Investigating Properties of a Family of Quantum Rényi Divergences

Simon M. Lin\textsuperscript{1} and Marco Tomamichel\textsuperscript{2,1}

\textsuperscript{1}Centre for Quantum Technologies, National University of Singapore, Singapore 117543, Singapore
\textsuperscript{2}School of Physics, The University of Sydney, Sydney 2006, Australia

Audenaert and Datta recently introduced a two-parameter family of relative Rényi entropies, known as the $\alpha$-$z$-relative Rényi entropies. The definition of the $\alpha$-$z$-relative Rényi entropy unifies all previously proposed definitions of the quantum Rényi divergence of order $\alpha$ under a common framework. Here we will prove that the $\alpha$-$z$-relative Rényi entropies are a proper generalization of the quantum relative entropy by computing the limit of the $\alpha$-$z$ divergence as $\alpha$ approaches one and $z$ is an arbitrary function of $\alpha$. We also show that certain operationally relevant families of Rényi divergences are differentiable at $\alpha = 1$. Finally, our analysis reveals that the derivative at $\alpha = 1$ evaluates to half the relative entropy variance, a quantity that has attained operational significance in second-order quantum hypothesis testing.

I. INTRODUCTION

In classical information theory, the Shannon entropy\textsuperscript{1} measures the average information content of a given message, while the Kullback-Leibler divergence\textsuperscript{2} (also known as relative entropy) is a non-symmetric measure of the difference between the probability distributions. Roughly speaking, the Shannon entropy is a measure of the uncertainty and unpredictability of the information content, while the Kullback-Leibler divergence is a measure of how well a probability distribution approximates another probability distribution, for the purposes of measuring information content. In addition, the Shannon entropy also gives a bound on the maximal lossless compression possible, and the performance of data compression algorithms are usually benchmarked against the Shannon entropy.

From a mathematical perspective, the Shannon entropy\textsuperscript{1} is uniquely characterized by a set of axioms, that is, the Shannon entropy satisfies the axioms, and the only functional on the set of probability distributions that satisfies the axioms is the Shannon entropy. By relaxing one of the axioms, we obtain a family of functionals on the set of probability distributions indexed by a parameter $\alpha$, called the Rényi entropies of order $\alpha$\textsuperscript{3}. In particular, when $\alpha$ tends to 1, we recover our usual notion of the Shannon entropy. In a similar spirit, we also have that the Kullback-Leibler divergence is uniquely characterized by a set of axioms, and that we may also obtain a family of functions on the set of pairs of probability distributions indexed by a parameter $\alpha$, called the Rényi divergences of order $\alpha$\textsuperscript{3}. In particular, when $\alpha$ tends to 1, we recover our usual notion of the Kullback-Leibler divergence.

The families of Rényi entropies of order $\alpha$ and Rényi divergences of order $\alpha$ contain several other entropies and divergences of operational significance, which are used e.g. in cryptography and information theory. For instance, when $\alpha$ tends to infinity, we obtain the minimum entropy and maximum relative entropy, where the minimum entropy quantifies how hard it is to guess a random variable. More generally, the Rényi entropies and divergences are a prominent tool in information theory\textsuperscript{4}, and have an operational interpretation, for example in hypothesis testing or channel coding.

A natural question to ask is if the notion of Rényi divergence has a natural extension in the quantum setting. Indeed, by replacing the notion of probability distributions with the notion of positive semi-definite operators on a finite-dimensional Hilbert space, we could obtain the quantum counterparts of Shannon entropy and Kullback-Leibler divergence, namely the von Neumann entropy\textsuperscript{5}, and the quantum relative entropy\textsuperscript{6}. Moreover, several definitions for the quantum Rényi divergences of order $\alpha$ have been proposed. The most widely used definition is based on Petz’ quasi-entropies (see, e.g.\textsuperscript{7}) but more recently a different definition has been proposed independently by Müller-Lennert \textit{et al.}\textsuperscript{8} and Wilde \textit{et al.}\textsuperscript{9}. Both the initial and the new definitions have found operational interpretation in quantum hypothesis testing. (See\textsuperscript{10} and references therein.) The non-uniqueness of the definition is clearly a result of the non-commutativity of general quantum states and it is possible to devise various different generalizations of the classical Rényi divergence.

\* email: a0067236@nus.edu.sg
Most recently, a two-parameter family for the quantum Rényi divergence, namely the α-z relative Rényi entropies, has been proposed by Audenaert and Datta [11], in the hope that it could shed some light on the problem of the definition of the quantum Rényi divergence of order α. In this work we establish some properties of the general α-z entropies, in particular we establish that they are continuous and differentiable at α = 1 under some weak constraints. We compute the derivative at α = 1 and find that it corresponds to the relative information variance that has recently found operational significance in second-order quantum hypothesis testing [12, 13].

The remainder of this paper is organized as follows. Section II introduces the relevant notation and definitions of the Rényi divergences used in this work. Section III then provides an overview and discussion of the main technical results. Their proofs are provided in Section IV.

II. NOTATIONS AND PRELIMINARIES

A. Notation

Throughout this paper, let $\mathcal{H}$ denote a finite-dimensional Hilbert space, which is a finite-dimensional vector space over the field of complex numbers, endowed with a complex inner product, and let $L(\mathcal{H})$ denote the set of linear operators on $\mathcal{H}$. Furthermore, let $T(\mathcal{H})$ denote the set of all Hermitian operators on $\mathcal{H}$, $P(\mathcal{H})$ denote the set of all positive semi-definite operators on $\mathcal{H}$, and let $P^+(\mathcal{H})$ denote the set of all positive definite operators on $\mathcal{H}$. It is an easy exercise to verify that $P^+(\mathcal{H}) \subseteq P(\mathcal{H}) \subseteq T(\mathcal{H}) \subseteq L(\mathcal{H})$.

Also, we note that a linear operator $\rho \in L(\mathcal{H})$ is Hermitian if and only if it is unitarily diagonalizable, and its eigenvalues are real. A Hermitian operator $\rho \in T(\mathcal{H})$ is positive semi-definite, denoted $\rho \geq 0$, (respectively positive definite, denoted $\rho > 0$) if and only if its eigenvalues are non-negative (respectively positive). Also, for all $\rho, \sigma \in P(\mathcal{H})$, we write $\rho \geq \sigma$ or $\sigma \leq \rho$ if and only if $\rho - \sigma \geq 0$. Next, let $D(\mathcal{H})$ denote the set of all density operators (all positive semi-definite operators of trace 1) on $\mathcal{H}$, and let $S(\mathcal{H})$ denote the set of all sub-normalized operators (all positive semi-definite operators of trace at most 1) on $\mathcal{H}$.

For a given $\rho \in L(\mathcal{H})$, we denote the kernel of $\rho$, the span of vectors $v$ of $\rho$ such that $\rho(v) = 0$, by $\ker \rho$. For a given $\rho \in T(\mathcal{H})$ and $v \in \mathcal{H}$, we have $v \in \ker \rho$ if and only if $v$ is a eigenvector of $\rho$ associated to the zero eigenvalue. Also, for a given $\rho \in T(\mathcal{H})$, we denote the support of $\rho$, the span of eigenvectors of $\rho$ associated to the non-zero eigenvalues of $\rho$ by $\text{supp} \rho$. It is an easy exercise to verify that $\mathcal{H} = \ker \rho \oplus \text{supp} \rho$ for all $\rho \in T(\mathcal{H})$. For any $\sigma, \tau \in T(\mathcal{H})$, we use the notation $\sigma \triangleright \tau$ to indicate the fact that $\ker \sigma \subseteq \ker \tau$ (or equivalently, $\supp \tau \subseteq \supp \sigma$), and we use the notation $\sigma \perp \tau$ to indicate that $\sigma$ and $\tau$ are orthogonal to each other. Also, for any $\rho \in T(\mathcal{H})$, we denote the projection operator of $\rho$ onto $\text{supp} \rho$ by $\Pi_\rho$. Finally, we shall denote the Schatten $p$-norm on the vector space $T(\mathcal{H})$, where $p \in [1, \infty)$, by $\| \cdot \|_p$.

B. Classical Divergences

Firstly, let us recall the definition of the Kullback-Leibler divergence, which is a non-symmetric measure of the difference between the probability distributions $X$ and $Y$ [2]:

**Definition 1.** Let $X$ and $Y$ be random variables with alphabet $\{a_1, a_2, \cdots, a_n\}$, and let $p(a_i)$ and $q(a_i)$ denote the probabilities that the outcomes $a_i$ occurs in the events $X$ and $Y$ respectively. The **Kullback-Leibler divergence** of $Y$ from $X$, denoted $D_{KL}(X \| Y)$, is defined as follows:

$$D_{KL}(X \| Y) = \sum_{i=1}^{n} p(a_i) \log \frac{p(a_i)}{q(a_i)}.$$ 

In 1961, Alfred Rényi produced a set of axioms that would uniquely characterize the Kullback-Leibler divergence. By relaxing one of the conditions in one of the axioms, Rényi introduced an one-parameter family of divergences, which are uniquely characterized by the relaxed set of axioms, and is a generalization of the Kullback-Leibler divergence [3]:

**Definition 2.** Let $X$ and $Y$ be random variables with alphabet $\{a_1, a_2, \cdots, a_n\}$, and let $p(a_i)$ and $q(a_i)$ denote the probabilities that the outcome $a_i$ occurs in the events $X$ and $Y$ respectively. The **Rényi divergence** of order $\alpha$, denoted $D_{\alpha}(X \| Y)$, is defined as follows:

$$D_{\alpha}(X \| Y) = \sum_{i=1}^{n} p(a_i) \left[ \log \frac{p(a_i)}{q(a_i)} \right]^{\frac{1}{\alpha}},$$ 

for $\alpha \in (0, \infty)$.
**Definition 6.** Let $\alpha \in (0, 1) \cup (1, \infty)$, the Renyi divergence of order $\alpha$ is defined as follows:

$$D_\alpha(X\|Y) = \frac{1}{1-\alpha} \left( \sum_{i=1}^n p(x_i)^\alpha q(x_i)^{1-\alpha} \right).$$

In particular, it is well known that $\lim_{\alpha \to 1} D_\alpha(X\|Y) = D_{KL}(X\|Y)$.

**C. Quantum Divergences**

In the quantum setting, the Renyi divergence of order $\alpha$, where $\alpha \in (0, 1) \cup (1, \infty)$, has a natural extension to the quantum setting, in the easier case where two positive semi-definite operators commute.

**Definition 3.** Let $\rho, \sigma \in \mathcal{P}(\mathcal{H})$ be given, with $\rho \neq 0$, such that $\sigma \gg \rho$, and $\rho$ and $\sigma$ commute. The quantum Renyi divergence of order $\alpha$, where $\alpha \in (0, 1) \cup (1, \infty)$, from $\rho$ to $\sigma$, denoted $D_\alpha(\rho\|\sigma)$, is defined as follows:

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

Umegaki found that the quantum relative entropy could be easily extended to the non-commuting case as well [8].

**Definition 4.** Let $\rho, \sigma \in \mathcal{P}(\mathcal{H})$ be given with $\rho \neq 0$. The quantum relative entropy from $\rho$ to $\sigma$, denoted $D(\rho\|\sigma)$, is defined as follows:

$$D(\rho\|\sigma) = \begin{cases} \frac{\text{Tr}(\rho \log \rho - \log \sigma)}{\text{Tr}(\rho)} & \text{if } \sigma \gg \rho, \\ \infty & \text{otherwise} \end{cases}$$

Müller-Lennert et al [8] and independently, Wilde et al [9], proposed a new definition for the quantum Renyi divergence of order $\alpha$ [10].

**Definition 5.** Let $\rho, \sigma \in \mathcal{P}(\mathcal{H})$ and $\alpha \in (0, 1) \cup (1, \infty)$ be given with $\rho \neq 0$. The quantum Renyi divergence of order $\alpha$ from $\rho$ to $\sigma$, denoted $\check{D}_\alpha(\rho\|\sigma)$, is defined as follows:

$$\check{D}_\alpha(\rho\|\sigma) = \left\{ \begin{array}{ll} \frac{1}{1-\alpha} \log \frac{\text{Tr}(\rho^{\frac{1}{1-\alpha}} \sigma^\alpha)}{\text{Tr}(\rho)} & \text{if } \rho \nleq \sigma \text{ and } \sigma \gg \rho \text{ or } \alpha < 1, \\ \infty & \text{otherwise} \end{array} \right.$$  

Here, we note that when $\rho$ and $\sigma$ commute, the two definitions for the quantum Renyi divergence of order $\alpha$, namely $D_\alpha(\rho\|\sigma)$ and $\check{D}_\alpha(\rho\|\sigma)$, coincide.

Furthermore, Mosonyi and Ogawa proposed a piecewise definition for the quantum Renyi divergence of order $\alpha$, based on its operational interpretation in quantum hypothesis testing [10].

**Definition 6.** Let $\rho \in \mathcal{D}(\mathcal{H})$, $\sigma \in \mathcal{P}(\mathcal{H})$ and $\alpha \in (0, 1) \cup (1, \infty)$ be given with $\sigma \gg \rho$. The quantum Renyi divergence of order $\alpha$ from $\rho$ to $\sigma$, denoted $\tilde{D}_\alpha(\rho\|\sigma)$, is defined as follows:

$$\tilde{D}_\alpha(\rho\|\sigma) = \left\{ \begin{array}{ll} D_\alpha(\rho\|\sigma) & \text{if } \alpha < 1, \\ \check{D}_\alpha(\rho\|\sigma) & \text{if } \alpha > 1 \end{array} \right.$$  

The definition $\check{D}_\alpha(\rho\|\sigma)$ for the quantum Renyi divergence of order $\alpha$, proposed by Mosonyi and Ogawa, satisfies the data processing inequality for all admissible values of $\alpha$ [8,14,15]. Furthermore, this definition for the quantum Renyi divergence of order $\alpha$ satisfies the axioms proposed by Müller-Lennert et al. for the quantum Renyi divergence of order $\alpha$ for all $\alpha \in (0, 1) \cup (1, \infty)$.

While the three proposed definitions for the quantum Renyi divergence of order $\alpha$ have found operational significances so far in literature, all three definitions do not possess perfect limiting properties. For instance, the relative max-, and collision entropies are shown to be not a specialization of $D_\alpha(\rho\|\sigma)$ for any value of $\alpha$ [12,17], while the relative min entropy is shown not to be a specialization of $\tilde{D}_\alpha(\rho\|\sigma)$ for any value of $\alpha$ [18].

---

1 In the latter paper, the divergence is called “sandwiched Renyi relative entropy”.

---

3
D. \( \alpha \)-z relative Rényi entropies

Finally, Audenaert and Datta proposed a two-parameter family of \( \alpha \)-z relative Rényi entropies [11], and along with the definition, proved some limiting properties of the \( \alpha \)-z relative Rényi entropies as \( z \) tends to infinity, and \( \alpha \) tends to 0 [11]. The aim of defining a two-parameter family of \( \alpha \)-z relative Rényi entropies is to seek to unite the different proposed definitions for the Rényi divergences.

Firstly, we shall recall the definition of the \( \alpha \)-z relative Rényi entropy from [11], and along with the definition, proved some limiting properties of the \( \alpha \)-z relative Rényi entropies as

\[
\lim_{\alpha \to \infty} D_{\alpha,z}(\rho \| \sigma) = D_{\alpha}(\rho \| \sigma) = \frac{1}{\alpha - 1} \log \text{Tr} \left( \left( \frac{1}{\alpha} \rho^z \frac{1}{\alpha} \sigma^z \right)^\alpha \right).
\]

Hereafter, we will refer the \( \alpha \)-z relative Rényi entropies as \( \alpha \)-z divergences for convenience.

In the cases \( z = 1 \) and \( z = \alpha \), it is easy to see that the quantities \( D_{\alpha}(\rho \| \sigma) \) and \( D_{\alpha}(\rho \| \sigma) \) are specializations of the \( \alpha \)-z divergences, in the case where \( \rho \in \mathcal{D}(\mathcal{H}) \) (that is, \( \text{Tr}(\rho) = 1 \)). Indeed, we have

\[
D_{\alpha,1}(\rho \| \sigma) = D_{\alpha}(\rho \| \sigma) = \frac{1}{\alpha - 1} \log \text{Tr} \left( \sigma^{1-\alpha} \rho^\alpha \right), \quad \text{and}
\]

\[
D_{\alpha,\alpha}(\rho \| \sigma) = \tilde{D}_{\alpha}(\rho \| \sigma) = \frac{1}{\alpha - 1} \log \text{Tr} \left( \left( \frac{1}{\alpha} \rho^z \frac{1}{\alpha} \sigma^z \right)^\alpha \right).
\]

III. MAIN RESULTS AND DISCUSSION

Here we present our main results. Their proofs, together with various auxiliary result, are presented in the following section. Our first result establishes that the \( \alpha \)-z divergence is a generalization of the relative entropy in the following strong sense:

**Theorem 8.** Let \( \rho \in \mathcal{D}(\mathcal{H}) \) and \( \sigma \in \mathcal{P}(\mathcal{H}) \) with \( \sigma \gg \rho \). Suppose \( J \) is an open interval containing 1, and \( g : J \to \mathbb{R} \) is a continuously differentiable function that satisfies \( g(1) \neq 0 \). Then we have

\[
\lim_{\alpha \to 1} D_{\alpha,g(\alpha)}(\rho \| \sigma) = D(\rho \| \sigma) = \text{Tr}(\rho(\log \rho - \log \sigma)).
\]

Here, \( \log \rho \) and \( \log \sigma \) are evaluated on their supports.

Henceforth, we may proceed to define

\[
D_{1,z}(\rho \| \sigma) := \lim_{\alpha \to 1} D_{\alpha,z}(\rho \| \sigma) = D(\rho \| \sigma) = \text{Tr}(\rho(\log \rho - \log \sigma))
\]

for all \( z \neq 0 \). In particular, our proof generalizes the arguments in [8, 9] which established that \( \lim_{\alpha \to 1} \tilde{D}_{\alpha}(\rho \| \sigma) = D(\rho \| \sigma) \). We also remark that we have

\[
\lim_{\alpha \to 1} \lim_{z \to \infty} D_{\alpha,z}(\rho \| \sigma) = D(\rho \| \sigma) = \text{Tr}(\rho(\log \rho - \log \sigma))
\]

in general, where this limit has been proved by Audenaert and Datta [11].

Finally, we remark that the condition \( g(1) \neq 0 \) is crucial in proving that

\[
\lim_{\alpha \to 1} D_{\alpha,g(\alpha)}(\rho \| \sigma) = D(\rho \| \sigma) = \text{Tr}(\rho(\log \rho - \log \sigma)).
\]

Indeed, the limit is not reproduced in general when \( g(1) = 0 \) (see [11, Theorem 2] for further details).

Our second main result establishes both \( \alpha \to D_{\alpha}(\rho \| \sigma) \) and \( \alpha \to \tilde{D}_{\alpha}(\rho \| \sigma) \) are continuously differentiable at \( \alpha = 1 \). Moreover, their derivatives agree and are proportional to the relative entropy variance that has found operational significance in the context of the second-order analysis in quantum hypothesis testing [12, 13].
Theorem 9. Let $\rho \in \mathcal{D}(\mathcal{H})$ and $\sigma \in \mathcal{P}(\mathcal{H})$ be given with $\sigma \gg \rho$. Then $D_{\alpha,1}(\rho\|\sigma)$ and $D_{\alpha,\alpha}(\rho\|\sigma)$ are differentiable at $\alpha = 1$, and we have
\[
\left.\frac{d}{d\alpha} D_{\alpha,1}(\rho\|\sigma)\right|_{\alpha=1} = \left.\frac{d}{d\alpha} D_{\alpha,\alpha}(\rho\|\sigma)\right|_{\alpha=1} = \frac{1}{2} V(\rho\|\sigma), \quad \text{where}
\]
\[
V(\rho\|\sigma) := \text{Tr} \left( \rho (\log \rho - \log \sigma)^2 \right) - \left( \text{Tr} (\rho (\log \rho - \log \sigma)) \right)^2
\]
is the relative entropy variance \cite{12,13}. Here, $\log \rho$ and $\log \sigma$ are evaluated on their supports.

The second result thus in particular establishes that the piecewise definition of Mosonyi and Ogawa, $D_{\alpha}(\rho\|\sigma)$ in Def. \cite{6}, is continuously differentiable in $\alpha$. An interesting open question to ask is if the function $\alpha \mapsto D_{\alpha}(\rho\|\sigma)$ is in fact smooth. We conjecture that the second derivative of $D_{\alpha}(\rho\|\sigma)$ does not exist at $\alpha = 1$, and hence the second derivative of the function $\alpha \mapsto D_{\alpha}(\rho\|\sigma)$ is not continuous, and hence not smooth. However, due to the non-commutativity of a matrix function and its derivative, it is difficult to find a sufficiently nice closed form expression for the second derivative of the function $\alpha \mapsto D_{\alpha}(\rho\|\sigma)$, and it remains an open problem to decide if the second derivative is continuous.

IV. PROOFS

A. Partial Derivative of the $\alpha$-z Divergence with Respect to $z$

It is generally well known that the functions $\alpha \mapsto D_{\alpha}(\rho\|\sigma)$ and $\alpha \mapsto \tilde{D}_{\alpha}(\rho\|\sigma)$ are monotonically increasing \cite{14,17}. A natural question to ask then is whether if the $\alpha$-$z$ divergences are monotone with respect to $\alpha$ or $z$. The following follows by a simple application of the Araki-Lieb-Thirring trace inequality \cite{13,20}.

Proposition 10. Let us fix a $\rho \in \mathcal{D}(\mathcal{H})$ and $\sigma \in \mathcal{P}(\mathcal{H})$ with $\sigma \gg \rho$, and an $\alpha \in \mathbb{R} \setminus \{1\}$. Then for all $z > 0$, we have $z \mapsto \frac{d}{dz} D_{\alpha,z}(\rho\|\sigma)$ to be monotonically decreasing if $\alpha > 1$, and monotonically increasing if $\alpha < 1$.

In general, we remark that proving monotonicity properties for the $\alpha$-$z$ divergences with respect to $\alpha$ is much harder than proving monotonicity properties for the $\alpha$-$z$ divergences with respect to $z$. However, numerical results lead us to conjecture that the $\alpha$-$z$ divergence is monotone with respect to $\alpha$ for a fixed $z$. Nevertheless, we shall prove some local monotonicity properties for the $\alpha$-$z$ divergences with respect to $\alpha$ at $\alpha = 1$ in the following subsections.

B. Continuity of $\alpha$-$z$ Divergence as $\alpha \to 1$

The main purpose of this section is to prove Theorem \cite{8}. It is generally well known (using the product rule) that if $\mathcal{H}$ be a finite-dimensional Hilbert space, $J$ an open interval in $\mathbb{R}$, and $F : J \to \mathcal{P}(\mathcal{H})$ is a differentiable function, then the function $x \mapsto \text{Tr}((F(x))^n)$ is a real-valued differentiable function for all $n \in \mathbb{Z}$, with derivative $\text{Tr}(n(F(x))^{n-1}F'(x))$.

Some of the main ingredients for this proof are the following generalizations of this statement, which are established in Appendix A and B respectively.

Proposition 11. Let $J$ be an open interval in $\mathbb{R}$, and suppose $F : J \to \mathcal{P}^+(\mathcal{H})$ is a continuously differentiable function. Fix a $z \in \mathbb{R}$. Then the function $x \mapsto \text{Tr}(F(x)^z)$ is continuously differentiable, and
\[
\frac{d}{dx} \text{Tr}(F(x)^z) = \text{Tr}(z(F(x))^{z-1}F'(x)).
\]

Proposition 12. Let $J$ be an open interval in $\mathbb{R}$, and suppose $F : J \to \mathcal{P}^+(\mathcal{H})$ is a continuously differentiable function. Then the function $x \mapsto \text{Tr}(F(x)^z)$ is continuously differentiable, and
\[
\frac{d}{dx} \text{Tr}(F(x)^z) = \text{Tr}(F'(x)^z \log F(x) + z F(x) F'(x) (F(x))^{z-1}).
\]

As an immediate consequence of the latter statement, we find the following:
Corollary 13. For all $\rho \in \mathcal{D}(\mathcal{H})$ and $\sigma \in \mathcal{P}(\mathcal{H})$ with $\sigma \gg \rho$, and $z \in \mathbb{R} \setminus \{0\}$, we have

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

Proof. Note that the function $F : z \mapsto \left( \sigma^{\frac{1-\alpha}{2z}} \rho^\frac{1}{2} \sigma^{\frac{1-\alpha}{2z}} \right)^z$ is continuously differentiable, and that the derivative of $F$ is continuous in $\alpha$ by Proposition 12. Hence, it suffices to evaluate $\frac{\partial}{\partial z}$ at $\alpha = 1$. Indeed, we have

$$\frac{\partial}{\partial z} \left( \sigma^{\frac{1-\alpha}{2z}} \rho^\frac{1}{2} \sigma^{\frac{1-\alpha}{2z}} \right) = -\frac{1-\alpha}{2z^2} \sigma^{\frac{1-\alpha}{2z}} (\log \sigma) \rho^\frac{1}{2} \sigma^{\frac{1-\alpha}{2z}} - \frac{1}{2z^2} \sigma^{\frac{1-\alpha}{2z}} \rho^\frac{1}{2} \sigma^{\frac{1-\alpha}{2z}} (\log \sigma).$$  

Hence, at $\alpha = 1$, we have

$$\frac{\partial}{\partial z} \left( \sigma^{\frac{1-\alpha}{2z}} \rho^\frac{1}{2} \sigma^{\frac{1-\alpha}{2z}} \right) = -\frac{1}{2z} \rho^\frac{1}{2} \log \rho.$$  

Thus, Proposition 12 at $\alpha = 1$ yields

$$\frac{\partial}{\partial z} \Tr \left( \sigma^{\frac{1-\alpha}{2z}} \rho^\frac{1}{2} \sigma^{\frac{1-\alpha}{2z}} \right)^z = \Tr \left( \rho \log \rho^\frac{1}{2} - z \left( \frac{1}{2z} \rho^\frac{1}{2} \log \rho \right) \rho^\frac{1}{2} \right)$$

$$= \Tr \left( \rho \log \rho^\frac{1}{2} - \frac{1}{z} \rho \log \rho \right) = 0.$$  

$\square$

Let us also recall l’Hôpital’s rule:

Theorem 14. (l’Hôpital’s rule. [21, Theorem 5.13]) Let $J$ be an open interval in $\mathbb{R}$, and $c \in J$. Furthermore, let $f$ and $g$ be differentiable functions on $J \setminus \{c\}$, such that $\lim_{x \to c} f(x) = \lim_{x \to c} g(x) = 0$ or $\pm \infty$, $\lim_{x \to c} \frac{f(x)}{g(x)}$ exists, and $g’(x) \neq 0$ for all $x \in J \setminus \{c\}$, then $\lim_{x \to c} \frac{f(x)}{g(x)}$ exists, and is equal to $\lim_{x \to c} \frac{f’(x)}{g’(x)}$.

Now, we are ready to prove the following special case of Theorem 8.

Proposition 15. Let us fix a $\rho \in \mathcal{D}(\mathcal{H})$ and $\sigma \in \mathcal{P}(\mathcal{H})$ with $\sigma \gg \rho$. Then for all $z \neq 0$, we have

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

Here, $\log \rho$ and $\log \sigma$ are evaluated on their supports.

Proof. By restricting $\rho$ to its support, we may assume that $\rho$ is invertible. Firstly, let us define $F : (0, \infty) \to \mathcal{P}(\mathcal{H})$ by $F(\alpha) = \sigma^{\frac{1-\alpha}{2z}} \rho^\frac{1}{2} \sigma^{\frac{1-\alpha}{2z}}$. Then it is clear that $F : (0, \infty) \to \mathcal{P}(\mathcal{H})$ is a well-defined continuously differentiable function. Furthermore, we see that $F(1) = \rho^\frac{1}{2}$ is invertible. Now, for each $d \times d$ Hermitian matrix $A$, let us denote the smallest eigenvalue of $A$ by $\lambda_d(A)$. Then $\lambda_d$ is continuous on the space of $d \times d$ Hermitian matrices. In particular, the function $x \mapsto |\lambda_d(F(x))|$ is continuous. As $|\lambda_d(F(1))| \neq 0$ (since $F(1)$ is invertible), there exists some $r \in (0,1/2)$, such that $F(x)$ is invertible for all $x \in [1-r,1+r]$. Then by Proposition 11, the function $g(\alpha) = \Tr(F(\alpha)^z)$ is differentiable at $\alpha = 1$, with derivative $\Tr(z(F(1))^{z-1}F’(1))$.

Now, since $g(1) = \Tr(F(1)^z) = \Tr(\rho) = 1$, we see that $\lim (\alpha - 1) = \lim \log g(\alpha) = 0$. Furthermore, we see that $\frac{\partial}{\partial \alpha}(\alpha - 1) = 1 \neq 0$. Thus, it remains to compute $\lim_{\alpha \to 1} \frac{\partial}{\partial \alpha} (\log g(\alpha))$.

Firstly, we note that $\frac{\partial}{\partial \alpha} (\log g(\alpha)) = \frac{g’(\alpha)}{g(\alpha)}$, so that $\lim_{\alpha \to 1} \frac{\partial}{\partial \alpha} (\log g(\alpha)) = \frac{g’(1)}{g(1)} = g’(1)$. Now, we have

$$F’(\alpha) = -\frac{1}{2z} (\log \sigma) \sigma^{\frac{1-\alpha}{2z}} \rho^\frac{1}{2} \sigma^{\frac{1-\alpha}{2z}} + \frac{1}{z} \sigma^{\frac{1-\alpha}{2z}} (\log \rho) \rho^\frac{1}{2} \sigma^{\frac{1-\alpha}{2z}} - \frac{1}{2z} \sigma^{\frac{1-\alpha}{2z}} \rho^\frac{1}{2} \sigma^{\frac{1-\alpha}{2z}} \log \sigma.$$
This implies that \( F'(1) = \frac{1}{2} \left( -\frac{1}{2} (\log \sigma) \rho^\frac{1}{\sigma} + (\log \rho) \rho^\frac{1}{\rho} - \frac{1}{2} \rho^\frac{1}{\rho} \log \sigma \right) \). Hence,

\[
g'(1) = \text{Tr}(z(F(1))^{\frac{1}{z} - 1} F'(1))
\]

\[
= \text{Tr} \left( \rho^{\frac{1}{z} - 1} \left( -\frac{1}{2} (\log \sigma) \rho^\frac{1}{\sigma} + (\log \rho) \rho^\frac{1}{\rho} - \frac{1}{2} \rho^\frac{1}{\rho} \log \sigma \right) \right)
\]

\[
= \text{Tr}(\rho (\log \rho - \log \sigma)).
\]

Finally, l'Hôpital’s rule yields

\[
\lim_{\alpha \to 1} D_{\alpha,z}(\rho \| \sigma) = \lim_{\alpha \to 1} \frac{\log g(\alpha)}{\alpha - 1} = \lim_{\alpha \to 1} \frac{g'(\alpha)}{g(\alpha)} = \text{Tr}(\rho (\log \rho - \log \sigma)).
\]

Finally, we are ready to prove Theorem 8.

**Theorem 8.**

By restricting \( \rho \) to its support, we may assume that \( \rho \) is invertible. Since \( g(1) \neq 0 \), there exists some \( r \in (0, 1) \) such that \( g(x) \neq 0 \) for all \( x \in [1-r, 1+r] \). Firstly, let us define \( H : (0, \infty) \times \mathbb{R} \rightarrow \mathcal{P}(\mathcal{H}) \) by \( H(\alpha, z) = \left( \sigma \frac{H(\alpha, z)}{\rho} \right)^{\frac{1}{z}} \), and \( F : [1-r, 1+r] \rightarrow \mathcal{P}(\mathcal{H}) \) by \( F(\alpha) = H(\alpha, g(\alpha)) \). Then we have \( \frac{d}{d\alpha}(F(\alpha)) = \frac{dg}{d\alpha}(H(\alpha, z)) + \frac{dg}{d\alpha} \frac{d}{d\alpha} (H(\alpha, z)) \), which implies that \( F : [1-r, 1+r] \rightarrow \mathcal{P}(\mathcal{H}) \) is a continuously differentiable function. By employing a similar argument as in Proposition 14, we have

\[
\lim_{\alpha \to 1} D_{\alpha,g(\alpha)}(\rho \| \sigma) = \lim_{\alpha \to 1} \frac{\text{Tr}(F'(\alpha))}{\text{Tr}(F(\alpha))}
\]

\[
= \lim_{\alpha \to 1} \text{Tr}(F'(\alpha))
\]

\[
= \lim_{\alpha \to 1} \text{Tr} \left( \frac{\partial}{\partial \alpha} (H(\alpha, z)) + \frac{dg}{d\alpha} \frac{\partial}{\partial \alpha} (H(\alpha, z)) \right)
\]

\[
= \lim_{\alpha \to 1} \text{Tr} \left( \frac{\partial}{\partial \alpha} (H(\alpha, z)) \right) + \lim_{\alpha \to 1} \text{Tr} \left( \frac{dg}{d\alpha} \frac{\partial}{\partial \alpha} (H(\alpha, z)) \right)
\]

\[
= D(\rho \| \sigma) + \lim_{\alpha \to 1} \frac{dg}{d\alpha} \text{Tr} \left( \lim_{\alpha \to 1} \frac{\partial}{\partial \alpha} (H(\alpha, z)) \right)
\]

\[
= D(\rho \| \sigma).
\]

\[\square\]

**C. Differentiability of \( \alpha \cdot z \) Divergence as \( \alpha \to 1 \)**

The main purpose of this section is to prove Theorem 8. We shall first recall a consequence of L’Hôpital’s Rule:

**Corollary 16.** Let \( J \) be an open interval containing \( \alpha \in \mathbb{R} \), and \( f \) be a continuous function on \( J \), and differentiable on \( J \setminus \{a\} \). Suppose furthermore that \( \lim_{x \to a} f'(x) \) exists. Then \( f \) is differentiable at \( x = a \), and \( f'(a) = \lim_{x \to a} f'(x) \).

The proof of Theorem 8 follows from the following two proposition.

**Proposition 17.** Let \( \rho \in \mathcal{D}(\mathcal{H}) \) and \( \sigma \in \mathcal{P}(\mathcal{H}) \) be given with \( \sigma \gg \rho \). Then \( D_{\alpha,1}(\rho \| \sigma) \) is differentiable at \( \alpha = 1 \), and we have

\[
\frac{d}{d\alpha} D_{\alpha,1}(\rho \| \sigma) = \frac{V(\rho \| \sigma)}{2}
\]

at \( \alpha = 1 \). Here, \( \log \rho \) and \( \log \sigma \) are evaluated on their supports.
Proof. Let us define \( f(\alpha) = \text{Tr} \left( \sigma^{1-\alpha} \rho^\alpha \sigma^{-\alpha} \right) = \text{Tr} \left( \rho^\alpha \sigma^{1-\alpha} \right) \) for all \( \alpha \in \mathbb{R} \). Then it is easy to see that \( f \) is non-zero and infinitely differentiable everywhere on \( \mathbb{R} \). This implies that the derivative of \( D_{\alpha,1}(\rho||\sigma) = \frac{\log f(\alpha)}{\alpha-1} \) exists for all \( \alpha \in \mathbb{R} \setminus \{1\} \), and is equal to \( \frac{(\alpha-1)\frac{d}{d\alpha}(\log f(\alpha)) - \log f(\alpha)}{(\alpha-1)^2} \). By Corollary 16 it suffices to show that \( \lim_{\alpha \to 1} \frac{d}{d\alpha} D_{\alpha,1}(\rho||\sigma) \) exists, and is equal to \( \frac{V(\rho||\sigma)}{2} \). Firstly, since

\[
\lim_{\alpha \to 1} \log f(\alpha) = \log f(1) = \log \text{Tr}(\rho) = 0,
\]

and \( \lim_{\alpha \to 1} \frac{d}{d\alpha}(\log f(\alpha)) \) exists, we have

\[
\lim_{\alpha \to 1} \left[ (\alpha - 1) \frac{d}{d\alpha}(\log f(\alpha)) - \log f(\alpha) \right] = 0.
\]

Furthermore, we see that

\[
\frac{d}{d\alpha} \left( (\alpha - 1) \frac{d}{d\alpha}(\log f(\alpha)) - \log f(\alpha) \right) = \frac{d}{d\alpha}(\log f(\alpha)) + (\alpha - 1) \frac{d^2}{d^2\alpha}(\log f(\alpha)) - \frac{d}{d\alpha}(\log f(\alpha)) = (\alpha - 1) \frac{d^2}{d^2\alpha}(\log f(\alpha)).
\]

This implies that

\[
\lim_{\alpha \to 1} \frac{(\alpha - 1) \frac{d}{d\alpha}(\log f(\alpha)) - \log f(\alpha)}{(\alpha - 1)^2} = \lim_{\alpha \to 1} \frac{d}{d\alpha} \left( (\alpha - 1) \frac{d}{d\alpha}(\log f(\alpha)) \right) \frac{2}{2(\alpha - 1)} = \lim_{\alpha \to 1} \frac{d^2}{d^2\alpha} \left( \frac{1}{2} \log f(\alpha) \right) = \lim_{\alpha \to 1} \frac{d}{d\alpha} \left( \frac{f'(\alpha)}{2f(\alpha)} \right) = \lim_{\alpha \to 1} \frac{d}{d\alpha} \left( \frac{f(\alpha)f''(\alpha) - (f'(\alpha))^2}{2(f(\alpha))^2} \right) = \frac{f(1)f''(1) - (f'(1))^2}{2(f(1))^2} = \frac{f''(1) - (f'(1))^2}{2}.
\]

Now, we note that

\[
f'(\alpha) = \text{Tr} \left( \rho^\alpha \log(\rho)\sigma^{1-\alpha} - \rho^\alpha \log(\sigma)\sigma^{1-\alpha} \right) = \text{Tr} \left( \rho^\alpha \sigma^{1-\alpha} \log(\rho - \log(\sigma)) \right),
\]

and

\[
f''(\alpha) = \text{Tr} \left( \rho^\alpha \log(\rho)\sigma^{1-\alpha} \log(\rho - \log(\sigma)) - \rho^\alpha \log(\sigma)\sigma^{1-\alpha} \log(\rho - \log(\sigma)) \right) = \text{Tr} \left( \rho^\alpha \log(\rho - \log(\sigma))\sigma^{1-\alpha} \log(\rho - \log(\sigma)) \right).
\]

This implies that \( f'(1) = \text{Tr} \left( \rho(\log(\rho - \log(\sigma)) \right) \) and \( f''(1) = \text{Tr} \left( \rho(\log(\rho - \log(\sigma))^2 \right) \), concluding the proof.

**Proposition 18.** Let \( \rho \in \mathcal{D}(\mathcal{H}) \) and \( \sigma \in \mathcal{P}(\mathcal{H}) \) be given with \( \sigma \gg \rho \). Then \( D_{\alpha,\sigma}(\rho||\sigma) \) is differentiable at \( \alpha = 1 \), and we have

\[
\frac{d}{d\alpha} D_{\alpha,\sigma}(\rho||\sigma) = \frac{V(\rho||\sigma)}{2}
\]

at \( \alpha = 1 \). Here, \( \log \rho \) and \( \log \sigma \) are evaluated on their supports.
Proof. Let us constrain ourselves to the line $\alpha = z$. Then we have
\[
\frac{d}{d\alpha} D_{\alpha,\alpha}(\rho||\sigma) = \frac{\partial}{\partial \alpha} D_{\alpha,z}(\rho||\sigma) + \frac{dz}{d\alpha} \frac{\partial}{\partial z} D_{\alpha,z}(\rho||\sigma) = \frac{\partial}{\partial \alpha} D_{\alpha,z}(\rho||\sigma) + \frac{\partial}{\partial z} D_{\alpha,z}(\rho||\sigma).
\]
The desired then follows since at $\alpha = z = 1$, we have
\[
\frac{\partial}{\partial \alpha} D_{\alpha,z}(\rho||\sigma) = \frac{\partial}{\partial \alpha} D_{\alpha,1}(\rho||\sigma) = \frac{\text{Tr} \left( (\rho \log \rho - \log \sigma)^2 \right) - (\text{Tr} (\rho (\log \rho - \log \sigma))^2)}{2},
\]
and $\frac{\partial}{\partial z} D_{\alpha,z}(\rho||\sigma) = 0$.

ACKNOWLEDGMENTS

The results in this paper were achieved as part of a final year project at the School of Computing, Department of Computer Science at the National University of Singapore under the supervision of Prof. Stephanie Wehner, and we want to thank her for discussions and for providing a conducive research environment at the Centre for Quantum Technologies. SL would also like to thank Jedrzej Kaniewski, Patrick Coles, and Mischa Woods for discussions leading to the results in this project. Finally, he would like to thank Nelly Ng, his tutor in the Introduction to Information Theory (CS3236) class, for getting him interested in information theory. SL acknowledges support from the Agency for Science, Technology and Research (A*STAR). MT is funded by the Ministry of Education (MOE) and National Research Foundation Singapore, as well as MOE Tier 3 Grant Random numbers from quantum processes (MOE2012-T3-1-009).

Appendix A: Proof of Proposition 11

Before we commence with the proof, we will need a few technical lemmas:

Lemma 19. Let $z \in \mathbb{R} \setminus \mathbb{Z}$, and $a > 0$. Then the Taylor series expansion of $x^z$ about the point $x = a$ has radius of convergence $a$.

Proof. Consider the Taylor series expansion $x^z = \sum_{n=0}^{\infty} a_n (x - a)^n$. We note that
\[
a_n = \frac{z(z - 1) \cdots (z - (n - 1)) a^{z-n}}{n!}, \quad \text{and} \quad \frac{a_n}{a_{n+1}} = \frac{a(n+1)}{z-n},
\]
for all positive integers $n$. Since $\lim_{n \to \infty} \left| \frac{a_n}{a_{n+1}} \right| = a$, the desired follows.

As a consequence of the preceding lemma, we have the following:

Corollary 20. Let $z \in \mathbb{R} \setminus \mathbb{Z}$, $A$ be a positive definite $d \times d$ matrix. Also, let us choose some $a > \|A\|_{\infty}$. Furthermore, for each $n \in \mathbb{N}$, we define
\[
a_n = \frac{z(z - 1) \cdots (z - (n - 1)) a^{z-n}}{n!}, \quad a_0 = a^z.
\]
Then the series $\sum_{n=0}^{\infty} a_n (A - aI)^n$ converges (in the Schatten $\infty$-norm) to $A^z$.

Proof. Since $A$ is positive definite, we have $A = U^{-1}DU$ for some unitary matrix $d \times d$ U and some diagonal $d \times d$ matrix $D$, where $D$ has positive eigenvalues. This implies that $A - aI = U^{-1} (D - AI)U$. Noting that the Schatten $\infty$-norm is unitarily invariant, we have
\[
\left\| \sum_{n=0}^{k} a_n (A - aI)^n - A^z \right\|_{\infty} = \left\| U^{-1} \left( \sum_{n=0}^{k} a_n (D - aI)^n - D^z \right) U \right\|_{\infty} = \left\| \sum_{n=0}^{k} a_n (D - aI)^n - D^z \right\|_{\infty}.
\]
for all \( k \in \mathbb{N} \). Since \( \alpha > \|A\|_{\infty} = \|D\|_{\infty} \), we see that the series \( \sum_{n=0}^{\infty} a_n(D - aI)^n \) converges in the Schatten \( \infty \)-norm to \( D^z \) by Lemma 19, so we must have

\[
\lim_{k \to \infty} \left\| \sum_{n=0}^{k} a_n(D - aI)^n - D^z \right\|_{\infty} = 0
\]

by definition. This concludes the proof.

Next, we shall recall a useful result about uniform convergence. The following theorem is a modified version of [21, Theorem 7.17]:

**Theorem 21.** Let \( \{f_n\}_{n=0}^{\infty} \) be a sequence of differentiable functions on \([a, b]\) such that \( \sum_{n=0}^{\infty} f_n \) converges to a function \( f \), and \( \sum_{n=0}^{\infty} f'_n \) converges uniformly on \([a, b]\) to a function \( g \). Then \( f \) is differentiable on \([a, b]\), and \( f' = g \).

We are now ready to prove Proposition 11.

**Proposition 11** Fix a \( c \in J \), and choose \( r > 0 \) such that \([c - r, c + r] \subseteq J \). We note that \( x \mapsto \text{Tr}(F(x)) \) is a continuous function on \( J \), so there exists a \( M > 0 \) such that \( \text{Tr}(F(x)) < M \) for all \( x \in [c - r, c + r] \).

Furthermore, since \( F' \) is continuous, it follows that \( \sup_{x \in [c - r, c + r]} \|F'(x)\|_{\infty} \) exists in \( \mathbb{R}_{\geq 0} \). Now, for each \( n \in \mathbb{N} \), we define \( a_n = \frac{z(z - (n - 1)M^z - n)}{n!} \), and \( a_0 = M^z \). By assumption, we have the series \( \sum_{n=0}^{\infty} a_n(F(x) - MI)^n \) converges in the Schatten \( \infty \)-norm to \( F(x)^z \) for all \( x \in [c - r, c + r] \).

Now, for each \( x \in [c - r, c + r] \), and \( n \in \mathbb{Z}_{\geq 0} \), we define \( g(x) = \text{Tr}(F(x)^z) \), and \( g_n(x) = \text{Tr}(a_n(F(x) - MI)^n) \).

Then it is clear that \( \sum_{n=0}^{\infty} g_n \) converges pointwise to \( g \), and each \( g_n \) is differentiable with derivative \( g'_n(x) = \text{Tr}(na_n(F(x) - MI)^{n-1}F'(x)) \). Furthermore, let us define \( h(x) = \text{Tr}(z(F(x))^z - 1 F'(x)) \).

We shall show that \( \sum_{n=0}^{\infty} g'_n \) converges uniformly on \([c - r, c + r]\) to \( h \).

Firstly, we note that by a similar argument as in the proofs of Corollary 20, we have the series \( \sum_{n=1}^{\infty} na_n(F(x) - MI)^{n-1} \) to converge in the Schatten \( \infty \)-norm to \( zF(x)^{z-1} \) for all \( x \in [c - r, c + r] \).

Furthermore, let us define \( h_n(x) = na_n(F(x) - MI)^{n-1} F'(x) \) for all \( n \in \mathbb{N} \), and \( h_{\infty}(x) = zF(x)^{z-1} F'(x) \) for all \( x \in [c - r, c + r] \). Then for all \( k \in \mathbb{N} \), we have

\[
\left\| \sum_{n=1}^{k} h_n(x) - h_{\infty}(x) \right\|_{\infty} = \left\| \sum_{n=1}^{k} na_n(F(x) - MI)^{n-1} F'(x) - zF(x)^{z-1} F'(x) \right\|_{\infty} \\
\leq \left\| \sum_{n=1}^{k} na_n(F(x) - MI)^{n-1} - zF(x)^{z-1} \right\|_{1} \|F'(x)\|_{\infty} \\
= \left\| \sum_{n=k+1}^{\infty} na_n(F(x) - MI)^{n-1} - \sum_{n=1}^{\infty} na_n(F(x) - MI)^{n-1} \right\|_{\infty} \|F'(x)\|_{\infty} \\
= \left\| \sum_{n=k+1}^{\infty} na_n(F(x) - MI)^{n-1} \right\|_{\infty} \|F'(x)\|_{\infty}
\]
Now, we note that the power series
\[ \lambda_n \]
for all \( k \) (and hence absolutely) on any closed interval in \((0, k)\). This implies that
\[ \lim_{n \to \infty} \left| |n| \right| = 0. \]
By definition of the Schatten \( \infty \)-norm, we see that if \( \lambda_{d,x} > 0 \) is the smallest eigenvalue of \( F(x) \), then
\[ \|(F(x) - MI)\|_\infty = M - \lambda_{d,x}. \]
Now, for each \( d \times d \) Hermitian matrix \( A \), let us denote the smallest eigenvalue of \( A \) by \( \lambda_d(A) \). Then \( \lambda_d \) is continuous on the space of \( d \times d \) Hermitian matrices. In particular, the function \( x \mapsto |\lambda_d(F(x) - MI)| = M - \lambda_{d,x} \) is continuous, and hence achieves a maximum on \([c - r, c + r]\). Let \( \lambda_{\max} := \max_{x \in [c - r, c + r]} |\lambda_d(F(x) - MI)| \). Since \( F(x) \) is invertible for all \( x \in J \), we must have \( \lambda_{\max} < M \). Hence, we have
\[
\left| \sum_{n=1}^{k} h_n(x) - h_\infty(x) \right|_\infty \leq \sum_{n=k+1}^{\infty} |na_n| \left| |(F(x) - MI)\|_1 \right|^{n-1} ||F'(x)||_\infty \]
\[
\leq \sum_{n=k+1}^{\infty} |na_n| \lambda_{\max}^{n-1} ||F'(x)||_\infty. \]
Now, we note that the power series \( \sum_{n=1}^{\infty} na_n(y - M)^{n-1} \) about the point \( y = M \) converges uniformly (and hence absolutely) on any closed interval in \((0, 2M)\) to \( y^{z-1} \). In particular, since \( 0 \leq \lambda_{\max} < M \), we must have the power series \( \sum_{n=1}^{\infty} |na_n| \lambda_{\max}^{n-1} \) to converge by assumption, which implies that
\[
\lim_{k \to \infty} \sum_{n=k+1}^{\infty} |na_n| \lambda_{\max}^{n-1} = 0. \]
As
\[
\left| \sum_{n=1}^{k} h_n(x) - h_\infty(x) \right|_\infty \leq \sum_{n=k+1}^{\infty} |na_n| \lambda_{\max}^{n-1} ||F'(x)||_\infty \]
for all \( k \in \mathbb{N} \) and \( x \in [c - r, c + r] \), this implies that \( \sum_{n=1}^{k} h_n \) converges uniformly in Schatten \( \infty \)-norm on \([c - r, c + r]\) to \( h_\infty \). Furthermore, since any two norms on a finite-dimensional Hilbert space are equivalent, this implies that \( \sum_{n=1}^{k} h_n \) converges uniformly in Schatten 1-norm on \([c - r, c + r]\) to \( h_\infty \). Finally, since
\[
\left| \sum_{n=0}^{k} g'_n(x) - h(x) \right| = \left| \text{Tr} \left( \sum_{n=1}^{k} h_n(x) - h_\infty(x) \right) \right| \]
\[
\leq \text{Tr} \left( \left| \sum_{n=1}^{k} h_n(x) - h_\infty(x) \right|_1 \right) \]
\[
= \left| \sum_{n=1}^{k} h_n(x) - h_\infty(x) \right|_1 \]
for all \( k \in \mathbb{N} \), this implies that \( \sum_{n=0}^{\infty} g'_n \) converges uniformly on \([c - r, c + r]\) to \( h \). Therefore, by Theorem 21 we have \( g \) to be differentiable, with derivative \( h \) as desired.  

Appendix B: Proof of Proposition 12

Lemma 22. For all \( x > 0 \), we have
\[
\log x = \int_{0}^{\infty} \frac{1}{1 + s} - \frac{1}{x + s} ds. \]
As a corollary, we deduce that:
Corollary 23. Let $A \in \mathcal{P}^+(\mathcal{H})$ be given. Then we have $\log A = \int_0^\infty \frac{1}{1+s} I - (A + sI)^{-1} ds$. Here, $\log A$ refers to the principal logarithm of $A$.

Next, we recall a useful result about the derivatives of the inverses of invertible matrices:

Lemma 24. \cite[Section 6.5]{22} Let $J$ be an open interval in $\mathbb{R}$, and suppose $F : J \to \mathcal{P}^+(\mathcal{H})$ is a differentiable function. Then we have

$$\frac{d}{dx} (f(x)^{-1}) = -f(x)^{-1} f'(x) f(x)^{-1}.$$  

In addition to the previous lemma, the following theorem, which is a modified version of \cite[Theorem 3.2, Section XIII]{23}, establishes the precise conditions under which we can interchange the differentiation sign with the integral sign:

Theorem 25. Let $F : [a, \infty) \times [c, d] \to \mathcal{L}(\mathcal{H})$ be a function of two variables $(t, x)$ defined for all $t \geq a$, and $x$ in some interval $J = [c, d]$, $c < d$. Assume that $\frac{\partial F}{\partial x}$ exists, and that both $F$ and $\frac{\partial F}{\partial x}$ are continuous. Assume that there are non-negative functions $\phi : [a, \infty) \to \mathbb{R}$ and $\psi : [a, \infty) \to \mathbb{R}$, such that $\|F(t, x)\|_\infty \leq \phi(t)$ and $\|\frac{\partial F}{\partial x}(t, x)\|_\infty \leq \psi(t)$ for all $t$ and $x$, and such that the integrals $\int_a^\infty \phi(t) dt$ and $\int_a^\infty \psi(t) dt$ converge. Let $g(x) = \int_a^x F(t, x) dt$. Then $g$ is differentiable, and

$$\frac{dg}{dx} = \int_a^x \frac{\partial F}{\partial x}(t, x) dt.$$  

With that, we are ready to prove the following proposition:

Proposition 26. Let $J$ be a closed and bounded interval in $\mathbb{R}$, and suppose $F : J \to \mathcal{P}^+(\mathcal{H})$ is a continuously differentiable function. Then the function $x \mapsto \log F(x)$ is continuously differentiable, and $\frac{d}{dx} (\log F(x)) = \int_0^\infty (F(x) + sI)^{-1} F'(x)(F(x) + sI)^{-1} ds$. Here, $\log F(x)$ refers to the principal logarithm of $F(x)$.

Proof. Let us define $H(s, x) = \frac{1}{1+s} I - (F(x) + sI)^{-1}$ for all $x \in J$ and $s > 0$. Then it is clear that $\frac{\partial H}{\partial s}$ exists and is equal to $(F(x) + sI)^{-1} F'(x)(F(x) + sI)^{-1}$. Also, it is clear that both $H$ and $\frac{\partial H}{\partial s}$ are continuous. Now, for each $d \times d$ Hermitian matrix $A$, let us denote the smallest eigenvalue of $A$ by $\lambda_d(A)$. Then it is clear that from the definition of the Schatten $\infty$-norm that $\| (F(x) + sI)^{-1} \|_\infty = \frac{1}{\lambda_d(F(x)) + s}$. Also, $\lambda_d$ is continuous on the space of $d \times d$ Hermitian matrices. In particular, the function $x \mapsto \lambda_d(F(x))$ is continuous. This implies that

$$\| (F(x) + sI)^{-1} \|_\infty = \left\| \frac{1}{1+s} I - (F(x) + sI)^{-1} \right\|_\infty$$

$$= \left\| (F(x) + sI)^{-1} ((F(x) + sI) - (1+s)I) \frac{1}{1+s} \right\|_\infty$$

$$\leq \frac{\| F(x) - I \|_\infty}{\lambda_d(F(x)) + s (1+s)}$$

$$\leq \frac{\| F(x) - I \|_\infty}{\inf_{x \in J} \lambda_d(F(x)) + s (1+s)}.$$

Also, we have

$$\| (F(x) + sI)^{-1} F'(x)(F(x) + sI)^{-1} \|_\infty \leq \left( \frac{1}{\inf_{x \in J} \lambda_d(F(x)) + s} \right)^2 \sup_{x \in J} \| F'(x) \|_\infty.$$

By setting

$$\phi(s) = \frac{\sup_{x \in J} \| F(x) - I \|_\infty}{\inf_{x \in J} \lambda_d(F(x)) + s (1+s)},$$

and

$$\phi^{-1}(t) = \frac{1}{\sup_{x \in J} \| F(x) - I \|_\infty} \frac{1}{\inf_{x \in J} \lambda_d(F(x)) + s (1+s)},$$

we have

$$\phi(s) \phi^{-1}(t) = \frac{1}{\sup_{x \in J} \| F(x) - I \|_\infty} \frac{1}{\inf_{x \in J} \lambda_d(F(x)) + s (1+s)}.$$
\[ \psi(s) = \left( \inf_{x \in J} \lambda_0(F(x)) + s \right)^2 \sup_{x \in J} \|F'(x)\|_\infty, \]

it is easy to check that \( \int_0^\infty \phi(s)ds \) and \( \int_0^\infty \psi(s)ds \) converge. As \( F(x) = \int_0^\infty H(s,x)ds \) for all \( x \in J \), it follows from the Theorem \[23\] that

\[ \frac{d}{dx} \log F(x) = \int_0^\infty \frac{\partial H}{\partial x}(s,x)ds = \int_0^\infty (F(x) + sI)^{-1} F'(x) (F(x) + sI)^{-1} ds. \]

Finally, since \( F \) and \( F' \) are continuous, so is \( \frac{d}{dx} (\log F(x)) \), and we are done. \( \square \)

Next, let us recall some useful technical lemmas:

**Lemma 27.** The Taylor series expansion of \( e^x \) about the point \( x = 0 \) has infinite radius of convergence, and \( e^x = \sum_{n=0}^\infty \frac{x^n}{n!} \) for all \( x \in \mathbb{R} \).

**Corollary 28.** Let \( A \) be a positive definite \( d \times d \) matrix. Then we have \( A^x = \sum_{n=0}^\infty \frac{x^n}{n!} (\log A)^n \) (Equivalently, \( \sum_{n=0}^\infty \frac{x^n}{n!} \) converges in Schatten \( \infty \)-norm to \( A^x \)).

**Proof.** By a similar argument as in Corollaries ?? and [20], we have \( A = e^{\log A} \). Let us show that \( A^x = e^{x \log A} \). Indeed, there exists an invertible \( d \times d \) matrix \( B \) and a diagonal \( d \times d \) matrix \( D \) with positive eigenvalues (diagonal entries), such that \( A = BDB^{-1} \). This implies that \( \log A = B(\log D)B^{-1} \), and hence we have

\[ A^x = (BDB^{-1})^x = BDB^{-1} = B e^{x \log D} B^{-1} = e^{B(x \log D)B^{-1}} = e^{x \log A}. \]

The desired statement now follows from the preceding lemma. \( \square \)

With that, we are ready to compute the partial derivative of \( D_{\alpha,z}(\rho\|\sigma) \) with respect to \( z \):

**Proof.** Firstly, we note that \( \text{Tr}(F(x)^x) = \sum_{n=0}^\infty x^n \text{Tr} \left( \frac{(\log F(x))^n}{n!} \right) \) for all \( x \in J \). Next, by denoting \( \frac{d}{dx} (\log F(x)) \) by \( G(x) \), we note that

\[ \frac{d}{dx} \left( x^n \text{Tr} \left( \frac{(\log F(x))^n}{n!} \right) \right) = x^{n-1} \text{Tr} \left( \frac{(\log F(x))^n}{(n-1)!} \right) + x^n \text{Tr} \left( \frac{(\log F(x))^{n-1}G(x)}{(n-1)!} \right) \]

for all positive integers \( n \) and \( x \in J \). By applying a similar argument as in the proof of Proposition [11], we deduce that for all \( x \in J \) and \( r_x > 0 \) such that \( [x - r_x, x + r_x] \subseteq J \), we have \( \sum_{n=1}^\infty \frac{x^{n-1}(\log F(x))^n}{(n-1)!} \) to converge uniformly to \( F(x)^x \log F(x) \)

on \( [x - r_x, x + r_x] \), and \( \sum_{n=1}^\infty \frac{x^n(\log F(x))^{n-1}G(x)}{(n-1)!} \) to converge uniformly to \( xF(x)^xG(x) \) on \( [x - r_x, x + r_x] \).

Consequently, we have \( \sum_{n=1}^\infty x^{n-1} \text{Tr} \left( \frac{(\log F(x))^n}{(n-1)!} \right) \) to converge uniformly to \( \text{Tr}(F(x)^x \log F(x)) \) on \( [x - r_x, x + r_x] \), and \( \sum_{n=1}^\infty x^n \text{Tr} \left( \frac{(\log F(x))^{n-1}G(x)}{(n-1)!} \right) \) to converge uniformly to \( \text{Tr}(xF(x)^xG(x)) \) on \( [x - r_x, x + r_x] \).
This implies that the map \( x \mapsto Tr(F(x)^x) \) is continuously differentiable, and \( \frac{d}{dx}(Tr(F(x)^x)) = Tr(F(x)^x \log F(x) + xF(x)^{x}G(x)) \). It remains to simplify the expression \( Tr(xF(x)^{x}G(x)) \).

Now, we note that \( G(x) = \int_0^{\infty} f(x + sI)^{-1} f'(x)(f(x + sI)^{-1} ds \) by Lemma [23]. This implies that

\[
Tr(xF(x)^{x}G(x)) = Tr \left( xF(x)^{x} \int_0^{\infty} (F(x) + sI)^{-1} F'(x)(F(x) + sI)^{-1} ds \right)
\]

\[
= \int_0^{\infty} Tr(xF(x)^{x}(F(x) + sI)^{-1} F'(x)(F(x) + sI)^{-1} ds
\]

\[
= \int_0^{\infty} Tr(xF'(x)(F(x) + sI)^{-2} ds
\]

\[
= Tr \left( xF'(x)(F(x)^{x} \int_0^{\infty} (F(x) + sI)^{-2} ds \right)
\]

\[
= Tr \left( xF'(x)(F(x)^{x} \right)^{-1}
\]

\[
= Tr \left( xF'(x)(F(x)^{x} \right)^{-1}
\]

Hence, we have \( \frac{d}{dx}(Tr(F(x)^x)) = Tr(F(x)^x \log F(x) + xF'(x)(F(x)^x)^{-1} \) as desired. Clearly,

\[
\frac{d}{dx}(Tr(F(x)^x))
\]

is continuous, and we are done. \( \square \)

[1] Shannon, C. A mathematical theory of communication. The Bell System Technical Journal, Volume 27, pp. 379–423 (1948).
[2] Kullback, S. and Leibler, R. A. On Information and Sufficiency. The Annals of Mathematical Statistics, vol. 22, Number 1, pp. 79-86 (1951).
[3] Rényi, A. On measures of information and entropy. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1: Contributions to the Theory of Statistics, University of California Press, pp. 547-561 (1961).
[4] Csiszár, I. Generalized Cutoff Rates and Rényis Information Measures. IEEE Transactions on Information Theory, vol. 41, no. 1, pp. 26-34 (1995).
[5] von Neumann, J. Mathematische Grundlagen der Quantenmechanik. Springer, Berlin, 1932; English Translation, Princeton University Press, Princeton.
[6] Umegaki, H. Conditional expectations in an operator algebra. IV. Entropy and information. Kodai Mathematical Seminar Reports, Volume 14, Number 2, pp. 5985 (1962).
[7] Ohya, M. and Petz, D. Quantum Entropy and Its Use. Springer, 1993.
[8] Müller-Lennert, M., Dupuis, F., Szehr, O., Fehr, S., and Tomamichel, M. On quantum Rényi entropies: a new generalization and some properties. Journal of Mathematical Physics, vol. 54, no. 12, p. 122203 (2013).
[9] Wilde, M. M., Winter, A. and Yang, D. Strong converse for the classical capacity of entanglement-breaking channels. June 2013. [arXiv:1306.1586]
[10] Mosonyi, M. and Ogawa, T. Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies. September 2013. [arXiv:1309.3228]
[11] Audenaert, K. M. R. and Datta, N. \( \alpha \)-Relative Rényi Entropies. October 2013. [arXiv:1313.7175]
[12] Tomamichel, M. and Hayashi, M. A Hierarchy of Information Quantities for Finite Block Length Analysis of Quantum Tasks. IEEE Transactions on Information Theory, vol. 59, no. 11, pp. 7693–7710 (2013).
[13] Li, K., Second-Order Asymptotics for Quantum Hypothesis Testing Annals of Statistics, vol. 42, no. 1, pp. 171–189 (2014).
[14] Beigi, S. Quantum Rényi divergence Satisfies Data Processing Inequality. June 2013. [arXiv:1306.5920]
[15] Frank, R. L. and Lieb, E. H. Monotonicity of a relative Rényi entropy. June 2013. [arXiv:1306.5358v2].
[16] Tomamichel, M., Colbeck, R. and Renner, R. A fully quantum asymptotic equipartition property. IEEE Transactions on Information Theory, vol. 55, no. 12, pp. 5840-5847 (2009).
[17] Tomamichel, M. A Framework for non-asymptotic quantum information theory. PhD thesis, ETH Zurich, 2012. [arXiv:1203.2142]
[18] Datta, N. and Leditzky, F. A limit of the quantum Rényi divergence. August 2013. [arXiv:1308.5961]
[19] Araki, H. On an Inequality of Lieb and Thirring. Letters in Mathematical Physics, vol. 19, pp. 167-170 (1990).
[20] Lieb, E. H. and Thirring, W. E. *Inequalities for the Moments of the Eigenvalues of the Schrödinger Hamiltonian and Their Relation to Sobolev Inequalities*. Edited by Lieb, E. H., Simon, B. and Wrightman, A. S. in Studies in Mathematical Physics, Princeton Press, pp. 269-303 (1976).

[21] Rudin, W. *Principles of Mathematical Analysis*. Third edition. International Series in Pure and Applied Mathematics. McGraw-Hill Book Co., New York-Auckland-Düsseldorf, 1976.

[22] Horn, R. A. and Johnson, C. R. *Topics in matrix analysis*. Corrected reprint of the 1991 original. Cambridge University Press, Cambridge, 1994.

[23] Lang, S. *Undergraduate Analysis*. Second edition. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1997.