On maximization of measured $f$-divergence between a given pair of quantum states

Keiji Matsumoto
National Institute of Informatics,
2-1-2, Hitotsubashi, Chiyoda-ku, Tokyo 101-8430
keiji@nii.ac.jp

March 6, 2015

Abstract

This paper deals with maximization of classical $f$-divergence between the distributions of a measurement outputs of a given pair of quantum states. $f$-divergence $D_f$ between the probability density functions $p_1$ and $p_2$ over a discrete set is defined as $D_f(p_1||p_2) := \sum_x p_2(x) f\left(\frac{p_1(x)}{p_2(x)}\right)$. For example, Kullback-Leibler divergence and Renyi type relative entropy are well-known examples with good operational meanings. Thus, finding the maximal value $D_{f_{\text{min}}}$ of measured $f$-divergence is also an interesting question. But so far the question is solved only for very restricted example of $f$. The purposes of the present paper is to advance the study further, by investigating its properties, rewriting the maximization problem to more tractable form, and giving closed formulas of the quantity in some special cases. One of the two core results is Theorem 9 if the convex conjugate $f^*$ of $f$ is operator convex, the optimal measured $f$-divergence is written as the supremum of a concave function of Hermitian operator. The other core result is Theorem 10 if $f^*$ is operator convex, $\text{dom } f^*$ is unbounded from below, and the first argument $\rho_1$ is not full-rank, one can reduce the optimization problem to the one in the support of $\rho_1$. An application of this theorem to the case where $\rho_1$ is pure leads to a closed formula for the optimal measured $f$-divergence. Using these results, we analyze $D_{f_{\text{min}}}$ between two very close states up to the second order of their distance. It have been a folklore that this limit equals the constant multiple of SLD (symmetric logarithmic derivative) Fisher information, which plays an important role in the asymptotic theory of statistics. We prove the folklore in the case where the rank of states is either full or one, and disprove it in other cases, giving the correct formula.

1 Introduction

This paper deals with maximization of classical $f$-divergence between the distributions of a measurement outputs of a given pair of quantum states. $f$-
divergence $D_f$ between the probability density functions $p_1$ and $p_2$ over a discrete set is defined as

$$D_f (p_1||p_2) := \sum_x p_2(x) f \left( \frac{p_1(x)}{p_2(x)} \right),$$

if $p_2(x) > 0$ for all $x$. (The definition for the general case is given later.)

This problem is important since $D_f$ has good operational meanings. For example,

$$f_{KL} (\lambda) := \lambda \ln \lambda, \quad f_{KL_2} (\lambda) = -\ln \lambda$$
correspond to Kullback-Leibler divergence. Also,

$$f_\alpha (\lambda) := \begin{cases} \lambda^\alpha, & (\alpha \leq 0, \alpha \geq 1) \\ -\lambda^\alpha, & (0 < \alpha < 1) \end{cases} \quad (1)$$
correspond to Rnyi-type relative entropy (or its monotone function). They play key role in the theory of large deviation, and thus extensively used in asymptotic analysis of error probability of decoding, hypothesis test, and so on.

Other $f$-divergences than above have also a good operational meaning. If $f$ is a proper closed convex function $f$ whose domain contains positive half-line, $D_f (p_1||p_2)$ is the optimal gain of a certain Bayes decision problem. In other words, for each $f$ satisfying above mentioned conditions, there is a pair of real valued functions $w_1$ and $w_2$ on decision space representing a gain of decision $d$, with

$$D_f (p_1||p_2) = \sup_{d(\cdot)} \sum_x (w_1 (d(x)) p_1(x) + w_2 (d(x)) p_2(x)). \quad (2)$$

Conversely, for each $\vec{w}(\cdot) = (w_1(\cdot), w_2(\cdot))$, there is a proper closed convex function $f$ satisfying the above identity. Also, by (2) and the celebrated randomization criterion [6][7], there is a Markov map which sends $(p, q)$ to $(p', q')$ if and only if $D_f (p||q) \geq D_f (p'||q')$ holds for any proper closed convex function $f$ with dom $f \supset (0, \infty)$.

In quantum information and statistics, we make decisions based on the data out of measurements performed on given quantum states. Thus, finding the optimal measured $f$-divergence $D_{\text{min}}^f$ is of interest:

$$D_{\text{min}}^f (\rho_1||\rho_2) := \sup_{M: \text{POVM}} D_f (P^M_{\rho_1}||P^M_{\rho_2}), \quad (3)$$

where POVM is the short for positive operator valued measures, and $P^M_{\rho_0}$ is the probability distribution obtained by applying the measurement $M$ to the state $\rho_0 (\theta \in \{1, 2\})$. (Underlying Hilbert space, denoted by $\mathcal{H}$, is finite dimensional throughout the paper.) A trouble with $D_{\text{min}}^f (\rho_1||\rho_2)$ is that the definition involves intractable maximization of non-linear functional of POVM. This optimization problem had been so far solved only for $f_{1/2}$, and $f (\lambda) = |1 - \lambda|$, The former and the latter corresponds to fidelity and statistical distance, respectively.
The purposes of the present paper is to advance the study further, by investigating its properties, rewriting the maximization problem to more tractable form, and giving closed formulas of the quantity in some special cases. One of the two core results is Theorem 6: if the convex conjugate $f^*$ of $f$ is operator convex, $D_{f}^{\min} (\rho_1 || \rho_2)$ is written as the supremum of a concave function of Hermitian operator in $H$ (If the operators could move extended Hilbert space, the result would not have been as useful), namely,

$$D_{f}^{\min} (\rho_1 || \rho_2) = \sup \{ \text{tr} \rho_1 T - \text{tr} \rho_2 f^*(T); \text{spec} T \subset \text{dom} f^* \}$$

This minimization problem is straightforwardly solved $f = f_2$ and $f_{-1}$.

The other core result is Theorem 10: if $f^*$ is operator convex, $\text{dom} f^*$ is unbounded from below, and the kernel of the first argument $\rho_1$ is non-trivial, one can reduce the optimization problem to the one in $\text{supp} \rho_1$, namely,

$$D_{f}^{\min} (\rho_1 || \rho_2) = D_{f}^{\min} (\rho_1 || \pi \rho_1 \rho_2 \pi \rho_1) + f^*(-\infty) (1 - \text{tr} \rho_2 \pi \rho_1)$$

where $\pi \rho_1$ is the projection onto $\text{supp} \rho_1$. An application of this leads to a closed formula for the case where $\rho_1$ is a pure state.

Using above results, we analyze $D_{f}^{\min}$ between two very close states, up to the second order of the distance between the states. More explicitly, we computed the limit

$$\lim_{\eta' \to \eta} \frac{1}{(\eta' - \eta)^2} D_{f}^{\min} (\rho_\eta || \rho_{\eta'})$$

where $\{ \rho_\eta \}_{\eta \in \mathbb{R}}$ is a family of parameterized states and $f$ has good properties. It have been a folklore that this limit equals the constant multiple of SLD (symmetric logarithmic derivative) Fisher information, which plays an important role in the asymptotic theory of statistics. We prove the folklore in the case where $\rho_\eta$'s rank is either full or one, and disprove it in other cases, giving the alternative correct formula.

2 Classical $f$-divergence

In this section, we summarize the known or trivial facts about classical $f$-divergence and convex functions. Following [5], we suppose that $f$ is a map from $\mathbb{R}$ to $\mathbb{R} \cup \{ \pm \infty \}$. Instead of saying that $f$ is not defined on a certain set, we say that $f(\lambda) = \infty$ on that set. The effective domain of $f$, denoted by $\text{dom} f$, is the set of all $\lambda$’s with $f(\lambda) < \infty$. $f$ is said to be convex if and only if the set $\{(\lambda_1, \lambda_2); \lambda_2 \geq f(\lambda_1)\}$ is convex. A convex function $f$ is proper if and only if $f$ is nowhere $-\infty$ and not $\infty$ everywhere, and is closed if and only if $\{(\lambda_1, \lambda_2); \lambda_2 \geq f(\lambda_1)\}$ is closed, or equivalently, $f$ is lower semicontinuous. Improper convex functions rarely appear in practical occasions, since they cannot be finite at any point. Also, any convex but non-closed function $f$ can be changed to a closed one by changing value at the boundary of its effective domain (recall that any convex function is continuous on the interior of its effective domain).
Thus, without damaging usefulness, only proper and closed functions are used in this paper.

Let $f$ be a proper closed convex function whose effective domain has overlap with $(0, \infty)$. Then $f$-divergence $D_f(P_1||P_2)$ between positive finite measures $P_1$ and $P_2$ is defined by

$$D_f(P_1||P_2) := \int_{\text{supp} P_2} p_1(x) f\left(\frac{p_1(x)}{p_2(x)} \right) \, d\mu(x) + P_1((\text{supp} P_2)^c) \lim_{\lambda_2 \downarrow 0} \lambda_2 f\left(\frac{1}{\lambda_2}\right),$$

where $p_1$ and $p_2$ are density function of $P_1$ and $P_2$ with respect to a measure $\mu$ which dominates them. $D_f$ apparently depends on the choice of $\mu$, but, as a matter of fact, it does not. Thus, without loss of generality, $\mu$ may be taken as $P_1 + P_2$, for example.

Below, we give a useful expression of $D_f$. Define $g(\lambda_1, \lambda_2)$ by

$$g(\lambda_1, \lambda_2) := \begin{cases} 
\lambda_2 f\left(\frac{\lambda_1}{\lambda_2}\right), & \text{if } \lambda_1 \in \text{dom } f, \lambda_2 > 0 \\
\lim_{\lambda_2 \downarrow 0} \lambda_2 f\left(\frac{\lambda_1}{\lambda_2}\right), & \text{if } \lambda_1 \in \text{dom } f, \lambda_2 = 0, \\
0, & \text{if } \lambda_1 = \lambda_2 = 0, \\
\infty, & \text{if } \lambda_1 \notin \text{dom } f \text{ or } \lambda_2 < 0.
\end{cases}$$

If $p_2(x) = 0$ and $p_1(x) > 0$,

$$g(p_1(x), p_2(x)) = \lim_{\lambda_2 \downarrow 0} \lambda_2 f\left(\frac{p_1(x)}{\lambda_2}\right),$$

and if $p_1(x) = 0$ and $p_2(x) = 0$, $g(p_1(x), p_2(x)) = 0$. Therefore, $f$-divergence $D_f(P_1||P_2)$ positive finite measures $P_1$ and $P_2$ is given by

$$D_f(P_1||P_2) = \int g(p_1(x), p_2(x)) \, d\mu(x).$$

Since $g$ is positively homogeneous, or

$$\forall a \geq 0, \ g(a\lambda_1, a\lambda_2) = ag(\lambda_1, \lambda_2),$$

it is obvious the value of $D_f$ does not depends on the choice of $\mu$.

In addition to being positively homogeneous, $g$ is proper, closed, and convex (see p.35 and p.67 of [5]). Therefore, by Corollary 13.5.1 of [5], it is the pointwise supremum of linear functions,

$$g(\lambda_1, \lambda_2) = \sup_{\vec{w} \in \mathcal{W}_f} \sum_{\theta \in \{1, 2\}} w_\theta \lambda_\theta, \quad (4)$$

where

$$\mathcal{W}_f = \left\{ \vec{w} = (w_\theta)_{\theta \in \{1, 2\}} : w_1 \leq w_1', w_2 \leq -f^{*}(w_1'), \exists w_1' \in \text{dom } f^{*} \right\},$$

4
and \( f^* \) is convex conjugate of \( f \),

\[
  f^* (t) := \sup_\lambda (t \lambda - f (\lambda)).
\]

Therefore,

\[
  D_f (P_1 || P_2)
  = \int \left( \sup_{\bar{w} \in W} \sum_{\theta \in \{1, 2\}} w_\theta p_\theta (x) \right) d\mu (x)
  = \sup \left\{ \int \sum_{\theta \in \{1, 2\}} w_\theta (x) p_\theta (x) d\mu (x) : \bar{w} (\cdot) : \text{measurable, into } W_f \right\}. \tag{5}
\]

**Remark 1** No indicates (2). To see this, use \( W_f \) as a decision space, and let the gain of the decision \( \bar{w} \) is \( w_\theta \) when true probability distribution is \( P_\theta \).

The correspondence between \( D_f \) and \( W_f \) is one-to-one. But two different convex functions \( f \) and \( f_0 \) may corresponds to the same \( f \)-divergence, \( D_f = D_{f_0} \), since the values of \( f \) in the negative half line does not affect the value of \( D_f \). To see this redundancy in terms of the correspondence between \( f \) and \( W_f \), observe that \( w_2 = -f^* (w_1) \) defines upper-right border of \( W_f \) and \( W_f \) is unbounded towards lower-left. Hence, the part of \( f^* \) which is not monotone increasing does not affect the definition of \( W_f \).

Thus, we call the choice of \( f \) is standard if \( f^* \) is (strictly) monotone increasing on \( \text{dom} f^* \). If \( f^* \) is not standard, we define \( f_0^* \) by

\[
  f_0^* (t) := \begin{cases} f^* (t), & \text{if } \forall t' > t, f^* (t') > f^* (t), \\ +\infty, & \text{otherwise}, \end{cases}
\]

and define \( f_0 \) by \( f_0 := f_0^{**} \). Recall that \( f \) used in the definition of \( f \)-divergence is not always \( +\infty \) on the positive half line. Therefore, \( \text{dom} f_0^* \) is not empty.

Below, \( f_0 \) is written directly in terms of \( f \), without refering to their cojugate functions. Denote by \( f'_+ \) the right derivative of \( f \). Then if \( f'_+ (0) \) is finite,

\[
  f_0 (\lambda) = \begin{cases} f (\lambda), & \text{if } \lambda \geq 0, \\ f'_+ (0) t + f (0), & \text{if } \lambda < 0. \end{cases}
\]

Otherwise,

\[
  f_0 (\lambda) = \begin{cases} f (\lambda), & \text{if } \lambda \geq 0, \\ +\infty, & \text{if } \lambda < 0. \end{cases}
\]

To see that \( f_0 \) and \( f \) coincide on the positive half-line, observe that \( f^* \) is monotone non-increasing in the region below \( \text{dom} f_0^* \). Therefore, in that region, \( t \to t\lambda - f^* (t) \) is monotone non-decreasing if \( \lambda \geq 0 \). Thus,

\[
  f (\lambda) = \sup_{t \in \text{dom} f_0^*} t\lambda - f^* (t)
  = \sup_{t \in \text{dom} f_0^*} t\lambda - f_0^* (t)
  = f_0 (\lambda).
\]
Let $\lambda < 0$. If $\text{dom} \ f^*$ is not bounded from below, $f^* = f^*_0$ is monotone increasing and

$$f_0(\lambda) = \sup_{t \in \text{dom} f^*_0} t \lambda - f^*(t) = -\inf_{t \in \text{dom} f^*_0} \{(\lambda) t + f^*(t)\}$$

$$= -\lim_{t \to -\infty} \{(\lambda) t + f^*(t)\} = \infty.$$ 

On the other hand, suppose $\text{dom} f^*_0$ is bounded from below, and let $t_0$ be the largest lower bound of $\text{dom} f^*_0$. Then since $f^*_0$ monotone increasing on its effective domain,

$$f^*_0(t_0) < \infty$$

and

$$f_0(\lambda) = \sup_{t \in \text{dom} f^*_0} t \lambda - f^*(t) = -\inf_{t \in \text{dom} f^*_0} \{(\lambda) t + f^*(t)\}$$

$$= \lambda t_0 - f^*(t_0) = \lambda t_0 + f(0).$$

It remains to show that $t_0$ equals right derivative of $f(\lambda)$ at 0,

$$t_0 = f'_{+}(0). \quad (6)$$

To show this, suppose the contrary is true. Since $f_0$ is convex, $t_0 < f'_{+}(0)$. Here note $f'_{+}(0)$ is finite since $t_0 > -\infty$ and $f$ is finite at some point on the positive half line. Let $t$ be a real number between $t_0$ and $f'_{+}(0)$. Then

$$f^*_0(t) = \max \left\{ \sup_{\lambda > 0} (t \lambda - f(\lambda)), \sup_{\lambda \leq 0} (t \lambda - (\lambda t_0 + f(0))) \right\}$$

$$\leq \max \left\{ \sup_{\lambda > 0} ((t - f'_{+}(0)) \lambda - f(0)), \sup_{\lambda \leq 0} ((t - t_0) \lambda - f(0)) \right\}$$

$$= -f(0) = f^*(t_0) = f^*_0(t_0).$$

This contradicts with the requirent that $f^*_0$ should be monotone increasing in its effective domain. Thus we should have (6).

For example, both of

$$f_{TV}(\lambda) := |1 - \lambda|,$$

$$f_{TV_2}(\lambda) := \begin{cases} |1 - \lambda|, & \lambda \geq 0, \\ \infty, & \lambda < 0, \end{cases}$$

correspond to the total variation distance,

$$D_{f_{TV}}(P||Q) = D_{f_{TV_2}}(P||Q) = \|P - Q\|_1.$$ 

The former is standard but the latter is not,

$$f^*_{TV}(t) = \begin{cases} \infty, & t < -1, \\ t, & -1 \leq t \leq 1, \\ \infty, & t > 1, \end{cases}$$

$$f^*_{TV_2}(t) = \begin{cases} -1, & t < -1, \\ t, & -1 \leq t \leq 1, \\ \infty, & t > 1, \end{cases}$$
Define
\[ \hat{f}(\lambda) := g(1, \lambda), \quad \lambda \geq 0, \]
and extend it to the negative half line so that it becomes standard. Also, define \( \hat{g} \) by in a parallel manner as the definition of \( g \), replacing \( f \) by \( \hat{f} \). Then,
\[ \hat{g}(\lambda_1, \lambda_2) = g(\lambda_2, \lambda_1), \quad (7) \]
holds for all \( \lambda_1 \geq 0 \), as will be checked below.

If \( \lambda_1 \in \text{dom} \hat{f} \) and \( \lambda_2 > 0 \), the relation is checked by easy computation. If \( \lambda_1 = \lambda_2 = 0 \), the both ends of (7) is 0. Finally, if \( \lambda_1 = 0 \notin \text{dom} \hat{f} \) and \( \lambda_2 > 0 \),
\[ \hat{g}(0, \lambda_2) = \infty = \hat{f}(0) = \lambda_2 \hat{f}(0) = \lambda_2 g(1, 0) = g(\lambda_2, 0), \]
and the relation is checked. (7) means
\[ D_f(P_1||P_2) = D_{\hat{f}}(P_2||P_1). \]

3 Necessary and sufficient conditions for \( D_f^{\min} < \infty \)

Now we proceed to a quantum analogue of \( f \)-divergence. Let \( \{\rho_\theta\}_{\theta \in \{1, 2\}} \) be a pair of positive operators in a finite dimensional Hilbert space \( \mathcal{H} \). We define \( D_f^{\min} \) by (2). The notation comes from the following fact. If \( D_f^{Q} \) is a real valued function of \( \{\rho_\theta\}_{\theta \in \{1, 2\}} \) which coincide with \( D_f \) on any commutative subalgebra and is monotone non-increasing by application of unital completely positive (CP) maps, then
\[ D_f^{\min}(\rho_1||\rho_2) \leq D_f^{Q}(\rho_1||\rho_2). \]
(If the statement is untrue, it obviously contradict with the monotonicity of \( D_f^{Q} \).)

**Remark 2** Here, \( D_f \) and \( D_f^{\max} \) have operational meanings only if the entities are normalized. But extending them to general positive finite measures and positive operators makes mathematical treatment easier. Since we sometimes consider submatrices of inputs states, especially in the case that measurement process is involved.

In this section we determine the case where \( D_f^{\min} \) stays finite. Note this result also give, due to the above mentioned fact, the sufficient condition that all the quantum versions of \( f \)-divergence becomes infinite. Before analyzing non-trivial cases, we first present the cases where \( D_f^{\min} \) is finite trivially.
The first trivial case is the one where $\mathcal{W}_f$ is "bounded from above", or there are finite numbers $a_1$ and $a_2$ with
\[ \mathcal{W}_f \subset (-\infty, a_1] \times (-\infty, a_2]. \] (This is the case if dom $f^*$ is a finite interval.) Then by (5),
\[ D_f (P_1||P_2) \leq \max \{ a_1, a_2 \}. \]
Thus, $D_f^{\text{max}}$ is also finite. Therefore, we consider the cases where the above assumption is false.

Another trivial case is the one where $\rho_1 > 0$ and $\rho_2 > 0$. Then there is a constant $b_\ast > 0$ and $b'_\ast > 0$ ($b_\ast \geq b'_\ast$) with
\[ \rho_1 - b_\ast \rho_2 \leq 0, \] \[ \rho_1 - b'_\ast \rho_2 \geq 0, \]
respectively. For latter use, $b_\ast$ and $b'_\ast$ is defined as the smallest number with (9) and the largest number with (10), respectively. Then, denoting by $p_{\rho_0}^M$ the density of $P_{\rho_0}^M$ with respect to $P_{\rho_1}^M + P_{\rho_2}^M$,
\[ \frac{p_{\rho_1}^M (x)}{p_{\rho_2}^M (x)} \leq \sup_{0 \leq M \leq 1} \frac{\text{tr} M_{\rho_1}}{\text{tr} M_{\rho_2}} \leq b_\ast, \]
\[ \frac{p_{\rho_1}^M (x)}{p_{\rho_2}^M (x)} \geq \inf_{0 \leq M \leq 1} \frac{\text{tr} M_{\rho_1}}{\text{tr} M_{\rho_2}} \geq b'_\ast. \]
Hence, using the assumption that $f$ is finite on $(0, \infty)$, we have the asserted statement. Thus, the question is the case where at least one of them is not full-rank.

Suppose ker $\rho_1$ is non trivial, and let $\pi$ be the projection onto ker $\rho_1$. Then by (4) and (3),
\[ D_f^{\text{max}} (\rho_1||\rho_2) \geq w_1 \text{tr} \rho_1 \pi + w_2 \text{tr} \rho_2 \pi + w'_1 \text{tr} \rho_1 (1 - \pi) + w'_2 \text{tr} \rho_2 (1 - \pi) \]
\[ = w_1 \text{tr} \rho_2 \pi + w'_1 \text{tr} \rho_1 (1 - \pi) + w'_2 \text{tr} \rho_2 (1 - \pi) \]
holds for any $w, w' \in \mathcal{W}_f$. Therefore, $D_f^{\text{max}} (\rho_1||\rho_2)$ can be finite only if $w_2$ stays finite, i.e.,
\[ \mathcal{W}_f \subset \mathbb{R} \times (-\infty, a_2]. \] (11)

On the other hand, suppose the above inclusion is true. To avoid the trivial case, suppose $\mathcal{W}_f$ is not bounded from above, and $\rho_2$ is full-rank. (If the latter is false, the parallel argument as above shows $D_f^{\text{max}} = \infty$.) By the latter assumption, there is $b_\ast > 0$ with (9). Let $t_\ast$ be the number such that $b_\ast$ is a subgradient of $f^*$. Then by the convexity of $f^*$ and (4), we have
\[ D_f^{\text{min}} (\rho_1||\rho_2) \leq \sup_{\{M_t\}} \sum_t (t \text{tr} \rho_1 M_t + (-b_\ast (t - t_\ast) - f^* (t_\ast)) \text{tr} \rho_2 M_t) \]
\[ = \sup_{\{M_t\}} \sum_t t \text{tr} (\rho_1 - b_\ast \rho_2) M_t + b_\ast t_\ast \text{tr} \rho_1 - f^* (t_\ast) \text{tr} \rho_2 \]
\[ \leq b_\ast t_\ast \text{tr} \rho_1 - f^* (t_\ast) \text{tr} \rho_2, \]
Exchanging $\rho_1$ and $f^*$ by $\rho_2$ and $f^{*\perp}$ (here, we suppose $f$ is standard without loss of generality), we have all the cases, summarized as below.

**Theorem 3** Suppose $f$ is a proper closed convex function with $\text{dom } f \supset (0, \infty)$. Then, $D_{\rho_1}^{\max} (\rho_1 || \rho_2) < \infty$ holds if $\rho_1 > 0$ and $\rho_2 > 0$ hold. Also, $D_{\rho_1}^{\max} (\rho_1 || \rho_2) < \infty$ is equivalent to (3) if $\ker \rho_1 \neq \{0\}$ and $\ker \rho_2 \neq \{0\}$ hold, to (11) if $\ker \rho_1 \neq \{0\}$ and $\rho_2 > 0$ hold, and to

$$W_f \subset (-\infty, a_1] \times \mathbb{R},$$

if $\rho_1 > 0$, $\ker \rho_2 \neq \{0\}$.

Note (11) and (12) is equivalent to $f^* (-\infty) > -\infty$

and

$$\text{dom } f^* \subset (-\infty, a_1],$$

respectively.

**4 Continuity of $D_{\rho}^{\min}$**

Since $D_{\rho}^{\min}$ is jointly convex almost by definition, it is continuous in the interior of $\text{dom } D_{\rho}^{\min}$, or at the points where $\rho_1 > 0$ and $\rho_2 > 0$. By applying well-known facts in convex analysis, a sort of continuity at the edge is easily proved. For a given convex function $f$, we denote $\text{epi } f := \{ (\lambda_1, \lambda_2) ; \lambda_2 \geq f (\lambda_1) \}$.

The definition (3) is rewritten as, using (5),

$$\sup_{M'} D_f (P_{\rho_1}^{M'} || P_{\rho_2}^{M'}) = \sup_{M', \tilde{w}} \left\{ \int \sum_{\theta \in \{1,2\}} w_\theta (x) \, dP_{\rho_\theta}^{M'} (x) ; \tilde{w} (\cdot) : \text{measurable, into } W_f \right\}$$

$$= \sup_{M', \omega} \left\{ \int \sum_{\theta \in \{1,2\}} w_\theta (x) \, \text{tr} \, \rho_\theta M' (dx) ; \tilde{w} (\cdot) : \text{measurable, into } W_f \right\}$$

$$= \sup_{M} \left\{ \sum_{\theta \in \{1,2\}} \text{tr} \left( \rho_\theta \int_{\omega \in W_f} w_\theta M (d\tilde{\omega}) \right) \right\},$$

where $M (B) := M' (\tilde{\omega}^{-1} (B))$ is a POVM over the topological $\sigma$-field on $W_f \subset \mathbb{R}^2$.

Let $\mathcal{M}_{\mathcal{H}, f}$ be the set of pairs of operators such that

$$(W_\theta)_{\theta \in \Theta} \in \mathcal{M}_{\mathcal{H}, f}$$

$$\Leftrightarrow \exists M : \text{POVM } W_\theta \leq \int_{\tilde{\omega} \in W_f} w_\theta M (d\tilde{\omega}).$$
Then

$$D_f^{\min}(\rho_1||\rho_2) := \sup_M D_f(P_{\rho_1}^M||P_{\rho_2}^M)$$

$$= \sup_{\vec{W} \in \mathcal{W}_{H,f}} \sum_{\theta \in \{1,2\}} \text{tr}(\rho_\theta W_\theta),$$

(13)

where

$$\vec{W} = (W_\theta)_{\theta \in \Theta} = (W_1, W_2).$$

**Lemma 4** $D_f^{\min}$ is a proper closed convex function which is positively homogeneous.

**Proof.** That $D_f^{\min}$ is proper, convex and positively homogeneous is trivial. So we prove it is closed. By (13),

$$\text{epi} D_f^{\min} = \text{epi} \sup_{\vec{W} \in \mathcal{W}_{H,f}} h_{\vec{W}} = \bigcap_{(W_\theta)_{\theta \in \Theta} \in \mathcal{W}_{H,f}} \text{epi} h_{\vec{W}},$$

where

$$h_{\vec{W}}(\rho_1, \rho_2) := \sum_{\theta \in \{1,2\}} \text{tr}(\rho_\theta W_\theta).$$

Since each $\text{epi} h_{\vec{W}}$ is closed, $\bigcap_{(W_\theta)_{\theta \in \Theta} \in \mathcal{W}_{H,f}} \text{epi} h_{\vec{W}}$ is also closed. Thus we have the assertion. ■

**Proposition 5** For any $\rho_1 \geq 0$ and $\rho_2 \geq 0$, and for any $X_1, X_2 \geq 0$,

$$\lim_{s \downarrow 0} D_f^{\min}(\rho_1 + sX_1||\rho_2 + sX_2) = D_f^{\min}(\rho_1||\rho_2).$$

**Proof.** Since $D_f^{\min}$ is closed, it is continuous on any (finite dimensional) simplex inside dom $D_f^{\min}$ by Theorem 10.2 of [5]. Applying this fact to the line segment connecting $(\rho_1 + sX_1, \rho_2 + sX_2)$ and $(\rho_1, \rho_2)$, we have the assertion. ■

5 Simplifying the optimization problem

Define $\mathcal{W}'_{H,f}$ as the set of pairs of operators such that $(W_\theta)_{\theta \in \Theta} \in \mathcal{W}'_{H,f}$ is equivalent to the existence of a measurable set and a projection valued measure (PVM) $E$ over it in the Hilbert space $\mathcal{H}$ with

$$W_\theta \leq \int_{w \in W_f} w_\theta E(dw).$$

Obviously, $\mathcal{W}'_{H,f} \subset \mathcal{W}_{H,f}$ and, by Naimark extension theorem, for any $(W_\theta)_{\theta \in \Theta} \in \mathcal{W}_{H,f}$, there is a separable Hilbert space $\mathcal{K}$ and an isometry from $\mathcal{K}$ onto $\mathcal{H}$ such that

$$\vec{W} \in V^\dagger \mathcal{W}'_{\mathcal{K},f} V.$$
Therefore, if
\[ V^\dagger \mathfrak{M}_{\mathcal{K},f} V \subset \mathfrak{M}_{\mathcal{H},f} \]  
for any separable Hilbert space \( \mathcal{K} \) and any isometry \( V \) from \( \mathcal{K} \) onto \( \mathcal{H} \), we have
\[ D_f^{\min}(\rho_1||\rho_2) = \sup_{\tilde{W} = \in \mathfrak{M}_{\mathcal{H},f}} \sum_{\theta \in \{1,2\}} \text{tr} (\rho_\theta W_\theta). \]

If one of the following two are true, it is easy to prove that this is the case:

**I**  \( f^\ast \) is operator convex on \( \text{dom} f^\ast \)

**II**  \( \hat{f}^\ast (t) := -f^\ast (-t) \) is operator convex on \( \text{dom} \hat{f}^\ast \)

Observe \( (W_\theta)_{\theta \in \Theta} \in V^\dagger \mathfrak{M}_{\mathcal{K},f} V \) if and only if
\[
W_1 \leq V^\dagger \int_{\vec{w} \in \mathcal{W}_f} w_1 E (d\vec{u}) V = V^\dagger \int_{t \in \text{dom} f^\ast} t E^\prime (dt) V = V^\dagger TV
\]
\[
W_2 \leq -V^\dagger \int_{\vec{w} \in \mathcal{W}_f} w_2 E (d\vec{u}) V
\]
\[
\leq -V^\dagger \int_{t \in \text{dom} f^\ast} f^\ast (t) E^\prime (dt) V = -V^\dagger f^\ast (T) V,
\]
where \( E \) is a PVM in \( \mathcal{K} \), \( E^\prime \) is the marginal of \( E \), and
\[
T := \int_{t \in \text{dom} f^\ast} t E^\prime (dt).
\]

If (I) holds, or \( f^\ast \) is operator convex, by Jensen-type inequality,
\[
-V^\dagger f^\ast (T) V \leq -f^\ast (V^\dagger TV).
\]

Therefore, defining a PVM \( E'' \) by the spectral decomposition \( V^\dagger TV = \int_{s \in \text{dom} f^\ast} s E'' (ds) \),
\[
W_1 \leq \int s E'' (ds), \quad W_2 \leq -\int f^\ast (s) E'' (ds).
\]

Also, observe
\[
\text{spec} V^\dagger W_1 V \subset \text{spec} W_1 \subset s \in \text{dom} f^\ast.
\]

Thus, (14) is verified. In the case that (II) holds, almost in the parallel manner (14) is checked.

\( \mathfrak{M}_{\mathcal{H},f} \) can be expressed in the following way:
\[
\mathfrak{M}_{\mathcal{H},f} = \left\{ \tilde{W} ; W_1 \leq T, W_2 \leq -f^\ast (T), \text{spec} T \subset \text{dom} f^\ast \right\}.
\]

This is checked by noticing that \( \tilde{W} \in \mathfrak{M}_{\mathcal{H},f} \) is equivalent to
\[
W_1 \leq T,
\]
\[
W_2 \leq \int_{\vec{w} \in \mathcal{W}_f} w_2 E (d\vec{u}) \leq -\int_{\vec{w} \in \mathcal{W}_f} f^\ast (w_1) E (d\vec{u}) = -f^\ast (T).
\]
Here, $E$ is an arbitrary PVM over $\mathcal{H}$.

As we analyze later, $f_\alpha$ ($\alpha \leq \frac{1}{2}$) and $f_{KL_2}$ are the examples where (I) holds, and $f_\alpha$ ($\alpha \geq \frac{1}{2}$) and $f_{KL}$, are the examples where (II) holds. Summarizing the argument so far, we have:

**Theorem 6** Let $f$ be a proper convex function with $\text{dom } f^* \supset (0, \infty)$. If (14) is true, then $W'_{H,f} = W_{H,f}$ and thus

$$D_f^{\text{min}} (\rho_1||\rho_2) = \sup \{ \text{tr} \rho_1 T - \text{tr} \rho_2 f^* (T) ; \text{spec } T \subset \text{dom } f^* \} \quad (15)$$

$$= \sup \{ \text{tr} \rho_1 f^{*-1} (S) - \text{tr} \rho_2 S ; \text{spec } S \subset \text{dom } f^{*-1} \}.$$

Also, instead of all POVM, we only have to optimize over all PVM’s. If the supremum is achieved, it is achieved by the spectrum decomposition of Hermitian operators $T$ or $S$ achieving the maximum of the above. (14) is true if (I) or (II) is true.

### 6 The stationary point

In the section, unless otherwise mentioned, we suppose that the condition (I) holds. This means that

$$G(T) := \text{tr} \rho_1 T - \text{tr} \rho_2 f^* (T) \quad (16)$$

is concave in $T$, and its supremum equals $D_f^{\text{min}}$. The dom $G$ is

$$\{ T ; \text{spec } T \subset \text{dom } f^* \}.$$

In this section we focus on an easy case where the stationary point $T_0$ of $G$, or $T_0$ with $\partial G (T_0) / \partial T = 0$ exists in dom $G$. (Note $f^*$ is necessarily differentiable, being operator convex.) In this case, the supremum of $G$ is achieved at $T_0$.

Below, we suppose $f^*$ is differentiable. Hence the Frechet derivative $D f^* (T)$ of $f^*$ i.e., a linear transform in $\mathcal{B}(\mathcal{H})$ with

$$\frac{\| f^* (T + X) - f^* (T) - D f^* (T) (X) \|_2}{\| X \|_2} \rightarrow 0, \text{ as } \| X \|_2 \rightarrow 0.$$

is given by, in the basis which diagonalizes $T$,

$$D f^* (T) (X) = \left[ f^{*[1]} (\mu_i, \mu_j) X_{i,j} \right], \quad (17)$$

where $\mu_i$ ($i = 1, \cdots$) are eigenvalues of $T$, and

$$f^{*[1]} (s,t) := \begin{cases} \frac{f(s) - f(t)}{s - t}, & (s \neq t), \\ f'(s), & (s = t). \end{cases}$$

12
An important consequence of this formula is $D f^* (T) \langle \cdot \rangle$ is self-adjoint with respect to the inner product $\text{tr} X Y$,

$$\text{tr} Y D f^* (T) (X) = \sum_{i,j} \rho_{2,i,j} f^* [1] (\mu_i, \mu_j) X_{i,j} = \sum_{i,j} f^* [1] (\mu_i, \mu_j) \rho_{2,i,j} X_{i,j} = \text{tr} X D f^* (T) (Y).$$

With all definitions and assumptions are made, we now proceed to the analysis of the maximal point of $G (T)$. Most tractable case is that there is a stationary point of $G (T)$ in $\text{dom} G$. Therefore, if

$$\frac{d G (T_0 + s X)}{d s} \bigg|_{s=0} = \text{tr} X \rho_1 - \text{tr} \{ \rho_2 D f^* (T_0) (X) \} = \text{tr} X (\rho_1 - D f^* (T_0) (\rho_2)) = 0$$

holds for any Hermitian matrix $X$, $T_0$ achieves maximum. (Here we used the fact that $D f^* (T_0) \langle \cdot \rangle$ is self-conjugate.) Thus, we have

$$\rho_1 = D f^* (T_0) (\rho_2) \quad (18)$$

Therefore,

$$D f^\min (\rho_1 | | \rho_2) = \text{tr} \rho_1 T_0 - \text{tr} \rho_2 f^* (T_0) = \text{tr} T_0 D f^* (T_0) (\rho_2) - \text{tr} \rho_2 f^* (T_0) = \text{tr} f (f^* (T_0)) \rho_2,$$

Here, in the third identity holds since $D f^* (T_0) \langle \cdot \rangle$ is self-adjoint, the fifth identity holds due to \((17)\), and the last identity is due to

$$f (x) = f^{**} (x) = \sup_t t x - f^* (t).$$

We give a sufficient condition that the supremum is achieved in the interior of the domain, or equivalently, $T_0$ exists:

$$\sup_{t \in \text{dom} f^*} f^{**} (t) > b_*, \quad (20)$$

$$\inf_{t \in \text{dom} f^*} f^{**} (t) < b_*, \quad (21)$$
where \( b_\ast \) and \( b'_\ast \) are as of [9] and [10]. We also define \( t_\ast \) and \( t'_\ast \) by

\[
\begin{align*}
  f^{*\prime}(t_\ast) &= b_\ast, \\
  f^{*\prime}(t'_\ast) &= b'_\ast,
\end{align*}
\]

respectively. Then for each given \( T \), we define \( \pi_\ast \) and \( \pi'_\ast \) as the projection onto the sum of eigenspaces corresponding to the eigenvalues more than \( t_\ast \) and less than \( t'_\ast \), respectively. Also, let \( \pi := 1 - \pi_\ast - \pi'_\ast \). Also, define

\[
\begin{align*}
  G_\ast &:= \text{tr} \rho_1 T \pi_\ast - \text{tr} \rho_2 f^*(T) \pi_\ast, \\
  G_0 &:= \text{tr} \rho_1 T \pi - \text{tr} \rho_2 f^*(T) \pi, \\
  G'_\ast &:= \text{tr} \rho_1 T \pi'_\ast - \text{tr} \rho_2 f^*(T) \pi'_\ast.
\end{align*}
\]

Then \( G(T) = G_\ast + G_0 + G'_\ast \),

\[
\begin{align*}
  G_\ast &\leq \text{tr} \rho_1 T \pi_\ast - \text{tr} \rho_2 (b_\ast (T \pi_\ast - t_\ast \pi_\ast) + f^*(t_\ast) \pi_\ast) \\
  &= \text{tr} \left( \rho_1 - \rho_2 \right) T \pi_\ast + \text{tr} \rho_2 (b_\ast (t_\ast - f^*(t_\ast)) \pi_\ast) \\
  &\leq \text{tr} \left( \rho_1 - \rho_2 \right) t_\ast \pi_\ast + \text{tr} \rho_2 (b_\ast (t_\ast - f^*(t_\ast)) \pi_\ast) \\
  &= \text{tr} \rho_1 (t_\ast \pi_\ast) - \text{tr} \rho_2 f^*(t_\ast) \pi_\ast,
\end{align*}
\]

and

\[
G'_\ast \leq \text{tr} \rho_1 (t'_\ast \pi'_\ast) - \text{tr} \rho_2 f^*(t'_\ast) \pi'_\ast.
\]

Therefore,

\[
G(T) \leq G(t_\ast \pi_\ast + T \pi + t'_\ast \pi'_\ast).
\]

To summarize:

**Lemma 7** Suppose \( f \) is a proper closed convex function with \( \text{dom} f^* \supset (0, \infty) \). Suppose also the assumption (I) is true. Then, the \( T \) in [13] can be restricted to the set of all the Hermitian operators with \( \text{spec} T \subset \left[ t'_\ast, t_\ast \right] \).

If \( H \) is finite dimensional and \( \rho_1 > 0, \rho_2 > 0 \), this lemma means that the supremum can be restricted to the interior of \( \text{dom} G \), and that the supremum of the differentiable concave function \( G(\cdot) \) is achieved by a point satisfying some point. Thus the solution \( T_0 \) to (18) achieve the supremum.

So far, we had supposed the assumption (I) is true. Now let us consider the case where (II) holds and (I) does not. A trivial approach is to exchange \( \rho_0 \) and \( \rho_1 \), and apply all the analysis replacing \( f \) by \( \hat{f} \). This means the change of variable from \( T \) to \( S := -f^*(T) \). But sometimes, the use of the variable \( T \) is more preferable by technical reasons. In such cases, still we can use [13], because of the following reason. Since the assumption (II) says that \( \hat{f}(t) = -f^{*-1}(t) \) is operator convex, it is continuously differentiable. Thus the stable point with respect to \( S \) is also a stable point with respect to \( T \).
7 Detailed analysis for non-full-rank case

In this case we treat the case where the condition (I) holds, \( \ker \rho_1 \) is non-trivial, and \( \dom f^* \) is not bounded from below.

By the following lemma, whose proof is given in Appendix A, then \( f^* \) in fact is operator monotone increasing.

**Lemma 8** Suppose \( f \) is proper, closed, convex, standard and \( \dom f \supset (0, \infty) \).
Suppose also \( f^* \) is operator convex, \( \dom f^* \) is not bounded from below, and \( f^*(-\infty) > -\infty \). Then \( f^* \) is operator monotone.

Below, \( \mathcal{H}_t \) is the restriction of \( \pi_{\mathcal{H}_t} \pi_{\mathcal{H}_t'} \) on \( \mathcal{H}_t \).

**Lemma 9** Suppose the domain of a function \( h \) is unbounded from below and \( h(-\infty) \) is finite. Then, if \( X \preceq 0 \),

\[
\lim_{s \to \infty} h(T + sX) = h(T_{\ker X}) \pi_{\ker X} + h(-\infty) \pi_{\supp X}.
\]  

(22)

**Proof.** Let \( r^\kappa_s \) and \(|\varphi^\kappa_s\rangle \) \((\|\varphi^\kappa_s\| = 1)\) be the \( \kappa \)-th eigenvalue and eigenvector of \( T + sX \), respectively. Since

\[
\lim_{s \to \infty} (s^{-1}T + X) = X,
\]

\[
\lim_{s \to \infty} r^\kappa_s / s := \xi^\kappa \quad \text{and} \quad \lim_{s \to \infty} \langle \varphi^\kappa_s \rangle := |\varphi^\kappa_\infty\rangle \quad \text{gives a complete set of eigenvalues and eigenvectors of} \ X.
\]

Without loss of generality, let \( \xi^\kappa \) be 0 if \( \kappa \leq \dim \ker X \), and negative otherwise.

Let \( t \) be an eigenvalue of \( T_{\ker X} \) and \(|e\rangle \) be a member of corresponding eigenspace, which is denoted by \( \mathcal{H}_t \). Observe, if \( \kappa \leq \dim \ker X \), we have

\[
\begin{align*}
    r^\kappa_s \pi_{\mathcal{H}_t} |\varphi^\kappa_s\rangle &= \pi_{\mathcal{H}_t} (T + sX) |\varphi^\kappa_s\rangle \\
    &= \pi_{\mathcal{H}_t} T \pi_{\ker X} |\varphi^\kappa_s\rangle + \pi_{\mathcal{H}_t} \pi_{\supp X} |\varphi^\kappa_s\rangle \\
    &= \pi_{\mathcal{H}_t} \pi_{\ker X} T \pi_{\ker X} |\varphi^\kappa_s\rangle + \pi_{\mathcal{H}_t} \pi_{\supp X} T \pi_{\ker X} |\varphi^\kappa_s\rangle + \pi_{\mathcal{H}_t} \pi_{\supp X} |\varphi^\kappa_s\rangle \\
    &= t \pi_{\mathcal{H}_t} |\varphi^\kappa_s\rangle + 0 + \pi_{\mathcal{H}_t} \pi_{\supp X} |\varphi^\kappa_s\rangle,
\end{align*}
\]

implying

\[
\lim_{s \to \infty} (r^\kappa_s - t) \pi_{\mathcal{H}_t} |\varphi^\kappa_s\rangle = \lim_{s \to \infty} \pi_{\mathcal{H}_t} \pi_{\supp X} |\varphi^\kappa_s\rangle = 0.
\]

Therefore, if \( \pi_{\mathcal{H}_t} |\varphi^\kappa_\infty\rangle \neq 0 \) holds, \( r^\kappa_s \to t \) holds, implying that \(|\varphi^\kappa_\infty\rangle\) is a member of \( \mathcal{H}_t \). Since \(|\varphi^\kappa_\infty\rangle\) \((\kappa \leq \dim \ker X)\) is a member of \( \ker X \), which is the direct sum of all \( \mathcal{H}_t \)'s, it overlaps with, and thus a member of, at least one of \( \mathcal{H}_t \)'s. Therefore, recalling that \( \{|\varphi^\kappa_\infty\rangle; \kappa \leq \dim \ker X\} \) is a CONS of \( \ker X \),

\[
\sum_{\kappa; \lim_{s \to \infty} r^\kappa_s = t} |\varphi^\kappa_\infty\rangle \langle \varphi^\kappa_\infty| = \pi_{\mathcal{H}_t}.
\]
Therefore,

$$\lim_{s \to \infty} h(T + sX) = \lim_{s \to \infty} \sum_{\kappa \leq \dim \ker X} h(r_s^\kappa) \langle \varphi_s^\kappa \rangle + \lim_{s \to \infty} \sum_{\kappa > \dim \ker X} h(r_s^\kappa) \langle \varphi_s^\kappa \rangle$$

$$= \sum_t h(t) \pi_{\mathcal{H}_t} + \sum_{\kappa > \dim \ker X} \lim_{s \to \infty} h(s\xi_s^\kappa) \langle \varphi_s^\kappa \rangle$$

$$= h(T_{\ker X}) \pi_{\ker X} + h(-\infty) \pi_{\supp X}.$$  

\[\square\]

**Theorem 10** Suppose \( f \) is proper, closed, convex, standard and \( \text{dom} \, f \supset (0, \infty) \). Suppose also \( f^* \) is operator convex and \( \text{dom} \, f^* \) is not bounded from below. Then if \( \rho_1 \) is not full-rank,

$$D^*_f (\rho_1 || \rho_2) = D^*_f (\pi_{\rho_1} \rho_2 \pi_{\rho_1}) - f^* (-\infty) \text{tr} \rho_2 (1 - \pi_{\rho_1}).$$  

(23)

Therefore, the problem reduces to optimization of \( G(T) \) restricting \( T \) to on the support of \( \rho_1 \). Especially, if \( \rho_1 \) is rank-1 state, \( \rho = |\varphi \rangle \langle \varphi| \), and \( \langle \varphi_1 | \rho_2 | \varphi_1 \rangle \neq 0 \),

$$D^*_f (|\varphi \rangle \langle \varphi| || \rho_2) = \sup_{t \in \text{dom} \, f^*} \{ t - \langle \varphi_1 | \rho_2 | \varphi_1 \rangle f^* (t) \} - f^* (-\infty) (1 - \langle \varphi_1 | \rho_2 | \varphi_1 \rangle)$$

$$= \hat{f} (\langle \varphi_1 | \rho_2 | \varphi_1 \rangle) - f^* (-\infty) (1 - \langle \varphi_1 | \rho_2 | \varphi_1 \rangle)$$  

(24)

**Proof.** By Theorem 3, we only have to prove the assertion for the case \(-\infty < f^* (-\infty) \) (otherwise, \( D^*_f (\rho_1 || \rho_2) = \infty \).) By Lemma 8 \( f^* \) is operator monotone. Thus, if \( X < 0 \) is supported on \( \ker \rho_1 \), \( G(T + sX) \), where \( G \) is as of (10), is non decreasing in \( s \). Therefore,

$$\sup_s G(T + sX) = \lim_{s \to \infty} G(T + sX).$$

Hence, by Lemma 9 we have

$$\sup_{X < 0: \text{supp} X = \ker \rho_1} G(T + X) = \text{tr} \rho_1 \pi_{\rho_1} T \pi_{\rho_1} - \text{tr} (\rho_2)_{\text{supp} \rho_1} f^* (T_{\text{supp} \rho_1})$$

$$- f^* (-\infty) \text{tr} \rho_2 (1 - \pi_{\rho_1}),$$

indicating (23).  

\[\square\]

8 **Examples**

8.1 **Renyi-type, \( f_\alpha (\lambda) \)**

Here we consider \( f_\alpha (\lambda) \) which is defined by (1) on the non-negative half line. On the negative half-line, we define it so that it becomes standard, i.e., \( f^*_\alpha \) is
monotone increasing on dom $f^*_\alpha$. Since $\alpha = 0$ and 1 give trivial functions, we omit these cases. The relation

$$\hat{f}_\alpha(\lambda) = f_{1-\alpha}(\lambda),$$

turns out to be quite useful.

If $0 < \alpha < 1$,

$$f^*_\alpha(t) = \begin{cases} 
\infty, & (t > 0), \\
(1-\alpha) \frac{\alpha}{1-\alpha} (t)^{\frac{\alpha}{1-\alpha}}, & (t \leq 0). 
\end{cases}$$

Thus, if $0 < \alpha \leq \frac{1}{2}$, the condition (I) is satisfied, and if $\frac{1}{2} \leq \alpha < 1$, the condition (II) is satisfied. If $\alpha < 0$,

$$f^*_\alpha(t) = \begin{cases} 
\infty, & (t \geq 0), \\
(\alpha-1) (-\alpha)^{\frac{\alpha}{1-\alpha}} (-t)^{\frac{\alpha}{1-\alpha}}, & (t < 0), 
\end{cases}$$

and the condition (I) is satisfied. Thus, if $\alpha > 1$, the condition (II) is satisfied, and

$$f^*_\alpha(t) = \begin{cases} 
(\alpha-1) \frac{\alpha}{1-\alpha} t^{\frac{\alpha}{1-\alpha}}, & (t > 0), \\
\infty, & (t \leq 0). 
\end{cases}$$

Thus, if $\alpha \geq 2$, in addition the condition (I) is satisfied.

For all the values of $\alpha$ (≠ 0, 1), $f^*_{\alpha'}(t)$ moves all over the positive half line $(0, \infty)$. Thus, by (21) and (20), the supremum is achieved by $T_0$ with (19).

In the case of $\alpha = -1$ and 2, we can solve the problem ”explicitly”. Observe $f^*_2$ is operator convex on dom $f^*_2$.

$$f^*_2(t) = \begin{cases} 
\frac{1}{4} t^2, & (t \geq 0), \\
\infty, & (t < 0). 
\end{cases}$$

By (18), we have a Lyapunov equation

$$Df^*_2(T_0)(\rho_2) = \frac{1}{4} (T_0 \rho_2 + \rho_2 T_0) = \rho_1.$$

If $\rho_2 > 0$, this can be solved about $T_0$,

$$T_0 = 4 \int_{-\infty}^{0} e^{\rho_2} \rho_1 e^{\rho_2} ds \geq 0,$$

and in the basis where $\rho_2$ is diagonal,

$$T_{0,i,j} = \frac{4}{\rho_{2,i,i} + \rho_{2,j,j}} \rho_{1,i,j}.$$
Thus, this solution $T_0$ has spectrum in $\text{dom } f^*_2$. By (19),

$$D_{f^*_1} (\rho_2 || \rho_1) = D_{f^*_2} (\rho_1 || \rho_2) = \text{tr } \rho_2 \left( \frac{1}{2} T_0 \right)^2 = \frac{1}{4} \text{tr } \rho_2 T_0^2$$

$$= \text{tr } \rho_1 T_0$$

$$= 4 \text{tr } \rho_1 \int_{-\infty}^{0} e^{\rho_2} \rho_1 e^{\rho_2} \text{d}s$$

$$= 4 \sum_{i,j} \frac{1}{\rho_{2,i,i} + \rho_{2,j,j}} |\rho_{1,i,j}|^2.$$

If $\alpha = \frac{1}{2}$,

$$f^*_2 (t) = \begin{cases} \infty, & (t > 0), \\ -\frac{1}{t} t^{-1}, & (t \leq 0), \end{cases}$$

and

$$D_{f^*_2} (T_0) (\rho_2) = \frac{1}{4} T_0^{-1} \rho_2 T_0^{-1} = \rho_1.$$

Thus

$$T_0^{-1} = -2 \rho_2^{-1/2} \sqrt{\rho_2^{-1/2} \rho_1 \rho_2^{-1/2}};$$

and

$$D_{f^*_1} (\rho_1 || \rho_2) = -\text{tr } \rho_2 \left( \frac{1}{4} T_0^{-2} \right)^{1/2} = -\frac{1}{2} \text{tr } \rho_2 \left( -T_0^{-1} \right)$$

$$= -\text{tr } \sqrt{\rho_2^{-1/2} \rho_1 \rho_2^{-1/2}},$$

which is $-1$ times the fidelity between $\rho_1$ and $\rho_2$, as expected.

By Theorem 3, $D_{f^*_1} (\rho_1 || \rho_2) < \infty$ for any $\rho_1 > 0$ if $\alpha < 0$, for any $\rho_2 > 0$ if $\alpha > 1$, and for any $\rho_1 \geq 0$, $\rho_2 \geq 0$ if $0 < \alpha < 1$.

If $0 < \alpha \leq \frac{1}{2}$, $f^*_\alpha$ is operator monotone, and thus by (24),

$$D_{f^*_\alpha} (|\varphi_1 \rangle \langle \varphi_1 | || \rho_2) = \hat{f}_\alpha (|\langle \varphi_1 | \rho_2 | \varphi_1 \rangle |) - f^*_\alpha (-\infty) (1 - \langle \varphi_1 | \rho_2 | \varphi_1 \rangle)$$

$$= -\langle \varphi_1 | \rho_2 | \varphi_1 \rangle^{1-\alpha},$$

and

$$D_{f^*_\alpha} (|\varphi_1 \rangle \langle \varphi_1 | || |\varphi_2 \rangle \langle \varphi_2 |) = - |\langle \varphi_1 | \varphi_2 \rangle|^{2(1-\alpha)}.$$

This means, if $\frac{1}{2} \leq \alpha < 1$, using $\hat{f}_\alpha = f_{1-\alpha}$,

$$D_{f^*_\alpha} (|\varphi_1 \rangle \langle \varphi_1 | || |\varphi_2 \rangle \langle \varphi_2 |) = - |\langle \varphi_1 | \varphi_2 \rangle|^{2\alpha}.$$
8.2 On Chernoff and Hoeffding bound

The results about Renyi type quantity, especially when two states are pure, gives another way to compute Chernoff bound and Hoeffding bound, whose classical counterpart is

\[
C(p_1 \parallel p_2) := \sup_{0 < \alpha < 1} \lim_{n \to \infty} -\frac{1}{n} \ln (\eta_{1,n} + \eta_{2,n})
\]

\[
= \sup_{0 < \alpha < 1} \{ -\ln (-D_{f_\alpha}(p_1 \parallel p_2)) \}
\]

and

\[
H_r(p_1 \parallel p_2) := \sup_{0 < \alpha < 1} \left\{ -\lim_{n \to \infty} \frac{1}{n} \ln \eta_{1,n}; \lim_{n \to \infty} \frac{1}{n} \eta_{2,n} \leq -r \right\}
\]

\[
= \sup_{0 < \alpha < 1} \left\{ -\frac{\alpha r}{1 - \alpha} - \frac{1}{(1 - \alpha)} \ln (-D_{f_\alpha}(p_1 \parallel p_2)) \right\},
\]

respectively. Here, \(\eta_{1,n}\) is the probability that the test mistakenly judges the true distribution (= \(p_1 \otimes n\), in fact,) as being \(p_2 \otimes n\). \(\eta_{2,n}\) the other direction of the error. Its quantum counterpart is defined by replacing distributions by states, and explicit form of these quantities are

\[
C(\rho_1 \parallel \rho_2) = \sup_{0 < \alpha < 1} \left\{ -\ln (-D_{f_\alpha}^{P}(\rho_1 \parallel \rho_2)) \right\},
\]

\[
H_r(p_1 \parallel p_2) = \sup_{0 < \alpha < 1} \left\{ -\alpha r - \frac{1}{(1 - \alpha)} \ln (-D_{f_\alpha}^{P}(\rho_1 \parallel \rho_2)) \right\},
\]

where

\[
D_{f_\alpha}^{P}(\rho_1 \parallel \rho_2) := -\text{tr} \rho_1^{\alpha} \rho_2^{1-\alpha}, \quad 0 < \alpha < 1.
\]

(See [1][3][4])

We confirm the achievability part of these celebrated results for the case where \(\rho_\theta = |\varphi_\theta\rangle \langle \varphi_\theta|\) (\(\theta = 1, 2\)). In fact, we have

\[
C(p_1 \parallel p_2) \geq \lim_{n \to \infty} \frac{1}{n} \sup_{0 < \alpha < 1} \left\{ -\ln (-D_{f_\alpha}^{\min}(\rho_1^{\otimes n} \parallel \rho_2^{\otimes n})) \right\},
\]

\[
H_r(p_1 \parallel p_2) \geq \lim_{n \to \infty} \sup_{0 < \alpha < 1} \left\{ -\frac{\alpha r}{1 - \alpha} - \frac{1}{n(1 - \alpha)} \ln (-D_{f_\alpha}^{\min}(\rho_1^{\otimes n} \parallel \rho_2^{\otimes n})) \right\}.
\]

These are not as trivial as they seem to be, since since the optimal measurement differs for each \(\alpha\) in general. However, since our states are pure, the supremum is infinitely approximated by the sequence of measurements independent of \(\alpha\), and thus the equality holds. Also, the RHS can be computed explicitly,

\[
-\frac{1}{n} \ln (-D_{f_\alpha}^{\min}(\rho_1^{\otimes n} \parallel \rho_2^{\otimes n})) = \begin{cases} 
-2(1 - \alpha) \ln |\langle \varphi_1 | \varphi_2 \rangle|, & 0 < \alpha \leq \frac{1}{2}, \\
-2\alpha \ln |\langle \varphi_1 | \varphi_2 \rangle|, & \frac{1}{2} \leq \alpha < 1.
\end{cases}
\]

The supremum is achieved at \(\alpha = 0, 1\), and the known result is confirmed. Interestingly, even though they give the same supremum, they differ for almost every value of \(\alpha\).
8.3 Kullback-Leibler divergence

Let
\[ f_{\text{KL}} (\lambda) := \begin{cases} \lambda \ln \lambda, & (\lambda \geq 0), \\ \infty, & (\lambda < 0), \end{cases} \]
\[ f_{\text{KL}}^2 (\lambda) := \hat{f}_{\text{KL}} (\lambda) = \begin{cases} -\ln \lambda, & (\lambda \geq 0), \\ \infty, & (\lambda < 0). \end{cases} \]

Then
\[ D_{f_{\text{KL}}} (p_1 || p_2) = D_{f_{\text{KL}}^2} (p_2 || p_1) = \int p_1 (x) \ln \frac{p_1 (x)}{p_2 (x)} d\mu (x) , \text{ if supp } p_1 \subset \text{supp } p_2 , \]
\[ \infty, \text{ otherwise}. \]
is the Kullback-Leibler divergence.

As easily seen, \[ f_{\text{KL}}^* (t) = e^{t-1} \]
is not operator convex but
\[ \hat{f}_{\text{KL}}^* (t) = \begin{cases} -1 - \ln (-t), & (t < 0), \\ \infty, & (t \geq 0), \end{cases} \]
is operator convex. Both of
\[ D_{f_{\text{KL}}} (T_0) (\rho_2) = \int_0^1 e^{sT_0} \rho_2 e^{(1-s)T_0} ds = \rho_1 , \]
\[ D_{\hat{f}_{\text{KL}}^2} (S_0) (\rho_1) = \int_0^\infty (sI - S_0)^{-1} \rho_1 (sI - S_0)^{-1} ds = \rho_2 \]
are difficult to solve explicitly. But using these solutions,
\[ D_{f_{\text{KL}}}^\text{min} (\rho_1 || \rho_2) = \text{tr } \rho_2 T_0' \ln T_0' = \text{tr } \rho_1 \ln S_0' , \]
where
\[ T_0' := e^{T_0 - 1}, S_0' = -S_0. \]
Also, applying Theorem A, \[ D_{f_{\text{KL}}}^\text{min} (\rho_1 || \rho_2) \] is finite only if \( \rho_2 > 0. \)

8.4 Total variation distance

Total variation distance \( \| \rho_1 - \rho_2 \|_1 \) equals, as is well known, \( D_{f_{TV}}^\text{min} \). We confirm this result using our method. Since \( f_{TV}^* \) is operator convex on \( \text{dom } f^* \), Here,
it is important to choose $f$ which is standard ($f^*_\text{TV}_2$ is not operator convex, for example). Then

$$D_{\text{min}}^{f_{\text{TV}}} (\rho_1||\rho_2) = \sup_{T: -1 \leq \text{spec} T \leq 1} \text{tr} \rho_1 T - \text{tr} \rho_2 T$$

$$= \sup_{T: -1 \leq \text{spec} T \leq 1} \text{tr} (\rho_1 - \rho_2) T = \|\rho_1 - \rho_2\|_1.$$

Note, $f^*_\text{TV}$ does not satisfy (21) nor (20). Indeed, the supremum is achieved by a $T$ whose eigenvalues are at the both end of the domain of $f^*_\text{TV}$.

9 Quantum Fisher information

Consider a parameterized family $\{p_\eta\}_{\eta \in \mathbb{R}}$ of probability density functions over a finite set, and suppose $\eta \to p_\eta$ is smooth, and $\text{ supp } p_\eta \subset \text{ supp } p_0$. Suppose also $f$ is a convex function with all good features. Then using Taylor expansion of $f$,

$$\lim_{\eta' \to \eta} \frac{1}{(\eta - \eta')^2} (D_f (p_\eta||p_{\eta'}) - D_f (p_\eta||p_\eta)) = \frac{1}{2} f''(1) J^2_\eta,$$

(25)

where

$$J_\eta := \sum_x \frac{(\partial p_\eta (x) / \partial \eta)^2}{p_\eta (x)}$$

is the Fisher information of the family $\{p_\eta\}_{\eta \in \mathbb{R}}$. This quantity characterizes asymptotic behavior of estimate of $\eta$, and the above mentioned relation is also important in deriving these results, and in relating estimation and hypothesis test.

Thus exploring its quantum analogue is also of interest. Let $\{\rho_\eta\}_{\eta \in \mathbb{R}}$ be a family of density operators, and suppose $\eta \to \rho_\eta$ is smooth. Then our task here is to evaluate $D_{\text{min}}^{f} (\rho_\eta||\rho_{\eta'})$ up to $(\eta - \eta')^2$ for small $|\eta - \eta'|$. Naively exchanging the order of limit and optimization, we have

$$\lim_{\eta' \to \eta} \frac{1}{(\eta - \eta')^2} (D_{\text{min}}^{f} (\rho_\eta||\rho_{\eta'}) - D_{\text{min}}^{f} (\rho_\eta||\rho_\eta)) = \frac{1}{2} f''(1) \max_M J^M_\eta (\eta - \eta')^2,$$

where $J^M_\eta$ is the Fisher information of the family $\{p^M_\eta\}_{\eta \in \mathbb{R}}$, and $p^M_\eta$ is the distribution of the data of the measurement $M$ applied to $\rho_\eta$. If this identity is true, we can use the well-known identity

$$\max_M J^M_\eta = J^S_\eta.$$

Here, $J^S_\eta$, called the SLD Fisher information, is defined by

$$J^S_\eta := \text{tr} \rho_\eta (L^S_\eta)^2,$$
where $L^S_\eta$, the symmetric logarithmic derivative of $\{\rho_\eta\}_{\eta \in \mathbb{R}}$, is defined as a Hermitian operator satisfying the equation

$$\frac{\partial \rho_\eta}{\partial \eta} = \frac{1}{2} \left( L^S_\eta \rho_\eta + \rho_\eta L^S_\eta \right).$$

SLD Fisher information, like its classical analogue, nicely characterize the asymptotic behavior of the optimal estimate of the unknown parameter $\eta$. Thus, by $D_{f}^{\min} (\rho_\eta || \rho_\eta) = f (1)$, we obtain

$$\lim_{\eta' \to \eta} \frac{1}{(\eta - \eta')^2} \left( D_{f}^{\min} (\rho_\eta || \rho_{\eta'}) - D_{f}^{\min} (\rho_\eta || \rho_\eta) \right) = \frac{1}{2} f'' (1) J^S_\eta. \quad (26)$$

If all members of $\{\rho_\eta\}_{\eta \in \mathbb{R}}$ have a common support, $p^M_\eta$ does not vanish and it is easy to make the above argument rigorous. If $\rho_\eta$ has non-trivial kernel, however, the remainder term of the Taylor expansion is not necessarily bounded due to $1/p^M_\eta$ factors. For example, if $\rho_\eta$ ($\eta \in \mathbb{R}$) are pure states, $D_{f}^{\min} (\alpha > 1, \alpha < 0)$, $D_{f,1}^{\min}$, and $D_{f,12}^{\min}$ diverge, and (26) is never true. On the other hand, in the case that $f^* (-\infty)$ is finite, $f = f_\alpha$ ($0 < \alpha \leq 1/2$) for example, it is easy to see the LHS of (26) equals constant multiple of $J^S_\eta$. Hence, the above naive argument is not completely false, though it is not rigorous. Below, we give deeper analysis on this issue. As it will turn out, (26) requires some non-trivial correction, when the rank of $\rho_\eta$ is neither full nor 1.

For the identity (25) for probability distribution families to be true, $f$ and $p_\eta$ should satisfy some conditions. Usually, it is supposed that $f$ is three times continuously differentiable in the neighborhood of 1, and

$$f'' (1) > 0.$$ 

(The last assumption is necessary for the $f$-divergence not to be constant at $\eta \approx \eta'$.) This means $t_0 := f' (1)$ lies in the interior of $\text{dom} f^*$, and

$$f^* (t) = t f'^{-1} (t) - f (f'^{-1} (t))$$

is three times continuously differentiable at the neighborhood of $t_0$. Using above identity,

$$f^* (t_0) = t_0 - f (1),$$

$$f'' (t_0) = 1,$$

$$f''' (t_0) = \frac{1}{f'' (1)} > 0.$$ 

Also, the map $\eta \to p_\eta$ is often three times continuously differentiable.

In quantum case, we impose the above conditions of $f$, and suppose that the map $\eta \to \rho_\eta$ is three times continuously differentiable, and the rank of $\rho_\eta$ is constant for all $\eta$. In addition, to use Theorem 10, we suppose the assumption
(I) holds and dom $f^*$ is unbounded from below. On these conditions,

$$
\lim_{\eta' \to \eta} \frac{1}{(\eta' - \eta)^2} \left( D^\text{min}_f (\rho_\eta||\rho_{\eta'}) - D^\text{min}_f (\rho_\eta||\rho_\eta) \right) \\
= \frac{1}{4} \left( f^* (t_0) - f^* (-\infty) \right) \left( \frac{f''(1)}{2} \text{tr} \rho_\eta (L^S_{\eta,1})^2 \right) + \frac{f''(1)}{2} \text{tr} \rho_\eta (L^S_{\eta,1})^2, \tag{27}
$$

where

$$
L^S_{\eta,1} := \pi_\eta L^S \pi_\eta, \quad L^S_{\eta,2} := (1 - \pi_\eta) L^S \pi_\eta.
$$

If $\rho_\eta$ is full rank (in this case, $L^S_{\eta,2} = 0$) or pure (in this case $L^S_{\eta,1} = 0$) the LHS of (27) equals a constant multiple of $J^S_\eta$, though the constant differs depending on which case it is. But if the rank of $\rho_\eta$ is full or 1, the result is a weighted sum of two components of SLD Fisher information, one is concerned with the change on the support of $\rho_\eta$ and the other is concerned with the change on the kernel of $\rho_\eta$.

To derive (27), by Theorem 10 we only have to maximize

$$
G (T; \eta, \eta') := \text{tr} \rho_\eta T - \text{tr} \pi_\eta \rho_{\eta'} \pi_\eta f^* (T) - f^* (-\infty) (1 - \text{tr} \pi_\eta \rho_{\eta'})
$$

where $T$ moves over all the Hermitian operators living in the support $\mathcal{H}_\eta$ of $\rho_\eta$, $\pi_\eta$ is the projector onto $\mathcal{H}_\eta$, and spec $T$ is a subset of dom $f^*$. Observe, if $\eta = 0$, the optimal $T$ is $t_0 \pi_\eta$. Thus, we put

$$
T_0 = t_0 \pi_\eta + Y.
$$

Then, using Lemma 7 we have, for small $|\eta' - \eta|$,

$$
\|Y_\eta\| \leq r_0^{-1} \|\rho_\eta - \pi_\eta \rho_{\eta'} \pi_\eta\| \sup_{t \in [t_0-c, t_0+c]} \frac{1}{f'''(t)} \tag{28}
$$

as detailed in Appendix B. Here, $r_0$ is the smallest eigenvalue of $\rho_\eta$ and $c$ is a positive number such that

$$
\sup_{t \in [t_0-c, t_0+c]} \frac{1}{f'''(t)} < \infty.
$$

Such $c > 0$ exists since $f^*$ is three times continuously differentiable at the neighborhood of $t_0$ and $f'''(t_0) > 0$.

Since $f^*$ is three times continuously differentiable in the neighborhood of $t_0$, by assumption,

$$
f^* (t_0 + \gamma) = f^* (t_0) + f''(t_0) \gamma + \frac{1}{2} f'''(t_0) \gamma^2 + \frac{1}{6} f^{(4)}(t_0) \gamma^3
$$

$$
= f^* (t_0) + \gamma + \frac{1}{2 f''(1)} \gamma^2 + \frac{1}{6 f'''(t')} \gamma^3,
$$

23
for a certain $t'$ between $t_0 + \gamma$ and $t_0$. Due to the fact that $t_0 \pi_\eta$ commute with $Y$ and $t_0 = f(1) + f^*(t_0)$,

$$G \left( t_0 \pi_\eta + Y; \eta, \eta' \right)$$

$$= \text{tr} \, \rho_\eta \left( t_0 \pi_\eta + Y \right) - \text{tr} \, \pi_\eta \rho_\eta' \pi_\eta f^* \left( t_0 \pi_\eta + Y \right) - f^* \left( -\infty \right) \left( 1 - \text{tr} \, \pi_\eta \rho_\eta' \right)$$

$$= \text{tr} \, \rho_\eta \left( t_0 \pi_\eta + Y \right) - \text{tr} \, \pi_\eta \rho_\eta' \pi_\eta \left( f^* \left( t_0 \right) + Y + \frac{1}{2 f'' \left( t_0 \right)} Y^2 \right) - f^* \left( -\infty \right) \left( 1 - \text{tr} \, \pi_\eta \rho_\eta' \right) + C$$

$$= f \left( 1 \right) + \left( f^* \left( t_0 \right) - f^* \left( -\infty \right) \right) \left( 1 - \text{tr} \, \pi_\eta \rho_\eta' \right) + \frac{1}{2 f'' \left( t_0 \right)} \left\{ \text{tr} \, \pi_\eta \rho_\eta' \pi_\eta Y_0^2 - \text{tr} \, \pi_\eta \rho_\eta' \pi_\eta \left( Y_0 - Y \right)^2 \right\} + C$$

$$\leq f \left( 1 \right) + \left( f^* \left( t_0 \right) - f^* \left( -\infty \right) \right) \left( 1 - \text{tr} \, \pi_\eta \rho_\eta' \right) + \frac{1}{2 f'' \left( t_0 \right)} \left\{ \text{tr} \, \pi_\eta \rho_\eta' \pi_\eta Y_0^2 \right\} + C,$$

where $C$ is bounded from above as

$$|C| \leq \frac{1}{6} \sup_{t' \in \left[ t_0 - \|Y\|, t_0 + \|Y\| \right]} f^{(3)} \left( t' \right) \|Y\|^3 = O \left( \eta' - \eta \right)^3,$$

and $Y_0$ is the solution to Lyapunov equation

$$- \left( \pi_\eta \rho_\eta' \pi_\eta - \rho_\eta \right) = \frac{1}{2 f'' \left( t_0 \right)} \left( \pi_\eta \rho_\eta' \pi_\eta Y_0 + Y_0 \pi_\eta \rho_\eta' \pi_\eta \right) = 0. \quad (29)$$

Since $G \left( t_0 \pi_\eta + Y; \eta, \eta' \right)$ coincide with the last end except for $O \left( \eta' - \eta \right)^3$ terms (here observe $\|Y_0\| = O \left( \eta' - \eta \right)$), we have

$$\sup_Y G \left( t_0 \pi_\eta + Y; \eta, \eta' \right) - f \left( 1 \right)$$

$$= \left( f^* \left( t_0 \right) - f^* \left( -\infty \right) \right) \left( 1 - \text{tr} \, \pi_\eta \rho_\eta' \right) + \frac{1}{2 f'' \left( t_0 \right)} \left\{ \text{tr} \, \pi_\eta \rho_\eta' \pi_\eta Y_0^2 \right\} + O \left( \eta' - \eta \right)^3.$$

Since the rank of $\rho_\eta$ is preserved while $\eta$ moves, we can write

$$\rho_\eta = A_\eta A_\eta^\dagger,$$

where the family $\{A_\eta\}_{\eta \in \mathbb{R}}$ satisfies

$$\frac{\partial}{\partial \eta} A_\eta = \frac{1}{2} L_\eta S A_\eta.$$

Then

$$A_\eta' = A_\eta + \frac{1}{2} L_\eta S A_\eta (\eta' - \eta) + C_1,$$

where $C_1$ is $O \left( \eta' - \eta \right)^2$. Then since

$$\pi_\eta \rho_\eta' \pi_\eta - \rho_\eta = \frac{1}{2} \left( \eta' - \eta \right) \left( L_\eta S_1^\dagger \rho_\eta + \rho_\eta L_\eta S_1 \right) + C_2,$$

24
we have, by (29),
\[ Y_0 = -f''(1)(\eta' - \eta)L_{\eta}^{S,1} + C_3, \]
where \(C_2\) and \(C_3\) are \(O((\eta' - \eta)^2).\) Also,
\[
(1 - \pi_\eta)\rho_\eta'(1 - \pi_\eta)
\]
\[
= (1 - \pi_\eta) \left( \frac{1}{2} L_{\eta}^S A_\eta (\eta' - \eta) + C_1 \right) \left( \frac{1}{2} L_{\eta}^S A_\eta (\eta' - \eta) + C_1 \right)^\dagger (1 - \pi_\eta)
\]
\[
= \left( \frac{1}{2} L_{\eta}^{S,2} A_\eta (\eta' - \eta) + (1 - \pi_\eta)C_1 \right) \left( \frac{1}{2} L_{\eta}^{S,2} A_\eta (\eta' - \eta) + (1 - \pi_\eta)C_1 \right)^\dagger
\]
\[
= \frac{1}{4} (\eta' - \eta)^2 L_{\eta}^{S,2} \rho_{\eta} L_{\eta}^{S,2} + C_4
\]
where \(C_4\) is \(O((\eta' - \eta)^3).\) After all, since \(D_f^{\min}(\rho_\eta||\rho_\eta) = f(1),\) we have (24).

10 Summary and questions
Using tools from convex analysis and matrix analysis, the maximization of the measured \(f\)-divergence is rewritten to a simpler form (15) on the assumption that (14) holds, and derived and proved some closed formulas and properties of \(D_f^{\min}\). Some questions are in order. First, what is the necessary and sufficient condition of \(f\) such that (14) holds? Second, the condition (I) is written in terms of \(f^*\), but it would be nicer to have some alternative condition written in terms of \(f\) itself, since in the study of other versions of quantum \(f\)-divergence, they often assume \(f\) to be operator convex (and not \(f^*\)). Third, it is easy to obtain a lower bound to \(D_f^{\min}\) by (15), but is there any good upper bound to the quantity, which is useful for the study of asymptotic theory?

References
[1] K.M.R. Audenaert, J. Calsamiglia, Ll. Masanes, R. Munoz-Tapia, A. Acin, E. Bagan, F. Verstraete, ” Discriminating states: the quantum Chernoff bound,” Physics Review Letters 98 160501, (2007)
[2] R. Bahatia, Matrix Analysis, Springer, (1997)
[3] F. Hiai, M. Mosonyi, D. Petz and C. Beny, ”Quantum \(f\)-divergences and error corrections,” Reviews in Mathematical Physics, 23, 691–747 (2011)
[4] H. Nagaoka, ”The converse part of the theorem for quantum Hoeffding bound,” quant-ph/0611289.
[5] R. Rockafellar, Convex analysis (Princeton, 1970).
[6] H. Strasser, ”Mathematical Theory of Statistics”, de Gruyter (1985).
A The proof of Lemma 8

If dom $f^*$ is the whole real line, any operator convex function $f^*$ is a quadratic function. Since in addition $f^*$ is monotone increasing (as a real valued function), it is an affine function with positive slope, contradicting $f^*(-\infty) > -\infty$. Thus dom $f^*$ = $(-\infty, a)$ or $(-\infty, a]$

Suppose dom $f^*$ = $(-\infty, a]$. Then, $h(t) := f^*(-t + a)$ is monotone non-increasing, dom $h$ = $[0, \infty)$ and $h(\infty) > -\infty$. Since $h$ is monotone non-increasing and proper, $h(\infty) < \infty$. Therefore, $h(\infty)$ is finite, and thus $\lim_{t \to -\infty} \frac{h(t)}{t} = 0$.

Therefore, by Proposition 8.4 of [3], $h$ can be written as

$$h(t) = h(0) + \alpha t - \int_{(0, \infty)} \frac{t}{t + \eta} d\nu(\eta)$$

using a non-negative measure $\nu$ with $\int_{(0, \infty)} \frac{1}{1 + \eta} d\nu(\eta) < \infty$. Since

$$\alpha = \frac{h(t) - h(0)}{t} + \int_{(0, \infty)} \frac{1}{t + \eta} d\nu(\eta)$$

and

$$\lim_{t \to \infty} \int_{(0, \infty)} \frac{1}{t + \eta} d\nu(\eta) = 0$$

by Lebesgue’s dominated convergence theorem, we have $\alpha = 0$. Since for each $\eta$ the function $t \to -t/(t + \eta)$ is operator monotone decreasing, $h(t)$ is operator monotone decreasing, implying the assertion.

If dom $f^*$ = $(-\infty, a)$, due to the above argument, $f^*$ is operator monotone increasing on $(-\infty, a - \varepsilon]$ for any $\varepsilon > 0$. Suppose the spectrum of $A_1$ and $A_2$ is a subset of $(-\infty, a]$, and $A_1 \geq A_2$. Then since $A_1 - \varepsilon 1 \geq A_2 - \varepsilon 1$,

$$f^*(A_1 - \varepsilon 1) \geq f^*(A_2 - \varepsilon 1).$$

Letting $\varepsilon \to 0$, we have $f^*(A_1) \geq f^*(A_2)$, meaning that $f^*$ is operator monotone increasing on $(-\infty, a)$.

B Proof of (28)

Let $b_*$ and $b'_*$ be as of (14) and (15), respectively, with $\rho_1 = \rho_\eta$ and $\rho_2 = \rho'_\eta$. 

[7] E. Torgersen, "Comparison of Statistical Experiments", Cambridge University Press (1991).

[8] D. Luenberger, Optimization by vector space method (Wiley, 1969).
\[1 - \frac{1}{b_*} = \left\|1 - \rho_\eta^{-1/2} (\pi_\eta \rho_\eta' \pi_\eta) \rho_\eta^{-1/2}\right\|
= \left\|\rho_\eta^{-1/2} (\rho_\eta - \pi_\eta \rho_\eta' \pi_\eta) \rho_\eta^{-1/2}\right\|
\leq r_0^{-1} \left\|\rho_\eta - \pi_\eta \rho_\eta' \pi_\eta\right\|\]

By rearranging terms, we have

\[
b_* \leq \frac{1}{1 - r_0^{-1} \left\|\rho_\eta - \pi_\eta \rho_\eta' \pi_\eta\right\|}
\leq 1 + r_0^{-1} \left\|\rho_\eta - \pi_\eta \rho_\eta' \pi_\eta\right\|.

Almost in parallel manner,

\[
b'_* \geq 1 - r_0^{-1} \left\|\rho_\eta - \pi_\eta \rho_\eta' \pi_\eta\right\|.

Then by Lemma 7,

\[
\|Y_\eta\| \leq \max \{|t_* - t_0|, |t'_0 - t'_*|\}
\leq \sup_{f^{*'}(t) \in [b'_*, b_*]} \frac{1}{f^{*'}(t)} \max \{|b_* - 1|, |b'_* - 1|\}
\leq r_0^{-1} \left\|\rho_\eta - \pi_\eta \rho_\eta' \pi_\eta\right\| \sup_{f^{*'}(t) \in [b'_*, b_*]} \frac{1}{f^{*'}(t)}.
\]

Since \(f^{*'}\) is continuous, when \(r_0^{-1} \left\|\rho_\eta - \pi_\eta \rho_\eta' \pi_\eta\right\|\) is small enough, the set of all \(t\) with \(f^{*'}(t) \in [b'_*, b_*]\) is the subset of \([t_0 - c, t_0 + c]\). Therefore, we have \(\text{(28)}\).