From Quasi-Entropy to Various Quantum Information Quantities

by

Fumio HIAI and Dénes PETZ

Abstract

The subject is applications of quasi-entropy in finite-dimensional spaces to many important quantities in quantum information. Operator monotone functions and relative modular operators are used. The origin is relative entropy, and f-divergence, monotone metrics, covariance and χ^2-divergence are the most important particular cases. Monotone metrics with two parameters are introduced and characterized by their joint convexity property.

Introduction

Quasi-entropy was introduced by Petz in 1985 as a quantum generalization of Csiszár's f-divergence in the setting of matrices or von Neumann algebras. An important special case was the relative entropy of Umegaki and Araki. In this paper applications are considered in the finite-dimensional setting. Quasi-entropy has some similarity to monotone metrics. In both cases a modular operator is involved, but there is an essential difference: in the quasi-entropy case two density matrices are involved, and the monotone metric has a single density matrix parameter. Indeed, the general form of quasi-entropy in the matrix algebra setting is

$$S_f^A(D_1\|D_2) := \langle A, f(\mathbb{L}_{D_1}\mathbb{R}_{D_2}^{-1})\mathbb{R}_{D_2}A\rangle,$$

F. Hiai: Graduate School of Information Sciences, Tohoku University, Aoba-ku, Sendai 980-8579, Japan;
e-mail: `hiai.fumio@gmail.com`
D. Petz: Alfréd Rényi Institute of Mathematics, H-1364 Budapest, POB 127, Hungary;
e-mail: `petz@math.bme.hu`

while the form of the monotone metric is

$$\gamma_D^f(A, B) := \langle A, (f(\mathbb{L}_D \mathbb{R}_D^{-1})\mathbb{R}_D)^{-1}B \rangle,$$

where D_1, D_2, and D are positive definite (density) matrices, A and B are (self-adjoint) matrices, and $\langle \cdot, \cdot \rangle$ is the Hilbert–Schmidt inner product for matrices. In the above, $\mathbb{L}_D$ and $\mathbb{R}_D$ are left and right multiplication on matrices, and $\mathbb{L}_{D_1}\mathbb{R}_{D_2}^{-1}$ is the relative modular operator in the matrix setting. For the monotone metric, an operator monotone function $f : \mathbb{R}^+ \to \mathbb{R}^+$ is used.

The aim of this paper is twofold. The first is to give a unified study of f-divergence, monotone metrics, the generalized covariance and χ^2-divergence in terms of quasi-entropy. Those important concepts in quantum information have often been discussed separately in their own languages. The second aim is to introduce a new concept of monotone metrics with two parameters and obtain their joint convexity, an extension of Lieb convexity.

In Section 1 the definition and basic properties of quasi-entropy are reviewed. In Section 2 we discuss the quantum concepts mentioned above to show that many important quantities in quantum information are special cases (or simple reformulations) of quasi-entropy. We show that rather recent results on generalized WYD information and χ^2-divergence can also be understood in terms of quasi-entropy. Finally in Section 3 we consider the extended monotone metrics $\gamma_f^A(D_1\|D_2)$ with two parameters D_1 and D_2. We prove that $\gamma_f^A(D_1\|D_2)$ is jointly convex in the three variables D_1, D_2 and A if (and only if) f is operator monotone.

§1. Quasi-entropy

Let $\mathbf{M}_n$ denote the algebra of $n \times n$ matrices with complex entries. For positive definite matrices $\rho_1, \rho_2 \in \mathbf{M}_n$, $A \in \mathbf{M}_n$ and a function $f : \mathbb{R}^+ \equiv [0, \infty) \to \mathbb{R}$, the *quasi-entropy* is defined as

$$(1) \quad S_f^A(\rho_1\|\rho_2) := \langle A\rho_2^{1/2}, f(\Delta(\rho_1/\rho_2))(A\rho_2^{1/2}) \rangle = \operatorname{Tr} \rho_2^{1/2} A^* f(\Delta(\rho_1/\rho_2))(A\rho_2^{1/2}),$$

where $\langle B, C \rangle := \operatorname{Tr} B^*C$ is the so-called *Hilbert–Schmidt inner product* and $\Delta(\rho_1/\rho_2) : \mathbf{M}_n \to \mathbf{M}_n$ is the linear mapping defined by

$$\Delta(\rho_1/\rho_2)B = \rho_1 B \rho_2^{-1}.$$

This concept was introduced by Petz in 1985; see [23, 24], or Chapter 7 in [22]. (The relative modular operator $\Delta(\rho_1/\rho_2)$ arose in the context of von Neumann algebras, and the paper of Araki [2] had a big influence even in the matrix case.) Quasi-entropy is a quantum generalization of the f-divergence of Csiszár used in

classical information theory (and statistics) [3, 20]. Therefore it could also be called quantum f-divergence as in [10].

The definition of quasi-entropy can be expressed in terms of a mean. For a function f the corresponding mean is defined as $m_f(x, y) = f(x/y)y$ for x, y positive numbers, or commuting positive definite matrices. (In fact, if $f : \mathbb{R}^+ \to \mathbb{R}^+$ satisfies $f(1) = 1$ and certain other conditions, then m_f is the mean discussed in [9].) The linear mappings

$$\mathbb{L}_{\rho_1} X = \rho_1 X \quad \text{and} \quad \mathbb{R}_{\rho_2} X = X \rho_2$$

are positive and commuting. The mean m_f makes sense for them and

$$(2) \qquad S_f^A(\rho_1 \| \rho_2) = \langle A, m_f(\mathbb{L}_{\rho_1}, \mathbb{R}_{\rho_2}) A \rangle.$$

Let $\alpha : \mathbf{M}_n \to \mathbf{M}_m$ be a linear mapping. Its dual $\alpha^* : \mathbf{M}_n \to \mathbf{M}_m$ with respect to the Hilbert–Schmidt inner product is positive if and only if α is positive. Moreover, α is unital if and only if α^* is trace-preserving. The mapping α is called a *Schwarz mapping* if

$$(3) \qquad \alpha(B^* B) \geq \alpha(B^*) \alpha(B) \quad \text{for every } B \in \mathbf{M}_n.$$

Quasi-entropy is monotone and jointly convex (under a suitable condition on f) [22, 24].

Theorem 1. *Assume that $f : \mathbb{R}^+ \to \mathbb{R}$ is an operator monotone function with $f(0) \geq 0$ and $\alpha : \mathbf{M}_n \to \mathbf{M}_m$ is a unital Schwarz mapping. Then*

$$(4) \qquad S_f^A(\alpha^*(\rho_1) \| \alpha^*(\rho_2)) \geq S_f^{\alpha(A)}(\rho_1 \| \rho_2)$$

for $A \in \mathbf{M}_n$ and for invertible density matrices $\rho_1, \rho_2 \in \mathbf{M}_m$.

It is remarkable that for α multiplicative (i.e., α is a $*$-homomorphism) we do not need the condition $f(0) \geq 0$. Moreover, since $V^* \Delta V = \Delta_0$, we do not need the operator monotony of the function f. In this case the operator concavity is the only condition needed to obtain a result analogous to Theorem 1. If we apply the monotonicity (4) (with $-f$ in place of f) to the embedding $\alpha(X) = X \oplus X$ of $\mathbf{M}_n$ into $\mathbf{M}_n \oplus \mathbf{M}_n \subset \mathbf{M}_n \otimes \mathbf{M}_2$ and to the densities $\rho_1 = \lambda E_1 \oplus (1 - \lambda) F_1$, $\rho_2 = \lambda E_2 \oplus (1 - \lambda) F_2$, then we obtain the joint convexity of quasi-entropy:

Theorem 2. *If $f : \mathbb{R}^+ \to \mathbb{R}$ is operator convex, then $S_f^A(\rho_1 \| \rho_2)$ is jointly convex in the variables ρ_1 and ρ_2.*

If we consider quasi-entropy in terms of means, then we can have another proof. The joint convexity of the mean is the inequality

$$f(\mathbb{L}_{(A_1+A_2)/2} \mathbb{R}_{(B_1+B_2)/2}^{-1}) \mathbb{R}_{(B_1+B_2)/2} \leq \tfrac{1}{2} f(\mathbb{L}_{A_1} \mathbb{R}_{B_1}^{-1}) \mathbb{R}_{B_1} + \tfrac{1}{2} f(\mathbb{L}_{A_2} \mathbb{R}_{B_2}^{-1}) \mathbb{R}_{B_2},$$

which can be simplified as

$$f(\mathbb{L}_{A_1+A_2}\mathbb{R}_{B_1+B_2}^{-1})$$
$$\leq \mathbb{R}_{B_1+B_2}^{-1/2}\mathbb{R}_{B_1}^{1/2}f(\mathbb{L}_{A_1}\mathbb{R}_{B_1}^{-1})\mathbb{R}_{B_1}^{1/2}\mathbb{R}_{B_1+B_2}^{-1/2} + \mathbb{R}_{B_1+B_2}^{-1/2}\mathbb{R}_{B_2}^{1/2}f(\mathbb{L}_{A_2}\mathbb{R}_{B_2}^{-1})\mathbb{R}_{B_2}^{1/2}\mathbb{R}_{B_1+B_2}^{-1/2}$$
$$= Cf(\mathbb{L}_{A_1}\mathbb{R}_{B_1}^{-1})C^* + Df(\mathbb{L}_{A_2}\mathbb{R}_{B_2}^{-1})D^*.$$

Here $CC^* + DD^* = I$ and

$$C(\mathbb{L}_{A_1}\mathbb{R}_{B_1}^{-1})C^* + D(\mathbb{L}_{A_2}\mathbb{R}_{B_2}^{-1})D^* = \mathbb{L}_{A_1+A_2}\mathbb{R}_{B_1+B_2}^{-1}.$$

So the joint convexity of quasi-entropy has the form

$$f(CXC^* + DYD^*) \leq Cf(X)C^* + Df(Y)D^*,$$

which is true for an operator convex function f [8, 28].

If $f : \mathbb{R}^+ \to \mathbb{R}^+$ is operator monotone, then it is operator concave and we have the joint concavity in the previous theorem. The book [28] contains information about operator monotone functions. Useful properties are integral representations. The Löwner theorem is

$$(5) \qquad f(x) = f(0) + \beta x + \int_0^\infty \frac{(\lambda+1)x}{\lambda+x}\,d\mu(\lambda),$$

where $\beta \geq 0$ and μ is a finite positive measure on $(0,\infty)$.

An operator monotone function $f : \mathbb{R}^+ \to \mathbb{R}^+$ will be called *standard* if f is symmetric and normalized, i.e., $xf(x^{-1}) = f(x)$ and $f(1) = 1$. A standard function f admits a canonical representation

$$(6) \qquad f(t) = \frac{1+t}{2}\exp\int_0^1 (1-t)^2\frac{\lambda^2-1}{(\lambda+t)(1+\lambda t)(\lambda+1)^2}h(\lambda)\,d\lambda,$$

where $h : [0,1] \to [0,1]$ is a measurable function [6].

§2. Applications

The concept of quasi-entropy includes many important special cases.

§2.1. f-divergence

If ρ_2 and ρ_1 are different and $A = I$, then we have a kind of relative entropy. For $f(x) = x\log x$ we have Umegaki's relative entropy $S(\rho_1\|\rho_2) = \mathrm{Tr}\,\rho_1(\log\rho_1 - \log\rho_2)$. (If we want an operator monotone function, then we can take $f(x) = \log x$ and then we get $S(\rho_2\|\rho_1)$.) This makes the probabilistic and non-commutative situation compatible as one can see in the next argument.

Let ρ_1 and ρ_2 be density matrices in $\mathbf{M}_n$. If in a certain basis they have diagonals $p = (p_1, \ldots, p_n)$ and $q = (q_1, \ldots, q_n)$, then the monotonicity theorem gives the inequality

$$(7) \qquad D_f(p\|q) \le S_f(\rho_1\|\rho_2)$$

for an operator convex function f, where $D_f(p\|q)$ is the f-divergence of p, q. If ρ_1 and ρ_2 commute, then we can take a common eigenbasis and equality holds in (7). It is non-trivial that otherwise the inequality is strict.

If ρ_1 and ρ_2 are different, then p and q can be chosen different as well. Then

$$0 < D_f(p\|q) \le S_f(\rho_1\|\rho_2)$$

as long as f is a non-linear operator convex function and $f(1) = 1$. Conversely, if $S_f(\rho_1\|\rho_2) = 0$, then $p = q$ for every basis and this implies $\rho_1 = \rho_2$. For relative entropy, a deeper result is known. The *Pinsker–Csiszár inequality* says that

$$(8) \qquad \|p - q\|_1^2 \le 2D(p\|q).$$

This extends to the quantum case as

$$(9) \qquad \|\rho_1 - \rho_2\|_1^2 \le 2S(\rho_1\|\rho_2)$$

(see [11] or [28, Chap. 3]).

Example 1. The f-divergence with $f(x) = x \log x$ is relative entropy. The rather familiar modification of the logarithm is

$$\log_\beta x := \frac{x^\beta - 1}{\beta} \qquad (\beta \in (0,1)),$$

whose limit as $\beta \to 0$ is log. If we take $f_\beta(x) = x \log_\beta x$, then

$$S_\beta(\rho_1\|\rho_2) = \frac{\operatorname{Tr} \rho_1^{1+\beta} \rho_2^{-\beta} - 1}{\beta}.$$

Since f_β is operator convex, this is a good generalized relative entropy. It appeared in [31] (see also [22, Chap. 3]) and it was proved that

$$S(\rho_1\|\rho_2) \le S_\beta(\rho_1\|\rho_2) \qquad (\beta \in (0,1)).$$

The *relative entropies of degree* α

$$S_\alpha(\rho_1\|\rho_2) := \frac{1}{\alpha(1-\alpha)} \operatorname{Tr} (I - \rho_2^\alpha \rho_1^{-\alpha})\rho_1$$

are essentially the same.

f-divergence is discued in detail in the recent papers [29, 10]. Recent results on the monotonicity of f-divergence in the case where f is operator convex are in [17, 10].

§2.2. WYD information

In [14] the functions

$$g_p(x) := \begin{cases} \dfrac{1}{p(1-p)}(x - x^p) & \text{if } p \neq 1, \\ x \log x & \text{if } p = 1, \end{cases}$$

were used, which is a reparametrization of $f_\beta(x)$ of Example 1 (up to a constant). (Note that g_p is well-defined for $x > 0$ and $p \neq 0$.) The case $p \in (0, 2]$ was considered; then g_p is operator concave.

For strictly positive A and B, Jenčová and Ruskai [14] defined

$$J_p(K, A, B) := \operatorname{Tr} \sqrt{B} K^* \, g_p\big(L_A R_B^{-1}\big)(K\sqrt{B}),$$

which is the particular case of $S_f^K(A\|B)$ with $f = g_p$.

The joint concavity of $J_p(K, A, B)$ is stated in Theorem 2 of [14], and this is a particular case of Theorem 2 above. For $K = K^*$, we have

$$J_p(K, A, A) = -\frac{1}{2p(1-p)} \operatorname{Tr}[K, A^p][K, A^{1-p}],$$

which is the Wigner–Yanase–Dyson information [33] (up to a constant) extended to the range $(0, 2]$.

§2.3. Monotone metrics

Let $\mathcal{M}_n$ be the set of positive definite density matrices in $\mathbf{M}_n$. This is a differentiable manifold and the set of tangent vectors is $\{A = A^* \in \mathbf{M}_n \; : \; \operatorname{Tr} A = 0\}$. A Riemannian metric is a family of real inner products $\gamma_D(A, B)$ on the tangent vectors [21]. By *monotone metrics* we mean a family of inner products for all manifolds $\mathcal{M}_n$ such that

$$(10) \qquad\qquad \gamma_{\beta(D)}(\beta(A), \beta(A)) \leq \gamma_D(A, A)$$

for every completely positive trace-preserving mapping $\beta : \mathbf{M}_n \to \mathbf{M}_m$.

Define $\mathbb{J}_D^f : \mathbf{M}_n \to \mathbf{M}_n$ as

$$(11) \qquad\qquad \mathbb{J}_D^f := f(\mathbb{L}_D \mathbb{R}_D^{-1})\mathbb{R}_D = \mathbb{L}_D \, m_f \, \mathbb{R}_D \,,$$

where $f : \mathbb{R}^+ \to \mathbb{R}^+$ and m_f is the mean induced by the function f.

It was shown in [26] that monotone metrics with the property

(12) $$\gamma_D(A, A) = \operatorname{Tr} D^{-1} A^2 \quad \text{if} \quad AD = DA$$

have the form

(13) $$\gamma_D(A, B) = \gamma_D^f(A, B) := \operatorname{Tr} A (\mathbb{J}_D^f)^{-1}(B)$$

where f is a standard operator monotone function. These monotone metrics are abstract Fisher informations [25]; the condition (12) tells us that in the commutative case the classical Fisher information is required. The familiar case in physics corresponds to $f(x) = (1 + x)/2$, which gives the SLD (or Bures–Uhlmann) Fisher information [13].

Since

$$\operatorname{Tr} A (\mathbb{J}_D^f)^{-1}(B) = \left\langle (AD^{-1}) D^{1/2}, \frac{1}{f} (\Delta(D/D)) (AD^{-1}) D^{1/2} \right\rangle,$$

we have

$$\gamma_D(A, A) = S_{1/f}^{AD^{-1}}(D \| D).$$

So the monotone metric can be reformulated in terms of quasi-entropy, but there is another relation. The next example is well-known.

Example 2. The Bogolyubov–Kubo–Mori Fisher information is induced by the function

$$f(x) = \frac{x - 1}{\log x} = \int_0^1 x^t \, dt.$$

Then

$$\mathbb{J}_D^f A = \int_0^1 (\mathbb{L}_D \mathbb{R}_D^{-1})^t \mathbb{R}_D A \, dt = \int_0^1 D^t A D^{1-t} \, dt$$

and computing the inverse we have

$$\gamma_D^{\mathrm{BKM}}(A, B) = \int_0^\infty \operatorname{Tr} (D + tI)^{-1} A (D + tI)^{-1} B \, dt.$$

A characterization is given in [5], and the relation with relative entropy is

$$\gamma_D^{\mathrm{BKM}}(A, B) = -\frac{\partial^2}{\partial t \partial s} S(D + tA \| D + sB)\big|_{t=s=0}.$$

Lesniewski and Ruskai [17] discovered that any monotone Fisher information is obtained from f-divergence by derivation:

$$\gamma_D^f(A, B) = -\frac{\partial^2}{\partial t \partial s} S_F(D + tA \| D + sB)\big|_{t=s=0}.$$

The relation of the function F to the function f in this formula is

$$(14) \qquad \frac{1}{f(t)} = \frac{F(t) + tF(t^{-1})}{(t-1)^2}.$$

When f is an operator monotone function, the monotone metric $\gamma_D^f(A, B)$ in (13) makes sense for all positive definite matrices D and general $A, B \in \mathbf{M}_n$, and the monotonicity (10) holds for such general D and A [26]. In this general situation, if (10) is assumed but condition (12) is not required, then we have the generalized monotone metrics characterized by Kumagai [16]. They have the form

$$K_\rho(A, B) := b(\operatorname{Tr} \rho) \operatorname{Tr} A^* \operatorname{Tr} B + c\langle A, (\mathbb{J}_\rho^f)^{-1}(B)\rangle,$$

where $f : \mathbb{R}^+ \to \mathbb{R}^+$ is operator monotone, $f(1) = 1$, $b : \mathbb{R}^+ \to \mathbb{R}^+$ and $c > 0$.

Let $\beta : \mathbf{M}_n \otimes \mathbf{M}_2 \to \mathbf{M}_n$ be defined as

$$\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \mapsto B_{11} + B_{22}.$$

This is completely positive and trace-preserving, the so-called partial trace. For

$$D = \begin{bmatrix} \lambda D_1 & 0 \\ 0 & (1-\lambda)D_2 \end{bmatrix}, \qquad A = \begin{bmatrix} \lambda A_1 & 0 \\ 0 & (1-\lambda)A_2 \end{bmatrix}$$

the inequality (10) gives

$$\gamma_{\lambda D_1 + (1-\lambda)D_2}(\lambda A_1 + (1-\lambda)A_2, \lambda A_1 + (1-\lambda)A_2)$$
$$\leq \gamma_{\lambda D_1}(\lambda A_1, \lambda A_1) + \gamma_{(1-\lambda)D_2}((1-\lambda)A_2, (1-\lambda)A_2).$$

Since $\gamma_{tD}(tA, tB) = t\gamma_D(A, B)$, we obtain joint convexity:

Theorem 3. *For an operator monotone function f, the monotone metric $\gamma_D^f(A, A)$ is a joint convex function for (D, A) of positive definite D and general $A \in \mathbf{M}_n$.*

In particular, the convexity of $\gamma_D^f(A, A)$ in D can be deduced from (13). We have the convexity of the operator $(\mathbb{J}_D^f)^{-1}$ in the positive definite D.

§2.4. Generalized covariance

If $\rho_1 = \rho_2 = \rho$ and $A, B \in \mathbf{M}_n$ are arbitrary, then one can considered the *generalized covariance* [27], which is defined as

$$(15) \qquad \operatorname{qCov}_\rho^f(A, B) := \langle A\rho^{1/2}, f(\Delta(\rho/\rho))(B\rho^{1/2})\rangle - (\operatorname{Tr} \rho A^*)(\operatorname{Tr} \rho B).$$

The first term is $\langle A, \mathbb{J}_\rho^f B \rangle$ and the covariance has some similarity to the monotone metrics.

If ρ, A and B commute, then (15) becomes $f(1) \operatorname{Tr} \rho A^* B - (\operatorname{Tr} \rho A^*)(\operatorname{Tr} \rho B)$. This shows that the normalization $f(1) = 1$ is natural. The generalized covariance $\operatorname{qCov}_\rho^f(A, B)$ is a sesquilinear form and it is determined by $\operatorname{qCov}_\rho^f(A, A)$ for $A \in \mathbf{M}_n$ with $\operatorname{Tr} \rho A = 0$. Formally, this is a quasi-entropy and Theorem 1 applies if f is operator monotone. If we require the symmetry condition $\operatorname{qCov}_\rho^f(A, A) = \operatorname{qCov}_\rho^f(A^*, A^*)$, then f should have the symmetry $x f(x^{-1}) = f(x)$.

Assume that $\operatorname{Tr} \rho A = \operatorname{Tr} \rho B = 0$ and $\rho = \operatorname{Diag}(\lambda_1, \ldots, \lambda_n)$. Then

$$(16) \qquad \operatorname{qCov}_\rho^f(A, B) = \sum_{ij} \lambda_i f(\lambda_j / \lambda_i) A_{ij}^* B_{ij}.$$

The usual *symmetrized covariance* corresponds to the function $f(t) = (t+1)/2$:

$$\operatorname{Cov}_\rho(A, B) := \frac{1}{2} \operatorname{Tr}(\rho(A^* B + B A^*)) - (\operatorname{Tr} \rho A^*)(\operatorname{Tr} \rho B).$$

It turns out that there is a one-to-one correspondence between generalized covariances and Fisher informations.

Theorem 4. *For a standard operator monotone function f the covariance $\operatorname{qCov}_\rho^f(A, A)$ is a concave function of ρ for each fixed self-adjoint A.*

In fact, $\langle A, \mathbb{J}_\rho^f A \rangle = S_f^A(\rho \| \rho)$ is concave in ρ by Theorem 2. The convexity of $(\operatorname{Tr} \rho A)^2$ is obvious.

§2.5. χ^2-divergence

The χ^2-*divergence*

$$\chi^2(p, q) := \sum_i \frac{(p_i - q_i)^2}{q_i} = \sum_i \left(\frac{p_i}{q_i} - 1 \right)^2 q_i$$

was first introduced by Karl Pearson in 1900. Since

$$\left(\sum_i |p_i - q_i| \right)^2 = \left(\sum_i \left| \frac{p_i}{q_i} - 1 \right| q_i \right)^2 \le \sum_i \left(\frac{p_i}{q_i} - 1 \right)^2 q_i,$$

we have

$$(17) \qquad \|p - q\|_1^2 \le \chi^2(p, q).$$

We also remark that the χ^2-divergence is the f-divergence of Csiszár with $f(x) = (x-1)^2$, which is an operator convex function. In the quantum case definition (1) gives

$$S_f(\rho, \sigma) = \operatorname{Tr} \rho^2 \sigma^{-1} - 1.$$

Another quantum generalization was introduced very recently in [32]:

$$\chi_\alpha^2(\rho, \sigma) = \text{Tr}((\rho - \sigma)\sigma^{-\alpha}(\rho - \sigma)\sigma^{\alpha-1}) = \text{Tr}\,\rho\sigma^{-\alpha}\rho\sigma^{\alpha-1} - 1,$$

where $\alpha \in [0,1]$. If ρ and σ commute, then this formula is independent of α. In the general case the above $S_f(\rho, \sigma)$ is obtained for $\alpha = 0$.

More generally, Temme et al. defined

$$\chi_k^2(\rho, \sigma) := \langle \rho - \sigma, \Omega_\sigma^k(\rho - \sigma)\rangle,$$

where $\Omega_\sigma^k = R_\sigma^{-1}k(\Delta(\sigma/\sigma))$ and $1/k$ is a standard operator monotone function. In the present notation $\Omega_\sigma^k = (\mathbb{J}_\sigma^{1/k})^{-1}$ and for density matrices we have

$$\chi_k^2(\rho, \sigma) = \langle \rho, \Omega_\sigma^k\rho\rangle - 1 = \langle \rho, (\mathbb{J}_\sigma^{1/k})^{-1}\rho\rangle - 1 = \gamma_\sigma^{1/k}(\rho, \rho) - 1.$$

Up to an additive constant this is a monotone metric. The monotonicity of χ^2-divergence follows from (10) and is stated as Theorem 4 in [32], where the important function k is

$$k_\alpha(x) = \frac{1}{2}(x^{-\alpha} + x^{\alpha-1}) \quad \text{and} \quad \chi_{k_\alpha}^2 = \chi_\alpha^2.$$

Note that $1/k_\alpha$ is a standard operator monotone function for $\alpha \in [0,1]$ and $k_\alpha(x)$ is convex in α. The latter implies that χ_α^2 is convex in α. The χ^2-divergence χ_α^2 is minimal if $\alpha = 1/2$. (It is interesting that this appeared in [30] as Example 4.)

When $1/k(x) = (1 + x)/2$ is the largest standard operator monotone function, then the corresponding χ^2-divergence is the smallest and in [32] the notation $\chi_{\text{Bures}}^2(\rho, \sigma)$ is used. Actually,

$$\chi_{\text{Bures}}^2(\rho, \sigma) = 2\int_0^\infty \text{Tr}\,\rho\exp(-t\sigma)\rho\exp(-t\sigma)\,dt - 1$$

(see Example 1 in [30]).

The monotonicity and the classical inequality (17) imply that

$$\|\rho - \sigma\|_1^2 \le \chi^2(\rho, \sigma).$$

Indeed, if E is the conditional expectation onto the commutative algebra generated by $\rho - \sigma$, then

$$\|\rho - \sigma\|_1^2 = \|E(\rho) - E(\sigma)\|_1^2 \le \chi^2(E(\rho), E(\sigma)) \le \chi^2(\rho, \sigma).$$

§3. Extension of monotone metrics

As an extension of the operator (11), define $\mathbb{J}_{D_1,D_2}^f : \mathbf{M}_n \to \mathbf{M}_n$ as

$$\mathbb{J}_{D_1,D_2}^f := f(\mathbb{L}_{D_1}\mathbb{R}_{D_2}^{-1})\mathbb{R}_{D_2} \equiv f(\Delta(D_1/D_2))\mathbb{R}_{D_2} = \mathbb{L}_{D_1}\,m_f\,\mathbb{R}_{D_2}\,,$$

where $f : \mathbb{R}^+ \to \mathbb{R}^+$. In this notation,

$$S_f^A(\rho_1 \| \rho_2) = \langle A, \mathbb{J}_{\rho_1,\rho_2}^f A \rangle.$$

Theorem 2 says that for an operator monotone function f, $\langle A, \mathbb{J}_{\rho_1,\rho_2}^f A \rangle$ is a jointly concave function of the variables ρ_1 and ρ_2.

Monotone metrics involve $(\mathbb{J}_{\rho,\rho}^f)^{-1}$; therefore we consider the inverse

$$(\mathbb{J}_{D_1,D_2}^f)^{-1} = f^{-1}(\Delta(D_1/D_2))\mathbb{R}_{D_2}^{-1},$$

whenever $f(x) > 0$ for $x > 0$. In this section, β is a completely positive trace-preserving mapping between matrix spaces.

Lemma 1. *Assume that D_1, D_2, $\beta(D_1)$, and $\beta(D_2)$ are positive definite and that $f(x) > 0$ for $x > 0$. Then the conditions*

$$(18) \qquad \beta^*(\mathbb{J}_{\beta(D_1),\beta(D_2)}^f)^{-1}\beta \leq (\mathbb{J}_{D_1,D_2}^f)^{-1}$$

and

$$(19) \qquad \beta\mathbb{J}_{D_1,D_2}^f\beta^* \leq \mathbb{J}_{\beta(D_1),\beta(D_2)}^f$$

are equivalent.

Proof. The following inequalities are equivalent forms of (18):

$$(\mathbb{J}_{D_1,D_2}^f)^{1/2}\beta^*(\mathbb{J}_{\beta(D_1),\beta(D_2)}^f)^{-1}\beta(\mathbb{J}_{D_1,D_2}^f)^{1/2} \leq I,$$

$$\|(\mathbb{J}_{\beta(D_1),\beta(D_2)}^f)^{-1/2}\beta(\mathbb{J}_{D_1,D_2}^f)^{1/2}\|^2$$
$$= \|(\mathbb{J}_{D_1,D_2}^f)^{1/2}\beta^*(\mathbb{J}_{\beta(D_1),\beta(D_2)}^f)^{-1}\beta(\mathbb{J}_{D_1,D_2}^f)^{1/2}\| \leq 1,$$

$$\|(\mathbb{J}_{D_1,D_2}^f)^{1/2}\beta^*(\mathbb{J}_{\beta(D_1),\beta(D_2)}^f)^{-1/2}\| \leq 1,$$

$$(\mathbb{J}_{\beta(D_1),\beta(D_2)}^f)^{-1/2}\beta(\mathbb{J}_{D_1,D_2}^f)\beta^*(\mathbb{J}_{\beta(D_1),\beta(D_2)}^f)^{-1/2} \leq I.$$

The last inequality is equivalent to (19). $\qquad\square$

Example 3. Let $f(x) = sx + 1$, where $s > 0$. Then

$$\langle A, (\mathbb{J}_{D_1,D_2}^f)^{-1}A \rangle = \langle A, (s\Delta(D_1/D_2) + 1)^{-1}\mathbb{R}_{D_2}^{-1}A \rangle = \langle A, (s\mathbb{L}_{D_1} + \mathbb{R}_{D_2})^{-1}A \rangle.$$

This was studied in [17], where the result

$$(20) \qquad \beta^*(s\mathbb{L}_{\beta(D_1)} + \mathbb{R}_{\beta(D_2)})^{-1}\beta \leq (s\mathbb{L}_{D_1} + \mathbb{R}_{D_2})^{-1}$$

was obtained. Another formulation is

$$(21) \qquad \beta^*(\mathbb{J}_{\beta(D_1),\beta(D_2)}^f)^{-1}\beta \leq (\mathbb{J}_{D_1,D_2}^f)^{-1},$$

which is equivalent to

$$(22) \qquad \beta \mathbb{J}^f_{D_1,D_2} \beta^* \le \mathbb{J}^f_{\beta(D_1),\beta(D_2)}$$

due to the previous lemma.

For $f(x) = sx + 1$ this is rather obvious:

$$\langle A, \beta \mathbb{J}^f_{D_1,D_2} \beta^* A \rangle = s \operatorname{Tr} D_1 \beta^*(A)\beta^*(A^*) + \operatorname{Tr} D_2 \beta^*(A^*)\beta^*(A)$$

and

$$\langle A, \mathbb{J}^f_{\beta(D_1),\beta(D_2)} A \rangle = s \operatorname{Tr} D_1 \beta^*(AA^*) + \operatorname{Tr} D_2 \beta^*(A^*A).$$

The Schwarz inequality

$$\beta^*(X)\beta^*(X^*) \le \beta^*(XX^*)$$

gives (22) and hence (20). $\hfill \square$

Theorem 5. *Let $\beta : \mathbf{M}_n \to \mathbf{M}_m$ be a completely positive trace-preserving mapping and $f : [0,\infty) \to (0,\infty)$ be an operator monotone function. Assume that D_1, D_2, $\beta(D_1)$, and $\beta(D_2)$ are positive definite. Then*

$$\beta^*(\mathbb{J}^f_{\beta(D_1),\beta(D_2)})^{-1}\beta \le (\mathbb{J}^f_{D_1,D_2})^{-1}.$$

Proof. Due to Lemma 1 it is enough to prove (19) for an operator monotone function. Based on the Löwner theorem (5), we may consider $f(x) = x/(\lambda + x)$ $(\lambda > 0)$. So

$$\mathbb{J}^f_{D_1,D_2} = \frac{\mathbb{L}_{D_1}}{\lambda I + \mathbb{L}_{D_1}\mathbb{R}_{D_2}^{-1}},$$

and the equivalent form of (19) is

$$\langle \beta(A), (\lambda I + \mathbb{L}_{\beta(D_1)}\mathbb{R}_{\beta(D_2)}^{-1})\mathbb{L}_{\beta(D_1)}^{-1}\beta(A) \rangle \le \langle A, (\lambda I + \mathbb{L}_{D_1}\mathbb{R}_{D_2}^{-1})\mathbb{L}_{D_1}^{-1}A \rangle$$

or

$$\lambda \operatorname{Tr} \beta(A^*)\beta(D_1)^{-1}\beta(A) + \operatorname{Tr} \beta(A)\beta(D_2)^{-1}\beta(A^*) \le \lambda \operatorname{Tr} A^*D_1^{-1}A + \operatorname{Tr} AD_2^{-1}A^*.$$

This inequality is true due to the matrix inequality

$$\beta(X^*)\beta(Y)^{-1}\beta(X) \le \beta(X^*Y^{-1}X) \quad (Y > 0)$$

(see [19]). $\hfill \square$

Generalized monotone metrics

$$(23) \qquad \gamma^f_{D_1,D_2}(A,B) := \langle A, (\mathbb{J}^f_{D_1,D_2})^{-1}B \rangle$$

are an extension of monotone metrics, which is the case $D_1 = D_2 = D$. We can call them also *monotone metrics with two parameters*. (The geometric meaning of these quantities is not clear at the moment, although the case $f(x) = 1 + sx$ appeared already in [17].)

Example 4. Let $f(x) = (x+1)/2$. Then

$$\mathbb{J}^f_{D_1,D_2}A = \tfrac{1}{2}(D_1A + AD_2)$$

and

$$(\mathbb{J}^f_{D_1,D_2})^{-1}B = \int_0^\infty \exp(-tD_1/2)B\exp(-tD_2/2)\,dt.$$

If D_1, D_2 and B commute, then

$$(\mathbb{J}^f_{D_1,D_2})^{-1}B = \left(\frac{D_1+D_2}{2}\right)^{-1}B.$$

Example 5. Let $f(x) = (x-1)/\log x$. Then similarly to $\mathbb{J}^f_D$ in Example 2, we have

$$\mathbb{J}^f_{D_1,D_2}A = \int_0^1 D_1^t A D_2^{1-t}\,dt.$$

When

$$D_1 = \sum_i \lambda_i P_i \quad \text{and} \quad D_2 = \sum_j \mu_j Q_j$$

are the spectral decompositions, then

(24) $$\mathbb{J}^f_{D_1,D_2}A = \sum_{i,j} m_f(\lambda_i,\mu_j)P_i A Q_j,$$

where m_f is the logarithmic mean. (The formula is general, it holds for all functions $f : \mathbb{R}^+ \to \mathbb{R}^+$.) To show that

$$(\mathbb{J}^f_{D_1,D_2})^{-1}B = \int_0^\infty (D_1+tI)^{-1}B(D_2+tI)^{-1}\,dt$$

is really the inverse, we may compute

$$\int_0^\infty (D_1+tI)^{-1}B(D_2+tI)^{-1}\,dt = \sum_{i,j} \frac{1}{m_f(\lambda_i,\mu_j)}P_i B Q_j.$$

If D_1, D_2 and B commute, then

$$(\mathbb{J}^f_{D_1,D_2})^{-1}B = \frac{D_1 - D_2}{\log D_1 - \log D_2}B.$$

We can recognize that in the commuting case

$$\mathbb{J}^f_{D_1,D_2}B = m_f(D_1,D_2)B, \quad (\mathbb{J}^f_{D_1,D_2})^{-1}B = \frac{1}{m_f(D_1,D_2)}B,$$

where m_f is the mean generated by the function f, i.e., $m_f(x,y) = xf(y/x)$.

Now let $f : (0,\infty) \to (0,\infty)$ be a continuous function; the definition of f at 0 is not necessary here. Define $g,h : (0,\infty) \to (0,\infty)$ by $g(x) := xf(x^{-1})$ and

$$(25) \qquad h(x) := \left(\frac{f(x)^{-1} + g(x)^{-1}}{2} \right)^{-1}, \quad x > 0.$$

Obviously, h is symmetric, i.e., $h(x) = xh(x^{-1})$ for $x > 0$, so we may call h the harmonic symmetrization of f.

Theorem 6. *In the above situation consider the following conditions:*

(i) *f is operator monotone,*

(ii) *$(D,A) \mapsto \langle A, (\mathbb{J}^f_D)^{-1}A \rangle$ is jointly convex in positive definite D and general A in $\mathbf{M}_n$ for every n,*

(iii) *$(D_1, D_2, A) \mapsto \langle A, (\mathbb{J}^f_{D_1,D_2})^{-1}A \rangle$ is jointly convex in positive definite D_1, D_2 and general A in $\mathbf{M}_n$ for every n,*

(iv) *$(D,A) \mapsto \langle A, (\mathbb{J}^f_D)^{-1}A \rangle$ is jointly convex in positive definite D and self-adjoint A in $\mathbf{M}_n$ for every n,*

(v) *h is operator monotone.*

Then (i)⇔(ii)⇔(iii)⇒(iv)⇔(v).

When f is symmetric, one can define the metric γ^f_D on the manifold of positive definite matrices in $\mathbf{M}_n$ by formula (13). Then the theorem says that $\gamma^f_D(A,A)$ is jointly convex in (D,A) for every matrix size n if and only if γ^f_D is a monotone metric. Thus the monotone metrics are characterized by joint convexity.

The difference between two parameters and one parameter is not essential if the matrix size can be changed. We need the next lemma.

Lemma 2. *For $D_1, D_2 > 0$ and general X in $\mathbf{M}_n$ let*

$$D := \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}, \quad Y := \begin{bmatrix} 0 & X \\ 0 & 0 \end{bmatrix}, \quad A := \begin{bmatrix} 0 & X \\ X^* & 0 \end{bmatrix}.$$

Then

$$(26) \qquad \langle Y, (\mathbb{J}^f_D)^{-1}Y \rangle = \langle X, (\mathbb{J}^f_{D_1,D_2})^{-1}X \rangle,$$

$$(27) \qquad \langle A, (\mathbb{J}^f_D)^{-1}A \rangle = 2\langle X, (\mathbb{J}^h_{D_1,D_2})^{-1}X \rangle.$$

Proof. First we show that

$$(28) \qquad (\mathbb{J}_D^f)^{-1} \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} = \begin{bmatrix} (J_{D_1}^f)^{-1} X_{11} & (\mathbb{J}_{D_1,D_2}^f)^{-1} X_{12} \\ (\mathbb{J}_{D_2,D_1}^f)^{-1} X_{21} & (\mathbb{J}_{D_2}^f)^{-1} X_{22} \end{bmatrix}.$$

Since continuous functions can be approximated by polynomials, it is enough to check (28) for $f(x) = x^k$, which is easy. From (28), (26) is obvious and

$$\langle A, (\mathbb{J}_D^f)^{-1} A \rangle = \langle X, (\mathbb{J}_{D_1,D_2}^f)^{-1} X \rangle + \langle X^*, (\mathbb{J}_{D_2,D_1}^f)^{-1} X^* \rangle.$$

By (24) we further have

$$(29) \qquad \langle X, (\mathbb{J}_{D_1,D_2}^g)^{-1} X \rangle = \sum_{i,j} m_g(\lambda_i, \mu_j) \operatorname{Tr} X^* P_i X Q_j$$

$$= \sum_{i,j} m_f(\mu_j, \lambda_i) \operatorname{Tr} X Q_j X^* P_i = \langle X^*, (\mathbb{J}_{D_2,D_1}^f)^{-1} X^* \rangle.$$

Therefore,

$$\langle A, (\mathbb{J}_D^f)^{-1} A \rangle = \langle X, (\mathbb{J}_{D_1,D_2}^g)^{-1} X \rangle + \langle X, (\mathbb{J}_{D_1,D_2}^g)^{-1} X \rangle = 2 \langle X, (\mathbb{J}_{D_1,D_2}^h)^{-1} X \rangle. \qquad \square$$

Proof of Theorem 6. (i)$\Rightarrow$(ii) is Theorem 3, and (ii)$\Rightarrow$(iii) follows from (26). We prove (iii)$\Rightarrow$(i). For each $\xi \in \mathbb{C}^n$ let $X_\xi := [\xi \ 0 \cdots 0] \in \mathbf{M}_n$, i.e., the first column of X_ξ is ξ and all other entries of X_ξ are zero. When $D_2 = I$ and $X = X_\xi$, we have for $D > 0$ in $\mathbf{M}_n$,

$$\langle X_\xi, (\mathbb{J}_{D,I}^f)^{-1} X_\xi \rangle = \langle X_\xi, f(D)^{-1} X_\xi \rangle = \langle \xi, f(D)^{-1} \xi \rangle.$$

Hence it follows from (iii) that $\langle \xi, f(D)^{-1} \xi \rangle$ is jointly convex in $D > 0$ in $\mathbf{M}_n$ and $\xi \in \mathbb{C}^n$. By a standard convergence argument we see that $(D, \xi) \mapsto \langle \xi, f(D)^{-1} \xi \rangle$ is jointly convex for positive invertible $D \in B(\mathcal{H})$ and $\xi \in \mathcal{H}$, where $B(\mathcal{H})$ is the set of bounded operators on a separable infinite-dimensional Hilbert space $\mathcal{H}$. Now Theorem 3.1 in [1] is used to conclude that $1/f$ is operator monotone decreasing, so f is operator monotone.

(ii)$\Rightarrow$(iv) is trivial. Assume (iv); then it follows from (27) that (iii) holds for h instead of f, so (v) holds thanks to (iii)$\Rightarrow$(i) for h. From (29) when $A = A^*$ and $D_1 = D_2 = D$, it follows that

$$\langle A, (\mathbb{J}_D^f)^{-1} A \rangle = \langle A, (\mathbb{J}_D^g)^{-1} A \rangle = \langle A, (\mathbb{J}_D^h)^{-1} A \rangle.$$

Hence (v) implies (iv) by applying (i)$\Rightarrow$(ii) to h. $\qquad \square$

It is worth noting that condition (iii) is strictly stronger than (iv) in Theorem 6. Consider any non-symmetric operator monotone function $f_0 : (0, \infty) \to$

$(0, \infty)$ (for example, $f_0(x) = x^\alpha$ with $\alpha \in (0,1)$, $\alpha \neq 1/2$) and define

$$f(x) := \begin{cases} f_0(x) & \text{if } 0 < x \leq 1, \\ xf_0(x^{-1}) & \text{if } x \geq 1. \end{cases}$$

Then f is not operator monotone but h given in (25) is operator monotone as the harmonic symmetrization of f_0. Hence the above fact follows from Theorem 6.

In the next theorem we prove the implication (i)$\Rightarrow$(iii) of Theorem 6 again in a more general setting.

Theorem 7. *For an operator monotone function f and $\theta \in (0,1]$, the mapping*

$$(D_1, D_2, A) \mapsto \langle A, (\mathbb{J}^f_{D_1, D_2})^{-\theta} A \rangle$$

is a jointly convex function of positive definite D_1, D_2 and general A in $\mathbf{M}_n$.

Proof. The joint concavity of operator means [15] implies that for all $D_1, D_2, D'_1, D'_2 > 0$ we have

$$\mathbb{J}^f_{\frac{D_1 + D'_1}{2}, \frac{D_2 + D'_2}{2}} \geq \frac{\mathbb{J}^f_{D_1, D_2} + \mathbb{J}^f_{D'_1, D'_2}}{2}.$$

Since $0 < x \mapsto x^{-\theta}$ is operator monotone decreasing, it is known (see Remark 4.6 in [7], also [1]) that $(A, \xi) \mapsto \langle \xi, A^{-\theta} \xi \rangle$ is jointly convex in positive invertible operators A and vectors ξ. Therefore,

$$\left\langle \frac{A + A'}{2}, \left(\mathbb{J}^f_{\frac{D_1 + D'_1}{2}, \frac{D_2 + D'_2}{2}} \right)^{-\theta} \left(\frac{A + A'}{2} \right) \right\rangle$$
$$\leq \left\langle \frac{A + A'}{2}, \left(\frac{\mathbb{J}^f_{D_1, D_2} + \mathbb{J}^f_{D'_1, D'_2}}{2} \right)^{-\theta} \left(\frac{A + A'}{2} \right) \right\rangle$$
$$\leq \frac{1}{2} \left(\langle A, (\mathbb{J}^f_{D_1, D_2})^{-\theta} A \rangle + \langle A', (\mathbb{J}^f_{D'_1, D'_2})^{-\theta} A' \rangle \right).$$

This is the desired joint convexity. $\square$

Example 6. Consider the operator monotone functions $f(x) = x^\alpha$, where $\alpha \in [0,1]$. Then

$$\langle A, (\mathbb{J}^f_{D_1, D_2})^{-\theta} A \rangle = \text{Tr } A^* D_1^{-\alpha\theta} A D_2^{-(1-\alpha)\theta}.$$

Hence Theorem 7 shows that if $p, q \geq 0$ and $p + q \leq 1$, then

$$(D_1, D_2, A) \mapsto \text{Tr } A^* D_1^{-p} A D_2^{-q}$$

is jointly convex for $D_1, D_2 > 0$ and general A in $\mathbf{M}_n$. This is the joint convexity theorem of Lieb [18].

Joint convexity of quasi-entropy type functions is discussed in more detail in the forthcoming paper [12].

References

[1] T. Ando and F. Hiai, Operator log-convex functions and operator means, Math. Ann. **350** (2011), 611–630.   Zbl 1227098 MR 2805638

[2] H. Araki, Relative entropy of states of von Neumann algebras, Publ. RIMS Kyoto Univ. **11** (1976), 809–833.   Zbl 0326.46031 MR 0425631

[3] I. Csiszár, Information-type measures of difference of probability distributions and indirect observations, Studia Sci. Math. Hungar. **2** (1967), 299–318.   Zbl 0157.25802 MR 0219345

[4] E. G. Effros, A matrix convexity approach to some celebrated quantum inequalities, Proc. Nat. Acad. Sci. USA **106** (2009), 1006–1008.   Zbl 1202.81018 MR 2475796

[5] M. Grasselli and R. F. Streater, Uniqueness of the Chentsov metric in quantum information theory, Infin. Dimens. Anal. Quantum Probab. Relat. Top. **4** (2001), 173–182.   Zbl 1039.81011 MR 1841616

[6] F. Hansen, Characterizations of symmetric monotone metrics on the state space of quantum systems, Quantum Inform. Comput. **6** (2006), 597–605.   Zbl 1152.81729 MR 2298673

[7] _____, Extensions of Lieb's concavity theorem, J. Statist. Phys. **124** (2006), 87–101.   Zbl 1157.47305 MR 2256619

[8] F. Hansen and G. K. Pedersen, Jensen's inequality for operators and Löwner's theorem, Math. Ann. **258** (1982), 229–241.   Zbl 0473.47011 MR 1513286

[9] F. Hiai and H. Kosaki, *Means of Hilbert space operators*, Lecture Notes in Math. 1820, Springer, 2003.   Zbl 1048.47001 MR 2005250

[10] F. Hiai, M. Mosonyi, D. Petz and C. Bény, Quantum f-divergences and error correction, Rev. Math. Phys. **23** (2011), 691–747.   Zbl 1230.81007 MR 2826462

[11] F. Hiai, M. Ohya and M. Tsukada, Sufficiency, KMS condition and relative entropy in von Neumann algebras, Pacific J. Math. **96** (1981), 99–109.   Zbl 0473.46044 MR 0634765

[12] F. Hiai and D. Petz, Convexity of quasi-entropy type functions: Lieb's and Ando's convexity theorems revisited, arXiv:1209.0546.

[13] A. S. Holevo, *Probabilistic and statistical aspects of quantum theory*, North-Holland, Amsterdam, 1982.   Zbl 0497.46053 MR 0681693

[14] A. Jenčová and M. B. Ruskai, A unified treatment of convexity of relative entropy and related trace functions, with conditions for equality, Rev. Math. Phys. **22** (2010), 1099–1121.   Zbl 1218.81025 MR 2733251

[15] F. Kubo and T. Ando, Means of positive linear operators, Math. Ann. **246** (1980), 205–224.   Zbl 0412.47013 MR 0563399

[16] W. Kumagai, A characterization of extended monotone metrics, Linear Algebra Appl. **434** (2011), 224–231.   Zbl 1206.94016 MR 2737244

[17] A. Lesniewski and M. B. Ruskai, Monotone Riemannian metrics and relative entropy on noncommutative probability spaces, J. Math. Phys. **40** (1999), 5702–5724.   Zbl 0968.81006 MR 1722334

[18] E. H. Lieb, Convex trace functions and the Wigner–Yanase–Dyson conjecture, Adv. Math. **11** (1973), 267–288.   Zbl 0267.46055 MR 0332080

[19] E. H. Lieb and M. B. Ruskai, Some operator inequalities of the Schwarz type, Adv. Math. **12** (1974), 269–273.   Zbl 0274.46045 MR 0336406

[20] F. Liese and I. Vajda, On divergences and informations in statistics and information theory, IEEE Trans. Inform. Theory **52** (2006), 4394–4412.   MR 2300826

[21] H. Nagaoka, On Fisher information on quantum statistical models, in *Asymptotic theory of quantum statistical inference*, M. Hayashi (ed.), World Sci., 2005, 113–124.

[22] M. Ohya and D. Petz, *Quantum entropy and its use*, Springer, Berlin, 1993; 2nd ed., 2004.

[23] D. Petz, Quasi-entropies for states of a von Neumann algebra, Publ. RIMS Kyoto Univ. **21** (1985), 787–800.   Zbl 0606.46039 MR 0817164

[24] ———, Quasi-entropies for finite quantum systems, Rep. Math. Phys. **23** (1986), 57–65. Zbl 0629.46061 MR 0868631

[25] ———, Geometry of canonical correlation on the state space of a quantum system, J. Math. Phys. **35** (1994), 780–795.   Zbl 0804.53101 MR 1257547

[26] D. Petz, Monotone metrics on matrix spaces, Linear Algebra Appl. **244** (1996), 81–96. Zbl 0856.15023 MR 1403277

[27] ———, Covariance and Fisher information in quantum mechanics, J. Phys. A **35** (2002), 929–939.   Zbl 1168.81317 MR 1890135

[28] ———, *Quantum information theory and quantum statistics*, Springer, Berlin, 2008. Zbl 1145.81002 MR 2363070

[29] ———, From f-divergence to quantum quasi-entropies and their use, Entropy **12** (2010), 304–325.   Zbl 1229.81054 MR 2608239

[30] D. Petz and C. Ghinea, Introduction to quantum Fisher information, in *Quantum probability and related topics*, QP–PQ: Quantum Probab. White Noise Anal. 27, R. Rebolledo and M. Ország (eds.), World Sci., 2011, 261–281.   Zbl pre06042766 MR 2799128

[31] M. B. Ruskai and F. H. Stillinger, Convexity inequalities for estimating free energy and relative entropy, J. Phys. A **23** (1990), 2421–2437.   Zbl 0709.60549 MR 1063572

[32] K. Temme, M. J. Kastoryano, M. B. Ruskai, M. M. Wolf and F. Verstraete, The χ^2-divergence and mixing times of quantum Markov processes, J. Math. Phys. **51** (2010), 122201.   MR 2779170

[33] E. P. Wigner and M. M. Yanase, Information content of distributions, Proc. Nat. Acad. Sci. USA **49** (1963), 910–918.   Zbl 0128.14104 MR 0151127