Information geometry of sandwiched Rényi α-divergence

Kaito Takahashi and Akio Fujiwara*
Department of Mathematics, Osaka University
Toyonaka, Osaka 560-0043, Japan

Abstract

Information geometrical structure $(g^{(D_\alpha)}, \nabla^{(D_\alpha)}, \nabla^{(D_\alpha)*})$ induced from the sandwiched Rényi α-divergence $D_\alpha(\rho\|\sigma) := \frac{1}{\alpha(\alpha-1)} \log \mathrm{Tr} \left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \, \sigma^{\frac{1-\alpha}{2\alpha}} \right)^\alpha$ on a finite quantum state space $\mathcal{S}$ is studied. It is shown that the Riemannian metric $g^{(D_\alpha)}$ is monotone if and only if $\alpha \in (-\infty, -1] \cup [\frac{1}{2}, \infty)$, and that the quantum statistical manifold $(\mathcal{S}, g^{(D_\alpha)}, \nabla^{(D_\alpha)}, \nabla^{(D_\alpha)*})$ is dually flat if and only if $\alpha = 1$.

1 Introduction

In his seminal paper, Rényi [19] introduced a new class of information divergence now usually referred to as the *Rényi relative entropy* of order α, where α is a positive number. Recently, Wilde *et al.* [20] and Müller-Lennert *et al.* [17] independently proposed an extension of the Rényi relative entropy to the quantum domain. Let $\mathcal{L}(\mathcal{H})$ and $\mathcal{L}_{\mathrm{sa}}(\mathcal{H})$ denote the sets of linear operators and selfadjoint operators on a finite dimensional complex Hilbert space $\mathcal{H}$, and let $\mathcal{L}_+(\mathcal{H})$ and $\mathcal{L}_{++}(\mathcal{H})$ denote the subsets of $\mathcal{L}_{\mathrm{sa}}(\mathcal{H})$ comprising positive operators and strictly positive operators. Given $\rho, \sigma \in \mathcal{L}_+(\mathcal{H})$ with $\rho \neq 0$, let

$$\psi(\alpha) := \log \mathrm{Tr} \left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \, \sigma^{\frac{1-\alpha}{2\alpha}} \right)^\alpha \tag{1}$$

for $\alpha \in (0, \infty)$, with the convention that $\psi(\alpha) := \infty$ if $\alpha > 1$ and $\ker \sigma \not\subset \ker \rho$. The first divided difference of ψ at $\alpha = 1$, i.e.,

$$\frac{\psi(\alpha) - \psi(1)}{\alpha - 1}, \qquad (\alpha \neq 1)$$

is called the *sandwiched Rényi relative entropy* [20] or the *quantum Rényi divergence* [17], and is denoted by $\tilde{D}_\alpha(\rho\|\sigma)$ in the present paper. It is explicitly written as

$$\tilde{D}_\alpha(\rho\|\sigma) := \frac{1}{\alpha - 1} \log \mathrm{Tr} \left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \, \sigma^{\frac{1-\alpha}{2\alpha}} \right)^\alpha - \frac{1}{\alpha - 1} \log \mathrm{Tr} \, \rho. \tag{2}$$

The sandwiched Rényi relative entropy is extended to $\alpha = 1$ by continuity, to obtain the *Umegaki relative entropy*:

$$\tilde{D}_1(\rho\|\sigma) := \lim_{\alpha \to 1} \tilde{D}_\alpha(\rho\|\sigma) = \mathrm{Tr} \left\{ \rho(\log \rho - \log \sigma) \right\}$$

*fujiwara@math.sci.osaka-u.ac.jp

with the convention that $\tilde{D}_1(\rho\|\sigma) = \infty$ if $\ker \sigma \not\subset \ker \rho$. The limiting cases $\alpha \downarrow 0$ and $\alpha \to \infty$ have also been studied in [8, 2] and [17], respectively.

The sandwiched Rényi relative entropy has several desirable properties: amongst others, it is monotone under completely positive trace preserving (CPTP) maps if $\alpha \geq \frac{1}{2}$ [17, 20, 4, 10]. This property was successfully used in studying the strong converse properties of the channel capacity [20, 16] and quantum hypothesis testing problems [15]. It is highly expected that the sandwiched Rényi relative entropy will also play an important role in other fields of quantum information science. In this paper, we will study the sandwiched Rényi relative entropy from the point of view of quantum information geometry [1], a rapidly growing field in quantum information science.

Let us confine our attention to the case when both ρ and σ are faithful density operators that belong to the *quantum state space*:

$$\mathcal{S} := \mathcal{S}(\mathcal{H}) := \{\rho \in \mathcal{L}_{++}(\mathcal{H}) \mid \operatorname{Tr} \rho = 1\}.$$

In this case there is no difficulty in extending the quantities (1) and (2) to the region $\alpha < 0$. In order to motivate our study, let us assume for now that ρ and σ commute. Then the quantity (1) is reduced to $\psi(\alpha) = \log \operatorname{Tr} \rho^\alpha \sigma^{1-\alpha}$: this is known as the potential function for the $\nabla^{(e)}$-geodesic connecting ρ and σ in classical information geometry [1], and is meaningful for all $\alpha \in \mathbb{R}$. On the other hand,

$$\tilde{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha - 1} \log \operatorname{Tr} \rho^\alpha \sigma^{1-\alpha}$$

for $\alpha < 0$ does not seem to be a reasonable measure of information [19], since it takes negative values. Meanwhile, there are also other types of divergence functions that have been found to be useful in classical information theory and statistics. For example, Csiszár [7] introduced a class of information divergence now usually referred to as the *Csiszár f-divergence*, a version of which is written as

$$D^f(\rho\|\sigma) := \operatorname{Tr} \{\sigma f(\rho \sigma^{-1})\},$$

where f is a real-valued, strictly convex, smooth function on the set $\mathbb{R}_{++}$ of positive real numbers satisfying $f(1) = 0$ and $f''(1) = 1$. It is easily seen from Jensen's inequality that $D^f(\rho\|\sigma) \geq 0$, and $D^f(\rho\|\sigma) = 0$ if and only if $\rho = \sigma$. Now let us consider a family of functions

$$f^{[\alpha]}(t) := \frac{1}{\alpha(1-\alpha)} (1 - t^\alpha)$$

having a one-dimensional parameter α with $\alpha \notin \{0, 1\}$. This family is known to play an important role in classical information geometry. For example, the corresponding $f^{[\alpha]}$-divergences are the well-known "alpha-divergences"

$$D^{f^{[\alpha]}}(\rho\|\sigma) = \frac{1}{\alpha(1-\alpha)} \left(1 - \operatorname{Tr} \rho^\alpha \sigma^{1-\alpha}\right), \tag{3}$$

although the parametrization "alpha" is different from the standard one [1]. Now from the Taylor expansion of $\log(1+x)$, the divergence (3) is related to the potential function $\psi(\alpha) = \log \operatorname{Tr} \rho^\alpha \sigma^{1-\alpha}$ as

$$\psi(\alpha) = \alpha(\alpha - 1)D^{f^{[\alpha]}}(\rho\|\sigma) + O\left(D^{f^{[\alpha]}}(\rho\|\sigma)^2\right). \tag{4}$$

2

In other words, the $f^{[\alpha]}$-divergence $D^{f^{[\alpha]}}(\rho\|\sigma)$, provided it is small enough, is not very different from $\frac{1}{\alpha}\tilde{D}_\alpha(\rho\|\sigma)$. Note that the quantity $\frac{1}{\alpha}\tilde{D}_\alpha(\rho\|\sigma)$ is nonnegative even for $\alpha < 0$.

Motivated by the above consideration, we aim at investigating the "rescaled" sandwiched Rényi relative entropy:

$$D_\alpha(\rho\|\sigma) := \frac{1}{\alpha(\alpha-1)} \log \mathrm{Tr}\,\left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho\, \sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha \tag{5}$$

on the quantum state space $\mathcal{S}$ for $\alpha \in \mathbb{R}\backslash\{0,1\}$, which shall be referred to as the *sandwiched Rényi α-divergence* in the present paper. It is continuously extended to $\alpha = 1$ as

$$D_1(\rho\|\sigma) := \lim_{\alpha\to 1} D_\alpha(\rho\|\sigma) = \mathrm{Tr}\,\{\rho(\log\rho - \log\sigma)\}\,.$$

However, we note that, unless ρ and σ commute, (5) cannot be extended to $\alpha = 0$ because $\lim_{\alpha\to 0} D_\alpha(\rho\|\sigma)$ does not always exist (cf., Appendix A). To put it differently, the sandwiched Rényi 0-divergence $D_0(\rho\|\sigma)$ is excluded on the quantum state space $\mathcal{S}$. This fact makes a striking contrast to classical information geometry.

A basic property of the sandwiched Rényi α-divergence (5) is the following:

$$D_\alpha(\rho\|\sigma) \geq 0, \quad \text{and} \quad D_\alpha(\rho\|\sigma) = 0 \quad \text{if and only if} \quad \rho = \sigma \tag{6}$$

for $\rho, \sigma \in \mathcal{S}$ and $\alpha \in \mathbb{R}\backslash\{0\}$ (cf., Appendix B). This fact enables us to introduce an information geometrical structure on the quantum state space $\mathcal{S}$ through Eguchi's method [9]. Firstly, the Riemannian metric $g^{(D_\alpha)}$ is defined by

$$g_\rho^{(D_\alpha)}(X,Y) := D_\alpha((XY)_\rho\|\sigma)\Big|_{\sigma=\rho} := XYD_\alpha(\rho\|\sigma)\Big|_{\sigma=\rho}. \tag{7}$$

In the last expression, the vector fields X and Y are regarded as acting only on ρ. Secondly, a pair of affine connections $\nabla^{(D_\alpha)}$ and $\nabla^{(D_\alpha)*}$ are defined by

$$g_\rho^{(D_\alpha)}(\nabla_X^{(D_\alpha)}Y, Z) := -D_\alpha((XY)_\rho\|(Z)_\sigma)\Big|_{\sigma=\rho}, \tag{8}$$

$$g_\rho^{(D_\alpha)}(\nabla_X^{(D_\alpha)*}Y, Z) := -D_\alpha((Z)_\rho\|(XY)_\sigma)\Big|_{\sigma=\rho}. \tag{9}$$

The right-hand sides of (8) and (9) are understood in an analogous way to (7). Since

$$D_\alpha((X)_\rho\|\sigma)\Big|_{\sigma=\rho} = 0$$

for any vector field X, which is a consequence of (6), the metric $g_\rho^{(D_\alpha)}$ is also written as

$$g_\rho^{(D_\alpha)}(X,Y) = -D_\alpha((X)_\rho\|(Y)_\sigma)\Big|_{\sigma=\rho}.$$

It is now straightforward to verify the duality:

$$Xg^{(D_\alpha)}(Y,Z) = g^{(D_\alpha)}(\nabla_X^{(D_\alpha)}Y, Z) + g^{(D_\alpha)}(Y, \nabla_X^{(D_\alpha)*}Z). \tag{10}$$

This property plays an essential role in information geometry [1].

3

A Riemannian metric g on a quantum state space is called *monotone* [18] if it satisfies

$$g_\rho(X, X) \geq g_{\gamma(\rho)}(\gamma_* X, \gamma_* X) \tag{11}$$

for all states $\rho \in \mathcal{S}$, tangent vectors $X \in T_\rho \mathcal{S}$, and CPTP maps $\gamma : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H}')$, with γ_* denoting the differential of γ. The monotonicity (11) implies that the distinguishability of two nearby states, measured by the metric g, cannot be enhanced by any physical process γ. This is a fundamental requirement for information processing, and hence, characterizing monotone metrics is important in quantum information theory.

The main result of the present paper is the following

Theorem 1. *The induced Riemannian metric $g^{(D_\alpha)}$ is monotone under CPTP maps if and only if $\alpha \in (-\infty, -1] \cup [\frac{1}{2}, \infty)$.*

As a by-product, we arrive at the following corollary, the latter part of which was first observed by numerical evaluation [17], (cf., [5]).

Corollary 2. *The sandwiched Rényi α-divergence $D_\alpha(\rho||\sigma)$ is not monotone under CPTP maps if $\alpha \in (-1, 0) \cup (0, \frac{1}{2})$. Consequently, the original sandwiched Rényi relative entropy $\tilde{D}_\alpha(\rho||\sigma)$ is not monotone if $\alpha \in (0, \frac{1}{2})$.*

We also study the dualistic structure $(g^{(D_\alpha)}, \nabla^{(D_\alpha)}, \nabla^{(D_\alpha)*})$ on the quantum state space $\mathcal{S}$, and obtain the following

Theorem 3. *The quantum statistical manifold $(\mathcal{S}, g^{(D_\alpha)}, \nabla^{(D_\alpha)}, \nabla^{(D_\alpha)*})$ is dually flat if and only if $\alpha = 1$.*

The paper is organized as follows. In Section 2, we compute the metric $g^{(D_\alpha)}$, and prove Theorem 1. In Section 3, we investigate the dualistic structure $(g^{(D_\alpha)}, \nabla^{(D_\alpha)}, \nabla^{(D_\alpha)*})$ on a quantum state space $\mathcal{S}$, and prove Theorem 3. Section 4 is devoted to concluding remarks. Some additional topics are discussed in Appendices A-D.

2 Proof of Theorem 1

In quantum information geometry, it is customary to use a pair of operator representations of tangent vectors called the m-representation and the e-representation [1]. The *m-representation* $X^{(m)}$ of a tangent vector $X \in T_\rho \mathcal{S}$ at $\rho \in \mathcal{S}$ is simply defined by

$$X^{(m)} := X\rho.$$

In order to introduce an e-representation, on the other hand, we need to specify a continuous monotone function $f : \mathbb{R}_{++} \to \mathbb{R}_{++}$ satisfying $f(1) = 1$ and $f(t) = tf(\frac{1}{t})$. Once such a function f is given, we define the corresponding *e-representation* $X_f^{(e)}$ of $X \in T_\rho \mathcal{S}$ by

$$X_f^{(e)} := f(\Delta_\rho)^{-1}\left\{(X\rho)\rho^{-1}\right\},$$

where Δ_ρ is the *modular operator* associated with $\rho \in \mathcal{S}$ defined by

$$\Delta_\rho : \mathcal{L}(\mathcal{H}) \longrightarrow \mathcal{L}(\mathcal{H}) : A \longmapsto \rho A \rho^{-1}.$$

4

A Riemannian metric g is then given by the pairing

$$g_\rho(X, Y) = \text{Tr}\left\{ X_f^{(e)} Y^{(m)} \right\}$$

between e- and m-representations. According to Petz's theorem [18], the metric g represented in this form is monotone if and only if the function f is operator monotone. Thus, to prove Theorem 1, we first derive the defining function $f = f^{(D_\alpha)}$ from each D_α (Lemma 4), and then verify that the function $f^{(D_\alpha)}$ is operator monotone if and only if $\alpha \in (-\infty, -1] \cup [\frac{1}{2}, \infty)$ (Lemma 5).

2.1 Computation of metric

We first note that, for a power function $f(x) = x^\lambda$ with $x > 0$ and $\lambda \in \mathbb{R}$, the directional derivative

$$Df(A)[B] := \lim_{t \to 0} \frac{f(A + tB) - f(A)}{t}, \qquad (A \in \mathcal{L}_{++}(\mathcal{H}),\ B \in \mathcal{L}_{\text{sa}}(\mathcal{H}))$$

is given by

$$Df(A)[B] = \lambda \int_0^1 dt \int_0^\infty ds\, A^{\lambda t}(sI + A)^{-1} B(sI + A)^{-1} A^{\lambda(1-t)}, \qquad (12)$$

where I is the identity (cf., (33) in Appendix C). The following formula

$$D(\text{Tr}\, A^\lambda)[B] = \text{Tr}\left\{ (\lambda A^{\lambda-1}) B \right\} \qquad (13)$$

for differentiation under the trace operation also follows from (12).

Lemma 4. *For each $\alpha \in \mathbb{R}\backslash\{0, 1\}$, the metric $g^{(D_\alpha)}$ is represented in the form*

$$g_\rho^{(D_\alpha)}(X, Y) = \text{Tr}\left\{ X_{f^{(D_\alpha)}}^{(e)} Y^{(m)} \right\}, \qquad (14)$$

where

$$f^{(D_\alpha)}(t) := (\alpha - 1)\frac{t^{\frac{1}{\alpha}} - 1}{1 - t^{\frac{1-\alpha}{\alpha}}}$$

with the convention that $f^{(D_\alpha)}(1) := \lim_{t \to 1} f^{(D_\alpha)}(t) = 1$.

Proof. Recall that the metric $g_\rho^{(D_\alpha)}$ was defined by

$$g_\rho^{(D_\alpha)}(X, Y) = XY D_\alpha(\rho\|\sigma)\Big|_{\sigma=\rho},$$

where X and Y act only on ρ. Since $D_\alpha(\rho\|\sigma)$ is written as

$$D_\alpha(\rho\|\sigma) = \frac{1}{\alpha(\alpha - 1)} \log \text{Tr}\, A^\alpha$$

where

$$A := \sigma^{\frac{1-\alpha}{2\alpha}} \rho\, \sigma^{\frac{1-\alpha}{2\alpha}},$$

5

we have

$$YD_\alpha(\rho\|\sigma) = \frac{1}{\alpha(\alpha-1)}\frac{Y(\operatorname{Tr}A^\alpha)}{\operatorname{Tr}A^\alpha}$$

$$= \frac{1}{\alpha(\alpha-1)}\frac{D(\operatorname{Tr}A^\alpha)[YA]}{\operatorname{Tr}A^\alpha}$$

$$= \frac{1}{\alpha(\alpha-1)}\frac{\operatorname{Tr}\left\{(\alpha A^{\alpha-1})(YA)\right\}}{\operatorname{Tr}A^\alpha}$$

$$= \frac{1}{\alpha-1}\frac{\operatorname{Tr}\left\{A^{\alpha-1}B_Y\right\}}{\operatorname{Tr}A^\alpha}.$$

Here

$$B_Y := YA = \sigma^{\frac{1-\alpha}{2\alpha}}(Y\rho)\,\sigma^{\frac{1-\alpha}{2\alpha}},$$

and the formula (13) was used in the third equality. Consequently,

$$XYD_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1}\left[\frac{X\operatorname{Tr}\left\{A^{\alpha-1}B_Y\right\}}{\operatorname{Tr}A^\alpha} - \frac{\left(\operatorname{Tr}\left\{A^{\alpha-1}B_Y\right\}\right)\left(\operatorname{Tr}\left\{(\alpha A^{\alpha-1})B_X\right\}\right)}{(\operatorname{Tr}A^\alpha)^2}\right]$$

$$= \frac{1}{\alpha-1}\frac{\operatorname{Tr}\left\{(XA^{\alpha-1})B_Y\right\} + \operatorname{Tr}\left\{A^{\alpha-1}C_{XY}\right\}}{\operatorname{Tr}A^\alpha}$$

$$\qquad - \frac{\alpha}{\alpha-1}\frac{\left(\operatorname{Tr}\left\{A^{\alpha-1}B_Y\right\}\right)\left(\operatorname{Tr}\left\{A^{\alpha-1}B_X\right\}\right)}{(\operatorname{Tr}A^\alpha)^2}, \tag{15}$$

where

$$B_X := \sigma^{\frac{1-\alpha}{2\alpha}}(X\rho)\,\sigma^{\frac{1-\alpha}{2\alpha}}, \qquad C_{XY} := XB_Y = \sigma^{\frac{1-\alpha}{2\alpha}}(XY\rho)\,\sigma^{\frac{1-\alpha}{2\alpha}}.$$

Since

$$\operatorname{Tr}A^\alpha\Big|_{\sigma=\rho} = \operatorname{Tr}\left(\rho^{\frac{1}{\alpha}}\right)^\alpha = \operatorname{Tr}\rho = 1,$$

$$\operatorname{Tr}\left\{A^{\alpha-1}B_X\right\}\Big|_{\sigma=\rho} = \operatorname{Tr}\left\{\rho^{\frac{\alpha-1}{\alpha}}\rho^{\frac{1-\alpha}{2\alpha}}(X\rho)\,\rho^{\frac{1-\alpha}{2\alpha}}\right\} = \operatorname{Tr}(X\rho) = X\operatorname{Tr}\rho = 0,$$

$$\operatorname{Tr}\left\{A^{\alpha-1}C_{XY}\right\}\Big|_{\sigma=\rho} = \operatorname{Tr}\left\{\rho^{\frac{\alpha-1}{\alpha}}\rho^{\frac{1-\alpha}{2\alpha}}(XY\rho)\,\rho^{\frac{1-\alpha}{2\alpha}}\right\} = \operatorname{Tr}(XY\rho) = XY\operatorname{Tr}\rho = 0,$$

we have

$$g_\rho^{(D_\alpha)}(X,Y) = XYD_\alpha(\rho\|\sigma)\Big|_{\sigma=\rho} = \frac{1}{\alpha-1}\operatorname{Tr}\left\{(XA^{\alpha-1})B_Y\right\}\Big|_{\sigma=\rho}. \tag{16}$$

Now we invoke the formula (12), with $\lambda = \alpha-1$, to obtain

$$XA^{\alpha-1} = Df(A)[XA]$$

$$= (\alpha-1)\int_0^1 dt\int_0^\infty ds\, A^{(\alpha-1)t}(sI+A)^{-1}B_X(sI+A)^{-1}A^{(\alpha-1)(1-t)}. \tag{17}$$

6

Combining (16) with (17), we have

$$
g_\rho^{(D_\alpha)}(X,Y)
$$
$$
= \mathrm{Tr}\left\{ B_Y \int_0^1 dt \int_0^\infty ds\, A^{(\alpha-1)t}(sI+A)^{-1}B_X(sI+A)^{-1}A^{(\alpha-1)(1-t)} \right\}\bigg|_{\sigma=\rho}
$$
$$
= \mathrm{Tr}\left\{ \rho^{\frac{1-\alpha}{2\alpha}}(Y\rho)\rho^{\frac{1-\alpha}{2\alpha}} \int_0^1 dt \right.
$$
$$
\left. \times \int_0^\infty ds \left(\rho^{\frac{\alpha-1}{\alpha}}\right)^t (sI+\rho^{\frac{1}{\alpha}})^{-1}\rho^{\frac{1-\alpha}{2\alpha}}(X\rho)\rho^{\frac{1-\alpha}{2\alpha}}(sI+\rho^{\frac{1}{\alpha}})^{-1}\left(\rho^{\frac{\alpha-1}{\alpha}}\right)^{(1-t)} \right\}
$$
$$
= \mathrm{Tr}\left\{ \rho^{\frac{1-\alpha}{\alpha}}(Y\rho)\rho^{\frac{1-\alpha}{\alpha}} \int_0^1 dt \int_0^\infty ds \left(\rho^{\frac{\alpha-1}{\alpha}}\right)^t (sI+\rho^{\frac{1}{\alpha}})^{-1}(X\rho)(sI+\rho^{\frac{1}{\alpha}})^{-1}\left(\rho^{\frac{\alpha-1}{\alpha}}\right)^{(1-t)} \right\}.
$$
$$
(18)
$$

Comparing (18) with (14), we see that the e-representation $X_{f^{(D_\alpha)}}^{(e)}$ of X is given by

$$
X_{f^{(D_\alpha)}}^{(e)} = \rho^{\frac{1-\alpha}{\alpha}} \int_0^1 dt \int_0^\infty ds \left(\rho^{\frac{\alpha-1}{\alpha}}\right)^t (sI+\rho^{\frac{1}{\alpha}})^{-1}(X\rho)(sI+\rho^{\frac{1}{\alpha}})^{-1}\left(\rho^{\frac{\alpha-1}{\alpha}}\right)^{(1-t)}\rho^{\frac{1-\alpha}{\alpha}}.
$$

In order to determine the function $f^{(D_\alpha)}$, we introduce an orthonormal basis $\{e_i\}_{1\le i\le n}$ of $\mathcal{H}$ comprising eigenvectors of ρ each corresponding to the eigenvalue p_i. Then

$$
\left\langle e_i \left| X_{f^{(D_\alpha)}}^{(e)} e_j \right.\right\rangle
$$
$$
= p_i^{\frac{1-\alpha}{\alpha}} \int_0^1 dt \int_0^\infty ds \left(p_i^{\frac{\alpha-1}{\alpha}}\right)^t \left(s+p_i^{\frac{1}{\alpha}}\right)^{-1}\langle e_i|(X\rho)e_j\rangle \left(s+p_j^{\frac{1}{\alpha}}\right)^{-1}\left(p_j^{\frac{\alpha-1}{\alpha}}\right)^{(1-t)} p_j^{\frac{1-\alpha}{\alpha}}
$$
$$
= \langle e_i|(X\rho)e_j\rangle\,(p_ip_j)^{\frac{1-\alpha}{\alpha}} \int_0^1 dt \left(p_i^{\frac{\alpha-1}{\alpha}}\right)^t \left(p_j^{\frac{\alpha-1}{\alpha}}\right)^{(1-t)} \int_0^\infty ds \left(s+p_i^{\frac{1}{\alpha}}\right)^{-1}\left(s+p_j^{\frac{1}{\alpha}}\right)^{-1}.
$$

Since $X_{f^{(D_\alpha)}}^{(e)}$ is continuous in ρ, we can assume without loss of generality that eigenvalues p_i are all different. Then by using the formulae

$$
\int_0^1 x^t y^{1-t} dt = \frac{x-y}{\log x - \log y}
$$

and

$$
\int_0^\infty \frac{ds}{(s+x)(s+y)} = \frac{\log x - \log y}{x-y}
$$

for $x \neq y$, we get

$$
\begin{aligned}
\left\langle e_i \left| X^{(e)}_{f^{(D_\alpha)}} e_j \right. \right\rangle &= \langle e_i | (X\rho) e_j \rangle \, (p_i p_j)^{\frac{1-\alpha}{\alpha}} \times \frac{p_i^{\frac{\alpha-1}{\alpha}} - p_j^{\frac{\alpha-1}{\alpha}}}{\log p_i^{\frac{\alpha-1}{\alpha}} - \log p_j^{\frac{\alpha-1}{\alpha}}} \times \frac{\log p_i^{\frac{1}{\alpha}} - \log p_j^{\frac{1}{\alpha}}}{p_i^{\frac{1}{\alpha}} - p_j^{\frac{1}{\alpha}}} \\
&= \langle e_i | (X\rho) e_j \rangle \, \frac{1 - \left(\dfrac{p_i}{p_j} \right)^{\frac{1-\alpha}{\alpha}}}{(\alpha - 1)\, p_j \left\{ \left(\dfrac{p_i}{p_j} \right)^{\frac{1}{\alpha}} - 1 \right\}} \\
&= \frac{\langle e_i | (X\rho) e_j \rangle}{p_j \, f^{(D_\alpha)} \left(\dfrac{p_i}{p_j} \right)},
\end{aligned}
$$

for all $i \neq j$, where

$$
f^{(D_\alpha)}(t) = (\alpha - 1) \frac{t^{\frac{1}{\alpha}} - 1}{1 - t^{\frac{1-\alpha}{\alpha}}}.
$$

This completes the proof. $\qquad\square$

Let us examine some special cases. When $\alpha = \frac{1}{2}$, the function $f^{(D_{1/2})}(t) = \frac{1+t}{2}$ corresponds to the SLD metric, and when $\alpha = -1$, the function $f^{(D_{-1})}(t) = \frac{2t}{1+t}$ corresponds to the real RLD metric. Furthermore, the limiting function $f^{(D_1)}(t) := \lim_{\alpha \to 1} f^{(D_\alpha)}(t) = \frac{t-1}{\log t}$ corresponds to the Bogoliubov metric: this is consistent with the fact that $D_1(\rho \| \sigma) := \lim_{\alpha \to 1} D_\alpha(\rho \| \sigma)$ is the Umegaki relative entropy. It is well known that these three functions are operator monotone. Incidentally, another limiting function $f^{(D_{\pm\infty})}(t) := \lim_{\alpha \to \pm\infty} f^{(D_\alpha)}(t) = \frac{t \log t}{t-1} = t / f^{(D_1)}(t)$ is also operator monotone.

2.2 Operator monotonicity

In what follows, we change the parameter α into $\beta := \frac{1}{\alpha}$, and denote the corresponding function $f^{(D_\alpha)}(t)$ by $f_\beta(t)$, i.e.,

$$
f_\beta(t) := \frac{\beta - 1}{\beta} \frac{t^\beta - 1}{t^{\beta-1} - 1}
$$

where $\beta \notin \{0, 1\}$. We extend this function to $\beta = 0$ and 1 by continuity, to obtain

$$
\begin{aligned}
f_0(t) &:= \lim_{\beta \to 0} f_\beta(t) = \frac{t \log t}{t - 1}, \\
f_1(t) &:= \lim_{\beta \to 1} f_\beta(t) = \frac{t - 1}{\log t}.
\end{aligned}
$$

Lemma 5. *The function $f_\beta(t)$ is operator monotone if and only if $\beta \in [-1, 2]$.*

Proof. We first prove the 'if' part[1]. A key observation is the identity

$$
f_{\frac{1}{2} - \delta}(t) = \frac{t}{f_{\frac{1}{2} + \delta}(t)}
$$

---

[1]After almost completing the paper, the authors became aware that the 'if' part had been proved in [11]. Our proof is slightly simpler.

for all $\delta \in \mathbb{R}$, which is easily verified by direct computation. It follows that if $f_{\frac{1}{2}+\delta}(t)$ is operator monotone, so is $f_{\frac{1}{2}-\delta}(t)$. It then suffices to prove that $f_\beta(t)$ is operator monotone if $\beta \in [\frac{1}{2}, 2]$. Firstly, operator monotonicity of $f_1(t)$ is well known. Secondly, for $\beta \in [\frac{1}{2}, 1)$, let us set $\gamma := 1 - \frac{1}{\beta}$, which satisfies $-1 \leq \gamma < 0$. Since $t \mapsto t^\gamma$ is operator convex, its first divided difference

$$\frac{t^\gamma - 1}{t - 1}$$

at $t = 1$ is operator monotone [6, Theorem V.3.10]. Moreover, since $x \mapsto -\frac{1}{x}$ is operator monotone, so is the function

$$h_\gamma(t) := \gamma \frac{t-1}{t^\gamma - 1}.$$

Furthermore, since the function $t \mapsto t^\beta$ is operator monotone, so is

$$f_\beta(t) = h_\gamma(t^\beta).$$

Finally, for $\beta \in (1, 2]$, rewrite $f_\beta(t)$ into

$$f_\beta(t) = \frac{\beta - 1}{\beta} \left(t + \frac{t - 1}{t^{\beta - 1} - 1} \right).$$

Since $\frac{\beta - 1}{\beta} > 0$ and the map $x \mapsto \frac{1}{x}$ is order reversing, it suffices to show that the function

$$t \longmapsto \frac{t^{\beta - 1} - 1}{t - 1}$$

is operator monotone decreasing, and this is true because the above function is the first divided difference of an operator concave function $t \mapsto t^{\beta - 1}$.

We next prove the 'only if' part. Suppose $f_\beta(t)$ is operator monotone. Then it must satisfy the inequalities

$$\frac{2t}{1 + t} \leq f_\beta(t) \leq \frac{1 + t}{2}$$

for all $t > 0$ [18]. In particular, by letting $t = e$, we have

$$\frac{2e}{1 + e} \leq \frac{\beta - 1}{\beta} \frac{e^\beta - 1}{e^{\beta - 1} - 1} \leq \frac{1 + e}{2}.$$

We shall prove that these inequalities hold only when $\beta \in [-1, 2]$. Since

$$\left. \frac{\beta - 1}{\beta} \frac{e^\beta - 1}{e^{\beta - 1} - 1} \right|_{\beta = -1} = \frac{2e}{1 + e},$$

and

$$\left. \frac{\beta - 1}{\beta} \frac{e^\beta - 1}{e^{\beta - 1} - 1} \right|_{\beta = 2} = \frac{1 + e}{2},$$

it suffices to prove that the function

$$\beta \longmapsto \frac{\beta - 1}{\beta} \frac{e^\beta - 1}{e^{\beta - 1} - 1}$$

9

is strictly increasing for $\beta \leq -1$ or $\beta \geq 2$. Taking the logarithm, this is rephrased that the function

$$\beta \longmapsto h(\beta) - h(\beta - 1),$$

where

$$h(\beta) := \log \frac{e^\beta - 1}{\beta},$$

is strictly increasing, or equivalently,

$$h'(\beta) - h'(\beta - 1)$$

is positive for $\beta \leq -1$ or $\beta \geq 2$. In view of the mean value theorem, we prove this by showing that $h''(\beta) > 0$ for $\beta \leq -1$ or $\beta \geq 1$. Since

$$h''(\beta) = \frac{1}{\beta^2} - \frac{1}{\left(e^{\frac{\beta}{2}} - e^{-\frac{\beta}{2}} \right)^2},$$

it suffices to prove that

$$\tilde{h}(\beta) := \frac{e^{\frac{\beta}{2}} - e^{-\frac{\beta}{2}}}{\beta} > 1 \tag{19}$$

for $\beta \neq 0$. Since $\tilde{h}(-\beta) = \tilde{h}(\beta)$, $\lim_{\beta \to 0} \tilde{h}(\beta) = 1$, and the derivative

$$\tilde{h}'(\beta) = \frac{e^{\frac{\beta}{2}}(\beta - 2) + e^{-\frac{\beta}{2}}(\beta + 2)}{2\beta^2} = \frac{e^{-\frac{\beta}{2}}}{2\beta^2} \int_0^\beta x\,(\beta - x)\,e^x\,dx$$

is positive for $\beta > 0$, the inequality (19) is verified. $\qquad \square$

3 Structure of quantum statistical manifold $\mathcal{S}$

In this section, we study the dualistic structure $(g^{(D_\alpha)}, \nabla^{(D_\alpha)}, \nabla^{(D_\alpha)*})$ on the quantum statistical manifold $\mathcal{S}$. In a quite similar way to the derivation of (16), it is proved (cf., Appendix D) that the affine connections $\nabla^{(D_\alpha)}$ and $\nabla^{(D_\alpha)*}$ defined by (8) and (9) are explicitly given by

$$g_\rho^{(D_\alpha)}(\nabla_X^{(D_\alpha)} Y, Z) = \frac{1}{\alpha - 1} \left(\operatorname{Tr}\left. \left\{ (ZXA^{\alpha-1})B_Y + (ZA^{\alpha-1})C_{XY} \right\} \right|_{\sigma=\rho} \right.$$
$$\left. - \operatorname{Tr}\left\{ (Y\rho)Z \left[\left. \rho^{\frac{1-\alpha}{2\alpha}} (XA^{\alpha-1}) \right|_{\sigma=\rho} \rho^{\frac{1-\alpha}{2\alpha}} \right] \right\} \right), \tag{20}$$

and

$$g_\rho^{(D_\alpha)}(\nabla_X^{(D_\alpha)*} Y, Z) = \frac{1}{\alpha - 1} \left(\operatorname{Tr}\left. \left\{ (XB_Z)(YA^{\alpha-1}) - (YXA^{\alpha-1})B_Z - (YA^{\alpha-1})C_{XZ} \right\} \right|_{\sigma=\rho} \right.$$
$$+ \operatorname{Tr}\left\{ (Z\rho)X \left[\left. \rho^{\frac{1-\alpha}{2\alpha}} (YA^{\alpha-1}) \right|_{\sigma=\rho} \rho^{\frac{1-\alpha}{2\alpha}} \right] \right\}$$
$$\left. + \operatorname{Tr}\left\{ (Z\rho)Y \left[\left. \rho^{\frac{1-\alpha}{2\alpha}} (XA^{\alpha-1}) \right|_{\sigma=\rho} \rho^{\frac{1-\alpha}{2\alpha}} \right] \right\} \right). \tag{21}$$

Now, if the quantum state space $\mathcal{S}$ is dually flat with respect to the dualistic structure $(g^{(D_\alpha)}, \nabla^{(D_\alpha)}, \nabla^{(D_\alpha)*})$, then there is a pair of affine coordinate systems that allows us a variety of information geometrical techniques on $\mathcal{S}$ [1]. It is therefore interesting to ask which value of α makes $\mathcal{S}$ dually flat. The answer is given by Theorem 3: the quantum statistical manifold $(\mathcal{S}, g^{(D_\alpha)}, \nabla^{(D_\alpha)}, \nabla^{(D_\alpha)*})$ is dually flat if and only if $\alpha = 1$.

Proof of Theorem 3. When $\alpha = 1$, the sandwiched Rényi α-divergence is reduced to the Umegaki relative entropy $D_1(\rho\|\sigma) = \mathrm{Tr}\,\{\rho(\log\rho - \log\sigma)\}$, and the dually flatness of $\mathcal{S}$ with respect to $(g^{(D_1)}, \nabla^{(D_1)}, \nabla^{(D_1)*})$ is well known [1].

To prove the necessity, let us take a submanifold $\mathcal{M}$ of $\mathcal{S}$ comprising commutative density operators, which can be regarded as the space of classical probability distributions. Since the sandwiched Rényi α-divergence restricted to $\mathcal{M}$ is identical, up to the first order, to the classical "alpha-divergence" as (4), the restricted metric $g^{(D_\alpha)}\big|_{\mathcal{M}}$ is identical to the classical Fisher metric for all α, and the restricted connections $\nabla^{(D_\alpha)}\big|_{\mathcal{M}}$ and $\nabla^{(D_\alpha)*}\big|_{\mathcal{M}}$ are the $(2\alpha - 1)$- and the $(1 - 2\alpha)$-connections, respectively, in the standard terminology of classical information geometry. Consequently, $\mathcal{M}$ is dually flat if and only if $\alpha = 0$ or 1 [1]. Now suppose that $\mathcal{S}$ is dually flat with respect to the structure $(g^{(D_\alpha)}, \nabla^{(D_\alpha)}, \nabla^{(D_\alpha)*})$. Then the submanifold $\mathcal{M}$ is also dually flat with respect to the restricted structure $(g^{(D_\alpha)}\big|_{\mathcal{M}}, \nabla^{(D_\alpha)}\big|_{\mathcal{M}}, \nabla^{(D_\alpha)*}\big|_{\mathcal{M}})$. Since the sandwiched Rényi 0-divergence is excluded on $\mathcal{S}$, as mentioned in Section 1, the only remaining possibility is $\alpha = 1$. $\qquad\square$

A closely related question is this: Is there a triad (α, β, γ) of real numbers for which $(\mathcal{S}, g^{(D_\alpha)}, \nabla^{(D_\beta)}, \nabla^{(D_\gamma)})$ becomes dually flat? The answer is negative. In fact, the connections $\nabla^{(D_\beta)}\big|_{\mathcal{M}}$ and $\nabla^{(D_\gamma)}\big|_{\mathcal{M}}$ restricted to a commutative submanifold $\mathcal{M}$ are the $(2\beta - 1)$- and the $(2\gamma - 1)$-connections, respectively. If $(\mathcal{S}, g^{(D_\alpha)}, \nabla^{(D_\beta)}, \nabla^{(D_\gamma)})$ is dually flat, then the pair (β, γ) must be either $(1, 0)$ or $(0, 1)$, as discussed above. Since the sandwiched Rényi 0-divergence is excluded on $\mathcal{S}$, we conclude that there is no triad (α, β, γ) that makes $(\mathcal{S}, g^{(D_\alpha)}, \nabla^{(D_\beta)}, \nabla^{(D_\gamma)})$ dually flat.

4   Concluding remarks

In the present paper, we studied information geometrical structure of the quantum state space $\mathcal{S}$ induced from the sandwiched Rényi α-divergence $D_\alpha(\rho\|\sigma)$, a variant of the quantum relative entropy recently proposed by Wilde *et al.* [20] and Müller-Lennert *et al.* [17]. We found that the induced Riemannian metric $g^{(D_\alpha)}$ is monotone if and only if $\alpha \in (-\infty, -1] \cup [\frac{1}{2}, \infty)$, and that the induced dualistic structure $(g^{(D_\alpha)}, \nabla^{(D_\alpha)}, \nabla^{(D_\alpha)*})$ makes the quantum state space $\mathcal{S}$ dually flat if and only if $\alpha = 1$.

Our results could be compared with the results by Hasegawa and Petz [12] who investigated the Riemannian metric induced from the plain quantum α-divergence (3), with ρ and σ being understood as elements of the quantum state space $\mathcal{S}$. They found that the induced metric is closely related to the so-called Wigner-Yanase-Dyson (WYD) skew information [21], and is monotone if and only if $\alpha \in [-1, 2]$. While the relationship to the WYD skew information was interesting from a purely mathematical point of view, a disappointing aspect was also pointed out; the plain quantum α-divergence (3) does not induce the SLD metric that plays an essential role in quantum state estimation [13, 14]. This is in

remarkable contrast to our Theorem 1, in which the metrics $g^{(D_\alpha)}$ form a one-parameter family of monotone metrics that interpolates between the SLD metric (the smallest monotone metric) and the real RLD metric (the largest monotone metric). Thus, the sandwiched Rényi α-divergence might involve richer geometrical structure and wider applicability in quantum information science.

The result about the monotonicity of $g^{(D_\alpha)}$, which is consistent with the known monotonicity of $D_\alpha(\rho\|\sigma)$ for $\alpha \in [\frac{1}{2}, \infty)$ [10], strongly suggests that $D_\alpha(\rho\|\sigma)$ might be monotone also for $\alpha \in (-\infty, -1]$. This problem raises another interesting question about reconstructing $D_\alpha(\rho\|\sigma)$ from a purely differential geometrical viewpoint. It is well known that the canonical divergence on a dually flat statistical manifold (M, g, ∇, ∇^*) is reconstructed by integrating the metric g along either ∇ or ∇^*-geodesic [1, 3]. Unfortunately, this method is not applicable to our problem because the quantum statistical manifold $(S, g^{(D_\alpha)}, \nabla^{(D_\alpha)}, \nabla^{(D_\alpha)*})$ is not dually flat unless $\alpha = 1$. If such a differential geometrical method of reconstructing a divergence function is successfully extended to non-flat statistical manifolds, then we may have a new, direct method of proving the monotonicity and/or the convexity of a global quantity $D_\alpha(\rho\|\sigma)$ from a local information $(g^{(D_\alpha)}, \nabla^{(D_\alpha)}, \nabla^{(D_\alpha)*})$ on the quantum statistical manifold $\mathcal{S}$.

Acknowledgments

Appendices

A Non-extendibility of D_α to $\alpha = 0$

The following Proposition is a special case of [2, Lemma 1].

Proposition 6. *For all $\rho, \sigma \in \mathcal{S}$,*

$$\lim_{\alpha \to 0} \left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \, \sigma^{\frac{1-\alpha}{2\alpha}} \right)^\alpha = \sigma.$$

Proof. Since $\dim \mathcal{H} < \infty$ and $\rho > 0$, there are positive numbers λ and μ that satisfy $\lambda I \leq \rho \leq \mu I$. This entails that

$$\lambda \sigma^{\frac{1-\alpha}{\alpha}} \leq \sigma^{\frac{1-\alpha}{2\alpha}} \rho \, \sigma^{\frac{1-\alpha}{2\alpha}} \leq \mu \sigma^{\frac{1-\alpha}{\alpha}}.$$

For $\alpha \in (0, 1)$, the function $f(t) = t^\alpha$ with $t > 0$ is operator monotone, and

$$\lambda^\alpha \sigma^{1-\alpha} \leq \left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \, \sigma^{\frac{1-\alpha}{2\alpha}} \right)^\alpha \leq \mu^\alpha \sigma^{1-\alpha}. \tag{22}$$

For $\alpha \in (-1, 0)$, on the other hand, the function $f(t) = t^\alpha$ with $t > 0$ is operator monotone decreasing, and

$$\mu^\alpha \sigma^{1-\alpha} \leq \left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \, \sigma^{\frac{1-\alpha}{2\alpha}} \right)^\alpha \leq \lambda^\alpha \sigma^{1-\alpha}. \tag{23}$$

Taking the limit $\alpha \downarrow 0$ in (22), and $\alpha \uparrow 0$ in (23), we have the assertion. $\qquad\square$

In view of Proposition 6, as well as the evaluation

$$-\log \mu - H(\sigma) \leq \liminf_{\alpha \to 0} D_\alpha(\rho\|\sigma) \leq \limsup_{\alpha \to 0} D_\alpha(\rho\|\sigma) \leq -\log \lambda - H(\sigma),$$

where $H(\sigma)$ is the von Neumann entropy, which follows from (22) and (23), it is natural to expect that $D_\alpha(\rho\|\sigma)$ could be continuously extended to $\alpha = 0$. In reality, it is in general untrue, as the following example shows.

Let $\mathcal{H} = \mathbb{C}^2$ and let

$$\rho = \begin{bmatrix} 1/2 & 1/4 \\ 1/4 & 1/2 \end{bmatrix}, \qquad \sigma = \begin{bmatrix} 3/4 & 0 \\ 0 & 1/4 \end{bmatrix}.$$

The eigenvalues of $\sigma^{\frac{1-\alpha}{2\alpha}} \rho \, \sigma^{\frac{1-\alpha}{2\alpha}}$ are

$$\frac{1}{3 \cdot 4^{1/\alpha}} \left(3 + 3^{1/\alpha} \pm \sqrt{9 - 3 \cdot 3^{1/\alpha} + 9^{1/\alpha}} \right),$$

and $\mathrm{Tr}\left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \, \sigma^{\frac{1-\alpha}{2\alpha}} \right)^\alpha$ is given by

$$\frac{1}{3^\alpha \cdot 4} \left\{ \left(3 + 3^{1/\alpha} + \sqrt{9 - 3 \cdot 3^{1/\alpha} + 9^{1/\alpha}} \right)^\alpha + \left(3 + 3^{1/\alpha} - \sqrt{9 - 3 \cdot 3^{1/\alpha} + 9^{1/\alpha}} \right)^\alpha \right\}.$$

By direct computation, we obtain

$$\lim_{\alpha \downarrow 0} D_\alpha(\rho\|\sigma) = \frac{1}{2} \log\left(\frac{3}{2} \right)$$

and

$$\lim_{\alpha \uparrow 0} D_\alpha(\rho\|\sigma) = \frac{1}{2} \log 2,$$

proving that

$$\lim_{\alpha \downarrow 0} D_\alpha(\rho\|\sigma) \neq \lim_{\alpha \uparrow 0} D_\alpha(\rho\|\sigma).$$

B   A basic property of D_α

The following Proposition is an extension of the results by Wilde *et al.* [20, Corollaries 7 and 8] for $1 < \alpha \leq 2$, Müller-Lennert *et al.* [17, Theorem 3] for $\alpha \geq \frac{1}{2}$, and Beigi [4, Theorem 5] for $\alpha > 0$.

Proposition 7. $D_\alpha(\rho\|\sigma) \geq 0$ *for all* $\alpha \in \mathbb{R}\backslash\{0\}$ *and* $\rho, \sigma \in \mathcal{S}$. *Moreover, the equality holds if and only if* $\rho = \sigma$.

Proof. Since the case $\alpha > 0$ has been treated in [4], we shall concentrate on the case $\alpha < 0$; however, we note that our method presented here is also applicable to the case $\alpha > 0$.

Let $\sigma = \sum_i s_i E_i$ be the spectral decomposition, where $\{s_i\}_i$ are distinct eigenvalues of σ and $\{E_i\}_i$ are the corresponding projection operators. The *pinching* operation $\mathcal{E}_\sigma : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H})$ associated with the state σ is defined by

$$\mathcal{E}_\sigma(A) := \sum_i E_i A E_i.$$

The pinching $\mathcal{E}_\sigma$ sends a state ρ to a state $\mathcal{E}_\sigma(\rho)$ that commutes with σ. To prove the first part of the claim, it suffices to show that $D_\alpha(\rho\|\sigma) \geq D_\alpha(\mathcal{E}_\sigma(\rho)\|\sigma)$ for all $\alpha \in \mathbb{R}\backslash\{0\}$, since the positivity $D_\alpha(\mathcal{E}_\sigma(\rho)\|\sigma) \geq 0$ is well known in classical information theory. Due to [6, Problem II.5.5], for any Hermitian matrix A, the vector $\lambda(\mathcal{E}_\sigma(A))$ comprising eigenvalues of $\mathcal{E}_\sigma(A)$ is majorised by the vector $\lambda(A)$ comprising eigenvalues of A; in symbol, $\lambda(\mathcal{E}_\sigma(A)) \prec \lambda(A)$. It follows that $\operatorname{Tr} f(\mathcal{E}_\sigma(A)) \leq \operatorname{Tr} f(A)$ for any convex function f (cf., (26) below). Applying this result to $A = \sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}}$ and $f(t) = t^\alpha$ with $t > 0$, which is convex for $\alpha < 0$, we have

$$\operatorname{Tr}\left(\mathcal{E}_\sigma(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}})\right)^\alpha \leq \operatorname{Tr}\left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha.$$

Taking the logarithm of both sides, dividing them by $\alpha(\alpha-1)$, which is positive for $\alpha < 0$, and noting that $\mathcal{E}_\sigma(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}}) = \sigma^{\frac{1-\alpha}{2\alpha}} \mathcal{E}_\sigma(\rho) \sigma^{\frac{1-\alpha}{2\alpha}}$, we have $D_\alpha(\mathcal{E}_\sigma(\rho)\|\sigma) \leq D_\alpha(\rho\|\sigma)$.

Let us proceed to the second part. The 'if' part is obvious. We show the 'only if' part. Since $D_\alpha(\rho\|\sigma) \geq D_\alpha(\mathcal{E}_\sigma(\rho)\|\sigma) \geq 0$, the condition $D_\alpha(\rho\|\sigma) = 0$ leads to a series of equalities $D_\alpha(\rho\|\sigma) = D_\alpha(\mathcal{E}_\sigma(\rho)\|\sigma)$ and $D_\alpha(\mathcal{E}_\sigma(\rho)\|\sigma) = 0$. The latter equality implies that $\mathcal{E}_\sigma(\rho) = \sigma$. The former equality is equivalent to

$$\operatorname{Tr} f(A) = \operatorname{Tr} f(\mathcal{E}_\sigma(A)),$$

where $A = \sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}}$ and $f(t) = t^\alpha$. Since $f(t)$ is strictly convex in $t > 0$ for $\alpha < 0$, Lemma 8 below shows that $\lambda^\downarrow(A) = \lambda^\downarrow(\mathcal{E}_\sigma(A))$. Here $\lambda^\downarrow(A)$ denotes the vector comprising eigenvalues of A arranged in the decreasing order. It then follows from Lemma 9 below that $A = \mathcal{E}_\sigma(A)$, or equivalently, $\rho = \mathcal{E}_\sigma(\rho)$. Putting these results together, we have $\rho = \mathcal{E}_\sigma(\rho) = \sigma$. $\qquad\square$

Lemma 8. *Let A and B be strictly positive $n \times n$ Hermitian matrices, and let $f : \mathbb{R}_{++} \to \mathbb{R}$ be a strictly convex function. If $\lambda(A) \succ \lambda(B)$ and $\operatorname{Tr} f(A) = \operatorname{Tr} f(B)$, then $\lambda^\downarrow(A) = \lambda^\downarrow(B)$.*

Proof. Let us denote the eigenvalues of A and B explicitly as follows:

$$\lambda^\downarrow(A) = (\lambda_1, \lambda_2, \ldots, \lambda_n), \qquad \lambda^\downarrow(B) = (\mu_1, \mu_2, \ldots, \mu_n),$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$ and $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n > 0$. Furthermore, let $(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_r)$ and $(\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_s)$ be the lists of distinct eigenvalues of A and B, respectively, labeled in the decreasing order, so that $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_r$ and $\hat{\mu}_1 > \hat{\mu}_2 > \cdots > \hat{\mu}_s$. In order to handle the multiplicity of eigenvalues, we introduce the subsets

$$I_\alpha := \{i \,|\, \lambda_i = \hat{\lambda}_\alpha\}, \qquad J_\beta := \{j \,|\, \mu_j = \hat{\mu}_\beta\}$$

of indices, each corresponding to distinct eigenvalue $\hat{\lambda}_\alpha$ or $\hat{\mu}_\beta$. Then the set $\{1, 2, \ldots, n\}$ is decomposed into disjoint unions of $\{I_\alpha\}_{1 \leq \alpha \leq r}$ and $\{J_\beta\}_{1 \leq \beta \leq s}$ as follows:

$$\{1, 2, \ldots, n\} = \bigsqcup_{\alpha=1}^{r} I_\alpha = \bigsqcup_{\beta=1}^{s} J_\beta.$$

We shall show that $r = s$, and that $\hat{\lambda}_\alpha = \hat{\mu}_\alpha$ and $I_\alpha = J_\alpha$ for each $\alpha = 1, \ldots, r$.

Since $\lambda(A) \succ \lambda(B)$, there is a doubly stochastic $n \times n$ matrix $Q = [Q_{ji}]$ that satisfies

$$\mu_j = \sum_{i=1}^{n} Q_{ji} \lambda_i, \qquad (j = 1, \ldots, n). \tag{24}$$

Thus, for a convex function f,

$$f(\mu_j) = f\left(\sum_{i=1}^{n} Q_{ji} \lambda_i\right) \leq \sum_{i=1}^{n} Q_{ji} f(\lambda_i) \tag{25}$$

and

$$\sum_{j=1}^{n} f(\mu_j) \leq \sum_{i=1}^{n} f(\lambda_i). \tag{26}$$

If there is some j for which the inequality (25) becomes strict, then the inequality (26) also becomes strict. Thus the condition $\operatorname{Tr} f(B) = \operatorname{Tr} f(A)$, which amounts to demanding equality in (26), leads us to equality in (25) for all $j = 1, \ldots, n$. Since f is strictly convex, equality in (25) holds if and only if there is an $\alpha \in \{1, \ldots, r\}$ such that the support set $\{i \,|\, Q_{ji} > 0\}$ is a subset of I_α, i.e.,

$$\lambda_i = \hat{\lambda}_\alpha \quad \text{for all} \quad i \in \{k \,|\, Q_{jk} > 0\}. \tag{27}$$

Combining (27) with (24), we also have

$$\mu_j = \hat{\lambda}_\alpha. \tag{28}$$

Put differently, for each $j \in \{1, \ldots, n\}$, there is a unique α that satisfies (27) and (28), and this correspondence defines a map $\Gamma : j \mapsto \alpha$.

Given $\beta \in \{1, \ldots, s\}$, let us choose $j_1, j_2 \in J_\beta$ arbitrarily. Then we see from (28) that

$$\hat{\lambda}_{\Gamma(j_1)} = \mu_{j_1} = \hat{\mu}_\beta = \mu_{j_2} = \hat{\lambda}_{\Gamma(j_2)}.$$

Consequently, $\Gamma(j_1) = \Gamma(j_2)$ for any $j_1, j_2 \in J_\beta$ and $\beta \in \{1, \ldots, s\}$. This implies that Γ naturally induces an injective map $\tilde{\Gamma} : \beta \mapsto \alpha$ for which $\hat{\mu}_\beta = \hat{\lambda}_\alpha$. In particular, we must have $s \leq r$.

Now, the above construction shows that $Q_{ji} > 0$ only if $(j, i) \in J_\beta \times I_\alpha$ with $\alpha = \tilde{\Gamma}(\beta)$. As a consequence, for any pair (α, β) satisfying $\alpha = \tilde{\Gamma}(\beta)$,

$$|J_\beta| \cdot \hat{\mu}_\beta = \sum_{j \in J_\beta} \mu_j = \sum_{j \in J_\beta} \left(\sum_{i \in I_\alpha} Q_{ji} \lambda_i\right) = \sum_{i \in I_\alpha} \lambda_i = |I_\alpha| \cdot \hat{\lambda}_\alpha.$$

Since $\hat{\mu}_\beta = \hat{\lambda}_\alpha > 0$, we have $|J_\beta| = |I_\alpha|$. This relation further concludes that $\tilde{\Gamma}$ is surjective; since otherwise $s < r$ from the injectivity of Γ, and

$$n = \sum_{\beta=1}^{s} |J_\beta| = \sum_{\beta=1}^{s} |I_{\tilde{\Gamma}(\beta)}| < \sum_{\alpha=1}^{r} |I_\alpha| = n,$$

which is a contradiction.

In summary, $\tilde{\Gamma}$ is bijective (in fact, the identity map), $s = r$, and $\hat{\mu}_\beta = \hat{\lambda}_\beta$ and $J_\beta = I_\beta$ for each $\beta = 1, \ldots, s \,(= r)$. Consequently, we have $\lambda^\downarrow(A) = \lambda^\downarrow(B)$. $\qquad \square$

Lemma 9. *Let $\mathcal{E}_\sigma$ be the pinching operation associated with a state $\sigma \in \mathcal{S}(\mathbb{C}^n)$. For an $n \times n$ Hermitian matrix A, the following conditions are equivalent.*

(i) $\lambda^\downarrow(A) = \lambda^\downarrow(\mathcal{E}_\sigma(A))$.

(ii) $A = \mathcal{E}_\sigma(A)$.

Proof. (ii) $\Rightarrow$ (i) is trivial. We show (i) $\Rightarrow$ (ii). Let us choose an appropriate basis of $\mathbb{C}^n$ comprising the eigenvectors of σ such that the effect of the pinching $\mathcal{E}_\sigma$ is represented as follows:

$$\mathcal{E}_\sigma : \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1r} \\ B_{21} & B_{22} & \cdots & B_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ B_{r1} & B_{r2} & \cdots & B_{rr} \end{pmatrix} \longmapsto \begin{pmatrix} B_{11} & & & 0 \\ & B_{22} & & \\ & & \ddots & \\ 0 & & & B_{rr} \end{pmatrix}, \tag{29}$$

where the diagonal blocks $B_{11}, B_{22}, \ldots, B_{rr}$ are square matrices. Now, the characteristic polynomial $\varphi_A(x) := \det(xI - A)$ of a Hermitian matrix $A = [a_{ij}]$ is expanded in x as

$$\varphi_A(x) = x^n - x^{n-1} \left(\sum_i a_{ii} \right) + x^{n-2} \left(\sum_{i<j}(a_{ii}a_{jj} - |a_{ij}|^2) \right) + \cdots + (-1)^n \det A.$$

Therefore, the difference between the characteristic polynomial of A and that of $\mathcal{E}_\sigma(A)$ is given by

$$\varphi_A(x) - \varphi_{\mathcal{E}_\sigma(A)}(x) = x^{n-2} \left(- \sum_{(i,j)\in\Delta} |a_{ij}|^2 \right) + \cdots, \tag{30}$$

where

$$\Delta := \{(i,j) \,|\, i < j, \text{ and } a_{ij} \text{ is replaced with zero by the operation (29)}\}.$$

Since (i) implies $\varphi_A(x) = \varphi_{\mathcal{E}_\sigma(A)}(x)$, the coefficient of x^{n-2} in (30) must vanish. As a consequence, we have $a_{ij} = 0$ for all $(i,j) \in \Delta$, proving that $A = \mathcal{E}_\sigma(A)$. $\qquad\square$

C Differential calculus

In calculating directional derivatives of functions on $\mathcal{L}(\mathcal{H})$, knowledge about the Gâteaux derivative is useful [6]. Let $\mathcal{X}$ and $\mathcal{Y}$ be real Banach spaces and let U be an open subset of $\mathcal{X}$. A continuous map $f : U \to \mathcal{Y}$ is said to be *Gâteaux differentiable* at $x \in U$ if, for every $v \in \mathcal{X}$, the limit

$$Df(x)[v] := \lim_{t \to 0} \frac{f(x + tv) - f(u)}{t}$$

exists in $\mathcal{Y}$. The quantity $Df(x)[v]$ is called the *Gâteaux derivative* of f at x in the direction v. If f is Gâteaux differentiable at every point of U, we say that f is Gâteaux differentiable on U. The following basic properties of the Gâteaux derivative are useful in applications.

(I) (Chain rule) If two maps $f : U \to \mathcal{Y}$ and $g : \mathcal{Y} \to \mathcal{Z}$ are Gâteaux differentiable, then their composition $g \circ f$ is also Gâteaux differentiable, and

$$D(g \circ f)(x)[v] = Dg(f(x))[Df(x)[v]]$$

holds for all $x \in U$ and $v \in \mathcal{X}$.

(II) (Product rule) Let τ be a bilinear map from the product of two Banach spaces $\mathcal{Y}_1$ and $\mathcal{Y}_2$ into $\mathcal{Z}$. If two maps $f : U \to \mathcal{Y}_1$ and $g : U \to \mathcal{Y}_2$ are Gâteaux differentiable, then $\tau(f, g)(x) := \tau(f(x), g(x))$ is also Gâteaux differentiable, and

$$D(\tau(f, g))(x)[v] = \tau(Df(x)[v], g(x)) + \tau(f(x), Dg(x)[v])$$

holds for all $x \in U$ and $v \in \mathcal{X}$. As a special case, let $\mathcal{Y}_1 = \mathcal{Y}_2 := \mathcal{L}(\mathcal{H})$ and let τ be the usual product of two operators denoted by $\cdot$. Then we obtain

$$D(f \cdot g)(x)[v] = Df(x)[v] \cdot g(x) + f(x) \cdot Dg(x)[v]$$

for all $x \in U$ and $v \in \mathcal{X}$.

Now we derive some formulae for the Gâteaux derivative. Firstly, let $f(A) = e^A$ with $A \in \mathcal{L}_{\mathrm{sa}}(\mathcal{H})$. Then

$$Df(A)[B] = \int_0^1 e^{(1-t)A} B e^{tA} dt \tag{31}$$

for all $A, B \in \mathcal{L}_{\mathrm{sa}}(\mathcal{H})$. In fact, integrating the identity

$$\frac{d}{dt}\left\{ e^{-tA} e^{t(A+B)} \right\} = e^{-tA} B e^{t(A+B)}$$

and operating e^A from the left, we have the *Dyson expansion*:

$$e^{A+B} - e^A = \int_0^1 e^{(1-t)A} B e^{t(A+B)} dt.$$

Replacing B in the above formula with uB, where $u \in \mathbb{R}$, we obtain

$$\begin{aligned}
\lim_{u \to 0} \frac{e^{(A+uB)} - e^A}{u} &= \lim_{u \to 0} \int_0^1 e^{(1-t)A} B e^{t(A+uB)} dt \\
&= \int_0^1 e^{(1-t)A} B e^{tA} dt.
\end{aligned}$$

Secondly, let $f(A) = \log A$ with $A \in \mathcal{L}_{++}(\mathcal{H})$. Then

$$Df(A)[B] = \int_0^\infty (sI + A)^{-1} B (sI + A)^{-1} ds \tag{32}$$

for all $A \in \mathcal{L}_{++}(\mathcal{H})$ and $B \in \mathcal{L}_{\mathrm{sa}}(\mathcal{H})$. In fact, using the integral representation

$$\log x = \int_0^\infty \left(\frac{1}{s+1} - \frac{1}{s+x} \right) ds,$$

17

we obtain

$$\lim_{u \to 0} \frac{\log(A + uB) - \log A}{u}$$

$$= \lim_{u \to 0} \int_0^\infty \frac{(s+1)^{-1}I - (sI + A + uB)^{-1} - (s+1)^{-1}I + (sI + A)^{-1}}{u} ds$$

$$= \lim_{u \to 0} \int_0^\infty \frac{(sI + A)^{-1} - (sI + A + uB)^{-1}}{u} ds$$

$$= \int_0^\infty (sI + A)^{-1} B (sI + A)^{-1} ds.$$

In the last equality, we used the resolvent identity:

$$(sI + Q)^{-1} - (sI + P)^{-1} = (sI + P)^{-1}(P - Q)(sI + Q)^{-1}.$$

Finally, let $f(A) = A^\lambda$ with $\lambda \in \mathbb{R}$ and $A \in \mathcal{L}_{++}(\mathcal{H})$. Then

$$Df(A)[B] = \lambda \int_0^1 dt \int_0^\infty ds\, A^{(1-t)\lambda}(sI + A)^{-1} B(sI + A)^{-1} A^{t\lambda} \qquad (33)$$

for all $A \in \mathcal{L}_{++}(\mathcal{H})$ and $B \in \mathcal{L}_{\mathrm{sa}}(\mathcal{H})$. In fact, since $f(A) = e^{\lambda \log A} = h(g(A))$, where $g(x) = \lambda \log x$ and $h(x) = e^x$, the chain rule yields

$$\begin{aligned}
Df(A)[B] &= Dh(g(A))[Dg(A)[B]] \\
&= Dh(g(A)) \left[\lambda \int_0^\infty (sI + A)^{-1} B (sI + A)^{-1} ds \right] \\
&= \int_0^1 e^{(1-t)(\lambda \log A)} \left[\lambda \int_0^\infty (sI + A)^{-1} B (sI + A)^{-1} ds \right] e^{t(\lambda \log A)}\, dt.
\end{aligned}$$

In the second and the third equalities, we used (32) and (31), respectively.

D Computation of affine connections

Let us derive the formulae (20) and (21) for the affine connections $\nabla^{(D_\alpha)}$ and $\nabla^{(D_\alpha)*}$. Since

$$\begin{aligned}
Xg^{(D_\alpha)}(Y, Z)\Big|_\rho &= X_\rho \left\{ D_\alpha((YZ)_\rho \| \sigma)\Big|_{\sigma=\rho} \right\} \\
&= D_\alpha((XYZ)_\rho \| \sigma)\Big|_{\sigma=\rho} + D_\alpha((YZ)_\rho \| (X)_\sigma)\Big|_{\sigma=\rho} \\
&= D_\alpha((XYZ)_\rho \| \sigma)\Big|_{\sigma=\rho} - g_\rho^{(D_\alpha)}(\nabla_Y^{(D_\alpha)} Z, X),
\end{aligned}$$

we have

$$g_\rho^{(D_\alpha)}(\nabla_X^{(D_\alpha)} Y, Z) = D_\alpha((ZXY)_\rho \| \sigma)\Big|_{\sigma=\rho} - Zg^{(D_\alpha)}(X, Y)\Big|_\rho. \qquad (34)$$

On the other hand, due to the duality (10), we have

$$\begin{aligned}
g_\rho^{(D_\alpha)}(\nabla_X^{(D_\alpha)*} Y, Z) &= Xg^{(D_\alpha)}(Y, Z)\Big|_\rho - g_\rho^{(D_\alpha)}(\nabla_X^{(D_\alpha)} Z, Y) \\
&= Xg^{(D_\alpha)}(Y, Z)\Big|_\rho + Yg^{(D_\alpha)}(X, Z)\Big|_\rho - D_\alpha((YXZ)_\rho \| \sigma)\Big|_{\sigma=\rho}. \qquad (35)
\end{aligned}$$

18

Equations (34) and (35) imply that computing the affine connections $\nabla^{(D_\alpha)}$ and $\nabla^{(D_\alpha)*}$ is reduced to computing $D_\alpha((XYZ)_\rho \| \sigma)|_{\sigma=\rho}$ and $Xg^{(D_\alpha)}(Y, Z)$.

We first compute $D_\alpha((XYZ)_\rho \| \sigma)|_{\sigma=\rho}$. We see from (15) that

$$
D_\alpha((YZ)_\rho \| \sigma) = \frac{1}{\alpha-1} \frac{\mathrm{Tr}\left\{(YA^{\alpha-1})B_Z\right\} + \mathrm{Tr}\left\{A^{\alpha-1}C_{YZ}\right\}}{\mathrm{Tr}\, A^\alpha}
$$
$$
- \frac{\alpha}{\alpha-1} \frac{\left(\mathrm{Tr}\left\{A^{\alpha-1}B_Z\right\}\right)\left(\mathrm{Tr}\left\{A^{\alpha-1}B_Y\right\}\right)}{(\mathrm{Tr}\, A^\alpha)^2}.
$$

Therefore

$$
D_\alpha((XYZ)_\rho \| \sigma)
$$
$$
= \frac{1}{\alpha-1} \frac{\mathrm{Tr}\left\{(XYA^{\alpha-1})B_Z\right\} + \mathrm{Tr}\left\{(YA^{\alpha-1})(XB_Z)\right\}}{\mathrm{Tr}\, A^\alpha}
$$
$$
+ \frac{1}{\alpha-1} \frac{\mathrm{Tr}\left\{(XA^{\alpha-1})C_{YZ}\right\} + \mathrm{Tr}\left\{A^{\alpha-1}(XC_{YZ})\right\}}{\mathrm{Tr}\, A^\alpha}
$$
$$
- \frac{1}{\alpha-1} \frac{\left(\mathrm{Tr}\left\{(YA^{\alpha-1})B_Z\right\} + \mathrm{Tr}\left\{A^{\alpha-1}C_{YZ}\right\}\right)\mathrm{Tr}\left\{\alpha A^{\alpha-1}B_X\right\}}{(\mathrm{Tr}\, A^\alpha)^2}
$$
$$
- \frac{\alpha}{\alpha-1} \frac{\left(\mathrm{Tr}\, X\left\{A^{\alpha-1}B_Z\right\}\right)\left(\mathrm{Tr}\left\{A^{\alpha-1}B_Y\right\}\right) + \left(\mathrm{Tr}\left\{A^{\alpha-1}B_Z\right\}\right)\left(\mathrm{Tr}\, X\left\{A^{\alpha-1}B_Y\right\}\right)}{(\mathrm{Tr}\, A^\alpha)^2}
$$
$$
+ \frac{\alpha}{\alpha-1} \frac{\left(\mathrm{Tr}\left\{A^{\alpha-1}B_Z\right\}\right)\left(\mathrm{Tr}\left\{A^{\alpha-1}B_Y\right\}\right) \cdot 2\,\mathrm{Tr}\left\{\alpha A^{\alpha-1}B_X\right\}}{(\mathrm{Tr}\, A^\alpha)^3}.
$$

Since

$$
\mathrm{Tr}\, A^\alpha \Big|_{\sigma=\rho} = 1, \qquad \mathrm{Tr}\left\{A^{\alpha-1}B_X\right\}\Big|_{\sigma=\rho} = 0, \qquad \mathrm{Tr}\left\{A^{\alpha-1}C_{YZ}\right\}\Big|_{\sigma=\rho} = 0,
$$

as in the proof of (16), and

$$
\mathrm{Tr}\{A^{\alpha-1}(XC_{YZ})\}\Big|_{\sigma=\rho} = \mathrm{Tr}\,(XYZ\rho) = XYZ\mathrm{Tr}\,\rho = 0,
$$

we have

$$
D_\alpha((XYZ)_\rho \| \sigma)\Big|_{\sigma=\rho}
$$
$$
= \frac{1}{\alpha-1}\,\mathrm{Tr}\left\{(XYA^{\alpha-1})B_Z + (YA^{\alpha-1})(XB_Z) + (XA^{\alpha-1})C_{YZ}\right\}\Big|_{\sigma=\rho}.
$$

$$(36)$$

We next compute $Xg^{(D_\alpha)}(Y, Z)$. We see from (16) that

$$
g_\rho^{(D_\alpha)}(Y, Z) = \frac{1}{\alpha-1}\mathrm{Tr}\left\{(YA^{\alpha-1})B_Z\right\}\Big|_{\sigma=\rho}
$$
$$
= \frac{1}{\alpha-1}\mathrm{Tr}\left\{(\rho^{\frac{1-\alpha}{2\alpha}}(Z\rho)\rho^{\frac{1-\alpha}{2\alpha}})(YA^{\alpha-1})\Big|_{\sigma=\rho}\right\}
$$
$$
= \frac{1}{\alpha-1}\mathrm{Tr}\left\{(Z\rho)\left[\rho^{\frac{1-\alpha}{2\alpha}}(YA^{\alpha-1})\Big|_{\sigma=\rho}\rho^{\frac{1-\alpha}{2\alpha}}\right]\right\}.
$$

19

Using this identity, we have

$$
\begin{aligned}
Xg^{(D_\alpha)}(Y,Z) &= \frac{1}{\alpha-1}\mathrm{Tr}\left\{(XZ\rho)\left[\rho^{\frac{1-\alpha}{2\alpha}}\left.(YA^{\alpha-1})\right|_{\sigma=\rho}\rho^{\frac{1-\alpha}{2\alpha}}\right]\right\} \\
&\quad +\frac{1}{\alpha-1}\mathrm{Tr}\left\{(Z\rho)\,X\left[\rho^{\frac{1-\alpha}{2\alpha}}\left.(YA^{\alpha-1})\right|_{\sigma=\rho}\rho^{\frac{1-\alpha}{2\alpha}}\right]\right\} \\
&= \frac{1}{\alpha-1}\mathrm{Tr}\left\{\rho^{\frac{1-\alpha}{2\alpha}}(XZ\rho)\rho^{\frac{1-\alpha}{2\alpha}}\left.(YA^{\alpha-1})\right|_{\sigma=\rho}\right\} \\
&\quad +\frac{1}{\alpha-1}\mathrm{Tr}\left\{(Z\rho)\,X\left[\rho^{\frac{1-\alpha}{2\alpha}}\left.(YA^{\alpha-1})\right|_{\sigma=\rho}\rho^{\frac{1-\alpha}{2\alpha}}\right]\right\} \\
&= \frac{1}{\alpha-1}\left.\mathrm{Tr}\left\{(XB_Z)(YA^{\alpha-1})\right\}\right|_{\sigma=\rho} \\
&\quad +\frac{1}{\alpha-1}\mathrm{Tr}\left\{(Z\rho)\,X\left[\rho^{\frac{1-\alpha}{2\alpha}}\left.(YA^{\alpha-1})\right|_{\sigma=\rho}\rho^{\frac{1-\alpha}{2\alpha}}\right]\right\}. \quad (37)
\end{aligned}
$$

Substituting (36) and (37) into (34) and (35), we obtain the formulae (20) and (21).

References

[1] S.-I. Amari and H. Nagaoka, *Methods of Information Geometry*, Translations of Mathematical Monographs 191 (AMS and Oxford, RI, 2000).

[2] K. M. R. Audenaert and N. Datta, "α-z-relative Rényi entropies," J. Math. Phys., **56**, 022202 (2015); arXiv:1310.7178.

[3] N. Ay and S.-I. Amari, "A novel approach to canonical divergences within information geometry," Entropy, **17**, 8111-8129 (2015).

[4] S. Beigi, "Sandwiched Rényi divergence satisfies data processing inequality," J. Math. Phys., **54**, 122202 (2013); arXiv:1306.5920.

[5] M. Berta, O. Fawzi, and M. Tomamichel, "On variational expressions for quantum relative entropies," arXiv:1512.02615.

[6] R. Bhatia, *Matrix Analysis*, Graduate Texts in Mathematics 169 (Springer, NY. 1997).

[7] I. Csiszár, "On topological properties of f-divergences, Studia Scientiarium Mathematicarium Hungaria, **2**, 329-339 (1967).

[8] N. Datta and F. Leditzky, "A limit of the quantum Rényi divergence," J. Phys. A: Math. Theor., **47**, 045304 (2014); arXiv:1308.5961.

[9] S. Eguchi, "Geometry of minimum contrast," Hiroshima Math. J., **22**, 631-647 (1992).

[10] R. L. Frank and E. H. Lieb, "Monotonicity of a relative Rényi entropy," J. Math. Phys., **54**, 122202 (2013): arXiv:1306.5358.

[11] T. Furuta, "Concrete examples of operator monotone functions obtained by an elementary method without appealing to Löwner integral representation," Lin. Alg. Appl., **429**, 972-980 (2008).

[12] H. Hasegawa and D. Petz, "Non-commutative extension of information geometry II," in *Quantum Communication, Computing, and Measurement*, Edited by Hirota *et al.*, 109–118 (Plenum, NY, 1997).

[13] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).

[14] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).

[15] M. Mosonyi and T. Ogawa, "Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies," Commun. Math. Phys., **334**, 1617-1648 (2015); arXiv:1309.3228.

[16] M. Mosonyi and T. Ogawa, "Strong converse exponent for classical-quantum channel coding," arXiv:1409.3562.

[17] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr and M. Tomamichel, "On quantum Rényi entropies: a new definition and some properties," J. Math. Phys., **54**, 122203 (2013); arXiv:1306.3142.

[18] D. Petz, "Monotone metrics on matrix spaces," Lin. Alg. Appl., **224**, 81-96 (1996).

[19] A. Rényi, "On measures of information and entropy," in *Proc. Symp. on Math. Stat. and Probability*, 547-561 (University of California Press, Berkeley, 1961).

[20] M. M. Wilde, A. Winter, and D. Yang, "Strong converse for the classical capacity of entanglement-breaking and Hadamard channels," Commun. Math. Phys., **331**, 593-622 (2014); arXiv:1306.1586.

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