Concavity of the Auxiliary Function for Classical-Quantum Channels

Hao-Chung Cheng and Min-Hsiu Hsieh

***Abstract*—The auxiliary function of a classical channel appears in two fundamental quantities, the random coding exponent and the sphere-packing exponent, which yield upper and lower bounds on the error probability of decoding, respectively. A crucial property of the auxiliary function is its concavity, and this property consequently leads to several important results in finite blocklength analysis. In this paper, we prove that the auxiliary function of a classical-quantum channel also enjoys the same concavity property, extending an earlier partial result to its full generality. We also prove that the auxiliary function satisfies the data-processing inequality, among various other important properties. Furthermore, we show that the concavity property of the auxiliary function enables a geometric interpretation of the random coding exponent and the sphere-packing exponent of a classical-quantum channel. The key component in our proof is an important result from the theory of matrix geometric means.**

I. Introduction

We consider a channel coding problem. Let $\mathcal{X} := \{1,2,\ldots,|\mathcal{X}|\}$ and $\mathcal{Y} = \{1,2,\ldots,|\mathcal{Y}|\}$ be the input and output alphabets of a discrete memoryless classical channel $Q(y|x)$. An n-length block code is a mapping from the message set $\mathcal{M} := \{1,2,\ldots,M\}$ to a sequence of n input symbols $\mathbf{x} = (x_1,x_2,\ldots,x_n) \in \mathcal{X}^n$. The probability of observing the sequence $\mathbf{y} = (y_1,y_2,\ldots,y_n) \in \mathcal{Y}^n$ at the output of the channel given input $\boldsymbol{x}$ is $Q(\mathbf{y}|\mathbf{x}) = \Pi_{i=1}^n Q(y_i|x_i)$. When message m is sent, the error probability of decoding is $\mathsf{P}_{\mathrm{e}|m} := 1 - \sum_{\mathbf{y}\in\mathcal{Y}_m} Q\left(\mathbf{y}|\mathbf{x}_m\right)$, where $\mathcal{Y}_m$ denotes the decoding region to the message m. The maximum error probability of the code is defined as

$$\mathsf{P}_{\mathrm{e,max}} := \max_{m\in\mathcal{M}} \mathsf{P}_{\mathrm{e}|m}.$$

Denote by $R := \frac{1}{n}\log M$ the rate of the code. Then we define $\mathsf{P}_{\mathrm{e}}(n,R)$ as the *smallest* maximum error probability among all codes of length n and rate at least R.

Let $\mathcal{P}(\mathcal{X})$ be the set of probability distributions on $\mathcal{X}$. For any fixed $P \in \mathcal{P}(\mathcal{X})$ and $s \geq 0$, the *auxiliary function* $E_0(s,P)$ of a classical communication channel $Q(y|x)$ is defined as

$$E_0(s,P) := -\log\left[\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} P(x)Q(y|x)^{\frac{1}{1+s}}\right)^{1+s}\right]. \tag{1}$$

This function appears in two fundamental quantities in classical channel coding: for any $R \geq 0$,

$$E_{\mathrm{r}}(R) := \max_{0\leq s\leq 1}\left\{\max_{P\in\mathcal{P}(\mathcal{X})} E_0(s,P) - sR\right\}, \tag{2}$$

and

$$E_{\mathrm{sp}}(R) := \sup_{s\geq 0}\left\{\max_{P\in\mathcal{P}(\mathcal{X})} E_0(s,P) - sR\right\}, \tag{3}$$

where $E_{\mathrm{r}}(R)$ is called the *random coding exponent* and $E_{\mathrm{sp}}(R)$ is called the *sphere-packing exponent* of the classical channel Q. These two quantities are critical since, for any blocklength n and any rate $0 \leq R \leq C$, where C denotes the capacity of the channel Q, the error probability $P_{\mathrm{e}}(n,R)$, minimized over all possible coding strategies, satisfies [1]–[3]

$$2^{-nE_{\mathrm{sp}}(R)} \lesssim P_{\mathrm{e}}(n,R) \lesssim 2^{-nE_{\mathrm{r}}(R)}, \tag{4}$$

where we write $f_n \lesssim g_n$ if $\limsup_{n\to\infty}\frac{1}{n}\log\frac{f_n}{g_n} \leq 0$. Consequently, properties of the auxiliary function $E_0(s,P)$ reveal important functional behaviour of the two exponents, and lead to a deeper understanding of the error probability of a given classical

Hao-Chung Cheng is with the Graduate Institute of Communication Engineering, National Taiwan University, Taiwan (R.O.C.) and the Centre for Quantum Computation & Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia. (email: Hao-Chung.Cheng@student.uts.edu.au)

Min-Hsiu Hsieh is with the Centre for Quantum Computation & Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia. (email: Min-Hsiu.Hsieh@uts.edu.au)

channel Q. It is well-known (and easy to show) [3]: $\forall s \geq 0$,

$$\begin{aligned} E_0(s,P) &\geq 0; & (5)\\ \frac{\partial E_0(s,P)}{\partial s} &> 0; & (6)\\ \frac{\partial^2 E_0(s,P)}{\partial s^2} &\leq 0. & (7) \end{aligned}$$

It turns out that $E_0(s,P)$ is concave in $s \geq 0$. In addition to other important contributions in finite blocklength analysis, this fact also provides an alternative proof to Shannon's noiseless channel coding theorem [4].

In recent years, much attention has been paid to understanding the reliable transmission of classical messages through a quantum channel. In this scenario, it suffices to consider a *classical-quantum channel*, which is a mapping $W : x \in \mathcal{X} \mapsto W_x \in \mathcal{S}(\mathcal{H})$ from the finite set $\mathcal{X}$ to $\mathcal{S}(\mathcal{H})$, the set of density operators (positive semi-definite operators with unit trace) on a fixed finite-dimensional Hilbert space $\mathcal{H}$. Given a (discrete memoryless)classical-quantum channel W and a distribution P on the input $\mathcal{X}$, we can similarly define the *auxiliary function* $E_0(s,P,W)$ [5], [6]: $\forall s \geq 0$,

$$E_0(s,P,W) := -\log \operatorname{Tr}\left[\left(\sum_{x\in\mathcal{X}} P(x)\cdot W_x^{\frac{1}{1+s}}\right)^{1+s}\right]. \quad (8)$$

This quantity is a quantum generalization of Eq. (1), and recovers Eq. (1) when all $\{W_x\}_{x\in\mathcal{X}}$ commute. When no confusion is possible, we ignore the argument W in $E_0(s,P,W)$.

The auxiliary function $E_0(s,P)$ in Eq. (8) also appears in the random coding exponent $E_\mathrm{r}(R)$ and the sphere-packing exponent $E_\mathrm{sp}(R)$ of a classical-quantum channel W, which can be similarly defined as that in Eqs. (2) and (3), respectively. However, relations between these two exponents and the error probability of the classical-quantum channel W are much harder to obtain. The random coding exponent $E_\mathrm{r}(R)$ was shown to be an upper bound to the error probability of a classical-quantum channel W when every W_x is pure (i.e. the density operator W_x is a rank-one matrix) in Ref. [5], and it is conjectured to hold for general quantum states. Furthermore, the sphere-packing bound that lower bounds the error probability of W was recently proved in Ref. [7]. These results are highly nontrivial due to the non-commutative nature of the density operators involved in their definitions. Furthermore, it was still unknown whether the auxiliary function $E_0(s,P)$ in Eq. (8) is concave for all $s \geq 0$. This might be one reason that the error probability of any finite blocklength n is less understood in the quantum regime. Note that $E_0(s,P)$ has been shown to be concave for $0 \leq s \leq 1$ in Ref. [8]. Its proof relies on an *ad-hoc* operator inequality in order to show that the second-order derivative of $E_0(s,P)$ is non-positive for $s \in [0,1]$. However, this method does not seem to extend for all $s \geq 0$.

In this paper, we prove that $E_0(s,P)$ of a classical-quantum channel W is concave for all $s \geq 0$. Our proof employs the recent developments in matrix algebra; in particular, the theory of matrix geometric means [9] (see [10], [11] for the general treatment). Our proof can be viewed as a direct generalization of the classical proof in Ref. [3, Theorem 5.6.3].

The paper is organized as follows. Section II presents the main technical tool, the "s-weighted geometric means". The main result and its proof are presented in Section III. We provide the properties of the auxiliary function and discuss how the concavity property of the auxiliary function enables a geometric interpretation of the random coding exponent and the sphere-packing exponent in Section IV. Finally we conclude this paper in Section V.

II. Technical Tools

Denote by $\mathbb{M}_d^+$ and $\mathbb{M}_d^{++}$ the set of $d\times d$ positive semi-definite matrices and positive definite matrices, respectively. For two $d\times d$ Hermitian matrices A and B, we denote by $A \succeq B$ if $A - B \in \mathbb{M}_d^+$. For $A, B \in \mathbb{M}_d^{++}$, the "$s$-weighted geometric mean" of A and B is defined as

$$A\#_s B := A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^s A^{1/2}. \quad (9)$$

The geometric means enjoy the following properties [9], [12], [13] (see also [10, Chapter 6], [14, Section 4] and [11, Chapter 5]).

Proposition 1 (Properties of Geometric Means). *Let $A, B, C, D \in \mathbb{M}_d^{++}$ and $s \in \mathbb{R}$. Then*

(a) Commutativity: $A\#_s B = A^{1-s}B^s$ *for* $AB = BA$*;*

(b) Joint homogeneity: $(aA)\#_s(bB) = a^{1-s}b^s(A\#_sB)$ *for* $a, b > 0$*;*

(c) Monotonicity: $A\#_sB \preceq C\#_sD$ *for* $A \preceq C$*,* $B \preceq D$ *and* $s \in [0,1]$*;*

(d) Congruence invariance: *For every non-singular matrix* M*,* $M(A\#_sB)M^\dagger = \left(MAM^\dagger\right)\#_s\left(MBM^\dagger\right)$*;*

(e) Self-duality: $A\#_sB = B\#_{1-s}A$*, and* $(A\#_sB)^{-1} = A^{-1}\#_sB^{-1}$*;*

(f) Concavity:

$$\begin{aligned}&(\theta A + (1-\theta)B)\#_s(\theta C + (1-\theta)D)\\ &\succeq \theta(A\#_sC) + (1-\theta)(B\#_sD)\end{aligned} \tag{10}$$

for all $\theta, s \in [0,1]$*;*

(g) HM-GM-AM inequality:

$$\begin{aligned}&\left((1-s)A^{-1} + sB^{-1}\right)^{-1}\\ &\preceq A\#_sB \preceq (1-s)A + sB\end{aligned}$$

for all $s \in [0,1]$.

(h) Continuity: $A\#_sB$ *is continuous in* A *and* B *with respect to the strong topology.*

Let $x := (x_1, \ldots, x_d) \in \mathbb{R}^d$ be a d-dimensional vector with positive elements. Denote by $x^\downarrow := (x_1^\downarrow, \ldots, x_d^\downarrow)$ the *decreasing arrangement* of x, i.e. $x_1^\downarrow \geq \cdots \geq x_d^\downarrow$. We say that x is *weak majorized* by y, denoted by $x \prec_{\text{w}} y$, if

$$\sum_{j=1}^k x_j^\downarrow \leq \sum_{j=1}^k y_j^\downarrow,\ 1 \leq k \leq d. \tag{11}$$

The *weak log-majorization* $x \prec_{\text{wlog}} y$ is defined when $\log x \prec_{\text{w}} \log y$, where we denote by $\log x$ the vector whose components equal to the logarithm of the components of x. It is well-known that $x \prec_{\text{wlog}} y$ implies $x \prec_{\text{w}} y$ [15, Example II.3.5]. Let $\lambda(X)$ denote the vector of eigenvalues of the matrix X. For two positive semi-definite matrices A and B, the weak majorization $\lambda(A) \prec_{\text{w}} \lambda(B)$ is equivalent to $|||A||| \leq |||B|||$ for all unitarily-invariant norm $|||\cdot|||$ [11, Theorem 6.23].

In the following, we collect a few lemmas that will be used in the main proof.

Lemma 2 **([16, Theorem 2.10]).** *For any* $A, B \in \mathbb{M}_d^{++}$*, and* $0 \leq \tau \leq 1$*. Then*

$$\lambda(A\#_\tau B) \prec_{\text{wlog}} \lambda\left(A^{1-\tau}B^\tau\right). \tag{12}$$

Lemma 3 **(Araki-Lieb-Thirring Inequality [17]; see also [15, Theorem IX.2.10]).** *Let* $A, B \in \mathbb{M}_d^+$*. Then, we have*

$$\lambda\left(B^tA^tB^t\right) \prec_{\text{w}} \lambda\left((BAB)^t\right),\ \textit{for}\ 0 \leq t \leq 1, \tag{13}$$

$$\lambda\left(B^tA^tB^t\right) \succ_{\text{w}} \lambda\left((BAB)^t\right),\ \textit{for}\ t \geq 1. \tag{14}$$

Lemma 4 **([15, Example II.3.5]).** *Let* $x, y \in \mathbb{R}_+^d$ *(the set of* d*-dimensional vectors of non-negative real numbers). Then*

$$x \prec_{\text{w}} y \quad \textit{implies} \quad x^t \prec_{\text{w}} y^t \tag{15}$$

for all $t \geq 1$.

Lemma 5 **(See, e.g. [18, Section 2.2]).** *Let* f *be a monotonically increasing function on the real line. Then* $A \preceq B$ *implies*

$$\operatorname{Tr}[f(A)] \leq \operatorname{Tr}[f(B)]. \tag{16}$$

Lemma 6 **(Matrix Hölder's Inequality [15, Corollary IV.2.6]).** *Let* $A, B \in \mathbb{M}_d^+$*. Then*

$$\operatorname{Tr}[AB] \leq \left(\operatorname{Tr}\left[A^{\frac{1}{\theta}}\right]\right)^\theta \left(\operatorname{Tr}\left[B^{\frac{1}{1-\theta}}\right]\right)^{1-\theta} \tag{17}$$

for all $0 \leq \theta \leq 1$.

III. Main Result

We first denote some notation. Let $\mathcal{X} = \{1, 2, \ldots, |\mathcal{X}|\}$ be a finite alphabet, and $\mathcal{H}$ be a Hilbert space of finite dimension. Denote by $\mathcal{P}(\mathcal{X})$ the set of probability distributions on $\mathcal{X}$. The set of density operators (i.e. positive semi-definite operators with unit trace) on $\mathcal{H}$ is defined as $\mathcal{S}(\mathcal{H})$. Denote the set of all (discrete memoryless) classical-quantum (c-q) channels $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ by $\mathcal{W}(\mathcal{X})$.

Theorem 7. *Given a classical-quantum channel* $W \in \mathcal{W}(\mathcal{X})$ *and a distribution* $P \in \mathcal{P}(\mathcal{X})$*, the auxiliary function* $E_0(s, P)$ *is concave in* $s \geq 0$.

Proof. Since the geometric means, Eq. (9), are defined for positive definite matrices, we first present the proof that only works when all $\{W_x\}_{x\in\mathcal{X}}$ are full rank. The proof can then be extended to include the non-invertible case.

Let X be a random variable with the distribution P, and denote by $\mathbb{E}_X$ the expectation with respect to P. Then it suffices to prove the convexity of the map:

$$f : t \mapsto \log \operatorname{Tr}\left[\left(\mathbb{E}_X W_X^{\frac{1}{t}}\right)^t\right] \tag{18}$$

for all $t \geq 1$.

Before starting the proof, we first give the following lemma that is crucial in our derivations.

Lemma 8. *Let $A, B \in \mathbb{M}_d^{++}$. Then, for every $t \geq 1$ and $0 \leq \tau \leq 1$, we have*

$$\operatorname{Tr}\left[(A \#_\tau B)^t\right] \leq \operatorname{Tr}\left[A^{t(1-\tau)} B^{t\tau}\right]. \tag{19}$$

Proof. From Lemma 2, we have

$$\begin{aligned}
\lambda\left(A \#_\tau B\right) &\prec_{\mathrm{w}} \lambda\left(A^{1-\tau} B^{\tau}\right) & (20)\\
&= \lambda\left(A^{\frac{1-\tau}{2}} B^{\tau} A^{\frac{1-\tau}{2}}\right) & (21)\\
&\prec_{\mathrm{w}} \lambda\left(\left(A^{\frac{t(1-\tau)}{2}} B^{t\tau} A^{\frac{t(1-\tau)}{2}}\right)^{\frac{1}{t}}\right), & (22)
\end{aligned}$$

where we employ the fact that $\lambda(XY) = \lambda(YX)$ for any two square matrices X, Y in Eq. (21) (see e.g. [11, Example 1.19]). The last inequality (22) follows from Eq. (13) in Lemma 3. Next, applying Lemma 4 on the above inequality yields

$$\lambda\left((A \#_\tau B)^t\right) \prec_{\mathrm{w}} \lambda\left(A^{\frac{t(1-\tau)}{2}} B^{t\tau} A^{\frac{t(1-\tau)}{2}}\right). \tag{23}$$

Finally, since the trace function is the summation of eigenvalues, the weak majorization in Eq. (23) implies the trace norm inequality in Eq. (19). □

We now begin the proof of Theorem 7. These steps follow closely with those in Ref. [3, Theorem 5.6.3]. Let l, r, and θ be arbitrary numbers $1 \leq l \leq r$, $0 \leq \theta \leq 1$, and define

$$t = \theta l + (1-\theta) r. \tag{24}$$

Let $t \equiv 1 + s \geq 1$. Then we prove the convexity of the map f from Eq. (18), i.e.

$$f(t) \leq \theta f(l) + (1-\theta) f(r). \tag{25}$$

Define the number $\tau \in [0,1]$ by

$$\tau = \frac{l\theta}{t}; \quad 1-\tau = \frac{r(1-\theta)}{t}. \tag{26}$$

Then it follows that

$$\frac{1}{t} = \frac{\theta}{t} + \frac{1-\theta}{t} = \frac{\tau}{l} + \frac{1-\tau}{r}. \tag{27}$$

The concavity of the geometric means (see item (f) in Proposition 1) implies that

$$\begin{aligned}
\mathbb{E}_X\left[W_X^{1/t}\right] &= \mathbb{E}_X\left[W_X^{\tau/l} W_X^{(1-\tau)/r}\right] & (28)\\
&= \mathbb{E}_X\left[W_X^{1/l} \#_{1-\tau} W_X^{1/r}\right] & (29)\\
&\preceq \mathbb{E}_X\left[W_X^{1/l}\right] \#_{1-\tau} \mathbb{E}_X\left[W_X^{1/r}\right]. & (30)
\end{aligned}$$

Now let $A \equiv \mathbb{E}_X\left[W_X^{1/l}\right]$ and $B \equiv \mathbb{E}_X\left[W_X^{1/r}\right]$. Since $x \mapsto x^t$ for $t \geq 1$ is a monotone function, Lemma 5 leads to

$$\begin{aligned}
\operatorname{Tr}\left[\left(\mathbb{E}_X\left[W_X^{1/t}\right]\right)^t\right] &\leq \operatorname{Tr}\left[(A \#_{1-\tau} B)^t\right] & (31)\\
&\leq \operatorname{Tr}\left[A^{t\tau} B^{t(1-\tau)}\right] & (32)\\
&= \operatorname{Tr}\left[A^{l\theta} B^{r(1-\theta)}\right], & (33)
\end{aligned}$$

where Eq. (32) follows from Lemma 8. Finally, applying the matrix Hölder's inequality, Lemma 6, on the right-hand side of Eq. (33), we have

$$\begin{aligned}
&\operatorname{Tr}\left[\left(\mathbb{E}_X\left[W_X^{1/t}\right]\right)^t\right] \leq \left(\operatorname{Tr}\left[A^l\right]\right)^\theta \left(\operatorname{Tr}\left[B^r\right]\right)^{1-\theta}\\
&= \left(\operatorname{Tr}\left(\mathbb{E}_X\left[W_X^{1/l}\right]\right)^l\right)^\theta \left(\operatorname{Tr}\left(\mathbb{E}_X\left[W_X^{1/r}\right]\right)^r\right)^{1-\theta}.
\end{aligned}$$

Taking the logarithm of the above inequality leads to $f(t) \leq \theta f(l) + (1-\theta) f(r)$. This completes the proof for the special case of invertible channel outputs.

The above proof assumes that every realization of the density operator W_x, $x \in \mathcal{X}$, is positive definite. Hence, each density operator $W_x^{\tau/l} W_x^{(1-\tau)/r}$ can be expressed as a geometric mean $W_x^{1/l} \#_{1-\tau} W_x^{1/r}$. However, if W_x is not invertible for some $x \in \mathcal{X}$, then consider a sequence of positive definite operators $W_{x,\epsilon} := W_x + \epsilon I$ that approximate W_x, i.e., $\lim_{\epsilon \searrow 0} W_{x,\epsilon} = W_x$. The geometric mean of $W_x^{1/l}$ and $W_x^{1/r}$ is defined as

$$\left(W_x^{1/l}\right) \#_s \left(W_x^{1/r}\right) := \lim_{\epsilon \searrow 0} \left(W_{x,\epsilon}^{1/l}\right) \#_s \left(W_{x,\epsilon}^{1/r}\right), \tag{34}$$

by the continuity of the geometric means (see item (h) in Proposition 1). Note that the concavity of the geometric means, and Lemmas 2 and 8 still hold if we use the definition in Eq. (34). We can thus obtain a complete proof.

□

IV. Properties of the Auxiliary Function

This section presents important properties of the auxiliary function. Most properties are obtained through the observation that the auxiliary function directly relates to the *quasi-arithmetic mean* [19, Section 4]. For a sequence of matrices $\boldsymbol{A} = (A_1, \ldots, A_M)$ and a probability vector $w = (w_1, \ldots, w_M)$, the quasi-arithmetic mean with parameter $t > 0$ is defined by

$$\mathfrak{m}^{\mathsf{QA}}(t, w, \boldsymbol{A}) := \left(\sum_{i=1}^{M} w_i A_i^t \right)^{1/t}. \tag{35}$$

Thus the auxiliary function can be expressed as

$$E_0(s, P, W) \equiv -\log \operatorname{Tr}\left[\mathfrak{m}^{\mathsf{QA}}\left(\frac{1}{1+s}, P, W \right) \right]. \tag{36}$$

Note that throughout the section we will explicitly include the classical-quantum channel W in the expression of the auxiliary function.

Proposition 9. *The auxiliary function $E_0(s, P, W)$ has the following properties.*

(a) Monotonicity: $E_0(s, P, W) \leq E_0(t, P, W)$ *for all* $0 \leq s \leq t$.

(b) Non-negativity: $E_0(s, P, W) \geq 0$ *for all* $s \geq 0$ *with* $E_0(0, P, W) = 0$.

(c) Relation with mutual information: $\partial E_0(s, P, W)/\partial s|_{s=0} = I(P, W)$*, where*

$$I(P, W) := \sum_{x \in \mathcal{X}} P(x) \operatorname{Tr}\left[W_x \log W_x \right] - \operatorname{Tr}\left[\left(\sum_{x \in \mathcal{X}} P(x) W_x \right) \log \left(\sum_{x \in \mathcal{X}} P(x) W_x \right) \right]$$

denotes the mutual information of the c-q channel W with the input distribution P [20].

(d) Concavity in s: $\partial^2 E_0(s, P, W)/\partial s^2 \leq 0$ *for all* $s \geq 0$.

(e) Convexity in W: *The map $W \mapsto E_0(s, P, W)$ is convex for all $W \in \mathcal{W}(\mathcal{X})$.*

(f) $\exp(-E_0(s, P, W))$ *is convex in* P.

(g) Tensor invariance: *For any quantum state ϱ on a Hilbert space $\mathcal{H}'$, we have*

$$E_0(s, P, W \otimes \varrho) = E_0(s, P, W), \tag{37}$$

where $W \otimes \varrho$ denotes the c-q channel that maps every x to $W_x \otimes \varrho$.

(h) Unitary invariance: *Let $W' : x \mapsto U W_x U^\dagger$, $x \in \mathcal{X}$ be the composition of the channel W with the unitary U. Then $E_0(s, P, W') = E_0(s, P, W)$.*

(i) Data-processing inequality: $E_0(s, P, \Phi \circ W) \leq E_0(s, P, W)$ *for any completely-positive and trace-preserving map Φ. Here, we denote the composite channel by $\Phi \circ W : x \mapsto \Phi(W_x)$.*

(j) Conditions for maximization over P: *The input distribution P attains $E_0(s, P, W)$ if and only if*

$$\operatorname{Tr}\left[W_x^{1/(1+s)} \left(\sum_{x \in \mathcal{X}} P(x) W_x^{1/(1+s)} \right)^s \right] \geq \operatorname{Tr}\left[\left(\sum_{x \in \mathcal{X}} P(x) W_x^{1/(1+s)} \right)^{1+s} \right], \quad \forall x \in \mathcal{X}. \tag{38}$$

Proof. (a) The monotonicity in s was first proved by Holevo [6, Appendix].

Recently, Bhatia and Grover established a stronger result: each eigenvalue $\lambda\left(\mathfrak{m}^{\mathsf{QA}}(t, P, W)\right)$ is an increasing function of $t \geq 0$ (see [21, Theorem 1]). By the relation (36), Bhatia and Grover's result directly implies the monotonicity of $E_0(s, P, W)$.

(b) It is clear that

$$E_0(0, P, W) = -\log \operatorname{Tr}\left(\sum_x P(x) W_x \right) = 0$$

when $s = 0$. The monotonicity in item (a) coupled with the identity $E_0(0, P) = 0$ thus yields the non-negative of $E_0(s, P, W)$.

(c) The relation to the mutual information was first discovered by Ogawa and Nagaoka [22, Eq. (12)].

(d) The concavity of the map $s \mapsto E_0(s, P)$ is our main result, Theorem 7.

(e) Hiai [23, Lemma 3.3] showed that the map $(A, B) \mapsto \operatorname{Tr}\left[(\theta A^t + (1-\theta) B^t)^{1/t} \right]$ is jointly concave for any $0 \leq \theta \leq 1$. The result can be easily extended to any set of finite matrices, leading to the concavity of the map $W \mapsto \operatorname{Tr}\left[\mathfrak{m}^{\mathsf{QA}}(1/(1+s), P, W) \right]$. Since the logarithm function preserves concavity, the convexity of $E_0(s, P, W)$ in W follows.

(f) Define

$$\begin{aligned} F(s,P,W) &:= \exp(-E_0(s,P,W)) \\ &= \operatorname{Tr}\left[\mathfrak{m}^{\mathsf{QA}}(1/(1+s),P,W)\right]. \end{aligned}$$

Let $P, Q \in \mathcal{P}(\mathcal{X})$ and let $0 \le t \le 1$. Simple algebra gives

$$F(s,tP+(1-t)Q,W) \tag{39}$$

$$= \operatorname{Tr}\left[\left(\sum_x (tP(x)+(1-t)Q(x)) \cdot W_x^{1/(1+s)}\right)^{1+s}\right] \tag{40}$$

$$= \operatorname{Tr}\left[\left(t\sum_x P(x)W_x^{1/(1+s)} + (1-t)\sum_x Q(x)W_x^{1/(1+s)}\right)^{1+s}\right] \tag{41}$$

Since the trace function preserves the convexity of the power function $u \mapsto u^{1+s}$ for $s \ge 0$ (see e.g. [18, Section 2.2]),

$$F(s,tP+(1-t)Q,W) \tag{42}$$

$$\le t\operatorname{Tr}\left[\left(\sum_x P(x)W_x^{1/(1+s)}\right)^{1+s}\right] + (1-t)\operatorname{Tr}\left[\left(\sum_x Q(x)W_x^{1/(1+s)}\right)^{1+s}\right] \tag{43}$$

$$= tF(s,P,W)+(1-t)F(s,Q,W). \tag{44}$$

Note that even in the classical case, the auxiliary function alone is not convex in the distribution P. For example, consider a binary symmetric channel: $\mathcal{X} = \{1,2\}$,

$$W_1 = \begin{pmatrix} 0.8 & 0 \\ 0 & 0.2 \end{pmatrix}, \quad W_2 = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.8 \end{pmatrix}$$

Let the two distributions be $P(1) = 1-P(2) = 3/4$, $Q(1) = 1-Q(2) = 1/4$, and let $s = 1$. Then, we find

$$\begin{aligned} &\frac{1}{2}E_0(s,P,W)+\frac{1}{2}E_0(s,Q,W) = 0.078 \\ &\not\ge E_0\left(s,\frac{P+Q}{2},W\right) = 0.1054. \end{aligned}$$

(g) The tensor invariance of $E_0(s,P,W)$ directly follows from the following property of the quasi-arithmetic mean, i.e.

$$\mathfrak{m}^{\mathsf{QA}}(1/(1+s),P,W\otimes\varrho) \tag{45}$$

$$= \left(\left[\sum_x P(x)W_x^{1/(1+s)}\right] \otimes \varrho^{1/(1+s)}\right)^{1+s} \tag{46}$$

$$= \mathfrak{m}^{\mathsf{QA}}(1/(1+s),P,W)\otimes\varrho. \tag{47}$$

(h) From the definition of the quasi-arithmetic mean, it is not hard to observe that

$$\begin{aligned} &\mathfrak{m}^{\mathsf{QA}}(1/(1+s),P,W') \\ &= U\mathfrak{m}^{\mathsf{QA}}(1/(1+s),P,W)U^\dagger \end{aligned}$$

for any unitary U. Then the unitary invariance of E_0 clearly follows, i.e.

$$E_0(s,P,W') = \ -\log\operatorname{Tr}\left[U\mathfrak{m}^{\mathsf{QA}}(1/(1+s),P,W)U^\dagger\right] =$$

(i) The data-processing inequality of the auxiliary function results from the argument proved by Frank and Lieb [24, Theorem 1]. That is, the map $W \mapsto E_0(s,P,W)$ satisfies the data-processing inequality if it is convex, tensor invariant, and invariant under unitary conjugation. Hence item (i) simply follows from item (e), item (g) and item (h).

(j) The sufficient and necessary condition (38) for the optimum distribution was proved by Holevo [6, Appendix] using the fact that the maximization of $E_0(s,P,W)$ over P is equivalent to the minimization of the function $\exp(-E_0(s,P,W))$ over P since the exponential function is monotonically increasing.

□

A. Relations to Random Coding Exponent and Sphere-Packing Exponent

The concavity property of the auxiliary function allows us to better characterize the random coding exponent and the sphere packing exponent. In the following, it is convenient to introduce the quantity[1]: $E_{\text{sp}}(R,P) := \sup_{s\ge0}\{E_0(s,P)-sR\}$ for any distribution $P \in \mathcal{P}(\mathcal{X})$.

Since $E_0(s,P)-sR$ is concave in s for all $s \ge 0$, the maximizer s to $E_0(s,P)-sR$ is hence the

[1] Since the classical-quantum channel W is fixed, it is omitted in the expression to improve the readability.

solution to the following equation:

$$\frac{\partial E_0(s,P)}{\partial s} = R. \tag{48}$$

We remark that the power function on matrices is continuously differentiable (see e.g. [25, Theorem 1.19]), the derivative of $E_0(s,P)$ in Eq. (48) is well-defined.

Moreover, due to the concavity of the auxiliary function, $\partial E_0(s,P)/\partial s$ is decreasing in s. The solution Eq. (48) exists if R is in the range:

$$\lim_{s\to\infty} \frac{\partial E_0(s,P)}{\partial s} \leq R \leq \left.\frac{\partial E_0(s,P)}{\partial s}\right|_{s=0} = I(P,W). \tag{49}$$

From this relation, it is not hard to observe that $E_0(s,P) - sR$ is maximized by $s=0$ for the region of $R > I(P,W)$. Hence, the quantity $E_{\text{sp}}(R,P)$ vanishes for all $P \in \mathcal{P}(\mathcal{X})$. On the other hand, we have $E_{\text{sp}}(R,P) = \infty$ if $R < \lim_{s\to\infty} \partial E_0(s,P)/\partial s$. This indicates that the sphere-packing exponent yields a very loose bound on the error probability in this range. Note that the above analysis also applies to counterpart of the quantity $E_{\text{r}}(R,P) := \max_{0\leq s\leq 1}\{E_0(s,P) - sR\}$ with the range of the parameter s being restricted to $[0,1]$.

In the following, we show that the sphere-packing exponent and the random coding exponent are convex in R, a result that follows from the concavity of the auxiliary function.

Proposition 10. *For any classical-quantum channel $W \in \mathcal{W}(\mathcal{X})$ with $\partial^2 E_0(s,P)/\partial s^2 < 0$, the sphere-packing exponent $E_{\text{sp}}(R)$ is decreasing and strictly convex in R within the range given in Eq. (49).*

The same holds for the random coding exponent $E_{\text{r}}(R)$ with $0 \leq s \leq 1$.

Proof. Fix a distribution $P \in \mathcal{P}(\mathcal{X})$. Given the rate R in the appropriate range, i.e.

$$\lim_{s\to\infty} \frac{\partial E_0(s,P)}{\partial s} \leq R \leq I(P,W),$$

Eq. (48) gives

$$E_{\text{sp}}(R,P) = E_0(s,P) - s\frac{\partial E_0(s,P)}{\partial s},$$

which means that given each R in Eq. (49) the quantity $E_{\text{sp}}(R,P)$ can be parameterized by some $s \geq 0$. Differentiating both sides with respect to s, we obtain

$$\frac{\partial E_{\text{sp}}(R,P)}{\partial s} = -s\frac{\partial^2 E_0(s,P)}{\partial s^2}. \tag{50}$$

Then,

$$\frac{\partial E_{\text{sp}}(R,P)}{\partial R} = \frac{\partial E_{\text{sp}}(R,P)}{\partial s}\frac{\partial s}{\partial R} = -s \leq 0 \tag{51}$$

follows from the chain rule, Eq. (50), and the fact that

$$\frac{\partial R}{\partial s} = \frac{\partial^2 E_0(s,P)}{\partial s^2}. \tag{52}$$

We have thus established that $E_{\text{sp}}(R,P)$ is decreasing in R.

Note that the right-hand side of Eq. (51) depends on the choice of R. Differentiating each side in Eq. (51) with respect to R, we obtain the strict convexity of $E_{\text{sp}}(R,P)$ in R since

$$\frac{\partial^2 E_{\text{sp}}(R,P)}{\partial R^2} = -\frac{\partial s}{\partial R} = -\left(\frac{\partial^2 E_0(s,P)}{\partial s^2}\right)^{-1} > 0. \tag{53}$$

Note that the results of $E_{\text{sp}}(R,P)$ hold for any $P \in \mathcal{P}(\mathcal{X})$. We conclude our claims for the sphere-packing exponent $E_{\text{sp}}(R) = \max_{P\in\mathcal{P}(\mathcal{X})} E_{\text{sp}}(R,P)$. □

V. Conclusion

In this paper, we settled an open question that was originally raised in Ref. [6]. A partial result to this question was obtained in Ref. [8]; however, we can extend the concavity of the auxiliary function $E_0(s,P)$ for all $s \geq 0$. Consequently, the definition of the auxiliary function Eq. (8) of a classical-quantum channel exactly recovers its classical counterpart [3], a quantity that plays a crucial role in classical information theory. We hope that this concave property will also allow us to better characterize the error probability of a classical-quantum channel in the finite blocklength regime.

Acknowledgements

We sincerely thank the three referees for their useful comments. MH is supported by an ARC Future Fellowship under Grant FT140100574.

References

[1] R. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Transaction on Information Theory*, vol. 11, no. 1, pp. 3–18, 1965. DOI: 10.1109/tit.1965.1053730.

[2] C. Shannon, R. Gallager, and E. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Information and Control*, vol. 10, no. 1, pp. 65–103, 1967. DOI: 10.1016/s0019-9958(67)90052-6.

[3] R. Gallager, *Information Theory and Reliable Communication*. Wiley, 1968, ISBN: 978-0-471-29048-3.

[4] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.

[5] M. V. Burnashev and A. S. Holevo, "On the reliability function for a quantum communication channel," *Problems of information transmission*, vol. 34, no. 2, pp. 97–107, 1998. [Online]. Available: http://mi.mathnet.ru/eng/ppi399.

[6] A. Holevo, "Reliability function of general classical-quantum channel," *IEEE Transaction on Information Theory*, vol. 46, no. 6, pp. 2256–2261, 2000. DOI: 10.1109/18.868501.

[7] M. Dalai, "Lower bounds on the probability of error for classical and classical-quantum channels," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 8027–8056, 2013. DOI: 10.1109/tit.2013.2283794.

[8] J. I. Fujii, R. Nakamoto, and K. Yanagi, "Concavity of the auxiliary function appearing in quantum reliability function," *IEEE Transaction on Information Theory*, vol. 52, no. 7, pp. 3310–3313, 2006. DOI: 10.1109/tit.2006.876248.

[9] F. Kubo and T. Ando, "Means of positive linear operators," *Mathematische Annalen*, vol. 246, no. 3, pp. 205–224, 1980. DOI: 10.1007/bf01371042.

[10] R. Bhatia, *Positive Definite Matrices*. Princeton University Press, 2009, ISBN: 978-1-4008-2778-7. DOI: 10.1515/9781400827787.

[11] F. Hiai and D. Petz, *Introduction to Matrix Analysis and Applications*. Springer International Publishing, 2014, ISBN: 978-3319041490. DOI: 10.1007/978-3-319-04150-6.

[12] G. Corach, H.Porta, and L. Recht, "Convexity of the geodesic distance on spaces of positive operators," *Illinois Journal of Mathematics*, vol. 38, pp. 87–94, 1994.

[13] J. Lawson and Y. Lim, "Metric convexity of symmetric cones," *Osaka Journal of Mathematics*, vol. 4, no. 4, pp. 795–816, 2007. [Online]. Available: http://projecteuclid.org/euclid.ojm/1199719405.

[14] F. Hiai, "Log-majorizations and norm inequalities for exponential operators," *Banach Center Publications*, vol. 38, no. 1, pp. 119–181, 1997.

[15] R. Bhatia, *Matrix Analysis*. Springer New York, 1997, ISBN: 978-1-4612-6857-4. DOI: 10.1007/978-1-4612-0653-8.

[16] J. S. Matharu and J. S. Aujla, "Some inequalities for unitarily invariant norm," *Linear Algebra and its Applications*, vol. 436, no. 6, pp. 1623–1631, 2012. DOI: 10.1016/j.laa.2010.08.013.

[17] H. Araki, "On an inequality of Lieb and Thirring," *Letters in Mathematical Physics*, vol. 19, no. 2, pp. 167–170, 1990. DOI: 10.1007/bf01045887.

[18] E. Carlen, "Trace inequalities and quantum entropy: An introductory course," in *Contemporary Mathematics*, vol. 529, American Mathematical Society, 2010, pp. 73–140. DOI: 10.1090/conm/529/10428.

[19] K. V. Bhagwat and R. Subramanian, "Inequalities between means of positive operators," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 83, no. 03, pp. 393–401, 1978. DOI: 10.1017/s0305004100054670.

[20] A. Holevo, "Bounds for the quantity of information transmitted by a quantum communication channel," *Problems of Information Transmission*, vol. 9, no. 3, 177–183, 1973.

[21] R. Bhatia and P. Grover, "Norm inequalities related to the matrix geometric mean," *Linear Algebra and its Applications*, vol. 437, no. 2, pp. 726–733, 2012. DOI: 10.1016/j.laa.2012.03.001.

[22] T. Ogawa and H. Nagaoka, “Strong converse to the quantum channel coding theorem,” *IEEE Transaction on Information Theory*, vol. 45, no. 7, pp. 2486–2489, 1999. DOI: 10.1109/18.796386.

[23] F. Hiai, “Concavity of certain matrix trace and norm functions,” *Linear Algebra and its Applications*, vol. 439, no. 5, pp. 1568–1589, 2013. DOI: 10.1016/j.laa.2013.04.020.

[24] R. L. Frank and E. H. Lieb, “Monotonicity of a relative Rényi entropy,” *Journal of Mathematical Physics*, vol. 54, no. 12, p. 122 201, 2013. DOI: 10.1063/1.4838835.

[25] N. J. Higham, *Functions of Matrices: Theory and Computation*. Society for Industrial & Applied Mathematics (SIAM), 2008, ISBN: 978-0898716467. DOI: 10.1137/1.9780898717778.