Information geometry of quantum states based on the symmetric logarithmic derivative

Akio Fujiwara

METR 94-11 August 1994

Information geometry of quantum states based on the symmetric logarithmic derivative

Akio Fujiwara,
University of Tokyo[*]

Abstract

The autoparallelity of quantum statistical models is investigated from an information geometrical viewpoint. The importance of canonical observables in quantum estimation theory is also stressed.

Keywords : quantum estimation theory, information geometry, symmetric logarithmic derivative, canonical observable

[*]Department of Mathematical Engineering and Information Physics, Faculty of Engineering, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan.

1 Introduction

In the classical estimation theory, geometrical methods have made great success in clarifying the invariant nature of estmation together with in offering some clear-cut methodologies. It was also found that it gives an important application of the so-called dualistic geometry, a generalization of the Riemannian geometry [1]. It is natural to ask whether some geometrical methods are also useful in quantum estimation theory [2][3]. Indeed, many authors have tried to find geometrical aspects of quantum estimation theory [4][5][6][7][8]. We should notice, however, that there exist a variety of manners to define quantum counterparts of geometrical notions which played essential roles in the classical estimation theory, such as logarithmic derivative, Fisher information, Cramér–Rao inequality, etc. Moreover, only a few of them have been proved crucial in solving concrete quantum estimation problems so far [9][10][11]. Therefore, if we invoke an easy analogy of a certain aspect of classical information geometry, we cannot expect fruitful geometrical viewpoints. In order not to construct abundance of useless imitations, we must stand upon a deeper understanding of inherent geometrical aspects of quantum estimation theory. In this paper, we offer a new geometrical structure of quantum states and investigate its significance in quantum estimation theory.

In Sec. 2, we define a dualistic geometrical structure of quantum states based on the symmetric logarithmic derivatives. This structure has a non-vanishing torsion field in general, so that there exists no divergence functions on the quantum space. In Sec. 3, the autoparallelity of a quantum state model is investigated in detail, which also clarify the difference between classical and various quantum criterions of estimation. In Sec. 4, a condition for a one-parameter unitary model to be autopallel, i.e., to have an efficient estimator, is derived. This condition indicates the importance of canonical observables in an estimation theoretical viewpoint. Some examples are also presented.

2 Dualistic structure on quantum states

For brevity, the readers are assumed familiar with some elementary knowledge of classical information geometry [1]. Let $\mathcal{H}$ be a separable Hilbert space which corresponds to a physical system with inner product $\langle\phi|\psi\rangle$, $(\phi, \psi \in \mathcal{H})$. A quantum state is represented by a *density operator* ρ which satisfies

$\rho = \rho^* \geq 0$ and $\mathrm{Tr}\,\rho = 1$. In this paper, we further assume $\rho > 0$ for simplicity. Denote by $\mathcal{P}$ the totality of such strictly positive density operators. Let us start with definitions of quantum analogues of some basic notions in classical theory. We first define the *mixture representation* of the tangent vector ∂ by the isomorphism $\partial \leftrightarrow \partial\rho$ and call $\partial\rho$ the *m-tangent vector*. Denote their totality by $T_\rho^{(m)}(\mathcal{P})$. On the other hand, as we mentioned in the previous section, there exist a variety of candidates of the quantum counterpart of the exponential representation of the tangent vector due to the non-commutativity. Nevertheless, we here adopt the isomorphism $\partial \leftrightarrow L$ defined by

$$\partial\rho = \frac{1}{2}[\rho L + L\rho], \quad L = L^*$$

as the *exponential representation*, since it gives the definition of the so-called symmetric logarithmic derivative (SLD) which leads us to the most informative Cramér–Rao bound in the one parameter estimation theory [12]. Let us call L the *e-tangent vector* and denote their totality by $T_\rho^{(e)}(\mathcal{P})$. These isomorphisms are characterized by the following schemata:

- m-tangent vector

$$\begin{array}{ccccl} T_\rho(\mathcal{P}) & \simeq & T_\rho^{(m)}(\mathcal{P}) & = & \{G\,;\ G = G^*,\ \mathrm{Tr}\,G = 0\} \\ \partial & \simeq & G & & \text{where} \quad \partial\rho = G \end{array}$$

- e-tangent vector

$$\begin{array}{ccccl} T_\rho(\mathcal{P}) & \simeq & T_\rho^{(e)}(\mathcal{P}) & = & \{H\,;\ H = H^*,\ \mathrm{Tr}\,\rho H = 0\} \\ \partial & \simeq & H & & \text{where} \quad \partial\rho = \frac{1}{2}[\rho H + H\rho], \quad H = H^* \end{array}$$

Correspondingly, we define two kinds of parallel translations:

- m-parallel translation

$$\begin{array}{ccccc} T_\rho(\mathcal{P}) & \simeq & T_\rho^{(m)}(\mathcal{P}) & \ni & G \\ \downarrow & & \downarrow & & \downarrow \\ T_{\rho'}(\mathcal{P}) & \simeq & T_{\rho'}^{(m)}(\mathcal{P}) & \ni & G \end{array}$$

- e-parallel translation

$$\begin{array}{ccccc} T_\rho(\mathcal{P}) & \simeq & T^{(e)}_\rho(\mathcal{P}) & \ni & H \\ \downarrow & & \downarrow & & \downarrow \\ T_{\rho'}(\mathcal{P}) & \simeq & T^{(e)}_{\rho'}(\mathcal{P}) & \ni & H - \mathrm{Tr}\,\rho' H \end{array}$$

These parallel translations naturally define two affine connections $\nabla^{(m)}$ and $\nabla^{(e)}$, which are called the mixture connection (m-connection) and the exponential connection (e-connection). Further, by using the symmetrized inner product $\langle A, B\rangle_\rho = \frac{1}{2}\mathrm{Tr}\,\rho(A^*B + BA^*)$, we define the Riemannian metric at ρ by

$$g_\rho(\partial_i, \partial_j) = \langle L_i, L_j\rangle_\rho = \frac{1}{2}\mathrm{Tr}\,\rho(L_iL_j + L_jL_i).$$

Since the metric can be written as $g_\rho(\partial_i, \partial_j) = \mathrm{Tr}\,(\partial_i\rho)L_j$, the two connections $\nabla^{(m)}$ and $\nabla^{(e)}$ are mutually dual with respect to the metric g in the following sense: For arbitrary vector fields X, Y, Z on $\mathcal{P}$,

$$X g(Y, Z) = g(\nabla^{(m)}_X Y, Z) + g(Y, \nabla^{(e)}_X Z)$$

holds. It is evident that both curvature tensors with respect to the connections $\nabla^{(m)}$ and $\nabla^{(e)}$ vanish, since the two parallel translations are defined independently of the choice of the path connecting ρ and ρ'. The torsion tensor with respect to the m-connection also vanishes, whereas the torsion tensor $T^{(e)}$ with respect to the e-connection does not vanish in general since it becomes (see Appendix A)

$$T^{(e)}(\partial_j, \partial_k)\rho = \frac{1}{4}[[L_j, L_k], \rho].$$

From this fact, there does not exist divergence functions on $\mathcal{P}$ in general.

3 Autoparallelity in quantum estimation theory

In classical estimation theory, one of the most important geometrical notion is the autoparallelity of a model with respect to the e-connection, which is an equivalent condition for the existance of the efficient estimator of the model [1]. An e-autoparallel model is also called an exponential family which takes the form

$$p_\theta(x) = \exp\left[\theta^i f_i(x) - \psi(\theta)\right],$$

where $\theta = (\theta^1, \cdots, \theta^n)$ is the n-dimensional e-affine parameter to be estimated statistically, $\psi(\theta)$ the normalization factor, and the Einstein's summation convention $\theta^i f_i(x) = \sum_i \theta^i f_i(x)$ is used. In this section, we investigate in detail the quantum counterpart of this notion.

We first give a brief summary of the conventional quantum estimation theory. For details, consult [2][3]. Suppose we are given a statistical parametric model composed of strictly positive density operators:

$$\mathcal{S} = \{\rho_\theta \,;\, \rho_\theta = \rho_\theta^* > 0,\ \mathrm{Tr}\,\rho_\theta = 1,\ \theta \in \Theta \subset \mathbf{R}^n\}. \tag{1}$$

Here, $\theta = (\theta^1, \cdots, \theta^n)$ is the parameter to be estimated statistically. An *estimator* for θ is identified to a generalized measurement which takes values on Θ. The expectation vector with respect to the measurement M at the state ρ_θ is defined as

$$E_\theta[M] = \int \hat{\theta} P_\theta^M(d\hat{\theta}).$$

The measurement M is called *unbiased* if $E_\theta[M] = \theta$ holds for all $\theta \in \Theta$, i.e.,

$$\int \hat{\theta}^j P_\theta^M(d\hat{\theta}) = \theta^j, \quad (j = 1, \cdots, n). \tag{2}$$

Differentiation yields

$$\int \hat{\theta}^j \frac{\partial}{\partial \theta^k} P_\theta^M(d\hat{\theta}) = \delta_k^j, \quad (j, k = 1, \cdots, n). \tag{3}$$

If (2) and (3) hold at a certain θ, M is called *locally unbiased* at θ. Obviously, M is unbiased iff M is locally unbiased at every $\theta \in \Theta$. Letting M be a locally unbiased measurement at θ, we define the covariance matrix $V_\theta[M] = [v_\theta^{jk}] \in \mathbf{R}^{n \times n}$ with respect to M at the state ρ_θ by

$$v_\theta^{jk} = \int (\hat{\theta}^j - \theta^j)(\hat{\theta}^k - \theta^k) P_\theta^M(d\hat{\theta}). \tag{4}$$

A lower bound for $V_\theta[M]$ is given by the following quantum Cramér–Rao inequality

$$V_\theta[M] \geq (J_\theta)^{-1}, \tag{5}$$

where J_θ is the SLD–Fisher information matrix whose i, j entry is identical to the metric g_{ij}. When the model is one dimensional, the inequalities (5) become scalar, and the corresponding lower bound can be attained by the estimator $T = \theta I + L_\theta / J_\theta$, where I is the identity. Thus, the one parameter quantum estimation theory is quite analogous to the classical one as long as the SLD is used. On the other hand, for the dimension $n \geq 2$, the matrix equalities in (5) cannot be attained in general, because of the impossibility of the exact simultaneous measurement of non–commuting observables. We must, therefore, abandon the strategy of finding the measurement that minimizes the covariance matrix itself. Rather, we often adopt another strategy as follows: Given a non-negative definite real matrix $G = [g_{jk}] \in \mathbf{R}^{n\times n}$, find the measurement M that minimizes the quantity

$$\mathrm{tr}\, GV_\theta[M] = \sum_{jk} g_{jk} v_\theta^{jk}. \tag{6}$$

If there is a constant C such that $\mathrm{tr}\, GV_\theta[M] \geq C$ holds for all M, C is called a Cramér–Rao type bound or simply a CR bound, which may depend on both G and θ.

We next consider some conditions relevant to the efficiency of the estimator. Let us first add some terminologies:

1. A locally unbiased measurement M is called *locally efficient* at θ if

   $$V_\theta[M] \leq V_\theta[M']$$

   holds for every locally unbiased measurement M' at θ. A measurement M is called *efficient* if M is locally efficient for all θ.

2. Given an arbitrary weight (real symmetric non-negative matrix) G, a locally unbiased measurement M is called *G–locally efficient* at θ if

   $$\mathrm{tr}\, GV_\theta[M] \leq \mathrm{tr}\, GV_\theta[M']$$

   holds for every locally unbiased measurement M' at θ. Given an arbitrary weight field $\mathcal{G} = \{G_\theta \mid \theta \in \Theta\}$, a measurement M is called *$\mathcal{G}$–efficient* if M is G_θ–locally efficient for all θ. In particular, if $G_\theta \equiv G$ for all θ, M is called *G-efficient*.

There exist some evident relations between these notions, which are listed in the following propositions to put the issues in order.

Proposition 1 *The following conditions for a measurement M are equivalent:*

(i) *M is locally efficient at θ.*

(ii) *For all* $G > 0$*,* M *is* G*–locally efficient at* θ*.*

(iii) *For all* $\mathbf{v} = (v_1, \cdots, v_n) \in \mathbf{R}^n$*,* M *is* $\mathbf{v}^T\mathbf{v}$*–locally efficient at* θ*.*

(iv) M *is locally unbiased at* θ *and* $V_\theta[M] = (J_\theta)^{-1}$ *holds at* θ*.*

Proposition 2 *The following conditions for a measurement* M *are equivalent:*

(i) M *is efficient.*

(ii) *For all* $\mathcal{G}$*,* M *is* $\mathcal{G}$*–efficient.*

(iii) *For all* $G > 0$*,* M *is* G*–efficient.*

(iv) *For all* $\mathbf{v} = (v_1, \cdots, v_n) \in \mathbf{R}^n$*,* M *is* $\mathbf{v}^T\mathbf{v}$*–efficient.*

(v) M *is unbiased and* $V_\theta[M] = (J_\theta)^{-1}$ *holds for all* θ*.*

Proposition 3 *Given a model* $\mathcal{S} = \{\rho_\theta \mid \theta \in \Theta\}$*, consider the following conditions:*

(i) $\mathcal{S}$ *has an efficient measurement.*

(ii) $\mathcal{S}$ *has (possibly* $\mathcal{G}$*-dependent)* $\mathcal{G}$*–efficient measurements* $M_\mathcal{G}$ *for all* $\mathcal{G}$*.*

(iii) $\mathcal{S}$ *has (possibly* G*-dependent)* G*–efficient measurements* M_G *for all* $G > 0$*.*

(iv) $\mathcal{S}$ *has (possibly* $\mathbf{v}^T\mathbf{v}$*-dependent)* $\mathbf{v}^T\mathbf{v}$*–efficient measurements* $M_{\{\mathbf{v}^T\mathbf{v}\}}$ *for all* $\mathbf{v} = (v_1, \cdots, v_n) \in \mathbf{R}^n$*.*

(v) *There exists a certain* $\mathcal{G}$ *for which* $\mathcal{S}$ *has a* $\mathcal{G}$*–efficient measurement.*

(vi) $\mathcal{S}$ *is e-autoparallel, all the SLD's commute, and* θ *is the m-affine coordinate system.*

In classical theory, all these conditions are equivalent. In quantum case, however, only the following relations hold in general:

$$\text{(i)} \Longleftrightarrow \text{(vi)}, \qquad \text{(i)} \Longrightarrow \text{(ii)} \Longrightarrow \begin{cases} \text{(iii)} \\ \text{(iv)} \\ \text{(v)} \end{cases}$$

Proposition 3 clarifies the difference between the classical and the quantum estimation theories. Indeed, since the SLD's do not commute in general, we cannot expect the existance of an efficient estimator. We therefore have adopted another strategy to minimize the weighted sum of the covariances (6) instead of the classical minimization strategy of the covariance matrix itself. Nevertheless, Proposition 3 also shows that there exist a variety of strategies which are not mutually equivalent in general. Indeed, the relations between the conditions (iii) (iv) (v) are not yet clear.

By the way, the condition (iv) in Proposition 3 is closely related to the one parameter quantum estimation theory which is rather well-established so far. Indeed, the restriction of the Cramér–Rao inequality (5) in the $\mathbf{v}$-direction

$$\mathbf{v} V_\theta[M] \mathbf{v}^T \geq \mathbf{v}(J_\theta)^{-1}\mathbf{v}^T \tag{7}$$

gives explicitly the locally attainable lower bound for $\mathrm{tr}\,(\mathbf{v}^T\mathbf{v})V_\theta[M]$. More precisely, the following lemma holds.

Lemma 1 *Given a model $\mathcal{S} = \{\rho_\theta\ ;\ \theta = (\theta^1, \ldots, \theta^n) \in \Theta\}$ and an arbitrary $\mathbf{v} \in \mathbf{R}^n$, consider the differential equation*

$$\frac{d\theta^i}{dt} = v_j J^{ji}, \tag{8}$$

where J^{ij} is the i, j entry of $(J_\theta)^{-1}$, and denote by $\theta(t)$ the solution of (8) for an arbitrarily fixed initial condition $\theta_0 \in \Theta$. Then the restricted Cramér–Rao inequality (7) can be regarded as the one dimensional Cramér–Rao inequality for the one parameter sub-model $\rho_{\theta(t)}$.

Proof Let L_j be the SLD for the parameter θ^j. For every locally unbiased measurement M of θ,

$$T(\mathbf{v}) = \int v_i \hat{\theta}^i M(d\hat{\theta})$$

becomes a locally unbiased estimator of the parameter $\theta(\mathbf{v}) = v_i\theta^i$ at θ and satisfies

$$\begin{aligned}\langle T(\mathbf{v}) - \theta(\mathbf{v}), L_j\rangle_{\rho_\theta} &= \mathrm{Re}\,\mathrm{Tr}\,\rho_\theta \int v_i\hat{\theta}^i M(d\hat{\theta}) L_j \\ &= v_i \mathrm{Re}\,\mathrm{Tr}\,L_j\rho_\theta \int \hat{\theta}^i M(d\hat{\theta}) \\ &= v_i \mathrm{Tr}\,\frac{\partial\rho_\theta}{\partial\theta^j}\int \hat{\theta}^i M(d\hat{\theta})\end{aligned}$$

$$
\begin{aligned}
&= v_i \frac{\partial}{\partial \theta^j} \mathrm{Tr}\, \rho_\theta \int \hat{\theta}^i M(d\hat{\theta}) \\
&= v_i \frac{\partial}{\partial \theta^j} \theta^i = v_j.
\end{aligned}
$$

Then the orthogonal projection of $T(\mathbf{v}) - \theta(\mathbf{v})$ onto the SLD–tangent space span $\{L_j\}_{j=1}^n$ with respect to the inner product $\langle \cdot, \cdot \rangle_{\rho_\theta}$ is $S(\mathbf{v}) = v_j J^{ji} L_i$ since, by expressing $S(\mathbf{v}) = c^i L_i$,

$$
v_j = \langle T(\mathbf{v}) - \theta(\mathbf{v}), L_j \rangle_{\rho_\theta} = \langle S(\mathbf{v}), L_j \rangle_{\rho_\theta} = c^i \langle L_i, L_j \rangle_{\rho_\theta} = c^i J_{ij},
$$

where J_{ij}, J^{ij} are the i,j entries of J_θ, $(J_\theta)^{-1}$, respectively. Therefore, the inequality

$$
\langle T(\mathbf{v}) - \theta(\mathbf{v}), T(\mathbf{v}) - \theta(\mathbf{v}) \rangle_{\rho_\theta} \geq \langle S(\mathbf{v}), S(\mathbf{v}) \rangle_{\rho_\theta}
$$

characteristic of the projection is nothing but (7). Now, let us determine the sub-model $\rho_{\theta(t)}$ whose Cramér–Rao inequality becomes (7), i.e., whose SLD is $S(\mathbf{v})$. Observing

$$
\frac{1}{2}[S(\mathbf{v})\rho_{\theta(t)} + \rho_{\theta(t)} S(\mathbf{v})] = \frac{d\rho_{\theta(t)}}{dt} = \frac{\partial \rho_\theta}{\partial \theta^i} \frac{d\theta^i}{dt} = \frac{1}{2}[L_i \rho_\theta + \rho_\theta L_i] \frac{d\theta^i}{dt},
$$

we have

$$
S(\mathbf{v}) = L_i \frac{d\theta^i}{dt}.
$$

Therefore, the desired sub-model is determined by the following differential equation

$$
\frac{d\theta^i}{dt} = v_j J^{ji}, \qquad \theta(0) = \theta_0.
$$

Since

$$
\frac{d}{dt}\left[v_i \theta^i \right] = v_i J^{ij} v_j > 0,
$$

we can take $t = v_i \theta^i$, which proves the lemma. ∎

This lemma immediately leads us to the following theorem, which shows that the condition (iv) in Proposition 3 is a sufficient condition for the model to be totally e-geodesic.

Theorem 1 *If a model* $\mathcal{S} = \{\rho_\theta \,;\, \theta = (\theta^1, \ldots, \theta^n) \in \Theta\}$ *has (possibly* $\mathbf{v}$*-dependent)* $\mathbf{v}^T\mathbf{v}$*–efficient measurements for all* $\mathbf{v} = (v_1, \cdots, v_n) \in \mathbf{R}^n$*, then* $\mathcal{S}$ *is totally e-geodesic and* θ *is the m-affine coordinate system.*

Proof From the assumption, the restricted Cramér–Rao bound (7) can be attained by a certain θ-independent measurement. Then, by invoking Lemma 1, the sub-model $\rho_{\theta(t)}$ which is determined by $\mathbf{v}$ and ρ_0 has an efficient estimator for t. Then $\rho_{\theta(t)}$ must be an e-geodesic and t is an m-affine parameter. Since $\mathbf{v}$ is arbitrary, the model is totally e-geodesic and θ is m-affine. ■

Theorem 1 asserts that the minimization strategy of $\operatorname{tr} GV_\theta[M]$, $(G \geq 0)$ is closely related to the geometric structure of the model. Indeed, if the model is not totally e-geodesic (hence not e-autoparallel), we may not expect the existance of G-efficient measurement in general. Therefore the geodesic nature of a model is an important notion in quantum estimation theory.

In general, an autoparallel submanifold is automatically a totally geodesic submanifold. Conversely, a totally geodesic submanifold becomes an autoparallel submanifold if the enveloping manifold is torsion free [13, II, p. 53]. In this sense, the assumption of Theorem 1 may have little to do with the condition for a model to be e-autoparallel. Though a necessary and sufficient condition for a model to be e-autoparallel has been obtained by Nagaoka [14], the relation between the assumption of Theorem 1 and the e-autoparallelity of the model is not yet clarified so far.

4 Autoparallelity in unitary models

Since a model ρ_θ is represented, in general, by a spectral decomposition

$$\rho_\theta = \int p_\theta(x) E_\theta(dx),$$

the parameter change is composed of two parts: eigenvalue part (classical part) and unitary part (purely quantum part). One extreme case is a model where E_θ is θ-independent (classical statistical model). Let us call another extreme a unitary model where p_θ is θ-independent. Owing to group theoretical symmetry of the physical system, we often encounter such unitary models. Therefore, it is expected that the study of unitary models may bring us some important suggestions toward the construction of general quantum estimation theory. In this section, we investigate the geodesic nature of unitary models.

Since Theorem 1 indicates the importance of the decomposition of the model into foliation of one-dimensional submanifolds, we consider here one-dimensional unitary models and derive a necessary and sufficient condition for the model to be an e-geodesic. Let us consider a one parameter unitary model of the form

$$\rho_t = e^{if(t)A}\rho_0 e^{-if(t)A}, \tag{9}$$

where A is a self-adjoint generator, ρ_0 a strictly positive initial density operator, $f(t)$ a real monotonic odd function with one dimensional real parameter t. Let us recall the commutation operator $\mathfrak{D}$ introduced by Holevo [3]. For an arbitrary state ρ and a self-adjoint operator X, $\mathfrak{D}$ is defined by

$$i(X\rho - \rho X) = \frac{1}{2}((\mathfrak{D}X)\rho + \rho(\mathfrak{D}X)), \quad (\mathfrak{D}X)^* = \mathfrak{D}X.$$

Lemma 2 *Denoting by $\mathfrak{D}_t$ the commutation operator with respect to the unitary model ρ_t, then*

$$\mathfrak{D}_t A = e^{if(t)A}(\mathfrak{D}_0 A)e^{-if(t)A},$$

holds, and the SLD becomes

$$L_t = f'(t)e^{if(t)A}(\mathfrak{D}_0 A)e^{-if(t)A}.$$

Proof

$$\begin{aligned}
\frac{d\rho_t}{dt} &= if'(t)[A\rho_t - \rho_t A] \\
&= if'(t)e^{if(t)A}[A\rho_0 - \rho_0 A]e^{-if(t)A} \\
&= f'(t)e^{if(t)A}\frac{1}{2}[\rho_0(\mathfrak{D}_0 A) + (\mathfrak{D}_0 A)\rho_0]e^{-if(t)A} \\
&= f'(t)\frac{1}{2}\left[\rho_t e^{if(t)A}(\mathfrak{D}_0 A)e^{-if(t)A} + e^{if(t)A}(\mathfrak{D}_0 A)e^{-if(t)A}\rho_t\right].
\end{aligned}$$

This lead to the lemma. ■

Lemma 3 *The unitary model (9) becomes an e-geodesic iff there exists a real function $J(t)$ such that the self-adjoint operator*

$$T = \frac{f'(0)}{J(0)}\mathfrak{D}_0 A$$

satisfies the following relation

$$e^{if(t)A}Te^{-if(t)A} = \frac{f'(0)J(t)}{f'(t)J(0)}(T - t).$$

In this case, T becomes the efficient estimator of the parameter t, and $J(t)$ becomes the SLD–Fisher information of the model.

Proof In general, a one-parameter model ρ_t has an efficient estimator T of the parameter t iff there exists a real function $J(t)$ such that

$$L_t = J(t)(T - t)$$

holds. Comparing this and Lemma 2, we have

$$f'(t)e^{if(t)A}(\mathfrak{D}_0 A)e^{-if(t)A} = J(t)(T - t).$$

Setting $t = 0$, we have the definition of T, which immedietly leads to the lemma. ■

Lemma 4 *Suppose we are given self-adjoint operators A, B, and real functions $f(t)$, $g(t)$ such that $f(-t) = -f(t)$, $g(0) = 1$. The equality of the form*

$$e^{if(t)A}Be^{-if(t)A} = g(t)(B - t) \tag{10}$$

holds iff there exists a non-zero real number μ such that

$$f(t) = \frac{t}{\mu}, \qquad g(t) = 1, \qquad [A, B] = i\mu.$$

Proof The sufficiency follows immedietly from the expansion formula

$$e^{itA/\mu}Be^{-itA/\mu} = B + \frac{it}{\mu}[A, B] + \frac{1}{2!}\left(\frac{it}{\mu}\right)^2 [A, [A, B]] + \cdots.$$

We show the necessity. Since the left-hand side of (10) is a equi-spectrum deformation, the right-hand side is also so only when B has continuous spectrum, so that the Hilbert space $\mathcal{H}$ on which B acts must be infinite dimensional. Denoting the eigen-equation for B by

$$B|b\rangle = b|b\rangle, \qquad (\ |b\rangle \in \mathcal{H}, \quad b \in \mathbf{R}\ ). \tag{11}$$

Then, from (10), we have

$$e^{if(t)A}Be^{-if(t)A}|b\rangle = g(t)(b - t)|b\rangle,$$

or

$$Be^{-if(t)A}|b\rangle = g(t)(b - t)e^{-if(t)A}|b\rangle,$$

which indicates that $e^{-if(t)A}|b\rangle$ is an eigenvector of B with eigenvalue $g(t)(b-t)$. On the other hand, operating $e^{if(t)A}$ to (11) from the left, and using (10), we have

$$\begin{aligned} e^{if(t)A}Be^{-if(t)A}e^{if(t)A}|b\rangle &= b\, e^{if(t)A}|b\rangle \\ &= g(t)(B-t)e^{if(t)A}|b\rangle, \end{aligned}$$

which leads to

$$Be^{if(t)A}|b\rangle = \left(\frac{b}{g(t)} + t\right) e^{if(t)A}|b\rangle.$$

This indicates that $e^{if(t)A}|b\rangle$ is an eigenvector of B with eigenvalue $(b/g(t)+t)$. Now, $e^{-if(t)A}|b\rangle$ and $e^{if(t)A}|b\rangle$ are one-parameter family of eigenvectors of B which start from a common eigenvector $|b\rangle$ and, since $f(t)$ is assumed odd, these eigenvectors must be related by

$$e^{-if(t)A}|b\rangle = e^{if(-t)A}|b\rangle.$$

Therefore, the corresponding eigenvalues must be identical:

$$g(t)(b-t) = \left(\frac{b}{g(-t)} - t\right),$$

or

$$b\left\{g(t) - \frac{1}{g(-t)}\right\} - t\{g(t)-1\} = 0.$$

Since this relation must hold for any b and t, we have $g(t) = 1$. In this case, (10) is reduced to

$$e^{if(t)A}Be^{-if(t)A} = B - t.$$

Expanding the left-hand side as

$$B + if(t)[A,B] + \frac{\{if(t)\}^2}{2!}[A,[A,B]] + \cdots = B - t,$$

and differentiating by t and setting $t = 0$, we have $[A,B] = i/f'(0)$, since $f(0) = 0$. Substituting this commutation relation into the above expansion, we see $f(t) = f'(0)t$. Setting $\mu = 1/f'(0)$, we have the conditions in the lemma. ■

These lemmas immedietly lead us to the following main theorem.

Theorem 2 *The one parameter unitary model (9) becomes e-geodesic iff there exist non-zero reals μ and J such that*

$$f(t) = \frac{t}{\mu}, \qquad [A, \mathfrak{D}_0 A] = i\mu^2 J$$

holds. In this case, $T = \mathfrak{D}_0 A/\mu J$ is the efficient estimator of the parameter t, and J is the SLD–Fisher information of the model.

Proof From Lemma 4, the conditions derived in Lemma 3 are rewritten as

$$f(t) = \frac{t}{\mu}, \qquad \frac{J(t)}{J(0)} = 1, \qquad [A, T] = i\mu, \qquad T = \frac{1}{\mu J(0)} \mathfrak{D}_0 A.$$

Setting $J = J(0)$, we have the theorem. ■

This theorem shows that, only when the generator is a canonical observable, the one parameter unitary model becomes an e-geodesic, i.e., it has an efficient estimator. Note that this fact is strongly indebted to the assumption where the model is strictly positive. For instance, this condition can be considerably loosened for pure state models, see [15].

Let us further determine ρ_0 which satisfy the relation $\mathfrak{D}_0 A = \mu J B$, ($[A, B] = i\mu$). Taking the non-commutative Fourier transformation (see Appendix B) of the identity

$$\rho_0 A - A\rho_0 = \frac{i}{2}\mu J[\rho_0 B + B\rho_0],$$

we have

$$-x\mathfrak{F}_{x,k}\{\rho_0\} = \mu^2 J \frac{\partial}{\partial x}\mathfrak{F}_{x,k}\{\rho_0\}.$$

Integrating this under the condition $\mathfrak{F}_{0,0}\{\rho_0\} = 1$, we have

$$\mathfrak{F}_{x,k}\{\rho_0\} = \exp(-\frac{x^2}{2\mu^2 J})F(k),$$

where $F(k)$ is an arbitrary function which satisfies $F(0) = 1$ and some regularity conditions. Then ρ_0 is written in the Weyl representation as

$$\rho_0 = \int \exp(-\frac{x^2}{2\mu^2 J})F(k) \exp\left[-\frac{i}{\mu}(kA + xB)\right] \frac{dxdk}{2\pi\mu}. \qquad (12)$$

For instance, it becomes a quantum Gaussian state if we take $F(k)$ as Gaussian. Here, the *quantum Gaussian state* of Boson with one degree of freedom is defined by the Glauber-Sudershan's P-representation as

$$\rho_\theta = \frac{1}{\pi}\int \varphi_\theta(z)|z\rangle\langle z|\, d^2 z, \tag{13}$$

where $\theta = (q+ip)/\sqrt{2\hbar}$ is the parameter with q, p being the expectation of position and momentum, respectively, $|z\rangle$ the Bozon coherent state with $z \in \mathbf{C}$, and

$$\varphi_\theta(z) = \frac{1}{\langle N\rangle}\exp\left[-\frac{|z-\theta|^2}{\langle N\rangle}\right], \tag{14}$$

with $\langle N\rangle$ a quantity which is related to the system temperature [3, Chap. V]. It can also be written in the form

$$\rho_\theta = U_\theta \rho_0 U_\theta^*, \qquad U_\theta = \exp\left[\frac{i}{\hbar}(pQ - qP)\right].$$

With a help of Proposition 4 in Appendix B, the corresponding quantum characteristic function becomes

$$\begin{aligned}\mathfrak{F}_{x,k}\{\rho_\theta\} &= \mathrm{Tr}\,\rho_\theta \exp\left[\frac{i}{\hbar}(kQ + xP)\right] \\ &= \exp\left[\frac{i}{\hbar}(kq + xp) - \frac{1}{2\hbar^2}(\sigma_Q^2 k^2 + \sigma_P^2 x^2)\right],\end{aligned} \tag{15}$$

where

$$\sigma_Q^2 = \hbar\left(\langle N\rangle + \frac{1}{2}\right), \qquad \sigma_P^2 = \hbar\left(\langle N\rangle + \frac{1}{2}\right) \tag{16}$$

are the variances of position Q and momentum P which satisfy

$$\sigma_Q \sigma_P = \hbar\left(\langle N\rangle + \frac{1}{2}\right).$$

Corollary 1 *Quantum Gaussian model is e-autoparallel.*

Proof Recall that for a Gaussian model, SLD's with respect to the real parameters q, p are [3, p. 277]

$$L_q = (Q-q)/\sigma_q^2, \qquad L_p = (P-p)/\sigma_p^2,$$

the commutation operator is [3, p. 250]

$$\mathfrak{D}Q = \hbar(P-p)/\sigma_p^2, \qquad \mathfrak{D}P = -\hbar(Q-q)/\sigma_q^2,$$

and the SLD Fisher metric becomes [3, p. 277]

$$J = \begin{bmatrix} 1/\sigma_p^2 & 0 \\ 0 & 1/\sigma_q^2 \end{bmatrix}.$$

The restricted one-dimensional sub-model in the direction $\mathbf{v} = (v_q, v_p)$ is then determined by the SLD

$$S = v_j J^{ji} L_i = v_q(Q - q) + v_p(P - p) = \frac{1}{\hbar}\mathfrak{D}R,$$

where $R = -v_q\sigma_q^2 P + v_p\sigma_p^2 Q$. Note R is independent on q and p. The sub-model is therefore written also in the form

$$\rho_t = e^{iRt/\hbar}\rho_0 e^{-iRt/\hbar}.$$

Observing

$$[R, \mathfrak{D}_0 R] = i\hbar^2\left(v_q^2\sigma_q^2 + v_p^2\sigma_p^2\right)$$

we can conclude, with the help of Theorem 2, that the sub-model is an e-geodesic. Since $\mathbf{v}$ is arbitrary, a Gaussian model is proved to be totally e-geodesic. Then it is also e-autoparallel since the torsion of the model vanish. ■

Here we give another example in the spin 2×2 representation. An e-geodesic (quantum exponential family) which has an efficient estimator σ_z is [12]

$$\rho_t = e^{\frac{1}{2}[t\sigma_z - \gamma(t)]}\ \rho_0\ e^{\frac{1}{2}[t\sigma_z - \gamma(t)]}, \qquad \gamma(t) = \log[\mathrm{Tr}\,\rho_0 e^{t\sigma_z}]. \tag{17}$$

If we set

$$\rho_0 = \frac{1}{2}\begin{bmatrix} 1 & x_0 \\ x_0 & 1 \end{bmatrix}, \qquad -1 \le x_0 \le 1$$

without loss of generality, then (17) becomes

$$x(t) = \frac{x_0}{\cosh t}, \qquad y(t) = 0, \qquad z(t) = \tanh t$$

in the Stokes' representation, i.e., $\rho_t = \frac{1}{2}(I + x(t)\sigma_x + y(t)\sigma_y + z(t)\sigma_z)$. It indicates that the e-geodesic is an ellipse of the form

$$\left\{\frac{x(t)}{x_0}\right\}^2 + \{z(t)\}^2 = 1,$$

which connects the north and south poles. Therefore, it cannot be an unitary model unless $x_0 = \pm 1$, i.e., ρ_0 is a pure state, since the equi-eigenvalue surfaces are spherical shells centered at the origin.

5 Conclusions

An information geometrical aspects of quantum statistical models were studied. We first introduced a natural dualistic structure on the quantum state space based on the symmetric logarithmic derivatives. We next investigated the autoparallelity and clarified the difference between the classical and the quantum estimation theory. The importance of canonical observables in quantum estimation theory was also stressed.

Acknowledgment

I am mostly indebted to Professor H. Nagaoka.

A Torsion of the e-connection

We first construct two e-parallel vector fields X_1 and X_2 by translating two arbitrarily fixed tangent vectors ∂_1, ∂_2 at ρ_0 with respect to the e-connection. Then

$$\nabla^{(e)}_{X_1} X_2 = \nabla^{(e)}_{X_2} X_1 = 0,$$

and the torsion becomes

$$T^{(e)}(X_1, X_2) = \nabla^{(e)}_{X_1} X_2 - \nabla^{(e)}_{X_2} X_1 - [X_1, X_2] = -[X_1, X_2].$$

Letting the e-representation of the tangent vector ∂_j ($j = 1, 2$) at ρ_0 be L_j, the e-representation of the tangent vector $(X_j)_\rho$ becomes

$$(X_j)_\rho \simeq (X_j^{(e)})_\rho = L_j - \mathrm{Tr}\, \rho L_j,$$

which acts on ρ as

$$X_j \rho = \frac{1}{2} \left[\rho (X_j^{(e)})_\rho + (X_j^{(e)})_\rho \rho \right] = \frac{1}{2} \left[(\rho L_j + L_j \rho) - 2\rho \mathrm{Tr}\, \rho L_j \right]. \qquad (18)$$

Since the quantuty $X_2 X_1 \rho$ describe the change of $X_1 \rho$ when ρ is slightly moved along X_2-direction,

$$\begin{aligned} X_2 X_1 \rho &= \frac{1}{2} \left[(X_2^{(e)} \rho)(X_1^{(e)})_\rho + (X_1^{(e)})_\rho (X_2^{(e)} \rho) \right] \\ &= \frac{1}{2} \left[((X_2 \rho) L_1 + L_1 (X_2 \rho)) - 2 (X_2 \rho) \mathrm{Tr}\, \rho L_1 - 2\rho \mathrm{Tr}\, (X_2 \rho) L_1 \right]. \end{aligned}$$

Substituting (18) into the above equation, we have

$$X_2 X_1 \rho = \frac{1}{4}(\rho L_2 L_1 + L_1 L_2 \rho) + \{\text{ symmetric terms with respect to } L_1, L_2 \}.$$

We can evaluate $X_1 X_2 \rho$ in the same way, yielding

$$T^{(e)}(X_1, X_2)\rho = [X_1, X_2]\rho = \frac{1}{4}[[L_1, L_2], \rho].$$

Now, since ρ_0 is arbitrary, the torsion at any point $\rho \in \mathcal{P}$ becomes

$$T^{(e)}(\partial_j, \partial_k)\rho = \frac{1}{4}[[L_j, L_k], \rho], \tag{19}$$

where L_j is the e-representation of the tangent vector ∂_j at ρ. For an arbitrary submanifold M, the torsion $T_M^{(e)}$ is obtained by projecting $T^{(e)}$ onto the tangent space of M with respect to the Riemannian metric.

B Non-commutative Fourier transform

In this appendix, we give a brief summary of the Weyl representation of Hilbert-Schmidt operators and the non-commutative Fourier transforms under the Fock space representation. For details, see [3, p. 223][16, p. 178]. For an arbitrary Hilbert-Schmidt operator A, define

$$\mathcal{F}_{x,k}\{A\} = \mathrm{Tr}\, A \exp\left[\frac{i}{\hbar}(kQ + xP)\right] = a(x,k), \tag{20}$$

where $[Q, P] = i\hbar$, then

$$A = \int a(x,k) \exp\left[-\frac{i}{\hbar}(kQ + xP)\right] \frac{dxdk}{2\pi\hbar} \tag{21}$$

holds. (20) is called the *non-commutative Fourier transform* of A, and (21) is called the *Weyl representation* of A. In particular, the non-commutative Fourier transform of a density operator $\mathcal{F}_{x,k}\{\rho\}$ is called the *quantum characteristic function*. With a help of Baker–Hausdorff formula, we have

$$\begin{aligned} \mathcal{F}_{x,k}\{\rho\} &= \mathrm{Tr}\,\rho\, \exp\left[\frac{i}{2\hbar}kx\right] \exp\left[\frac{i}{\hbar}kQ\right] \exp\left[\frac{i}{\hbar}xP\right] \\ &= \mathrm{Tr}\,\rho\, \exp\left[-\frac{i}{2\hbar}kx\right] \exp\left[\frac{i}{\hbar}xP\right] \exp\left[\frac{i}{\hbar}kQ\right]. \end{aligned}$$

Differentiating these equations, we have useful formulae. For instance,

$$\begin{aligned}\frac{\partial}{\partial k}\mathfrak{F}_{x,k}\{\rho\} &= \mathrm{Tr}\,\rho\left(\frac{i}{2\hbar}x+\frac{i}{\hbar}Q\right)\exp\left[\frac{i}{2\hbar}kx\right]\exp\left[\frac{i}{\hbar}kQ\right]\exp\left[\frac{i}{\hbar}xP\right]\\ &= \frac{i}{2\hbar}x\mathfrak{F}_{x,k}\{\rho\}+\frac{i}{\hbar}\mathfrak{F}_{x,k}\{\rho Q\},\end{aligned}$$

yields

$$\mathfrak{F}_{x,k}\{\rho Q\} = \left[-\frac{x}{2}-i\hbar\frac{\partial}{\partial k}\right]\mathfrak{F}_{x,k}\{\rho\}.$$

In the same way,

$$\begin{aligned}\mathfrak{F}_{x,k}\{Q\rho\} &= \left[+\frac{x}{2}-i\hbar\frac{\partial}{\partial k}\right]\mathfrak{F}_{x,k}\{\rho\}\\ \mathfrak{F}_{x,k}\{\rho P\} &= \left[+\frac{k}{2}-i\hbar\frac{\partial}{\partial x}\right]\mathfrak{F}_{x,k}\{\rho\}\\ \mathfrak{F}_{x,k}\{P\rho\} &= \left[-\frac{k}{2}-i\hbar\frac{\partial}{\partial x}\right]\mathfrak{F}_{x,k}\{\rho\}.\end{aligned}$$

Then we obtain the following formulae

$$\mathfrak{F}_{x,k}\{\rho Q-Q\rho\} = -x\mathfrak{F}_{x,k}\{\rho\},\qquad \mathfrak{F}_{x,k}\{\rho Q+Q\rho\} = -2i\hbar\frac{\partial}{\partial k}\mathfrak{F}_{x,k}\{\rho\}$$

$$\mathfrak{F}_{x,k}\{\rho P-P\rho\} = +k\mathfrak{F}_{x,k}\{\rho\},\qquad \mathfrak{F}_{x,k}\{\rho P+P\rho\} = -2i\hbar\frac{\partial}{\partial x}\mathfrak{F}_{x,k}\{\rho\}$$

The next formula is also useful, which translates the Weyl representation into the Glauber-Sudarshan's P-representation, and *vice versa* [16, p. 178].

Proposition 4 *The two representations of a trace class operator* T

$$\begin{aligned}T &= \int t(x,k)\exp\left[-\frac{i}{\hbar}(kQ+xP)\right]\frac{dxdk}{2\pi\hbar}\\ &= \int \varphi(p,q)|p,q\rangle\langle p,q|\frac{dpdq}{2\pi\hbar}\end{aligned}$$

are related by

$$\tilde{\varphi}(x,k) = t(x,k)\exp[\frac{1}{4\hbar}(x^2+k^2)].$$

where $\tilde{\varphi}(x,k)$ *is the Fourier transform of* $\varphi(p,q)$, *i.e.,*

$$\tilde{\varphi}(x,k) = \int \varphi(p,q)\exp\left[\frac{i}{\hbar}(kq+xp)\right]\frac{dpdq}{2\pi\hbar}.$$

References

[1] S. Amari, *Differential-Geometrical Methods in Statistics*, Lecture Notes in Statistics, Vol. 28 (Springer, Berlin, 1985).

[2] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).

[3] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982) (in Russian, 1980).

[4] W. K. Wootters, "Statistical distance and Hilbert space," Phys. Rev. **D23**, 357–362 (1981).

[5] R. S. Ingarden, H. Janyszek, A. Kossakowski, and T. Kawaguchi, "Information Geometry of Quantum Statistical Systems," Tensor, N. S., **37**, 105-111 (1982).

[6] E. R. Caianiello and W. Guz, "Quantum Fisher metric and uncertainty relations," Phys. Lett. **A126**, 223–225 (1988).

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

[8] S. L. Braunstein and C. M. Caves, "Statistical distance and the geometry of quantum states," Phys. Rev. Lett. **72**, 3439–3443 (1994).

[9] C. W. Helstrom, "Minimum Mean-Square Error Estimation in Quantum Statistics," Phys. Lett., **25A**, 101-102 (1967).

[10] H. P. H. Yuen and M. Lax, "Multiple-Parameter Quantum Estimation and Measurement of Non-Selfadjoint Observables," IEEE Trans., **IT-19**, 740-750 (1973).

[11] H. Nagaoka, "A new approach to Cramér-Rao bounds for quantum state estimation," IEICE Technical Report, **IT89–42**,9–14(1989).

[12] H. Nagaoka, "On Fisher information of quantum statistical models," SITA '87, 241–246 (1987) (in Japanese).

[13] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry, I, II* (John Wiley, New York, 1963, 1969).

[14] H. Nagaoka, private communication.

[15] A. Fujiwara, "One-parameter pure state estimation based on the symmetric logarithmic derivative," METR **94-8**, Univ. Tokyo (1994).

[16] J. R. Klauder and E. C. G. Sudarshan, *Fundamentals of Quantum Optics* (Benjamin, New York, 1968).