mapsto \langle A \rangle_\rho$ by D is written as

$$D\langle A \rangle = \mathrm{Tr}\,(D^{(\mathrm{m})} A) = \langle D^{(\mathrm{e})}, A \rangle_\rho. \tag{10}$$

For the symmetrized product, the equation (9) turns out

$$D\rho = \frac{1}{2}(\rho D^{(\mathrm{e})} + D^{(\mathrm{e})} \rho), \tag{11}$$

which means that $D^{(\mathrm{e})}$ is the SLD (symmetric logarithmic derivative) in the direction D. On the other hand, for the Bogoliubov product we have

$$D\rho = \int_0^1 \rho^\lambda D^{(\mathrm{e})} \rho^{1-\lambda} d\lambda. \tag{12}$$

By a standard calculation in the operator calculus, $D^{(\mathrm{e})}$ in this case is shown to turn out the derivative of the mapping $\rho \mapsto \log \rho$ from S to $\mathcal{A}$, which may be written as

$$D^{(\mathrm{e})} = D \log \rho. \tag{13}$$

Although the e-representations depend on the choice of generalized covariance, the range $T_\rho^{(\mathrm{e})}(S) \stackrel{\mathrm{def}}{=} \{D^{(\mathrm{e})} \mid D \in T_\rho(S)\}$ is simply written as

$$T_\rho^{(\mathrm{e})}(S) = \{A \mid A \in \mathcal{A},\ \langle A \rangle_\rho = 0\}. \tag{14}$$

This is verified as follows. For $\forall D \in T_\rho(S)$, we have $\langle D^{(\mathrm{e})} \rangle_\rho = \langle D^{(\mathrm{e})}, 1 \rangle_\rho = D \langle 1 \rangle = 0$ from (4), (10) and $\langle 1 \rangle \equiv 1$. This means that LHS $\subset$ RHS in (14). Since LHS and RHS are both of dimension n, we have (14).

From this fact, we can define a linear isomorphism $D \mapsto D'$ from $T_\rho(S)$ to $T_{\rho'}(S)$ by $D'^{(\mathrm{e})} = D^{(\mathrm{e})} - \langle D^{(\mathrm{e})} \rangle_{\rho'}$. We write this correspondence as $D' = [D]_{\rho'}$, $D = [D']_\rho$. Now we define the *exponential connection* or the *1-connection* $\nabla^{(\mathrm{e})}$ by

$$(\nabla_X^{(\mathrm{e})} Y)_\rho = X_\rho [Y]_\rho$$

for $\forall \rho \in S, \forall X, \forall Y \in \mathcal{X}(S)$, where the RHS means the derivative by X_ρ of $[Y]_\rho :$ $S \to T_\rho(S),\ \sigma \mapsto [Y_\sigma]_\rho$. The correspondence $D \mapsto D' = [D]_{\rho'}$ can be regarded as the parallel displacement w.r.t. $\nabla^{(\mathrm{e})}$, and hence the curvature tensor $\mathcal{R}^{(\mathrm{e})}$ of $\nabla^{(\mathrm{e})}$ vanishes. Note that $\nabla^{(\mathrm{e})}$ is not necessarily flat because the torsion tensor $\mathcal{T}^{(\mathrm{e})}$ does not necessarily vanish.

Define the inner product g_ρ on $T_\rho(S)$ by

$$g_\rho(D_1, D_2) \stackrel{\mathrm{def}}{=} \langle D_1^{(\mathrm{e})}, D_2^{(\mathrm{e})} \rangle_\rho = \mathrm{Tr}\,(D_1^{(\mathrm{m})} D_2^{(\mathrm{e})}). \tag{15}$$

Then $g = \{g_\rho \mid \rho \in S\}$ forms a Riemannian metric on S, which we call the *Fisher metric*.

Suppose that a coordinate system (ξ^i) of S is given and that each element ρ of S is specified by the coordinate $\xi \in \boldsymbol{R}^n$ as $\rho = \rho_\xi$. The mixture representation $\partial_i^{(\mathrm{m})}$ of $\partial_i = \partial/\partial\xi^i$ is written as $\partial_i\rho = \partial_i\rho_\xi$. Denoting the e-representation $\partial_i^{(\mathrm{e})}$ by L_i, we have

$$\begin{aligned} g_{ij} &\stackrel{\mathrm{def}}{=} g(\partial_i, \partial_j) = \langle L_i, L_j \rangle = \mathrm{Tr}\,(\partial_i\rho \cdot L_j) \\ \Gamma^{(\mathrm{m})}_{ij,k} &\stackrel{\mathrm{def}}{=} g(\nabla^{(\mathrm{m})}_{\partial_i}\partial_j, \partial_k) = \mathrm{Tr}\,(\partial_i\partial_j\rho \cdot L_k) \\ \Gamma^{(\mathrm{e})}_{ij,k} &\stackrel{\mathrm{def}}{=} g(\nabla^{(\mathrm{e})}_{\partial_i}\partial_j, \partial_k) = \langle \partial_i L_j, L_k \rangle = \mathrm{Tr}\,(\partial_i L_j \cdot \partial_k\rho), \end{aligned} \tag{16}$$

which may be more tractable than the original definitions. For the Bogoliubov product, in particular, we have from (13)

$$g_{ij} = \mathrm{Tr}\,(\partial_i\rho \cdot \partial_j \log\rho) \tag{17}$$

$$\Gamma^{(\mathrm{m})}_{ij,k} = \mathrm{Tr}\,(\partial_i\partial_j\rho \cdot \partial_k \log\rho), \quad \Gamma^{(\mathrm{e})}_{ij,k} = \mathrm{Tr}\,(\partial_i\partial_j \log\rho \cdot \partial_k\rho). \tag{18}$$

Let M be a submanifold of S. Then the Fisher metric and the e,m-connections on M are defined by taking the restriction and the g-projection of $(g, \nabla^{(\mathrm{m})}, \nabla^{(\mathrm{e})})$ onto M. When a coordinate system of M is given, these quantities are represented in the same form as (16).

4 Duality of e,m-connections

In general, if a Riemannian metric g and two affine connections ∇ and ∇^* on a manifold M satisfy

$$Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla^*_X Z) \tag{19}$$

for $\forall X, \forall Y, \forall Z \in \mathcal{X}(M)$, or equivalently if $\partial_i g_{jk} = \Gamma_{ij,k} + \Gamma^*_{ik,j}$ for $\forall i, \forall j, \forall k$, we say that ∇ and ∇^* are mutually *dual* (or *conjugate*) w.r.t. g ([9] [2] [10]), and (g, ∇, ∇^*)

is called a *dualistic triplet* . We can easily see from (16) that the e-connection and the m-connection are dual w.r.t. the Fisher metric on an arbitrary submanifold M of S (including the case $M = S$). The geometrical meaning of the duality is that, if $t \mapsto X_t$ and $t \mapsto Y_t$ are vector fields defined along a curve $t \mapsto \rho_t$ in M and if X and Y are ∇-parallel and ∇^*-parallel, respectively, then $g(X_t, Y_t)$ is constant on the curve.

Suppose that we are given a two-variable smooth function $K : M \times M \to \boldsymbol{R}$ satisfying $K(\rho, \sigma) > 0$ and $K(\rho, \rho) = 0$ for $\forall \rho \neq \forall \sigma$. For an arbitrary coordinate system (ξ^i), let

$$g_{ij} = -K(\partial_i | \partial_j) = K(\partial_i \partial_j | \cdot) = K(\cdot | \partial_i \partial_j) \tag{20}$$

$$\Gamma_{ij,k} = -K(\partial_i \partial_j | \partial_k), \quad \Gamma^*_{ij,k} = -K(\partial_k | \partial_i \partial_j), \tag{21}$$

where we have used the Eguchi's notation ([5]) such as

$$K(\partial_i | \partial_j) : \rho \mapsto \frac{\partial}{\partial \xi^i} \frac{\partial}{\partial \xi'^j} K(\rho, \rho')\big|_{\rho = \rho'}, \quad K(\partial_i \partial_j | \cdot) : \rho \mapsto \frac{\partial^2}{\partial \xi^i \partial \xi^j} K(\rho, \rho')\big|_{\rho = \rho'},$$

etc. If $[g_{ij}]$ forms a positive definite matrix, these quantities define a dualistic triplet (g, ∇, ∇^*) by $g_{ij} = g(\partial_i, \partial_j)$, $\Gamma_{ij,k} = g(\nabla_{\partial_i} \partial_j, \partial_k)$, $\Gamma^*_{ij,k} = g(\nabla^*_{\partial_i} \partial_j, \partial_k)$, being independent of the choice of (ξ^i) ([4] [5]). This is an important origin of the duality. For instance, the equations in (17) (18) for the Bogoliubov triplet $(g, \nabla^{(\mathrm{m})}, \nabla^{(\mathrm{e})})$ is written as (20) (21) if K is chosen to be the relative entropy (1). It should be noted, however, that a triplet (g, ∇, ∇^*) is obtained in this manner only when both ∇ and ∇^* are torsion free, because (21) implies $\Gamma_{ij,k} = \Gamma_{ji,k}$ and $\Gamma^*_{ij,k} = \Gamma^*_{ji,k}$. (Conversely, it can be shown that an arbitrary torsion free dualistic triplet is obtained in this manner.) Note also that such a function K is not uniquely determined by given (g, ∇, ∇^*).

5 Divergence of a dually flat space

First, we give a brief review of the results of Nagaoka and Amari [9] [2]. Suppose that ∇ and ∇^* are dual w.r.t. g on M, and let $\mathcal{R}, \mathcal{R}^*$ be the curvature tensors and $\mathcal{T}, \mathcal{T}^*$ be the torsion tensors of ∇, ∇^*, respectively. Then it generally holds that $\mathcal{R} = 0$ iff $\mathcal{R}^* = 0$, while the corresponding property does not hold for $\mathcal{T}, \mathcal{T}^*$. If (M, g, ∇, ∇^*) is *dually flat* in the sense that both ∇ and ∇^* are flat (i.e., $\mathcal{R} = \mathcal{R}^* = 0$ and $\mathcal{T} = \mathcal{T}^* = 0$), there exist a pair of coordinate systems $((\xi^i), (\zeta_i))$ of M and a pair of functions (φ, ψ) on S such that (ξ^i) and (ζ_i) are ∇-affine and ∇^*-affine, respectively, and that

$$\zeta_i = \partial_i \varphi, \quad \xi^i = \partial^i \psi, \quad \varphi + \psi = \sum_i \xi^i \zeta_i, \tag{22}$$

$$g(\partial_i, \partial^j) = \delta_i^{\,j}, \quad g_{ij} \stackrel{\text{def}}{=} g(\partial_i, \partial_j) = \partial_i \partial_j \varphi, \quad g^{ij} \stackrel{\text{def}}{=} g(\partial^i, \partial^j) = \partial^i \partial^j \psi, \tag{23}$$

where $\partial_i \stackrel{\text{def}}{=} \partial/\partial\xi^i$, $\partial^i \stackrel{\text{def}}{=} \partial/\partial\zeta_i$. These quantities have some degrees of freedom, but the following function K defined on $M \times M$ is uniquely determined by (M, g, ∇, ∇^*):

$$K(\rho, \sigma) = \varphi(\rho) + \psi(\sigma) - \sum_i \xi^i(\rho)\zeta_i(\sigma). \tag{24}$$

Indeed, this function is characterized by the condition that $K(\rho, \sigma) > 0$ and $K(\rho, \rho) = 0$ for $\forall\rho \neq \forall\sigma$ together with one of the following mutually equivalent properties:

$$\frac{\partial^2}{\partial\xi_1^i \partial\xi_1^j} K(\rho_1, \rho_2) = g_{ij}(\rho_1), \quad \forall\rho_1, \forall\rho_2 \in M$$
$$\frac{\partial^2}{\partial\zeta_{2i} \partial\zeta_{2j}} K(\rho_1, \rho_2) = g^{ij}(\rho_2), \quad \forall\rho_1, \forall\rho_2 \in M$$

where (ξ^i) ((ζ_i), resp.) is an arbitrary ∇-affine (∇^*-affine, resp.) coordinate system. We call K the *divergence* of the dually flat space (M, g, ∇, ∇^*).

Clearly, the above properties characterizing the divergence K imply (20) (21) and is much stronger than (20) (21). Indeed, the divergence K has the following generalized Pythagorean property. Let ρ_1, ρ_2, ρ_3 be arbitrary points in M, γ_{12} be the ∇-geodesic connecting ρ_1 and ρ_2, and γ^*_{23} be the ∇^*-geodesic connecting ρ_2 and ρ_3. If γ_{12} and γ^*_{23} intersect orthogonally (w.r.t. g) at ρ_2, then we have $K(\rho_1, \rho_3) = K(\rho_1, \rho_2) + K(\rho_2, \rho_3)$.

Let us return to the triplet $(g, \nabla^{(\mathrm{m})}, \nabla^{(\mathrm{e})})$ on S.

Theorem 1 *For the Bogoliubov product, $(S, g, \nabla^{(\mathrm{m})}, \nabla^{(\mathrm{e})})$ is dually flat, and its divergence is the relative entropy (1).*

Proof We have already seen that $\nabla^{(\mathrm{m})}$ is flat and $\mathcal{R}^{(\mathrm{e})} = 0$ on S for an arbitrary generalized covariance. For the Bogoliubov product, $\mathcal{T}^{(\mathrm{e})}$ vanishes as is seen from (18), and $\nabla^{(\mathrm{e})}$ is also flat.

Let $\{F^1, \ldots, F^n\}$ be an arbitrary set of operators such that $\{1, F^1, \ldots, F^n\}$ forms a basis of $\mathcal{A}$, and let $\xi^i(\rho) \stackrel{\mathrm{def}}{=} \mathrm{Tr}\,(\rho F^i) = \langle F^i \rangle_\rho$. Then (ξ^i) becomes a $\nabla^{(\mathrm{m})}$-affine coordinate system of S as mentioned in chap.3. Since $\log \rho$ belongs to $\mathcal{A}$ for $\forall \rho \in S$, it is written as $\log \rho = \sum_i \zeta_i F^i - \psi$ by $(\zeta_1, \ldots, \zeta_n, \psi) \in \boldsymbol{R}^{n+1}$. Since $\dim S = n$, we can choose $(\zeta_1, \ldots, \zeta_n)$ as a coordinate system of S and regard ψ as a function on S. Thus we have

$$\rho = \exp\left[\sum_{i=1}^{n} \zeta_i F^i - \psi(\rho)\right]. \tag{25}$$

Owing to (13), the e-representation of $\partial^i = \partial/\partial\zeta_i$ is written as $\partial^i \log \rho = F^i - \partial^i \psi$. Recalling (14), we have $\xi^i = \partial^i \psi$ in (22) and $(\partial^i)^{(\mathrm{e})} = F^i - \langle F^i \rangle$, the latter of which means that ∂^i is $\nabla^{(\mathrm{e})}$-parallel. Hence (ζ_i) is $\nabla^{(\mathrm{e})}$-affine. Next, let $\varphi(\rho) \stackrel{\mathrm{def}}{=} \mathrm{Tr}\,(\rho \log \rho) = -H(\rho)$, where $H(\rho)$ is the von Neumann entropy. Then the equation

$\varphi + \psi = \sum_i \xi^i \zeta_i$ in (22) follows from (25) and $\xi^i = \partial^i \psi$. Furthermore, using (17) we have

$$g(\partial_i, \partial^j) = \mathrm{Tr}\,(\partial_i \rho \cdot \partial^j \log \rho) = \mathrm{Tr}\,(\partial_i \rho \cdot (F^j - \xi^j)) = \mathrm{Tr}\,(\partial_i \rho \cdot F^j) = \partial_i \xi^j = \delta_i^{\ j}.$$

The remaining equations in (22) (23) are derived by simple calculations from the equations obtained so far. Finally, it is easily shown that (24) is reduced to (1) in this case.

(QED)

6 Torsion of e-connection

Suppose that an element A of $\mathcal{A}$ is arbitrarily fixed, and consider the mapping $\Phi_{\cdot}(A) : \rho \mapsto \Phi_\rho(A)$ from S to $\mathcal{A}$. Since $\mathcal{A}$ is a linear space, the differential of this mapping at a point ρ is regarded as a linear mapping from $T_\rho(S)$ to $\mathcal{A}$. We denote by $\Phi_D(A)$ $(\in \mathcal{A})$ the value (derivative) of the differential for $D \in T_\rho(S)$.

The following theorem gives a representation of the torsion tensor $\mathcal{T}^{(\mathrm{e})}$ of the e-connection on S for an arbitrary generalized covariance. Note that, although $\mathcal{T}^{(\mathrm{e})}$ was introduced in (3) as a mapping : $\mathcal{X}(S) \times \mathcal{X}(S) \to \mathcal{X}(S)$, it also induces a bilinear mapping : $T_\rho(S) \times T_\rho(S) \to T_\rho(S)$ at each point $\rho \in S$.

Theorem 2 *For $\forall D_1, \forall D_2 \in T_\rho(S)$, the m-representation of $\mathcal{T}^{(\mathrm{e})}(D_1, D_2) \in T_\rho(S)$ is given by*

$$\{\mathcal{T}^{(\mathrm{e})}(D_1, D_2)\}^{(\mathrm{m})} = \Phi_{D_2}(D_1^{(\mathrm{e})}) - \Phi_{D_1}(D_2^{(\mathrm{e})}).$$

Proof For given $D_1, D_2 \in T_\rho(S)$, define $X_1, X_2 \in \mathcal{X}(S)$ by

$$(X_i)_\sigma^{(\mathrm{e})} = D_i^{(\mathrm{e})} - \langle D_i^{(\mathrm{e})} \rangle_\sigma \ \in T_\sigma^{(\mathrm{e})}(S), \quad \forall \sigma \in S.$$

Since these vector fields are $\nabla^{(\mathrm{e})}$-parallel on S, we have $\nabla^{(\mathrm{e})}_Y X_i = 0$ for $\forall Y \in \mathcal{X}(S)$. Applying this to the definition of the torsion tensor (3), we obtain $\mathcal{T}^{(\mathrm{e})}(X_1, X_2) = X_2X_1 - X_1X_2$, which is evaluated at ρ as $\mathcal{T}^{(\mathrm{e})}(D_1, D_2) = D_2X_1 - D_1X_2$. This implies that

$$\{\mathcal{T}^{(\mathrm{e})}(D_1, D_2)\}^{(\mathrm{m})} = D_2X_1^{(\mathrm{m})} - D_1X_2^{(\mathrm{m})}, \tag{26}$$

where $D_jX_i^{(\mathrm{m})}$ is the derivative of $X_i^{(\mathrm{m})} : \sigma \mapsto (X_i)_\sigma^{(\mathrm{m})}$ by D_j. On the other hand, from (9) we have

$$(X_i)_\sigma^{(\mathrm{m})} = \Phi_\sigma((X_i)_\sigma^{(\mathrm{e})}) = \Phi_\sigma(D_i^{(\mathrm{e})}) - \langle D_i^{(\mathrm{e})}\rangle_\sigma \Phi_\sigma(1).$$

Since $\Phi_\rho(1) = \rho$ from the assumption (4), we obtain

$$\begin{aligned} D_jX_i^{(\mathrm{m})} &= \Phi_{D_j}(D_i^{(\mathrm{e})}) - (D_j\langle D_i^{(\mathrm{e})}\rangle)\rho - \langle D_i^{(\mathrm{e})}\rangle_\rho D_j^{(\mathrm{m})} \\ &= \Phi_{D_j}(D_i^{(\mathrm{e})}) - g(D_i, D_j)\rho, \end{aligned} \tag{27}$$

where the second equality is derived from (10) (15) and (14). The proof is completed by (26) and (27).

(QED)

Corollary 1 *For the symmetrized product, we have*

$$\{\mathcal{T}^{(\mathrm{e})}(D_1, D_2)\}^{(\mathrm{m})} = \frac{1}{4}[[D_1^{(\mathrm{e})}, D_2^{(\mathrm{e})}], \rho],$$

where $[\cdot, \cdot]$ *is the commutator for operators.*

Proof From $\Phi_\rho(A) = \frac{1}{2}(\rho A + A\rho)$ and (11), we have

$$\Phi_{D_j}(D_i^{(\mathrm{e})}) = \frac{1}{4}(D_i^{(\mathrm{e})}D_j^{(\mathrm{e})}\rho + D_i^{(\mathrm{e})}\rho D_j^{(\mathrm{e})} + D_j^{(\mathrm{e})}\rho D_i^{(\mathrm{e})} + \rho D_j^{(\mathrm{e})}D_i^{(\mathrm{e})}),$$

which leads to the desired equation.

(QED)

This corollary means that the e-connection for the symmetrized product is not torsion free, and in this case we cannot expect the existence of the divergence such as the relative entropy or even the existence of a function $K : S \times S \to \boldsymbol{R}$ for which (20) and (21) hold.

7 Cramér-Rao inequality and e-autoparallel submanifolds

let M be an arbitrary submanifold of S of dimension m $(\leq n)$ and $(\xi^i) = (\xi^1, \ldots, \xi^m)$ be an arbitrary coordinate system of M. For an arbitrary operator $F \in \mathcal{A}$, denote the restriction of $\langle F \rangle : S \to \boldsymbol{R}$ on M by $f \stackrel{\text{def}}{=} \langle F \rangle|_M : M \to \mathcal{R}$. Its differential $(df)_\rho$ at $\rho \in M$ belongs to the cotangent space $T^*_\rho(M)$, on which an inner product is induced from g_ρ , and so the norm $\| (df)_\rho \|_\rho$ is defined. Letting $G^{-1} = [g^{ij}]$ be the inverse of the $m \times m$ matrix $G = [g_{ij}] = [g(\partial_i, \partial_j)]$ $(\partial_i \stackrel{\text{def}}{=} \partial/\partial\xi^i)$, it is written as $\| (df)_\rho \|^2_\rho = \sum_{i,j} (g^{ij} \partial_i f \partial_j f)_\rho$.

Lemma 1 *For $\forall\rho \in M$ and $\forall c \in \boldsymbol{R}$, we have*

$$\langle F - c,\, F - c \rangle_\rho \ \geq \| (df)_\rho \|^2_\rho ,$$

where the equality holds iff $F - c \in T^{(\mathrm{e})}_\rho(M) \stackrel{\text{def}}{=} \{D^{(\mathrm{e})} \mid D \in T_\rho(M)\}$ $(\subset T^{(\mathrm{e})}_\rho(S))$.

Proof Let D be the vector in $T_\rho(M)$ whose e-representation $D^{(\mathrm{e})}$ is the orthogonal projection of $F - c$ onto $T^{(\mathrm{e})}_\rho(M)$ w.r.t. $\langle \cdot, \cdot \rangle_\rho$. Then, for $\forall D' \in T_\rho(M)$

$$g_\rho(D, D') = \langle D^{(\mathrm{e})}, D'^{(\mathrm{e})} \rangle_\rho = \langle F - c, D'^{(\mathrm{e})} \rangle_\rho \stackrel{(*)}{=} \langle F, D'^{(\mathrm{e})} \rangle_\rho \stackrel{(**)}{=} D' \langle F \rangle = (df)_\rho(D'),$$

where the equalities $(*)$ and $(**)$ follow from (14) and (10), respectively. The above equation means that D is the gradient vector of f at ρ, and we have $\| (df)_\rho \|^2_\rho = g_\rho(D, D) = \langle D^{(\mathrm{e})}, D^{(\mathrm{e})} \rangle_\rho$. This proves the lemma, because $D^{(\mathrm{e})}$ is the orthogonal projection of $F - c$.

(QED)

The following theorem ([11]) is a generalization of the quantum Cramér-Rao inequality based on the symmetric logarithmic derivatives (Helstrom, 1967), which corresponds to the case of the symmetrized inner product. For other generalized covariances, however, the meaning of the theorem is not clear in view of the parameter estimation. Note also that we do not take into consideration the simultaneous estimation problem of the multi-dimensional parameter (ξ^i). Since the theorem is a direct consequence of the previous lemma, we omit the proof.

Theorem 3 *If an m-tuple $F = (F^1, \ldots, F^m) \in \mathcal{A}^m$ of operators satisfies the unbiasedness condition*

$$\langle F^i \rangle_\rho = \xi^i(\rho), \qquad \forall \rho \in M, \forall i, \tag{28}$$

then the matrix $V(\rho) - G(\rho)^{-1}$ is nonnegative definite, where $V(\rho) = [v^{ij}(\rho)]$ is the $m \times m$ matrix defined by $v^{ij}(\rho) = \langle F^i - \xi^i(\rho), F^j - \xi^j(\rho) \rangle_\rho$.

Next, we show the theorem which describes the condition for the existence of $(F^1, \ldots, F^m)$ achieving the bound in the above theorem. This corresponds to the classical result giving the condition for the existence of the efficient estimator. In the sequel, we write the m,e-connections on M as $\nabla_M^{(\mathrm{m})}$, $\nabla_M^{(\mathrm{e})}$ to distinguish them from the m,e-connections $\nabla^{(\mathrm{m})}$, $\nabla^{(\mathrm{e})}$ on S.

Theorem 4 *For given $(M, (\xi^i))$, there exists an m-tuple $(F^1, \ldots, F^m) \in \mathcal{A}^m$ satisfying the unbiasedness condition (28) and $V(\rho) = G(\rho)^{-1}$ for $\forall \in M$, if and only if M is $\nabla^{(\mathrm{e})}$-autoparallel in S and (ξ^i) is a $\nabla_M^{(\mathrm{m})}$-affine coordinate system of M.*

Proof It is clear from Lemma 1 that the necessary and sufficient condition for $(F^1, \ldots, F^m) \in \mathcal{A}^m$ to satisfy (28) and $V(\rho) = G(\rho)^{-1}$ $(\forall \rho \in M)$ is that

$$F^i - \xi^i(\rho) \in T^{(\mathrm{e})}_\rho(M) \quad \text{for} \quad \forall \rho \in M, \forall i. \tag{29}$$

Assume this, and define $X^i \in \mathcal{X}(M)$ by $(X^i_\rho)^{(\mathrm{e})} = F^i - \xi^i(\rho)$ $(\forall \rho \in M)$ for $i = 1, \ldots, m$. Then the vector fields $\{X^1, \ldots, X^m\}$ are all $\nabla^{(\mathrm{e})}$-parallel on S and are linearly independent, which implies that M is $\nabla^{(\mathrm{e})}$-autoparallel in S. Furthermore, since $\{X^1, \ldots, X^m\}$ are also $\nabla^{(\mathrm{e})}_M$-parallel, and since

$$g(X^j, \partial_i) = \langle F^j - \xi^j, \partial^{(\mathrm{e})}_i \rangle = \langle F^j, \partial^{(\mathrm{e})}_i \rangle = \partial_i \langle F^j \rangle = \partial_i \xi^j = \delta^j_i,$$

it is concluded from the duality between $\nabla^{(\mathrm{e})}_M$ and $\nabla^{(\mathrm{m})}_M$ that $\{\partial_1, \ldots, \partial_m\}$ are all $\nabla^{(\mathrm{m})}_M$-parallel. Hence (ξ^i) is $\nabla^{(\mathrm{m})}_M$-affine.

Conversely, assume that M is $\nabla^{(\mathrm{e})}$-autoparallel and that (ξ^i) is $\nabla^{(\mathrm{m})}_M$-affine. For each $j = 1, \ldots, m$, define $X^j \in \mathcal{X}(M)$ by the condition $g(X^j, \partial_i) = \delta_i{}^j, \forall i$. Since $\{\partial_1, \ldots, \partial_m\}$ are all $\nabla^{(\mathrm{m})}_M$-parallel, $\{X^1, \ldots, X^m\}$ turn out to be $\nabla^{(\mathrm{e})}_M$-parallel due to the duality. Furthermore, $\{X^1, \ldots, X^m\}$ are $\nabla^{(\mathrm{e})}$-parallel because M is $\nabla^{(\mathrm{e})}$-autoparallel in S, and therefore there exists $(F^1, \ldots, F^m) \in \mathcal{A}^m$ such that $(X^j)^{(\mathrm{e})} = F^j - \langle F^j \rangle$. It is easily shown that $\partial_i \langle F^j \rangle = g(\partial_i, X^j) = \delta_i{}^j = \partial_i \xi^j$, and hence we can choose $(F^1, \ldots, F^m)$ such that $\langle F^j \rangle = \xi^j$ $(\forall j)$, which satisfies (29).

(QED)

Suppose that M is $\nabla^{(\mathrm{e})}$-autoparallel in S. Then $\nabla^{(\mathrm{e})}_M$ inherits the vanishing curvature from $\nabla^{(\mathrm{e})}$, and by the duality, $\nabla^{(\mathrm{m})}_M$ also turns out curvature free. Since the m-connection is always torsion free, we see that $\nabla^{(\mathrm{m})}_M$ is flat. Thus every $\nabla^{(\mathrm{e})}$-autoparallel submanifold has a $\nabla^{(\mathrm{m})}_M$-affine coordinate system. On the other hand,

M does not have a $\nabla_M^{(\mathrm{e})}$-affine coordinate system in general, unless the torsion $\mathcal{T}^{(\mathrm{e})}$ of $\nabla^{(\mathrm{e})}$ vanishes.

For the Bogoliubov product, M always has a $\nabla_M^{(\mathrm{e})}$-affine coordinate system (ζ_i). In this case, M is shown to be written in the form

$$\rho = \exp\left[F^0 + \sum_{i=1}^{m} \zeta_i F^i - \psi(\zeta)\right], \tag{30}$$

which includes (25) for S as a special case, and that the $\nabla^{(\mathrm{e})}$-autoparallelness is characterized by the form (30). We can see that there are close analogies between the structure of this form and that of exponential families in the classical case ([1][2]).

For the symmetrized product, M does not generally have a $\nabla_M^{(\mathrm{e})}$-affine coordinate system or such a representation as (30) characterizing the $\nabla^{(\mathrm{e})}$-autoparallelness. In particular, S itself is a $\nabla^{(\mathrm{e})}$-autoparallel submanifold of S with no $\nabla^{(\mathrm{e})}$-affine coordinate system. However, there are $\nabla^{(\mathrm{e})}$-autoparallel submanifolds having vanishing torsion, although the torsion of S does not vanish. For instance, since every 1-dimensional manifold is torsion free, an arbitrary e-geodesic M has a $\nabla_M^{(\mathrm{e})}$-affine coordinate system ζ. We can show that M is generally written in the form:

$$\rho = \exp\left[\{\zeta F - \psi(\zeta)\}/2\right]\,\rho_0\,\exp\left[\{\zeta F - \psi(\zeta)\}/2\right]. \tag{31}$$

It is noted that $(M, g, \nabla_M^{(\mathrm{m})}, \nabla_M^{(\mathrm{e})})$ is dually flat and has the divergence. For arbitrary points $\rho, \sigma \in S$, let K be the divergence of the e-geodesic connecting these points. Then we have

$$K(\rho, \sigma) = \mathrm{Tr}\,[\rho \log A],$$

where A is the positive operator in $\mathcal{A}$ satisfying $A^{1/2}\sigma A^{1/2} = \rho$, or is explicitly written as

$$A = \left\{\sigma^{-\frac{1}{2}}(\sigma^{\frac{1}{2}}\rho\sigma^{\frac{1}{2}})^{\frac{1}{2}}\sigma^{-\frac{1}{2}}\right\}^2.$$

This K is another quantum analogue of the Kullback divergence.

References

[1] Amari, S.: "Differential geometry of curved exponential families – curvatures and information loss", Ann. Statist., 10, pp.357-387 (1982).

[2] Amari, S.: "Differential-geometrical methods in statistics", Lecture Notes in Statistics 28, Springer-Verlag (1985).

[3] Chentsov, N. N.: "Statistical Decision Rules and Optimal Inference" (in Russian), Nauka Moscow (1972); translated in English, AMS, Rhode Island (1982).

[4] Eguchi, S.: "Second order efficiency of minimum contrast estimators in a curved exponential family", Ann. Statist., 11, pp.793-803 (1983).

[5] Eguchi, S.: "Geometry of minimum contrast", Hiroshima Mathematical Journal, 22, pp.631-647 (1992).

[6] Helstrom, C. W.: "Quantum Detection and Estimation Theory," Academic Press, New York (1976).

[7] Holevo, A. S.: "Probabilistic and Statistical Aspects of Quantum Theory," North-Holland (1982).

[8] Kobayashi, S., Nomizu, K.: "Foundations of differential geometry, vol.I, II", John Wiley & Sons (1963, 1969).

[9] Nagaoka, H., Amari, S.: "Differential geometry of smooth families of probability distributions", METR 82-7, Univ. Tokyo. (1982).

[10] Nomizu, K., Simon, U.: "Notes on conjugate connections", Geometry and topology of submanifolds, IV, World Scientific (1992).

[11] Petz, D., Toth, G.: "The Bogoliubov inner product in quantum statistics", Lett. Math. Phys., 27, pp.205-216 (1993).

[12] Rao, C. R.: "Information and accuracy attainable in the estimation of statistical parameters", Bull. Calcutta. Math. Soc., 37, pp.81-91 (1945).

[13] Yuen, H. P., and Lax, M.: "Multiple-parameter Quantum Estimation and Measurement of Nonselfadjoint Observables," Trans. IEEE, IT-19 , pp.740-750 (1973).