

Geometry of Canonical Correlation on the State Space of a Quantum System

Dénes Petz *

Research Institute for Mathematical Sciences, Kyoto University
Kyoto 606-01, Japan

Abstract

A Riemannian metric is defined on the state space of a finite quantum system by the canonical correlation (or Kubo-Mori/Bogoliubov scalar product). This metric is infinitesimally induced by the (nonsymmetric) relative entropy functional or the von Neumann entropy of density matrices. Hence its geometry expresses maximal uncertainty. We prove that the metric is monotone under stochastic mappings, however an example shows that it is not the only such Riemannian metric. This fact is remarkable because in the probabilistic case, the Markovian monotonicity property characterizes the Fisher information-metric. Essential difference appears in the curvatures of a classical state space and a quantum one. We make a conjecture that the scalar curvature is monotone with respect to the “more mixed” (statistical) partial order of density matrices. Furthermore, we establish an information inequality resembling the Cramér-Rao inequality of classical statistics. The inequality provides a lower bound for the canonical correlation-matrix (of an unbiased observable) and it is saturated when a (partial) observation level and the corresponding family of coarse-grained states are considered.

* On leave from the Technical University of Budapest.
E-mail: h1128pet@huella.binet

1. Introduction

Two draw a picture of the background of this paper, we have to mention some facts both from statistics and mathematical physics. On the one hand, there has been a great interest recently in the use of differential geometry to provide insight into some problems of probability and statistics. The idea originates from the paper [1] and a standard reference is [2]. The basic aspects are contained in the easier-to-read few pages of [3] and the review [4] is suggested to the ambitious readers to get full information about the directions of current research in the field of inference and geometry. On the other hand, the idea of statistical manifold appeared in thermodynamics as well [5] and the paper [6] initiated the study of the particular Riemannian geometry which is the main subject of the present discussion. The idea is essentially to use the maximum entropy principle infinitesimally and to arrive at a Riemannian metric. In this way the maximum entropy principle may get a geometric interpretation, and the information geometric ideas have been applied to many-body systems [7].

Contrary to classical probability, quantum probability is used to describe the micro-world. (In fact, the room for quantum probability is larger, because there are macroscopic quantum systems.) In the Hilbert space formalism of quantum mechanics a pure state of a system is described by a state vector $|\Phi\rangle$ belonging to a complex Hilbert space \mathcal{H} . The mean (or expectation) value of an observable $A = A^* \in B(\mathcal{H})$ is the scalar product

$$\langle A \rangle = \langle \Phi | A | \Phi \rangle. \quad (1.1)$$

The statistical operator (or density matrix) of a mixed state is a positive compact operator with trace 1. We denote by \mathfrak{S} the set of all such operators acting on the basic Hilbert space \mathcal{H} . Sometimes \mathfrak{S} is called state space. In a mixed state $D \in \mathfrak{S}$ the mean value of the observable A is

$$\langle A \rangle = \text{Tr } DA \quad (1.2)$$

which extends formula (1.1).

In the present paper, we mostly assume that the underlying complex Hilbert space \mathcal{H} is finite dimensional and we write n for its dimension. So operators on \mathcal{H} may be represented by matrices and a statistical operator corresponds to a positive semidefinite matrix of trace 1. Physically, a finite dimensional Hilbert space appears, for example, when one deals with a spin. Our aim is to endow the state space with a differentiable structure. Actually, we are going to study a Riemannian geometry on the space \mathcal{S} of all invertible density matrices.

Statistical and probabilistic ideas often find their ways to quantum mechanics, however, they have to be adapted to density matrices. Formula (1.2) of the expectation value and the von Neumann entropy of a density matrix are the first successful adaptations. The Riemannian metric we deal with is given by the Hessian matrix of the von Neumann

entropy, hence it is strongly related to entropy maximization [8]. Its study is motivated by quantum statistical mechanics (in particular, linear response theory) as well as quantum extension of the differential geometric approach to classical probability and classical statistics. The term “canonical correlation” is borrowed from linear response theory [9].

The paper is organized as follows. In Section 2 we show how to construct different information quantities by mean of an operator monotone function. Quasi-entropies of [10] and [11] are summarized here but the important special case of α -divergences is spelled out. Concerning the details of quasi-entropies, [12] is a reference.

In Section 3 the Kubo-Mori product appears as Riemannian metric induced by the relative entropy functional but one can reach it from the von Neumann entropy, too. The exponential and affine parameterizations of the state space are explained. The metric is explicitly given in both coordinate systems.

The main result of Section 4 is the monotonicity of the metric under stochastic mappings. Note that monotonicity determines Fisher’s information in classical probability. An example of another monotone metric is given, so uniqueness does not hold in the quantum case.

The topic of Section 5 is an analogue of the Pythagorean theorem concerning a triangle determined by orthogonal exponential and mixture geodesics. Possibility of extension to arbitrary von Neumann algebras is commented.

In Section 6 sectional, Ricci, and scalar curvatures are computed, mainly at the tracial state where the Ricci curvature is found to be constant. It is explained that the shape of the state space of a spin is very different from the geometry of the simplex of all probabilities of a finite set. It is conjectured that the scalar curvature is monotone with respect to the “more mixed” partial order of density matrices and, in particular, it is maximal at the tracial trace.

In Section 7 the Cramér-Rao pattern is followed and a lower bound is obtained for the canonical correlation-matrix of an unbiased estimator. The inequality is saturated by an observation level with respect to the coarsened states. The metric of an observation level is subject of Section 8. In this setting the Hilbert space can be infinite dimensional (or an operator algebra may be taken) and a class of states is endowed with Riemannian structure.

2. Some information quantities

In this section we recall how to construct a garden of information quantities for density matrices D_1 and D_2 , from a function $f : \mathbb{R}^+ \rightarrow \mathbb{R}$, and from an operator K . (For simplicity only invertible densities are considered below.) Before entering that subject, we fix a convention. Linear transformations of the space of operators will be called superoperators as it is rather frequent in physics literature. Superoperators will be denoted by bold face or Greek. We often use the Hilbert-Schmidt inner product on the space of matrices:

$$\langle A, B \rangle_{\text{HS}} = \text{Tr } A^* B.$$

For the densities D_1 and D_2 , the relative modular (super)operator is defined as

$$\Delta(D_2, D_1) : a \mapsto D_2 a D_1^{-1}.$$

The quantity

$$S_f^K(D_1, D_2) = \langle K D_1^{1/2}, f(\Delta(D_2, D_1)) K D_1^{1/2} \rangle_{\text{HS}} \quad (2.1)$$

was called quasi-entropy in [10], [11]. Physically, most of the functions f are irrelevant but some show up in the literature. Set

$$f_\alpha(t) = \begin{cases} \frac{4}{1-\alpha^2} (1 - t^{\frac{1+\alpha}{2}}) & \text{if } \alpha \neq \pm 1, \\ t \log t & \text{for } \alpha = 1, \\ -\log t & \text{for } \alpha = -1. \end{cases} \quad (2.2)$$

By the choice of this f_α , (2.1) supplies the following.

$$D_\alpha^K(D_1, D_2) = \begin{cases} \frac{4}{1-\alpha^2} \text{Tr } D_1 K^* (K - D_2^{\frac{1+\alpha}{2}} K D_1^{-\frac{1+\alpha}{2}}) & \text{if } \alpha \neq \pm 1, \\ \text{Tr } K^* D_2 (\log D_2 K - K \log D_1) & \text{if } \alpha = 1, \\ \text{Tr } D_1 K^* (K \log D_1 - \log D_2 K) & \text{if } \alpha = -1. \end{cases} \quad (2.3)$$

One may now decrease the number of parameters by taking $K = I$. So the α -divergences are obtained:

$$D_\alpha(D_1, D_2) = \frac{4}{1-\alpha^2} \text{Tr} (I - D_2^{\frac{1+\alpha}{2}} D_1^{-\frac{1+\alpha}{2}}) D_1 \quad (\alpha \neq \pm 1) \quad (2.4)$$

and the limit $\alpha \rightarrow -1$ yields Umegaki's relative entropy

$$S(D_1, D_2) = \text{Tr } D_1 (\log D_1 - \log D_2). \quad (2.5)$$

On the other hand, keeping K but choosing $D_1 = D_2 = D$, we get to the skew information

$$I_p(D, K) = \operatorname{Tr} DK^*K - \operatorname{Tr} D^{1-p}K^*D^pK \quad (2.6)$$

(due to Wigner, Yanase and Dyson, [13]) apart from a normalization factor and with parameter $p = (1 - \alpha)/2$ (see also [14]).

The quasi-entropies (2.1) are useful when $-f$ is operator monotone and then they are regarded as information distances. They show analogy with some statistical distances, cf. [15]. We do not enter the details of the general theory and operator algebraic extension.

Theorem 2.1. If $-f$ is operator monotone, then $S_f^k(D_1, D_2)$ is a convex function of (D_1, D_2) and $S_f^I(D_1, D_2) \geq f(1)$. Moreover, if f is not linear, then $S_f^I(D_1, D_2) = f(1)$ implies that $D_1 = D_2$.

A superoperator $\gamma : B(\mathcal{H}) \rightarrow B(\mathcal{H})$ will be called stochastic if it is unital and completely positive. (The later condition implies that the Schwarz inequality $\gamma(X)^*\gamma(X) \leq \gamma(X^*X)$ holds for every X .) The dual γ^* of a stochastic mapping γ with respect to the Hilbert-Schmidt inner product is completely positive and trace preserving. In particular, γ^* sends a density matrix into a density matrix.

Theorem 2.2. Let $f : [0, \infty) \rightarrow \mathbb{R}$ be an operator monotone function with $f(0) \geq 0$. If γ is a stochastic mapping, then for every $K \in B(\mathcal{H})$ and for densities D_1, D_2 we have

$$S_f^K(\gamma^*(D_1), \gamma^*(D_2)) \geq S_f^{\gamma(K)}(D_1, D_2).$$

These theorems apply to f_α if $-1 < \alpha < 1$. Concerning the proof and similar general results we refer to [10], [11]. (Two decades ago convexity of (2.6) received a lot of attention, and it should be cited as Lieb's convexity [16, 12].)

Stochastic mappings are the appropriate ‘‘morphisms’’ between C*-algebras. In particular, the dynamical evolution of a quantum system is described by a semigroup of stochastic mappings, and measurements correspond to positive unital mappings as well. The content of Theorem 2.2 may be explained as follows. The information distances are decreasing under the transformation γ^* of states associated with a stochastic mapping γ . All reasonable information distances must have this monotonicity property. (In some sense, stochastic mapping is the quantum analogue of a Markov kernel.)

3. The Riemannian metric of the canonical correlation

The state space \mathfrak{S} is easily visualized for spin $1/2$. In this case \mathfrak{S} is the space of all 2×2 density matrices. They are in the form

$$\frac{1}{2} \begin{pmatrix} 1+x & y+iz \\ y-iz & 1-x \end{pmatrix},$$

where x, y, z are real parameters and $x^2 + y^2 + z^2 \leq 1$. (This is the well-known polarization representation of a spin $1/2$.) Hence \mathfrak{S} is regarded as the unit ball in the Euclidean 3-space. The extreme boundary, i.e. the set of all extremal points, coincides with the topological boundary. This simple description does not generalize to higher dimension, and in fact, the spin $1/2$ does not show several features typical for the the general (quantum) case. For example in higher dimension, the extreme boundary will be much smaller than the topological one (cf. [17]).

Let \mathcal{S} be the set of all invertible density matrices acting on the (n -dimensional complex) Hilbert space \mathcal{H} . (When the dependence on n will be emphasized we use the notation \mathcal{S}_n .) Motivated by quantum statistical mechanics, we write elements of \mathcal{S} in the form

$$R_\beta(H) = \frac{e^{-\beta H}}{\text{Tr } e^{-\beta H}}, \quad (3.1)$$

where $\beta > 0$ is a real constant and H is a self-adjoint operator. ($R_\beta(H)$ is the density of the so-called Gibbs state at the inverse temperature β for the Hamiltonian H .) If the difference $H - H'$ is a multiple of the identity then H and H' give the same density matrix in (3.1). Therefore, to get a suitable parameterization we have to identify self-adjoint operators differing in a constant multiple of the identity I . For example, we may consider only traceless self-adjoint matrices, their space is denoted by \mathcal{K}_0^0 . The real linear space \mathcal{K}_0^0 has dimension $n^2 - 1$ and will be identified with the Euclidean space \mathbb{R}^{n^2-1} by means of the following linearly independent matrices.

$$\begin{aligned} \sigma_1^{kl} &= \frac{1}{\sqrt{2}}(E_{kl} + E_{lk}) & (1 \leq k < l \leq n), \\ \sigma_2^{kl} &= \frac{1}{\sqrt{2}}(-iE_{kl} + iE_{lk}) & (1 \leq k < l \leq n), \\ \sigma_3^m &= \frac{1}{\sqrt{m^2 + m}} \sum_{i=1}^m E_{ii} - \frac{m}{\sqrt{m^2 + m}} E_{m+1, m+1} & (1 \leq m \leq n-1), \end{aligned} \quad (3.2)$$

where (E_{ij}) is a system of matrix units. Observe that for $n = 2$ exactly the Pauli matrices show up here. (Of course, one may choose other basis matrices; this choice is convenient because the matrices (3.2) are orthonormal with respect to the Hilbert-Schmidt scalar product.) The homeomorphism

$$h : D \mapsto -\frac{1}{\beta} \log D + \frac{1}{\beta} \text{Tr } \log D \quad (3.3)$$

maps onto $\mathcal{K}_0^0 \equiv \mathbb{R}^{n^2-1}$ and endows \mathcal{S} with a differentiable structure. The mapping h is called the logarithmic coordinate system at the inverse temperature β . \mathcal{S} becomes a differentiable manifold with an atlas containing a single chart. Differentiation of a function $f : \mathcal{S} \rightarrow \mathbb{R}$ along the curve

$$t \mapsto D_t = \frac{e^{-\beta(H+tA)}}{\text{Tr} e^{-\beta(H+tA)}} \equiv R_\beta(H+tA) \quad (A, H \in \mathcal{K}_0^0) \quad (3.4)$$

is the same as differentiation of $f \circ h^{-1}$ in the direction A . The curves like (2.4) will be called e -geodesic. They correspond to straight lines in the parameter space when the exponential parameterization of the state space is considered.

Differentiating the relative entropy (2.5) we have

$$\begin{aligned} & - \left. \frac{\partial^2}{\partial t \partial u} \right|_{t=u=0} \frac{1}{\beta^2} S(R_\beta(H+tA), R_\beta(H+uB)) \\ & = \frac{1}{\beta} \int_0^\beta \text{Tr} (R(H) e^{xH} A e^{-xH} B) dx - \text{Tr} R(H) A \text{Tr} R(H) B, \end{aligned} \quad (3.5)$$

which is often called the canonical correlation of A and B . For the rest we fix $\beta = -1$. The scalar product

$$\langle A, B \rangle_H = \int_0^1 \text{Tr} R(H) e^{xH} A^* e^{-xH} B dx. \quad (3.6)$$

is an important ingredient of linear response theory and bears the names Kubo-Mori scalar product, canonical correlation, Bogoliubov inner product or Duhamel two-point function ([18, 9, 19, 20]). We choose it for the Riemannian metric.

We identify the tangent space $\mathbf{T}_H(\mathcal{S})$ of the manifold \mathcal{S} at the density matrix $R(H)$ by the space of all fluctuations

$$\mathbf{T}_H(\mathcal{S}) = \{A = A^* : \text{Tr} R(H) A = 0\} \quad (3.7)$$

and the Riemannian metric g is given by the Kubo-Mori product. If $A = A^*$ is an observable, then $\tilde{A} \equiv A - (\text{Tr} R(H) A) I$ is its fluctuation around the mean value $\text{Tr} R(H) A$. So we use the tilde when we want to stress that a self-adjoint matrix is a fluctuation at some density $R(H)$. Accordingly,

$$g(\tilde{A}, \tilde{B})(H) = \langle \tilde{A}, \tilde{B} \rangle_H \quad (\tilde{A}, \tilde{B} \in \mathbf{T}_H(\mathcal{S})). \quad (3.8)$$

Note that g is also obtained by differentiations of the generalized partition function:

$$\left. \frac{\partial^2}{\partial t \partial u} \right|_{t=u=0} \text{Tr} e^{H+t\tilde{A}+s\tilde{B}} \quad (\tilde{A}, \tilde{B} \in \mathbf{T}_H(\mathcal{S})). \quad (3.9)$$

Density matrices may be parameterized directly by their entries. More precisely, $D \mapsto D - I/n$ embeds \mathcal{S}_n into $\mathcal{K}_0^0 \equiv \mathbb{R}^{n^2-1}$ as a convex open subset and \mathcal{S}_n inherits a differentiable structure. This affine parameterization is equally natural (if not more natural, cf. the beginning of this section). The segment

$$D_t = tD_1 + (1-t)D_0 \quad (t \in [0, 1]) \quad (3.10)$$

joining $D_0, D_1 \in \mathcal{S}$ is called m -geodesic. (This terminology is from [2] and m comes from the word mixture.) By means of the identity

$$\frac{\partial}{\partial t} \log(L + tK) \Big|_{t=0} = \int_0^\infty (L + s)^{-1} K (L + s)^{-1} ds \quad (3.11)$$

one can compute

$$-\frac{\partial^2}{\partial t \partial u} \Big|_{t=u=0} S(D + tA, D + uB) = \int_0^\infty \text{Tr} A(D + s)^{-1} B(D + s)^{-1} ds. \quad (3.12)$$

This shows that in the affine parameterization the tangent space $\mathbf{T}_D(\mathcal{S})$ is regarded as the set \mathcal{K}_0^0 of all traceless self-adjoint matrices and the Riemannian metric is in the form

$$g(A, B)(D) = \int_0^\infty \text{Tr} A(D + s)^{-1} B(D + s)^{-1} ds \quad (A, B \in \mathcal{K}_0^0). \quad (3.13)$$

Above the Riemannian metric was deduced from the relative entropy functional (2.5) (cf. [6], [19]). It is worthwhile to note that one can arrive at the same metric starting from the von Neumann entropy $S(D) = -\text{Tr} D \log D$ (cf. [7]). Indeed,

$$g(A, B)(D) = -\frac{\partial^2}{\partial t \partial u} \Big|_{t=u=0} S(D + tA + uB) \quad (A, B \in \mathcal{K}_0^0). \quad (3.14)$$

There are some other Riemannian metrics occurring in the literature. The metric

$$g_s(\tilde{A}, \tilde{B})(H) = \frac{1}{2} \text{Tr} R(H)(\tilde{A}\tilde{B} + \tilde{B}\tilde{A}) \quad (3.15)$$

has the form of a Hessian, too:

$$g_s(\tilde{A}, \tilde{B})(H) = \frac{\partial^2}{\partial t \partial u} \Big|_{t=u=0} \text{Tr} R(H + t\tilde{A})(\log R(H + t\tilde{A}) - \log R(H + s\tilde{B}))^2, \quad (3.16)$$

see [21]. It is noteworthy that

$$g(\tilde{A}, \tilde{A})(H) \leq g_s(\tilde{A}, \tilde{A})(H) \quad (\tilde{A} \in \mathbf{T}_H(\mathcal{S})) \quad (3.17)$$

due to the Bogoliubov inequality [18] and the right-hand side of (3.16) is a quasi-entropy (corresponding to the function $t \mapsto (\log t)^2$.) Another metric comes from the Bures distance [22]. Both metrics do not seem to be in a direct relation to the subject of this paper.

4. Monotonicity of the metric

Every stochastic mapping γ induces a transformation of \mathfrak{S} . Indeed, the restriction of γ^* to \mathfrak{S} . When $\gamma^*(\mathcal{S}) \subset \mathcal{S}$, then γ^* is a diffeomorphism of \mathcal{S} and its differential $(\gamma^*)_\star$ in the affine coordinate system is the restriction of γ^* to the traceless self-adjoint matrices. The next results is the monotonicity of the Riemannian metric g under stochastic mappings.

Theorem 4.1. Let γ be a stochastic mapping and $D \in \mathcal{S}$. If $\gamma^*(D) \in \mathcal{S}$ then γ^* is a diffeomorphism in a neighborhood of D and its differential is a contraction with respect to the Riemannian metric g .

Proof : The differential $(\gamma^*)_\star$ has simple form in affine coordinates. As it was noted above, at the point $D \in \mathcal{S}$ $(\gamma^*)_\star : \mathcal{K}_0^0 \rightarrow \mathcal{K}_0^0$ is the restriction of γ^* . So we have to show the inequality

$$\begin{aligned} & \int_0^\infty \text{Tr} (\gamma^*(D) + s)^{-1} \gamma^*(K) (\gamma^*(D) + s)^{-1} \gamma^*(K) ds \\ & \leq \int_0^\infty \text{Tr} (D + s)^{-1} K (D + s)^{-1} K ds \end{aligned}$$

for every $K \in \mathcal{K}_0^0$. By means of the superoperator (5.2) our claim is equivalent to

$$\gamma \mathbf{T}_{\gamma^*(D)} \gamma^* \leq \mathbf{T}_D. \quad (4.1)$$

Theorem 2.2 yields that

$$\int_0^1 \text{Tr} \gamma^*(D)^{1-t} K \gamma^*(D)^t K dt \geq \int_0^1 \text{Tr} D^{1-t} \gamma(K) D^t \gamma(K) dt. \quad (4.2)$$

Introducing the superoperator

$$\mathbf{T}_Q^{-1} : K \mapsto \int_0^1 Q^{1-t} K Q^t dt, \quad (4.3)$$

(which is the inverse of \mathbf{T}_Q from (5.2)), we rewrite (4.2) as

$$\mathbf{T}_{\gamma^*(D)}^{-1} \geq \gamma^* \mathbf{T}_D^{-1} \gamma. \quad (4.4)$$

Since (4.1) and (4.4) are equivalent, the proof is complete.

Note that in logarithmic coordinates $(\gamma^*)_\star(\tilde{A}) = \mathbf{T}_{\gamma^*(D)} \circ \gamma^* \circ \mathbf{T}_D^{-1}(\tilde{A})$. \square

The monotonicity of the Riemannian metric is crucial when one likes to imitate the geometrical approach of [23]. An infinitesimal statistical distance has to be monotone

under stochastic mappings. We note that the monotonicity of g is a strengthening of the concavity of the von Neumann entropy. Indeed, positive definiteness of g is equivalent to the strict concavity of the von Neumann entropy due to (3.2) and monotonicity is much more than positivity.

In the probabilistic case the monotone metric is unique and is identical to Fisher's information. In the quantum case the uniqueness is certainly not true. Set a metric $g_{\mathbf{r}}$ as

$$g_{\mathbf{r}}(A, B)(D) = \frac{1}{2} \text{Tr} D^{-1}(AB + BA) \quad (A, B \in \mathcal{K}_0^0). \quad (4.5)$$

Under the conditions and with the notation of Theorem 3.1, monotonicity of $g_{\mathbf{r}}$ is the inequality

$$\text{Tr} \gamma^*(D)^{-1} \gamma^*(A)^2 \leq \text{Tr} D^{-1} A^2 \quad (A \in \mathcal{K}_0^0) \quad (4.6)$$

which is a consequence of the operator inequality

$$\gamma^*(K) \gamma^*(T)^{-1} \gamma^*(K)^* \leq \gamma^*(KT^{-1}K^*), \quad (4.7)$$

due to Choi [24], for positive invertible T . Indeed, taking the trace of both sides of (4.7), we conclude (4.6) if $K = A$ because $\text{Tr} \circ \gamma^* = \text{Tr}$. Observe that for commuting D and A we have $g(A, A)(D) = g_{\mathbf{r}}(A, A)(D)$ in accordance with the uniqueness of the monotone metric in the commutative case. The example shows that in the quantum case the monotone metric is not unique.

The question arises naturally: What is the class of the monotone Riemannian metrics on matrix algebras? The paper [25] attempted to describe this class but the conclusion of its analysis is not clear to the present author. Let us summarize our understanding of [25]. Let \tilde{g} be first a unitarily invariant metric whose matrix is to be described. Then it is enough to consider points where D is diagonal, say $D = \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ and we take the basis (3.2). It is established in [25] first that up to a constant factor the metric is of the following kind.

$$\begin{aligned} \tilde{g}(\sigma_3^m, \sigma_3^l)(D) &= \text{Tr} D^{-1} \sigma_3^m \sigma_3^l, \\ \tilde{g}(\sigma_1^{kl}, \sigma_2^{k'l'})(D) &= 0, \\ \tilde{g}(\sigma_1^{kl}, \sigma_1^{kl})(D) &= \tilde{g}(\sigma_2^{kl}, \sigma_2^{kl}), \\ \tilde{g}(\sigma_1^{kl}, \sigma_1^{kl})(D) &= c(\lambda_k, \lambda_l), \text{ where } c \text{ is a real function.} \end{aligned}$$

Therefore, the prescription of a unitarily invariant metric is done by giving the function c . Then the authors of [25] assume that \tilde{g} is monotone and possesses also a certain property concerning tensor product, and they show that the function c of such a metric satisfies a functional equation. They find 8 solutions of the equation and they arrive at 8 corresponding metrics. However, it is not stated in [25] that those metrics are monotone or only those functions satisfy the functional equation. In any case, g and $g_{\mathbf{r}}$ are on the list as items 3 and 8. In our opinion, it would be worthwhile to complete the analysis of [25] but in this paper we do not attack the problem of characterization of monotone metrics. Our aim was only to expose the state of affair.

5. The Pythagorean theorem

Let P , Q , and R be three points in \mathcal{S} . The e -geodesic connecting Q and R is the curve

$$\gamma_{QR}^e(t) = \frac{\exp(H + tA)}{\text{Tr} \exp(H + tA)} \quad (t \in [0, 1]),$$

where $H = \log Q$ and $A = \log R - \log Q$. Then $\gamma_{QR}^e(0) = Q$, $\gamma_{QR}^e(1) = R$ and $\dot{\gamma}_{QR}^e(0) = A$. The m -geodesic connecting Q and P is the curve

$$\gamma_{QP}^m(t) = (1 - t)Q + tP = Q + t(P - Q) \quad (t \in [0, 1])$$

which has a tangent at $t=0$

$$\dot{\gamma}_{QP}^m(0) = \left. \frac{\partial}{\partial t} \log(Q + t(Q - P)) \right|_{t=0} = \int_0^\infty (Q + s)^{-1} (P - Q) (Q + s)^{-1} ds.$$

The following result is called the (matricial) analogue of the Pythagorean theorem (in the sense of [2]).

$$S(P, Q) + S(Q, R) = S(P, R) + \langle \dot{\gamma}_{QR}^e(0), \dot{\gamma}_{QP}^m(0) \rangle_Q \quad (5.1)$$

In particular, we have

Theorem 5.1 If the e -geodesic connecting Q and R is orthogonal to the m -geodesic connecting Q and P at the point Q , then

$$S(P, Q) + S(Q, R) = S(P, R).$$

Proof : Plain computation yields

$$S(P, Q) + S(Q, R) - S(P, R) = \text{Tr} A(P - Q) = \langle A, P - Q \rangle_{\text{HS}},$$

where HS indicates the Hilbert-Schmidt inner product. For the superoperator

$$\mathbf{T}_Q : X \mapsto \int_0^\infty (Q + s)^{-1} X (Q + s)^{-1} ds \quad (5.2)$$

we have

$$\langle X, Y \rangle_{\text{HS}} = \langle \mathbf{T}_Q(X), Y \rangle_Q. \quad (5.3)$$

Therefore,

$$\langle A, P - Q \rangle_{\text{HS}} = \langle \mathbf{T}_Q(A), P - Q \rangle_Q = \langle \dot{\gamma}_{QR}^e(0), \dot{\gamma}_{QP}^m(0) \rangle_Q$$

and this completes the proof of (4.1) and that of the theorem. \square

The rest of this section is devoted to the extension of Theorem 5.1 to the context of von Neumann algebras. Since this part of the paper is rather independent of the differential geometric main line, we use the operator algebraic formalism of relative entropy developed in [26] freely. We refer to the monograph [12] concerning details. The discussion below shows that in the background of (5.1) one can meet the duality between states and relative Hamiltonians [29].

Let \mathcal{M} be a von Neumann algebra and ω, ψ, φ faithful normal states. For the sake of simplicity, we assume that $\lambda^{-1}\psi \leq \varphi \leq \lambda\psi$ for some positive $\lambda \in \mathbb{R}$. This majorization condition guarantees the existence of a relative Hamiltonian, that is, there is $h = h^* \in \mathcal{M}$ such that ψ is the minimizer of the functional

$$\nu \mapsto S(\nu, \varphi) + \nu(h),$$

in notation $\psi = [\varphi^h]$ [30]. Suppose, further more, that the relative entropy $S(\omega, \varphi)$ is finite. Then it can be obtained by differentiation of the Connes' cocycle $[D\omega, D\varphi]_t$ as

$$\begin{aligned} S(\omega, \varphi) &= i \frac{\partial}{\partial t} \Big|_{t=0} \omega([D\omega, D\varphi]_t) \\ &= i \frac{\partial}{\partial t} \Big|_{t=0} \omega([D\omega, D\psi]_t [D\psi, D\varphi]_t) \end{aligned}$$

due to the chain rule. Performing the differentiation we get

$$S(\omega, \varphi) = S(\omega, \psi) + \omega(h)$$

which together with $S(\psi, \varphi) = \psi(h)$ yields

$$S(\omega, \varphi) = S(\omega, \psi) + S(\psi, \varphi) + (\omega - \psi)(h) \tag{5.4}$$

and this corresponds to (5.1). When the orthogonality condition $(\omega - \psi)(h) = 0$ is fulfilled, we have the analogue of the Pythagorean theorem.

6. Curvature

In this section curvatures of \mathcal{S} will be discussed on the basis of [31]. The starting point is the Levi-Civita connection ∇ induced by the metric. (What we need here from differential geometry is standard, and we follow the book [32].) The Kostant formula says that

$$2g(Y, \nabla_Z X) = Xg(Y, Z) + Zg(X, Y) - Yg(X, Z) - g([X, Y], Z) - g([Z, Y], X) - g([X, Z], Y). \quad (6.1)$$

When the Lie brackets $[X, Y]$, $[Y, Z]$, $[Z, X]$ vanish, as it happens when X, Y, Z are coordinate fields, (6.1) becomes simpler. Since we want to have the convenience of vanishing Lie brackets, we restrict our discussion to linear combinations of coordinate fields (with constant coefficients). A tangent vector can uniquely extend to a vector field which is the linear combination of coordinate fields with constant coefficients. From now on, every vector field, usually denoted by X, Y, Z , is supposed to be a linear combinations of coordinate fields.

If $X(H) = \tilde{A}$, $Y(H) = \tilde{B}$, and $Z(H) = \tilde{C}$, then we understand $\tilde{A}\langle\tilde{B}, \tilde{C}\rangle$ as $X\langle Y, Z\rangle(H)$. Assuming that $\text{Tr} e^{-H} = 1$, we infer

$$\begin{aligned} \tilde{A}\langle\tilde{B}, \tilde{C}\rangle &= \int_{x=0}^1 \int_{v=0}^x \text{Tr} e^{-(1-x)H} \tilde{A} e^{-(x-v)H} \tilde{C} e^{-vH} \tilde{B} dv dx \\ &+ \int_{x=0}^1 \int_{v=0}^x \text{Tr} e^{-(x-v)H} \tilde{C} e^{-vH} \tilde{A} e^{-(1-x)H} \tilde{B} dv dx \end{aligned} \quad (6.2)$$

from the perturbation series of the exponential function. Observe that the two terms on the right-hand side are similar, exchanging \tilde{A} with \tilde{B} one gets from each term the other one. Moreover, both terms are invariant under cyclic permutations of \tilde{A}, \tilde{B} and \tilde{C} . Hence the full symmetry of their sum is follows. Due to the symmetry, we have

$$\langle X, \nabla_Y Z\rangle(H) = \frac{1}{2} \tilde{A}\langle\tilde{B}, \tilde{C}\rangle. \quad (6.3)$$

We are interested in the sectional curvatures of \mathcal{S} . Let $\tilde{A}, \tilde{B} \in \mathbf{T}_H(\mathcal{S})$ and let X, Y be vector fields such that $X(H) = \tilde{A}$ and $Y(H) = \tilde{B}$. The sectional curvature for the plane spanned by \tilde{A} and \tilde{B} is

$$K(\tilde{A}, \tilde{B}) = \frac{\langle \nabla_Y \nabla_X X, Y \rangle - \langle \nabla_X \nabla_Y X, Y \rangle}{\langle X, X \rangle \langle Y, Y \rangle - \langle X, Y \rangle^2} \quad (6.4)$$

by definition. Since

$$\begin{aligned} \langle \nabla_Y \nabla_X X, Y \rangle &= Y\langle \nabla_X X, Y \rangle - \langle \nabla_X X, \nabla_Y Y \rangle \\ \langle \nabla_X \nabla_Y X, Y \rangle &= X\langle \nabla_Y X, Y \rangle - \langle \nabla_Y X, \nabla_X Y \rangle \end{aligned}$$

and

$$Y\langle\nabla_X X, Y\rangle = \frac{1}{2}YX\langle X, Y\rangle = \frac{1}{2}XY\langle X, Y\rangle = X\langle\nabla_Y X, Y\rangle$$

due to $[X, Y] = 0$ and (6.3), we have

$$K(\tilde{A}, \tilde{B}) = \frac{\langle\nabla_Y X, \nabla_X Y\rangle - \langle\nabla_X X, \nabla_Y Y\rangle}{\langle X, X\rangle\langle Y, Y\rangle - \langle X, Y\rangle^2}. \quad (6.5)$$

If $\tilde{A}H = H\tilde{A}$ and $\tilde{B}H = H\tilde{B}$, then from (6.3)

$$\tilde{A}\langle\tilde{B}, \tilde{C}\rangle = \frac{1}{2}\langle\tilde{A}\tilde{B} + \tilde{B}\tilde{A}, \tilde{C}\rangle,$$

where ordinary matrix multiplications stand on the right-hand side. We conclude that under the above commutations

$$\nabla_X Y = \frac{1}{4}(\tilde{A}\tilde{B} + \tilde{B}\tilde{A}) - \frac{1}{4}\text{Tr} R(H)(\tilde{A}\tilde{B} + \tilde{B}\tilde{A}) \quad (6.6)$$

and we arrive at the following

Theorem 6.1. Let $\tilde{A}, \tilde{B} \in \mathbf{T}_H(\mathcal{S})$ be orthonormal. If $\tilde{A}H = H\tilde{A}$ and $\tilde{B}H = H\tilde{B}$, then

$$K(\tilde{A}, \tilde{B})(H) = \frac{4 + \text{Tr}(R(H)(\tilde{A}\tilde{B} - \tilde{B}\tilde{A})^2)}{16}.$$

Since $(\tilde{A}\tilde{B} - \tilde{B}\tilde{A})^2 \leq 0$, the sectional curvatures do not exceed $1/4$ at the tracial state $H = 0$. Positive sectional curvature will appear if $n \geq 3$ because then $\tilde{A}\tilde{B} - \tilde{B}\tilde{A} = 0$ may happen. It seems that the 2×2 case is exceptional from the point of view of curvature. Recall that for $n = 2$ all sectional curvatures are strictly negative except for the tracial state where they vanish [19].

Next we compute the Ricci and scalar curvatures at the tracial state τ . Let $\tilde{A}_1, \tilde{A}_2, \dots, \tilde{A}_{n^2-1}$ be an orthonormal basis in $\mathbf{T}_\tau(\mathcal{S})$ and write \tilde{A} for \tilde{A}_1 . The Ricci curvature of the direction \tilde{A} is expressed by means of sectional curvatures as

$$\text{Ric}(\tilde{A}, \tilde{A}) = \sum_{i=2}^{n^2-1} K(\tilde{A}, \tilde{A}_i). \quad (6.7)$$

We shall see that the Ricci curvature is independent of \tilde{A} (at the tracial state). The reason for this fact is the invariance of the tracial state under unitary transformations. To compute

$$\sum_{i=2}^{n^2-1} K(\tilde{A}, \tilde{A}_i) = \frac{n^2 - 2}{4} - \frac{1}{16} \sum_{i=2}^{n^2-1} 2\tau(\tilde{A}^2 \tilde{A}_i^2 - \tilde{A} \tilde{A}_i \tilde{A} \tilde{A}_i)$$

we use the Liouville operator $\mathbf{L} : K \mapsto \tilde{A}K - K\tilde{A}$ and the trace functional \mathbf{Tr} of superoperators. We have $\langle \mathbf{L}^2(\tilde{A}_i), \tilde{A}_i \rangle = 2\tau(\tilde{A}^2\tilde{A}_i^2 - \tilde{A}\tilde{A}_i\tilde{A}\tilde{A}_i)$ and

$$\sum_{i=2}^{n^2-1} \langle \mathbf{L}^2(\tilde{A}_i), \tilde{A}_i \rangle = \mathbf{Tr} \mathbf{L}^2 = 2n^2.$$

because $\mathbf{L}(I) = \mathbf{L}(\tilde{A}) = 0$. Hence the value of the Ricci curvature is

$$\text{Ric}(\tilde{A}, \tilde{A}) = \frac{n^2 - 4}{8}. \quad (6.8)$$

The scalar curvature is the sum of the sectional curvatures for all pairs of basis vectors and in this special case it equals to $n^2 - 1$ times the Ricci curvature.

Theorem 6.2. The scalar curvature of \mathcal{S}_n is given by the formula

$$\text{Scal}(0) = \frac{n^2 - 1}{8}(n^2 - 4)$$

at the tracial state corresponding to $H = 0$.

Hence the scalar curvature at the tracial state is positive for $n \geq 3$. It was conjectured in [31] that the scalar curvature of \mathcal{S}_n is maximal at the tracial trace (corresponding to $H = 0$). This is certainly true for $n = 2$ [19] and there are some numerical evidences for $n = 3$ and 4 [31]. For $n = 3$:

$R(H) = \text{Diag}(0.335, 0.333, 0.332)$	$\text{Scal}(H) = 4.9999$
$(0.339, 0.336, 0.325)$	4.99769
$(0.36, 0.34, 0.3)$	4.95935
$(0.35, 0.4, 0.25)$	4.72556
$(0.35, 0.5, 0.15)$	3.13543
$(0.5, 0.4, 0.1)$	1.16438
$(0.7, 0.2, 0.1)$	- 0.668265
$(0.99, 0.001, 0.009)$	- 568.425
$(0.999, 0.0001, 0.0009)$	- 5212.51
$(0.99999, 0.000001, 0.000009)$	- 459501

The paper [31] contains some results on the sectional curvatures, too.

When the (full) state spaces of a classical probabilistic and a quantum spin system are compared, the difference in curvature is easy to describe. While the classical state space has positive constant curvature, the scalar (or sectional) curvature of a quantum system

can be both positive and negative. This statement is independent of the validity of the conjecture. In the state space of an algebra with nontrivial center (i.e., in the presence of “superselection rules”), there are several tracial traces. By the above method, we can compute the scalar curvature at them. We find that the scalar curvature is constant on the set of tracial states.

Let us recall that a density D_1 is called more mixed than D_2 if there exists a trace preserving stochastic map α such that $D_1 = \alpha(D_2)$, in notation $D_1 \succ D_2$ (see [27], for example). We claim that the following conditions are equivalent.

- (i) The function $t \mapsto \text{Scal}(tH)$ is decreasing for $t \in \mathbb{R}^+$ and for selfadjoint H .
- (ii) If $R(H) \succ R(K)$ then $\text{Scal}(H) \geq \text{Scal}(K)$.

The implication (ii) \Rightarrow (i) follows from the fact that $R(t_2H) \succ R(t_1H)$ when $t_1 \leq t_2$. (A result of [33] should be quoted here, see also Prop. 3.18 in [12].) To see (i) \Rightarrow (ii), we take H and K such that $R(H) \succ R(K)$. We can find a sequence $H_1 \equiv H, H_2, \dots, H_m \equiv K$ such that the spectrums of each consecutive pair differ only in two eigenvalues and $R(H_k) \succ R(H_{k+1})$. It is straightforward that (i) yields

$$\text{Scal}(R(H_k)) \leq \text{Scal}(R(H_{k+1}))$$

and (ii) follows.

It is natural to guess that (i) and (ii) hold true.

Conjecture 6.3. The scalar curvature is monotone with respect to the “more mixed” partial ordering of density matrices.

The conjecture implies that the scalar curvature maximal at the tracial state because that is the most mixed. The parameter t in (i) is rather physical, the negative inverse temperature. Hence (i) says that the scalar curvature is an increasing function of the temperature. The conjecture is true in the 2-by-2 case, where the “more mixed” ordering is very simple. It was noted in [19] that the von Neumann entropy and the scalar curvature are monotone functions of each other.

In the paper [21], the Gaussian curvature was computed for the metric (3.15) in case of the quantum continuous ideal Boson and Fermion gases (two dimensional manifold) and physical meaning was attributed to the curvature.

7. An information inequality

In this section we follow the familiar Cramér-Rao pattern of mathematical statistics [34], [35]. Let (ψ_t) be a family of states of the quantum system \mathcal{M} and let the parameter set be an open subset of \mathbb{R}^m . By an estimator we mean an m -tuple (M_1, M_2, \dots, M_m) of observables, that is, self-adjoint elements of \mathcal{M} . The estimator M is called unbiased if

$$\psi_t(M_i) = t_i \quad (1 \leq i \leq m). \quad (7.1)$$

Since we are interested in the local behavior of the estimation, we assume that $t \mapsto \psi_t$ is sufficiently differentiable and the less restrictive condition

$$\left. \frac{\partial \psi_t(M_i)}{\partial t_j} \right|_{t=u} = \delta_{ij} \quad (1 \leq i, j \leq m) \quad (7.2)$$

will also work. If M satisfies (7.2) then it is called locally unbiased at s (being unbiased means locally unbiased at each point).

The Cramér-Rao pattern consists of finding a lower bound for the variance of the estimator. More precisely, we shall find lower bound for the canonical correlation-matrix

$$[[\langle M_i - \psi_t(M_i), M_j - \psi_t(M_j) \rangle_{\psi_t}]_{ij}] \equiv [[\langle M_i, M_j \rangle_{\psi_t} - \psi_t(M_i)\psi_t(M_j)]_{ij}] \quad (7.3)$$

of the estimator. The following concept of logarithmic derivative will be involved. If $t \mapsto \psi_t$ is a differentiable family of states then its logarithmic derivative (with respect to the canonical correlation and taken at s) is an m -tuple (L_1, L_2, \dots, L_m) of operators from \mathcal{M} such that

$$\left. \frac{\partial \psi_t(A)}{\partial t_i} \right|_{t=s} = \langle L_i, A \rangle_{\psi_s} - \psi_s(L_i)\psi_s(A) \quad (7.4)$$

for every $A \in \mathcal{M}$ and for $1 \leq i \leq m$.

Theorem 7.1. Let $M = (M_1, M_2, \dots, M_m)$ be a locally unbiased estimator for the family ψ_t of states at $t = s$. Assume that the logarithmic derivative $L = (L_1, L_2, \dots, L_m)$ exists. Then

$$[[\langle M_i - \psi_s(M_i), M_j - \psi_s(M_j) \rangle_{\psi_s}]_{ij}] \geq [[\langle L_i, L_j \rangle_{\psi_s}]_{ij}]^{-1}$$

in the sense that the difference is positive semidefinite.

Proof: The usual proof of the Cramér-Rao pattern may be followed which is linear algebra based on the relation

$$\langle M_i - \psi_s(M_i), L_j \rangle_{\psi_s} = \delta_{ij}.$$

This implies that the Gram matrices are in the stated relation (cf. [34], p. 276). \square

The information inequality of the previous theorem is tailored to the coarsened-grained states of an observation level, the subject of the next section.

It is in order to note the relation

$$[(\psi_s(M_i M_j) - \psi_s(M_i)\psi_s(M_j))_{ij}] \geq [(\langle M_i - \psi_s(M_i), M_j - \psi_s(M_j) \rangle_{\psi_s})_{ij}] \quad (7.5)$$

which is the consequence of the inequality $\langle K, K \rangle_\psi \leq \psi(K^2)$ for a self-adjoint K (see [18] or [19] for details of the Bogoliubov inequality). (7.5) is an inequality between the canonical correlation-matrix and the more standard covariance matrix. Hence the theorem yields a lower bound for the covariance matrix as well but this seems less useful. Note that the canonical correlation-matrix appeared already in [9].

Extension of the Cramér-Rao inequality was a reason to introduce Riemannian metrics on the state space. For example, the metric g_r of (4.5) corresponds to the so-called right logarithmic derivative, while g_s of (3.15) is the symmetrized correlation, see [34, 35, 19].

8. Geometry of an observation level

We start this section by a concise summary of some ideas of [7]. Let $H = (H_1, H_2, \dots, H_m)$ be an m -tuple of observables of a quantum system. We may think that those are the observables whose expectation values are supposed to be known from measurements. H means an “observation level”, dealing with only these observable we regard the collection (H_1, H_2, \dots, H_m) as “relevant observables”. The expectation values $t = (t_1, t_2, \dots, t_m)$ of the relevant observables yield an incomplete description of the actual state of the system, there are several states φ satisfying the constraints

$$\varphi(H_i) = t_i \text{ for } 1 \leq i \leq m, \text{ or shortly } \varphi(H) = t. \quad (8.1)$$

Since the von Neumann entropy S is an uncertainty measure, the state containing the least possible information about the irrelevant observables is the one which maximizes the von Neumann entropy under the constraint (8.1). This is the least biased prediction for the actual state of the total system. (What we follow here is the maximum entropy principle advocated by Jaynes and others, see [8, 36]). When states correspond to density matrices, the solution of the above extremum problem is among the generalized canonical statistical operators

$$R(\lambda H) \equiv \frac{\exp\left(\sum_{i=1}^m \lambda_i H_i\right)}{\text{Tr} \exp\left(\sum_{i=1}^m \lambda_i H_i\right)} \quad (\lambda \in \mathbb{R}^m), \quad (8.3)$$

cf. [37]. If $\text{Tr} R(\lambda H) H = t$ then $R(\lambda H)$ is the unique maximizer and we write this λ as $\lambda(t)$. Assuming that $R(\lambda(t) H)$ is defined when t is an open set $\mathcal{S} \subset \mathbb{R}^m$, we obtain a

parameterized family of states φ_t . φ_t is the reduced or coarse-grained state, its entropy $S(\varphi_t)$ is called the relevant entropy (for the observation level H). Now a Riemannian metric is supplied by the Hessian matrix of the relevant entropy function $t \mapsto S(\varphi_t)$:

$$g_{ij}(t) = -\frac{\partial^2 S(\varphi_t)}{\partial t_i \partial t_j} = \frac{\partial \lambda_j(t)}{\partial t_i}. \quad (8.4)$$

Concavity of the relevant entropy guarantees that (g_{ij}) is positive semidefinite. (Actually, it is definite, its inverse appears below.) This Riemannian metric defines the geometry of maximal uncertainty for the given level of observation. Next we explain how it is related to the content of Section 3.

We consider the case of complete observation: (H_1, H_2, \dots, H_m) will be the collection of matrices (3.2). There is only one density D_t which satisfies $\text{Tr } D_t H = t$, namely $D_t = I/n + tH$ and this must be the density of the “reduced” state φ_t . The correspondence $t \mapsto D_t$ what we have arrived at is just the affine parameterization of the state space. The relevant entropy coincides with the von Neumann entropy and (8.4) reduces to (3.14).

Now we have a look at the coarse-grained states φ_t from the viewpoint of estimation. Since $\psi_t(H_i) = t_i$ ($1 \leq i \leq m$), H is an unbiased estimator for the family ψ_t . What is the corresponding logarithmic derivative? Due to simple computation

$$\frac{\partial}{\partial t_i} \frac{\text{Tr} \exp A(\sum_{j=1}^m \lambda_j(t) H_j)}{\text{Tr} \exp(\sum_{j=1}^m \lambda_j(t) H_j)} = \sum_{j=1}^m \langle H_j - \psi_t(H_j), A \rangle_{\varphi_t} \frac{\partial \lambda_j(t)}{\partial t_i}$$

and we find that

$$L_i(t) = \sum_{j=1}^m (H_j - \psi_t(H_j)) \frac{\partial \lambda_j(t)}{\partial t_i}. \quad (8.5)$$

In this case the inequality of Theorem 7.1 becomes an equality. The matrix of the canonical correlation is inverse to g . From the point of view of statistics, the observation level is an efficient estimator for the family of coarse-grained states, which look like an exponential family.

The maximum entropy method is restricted to the case when (8.3) has a meaning, so, for example, all the observables H_i must have discrete spectrum. (This point is mostly overlooked in the literature.) When the relevant entropy is infinite, one has to have an “a priori” state to make prediction. (Even when the relevant entropy is finite, one can have a “bias”, for example, the state of the system known from a previous observation.) The relative entropy is invented for use in those cases when an a priori state is at disposal. Details are out of the scope of our present discussion but they will appear in a future publication. We note that the relative entropy principle is approached in an operator algebraic treatment in [38].

Acknowledgements. The author is grateful to Professor Shun-Ichi Amari for inviting him to the interdisciplinary workshop “Information and Geometry” and thanks to several participants for stimulating discussions. This work was partially supported by the Hungarian National Foundation for Scientific Research, grant No. 1900.

References

- [1] C.R. Rao, “Information and accuracy attainable in the estimation of statistical parameters”, Bull. Calcutta. Math. Soc. **37**(1945), pp. 81–91
- [2] S. Amari, *Differential-geometrical methods in statistics*, Lecture Notes in Stat. **28** (Springer, Berlin, Heidelberg, New York) 1985
- [3] L.L. Campbell, “The relation between information theory and differential geometry approach to statistics”, Inf. Sciences, **35**(1985), pp. 199–210
- [4] R.E. Kass, The geometry of asymptotic inference, Stat. Sci. **4**(1989), pp. 184–234
- [5] R.S. Ingarden, Y. Sato, K. Sugawa, M. Kawaguchi, “Information thermodynamics and differential geometry”, Tensor, **33**(1979). pp. 347–353
- [6] R.S. Ingarden, H. Janyszek, A. Kossakowski, T. Kawaguchi, “Information geometry of quantum statistical systems”, Tensor, **37**(1982). pp. 105–111
- [7] R. Balian, Y. Alhassid, H. Reinhardt, “Dissipation in many-body systems: A geometric approach based on information theory”, Phys. Rep. **131** (1986), 1–146
- [8] E.T. Jaynes, “Information theory and statistical mechanics. II”, Phys. Rev. **108**(1956), pp. 171–190
- [9] R. Kubo, The fluctuation-dissipation theorem, Rep. Prog. Phys. **29**(1966), 255-284
- [10] D. Petz, “Quasi-entropies for states of a von Neumann algebra”, Publ. RIMS Kyoto Univ. **23**(1985), pp. 787–800
- [11] D. Petz, “Quasi-entropies for finite quantum systems”, Rep. Math. Phys. **21** (1986), pp. 57–65
- [13] E.P. Wigner, M.M. Yanase, “Information content of distributions”, Proc. Nat. Acad. Sci. USA **49**(1963), pp. 910-918
- [14] H. Hasegawa, “ α -divergence of the non-commutative information geometry”, manuscript

- [15] I. Vajda, *Theory of Statistical Inference and Information* (Kluwer, Dordrecht) 1989
- [16] E.H. Lieb, “Convex trace functions and the Wigner-Yanase-Dyson conjecture”, *Advances in Math.* **11**(1973), pp. 267–288
- [17] M. Adelman, J.V. Corbett, C.A. Hurst, “The geometry of the state space”, *Found. Phys.* **23**(1993), pp. 211–223
- [18] O. Bratteli, D.W. Robinson, *Operator algebras and quantum statistical mechanics II*, (Springer, New York, Berlin, Heidelberg) 1981
- [19] D. Petz, G. Toth, “The Bogoliubov inner product in quantum statistics”, *Lett. Math. Phys.* **27**(1993), pp. 205–216
- [20] E. Fick, G. Sauer mann, *The quantum statistics of dynamic processes* (Springer, Berlin, Heidelberg) 1990
- [21] H. Janyszek, “On the geometrical structure of the generalized quantum Gibbs states”, *Rep. Math. Phys.* **24**(1986), pp. 11–19
- [22] A. Uhlmann, “The metric of Bures and the geometric phase”, in *Quantum Groups and Related Topics, Wroclaw, 1991*, pp. 267–274, eds. R. Gielerak et al. (Kluwer, Dordrecht) 1992
- [23] N.N. Čencov, *Statistical decision rules and optimal inferences* (Amer. Math. Society, Providence) 1982
- [24] M.D. Choi, “Some assorted inequalities for positive linear maps on C*-algebras”, *J. Operator Theory* **4**(1980), pp. 271–285
- [25] E.A. Morozova, N.N. Čencov, “Markovian invariant geometry on manifolds of states” (Russian), *Itogi Nauki i Tehniki* **36**(1990), pp. 69–102
- [26] H. Araki, “Relative Hamiltonian for faithful normal states of a von Neumann algebra”, *Publ. RIMS Kyoto Univ.* **9**(1973), pp. 165–209
- [27] P.M. Alberti, A. Uhlmann, *Stochasticity and partial order. Doubly stochastic maps and unitary mixing* (VEB Deutscher Verlag Wiss., Berlin) 1981
- [28] H. Araki, “Relative entropy for states of von Neumann algebras, *Publ. RIMS Kyoto Univ.* **11**(1976), pp. 809–833
- [29] D. Petz, “A variational expression for the relative entropy”, *Commun. Math. Phys.* **114**(1988), pp. 345–348

- [30] M.J. Donald, “Relative Hamiltonian which are not bounded from above”, *J. Functional Anal.* **91**(1990), pp. 143–173
- [12] M. Ohya, D. Petz, *Quantum Entropy and Its Use* (Springer-Verlag, Heidelberg, Berlin, New York) 1993
- [31] F. Hiai, D. Petz, G. Toth, “Curvature in the geometry of canonical correlation”, to be published
- [32] V.M. Boothby, *An introduction to differential manifolds and Riemannian geometry* (Academic Press, Orlando) 1986
- [33] A. Wehrl, “How chaotic is a state of a quantum system”, *Rep. Math. Phys.* **6**(1974), pp. 15–28
- [34] A.S. Holevo, *Probabilistic and statistical aspects of quantum theory* (North-Holland, Amsterdam) 1982
- [35] H. Nagaoka, “A new approach to Cramér-Rao bounds for quantum state estimation”, *IEICE Technical Report*, **89**(1989), No.228, IT89-42, pp. 9–14
- [36] R. Balian, N.L. Balazs, “Equiprobability, inference and entropy in quantum theory”, *Ann. Phys.* **179**(1986), pp. 97–144
- [37] M.B. Ruskai, “Extremal properties of relative entropy in quantum statistical mechanics”, *Rep. Math. Phys.* **26**(1988), pp. 143–150
- [38] J.P. Neirotti, G.A. Raggio, “Maximizing relative entropy at given energies”, to appear in *Lett. Math. Phys.*