数理解析研究所講究録
923 巻 1995 年 52-67

EXPONENTIAL AND MIXTURE FAMILIES IN QUANTUM STATISTICS
— dual structure and unbiased parameter estimation

Hiroshi HASEGAWA
Department of Electronics Engineering, Fukui University, Fukui 910, Japan

A problem of information geometry (differential-geometrical approach in statistics), namely to construct an *exponential family* and a *mixture family* on a smooth manifold in the parameter space Θ of states (density operators) for finite quantum systems, is discussed: In $\mathcal{B}(H^N)$ (N by N matrix algebra), our statistical model is $\mathcal{S} = \{\rho \in \mathcal{B}^{++}(H^N) =$ all strictly positive hermitian matrices, with $\mathrm{Tr}\rho = 1\}$, and we investigate two families of $\rho(\theta)$ defined by

$$\text{(e)} \qquad \rho(\theta) = \exp(\theta^i A_i - \psi(\theta)) \quad \theta \in \Theta = \mathrm{R}^n, \quad A_i \in \mathcal{B}^s(H^N)$$

$$\text{(m)} \qquad \rho(\theta) = \theta^i A_i + \theta^0 A_0 \quad \theta \in \Theta = (0,1)^{n+1} \quad \Sigma_{i=0}^n \theta^i = 1, \quad A_i(\mathrm{Tr}A_i = 1) \in \mathcal{B}^+(H^N)$$

(The tensorial summation convention for repeated indices is used.)

We prove some of basic theorems known in the classical information geometry by extending the formulation to such a non-commutative smooth manifold, and establish (1) existence of a pair of dual affine coordinate systems in (e) and (m) indicating that these constitute a single, dually flat manifold, (2) a projection theorem to insure the Cramer-Rao inequality with an identification of the efficient estimator.

1. INTRODUCTION

The purpose of the present paper is to establish the answer to a question in mathematical theory of statistics which has a rather long history[1][2][3] and has been informed recently to the community of mathematical physicists by a good review article, Amari's *Differential-Geometrical Methods in Statistics*[4]. The question is situated in a central part of mathematical statistics i.e. *parameter estimation theory* for a smooth manifold of probability distributions where one desires to find out a best choice of the parameter values from given data by observation. To be concrete, two important families of distributions which are hitherto studied intensively should be mentioned; *exponential family* and *mixture family*.

When a statistical physicists touches upon the above framework, specifically, upon the geometrical aspect of the exponential family defined by

$$\text{(e)} \qquad p(x,\theta) = \exp(\theta^i C_i(x) - \psi(\theta)) \ \ (-\infty < \theta^i < \infty, \ \ i = 1 \dots n), \tag{1}$$

he should immediately understand the common idea that exists between the framework and the usual statistical mechanics or statistical thermodynamics[5]. Let us consider the simplest example of a one-parameter family of the manifold in (1) $\mathcal{S} = \{p(\cdot\ \theta) = \exp(\theta C_1(x) - \psi(\theta))\}$. In Physics, this is the familiar Boltzmann distribution for a mechanical system with Hamiltonian $-C_1(x)$ immersed in a heat bath with temperature θ^{-1}, and $\psi(\theta)$ represents its free energy. Its extension to a multi-dimensional manifold of the form (1) also appears in statistical thermodynamics, when the mechanical system has several constants of motion other than the Hamiltonian. Even without such constants, the form (1) may be used to improve the thermodynamic description of the system in terms of the higher order temperatures[5]. In any case, the condition of normalization for $p(x,\theta)$ (generally, in a measure space $\mu(x)$) is to specify the function $\psi(\theta)$ such that

$$\psi(\theta) = \log \int \exp\left(\theta^i C_i(x)\right) d\mu(x) \tag{2}$$

which can be identified with the (dimensionless) free-energy.

All that stated in the above is well-known, and every physicist also knows how to replace the classical distribution (1) and the related quantity (2) by quantum mechanical expressions in terms of generally non-commuting operators. What we are going to discuss in the sequel is a special geometrical aspect associated with the exponential family (1) that would be totally unfamiliar to physicists but that has been fully elucidated in ref.[4], and we aim at establishing the same aspect in a quantum mechanical framework. For this purpose, let us first outline briefly the known result according to Amari's description which relates the exponential family (1) to another family i.e. the mixture family

$$\text{(m)} \qquad p(x,\theta) = \theta^i C_i(x) + \theta^0 C_0(x) \quad (C_i(x) > 0, \ 0 < \theta^i < 1, \ \Sigma_{i=1}^n \theta^i < 1, \ \ i = 1, 2 \cdots n) \tag{3}$$

$$= \theta^i C_i(x) + \left(1 - \Sigma_{i=1}^n \theta^i\right) C_0(x), \quad \int C_i(x) d\mu(x) = 1 \ \ (i = 0, 1, \cdots n), \tag{4}$$

satisfying the normalization $\int p(x,\theta) d\mu(x) = 1$. This is the existence of a pair of dual coordinate systems (θ, η) by means of which the two families (e) and (m) become *one and identical distribution*, and is explained below more in detail.

Case for constituting the m-family from a given e-family (e)→(m) in terms of $\eta = \eta(\theta)$

An inspection of expression (1) shows that the logarithm of $p(x,\theta)$ denoted by $l(x,\theta)$ is a linear function in θ's except the last term $-\psi(\theta)$ which is independent of the random variable x. Then,

$$E\left[\partial_i \partial_j l(x,\theta)\right] (= \text{minus of the Fisher information metric tensor } g_{ij})$$

$$= -\partial_i \partial_j \psi(\theta) \qquad (\partial_i \text{ denotes } \partial/\partial\theta^i)$$

$$E\left[\partial_i \partial_j l(x,\theta) \partial_k l(x,\theta)\right] = 0 \ \text{ because } \ E\left[\partial_k l(x,\theta)\right] = 0.$$

These relations yield two important geometrical results for the e-family, namely

$$g_{ij}(\theta) = E\left[\partial_i l(x,\theta) \partial_j l(x,\theta)\right] = E\left[(C_i(x) - \partial_i \psi)(C_j(x) - \partial_j \psi)\right] = \partial_i \partial_j \psi(\theta) \tag{5}$$

(The ψ-function plays a role of the so-called potential function),

and

$$\Gamma^{(e)}_{ijk}(\theta) \equiv E\left[\partial_i\partial_j l(x,\theta)\partial_k l(x,\theta)\right] = 0 \tag{6}$$

(Coefficients of the e-connection vanish identically).

Eq.(5) implies that, according to a theorem of Amari (Th. 3.4 in ref.[4]) by taking another basis of the tangent space T_θ which is biorthogonal to the starting one such that $\langle \partial_j l, \partial^k p\rangle = \delta^k_j$, we can choose a new coordinate system $\eta_i = \eta_i(\theta) = \partial_i\psi(\theta)$ $(i = 1, 2, \cdots n)$ which together with θ forms a pair of mutually dual coordinate systems. Eq.(6), on the other hand, implies that the starting coordinate is affine, which ensures (by virtue of Th.3.5 in ref. [4]) that the above new coordinate η is also affine in the dual system, since

$$\Gamma^{(e)}_{ijk} + \Gamma^{(m)}_{ikj} = \partial_i g_{jk} = 0 \;\; \text{(by the choice of the dual pair } (\theta,\eta)) \tag{7}$$

holds. Therefore, the same but dually written family $\bar{p}(x,\eta) \equiv p(x,\theta(\eta))$ may be represented in the form (4) with θ's being replaced by η's and with an appropriately chosen set of $C_i(x)$'s.

Case for constituting the e-family from a given m-family (m)→(e) in terms of $\eta = \eta(\theta)$

We proceed to a similar analysis on (m), eqs.(3) and (4), on the basis of two formulas corresponding to (5) and (6) i.e.

$$\begin{aligned} g_{ij}(\theta) &= E\left[\partial_i l(x,\theta)\partial_j l(x,\theta)\right] = \int \frac{1}{p(x,\theta)}\partial_i p(x,\theta)\partial_j p(x,\theta)d\mu \\ &= \int \frac{1}{p(x,\theta)}(C_i(x) - C_0(x))(C_j(x) - C_0(x))d\mu \end{aligned} \tag{8}$$

$$\begin{aligned} \Gamma^{(m)}_{ijk} &= E\left[\{\partial_i\partial_j l(x,\theta) + \partial_i l(x,\theta)\partial_j l(x,\theta)\}\,\partial_k l(x,\theta)\right] \\ &= \int \frac{1}{p(x,\theta)}\left(\partial_i\partial_j p(x,\theta)\right)\partial_k p(x,\theta) = 0. \end{aligned} \tag{9}$$

As before, the result (8) guarantees the existence of the potential function

$$\psi(\theta) \equiv \int p(x,\theta)\log p(x,\theta)d\mu$$

which satisfies

$$\partial_i\partial_j\psi(\theta) = \int \frac{1}{p(x,\theta)}(C_i(x) - C_0(x))(C_j(x) - C_0(x))d\mu = g_{ij}(\theta), \tag{10}$$

indicating that the new coordinate $\eta(\eta_i \equiv \partial_i\psi)$ together with θ forms a pair of dual coordinate systems. Also, the result (9) assures that this η must be an affine coordinate for the dually written family $\bar{p}(x,\eta) \equiv p(x,\theta(\eta))$ to be represented as in (1), namely

$$\bar{p}(x,\eta) = \exp\left(\eta_i\bar{C}_i(x) - \psi(\eta)\right) \;\; \text{with some } \bar{C}_i(x) \text{ and } \bar{\psi}(\eta).$$

From the above two-fold account, therefore, it can be observed that we are dealing with just a single, smooth and flat manifold of distributions $\mathcal{S} = \{p(x)\}$ which is represented merely in two different coordinates. Our task in this paper, then, is to show the precisely same geometrical structure in non-commutative algebras. This analysis is contained in the next two sections, and the last section is devoted to the problem of estimation bound in quantum parameter-estimation theory.

2. QUANTUM STATISTICAL FORMULATION WITH CANONICAL METRIC

The first attempt to extend the Riemannian metric structure for information geometry to a quantum (non-commutative)framework was made by Ingarden et al [7] who noticed that the classical expression (5), in its first equality, for the information metric tensor is inapplicable but that the following expression still holds

$$g_{ij}(\theta) = -E\left(\partial_i\partial_j \log\rho\right)(\neq E\left(\partial_i \log\rho\partial_j \log\rho\right)). \tag{11}$$

They were able to remedy the unequated expression by means of an expansion formula for the exponential operator $\exp\left\{A(\theta) - \log(\mathrm{Tr}e^{A(\theta)})\right\}$ such that

$$g_{ij}(\theta) = \int_0^1 d\lambda\langle e^{-\lambda A}\left(\partial_i A - \langle\partial_i A\rangle\right)e^{\lambda A}\left(\partial_j A - \langle\partial_j A\rangle\right)\rangle \tag{12}$$

where the usual notation $\langle\cdot\rangle \equiv \mathrm{Tr}(\rho\cdot)$ is used in place of $E(\cdot)$ for a quantum-statistical expectation. The expression of the right-hand side in (12) was frequently used first in Kubo's theory[8], and its scalar-product nature was explicitly exploited in Mori's work[9] i.e.

$$\int_0^1 d\lambda \mathrm{Tr}\left(e^{(1-\lambda)A}Xe^{\lambda A}Y\right) = \int_0^1 d\lambda \mathrm{Tr}\left(e^{(1-\lambda)A}Ye^{\lambda A}X\right), \tag{13}$$

which is obtained by a change of integration variable $\lambda \to 1-\lambda$ and by virtue of the identity $\mathrm{Tr}(XYZ) = \mathrm{Tr}(YZX)$. We shall denote the above expression by $\ll X, Y \gg$ in contrast with the Hilbert-Schmidt inner product $\langle X, Y\rangle = \mathrm{Tr}(XY)$ for all $X, Y \in \mathcal{B}^s(H)$ (and also its complexitized generalization). Sometimes it is called Kubo-Mori/Bogoliubov inner-product. We prefer the naming *canonical inner-product* for this and *canonical metric* for the metric induced by this product. The reason will be made clear later.

An extensive study of the geometry in quantum statistics involving this inner product has been made recently by Petz[10]. Our present paper certainly has an intimate relation to this work and another[11]: On reading carefully, we find that his context does not include a comprehensive analysis of Amari's notion of *duality*, to which all of our efforts will be devoted for elucidating the points (in particular, the point stated in the last section of ref. [10]).

With above remarks about the previous papers important for us, we now state our formulation of the problem as follows. Our object statistical model $\mathcal{S}$ is a C^∞ manifold in the parameter space $\Theta(\theta^1, \theta^2, \cdots, \theta^n)$ of all N-dimensional, invertible density matrices in $\mathcal{B}^{++}(H^N)$, and define an exponential family

$$(e)\quad \rho(\theta) = \exp\left(\theta^i A_i - \psi(\theta)\right),\quad A_i \in \mathcal{B}^s(H^N)\ \text{ all hermitians } \in \mathcal{B}(H^N)$$

$$\text{with } \psi(\theta) = \log\left(\mathrm{Tr}e^{\theta^i A_i}\right) \text{ for } \mathrm{Tr}\rho(\theta) = 1, \quad \text{where the space } \Theta \text{ is the open set } \mathrm{R}^n, \tag{14}$$

also a mixture family

$$(m)\quad \rho(\theta) = \theta^i A_i + (1 - \Sigma_{i=1}^n \theta^i)A_0,\quad A_i \in \mathcal{B}^+(H^N),\quad (i = 0, 1, \cdots, n)$$

$$\text{with } \mathrm{Tr}A_i(i = 0, 1, \cdots, n) = 1,\ \text{ where the space } \Theta \text{ is the open set } (0,1)^n,\ \sum_{i=1}^n \theta^i < 1. \tag{15}$$

For simplicity, our analysis is restricted to real fields in the algebra, where the dimensionality n of the parameter space must be $n \leq \frac{1}{2}N(N+1)-1$ (cf. ref.[6]). Then, the problem is to answer to the question whether a pair of dual coordinate systems (θ, η) exists in the family (e) or (m), whether these coordinates are affine, and finally whether these two families are identical with each other. The affirmative answers should be given in parallel with the analysis outlined in Section 1 for the classical formulation.

First we show an essential ingredient about non-commutative differentiations which distinguish the operation on c-number and on matrix-valued manifolds: if $X(\theta)$ denotes a C^∞ function in a commutative manifold, the meaning of partial derivation $\partial_i X(\theta)$ is well-known to satisfy locally $\partial_j \partial_i X(\theta) = \partial_i \partial_j X(\theta)$. This can be modified, once $X(\theta)$ belongs to a matrix manifold $\mathcal{S}(\subset \mathcal{B}(H^N))$, such that only the commutative part of derivatives with $X(\theta)$ satisfies this relation. Our prescription for this problem is given in ref. [6] and summarized as follows.

Suppose that $X(\theta)$ is hermitian and invertible for every fixed values of θ. Then, a partial differentiation ∂_i on $X(\theta)$ should be replaced by

$$\delta_i X = \partial_i X + [X, \Delta_i], \quad \partial_i X \in \mathcal{C}(X), \quad \sqrt{-1}\Delta_i \in \mathcal{B}^s(H^N), \tag{16}$$

where $\mathcal{C}(X)$ denotes the commutant of X i.e. $\mathcal{C}(X) = \{A \in \mathcal{B}(H^N), [A, X] = 0\}$ and $[X, \Delta_i] \in \mathcal{C}(X)^\perp$ with respect to the Hilbert-Schmidt inner product: Representation (16) exhibits a unique orthogonal decomposition of $\delta_i X$ by this inner product into a commutative part and a commutator part with X.

The n-tuple $(\delta_1, \delta_2, \cdots \delta_n)$ forms a vector of super-operators on $\mathcal{S}$ (a basis of the non-commutative tangent space T_θ) whose basic property reads

i) it is a covariant vector: by a c-number transformation of parameter $\theta \to \bar{\theta}$

$$\delta_i = \frac{\partial \bar{\theta}^j}{\partial \theta^i} \bar{\delta}_j \ \text{(more precisely, } \partial_i = \frac{\partial \bar{\theta}^j}{\partial \theta^i} \bar{\partial}_j \ \text{ and } \ \Delta_i = \frac{\partial \bar{\theta}^j}{\partial \theta^i} \bar{\Delta}_j \ \text{ hold)} \tag{17}$$

ii) derivation property under a fixed $\mathcal{C}(X)$ with some $X \in \mathcal{B}^s(H^N)$

$$\delta_i (c_1 X_1 + c_2 X_2) = c_1 \delta_i X_1 + c_2 \delta_i X_2 \quad (c_{1,2} \ \text{ are } c \text{ number constants})$$

$$\delta_i (X_1 X_2) = (\delta_i X_1) X_2 + X_1 (\delta_i X_2) \tag{18}$$

iii) for any differentiable function of X i.e. $F(X)$, together with (16),

$$\delta_i F(X) = \partial_i F(X) + [F(X), \Delta_i], \quad \partial_i F(X) \in \mathcal{C}(X) \tag{19}$$

implying that the commutativity holds for ∂_i's; $\partial_i \partial_j = \partial_j \partial_i$.

iv)

$$\delta_i \delta_j X = \partial_i \partial_j X + [\partial_i X, \Delta_j] + [\partial_j X, \Delta_i] + [X, \partial_i \Delta_j] + [[X, \Delta_i], \Delta_j] \tag{20}$$

and hence

$$\delta_i \delta_j - \delta_j \delta_i = [\cdot\ , \Sigma_{ij}]$$

where

$$\Sigma_{ij} = \partial_i \Delta_j - \partial_j \Delta_i - [\Delta_i, \Delta_j] \tag{21}$$

which characterizes the non-commutativity of the derivation δ_i's.

We now have our central proposition for connecting a derivative of a positive hermitian and its logarithmic derivative.

Proposition 1. If $X(\theta)(\in \mathcal{S})$ is a strictly positive hermitian ($\in \mathcal{B}^{++}(H^N)$), the two kinds of derivatives $\delta_i X$ and $\delta_i \log X$ are related through

$$\delta_i X = \int_0^1 d\lambda X^{1-\lambda} (\delta_i \log X) X^{\lambda} = \int_0^1 d\lambda X^{\lambda} (\delta_i \log X) X^{1-\lambda}. \tag{22}$$

When applied this proposition to an invertible density matrix ρ, we immediately obtain, for the question of Ingarden et al. in eq.(11),

Proposition 2. The tensor $g_{ij}(\theta)$ of the canonical metric for the density matrix $\rho(\in \mathcal{S})$, which is defined by

$$g_{ij}(\theta) = \langle \delta_i \log \rho, \delta_j \rho \rangle \quad (= \langle \delta_i \rho, \delta_j \log \rho \rangle),$$

can be written as

$$\begin{aligned} g_{ij}(\theta) &= \int_0^1 d\lambda \mathrm{Tr} \left(\rho^{1-\lambda} (\delta_i \log \rho) \rho^{\lambda} (\delta_j \log \rho) \right) \\ &= \ll \delta_i \log \rho, \delta_j \log \rho \gg \quad (= g_{ji}(\theta)). \end{aligned} \tag{23}$$

Proof of (22). All what we need is to relate $\delta_i X$ in (16) and $\delta_i \log X$ in (19) (i.e. $F(X) \equiv \log X$). For this we can use

$$[X, \Delta] = \int_0^1 d\lambda X^{\lambda} [\log X, \Delta] X^{1-\lambda}$$

obtainable from the integration of both sides of the identity $\frac{d}{d\lambda}\left(X^{\lambda} \Delta X^{-\lambda}\right) = X^{\lambda}[\log X, \Delta] X^{-\lambda}$ from $\lambda = 0$ to $\lambda = 1$.

As to **Proposition 2**, we should remark that the starting definition of the canonical metric, $\langle \delta_i \log \rho, \delta_j \rho \rangle$, stems from a more general scope of dual metrics which we will discuss in the next section (cf. ref.[6]).

Combining **Proposition 2** with property iv) for a repeated derivation $\delta_i \delta_j$, we have, for answering to the question in (11),

Theorem 1. The canonical metric can be written as

$$g_{ij}(\theta) = -\langle \delta_i \delta_j \log \rho(\theta) \rangle = \ll \delta_i \log \rho, \delta_j \log \rho \gg, \tag{24}$$

where the derivation δ_i on the matrix space is defined in (16) or (19).

Proof. Using iv) with $X = \log \rho$, we compute $\langle \rho, \delta_i \delta_j \log \rho \rangle$ to obtain

$$\mathrm{Tr}(\rho \delta_i \delta_j \log \rho) = \mathrm{Tr} \rho \partial_i \partial_j \log \rho + \mathrm{Tr} \rho [[\log \rho, \Delta_i], \Delta_j]$$

(the rest three terms in (20) vanish because $\partial_i X, \partial_i \Delta_j$ etc. $\in \mathcal{C}(\rho)$)

$$= -\mathrm{Tr} \partial_i \log \rho \partial_j \rho - \mathrm{Tr}[\log \rho, \Delta_i][\rho, \Delta_j] = -\langle \delta_i \log \rho, \delta_j \rho \rangle.$$

It should be remarked that the above two propositions and hence **Theorem 1** have the validity under very general conditions: it does not require the affine entrance of the parameter θ in the exponential operators. This remark holds also to the following basic theorem about the Umegaki relative entropy, namely

Theorem 2. Let $D(\rho,\sigma)$ denotes the relative entropy $\mathrm{Tr}\rho(\log\rho - \log\sigma)$ $(\rho,\sigma \in \mathcal{S})$. Then,

$$D(\rho,\rho+d\rho) = D(\rho+d\rho,\rho) = \frac{1}{2}g_{ij}(\theta)d\theta^i d\theta^j + O(d\theta^3)$$

where

$$d\rho = \delta_i \rho d\theta^i. \tag{25}$$

The proof was given in refs.[7] and [12]. Thus, the general validity of the context in this section is the very reason of our naming *canonical metric* rather than Kubo-Mori/Bogoliubov. Discussions about *non-canonical metrics* must be postponed for later publications.

3. THE MAIN THEOREMS

Two theorems to be presented, Theorem 3 and 4, correspond to Amari's Th.3.4 and 3.5, respectively, adapted to the present non-commutative version.

Definition 1. For a given basis vector $\{\delta_i\}$ in the tangent space T_θ, another basis vector $\{\delta^i\}$ which satisfies

$$\langle \delta_i\rho, \delta^j \log\rho\rangle = \ll \delta_i \log\rho, \delta^j \log\rho \gg = \delta_i^j \ \ (\text{Kronecker}\delta) \tag{26}$$

is said to be biorthogonal to $\{\delta_i\}$. Then, this $\{\delta_i\}$ is biorthogonal to $\{\delta^i\}$ so that the relation is mutual. (The δ_i's transform covariantly, whereas δ^i's do contravariantly, hence, $\{\delta^i\}$ may be called a basis of cotangent space denoted by T^*_θ.) The biorthogonal bases exist in $\mathcal{S}$ as far as the metric g_{ij} is non-degenerate, since then

$$\delta^j = g^{jk}\delta_k \ \text{ where } \ g^{jk} = (G^{-1})^{jk} \ \text{ with } G = (g_{ij}). \tag{27}$$

A new coordinate system $\{\eta_i, \eta = \eta(\theta) \text{ and } \theta = \theta(\eta)\}$ by which $\{\delta_i\}$ is transformed into $\{\delta^i\}$ and conversely, namely

$$\delta_i = \frac{\partial \eta_k}{\partial \theta^i}\delta^k \ \text{ and } \ \delta^i = \frac{\partial \theta^k}{\partial \eta_i}\delta_k, \tag{28}$$

is said to be dual to the original coordinate system $\{\theta^i\}$, and by virtue of its mutuality, the two systems are said to be mutually dual.

Theorem 3. A necessary and sufficient condition for a pair of dual coordinate systems (θ,η) to exist in the smooth manifold $\mathcal{S}$ with a non-degenerate metric is that at least one of the metric tensor $g_{ij}(\theta)$ or $g^{ij}(\eta)$ can be given in terms of a scalar function $\psi(\theta)$ or $\phi(\eta)$ (called the potential function) by

$$g_{ij}(\theta) = \frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}, \ \text{ or } \ g^{ij}(\eta) = \frac{\partial^2\phi}{\partial\eta_i\partial\eta_j}, \tag{29}$$

and once this is satisfied, the other is automatically satisfied by the Legendre transformation

$$\eta_i = \frac{\partial\psi}{\partial\theta^i} \ \text{ or } \ \theta^i = \frac{\partial\psi}{\partial\eta_i} \ \text{ with the identity } \ \psi(\theta) + \phi(\eta) - \theta^i\eta_i = 0. \tag{30}$$

Proof of this theorem is precisely same as that provided by Amari for his Th. 3.4 in ref.[4].

Definition 2.* Given a Riemaniann metric $g_{ij}(\theta)$ in the smooth manifold $\mathcal{S}$ in terms of two differentiable functions $L(\rho)$ and $R(\rho)$ of $\rho = \rho(\theta)$ such that

$$g_{ij}(\theta) = \langle \delta_i L(\rho), \delta_j R(\rho) \rangle \quad (= \mathrm{Tr} \delta_i L(\rho) \delta_j R(\rho)) = \langle \delta_i R(\rho), \delta_j L(\rho) \rangle. \tag{31}$$

A coefficient of the affine connection associated with L and that associated with R is defined, respectively, by

$$\Gamma^L_{ijk}(\theta) = \langle \delta_j \delta_i L(\rho), \delta_k R(\rho) \rangle, \quad \Gamma^R_{ijk}(\theta) = \langle \delta_j \delta_i R(\rho), \delta_k L(\rho) \rangle, \tag{32}$$

which satisfy

$$\partial_i g_{jk}(\theta) = \Gamma^L_{ijk}(\theta) + \Gamma^R_{ikj}(\theta). \tag{33}$$

These are called the dual affine connections with respect to the Riemannian metric $g_{ij}(\theta)$. Best example of such is the α- connection with respect to the Fisher metric in the classical framework discussed in ref.[4] (cf. ref.[6] for the non-commutative extension). A vanishing of $\Gamma^L_{ijk}(\theta)$ for all i, j, k implies that the coordinate θ along a path satisfies $\ddot{\theta} = 0$ whose solution is a geodesic with respect to the metric g_{ij}, and the (every component of) θ is called an L-affine coordinate (or, the manifold $\mathcal{S}$ is L-flat with respect to g_{ij}).

Theorem 4. Suppose that the manifold $\mathcal{S}$ is equipped with a Riemannian metric $g_{ij}(\theta)$ as identified in **Definition 2** i.e. in the form (31), and that the coordinate θ is L-affine, then there exists in $\mathcal{S}$ a pair of dual coordinate systems (θ, η) such that the θ is L-affine and the η is R-affine coordinate system.

Before going into **proof** of this theorem, it would be worthwhile to remark about the problem of *torsion* (Amari's procedure of proving his Th.3.5 assumes a torsion-free manifold, and hence cannot be applied here straightforwardly). Consider the two expressions of connections $\Gamma^L_{ijk}(\theta)$ and $\Gamma^R_{ijk}(\theta)$ in (32), where these are not in general symmetric with respect to the first two indices i and j because of the non-commutativity of δ_i and δ_j indicated in (20).

Definition 3. The torsion $S^{L,R}_{ijk}$ associated with $\Gamma^{L,R}_{ijk}$ is defined by

$$S^{L,R}_{ijk} = \Gamma^{L,R}_{ijk} - \Gamma^{L,R}_{jik}. \tag{34}$$

Lemma. The two torsions associated with $L(\rho)$ and $R(\rho)$ are given in terms of Δ_i's in (20) as follows.

$$S^L_{ijk} = \mathrm{Tr}\,[L(\rho), [\Delta_i, \Delta_j]]\,[R(\rho), \Delta_k], \quad S^R_{ijk} = \mathrm{Tr}\,[R(\rho), [\Delta_i, \Delta_j]]\,[L(\rho), \Delta_k]. \tag{35}$$

When both $L(\rho)$ and $R(\rho)$ belong to $\mathcal{C}(\rho)$ for a fixed value of θ, then these are identical to each other by virtue of $[L(\rho), R(\rho)] = 0$, i.e.

$$S^L_{ijk} = S^R_{ijk}. \tag{36}$$

This Lemma can be proved straightforwardly by using the basic property of non-commutative derivations iv) to compute $\Gamma^{L,R}_{ijk}$.

The identity (36) yields now an important result which we list in

* We avoid the standard definition of *connection* in terms of covariant derivatives which is still not yet fully established on non-commutative manifolds.

Theorem 5. If the manifold $\mathcal{S}$ of density matrices $\rho(\theta)$ is L-torsion free ($S^L_{ijk} = 0\ \forall i,j,k$), then it is also R-torsion free. In other words, if $\Gamma^L_{ijk} = \Gamma^L_{jik}$ holds for all i,j,k's, then also $\Gamma^R_{ijk} = \Gamma^R_{jik}$.

Proof of Theorem 4. The condition that the coordinate θ is L-affine means $\Gamma^L_{ijk} = 0$ for all i,j,k's, and of course $S^L_{ijk} = 0$. Hence, also $S^R_{ijk} = 0$ which yields the symmetry $\Gamma^R_{ijk} = \Gamma^R_{jik}$. By virtue of the duality relation (33),

$$\partial_i g_{jk}(\theta) = \Gamma^R_{ikj}(\theta) = \Gamma^R_{kij} = \partial_k g_{ji}(\theta), \quad \text{or} \quad \partial_i g_{kj} = \partial_k g_{ji}$$

which shows the existence of another coordinate system η such that

$$g_{ij} = \frac{\partial \eta_j}{\partial \theta^i} = \frac{\partial \eta_i}{\partial \theta^j} \quad (\text{because } g_{ij} = g_{ji}).$$

Thus, there should exist a scalar function $\psi(\theta)$ such that $g_{ij}(\theta) = \partial_i \partial_j \psi(\theta)$ i.e. the potential function. From **Theorem 3**, therefore, the two coordinate systems (θ, η) in which the basis of η is chosen as biorthogonal to δ_i's are a pair of dual coordinate systems, and also this η becomes the R-affine coordinate.

It is now possible to examine our problem about non-commutative exponential and mixture families, and to get an affirmative answer to the starting questions at the beginning.

Theorem 6. A pair of dual affine coordinate systems exists both in the e-family (14) and in the m-family (15).

Proof. First, we assign

$$L(\rho) = \log \rho, \quad R(\rho) = \rho \tag{37}$$

to the exponential family, and

$$L(\rho) = \rho, \quad R(\rho) = \log \rho \tag{38}$$

to the mixture family. The reason for this assignment may be seen from a more general setting of α-family, $L(\rho) = L^\alpha(\rho) \equiv \frac{2}{1-\alpha}\left(\rho^{\frac{1-\alpha}{2}} - 1\right)$ and $R(\rho) = L^{-\alpha}(\rho) \equiv \frac{2}{1+\alpha}\left(\rho^{\frac{1+\alpha}{2}} - 1\right)$, a central object in Amari's context [4] (cf. ref.[6] for the non-commutative extension). Expression (37) is reduced by taking the limit $\alpha \to 1$, while (38) by the limit $\alpha \to -1$, yielding the common canonical metric.

Case for the exponential family (e) $\log \rho = \theta^i A_i - \psi(\theta)$ in (14).

For the c-number part, $\psi(\theta) = \log \mathrm{Tr} e^{\theta^i A_i}$ yields

$$\begin{aligned} \partial_i \psi(\theta) = \langle A_i \rangle, \quad \text{and} \quad \partial_i \partial_j \psi(\theta) &= \int_0^1 \lambda \langle \rho^{-\lambda} A_i \rho^\lambda A_j \rangle - \langle A_i \rangle \langle A_j \rangle \\ &= \ll A_i - \langle A_i \rangle, A_j - \langle A_j \rangle \gg . \end{aligned} \tag{39}$$

For the non-commutative part, on the other hand, the differentiation can be performed in two ways: one way is just by setting

$$\delta_i \left(\theta^j A_j\right) = A_i \quad (= \lim_{\epsilon \to 0} \frac{1}{\epsilon}(\theta + \epsilon - \theta) A_i), \tag{40}$$

and the other way is

$$\delta_i \left(\theta^j A_j\right) = \mathrm{proj}(A_i)|_{\mathcal{C}(\rho)} + [\theta^j A_j, \Delta_i] = A_i^0 + [\log \rho, \Delta_i] \tag{41}$$

in accordance with the prescription (16). In (41), A_i^0 stands for $\mathrm{proj}(A_i)|_{\mathcal{C}(\rho)}$ which means the diagonal part of A_i in the ρ-diagonal representation, satisfying $\langle A_i^0 \rangle = \langle A_i \rangle$. Hence expression

(41) just yields the decomposition of A_i into the diagonal and off-diagonal parts with respect to this representation.

We obtain, by equating (40) and (41),

$$[\log\rho, \Delta_i] = A_i - A_i^0, \tag{42}$$

which is inserted into the expression of the Riemannian metric

$$\begin{aligned}
g_{ij}(\theta) &=\ll \delta_i \log\rho, \delta_j \log\rho \gg \\
&= \mathrm{Tr}\rho\left(A_i^0 - \langle A_i\rangle\right)\left(A_j^0 - \langle A_j\rangle\right) + \ll [\log\rho, \Delta_i], [\log\rho, \Delta_j] \gg \\
&= \mathrm{Tr}\rho\left(A_i^0 - \langle A_i\rangle\right)\left(A_j^0 - \langle A_j\rangle\right) + \ll A_i - A_i^0, A_j - A_j^0 \gg \\
&=\ll A_i - \langle A_i\rangle, A_j - \langle A_j\rangle \gg \quad (\text{because } A_i^0 - \langle A_i\rangle \perp A_i - A_i^0 \text{ in both } \langle\ \rangle \text{ and } \ll\gg) \\
&= \partial_i\partial_j\psi(\theta) \qquad \text{from (39).}
\end{aligned} \tag{43}$$

Therefore, there should exist a pair of dual coordinate systems (θ, η). That the coordinate θ is affine can be assured from (40) i.e. $\delta_j\delta_i \log\rho = -\partial_i\partial_j\psi(\theta)$ so that $\Gamma_{ijk}^{(e)} = \langle\delta_j\delta_i \log\rho, \delta_k\rho\rangle = -g_{ij}(\theta)\mathrm{Tr}\delta_k\rho = 0$. Thus, the pair of dual coordinate systems $(\theta, \eta = \partial\psi)$ is e-affine for θ and m-affine for η, implying that $\bar\rho(\eta) \equiv \rho(\theta(\eta))$ may be represented as an m-family. The problem of precise identification of this $\bar\rho(\eta)$ with the form of m-family (15) will be treated separately after the reverse analysis (m) $\rightarrow$ (e).

Case for the mixture family (m) $\rho = \theta^i A_i + (1 - \sum\theta^i)A_0$ in (15).

Consider a scalar function $\psi(\theta)$ defined by

$$\psi(\theta) = \mathrm{Tr}\rho(\theta)\log\rho(\theta) \text{ (minus of the von Neumann entropy).}$$

$$\begin{aligned}
\text{Then,}\quad \partial_i\psi(\theta) &= \mathrm{Tr}(\delta_i\rho(\theta))\log\rho(\theta) + \mathrm{Tr}\rho(\theta)\delta_i\log\rho(\theta) \\
&= \mathrm{Tr}(\delta_i\rho(\theta))\log\rho(\theta) = \mathrm{Tr}(A_i - A_0)\log\rho(\theta),
\end{aligned}$$

because $\mathrm{Tr}\rho(\theta)\delta_i\log\rho(\theta) = \mathrm{Tr}\rho(\theta)\partial_i\log\rho(\theta) + \mathrm{Tr}\rho(\theta)[\log\rho(\theta), \Delta_i] = \mathrm{Tr}\delta_i\rho(\theta) = 0$.

$$\begin{aligned}
\text{Consequently,}\quad \partial_i\partial_j\psi(\theta) &= \mathrm{Tr}(\delta_i\rho(\theta))\delta_j\log\rho(\theta) \quad (\delta_j\delta_i\rho(\theta) = \delta_j(A_i - A_0) = 0) \\
&=\ll \delta_i\log\rho(\theta), \delta_j\log\rho(\theta) \gg
\end{aligned} \tag{44}$$

which means $\partial_i\partial_j\psi(\theta) = g_{ij}(\theta)$ that $\psi(\theta)$ is the potential function, and also $\Gamma_{ijk}^{(m)} = \langle\delta_j\delta_i\rho, \delta_k\log\rho\rangle = 0$. Therefore, again $(\theta, \eta = \partial\psi)$ is m-affine for θ and e-affine for η so that $\bar\rho(\eta) \equiv \rho(\theta(\eta))$ may be represented as an e-family.

Theorem 7 Bijective property of the dual transformation between (14) and (15). Let $\mathrm{H}(\eta) = \mathrm{H}(\eta_1, \cdots \eta_n)$ denote the image of the parameter space $\Theta(\theta^1, \cdots \theta^n)$ mapped by the dual transformation $\theta \rightarrow \eta$; $\eta_i = \partial_i\psi(\theta)$ and $\partial_i\partial_j\psi(\theta) = g_{ij}(\theta)$ with $G(\theta) = (g_{ij}(\theta)) > 0$. Then, the mapping $\Theta \rightarrow \mathrm{H}$ is one-to-one, and the image H can be identified with the parameter space of the inverse dual transformation $\eta \rightarrow \theta$; $\theta^i = \partial^i\phi(\eta)$ and $\partial^i\partial^j\phi(\eta) = g^{ij}(\eta)$, $(g^{ij}) = G^{-1}$, with the Legendre identity $\psi(\theta) + \phi(\eta) - \theta^i\eta_i = 0$ so that the mapping is a bijection between $\Theta = \mathrm{R}^n$ in (e) and $\Theta = (0,1)^n$, $\sum_{i=1}^n \eta < 1$, in (m).

Proof. Consider first the e-family (14) whose parameter space Θ is R^n. By **Theorem 6**, we know that every density matrix $\rho(\theta)$ can be represented in the form $\eta_i A^i + B$ with a fixed set

$\{A^i(i=1,\cdots n), B \in \mathcal{B}^s(H^N\}$ and with n parameters $\{\eta_i\}$ which are bounded because $\|\rho(\theta)\| < 1$. The η is determined by the condition $\eta_i = \sup_{\theta\in\Theta}(\eta_i\theta^i - \psi(\theta))$ with sup being replaced by max because of the convexity of the potential function $\psi(\theta)$, $\partial_i\partial_j\psi = g_{ij}$, $G = \{g_{ij}\} > 0$. This assures that the mapping $\Theta \to \mathrm{H}$ is injective. For each fixed θ, the set $\{\eta; \phi(\eta) \le \theta^i\eta_i - \psi(\theta)\}$ is a bounded convex set in R^n decomposable into convex polyhedrons, and we show that it can be a single unit: There exists an affine transformation $\{\eta_i\} \to \{\bar{\eta}_i\}$ such that every $\bar{\eta}_i$ satisfies $0 < \bar{\eta}_i < 1$ and

$$\eta_i A^i = \bar{\eta}_i \bar{A}^i + \eta_0 A^0. \tag{45}$$

Our choice is $\bar{\eta}_i = \frac{1}{\alpha}(\eta_i - \eta_0)$ and $\bar{A}^i = \alpha A^i, A^0 = \sum_{i=1}^n A^i$, where $\eta_0 = \min_{i,\theta\in\Theta}\partial_i\psi(\theta)$ and $\alpha = \max_{i,\theta\in\Theta}(\partial_i\psi(\theta) - \eta_0) > 0$. The corresponding Legendre identity becomes

$$\psi(\theta) + \phi(\eta) - \theta^i\eta_i = \bar{\psi}(\bar{\theta}) + \bar{\phi}(\bar{\eta}) - \bar{\theta}^i\bar{\eta}_i = 0$$

with $\bar{\theta} = \alpha\theta$, $\bar{\psi}(\bar{\theta}) = \psi(\theta) - \eta_0\sum_{i=1}^n\theta^i$, $\bar{\phi}(\bar{\eta}) = \phi(\eta)$, where the convex property of $\bar{\psi}(\bar{\theta})$ and $\bar{\phi}(\bar{\eta})$ is the same as before. Furthermore, we can assume that $\sum_{i=1}^n\bar{\eta}_i(\theta) < 1$ by a renormalization $\bar{\eta}_i(\theta) \to \frac{1}{\beta}\bar{\eta}_i(\theta)$ with $\beta = \max_{\theta\in\Theta}\sum_{i=1}^n\bar{\eta}_i(\theta)$ and $\bar{A}^i \to \beta\bar{A}^i$.

Therefore, without loss of generality, we may assume that

$$\rho(\theta) = \eta_i(\theta)A^i + B, \qquad 0 < \eta_i < 1 \;\; i = 1,\cdots n, \;\; \text{and} \;\; \Sigma_{i=1}^n\eta_i(\theta) < 1. \tag{46}$$

We enlarge the manifold $\mathcal{S}$ by introducing $n+1$ th component $\eta_0 > 0$ of η by which B is multiplied to constitute a homogenous coordinate system $\{\eta_i\}_{i=0}^n$, i.e. $\rho = \eta_i A^i$, and then the normalization is conditioned by $\sum_{i=0}^n\eta_i = 1$ so that

$$\rho(\theta) = \eta_i A^i, \qquad \Sigma_{i=0}^n\eta_i(\theta) = 1. \tag{47}$$

We now have

$$\rho(\theta) = \eta_i(\theta)A^i + \left(1 - \sum_{i=1}^n \eta_i(\theta)\right)A^0, \quad \mathrm{Tr}\rho(\theta) = 1. \tag{48}$$

If we take the closure $\overline{\mathrm{H}(\eta)}$ where the range of η's becomes $0 \le \eta_i \le 1$, the resulting $\rho = \sum_{i=0}^n\eta_i A^i$ may get a pure state, but still $\rho \ge 0$ and $\mathrm{Tr}\rho = 1$ holds. This should provide each A^i with positiveness and normalization such that

$$A^i \ge 0 \qquad \text{and} \qquad \mathrm{Tr}A^i = 1 \qquad i = 0, 1, \cdots n, \tag{49}$$

and also

$$\sum_{i=0}^n \eta_i(\theta) = 1. \tag{50}$$

This shows that the mapping $\Theta \to \mathrm{H}$ is surgective onto the space of the η-coordinate system.

A similar reasoning can be made, when we start from the m-family (15). In order to show that the image $\mathrm{H}(\eta)$ now is identical with R^n, it suffices to see that here $\log\rho(\theta)$ is unbounded when each θ^i tends to the end point 0 or 1.

4. PROJECTION THEOREM AND PARAMETER ESTIMATION INEQUALITY

The problem of parameter estimation in classical statistics is stated as follows [4].

Let X^i denote a (contravariant) vector of random variables whose expectation over a family of distributions parametrized by $\{\theta^i\} \in \Theta$ is restricted by the unbiasedness condition

$$E[X^i] = \theta^i, \quad \text{hence} \quad E[\partial_i \log \rho X^j] = \delta_i^j. \tag{51}$$

This vector random variable X^i is called an unbiased estimator. Then, the covariance matrix V of the unbiased estimator $V = (V^{ij})$, $V^{ij} \equiv E\left[(X^i - \theta^i)(X^j - \theta^j)\right]$, is lower bounded by the inverse Fisher metric tensor g^{ij} at each fixed θ-value so that the following inequality holds:

$$V \geq G^{-1}_{Fisher}, \quad \text{or} \quad V^{ij}(\theta) \geq g^{ij}(\theta) \quad ((V^{ij} - g^{ij})\xi_i\xi_j \geq 0 \quad \text{for any real } \xi_i's). \tag{52}$$

The choice of unbiased estimators is desired such that their fluctuations around θ be as small as possible; the smallest possible is the one whose covariance V^{ij} is just equal to g^{ij}, and is called *efficient estimator*. A search for this estimator is an important subject in each statistical model. One of merits of the exponential and mixture families (1) (2) and (3) (4), respectively, is that the efficient estimator can be explicitly constructed by virtue of their duality [13] (see below).

A new problem arises in quantum parameter estimation theory that the information metric defined on non-commutative tangent spaces is not unique so that the estimation inequality (52) must be set up depending on each metric tensor [15][10][11]. We establish the inequality with respect to the *canonical metric* on the best possible standpoint by using the projection theorem of Csiszar's form[14].

Theorem 8. Let $\mathcal{C}$ denote a closed convex set of the Hilbert space $\mathcal{H}$ defined by the canonical inner product $\ll \cdot, \cdot \gg$ introduced in $\mathcal{B}^s(H)$ (here $H = H^N$ but with possible generalizations to infinite N cases) and assume $\mathcal{C} \subset \bar{\mathcal{S}} =$ (all density operators $\subset \mathcal{J}(H) \cap \mathcal{B}^+(H)$). The relative entropy involving two density operators ρ and σ is defined by

$$\begin{aligned} D(\rho, \sigma) &= \mathrm{Tr}\rho(\log\rho - \log\sigma) \quad (\text{support of } \rho, \ \ s(\rho) \subset s(\sigma)) \\ &= +\infty \qquad\qquad\qquad\quad (s(\rho), \ s(\sigma) \ \text{otherwise}). \end{aligned} \tag{53}$$

For any σ in $\bar{\mathcal{S}}$ but not in $\mathcal{C}$, there exists a projection of σ onto $\mathcal{C}$ denoted by $\sigma_\perp$ such that

$$D(\sigma_\perp, \sigma) = \min_{\rho \in \mathcal{C}} D(\rho, \sigma), \quad \sigma_\perp \in \mathcal{C}. \tag{54}$$

The projection $\sigma_\perp$ is unique, and satisfies

$$D(\rho, \sigma) \geq D(\rho, \sigma_\perp) + D(\sigma_\perp, \sigma) \quad \rho \in \mathcal{C}, \tag{55}$$

where the equality (Pythagorean relation) holds for all $\rho \in \mathcal{C}$, iff $\mathcal{C}$ is affine i.e. a flat manifold.

Proof. Csiszar's version for the classical statistics is applicable in the present non-commutative case, since the parallelogram identity about the relative entropy holds here, and we present an outline of the proof. Let $\{\rho_n\} \subset \mathcal{C}$ be a sequence such that $D(\rho_n, \sigma) < \infty$ (valid, if σ is strictly positive) and

$$\lim_{n\to\infty} D(\rho_n, \sigma) = \inf_{\rho \in \mathcal{C}} D(\rho, \sigma) \equiv D_0.$$

Apply the parallelogram identity to this sequence, i.e.

$$D(\rho_m, \sigma) + D(\rho_n, \sigma) - 2D\left(\frac{\rho_m + \rho_n}{2}, \sigma\right) = D\left(\rho_m, \frac{\rho_m + \rho_n}{2}\right) + D\left(\rho_n, \frac{\rho_m + \rho_n}{2}\right), \tag{56}$$

and show that

$$\lim_{m,n\to\infty} D\left(\rho_{m,n}, \frac{\rho_m+\rho_n}{2}\right) = 0.$$

This is due to the lower semi-continuity of $D(\rho,\sigma)$ (for any $\epsilon > 0$, an integer N can be chosen in such a way that all $m, n > N$, $D(\rho_{m,n},\sigma) - D_0 < \epsilon$ and $D(\frac{\rho_m+\rho_n}{2},\sigma) - D_0 < \epsilon$ (due to the convexity of $\mathcal{C}$) so that the left-hand side and hence the right-hand side of (56) converges to 0). Thus, for sufficiently large m and n, the right-hand side of (56) can be replaced by the canonical metric $\frac{1}{4} \ll \rho_m - \rho_n\ ,\ \rho_m - \rho_n \gg$, implying that a Cauchy subsequence can be chosen from $\{\rho_n\}$ to show the $\lim_{n\to\infty}\rho_n$ to be identified with $\sigma_\perp$ in (54). That this $\sigma_\perp$ is unique can be assured also from the identity (56) with one of ρ_n being replaced by the limit $\sigma_\perp$ and with inequality $D\left(\frac{\rho+\sigma_\perp}{2},\sigma\right) \geq D(\sigma_\perp,\sigma)$ i.e.,

$$D(\rho,\sigma) \geq D(\sigma_\perp,\sigma) + D\left(\rho, \frac{\rho+\sigma_\perp}{2}\right) + D\left(\sigma_\perp, \frac{\rho+\sigma_\perp}{2}\right), \\ > D(\sigma_\perp,\sigma) \quad \text{for} \quad \rho \neq \sigma_\perp. \tag{57}$$

With this projection $\sigma_\perp$, we wish to prove the inequality (55). In order to do this, let us consider a convex combination of ρ and $\sigma_\perp$ in $\mathcal{C}$

$$\rho_t \equiv t\rho + (1-t)\sigma_\perp \in \mathcal{C}, \quad 0 \leq t \leq 1,$$

and

$$D(\rho_t,\sigma) = \mathrm{Tr}\rho_t(\log\rho_t - \log\sigma),$$

which yields

$$\frac{d}{dt}D(\rho_t,\sigma)|_{t=0} = \mathrm{Tr}\rho(\log\sigma_\perp - \log\sigma) - D(\sigma_\perp,\sigma) \\ = D(\rho,\sigma) - D(\rho,\sigma_\perp) - D(\sigma_\perp,\sigma).$$

Suppose that (55) does not hold for the given $\rho(\neq\sigma_\perp)$ in $\mathcal{C}$ i.e.

$$\frac{d}{dt}D(\rho_t,\sigma)|_{t=0} < 0.$$

Then, for some t, $0 < t < 1$, $D(\rho_t,\sigma) < D(\rho_{t=0},\sigma) = D(\sigma_\perp,\sigma)$, which contradicts with the fact that $\sigma_\perp$ is the projection of σ onto $\mathcal{C}$. This proves the validity of (55). The last statement that the equality in (55) is the only case of flat $\mathcal{C}$ is in accordance with Amari's formulation of the projection theorem [4] (also cf. ref.[14]).

We are in a position to formulate a quantum version of the parameter estimation inequality on the basis of the above theorem. A quantum mechanical unbiased estimator is a (contravariant) vector operator $X^i \in \mathcal{B}^s(H)$ which satisfies

$$\langle X^i\rangle (= \mathrm{Tr}\rho(\theta)X^i) = \theta^i \;\text{ so that }\; \langle\delta_i\rho, X^j\rangle = \ll \delta_i\log\rho, X^j \gg = \delta_i^j \text{ (Kronecker } \delta). \tag{58}$$

We also define its covariant version $X_i \equiv g_{ik}(\theta^k - X^k) + \eta_i(\theta)$, $\langle X_i\rangle = \eta_i(\theta)$

$$(= g_{ik}\partial^k\psi\{\theta(\eta)\} = \partial_i\psi(\theta) = \eta_i, \text{ if } (\theta,\eta) \text{ is a pair of dual coordinate systems}). \tag{59}$$

Theorem 9. For any unbiased estimator X^i which satisfies (58), the following two sets of inequalities hold:

$$\text{contravariant version} \qquad \ll X^i - \theta^i, X^j - \theta^j \gg \; \geq g^{ij}(\theta) \tag{60}$$

$$\text{covariant version} \qquad \ll X_i - \eta_i(\theta), X_j - \eta_j(\theta) \gg \; \geq g_{ij}(\theta), \tag{61}$$

where $g^{ij}(\theta) = (g_{ij}(\theta))^{-1}_{ij}$ and the inequalities imply that the tensor $V - G$ is a positive definite tensor i.e.

$$(V-G)^{ij}\xi_i\xi_j \geq 0 \quad \text{for any real} \;\; \xi_i\,'s.$$

Proof. In **Theorem 8**, we first choose an exponential family represented in the equivalent mixture family as the convex set $\mathcal{C}$

$$\mathcal{C} = \{\rho(\theta) = \eta_i(\theta)A^i + (1 - \Sigma\eta_i(\theta))A^0\}. \tag{62}$$

Then it is a flat manifold in terms of the dual coordinate system $\{\eta_i\}$. Let σ be an arbitrary density operator in $\mathcal{S}$ specified in terms of a real, free covariant vector ξ_i such that

$$\sigma = e^{\log\rho(\theta)+X-\psi(\theta)}, \;\; X = \xi_i(X^i - \theta^i), \text{ and } \psi(\theta) \text{ is the normalization factor.}$$

This function $\psi(\theta)$ is determined by

$$\psi(\theta) = \log\rho(\theta) - \log\sigma + X, \;\; \text{hence taking } \langle\;\rangle_\rho \text{ of both hand side,}$$

$$= D(\rho(\theta),\sigma) + \xi_i\langle X^i - \theta^i\rangle_\rho = D(\rho(\theta),\sigma)$$

by virtue of the unbiasedness condition (58). Therefore, the projection $\sigma_\perp$ of σ onto $\mathcal{C}$ is indicated by

$$\min_{\rho\in\mathcal{C}} D(\rho,\sigma) = \min_{\theta\in\Theta}\psi(\theta) = D(\sigma_\perp,\sigma),$$

in which $\sigma_\perp$ can be represented as

$$\sigma_\perp = \rho(\theta) + \eta_i\delta^i\rho(\theta) \;\; \text{with } \eta_i \text{ to be determined by the minimality condition.}$$

In terms of the biorthogonal bases (δ_i, δ^j), the condition is given by

$$\log\sigma - \log\sigma_\perp \perp \mathcal{C}, \;\; \text{i.e. } \langle\delta_i\rho, \log\sigma - \log\sigma_\perp\rangle (= \ll \delta_i\log\rho, \log\sigma - \log\sigma_\perp \gg) = 0 \;\; \text{or,}$$

$$\langle\delta_i\rho, X\rangle = \langle\delta_i\rho, L\rangle \;\; \text{where} \;\; L = \eta_i\delta^i\log\rho(\theta),$$

which yields

$$\eta_i = \xi_i, \;\; \text{and } D(\sigma_\perp,\rho) = \mathrm{Tr}\sigma_\perp(\log\sigma_\perp - \log\rho) = \frac{1}{2}\mathrm{Tr}(\delta^i\rho\delta^j\log\rho)\xi_i\xi_j = D(\rho,\sigma_\perp) \text{ in } O(\xi^2). \tag{63}$$

Now, the inequality (55) (here the Pythagorean equality because of the flat $\mathcal{C}$) can be used in the form $D(\rho,\sigma) \geq D(\rho,\sigma_\perp)$, and up to $O(\xi^2)$ expressed as

$$\ll X^i - \theta^i \;,\; X^j - \theta^j \gg \xi_i\xi_j \geq \; \ll \delta^i\log\rho, \delta^j\log\rho \gg \xi_i\xi_j$$

which is the desired inequality (60), and is transformable into its covariant version (61).

In the above proof, we have assumed the set $\mathcal{C}$ to be an exponential family. This assumption can be removed for those families which are restricted only by $g_{ij}(\theta) > 0$, where $\mathcal{C}$ is chosen as the tangent space of $\rho(\theta)$ at a fixed θ and σ is near to this ρ in $O(\xi^2)$.

Our final result is concerning the efficient estimator of the exponential and mixture families. To be comprehensive, we first give an inverse map of two derivations in Proposition 1, namely

Proposition 3. The inversion of the map $\delta_i \log X \rightarrow \delta_i X$ given by (22) is expressed as

$$\delta_i \log X = \int_0^\infty dt \frac{1}{t+X} \delta_i X \frac{1}{t+X} \equiv \Phi(\delta_i X). \tag{64}$$

This was devised by Petz [10], and can be verified also by means of our decomposition formula (16).

Theorem 10. Let an exponential family $\rho(\theta) = \exp(\theta^i A_i - \psi(\theta))$ be represented in the corresponding mixture family

$$\rho(\theta) = \eta_i(\theta)(A^i - A^0) + A^0, \quad \text{where} \quad \eta_i(\theta) = \partial_i \psi(\theta) = \langle A_i \rangle. \tag{65}$$

Then, the vector operator A_i yields the unique efficient estimator X_i in the covariant version such that

$$\ll A_i - \eta_i(\theta) \ , \ A_j - \eta_j(\theta) \gg = g_{ij}(\theta). \tag{66}$$

Similarly, the inverse map Φ of the vector operator $A^i - A^0$ yields the fluctuating part of the efficient estimator X^i in the contravariant version such that

$$\ll \partial^i \phi + \Phi(A^i - A^0) - \theta^i, \partial^j \phi + \Phi(A^j - A^0) - \theta^j \gg = g^{ij}(\theta), \tag{67}$$

where

$$\partial^i \phi = \mathrm{Tr}(A^i - A^0) \log \rho \quad \text{is the systematic part of } X^i.$$

Proof. Expression (66) is just the result of analysis from eq.(39) to (43) in the proof of **Theorem 6, Case for the exponential family**. The same analysis can be made and added to **Case for the mixture family** in terms of the η-coordinate system. This yields the explicit form for $g^{ij}(\eta)$ with

$$\delta^i \log \rho(\eta) = \Phi(\delta^i \rho(\eta)), \tag{68}$$

where

$$\delta^i \rho(\eta) = A^i - A^0. \tag{69}$$

Thus, by inserting this into $g^{ij}(\eta) = \ll \delta^i \log \rho, \delta^j \log \rho \gg$ and noting that $\langle \Phi(A^i - A^0) \rangle = 0$, we obtain expression (67).

Consequently, the starting question what is the efficient estimator for the parameter θ in the exponential family (65) is now answered:

$$X^i_{efficient} = \partial^i \phi + \Phi(A^i - A^0), \quad \langle X^i_{efficient} \rangle = \partial^i \phi = \theta^i, \tag{70}$$

where A^i $i = 1, \cdots n$ and A^0 are given in the mixture representation of the family (65).

Remark. In **Theorem 9**, the covariant version and the contravariant version of an estimator can be defined under the most general condition of a nondegenerate metric, which includes the case where a pair of dual coordinates exists, as specified in (59). Such a general condition does not warrant the existence of each efficient estimator. **Theorem 10** implies the speciality of the (e) and (m) families for which the efficient estimator can be identified in both versions.

Acknowledgements.
The author wishes to thank D. Petz and H. Nagaoka for providing him with their reprints having discussions about a number of crucial points during the year of 1994.

REFERENCE

[1] S. Kullback, *Information Theory and Statistics*, Wiley, New York (1959).

[2] N.N. Chentsov, *Statistical Decision Rules and Optimal Inference* (in Russian) Nauka Moscow (1972); English translation, AMS, Rhode Island (1982).

[3] R.S. Ingarden, Y. Sato, K. Sugawa and M. Kawaguchi, *Information thermodynamics and differential geometry, Tensor* N. S. **33**(1979) 347-353.

[4] S. Amari, *Differential Geometrical Method in Statistics: Lecture Notes in Statistics* Vol 28 (Springer 1985).

[5] R.S. Ingarden, *Information Theory and Variational Principles in Statistical Physics, Bull. Acad. Polon. Sci. Sér. math. phys. astr.* **11** (1963), 541-547.

[6] Preliminary note of the present result is given in H. Hasegawa, *Non-commutative Extension of the Information Geometry* Proc. Intern. Workshop on *Quantum Communications and Measurement*, Nottingham 1994 ed. by V. P. Belavkin, O. Hirota and R. l. Hudson, Plenum Publishing Corp. (1995).

[7] R.S. Ingarden, H. Janyszek, A. Kossakowski and T. Kawaguchi, *Information geometry and quantum statistical systems, Tensor*, N. S. **37**(1982) 105-111.

[8] T. Kubo, *Statistical mechanical theory of irreversible processes, J. Phys. Soc. Japan* **12**(1957), 570.

[9] H. Mori, *Transport, collective motion and Brownian motion, Prog. Theor. Phys.* **33** (1965), 423-455.

[10] D. Petz, *Geometry of canonical correlation on the state space of a quantum system, J. Math. Phys.* **35** (1994), 780-795.

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

[12] H. Hasegawa, *α-divergence of the non-commutative information geometry, Rep. Math. Phys.* **33** (1993) 87-93.

[13] S. Amari, *Differential geometry of curved exponential families — curvatures and information loss, Ann. Statistics* **10** (1982), 357-385.

[14] I. Csiszar, *I-divergence geometry of probability distributions and minimization problems, Ann. Probability* **3** (1975), 146-158.

[15] H.Nagaoka, *On the relation between Kullback divergence and Fisher information — from classical systems to quantum systems;* an outline of *IEICE Tech. Rep.* **89**(1989), 9-14.