Strong Converse and Stein's Lemma
in the Quantum Hypothesis Testing

Tomohiro Ogawa and Hiroshi Nagaoka[*]

Abstract

The hypothesis testing problem of two quantum states is treated. We show a new inequality between the error of the first kind and the second kind, which complements the result of Hiai and Petz to establish the quantum version of Stein's lemma. The inequality is also used to show a bound on the first kind error when the power exponent for the second kind error exceeds the quantum relative entropy, and the bound yields the strong converse in the quantum hypothesis testing. Finally, we discuss the relation between the bound and the power exponent derived by Han and Kobayashi in the classical hypothesis testing.

Keywords

Quantum hypothesis testing, Stein's lemma, strong converse, quantum relative entropy.

1   Introduction

Let $\mathcal{H}$ be a Hilbert space which represents a physical system in interest. We suppose $\dim \mathcal{H} < \infty$ for mathematical simplicity. Let $\mathcal{B}(\mathcal{H})$ be the set of linear operators on $\mathcal{H}$ and put

$$\mathcal{S}(\mathcal{H}) \stackrel{\text{def}}{=} \{\rho \in \mathcal{B}(\mathcal{H}) \,|\, \rho = \rho^* \geq 0, \operatorname{Tr} \rho = 1\},$$

which is the set of density operators on $\mathcal{H}$.

We treat the problem of hypothesis testing a null hypothesis $\rho \in \mathcal{S}(\mathcal{H})$ versus an alternative hypothesis $\sigma \in \mathcal{S}(\mathcal{H})$. Here, we assume $\operatorname{Im} \rho \subset \operatorname{Im} \sigma$. To consider an asymptotic situation, suppose that either $\rho^{\otimes n} \in \mathcal{S}(\mathcal{H}^{\otimes n})$ or $\sigma^{\otimes n} \in \mathcal{S}(\mathcal{H}^{\otimes n})$ is given. The problem is to decide which hypothesis is true, and the decision is given by a two-valued quantum measurement $\{A_n, 1 - A_n\} \, (A_n \in \mathcal{B}(\mathcal{H}^{\otimes n}), 0 \leq A_n \leq 1)$, where A_n corresponds to the acceptance of $\rho^{\otimes n}$ and $1 - A_n$ corresponds to the acceptance of $\sigma^{\otimes n}$. We call $A_n \in \mathcal{B}(\mathcal{H}^{\otimes n})$ $(0 \leq A_n \leq 1)$ a test in the sequel.

---

[*]The authors are with the Graduate School of Information Systems, University of Electro-Communications, 1–5–1 Chofugaoka, Chofu, Tokyo 182–8585, Japan. (E-mail: ogawa@hn.is.uec.ac.jp, nagaoka@is.uec.ac.jp)

For a test A_n, define the error probability of the first kind and the second kind by

$$\alpha_n(A_n) \stackrel{\text{def}}{=} \operatorname{Tr}\rho^{\otimes n}(1 - A_n),$$
$$\beta_n(A_n) \stackrel{\text{def}}{=} \operatorname{Tr}\sigma^{\otimes n}A_n,$$

respectively. We see that $\alpha_n(A_n)$ is the error probability of the acceptance of $\sigma^{\otimes n}$ when $\rho^{\otimes n}$ is true and $\beta_n(A_n)$ is the error probability of the converse situation. Since we can not have $\alpha_n(A_n)$ and $\beta_n(A_n)$ arbitrarily small simultaneously, we will make $\beta_n(A_n)$ as small as possible under the constraint $\alpha_n(A_n) \le \varepsilon$. In other words, the problem is to examine the asymptotic behavior of the following quantity:

$$\beta_n^*(\varepsilon) \stackrel{\text{def}}{=} \min\{\beta_n(A_n) \mid A_n \in \mathcal{B}(\mathcal{H}^{\otimes n}),\, 0 \le A_n \le I,\, \alpha_n(A_n) \le \varepsilon\}.$$

Concerning $\beta_n^*(\varepsilon)$, Hiai and Petz [1] showed

$$\limsup_{n\to\infty} \frac{1}{n}\log\beta_n^*(\varepsilon) \le -D(\rho||\sigma), \tag{1}$$

and

$$-\frac{1}{1-\varepsilon}D(\rho||\sigma) \le \liminf_{n\to\infty}\frac{1}{n}\log\beta_n^*(\varepsilon), \tag{2}$$

where

$$D(\rho||\sigma) \stackrel{\text{def}}{=} \operatorname{Tr}\rho(\log\rho - \log\sigma),$$

is the quantum relative entropy. As for (2), they used the monotonicity of the quantum relative entropy [2, 3] as follows:

$$
\begin{aligned}
&D(\rho^{\otimes n}||\sigma^{\otimes n})\\
\ge\ & \alpha_n(A_n)\log\frac{\alpha_n(A_n)}{1-\beta_n(A_n)} + (1-\alpha_n(A_n))\log\frac{1-\alpha_n(A_n)}{\beta_n(A_n)}\\
=\ & -h(\alpha_n(A_n)) - \alpha_n(A_n)\log(1-\beta_n(A_n)) - (1-\alpha_n(A_n))\log\beta_n(A_n)\\
\ge\ & -\log 2 - (1-\alpha_n(A_n))\log\beta_n(A_n),
\end{aligned}
$$

where $h(x)$ is the binary entropy. Thus it holds that

$$(1 - \alpha_n(A_n))\frac{1}{n}\log\beta_n(A_n) \ge -\frac{\log 2}{n} - D(\rho||\sigma), \tag{3}$$

which immediately yields (2). Note that (3) also leads the weak converse property, which means that if $\beta_n(A_n) \le e^{-nr}$ $(r > D(\rho||\sigma))$ then $\alpha_n(A_n)$ does not go to zero as $n \to \infty$.

In this paper, we will show a fundamental inequality, which complements (1) by Hiai and Petz to show the quantum version of Stein's lemma (see $e.g.$ [4], p.115). We will also show a bound on $1 - \alpha_n(A_n)$ under the exponential-type constraint $\beta_n(A_n) \le e^{-nr}$. The bound leads to the strong converse property [5, 6] in the quantum hypothesis testing, i.e., if $\beta_n(A_n) \le e^{-nr}$ $(r > D(\rho||\sigma))$ then $\alpha_n(A_n)$ goes to one as $n \to \infty$. Finally, we discuss the relation with the result of Han and Kobayashi [6] in the classical hypothesis testing.

2 A Fundamental Bound on the Error Probabilities

In this section, we show a fundamental inequality between the error probabilities of the first kind and the second kind.

Let λ be a real number and

$$\rho^{\otimes n} - e^{n\lambda}\sigma^{\otimes n} = \sum_j \mu_{n,j} E_{n,j}, \tag{4}$$

be the spectral decomposition. Define a test $X_{n,\lambda}$ by

$$X_{n,\lambda} \stackrel{\text{def}}{=} \sum_{j \in D_n} E_{n,j},$$

where $D_n = \{j \mid \mu_{n,j} \geq 0\}$. Then, the following lemma holds, which corresponds to the quantum version of the Neyman-Pearson lemma (see [7], p.108).

Lemma 1 *For any test A_n, we have*

$$\text{Tr}\,(\rho^{\otimes n} - e^{n\lambda}\sigma^{\otimes n})X_{n,\lambda} \geq \text{Tr}\,(\rho^{\otimes n} - e^{n\lambda}\sigma^{\otimes n})A_n. \tag{5}$$

Proof:

$$
\begin{aligned}
\text{Tr}\,(\rho^{\otimes n} - e^{n\lambda}\sigma^{\otimes n})A_n &= \sum_j \mu_{n,j}\text{Tr}\,E_{n,j}A_n \\
&\leq \sum_{j \in D_n} \mu_{n,j}\text{Tr}\,E_{n,j}A_n \\
&\leq \sum_{j \in D_n} \mu_{n,j}\text{Tr}\,E_{n,j} \\
&= \text{Tr}\,(\rho^{\otimes n} - e^{n\lambda}\sigma^{\otimes n})X_{n,\lambda}.
\end{aligned}
$$

∎

Theorem 1 *For any test A_n and any $\lambda \in \mathbf{R}$, we have*

$$1 - \alpha_n(A_n) \leq e^{-n\varphi(\lambda)} + e^{n\lambda}\beta_n(A_n), \tag{6}$$

where

$$\varphi(\lambda) \stackrel{\text{def}}{=} \max_{0 \leq s \leq 1}\{\lambda s - \psi(s)\}, \tag{7}$$

$$\psi(s) \stackrel{\text{def}}{=} \log \text{Tr}\,\rho^{1+s}\sigma^{-s}. \tag{8}$$

Here, note that $\varphi(\lambda)$ is the Legendre transformation of a convex function $\psi(s)$ (see Fig. 1 and 2). Putting $A = \log \rho - \log \sigma - \psi'(s)$, the convexity of $\psi(s)$ is verified as

$$
\begin{aligned}
\psi'(s) &= e^{-\psi(s)}\,\text{Tr}\,\rho^{1+s}\sigma^{-s}(\log \rho - \log \sigma), \\
\psi''(s) &= e^{-\psi(s)}\,\text{Tr}\,\rho^{1+s}A\sigma^{-s}A \\
&= e^{-\psi(s)}\,\text{Tr}\,\left(\rho^{\frac{1+s}{2}}A\sigma^{-\frac{s}{2}}\right)\left(\rho^{\frac{1+s}{2}}A\sigma^{-\frac{s}{2}}\right)^* \\
&> 0.
\end{aligned}
$$

Observing that $\psi(0) = 0$ and $\psi'(0) = D(\rho||\sigma)$, we can see that if $\lambda > D(\rho||\sigma)$ then $\varphi(\lambda) > 0$. It is important to note that for any λ satisfying $D(\rho||\sigma) \leq \lambda \leq \psi'(1)$, we have

$$s^* = \underset{0 \leq s \leq 1}{\operatorname{argmax}}\{\lambda s - \psi(s)\} \iff \psi'(s^*) = \lambda. \tag{9}$$

Proof of Theorem 1: Define probability distributions $p_n = \{p_{n,j}\}$ and $q_n = \{q_{n,j}\}$ by

$$p_{n,j} = \operatorname{Tr} \rho^{\otimes n} E_{n,j}, \quad q_{n,j} = \operatorname{Tr} \sigma^{\otimes n} E_{n,j}.$$

From (4), we have $\mu_{n,j}\operatorname{Tr} E_{n,j} = p_{n,j} - e^{n\lambda}q_{n,j}$, and hence,

$$D_n = \{j \,|\, 0 \leq \forall s \leq 1, \, e^{-n\lambda s}p_{n,j}^s q_{n,j}^{-s} \geq 1\}.$$

Thus, we obtain

$$
\begin{aligned}
\operatorname{Tr} \rho^{\otimes n} X_{n,\lambda} &= \sum_{j \in D_n} \operatorname{Tr} \rho^{\otimes n} E_{n,j} \\
&= \sum_{j \in D_n} p_{n,j} \\
&\leq \sum_{j \in D_n} p_{n,j} \cdot e^{-n\lambda s}p_{n,j}^s q_{n,j}^{-s} \\
&\leq e^{-n\lambda s} \sum_j p_{n,j}^{1+s} q_{n,j}^{-s} \\
&\leq e^{-n\lambda s} \operatorname{Tr} (\rho^{\otimes n})^{1+s}(\sigma^{\otimes n})^{-s},
\end{aligned}
$$

where we used the monotonicity of the quantum f-divergence [8] for an operator convex function $f(u) = u^{-s}$ $(0 \leq s \leq 1)$ (see *e.g.* [9], p.123). Therefore, we have

$$\operatorname{Tr} \rho^{\otimes n} X_{n,\lambda} \leq \exp\left[-n\{\lambda s - \psi(s)\}\right],$$

and hence,

$$\operatorname{Tr} \rho^{\otimes n} X_{n,\lambda} \leq e^{-n\varphi(\lambda)},$$

by taking the maximum. Now, from (5), the theorem is proved as follows:

$$
\begin{aligned}
1 - \alpha_n(A_n) &= \operatorname{Tr} \rho^{\otimes n} A_n \\
&\leq \operatorname{Tr} (\rho^{\otimes n} - e^{n\lambda}\sigma^{\otimes n})X_{n,\lambda} + e^{n\lambda}\operatorname{Tr} \sigma^{\otimes n} A_n \\
&\leq \operatorname{Tr} \rho^{\otimes n} X_{n,\lambda} + e^{n\lambda}\operatorname{Tr} \sigma^{\otimes n} A_n \\
&\leq e^{-n\varphi(\lambda)} + e^{n\lambda}\beta_n(A_n).
\end{aligned}
$$

∎

3   The Quantum Stein's Lemma

Theorem 2 *For any $0 \leq \varepsilon < 1$, it holds that*

$$\lim_{n \to \infty} \frac{1}{n} \log \beta_n^*(\varepsilon) = -D(\rho||\sigma). \tag{10}$$

4

Proof: From (1) by Hiai and Petz, we have only to show that

$$\liminf_{n\to\infty} \frac{1}{n} \log \beta_n^*(\varepsilon) \geq -D(\rho||\sigma). \tag{11}$$

Let A_n be an arbitrary test which satisfies $\alpha_n(A_n) \leq \varepsilon$. From (6), we have

$$1 - \varepsilon \leq 1 - \alpha_n(A_n) \leq e^{-n\varphi(\lambda)} + e^{n\lambda}\beta_n(A_n),$$

and hence,

$$\beta_n(A_n) \geq e^{-n\lambda}(1 - \varepsilon - e^{-n\varphi(\lambda)}).$$

By taking the minimum, we obtain

$$\beta_n^*(\varepsilon) \geq e^{-n\lambda}(1 - \varepsilon - e^{-n\varphi(\lambda)}). \tag{12}$$

Now, let $\lambda = D(\rho||\sigma) + \delta$ ($\delta > 0$), then $\varphi(\lambda) > 0$ and $1 - \varepsilon - e^{-n\varphi(\lambda)} > 0$ for sufficiently large n. Thus, (12) yields

$$\frac{1}{n} \log \beta_n^*(\varepsilon) \geq -\lambda + \frac{1}{n} \log(1 - \varepsilon - e^{-n\varphi(\lambda)}),$$

and hence,

$$\liminf_{n\to\infty} \frac{1}{n} \log \beta_n^*(\varepsilon) \geq -D(\rho||\sigma) - \delta.$$

Since $\delta > 0$ is arbitrary, the theorem has been proved. ∎

4   Strong Converse

Theorem 3 *For any test A_n, if*

$$\limsup_{n\to\infty} \frac{1}{n} \log \beta_n(A_n) \leq -r, \tag{13}$$

then

$$\limsup_{n\to\infty} \frac{1}{n} \log(1 - \alpha_n(A_n)) \leq -\varphi(\lambda^*), \tag{14}$$

where λ^* is a real number which satisfies $\varphi(\lambda^*) = r - \lambda^*$. Moreover, $\varphi(\lambda^*)$ is represented as

$$\varphi(\lambda^*) = \max_{0 \leq s \leq 1} \left\{ \frac{s}{1+s} r - \frac{1}{1+s} \psi(s) \right\}. \tag{15}$$

Proof: For all $\delta > 0$, there exists n_0 such that

$$\beta_n(A_n) \leq e^{-n(r-\delta)}, \quad \forall n \geq n_0,$$

from (13). Putting $\lambda = \lambda^*$ in (6), we have

$$1 - \alpha_n(A_n) \le e^{-n\varphi(\lambda^*)} + e^{-n(r - \lambda^* - \delta)}, \quad \forall n \ge n_0,$$

and hence,

$$\limsup_{n \to \infty} \frac{1}{n} \log(1 - \alpha_n(A_n)) \le -\varphi(\lambda^*) + \delta.$$

Since $\delta > 0$ is arbitrary, (14) has been proved.

To show (15), suppose that $\psi'(0) \le r \le 2\psi'(1) - \psi(1)$ firstly (see Fig. 2), and s^* attains the maximum in the equation

$$u(r) \overset{\text{def}}{=} \varphi(\lambda^*) = \max_{0 \le s \le 1} \{s\lambda^* - \psi(s)\} = r - \lambda^*.$$

Then, taking (9) into account, $u(r)$ is represented parametrically as

$$u(r) = s^* \psi'(s^*) - \psi(s^*), \tag{16}$$
$$\text{where,} \quad r = (s^* + 1)\psi'(s^*) - \psi(s^*). \tag{17}$$

By using (17) to eliminate $\psi'(s^*)$ from (16), we have

$$u(r) = \frac{s^*}{s^* + 1} r - \frac{1}{s^* + 1}\psi(s^*).$$

On the other hand, let

$$g(s) \overset{\text{def}}{=} \frac{s}{s + 1} r - \frac{1}{s + 1}\psi(s),$$

then we have

$$g'(s) = \frac{1}{(s + 1)^2} \{r + \psi(s) - (1 + s)\psi'(s)\}.$$

To examine the sign of $g'(s)$, put $h(s) \overset{\text{def}}{=} r + \psi(s) - (1 + s)\psi'(s)$, and we have $h'(s) = -(1 + s)\psi''(s) \le 0$, which indicates that the sign of $g'(s)$ changes at most once. Therefore $g(s)$ takes its maximum at $s = \hat{s}$ if and only if

$$r + \psi(\hat{s}) - (1 + \hat{s})\psi'(\hat{s}) = 0.$$

This is nothing but the condition (17), and hence, we obtain $u(r) = \max_{0 \le s \le 1} g(s)$.

In the other cases, it is clear that

$$\varphi(\lambda^*) = \frac{1}{2} r - \frac{1}{2}\psi(1) = g(1) = \max_{0 \le s \le 1} g(s), \quad \text{if} \quad r \ge 2\psi'(1) - \psi(1),$$

and

$$\varphi(\lambda^*) = 0 = g(0) = \max_{0 \le s \le 1} g(s), \quad \text{if} \quad r \le \psi'(0).$$

∎

It should be noted that (15) corresponds to the representation which Blahut [5] derived, in the classical hypothesis testing (i.e., when ρ and σ commute), concerning the power exponent for $\alpha_n(A_n)$ when $r < D(\rho||\sigma)$. We can easily see that if $r > D(\rho||\sigma)$ then $\varphi(\lambda^*) > 0$ (see Fig. 2). Hence, the following corollary holds.

Corollary 1 *For any test A_n, if*

$$\limsup_{n \to \infty} \frac{1}{n} \log \beta_n(A_n) < -D(\rho||\sigma), \tag{18}$$

then $\alpha_n(A_n)$ goes to one exponentially.

5   Relation with the Classical Hypothesis Testing

In this section, we discuss the relation between $\varphi(\lambda^*)$ and the power exponent derived by Han and Kobayashi [6] in the classical hypothesis testing.

Let p and q be probability distributions on a finite set $\mathcal{X}$, a null hypothesis and an alternative hypothesis respectively. And define $\alpha_n(A_n) \stackrel{\text{def}}{=} p^n(A_n^c)$ and $\beta_n(A_n) \stackrel{\text{def}}{=} q^n(A_n)$, where p^n and q^n are the i.i.d. extensions of p and q, and $A_n \subset \mathcal{X}^n$ is an acceptance region of p^n. Blahut [5] proved that if $\beta_n(A_n) \leq e^{-nr}$ ($r > D(p||q)$) then $\alpha_n(A_n)$ tends to one as $n \to \infty$ for all $A_n \subset \mathcal{X}^n$. Although Blahut showed that $1 - \alpha_n(A_n)$ converges at one in the polynomial order, Han and Kobayashi [6] proved a stronger result. Putting

$$\alpha_n^*(r) \stackrel{\text{def}}{=} \min\{\alpha_n(A_n) \,|\, A_n \subset \mathcal{X}^n, \beta_n(A_n) \leq e^{-nr}\},$$

they derived the power exponent for $1 - \alpha_n^*(r)$, namely, they proved

$$\liminf_{n \to \infty} \frac{1}{n} \log(1 - \alpha_n^*(r)) = -\tilde{u}(r), \tag{19}$$

where

$$\tilde{u}(r) \stackrel{\text{def}}{=} \min_{\hat{p}:D(\hat{p}||q) \leq r} \{D(\hat{p}||p) + r - D(\hat{p}||q)\}. \tag{20}$$

As remarked in Han and Kobayashi [6], when $r > D(p||q)$ is not so large, the minimum of (20) is attained with equality, which we suppose here. Applying the method used in Appendix, (20) is rewritten as

$$\tilde{u}(r) = \max_{s \geq 0} \left\{ \frac{s}{1+s}r - \frac{1}{1+s} \log \sum_{j \in \mathcal{X}} p_j^{1+s} q_j^{-s} \right\}. \tag{21}$$

Moreover, when $r > D(p||q)$ is sufficiently small, (21) yields

$$\tilde{u}(r) = \max_{0 \leq s \leq 1} \left\{ \frac{s}{1+s}r - \frac{1}{1+s} \log \sum_{j \in \mathcal{X}} p_j^{1+s} q_j^{-s} \right\},$$

which corresponds to (15).

It is interesting to observe that in the quantum case we have

$$\min_{\hat{\rho}:D(\hat{\rho}||\sigma)\leq r}\{D(\hat{\rho}||\rho)+r-D(\hat{\rho}||\sigma)\}=\max_{s\geq 0}\left\{\frac{s}{1+s}r-\frac{1}{1+s}\overline{\psi}(s)\right\},$$

where we put

$$\overline{\psi}(s)=\log\operatorname{Tr}e^{(1+s)\log\rho-s\log\sigma}.$$

By the Golden-Thompson inequality (see *e.g.* [9], p.261), we can see that $\psi(s)\geq\overline{\psi}(s)$ and the equality holds if and only if ρ and σ commute. Thus, we have

$$\min_{\hat{\rho}:D(\hat{\rho}||\sigma)\leq r}\{D(\hat{\rho}||\rho)+r-D(\hat{\rho}||\sigma)\}\geq\varphi(\lambda^*). \tag{22}$$

6   Concluding Remarks

So far we have shown the fundamental inequality, and seen that the quantum Stein's lemma and the strong converse in the quantum hypothesis testing are obtained as applications of the inequality.

In the classical hypothesis testing, $\tilde{u}(r)$ is shown to be the optimal exponent. However, whether $\varphi(\lambda^*)$ in the quantum hypothesis testing is optimal or not is left open.

Appendix

We show (21) for readers' convenience. Here, we will derive (21) by the information geometrical method (see e.g., [10]), although (21) can be shown by using the Lagrange multiplier method as given in [5].

Suppose that $r>D(p||q)$ is not so large that there exists a probability distribution of the form:

$$p(s)_j\overset{\text{def}}{=}e^{-\tilde{\psi}(s)}p_j^{1+s}q_j^{-s},\quad(j\in\mathcal{X},\,s>0),$$
$$\text{where,}\quad\tilde{\psi}(s)\overset{\text{def}}{=}\log\sum_{j\in\mathcal{X}}p_j^{1+s}q_j^{-s},$$

such that $D(p(s)||q)=r$. Note that

$$\tilde{\psi}'(s)\ =\ E_{p(s)}\left[\log\frac{p}{q}\right]\overset{\text{def}}{=}\eta(s),$$
$$\tilde{\psi}''(s)\ =\ E_{p(s)}\left[\left(\log\frac{p}{q}-\eta(s)\right)^2\right]>0.$$

Firstly, we will show that for all probability distribution $\hat{p}$ it holds that

$$D(\hat{p}||q)=D(p(s)||q)\implies D(\hat{p}||p)\geq D(p(s)||p). \tag{23}$$

To this end, suppose that there exists a probability distribution $\hat{p}$ which satisfies $D(\hat{p}||q)=D(p(s)||q)$ and

$$E_{p(s)}\left[\log\frac{p}{q}\right]<E_{\hat{p}}\left[\log\frac{p}{q}\right].$$

8

Then, we have

$$
\begin{aligned}
D(\hat{p}||q) &= D(\hat{p}||p(s)) + D(p(s)||q) + \sum_{j\in\mathcal{X}}(\hat{p}_j - p(s)_j)(\log p(s)_j - \log q_j) \\
&= D(\hat{p}||p(s)) + D(p(s)||q) + \sum_{j\in\mathcal{X}}(\hat{p}_j - p(s)_j)\left((1+s)\log\frac{p_j}{q_j} - \tilde{\psi}(s)\right) \\
&= D(\hat{p}||p(s)) + D(p(s)||q) + (1+s)\left(E_{\hat{p}}\left[\log\frac{p}{q}\right] - E_{p(s)}\left[\log\frac{p}{q}\right]\right) \\
&> D(p(s)||q),
\end{aligned}
$$

which contradicts with the assumption. Therefore, for all probability distribution with $D(\hat{p}||q) = D(p(s)||q)$, we have

$$
E_{p(s)}\left[\log\frac{p}{q}\right] \geq E_{\hat{p}}\left[\log\frac{p}{q}\right],
$$

and hence, there exists $t \leq s$ such that

$$
E_{p(t)}\left[\log\frac{p}{q}\right] = E_{\hat{p}}\left[\log\frac{p}{q}\right], \tag{24}
$$

since $\eta(s)$ is continuous and monotone increasing. Now, from (24) and the Pythagorean relation for the Kullback-Leibler divergence we have

$$
\begin{aligned}
D(\hat{p}||q) &= D(\hat{p}||p(t)) + D(p(t)||q) \\
&= D(p(s)||q). \tag{25}
\end{aligned}
$$

Using the Pythagorean relation one more times and from (25), (23) is proved as follows:

$$
\begin{aligned}
D(\hat{p}||p) - D(p(s)||p) &= D(\hat{p}||p(t)) + D(p(t)||p) - D(p(s)||p) \\
&= D(p(s)||q) - D(p(t)||q) + D(p(t)||p) - D(p(s)||p) \\
&= \{D(p(s)||q) - D(p(s)||p)\} - \{D(p(t)||q) - D(p(t)||p)\} \\
&= \eta(s) - \eta(t) \\
&\geq 0.
\end{aligned}
$$

Now, taking (23) into account, (20) is represented as

$$
\begin{aligned}
\tilde{u}(r) &= \min_{s:D(p(s)||q)\leq r} \{D(p(s)||p) + r - D(p(s)||q)\} \\
&= \min_{s:D(p(s)||q)\leq r} \{r - \eta(s)\}. \tag{26}
\end{aligned}
$$

Here, we can see that $d(s) \overset{\text{def}}{=} D(p(s)||q)$ is a monotone increasing function of s, which is verified by $d'(s) = (1+s)\tilde{\psi}''(s) > 0$. Thus, the minimum of (26) is attained with equality, and $\tilde{u}(r)$ is represented parametrically as

$$
\tilde{u}(r) = D(p(s)||p) = s\,\eta(s) - \tilde{\psi}(s),
$$
$$
\text{where,} \quad r = D(p(s)||q) = (1+s)\eta(s) - \tilde{\psi}(s).
$$

This representation corresponds to (16) (17) and we obtain (21) by following the same procedure as the proof of Theorem 3.

Acknowledgment

References

[1] F. Hiai and D. Petz, "The proper formula for relative entropy and its asymptotics in quantum probability," *Commun. Math. Phys.*, vol. 143, pp. 99–114, 1991.

[2] G. Lindblad, "Completely positive maps and entropy inequalities," *Commun. Math. Phys.*, vol. 40, pp. 147–151, 1975.

[3] A. Uhlmann, "Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory," *Commun. Math. Phys.*, vol. 54, pp. 21–32, 1997.

[4] R. E. Blahut, *Principles and Practice of Information Theory*, Addison-Wesley, Massachusetts, 1991.

[5] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 405–417, 1974.

[6] T. S. Han and K. Kobayashi, "The strong converse theorem for hypothesis testing," *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 178–180, 1989.

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

[8] D. Petz, "Quasi-entropies for finite quantum systems," *Rep. Math. Phys.*, vol. 23, pp. 57–65, 1986.

[9] R. Bhatia, *Matrix Analysis*, Springer, New York, 1997.

[10] S. Amari, *Differential-Geometrical Methods in Statistics*, Springer, New York, 1985.
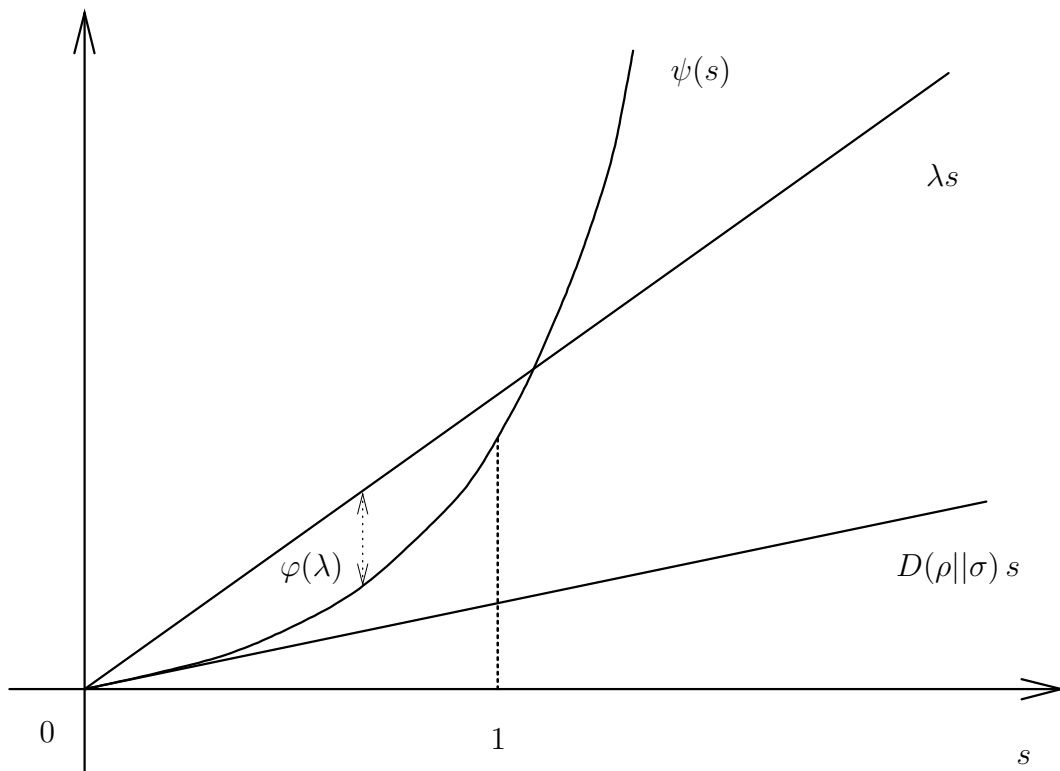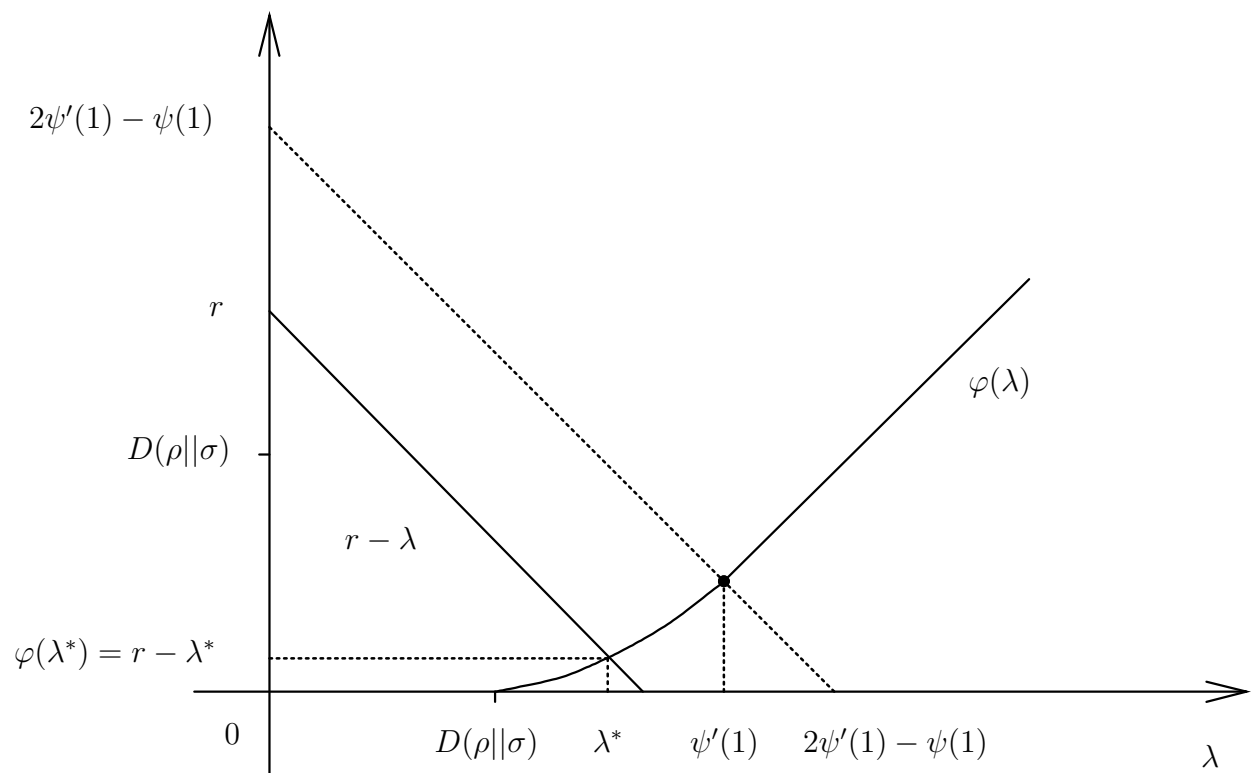
Figure 1: The graph of $\psi(s)$

Figure 2: The graph of $\varphi(\lambda)$