From quasi-entropy

Dénes Petz[1]
Alfréd Rényi Institute of Mathematics
H-1364 Budapest, POB 127, Hungary

Abstract

The subject is the overview of the use of quasi-entropy in finite dimensional spaces. Matrix monotone functions and relative modular operators are used. The origin is the relative entropy and the f-divergence, monotone metrics, covariance and the χ^2 divergence are the most important particular cases. The extension of the monotone metric to two variables is a new concept.

Key words and phrases: f-divergence, quasi-entropy, von Neumann entropy, relative entropy, monotonicity property, Fisher information, χ^2-divergence.

Quasi-entropy was introduced by Petz in 1985 as the quantum generalization of Csiszár's f-divergence in the setting of matrices or von Neumann algebras. The important special case was the relative entropy of Umegaki and Araki. In this paper the applications are overviewed in the finite dimensional setting. Quasi-entropy has some similarity to the monotone metrics, in both cases the modular operator is included, but there is an essential difference: In the quasi-entropy two density matrices are included and for the monotone metric on foot-point density matrices. In this paper two density matrices are introduced in the monotone metric style.

1 Quasi-entropy

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

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

[1]E-mail: petz@math.bme.hu. Partially supported by the Hungarian Research Grant OTKA T068258 and the Mittag-Leffler Institute in Stockholm.

where $\langle B, C\rangle := \mathrm{Tr}\, B^*C$ is the so-called *Hilbert-Schmidt inner product* and $\Delta(\rho_1/\rho_2)$: $\mathcal{M} \to \mathcal{M}$ is a linear mapping acting on matrices:

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

This concept was introduced by Petz in 1985, see [19, 20], or Chapter 7 in [18]. (The relative modular operator $\Delta(\rho_1/\rho_2)$ was born in the context of von Neumann algebras and the paper of Araki [1] had a big influence even in the matrix case.) The quasi-entropy is the quantum generalization of the f-divergence of Csiszár used in classical information theory (and statistics) [2, 16]. Therefore the quantum f-divergence could be another terminology as in [10].

The definition of quasi-entropy can be formulated with mean. For a function f the corresponding mean is defined as $m_h(x, y) = f(x/y)y$ for positive numbers, or for commuting positive definite matrices. The linear mappings

$$L_{\rho_1}X = \rho_1 X \quad \text{and} \quad R_{\rho_2}X = X\rho_2$$

are positive and commuting. The mean m_f makes sense and

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

Let $\alpha : \mathcal{M}_0 \to \mathcal{M}$ be a mapping between two matrix algebras. The dual $\alpha^* : \mathcal{M} \to \mathcal{M}_0$ 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. $\alpha : \mathcal{M}_0 \to \mathcal{M}$ is called a *Schwarz mapping* if

$$\alpha(B^*B) \geq \alpha(B^*)\alpha(B) \tag{3}$$

for every $B \in \mathcal{M}_0$.

The quasi-entropies are monotone and jointly convex [18, 20].

Theorem 1 *Assume that $f : \mathbb{R}^+ \to \mathbb{R}$ is an operator monotone function with $f(0) \geq 0$ and $\alpha : \mathcal{M}_0 \to \mathcal{M}$ is a unital Schwarz mapping. Then*

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

holds for $A \in \mathcal{M}_0$ and for invertible density matrices ρ_1 and ρ_2 from the matrix algebra $\mathcal{M}$.

Proof: The proof is based on inequalities for operator monotone and operator concave functions. First note that

$$S_{f+c}^A(\alpha^*(\rho_1)\|\alpha^*(\rho_2)) = S_f^A(\alpha^*(\rho_1)\|\alpha^*(\rho_2)) + c\,\mathrm{Tr}\,\rho_1\alpha(A^*A))$$

and

$$S_{f+c}^{\alpha(A)}(\rho_1\|\rho_2) = S_f^{\alpha(A)}(\rho_1\|\rho_2) + c\,\mathrm{Tr}\,\rho_1(\alpha(A)^*\alpha(A))$$

for a positive constant c. Due to the Schwarz inequality (3), we may assume that $f(0) = 0$.

Let $\Delta := \Delta(\rho_1/\rho_2)$ and $\Delta_0 := \Delta(\alpha^*(\rho_1)/\alpha^*(\rho_2))$. The operator

$$V X \alpha^*(\rho_2)^{1/2} = \alpha(X)\rho_2^{1/2} \qquad (X \in \mathcal{M}_0) \tag{5}$$

is a contraction:

$$\begin{aligned} \|\alpha(X)\rho_2^{1/2}\|^2 &= \operatorname{Tr} \rho_2(\alpha(X)^*\alpha(X)) \\ &\leq \operatorname{Tr} \rho_2(\alpha(X^*X) = \operatorname{Tr} \alpha^*(\rho_2)X^*X = \|X\alpha^*(\rho_2)^{1/2}\|^2 \end{aligned}$$

since the Schwarz inequality is applicable to α. A similar simple computation gives that

$$V^*\Delta V \leq \Delta_0\,. \tag{6}$$

Since f is operator monotone, we have $f(\Delta_0) \geq f(V^*\Delta V)$. Recall that f is operator concave, therefore $f(V^*\Delta V) \geq V^* f(\Delta)V$ and we conclude

$$f(\Delta_0) \geq V^* f(\Delta) V\,. \tag{7}$$

Application to the vector $A\alpha^*(\rho_2)^{1/2}$ gives the statement. □

It is remarkable that for a multiplicative α we do not need the condition $f(0) \geq 0$. Moreover, $V^*\Delta V = \Delta_0$ and we do not need the matrix monotonicity of the function f. In this case the only condition is the matrix concavity, analogously to Theorem 1. If we apply the monotonicity (4) to the embedding $\alpha(X) = X \oplus X$ of $\mathcal{M}$ into $\mathcal{M} \oplus \mathcal{M}$ 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 concavity of the quasi-entropy:

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

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

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

which can be simplified as

$$\begin{aligned} &f(L_{A_1+A_2}R^{-1}_{B_1+B_2}) \\ &\quad \leq R^{-1/2}_{B_1+B_2}R^{1/2}_{B_1} f(L_{A_1}R^{-1}_{B_1})R^{1/2}_{B_1}R^{-1/2}_{B_1+B_2} + R^{-1/2}_{B_1+B_2}R^{1/2}_{B_2} f(L_{A_2}R^{-1}_{B_2})R^{1/2}_{B_2}R^{-1/2}_{B_1+B_2} \\ &\quad \leq C f(L_{A_1}R^{-1}_{B_1})C^* + D f(L_{A_2}R^{-1}_{B_2})D^*. \end{aligned}$$

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

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

So the joint convexity of the quasi-entropy has the form

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

which is true for an operator convex function f [5, 24].

If f is operator monotone function, then it is operator concave and we have joint concavity in the previous theorem. The book [24] contains information about operator monotone functions. The standard useful properties are integral representations. The Löwner theorem is

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

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

$$f(t) = \frac{1+t}{2} \exp \int_0^1 (1-t)^2 \frac{\lambda^2 - 1}{(\lambda + t)(1 + \lambda t)(\lambda + 1)^2} h(\lambda)\, d\lambda, \tag{8}$$

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-divergences

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 $S(\rho_1 \| \rho_2) = \operatorname{Tr} \rho_1 (\log \rho_1 - \log \rho_2)$. (If we want a matrix monotone function, then we can take $f(x) = \log x$ and then we get $S(\rho_2 \| \rho_1)$.) Umegaki's relative entropy is the most important example, therefore the function f will be chosen to be matrix convex. 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 $\mathcal{M}$. If in certain basis they have diagonal $p = (p_1.p_2, \dots, p_n)$ and $q = (q_1, q_2, \dots, q_n)$, then the monotonicity theorem gives the inequality

$$D_f(p \| q) \le S_f(\rho_1 \| \rho_2) \tag{9}$$

for a matrix convex function f. If ρ_1 and ρ_2 commute, them we can take the common eigenbasis and in (9) the equality appears. It is not trivial that otherwise the inequality is strict.

If ρ_1 and ρ_2 are different, then there is a choice for p and q such that they are different as well. Then

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

Conversely, if $S_f(\rho_1\|\rho_2) = 0$, then $p = q$ for every basis and this implies $\rho_1 = \rho_2$. For the relative entropy, a deeper result is known. The *Pinsker-Csiszár inequality* says that

$$\|p - q\|_1^2 \le 2D(p\|q). \tag{10}$$

This extends to the quantum case as

$$\|\rho_1 - \rho_2\|_1^2 \le 2S(\rho_1\|\rho_2), \tag{11}$$

see [8], or [24, Chap. 3].

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

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

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

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

Since f_β is operator convex, this is a good generalized entropy. It appeared in the paper [27], see also [18, Chap. 3], there γ is written instead of β and

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

is proven.

The *relative entropies of degree* α

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

are essentially the same. □

The f-divergence is contained in details in the recent papers [25, 10].

2.2 WYD information

In the paper [12] the functions

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

are used, this is a reparametrization of Example 1. (Note that g_p is well-defined for $x > 0$ and $p \neq 0$.) The considered case is $p \in [1/2, 2]$, then g_p is operator concave.

For strictly positive A and B, Jenčová and Ruskai define

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

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

The joint concavity of $J_p(K, A, B)$ is stated in Theorem 2 in [12] and this is a particular case of Theorem 2 above. For $K = K^*$, we have

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

which is the Wigner-Yanase-Dyson information (up to a constant) and extends it 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 manifold and the set of tangent vectors is $\{A = A^* \in \mathbf{M}_n \,:\, \mathrm{Tr}\, A = 0\}$. A Riemannian geometry is a set of real inner products $\gamma_D(A, B)$ on the tangent vectors [17]. By monotone metrics we mean inner product for all matrix spaces such that

$$\gamma_{\beta(D)}(\beta(A), \beta(A)) \leq \gamma_D(A, A) \tag{12}$$

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

$$\mathbb{J}_D^f = f(\mathbb{L}_D \mathbb{R}_D^{-1})\mathbb{R}_D = \mathbb{L}_D\, m_f\, \mathbb{R}_D\,, \tag{13}$$

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

It was obtained in the paper [22] that monotone metrics with the property

$$\gamma_D(A, A) = \mathrm{Tr}\, D^{-1}A^2 \quad \text{if} \quad AD = DA \tag{14}$$

has the form

$$\gamma_D(A, B) = \mathrm{Tr}\, A(\mathbb{J}_D^f)^{-1}(B) \tag{15}$$

where f is a standard matrix monotone function. These monotone metrics are abstract Fisher informations, the condition (14) tells that in the commutative case the classical Fisher information is required. The popular case in physics corresponds to $f(x) = (1+x)/2$, this gives the SSA Fisher information.

Since

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

we have

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

So the monotone metric is a particular case of the quasi-entropy, but there is another relation. The next example has been well-known.

Example 2 The Boguliubov-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

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

A characterization is in the paper [4] and the relation with the relative entropy is

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

□

Ruskai and Lesniewski discovered that all monotone Fisher informations are obtained from an f-divergence by derivation [14]:

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

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

$$\frac{1}{f(t)} = \frac{F(t) + tF(t^{-1})}{(t-1)^2}. \tag{16}$$

If D runs on all positive definite matrices, conditions $\gamma_D(A,A) \in \mathbb{R}$ for self-adjoint A and (14) are not required, but the monotonicity (12) is assumed, then we have the generalized monotone metric characterized by Kumagai [13]. They have the form

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

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

Let $\beta : \mathbf{M}_n \otimes \mathbf{M}_2 \to \mathbf{M}_m$ 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, it is a so-called partial trace. For

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

the inequality (12) gives

$$\gamma_{\lambda D_1+(1-\lambda)D_2}(B,B) \le \gamma_{\lambda D_1}(\lambda B, \lambda B) + \gamma_{(1-\lambda)D_2}((1-\lambda)B, (1-\lambda)B).$$

Since $\gamma_{tD}(tA, tB) = t\gamma_D(A,B)$, we obtained the convexity.

Theorem 3 *For a standard matrix monotone function f and for a self-adjoint matrix A the monotone metric $\gamma_D^f(A,A)$ is a convex function of D.*

This convexity relation can be reformulated from formula (15). We have the convexity of the operator $(\mathbb{J}_D^f)^{-1}$ in the positive definite D.

2.4 Generalized covariance

If $\rho_2 = \rho_1 = \rho$ and $A, B \in \mathcal{M}$ are arbitrary, then one can approach to the *generalized covariance* [23].

$$\mathrm{qCov}_\rho^f(A,B) := \langle A\rho^{1/2}, f(\Delta(\rho/\rho))(B\rho^{1/2})\rangle - (\mathrm{Tr}\,\rho A^*)(\mathrm{Tr}\,\rho B). \tag{17}$$

is a generalized covariance. 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 this becomes $f(1)\mathrm{Tr}\,\rho A^*B - (\mathrm{Tr}\,\rho A^*)(\mathrm{Tr}\,\rho B)$. This shows that the normalization $f(1) = 1$ is natural. The generalized covariance $\mathrm{qCov}_\rho^f(A,B)$ is a sesquilinear form and it is determined by $\mathrm{qCov}_\rho^f(A,A)$ when $\{A \in \mathcal{M} : \mathrm{Tr}\,\rho A = 0\}$. Formally, this is a quasi-entropy and Theorem 1 applies if f is matrix monotone. If we require the symmetry condition $\mathrm{qCov}_\rho^f(A,A) = \mathrm{qCov}_\rho^f(A^*,A^*)$, then f should have the symmetry $xf(x^{-1}) = f(x)$.

Assume that $\mathrm{Tr}\,\rho A = \mathrm{Tr}\,\rho B = 0$ and $\rho = \mathrm{Diag}\,(\lambda_1, \lambda_2, \dots, \lambda_n)$. Then

$$\mathrm{qCov}_\rho^f(A,B) = \sum_{ij} \lambda_i f(\lambda_j/\lambda_i) A_{ij}^* B_{ij}. \tag{18}$$

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

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

The interpretation of the covariances is not at all clear. In the next section they will be called *quadratic cost functions*. It turns out that there is a one-to-one correspondence between quadratic cost functions and Fisher informations.

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

Proof: The argument similar to the proof of Theorem 3. Instead of the inequality $\beta^*(\mathbb{J}^f_{\beta(D)})^{-1}\beta \le (\mathbb{J}^f_D)^{-1}$ we use the inequality $\beta \mathbb{J}^f_D \beta^* \le \mathbb{J}^f_{\beta(D)}$ (see Theorem 1.2 in [26] or [23]). This gives the concavity of $\langle A, \mathbb{J}^f_\rho a\rangle$. The convexity of $(\mathrm{Tr}\,\rho A)^2$ 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

$$\|p - q\|_1^2 \le \chi^2(p,q). \tag{19}$$

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

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

Another quantum generalization was introduced very recently in [28]:

$$\chi^2_\alpha(\rho,\sigma) = \mathrm{Tr}\,\left((\rho-\sigma)\sigma^{-\alpha}(\rho-\sigma)\sigma^{\alpha-1}\right) = \mathrm{Tr}\,\rho\sigma^{-\alpha}\rho\sigma^{\alpha-1} - 1$$

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

More generally, they defined

$$\chi^2_k(\rho,\sigma) := \left\langle \rho - \sigma, \Omega^k_\sigma(\rho - \sigma)\right\rangle,$$

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

$$\chi^2_k(\rho,\sigma) = \langle \rho, \Omega^k_\sigma \rho\rangle - 1 = \langle \rho, (\mathbb{J}^{1/k}_\sigma)^{-1}\rho\rangle - 1 = \gamma^{1/k}_\sigma(\rho,\rho) - 1.$$

Up to the additive constant this is a monotone metric. The monotonicity of the χ^2-divergence follows from (12) and monotonicity is stated as Theorem 4 in the paper [28], where the important function k is

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

$1/k_\alpha$ is a standard matrix monotone function for $\alpha \in [0,1]$ and $k_\alpha(x)$ is convex in the variable α. The latter implies that χ^2_α is convex in α. The χ^2-divergence χ^2_α is minimal if $\alpha = 1/2$. (It is interesting that this appeared in [26] as Example 4.)

When $1/k(x) = (1+x)/2$ is the largest standard matrix monotone function, then the corresponding χ^2-divergence is the smallest and in the paper [28] the notation $\chi^2_{Bures}(\rho,\sigma)$ is used. Actually,

$$\chi^2_{Bures}(\rho,\sigma) = 2\int_0^\infty \mathrm{Tr}\, \rho \exp(-t\omega)\rho \exp(-t\omega)\, dt - 1,$$

see Example 1 in [26].

The monotonicity and the classical inequality (19) imply

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

(when the conditional expectation onto the commutative algebra generated by $\rho-\sigma$ is used).

3 Extension of monotone metric

As an extension of the operator (13), define $\mathbb{J}^f_{D_1,D_2} : \mathbf{M}_n \to \mathbf{M}_n$ as

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

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

$$S^A_f(\rho_1\|\rho_2) = \langle A, \mathbb{J}^f_{\rho_1,\rho_2} A\rangle.$$

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

The monotone metrics contains $(\mathbb{J}^f_{\rho,\rho})^{-1}$, therefore we consider the inverse

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

In this chapter β is completely positive trace preserving mapping between matrix spaces.

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

$$\beta^*(\mathbb{J}^f_{\beta(D_1),\beta(D_2)})^{-1}\beta \le (\mathbb{J}^f_{D_1,D_2})^{-1} \tag{20}$$

and

$$\beta \mathbb{J}^f_{D_1,D_2}\beta^* \le \mathbb{J}^f_{\beta(D_1),\beta(D_2)} \tag{21}$$

are equivalent.

Proof: The following inequalities are equivalent forms of (20):

$$(\mathbb{J}^f_{D_1,D_2})^{1/2}\beta^*(\mathbb{J}^f_{\beta(D_1),\beta(D_2)})^{-1}\beta(\mathbb{J}^f_{D_1,D_2})^{1/2} \le I$$

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

$$\|(\mathbb{J}^f_{D_1,D_2})^{1/2}\beta^*(\mathbb{J}^f_{\beta(D_1),\beta(D_2)})^{-1/2}\| \le 1$$

$$(\mathbb{J}^f_{\beta(D_1),\beta(D_2)})^{-1/2}\beta(\mathbb{J}^f_{D_1,D_2})\beta^*(\mathbb{J}^f_{\beta(D_1),\beta(D_2)})^{-1/2} \le I$$

The last inequality is equivalent to (21). □

Example 3 Let $f(x) = sx + 1$. Then

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

This was studied in the paper [14], where the result

$$\beta^*(s\mathbb{L}_{\beta(D_1)} + \mathbb{R}_{\beta(D_2)})^{-1}\beta \le (s\mathbb{L}_{D_1} + \mathbb{R}_{D_2})^{-1} \tag{22}$$

was obtained. Another formulation is

$$\beta^*(\mathbb{J}^f_{\beta(D_1),\beta(D_2)})^{-1}\beta \le (\mathbb{J}^f_{D_1,D_2})^{-1} \tag{23}$$

which is equivalent to

$$\beta\mathbb{J}^f_{D_1,D_2}\beta^* \le \mathbb{J}^f_{\beta(D_1),\beta(D_2)} \tag{24}$$

due to the previous Lemma.

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

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

and

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

The Schwarz inequality

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

is needed and gives (22) and (24). □

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 $D_1, D_2, \beta(D_1)$, $\beta(D_2)$ 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 the Lemma it is enough to prove (21) for an operator monotone function. Based on the Löwner theorem, we consider $f(x) = x/(\lambda + x)$ $(\lambda > 0)$. So

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

and we need (21). The equivalent form (21) is

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

or

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

This inequality is true due to the matrix inequality

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

see [15]. □

The generalized monotone metric

$$\gamma^f_{D_1,D_2}(A, B) := \langle A, (\mathbb{J}^f_{D_1,D_2})^{-1} B\rangle \tag{25}$$

is an extension of the monotone metric which is the case $D = D_1 = D_2$. We can call it also as *monotone metric with two parameters*. (The use of this quantity is not clear to me in the moment, although the case $f(x) = 1 + sx$ appeared already in the paper [14].)

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

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

and

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

If D_1, D_2 and C commute, then

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

□

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

$$\mathbb{J}^f_{D_1,D_2} A = \sum_{i,j} m_f(\lambda_i, \mu_j) P_i A Q_j, \tag{26}$$

where m_f is the logarithmic mean. (The formula is general, it holds for all standard matrix monotone functions f.) To show that

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

is really the inverse, we compute

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

If D_1, D_2 and C commute, then

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

We can recognize that in the commuting case

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

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

Corollary 1 *For a matrix monotone function f the generalized monotone metric*

$$\langle A, (\mathbb{J}^f_{D_1,D_2})^{-1} A \rangle$$

is jointly convex function of the variables D_1 and D_2.

The difference between two parameters and one parameter is not essential if the matrix size can be changed. Let

$$D = \begin{bmatrix} D_2 & 0 \\ 0 & D_1 \end{bmatrix}, \qquad \text{and} \qquad A = \begin{bmatrix} 0 & B \\ B & 0 \end{bmatrix}. \tag{27}$$

We show that

$$\langle A, \mathbb{J}^f_D A \rangle = \langle B, \mathbb{J}^f_{D_1,D_2} B \rangle + \langle B, \mathbb{J}^g_{D_1,D_2} B \rangle$$

where $g(x) = x f(x^{-1})$. Since continuous functions can be approximated by polynomials, it is enough to check $f(x) = x^k$ trivially. The case of inverse functions is similar.

Lemma 2 *For standard operator monotone function f, we have*

$$\langle A, \mathbb{J}^f_D A \rangle = 2 \langle B, \mathbb{J}^f_{D_1,D_2} B \rangle \qquad \textit{and} \qquad \langle A, (\mathbb{J}^f_D)^{-1} A \rangle = 2 \langle B, (\mathbb{J}^f_{D_1,D_2})^{-1} B \rangle$$

for the matrices (27).

It follows that the monotonicity, Theorem 5, and the joint convexity, Corollary 1, are consequences of the one parameter case.

References

[1] H. Araki, Relative entropy of state of von Neumann algebras, Publ. RIMS Kyoto Univ. **9**(1976), 809 – 833.

[2] I. Csiszár, Information type measure of difference of probability distributions and indirect observations, Studia Sci. Math. Hungar. **2**(1967), 299–318.

[3] P. Gibilisco, F. Hiai and D. Petz, Quantum covariance, quantum Fisher information and the uncertainty principle, IEEE Trans. Inform. Theory **55**(2009), 439–443.

[4] 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.

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

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

[7] F. Hansen, Metric adjusted skew information, Proc. Natl. Acad. Sci. USA. **105**(2008), 9909–9916.

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

[9] F. Hiai and D. Petz, Riemannian geometry on positive definite matrices related to means, Lin. Alg. Appl. **430**(2009), 3105–3130.

[10] F. Hiai, M. Mosonyi and D. Petz, Monotonicity of f-divergences: A review with new results, arXiv:1008.2529.

[11] A. S. Holevo, *Probabilistic and statistical aspects of quantum theory*, North-Holland, Amsterdam, 1982.

[12] A. Jenčová and M.B. Ruskai, A unified treatment of convexity of relative entropy and related trace functions, with conditions for equality, arXiv:0903.2895

[13] W. Kumagai, A characterization of extended monotone metrics, to be published in Lin. Alg. Appl.

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

[15] E.H. Lieb and M.B. Ruskai, Some operator inequalities of the Schwarz type. Advances in Math. **12**(1974), 269–273.

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

[17] H. Nagaoka, On Fisher information on quantum statistical models, in *Asymptotic Theory of Quantum Statistical Inference*, 113–124, ed. M. Hayashi, World Scientific, 2005.

[18] M. Ohya and D. Petz, *Quantum Entropy and Its Use*, Springer-Verlag, Heidelberg, 1993. Second edition 2004.

[19] D. Petz, Quasi-entropies for states of a von Neumann algebra, Publ. RIMS. Kyoto Univ. **21**(1985), 781–800.

[20] D. Petz, Quasi-entropies for finite quantum systems, Rep. Math. Phys., **23**(1986), 57-65.

[21] D. Petz, Geometry of canonical correlation on the state space of a quantum system, J. Math. Phys. **35**(1994), 780–795.

[22] D. Petz, Monotone metrics on matrix spaces, Linear Algebra Appl. **244**(1996), 81–96.

[23] D. Petz, Covariance and Fisher information in quantum mechanics. J. Phys. A: Math. Gen. **35**(2003), 79–91.

[24] D. Petz, *Quantum Information Theory and Quantum Statistics*, Springer, Berlin, Heidelberg, 2008.

[25] D. Petz, From f-divergence to quantum quasi-entropies and their use, Entropy **12**(2010), 304–325.

[26] D. Petz and C. Ghinea, Introduction to quantum Fisher information, arXiv:1008.2417, 2010.

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

[28] 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, arXiv:1005.2358

[29] E.P. Wigner, M.M. Yanase, Information content of distributions, Proc. Nat. Acad. Sci. USA **49**(1963), 910–918.