Quantum Fisher Information and $q$-Deformed Relative Entropies

--- Additivity vs Nonadditivity ---

Hiroshi HASEGAWA

Institute of Quantum Science, Nihon University, Tokyo 101-8308, Japan

It is shown by way of examples that the quantum correspondent to Fisher information in statistics named quantum information metric is an extensive quantity in spite of the nonextensivity of the quantum $q$-deformed relative entropy. A general proof is presented, and some related topics on the quantum relative entropies vs the metrics are discussed.

§1. Introduction

Nonextensive or nonadditive statistical physics started with Tsallis’ paper in 1988\(^1\) with central keyword $q$-deformed entropy, aiming the Shannon entropy to be replaced by this entropy. A closely related is $q$-deformed relative entropy.

This concept is seen to be identical to $\alpha$-divergence introduced by Amari in geometrical method in statistics, 1985,\(^2\) now called Information Geometry. Its quantum mechanical framework is Quantum Information Geometry.

In this presentation, therefore, it is intended to communicate a latest result, namely extensivity of quantum information metrics and its related topics to the community of Tsallis statistics, specifically, Quantum Tsallis statistics.

Reference Papers in the Presentation


(c) S. Abe:\(^5\) “Nonadditive generalization of the quantum Kullback-Leibler divergence for measuring the degree of purification” Phys. Rev. A 68 (2003), 032302.

(d) S. Abe:\(^6\) “Monotonic decrease of the quantum nonadditive divergence by projective measurement” Phys. Lett. A 312 (2003), 336.


—These will be cited relevantly with maximum hope of communication for basic matters of common interest from the side of quantum information geometry.
§2. Additivity of quantum information metrics

Example 1. Fisher metrics on commutative spaces by Abe

$$K_{\mu\nu}^F(\xi) = \sum_i p_i^{-1}(\xi) \partial_\mu p_i(\xi) \partial_\nu p_i(\xi) = \text{Tr} \rho(\xi) \partial_\mu \log \rho(\xi) \partial_\nu \log \rho(\xi), \quad (2.1)$$

$$\rho(\xi) = \text{density operator as a function of parameter set } \xi = (\xi_1, \xi_2, ..., \xi_m)$$

under the assumption that all matrix realizations of density operator with different parameter values are commutative with each other, i.e.,

$$[\rho(\xi), \rho(\eta)] = 0 \quad \text{for any } \xi, \eta.$$

In Eq. (2.1), $p_i(\xi); i = 1, ..., N,$ denotes all eigenvalues of $\rho; N \times N$ density matrix.

Now, assume that $\rho(\xi) \in \mathcal{M}^{(1)} \otimes \mathcal{M}^{(2)}$ such that $\rho(\xi) = \rho^{(1)}(\xi) \otimes \rho^{(2)}(\xi).$ Then, by virtue of the direct product matrix and the identity for log functions $\log X = \log X + \log Y; [X, Y] = 0$ (applied to $X = \rho^{(1)} \otimes 1^{(2)}, Y = 1^{(1)} \otimes \rho^{(2)}$)

$$\log(\rho(\xi)) = \log(\rho^{(1)}(\xi) \otimes 1^{(2)}) + 1^{(1)} \otimes \log(\rho^{(2)}(\xi)) \quad \text{and} \quad \partial_\mu \log(\rho^{(1)} \otimes \rho^{(2)}) = \partial_\mu \log(\rho^{(1)} \otimes 1^{(2)}) + 1^{(1)} \otimes \partial_\mu \log \rho^{(2)} \quad \text{and} \quad (\mu \leftrightarrow \nu) \quad \text{i.e.}$$

$$\partial_\nu \log(\rho^{(1)} \otimes \rho^{(2)}) = \partial_\nu \log(\rho^{(1)} \otimes 1^{(2)}) + 1^{(1)} \otimes \partial_\nu \log \rho^{(2)}$$

so that for definition (2.1) with a direct-product density matrix $\rho^{(1)}(\xi) \otimes \rho^{(2)}(\xi)$

$$\text{Tr}(\rho^{(1)} \otimes \rho^{(2)}) \partial_\mu (\log \rho^{(1)} \otimes \log \rho^{(2)}) \partial_\nu (\log \rho^{(1)} \otimes \log \rho^{(2)})$$

$$= \text{Tr}_1 \rho^{(1)} \partial_\mu \log \rho^{(1)} \partial_\nu \log \rho^{(1)} + \text{Tr}_2 \rho^{(2)} \partial_\mu \log \rho^{(2)} \partial_\nu \log \rho^{(2)} = K^{F(1)}_{\mu\nu}(x) + K^{F(2)}_{\mu\nu}(x)$$

holds, where the other two cross products vanish by virtue of $\text{Tr}^{(i)} \rho^{(i)} \partial_{\mu_1 \nu_1} \log \rho^{(i)} = \text{Tr}^{(i)} \partial_{\mu_1 \nu_1} \rho^{(i)} = 0 \quad (i = 1 \text{ or } 2)$ under normalization

$$\text{Tr}^{(i)} \rho^{(i)}(x) = 1 \Leftrightarrow \text{Tr} \rho(\xi) = 1. \quad (2.2)$$

This establishes that the Fisher metric on a direct-product matrix space is additive. We note that by the assumed commutativity $[\rho(\xi), \rho(\eta)] = 0$ no ambiguity exists about (partial) derivatives of matrix functions. This is not the case when the commutativity does not hold: then we need a systematic noncommutative analysis.

Example 2. The Wigner-Yanase-Dyson metrics (Wigner and Yanase)

$$K^{WYD}_{\mu\nu}(\xi) = -c(q)\text{Tr}[\rho^{\delta}(\xi), A_\mu(\xi)][\rho^{1-q}(\xi), A_\nu(\xi)] \quad (c(q) \text{ real to be fixed}),$$

$$\rho, A_\mu, s \quad \text{all hermitians, in particular, } \rho \text{ positive-definite}. \quad (2.3)$$

Short history 1 about Wigner-Yanase-Dyson: Wigner and Yanase stated that expression (2.3) for $q = 1/2$ represents an information quantity, as it satisfies convexity with respect to $\rho$ like $\text{Tr} \rho \log \rho,$ and Dyson commented that the convexity would extend to $q \neq 1/2.$ This was called the WYD conjecture by Lieb; he gave a proof of this conjecture for the first time in the framework of convex trace functions.
Remark about the convexity/concavity of the WYD quantity  The quantity $\text{Tr}[\rho^q, A_\mu] [\rho^{1-q}, A_\nu]$; $0 < q < 1$ is a concave function of $\rho$, which is called sometimes WYDL concavity. Instead of hermitians for $A_\mu$'s, it is more convenient to define $A_\mu = \text{antihermitian}(A^\dagger = -A)$. With this alteration, let us redefine

$$K_{\mu\nu}^{WYD} = c(q)\text{Tr}[\rho^q, \Delta A_\mu][\rho^{1-q}, \Delta A_\nu] \quad \text{with } \Delta A \text{ antihermitian.} \quad (2.4)$$

Then, Lieb's theorem\textsuperscript{9}) states that $K_{\mu\nu}^{WYD}$ in Eq. (2.4) is convex with respect to $\rho$ for $0 < q < 1$ with $c(q) > 0$; it is also convex for $-1 < q < 0$, or $1 < q < 2$ with $c(q) < 0$ (the latter case should be related to the issue of the possible range of $q$: discussion in §5).

Short history 2 about Wigner-Yanase-Dyson after 1990: Hasegawa\textsuperscript{10}) discussed this quantity for the first time as quantum information metric by showing its derivation from quantum $q$-divergence (notation follows Abe\textsuperscript{5},6).

$$K_q[\rho||\sigma] \equiv \frac{1}{1 - q} (1 - \text{Tr}(\rho^q\sigma^{1-q})); \partial_\rho \partial_\sigma K_q[\rho||\sigma] \text{derivatives for } \sigma(\xi) \text{ and } |\sigma = \rho$$

$$(K_{\mu\nu}^F + K_{\mu\nu}^{WYD})$, if $c(q) = \frac{1}{q(1 - q)}$, it confirms to the above Lieb theorem in particular, the range of $q$ (also Hasegawa-Petz;\textsuperscript{11}Hasegawa\textsuperscript{12}).

Additivity of the WYD metric  The following expressions are shown to give the additivity $K^{WYD}(\rho^{(1)} \otimes \rho^{(2)}; \xi) = K^{WYD}(\rho^{(1)}; \xi) + K^{WYD}(\rho^{(2)}; \xi)$ by setting

$$\rho = \rho^{(1)} \otimes \rho^{(2)} (\text{Tr}_1\rho^{(1)} = \text{Tr}_2\rho^{(2)} = 1)$$

$$\Delta_A = \Delta_A^{(1)} \otimes I^{(2)} + I^{(1)} \otimes \Delta_A^{(2)} \quad (2.5).$$

Accordingly,

$$\text{Tr}[(\rho^{(1)} \otimes \rho^{(2)})^q, \Delta A][(\rho^{(1)} \otimes \rho^{(2)})^{1-q}, \Delta A] = \text{Tr}^{(1)}[\rho^{(1)q}, \Delta_A^{(1)}][\rho^{1-q}, \Delta_A^{(1)}] + (1 \leftrightarrow 2).$$

Example 3.  The Bures (nonparametric) metric form (Helstrom's symmetric logarithmic-derivative metric Petz-Sudár\textsuperscript{15})

$$K_\rho^{\text{Bures}}(A, A) = \langle A, 2(L_\rho + R_\rho)^{-1}A \rangle \equiv \sum_i p_i A_{ii}^2 + 2 \sum_{j<k} \frac{2}{p_j + p_k} |A_{jk}|^2 \quad (2.6)$$

where $\rho = \text{Diag}(p_1, ..., p_N)$ with the set of $N$-eigenvalues of the density matrix $\rho$. This is a metric form represented as an inner product of two matrix vectors, called tangent vector, $\langle A, B(= K_\rho A) \rangle \equiv \text{Tr}AKA$, where the superoperator $K_\rho$ is given by

$$K_\rho \equiv 2(L_\rho + R_\rho)^{-1} = \sum_{n=0}^\infty 2(-1)^n L_\rho^n R_\rho^{-(n+1)} \quad \text{with } L_\rho R_\rho^{-1} A = R_\rho^{-1} L_\rho A$$

$$\text{Diag} \rho A \rho$$

Hasegawa-Petz;\textsuperscript{11}Hasegawa\textsuperscript{12})
which is called left-right multiplication operator acting on $A$. Accordingly,

$$K_{\rho}^{\text{Bures}}(A, A) = \sum_{n=0}^{\infty} 2(-1)^n \langle A, \rho^n A \rho^{-(n+1)} \rangle. \quad (2.7)$$

Theory of symmetric monotone metrics on matrix spaces was developed by Petz,\textsuperscript{14} also Petz-Sudár\textsuperscript{15} where the Bures metric was explicitly discussed as Eq. (2.6). A general representation of metric form is

$$K_{\rho}^f(A, A) = \langle A, K_{\rho}^f A \rangle = \langle A, \rho^{-1/2}(f(L_{\rho} R_{\rho})^{-1})\rho^{-1/2}A \rangle \quad (2.8)$$
in terms of a real function $f : \mathbb{R}^+ \mapsto \mathbb{R}^+$ with properties.

a) uniformity in 1st order $f(\lambda x) = \lambda f(x) \Rightarrow$ description by scaled $x$.

b) operator monotonicity $x \leq y \Rightarrow f(x) \leq f(y)$, it is related to the monotone-decreasing by projection (Abe\textsuperscript{6}); more generally by any coarse-graining map on the metric form (2.8): $K_{\rho}^f(TA, TA) \leq K_{\rho}^f(A, A)$.

c) symmetry $xf(x^{-1}) = f(x)$.

A possible range of all monotone functions satisfying a)-c) was shown to be

$$f_{\text{Bures}}(x) = f_{\text{max}}(x) = \frac{1}{2} \left( 1 + \frac{1}{x} \right) \geq f(x) \geq \frac{x}{1 + x} = f_{\text{min}}(x) \quad \text{cf. Fig. 1.} \quad (2.9)$$

The power-series expression (2.7) enables us to prove the additivity by using the relation between the tangent vector $A$ (which appears in a Fréchet differentiation (2.11) below) and $\Delta A$ in Eq. (2.4)(Hasegawa\textsuperscript{12}) such that

$$A = A^c + [\rho, \Delta A](\langle A, \rho^n A \rho^{-(n+1)} \rangle \Rightarrow \langle \Delta A, \rho^n \Delta A \rho^{1-n} \rangle), \quad (2.10)$$

$$D_{\rho} \varphi(\rho)(A) \equiv \lim_{t \to 0} \frac{\varphi(\rho + tA) - \varphi(\rho)}{t} \quad \text{(Fréchet differential)} \quad (2.11)$$

Equation (2.10) implies that by replacing $A$ by $\Delta A$ each term of the power-series is of the form $\langle \rho^p \Delta \rho^{1-p} \Delta \rangle$, which provides an insight: how additivity arises (cf. Petz and Jencová\textsuperscript{17} for another proof of the additivity).

§3. Relation between symmetric monotone metrics and relative entropies—a general proof of additivity

$$K_{\rho}(A, B) = -D_{\rho} D_{\sigma} \text{Tr}(\rho - \sigma) \sum_{n=0}^{\infty} c_n \sigma^{pn} (1 - \sigma^{-1})\rho^{-pn} (A, B)|_{\sigma=\rho}$$

$$= \langle A, \sum_{n=0}^{\infty} c_n \rho^{pn} B \rho^{-pn-1} \rangle, \text{ with } 0 < p < 1 \text{ and unit convergence radius.} \quad (3.1)$$
This shows that each term of the series is of the form \( \text{Tr}(A\rho^m B\rho^{-m-1}) \) which satisfies, as \( A = [\rho, \Delta_A] ; B = [\rho, \Delta_B](A^c, B^c \text{ contributing only to } K^F) \)

\[
\text{Tr}[\rho, \Delta_A]\rho^m [\rho, \Delta_B]\rho^{-m-1} = (\text{linear combination of } \text{Tr}\Delta_A\rho^m \Delta_B\rho^{1-m})
= \text{Tr}_1[(\rho^{(1)}), \Delta_A^{(1)}](\rho^{(1)})^m [(\rho^{(1)}), \Delta_B^{(1)}](\rho^{(2)})^{-m-1} + (\text{term}(1) \leftrightarrow (\text{term}(2)))
\]

again by setting \( \rho = \rho^{(1)} \otimes \rho^{(2)}, \) and \( \Delta_A = \Delta_A^{(1)} \otimes I^{(2)} + I^{(1)} \otimes \Delta_A^{(2)} \).

It verifies the expected general additivity for the symmetric monotone metrics, provided the power-series expansion (3.1) holds. Then, how this series arises?

**Outline of deriving the power-series expansion (3.1) by a use of correspondence theorem of Lesniewski-Ruskai** \(^{16}\) (prescription of the precise relation between metric and relative entropy):

an operator convex function \( g(x) = b(x-1)^2 + c\frac{(x-1)^2}{x} + \int_0^\infty \frac{(x-1)^2}{x+s} dm(s) \)
and its dual function \( g^\text{dual}(x) \equiv xg(x^{-1}) \),

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

where \( \bar{m}(s) = sm(1/s) \).

There exists one-to-one correspondence between a monotone metric \( K_\rho \) with operator-monotone decreasing function denoted by \( k(x) = 1/f(x) \), and a symmetrized quasi-entropy \( (S_g(\rho, \sigma) + S_{g^\text{dual}}(\rho, \sigma)) \), which is written as

\[
k(x) = \frac{g(x) + g^\text{dual}(x)}{(x-1)^2} \quad \text{for the monotone metric, i.e.} \quad (3.2)
\]

\[
K_\rho(A, B) = -D_\sigma D_\rho S_h(\rho, \sigma)(A, B)|_{\sigma = \rho} = (A, R_\rho^{-1}h^{(2)}(L_\sigma R_\rho^{-1})B). (3.3)
\]

Then, \( k(x) = \sum_{n=0}^\infty c_n x^n \) by which \( c_n \) can be identified with that in (3.1).

**§4. Classification of general relative entropies**

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

\[
\mathcal{G}_\text{sym} = \{ g; g(x) = g^\text{dual}(x) \}, \quad \mathcal{G}_\text{asym} = \{ g; g(x) \neq g^\text{dual}(x) \}. \quad (4.1)
\]

Examples of selfdual and non-selfdual relative entropies

\[
g_{\text{Bures}}(x) = \frac{2(1-x)^2}{1+x} \in \mathcal{G}_\text{sym}, (\text{expandable in a power series for } |x| < 1) \quad (4.2)
\]

\[
g_{\text{WYD}}^q(x) = \varphi_q(x) \equiv \frac{1}{q(1-q)}(1-x^q); \quad \chi_q(x) = 1 - x^{1-q} \quad \text{both } \in \mathcal{G}_\text{asym} (4.3)
\]
an equivalent class of the single dual pair \( (\varphi_q(x) = \frac{1}{q} x^q, \chi_q(x) = \frac{1}{1-q} x^{1-q}) \):

it satisfies \( \text{Tr} \varphi(\sigma^{(1)} \otimes \sigma^{(2)}) \chi(\sigma^{(1)} \otimes \sigma^{(2)}) = \text{Tr}^{(1)} \varphi(\sigma^{(1)}) \chi(\sigma^{(1)}) \times \text{Tr}^{(2)} \varphi(\sigma^{(2)}) \chi(\sigma^{(2)}) \)

which leads to the additivity of Rényi’s entropy (Petz\(^{13}\)).

Relative entropy for the power-mean metrics (Bures-WYD interpolation)

\[
g_{\text{power}}^\nu(x) = \frac{2^\nu (1 - x)^2}{(1 + x^{1/\nu})^\nu} \in \mathcal{G}_{\text{sym}} \quad 1 \leq \nu \leq 2 \quad \text{“gap region” (Fig. 1), (4.4)}
\]

where no \( q \)-entropy is shown to exist.

This gives rise to two basic questions:

**Question 1. Uniqueness of Fisher metric\(^{18}\)** We have seen two existing classes of symmetric monotone metrics: the WYD class and the power-mean class. What are the possible symmetric monotone metrics?

**Question 2.** The correspondence relation (3-3) states that the \( q \)-deformed class corresponds to the WYD class, hence how the existence of the hitherto unknown relative entropy for the power-mean class (4-4) can be characterized?

§5. Discussions

(1) the range of the nonadditivity parameter In Ref. 19) it was asserted that the allowed range of \( q \) in the expression \( K_q[\rho||\rho_0] \) must be \( 0 < q < 1 \) on the ground

\[
\begin{align*}
\text{Power-mean metrics indexed by } \nu & : \\
\alpha & = 0 \\
\alpha & = \pm 0.6 \\
\alpha & = \pm 1 \\
\alpha & = \pm 2 \\
\alpha & = \pm 3
\end{align*}
\]

\( f_{\text{power}}^\nu(x) = \left( \frac{1 + x^{1/\nu}}{2} \right)^\nu \quad 1 \leq \nu \leq 2; \\ f_{\text{WYD}}^q(x) = \frac{q(1 - q)(1 - x)^2}{(1 - x^q)(1 - x^{1-q})} \quad -1 \leq q \leq 2
\]

Fig. 1. Order structure of monotone metrics on matrix spaces in terms of the \( f \)-functions of Petz\(^{14}\) to cover the entire monotone metrics.
of “form invariance” in the formulation of maximum-entropy principle (that is; in going from $q = 1$ to $q \neq 1$ in the $q$-deformed divergence $K_q[\rho || \rho_0]$, the standard form of the maximum principle be retained). This was said to renormalize the density matrix $\rho \mapsto \rho^q / \text{Tr} \rho^q$. However, it contradicts Fig. 1: we judge the correct range to be $q \in \mathbb{R}$ for the classical framework, and $-1 \leq q \leq 2$ for the quantum one, at least, within the satisfaction of convexity condition. Later in 7), Abe examined the question further on the basis of stability of a distribution, obtaining a modified conclusion that the Tsallis distribution without renormalization is stable (for all $q$ values) but the renormalized one is unstable against small perturbations. So, by this result the question seems to be reconciled with the above judgement.

(2) $q$ vs $1 - q$ duality As we have seen in the non-selfdual example of a pair of convex functions $\varphi_q(x), \chi_q(x)$, the $q$-divergence has the dual structure as regards interchange of the parameter $q$ and $1 - q$. The situation can better be described in terms of $\alpha (= 2q - 1 or -2q + 1)$, which was devised by Amari who called it “$\alpha$-divergence” (Amari\textsuperscript{2}). See Fig. 1.

(3) question 2 of uniqueness of $q$-divergence It may well be conjectured that the only convex divergence function of duality would be the $q$-divergence (uniqueness of Tsallis relative entropy not Tsallis entropy Abe.\textsuperscript{4}) As emphasized in 7), the uniqueness of $q$-divergence is a much harder problem; especially in the quantum framework. This is left to a later exposition.

References

7) S. Abe, Phys. Lett. A 312 (2003), 336 [corrigendum; 324 (2004), 507].
12) H. Hasegawa, IDAQP 6 (2003), 413.
18) N. N. Chentsov, Translations of Mathematical Monograph 53, AMS, Providence, R. I. (1982), 159; Theorem 11.1; original paper in Russia 1972.

The conclusion concerning the range of $q$ (i.e. $0 < q < 1$) was withdrawn in Ref. 7).