SUGAKU EXPOSITIONS
Volume 30, Number 1, June 2017, Pages 35–63
http://dx.doi.org/10.1090/suga/417
Article electronically published on March 17, 2017

# MATRIX ANALYSIS AND QUANTUM INFORMATION

FUMIO HIAI

Abstract. This survey article treats, by highlighting the role of matrix analysis, some topics from quantum information, mostly related to quantum (relative) entropies and divergences.

## Introduction

This article is a survey on (relative) entropy and its connections with several topics from the area of quantum information. What we should stress here is that a mathematical discipline called matrix analysis plays an important role both theoretically and technically in the study of quantum information. In Section 1, as an introduction to quantum information, we recall two famous inequalities for the von Neumann entropy and also a related topic on entangled states. In Section 2 we show recent results on the monotonicity and its equality conditions for quantum $f$-divergences that are generalizations of the relative entropy. In Section 3 we discuss several distances such as fidelity used in the state discrimination and their relations with quantum hypothesis testing. Section 4 is an appendix on some results from matrix analysis used in the main body of the article, which might be of independent interest as materials from matrix analysis. The present article covers only a few topics in the broad area of quantum information as well as it treats very limited topics in matrix analysis, so the title might be too large though simple and clear. The monographs [54, 67, 24] are good sources on the present status of quantum information, and the reader may consult [12, 13, 27] on the subject matter of matrix analysis.

Throughout the article, we assume that a quantum system $\mathcal{A}$ is given over a finite-dimensional Hilbert space $\mathcal{H}$; more precisely, $\mathcal{A}$ is the ($C^*$-)algebra $B(\mathcal{H})$ of all linear operators (automatically bounded due to the finite-dimensionality) on $\mathcal{H}$. Slightly more generally, $\mathcal{A}$ can be a $*$-subalgebra of $B(\mathcal{H})$ (i.e., $\mathcal{A}$ is a finite-dimensional $C^*$-algebra) containing the identity operator $I$. However, this generalization is rather formal and not so meaningful in the study of quantum information. When $d = \dim \mathcal{H}$, $B(\mathcal{H})$ is identified with the $d \times d$ matrix algebra $M_d(\mathbb{C})$, so hereafter we always assume that a quantum system is given in the framework of $\mathcal{A} = B(\mathcal{H}) = M_d(\mathbb{C})$.

This article originally appeared in Japanese in Sūgaku **65** (2013), no. 2, 133–159.
2010 *Mathematics Subject Classification.* Primary 15A45, 94A17; Secondary 46L53.
*Key words and phrases.* von Neumann entropy, relative entropy, strong subadditivity, entangled state, quasi-entropy, $f$-divergence, fidelity, Chernoff distance, Hoeffding distance, operator monotone function, operator convex function, completely positive map, Schwarz map, decomposable map.

We write Tr for the canonical (i.e., $\mathrm{Tr}\, E = 1$ for one-dimensional projections $E \in \mathcal{A}$) trace on $\mathcal{A}$; then $\mathcal{A}$ becomes a Hilbert space with respect to the Hilbert-Schmidt inner product $\langle X, Y\rangle_{\mathrm{HS}} := \mathrm{Tr}\, X^*Y$. (We adopt the convention that the inner product is conjugate-linear in the first component.) The absolute value of an $X \in \mathcal{A}$ is $|X| := (X^*X)^{1/2}$. When $f$ is a (real continuous) function on an interval $J$ of the real numbers, $f(A) \in \mathcal{A}$ is the usual functional calculus of a self-adjoint operator $A = A^*$ in $\mathcal{A}$ whenever the eigenvalues of $A$ are in $J$. We write $\mathcal{A}^+$ for the positive operators (or positive semidefinite matrices) in $\mathcal{A}$, and $\mathcal{A}^{++}$ for the invertible positive operators (or positive definite matrices) in $\mathcal{A}$. A state $\varphi$ on $\mathcal{A}$ is a positive linear functional $\varphi : \mathcal{A} \to \mathbb{C}$ with $\varphi(I) = 1$, which is uniquely represented as $\varphi(X) = \mathrm{Tr}\,\rho X$ ($X \in \mathcal{A}$) by a density matrix (or density operator) $\rho$ in $\mathcal{A}$. Below, a state $\varphi$ is always identified with its density matrix $\rho$.

We denote by $\mathcal{S} = \mathcal{S}(\mathcal{H})$ the set of all density matrices on a finite-dimensional Hilbert space $\mathcal{H}$. Various norms can be defined on $\mathcal{A} = B(\mathcal{H})$, such as the operator norm $\|X\|_\infty := \max\{\|X\xi\| : \xi \in \mathcal{H},\ \|\xi\| = 1\}$ and the Schatten $p$-norm $\|X\|_p := (\mathrm{Tr}\,|X|^p)^{1/p}$ ($1 \le p < \infty$). As for distances between density matrices $\rho, \sigma \in \mathcal{S}$, it is most natural to use the distance $\|\rho - \sigma\|_1$ via the trace-norm (= the Schatten 1-norm).

## 1. Entropy inequalities

Entropy is the most fundamental quantity in information theory. A basic idea in information theory is that entropy represents the quantity of uncertainty or ambiguity, whose quantity is gained when the uncertainty is cleared up. This section deals with two old and famous topics on the von Neumann entropy, the Fannes inequality and the strong subadditivity, whose developments up to recent investigations will be explained. A related topic on entangled states will be treated as well, which has been significant in the development of quantum information theory. Various types of entropies, including the von Neumann entropy, show up in quantum information theory. But the relative entropy, sometimes called the "mother entropy", is mathematically the most basic one, as will be seen from some essential arguments based on the relative entropy in this section.

### 1.1. **von Neumann entropy and relative entropy.**

The function $\eta(x) := -x\log x$ ($x \ge 0$) is often called the *information function*. The *von Neumann entropy* of a density matrix $\rho \in \mathcal{S}$ is defined as

$$S(\rho) := \mathrm{Tr}\,\eta(\rho) = -\mathrm{Tr}\,\rho\log\rho.$$

When $\lambda_1, \dots, \lambda_d$ are the eigenvalues of $\rho$ with multiplicities, $S(\rho)$ is nothing but the *Shannon entropy* of the finite probability distribution $(\lambda_1, \dots, \lambda_n)$. Since the information function $\eta$ is continuous and concave (furthermore, operator concave, see Section 4.1) on $[0, \infty)$, the von Neumann entropy $S(\rho)$ is continuous and concave on $\mathcal{S}(\mathcal{H})$. A quite useful estimate of the von Neumann entropy from both sides is

$$(1.1)\qquad (1-t)S(\rho) + tS(\sigma) \le S((1-t)\rho + t\sigma) \le (1-t)S(\rho) + tS(\sigma) + h(t)$$

for every $\rho, \sigma \in \mathcal{S}$ and $t \in [0, 1]$, where $h(t) := -t\log t - (1-t)\log(1-t)$ ($0 \le t \le 1$). The lower estimate of $S((1-t)\rho + t\sigma)$ is the concavity of $S(\rho)$ but adding $h(t)$ ($\le \log 2$) gives the upper estimate. (See [59, Proposition 1.6] for the proof of (1.1).)

On the other hand, the *relative entropy* of two $\rho, \sigma \in \mathcal{S}$ is defined as

$$S(\rho\|\sigma) := \begin{cases} \operatorname{Tr}\rho(\log\rho - \log\sigma) & \text{if } \operatorname{supp}\rho \le \operatorname{supp}\sigma, \\ +\infty & \text{if } \operatorname{supp}\rho \not\le \operatorname{supp}\sigma, \end{cases} \tag{1.2}$$

where $\operatorname{supp}\rho$ is the support projection of $\rho$, i.e., the orthogonal projection onto the range space (or the orthogonal complement of the null space) of $\rho$. When $\rho$ or $\sigma$ is not invertible, the above expression $\operatorname{Tr}\rho(\log\rho - \log\sigma)$ is a bit ambiguous since $\log\rho$ or $\log\sigma$ itself is not well defined. To avoid this, one can define as $\operatorname{Tr}\rho(\log^*\rho - \log^*\sigma)$ with $\log^* x := \log x$ $(x > 0)$, $\log^* 0 := 0$. In particular, when $\rho$ and $\sigma$ are diagonal matrices with diagonal entries $\lambda_1, \dots, \lambda_n$ and $\mu_1, \dots, \mu_n$, respectively, $S(\rho\|\sigma)$ coincides with the *Kullback-Leibler divergence* $\sum_{i=1}^n \lambda_i \log(\lambda_i/\mu_i)$, useful in the classical information theory. For general $A, B \in \mathcal{A}^+$ the relative entropy $S(A\|B)$ can be defined by just replacing $\rho, \sigma$ with $A, B$ in (1.2). The relative entropy was first introduced by Umegaki [75] for normal states on a semifinite von Neumann algebra, whose definition is the same as (1.2) where Tr is replaced with a semifinite normal trace. In 1976 Araki [6] extended the relative entropy to the general von Neumann algebra setting by using the concept of the relative modular operator and he showed properties of the relative entropy such as positivity, lower semicontinuity, joint convexity, monotonicity and martingale convergence. Later, Kosaki [43] gave the variational expression of the relative entropy to simplify the proofs of the above properties. A stronger version of positivity is the *Pinsker-Csiszár inequality*

$$2S(\rho\|\sigma) \ge \|\rho - \sigma\|_1^2 \tag{1.3}$$

shown in [32].

1.2. **Fannes inequality and Alicki-Fannes inequality.** An important inequality of $S(\rho)$, used in quantum information theory, is the *Fannes inequality* [18] saying that if $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ satisfy $\varepsilon := \|\rho - \sigma\|_1 \le 1/e$, then

$$|S(\rho) - S(\sigma)| \le \varepsilon \log d + \eta(\varepsilon), \tag{1.4}$$

where $d := \dim\mathcal{H}$. This means that $S(\rho)$ is almost Lipschitz continuous with respect to the trace-norm distance, giving a good estimate for the uniform continuity of $S(\rho)$. The next theorem due to Audenaert refines inequality (1.4) to a sharp inequality. The proof of (1.5) is much harder than that of (1.4).

**Theorem 1.1** ([8]). *For every $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, if $\varepsilon := \|\rho - \sigma\|_1$, then*

$$|S(\rho) - S(\sigma)| \le \frac{\varepsilon}{2}\log(d-1) + h(\varepsilon/2). \tag{1.5}$$

The bipartite quantum system combining two quantum systems $\mathcal{A}_1 = B(\mathcal{H}_1)$ and $\mathcal{A}_2 = B(\mathcal{H}_2)$ is the ($C^*$-algebra) tensor product $\mathcal{A}_{12} := \mathcal{A}_1 \otimes \mathcal{A}_2$, which is also represented as $\mathcal{A}_{12} = B(\mathcal{H}_1 \otimes \mathcal{H}_2)$ over the Hilbert space tensor product $\mathcal{H}_1 \otimes \mathcal{H}_2$. For simplicity we denote by the same Tr the canonical traces on $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{A}_{12}$, etc. Thus, Tr on $\mathcal{A}_{12}$ becomes the tensor product of the two Tr's on $\mathcal{A}_1$ and on $\mathcal{A}_2$. Then, the *partial trace* $X_1 = \operatorname{Tr}_2 X \in \mathcal{A}_1$ of $X \in \mathcal{A}_{12}$ is defined by

$$\operatorname{Tr} AX_1 = \operatorname{Tr}(A \otimes I_2)X, \qquad A \in \mathcal{A}_1.$$

That is, $\operatorname{Tr}_2 : \mathcal{A}_{12} \to \mathcal{A}_1$ is the dual map of the injection $\iota : \mathcal{A}_1 \to \mathcal{A}_{12}$, $\iota(A) := A \otimes I_2$, and $(\operatorname{Tr} I_2)^{-1}\operatorname{Tr}_2$ is the conditional expectation from $\mathcal{A}_{12}$ onto $\mathcal{A}_1$ (more precisely, $\mathcal{A}_1 \otimes I_2$) with respect to Tr. Similarly, $\operatorname{Tr}_1 : \mathcal{A}_{12} \to \mathcal{A}_2$ is defined. Here,

when $\mathcal{A}_0$ is a $*$-subalgebra of $\mathcal{A}$ containing $I$, the *conditional expectation* from $\mathcal{A}$ onto $\mathcal{A}_0$ with respect to Tr is the map $\mathcal{E}:\mathcal{A}\to\mathcal{A}_0$ defined by $\operatorname{Tr}\mathcal{E}(A)B=\operatorname{Tr}AB$ ($A\in\mathcal{A}$, $B\in\mathcal{A}_0$). When regarded as an operator on the Hilbert space $(\mathcal{A},\langle\cdot,\cdot\rangle_{\mathrm{HS}})$, $\mathcal{E}$ is the orthogonal projection from $\mathcal{A}$ onto $\mathcal{A}_0$.

To define the conditional entropy in quantum systems, we first recall the case in classical systems. Classical systems $\mathcal{A}_1$ and $\mathcal{A}_2$ are given by finite subsets $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively. For a probability distribution $p_{12}$ on the product set $\mathcal{X}_1\times\mathcal{X}_2$, the conditional entropy of $p_{12}$ (conditioned to $\mathcal{A}_2$) is defined as

$$\begin{aligned} S(p_{12}|\mathcal{A}_2) &:= \sum_{b\in\mathcal{X}_2} p_2(b)S(p_1(\cdot|b)) = \sum_{(a,b)\in\mathcal{X}_1\times\mathcal{X}_2} p_{12}(a,b)\log\frac{p_2(b)}{p_{12}(a,b)} \\ &= S(p_{12})-S(p_2), \end{aligned} \tag{1.6}$$

where $p_2$ is the marginal distribution of $p_{12}$ onto $\mathcal{X}_2$, and $p_1(a|b):=p_{12}(a,b)/p_2(b)$ is the conditional probability. Now, for a density matrix $\rho_{12}\in\mathcal{S}(\mathcal{H}_1\otimes\mathcal{H}_2)$, the *conditional entropy* of $\rho_{12}$ (conditioned to $\mathcal{A}_2=B(\mathcal{H}_2)$) is defined, similarly to the last expression of (1.6), as

$$S(\rho_{12}|\mathcal{A}_2) := S(\rho_{12})-S(\rho_2), \tag{1.7}$$

where $\rho_2:=\operatorname{Tr}_1\rho_{12}$ (i.e., $\rho_2$ is the restriction of the state $\rho_{12}$ onto $\mathcal{A}_2$). Here, note that the conditional entropy (1.7) can be both positive and negative while (1.6) in the classical case is always non-negative. Indeed, if $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent with respect to $\rho_{12}$ (i.e., $\rho_{12}=\rho_1\otimes\rho_2$), then $S(\rho_{12})=S(\rho_1)+S(\rho_2)$ (additive) so that $S(\rho_{12}|\mathcal{A}_2)=S(\rho_1)\geq 0$. On the other hand, if $\rho_{12}$ is a one-dimensional projection (i.e., $\rho_{12}$ is a pure state on $\mathcal{A}_{12}$), then $S(\rho_{12})=0$ so that $S(\rho_{12}|\mathcal{A}_2)=-S(\rho_2)$ is negative as far as $\rho_2$ is not pure. For example, when $\{e_1,e_2\}$ is an orthonormal basis of two-dimensional $\mathcal{H}_1=\mathcal{H}_2$, let

$$\xi := \frac{1}{\sqrt{2}}(e_1\otimes e_2+e_2\otimes e_1) \tag{1.8}$$

and $\rho_{12}:=|\xi\rangle\langle\xi|$, that is, a one-dimensional projection given by $(|\xi\rangle\langle\xi|)\eta:=\langle\xi,\eta\rangle\xi$. Then $\rho_1:=\operatorname{Tr}_2\rho_{12}$ and $\rho_2:=\operatorname{Tr}_1\rho_{12}$ are both equal to $I/2$ so that $S(\rho_2)=\log 2$. This $\rho_{12}$ is called the *Bell state*, which is a typical example of entangled states (as discussed in Section 1.4). Furthermore, when $\rho_{12}$ is a pure state, the positive eigenvalues of $\rho_1$ and $\rho_2$ are the same (including multiplicities), and hence $S(\rho_1)=S(\rho_2)$. Also, note that the *triangle inequality*

$$|S(\rho_1)-S(\rho_2)|\leq S(\rho_{12})\leq S(\rho_1)+S(\rho_2) \tag{1.9}$$

holds for a general state $\rho_{12}$ ([7]). The right-hand side inequality of (1.9) is the so-called *subadditivity*, which is nothing but the restatement of the non-negativity of the *mutual information*

$$I_{\rho_{12}}(\mathcal{A}_1,\mathcal{A}_2) := S(\rho_{12}\|\rho_1\otimes\rho_2)$$

of $\mathcal{A}_1,\mathcal{A}_2$ with respect to $\rho_{12}$. Thanks to (1.3), the equality condition for the subadditivity is $\rho_{12}=\rho_1\otimes\rho_2$ (i.e., the independence of $\mathcal{A}_1,\mathcal{A}_2$ with respect to $\rho_{12}$).

Alicki and Fannes [2] proved an inequality of Fannes type for the conditional entropy as follows: If $\rho_{12},\sigma_{12}\in\mathcal{S}(\mathcal{H}_1\otimes\mathcal{H}_2)$ satisfies $\varepsilon:=\|\rho_{12}-\sigma_{12}\|_1\leq 1$, then

$$|S(\rho_{12}|\mathcal{A}_2)-S(\sigma_{12}|\mathcal{A}_2)|\leq 4\varepsilon\log d_1+2h(\varepsilon), \tag{1.10}$$

where $d_1:=\dim\mathcal{H}_1$. The next result is a slight improvement of (1.10).

**Proposition 1.2.** *Assume that a function $F : \mathcal{S}(\mathcal{H}) \to \mathbb{C}$ satisfies*

$$|F((1-\varepsilon)\rho + \varepsilon\sigma) - (1-\varepsilon)F(\rho) - \varepsilon F(\sigma)| \le h(\varepsilon)$$

*for every $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ and $\varepsilon \in (0,1)$. Then, for any $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, if*

$$\varepsilon := \|\rho - \sigma\|_1/(2 + \|\rho - \sigma\|_1),$$

*then*

$$|F(\rho) - F(\sigma)| \le 4\varepsilon M + 2h(\varepsilon)$$

*holds, where $M := \sup_{\rho \in \mathcal{S}(\mathcal{H})} |F(\rho)|$.*

*Proof.* Let $\rho - \sigma = (\rho-\sigma)_+ - (\rho-\sigma)_-$ be the Jordan decomposition for a matrix (or operator) $\rho - \sigma$. Since $\|(\rho-\sigma)_+\|_1 = \|(\rho-\sigma)_-\|_1 = \frac{1}{2}\|\rho-\sigma\|_1$,

$$\tau^* := \frac{\rho + (\rho-\sigma)_-}{1 + \frac{1}{2}\|\rho-\sigma\|_1} = \frac{\sigma + (\rho-\sigma)_+}{1 + \frac{1}{2}\|\rho-\sigma\|_1}, \quad \tau_1 := \frac{(\rho-\sigma)_-}{\frac{1}{2}\|\rho-\sigma\|_1}, \quad \tau_2 := \frac{(\rho-\sigma)_+}{\frac{1}{2}\|\rho-\sigma\|_1},$$

belong to $\mathcal{S}(\mathcal{H})$. For $\varepsilon := \|\rho-\sigma\|_1/(2+\|\rho-\sigma\|_1)$, since $\tau^* = (1-\varepsilon)\rho + \varepsilon\tau_1 = (1-\varepsilon)\sigma + \varepsilon\tau_2$ (such $\{\tau_1, \tau_2\}$ is called a Helström family [41]), we have

$$\begin{aligned}
|F(\rho) - F(\sigma)| &\le |F(\tau^*) - F(\rho)| + |F(\tau^*) - F(\sigma)| \\
&\le |F(\tau^*) - (1-\varepsilon)F(\rho) - \varepsilon F(\tau_1)| + |F(\tau^*) - (1-\varepsilon)F(\sigma) - \varepsilon F(\tau_2)| \\
&\quad + \varepsilon(|F(\rho)| + |F(\tau_1)| + |F(\sigma)| + |F(\tau_2)|) \\
&\le 2h(\varepsilon) + 4\varepsilon M.
\end{aligned}$$

□

In the case of the conditional entropy $S(\rho_{12}|\mathcal{A}_2)$ ($\rho_{12} \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$), since it is represented as

$$S(\rho_{12}|\mathcal{A}_2) = -S(\rho_{12}\|I_1 \otimes \rho_2)$$

(though $I_1 \otimes \rho_2 \in \mathcal{A}_{12}^+$ is not a density matrix), the joint convexity of the relative entropy (see Theorem 2.4 and Example 2.5 in the next section) implies that $S(\rho_{12}|\mathcal{A}_2)$ is a concave function of $\rho_{12}$. Hence, for every $\rho_{12}, \sigma_{12} \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ and $\varepsilon \in (0,1)$,

$$S((1-\varepsilon)\rho_{12} + \varepsilon\sigma_{12}|\mathcal{A}_2) \ge (1-\varepsilon)S(\rho_{12}|\mathcal{A}_2) + \varepsilon S(\sigma_{12}|\mathcal{A}_2).$$

Moreover, from the estimates from both sides in (1.1),

$$\begin{aligned}
&S((1-\varepsilon)\rho_{12} + \varepsilon\sigma_{12}|\mathcal{A}_2) \\
&\quad \le \{(1-\varepsilon)S(\rho_{12}) + \varepsilon S(\sigma_{12}) + h(\varepsilon)\} - \{(1-\varepsilon)S(\rho_2) + \varepsilon S(\sigma_2)\} \\
&\quad = (1-\varepsilon)S(\rho_{12}|\mathcal{A}_2) + \varepsilon S(\sigma_{12}|\mathcal{A}_2) + h(\varepsilon),
\end{aligned}$$

and therefore the assumption of Proposition 1.2 is satisfied. Since $M = \log d_1$ due to (1.9), inequality (1.10) holds true with $\varepsilon = \|\rho_{12} - \sigma_{12}\|_1/(2 + \|\rho_{12} - \sigma_{12}\|_1)$.

The following is also a corollary of Proposition 1.2, which slightly improves the estimate given in [70].

**Corollary 1.3.** *Let $\mathcal{D}$ be a convex subset of $\mathcal{S}(\mathcal{H})$ containing $I/d$ (the maximally mixed state). Define the relative entropy distance from $\rho \in \mathcal{S}(\mathcal{H})$ to $\mathcal{D}$ by*

$$E_{\mathrm{re}}^{\mathcal{D}}(\rho) := \inf\{S(\rho\|\tau) : \tau \in \mathcal{D}\}. \tag{1.11}$$

*Then, for every $\rho, \sigma \in \mathcal{S}(\mathcal{H})$,*

$$|E_{\mathrm{re}}^{\mathcal{D}}(\rho) - E_{\mathrm{re}}^{\mathcal{D}}(\sigma)| \le 4\varepsilon \log d + 2h(\varepsilon),$$

*where* $\varepsilon := \|\rho-\sigma\|_1/(2+\|\rho-\sigma\|_1)$.

*Proof.* For every $\rho,\sigma\in\mathcal{S}(\mathcal{H})$, $\varepsilon\in(0,1)$, and $\tau_1,\tau_2\in\mathcal{D}$, we have

$$E_{\rm re}^{\mathcal{D}}((1-\varepsilon)\rho+\varepsilon\sigma)\le S((1-\varepsilon)\rho+\varepsilon\sigma\|(1-\varepsilon)\tau_1+\varepsilon\tau_2)\le(1-\varepsilon)S(\rho\|\tau_1)+\varepsilon S(\sigma\|\tau_2)$$

so that

$$E_{\rm re}^{\mathcal{D}}((1-\varepsilon)\rho+\varepsilon\sigma)\le(1-\varepsilon)E_{\rm re}^{\mathcal{D}}(\rho)+\varepsilon E_{\rm re}^{\mathcal{D}}(\sigma). \tag{1.12}$$

On the other hand, if $\tau\in\mathcal{D}$ satisfies $S((1-\varepsilon)\rho+\varepsilon\sigma\|\tau)<+\infty$, then from (1.1)

$$\begin{aligned}S((1-\varepsilon)\rho+\varepsilon\sigma\|\tau)&=-S((1-\varepsilon)\rho+\varepsilon\sigma)-(1-\varepsilon)\mathrm{Tr}\,\rho\log\tau-\varepsilon\mathrm{Tr}\,\sigma\log\tau\\&\ge-(1-\varepsilon)S(\rho)-\varepsilon S(\sigma)-h(\varepsilon)-(1-\varepsilon)\mathrm{Tr}\,\rho\log\tau-\varepsilon\mathrm{Tr}\,\sigma\log\tau\\&=(1-\varepsilon)S(\rho\|\tau)+\varepsilon S(\sigma\|\tau)-h(\varepsilon)\\&\ge(1-\varepsilon)E_{\rm re}^{\mathcal{D}}(\rho)+\varepsilon E_{\rm re}^{\mathcal{D}}(\sigma)-h(\varepsilon).\end{aligned}$$

Therefore,

$$E_{\rm re}^{\mathcal{D}}((1-\varepsilon)\rho+\varepsilon\sigma)\ge(1-\varepsilon)E_{\rm re}^{\mathcal{D}}(\rho)+\varepsilon E_{\rm re}^{\mathcal{D}}(\sigma)-h(\varepsilon). \tag{1.13}$$

Since (1.12) and (1.13) say that $E_{\rm re}^{\mathcal{D}}(\rho)$ satisfies the assumption of Proposition 1.2, the required inequality follows. □

1.3. **Strong subadditivity of von Neumann entropy.** In the tripartite quantum system $\mathcal{A}_{123}=\mathcal{A}_1\otimes\mathcal{A}_2\otimes\mathcal{A}_3=B(\mathcal{H}_1\otimes\mathcal{H}_2\otimes\mathcal{H}_3)$ one can define $\mathrm{Tr}_3:\mathcal{A}_{123}\to\mathcal{A}_{12}$, $\mathrm{Tr}_1:\mathcal{A}_{123}\to\mathcal{A}_{23}$, $\mathrm{Tr}_{13}:\mathcal{A}_{123}\to\mathcal{A}_2$, etc. as in the bipartite case. The most famous inequality concerning the von Neumann entropy is the *strong subadditivity* due to Lieb and Ruskai.

**Theorem 1.4** ([48])**.** *For every* $\rho_{123}\in\mathcal{S}(\mathcal{H}_1\otimes\mathcal{H}_2\otimes\mathcal{H}_3)$ *let* $\rho_{12}:=\mathrm{Tr}_3\,\rho_{123}$, $\rho_{23}:=\mathrm{Tr}_1\,\rho_{123}$, *and* $\rho_2:=\mathrm{Tr}_{13}\,\rho_{123}$. *Then*

$$S(\rho_{123})+S(\rho_2)\le S(\rho_{12})+S(\rho_{23}). \tag{1.14}$$

This can easily be proved by using the relative entropy. First, note that

$$\begin{aligned}&S(\rho_{123}\|\rho_1\otimes\rho_{23})-S(\rho_{12}\|\rho_1\otimes\rho_2)\\&\quad=\mathrm{Tr}\,\rho_{123}(\log\rho_{123}-\log\rho_1-\log\rho_{23})-\mathrm{Tr}\,\rho_{12}(\log\rho_{12}-\log\rho_1-\log\rho_2)\\&\quad=-S(\rho_{123})+S(\rho_{23})+S(\rho_{12})-S(\rho_2).\end{aligned}$$

In the above, for simplicity we have written, for instance, $\log\rho_1$ for $\log\rho_1\otimes I_{23}$. Since $\rho_{12}=\mathrm{Tr}_3\,\rho_{123}$ and $\rho_1\otimes\rho_2=\mathrm{Tr}_3(\rho_1\otimes\rho_{23})$, the monotonicity of the relative entropy (see Theorem 2.2 or rather Theorem 2.7 with Example 2.5) implies that

$$S(\rho_{123}\|\rho_1\otimes\rho_{23})\ge S(\rho_{12}\|\rho_1\otimes\rho_2), \tag{1.15}$$

from which (1.14) follows.

When $\mathcal{A}_2=\mathbb{C}$ and $\mathcal{A}_3$ is replaced with $\mathcal{A}_2$, the strong subadditivity reduces to the subadditivity of $\rho_{12}\in\mathcal{S}(\mathcal{H}_1\otimes\mathcal{H}_2)$ in (1.9), and its equality condition is $\rho_{12}=\rho_1\otimes\rho_2$. On the other hand, a long-standing problem on the equality condition for the strong subadditivity in (1.14) was settled by Petz [65, 66], Ruskai [68], and Hayden et al. [26] as follows. Here, note that the equality case of (1.14) is equivalent to that of the monotonicity of the relative entropy in (1.15).

**Theorem 1.5** ([68, 66, 26])**.** *Let* $\mathcal{H}_j$ $(j=1,2,3)$ *be finite-dimensional Hilbert spaces and let* $\rho_{123}\in\mathcal{S}(\mathcal{H}_1\otimes\mathcal{H}_2\otimes\mathcal{H}_3)$. *Then the following conditions are equivalent:*

(i) *equality holds in the strong subadditivity in* (1.14)*;*

(ii) $\log\rho_{123} - \log\rho_{12} = \log\rho_{23} - \log\rho_2$;
(iii) *there are a decomposition of* $\mathcal{H}_2$

$$\mathcal{H}_2 = \bigoplus_{i=1}^{k} \mathcal{H}_i^L \otimes \mathcal{H}_i^R,$$

*density operators* $\rho_{1i}^L \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_i^L)$, $\rho_{i3}^R \in \mathcal{S}(\mathcal{H}_i^R \otimes \mathcal{H}_3)$, *and a probability vector* $(p_1, \dots, p_k)$ *such that*

$$\rho_{123} = \bigoplus_{i=1}^{k} p_i \rho_{1i}^L \otimes \rho_{i3}^R.$$

(i) $\Leftrightarrow$ (ii) was proved in [68, 66] and (i) $\Leftrightarrow$ (iii) was in [26]. The equality condition for the monotonicity of the relative entropy (corresponding to sufficient statistics in classical probability theory) was proved in [66], which played an essential role in the proof in [26]. In Section 2 we shall state recent results in [31] extending the monotonicity of the relative entropy and its equality condition.

It is worth noting that the same construction as (iii) above appeared in the characterization theorem in [1, 58] for quantum Markov states of the one-dimensional spin chain (infinite tensor product $C^*$-algebra) $\bigotimes_1^\infty M_d(\mathbb{C})$. Indeed, condition (iii) can be considered as a definition of Markovianity for short tripartite chains $\mathcal{A}_1 \otimes \mathcal{A}_2 \otimes \mathcal{A}_3$.

1.4. **Entangled states.** In this section we are given a bipartite quantum system $\mathcal{A}_{12} = \mathcal{A}_1 \otimes \mathcal{A}_2 = B(\mathcal{H}_1 \otimes \mathcal{H}_2)$. A state (density operator) $\rho \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ is said to be *separable* if $\rho$ is represented as

$$\rho = \sum_{i=1}^{k} A_{1i} \otimes A_{2i} \qquad (A_{1i} \in \mathcal{A}_1^+,\ A_{2i} \in \mathcal{A}_2^+), \tag{1.16}$$

that is, $\rho$ is a convex combination of tensor product states $\rho_1 \otimes \rho_2$ ($\rho_1 \in \mathcal{S}(\mathcal{H}_1)$, $\rho_2 \in \mathcal{S}(\mathcal{H}_2)$). Therefore, the set $\mathcal{S}_{\rm sep}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ of all separable states is a closed convex subset of $\mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ given as

$$\mathcal{S}_{\rm sep}(\mathcal{H}_1 \otimes \mathcal{H}_2) = \mathrm{conv}\{\rho_1 \otimes \rho_2 : \rho_1 \in \mathcal{S}(\mathcal{H}_1),\ \rho_2 \in \mathcal{S}(\mathcal{H}_2)\}$$

(conv stands for convex hull). If $\rho \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ does not belong to $\mathcal{S}_{\rm sep}(\mathcal{H}_1 \otimes \mathcal{H}_2)$, then $\rho$ is called an *entangled state*. In particular, a pure state $\rho = |\xi\rangle\langle\xi|$ is separable only if $\xi \in \mathcal{H}_1 \otimes \mathcal{H}_2$ is of the form $\xi = \xi_1 \otimes \xi_2$ ($\xi \in \mathcal{H}_1$, $\xi_2 \in \mathcal{H}_2$). Therefore, $|\xi\rangle\langle\xi|$ is an entangled state unless $\xi$ is of this form.

Entangled states play important roles in quantum computation and quantum communication; however, we do not touch this direction here. But from the point of view of mathematics we are interested in criterions for the entanglement and measures for the degree of entanglement. Some simple criterions for entangled states have been known so far as well as equivalent restatements of the entanglement property. However, those simple criterions are not so efficient when $\dim\mathcal{H}_1$ and $\dim\mathcal{H}_2$ are not small (see [79] for example). Here we give a brief exposition on the *PPT* (*positive partial transpose*) criterion due to Peres [62] and Horodecki[3] [35]. Define the transpose map $TA := A^t$ ($A \in \mathcal{A}_1$) under identification of $\mathcal{A}_1 = B(\mathcal{H}_1)$ with $M_{d_1}(\mathbb{C})$, and consider the partial transpose map $T \otimes \mathrm{id}$ on $\mathcal{A}_1 \otimes \mathcal{A}_2$. Since $(T \otimes \mathrm{id})\rho = \sum_{i=1}^k A_{1i}^t \otimes A_{2i} \geq 0$ for $\rho$ in (1.16), it is obvious that $(T \otimes \mathrm{id})\rho \geq 0$

whenever $\rho \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ is separable. Therefore, if $(T \otimes \mathrm{id})\rho \geq 0$ does not hold, then it is shown that $\rho$ is an entangled state (the converse is of course not true).

The PPT criterion is closely related to the decomposability of linear maps $\Phi : \mathcal{A}_1 \to \mathcal{A}_2$ (see Section 4.2). Now, let us consider the following two convex subcones of the $C^*$-algebraic positive cone $(\mathcal{A}_1 \otimes \mathcal{A}_2)^+$ of $\mathcal{A}_1 \otimes \mathcal{A}_2$:

$$\mathcal{V}_{\mathrm{sep}} := \{\alpha\rho : \rho \in \mathcal{S}_{\mathrm{sep}}(\mathcal{H}_1 \otimes \mathcal{H}_2),\ \alpha \geq 0\}, \tag{1.17}$$

$$\mathcal{V}_{\mathrm{PPT}} := \{A \in \mathcal{A}_1 \otimes \mathcal{A}_2 : A \geq 0,\ (T \otimes \mathrm{id})A \geq 0\}. \tag{1.18}$$

The PPT criterion is based on the fact that $\mathcal{V}_{\mathrm{sep}} \subset \mathcal{V}_{\mathrm{PPT}}$, and it becomes perfect if $\mathcal{V}_{\mathrm{sep}} = \mathcal{V}_{\mathrm{PPT}}$ would hold. To compare the two subcones, we look at their dual cones. First, note that there is a one-to-one correspondence between linear functionals $\phi$ on $\mathcal{A}_1 \otimes \mathcal{A}_2$ and linear maps $\Phi : \mathcal{A}_1 \to \mathcal{A}_2$ determined by

$$\phi(A_1 \otimes A_2) = \mathrm{Tr}\,\Phi(A_1^t)A_2, \qquad A_1 \in \mathcal{A}_1,\ A_2 \in \mathcal{A}_2.$$

Then, $\phi$ is represented by the Choi matrix of $\Phi$ (see Section 4.2) as follows:

$$\phi(A_1 \otimes A_2) = \mathrm{Tr}\left(\sum_{i,j} E_{ij} \otimes \Phi(E_{ij})\right)(A_1 \otimes A_2), \qquad A_1 \in \mathcal{A}_1,\ A_2 \in \mathcal{A}_2.$$

Hence by Theorem 4.1, $\phi$ is positive on $(\mathcal{A}_1 \otimes \mathcal{A}_2)^+$ if and only if $\Phi$ is completely positive. Also, $\phi$ is positive on $(T \otimes \mathrm{id})(\mathcal{A}_1 \otimes \mathcal{A}_2)^+$ if and only if $\Phi$ is completely copositive, i.e., $\Phi \circ T$ is completely positive. Since $\mathcal{V}_{\mathrm{PPT}} = (\mathcal{A}_1 \otimes \mathcal{A}_2)^+ \cap (T \otimes \mathrm{id})(\mathcal{A}_1 \otimes \mathcal{A}_2)^+$, the bipolar theorem implies that $\phi$ is positive on $\mathcal{V}_{\mathrm{PPT}}$ if and only if $\Phi$ is a decomposable positive map. On the other hand, as easily verified, $\phi$ is positive on $\mathcal{V}_{\mathrm{sep}}$ if and only if $\Phi$ is a positive map. Therefore, when we describe the dual cones of (1.17) and (1.18) as sets of linear maps $\Phi : \mathcal{A}_1 \to \mathcal{A}_2$, they are the set of positive maps and the set of decomposable positive maps, respectively. Thus, the next proposition is exactly a restatement of Theorem 4.2.

**Proposition 1.6** ([78]). *Assume that* $\dim \mathcal{H}_1 \cdot \dim \mathcal{H}_2 \leq 6$. *Then* $\rho \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ *is separable if and only if* $(T \otimes \mathrm{id})\rho \geq 0$. *That is, the PPT criterion is perfect in this case. If* $\dim \mathcal{H}_1 \cdot \dim \mathcal{H}_2 > 6$, *then* $(T \otimes \mathrm{id})\rho \geq 0$ *is only necessary for the separability of* $\rho$.

**Example 1.7** ([76]). When $\dim \mathcal{H}_1 = \dim \mathcal{H}_2 = 2$ (called two qubits), the *Werner states* are

$$\rho_p := p|\xi\rangle\langle\xi| + (1-p)\tau_0 \qquad (0 \leq p \leq 1),$$

where $\xi$ is a vector given in (1.8) and $\tau_0$ is the maximally mixed state $I/4$. Since the $4 \times 4$ matrix representations of $\rho_p$ and $(T \otimes \mathrm{id})\rho_p$ are

$$\rho_p = \frac{1}{4}\begin{bmatrix} 1-p & 0 & 0 & 0 \\ 0 & 1+p & 2p & 0 \\ 0 & 2p & 1+p & 0 \\ 0 & 0 & 0 & 1-p \end{bmatrix},$$

$$(T \otimes \mathrm{id})\rho_p = \frac{1}{4}\begin{bmatrix} 1-p & 0 & 0 & 2p \\ 0 & 1+p & 0 & 0 \\ 0 & 0 & 1+p & 0 \\ 2p & 0 & 0 & 1-p \end{bmatrix}$$

it follows that $(T \otimes \mathrm{id})\rho_p \geq 0$ if and only if $(1-p)^2 - 4p^2 \geq 0$, i.e., $p \leq 1/3$. Hence Proposition 1.6 says that $\rho_p$ is an entangled state if (and only if) $p > 1/3$.

More discussions on the relation between the entangled states and various cones of positive linear maps are found, for example, in the recent expository paper [46] (and the references therein).

Among many other proposals here we discuss two typical entanglement measures (see, for example, [34, 70] for comprehensive accounts). First, we give a useful lemma for discussions on bipartite quantum systems.

**Lemma 1.8.** *For every unit vector $\xi \in \mathcal{H}_1 \otimes \mathcal{H}_2$ there exist orthonormal sets $\{\eta_{1i}\}_{i=1}^k \subset \mathcal{H}_1$, $\{\eta_{2i}\}_{i=1}^k \subset \mathcal{H}_2$, and a probability vector $(p_1, \dots, p_k)$ such that*

$$\xi = \sum_{i=1}^k \sqrt{p_i}\, \eta_{1i} \otimes \eta_{2i}. \tag{1.19}$$

*Proof.* Define a conjugate-linear map $\Lambda : \mathcal{H}_1 \to \mathcal{H}_2$ and its adjoint $\Lambda^* : \mathcal{H}_2 \to \mathcal{H}_1$ by

$$\langle \Lambda\zeta_1, \zeta_2\rangle = \langle \Lambda^*\zeta_2, \zeta_1\rangle = \langle \xi, \zeta_1 \otimes \zeta_2\rangle, \qquad \zeta_1 \in \mathcal{H}_1,\ \zeta_2 \in \mathcal{H}_2.$$

Let $\{f_j\}_{j=1}^{d_2}$ be an orthonormal basis of $\mathcal{H}_2$. For every $\zeta_1, \zeta_1' \in \mathcal{H}_1$,

$$\begin{aligned}
\operatorname{Tr} |\xi\rangle\langle\xi|(|\zeta_1\rangle\langle\zeta_1'| \otimes I_2) &= \sum_{j=1}^{d_2} \operatorname{Tr} |\xi\rangle\langle\xi|\, |\zeta_1 \otimes f_j\rangle\langle\zeta_1' \otimes f_j| \\
&= \sum_{j=1}^{d_2} \langle \xi, \zeta_1 \otimes f_j\rangle\langle\zeta_1' \otimes f_j, \xi\rangle = \sum_{j=1}^{d_2} \langle \Lambda\zeta_1, f_j\rangle\langle f_j, \Lambda\zeta_1'\rangle \\
&= \langle \Lambda\zeta_1, \Lambda\zeta_1'\rangle = \langle \zeta_1', \Lambda^*\Lambda\zeta_1\rangle = \operatorname{Tr} \Lambda^*\Lambda\, |\zeta_1\rangle\langle\zeta_1'|.
\end{aligned}$$

Hence, letting $\rho := |\xi\rangle\langle\xi|$ and $\rho_1 := \Lambda^*\Lambda$, we have $\rho_1 = \operatorname{Tr}_2 \rho$. Let

$$\rho_1 = \sum_{i=1}^k p_i |\eta_{1i}\rangle\langle\eta_{1i}|$$

be the spectral decomposition of $\rho_1$ where $p_i > 0$ and $\{\eta_{1i}\}_{i=1}^k$ is an orthonormal set, and define $\eta_{2i} := p_i^{-1/2}\Lambda\eta_{1i}$. Then we have $\langle \eta_{2i}, \eta_{2j}\rangle = (p_i p_j)^{-1/2}\langle \eta_{1j}, \Lambda^*\Lambda\eta_{1i}\rangle = \delta_{ij}$ so that $\{\eta_{2i}\}_{i=1}^k$ is orthonormal. Expand $\{\eta_{1i}\}_{i=1}^k$ and $\{\eta_{2i}\}_{i=1}^k$ into orthonormal bases $\{\eta_{1i}\}_{i=1}^{d_1}$ of $\mathcal{H}_1$ and $\{\eta_{2i}\}_{i=1}^{d_2}$ of $\mathcal{H}_2$, respectively. We have for $i, j = 1, \dots, k$

$$\langle \xi, \eta_{1i} \otimes \eta_{2j}\rangle = \langle \Lambda\eta_{1i}, \eta_{2j}\rangle = \frac{1}{\sqrt{p_j}}\langle \Lambda^*\Lambda\eta_{1i}, \eta_{1j}\rangle = \sqrt{p_i}\delta_{ij},$$

and moreover it is easy to see that $\langle \xi, \eta_{1i} \otimes \eta_{2j}\rangle = 0$ if $i > k$ or $j > k$. Hence (1.19) follows. □

Expression (1.19) is called the *Schmidt decomposition* of $\xi$. Then, the restrictions of the pure state $\rho = |\xi\rangle\langle\xi|$ to $\mathcal{A}_1$ and $\mathcal{A}_2$ are given as $\rho_1 = \sum_{i=1}^k p_i |\eta_{1i}\rangle\langle\eta_{1i}|$ and $\rho_2 = \sum_{i=1}^k p_i |\eta_{2i}\rangle\langle\eta_{2i}|$, respectively. Hence, as mentioned in Section 1.2 (just above (1.9)), $\rho_1$ and $\rho_2$ have the same positive eigenvalues. When $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}$ and $p_i = 1/d$ for $1 \le i \le d := \dim \mathcal{H}$ in the Schmidt decomposition of $\xi$, the pure state $|\xi\rangle\langle\xi|$ is called a *maximally entangled state*. In this case, $\rho_1$ and $\rho_2$ are maximally mixed states. The Bell state mentioned in Section 1.2 is a maximally entangled state.

The most natural entanglement measure is the *entanglement relative entropy* given by taking $\mathcal{D} = \mathcal{S}_{\rm sep}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ in (1.11) as

$$E_{\rm re}(\rho_{12}) := \inf\{S(\rho_{12}\|\tau) : \tau \in \mathcal{S}_{\rm sep}(\mathcal{H}_1 \otimes \mathcal{H}_2)\}, \qquad \rho_{12} \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2).$$

The following entanglement measure is called the *squashed entanglement*: for $\rho_{12} \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$,

$$E_{\rm sq}(\rho_{12}) := \frac{1}{2}\inf\{S(\rho_{13}|\mathcal{A}_3) - S(\rho_{123}|\mathcal{A}_{23}) : \rho_{123} \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3),\ \rho_{12} = {\rm Tr}_3\, \rho_{123}\}.$$

These two have the following common properties:

**Proposition 1.9.** *In either case of $E = E_{\rm re}$ and $E = E_{\rm sq}$ the following hold:*

(1) *$E$ is a non-negative convex function on $\mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$.*
(2) *If $\rho_{12} \in \mathcal{S}_{\rm sep}(\mathcal{H}_1 \otimes \mathcal{H}_2)$, then $E(\rho_{12}) = 0$.*
(3) *If $\rho_{12}$ is a pure state, then $E(\rho_{12}) = S(\rho_1) = S(\rho_2)$.*
(4) *$E(\rho_{12})$ is continuous in $\rho_{12}$.*

*Proof.* (1) It is obvious that $E_{\rm re}(\rho_{12}) \geq 0$. The convexity of $E_{\rm re}$ is easily seen from the joint convexity of the relative entropy. Strong subadditivity (1.14) implies that

$$S(\rho_{13}|\mathcal{A}_3) - S(\rho_{123}|\mathcal{A}_{23}) = S(\rho_{13}) + S(\rho_{23}) - S(\rho_{123}) - S(\rho_3) \geq 0$$

so that $E_{\rm sq}(\rho_{12}) \geq 0$. To prove the convexity of $E_{\rm sq}$, for every $\rho_{12}, \sigma_{12} \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ and $\lambda \in (0,1)$, let $\kappa_{12} := \lambda\rho_{12} + (1-\lambda)\sigma_{12}$ and choose any extensions $\rho_{123}$ of $\rho_{12}$ and $\sigma_{123}$ of $\sigma_{12}$ (here $\rho_{123}$ and $\sigma_{123}$ can be assumed to be on the same $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$). Moreover, choosing orthogonal pure states $|\xi\rangle\langle\xi|$ and $|\eta\rangle\langle\eta|$ on $\mathcal{H}_4$, define an extension of $\kappa_{12}$ as $\kappa_{1234} := \lambda\rho_{123} \otimes |\xi\rangle\langle\xi| + (1-\lambda)\sigma_{123} \otimes |\eta\rangle\langle\eta|$. Then a direct computation gives

$$\begin{aligned} &S(\kappa_{134}|\mathcal{A}_{34}) - S(\kappa_{1234}|\mathcal{A}_{234}) \\ &\quad = \lambda(S(\rho_{13}|\mathcal{A}_3) - S(\rho_{123}|\mathcal{A}_{23})) + (1-\lambda)(S(\sigma_{13}|\mathcal{A}_3) - S(\sigma_{123}|\mathcal{A}_{23})), \end{aligned}$$

from which $E_{\rm sq}(\kappa_{12}) \leq \lambda E_{\rm sq}(\rho_{12}) + (1-\lambda)E_{\rm sq}(\sigma_{12})$ follows.

(2) Assume that $\rho_{12} \in \mathcal{S}_{\rm sep}(\mathcal{H}_1 \otimes \mathcal{H}_2)$. Hence $E_{\rm re}(\rho_{12}) = 0$ is obvious. One can further decompose expression (1.16) to have a representation $\rho_{12} = \sum_{i=1}^m p_i\rho_{1i} \otimes \rho_{2i}$ with pure states $\rho_{1i} \in \mathcal{S}(\mathcal{H}_1)$ and $\rho_{2i} \in \mathcal{S}(\mathcal{H}_2)$. Moreover, choose mutually orthogonal pure states $\rho_{3i} \in \mathcal{S}(\mathcal{H}_3)$ $(1 \leq i \leq m)$ and define $\rho_{123} := \sum_{i=1}^m p_i\rho_{1i} \otimes \rho_{2i} \otimes \rho_{3i}$. Then we have $S(\rho_{13}) = S(\rho_{23}) = S(\rho_{123}) = S(\rho_3) = S(p)$ so that $E_{\rm sq}(\rho_{12}) = 0$.

(3) For any $\tau = \sum_i p_i\tau_{1i} \otimes \tau_{2i} \in \mathcal{S}_{\rm sep}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ we have $\tau \leq \sum_i p_i\tau_{1i} \otimes I_2 = \tau_1 \otimes I_2$, where $\tau_1 = {\rm Tr}_2\, \tau$. Since $\log x$ $(x > 0)$ is operator monotone (see Section 4.1), we have $\log\tau \leq (\log\tau_1) \otimes I_2$ (as long as $\tau$ is invertible). Therefore,

$$\begin{aligned} S(\rho_{12}\|\tau) &\geq -S(\rho_{12}) - {\rm Tr}\,\rho_{12}((\log\tau_1) \otimes I_2) = -S(\rho_{12}) - {\rm Tr}\,\rho_1 \log\tau_1 \\ &= -S(\rho_{12}) + S(\rho_1) + S(\rho_1\|\tau_1) \geq -S(\rho_{12}) + S(\rho_1). \end{aligned}$$

This implies that $E_{\rm re}(\rho_{12}) \geq -S(\rho_{12}) + S(\rho_1)$ holds in general. Now, let $\rho_{12} = |\xi\rangle\langle\xi|$ be a pure state; hence $E_{\rm re}(\rho_{12}) \geq S(\rho_1)$. Take the Schmidt decomposition of $\xi$ in (1.19) and define

$$\tau := \sum_{i=1}^k p_i|\eta_{1i} \otimes \eta_{2i}\rangle\langle\eta_{1i} \otimes \eta_{2i}| = \sum_{i=1}^k p_i|\eta_{1i}\rangle\langle\eta_{1i}| \otimes |\eta_{2i}\rangle\langle\eta_{2i}|.$$

Then $S(\rho_{12}\|\tau) = -\sum_i p_i \log p_i = S(\rho_1)$ and hence $E_{\rm re}(\rho_{12}) = S(\rho_1)$. Next, if $\rho_{12} = |\xi\rangle\langle\xi|$ and $\rho_{123}$ is an extension of $\rho_{12}$, then we have $\operatorname{Tr}\rho_{123}((I_{12} - |\xi\rangle\langle\xi|) \otimes I_3) = 0$ so that

$$\rho_{123} = \rho_{123}(|\xi\rangle\langle\xi| \otimes I_3) = (|\xi\rangle\langle\xi| \otimes I_3)\rho_{123}(|\xi\rangle\langle\xi| \otimes I_3).$$

This implies that $\rho_{123}$ is of the form $\rho_{123} = \rho_{12} \otimes \rho_3$ so that

$$\begin{aligned} &S(\rho_{13}) - S(\rho_3) - S(\rho_{123}) + S(\rho_{23}) \\ &\quad = (S(\rho_1) + S(\rho_3)) - S(\rho_3) - S(\rho_3) + (S(\rho_2) + S(\rho_3)) = 2S(\rho_1). \end{aligned}$$

Hence $E_{\rm sq}(\rho_{12}) = S(\rho_1)$.

(4) Corollary 1.3 shows that $E_{\rm re}(\rho_{12})$ is continuous in $\rho_{12}$ (a direct proof is also easy). Finally, the continuity of $E_{\rm sq}(\rho_{12})$ can be proved by using the Alicki-Fannes inequality in (1.10) and Theorem 3.6 below (see [16] for the details of the proof). □

## 2. Quasi-entropies and quantum $f$-divergences

Monotonicity is the most important property of the relative entropy. This was effectively used, for instance, in the previous section in the proof of strong subadditivity of the von Neumann entropy. The equality condition for the monotonicity of the relative entropy is worth investigating as an analogy in quantum probability of the characterization theorem for sufficient statistics in classical probability. Moreover, joint convexity of the relative entropy is closely related to the WYDL concavity (Wigner-Yanase-Dyson-Lieb concavity) (see [77, 47, 6, 74, 4]). (Here, the difference between convexity and concavity simply comes from the choice of the plus or minus sign.) Kosaki [42] proved the WYDL concavity for the expression extending the relative entropy by use of the interpolation theory. Petz [63, 64] introduced quasi-entropies in the same expression and showed their monotonicity and joint convexity. In this section we will explain monotonicity and its equality condition shown in a recent paper [31] in the restricted situation of quantum $f$-divergences in finite-dimensional quantum systems. Although our main discussions here are restricted to quantum $f$-divergences (a special case of quasi-entropies), we will present our results under assumptions as weak as possible.

Consider finite-dimensional $\mathcal{A} = B(\mathcal{H}) = M_d(\mathbb{C})$ as in the previous section, and $\mathcal{A}_1, \mathcal{A}_2$ are similar. For (not necessarily invertible) $A \in \mathcal{A}^+$, write $A^0$ for the support projection of $A$ and $A^{-1}$ for the generalized inverse of $A$, i.e., $A^{-1} := (A|_{A^0\mathcal{H}})^{-1}A^0$. Furthermore, we write $L_A$ and $R_A$ for the left and right multiplications of $A$, i.e., $L_A X := AX$ and $R_A X := XA$ ($X \in \mathcal{A}$). Let $f$ be a general real function on $[0,\infty)$. For $A, B \in \mathcal{A}^+$, since $L_A$ and $R_{B^{-1}}$ commute, one can define $f(L_A R_{B^{-1}})$. More concretely, with the spectral decompositions $A = \sum_{a\in{\rm Sp}(A)} aP_a$ and $B = \sum_{b\in{\rm Sp}(B)} bQ_b$ where ${\rm Sp}(A)$ is the set of eigenvalues (spectra) of $A$, we define

$$f(L_A R_{B^{-1}})X = \sum_{a\in{\rm Sp}(A)} \sum_{b\in{\rm Sp}(B)} f(ab^{-1})P_a X Q_b, \qquad X \in \mathcal{A},$$

with the convention that $f(ab^{-1}) = f(0)$ if $b = 0$. When $\varphi(X) := \operatorname{Tr} AX$ and $\psi(X) := \operatorname{Tr} BX$ ($X \in \mathcal{A}$), note that $L_A R_{B^{-1}}$ is the *relative modular operator* $\Delta_{\varphi,\psi}$ of $\varphi, \psi$ due to Araki [6]. Indeed, the so-called standard representation of $\mathcal{A}$ is the left multiplication of the elements of $\mathcal{A}$ on the Hilbert space $(\mathcal{A}, \langle\cdot,\cdot\rangle_{\rm HS})$, and the commutant of $\mathcal{A}$ is represented as the right multiplication of the elements of $\mathcal{A}$. Since the representing vectors of $\varphi$ and $\psi$ are $A^{1/2}$ and $B^{1/2}$, respectively,

the conjugate-linear operator $S_{\varphi,\psi}$ and $F_{\varphi,\psi} = S_{\varphi,\psi}^*$ defined by Araki are given as follows:

$$S_{\varphi,\psi}(XB^{1/2}+Y) := B^0X^*A^{1/2}, \qquad X,Y \in \mathcal{A},\ YB^0 = 0,$$
$$F_{\varphi,\psi}(B^{1/2}X+Z) := A^{1/2}X^*B^0, \qquad X,Z \in \mathcal{A},\ B^0Z = 0.$$

Hence $\Delta_{\varphi,\psi} := F_{\varphi,\psi}S_{\varphi,\psi}$ is given as

$$\Delta_{\varphi,\psi}(XB^{1/2}+Y) = F_{\varphi,\psi}(B^{1/2}B^{-1/2}X^*A^{1/2}) = AXB^{-1/2} = L_AR_{B^{-1}}(XB^{1/2}+Y).$$

Now, the *quasi-entropy* of $A, B$ with respect to $f$ and $K \in \mathcal{A}$ is defined by

$$S_f^K(A\|B) := \langle KB^{1/2}, f(L_AR_{B^{-1}})(KB^{1/2})\rangle_{\mathrm{HS}} = \operatorname{Tr} B^{1/2}K^*f(L_AR_{B^{-1}})(KB^{1/2}).$$

Monotonicity and joint concavity/convexity of quasi-entropies due to Petz (also Kosaki) are summarized as follows:

**Theorem 2.1** ([63, 64])**.** *Let $f$ be an operator monotone function (see Section 4.1) on $[0,\infty)$ with $f(0) \geq 0$, and $\Phi : \mathcal{A}_1 \to \mathcal{A}_2$ be a 2-positive map (see Section 4.2) preserving traces. If $A_1, B_1 \in \mathcal{A}_1$ and $A_2, B_2 \in \mathcal{A}_2$ satisfy $\Phi(A_1) \leq A_2$ and $\Phi(B_1) \leq B_2$, then for every $K \in \mathcal{A}_2$*

$$S_f^K(A_2\|B_2) \geq S_f^{\Phi^*(K)}(A_1\|B_1).$$

*Furthermore, when $\Phi : \mathcal{A}_1 \to \mathcal{A}_2$ preserves traces and $\Phi^*$ is a Schwarz map (see Section 4.2), the same result holds if $A_1, B_1 \in \mathcal{A}_1^{++}$ and $A_2, B_2 \in \mathcal{A}_2^{++}$ (i.e., $A_j, B_j$ are invertible).*

**Theorem 2.2** ([64])**.** *Let $f$ be an operator convex function on $[0,\infty)$, $\mathcal{A}_0$ be a $*$-subalgebra of $\mathcal{A}$ and $\mathcal{E} : \mathcal{A} \to \mathcal{A}_0$ be the conditional expectation preserving traces (i.e., the dual map of the injection $\mathcal{A}_0 \hookrightarrow \mathcal{A}$). Then, for every $A \in \mathcal{A}^+$, $B \in \mathcal{A}^{++}$ and $K \in \mathcal{A}_0$,*

$$S_f^K(\mathcal{E}(A)\|\mathcal{E}(B)) \leq S_f^K(A\|B).$$

**Theorem 2.3** ([42, 63, 64])**.** *Let $f$ be the same as in Theorem 2.1. If $A_1, A_2, A, B_1, B_2, B \in \mathcal{A}^+$ and $\lambda, \mu \geq 0$ satisfy $\lambda A_1 + \mu A_2 \leq A$ and $\lambda B_1 + \mu B_2 \leq B$, then for every $K \in \mathcal{A}$*

$$\lambda S_f^K(A_1\|B_1) + \mu S_f^K(A_2\|B_2) \leq S_f^K(A\|B).$$

**Theorem 2.4** ([63])**.** *Let $f$ be the same as in Theorem 2.2. Then, for every $K \in \mathcal{A}$, $S_f^K(A\|B)$ is jointly convex in $(A,B) \in \mathcal{A}^+ \times \mathcal{A}^{++}$.*

Since an operator monotone function on $[0,\infty)$ is operator concave, the hypothesis on $f$ in Theorem 2.1 is stronger than that on $-f$ in Theorem 2.2. On the other hand, a map $\mathcal{E}$ in Theorem 2.2 is very special compared with that in Theorem 2.1. Our target is to combine both theorems to show

$$S_f^K(\Phi(A)\|\Phi(B)) \leq S_f^{\Phi^*(K)}(A\|B)$$

for $A, B \in \mathcal{A}_1^+$ and $K \in \mathcal{A}_2$ when $f$ is operator convex and $\Phi : \mathcal{A}_1 \to \mathcal{A}_2$ is a certain map (preserving traces). However, due to technical difficulties, we have confined ourselves to the case where $K = I_2$ (and also $\Phi^*(K) = I_1$). To stress this restriction, we repeat the definition of $S_f(A\|B)$ in the name of (quantum) $f$-divergence. In the rest of this section we assume that $f$ is a real function on $[0,\infty)$ such that it is continuous on $(0,\infty)$ and the limit

$$\omega(f) := \lim_{x\to\infty} \frac{f(x)}{x} \in [-\infty, +\infty]$$

exists. For $A, B \in \mathcal{A}^+$ such that $\operatorname{supp} A \le \operatorname{supp} B$ we define

$$S_f(A\|B) := \langle B^{1/2}, f(L_A R_{B^{-1}})B^{1/2}\rangle_{\mathrm{HS}},$$

and further define the *$f$-divergence* of general $A, B \in \mathcal{A}^+$ by

$$S_f(A\|B) := \lim_{\varepsilon \searrow 0} S_f(A\|B + \varepsilon I).$$

Consequently, it turns out that $S_f(A\|B)$ is generally expressed as

$$S_f(A\|B) = \langle B^{1/2}, f(L_A R_{B^{-1}})B^{1/2}\rangle_{\mathrm{HS}} + \omega(f)\mathrm{Tr}\, A(I - B^0).$$

**Example 2.5.** The function $f(x) := x \log x$ on $[0, \infty)$ is operator convex. Noting that $\omega(f) = +\infty$, it is easy to see that $S_f(A\|B)$ coincides with the relative entropy $S(A\|B)$ in (1.2). Next, for each $\alpha > 0$ define $f_\alpha(x) := x^\alpha$ $(x \ge 0)$ and $f_0(0) := 1$, $f_0(x) := 0$ $(x > 0)$. Then we have

$$S_{f_\alpha}(A\|B) = \begin{cases} \operatorname{Tr} A^\alpha B^{1-\alpha} & \text{if } 0 \le \alpha < 1 \text{ or } \operatorname{supp} A \le \operatorname{supp} B, \\ +\infty & \text{if } \alpha > 1 \text{ and } \operatorname{supp} A \not\le \operatorname{supp} B. \end{cases}$$

For $\alpha \in [0, \infty) \setminus \{1\}$,

$$S_\alpha(A\|B) := \frac{1}{\alpha - 1} \log S_{f_\alpha}(A\|B)$$

is called the *Rényi $\alpha$-relative entropy* of $A, B$.

It is well known that the relative entropy $S(A\|B)$ is jointly lower semicontinuous in $A, B$ but it is not continuous in $A$. However, it seems that it is not so well known that $S(A\|B)$ is continuous in $B$ in the finite-dimensional case. Therefore, the next proposition might be rather remarkable, saying that any $f$-divergence is continuous in the second variable.

**Proposition 2.6** ([31])**.** *$S_f(A\|B)$ is continuous in $B$, that is, $\lim_{k\to\infty} S_f(A\|B_k) = S_f(A\|B)$ if $A, B, B_k \in \mathcal{A}^+$ and $B_k \to B$.*

The next theorem is concerned with the monotonicity of $f$-divergences.

**Theorem 2.7** ([31])**.** *Let $f$ be an operator convex function on $[0, \infty)$ and $\Phi : \mathcal{A}_1 \to \mathcal{A}_2$ be such that $\Phi^*$ is a Schwarz map. Let $A, B \in \mathcal{A}_1^+$ and assume that $\operatorname{Tr}\Phi(B) = \operatorname{Tr} B$. Furthermore, assume that either $\operatorname{Tr}\Phi(A) = \operatorname{Tr} A$ or $\omega(f) \ge 0$. Then*

$$S_f(\Phi(A)\|\Phi(B)) \le S_f(A\|B).$$

To prove this, when $f$ is continuous at 0, one can apply integral expression (4.6) in Section 4.1 to have

$$S_f(A\|B) = f(0)\mathrm{Tr}\, B + a\mathrm{Tr}\, AB^0 + b\mathrm{Tr}\, A^2B^{-1} + \int_{(0,\infty)} \left( \frac{\mathrm{Tr}\, AB^0}{1+\lambda} + S_{\varphi_\lambda}(A\|B) \right) d\mu(\lambda) + \omega(f)\mathrm{Tr}\, A(I_1 - B^0)$$

and a similar expression of $S_f(\Phi(A)\|\Phi(B))$, where for $\lambda > 0$ the function

$$\varphi_\lambda(x) := -x/(x+\lambda) = -1 + \lambda/(x+\lambda) \tag{2.1}$$

is operator convex. By assumption, $\operatorname{Tr}\Phi(B) = \operatorname{Tr} B$. Furthermore, $S_{\varphi_\lambda}(\Phi(A)\|\Phi(B)) \le S_{\varphi_\lambda}(A\|B)$ can be proved. Hence it suffices to show the monotonicity of other terms in the above expression, which can be done by dividing into the cases of $\operatorname{Tr}\Phi(A) = \operatorname{Tr} A$, $\omega(f) = +\infty$, and $0 \le \omega(f) < +\infty$. Here, Proposition 2.6 is useful.

When $f$ is not continuous at 0, $\tilde{f} := f - \alpha \mathbf{1}_{\{0\}}$ is operator convex and continuous at 0, where $\alpha := f(0) - f(0+) > 0$. Hence, to complete the proof, it remains to show that $S_{\mathbf{1}_{\{0\}}}(\Phi(A)\|\Phi(B)) \le S_{\mathbf{1}_{\{0\}}}(A\|B)$.

The joint convexity of $f$-divergences can easily be shown from the monotonicity in Theorem 2.7.

**Corollary 2.8** ([31]). *If $f$ is operator convex on $[0,\infty)$, then $S_f(A\|B)$ is jointly convex in $A, B \in \mathcal{A}^+$.*

Indeed, $\Phi : \mathcal{A} \otimes M_2(\mathbb{C}) \to \mathcal{A}$, $\begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} \mapsto X_{11} + X_{22}$ is a completely positive map preserving traces. For $A_1, A_2, B_1, B_2 \in \mathcal{A}^+$ and $0 < p < 1$, with

$$A := \begin{bmatrix} pA_1 & 0 \\ 0 & (1-p)A_2 \end{bmatrix}, \quad B := \begin{bmatrix} pB_1 & 0 \\ 0 & (1-p)B_2 \end{bmatrix},$$

Theorem 2.7 implies that

$$\begin{aligned} &S_f(pA_1 + (1-p)A_2 \| pB_1 + (1-p)B_2) \\ &\quad = S_f(\Phi(A)\|\Phi(B)) \le S_f(A\|B) = pS_f(A_1\|B_1) + (1-p)S_f(A_2\|B_2). \end{aligned}$$

Concerning the equality condition for the monotonicity of $f$-divergences given in Theorem 2.7 we have the following comprehensive result, extending the results [65, 38, 39] for the relative entropy and the Rényi relative entropy to a wide class of operator convex functions. The assumption $\operatorname{supp} A \le \operatorname{supp} B$ added to Theorem 2.7 is essential to avoid the case $S_f(\Phi(A)\|\Phi(B)) = S_f(A\|B) = +\infty$.

**Theorem 2.9** ([31]). *Let $\Phi : \mathcal{A}_1 \to \mathcal{A}_2$ be such that $\Phi^*$ is a Schwarz map. Let $A, B \in \mathcal{A}_1^+$ and assume that $\operatorname{supp} A \le \operatorname{supp} B$ and $\operatorname{Tr} \Phi(B) = \operatorname{Tr} B$ (then $\operatorname{Tr} \Phi(A) = \operatorname{Tr} A$ automatically holds as well). Then, for the conditions listed below, we have*

$$\text{(i)} \Longrightarrow \text{(ii)} \Longrightarrow \text{(iii)} \Longrightarrow \text{(iv)} \Longleftrightarrow \text{(v)} \Longleftrightarrow \text{(vi)} \Longleftrightarrow \text{(vii)} \Longleftrightarrow \text{(viii)} \Longleftrightarrow \text{(ix)} \Longrightarrow \text{(x)}.$$

*Furthermore, if $\Phi$ is 2-positive, then (i)–(x) are all equivalent.*

(i) *There exists a trace-preserving $\Psi : \mathcal{A}_2 \to \mathcal{A}_1$ such that $\Psi^*$ is a Schwarz map, $\Psi(\Phi(A)) = A$ and $\Psi(\Phi(B)) = B$.*

(ii) *There exists a map $\Psi : \mathcal{A}_2 \to \mathcal{A}_1$ such that $\Psi^*$ is a Schwarz map, $\Psi(\Phi(A)) = A$ and $\Psi(\Phi(B)) = B$.*

(iii) *$S_f(\Phi(A)\|\Phi(B)) = S_f(A\|B)$ for every operator convex function $f$ on $[0,\infty)$.*

(iv) *$S_f(\Phi(A)\|\Phi(B)) = S_f(A\|B)$ for some operator convex function $f$ on $[0,\infty)$ such that $|\operatorname{supp} \mu_f| \ge |\operatorname{Sp}(L_A R_{B^{-1}}) \cup \operatorname{Sp}(L_{\Phi(A)} R_{\Phi(B)^{-1}})|$, where $\mu_f$ is the representing measure $\mu$ of integral expression (4.6) for $\tilde{f}$ given after (2.1), and $|H|$ is the cardinality of a set $H$.*

(v) *There exists a set $\Lambda \subset (0,\infty)$ such that*

$$|\Lambda| \ge |\operatorname{Sp}(L_A R_{B^{-1}}) \cup \operatorname{Sp}(L_{\Phi(A)} R_{\Phi(B)^{-1}})|$$

*and $S_{\varphi_\lambda}(\Phi(A)\|\Phi(B)) = S_{\varphi_\lambda}(A\|B)$ holds for every $\lambda \in \Lambda$, where $\varphi_\lambda$ is an operator convex function on $[0,\infty)$ defined by (2.1).*

(vi) *$B^0 \Phi^*(\Phi(B)^{-z}\Phi(A)^z) = B^{-z}A^z$ for every $z \in \mathbb{C}$.*

(vii) *$B^0 \Phi^*(\Phi(B)^{-\alpha}\Phi(A)^\alpha) = B^{-\alpha}A^\alpha$ for some $\alpha \in (0,2) \setminus \{1\}$.*

(viii) *$B^0 \Phi^*(\Phi(B)^{-it}\Phi(A)^{it}) = B^{-it}A^{it}$ for every $t \in \mathbb{R}$.*

(ix) *$B^0 \Phi^*(\log^* \Phi(A) - (\log^* \Phi(B))\Phi(A)^0) = \log^* A - (\log^* B)A^0$ ($\log^*$ was defined just after (1.2)).*

(x) *The map* $\Phi_B^* : \mathcal{A}_2 \to \mathcal{A}_1$ *defined by*

$$\Phi_B^*(Y) := B^{1/2}\Phi^*(\Phi(B)^{-1/2}Y\Phi(B)^{-1/2})B^{1/2} \ (Y \in \mathcal{A}_2)$$

*satisfies* $\Phi_B^*(\Phi(A)) = A$ *(*$\Phi_B^*(\Phi(B)) = B$ *automatically holds).*

**Corollary 2.10** ([31])**.** *Let* $A_1, A_2, B_1, B_2 \in \mathcal{A}^+$ *and* $0 < p < 1$*. Assume that* $\operatorname{supp} A_j \le \operatorname{supp} B_j$ $(j = 1, 2)$*. A necessary and sufficient condition for every operator convex function (or an operator convex function satisfying the assumption in (iv) above)* $f$ *on* $[0, \infty)$ *to satisfy equality condition*

$$S_f(pA_1 + (1-p)A_2 \| pB_1 + (1-p)B_2) = pS_f(A_1\|B_1) + (1-p)S_f(A_2\|B_2)$$

*in the joint convexity of Corollary* 2.8 *is*

$$B_j^{-1/2}A_jB_j^{-1/2} = (pB_1 + (1-p)B_2)^{-1/2}(pA_1 + (1-p)A_2)(pB_1 + (1-p)B_2)^{-1/2}, \quad j = 1, 2.$$

Indeed, the above condition is a rephrase of $\Psi_B^*(\Phi(A)) = A$ of (x) for $\Phi$, $A$ and $B$ given in Corollary 2.8.

*Remark* 2.11. An operator convex function $f$ in Theorem 2.9 (iv) cannot be general. For example, for a linear function $f(x) = ax+b$, $S_f(\Phi(A)\|\Phi(B)) = S_f(A\|B)$ always holds whenever $\operatorname{supp} A \le \operatorname{supp} B$ and $\operatorname{Tr}\Phi(B) = \operatorname{Tr} B$; however, we do not have property (x) of Theorem 2.9. Moreover, for $f(x) = x^2$, [40, Example 2.2] provided a conditional expectation $\Phi : \mathcal{A}_1 := B(\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3) \to \mathcal{A}_2 := B(\mathcal{H}_1 \otimes \mathcal{H}_2)$ for which $S_f(\Phi(A)\|\Phi(B)) = S_f(A\|B)$ holds but (viii) of Theorem 2.9 is not satisfied.

## 3. State discrimination distances and quantum hypothesis testing

The notion of $f$-divergences (in particular, the relative entropy) discussed in the preceding section is useful as certain distances to discriminate states (though they do not satisfy the triangle inequality so that they are not distances in the strict sense). Of course, the most natural distance between density matrices is the trace-norm distance $\|\rho - \sigma\|_1$. When $\Phi$ is a positive map satisfying $\operatorname{Tr} \circ \Phi \le \operatorname{Tr}$, the monotonicity $\|\Phi(\rho) - \Phi(\sigma)\|_1 \le \|\rho - \sigma\|_1$ in the trace-norm distance is easily verified. What is called fidelity is another important quantity in quantum information theory to measure the distance between states. Moreover, the Chernoff and Hoeffding distances are notable, as they appeared as asymptotic error exponents in a recent development of quantum hypothesis testing. In this section we will discuss those state discrimination distances and explain their monotonicity property and related inequalities. In the last section we will briefly touch quantum hypothesis testing as well.

For states $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ the *fidelity* introduced by Uhlmann [73] is

$$F(\rho, \sigma) := \operatorname{Tr}\left((\rho^{1/2}\sigma\rho^{1/2})^{1/2}\right) = \|\sigma^{1/2}\rho^{1/2}\|_1.$$

This is a non-commutative version of the *Hellinger affinity* $B(p, q) := \sum_{i=1}^d \sqrt{p_i q_i}$ for probability vectors $p = (p_1, \ldots, p_d)$ and $q = (q_1, \ldots, q_d)$ in classical probability. In the case of pure states $\rho = |\xi\rangle\langle\xi|$ and $\sigma = |\eta\rangle\langle\eta|$, $F(\rho, \sigma) = |\langle\xi, \eta\rangle|$. Note that $|\langle\xi, \eta\rangle|^2$ is an old notion called *transition probability* so that fidelity is also called *Uhlmann's transition probability* though they are different in powers. Basic properties are summarized as follows:

**Proposition 3.1.** (1) $F(\rho, \sigma) = F(\sigma, \rho)$.

(2) $0 \le \operatorname{Tr}\rho^{1/2}\sigma^{1/2} \le F(\rho,\sigma) \le 1$.
(3) $F(\rho,\sigma) = 0 \Leftrightarrow \operatorname{supp}\rho \perp \operatorname{supp}\sigma$.
(4) $F(\rho,\sigma) = 1 \Leftrightarrow \rho = \sigma$.

*Proof.* (1) is obvious since $\|\rho^{1/2}\sigma^{1/2}\|_1 = \|\sigma^{1/2}\rho^{1/2}\|_1$. The inequalities $\operatorname{Tr}\rho^{1/2}\sigma^{1/2} \ge 0$ and $\operatorname{Tr}\rho^{1/2}\sigma^{1/2} \le F(\rho,\sigma)$ in (2) are also obvious. By Hölder's inequality, $F(\rho,\sigma) \le \operatorname{Tr}\rho\cdot\operatorname{Tr}\sigma = 1$, and equality holds if and only if $\rho$ and $\sigma$ are proportional, which reduces to $\rho = \sigma$ for density operators $\rho,\sigma$. Hence (4) holds (this can also be seen from Theorem 3.6 below). If $\operatorname{supp}\rho \perp \operatorname{supp}\sigma$, then $\rho^{1/2}\sigma^{1/2} = 0$ so that $F(\rho,\sigma) = 0$. Conversely, since there are $\alpha,\beta > 0$ such that $\rho^{1/2} \ge \alpha\rho^0$, $\sigma^{1/2} \ge \beta\sigma^0$, we have $\operatorname{Tr}\rho^{1/4}\sigma^{1/2}\rho^{1/4} \ge \alpha\beta\operatorname{Tr}\rho^0\sigma^0$. This implies that if $\operatorname{Tr}\rho^{1/2}\sigma^{1/2} = 0$, then $\operatorname{Tr}\rho^0\sigma^0 = 0$ so that $\rho^0\sigma^0 = 0$. Hence (3) holds. □

From the properties in the above proposition we may regard $1 - F(\rho,\sigma)$ as a function measuring a distance between $\rho,\sigma$.

The following variational expression was shown by Uhlmann in the $C^*$-algebra setting while the form here in the finite-dimensional case is based on [67, Theorem 6.1].

**Theorem 3.2** ([73])**.** *For every $\rho,\sigma \in \mathcal{S}$,*

$$F(\rho,\sigma) = \inf\{\sqrt{\operatorname{Tr}\rho G\cdot\operatorname{Tr}\sigma G^{-1}} : G \in \mathcal{A}^{++}\}.$$

*Proof.* By continuity we may assume that $\rho$ and $\sigma$ are invertible. Take the polar decomposition $\sigma^{1/2}\rho^{1/2} = V|\sigma^{1/2}\rho^{1/2}|$ (with a unitary $V$). For every $G \in \mathcal{A}^{++}$, by the Schwarz inequality,

$$\begin{aligned} F(\rho,\sigma) = \operatorname{Tr}V^*\sigma^{1/2}\rho^{1/2} &= \operatorname{Tr}G^{1/2}\rho^{1/2}V^*\sigma^{1/2}G^{-1/2} = \langle G^{1/2}\rho^{1/2}, G^{-1/2}\sigma^{1/2}V\rangle_{\mathrm{HS}} \\ &\le \langle G^{1/2}\rho^{1/2}, G^{1/2}\rho^{1/2}\rangle_{\mathrm{HS}}^{1/2}\,\langle G^{-1/2}\sigma^{1/2}V, G^{-1/2}\sigma^{1/2}V\rangle_{\mathrm{HS}}^{1/2} \\ &= \sqrt{\operatorname{Tr}\rho G\cdot\operatorname{Tr}\sigma G^{-1}}. \end{aligned}$$

Furthermore, let $G := \rho^{-1/2}V^*\sigma^{1/2} = \rho^{-1/2}|\sigma^{1/2}\rho^{1/2}|\rho^{-1/2}$. Then $G \in \mathcal{A}^{++}$ and

$$\operatorname{Tr}\rho G = \operatorname{Tr}\sigma G^{-1} = \operatorname{Tr}|\sigma^{1/2}\rho^{1/2}|$$

so that the required equality follows. □

The above theorem immediately implies the (inverse) monotonicity of fidelity.

**Theorem 3.3.** (1) *If $\Phi : \mathcal{A}_1 = B(\mathcal{H}_1) \to \mathcal{A}_2 = B(\mathcal{H}_2)$ is a trace-preserving positive map, then $F(\Phi(\rho),\Phi(\sigma)) \ge F(\rho,\sigma)$ for every $\rho,\sigma \in \mathcal{S}(\mathcal{H}_1)$.*
(2) *$F(\rho,\sigma)$ is jointly concave in $\rho,\sigma \in \mathcal{S}$.*

*Proof.* Since $\Phi^*$ is a unital positive map, it follows (see, for example, [13, Theorem 2.3.6]) that $\Phi^*(G^{-1}) \ge \Phi^*(G)^{-1}$ for every $G \in \mathcal{A}_2^{++}$. Hence, by Theorem 3.2,

$$\begin{aligned} \operatorname{Tr}\Phi(\rho)G\cdot\operatorname{Tr}\Phi(\sigma)G^{-1} = \operatorname{Tr}\rho\Phi^*(G)\cdot\operatorname{Tr}\sigma\Phi^*(G^{-1}) &\ge \operatorname{Tr}\rho\Phi^*(G)\cdot\operatorname{Tr}\sigma\Phi^*(G)^{-1} \\ &\ge F(\rho,\sigma)^2, \end{aligned}$$

which gives monotonicity.

(2) can be shown from (1) similarly to Corollary 2.8. □

Let $\rho \in \mathcal{S}(\mathcal{H})$. When $\tilde{\mathcal{H}}$ is a (finite-dimensional) Hilbert space and $\xi \in \mathcal{H}\otimes\tilde{\mathcal{H}}$ is a unit vector such that $\operatorname{Tr}\rho X = \langle\xi,(X\otimes I)\xi\rangle_{\mathrm{HS}}$ $(X \in \mathcal{A})$, i.e., $\rho = \operatorname{Tr}_2|\xi\rangle\langle\xi|$ holds, $\xi$ (also $|\xi\rangle\langle\xi|$) is called a *purification* of $\rho$. Indeed, let $\tilde{\mathcal{H}} = \mathcal{H}$ and $\{e_i\}_{i=1}^d$ be an

orthonormal basis of $\mathcal{H}$. Then $\xi := \sum_{i=1}^d \rho^{1/2} e_i \otimes e_i$ becomes a purification of $\rho$. Also, let $\rho = \sum_{i=1}^k p_i |f_i\rangle\langle f_i|$ be the spectral decomposition of $\rho$ with $p_i > 0$ and an orthonormal set $\{f_i\}_{i=1}^k$. Then $\xi := \sum_{i=1}^k p_i f_i \otimes e_i$ is another purification of $\rho$. The next variational expression of fidelity in terms of purifications is due to Uhlmann, which is quite useful in quantum information theory.

**Theorem 3.4** ([73]). *For every $\rho, \sigma \in \mathcal{S}(\mathcal{H})$,*

$$F(\rho,\sigma) = \max\{|\langle \xi, \eta\rangle| : \xi, \eta \in \mathcal{H} \otimes \tilde{\mathcal{H}} \text{ are respective purifications of } \rho, \sigma\}.$$

*Moreover, the above* max *is attained in the case $\tilde{\mathcal{H}} = \mathcal{H}$.*

*Proof.* If $\xi$ and $\eta$ are respective purifications of $\rho$ and $\sigma$, then by Theorem 3.3 (1),

$$F(\rho,\sigma) = F(\mathrm{Tr}_2 |\xi\rangle\langle\xi|, \mathrm{Tr}_2 |\eta\rangle\langle\eta|) \geq F(|\xi\rangle\langle\xi|, |\eta\rangle\langle\eta|) = |\langle\xi,\eta\rangle|.$$

Let $\{e_i\}_{i=1}^d$ be an orthonormal basis of $\mathcal{H}$ and $U \in \mathcal{A}$ be a unitary. Define

$$\xi := \sum_{i=1}^d \rho^{1/2} e_i \otimes e_i, \quad \eta := \sum_{i=1}^d \sigma^{1/2} U e_i \otimes e_i.$$

Then $\xi$ and $\eta$ are purifications of $\rho$ and $\sigma$, respectively, and we have $|\langle\xi,\eta\rangle| = |\mathrm{Tr}\, \rho^{1/2}\sigma^{1/2}U|$. Maximizing this on $U$ gives $\mathrm{Tr}\,|\rho^{1/2}\sigma^{1/2}|$. □

The following inequality was proved by Audenaert et al. [10], and another proof was given in [28]. Another proof, shown below, is based on the monotonicity of $f$-divergences (Theorem 2.7) and Theorem 3.4.

**Corollary 3.5.** *For every $\rho, \sigma \in \mathcal{S}$,*

$$F(\rho,\sigma)^2 \leq \min_{0\leq\alpha\leq 1} \mathrm{Tr}\, \rho^\alpha \sigma^{1-\alpha}.$$

*Proof.* By Theorem 3.4 there exist purifications $\xi, \eta \in \mathcal{H}\otimes\mathcal{H}$ of $\rho, \sigma$ such that

$$F(\rho,\sigma)^2 = |\langle\xi,\eta\rangle|^2 = \mathrm{Tr}\,|\xi\rangle\langle\xi|\,|\eta\rangle\langle\eta| = \mathrm{Tr}\,(|\xi\rangle\langle\xi|)^\alpha (|\eta\rangle\langle\eta|)^{1-\alpha}, \qquad 0 \leq \alpha \leq 1.$$

Since $\rho = \mathrm{Tr}_2 |\xi\rangle\langle\xi|$ and $\sigma = \mathrm{Tr}_2 |\eta\rangle\langle\eta|$, for each $\alpha \in [0,1]$ one can apply Theorem 2.7 to $f(x) = -x^\alpha$ from Example 2.5 to have

$$\mathrm{Tr}\,(|\xi\rangle\langle\xi|)^\alpha (|\eta\rangle\langle\eta|)^{1-\alpha} \leq \mathrm{Tr}\,\rho^\alpha\sigma^{1-\alpha}.$$

Hence the required inequality follows. □

The following neat relation between trace-norm distance and fidelity is due to Fuchs and van de Graaf.

**Theorem 3.6** ([20]). *For every $\rho, \sigma \in \mathcal{S}$,*

$$1 - F(\rho,\sigma) \leq \frac{1}{2}\|\rho - \sigma\|_1 \leq \sqrt{1 - F(\rho,\sigma)^2}.$$

*Proof.* By continuity we may assume that $\rho$ and $\sigma$ are invertible. By the proof of Theorem 3.2 there exists a $G \in \mathcal{A}^{++}$ such that $\mathrm{Tr}\,\rho G = \mathrm{Tr}\,\sigma G^{-1} = F(\rho,\sigma)$. Let $G = \sum_{i=1}^d \alpha_i |e_i\rangle\langle e_i|$ be the spectral decomposition with $\alpha_i > 0$ and an orthonormal basis $\{e_i\}_{i=1}^d$. Define a trace-preserving positive map $\Phi : \mathcal{A} \to \mathbb{C}^d$ ($\subset M_d(\mathbb{C})$ and the trace on $M_d(\mathbb{C})$ is restricted on $\mathbb{C}^d$) by $\Phi(X) := (\langle e_i, Xe_i\rangle)_{i=1}^d$ ($X \in \mathcal{A}$). Then

$p=(p_i)_{i=1}^d := \Phi(\rho)$ and $q=(q_i)_{i=1}^d := \Phi(\sigma)$ are probability vectors and we have $\sum_i \alpha_i p_i = \sum_i \alpha_i^{-1} q_i = F(\rho,\sigma)$. By Theorem 3.3 (1) and the Schwarz inequality,

$$F(\rho,\sigma) \le F(\Phi(\rho),\Phi(\sigma)) = \sum_i \sqrt{p_i q_i} \le \left(\sum_i \alpha_i p_i\right)^{1/2} \left(\sum_i \alpha_i^{-1} q_i\right)^{1/2} = F(\rho,\sigma).$$

Hence $F(\rho,\sigma) = \sum_i \sqrt{p_i q_i}$ so that

$$1 - F(\rho,\sigma) = 1 - \sum_i \sqrt{p_i q_i} = \frac{1}{2}\sum_i (\sqrt{p_i} - \sqrt{q_i})^2 \le \frac{1}{2}\sum_i |p_i - q_i| \le \frac{1}{2}\|\rho - \sigma\|_1.$$

Next, let $P := \operatorname{supp}(\rho-\sigma)_+$ and $P^\perp := I - P$, and define a trace-preserving positive map $\Psi : \mathcal{A} \to \mathbb{C}^2$ $(\subset M_2(\mathbb{C}))$ by $\Psi(X) := (\operatorname{Tr} PX, \operatorname{Tr} P^\perp X)$. Letting $(\lambda_1,\lambda_2) := \Psi(\rho)$ and $(\mu_1,\mu_2) := \Psi(\sigma)$ we have

$$\begin{aligned}
\frac{1}{2}\|\rho-\sigma\|_1 &= \frac{1}{2}(|\lambda_1-\mu_1| + |\lambda_2-\mu_2|) \\
&= \frac{1}{2}\big\{\big|\sqrt{\lambda_1}+\sqrt{\mu_1}\big|\,\big|\sqrt{\lambda_1}-\sqrt{\mu_1}\big| + \big|\sqrt{\lambda_2}+\sqrt{\mu_2}\big|\,\big|\sqrt{\lambda_2}-\sqrt{\mu_2}\big|\big\} \\
&\le \frac{1}{2}\big\{\big(\sqrt{\lambda_1}+\sqrt{\mu_1}\big)^2 + \big(\sqrt{\lambda_2}+\sqrt{\mu_2}\big)^2\big\}^{1/2} \\
&\qquad \times \big\{\big(\sqrt{\lambda_1}-\sqrt{\mu_1}\big)^2 + \big(\sqrt{\lambda_2}-\sqrt{\mu_2}\big)^2\big\}^{1/2} \\
&= \frac{1}{2}\big\{2 + 2\big(\sqrt{\lambda_1\mu_1}+\sqrt{\lambda_2\mu_2}\big)\big\}^{1/2}\big\{2 - 2\big(\sqrt{\lambda_1\mu_1}+\sqrt{\lambda_2\mu_2}\big)\big\}^{1/2} \\
&= \{1 - F(\Psi(\rho),\Psi(\sigma))^2\}^{1/2} \le \{1-F(\rho,\sigma)^2\}^{1/2}.
\end{aligned}$$

In the last inequality above, Theorem 3.3 (1) has been used again. □

In the rest of this section we will explain the Chernoff distance and the Hoeffding distance. For $A, B \in \mathcal{A}^+$ that are not necessarily states, define

$$\psi(\alpha|A\|B) := \log \operatorname{Tr} A^\alpha B^{1-\alpha}, \qquad 0 \le \alpha \le 1.$$

Moreover, for $0 \le \alpha \le 1$ let $S_\alpha(A\|B)$ be the Rényi $\alpha$-relative entropy introduced in Example 2.5. Then the *Chernoff distance* between $A$ and $B$ is defined by

$$C(A\|B) := \sup_{0\le\alpha<1}\{(1-\alpha)S_\alpha(A\|B)\} = -\min_{0\le\alpha\le1}\psi(\alpha|A\|B). \tag{3.1}$$

For any $r \in \mathbb{R}$ the *Hoeffding distance* between $A$ and $B$ is defined by

$$H_r(A\|B) := \sup_{0\le\alpha<1}\left\{-\frac{\alpha r}{1-\alpha} + S_\alpha(A\|B)\right\} = \sup_{0\le\alpha<1}\frac{-\alpha r - \psi(\alpha|A\|B)}{1-\alpha}. \tag{3.2}$$

For simplicity we write $\psi(\alpha) = \psi(\alpha|A\|B)$. Except in the case where $\operatorname{supp} A \perp \operatorname{supp} B$, $\psi(\alpha) > -\infty$ $(0 \le \alpha \le 1)$ and $\psi$ is a convex function on $[0,1]$ (as seen by computing the second derivative). We have $\psi'(0) = -S(B\|A)$ if $\operatorname{supp} A \ge \operatorname{supp} B$, and $\psi'(1) = S(A\|B)$ if $\operatorname{supp} A \le \operatorname{supp} B$. Furthermore, we have

$$H_r(A\|B) = \begin{cases} -\psi(0) & \text{if } r > -\psi(0) - \psi'(0), \\ +\infty & \text{if } r < -\psi(1), \end{cases}$$

and

$$-S_\alpha(A\|B) = \sup_{r\in\mathbb{R}}\left\{\frac{-r\alpha}{1-\alpha} - H_r(A\|B)\right\} = \sup_{-\psi(1)\le r\le -\psi(0)-\psi'(0)}\left\{\frac{-r\alpha}{1-\alpha} - H_r(A\|B)\right\}.$$

Hence, the Rényi $\alpha$-relative entropy and the Hoeffding distance are written from each other in terms of an expression similar to the Legendre transform. Also, note that if $\operatorname{Tr} A = 1$, then

$$H_0(A\|B) = \lim_{\alpha \nearrow 1} S_\alpha(A\|B) = S(A\|B).$$

When $\rho, \sigma \in \mathcal{S}$, $C(\rho\|\sigma) \geq 0$, and $C(\rho\|\sigma) = 0$ if and only if $\rho = \sigma$ (similarly to the proof of Proposition 3.1 (4)). Also, for any $r \in \mathbb{R}$, $H_r(\rho\|\sigma) \geq 0$, and $H_r(\rho\|\sigma) = 0$ if and only if $\rho = \sigma$, or, $\operatorname{supp}\rho \geq \operatorname{supp}\sigma$ and $r \geq S(\sigma\|\rho)$. The next proposition is the monotonicity of the Chernoff and Hoeffding distances, which is immediately seen from Theorems 2.7, 2.9 and definitions (3.1), (3.2).

**Proposition 3.7** ([31]). *Let $\Phi : \mathcal{A}_1 \to \mathcal{A}_2$ be such that $\Phi^*$ is a Schwarz map. Let $A, B \in \mathcal{A}_1^+$ and assume that $\operatorname{Tr}\Phi(B) = \operatorname{Tr} B$. Then*

$$C(\Phi(A)\|\Phi(B)) \leq C(A\|B), \quad H_r(\Phi(A)\|\Phi(B)) \leq H_r(A\|B), \quad r \in \mathbb{R}.$$

*Furthermore, equality holds in the above inequalities if $\Psi(\Phi(A)) = A$ and $\Psi(\Phi(B)) = B$ for some $\Psi : \mathcal{A}_2 \to \mathcal{A}_1$ such that $\Psi^*$ is a Schwarz map.*

The next theorem is concerned with the equality condition of the above monotonicity (or the converse direction of the latter assertion). Since the Chernoff and Hoeffding distances cannot be represented as an $f$-divergence, the result is not contained in Theorem 2.9.

**Theorem 3.8** ([31]). *Let $\Phi : \mathcal{A}_1 \to \mathcal{A}_2$ be such that $\Phi^*$ is a Schwarz map. Let $A, B \in \mathcal{A}_1^+$ and assume that $\operatorname{supp} A \leq \operatorname{supp} B$ and $\operatorname{Tr}\Phi(B) = \operatorname{Tr} B$. Assume that one of the following holds:*

(i) $C(\Phi(A)\|\Phi(B)) \neq S_0(\Phi(A)\|\Phi(B))$, $C(\Phi(A)\|\Phi(B)) \neq S_0(\Phi(B)\|\Phi(A))$, $C(\Phi(A)\|\Phi(B)) = C(A\|B)$.

(ii) $\operatorname{supp} A = \operatorname{supp} B$, $\operatorname{Tr} A = \operatorname{Tr} B$, $C(\Phi(A)\|\Phi(B)) = C(A\|B)$.

(iii)
$$H_r(\Phi(A)\|\Phi(B)) = H_r(A\|B)$$
*for some*
$$r \in (-\psi(1|\Phi(A)\|\Phi(B)), -\psi(0|\Phi(A)\|\Phi(B)) - \psi'(0|\Phi(A)\|\Phi(B))).$$

*Then $\Phi_B^*(\Phi(A)) = A$ ($\Phi_B^*$ was defined in Theorem 2.9 (x)). Moreover, if $\Phi$ is 2-positive, then there exists a trace-preserving $\Psi : \mathcal{A}_2 \to \mathcal{A}_1$ such that $\Psi^*$ is a Schwarz map, $\Psi(\Phi(A)) = A$ and $\Psi(\Phi(B)) = B$.*

Finally in this section we will give a survey on three kinds of hypothesis testing where the relative entropy, the Chernoff distance and the Hoeffding distance show up as asymptotic bounds for the exponential rates of error probabilities. Assume that the state of $\mathcal{A}$ is one of two $\rho, \sigma \in \mathcal{S}(\mathcal{H})$. For each natural number $n$, the compound system of $n$-copies of $\mathcal{A}$ is the $n$-fold tensor product system $\mathcal{A}^{\otimes n} = B(\mathcal{H}^{\otimes n})$, where we are given a state from two options of tensor product states $\rho^{\otimes n}$ and $\sigma^{\otimes n}$. That is, our situation here is that of independent and identically distributed (i.i.d.), familiar in classical probability. We decide a state between $\rho$ and $\sigma$ based on binary measurements $(T, I-T)$ where $T$ is an operator on $\mathcal{H}^{\otimes n}$ with $0 \leq T \leq I$. We adopt $\rho$ when the outcome belonging to $T$ occurs, and adopt $\sigma$ (or reject $\rho$) when the outcome belonging to $I-T$ occurs. There are two kinds of error probabilities. One is the *error probability of the first kind* $\alpha_n(T) := \operatorname{Tr}\rho^{\otimes n}(I-T)$ in

the case where we reject $\rho$ though it is a true state. The other is the *error probability of the second kind* $\beta_n(T) := \operatorname{Tr}\sigma^{\otimes n}T$ in the case where we adopt $\rho$ though it is not a true state (or $\sigma$ is true). Since there is a trade-off between two error probabilities $\alpha_n(T)$ and $\beta_n(T)$, it is impossible except in some trivial case to make them together as small as we want. In this situation, we may consider the following three methods in minimizing the error probabilities:

(1) Stein type: Given $\varepsilon \in (0,1)$ minimize $\beta_n(T)$ under the condition that $\alpha_n(T) \le \varepsilon$:

$$\beta^{\mathrm{S}}_{n,\varepsilon} := \min\{\beta_n(T) : \alpha_n(T) \le \varepsilon,\, T \in \mathcal{A}^{\otimes n},\, 0 \le T \le I\}.$$

(2) Chernoff type: Given a prior probability $(p, 1-p)$ of $(\rho, \sigma)$ minimize the average of the two error probabilities:

$$\beta^{\mathrm{C}}_{n,p} := \min\{p\alpha_n(T) + (1-p)\beta_n(T) : T \in \mathcal{A}^{\otimes n},\, 0 \le T \le I\}.$$

(3) Hoeffding type: Given $r > 0$ minimize $\beta_n(T)$ under the condition that $\alpha_n(T)$ decreases in the exponential order $e^{-nr}$:

$$\beta^{\mathrm{H}}_{n,r} := \min\{\beta_n(T) : \alpha_n(T) \le e^{-nr},\, T \in \mathcal{A}^{\otimes n},\, 0 \le T \le I\}.$$

The three results on quantum hypothesis testing, corresponding to the above three methods, are known as follows. The theorem (1) is called the *quantum Stein's lemma*, whose direct part was shown by Hiai and Petz and whose converse part was shown by Ogawa and Nagaoka. The theorem (2) for the quantum Chernoff bound was shown by Audenaert et al. and Nussbaum and Szkła, and the theorem (3) for the quantum Hoeffding bound was shown by Hayashi and Nagaoka.

**Theorem 3.9.** (1) [33, 57] *For every* $\varepsilon \in (0,1)$,

$$\lim_{n\to\infty} \frac{1}{n} \log \beta^{\mathrm{S}}_{n,\varepsilon} = -S(\rho\|\sigma).$$

(2) [9, 55] *For every* $p \in (0,1)$,

$$\lim_{n\to\infty} \frac{1}{n} \log \beta^{\mathrm{C}}_{n,p} = -C(\rho\|\sigma).$$

(3) [25, 53] *For every* $r > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log \beta^{\mathrm{H}}_{n,r} = -H_r(\rho\|\sigma).$$

For the quantum Chernoff bound, a key of the proof of the direct part

$$\limsup_n \frac{1}{n} \log \beta^{\mathrm{C}}_{n,p} \le -C(\rho\|\sigma)$$

is the following trace inequality: For every $A, B \in \mathcal{A}^+$,

$$\frac{1}{2}\operatorname{Tr}(A + B - |A - B|) \le \operatorname{Tr} A^s B^{1-s}, \qquad 0 \le s \le 1. \tag{3.3}$$

Indeed, it is easy to verify that

$$\min\{\operatorname{Tr} A(I-T) + \operatorname{Tr} BT : T \in \mathcal{A},\, 0 \le T \le I\} = \frac{1}{2}\operatorname{Tr}(A + B - |A - B|)$$

so that $\beta^{\mathrm{C}}_{n,p} = \frac{1}{2}\{1 - \|p\rho - (1-p)\sigma\|_1\}$. Hence by (3.3) we have

$$\beta^{\mathrm{C}}_{n,p} \le p^s(1-p)^{1-s}\operatorname{Tr}\rho^s\sigma^{1-s}, \qquad 0 \le s \le 1,$$

which immediately implies the direct part of (2). The original proof of (3.3) in [9] is somewhat complicated but a short proof was given by N. Ozawa as included in [36]. An extension of (3.3) to the general von Neumann algebra setting was in [56]. On the other hand, the large deviation principle was efficiently used in the proof of the converse part $\liminf_n \frac{1}{n}\log\beta^{\mathrm{C}}_{n,p} \geq -C(\rho\|\sigma)$ in [55]. Since $2\log F(\rho,\sigma) \leq -C(\rho\|\sigma)$ by Corollary 3.5, we have $\lim_{n\to\infty}\frac{1}{n}\log\beta^{\mathrm{C}}_{n,p} \geq 2\log F(\rho,\sigma)$. The lower bound $2\log F(\rho,\sigma)$ may be convenient since taking min is unnecessary to compute.

In the papers [29, 30, 52, 50, 28], results on quantum hypothesis testing extending Theorem 3.9 were obtained for states of non-i.i.d. type (i.e., states having correlations among sites) such as Gibbs states and quasi-free states on a spin chain $\bigotimes_{-\infty}^{\infty}\mathcal{A}$.

## 4. Appendices from matrix analysis

In this section, for the convenience of the reader, we will survey two topics in matrix analysis, which are selected from materials used in the main body of the paper and have their own mathematical interest and importance. We will give more comprehensive exposition than necessary in the main body to make the section interesting by itself.

4.1. **Operator monotone and convex functions: Integral expressions.** The theory of operator monotone and convex functions was initiated by Löwner [49] and Kraus [44], whose modern treatment was, after studies of Bendat and Sherman [11] and others, established by Hansen and Pedersen [23]. The theory has played quite an important role in the study of operator and matrix analysis. A real function $f$ on an interval $J$ is said to be *operator monotone* if, for every natural number $n$ and every $A, B \in M_n(\mathbb{C})^{sa}$ (the $n\times n$ Hermitian matrices) with the spectra $\mathrm{Sp}(A), \mathrm{Sp}(B) \subset J$,

$$A \leq B \implies f(A) \leq f(B).$$

Also, $f$ is said to be *operator convex* if, for every natural number $n$ and every $A, B \in M_n(\mathbb{C})^{sa}$ with $\mathrm{Sp}(A), \mathrm{Sp}(B) \subset J$,

$$f(\lambda A + (1-\lambda)B) \leq \lambda f(A) + (1-\lambda)f(B), \qquad 0 < \lambda < 1.$$

Furthermore, $f$ is said to be *operator monotone decreasing* if $-f$ is operator monotone, and *operator concave* if $-f$ is operator convex. Each of the above definitions is equivalent to saying that the same property holds for every $A, B \in B(\mathcal{H})^{sa}$ with a fixed infinite-dimensional Hilbert space $\mathcal{H}$. A power function $x^p$ on $(0,\infty)$ is operator monotone and operator concave if $p \in [0,1]$, operator convex if $p \in [1,2]$, and operator convex and operator monotone decreasing if $p \in [-1,0]$. When $p \in (-\infty,-1)\cup(2,\infty)$, $x^p$ is convex but not operator convex.

A complex function $f$ on $\mathbb{C}^+ := \{z \in \mathbb{C} : \mathrm{Im}\, z > 0\}$ is called a *Pick function* if $f$ is analytic in $\mathbb{C}^+$ and maps $\mathbb{C}^+$ into $\{z \in \mathbb{C} : \mathrm{Im}\, z \geq 0\}$. The famous theorem of Löwner says that a real function $f$ on an open interval $(a,b)$ where $-\infty \leq a < b \leq \infty$ is operator monotone if and only if $f$ is continuously extended to a Pick function (hence by reflection $f$ can be extended to an analytic function in $(\mathbb{C}\setminus\mathbb{R})\cup(a,b)$). Moreover, it is known from Kraus' theorem [44] that a real function $f$ on $(a,b)$ is operator convex if and only if, for every (equivalently for some) $c \in (a,b)$, $(f(x)-f(c))/(x-c)$ is operator monotone on $(a,b)$. See [17, 3, 12, 27] for detailed expositions on operator monotone and convex functions including the

above-mentioned results. A main goal in the theory of those functions is to obtain their integral expressions. The standard way to do so is to first obtain the integral expression of operator monotone functions on $(-1,1)$, as done in [23] by use of the Krein-Milman theorem, and then to transform it to operator monotone functions on, for example, $(0,\infty)$ or $[0,\infty)$. It is rather easy to show Löwner's theorem mentioned above from the integral expression of operator monotone functions. Conversely, once Löwner's theorem is at our disposal, we have the integral expression by modifying Nevanlinna's integral expression for general Pick functions.

In the rest of this section, we restrict our exposition to integral expressions of operator monotone and convex functions on the positive half line $[0,\infty)$ or $(0,\infty)$, which are of wide use in quantum information theory.

**(1)** A continuous real function $f$ on the half line $[0,\infty)$ is operator monotone if and only if there exist a constant $b \geq 0$ and a finite positive measure $\mu$ on $(0,\infty)$ such that

$$f(x) = f(0) + bx + \int_{(0,\infty)} \frac{x(1+\lambda)}{x+\lambda}\, d\mu(\lambda), \qquad x \in [0,\infty). \tag{4.1}$$

Here, $b$ and $\mu$ are unique with $b = \lim_{x\to\infty} f(x)/x$. The measure $\mu$ is called the representing measure of $f$. It is well known that an operator monotone function on $[0,\infty)$ is operator concave, which is seen from the above integral expression since $x/(x+\lambda) = 1-\lambda(x+\lambda)^{-1}$ is operator concave. The well-known integral expression of a typical operator monotone function $x^p$ $(0<p<1)$ on $[0,\infty)$ is

$$x^p = \frac{\sin p\pi}{\pi} \int_0^\infty \frac{x\lambda^{p-1}}{x+\lambda}\, d\lambda, \qquad x \in [0,\infty),$$

so the representing measure of $x^p$ is $(\sin p\pi/\pi)(\lambda^{p-1}/(1+\lambda))\, d\lambda$.

**(2)** The operator monotone function $\log x$ on the half open interval $(0,\infty)$ does not admit an integral expression of the form (4.1) but instead it is represented as

$$\log x = \int_0^\infty \left(\frac{1}{\lambda+1} - \frac{1}{x+\lambda}\right) d\lambda = \int_0^\infty \left(\frac{\lambda}{\lambda^2+1} - \frac{1}{x+\lambda}\right) d\lambda.$$

It can be seen from Nevanlinna's integral representation theorem that a real function on $(0,\infty)$ is operator monotone if and only if there exist constants $a \in \mathbb{R}$, $b \geq 0$ and a positive measure $\mu$ on $[0,\infty)$ with

$$\int_{[0,\infty)} \frac{1}{(1+\lambda)^2}\, d\mu(\lambda) < +\infty$$

such that

$$f(x) = a + bx + \int_{[0,\infty)} \left(\frac{\lambda}{\lambda^2+1} - \frac{1}{x+\lambda}\right) d\mu(\lambda), \qquad x \in (0,\infty). \tag{4.2}$$

Noting that

$$f(1) = a + b + \int \left(\frac{\lambda}{\lambda^2+1} - \frac{1}{\lambda+1}\right) d\mu(\lambda),$$

one can rewrite (4.2) as

$$f(x) = f(1) + b(x-1) + \int_{[0,\infty)} \left(\frac{1}{\lambda+1} - \frac{1}{x+\lambda}\right) d\mu(\lambda), \qquad x \in (0,\infty). \tag{4.3}$$

Furthermore, for each $\alpha \in (0,\infty)$, writing down $f(x) - f(\alpha)$ we have also

$$f(x) = f(\alpha) + b(x-\alpha) + \int_{[0,\infty)} \frac{x-\alpha}{(x+\lambda)(\alpha+\lambda)}\, d\mu(\lambda), \qquad x \in (0,\infty), \tag{4.4}$$

where the measure $\mu$ is independent of the choice of $\alpha$. Integral expression (4.4) was used in a recent paper [19]. Recently in [22], Hansen gave a proof of (4.2) that is more direct without transforming from the integral expression on $(-1,1)$ and shorter than the existing ones.

**(3)** A non-negative function $f$ on $(0,\infty)$ is operator monotone decreasing if and only if there exist a constant $c \geq 0$ and a finite positive measure $\mu$ on $[0,\infty)$ such that

$$f(x) = c + \int_{[0,\infty)} \frac{1+\lambda}{x+\lambda}\, d\mu(\lambda), \qquad x \in (0,\infty).$$

Here, $c$ and $\mu$ are unique with $c = \lim_{x\to\infty} f(x)$. The proof is found in [21] (also [5]).

**(4)** A real function $f$ on $[0,\infty)$ is operator convex and has finite $f'(0)$ (more precisely, the right derivative $f'_+(0)$) if and only if there exist constants $a \in \mathbb{R}$, $b \geq 0$ and a finite positive measure $\mu$ on $(0,\infty)$ such that

$$f(x) = f(0) + ax + bx^2 + \int_{(0,\infty)} \frac{x^2(1+\lambda)}{x+\lambda}\, d\mu(\lambda), \qquad x \in [0,\infty). \tag{4.5}$$

Here, $a$, $b$ and $\mu$ are unique with $a = f'(0)$, $b = \lim_{x\to\infty} f(x)/x^2$. This can be proved by applying integral expression (4.1) to the operator monotone function $(f(x) - f(0))/x$ on $[0,\infty)$.

**(5)** Integral expression (4.5) is not available unless $f'(0)$ exists. As shown in [31], a real continuous function $f$ on $[0,\infty)$ is operator convex if and only if there exist constants $a \in \mathbb{R}$, $b \geq 0$ and a positive measure $\mu$ on $(0,\infty)$ with $\int_{(0,\infty)} (1+\lambda)^{-2}\, d\mu(\lambda) < +\infty$ such that

$$f(x) = f(0) + ax + bx^2 + \int_{(0,\infty)} \left( \frac{x}{1+\lambda} - \frac{x}{x+\lambda} \right) d\mu(\lambda), \qquad x \in [0,\infty). \tag{4.6}$$

Here, $a$, $b$ and $\mu$ are unique with $b = \lim_{x\to\infty} f(x)/x^2$, $a = f(1) - f(0) - b$. As, for instance, the integral representation of $x \log x$ $(x \geq 0)$

$$x \log x = \int_{(0,\infty)} \left( \frac{x}{1+\lambda} - \frac{x}{x+\lambda} \right) d\lambda$$

shows, it is not necessarily true that $\int_{(0,\infty)} (1+\lambda)^{-1}\, d\mu(\lambda) < +\infty$. Hence, we cannot combine $ax$ and $\int_{(0,\infty)} x(1+\lambda)^{-1}\, d\mu(\lambda)$ together in expression (4.6). This is possible if and only if $\lim_{x\to\infty} f(x)/x < +\infty$. Integral expression (4.6) played an important role to prove Theorems 2.7 and 2.9 in [31]. To prove (4.6), apply integral expression (4.3) to $(f(x) - f(0))/x$ (the proof in [31] is a bit more complicated).

4.2. **Positive maps between $C^*$-algebras.** Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be $C^*$-algebras, and $\Phi : \mathcal{A}_1 \to \mathcal{A}_2$ be a linear map. If $\Phi(A^*A) \geq 0$ $(A \in \mathcal{A}_1)$, then $\Phi$ is said to be *positive*. For each natural number $n$, if the map $\Phi_n : \mathcal{A}_1 \otimes M_n(\mathbb{C}) \to \mathcal{A}_2 \otimes M_n(\mathbb{C})$ defined as

$$\Phi_n([A_{ij}]_{i,j=1}^n) := [\Phi(A_{ij})]_{i,j=1}^n \qquad (A_{ij} \in \mathcal{A}_1)$$

is positive, then $\Phi$ is said to be *$n$-positive*. The map $\Phi$ is said to be *completely positive* if it is $n$-positive for every $n$. We also say that $\Phi$ is a *Schwarz map* if $\Phi(A^*A) \geq \Phi(A)^*\Phi(A)$ $(A \in \mathcal{A}_1)$. Among these positivity conditions we have

$$\text{completely positive} \Rightarrow \cdots \Rightarrow \text{3-positive} \Rightarrow \text{2-positive} \Rightarrow \text{Schwarz} \Rightarrow \text{positive}$$

(the converse does not hold for each implication, see [72]).

In quantum information theory, a (quantum) information channel is usually defined to be a trace-preserving and completely positive map between finite-dimensional $C^*$-algebras. Below we discuss in the setting of finite-dimensional $\mathcal{A}_1 = B(\mathcal{H}_1) = M_{d_1}(\mathbb{C})$ and $\mathcal{A}_2 = B(\mathcal{H}_2) = M_{d_2}(\mathbb{C})$. For a linear map $\Phi : \mathcal{A}_1 \to \mathcal{A}_2$ the dual map $\Phi^* : \mathcal{A}_2 \to \mathcal{A}_1$ is defined as

$$\langle \Phi(A_1), A_2 \rangle_{\mathrm{HS}} = \langle A_1, \Phi^*(A_2) \rangle_{\mathrm{HS}} \qquad (A_1 \in \mathcal{A}_1,\ A_2 \in \mathcal{A}_2).$$

Note that $\Phi$ preserves traces (i.e., $\mathrm{Tr} \circ \Phi = \mathrm{Tr}$) if and only if $\Phi^*$ preserves units (i.e., $\Phi^*(I_2) = I_1$), and that $\Phi$ is positive (resp., $n$-positive, completely positive) if and only if $\Phi^*$ is positive (resp., $n$-positive, completely positive). The next theorem summarizes the characterization and representation results for completely positive maps between finite-dimensional $C^*$-algebras (or matrix algebras). Let $\{E_{ij} : i, j = 1, \ldots, d_1\}$ be the matrix units of $M_{d_1}(\mathbb{C})$ (i.e., $E_{ij}$ is the matrix of $(i, j)$-entry 1 and all others 0).

**Theorem 4.1** ([15, 37, 45, 69])**.** *For a linear map $\Phi : \mathcal{A}_1 \to \mathcal{A}_2$ the following conditions are equivalent:*

(i) *$\Phi$ is completely positive;*
(ii) *$\sum_{i,j=1}^{d_1} E_{ij} \otimes \Phi(E_{ij}) \geq 0$ as an element of $\mathcal{A}_1 \otimes \mathcal{A}_2$;*
(iii) *there exists a linear map $V_i : \mathcal{H}_1 \to \mathcal{H}_2$ $(i = 1, \ldots, r)$ such that $\Phi(A) = \sum_{i=1}^{r} V_i A V_i^*$ $(A \in \mathcal{A}_1)$;*
(iv) *there exist a (finite-dimensional) Hilbert space $\mathcal{K}$, a representation ($*$-homomorphism) $\pi : B(\mathcal{H}_1) \to B(\mathcal{K})$ and an operator $V : \mathcal{H}_2 \to \mathcal{K}$ such that $\Phi(A) = V^*\pi(A)V$ $(A \in \mathcal{A}_1)$.*

The characterization in (ii) is useful to check the complete positivity. The element $\sum_{i,j=1}^{d_1} E_{ij} \otimes \Phi(E_{ij})$ in $\mathcal{A}_1 \otimes \mathcal{A}_2$ is called the *Choi matrix* or the *Choi-Jamiołkowski correspondence* ([14, 37]), which is represented as a block matrix ($d_1 d_2 \times d_1 d_2$ matrix) $[\Phi(E_{ij})]_{i,j=1}^{d_1}$ of $(i, j)$-entry $\Phi(E_{ij})$. The representation in (iii) is called the *Kraus representation* or the *Choi-Kraus representation* ([14, 45]). (iv) is called the *Stinespring representation*, a well-known representation theorem in operator algebras, that holds for completely positive maps between general $C^*$-algebras though $\mathcal{K}$ is no longer finite-dimensional. Moreover, (ii) shows that $\Phi$ is completely positive if it is $d_1$-positive. Since the same holds for $\Phi^*$ as well, it is seen that $\Phi$ is completely positive if it is $n$-positive with $n := \min\{d_1, d_2\}$. See [13, 67] for completely positive maps in matrix algebras, including the proof of the above theorem. Also, a detailed account on completely positive (and completely bounded) maps in general operator algebras is found in [61] for example.

A linear map $\Phi : \mathcal{A}_1 \to \mathcal{A}_2$ is said to be *completely copositive* if $A \in \mathcal{A}_1 \mapsto \Phi(A^t)$ ($A^t$ is the transpose of $A$). A completely copositive map is positive. Furthermore, a positive map $\Phi : \mathcal{A}_1 \to \mathcal{A}_2$ is said to be *decomposable* if it is given as the sum $\Phi(A) = \Psi(A) + \hat{\Psi}(A)$ $(A \in \mathcal{A}_1)$ of a completely positive map $\Psi : \mathcal{A}_1 \to \mathcal{A}_2$ and a completely copositive map $\hat{\Psi} : \mathcal{A}_1 \to \mathcal{A}_2$. Concerning decomposable maps between

low-dimensional matrix algebras, the next theorem combining results of Choi and of Woronowicz is fundamental (there are many other references, for example, [71, 60]).

**Theorem 4.2** ([15, 78]). *When $d_1 \cdot d_2 \leq 6$, any positive map $\Phi : M_{d_1}(\mathbb{C}) \to M_{d_2}(\mathbb{C})$ is decomposable. But when $d_1 \cdot d_2 > 6$ (in particular, $d_1 = d_2 = 3$), there is a non-decomposable positive map from $M_{d_1}(\mathbb{C})$ to $M_{d_2}(\mathbb{C})$.*

## Acknowledgments

The author would like to thank Dr. Milán Mosonyi who pointed out a number of relevant references in quantum information, and Dr. Hiromichi Ohno for his guidance on quantum Markov states.

**Added in proof.** The original Japanese version of this article was published in 2013. Since then, quite a few new developments have been made in the subject of the article. The most relevant issue is new types of Rényi divergences, which we briefly survey for the reader's convenience. For simplicity we here assume that $A, B$ are positive invertible operators (or positive definite matrices) in $\mathcal{A} = M_d(\mathbb{C})$. For $\alpha > 0$ with $\alpha \neq 1$, the traditional Rényi divergence (i.e., Rényi $\alpha$-relative entropy) is

$$D_\alpha(A\|B) := \frac{1}{\alpha - 1} \log \frac{\operatorname{Tr} A^\alpha B^{1-\alpha}}{\operatorname{Tr} A}$$

by slightly modifying the definition of Example 2.5, with a notation $D_\alpha$ different from $S_\alpha$ in accordance with other new ones below. A new version called the *sandwiched Rényi divergence* is

$$D^*_\alpha(A\|B) := \frac{1}{\alpha - 1} \log \frac{\operatorname{Tr}(B^{\frac{1-\alpha}{2\alpha}} A B^{\frac{1-\alpha}{2\alpha}})^\alpha}{\operatorname{Tr} A},$$

whose detailed study was first made in [A11, A12]. Unlike $D_\alpha$, note that $D^*_\alpha$ is not in the class of quantum $f$-divergences. Also, note that $\lim_{\alpha\to 1} D^*_\alpha(A\|B) = S(A\|B)$. Another quantum divergence called the *$\alpha$-$z$-Rényi divergence* is

$$D_{\alpha,z}(A\|B) := \frac{1}{\alpha - 1} \log \frac{\operatorname{Tr}(B^{\frac{1-\alpha}{2z}} A^{\frac{\alpha}{z}} B^{\frac{1-\alpha}{2z}})^z}{\operatorname{Tr} A},$$

which was first introduced in [36] and studied in further detail in [A1]. See [A11, A12] and [A1] for the precise definitions of $D^*_\alpha$ and $D_{\alpha,z}$ for general positive operators $A, B$ (with $A \neq 0$). The $\alpha$-$z$-Rényi divergence is the generalization of $D_\alpha$ and $D^*_\alpha$ together; indeed, $D_\alpha = D_{\alpha,1}$, $D^*_\alpha = D_{\alpha,\alpha}$, and $D^*_{1/2}(\rho\|\sigma) = -2 \log F(\rho, \sigma)$ for density operators $\rho, \sigma$.

The most important problem on the sandwiched divergence as well as the $\alpha$-$z$-divergence is to determine the range of the parameter $\alpha$ (resp. $\alpha, z$) for which $D^*_\alpha$ (resp. $D_{\alpha,z}$) satisfies monotonicity under quantum channels similarly to $f$-divergences (see Theorem 2.7). Monotonicity inequality or DPI (Data processing inequality) of $D^*_\alpha$ was discussed in several papers [A11, A12, A4, A2, A10] and the result is that $D^*_\alpha$ satisfies monotonicity if and only if $\alpha \geq 1/2$, while $D_\alpha$ does if and only if $0 < \alpha \leq 2$. On the other hand, some range of $\alpha, z$ for which $D_{\alpha,z}$ satisfies monotonicity was found in [A1, A3] (also [A5]) but a complete characterization of all $\alpha, z$ values for monotonicity is still missing.

In [31] we showed several conditions for reversibility of quantum operations via the equality in the monotonicity inequality of $f$-divergences (see Theorem 2.9),

which was complemented in [A7] and furthermore in a recent paper [A6]. An algebraic characterization of the equality in the monotonicity inequality of $D_\alpha^*$ for $\alpha > 1/2$ was given in [A9]. It was shown in [A8] that the equality case in the monotonicity inequality of $D_\alpha^*$ for $\alpha > 1$ implies reversibility. Moreover, reversibility via the equality in the monotonicity of $D_{\alpha,z}$ was discussed in [A6] in certain restricted situations.

More interestingly, it was shown in [A10] that in the strong converse problem of quantum hypothesis testing, the optimal exponent of success probabilities $1 - \alpha_n(T_n) = \operatorname{Tr} \rho^{\otimes n} T_n$ under $\beta_n(T_n) = \operatorname{Tr} \sigma^{\otimes n} T_n < e^{-nr}$ for $r > S(\rho\|\sigma)$ and $0 \le T_n \le I$ in $\mathcal{A}^{\otimes n}$ can be described in terms of the sandwiched divergence $D_\alpha^*$ for $\alpha > 1$. On the other hand, the exponent of the error probability in the direct part of similar quantum hypothesis testing, known as the quantum Hoeffding bound, is given in terms of the traditional Rényi divergence $D_\alpha$ for $\alpha < 1$ (see Theorem 3.9 (3) and (3.2)). This suggests that the operationally relevant definition of the quantum Rényi divergences depends on the parameter $\alpha$ in such a way that the right choice for $\alpha > 1$ should be $D_\alpha^*$ though for $\alpha < 1$ the right choice is $D_\alpha$.

In this way, it seems reasonable to state that the quantum divergences most relevant to recent developments of quantum information are old and new quantum Rényi divergences, including the usual relative entropy as a special case.

Finally, related to Theorem 4.2, it is worth noting that an advance on decomposable positive maps between matrix algebras was recently made in [A13], giving an affirmative answer to the conjecture in [46] that every 2-positive (also 2-copositive) map from $M_3(\mathbb{C})$ to $M_3(\mathbb{C})$ is decomposable.

## References


[1] L. Accardi and V. Liebscher, *Markovian KMS-states for one-dimensional spin chains*, Infin. Dimens. Anal. Quantum Probab. Relat. Top. **2** (1999), 645–661. MR1810817 (2002d:82006)

[2] R. Alicki and M. Fannes, *Continuity of quantum conditional information*, J. Phys. A: Math. Gen. **37** (2004), L55–L57. MR2043448 (2004m:82005)

[3] T. Ando, *Topics on Operator Inequalities*, Lecture notes (mimeographed), Hokkaido Univ., Sapporo, 1978 MR0482378 (58:2451)

[4] T. Ando, *Concavity of certain maps on positive definite matrices and applications to Hadamard Products*, Linear Algebra Appl. **26** (1979), 203–241. MR535686 (80f:15023)

[5] T. Ando and F. Hiai, *Operator log-convex functions and operator means*, Math. Ann. **350** (2011), 611–630. MR2805638 (2012d:47048)

[6] H. Araki, *Relative entropy of states of von Neumann algebras I, II*, Publ. Res. Inst. Math. Sci. **11** (1976), 809–833; **13** (1977), 173–192. MR0425631 (54:13585); MR0454656 (56:12905)

[7] H. Araki and E. H. Lieb, *Entropy Inequalities*, Comm. Math. Phys. **18** (1970), 160–170. MR0266563 (42:1466)

[8] K. M. R. Audenaert, *A sharp continuity estimate for the von Neumann entropy*, J. Phys. A: Math. Theor. **40** (2007), 8127. MR2344161 (2008m:82009)

[9] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, Ll. Masanes, A. Acin and F. Verstraete, *Discriminating states: The quantum Chernoff bound*, Phys. Rev. Lett. **98** (2007), 160501.

[10] K. M. R. Audenaert, M. Nussbaum, A. Szkoła and F. Verstraete, *Asymptotic error rates in quantum hypothesis testing*, Comm. Math. Phys. **279** (2008), 251–283. MR2377635 (2009a:81014)

[11] J. Bendat and S. Sherman, *Monotone and convex operator functions*, Trans. Amer. Math. Soc. **79** (1955), 58–71. MR0082655 (18:588b)

[12] R. Bhatia, *Matrix Analysis*, Springer, New York, 1996. MR1477662 (98i:15003)

[13] R. Bhatia, *Positive Definite Matrices*, Princeton Univ. Press, Princeton, 2007. MR2284176 (2007k:15005)

[14] M.-D. Choi, *Completely positive linear maps on complex matrices*, Linear Algebra Appl. **10** (1975), 285–290. MR0376726 (51:12901)
[15] M.-D. Choi, *Positive semidefinite biquadratic forms*, Linear Algebra Appl. **12** (1975), 95–100. MR0379365 (52:270)
[16] M. Christandl and A. Winter, *"Squashed entanglement": An additive entanglement measure*, J. Math. Phys. **45** (2004), 829–840. MR2036165 (2004m:81028)
[17] W. F. Donoghue, Jr., *Monotone Matrix Functions and Analytic Continuation*, Springer, Berlin-Heidelberg-New York, 1974 MR0486556 (58:6279)
[18] M. Fannes, *A continuity property of entropy density for spin lattice systems*, Comm. Math. Phys. **31** (1973), 291–294. MR0345574 (49:10309b)
[19] U. Franz, F. Hiai and É. Ricard, *Higher order extension of Löwner's theory: Operator k-tone functions*, Trans. Amer. Math. Soc., to appear (arXiv:1105.3881). MR3180739
[20] C. A. Fuchs and J. van de Graaf, *Cryptographic distinguishability measures for quantum mechanical states*, IEEE Trans. Inf. Theory **45** (1999), 1216–1227. MR1686254 (2000g:94031)
[21] F. Hansen, *Trace functions as Laplace transforms*, J. Math. Phys. **47**, (2006), 043504, 1–11. MR2226341 (2007g:47024)
[22] F. Hansen, The fast track to Löwner's theorem, *Linear Algebra Appl.* **438** (2013), 4557–4571. MR3034551
[23] F. Hansen and G. K. Pedersen, *Jensen's inequality for operators and Löwner's theorem*, Math. Ann. **258** (1982), 229–241. MR1513286
[24] M. Hayashi, *Quantum Information: An Introduction*, Springer, Berlin, 2006. MR2228302 (2007b:81050)
[25] M. Hayashi, *Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding*, Phys. Rev. A **76** (2007), 062301.
[26] P. Hayden, R. Jozsa, D. Petz and A. Winter, *Structure of states which satisfy strong subadditivity of quantum entropy with equality*, Comm. Math. Phys. **246** (2004), 359–374. MR2048562 (2005a:82016)
[27] F. Hiai, *Matrix Analysis: Matrix Monotone Functions, Matrix Means, and Majorization (GSIS selected lectures)*, Interdisciplinary Information Sciences **16** (2010), 139–248. MR2761752 (2011k:15040)
[28] F. Hiai, M. Mosonyi and M. Hayashi, *Quantum hypothesis testing with group symmetry*, J. Math. Phys. **50** (2009), 103304, 1–31. MR2573132 (2010m:81073)
[29] F. Hiai, M. Mosonyi and T. Ogawa, *Large deviations and Chernoff bound for certain correlated states on the spin chain*, J. Math. Phys. **48** (2007), 123301. MR2377831 (2009d:82020)
[30] F. Hiai, M. Mosonyi and T. Ogawa, *Error exponents in hypothesis testing for correlated states on a spin chain*, J. Math. Phys. **49** (2008), 032112. MR2406780 (2009g:81034)
[31] F. Hiai, M. Mosonyi, D. Petz and C. Bény, *Quantum f-divergences and error correction*, Rev. Math. Phys. **23** (2011), 691–747. MR2826462 (2012j:81020)
[32] F. Hiai, M. Ohya and M. Tsukada, *Sufficiency, KMS condition and relative entropy in von Neumann algebras*, Pacific J. Math. **96** (1981), 99–109. MR634765 (84f:46082)
[33] F. Hiai and D. Petz, *The proper formula for relative entropy and its asymptotics in quantum probability*, Comm. Math. Phys. **143** (1991), 99–114. MR1139426 (93b:46122)
[34] M. Horodecki, Entanglement measures, *Quantum Information and Computation* **1** (2001), 3–26. MR1910008 (2003c:81026)
[35] M. Horodecki, P. Horodecki and R. Horodecki, *Separability of mixed states: necessary and sufficient conditions*, Physics Letters A **223** (1996), 1–8. MR1421501 (97k:81009)
[36] V. Jakšić, Y. Ogata, Y. Pautrat and C.-A. Pillet, *Entropic fluctuations in quantum statistical mechanics. an introduction*, in: Quantum Theory from Small to Large Scales (École de Physique des Houches Session XCV 2010), J. Fröhlich et al. (eds.), Oxford University Press, 2012.
[37] A. Jamiołkowski, *Linear transformations which preserve trace and positive semidefiniteness of operators*, Rep. Math. Phys. **3** (1972), 275–278. MR0342537 (49:7283)
[38] A. Jenčová and D. Petz, *Sufficiency in quantum statistical inference*, Comm. Math. Phys. **263** (2006), 259–276. MR2207329 (2006m:81154)
[39] A. Jenčová and D. Petz, *Sufficiency in quantum statistical inference. A survey with examples*, Infin. Dimens. Anal. Quantum Probab. Relat. Top. **9** (2006), 331–351. MR2256497 (2007g:81046)

[40] A. Jenčová, D. Petz and J. Pitrik, *Markov triplets on CCR algebras*, Acta Sci. Math. (Szeged) **76** (2010), 27–50. MR2668410 (2012c:46178)
[41] G. Kimura, T. Miyadera and H. Imai, *Optimal state discrimination in general probabilistic theories*, Phys. Rev. A **79** (2009), 062306.
[42] H. Kosaki, *Interpolation theory and the Wigner-Yanase-Dyson-Lieb concavity*, Comm. Math. Phys. **87** (1982), 315–329. MR682110 (84h:46101)
[43] H. Kosaki, *Relative entropy of states: a variational expression*, J. Operator Theory **16** (1986), 335–348. MR860352 (87j:46110)
[44] F. Kraus, *Über konvexe Matrixfunktionen*, Math. Z. **41** (1936), 18–42. MR1545602
[45] K. Kraus, *General state changes in quantum theory*, Annals of Physics **64** (1971), 311–335. MR0292434 (45:1520)
[46] S.-H. Kye, *Facial structures for various notions of positivity and applications to the theory of entanglement*, Rev. Math. Phys. **25** (2013), 1330002 (52 pages). MR3040810
[47] E. H. Lieb, *Convex trace functions and the Wigner-Yanase-Dyson conjecture*, Advances in Math. **11** (1973), 267–288. MR0332080 (48:10407)
[48] E. H. Lieb and M. B. Ruskai, *Proof of the strong subadditivity of quantum-mechanical entropy*, J. Math. Phys. **14** (1973), 1938–1941. MR0345558 (49:10294)
[49] K. Löwner, *Über monotone Matrixfunctionen*, Math. Z. **38** (1934), 177–216. MR1545446
[50] M. Mosonyi, *Hypothesis testing for Gaussian states on bosonic lattices*, J. Math. Phys. **50** (2009), 032105. MR2510896 (2010j:82012)
[51] M. Mosonyi and F. Hiai, *On the quantum Rényi relative entropies and related capacity formulas*, IEEE Trans. Inform. Theory **57** (2011), 2474–2487. MR2809103 (2012f:81050)
[52] M. Mosonyi, F. Hiai, T. Ogawa and M. Fannes, *Asymptotic distinguishability measures for shift-invariant quasi-free states of fermionic lattice systems*, J. Math. Phys. **49** (2008), 072104. MR2432028 (2009g:81020)
[53] H. Nagaoka, *The converse part of the theorem for quantum Hoeffding bound*, Preprint, 2006 (arXiv:quant-ph/0611289).
[54] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge, 2000. MR1796805 (2003j:81038)
[55] M. Nussbaum and A. Szkoła, *A lower bound of Chernoff type for symmetric quantum hypothesis testing*, Ann. Statist. **37** (2009), 1040–1057. MR2502660 (2010i:62306)
[56] Y. Ogata, *A generalization of Powers-Størmer inequality*, Lett. Math. Phys. **97** (2011), 339–346. MR2826816 (2012f:46117)
[57] T. Ogawa and H. Nagaoka, *Strong converse and Stein's lemma in quantum hypothesis testing*, IEEE Trans. Inform. Theory **47** (2000), 2428–2433. MR1806811 (2001m:94025)
[58] H. Ohno, *Translation-invariant quantum Markov states*, Interdisciplinary Information Sciences **10** (2004), 53–58. MR2062192 (2005b:81084)
[59] M. Ohya and D. Petz, *Quantum Entropy and Its Use*, 2nd ed., Springer, Heidelberg, 2004. MR1230389 (94k:81002)
[60] H. Osaka, *A class of extremal positive maps in* $3 \times 3$ *matrix algebras*, Publ. Res. Inst. Math. Sci. **28** (1992), 747–756. MR1195997 (94a:46076)
[61] V. Paulsen, *Completely Bounded Maps and Operator Algebras*, Cambridge Studies in Advanced Mathematics 78, Cambridge Univ. Press, Cambridge, 2002. MR1976867 (2004c:46118)
[62] A. Peres, *Separability criterion for density matrices*, Phys. Rev. Lett. **77** (1996), 1413–1415. MR1401726 (97d:82004)
[63] D. Petz, *Quasi-entropies for states of a von Neumann algebra*, Publ. Res. Inst. Math. Sci. **21** (1985), 787–800. MR817164 (87e:46092)
[64] D. Petz, *Quasi-entropies for finite quantum systems*, Rep. Math. Phys. **23** (1986), 57–65. MR868631 (88a:46079)
[65] D. Petz, *Sufficiency of channels over von Neumann algebras*, Quart. J. Math. Oxford, **39** (1988), 907–1008. MR929798 (89f:46124)
[66] D. Petz, *Monotonicity of quantum relative entropy revisited*, Rev. Math. Phys. **15** (2003), 79–91. MR1961186 (2003m:82007)
[67] D. Petz, *Quantum Information Theory and Quantum Statistics*, Springer, Berlin-Heidelberg, 2008. MR2363070 (2009c:81026)
[68] M. B. Ruskai, *Inequalities for quantum entropy: A Review with conditions for equality*, J. Math. Phys. **43** (2002), 4358–4375. MR1924445 (2004e:82007)

[69] W. F. Stinespring, *Positive functions on $C^*$-algebras*, Proc. Amer. Math. Soc. **6** (1955), 211–216. MR0069403 (16:1033b)

[70] B. Synak-Radtke and M. Horodecki, *On asymptotic continuity of functions of quantum states*, J. Phys. A **39** (2006), L423–L437. MR2238498 (2007d:81031)

[71] K. Tanahasi and J. Tomiyama, *Indecomposable positive maps in matrix algebras*, Canad. Math. Bull. **31** (1988), 308–317. MR956361 (90a:46156)

[72] J. Tomiyama, *On the geometry of positive maps in matrix algebras. II*, Linear Algebra Appl. **69** (1985), 169–177. MR798371 (87b:46064)

[73] A. Uhlmann, *The "transition probability" in the state space of a *-algebra*, Rep. Math. Phys. **9** (1976), 273–279. MR0423089 (54:11072)

[74] A. Uhlmann, *Relative entropy and Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory*, Comm. Math. Phys. **54** (1977), 21–32. MR0479224 (57:18671)

[75] H. Umegaki, *Conditional expectation in an operator algebra, IV (entropy and information)*, Kōdai Math. Sem. Rep. **14** (1962), 59–85. MR0142006 (25:5401)

[76] R. F. Werner, *Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model*, Phys. Rev. A **40** (1989), 4277–4281.

[77] E. P. Wigner and M. M. Yanase, *Information contents of distributions*, Proc. Nat. Acad. Sci. USA **49** (1963), 910–918. MR0151127 (27:1113)

[78] S. L. Woronowicz, *Positive maps of low dimensional matrix algebras*, Rep. Math. Phys. **10** (1976), 165–183. MR573218 (81m:15014)

[79] D. Ye, *On the Bures volume of separable quantum states*, J. Math. Phys. **50** (2009), 083502. MR2554430 (2010j:81048)

[A1] K. M. R. Audenaert and N. Datta, *$\alpha$-$z$-Rényi relative entropies*, J. Math. Phys. **56** (2015), 022202.

[A2] S. Beigi, *Sandwiched Rényi divergence satisfies data processing inequality*, J. Math. Phys. **54** (2013), 122202.

[A3] E. A. Carlen, R. L. Frank and E. H. Lieb, *Some operator and trace function convexity theorems*, Linear Algebra Appl. **490** (2016), 174–185.

[A4] R. L. Frank and E. H. Lieb, *Monotonicity of a relative Rényi entropy*, J. Math. Phys. **54** (2013), 122201.

[A5] F. Hiai, *Concavity of certain matrix trace and norm functions*, Linear Algebra Appl. **439** (2013), 1568–1589.

[A6] F. Hiai and M. Mosonyi, *Different quantum $f$-divergences and the reversibility of quantum operations*, arXiv:1604.03089.

[A7] A. Jenčová, *Reversibility conditions for quantum operations*, Rev. Math. Phys. **24** (2012), 1250016, 26 pp.

[A8] A. Jenčová, *Preservation of a quantum Rényi relative entropy implies existence of a recovery map*, J. Phys. A: Math. Theor., to appear, 18 pp.

[A9] F. Leditzky, C. Rouzé and N. Datta, *Data processing for the sandwiched Rényi divergence: a condition for equality*, Lett. Math. Phys., published online: 15 November 2016, 20 pp.

[A10] M. Mosonyi and T. Ogawa, *Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies*, Comm. Math. Phys. **334** (2015), 1617–1648.

[A11] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr and M. Tomamichel, *On quantum Rényi entropies: A new generalization and some properties*, J. Math. Phys. **54** (2013), 122203.

[A12] M. M. Wilde, A. Winter and D. Yang, *Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy*, Comm. Math. Phys. **331** (2014), 593–622.

[A13] Y. Yang, D. H. Leung and W.-S. Tang, *All 2-positive linear maps from $M_3(\mathbb{C})$ to $M_3(\mathbb{C})$ are decomposable*, Linear Algebra Appl., **503** (2016), 233–247.

Translated by FUMIO HIAI

Tohoku University (Emeritus), Hakusan 3-8-16-303, Abiko 270-1154, Japan
*E-mail address*: hiai.fumio@gmail.com