The Annals of Statistics
2014, Vol. 42, No. 1, 171–189
DOI: 10.1214/13-AOS1185

SECOND-ORDER ASYMPTOTICS FOR QUANTUM HYPOTHESIS TESTING

BY KE LI[1]

IBM TJ Watson Research Center, Massachusetts Institute of Technology and National University of Singapore

In the asymptotic theory of quantum hypothesis testing, the minimal error probability of the first kind jumps sharply from zero to one when the error exponent of the second kind passes by the point of the relative entropy of the two states in an increasing way. This is well known as the direct part and strong converse of quantum Stein's lemma.

Here we look into the behavior of this sudden change and have make it clear how the error of first kind grows smoothly according to a lower order of the error exponent of the second kind, and hence we obtain the second-order asymptotics for quantum hypothesis testing. This actually implies quantum Stein's lemma as a special case. Meanwhile, our analysis also yields tight bounds for the case of finite sample size. These results have potential applications in quantum information theory.

Our method is elementary, based on basic linear algebra and probability theory. It deals with the achievability part and the optimality part in a unified fashion.

1. Introduction.

We are interested in the asymptotic theory of hypothesis testing with two hypotheses. Suppose there are many identical physical systems, each independently being in some random states, subject to the same statistical description. Here the statistical description is probability distribution in classical world and quantum state which is positive semi-definite matrix with trace 1 in quantum mechanics. However, the statistical description is not fixed: it has two possibilities, say, either ρ (the null hypothesis) or σ (the alternative hypothesis). Thus the task is to identify which statistical description is the true one, based on the instances of the physical systems.

It is the central problem in asymptotic hypothesis testing to characterize the behavior of errors. An intuitive understanding is that the probabilities of mistaking one hypothesis for the other can be made arbitrarily small when the sample size n is big enough, except for the trivial case that ρ and σ are the same. However,

Received January 2013; revised July 2013.

[1] Supported by NSF Grant CCF-1111382. The Centre for Quantum Technologies is funded by the Singapore Ministry of Education and the National Research Foundation as part of the Research Centres of Excellence program.

MSC2010 subject classifications. 62P35, 62G10.

Key words and phrases. Quantum hypothesis testing, quantum Stein's lemma, second-order asymptotics, finite sample size.

assuming exponential decay, we want to optimize the rate exponent with which the error of concern, under certain reasonable preconditions, converges to zero. In the classical setting, this problem has been well understood, featured with a list of famous results [7, 9–11, 13, 20], including the celebrated Stein's lemma, Chernoff distance and Hoeffding bound. These results are all obtainable using the likelihood ratio tests.

In contrast to its classical counterpart, the problem of quantum hypothesis testing becomes very difficult due to the noncommutativity of the two quantum states ρ and σ, and the more complicated mechanics for observing the underlying physical systems, that is, quantum measurement. Although the quantum generalization of the likelihood ratio test was obtained in the 1970s [18, 21], its structure is not clear in the aymptotic limit. Yet, substantial achievements have been made since.

In 1991, Hiai and Petz established the quantum Stein's lemma, providing rigorous operational interpretation for the quantum relative entropy, or quantum Kullback–Leibler divergence [19]. Then its optimality part was strengthened by Ogawa and Nagaoka, with a strong converse theorem [28]. More recently, quantum Chernoff distance, the optimal rate exponent under which the average error tends to 0 in the setting of symmetric hypothesis testing, has been identified in two seminal papers. The achievability part was due to Audenaert et al. [1], and the optimality part was by Nussbaum and Szkoła [26]. The methods invented in these two papers were subsequently used to derive the quantum Hoeffding bound [3, 15, 24].

The quantum Stein's lemma characterizes the optimal error exponent in asymmetric hypothesis testing. Besides the breakthroughs mentioned above, some other important progresses in this regime can be found in [3–6, 8, 14, 25]. To state this result, we define two types of errors. Type I error (or the error of the first kind) is the probability that we incorrectly accept the alternative hypothesis $\sigma^{\otimes n}$ while it is actually the null hypothesis $\rho^{\otimes n}$, and type II error (or the error of the second kind) is the probability of the opposite situation. In an asymmetric setting, we want to minimize the type II error while only simply requiring that the type I error converges to 0. Let $\operatorname{supp}(X)$ be the support of the operator X. The quantum Stein's lemma states that the maximal exponent of type II error is the quantum relative entropy [19], given by

$$D(\rho\|\sigma) = \begin{cases} \operatorname{Tr}\big(\rho(\log\rho - \log\sigma)\big), & \text{if } \operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma), \\ +\infty, & \text{otherwise.} \end{cases}$$

It also asserts that if the type II error goes to 0 with an exponent larger than $D(\rho\|\sigma)$, then the type I error inevitably converges to 1 [28].

However, the drawback of the quantum Stein's lemma is that it characterizes the asymptotic behavior of errors in a relatively coarse-grained way. To be precise, it considers only the linear term of the type II error exponent, which is of the order n (we call it the first order). As a result, the optimal type I error jumps sharply from 0 to 1 when the rate exponent of type II error—quantified by its first order—passes by the relative entropy $D(\rho\|\sigma)$ from the smaller side to the larger side.

In this paper, we prove the second-order asymptotic theorem, and thus fundamentally refine the quantum Stein's lemma. Specifically, we track the exponent of the type II error in depth, to the order $\sqrt{n}$ (we call it the second order), and clarify how the type I error varies smoothly as a function of this second-order exponent. A variance-like quantity, defined as

$$V(\rho\|\sigma) := \operatorname{Tr}\rho(\log\rho - \log\sigma)^2 - \bigl(D(\rho\|\sigma)\bigr)^2, \tag{1}$$

will play an important role, and we name it the *quantum relative variance* of ρ and σ. Write the second-order rate exponent of the type II error as E_2. Then our result shows that, asymptotically, the minimal type I error is given by $\Phi(E_2/\sqrt{V(\rho\|\sigma)})$, which grows smoothly from 0 to 1 when E_2 increases from $-\infty$ to $+\infty$. Here Φ is the cumulative distribution function of standard normal distribution, and its appearance in our result comes from the use of the central limit theorem in the proof.

We also obtain very tight bounds for the case of finite sample size n. Supposing that the type I error is no larger than a constant ε, we minimize the type II error and consider its negative logarithm. Then we derive upper and lower bounds for this quantity, based on the method for proving our second-order asymptotic theorem. This enables us to establish that this quantity can be written as

$$nD(\rho\|\sigma) + \sqrt{n}\sqrt{V(\rho\|\sigma)}\Phi^{-1}(\varepsilon) + O(\log n).$$

The first two terms coincide with the results of the quantum Stein's lemma and our second-order asymptotic theorem, respectively. Furthermore, the next leading term (this is the term of the third order), included in $O(\log n)$ of the above formula, lies between a constant and $2\log n$.

Our results have potential applications in quantum information theory. There is a deep connection between hypothesis testing and other topics in information theory (e.g., channel capacity), both in the classical regime [12, 31] and in the quantum regime [17]. Recently, this connection has been generalized to the one-shot scenario as well [23, 32]. Indeed, such a connection is very helpful in the derivation of the second-order coding rate and finite blocklength analysis in classical channel coding [16, 29]. Our results make it possible to investigate the second-order and finite blocklength analysis for classical information transmission over quantum channels.

We point out that the results presented here are independently and concurrently obtained by Tomamichel and Hayashi [30], using a different method. In [30], such analysis is conducted in the context of one-shot entropies and has been applied to the tasks of data compression with quantum side information and randomness extraction against quantum side information. The bounds for finite sample size in these two works are slightly different; see Section 4 for details.

The remainder of this paper is organized as follows. In Section 2, we present our main result of second-order asymptotics. Then we prove it in Section 3. In Section 4, we treat the case of finite sample size. In Section 5, we note a few remarks. Finally, we give the proofs to technical lemmas in the Appendix.

2. Second-order asymptotics. Every quantum system is associated with a complex Hilbert space. The state of the quantum system is described by a density matrix ϖ, which is a nonnegative definite matrix in the Hilbert space and satisfies the normalization condition $\mathrm{Tr}\,\varpi = 1$. To detect the quantum system, we have to do quantum measurement, which, in the most general form, is formulated as positive operator-valued measurement (POVM) $\mathcal{M} = \{M_i\}_i$, with $0 \le M_i \le \mathbb{1}$ and $\sum_i M_i = \mathbb{1}$. Then the measurement outcome i is obtained with probability $\mathrm{Tr}(\varpi M_i)$.

We consider a large number n of identical quantum systems, each of which has finite level and is associated with the Hilbert space $\mathcal{H}$ of finite dimension $|\mathcal{H}|$. Given that the quantum systems are either of the state $\rho^{\otimes n}$ (the null hypothesis) or of the state $\sigma^{\otimes n}$ (the alternative hypothesis), we want to identify which state the systems belong to. Without loss of generality, this can be done by applying a two-outcome POVM $(A_n, \mathbb{1} - A_n)$, with $0 \le A_n \le \mathbb{1}$, on the joint Hilbert space $\mathcal{H}^{\otimes n}$ of the quantum systems. If we obtain the outcome associated to A_n, then we conclude that the state is $\rho^{\otimes n}$. Similarly, the outcome associated to $(\mathbb{1} - A_n)$ corresponds to the state $\sigma^{\otimes n}$. The error probabilities of the first kind and the second kind are, respectively, given by $\alpha_n(A_n) = \mathrm{Tr}(\rho^{\otimes n}(\mathbb{1} - A_n))$ and $\beta_n(A_n) = \mathrm{Tr}(\sigma^{\otimes n} A_n)$.

The quantum Stein's lemma shows that the relative entropy $D(\rho\|\sigma)$ is a critical jump point in the asymptotics of asymmetric hypothesis testing. Explicitly, it is stated in two parts as follows:

- Direct part [19]: for arbitrary $R \le D(\rho\|\sigma)$, there exist tests $\{(A_n, \mathbb{1} - A_n)\}_n$ satisfying

$$\liminf_{n\to\infty} \frac{-1}{n} \log \beta_n(A_n) \ge R \quad \text{and} \quad \lim_{n\to\infty} \alpha_n(A_n) = 0.$$

- Strong converse [28]: if a sequence of tests $\{(A_n, \mathbb{1} - A_n)\}_n$ is such that

$$\liminf_{n\to\infty} \frac{-1}{n} \log \beta_n(A_n) > D(\rho\|\sigma),$$

then $\lim_{n\to\infty} \alpha_n(A_n) = 1$.

Instead of the rate exponent $\frac{-1}{n} \log \beta_n(A_n)$ considered in the quantum Stein's lemma, we are concerned with a smaller order of the type II error exponent, that is, $\frac{1}{\sqrt{n}}(-\log \beta_n(A_n) - nD(\rho\|\sigma))$. Then we think about the optimal tradeoff between the asymptotic limit of this quantity and the type I error $\alpha_n(A_n)$. In an equivalent way, we define the error-dependency functions as follows and present our result subsequently.

DEFINITION 1. Let $E_1, E_2 \in \mathbb{R}$, and $f(n)$ be a fixed function of some order other than n or $\sqrt{n}$, which is to be specified when necessary. We define a sequence of functions $\{\alpha_n(E_1, E_2|f) : n \in \mathbb{N}\}$, which reflects the dependency of the minimal error probability of the first kind on the error exponent of the second kind, up to the order n and $\sqrt{n}$, as

$$\alpha_n(E_1, E_2|f) := \min_{A_n}\{\alpha_n(A_n)|\beta_n(A_n) \le \exp(-(E_1 n + E_2\sqrt{n} + f(n)))\}.$$

If $\operatorname{supp}(\rho) \nsubseteq \operatorname{supp}(\sigma)$, we have $D(\rho\|\sigma) = +\infty$. Asymptotically, the optimal error probability of the first kind is always 0, while the error exponent of the second kind can be arbitrarily large. In such a case, the second-order asymptotics makes no sense. So, in this paper, we suppose $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$, and without loss of generality, we further suppose σ is of full rank.

Our main result is the following theorem.

THEOREM 2. *Let $\{\alpha_n(E_1, E_2|f)\}_n$, the sequence of error-dependency functions, be as defined in Definition 1, and let $V(\rho\|\sigma)$, the quantum relative variance of ρ and σ, be as defined by equation (1). We have*

$$\lim_{n\to\infty} \alpha_n(E_1, E_2|f) = \begin{cases} 0, & \text{if } E_1 < D(\rho\|\sigma),\ f \in o(n), \\ \Phi\left(\dfrac{E_2}{\sqrt{V(\rho\|\sigma)}}\right), & \text{if } E_1 = D(\rho\|\sigma),\ f \in o(\sqrt{n}), \\ 1, & \text{if } E_1 > D(\rho\|\sigma),\ f \in o(n), \end{cases} \tag{2}$$

where $\Phi(x)$ is the cumulative distribution function of the standard normal distribution, that is, $\Phi(x) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,\mathrm{d}t$.

The second case of equation (2) is our second-order asymptotics. In fact, it implies the first and third cases, which are nothing else but the direct part and strong converse of quantum Stein's lemma, respectively. We include them here such that one easily gets the full information at first sight. To see this, we take the first case, for example. It is obvious from Definition 1 that, for arbitrary $E_1 < D(\rho\|\sigma)$, $E_2 \in \mathbb{R}$, $E_2' \in \mathbb{R}$, $f(n) \in o(n)$ and $f'(n) \in o(\sqrt{n})$,

$$\lim_{n\to\infty} \alpha_n(E_1, E_2|f) \le \lim_{n\to\infty} \alpha_n(D(\rho\|\sigma), E_2'|f'). \tag{3}$$

Assuming the second case of equation (2), the right-hand side of equation (3) equals $\Phi(\frac{E_2'}{\sqrt{V(\rho\|\sigma)}})$. Now letting $E_2' \to -\infty$, the first case of equation (2) follows immediately since $\alpha_n(E_1, E_2|f)$ is always nonnegative.

We divide Theorem 2 (precisely, its second case) into the achievability part and optimality part, and equivalently reformulate it below. This reformulation form corresponds to the structure of the proof in the next section.

REFORMULATION OF THEOREM 2. *For quantum hypothesis testing with the null hypothesis $\rho^{\otimes n}$ and the alternative hypothesis $\sigma^{\otimes n}$ and the error probabilities of the first and second kinds denoted as $\alpha_n(A_n)$ and $\beta_n(A_n)$, respectively, we have:*

Achievability: *for any* $E_2 \in \mathbb{R}$ *and* $f(n) \in o(\sqrt{n})$, *there exists a sequence of measurements* $\{(A_n, \mathbb{1} - A_n)\}_n$, *such that*

$$\beta_n(A_n) \leq \exp\{-(nD(\rho\|\sigma) + E_2\sqrt{n} + f(n))\}, \tag{4}$$

$$\limsup_{n\to\infty} \alpha_n(A_n) \leq \Phi\left(\frac{E_2}{\sqrt{V(\rho\|\sigma)}}\right). \tag{5}$$

Optimality: *if there is a sequence of measurements* $\{(A_n, \mathbb{1} - A_n)\}_n$ *such that*

$$\beta_n(A_n) \leq \exp\{-(nD(\rho\|\sigma) + E_2\sqrt{n} + f(n))\} \tag{6}$$

holds for given $E_2 \in \mathbb{R}$ *and* $f(n) \in o(\sqrt{n})$, *then*

$$\liminf_{n\to\infty} \alpha_n(A_n) \geq \Phi\left(\frac{E_2}{\sqrt{V(\rho\|\sigma)}}\right). \tag{7}$$

The equivalence is obvious. By the definition of $\alpha_n(E_1, E_2|f)$, it is straightforward to see that the achievability part of the above reformulation is equivalent to

$$\limsup_{n\to\infty} \alpha_n(D(\rho\|\sigma), E_2|f) \leq \Phi\left(\frac{E_2}{\sqrt{V(\rho\|\sigma)}}\right) \qquad \forall f(n) \in o(\sqrt{n}) \tag{8}$$

and the optimality part of this reformulation is equivalent to

$$\liminf_{n\to\infty} \alpha_n(D(\rho\|\sigma), E_2|f) \geq \Phi\left(\frac{E_2}{\sqrt{V(\rho\|\sigma)}}\right) \qquad \forall f(n) \in o(\sqrt{n}). \tag{9}$$

Equations (8) and (9), in turn, are equivalent to the second case of equation (2).

3. Proof of main result.

This section is devoted to the proof of our second-order asymptotics presented in Section 2. The proof goes along the line of the reformulation of Theorem 2. At first we make some necessary preparations, and then we accomplish the proof by showing the achievability part and the optimality part sequentially.

3.1. *Preparations.*

Write $\rho = \sum_x \lambda(x)|a_x\rangle\langle a_x|$ and $\sigma = \sum_y \mu(y)|b_y\rangle\langle b_y|$ in their diagonal form, where $\{|a_x\rangle\}_x$ and $\{|b_y\rangle\}_y$, each being an orthonormal basis of the underlying Hilbert space $\mathcal{H}$, are the eigenvectors of ρ and σ, respectively. $\lambda(x)$ and $\mu(y)$ are the corresponding eigenvalues, which satisfy $0 \leq \lambda(x) \leq 1$, $0 < \mu(y) \leq 1$ and $\sum_x \lambda(x) = \sum_y \mu(y) = 1$. Recall that we suppose σ is of full rank, and thus $\mu(y) \neq 0$. Let x^n denote the sequence $x_1x_2\ldots x_n$ and y^n denote $y_1y_2\ldots y_n$. For n copies of the states ρ, we can write

$$\rho^{\otimes n} = \sum_{x^n} \lambda^n(x^n)|a^n_{x^n}\rangle\langle a^n_{x^n}| \tag{10}$$

with $\lambda^n(x^n) = \prod_{i=1}^n \lambda(x_i)$ and $|a^n_{x^n}\rangle = |a_{x_1}\rangle \otimes |a_{x_2}\rangle \otimes \cdots \otimes |a_{x_n}\rangle$. Similarly,

$$\sigma^{\otimes n} = \sum_{y^n} \mu^n(y^n)|b^n_{y^n}\rangle\langle b^n_{y^n}| \tag{11}$$

with $\mu^n(y^n) = \prod_{i=1}^n \mu(y_i)$ and $|b^n_{y^n}\rangle = |b_{y_1}\rangle \otimes |b_{y_2}\rangle \otimes \cdots \otimes |b_{y_n}\rangle$. The subscripts of x and y indicate which systems they belong to. We further write $|a_x\rangle$'s as superpositions of the vectors $\{|b_y\rangle\}_y$, namely, $|a_x\rangle = \sum_y \gamma_{xy}|b_y\rangle$, with $\gamma_{xy} = \langle b_y|a_x\rangle \in \mathbb{C}$ and $\sum_x |\gamma_{xy}|^2 = \sum_y |\gamma_{xy}|^2 = 1$. In such a way, we have

$$|a^n_{x^n}\rangle = \sum_{y^n} \gamma^n_{x^n y^n}|b^n_{y^n}\rangle \qquad \text{with } \gamma^n_{x^n y^n} = \prod_{i=1}^n \gamma_{x_i y_i}. \tag{12}$$

Define a pair of random variables (X, Y), with alphabet $\{(x, y)\}_{x,y=1}^{|\mathcal{H}|}$ and joint distribution $P_{X,Y}(x, y) = \lambda(x)|\gamma_{xy}|^2$. Operationally, this is the probability of obtaining (x, y) when we measure the quantum state ρ, sequentially in the bases $\{|a_x\rangle\}_x$ and $\{|b_y\rangle\}_y$. Let $(X^n, Y^n) := (X_1, Y_1)(X_2, Y_2)\cdots(X_n, Y_n)$ be a sequence of independent and identically distributed random variable pairs, and each (X_i, Y_i) has the same distribution as (X, Y). Then

$$P_{X^n,Y^n}(x^n, y^n) = \prod_{i=1}^n \lambda(x_i)|\gamma_{x_i y_i}|^2 = \lambda^n(x^n)|\gamma^n_{x^n y^n}|^2. \tag{13}$$

As functions of X and Y, $\lambda(X)$ and $\mu(Y)$ are also random variables, and so are $\lambda^n(X^n)$ and $\mu^n(Y^n)$. Using the idea of Nussbaum and Szkoła [26], we are able to express the quantum relative entropy and quantum relative variance as statistical quantities of classical random variables, as follows.

LEMMA 3. *We have*

$$D(\rho\|\sigma) = \mathrm{E}_{(X,Y)} \log \frac{\lambda(X)}{\mu(Y)}, \tag{14}$$

$$V(\rho\|\sigma) = \mathrm{Var}_{(X,Y)} \log \frac{\lambda(X)}{\mu(Y)}. \tag{15}$$

Note again that we are only interested in the case that σ has full rank, so $\mu(Y) > 0$. During the computation of the right-hand sides of equations (14) and (15), if $\lambda(x) = 0$, we let $\lambda(x)\log\lambda(x) := \lim_{z\to 0} z\log z = 0$, and $\lambda(x)\log^2\lambda(x) := \lim_{z\to 0} z\log^2 z = 0$.

We also present below another technical lemma, which will be used in Section 3.3 in the proof of the optimality part.

LEMMA 4. *Let $|\phi\rangle$ and $|\varphi\rangle$ be normalized vectors in some Hilbert space. Let π be a projector and $\|\cdot\|$ the* 2*-norm, that is, $\||\psi\rangle\| := \sqrt{\langle\psi|\psi\rangle}$. If $\||\phi\rangle - \pi|\phi\rangle\| \le \varepsilon$, then*

$$\|(|\phi\rangle\langle\phi|)|\varphi\rangle\|^2 - \|(\pi|\phi\rangle\langle\phi|\pi)|\varphi\rangle\|^2 \le 2\sqrt{2}\varepsilon. \tag{16}$$

The proofs of Lemmas 3 and 4 are given in the Appendix.

3.2. *Proof of the achievability part.* For any fixed $E_2 \in \mathbb{R}$ and $f(n) \in o(\sqrt{n})$, let

$$L_n := \exp\{nD(\rho\|\sigma) + E_2\sqrt{n} + f(n)\}.$$

Associated with every x^n, we define a projector $Q^n_{x^n}$ as

$$Q^n_{x^n} := \sum_{y^n : \lambda^n(x^n)/\mu^n(y^n) \geq L_n} |b^n_{y^n}\rangle\langle b^n_{y^n}|.$$

Write $|\xi^n_{x^n}\rangle := Q^n_{x^n}|a^n_{x^n}\rangle$. Referring to equation (12), we have

$$|\xi^n_{x^n}\rangle = \sum_{y^n : \lambda^n(x^n)/\mu^n(y^n) \geq L_n} \gamma^n_{x^n y^n}|b^n_{y^n}\rangle. \tag{17}$$

Let A_n be the projector onto the space S_n that is spanned by $\{|\xi^n_{x^n}\rangle\}_{x^n}$. We claim that the sequence of measurements $\{(A_n, \mathbb{1} - A_n)\}_n$ is what we needed: it satisfies equations (4) and (5).

Arrange all the values of x^n in such a way that the eigenvalues of $\rho^{\otimes n}$, $\lambda^n(x^n)$'s are in an increasing order. This gives an ordering to the vectors $\{|\xi^n_{x^n}\rangle\}_{x^n}$ as well. Let $g : \{i\}_{i=1}^{|\mathcal{H}|^n} \mapsto \{x^n\}$ be the bijection mapping the position of x^n to x^n itself, that is, x^n is at the $g^{-1}(x^n)$th position in the above ordering. Then we have

$$\lambda^n(g(1)) \leq \lambda^n(g(2)) \leq \cdots \leq \lambda^n(g(|\mathcal{H}|^n)). \tag{18}$$

Applying a modified Gram–Schmidt orthonormalization process to the sequence of vectors

$$|\xi^n_{g(1)}\rangle, |\xi^n_{g(2)}\rangle, |\xi^n_{g(3)}\rangle, \ldots, |\xi^n_{g(|\mathcal{H}|^n)}\rangle,$$

we obtain a new sequence of vectors

$$|\hat{\xi}^n_{g(1)}\rangle, |\hat{\xi}^n_{g(2)}\rangle, |\hat{\xi}^n_{g(3)}\rangle, \ldots, |\hat{\xi}^n_{g(|\mathcal{H}|^n)}\rangle. \tag{19}$$

The modification is that if $|\xi^n_{g(i)}\rangle \in \operatorname{Span}(\{|\xi^n_{g(j)}\rangle\}_{j=1}^{i-1})$ (this includes the case that $|\xi^n_{g(i)}\rangle = 0$), we let $|\hat{\xi}^n_{g(i)}\rangle = 0$. As a result, the set of vectors $\{|\hat{\xi}^n_{x^n}\rangle\}_{x^n}$ consists of an orthonormal basis of the space S_n, plus some zero vectors. Thus

$$A_n = \sum_{x^n}|\hat{\xi}^n_{x^n}\rangle\langle\hat{\xi}^n_{x^n}|. \tag{20}$$

The vectors $\{|\hat{\xi}^n_{x^n}\rangle\}_{x^n}$ have another property as follows. From the Gram–Schmidt process, we know that

$$|\hat{\xi}^n_{g(i)}\rangle = \sum_{j=1}^{i} s^n_{ij}|\xi^n_{g(j)}\rangle \tag{21}$$

for all $1 \leq i \leq |\mathcal{H}|^n$, with the coefficients $s_{ij}^n \in \mathbb{C}$. Further, from equations (17), (18), (21), and paying attention to the definition of g, we conclude that

$$|\hat{\xi}_{x^n}^n\rangle = \sum_{y^n : \lambda^n(x^n)/\mu^n(y^n) \geq L_n} t_{x^n y^n}^n |b_{y^n}^n\rangle, \tag{22}$$

where $t_{x^n y^n}^n \in \mathbb{C}$ and

$$\sum_{y^n : \lambda^n(x^n)/\mu^n(y^n) \geq L_n} \left|t_{x^n y^n}^n\right|^2 = 1. \tag{23}$$

Equations (11), (22), (23) lead to

$$\operatorname{Tr}\big(\sigma^{\otimes n} |\hat{\xi}_{x^n}^n\rangle\langle\hat{\xi}_{x^n}^n|\big) = \sum_{y^n : \lambda^n(x^n)/\mu^n(y^n) \geq L_n} \left|t_{x^n y^n}^n\right|^2 \mu^n(y^n) \leq \frac{\lambda^n(x^n)}{L_n}. \tag{24}$$

So, making use of equations (20) and (24), we arrive at

$$\beta_n(A_n) = \operatorname{Tr} \sigma^{\otimes n} A_n \leq \frac{1}{L_n} = \exp\big\{-\big(nD(\rho\|\sigma) + E_2\sqrt{n} + f(n)\big)\big\}, \tag{25}$$

which is exactly equation (4).

On the other hand, equation (5) is confirmed as follows. Let

$$|\bar{\xi}_{x^n}^n\rangle := \begin{cases} 0, & \text{if } |\xi_{x^n}^n\rangle = 0, \\ \dfrac{|\xi_{x^n}^n\rangle}{\sqrt{\langle\xi_{x^n}^n|\xi_{x^n}^n\rangle}}, & \text{if } |\xi_{x^n}^n\rangle \neq 0. \end{cases}$$

Obviously, $|\bar{\xi}_{x^n}^n\rangle \in S_n$. So

$$|\bar{\xi}_{x^n}^n\rangle\langle\bar{\xi}_{x^n}^n| \leq A_n. \tag{26}$$

Then we have

$$\begin{aligned} \alpha_n(A_n) &= 1 - \operatorname{Tr}\big(\rho^{\otimes n} A_n\big) \\ &\leq 1 - \sum_{x^n} \lambda^n(x^n) \operatorname{Tr}\big(\big(|a_{x^n}^n\rangle\langle a_{x^n}^n|\big)\big(|\bar{\xi}_{x^n}^n\rangle\langle\bar{\xi}_{x^n}^n|\big)\big) \\ &= 1 - \sum_{x^n} \lambda^n(x^n) \langle\xi_{x^n}^n|\xi_{x^n}^n\rangle \\ &= \Pr\left\{\frac{\lambda^n(X^n)}{\mu^n(Y^n)} < L_n\right\}, \end{aligned}$$

where the second line is by equations (10) and (26), the third line can be seen from the definitions of $|\xi_{x^n}^n\rangle$ and $|\bar{\xi}_{x^n}^n\rangle$ and the fourth line follows from equations (17) and (13). Recalling that $\lambda^n(X^n) = \prod_{i=1}^n \lambda(X_i)$ and $\mu^n(Y^n) = \prod_{i=1}^n \mu(Y_i)$, and by taking logarithms at both sides of $\frac{\lambda^n(X^n)}{\mu^n(Y^n)} < L_n$, we further obtain

$$\alpha_n(A_n) \leq \Pr\left\{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \log\frac{\lambda(X_i)}{\mu(Y_i)} - D(\rho\|\sigma)\right) < E_2 + \frac{f(n)}{\sqrt{n}}\right\}. \tag{27}$$

Since $f(n) \in o(\sqrt{n})$, due to the central limit theorem and also by Lemma 3, the limit of right-hand side of equation (27) equals

$$\Phi\Big(\frac{E_2}{\sqrt{V(\rho\|\sigma)}}\Big);$$

thus equation (5) follows, and we are done.

3.3. *Proof of the optimality part.* Suppose that the sequence of measurements $\{(A_n, \mathbb{1} - A_n)\}_n$ satisfies equation (6). We will prove equation (7). Let

$$L_n := \exp\big\{\big(nD(\rho\|\sigma) + E_2\sqrt{n} + f(n)\big) - f'(n)\big\} \tag{28}$$

with some fixed

$$f'(n) \in o(\sqrt{n}) \cap \omega(1). \tag{29}$$

Here $\omega(1)$ is the family of functions that are defined on $\mathbb{N}$ and diverge to $+\infty$. Associated with every x^n, we define the projector $Q^n_{x^n}$ as

$$Q^n_{x^n} := \sum_{y^n : \lambda^n(x^n)/\mu^n(y^n) \geq L_n} |b^n_{y^n}\rangle\langle b^n_{y^n}|. \tag{30}$$

Inserting equation (10) into the definition of $\alpha_n(A_n)$, namely, $\alpha_n(A_n) := \mathrm{Tr}(\rho^{\otimes n}(\mathbb{1} - A_n))$, and after a few calculations, we write

$$\alpha_n(A_n) = 1 - C_n - D_n, \tag{31}$$

where C_n and D_n are

$$C_n := \sum_{x^n} \lambda^n(x^n)\,\mathrm{Tr}\big(Q^n_{x^n}\sqrt{A_n}|a^n_{x^n}\rangle\langle a^n_{x^n}|\sqrt{A_n}Q^n_{x^n}\big), \tag{32}$$

$$D_n := \sum_{x^n} \lambda^n(x^n)\,\mathrm{Tr}\big((\mathbb{1} - Q^n_{x^n})\sqrt{A_n}|a^n_{x^n}\rangle\langle a^n_{x^n}|\sqrt{A_n}(\mathbb{1} - Q^n_{x^n})\big). \tag{33}$$

The basic difficulty in bounding C_n and D_n is that the POVM element A_n is very general, except for the constraint of equation (6). Nevertheless, we will be able to show that the D_n term is asymptotically negligible, due to the constraint of equation (6) and our choice of L_n. This in turn, ensures that the C_n term can be upper bounded by removing the operator "$\sqrt{A_n}$" from its expression, with only an infinitesimal correction; cf. equation (41).

Now we show that the D_n term is asymptotically negligible. Because

$$\sigma^{\otimes n} \geq (\mathbb{1} - Q^n_{x^n})\sigma^{\otimes n}(\mathbb{1} - Q^n_{x^n}) \geq \frac{\lambda^n(x^n)}{L_n}(\mathbb{1} - Q^n_{x^n}),$$

where the first inequality is owing to the commutativity of $\sigma^{\otimes n}$ and the projector $(\mathbb{1} - Q^n_{x^n})$, and the second one can be seen from the definition of $Q^n_{x^n}$, we obtain

$$\begin{aligned}\beta_n(A_n) &= \mathrm{Tr}(\sigma^{\otimes n} A_n) \\ &= \sum_{x^n} \mathrm{Tr}(\sigma^{\otimes n}(\sqrt{A_n}|a^n_{x^n}\rangle\langle a^n_{x^n}|\sqrt{A_n})) \\ &\geq \sum_{x^n} \mathrm{Tr}\Big(\frac{\lambda^n(x^n)}{L_n}(\mathbb{1} - Q^n_{x^n})(\sqrt{A_n}|a^n_{x^n}\rangle\langle a^n_{x^n}|\sqrt{A_n})\Big) \\ &= \frac{D_n}{L_n}.\end{aligned}$$

This result, together with equations (6), (28) and (29), tells us that

$$D_n \leq L_n \beta_n(A_n) \leq \exp\{-f'(n)\} \to 0. \tag{34}$$

The evaluation of the C_n term will be a bit more complicated. For simplicity, we use the notation of norm, $\||\psi\rangle\| = \sqrt{\langle\psi|\psi\rangle} = \sqrt{\mathrm{Tr}\,|\psi\rangle\langle\psi|}$, with $|\psi\rangle$ being a vector of some Hilbert space. Thus C_n is rewritten as

$$C_n = \sum_{x^n} \lambda^n(x^n) \| Q^n_{x^n}\sqrt{A_n}|a^n_{x^n}\rangle\|^2. \tag{35}$$

Our strategy is to divide the terms in the sum of the above expression into different classes, each satisfying some special conditions. Then we evaluate them individually under these conditions. For such a purpose, we define index sets

$$\begin{aligned}\mathcal{O}^n_1 &:= \{x^n | \|\sqrt{A_n}|a^n_{x^n}\rangle\| \geq \epsilon_1\}, \\ \mathcal{O}^n_2 &:= \{x^n | \|(\mathbb{1} - Q^n_{x^n})\sqrt{A_n}|a^n_{x^n}\rangle\| \leq \epsilon_1\epsilon_2\}\end{aligned}$$

with sufficiently small $\epsilon_1, \epsilon_2 > 0$. Denote the full set of all the x^n's as $\mathcal{O}^n$, and the complementary sets of $\mathcal{O}^n_1$ and $\mathcal{O}^n_2$ as $\overline{\mathcal{O}^n_1}$ and $\overline{\mathcal{O}^n_2}$, respectively. Since $\mathcal{O}^n$ is the union of three disjoint subsets

$$\mathcal{O}^n = \overline{\mathcal{O}^n_1} \cup (\mathcal{O}^n_1 \cap \overline{\mathcal{O}^n_2}) \cup (\mathcal{O}^n_1 \cap \mathcal{O}^n_2),$$

we deal with equation (35) under distinct cases that x^n belongs to these subsets, respectively, and then sum them up.

The first case is that $x^n \in \overline{\mathcal{O}^n_1}$. Noting that a projector (more generally, any contraction whose singular values are no larger than 1) acting on a vector will not increase its norm, we have

$$\sum_{x^n \in \overline{\mathcal{O}^n_1}} \lambda^n(x^n) \| Q^n_{x^n}\sqrt{A_n}|a^n_{x^n}\rangle\|^2 \leq \sum_{x^n \in \overline{\mathcal{O}^n_1}} \lambda^n(x^n)\epsilon_1^2 \leq \epsilon_1^2. \tag{36}$$

The second case is that $x^n \in \mathcal{O}_1^n \cap \overline{\mathcal{O}_2^n}$. We upper bound it as

$$\begin{aligned} &\sum_{x^n \in \mathcal{O}_1^n \cap \overline{\mathcal{O}_2^n}} \lambda^n(x^n) \| Q_{x^n}^n \sqrt{A_n} |a_{x^n}^n\rangle\|^2 \\ &\leq \sum_{x^n \in \overline{\mathcal{O}_2^n}} \lambda^n(x^n) \leq \sum_{x^n \in \overline{\mathcal{O}_2^n}} \lambda^n(x^n) \frac{1}{\epsilon_1^2 \epsilon_2^2} \|(\mathbb{1} - Q_{x^n}^n)\sqrt{A_n}|a_{x^n}^n\rangle\|^2 \\ &\leq \frac{1}{\epsilon_1^2\epsilon_2^2} \sum_{x^n} \lambda^n(x^n) \|(\mathbb{1} - Q_{x^n}^n)\sqrt{A_n}|a_{x^n}^n\rangle\|^2 = \frac{D_n}{\epsilon_1^2\epsilon_2^2}, \end{aligned} \tag{37}$$

where for the first inequality we use $\|Q_{x^n}^n\sqrt{A_n}|a_{x^n}^n\rangle\| \leq \|\sqrt{A_n}|a_{x^n}^n\rangle\| \leq \||a_{x^n}^n\rangle\| = 1$, the second inequality is by definition of $\overline{\mathcal{O}_2^n}$ and the last equality can be easily seen from equation (33) and the definition of norm.

The last case, which will turn out to be the dominant part, is that $x^n \in \mathcal{O}_1^n \cap \mathcal{O}_2^n$. In such a case, paying attention to the definition of $\mathcal{O}_1^n$ and $\mathcal{O}_2^n$, we see that

$$\left\| \frac{\sqrt{A_n}|a_{x^n}^n\rangle}{\|\sqrt{A_n}|a_{x^n}^n\rangle\|} - Q_{x^n}^n \frac{\sqrt{A_n}|a_{x^n}^n\rangle}{\|\sqrt{A_n}|a_{x^n}^n\rangle\|} \right\| = \frac{\|(\mathbb{1} - Q_{x^n}^n)\sqrt{A_n}|a_{x^n}^n\rangle\|}{\|\sqrt{A_n}|a_{x^n}^n\rangle\|} \leq \frac{\epsilon_1\epsilon_2}{\epsilon_1} = \epsilon_2.$$

Then, directly applying Lemma 4, we get

$$\begin{aligned} &\left\| \frac{\sqrt{A_n}|a_{x^n}^n\rangle\langle a_{x^n}^n|\sqrt{A_n}}{\|\sqrt{A_n}|a_{x^n}^n\rangle\|^2} |a_{x^n}^n\rangle \right\|^2 \\ &\leq \left\| \left(Q_{x^n}^n \frac{\sqrt{A_n}|a_{x^n}^n\rangle\langle a_{x^n}^n|\sqrt{A_n}}{\|\sqrt{A_n}|a_{x^n}^n\rangle\|^2} Q_{x^n}^n \right) |a_{x^n}^n\rangle \right\|^2 + 2\sqrt{2}\epsilon_2. \end{aligned} \tag{38}$$

Since $0 \leq A_n \leq \mathbb{1}$, it holds that $A_n \leq \sqrt{A_n}$. As a result,

$$\|Q_{x^n}^n\sqrt{A_n}|a_{x^n}^n\rangle\|^2 \leq \|\sqrt{A_n}|a_{x^n}^n\rangle\|^2 \leq \left\| \frac{\langle a_{x^n}^n|\sqrt{A_n}|a_{x^n}^n\rangle}{\langle a_{x^n}^n|A_n|a_{x^n}^n\rangle} \sqrt{A_n}|a_{x^n}^n\rangle \right\|^2. \tag{39}$$

The last term of equation (39) and the left-hand side of equation (38) are actually the same. So, combining these two equations together, and noting that the right-hand side of equation (38) is obviously upper bounded by

$$\|Q_{x^n}^n|a_{x^n}^n\rangle\|^2 + 2\sqrt{2}\epsilon_2,$$

we arrive at

$$\begin{aligned} &\sum_{x^n \in \mathcal{O}_1^n \cap \mathcal{O}_2^n} \lambda^n(x^n) \|Q_{x^n}^n\sqrt{A_n}|a_{x^n}^n\rangle\|^2 \\ &\leq \sum_{x^n \in \mathcal{O}_1^n \cap \mathcal{O}_2^n} \lambda^n(x^n) (\|Q_{x^n}^n|a_{x^n}^n\rangle\|^2 + 2\sqrt{2}\epsilon_2) \\ &\leq \sum_{x^n} \lambda^n(x^n) \|Q_{x^n}^n|a_{x^n}^n\rangle\|^2 + 2\sqrt{2}\epsilon_2. \end{aligned} \tag{40}$$

Now, adding equations (36), (37) and (40) together, we obtain from equation (35) that

$$C_n \leq \sum_{x^n} \lambda^n(x^n) \| Q^n_{x^n} |a^n_{x^n}\rangle\|^2 + \frac{D_n}{\epsilon_1^2 \epsilon_2^2} + \epsilon_1^2 + 2\sqrt{2}\epsilon_2. \tag{41}$$

In analogy to the process in the derivation of equation (5) in Section 3.2, making use of equations (12), (30) and then (13), we can check that the first term of the right-hand side of equation (41) is equal to the probability of the event $\{\lambda^n(X^n)/\mu^n(Y^n) \geq L_n\}$, which is equivalent to

$$\left\{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \log\frac{\lambda(X_i)}{\mu(Y_i)} - D(\rho\|\sigma)\right) \geq E_2 + \frac{f(n) - f'(n)}{\sqrt{n}}\right\}.$$

So inserting equations (34) and (41) into equation (31), we eventually obtain

$$\begin{aligned}\alpha_n(A_n) \geq \Pr&\left\{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \log\frac{\lambda(X_i)}{\mu(Y_i)} - D(\rho\|\sigma)\right) \leq E_2 + \frac{f(n) - f'(n)}{\sqrt{n}}\right\} \\ &- \left(\frac{1}{\epsilon_1^2\epsilon_2^2} + 1\right)\exp\{-f'(n)\} - \epsilon_1^2 - 2\sqrt{2}\epsilon_2.\end{aligned} \tag{42}$$

Recalling that $f(n) \in o(\sqrt{n})$ and $f'(n) \in o(\sqrt{n}) \cap \omega(1)$, and then making use of the central limit theory and Lemma 3, we see that the right-hand side of equation (42) converges to

$$\Phi\left(\frac{E_2}{\sqrt{V(\rho\|\sigma)}}\right) - \epsilon_1^2 - 2\sqrt{2}\epsilon_2,$$

when $n \to \infty$. Thus equation (7) follows, since ϵ_1 and ϵ_2 can be arbitrarily small, and we are done.

4. Finite sample size analysis. In Section 3, we proved the second-order asymptotics. Here we show that our method is able to provide tight bounds for the case of finite sample size as well. The basic idea is to use the Berry–Esseen theorem instead of the central limit theorem.

The Berry–Esseen theorem quantifies how fast the standardized mean of a random sample converges to a normal distribution. Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables, with $\mathrm{E}(X_i) = \overline{X}$, $\mathrm{E}(X_i - \overline{X})^2 = \varrho^2 > 0$, and $\mathrm{E}|X_i - \overline{X}|^3 = \varsigma^3 < +\infty$. Then it asserts

$$\left|\Pr\left\{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n X_i - \overline{X}\right) \leq x\right\} - \Phi\left(\frac{x}{\varrho}\right)\right| \leq \frac{C\varsigma^3}{\sqrt{n}\varrho^3}, \tag{43}$$

where $0.40973 \leq C \leq 0.4784$ is a constant [22].

Consider the minimal type II error given that the type I error is no larger than some constant, and define $\beta_n(\varepsilon) := \min_{A_n}\{\beta_n(A_n)|\alpha_n(A_n) \le \varepsilon\}$. Theorem 5 provides this quantity with tight upper and lower bounds. It has fixed the second-order term in the asymptotic expansion of $-\log \beta_n(\varepsilon)$, and also indicates that the third-order term lies between a constant and $2\log n$. Note that previously, only the first-order term ($nD(\rho\|\sigma)$) is known exactly [19, 28], and the second-order term is known to be of the order $\sqrt{n}$ [2]. Compared to the upper and lower bounds obtained in the independent work of Tomamichel and Hayashi [30], those presented here are tighter in the third-order term and are also relatively cleaner because those in [30] depend on more parameters, such as the number of distinct eigenvalues and the ratio between the maximum and minimum eigenvalues of the quantum state.

THEOREM 5. *Let C be the constant in the Berry–Esseen theorem, and let $T^3 = \mathrm{E}_{(X,Y)}|\log \frac{\lambda(X)}{\mu(Y)} - D(\rho\|\sigma)|^3$; cf. Section 3.1. Then for n sufficiently large such that $\varepsilon - \frac{1}{\sqrt{n}}\frac{CT^3}{\sqrt{V(\rho\|\sigma)}^3} \ge 0$, we have*

$$\begin{aligned} -\log\beta_n(\varepsilon) &\ge nD(\rho\|\sigma) + \sqrt{n}\sqrt{V(\rho\|\sigma)}\Phi^{-1}\biggl(\varepsilon - \frac{1}{\sqrt{n}}\frac{CT^3}{\sqrt{V(\rho\|\sigma)}^3}\biggr) \\ &= nD(\rho\|\sigma) + \sqrt{n}\sqrt{V(\rho\|\sigma)}\Phi^{-1}(\varepsilon) + O(1) \end{aligned} \tag{44}$$

and for n sufficiently large such that $\varepsilon + \frac{1}{\sqrt{n}}(\frac{CT^3}{\sqrt{V(\rho\|\sigma)}^3} + 2) \le 1$, we have

$$\begin{aligned} -\log\beta_n(\varepsilon) &\le nD(\rho\|\sigma) + \sqrt{n}\sqrt{V(\rho\|\sigma)}\Phi^{-1}\biggl(\varepsilon + \frac{1}{\sqrt{n}}\biggl(\frac{CT^3}{\sqrt{V(\rho\|\sigma)}^3} + 2\biggr)\biggr) \\ &\qquad + \log(2^9 n^2) \\ &= nD(\rho\|\sigma) + \sqrt{n}\sqrt{V(\rho\|\sigma)}\Phi^{-1}(\varepsilon) + 2\log n + O(1). \end{aligned} \tag{45}$$

PROOF. The equalities in equations (44) and (45) are easy to see by expanding Φ^{-1} at the point ε using Lagrange's mean value theorem. So it suffices to prove the two inequalities.

Applying the Berry–Esseen theorem to the right-hand side of equation (27), and then following from the argument in Section 3.2 [cf. equations (25) and (27)], we have for any $E_2 \in \mathbb{R}$ and $f(n) \in o(\sqrt{n})$, there exists a sequence of measurements $\{(A_n, \mathbb{1} - A_n)\}_n$, such that

$$\alpha_n(A_n) \le \Phi\biggl(\frac{E_2 + f(n)/\sqrt{n}}{\sqrt{V(\rho\|\sigma)}}\biggr) + \frac{1}{\sqrt{n}}\frac{CT^3}{\sqrt{V(\rho\|\sigma)}^3}, \tag{46}$$

$$\beta_n(A_n) \le \exp\{-(nD(\rho\|\sigma) + E_2\sqrt{n} + f(n))\}. \tag{47}$$

When $\varepsilon - \frac{1}{\sqrt{n}} \frac{CT^3}{\sqrt{V(\rho\|\sigma)}^3} \geq 0$, letting the right-hand side of equation (46) be equal to ε, then eliminating $E_2\sqrt{n} + f(n)$ from equation (47), we get

$$\begin{aligned}
&\alpha_n(A_n) \leq \varepsilon, \\
&\beta_n(A_n) \leq \exp\biggl\{-\biggl(nD(\rho\|\sigma) + \sqrt{n}\sqrt{V(\rho\|\sigma)}\Phi^{-1}\biggl(\varepsilon - \frac{1}{\sqrt{n}} \frac{CT^3}{\sqrt{V(\rho\|\sigma)}^3}\biggr)\biggr)\biggr\}.
\end{aligned}$$

This, by the definition of $\beta_n(\varepsilon)$, leads to the first inequality in equation (44).

On the other hand, applying the Berry–Esseen theorem to the first term of the right-hand side of equation (42), then the argument in Section 3.3 implies the following [cf. the precondition and equation (42)]: if there is a sequence of measurements $\{(A_n, \mathbb{1} - A_n)\}_n$ such that

$$\beta_n(A_n) \leq \exp\{-(nD(\rho\|\sigma) + E_2\sqrt{n} + f(n))\},$$

then

$$\alpha_n(A_n) \geq \Phi\biggl(\frac{E_2 + (f(n) - f'(n))/\sqrt{n}}{\sqrt{V(\rho\|\sigma)}}\biggr) - F$$

with $F = \frac{1}{\sqrt{n}} \frac{CT^3}{\sqrt{V(\rho\|\sigma)}^3} + (\frac{1}{\epsilon_1^2\epsilon_2^2} + 1)\exp\{-f'(n)\} + \epsilon_1^2 + 2\sqrt{2}\epsilon_2$. Since α_n and β_n are continuous functionals of A_n, this equivalently states that if

$$\alpha_n(A_n) \leq \Phi\biggl(\frac{E_2 + (f(n) - f'(n))/\sqrt{n}}{\sqrt{V(\rho\|\sigma)}}\biggr) - F, \tag{48}$$

then

$$\beta_n(A_n) \geq \exp\{-(nD(\rho\|\sigma) + E_2\sqrt{n} + f(n))\}. \tag{49}$$

When $\varepsilon + F \leq 1$, let E_2 and $f(n)$ be such that the right-hand side of equation (48) equals ε, then we eliminate $E_2\sqrt{n} + f(n)$ from equation (49) using this equality. Thus the above statement implies, by the definition of $\beta_n(\varepsilon)$,

$$\beta_n(\varepsilon) \geq \exp\{-(nD(\rho\|\sigma) + \sqrt{n}\sqrt{V(\rho\|\sigma)}\Phi^{-1}(\varepsilon + F) + f'(n))\}. \tag{50}$$

At last, to optimize over the parameters, let $\epsilon_1 = 2^{1/8}\exp\{-\frac{1}{8}f'(n)\}$, $\epsilon_2 = 2^{-1/4}\exp\{-\frac{1}{4}f'(n)\}$ and $f'(n) = \log(2^9 n^2)$. Thus $F \leq \frac{1}{\sqrt{n}}(\frac{CT^3}{\sqrt{V(\rho\|\sigma)}^3} + 2)$. Inserting these into equation (50) results in the inequality of equation (45), and we are done. □

5. Concluding remarks. The relation between our second-order asymptotics and the quantum Stein's lemma is similar in spirit to the relation between the central limit theorem and the week law of large numbers. Indeed, we have employed the central limit theorem and the Berry–Esseen theorem, to derive our results.

We have succeeded in proving the results using elementary linear algebra and probability theory in a unified fashion for the achievability part and optimality part. Specifically, we have explicitly constructed a sequence of asymptotically optimal tests for our problem, specifying the bases of spaces onto which the projective measurements are applied by employing a modified Gram–Schmidt orthonormalization process. In [27], the Gram–Schmidt orthonormalization process has already been used in order to find the asymptotically optimal tests for testing multiple hypotheses in the symmetric setting (regarding the Chernoff bound). The attempt in [27] is successful in some special cases, and is successful in general up to a constant factor 1/3. We notice that even for two hypotheses, such an elementary method is not known for fully proving the achievability of the Chernoff bound; recall that the original—and hitherto unique—proof in [1] was based on the nontrivial matrix inequality $\mathrm{Tr}(\rho^s\sigma^{1-s}) \geq \mathrm{Tr}(\rho+\sigma-|\rho-\sigma|)/2$, for all $0 \leq s \leq 1$.

The case that $V(\rho\|\sigma)=0$ is a singular point in Theorem 2 and Theorem 5; however, we will see that this represents a very trivial case within classical hypothesis testing. Using Lemma 3, we check that the equivalent conditions of $V(\rho\|\sigma)=0$ is as follows: (i) ρ and σ commute. This means that ρ and σ can be simultaneously diagonalized as $\rho=\sum_x \lambda(x)|a_x\rangle\langle a_x|$ and $\sigma=\sum_x \mu(x)|a_x\rangle\langle a_x|$, and our problem reduces to a classical one with probability laws $\{\lambda(x)\}_x$ and $\{\mu(x)\}_x$. (ii) There is a constant k such that for all x with $\lambda(x)\neq 0$, we have $\lambda(x)=k\mu(x)$; and actually $\log k = D(\rho\|\sigma)$. Assigning arbitrarily any x^n with nonzero $\lambda^n(x^n)$ to the null hypothesis $\rho^{\otimes n}$, we obtain the best tradeoff between the type I error α_n and type II error β_n, and this is expressed as $\alpha_n = 1-\beta_n \exp\{nD(\rho\|\sigma)\}$.

APPENDIX: PROOF OF LEMMAS

We give proofs to the two lemmas presented in Section 3.1.

PROOF OF LEMMA 3. This is done by direct calculation. For functions v and w, it is obvious that

$$\mathrm{Tr}\, v(\rho) = \sum_x v(\lambda(x)) = \sum_{xy} v(\lambda(x))|\gamma_{xy}|^2$$

and

$$\begin{aligned}\mathrm{Tr}\, v(\rho)w(\sigma) &= \mathrm{Tr}\Bigl(\sum_x v(\lambda(x))|a_x\rangle\langle a_x|\Bigr)\Bigl(\sum_y w(\mu(y))|b_y\rangle\langle b_y|\Bigr)\\ &= \sum_{xy} v(\lambda(x))w(\mu(y))|\gamma_{xy}|^2.\end{aligned}$$

Using these two equations with proper v and w at every step when needed, we get

$$\begin{aligned}\mathrm{Tr}\,\rho(\log\rho - \log\sigma) &= \sum_{xy}\bigl(\lambda(x)\log\lambda(x)|\gamma_{xy}|^2 - \lambda(x)\log\mu(y)|\gamma_{xy}|^2\bigr)\\ &= \sum_{xy} P_{X,Y}(x,y)\log\frac{\lambda(x)}{\mu(y)} = \mathrm{E}\Bigl(\log\frac{\lambda(X)}{\mu(Y)}\Bigr)\end{aligned} \tag{51}$$

and

$$
\begin{aligned}
&\operatorname{Tr}\rho(\log\rho-\log\sigma)^2\\
&\quad=\operatorname{Tr}\rho\log^2\rho-2\operatorname{Tr}(\rho\log\rho)\log\sigma+\operatorname{Tr}\rho\log^2\sigma\\
&\quad=\sum_{xy}\bigl(\lambda(x)\log^2\lambda(x)|\gamma_{xy}|^2\\
&\qquad\quad-2\lambda(x)\log\lambda(x)\log\mu(y)|\gamma_{xy}|^2+\lambda(x)\log^2\mu(y)|\gamma_{xy}|^2\bigr)\\
&\quad=\sum_{xy}P_{X,Y}(x,y)\biggl(\log\frac{\lambda(x)}{\mu(y)}\biggr)^2=\mathrm{E}\biggl(\log\frac{\lambda(X)}{\mu(Y)}\biggr)^2.
\end{aligned}
\tag{52}
$$

Equation (51) confirms equation (14), and equations (51) and (52) together lead to equation (15). Thus we finish the proof of Lemma 3. □

PROOF OF LEMMA 4. We show equation (16) as follows:

$$
\begin{aligned}
&\bigl\|\bigl(|\phi\rangle\langle\phi|\bigr)|\varphi\rangle\bigr\|^2-\bigl\|\bigl(\pi|\phi\rangle\langle\phi|\pi\bigr)|\varphi\rangle\bigr\|^2\\
&\quad=\bigl(\bigl\|\bigl(|\phi\rangle\langle\phi|\bigr)|\varphi\rangle\bigr\|+\bigl\|\bigl(\pi|\phi\rangle\langle\phi|\pi\bigr)|\varphi\rangle\bigr\|\bigr)\times\bigl(\bigl\|\bigl(|\phi\rangle\langle\phi|\bigr)|\varphi\rangle\bigr\|-\bigl\|\bigl(\pi|\phi\rangle\langle\phi|\pi\bigr)|\varphi\rangle\bigr\|\bigr)\\
&\quad\le 2\bigl(\bigl\|\bigl(|\phi\rangle\langle\phi|\bigr)|\varphi\rangle\bigr\|-\bigl\|\bigl(\pi|\phi\rangle\langle\phi|\pi\bigr)|\varphi\rangle\bigr\|\bigr)\\
&\quad\le 2\bigl\|\bigl(|\phi\rangle\langle\phi|\bigr)|\varphi\rangle-\bigl(\pi|\phi\rangle\langle\phi|\pi\bigr)|\varphi\rangle\bigr\|\\
&\quad=2\bigl\|\bigl(\langle\phi|(\mathbb{1}-\pi)|\varphi\rangle\bigr)\pi|\phi\rangle+\bigl(\langle\phi|\varphi\rangle\bigr)(\mathbb{1}-\pi)|\phi\rangle\bigr\|\\
&\quad=2\sqrt{\bigl|\langle\phi|(\mathbb{1}-\pi)|\varphi\rangle\bigr|^2\cdot\bigl\|\pi|\phi\rangle\bigr\|^2+\bigl|\langle\phi|\varphi\rangle\bigr|^2\cdot\bigl\|(\mathbb{1}-\pi)|\phi\rangle\bigr\|^2}\\
&\quad\le 2\sqrt{\bigl\||\phi\rangle-\pi|\phi\rangle\bigr\|^2\cdot 1+1\cdot\bigl\||\phi\rangle-\pi|\phi\rangle\bigr\|^2}\\
&\quad\le 2\sqrt{\varepsilon^2+\varepsilon^2}=2\sqrt{2}\varepsilon,
\end{aligned}
$$

where the fourth line is by the triangle inequality, the sixth line is due to Pythagoras' theorem and the other lines are trivially by direct calculations and the conditions stated in the lemma. □

Acknowledgments. The author would like to thank Fernando Brandão, Shunlong Luo, William Matthews, Milán Mosonyi and Andreas Winter for interesting discussions. He is especially grateful to Masahito Hayashi for many helpful discussions.

REFERENCES

[1] AUDENAERT, K. M. R., CASAMIGLIA, J., MUNOZ-TAPIA, R., BAGAN, E., MASANES, L., ACIN, A. and VERSTRAETE, F. (2007). Discriminating states: The quantum Chernoff bound. *Phys. Rev. Lett.* **98** 160501.

[2] AUDENAERT, K. M. R., MOSONYI, M. and VERSTRAETE, F. (2012). Quantum state discrimination bounds for finite sample size. *J. Math. Phys.* **53** 122205. MR3058178

[3] AUDENAERT, K. M. R., NUSSBAUM, M., SZKOŁA, A. and VERSTRAETE, F. (2008). Asymptotic error rates in quantum hypothesis testing. *Comm. Math. Phys.* **279** 251–283. MR2377635

[4] BJELAKOVIĆ, I., DEUSCHEL, J.-D., KRÜGER, T., SEILER, R., SIEGMUND-SCHULTZE, R. and SZKOŁA, A. (2005). A quantum version of Sanov's theorem. *Comm. Math. Phys.* **260** 659–671. MR2183961

[5] BJELAKOVIĆ, I., DEUSCHEL, J.-D., KRÜGER, T., SEILER, R., SIEGMUND-SCHULTZE, R. and SZKOŁA, A. (2008). Typical support and Sanov large deviations of correlated states. *Comm. Math. Phys.* **279** 559–584. MR2383599

[6] BJELAKOVIĆ, I. and SIEGMUND-SCHULTZE, R. (2004). An ergodic theorem for the quantum relative entropy. *Comm. Math. Phys.* **247** 697–712. MR2062648

[7] BLAHUT, R. E. (1974). Hypothesis testing and information theory. *IEEE Trans. Inform. Theory* **20** 405–417. MR0396072

[8] BRANDÃO, F. G. S. L. and PLENIO, M. B. (2010). A generalization of quantum Stein's lemma. *Comm. Math. Phys.* **295** 791–828. MR2600034

[9] CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507. MR0057518

[10] COVER, T. M. and THOMAS, J. A. (1991). *Elements of Information Theory*. Wiley, New York. MR1122806

[11] CSISZÁR, I. and LONGO, G. (1971). On the error exponent for source coding and for testing simple statistical hypotheses. *Studia Sci. Math. Hungar.* **6** 181–191. MR0343992

[12] HAN, T. S. (2003). *Information-Spectrum Methods in Information Theory*. Springer, Berlin. MR1938683

[13] HAN, T. S. and KOBAYASHI, K. (1989). The strong converse theorem for hypothesis testing. *IEEE Trans. Inform. Theory* **35** 178–180. MR0995334

[14] HAYASHI, M. (2006). *Quantum Information*: *An Introduction*. Springer, Berlin. MR2228302

[15] HAYASHI, M. (2007). Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding. *Phys. Rev. A* (3) **76** 062301.

[16] HAYASHI, M. (2009). Information spectrum approach to second-order coding rate in channel coding. *IEEE Trans. Inform. Theory* **55** 4947–4966. MR2596952

[17] HAYASHI, M. and NAGAOKA, H. (2003). General formulas for capacity of classical-quantum channels. *IEEE Trans. Inform. Theory* **49** 1753–1768. MR1985576

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

[19] HIAI, F. and PETZ, D. (1991). The proper formula for relative entropy and its asymptotics in quantum probability. *Comm. Math. Phys.* **143** 99–114. MR1139426

[20] HOEFFDING, W. (1965). Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist.* **36** 369–408. MR0173322

[21] HOLEVO, A. S. (1978). On asymptotically optimal hypothesis testing in quantum statistics. *Theory Probab. Appl.* **23** 411–415.

[22] KOROLEV, V. and SHEVTSOVA, I. (2012). An improvement of the Berry–Esseen inequality with applications to Poisson and mixed Poisson random sums. *Scand. Actuar. J.* **2** 81–105. MR2929524

[23] MOSONYI, M. and DATTA, N. (2009). Generalized relative entropies and the capacity of classical-quantum channels. *J. Math. Phys.* **50** 072104. MR2548615

[24] NAGAOKA, H. (2006). The converse part of the theorem for quantum Hoeffding bound. Preprint. Available at arXiv:quant-ph/0611289.

[25] NAGAOKA, H. and HAYASHI, M. (2007). An information-spectrum approach to classical and quantum hypothesis testing for simple hypotheses. *IEEE Trans. Inform. Theory* **53** 534–549. MR2302768

[26] NUSSBAUM, M. and SZKOŁA, A. (2009). The Chernoff lower bound for symmetric quantum hypothesis testing. *Ann. Statist.* **37** 1040–1057. MR2502660

[27] NUSSBAUM, M. and SZKOŁA, A. (2011). An asymptotic error bound for testing multiple quantum hypotheses. *Ann. Statist.* **39** 3211–3233. MR3012406

[28] OGAWA, T. and NAGAOKA, H. (2000). Strong converse and Stein's lemma in quantum hypothesis testing. *IEEE Trans. Inform. Theory* **46** 2428–2433. MR1806811

[29] POLYANSKIY, Y., POOR, H. V. and VERDÚ, S. (2010). Channel coding rate in the finite blocklength regime. *IEEE Trans. Inform. Theory* **56** 2307–2359. MR2729787

[30] TOMAMICHEL, M. and HAYASHI, M. (2012). A hierarchy of information quantities for finite block length analysis of quantum tasks. Preprint. Available at arXiv:1208.1478 [quant-ph].

[31] VERDÚ, S. and HAN, T. S. (1994). A general formula for channel capacity. *IEEE Trans. Inform. Theory* **40** 1147–1157.

[32] WANG, L. and RENNER, R. (2012). One-shot classical-quantum capacity and hypothesis testing. *Phys. Rev. Lett.* **108** 200501.

IBM T. J. WATSON RESEARCH CENTER
1101 KITCHAWAN ROAD
YORKTOWN HEIGHTS, NEW YORK 10598
USA
E-MAIL: carl.ke.lee@gmail.com