Convexity Properties of the Quantum Rényi Divergences, with Applications to the Quantum Stein's Lemma *

Milán Mosonyi[1,2]

1 **Física Teòrica: Informació i Fenomens Quàntics, Universitat Autònoma de Barcelona, ES-08193 Bellaterra (Barcelona), Spain**
2 **Mathematical Institute, Budapest University of Technology and Economics, Egry József u 1., Budapest, 1111 Hungary**

Abstract

We show finite-size bounds on the deviation of the optimal type II error from its asymptotic value in the quantum hypothesis testing problem of Stein's lemma with composite null-hypothesis. The proof is based on some simple properties of a new notion of quantum Rényi divergence, recently introduced in [Müller-Lennert, Dupuis, Szehr, Fehr and Tomamichel, J. Math. Phys. **54**, 122203, (2013)], and [Wilde, Winter, Yang, arXiv:1306.1586].

1998 ACM Subject Classification E.4 Coding and information theory, H.1.1 Information theory

Keywords and phrases Quantum Rényi divergences, Stein's lemma, composite null-hypothesis, second-order asymptotics

Digital Object Identifier 10.4230/LIPIcs.TQC.2014.88

1 Introduction

Rényi defined the α-divergence [36] of two probability distributions p, q on a finite set $\mathcal{X}$ as

$$D_\alpha(p\|q) := \frac{1}{\alpha-1}\log\sum_{x\in\mathcal{X}} p(x)^\alpha q(x)^{1-\alpha},$$

where $\alpha \in (0,+\infty)\setminus\{1\}$. These divergences have various desirable mathematical properties; they are strictly positive, non-increasing under stochastic maps, and jointly convex for $\alpha \in (0,1)$ and jointly quasi-convex for $\alpha > 1$. For fixed p and q, $D_\alpha(p\|q)$ is a monotone increasing function of α, and the limit $\alpha \to 1$ yields the relative entropy (a.k.a. Kullback-Leibler divergence), probably the single most important quantity in information theory. Even more importantly, the Rényi divergences have great operational significance, as quantifiers of the trade-off between the relevant operational quantities in many information theoretic tasks, including hypothesis testing, source compression, and information transmission through noisy channels [12]. A direct operational interpretation of the Rényi divergences as generalized cutoff rates has been shown in [12].

In the view of the above, it is natural to look for an extension of the Rényi divergences for pairs of quantum states. One such extension has been known in quantum information theory for quite some time, defined for states ρ and σ as [34]

$$D_\alpha^{(\mathrm{old})}(\rho\|\sigma) := \frac{1}{\alpha-1}\log \mathrm{Tr}\,\rho^\alpha\sigma^{1-\alpha}.$$

* This work was partially supported by the European Research Council Advanced Grant "IRQUAT".

© Milán Mosonyi;
licensed under Creative Commons License CC-BY

9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC'14).
Editors: Steven T. Flammia and Aram W. Harrow; pp. 88–98

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

These divergences also form a monotone increasing family, with the Umegaki relative entropy $D_1(\rho\|\sigma) := \operatorname{Tr}\rho(\log\rho - \log\sigma)$ as their limit at $\alpha \to 1$. They are also strictly positive; however, monotonicity under stochastic (i.e., completely positive and trace-preserving) maps only holds for $\alpha \in [0,2]$. Recently, a new quantum Rényi divergence has been introduced in [28, 41], defined as

$$D_\alpha^{(\mathrm{new})}(\rho\|\sigma) := \frac{1}{\alpha-1}\log\operatorname{Tr}\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha.$$

Again, these new divergences yield the Umegaki relative entropy in the limit $\alpha \to 1$, and monotonicity only holds on a restricted domain, in this case for $\alpha \in [1/2, +\infty)$.

Operational interpretation has been found for both definitions in the setting of binary hypothesis testing for different and matching domains of α. The goal in hypothesis testing is to decide between two candidates, ρ and σ, for the true state of a quantum system, based on a measurement on many identical copies of the system. The quantum Stein's lemma [19, 32] states that it is possible to make the probability of erroneously choosing ρ (type II error) to vanish exponentially fast in the number of copies, with the exponent being the relative entropy $D_1(\rho\|\sigma)$, while the probability of erroneously choosing σ (type I error) goes to zero asymptotically. If the type II error is required to vanish with a suboptimal exponent $r < D_1(\rho\|\sigma)$ (this is called the direct domain) then the type I error can also be made to vanish exponentially fast, with the optimal exponent being the Hoeffding divergence $H_r := \sup_{\alpha\in(0,1)} \frac{\alpha-1}{\alpha}[r - D_\alpha^{(\mathrm{old})}(\rho\|\sigma)]$ [4, 18, 30]. Thus, the $D_\alpha^{(\mathrm{old})}$ with $\alpha \in (0,1)$ quantify the trade-off between the rates of the type I and the type II error probabilities in the direct domain. Based on this trade-off relation, a more direct operational interpretation was obtained in [25] as generalized cutoff rates in the sense of Csiszár [12]. On the other hand, if the type II error is required to vanish with an exponent $r > D_1(\rho\|\sigma)$ (this is called the strong converse domain) then the type I error goes to 1 exponentially fast, with the optimal exponent being the converse Hoeffding divergence $H_r^* := \sup_{\alpha>1} \frac{\alpha-1}{\alpha}[r - D_\alpha^{(\mathrm{new})}(\rho\|\sigma)]$ [26]. Thus, the $D_\alpha^{(\mathrm{new})}$ with $\alpha > 1$ quantify the trade-off between the rates of the type I success probability and the type II error probability in the strong converse region. Based on this, a direct operational interpretation of the $D_\alpha^{(\mathrm{new})}$ as generalized cutoff rates was also given in [26] for $\alpha > 1$.

In the view of the above results, it seems that the old and the new definitions provide the operationally relevant quantum extension of Rényi's divergences in different domains: for $\alpha \in (0,1)$, the operationally relevant definition seems to be the old one, corresponding to the direct domain of hypothesis testing, whereas for $\alpha > 1$, the operationally relevant definition seems to be the new one, corresponding to the strong converse domain of hypothesis testing.

This is the picture at least when one wants to describe the full trade-off curve; most of the time, however, one is interested in one single point of this curve, corresponding to $\alpha = 1$, where the transition from exponentially vanishing error probability to exponentially vanishing success probability happens. It is known that using the "wrong" divergence can be beneficial to obtaining coding theorems at this point. Indeed, the strong converse property for hypothesis testing and classical-quantum channel coding has been proved using $D_\alpha^{(\mathrm{old})}$ for $\alpha > 1$ in [29, 32, 33] ("wrong" divergence with the "right" values of α), while a proof for the direct part of these problems was obtained recently in [8], using $D_2^{(\mathrm{new})}$ ("'wrong" divergence with a "wrong" value of α).

Further examples of coding theorems based on the "wrong" Rényi divergence were given in [27], where it was shown that a certain concavity property of the new Rényi divergences, which the old ones don't have, make them a very convenient tool to prove the direct part of various coding theorems in composite/compound settings. This was demonstrated by giving

short and simple proofs for the direct part of Stein's lemma with composite null-hypothesis and for classical-quantum channel coding with compound channels. Although the optimal rates for these problems have already been known [10, 11, 13, 31], the proofs in [27] are different from the previous ones, and offer considerable simplifications. The general approach is the following:

1. We start with a single-shot coding theorem that gives a trade-off relation between the relevant quantities of the problem in terms of Rényi divergences. For Stein's lemma, this is Audenaert's trace inequality [3], while for channel coding we use the Hayashi-Nagaoka random coding theorem from [17].

2. We then use general properties of the Rényi divergences to decouple the upper bounds from multiple to a single null-hypothesis/channel and to derive the asymptotics.

The main advantage of this approach is that the second step only relies on universal properties of the Rényi divergences and is largely independent of the concrete problem at hand. In particular, the coding theorems for the composite/compound settings can be obtained with the same amount of effort as for a simple null-hypothesis/single channel.

In this paper we present a variant for the proof of Stein's lemma with composite null-hypothesis. While in [27] exponential bounds on the error probabilities were given, here we study the asymptotics of the optimal type II error probability for a given threshold ε on the type I error probability. Building on results from [6] and [27], we derive finite-size bounds on the deviation of the optimal type II error from its asymptotic value. Such bounds are of practical importance, since in real-life scenarios one always works with finitely many copies.

The structure of the paper is as follows. Section 2 is a summary of notations. In Section 3 we review some properties of the quantum Rényi divergences, including two inequalities from [27]: Lemma 4, which gives quantitative bounds between the old and the new definitions of the quantum Rényi divergences, and Corollary 6, which shows that the convexity of the new Rényi divergence in its first argument can be complemented in the form of a weak quasi-concavity inquality. For readers' convenience, we include the proof of these inequalities. In Section 4 we prove the above mentioned finite-size version of Stein's lemma.

2 Notations

For a finite-dimensional Hilbert space $\mathcal{H}$, let $\mathcal{B}(\mathcal{H})_+$ denote the set of all non-zero positive semidefinite operators on $\mathcal{H}$, and let $\mathcal{S}(\mathcal{H}) := \{\rho \in \mathcal{B}(\mathcal{H})_+ \,;\, \operatorname{Tr}\rho = 1\}$ be the set of all density operators (states) on $\mathcal{H}$.

We define the powers of a positive semidefinite operator A only on its support; that is, if $\lambda_1, \dots, \lambda_r$ are the strictly positive eigenvalues of A, with corresponding spectral projections $P_1, \dots, P_r$, then we define $A^\alpha := \sum_{i=1}^r \lambda_i^\alpha P_i$ for all $\alpha \in \mathbb{R}$. In particular, $A^0 = \sum_{i=1}^r P_i$ is the projection onto the support of A, and we use $A^0 \le B^0$ as a shorthand for $\operatorname{supp} A \subseteq \operatorname{supp} B$.

By a *POVM (positive operator-valued measure)* T on a Hilbert space $\mathcal{H}$ we mean a map $T : \mathcal{Y} \to \mathcal{B}(\mathcal{H})$, where $\mathcal{Y}$ is some finite set, $T(y) \ge 0$ for all y, and $\sum_{y\in\mathcal{Y}} T(y) = I$. In particular, a binary POVM is a POVM with $\mathcal{Y} = \{0, 1\}$.

We denote the natural logarithm by log, and use the convention $\log 0 := -\infty$ and $\log +\infty := +\infty$.

3 Rényi divergences

For non-zero positive semidefinite operators ρ, σ, the *Rényi α-divergence* of ρ w.r.t. σ with parameter $\alpha \in (0, +\infty) \setminus \{1\}$ is traditionally defined as [34]

$$D_\alpha^{(\mathrm{old})}(\rho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log \operatorname{Tr}\rho^\alpha\sigma^{1-\alpha} - \frac{1}{\alpha-1}\log\operatorname{Tr}\rho, & \alpha\in(0,1) \text{ or } \rho^0\le\sigma^0, \\ +\infty, & \text{otherwise.} \end{cases}$$

For the mathematical properties of $D_\alpha^{(\mathrm{old})}$, see, e.g. [22, 25, 35]. Recently, a new notion of Rényi divergence has been introduced in [28, 41], defined as

$$D_\alpha^{(\mathrm{new})}(\rho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log \operatorname{Tr}\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha - \frac{1}{\alpha-1}\log\operatorname{Tr}\rho, & \alpha\in(0,1) \text{ or } \rho^0\le\sigma^0, \\ +\infty, & \text{otherwise.} \end{cases}$$

For the mathematical properties of $D_\alpha^{(\mathrm{new})}$, see, e.g. [7, 15, 26, 28, 41].

An easy calculation shows that for fixed ρ and σ, the function $\alpha \mapsto \log\operatorname{Tr}\rho^\alpha\sigma^{1-\alpha}$ is convex, which in turn yields immediately that $\alpha \mapsto D_\alpha^{(\mathrm{old})}(\rho\|\sigma)$ is monotone increasing. Moreover, the limit at $\alpha = 1$ can be easily calculated as

$$D_1(\rho\|\sigma) := \lim_{\alpha\to1} D_\alpha^{(\mathrm{old})}(\rho\|\sigma) = \begin{cases} \frac{1}{\operatorname{Tr}\rho}\operatorname{Tr}\rho(\log\rho-\log\sigma), & \rho^0\le\sigma^0, \\ +\infty, & \text{otherwise,} \end{cases} \tag{1}$$

where the latter expression is *Umegaki's relative entropy* [40]. The same limit relation for $D_\alpha^{(\mathrm{new})}(\rho\|\sigma)$ has been shown in [28, Theorem 5]. The following Lemma, due to [37] and [38], complements the above monotonicity property around $\alpha = 1$, and in the same time gives a quantitative version of (1):

▶ **Lemma 1.** *Let $\rho, \sigma \in \mathcal{B}(\mathcal{H})_+$ be such that $\rho^0 \le \sigma^0$, let $\kappa := \log(1 + \operatorname{Tr}\rho^{3/2}\sigma^{-1/2} + \operatorname{Tr}\rho^{1/2}\sigma^{1/2})$, let $c > 0$, and $\delta := \min\left\{\frac{1}{2}, \frac{c}{2\kappa}\right\}$. Then*

$$D_1(\rho\|\sigma) \ge D_\alpha^{(\mathrm{old})}(\rho\|\sigma) \ge D_1(\rho\|\sigma) - 4(1-\alpha)\kappa^2\cosh c, \qquad 1-\delta<\alpha<1,$$

and the inequalities hold in the converse direction for $1 < \alpha < 1 + \delta$.

▶ **Remark 2.** *Assume that ρ and σ are states. The function $f(\alpha) := \operatorname{Tr}\rho^\alpha\sigma^{1-\alpha}$ is convex in α, and $\rho^0 \le \sigma^0$ implies that $f(1) = 1$. Hence, $\alpha \mapsto (f(\alpha)-1)/(\alpha-1)$ is monotone increasing. Comparing the values at $1/2$ and $3/2$, we see that $\operatorname{Tr}\rho^{3/2}\sigma^{-1/2} + \operatorname{Tr}\rho^{1/2}\sigma^{1/2} \ge 2$, and thus $\kappa > 1$.*

▶ **Remark 3.** *The Rényi entropy of a positive semidefinite operator $\rho \in \mathcal{B}(\mathcal{H})_+$ with parameter $\alpha \in (0, +\infty)$ is defined as*

$$S_\alpha(\rho) := -D_\alpha^{(\mathrm{old})}(\rho\|I) = -D_\alpha^{(\mathrm{new})}(\rho\|I) = \frac{1}{1-\alpha}\log\operatorname{Tr}\rho^\alpha - \frac{1}{1-\alpha}\log\operatorname{Tr}\rho.$$

By the above considerations, $\alpha \mapsto S_\alpha(\rho)$ is monotone decreasing, and comparing its values at α and at 0, we get

$$\operatorname{Tr}\rho^\alpha \le (\operatorname{Tr}\rho^0)^{(1-\alpha)}(\operatorname{Tr}\rho)^\alpha, \qquad \alpha\in(0,1). \tag{2}$$

According to the Araki-Lieb-Thirring inequality [2, 23], for any positive semidefinite operators A, B, $\operatorname{Tr} A^\alpha B^\alpha A^\alpha \le \operatorname{Tr}(ABA)^\alpha$ for $\alpha \in (0,1)$, and the inequality holds in the converse direction for $\alpha > 1$. A converse to the Araki-Lieb-Thirring inequality was given in [5], where it was shown that $\operatorname{Tr}(ABA)^\alpha \le \left(\|B\|^\alpha \operatorname{Tr} A^{2\alpha}\right)^{1-\alpha} (\operatorname{Tr} A^\alpha B^\alpha A^\alpha)^\alpha$ for $\alpha \in (0,1)$, and the inequality holds in the converse direction for $\alpha > 1$. Applying these inequalities to $A := \rho^{\frac{1}{2}}$ and $B := \sigma^{\frac{1-\alpha}{\alpha}}$, we get

$$\operatorname{Tr} \rho^\alpha \sigma^{1-\alpha} \le \operatorname{Tr}\left(\rho^{\frac{1}{2}} \sigma^{\frac{1-\alpha}{\alpha}} \rho^{\frac{1}{2}}\right)^\alpha \le \|\sigma\|^{(1-\alpha)^2} (\operatorname{Tr} \rho^\alpha)^{1-\alpha} \left(\operatorname{Tr} \rho^\alpha \sigma^{1-\alpha}\right)^\alpha \tag{3}$$

for $\alpha \in (0,1)$, and the inequalities hold in the converse direction for $\alpha > 1$. In terms of the Rényi divergences, the above inequalities yield the ones in the following Lemma, the first of which has already been pointed out in [41] and [14].

▶ **Lemma 4.** *Let $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ be states. For any $\alpha \in (0, +\infty)$,*

$$D_\alpha^{(\mathrm{old})}(\rho\|\sigma) \ge D_\alpha^{(\mathrm{new})}(\rho\|\sigma) \ge \alpha D_\alpha^{(\mathrm{old})}(\rho\|\sigma) - |\alpha - 1| \log \dim \mathcal{H}. \tag{4}$$

Proof. The first inequality is immediate from the first inequality in (3). Taking into account (2), and that $\|\sigma\| \le 1$, the second inequality in (3) yields the second inequality in (4) for $\alpha \in (0,1)$. For $\alpha > 1$, we have $\operatorname{Tr}(\rho/\|\rho\|)^\alpha \le \operatorname{Tr}(\rho/\|\rho\|$, and hence we get $\operatorname{Tr}\left(\rho^{\frac{1}{2}} \sigma^{\frac{1-\alpha}{\alpha}} \rho^{\frac{1}{2}}\right)^\alpha \ge \|\sigma\|^{(1-\alpha)^2} \|\rho\|^{-(\alpha-1)^2} \left(\operatorname{Tr} \rho^\alpha \sigma^{1-\alpha}\right)^\alpha$. Using that $\|\rho\| \le 1$ and that $\|\sigma\| \ge 1/\dim \mathcal{H}$, we get the second inequality in (4) for $\alpha > 1$. ◀

For $\rho, \sigma \in \mathcal{B}(\mathcal{H})_+$, let

$$Q_\alpha^{(\mathrm{old})}(\rho\|\sigma) := \operatorname{Tr} \rho^\alpha \sigma^{1-\alpha}, \qquad Q_\alpha^{(\mathrm{new})}(\rho\|\sigma) := \operatorname{Tr}\left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha \tag{5}$$

be the core quantities of the Rényi divergences $D_\alpha^{(\mathrm{old})}$ and $D_\alpha^{(\mathrm{new})}$, respectively. $Q_\alpha^{(\mathrm{old})}$ is jointly concave in (ρ, σ) for $\alpha \in [0,1]$ (see [22, 35]) and jointly convex for $\alpha \in [1,2]$ (see [1, 35]). The general concavity result in [20, Theorem 2.1] implies as a special case that $Q_\alpha^{(\mathrm{new})}(\rho\|\sigma)$ is jointly concave in (ρ, σ) for $\alpha \in [1/2, 1)$. (See also [15] for a different proof of this). In [28, 41], joint convexity of $Q_\alpha^{(\mathrm{new})}$ was shown for $\alpha \in [1,2]$, which was later extended in [15], using a different proof method, to all $\alpha > 1$. These results are equivalent to the monotonicity of the Rényi divergences under completely positive trace-preserving maps, for $\alpha \in [0,2]$ in the case of $D_\alpha^{(\mathrm{old})}$, and for $\alpha \ge 1/2$ in the case of $D_\alpha^{(\mathrm{new})}$.

The next lemma shows that the concavity of $Q_\alpha^{(\mathrm{new})}$ in its first argument can be complemented by a subadditivity inequality for $\alpha \in (0,1)$:

▶ **Lemma 5.** *Let $\rho_1, \ldots, \rho_r \in \mathcal{S}(\mathcal{H})$ be states and $\sigma \in \mathcal{B}(\mathcal{H})_+$, and let $\gamma_1, \ldots, \gamma_r$ be a probability distribution. For every $\alpha \in (0,1)$,*

$$\sum_i \gamma_i Q_\alpha^{(\mathrm{new})}(\rho_i\|\sigma) \le Q_\alpha^{(\mathrm{new})}\left(\sum_i \gamma_i \rho_i \middle\| \sigma\right) \le \sum_i \gamma_i^\alpha Q_\alpha^{(\mathrm{new})}(\rho_i\|\sigma). \tag{6}$$

Proof. The function $x \mapsto x^\alpha$ is operator concave on $[0, +\infty)$ for $\alpha \in (0,1)$ (see Theorems V.1.9 and V.2.5 in [9]), from which the first inequality in (6) follows immediately. To prove the second inequality, we use a special case of the Rotfel'd inequality, for which we provide a proof below. First let $A, B \in \mathcal{B}(\mathcal{H})_+$ be invertible. Then

$$\begin{aligned} \operatorname{Tr}(A+B)^\alpha - \operatorname{Tr} A^\alpha &= \int_0^1 \frac{d}{dt} \operatorname{Tr}(A+tB)^\alpha \, dt = \int_0^1 \alpha \operatorname{Tr} B(A+tB)^{\alpha-1} \, dt \\ &\le \int_0^1 \alpha \operatorname{Tr} B(tB)^{\alpha-1} \, dt = \operatorname{Tr} B^\alpha \int_0^1 \alpha t^{\alpha-1} \, dt = \operatorname{Tr} B^\alpha, \end{aligned} \tag{7}$$

where in the first line we used the identity $(d/dt)\operatorname{Tr} f(A+tB) = \operatorname{Tr} Bf'(A+tB)$, and the inequality follows from the fact that $x \mapsto x^{\alpha-1}$ is operator monotone decreasing on $(0,+\infty)$ for $\alpha \in (0,1)$. By continuity, we can drop the invertibility assumption, and (7) holds for any $A, B \in \mathcal{B}(\mathcal{H})_+$. Obviously, (7) extends to more than two operators, i.e., $\operatorname{Tr}(A_1 + \ldots + A_r)^\alpha \leq \operatorname{Tr} A_1^\alpha + \ldots + \operatorname{Tr} A_r^\alpha$ for any $A, \ldots, A_r \in \mathcal{B}(\mathcal{H})_+$ and $\alpha \in (0,1)$. Choosing now $A_i := \sigma^{\frac{1-\alpha}{2\alpha}} \gamma_i \rho_i \sigma^{\frac{1-\alpha}{2\alpha}}$ yields the second inequality in (6). ◀

▶ **Corollary 6.** *Let $\rho_1, \ldots, \rho_r \in \mathcal{S}(\mathcal{H})$ be states and $\sigma \in \mathcal{B}(\mathcal{H})_+$, and let $\gamma_1, \ldots, \gamma_r$ be a probability distribution. For every $\alpha \in (0,1)$,*

$$\min_i D_\alpha^{(\mathrm{new})}(\rho_i\|\sigma) + \log \min_i \gamma_i \leq D_\alpha^{(\mathrm{new})}\left(\sum_i \gamma_i \rho_i \middle\| \sigma\right) \leq \sum_i \gamma_i D_\alpha^{(\mathrm{new})}(\rho_i\|\sigma).$$

Proof. Immediate from Lemma 5. ◀

4 Stein's lemma with composite null-hypothesis

In the general formulation of binary quantum hypothesis testing, we assume that for every $n \in \mathbb{N}$, a quantum system with Hilbert space $\mathcal{H}_n$ is given, together with two subsets $H_{0,n}$ and $H_{1,n}$ of the state space of $\mathcal{H}_n$, corresponding to the *null-hypothesis* and the *alternative hypothesis*, respectively. Our aim is to guess, based on a binary POVM, which set the true state of the system falls into. Here we consider the i.i.d. case with composite null-hypothesis and simple alternative hypothesis. That is, for every $n \in \mathbb{N}$, $\mathcal{H}_n = \mathcal{H}^{\otimes n}$ for some finite-dimensional Hilbert space $\mathcal{H}$; the null-hypothesis is represented by a set of states $\mathcal{N} \subseteq \mathcal{S}(\mathcal{H})$, and the alternative hypothesis is represented by a single state $\sigma \in \mathcal{S}(\mathcal{H})$. For every $n \in \mathbb{N}$, we have $H_{0,n} = \mathcal{N}^{(\otimes n)} := \{\rho^{\otimes n} : \rho \in \mathcal{N}\}$, and $H_{1,n} = \{\sigma^{\otimes n}\}$.

Given a binary POVM $T_n = (T_n(0), T_n(1))$, with $T_n(0)$ corresponding to accepting the null-hpothesis and $T_n(1)$ to accepting the alternative hypothesis, there are two possible ways of making an erroneous decision: accepting the alternative hypothesis when the null-hypothesis is true, called the type I error, or the other way around, called the type II error. The probabilities of these two errors are given by

$$\alpha_n(T_n) := \sup_{\rho\in\mathcal{N}} \operatorname{Tr} \rho^{\otimes n} T_n(1), \quad \text{(type I)} \qquad \text{and} \qquad \beta_n(T_n) := \operatorname{Tr} \sigma^{\otimes n} T_n(0), \quad \text{(type II)}.$$

Note that in the definition of α_n, we used a worst-case error probability.

In the setting of Stein's lemma, one's aim is to keep the type I error below a threshold ε, and to optimize the type II error under this condition. For any set $\mathcal{M} \subseteq \mathcal{S}(\mathcal{H}^{\otimes n})$ and any $\varepsilon \in (0,1)$, let

$$\beta_\varepsilon(\mathcal{M}\|\sigma^{\otimes n}) := \inf\left\{\operatorname{Tr} \sigma^{\otimes n} T_n(0) : \sup_{\omega\in\mathcal{M}} \operatorname{Tr} \omega T_n(1) \leq \varepsilon\right\},$$

where the infimum is taken over all binary POVM T_n on $\mathcal{H}^{\otimes n}$. When $\mathcal{M}$ consists of one single element ω, we simply write $\beta_\varepsilon(\omega\|\sigma^{\otimes n})$. The quantum Stein's lemma states that

$$\lim_{n\to+\infty} \frac{1}{n} \log \beta_\varepsilon\left(\mathcal{N}^{(\otimes n)}\|\sigma^{\otimes n}\right) = -D_1(\mathcal{N}\|\sigma) := -\inf_{\rho\in\mathcal{N}} D_1(\rho\|\sigma). \tag{8}$$

This has been shown first in [19, 33] for the case where $\mathcal{N}$ consists of one single element ρ. Theorem 2 in [16] uses group representation techniques to give an approximation of the relative entropy in terms of post-measurement relative entropies, which, when combined with

Stein's lemma for probability distributions, yields (8) for finite $\mathcal{N}$. A direct proof for the case of infinite $\mathcal{N}$, also based on group representation theory, has recently been given in [31]. A version of Stein's lemma with infinite $\mathcal{N}$ has been previously proved in [10], however, with a weaker error criterion.

Here we give a different proof of the quantum Stein's lemma with possibly infinite composite null-hypothesis. Our proof is based on the results of [6], where bounds on β_ε were obtained in terms of Rényi divergences, and general properties of the Rényi divergences from Section 3. Moreover, we give a refined version of (8) in Theorem 9 by providing finite-size corrections to the deviation of $\frac{1}{n}\log\beta_\varepsilon\left(\mathcal{N}^{(\otimes n)}\|\sigma^{\otimes n}\right)$ from its asymptotic value $-D_1(\mathcal{N}\|\sigma)$ for every $n\in\mathbb{N}$.

We will need the following results from [6]:

▶ **Lemma 7.** *Let $\rho,\sigma\in\mathcal{S}(\mathcal{H})$. For every $\varepsilon\in(0,1)$ and every $\alpha\in(0,1)$,*

$$\log\beta_\varepsilon(\rho\|\sigma)\leq -D_\alpha^{(\mathrm{old})}(\rho\|\sigma)+\frac{\alpha}{1-\alpha}\log\varepsilon^{-1}-\frac{h_2(\alpha)}{1-\alpha}, \tag{9}$$

where $h_2(\alpha):=-\alpha\log\alpha-(1-\alpha)\log(1-\alpha)$ is the binary entropy function. Moreover, for every $n\in\mathbb{N}$,

$$\frac{1}{n}\log\beta_\varepsilon\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right)\geq -D_1(\rho\|\sigma)-\frac{1}{\sqrt{n}}4\sqrt{2}\kappa\log(1-\varepsilon)^{-1}, \tag{10}$$

where κ is given in Lemma 1.

Proof. The upper bound (9) is due to [6, Proposition 3.2], while the lower bound in (10) is formula (19) in [6, Theorem 3.3]. ◀

When $\mathcal{N}$ is infinite, we will need the following approximation lemma, which is a special case of [24, Lemma 2.6]:

▶ **Lemma 8.** *For every $\delta>0$, let $\mathcal{N}_\delta\subset\mathcal{N}$ be a set of minimal cardinality such that $\sup_{\rho\in\mathcal{N}}\inf_{\rho'\in\mathcal{N}_\delta}\|\rho-\rho'\|_1\leq\delta$. Then $|\mathcal{N}_\delta|\leq\min\{|\mathcal{N}|,(1+2\delta^{-1})^D\}$, where $D=(\dim\mathcal{H}+1)(\dim\mathcal{H})/2$, and*

$$\sup_{\rho\in\mathcal{N}}\inf_{\rho'\in\mathcal{N}_\delta}\left\|\rho^{\otimes n}-(\rho')^{\otimes n}\right\|_1\leq n\sup_{\rho\in\mathcal{N}}\inf_{\rho'\in\mathcal{N}_\delta}\|\rho-\rho'\|_1\leq n\delta,\qquad n\in\mathbb{N}. \tag{11}$$

Now we are ready to prove our main result:

▶ **Theorem 9.** *Let $\varepsilon\in(0,1)$, and for every $n\in\mathbb{N}$, let $0\leq\delta_n\leq\varepsilon/(2n)$. Then*

$$\begin{aligned}\frac{1}{n}\log\beta_\varepsilon\left(\mathcal{N}^{(\otimes n)}\|\sigma^{\otimes n}\right)\leq&-D_1(\mathcal{N}\|\sigma)\\&+\sqrt{\frac{\log\left(2|\mathcal{N}_{\delta_n}|\varepsilon^{-1}\right)}{n}}\cdot 2\left[8\kappa_{\max}^2+\log\dim\mathcal{H}+D_1(\mathcal{N}\|\sigma)\right]^{\frac{1}{2}}\\&+\frac{\log\left(2|\mathcal{N}_{\delta_n}|\varepsilon^{-1}\right)}{n}\cdot 4\kappa_{\max},\end{aligned} \tag{12}$$

$$\frac{1}{n}\log\beta_\varepsilon\left(\mathcal{N}^{(\otimes n)}\|\sigma^{\otimes n}\right)\geq -D_1(\mathcal{N}\|\sigma)-\frac{1}{\sqrt{n}}4\sqrt{2}\log(1-\varepsilon)^{-1}\kappa_{\max}, \tag{13}$$

where $\kappa_{\max}:=\sup_{\rho\in\mathcal{N}}\{\log(1+\operatorname{Tr}\rho^{3/2}\sigma^{-1/2}+\operatorname{Tr}\rho^{1/2}\sigma^{1/2})\}\leq\log(2+\operatorname{Tr}\sigma^{-1/2})<+\infty$.

In (12), the slowest decaying term after $-D_1(\mathcal{N}\|\sigma)$ is of the order $1/\sqrt{n}$ when $\mathcal{N}$ is finite, and when $\mathcal{N}$ is infinite, it can be chosen to be of the order $\sqrt{\frac{\log n}{n}}$.

Proof. The lower bound in (13) is immediate from (10), and hence we only have to prove (12). We have

$$\begin{aligned}\log\beta_\varepsilon\left(\mathcal{N}^{(\otimes n)}\|\sigma^{\otimes n}\right) &\le \log\beta_{\varepsilon-n\delta_n}\left(\mathcal{N}_{\delta_n}^{(\otimes n)}\|\sigma^{\otimes n}\right) \le \log\beta_{\frac{\varepsilon-n\delta_n}{|\mathcal{N}_{\delta_n}|}}\left(\sum_{\rho\in\mathcal{N}_{\delta_n}}\frac{1}{|\mathcal{N}_{\delta_n}|}\rho^{\otimes n}\middle\|\sigma^{\otimes n}\right)\\ &\le -D_\alpha^{(\mathrm{old})}\left(\sum_{\rho\in\mathcal{N}_{\delta_n}}\frac{1}{|\mathcal{N}_{\delta_n}|}\rho^{\otimes n}\middle\|\sigma^{\otimes n}\right)+\frac{\alpha}{1-\alpha}\log\frac{|\mathcal{N}_{\delta_n}|}{\varepsilon-n\delta_n}\\ &\le -D_\alpha^{(\mathrm{new})}\left(\sum_{\rho\in\mathcal{N}_{\delta_n}}\frac{1}{|\mathcal{N}_{\delta_n}|}\rho^{\otimes n}\middle\|\sigma^{\otimes n}\right)+\frac{\alpha}{1-\alpha}\log\frac{|\mathcal{N}_{\delta_n}|}{\varepsilon-n\delta_n},\end{aligned}$$

where the first inequality is due to (11), the second inequality is obvious, the third one follows from (9), and the last one is due to Lemma 4. Note that $\varepsilon-n\delta_n\ge\varepsilon/2$ by assumption. Using Corollary 6, we can continue the above upper bound as

$$\begin{aligned}&\log\beta_\varepsilon\left(\mathcal{N}^{(\otimes n)}\|\sigma^{\otimes n}\right)\\ &\le -\min_{\rho\in\mathcal{N}_{\delta_n}} D_\alpha^{(\mathrm{new})}\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right)+\log|\mathcal{N}_{\delta_n}|++\frac{\alpha}{1-\alpha}\log|\mathcal{N}_{\delta_n}|+\frac{\alpha}{1-\alpha}\log\frac{2}{\varepsilon}\\ &\le -n\inf_{\rho\in\mathcal{N}} D_\alpha^{(\mathrm{new})}\left(\rho\|\sigma\right)+\frac{1}{1-\alpha}\log|\mathcal{N}_{\delta_n}|+\frac{1}{1-\alpha}\log\frac{2}{\varepsilon},\end{aligned}$$

where in the last line we used the additivity property $D_\alpha^{(\mathrm{new})}\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right)=nD_\alpha^{(\mathrm{new})}\left(\rho\|\sigma\right)$.

By Lemmas 4 and 1, for every $\alpha\in(1/2,1)$ such that $\alpha>1-\frac{c}{2\kappa_{\max}}$,

$$\begin{aligned}\inf_{\rho\in\mathcal{N}} D_\alpha^{(\mathrm{new})}\left(\rho\|\sigma\right) &\ge \alpha\inf_{\rho\in\mathcal{N}} D_\alpha^{(\mathrm{old})}\left(\rho\|\sigma\right)-(1-\alpha)\log\dim\mathcal{H}\\ &\ge \alpha\inf_{\rho\in\mathcal{N}} D_1\left(\rho\|\sigma\right)-4\alpha(1-\alpha)\kappa_{\max}^2\cosh c-(1-\alpha)\log\dim\mathcal{H},\end{aligned}$$

where c is an arbitrary positive constant. Now choose $\alpha:=1-a/\sqrt{n}$. Then

$$\begin{aligned}\frac{1}{n}\log\beta_\varepsilon\left(\mathcal{N}^{(\otimes n)}\|\sigma^{\otimes n}\right) \le &-\left(1-\frac{a}{\sqrt{n}}\right)D_1(\mathcal{N}\|\sigma)+\frac{a}{\sqrt{n}}\left(4\kappa_{\max}^2\cosh c+\log\dim\mathcal{H}\right)\\ &+\frac{1}{a\sqrt{n}}\left(\log|\mathcal{N}_{\delta_n}|+\log\frac{2}{\varepsilon}\right).\end{aligned}$$

Optimizing over a yields

$$\begin{aligned}&\frac{1}{n}\log\beta_\varepsilon\left(\mathcal{N}^{(\otimes n)}\|\sigma^{\otimes n}\right)\\ &\quad\le -D_1(\mathcal{N}\|\sigma)+\frac{2}{\sqrt{n}}\left[4\kappa_{\max}^2\cosh c+\log\dim\mathcal{H}+D_1(\mathcal{N}\|\sigma)\right]^{\frac{1}{2}}\cdot\left[\log(2|\mathcal{N}_{\delta_n}|\varepsilon^{-1})\right]^{\frac{1}{2}}.\end{aligned}\tag{14}$$

The optimum is reached at

$$a^*=\left[\log(2|\mathcal{N}_{\delta_n}|\varepsilon^{-1})\right]^{\frac{1}{2}}\cdot\left[4\kappa_{\max}^2\cosh c+\log\dim\mathcal{H}+D_1(\mathcal{N}\|\sigma)\right]^{-\frac{1}{2}},$$

and we need $a^*/\sqrt{n}\le 1/2$ and $a^*/\sqrt{n}\le c/(2\kappa_{\max})$, which is satisfied if

$$\kappa_{\max}^2\cosh c\ge\frac{1}{n}\log(2|\mathcal{N}_{\delta_n}|\varepsilon^{-1})\qquad\text{and}\qquad c^2\cosh c\ge\frac{1}{n}\log(2|\mathcal{N}_{\delta_n}|\varepsilon^{-1}).$$

Let us choose $c > 0$ such that $\cosh c = 2 + \frac{1}{n}\log(2|\mathcal{N}_{\delta_n}|\varepsilon^{-1})$. By Remark 2, $\kappa_{\max} > 1$, and hence the first inequality is satisfied. Moreover, with this choice $c > 1$, and thus the second inequality is satisfied as well.

Substituting this choice of c into (14), and using the subadditivity of the square root, we get (12).

When $\mathcal{N}$ is finite, we can choose $\delta_n = 0$, and hence $\mathcal{N}_{\delta_n} = \mathcal{N}$, for every n. This shows that the second term in (12) is of the order $1/\sqrt{n}$, while the third term is of the order $1/n$. When $\mathcal{N}$ is infinite, we can choose $\delta_n = \varepsilon/(2n^2)$, whence the order of the second term in (12) is $\sqrt{\frac{\log n}{n}}$, and the order of the third term is $\frac{\log n}{n}$. ◀

▶ **Remark 10.** *In the case of a simple null-hypothesis $\mathcal{N} = \{\rho\}$, the limit*

$$\lim_{n\to+\infty} \sqrt{n}\left(\frac{1}{n}\log\beta_\varepsilon(\mathcal{N}^{(\otimes n)}\|\sigma^{\otimes n}) + D_1(\mathcal{N}\|\sigma)\right), \tag{15}$$

called the second-order asymptotics, has been determined in [21, 39]. Their results show that the finite-size bounds of [6] are not asymptotically optimal, and hence the same holds for the bounds in Theorem 9. The merit of these latter results, on the other hand, is that the correction terms are easily computable, and the bounds are valid for any finite n. To the best of our knowledge, the value of the limit (15) *has not yet been determined when $|\mathcal{N}| > 1$, and our bounds in Theorem 9 give bounds on the second-order asymptotics in this case.*

Acknowledgements. The author is grateful to Professor Fumio Hiai and Nilanjana Datta for discussions.

References

1 T. Ando. *Concavity of certain maps and positive definite matrices and applications to Hadamard products*. Linear Algebra Appl. **26**, 203–241 1979

2 H. Araki. *On an inequality of Lieb and Thirring*. Letters in Mathematical Physics, Volume 19, Issue 2, pp. 167–170, 1990

3 K.M.R. Audenaert, J. Calsamiglia, Ll. Masanes, R. Munoz-Tapia, A. Acin, E. Bagan, F. Verstraete. *Discriminating states: the quantum Chernoff bound*. Phys. Rev. Lett. **98** 160501, 2007

4 K.M.R. Audenaert, M. Nussbaum, A. Szkoła, F. Verstraete. *Asymptotic error rates in quantum hypothesis testing*. Commun. Math. Phys. **279**, 251–283, 2008

5 K.M.R. Audenaert. *On the Araki-Lieb-Thirring inequality*. Int. J. of Information and Systems Sciences **4**, pp. 78–83, 2008)

6 Koenraad M.R. Audenaert, Milan Mosonyi, Frank Verstraete. *Quantum state discrimination bounds for finite sample size*. J. Math. Phys. **53**, 122205, 2012

7 Salman Beigi. *Quantum Rényi divergence satisfies data processing inequality*. J. Math. Phys., **54**, 122202, 2013

8 Salman Beigi, Amin Gohari. *Quantum Achievability Proof via Collision Relative Entropy*. arXiv:1312.3822, 2013

9 R. Bhatia. *Matrix Analysis*. Graduate Texts in Mathematics **169**, Springer, 1997

10 I. Bjelakovic, J.-D. Deuschel, T. Krüger, R. Seiler, R. Siegmund-Schultze, A. Szkoła. *A quantum version of Sanov's theorem*. Commun. Math. Phys. **260**, pp. 659–671, 2005

11 I. Bjelakovic, H. Boche. *Classical capacities of compound and averaged quantum channels*. IEEE Trans. Inform. Theory **55**, 3360–3374, 2009

12 I. Csiszár. *Generalized cutoff rates and Rényi's information measures*. IEEE Trans. Inf. Theory **41**, 26–34, 1995

13 N. Datta, T.C. Dorlas. *The Coding Theorem for a Class of Quantum Channels with Long-Term Memory*. Journal of Physics A: Mathematical and Theoretical, vol. 40, 8147, 2007

14 Nilanjana Datta and Felix Leditzky. *A limit of the quantum Rényi divergence*. J. Phys. A: Math. Theor. **47** 045304, 2014

15 Rupert L. Frank and Elliott H. Lieb. *Monotonicity of a relative Rényi entropy*. J. Math. Phys. 54 , 122201, 2013

16 Masahito Hayashi. *Asymptotics of quantum relative entropy from a representation theoretical viewpoint*. J. Phys. A: Math. Gen. **34** 3413, (2001)

17 M. Hayashi, H. Nagaoka. *General Formulas for Capacity of Classical-Quantum Channels*. IEEE Trans. Inf. Theory **49**, 2003

18 M. Hayashi. *Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding*. Phys. Rev. A **76**, 062301, 2007

19 F. Hiai, D. Petz. *The proper formula for relative entropy and its asymptotics in quantum probability*. Comm. Math. Phys. **143**, 99–114, 1991

20 F. Hiai. *Concavity of certain matrix trace and norm functions*. Linear Algebra and Appl. **439**, 1568–1589, 2013

21 Ke Li. *Second-order asymptotics for quantum hypothesis testing*. Annals of Statistics, Vol. 42, No. 1, pp. 171–189, 2014

22 E.H. Lieb. *Convex trace functions and the Wigner-Yanase-Dyson conjecture*. Adv. Math. **11**, 267–288, 1973

23 E.H. Lieb, W. Thirring. *Studies in mathematical physics*. pp. 269–297. Princeton University Press, Princeton, 1976

24 Vitali D. Milman, Gideon Schechtman. *Asymptotic Theory of Finite Dimensional Normed Spaces*. Lecture Notes in Mathematics, Springer-Verlag Berlin Heidelberg, 1986

25 M. Mosonyi, F. Hiai. *On the quantum Rényi relative entropies and related capacity formulas*. IEEE Trans. Inf. Theory, **57**, 2474–2487, 2011

26 M. Mosonyi, T. Ogawa. *Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies*. arXiv:1308.3228, 2013

27 M. Mosonyi. *Inequalities for the quantum Rényi divergences with applications to compound coding problems*. arXiv:1310.7525; submitted to IEEE Transactions on Information Theory

28 M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, M. Tomamichel. *On quantum Renyi entropies: a new definition and some properties*. J. Math. Phys. **54**, 122203, 2013

29 H. Nagaoka. *Strong converse theorems in quantum information theory*. in the book "Asymptotic Theory of Quantum Statistical Inference" edited by M. Hayashi, World Scientific, 2005

30 H. Nagaoka. *The converse part of the theorem for quantum Hoeffding bound*. quant-ph/0611289, 2006

31 J. Nötzel. *Hypothesis testing on invariant subspaces of the symmetric group, part I - quantum Sanov's theorem and arbitrarily varying sources*. arXiv:1310.5553, 2013

32 T. Ogawa, H. Nagaoka. *Strong converse to the quantum channel coding theorem*. IEEE Transactions on Information Theory, vol. 45, no. 7, pp. 2486-2489, 1999

33 T. Ogawa, H. Nagaoka. *Strong converse and Stein's lemma in quantum hypothesis testing*. IEEE Trans. Inform. Theory **47**, 2428–2433, 2000

34 M. Ohya, D. Petz. *Quantum Entropy and its Use*. Springer, 1993

35 D. Petz. *Quasi-entropies for finite quantum systems*. Rep. Math. Phys. **23**, 57–65, 1986

36 A. Rényi. *On measures of entropy and information*. Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I, pp. 547–561, Univ. California Press, Berkeley, California, 1961

37 M. Tomamichel, R. Colbeck, R. Renner. *A fully quantum asymptotic equipartition property*. IEEE Trans. Inform. Theory **55**, 5840–5847, 2009

38 M. Tomamichel. *A framework for non-asymptotic quantum information theory*. PhD thesis, ETH Zürich, 2012

39 M. Tomamichel, M. Hayashi. *A Hierarchy of Information Quantities for Finite Block Length Analysis of Quantum Tasks*. IEEE Transactions on Information Theory **59**, pp. 7693–7710, 2013

40 H. Umegaki. *Conditional expectation in an operator algebra*. Kodai Math. Sem. Rep. **14**, 59–85, 1962

41 Mark M. Wilde, Andreas Winter, Dong Yang. *Strong converse for the classical capacity of entanglement-breaking and Hadamard channels*. arXiv:1306.1586, 2013