Convexity of quantum χ^2-divergence

Frank Hansen[1]

Institute for International Education, Tohoku University, 41 Kawauchi, Aoba-ku, Sendai 980-8576, Japan

Edited by Richard V. Kadison, University of Pennsylvania, Philadelphia, PA, and approved April 29, 2011 (received for review April 21, 2011)

The general quantum χ^2-divergence has recently been introduced by Temme et al. [Temme K, Kastoryano M, Ruskai M, Wolf M, Verstrate F (2010) *J Math Phys* 51:122201] and applied to quantum channels (quantum Markov processes). The quantum χ^2-divergence is not unique, as opposed to the classical χ^2-divergence, but depends on the choice of quantum statistics. It was noticed that the elements in a particular one-parameter family $\chi^2_\alpha(\rho,\sigma)$ of quantum χ^2-divergences are convex functions in the density matrices (ρ,σ), thus mirroring the convexity of the classical $\chi^2(p,q)$-divergence in probability distributions (p,q). We prove that any quantum χ^2-divergence is a convex function in its two arguments.

quantum mechanics | monotone metric | operator monotone function

1 Introduction

The geometrical formulation of quantum statistics originates in a study by Chentsov of the classical Fisher information. Chentsov proved (1) that the Fisher–Rao metric is the only Riemannian metric, defined on the tangent space, that is decreasing under Markov morphisms. Because Markov morphisms represent coarse graining or randomization, it means that the Fisher information is the only Riemannian metric possessing the attractive property that distinguishability of probability distributions becomes more difficult when they are observed through a noisy channel.

Morozova (2) extended the analysis to quantum mechanics by replacing Riemannian metrics defined on the tangent space of the simplex of probability distributions with positive definite sesquilinear (originally bilinear) forms K_ρ defined on the tangent space of a quantum system, where ρ is a positive definite state. Customarily, K_ρ is extended to all operators (matrices) supported by the underlying Hilbert space; cf. refs. 3 and 4 for details. Noisy channels are in this setting represented by stochastic (completely positive and trace preserving) mappings, and the contraction property is replaced by the monotonicity requirement

$$K_{T(\rho)}(T(A),T(A)) \le K_\rho(A,A)$$

for every stochastic mapping $T: M_n(\mathbb{C}) \to M_m(\mathbb{C})$. Unlike the classical situation, these requirements no longer uniquely determine the metric.

We consider the following class of functions that is used to characterize monotone metrics.

Definition 1.1: $\mathscr{F}_{\rm op}$ is the class of functions $f: (0,+\infty) \to (0,+\infty)$ such that

i. f is operator monotone,
ii. $f(t) = tf(t^{-1})$ for $t > 0$,
iii. $f(1) = 1$.

By the combined efforts of Chentsov and Morozova (2), and Petz (3) it was established that a monotone metric is given on the canonical form

$$K_\rho(A,B) = \mathrm{Tr}\, A^* c(L_\rho, R_\rho) B, \qquad [1]$$

where the so-called Morozova–Chentsov function c is of the form

$$c(x,y) = \frac{1}{yf(xy^{-1})} \qquad f: \mathbf{R}_+ \to \mathbf{R}_+,$$

for a function $f \in \mathscr{F}_{\rm op}$ and c is taken in the two commuting positive definite (super) operators L_ρ and R_ρ defined by setting

$$L_\rho A = \rho A \quad \text{and} \quad R_\rho A = A\rho.$$

It is condition (ii) in the definition above that ensures symmetry of the metric in the sense that $K_\rho(A,B) = K_\rho(B,A)$ for self-adjoint A and B.

Lesniewski and Ruskai (5) gave equivalent descriptions in terms of operator convex functions and operator monotone decreasing functions. In particular

$$K_\rho(A,B) = \mathrm{Tr}\, A^* R_\rho^{-1} k(L_\rho R_\rho^{-1}) B,$$

where $k: \mathbf{R}_+ \to \mathbf{R}_+$ is an operator monotone decreasing function satisfying $k(t^{-1}) = tk(t)$ for $t > 0$ and $k(1) = 1$. The Morozova–Chentsov–Petz formalism is then recovered by inserting the operator monotone function $f(t) = 1/k(t)$ in Eq. **1**.

Definition 1.2: The χ^2_f-divergence (relative to a choice of monotone metric) is given by

$$\chi^2_f(\rho,\sigma) = K^c_\sigma(\rho-\sigma, \rho-\sigma),$$

where

$$c(x,y) = \frac{1}{yf(xy^{-1})} \qquad x,y > 0$$

is the Morozova–Chentsov function specified by a function $f \in \mathscr{F}_{\rm op}$.

The functions

$$f_\alpha(t) = \frac{2t^\alpha}{t^{2\alpha-1}+1} \qquad t > 0$$

with parameter $\alpha \in [0,1]$ are elements in $\mathscr{F}_{\rm op}$ and correspond to the functions

$$k_\alpha(t) = \frac{1}{2}(t^{-\alpha} + t^{\alpha-1}) \qquad t > 0$$

in the Ruskai–Lesniewski formalism. The associated Morozova–Chentsov functions are given by

$$c_\alpha(x,y) = \frac{1}{yf_\alpha(xy^{-1})} = \frac{x^{\alpha-1}y^{-\alpha} + x^{-\alpha}y^{\alpha-1}}{2},$$

and the divergences for positive definite density matrices are

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

This expression is convex in (ρ,σ) as pointed out in ref. 6.

Author contributions: F.H. designed research, performed research, and wrote the paper.

The author declares no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: frank.hansen@m.tohoku.ac.jp.

 www.pnas.org/cgi/doi/10.1073/pnas.1106423108

2 Convexity

Consider a function $f\colon \mathbf{R}_+ \to \mathbf{R}_+$. The perspective (function) of f is the function g of two variables defined by setting

$$g(t,s) = sf(ts^{-1}) \qquad t,s > 0.$$

Effros showed (7) that if f is operator convex then, whenever meaningful, the perspective is a convex operator function of two variables. In particular, it becomes operator convex in the sense of Korányi (8). It is also convex in functions of commuting operators. In particular, the function

$$(\rho,\sigma) \to g(L_\rho, R_\sigma)$$

is convex in pairs of positive definite $n \times n$ matrices, equivalent to the statement that the function

$$(\rho,\sigma) \to \operatorname{Tr} A^* g(L_\rho, R_\sigma) A$$

is convex for any $n \times n$ matrix A. Similar statements are valid also for operator concave functions.

Theorem 2.1. *The $\chi_f^2(\rho,\sigma)$-divergence is convex in (ρ,σ) for any f in $\mathscr{F}_{op}$.*

Proof: Let us consider a function f in $\mathscr{F}_{op}$ with Morozova–Chensov function

$$c(x,y) = \frac{1}{yf(xy^{-1})} = F(x,y)^{-1} \qquad x,y > 0,$$

where $F(x,y) = yf(xy^{-1})$ is the perspective of f. Because f is operator concave we obtain that $F(x,y)$ is operator concave as a function of two variables. Inversion (of super operators) is decreasing. By using linearity of the mappings $\sigma \to L_\sigma$ and $\sigma \to R_\sigma$ we therefore obtain

$$\begin{aligned} F(L_{\lambda\sigma_1+(1-\lambda)\sigma_2}, R_{\lambda\sigma_1+(1-\lambda)\sigma_2})^{-1} &= F(\lambda L_{\sigma_1} + (1-\lambda)L_{\sigma_2}, \lambda R_{\sigma_1} \\ &\quad + (1-\lambda)R_{\sigma_2})^{-1} \\ &\le (\lambda F(L_{\sigma_1}, R_{\sigma_1}) \\ &\quad + (1-\lambda)F(L_{\sigma_2}, R_{\sigma_2}))^{-1} \end{aligned}$$

for states (density matrices) ρ_1, ρ_2 and σ_1, σ_2, and real numbers $\lambda \in [0,1]$. The divergence $\chi_f^2(\rho,\sigma)$ is given on the form

$$\chi_f^2(\rho,\sigma) = \operatorname{Tr}(\rho-\sigma)c(L_\sigma, R_\sigma)(\rho-\sigma),$$

and by setting

$$\rho = \lambda\rho_1 + (1-\lambda)\rho_2 \quad \text{and} \quad \sigma = \lambda\sigma_1 + (1-\lambda)\sigma_2$$

we obtain the inequality

$$\begin{aligned} \chi_f^2(\rho,\sigma) &= \operatorname{Tr}(\rho-\sigma)F(L_\sigma, R_\sigma)^{-1}(\rho-\sigma) \\ &= \operatorname{Tr}(\rho-\sigma)F(L_{\lambda\sigma_1+(1-\lambda)\sigma_2}, R_{\lambda\sigma_1+(1-\lambda)\sigma_2})^{-1}(\rho-\sigma) \\ &\le \operatorname{Tr}(\rho-\sigma)(\lambda F(L_{\sigma_1}, R_{\sigma_1}) + (1-\lambda)F(L_{\sigma_2}, R_{\sigma_2}))^{-1}(\rho-\sigma) \\ &= \operatorname{Tr}(\lambda(\rho_1-\sigma_1) + (1-\lambda)(\rho_2-\sigma_2)) \\ &\quad \times (\lambda F(L_{\sigma_1}, R_{\sigma_1}) + (1-\lambda)F(L_{\sigma_2}, R_{\sigma_2}))^{-1} \\ &\quad \times (\lambda(\rho_1-\sigma_1) + (1-\lambda)(\rho_2-\sigma_2)) \\ &\le \lambda \operatorname{Tr}(\rho_1-\sigma_1)F(L_{\sigma_1}, R_{\sigma_1})^{-1}(\rho_1-\sigma_1) \\ &\quad + (1-\lambda)\operatorname{Tr}(\rho_2-\sigma_2)F(L_{\sigma_2}, R_{\sigma_2})^{-1}(\rho_2-\sigma_2) \\ &= \lambda\chi_f^2(\rho_1,\sigma_1) + (1-\lambda)\chi_f^2(\rho_2,\sigma_2), \end{aligned}$$

where we in the second inequality, applied on super operators, used that the mapping

$$(A,\xi) \to (\xi \mid A^{-1}\xi)$$

is jointly convex for positive invertible operators A on a Hilbert space H, and vectors $\xi \in H$; cf. Proposition 4.3 in ref. 9. This is also a direct consequence of convexity of the mapping

$$(A,B) \to B^* A^{-1} B$$

for B arbitrary and A positive definite (cf. Remark after Theorem 1 in ref. 10 and Remark 4.5 in ref. 9). Furthermore, it is related to Theorem 3.1 in ref. 11.

Any function f in $\mathscr{F}_{op}$ satisfies the inequalities

$$\frac{2t}{t+1} \le f(t) \le \frac{t+1}{2} \qquad t > 0,$$

where the smallest function in $\mathscr{F}_{op}$ corresponds to the Bures metric. We mention the following characterization (4, 12, 13) of the functions in $\mathscr{F}_{op}$.

Theorem 2.2. *A function f in $\mathscr{F}_{op}$ admits a canonical representation*

$$f(t) = \frac{1+t}{2}\exp\left[-\int_0^1 \frac{(1-\lambda)(1-t)^2}{(\lambda+t)(1+\lambda t)(1+\lambda)} h(\lambda)\, d\lambda\right], \qquad \textbf{[2]}$$

where the weight function $h\colon [0,1] \to [0,1]$ is measurable. The equivalence class containing h is uniquely determined by f. Any function on the given form is in $\mathscr{F}_{op}$.

Notice that the integral kernel is nonnegative for every $t > 0$. The representation induces an order relation $\preceq$ in $\mathscr{F}_{op}$ stronger than the pointwise order by setting $f \preceq g$ if the representing weight functions h_f and h_g satisfy $h_f \ge h_g$ almost everywhere. With this order relation $(\mathscr{F}_{op}, \preceq)$ becomes a lattice, inducing a lattice structure on the set of quantum χ^2-divergences. It is compatible with the parametrization of the Wigner–Yanase–Dyson metrics; cf. Theorem 2.8 in ref. 13.

The representation in **[2]** may be used to construct families of metrics that increase monotonously from the smallest (Bures) metric to the largest. In fact any family of weight functions that decreases monotonously from the constant 1 to the zero function will induce this property. In Proposition 3 in ref. 4, we considered the constant weight functions $h_\alpha = \alpha$ for $0 \le \alpha \le 1$ and obtained in this way a family of metrics

$$f_\alpha(t) = t^\alpha \left(\frac{1+t}{2}\right)^{1-2\alpha} \qquad t > 0$$

that decreases monotonously from the largest monotone metric down to the Bures metric for α increasing from 0 to 1. In the Lesniewski–Ruskai representation that corresponds to the functions

$$k_\alpha(\omega) = \omega^{-\alpha}\left(\frac{1+\omega}{2}\right)^{2\alpha-1},$$

as mentioned in Appendix A of ref. 6.

1. Censov N (1982) *Statistical Decision Rules and Optimal Inferences, Transl. Math. Monogr,* (American Mathematical Society, Providence, RI), Vol 53, pp 117–120.
2. Morozova E, Chentsov N (1990) Markov invariant geometry on state manifolds (Translated from Russian). *Itogi Nauki i Techniki* 36:69–102 Translated in *J Soviet Math* 56:2648–2669 (1991).
3. Petz D (1996) Monotone metrics on matrix spaces. *Linear Algebra Appl* 244:81–96.
4. Hansen F (2006) Characterizations of symmetric monotone metrics on the state space of quantum systems. *Quantum Inform Compu* 6:597–605.
5. Lesniewski A, Ruskai M (1999) Monotone Riemannian metrics and relative entropy on non-commutative probability spaces. *J Math Phys* 40:5702–5724.
6. Temme K, Kastoryano M, Ruskai M, Wolf M, Verstrate F (2010) The χ^2-divergence and mixing times of quantum Markov processes. *J Math Phys* 51:122201.
7. Effros E (2009) A matrix convexity approach to some celebrated quantum inequalities. *Proc Natl Acad Sci USA* 106:1006–1008.
8. Korányi A (1961) On some classes of analytic functions of several variables. *T Am Math Soc* 101:520–554.
9. Hansen F (2006) Extensions of Lieb's concavity theorem. *J Stat Phys* 124:87–101.
10. Lieb E, Ruskai M (1974) Some operator inequalities of the Schwarz type. *Adv Math* 12:269–273.
11. Ando T, Hiai F (2010) Operator log-convex functions and operator means. *Math Ann*, 10.1007/s00208-010-0577-4.
12. Hansen F (2008) Metric adjusted skew information. *Proc Natl Acad Sci USA* 105:9909–9916.
13. Audenaert K, Cai L, Hansen F (2008) Inequalities for quantum skew information. *Lett Math Phys* 85:135–146.