Proceedings of the 9th WSEAS International Conference on Applied Mathematics, Istanbul, Turkey, May 27-29, 2006 (pp219-226)

Noncommutative Analysis: Application to Quantum Information Geometry

HIROSHI HASEGAWA
Institute of Quantum Science, College of Science and Technology
Nihon University,Chiyoda-ku,Tokyo 101-8308,
JAPAN
h-hase@mxj.mesh.ne.jp

Abstract: A simplified version of differential analysis for smooth functions $\varphi(X)$ of a hermitian matrix variable X with an increment A in an n-dimensional matrix manifold, $\mathcal{M}_n$, is established by proving the representation

$$D_X\varphi(X)(A) = D^c_X\varphi(X)(A^c) + [\varphi(X), \Delta_A], \qquad (D)$$

where $D_X \cdot (A)$ denotes the Fréchet derivative with an increment A of X, $D^c_X \cdot (A^c)$ the projection of $D_X \cdot (A)$ onto the commutative submanifold of $D_X \cdot (A)$, and Δ_A a skew-symmetric matrix which depends only on A but independent of φ in the commutator $[\varphi(X), \cdot]$. As an application, we discuss the monotone Riemannian metrics, in particular, *quantum Fisher information* and *generalized relative entropy*.

Key–Words: Fréchet differentiation, quantum Fisher information, monotonicity, convex operator function, quasi-entropy, additivity vs nonadditivity.

1 Introduction

Since around 1996, considerable efforts have been placed on studies of linear algebra and applications to put a systematic formulation of matrix analysis i.e. differential analysis with a matrix-valued variable X(here, it is restricted to hermitian) and its real, smooth functions $\varphi(X)$ which is also matrix-valued. (Readers are referred to Bhatia's monogragh(1997)[1].) Such a differential analysis was discussed early in 1980 by Flett in another monograph[2] by means of Fréchet differentiations. An elementary definition of Fréchet differential of a function $\varphi(X)$ is as follows.

$$D_X(\varphi(X))(A) \equiv \lim_{t\to 0} \frac{\varphi(X+tA)-\varphi(X)}{t} \qquad (1.1)$$

for any $A \in \mathcal{M}_n$ under the satisfaction $\mathrm{Tr}A = 0$: The hermitian subspace of $\mathcal{M}^h_n$ with property $\mathrm{Tr}A = 0$ is called, from a geometrical motivation, the *tangent space* of $\varphi(X)$ at X denoted by $T_X\mathcal{M}$. We shall show in the sequel(**Sec.2**) that a combination of the concept *derivation in operator algebras* and the Fréchet derivative gives rise to a new standpoint of noncommutative differential analysis, as indicated by *(D)* in the *abstract*.

In 1996, Petz proposed a new approach to geometrical studies of information theory by the name of *monotone metrics* in the framework of linear algebra and applications, which revised the classical Fisher information developed by Chentzov[3] and Amari[4] during 1970-1980's. Chentzov aimed to extend his result to noncommutative manifolds which was inherited to Morozova(his spouth) in the form of a paper by Morozova and Chentzov[5].

A remarkable feature disclosed by Morozova-Chentzov and Petz[6] was that, unlike the classical information geometry where the monotonicity on metrics fixes the object uniquely that is the Fisher information, whereas a variety of Fisher type metrics may exist that are defined on matrix spaces, if the only condition of monotonicity is imposed on the metrics.

The prediction of Morozova-Chentzov and Petz on noncommutative information metrics received a revised interest in geometrical aspect of information theory, called now *quantum information geometry*, whose basis can be seen to stem from the noncommutative differential formula (*D*). It can be realized that the skew information, first arose in the decade of 1960-1970 thanks to Wigner-Yanase-Dyson[7] and to Lieb[8], can be understood transparently only by using this formula, which will be discussed in **Sec.3** based on Hasegawa[9] with the**Appendix**. It is shown in **Sec.4** that the partial-derivative version of on Hasegawa[9] with the**Appendix**. It is shown in **Sec.4** that the partial-derivative version of (D) yields the framework that is quite parallel to the cllassical information geometry[4].

Another aspect worth to be discussed is the relationship between the information metrics and the *relative entropy*. This subject was first pursued by Amari[4] by the name of α-divergence(its quantum version by Petz[12] and Hasegawa[13]). The relation has another context which arose in the decade of 1990 to the community of statistical physics for the aspect of the nonextensive generalization of Gibbs entropy, first pointed out by Tsallis[19]. We reconsider the Lesniewski-Ruskai theorem[14] in **Sec.5** to verify the general extensivity of monotone metrics with formula (D).

The present article will continue to the second version **II**, where detailed proofs of the characterization theorem for the Wigner-Yanase-Dyson metrics and for the α-divergence are given; also, all references throughout are at the end of **II**.

2 Derivation and Fréchet differentiation

2.1 Lemma on derivation of analytic functions of a hermition variable X

$$d\varphi(X) = \varphi'(X)d^cX + [\varphi(X), \Delta_\varphi] \qquad (2.1)$$

under the hypothesis

$$dX \equiv d^cX + [X, \Delta_A], \quad \text{where} \qquad (2.2)$$

$$d^cX \in \mathcal{C}_X = \{Y \in \mathcal{M}_n\}; [Y, X] = 0\}.$$

The above writing $d\varphi(X) = d^c\varphi(X) + [\varphi(X), \Delta_\varphi]$ is an orthogonal decomposition of the derivation $d\varphi = d^c\varphi + d^\perp\varphi$ in which $\Delta_\varphi = \Delta$ for X that is irrespective of φ: $\Delta_\varphi = \Delta_A$ holds.

proof. Let $\varphi(z)$ denote a holomorphic function(we say simply an analytic function hereafter) defined on the complex z-plane. In writing $\varphi(X)$(which we call "analytic function of X"), it can be regarded as a C^∞ function on $\mathcal{M}_n; n \in I$. Its 1st order derivation, which is shown to arise in a contour integration of the resolvent for X i.e. $(z - X)^{-1}$ to compute $\varphi(X + dX) - \varphi(X)$, where the contour Γ encircles the spectrum of X which lies on the real axis so that $\varphi(X) = \int_\Gamma \frac{\varphi(z)dz}{z-X}$. It gives rise to the above decomposition as follows.

$$\varphi(X + dX) - \varphi(X) =$$

$$\int_\Gamma \left(\frac{\varphi(z)}{z - (X + dX)} - \frac{\varphi(z)}{z - X} \right) \frac{dz}{2\pi i}$$

$$= \int_\Gamma \left(\frac{1}{z - X} dX \frac{1}{z - X} \right) \frac{\varphi(z)dz}{2\pi i} + \mathcal{O}(dX)^2, \quad \text{to which}$$

$dX = d^cX + [X, \Delta_A]$ is inserted. The first term yields, since d^cX commutes with X,

$$\int_\Gamma \frac{1}{(z - X)^2} \frac{\varphi(z)dz}{2\pi i} = \varphi'(X)d^cX, \qquad (2.3)$$

and the second term

$$\int_\Gamma \frac{1}{z - X}[X, \Delta_X]\frac{1}{z - X} \frac{\varphi(z)dz}{2\pi i}$$

$$= \varphi(X)\Delta_A - \Delta_X\varphi(X)$$

$$= [\varphi(X), \Delta_A], \qquad (2.4)$$

showing that the structure of decomposition is inherited from dX to $d\varphi$ without any modification of the skew-symmetric matrix Δ_A i.e. common to all the analytic functions φ. This decomposition is shown to be an othogonal decomposition of the Hilbert-Schmidt type, as $\mathrm{Tr}d^c\varphi(X)[\varphi(X), \Delta_A] = 0$ by virtue of $d^c\varphi(X) \in \mathcal{C}_X$ i.e. it commutes with $\varphi(X)$. **end of proof.**

Remark 1. Tangent space at a fixed value of X is denoted by $T_X\mathcal{M} = \{A \in \mathcal{M}^h; \mathrm{Tr}A = 0\}$, $\dim T_X\mathcal{M}_n = n^2 - 1$. The subset of all analytic functions of X denoted by $\mathcal{A}_X$ constititute a commutative subalgebra in $T_X\mathcal{M}$. An element of $T_X\mathcal{M}_n$ denoted by $A, B, ..$ is called a tangent vector: this is shown to be related to $\Delta_A, \Delta_B, ..$ in *derivation*(2.2) in **Sec. 2.2** below.

Remark 2. Definition of *Commutant* associated with X: $\mathcal{C}_X = \{Y \in \mathcal{M}_n; [X, Y] = 0\}$. Thus, $d^c\varphi(X) \in \mathcal{C}_X$, and $T_X\mathcal{A}_X \subset \mathcal{C}_X \subset T_X\mathcal{M}_n$ holds.

Remark 3. A Fréchet derivative $D_X(\varphi(X))(A) \equiv \lim_{t\to 0} \frac{\varphi(X+tA)-\varphi(X)}{t}$ eq.(1.1) satisfies

$$\lim_{h\to 0} \frac{\| \varphi(X + h) - \varphi(X) - D_X(\varphi(X))(h) \|}{\| h \|} = 0$$

for any $h \in T_X\mathcal{M}_n[2]$

2.2 Two kinds of a tangent vector A and Δ_A–the representation(D)

$[X, \Delta_A] = A, \quad A \in T_X\mathcal{M}$ and $\Delta_A \in T_X\phi$.
We can express the map of the nonparametric

tangent space to a parametrized tangent space with eigenvalues and matrix elements, which is denoted by $\phi \circ \varphi$[11], as

$$A \mapsto \Delta_A, \text{ where } X = \sum_i \lambda_i e_{ii}; A = \sum_{i \neq j} A_{ij} e_{ij}$$
$$\Delta_A = \sum_{i \neq j} \frac{A_{ij}}{\lambda_i - \lambda_j} e_{ij} \in T_X \phi \circ \varphi \subset T_X \mathcal{M}. \quad (2.5)$$

Then, together with $A^c = \text{diag}A \in \mathcal{C}_X$,

$$D_X(\varphi(X))(A) = D_X^c \varphi(X)(A^c) + [\varphi(X), \Delta_A]$$

which establishes the aimed representation (D).

3 Monotone metrics on matrix spaces—nonparametric version of quantum information geometry

3.1 Axiomatic approach of Morozova-Chentsov and Petz

These authors initiated Riemannian metric formulation for noncommutative geometry on matrix spaces as $K(B, A) \equiv \text{Tr}B^* \mathbf{K}(A)$. Our concern hereafter with matrix variable X is a density matrix denoted by ρ by which quantum expectation can be made: in the present context we make a strong assumption of positive definite $\rho \in \mathcal{M}^{++}$ to satisfy the following.

(a) $(A, B) \mapsto K_\rho(A, B)$ or, in a bracket form $\langle A, K_\rho B \rangle$, is sesquilinear

(b) $K_\rho(A, A) \geq 0$ and the equality holds if and only if $A = 0$

(c) $\rho \mapsto K_\rho(A, A)$ is continuous on $\mathcal{M}_n^+$(all $n \times n$ positive matrices) for every fixed A

(d) monotonicity condition: $K_{T(\rho)}(T(A), T(A)) \leq K_\rho(A, A)$ for every stochastic map T(linear, completely positive and trace preserving) $\mathcal{M}_n(\boldsymbol{C}) \mapsto \mathcal{M}_m(\boldsymbol{C})$, $T\mathcal{M}_n \subset \mathcal{M}_m$ and for every $\rho \in \mathcal{M}_n^{++}$(all $n \times n$ positiv definite matrices) and $A \in \mathcal{M}_n^h$.

Instead of condition(d) we could require the weaker condition(Chentsov's Markovian invariance)

(d′) $K_{U^* \rho U}(U^* A U, U^* A U) = K_\rho(A, A)$.

In the sequel, we always consider the symmetric case of the sesquilinear form

$$K'_\rho(B, A) \equiv \frac{1}{2}(K_\rho(B, A) + K_\rho(A, B)) = K'_\rho(A, B),$$
$$A, B \in \mathcal{M}_n^h \text{ (all } n \times n \text{ hermitians)}, \quad (3.1)$$

the prime being dropped for $K_\rho(A, B)$ hereafter.

By taking the ρ-diagonalized basis, it can be expressed in terms of a two-variable function $c(\lambda, \mu)$ on $\mathbf{R}^+ \times \mathbf{R}^+$ and a single variable one $c(\lambda) \equiv c(\lambda, \lambda)(= 1/\lambda$ that identifies the Fisher term; Theorem 3.1 below) in the above bilinear form as

$$K_\rho(B, A) = \sum_i c(\lambda_i) B_{ii} A_{ii} + 2 \sum_{i<j} c(\lambda_i, \lambda_j) B^*_{ij} A_{ij} \quad (3.2)$$

$$c(\mu, \lambda) = c(\lambda, \mu). \quad (3.3)$$

3.2 Extended Chentsov theorem for noncommutative manifolds– MC function $c(\lambda, \mu)$ and reprentation function $f(x)$

Theorem 3.1(Morozova-Chentsov[5]). *Suppose that(a),(b),(c)and(d′)hold for a real, bilinear form $K_\rho(A, B)$ on self-adjoint elements in $\mathcal{M}_n$. Then, Morozova-Chentsov function* (MC-function) *to represent the symmetric monotone metric $K_\rho(A, B)$ satisfies:*

(i) $c(\lambda)$, $c(\lambda, \mu)(= c(\mu, \lambda))$ *are continuous, positive functions*

(ii) $\lim_{\mu \to \lambda} c(\lambda, \mu) = c(\lambda) = \lambda^{-1}$, $c = 1$ (the Fisher term of the metric)

(iii) $c(\lambda, \mu)$ *is homogeneous of order -1 in λ and μ, implying $c(t\lambda, t\mu) = t^{-1} c(\lambda, \mu)$ for any $t > 0$.*

Petz's representation of a symmetric monotone metric

A positive function is operator monotone, if $f(x); \mathcal{M}^{++} \mapsto \mathbf{R}^+$; satisfies that for any x, y, $x \leq y$ implies $f(x) \leq f(y)$. The corresponding symmetric monotone metric can be expressed as

$$\mathbf{K}_\rho(A, B) = \langle A R_\rho^{-1/2} f(L_\rho R_\rho^{-1})^{-1} R_\rho^{-1/2} B \rangle, \quad (3.4)$$

where $L_\rho R_\rho^{-1} A = \rho A \rho^{-1}$ in term of *left vs right multiplication operator*, respectively, defined by

$$L_\rho(A) \equiv \rho A, \quad R_\rho(A) \equiv A\rho. \quad (3.5)$$

Theorem 3.2 (Petz [6]). *There exists a one-to-one correspondence between the MC function $c(\lambda, \mu)$ subject to* (i),(ii),(iii) *for a symmetric monotone metric $K_\rho(A, B)$ and a metric-characterizing function $f(x); f(1) = 1$ with monotonicity as follows.*

the relation between MC and f functions

i) $f(x) = \dfrac{1}{c(x,1)}$

ii) $c(\lambda,\mu) = \dfrac{1}{\mu f(\lambda/\mu)}$, with symmetry

iii) $c(\mu,\lambda) = c(\lambda,\mu) \Leftrightarrow x f(x^{-1}) = f(x)$,

and every normalized ($f(1) = 1$) symmetric monotone function $f(x)$ lies in a narrow range as

$$\frac{1+x}{2}(\text{minimum}) \geq f(x) \geq \frac{2x}{1+x}(\text{maximum}).$$

Note that *mini-max* corresponds to the minimum vs maximum metric: the monotone metrics are inversely proportional to the f-functions, motivated by the theory of *positive operator means*[10].

3.3 Concrete examples of symmetric monotone metrics

the Wigner-Yanase-Dyson metrics

The original definition of the skew information was $-\frac{1}{2}\mathrm{Tr}[\rho^{1/2}, k]^2$(Wigner-Yanase[7]), where Dyson suggested that it could be generalized to the one exponent ϑ and the other $1-\vartheta$ for the power of ρ. Lieb, in proving thisWigner-Yanase-Dyson(WYD) conjecture[8], expressed it as

$$I_p(\rho,\Delta) \equiv c_p \times \mathrm{Tr}[\rho^p,\Delta][\rho^{1-p},\Delta] \ (-1 \leq p \leq 2)$$

$$\Delta \in \{\text{skew hermitians}\},\ \text{i.e.}\, i\Delta \in \mathcal{M}^h. \quad (3.6)$$

It can be shown that this form is added to the classical Fisher term to yield a full metric form in terms of a Fréchet differential denoted by $D_\rho\cdot(A)$:

$$A = A^c + [\rho,\Delta_A];\quad \mathrm{Tr}A^c g^F A^c + I_p(\rho,\Delta_A),$$

$$= \frac{1}{p(1-p)}\mathrm{Tr}D_\rho(\rho^p)(A)D_\rho(\rho^{1-p})(A), \qquad (3.7)$$

where the factor $[p(1-p)]^{-1}$ is to normalize the form so that in the commutative limit it agrees with the Fisher term ρ^{-1}. Thus, the quantum correspondent to the Fisher term shows a structure, which could be derived from the corresponding quantum divergence(**Sec.5**).

Another remark is the origin of "α" in the α-divergence. This stems from Amari's device[4].

$p = \dfrac{1+\alpha}{2}$, $1-p = \dfrac{1-\alpha}{2}$: Eq.(3.7) changes to

$$K_\rho(A,B) = \frac{4}{1-\alpha^2}\mathrm{Tr}D_\rho(\rho^{\frac{1+\alpha}{2}})(A)D_\rho(\rho^{\frac{1-\alpha}{2}})(B)$$

($p \leftrightarrow 1-p$ *duality* becomes $\pm\,\alpha$ *duality*). (3.8)

Theorem 3.3(Hasegawa[9]).
The Wigner-Yanase-Dyson information is a symmetric monotone metric defined on matrix spaces. Conversely, if a metric is defined on matrix spaces of the form in terms of a pair (φ,χ), $\mathrm{Tr}D_\rho\varphi(\rho)(A)D_\rho\chi(\rho)(B)$, and is monotone with respect to stochstic maps to satisfy **(d)**, *then the resulting MC-function is given by*

$$c(\lambda,\mu) = \frac{(\varphi(\lambda)-\varphi(\mu))(\chi(\lambda)-\chi(\mu))}{(\lambda-\mu)^2}, \qquad (3.9)$$

where the product $\varphi(x)\chi(x)$ at $x \to 0+$ is 0, and it is identical to one of the WYD metrics.

Its representing function is given by

$$f^{\alpha}_{WYD}(x) = \begin{cases} \frac{1-\alpha^2}{4}\frac{(1-x)^2}{(1-x^{\frac{1-\alpha}{2}})(1-x^{\frac{1+\alpha}{2}})} & |\alpha| \neq 1 \\ \frac{x-1}{\log x} & |\alpha| = 1. \end{cases}$$

$$(-1 \leq p \leq 2 \Leftrightarrow 0 \leq |\alpha| \leq 3) \qquad (3.10)$$

The power-mean metrics There exists another region of forbiddenness which we may call "gap region", that is, the maximum of the f_{WYD}-functions having $\alpha = 0$ and the $f_{Bures} = \frac{1+x}{2}$(the Bures metric) was found to be unfilled by any f_{WYD}(Hasegawa[9]): a possible interpolation to fill this gap is

$$f^{\nu}_{power}(x) = \left(\frac{1+x^{1/\nu}}{2}\right)^{\nu} \quad (1 \leq \nu \leq 2). \quad (3.11)$$

Let us denote the set of all the f functions in Theorem 3.2 by $\mathcal{F}$. Similarly, $\{f_{power}\} = \mathcal{F}_{power}$ and$\{f_{WYD}\} = \mathcal{F}_{WYD}$. Then, the combination of $\mathcal{F}_{power}$ and $\mathcal{F}_{WYD}$ forms a linearly ordered subset of $\mathcal{F}$ in the sense that for an order of two parameter sets $1 \leq \nu_1 \leq \nu_2 \leq 2; 0 \leq |\alpha_1| \leq |\alpha_2| \leq 3$, a series of inequalities

$$f_{max}(= f_{Bures}) \geq f^{\nu_1}_{power} \geq f^{\nu_2}_{power} \geq$$

$$f^{|\alpha_1|}_{WYD} \geq f^{|\alpha_2|}_{WYD} \geq f^{3}_{WYD} = f_{min} \ \text{holds[9]} \qquad (3.12)$$

(**Fig.1 in the Appendix**).

4 Parametrization of matrix manifolds –Fréchet partial derivatives

In the classical framework of information geometry, one deals with a manifold in the parameter soace $\mathcal{S}(\{\theta_i; i = 1,2,..,r\})$, in which all variables are assumed to be commutative to each other.

Nevertheless, one observes that "noncommutativity" arises in the theory(see, recent monograph by Amari and Nagaoka[4]), because the concept of *vector field* appears as an important ingredient so that a basic commutation rule exists. Namely,

$$[\partial_i, \partial_j] = 0, \quad [\theta^i, \theta^j] = 0, \quad [\partial_i, \theta^j] = \delta_i^j, \\ \text{where} \quad \partial_i \equiv \frac{\partial}{\partial \theta^i} \quad i = 1, .., r. \qquad (4.1)$$

Here, we discuss briefly how the above noncommutativity can be incorporated without loosing consistency into the present "quantum information geometry". For this purpose we extend the foregoing nonparametric version to that for parametrized matrix manifold, where every density matrix is looked as an analytic function of $\{\theta^i\}$ to form a manifold. We proceed to redefining *tangent vector* introduced in **Sec.2** to conform to the definition in classical geometry: we sketch the essence in two propositions as follows.

4.1 Propositions for parametrizing steps

Proposition 4.1. *Let* $T_\rho \mathcal{M}_n^h \mapsto \mathrm{R}^m (m \leq n^2 - 1)$*denote the derivative map of the tangent space* $T_\rho \mathcal{M}^h$ *expressed as* $\phi \circ \varphi$[11](ϕ to symbolize the parametrization). *There exists a set of linear independent matrix vectors* $\{\Delta_{A_i}\} i = 1, .., m \in T_\rho \phi \circ \varphi \subset T_\rho \mathcal{M}^h$ *by which a tangent vector of classical geometry can be given by a derivative map of projection of* $T_\rho \mathcal{M}^h$ *onto* $T_\rho \mathcal{A}_\rho$ *as*(cf. **Remark 1** and **2, Sec.2** for definitions.)

$$\partial_i \varphi(\rho(\theta)) \equiv D_{\theta^i} \varphi(\rho(\theta)))(A_i) \\ = \frac{\partial^c \varphi}{\partial \theta^i} A_i^c + [\varphi(\rho), \Delta_{A_i}]; by \qquad (4.2)$$

$$Proj(D_{\theta^i}) T_\rho \mathcal{M}_h \mapsto T_\rho \mathcal{A}_\rho. \qquad (4.3)$$

This generalizes a classically defined tangent vector in the sense that, when the tangent space at a fixed ρ, $T_\rho \mathcal{M}^h$, is identical to $T_\rho \mathcal{A}_\rho$, $\partial_i \varphi(\rho(\theta))$ reduces to the first term $\partial^c \varphi / \partial \theta^i$; partial-derivative version of (*D*).

Proposition 4.2. *In the parametrized tangent space at a fixed* ρ*, there exist* m *linearly independent tangent vectors* $\partial_i = \partial_i^c + [\cdot, \Delta_{A_i}]$ *that form a natural basis of a vector field defined by* $\rho \mapsto X_\rho \equiv X^i \partial_i$ *in terms of* $\mathcal{A}_\rho$ *functions*

$$X_i(\rho) \in \mathcal{A}_\rho, \textit{for all } i \textit{ which satisfy}$$

$$[X_i, X_j] = 0, \quad [\partial_i, \partial_j] = 0; \quad [\partial_i, X_j] = \frac{\partial X_j}{\partial \theta^i} \in \mathcal{A}_\rho. \qquad (4.4)$$

Remark 4. The first and second commutativities are due to the definition of the set $\mathcal{A}_\rho \subset \mathcal{C}_\rho$ and the torsionless property of the Fréchet differentials[2], respectively. By virtue of these facts, commutation relations among a set of vector fields at a fixed ρ can be described as if it were in the commutative framework(although this no more holds, if one goes outside $T_\rho \mathcal{A}_\rho$).

4.2 Application to monotone metrics

We compare the two representations of the noncomutative differential in eqs.(2.6) and (4.2), nonparametric vs parametric versions, for a pair of analytic functions (φ, χ):

$$D_\rho(\varphi(\rho))(A) = D_\rho^c \varphi(\rho)(A^c) + [\varphi(\rho), \Delta_A]; \qquad (4.5)$$

$$\partial_i \varphi(\rho(\theta))(A) = \frac{\partial^c \varphi}{\partial \theta^i} A_i^c + [\varphi(\rho), \Delta_{A_i}], \qquad (4.6)$$

and a similar comparison for $D_\rho(\chi)(\rho)(B)$ vs $\partial_i \varphi(\rho(\theta))(B)$. The former expression yields the information metric of the Morozova-Chenzov and Petz form

$$K_\rho(A, B) = \langle A^c \rho^{-1} B^c \rangle + \mathrm{Tr}[\varphi(\rho), \Delta_A][\chi(\rho), \Delta_B]$$

in accordance with Theorem 3.3, whereas the latter expression yields

$$K(A, B)_{ij} = \langle A_i^c \rho^{-1} \partial_i \rho \partial_j \rho B_j^c \rangle \\ + \mathrm{Tr}([\varphi(\rho), \Delta_{A_i}][\chi(\rho), \Delta_{B_j}]) \qquad (4.7)$$

which is in parallel with the classical expression apart from the presence of noncommutative part.

5 Operator convex functions and quasi-entropy

An important aspect of the classical information geometry is that the Fisher information is an object which is derived from *relative entropy*. In quantum case, this subject was studied by Petz[12] by the name of *quasi-entropy*. The direct derivation of the WYD metrics from the noncommutative α-divergence was by Hasegawa[13]. Here, we present a comprehensive discussion.

5.1 Representations and general properties of quasi-entropy

We consider an operator function of a pair of density matrices ρ, σ (both invertible and $\mathrm{Tr}\rho, \sigma = 1$) in terms of an operator convex function $g(x); x \in \mathbf{R}^+$, and satisfies $g(1) = 0$ (See Fig.2A and 2B).

Theorem 5.1(Lesniewski-Ruskai[14]).
The quasi-entropy on finite quantum states $S_g(\rho,\sigma) = \langle \rho^{1/2} g(L_\sigma R_\rho^{-1})\rho^{1/2}\rangle$ admits a general representation in terms of the above $g(x)$:

$$S_g(\rho,\sigma) = \mathrm{Tr}(\sigma-\rho) g^{(2)}(L_\sigma R_\rho^{-1}) R_\rho^{-1}(\sigma-\rho), \quad \text{where } g^{(2)}(x) \equiv \frac{g(x)}{(x-1)^2}, \quad \text{and} \tag{5.1}$$

$$g(x) = b(x-1)^2 + c\frac{(x-1)^2}{x} + \int_0^\infty \frac{(x-1)^2}{x+s} dm(s), \text{ with two constants} \tag{5.2}$$

$b, c \geq 0$ and a finite measure m defined on $(0,\infty)$.

General properties of $S_g(\rho,\sigma)$ are as follows.

(a) $S_g(\rho,\sigma) \geq 0$, and the equality holds if and only if $\sigma = \rho$ by $g(1) = 0$

(b) $S_g(t\rho, t\sigma) = tS_g(\rho,\sigma)$; $t > 0$ (homogenuity of order 1 with respect jointly to ρ and σ)

(c) $S_g(T\rho, T\sigma) \leq S_g(\rho,\sigma)$ with every completely, trace preserving(i.e. stochastic) map T. Equivalently, $S_g(\rho,\sigma)$ is jointly convex in ρ and σ.

(d) $S_g(\rho,\sigma)$ is Fréchet differentiable with respect independently to ρ and σ.

5.2 Quasi-entropy of selfdual and non-selfdual classes

Let $g^{dual}(x)$ denote the function $xg(x^{-1})$, and then $S_{g^{dual}}(\rho,\sigma) = S_g(\sigma,\rho)$ holds. Taking the *dual* of the representation (5.2), we have

$$g^{dual}(x) = xg(x^{-1}) = c(x-1)^2 + b\frac{(x-1)^2}{x} + \int_0^\infty \frac{(x-1)^2}{x+s} d\bar{m}(s), \int_0^\infty d\bar{m}(s) < \infty,$$

$$\text{where} \quad \bar{m}(s) \equiv sm(1/s). \tag{5.3}$$

Theorem 5.2(Lesniewski-Ruskai[14]).
*There exists one-to-one correspondence between a monotone metric K_ρ with operator-monotone decrasing function which is denoted by $k(x)(= 1/f(x)$**Sec.3**$)$, and a symmetrized quasi-entropy $S_g(\rho,\sigma) + S_{g^{dual}}(\rho,\sigma)$, which is written as*

$$k(x) = \frac{g(x) + g^{dual}(x)}{(x-1)^2} \tag{5.4}$$

for the monotone metric, i.e.

$$K_\rho(A,B) = -D_\sigma D_\rho S_g(\rho,\sigma)(A,B)|_{\sigma=\rho} = \langle A, R_\rho^{-1} k(L_\sigma R_\rho^{-1})(B)\rangle, \tag{5.5}$$

where $S_{g^{dual}}(\rho,\sigma) = S_g(\sigma,\rho)$ holds.

Let $\mathcal{G}_{sym}$ and $\mathcal{G}_{asym}$ denote the set of all quasi-entropy g-functions $x \in \mathbf{R}^+ \mapsto \mathbf{R}^+$, defined for symmetric and asymmetric class, respectively, by

$$\mathcal{G}_{sym} = \{g; g(x) = g^{dual}(x)\}, \quad \mathcal{G}_{asym} = \{g; g(x) \neq g^{dual}(x)\}. \tag{5.6}$$

Definition 5.1: selfdual/non-selfdual class

(1) Selfdual class: $S_{s.dual}(\rho,\sigma) = S_{s.dual}(\sigma,\rho)$ with a symmetric $g \in \mathcal{G}_{sym}$. In terms of the measure representation (5.2), (5.3), both $b = c$ and $\bar{m}(s) = m(s)$ hold.
(2) Non-selfdual class: $S_{ns.dual}(\rho,\sigma) \neq S_{ns.dual}(\sigma, \rho)$ with an asymmetric $g \in \mathcal{G}_{asym}$.

Definition 5.2: equivalent class of a pair $(\varphi(x), \chi(x))$ (Gibilisco and Isola[15]). A pair of functions $(\varphi(x) = A\varphi_0(x)+B, \chi(x) = C\chi_0(x)+D)$ conditioned by $AC = 1$ is said to be an equivalent class. Then, the WYD metrics are characterized by the equivalent class of a pair of power functions (x^p, x^{1-p}); $-1 \leq p \leq 2$ called *dual pair.*

A typical example of *selfdual* and *non-selfdual* class for quasi-entropy is that of *Bures metric* and of *WYD metric* $\alpha \neq 0$, respectively, given by

$$g_{Bures}(x) = \frac{(1-x)^2}{1+x} \in \mathcal{G}_{sym}, \tag{5.7}$$

$$g^\alpha_{WYD}(x) = \varphi_\alpha(x) \equiv \frac{4}{1-\alpha^2}(1 - x^{\frac{1+\alpha}{2}}); \tag{5.8}$$

$$\in \mathcal{G}_{asym}, \text{ satisfying } \quad \varphi''(1) = 1, \text{ and } \quad \varphi^{dual}_\alpha(x) = \frac{4}{1-\alpha^2}(x - x^{\frac{1-\alpha}{2}}). \tag{5.9}$$

Quasi-entropy for the power-mean metrics (Fig.2A)

$$g^\nu_{power}(x) = \frac{2^{\nu-1}(1-x)^2}{(1+x^{1/\nu})^\nu} \in \mathcal{G}_{sym}; 1 \leq \nu \leq 2$$

$$S_{g^\nu}(\rho,\sigma) = S_{g^\nu}(\sigma,\rho) = \sum_n c_n \mathrm{Tr}\sigma^{n/\nu}\rho^{1-n/\nu} \tag{5.10}$$

with unit radius of convergence. This is an infinite sum of trace functions, each being of the form $\sigma^p\rho^{1-p}$ (*operator dual pair*). For $\nu = 1$, $S_{g_{Bures}}(\rho,\sigma)$

$$= \sum_{n=1}^\infty \mathrm{Tr}(-1)^n \sigma^n (1 - L_\sigma R_\rho^{-1})^2(\sigma,\rho)\,\rho^{1-n}$$

$$((1 - L_\sigma R_\rho^{-1})(\sigma,\rho) \equiv 1 - 2\sigma\rho^{-1} + \sigma^2\rho^{-2}).$$

There is no member of $\mathcal{G}_{asym}$ in the gap region.

Quasi-entropy for the WYD metrics (Fig.2B)

For $\alpha \neq \pm 1$, $g_\alpha(x) = \frac{4}{1-\alpha^2}(1 - x^{\frac{1+\alpha}{2}})$

$$S_{g_\alpha}(\rho,\sigma) = \frac{4}{1-\alpha^2}(1 - \mathrm{Tr}\sigma^{\frac{1+\alpha}{2}}\rho^{\frac{1-\alpha}{2}}), \quad (5.11)$$

and for $\alpha = \pm 1 \begin{cases} -\log x & \alpha = -1 \\ x \log x & \alpha = 1, \end{cases}$ obtainable from two limits for $\alpha \to -1$ i.e.
$\lim g_\alpha(x) = -\log x;\quad \lim g_\alpha^{dual}(x) = x\log x$:

the metric is called BKM(Bogoliubov-Kubo-Mori)[23], for which the quasi-entropy is

$$S_{g_{-1}}(\rho,\sigma) = \mathrm{Tr}\rho(\log\rho - \log\sigma) \quad \text{and} \quad (5.12)$$
$$S_{g_1}(\rho,\sigma) = \mathrm{Tr}\sigma(\log\sigma - \log\rho) \text{ (Umegaki entropy)}.$$

5.3 Lesniewski-Ruskai theorem by the second Fréchet derivative

In this section we establish another basic property of the information metric form expressed in terms of second derivative of the quasi-entropy.

Theorem 5.3(Jenčová-Hasegawa[27])
Lesniewski-Ruskai formula in Theorem 5.2 can be added by another expression for the Riemannian metric as follows.

$$D_\rho^2 S_g(\rho,\sigma)|_{\sigma=\rho}(A,B) =$$
$$-D_\rho D_\sigma S_g(\rho,\sigma)|_{\sigma=\rho}(A,B) = K_\rho(A,B) \quad (5.13)$$

where

$$S_g(\rho,\sigma = \mathrm{Tr}(\rho-\sigma)R_\rho^{-1}g^{(2)}(L_\sigma R_\rho^{-1})(\rho-\sigma)$$
$$\text{with} \quad g^{(2)}(x) = \frac{g(x)}{(x-1)^2}. \quad (5.14)$$

It implies that the first part of eq.(5.13) can be verified by the Lesniewski-Ruskai formula(5.4). Here, we prove both equalities on equal footing.

proof. We may assume that the function $g^{(2)}(x)$ in eq.(5.1) is analytic in the unit circle of a complex plane except $x \in (0,1)$ which can be expanded as a power series of x^p with unit convergence radius, or more generally, a linear combination of such series with different p's: then we consider its prototype form

$$g^{(2)}(x) = \sum_{n=0}^{\infty} c_n x^{pn} \quad (0 < p \leq 1,\ c_n \in \mathbf{R}).$$

The linear operation $R_\rho^{-1}g^{(2)}(L_\sigma R_\rho^{-1})$ can be written as for any $X \in \mathcal{M}^h$

$$R_\rho^{-1}g^{(2)}(L_\sigma R_\rho^{-1})X = \sum c_n \sigma^{pn} X \rho^{-pn-1}$$

in accordance with the prescription of the left-right multiplication operator $L_\sigma R_\rho^{-1}$ (3.5). By taking $X = \rho - \sigma \in \mathcal{M}$, we have

$$S_g(\rho,\sigma) =$$
$$\mathrm{Tr}(\rho-\sigma)\sum_{n=0}^{\infty} c_n \sigma^{pn}(1-\sigma\rho^{-1})\rho^{-pn}. \quad (5.15)$$

This expression applies to the pertinent cases, as can be seen from eqs.(5.10) and (5.11), respectively. Then, we can proceed to twice Fréchet differentiations on $S_g(\rho,\sigma)$. We may use two elementary derivative formulas[1], namely,

i) $D_\sigma(\sigma)(A) = A$
ii) $D_\rho(\rho^{-1})(B) = -\rho^{-1}B\rho^{-1}\ (A, B \in \mathcal{M}_h)$.

$-D_\rho D_\sigma$ operation on $S_g(\rho,\sigma)$ in eq.(5.15).

Two differentiations D_ρ and D_σ can be made by simple setting $\rho - \sigma = A$ for $D_\rho(\cdot)(A)$, and $\sigma - \rho = B$ for $D_\sigma(\cdot)(B)$, respectively, according to **i)**, because the subsequent setting $\sigma = \rho$ makes any derivatives other than on these two $(\rho-\sigma)$'s to vanish. Therefore,

$$-D_\rho D_\sigma \mathrm{Tr}(\rho-\sigma)\sum_{n=0}^{\infty} c_n\sigma^{pn}(1-\sigma\rho^{-1})\rho^{-pn}(A,B)$$
$$= \langle A, \sum_{n=0}^{\infty} c_n \sigma^{pn} B\rho^{-1}\cdot\rho^{-pn}\rangle \text{ by definition}$$

of L_σ, R_ρ^{-1} in eq.(3.5), and setting $\sigma = \rho$,

$$= \langle A, R_\rho^{-1}g^{(2)}(L_\rho R_\rho^{-1})B\rangle = K^g(A,B). \quad (5.16)$$

This identifies the latter half of eq.(5.13).

D_ρ^2 operation on $S_g(\rho,\sigma)$ in eq.(5.15).

The first operation $D_\rho(\cdot)(A)$ on $S_g(\rho,\sigma) = \mathrm{Tr}(\rho-\sigma)g^{(2)}(L_\sigma R_\rho^{-1})(1-\sigma\rho^{-1})$ is as before, but the second operation $D_\rho(\cdot)(B)$ requires that it be operated on $(1-\sigma\rho^{-1})$. Accordingly,

$$D_\rho^2 S_g(\rho,\sigma)(A,B) =$$
$$D_\rho\langle A, \sum_{n=0}^{\infty} c_n\sigma^{pn}(1-\sigma\rho^{-1})\rho^{-pn}\rangle(B) + \mathcal{O}(\sigma-\rho)^2$$
$$= \langle A, \sum_{n=0}^{\infty} c_n\sigma^{pn}(\sigma\rho^{-1}B\rho^{-1})\rho^{-pn}\rangle \text{ by using } \mathbf{ii)}$$
$$= \langle A, \sum_{n=0}^{\infty} c_n\sigma^{pn}(\sigma\rho^{-1}B)\rho^{-pn-1})\rangle$$
$$\xrightarrow{\sigma=\rho} \langle A, R_\rho^{-1}g^{(2)}(L_\rho R_\rho^{-1})B\rangle. \quad (5.17)$$

This establises the major relation in eq.(5.13).

end of proof.

5.4 Nonextensive vs extensive characteristics of the quasi-(or, relative) entropy and the monotone metric

There has been a new trend in statistical mechanics community, namely *nonextensive generalization of Boltzmann-Gibbs entropy* initiated by Tsallis[19]. We see a connection between this and Information Geometry via divergence functions. See, for example, papers by Abe[20][21] who discusses *non-additive generalization of the Kullback-Leibler divergence*, where this divergence is an additive(extensive) object, whereas nonzero-q modified divergence is nonadditive. We show a general proof that the information metrics retain the additivity[25]. The following argument based on [26] is an application of the foregoing **Theorem 5.3**:

$$K_\rho(A,B) = \langle A, \sum_{n=0}^{\infty} c_n \rho^{pn} B \rho^{-pn-1} \rangle, \qquad (5.18)$$

with $0 < p \leq 1$ and unit convergent radius.

This shows that each term of the series is of the form $\mathrm{Tr}(A\rho^{pn}B\rho^{-pn-1})$ and, as $A = [\rho, \Delta_A]; B = [\rho, \Delta_B]$, linear combinations of $\mathrm{Tr}(\rho^q \Delta_A \rho^{1-q} \Delta_B)$. We can prove that this satisfies *additivity*, and then the series itself:

$$\mathrm{Tr}(\rho^q \Delta_A \rho^{1-q} \Delta_B) = \mathrm{Tr}^{(1)}(\rho^{(1)q} \Delta_{A^{(1)}} \rho^{(1)(1-q)} \Delta_{B^{(1)}}$$

$$+\mathrm{Tr}^{(2)}(\rho^{(2)q} \Delta_{A^{(2)}} \rho^{(2)(1-q)} \Delta_{B^{(2)}} \qquad (5.19)$$

with $\rho = \rho^{(1)} \otimes \rho^{(2)}$ $(\mathrm{Tr}\rho^{(i)} = 1;\ i = 1,2)$ and

$$\Delta_A = \Delta_{A^{(1)}} \otimes 1^{(2)} + 1^{(1)} \otimes \Delta_{A^{(2)}}. \qquad (5.20)$$

(A more detail will be shown in a WSEAS journal.) It verifies the expected general additivity for the symmetric monotone metrics, provided the power-series expansion(5.18) holds. The WYD metric provides its typical example.

Acknowledgements The author thanks A.Jenčová and S.Abe for exchange of a number of information on the present subjects.

Appendix Monotone metric f-functions and quasi-entropy g-functions.

the gap region **the WYD region**

Fig.1

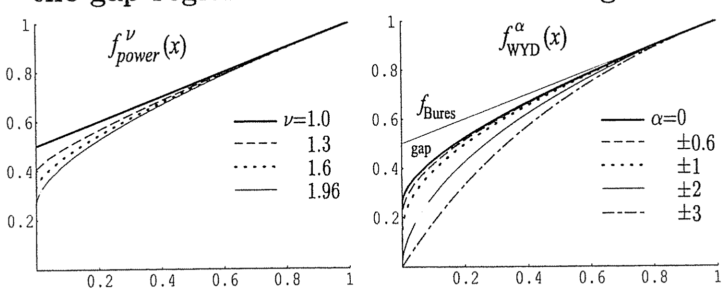

Fig.2A $g^{\nu}_{power}(x); 1 \leq \nu \leq 2$

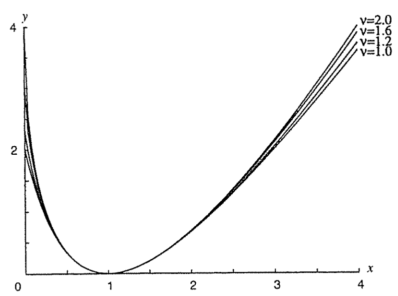

Fig.2B $g_{WYD}(x);\ -3 \leq \alpha < 0$ $g^{\alpha}_{WYD}(x);\ 0 \leq \alpha \leq 3$