A New Characterization Of Augustin Information And Mean

Hao-Chung Cheng
Department of Electrical Engineering
National Taiwan University
Taipei 10617, Taiwan (R.O.C.)
Hon Hai (Foxconn) Quantum Computing Centre
0000-0003-4499-4679

Barış Nakiboğlu
Department of Electrical and Electronics Engineering
Middle East Technical University
06800 Ankara, Turkey
0000-0001-7737-5423

***Abstract*—A new method to characterize Augustin information and mean is proposed. The proposed method allows for briefer and more direct proofs for the previously known results and leads to new observations related to this new characterization for classical channels. For classical-quantum channels, the proposed method extends the results, such as the existence of Augustin mean and Augustin fixed point property, to models with separable Hilbert spaces at the output.**

I. Introduction

The sphere packing, random coding, and strong converse exponents for channel coding problem have been characterized in terms of the Augustin information and capacity [1]–[14]. Similar to the characterizations in Haroutunian form [15]–[17], these characterizations work for the cost-constrained and the composition-constrained cases without the help of Lagrange multipliers techniques, see [1]–[10]. Furthermore, with the help of Augustin mean, they have been used to establish refined sphere packing bounds [5], [6] and refined strong converses [7], both for constant composition codes on general channels and for general channel codes on channels with certain symmetries.

For a channel W and input distribution p, the order α Augustin information is defined in terms of the order α conditional Rényi divergence as

$$I_{\alpha}(p;W) := \inf\nolimits_{q\in\mathcal{P}(\mathcal{Y})} D_{\alpha}(W\|q|p), \tag{1}$$

where $\mathcal{P}(\mathcal{Y})$ is the set of all probability measures on the output space $(\mathcal{Y},\mathcal{Y})$, which is a measurable space, and $D_{\alpha}(W\|q|p)$ is the expected value of $D_{\alpha}(W(\mathsf{X})\|q)$ for the case when X has the distribution p. The primary challenge in working with Augustin information is that neither the minimizer nor the infimum value in (1), has general closed form expressions that are valid for all channels and input distributions, except for the $\alpha=1$ case, in which case the infimum is the mutual information and the unique minimizer is the output distribution of the channel W for the input distribution p.

The minimizer in (1) is called an Augustin mean; its existence and uniqueness is established for input distributions with a finite support set on arbitrary classical channels in [1], [13], and on classical-quantum channels with finite-dimensional Hilbert spaces at the output in [3], [9]. For arbitrary input distributions on classical channels with countably-generated output σ-algebras, the existence of a unique Augustin mean is established in [18] under the finite Augustin information hypothesis. In all of the cases mentioned above, the Augustin mean satisfies a fixed point property, that not only provides an alternative characterization that does not refer to (1) explicitly, but also plays a critical role in establishing refined converses, [5]–[7].

If the output set $\mathcal{Y}$ is finite, then $\mathcal{P}(\mathcal{Y})$ is a compact set and the existence of the order α Augustin mean follows from the extreme value theorem and the lower semicontinuity of the Rényi divergence. In [1], [13], to establish the existence of the Augustin mean for arbitrary output spaces, first, a compact subset of $\mathcal{P}(\mathcal{Y})$ on which the infimum in (1) is achieved is found. To address the same technical challenge, instead of finding the minimizer in (1), we characterize the maximizer of another optimization problem through a fixed point property and show that the maximizer of the new problem determines the minimizer in (1). Furthermore, the new problem is a finite-dimensional optimization problem as long as the support of the input distribution p is a finite set, no matter what the output space of the channel is. This feature of the new characterization allows us to apply it, not only to the classical channels with arbitrary output spaces, but also to the classical-quantum channels with separable Hilbert spaces at the output.

We will present the alternative characterization of the Augustin information and mean and associated new results first for classical channels in §II, then for classical-quantum channels in §III. Before we start our discussion in earnest, let us point out a few notational conventions we will adhere to throughout the manuscript.

For any set $\mathcal{X}$, we denote the set of all mass (i.e., non-negative) functions on $\mathcal{X}$ with a countable support set, and the set of all probability mass function on $\mathcal{X}$ by $\mathcal{M}^{+}(\mathcal{X})$, and $\mathcal{P}(\mathcal{X})$. For any $p\in\mathcal{M}^{+}(\mathcal{X})$, we denote the set of all extended real valued functions that are finite on the support of p by $\mathcal{F}_{p}$. The constant function with the value one is denoted by $\mathbb{1}$. For any $|\mathcal{X}|$-by-$|\mathcal{X}|$ positive semidefinite matrix Λ, we define the inner product $\langle\cdot,\cdot\rangle_{\Lambda}$ and the norm $\|\cdot\|_{\Lambda}$ as

$$\langle f,g\rangle_{\Lambda} := f^{T}\Lambda g \qquad \forall f,g\in\mathbb{R}^{\mathcal{X}},$$
$$\|f\|_{\Lambda} := (\langle f,f\rangle_{\Lambda})^{1/2} \qquad \forall f\in\mathbb{R}^{\mathcal{X}}.$$

We employ the same notation when either f, or g, or both are parametrized by another variable, say α or y. When Λ is the identity matrix, we denote the inner product by $\langle\cdot,\cdot\rangle$ and the norm by $\|\cdot\|$. We denote ℓ^{1} and L^{1} norms by $\|\cdot\|_{1}$. We use $a\wedge b$ to denote the minimum between scalars a and b.

II. Augustin Dual Stationarity: A New Characterization Of Augustin Information and Mean

Definition 1. For any order $\alpha\in\mathbb{R}_{+}$, channel $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, mass function $p\in\mathcal{M}^{+}(\mathcal{X})$ with a finite support set, and function $f\in\mathcal{F}_{p}$, the (adjusted) power mean $\mu_{\alpha,p}^{f}$ is defined as

$$\tfrac{\mathrm{d}\mu_{\alpha,p}^{f}}{\mathrm{d}\nu}:=\left(\left\langle p, e^{(1-\alpha)f}\left(\tfrac{\mathrm{d}W}{\mathrm{d}\nu}\right)^{\alpha}\right\rangle\right)^{1/\alpha}, \tag{2}$$

where ν is any σ-finite reference measure such that $W(x)$ is absolutely continuous with respect to ν, p-almost everywhere.

In [19], the power mean is defined for the case when $f=0$ and denoted by $\mu_{\alpha,p}$. We will adhere to the same notational convention and use $\mu_{\alpha,p}$ for the case when $f=0$ holds p-.a.e..

One can confirm by substitution that

$$\mu_{\alpha,\gamma p}^{f+\tau}=\gamma^{\frac{1}{\alpha}}e^{\frac{1-\alpha}{\alpha}\tau}\mu_{\alpha,p}^{f}, \qquad \forall\gamma\in\mathbb{R}_{+},\ \tau\in\mathbb{R}. \tag{3}$$

Furthermore, the log of the Radon-Nikodym derivative of the power mean with respect to any σ-finite reference measure ν satisfies the following pointwise convexity by the Hölder's inequality:

$$\ln\tfrac{\mathrm{d}\mu_{\alpha,p}^{f_{\beta}}}{\mathrm{d}\nu}\leq\beta\ln\tfrac{\mathrm{d}\mu_{\alpha,p}^{f_{1}}}{\mathrm{d}\nu}+(1-\beta)\ln\tfrac{\mathrm{d}\mu_{\alpha,p}^{f_{0}}}{\mathrm{d}\nu} \qquad \nu\text{-a.e.}, \tag{4}$$

for all $\beta\in(0,1)$ where $f_{\beta}=\beta f_{1}+(1-\beta)f_{0}$. Furthermore, (4) holds as an equality for a $y\in\mathcal{Y}$ iff there exists $\tau\in\mathbb{R}$ for which $f_{1}(x)=f_{0}(x)+\tau$ holds for all x satisfying $\left(\tfrac{\mathrm{d}W(x)}{\mathrm{d}\nu}(y)\right)p(x)>0$.

Using Hölder's inequality once more one can extend the log-convexity to the L^{1} norm of the power mean:

$$\ln\left\|\mu_{\alpha,p}^{f_{\beta}}\right\|_{1}\leq\beta\ln\left\|\mu_{\alpha,p}^{f_{1}}\right\|_{1}+(1-\beta)\ln\left\|\mu_{\alpha,p}^{f_{0}}\right\|_{1}, \tag{5}$$

for all $\beta\in(0,1)$ and the inequality is strict unless $f_{1}=f_{0}+\tau$ holds p-a.e. for some $\tau\in\mathbb{R}$.

Definition 2. For any $\alpha\in\mathbb{R}_{+}$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, $p\in\mathcal{M}^{+}(\mathcal{X})$ with a finite support set, and $f\in\mathcal{F}_{p}$, the (adjusted) Gallager's function $A_{\alpha}(p,f)$ and the maximal (adjusted) Gallager's function $A_{\alpha}(p)$ are defined as

$$A_{\alpha}(p,f):=\tfrac{1-\alpha}{\alpha}\langle p,f\rangle-\langle p,\mathbb{1}\rangle\ln\left\|\mu_{\alpha,p}^{f}\right\|_{1}, \tag{6}$$

$$A_{\alpha}(p):=\sup\nolimits_{f\in\mathcal{F}_{p}}A_{\alpha}(p,f). \tag{7}$$

Then $A_{\alpha}(p,f)$ satisfies the following shift invariance property

$$A_{\alpha}(p,f)=A_{\alpha}(p,f+\tau) \qquad \forall\tau\in\mathbb{R}. \tag{8}$$

The norm log-convexity of the power mean $\mu_{\alpha,p}^{f}$ in f implies the concavity of the Gallager's function in f, i.e., (5) implies

$$A_{\alpha}(p,f_{\beta})\geq\beta A_{\alpha}(p,f_{1})+(1-\beta)A_{\alpha}(p,f_{0}) \tag{9}$$

for all $\beta\in(0,1)$ and the inequality is strict unless $f_{1}=f_{0}+\tau$ holds p-a.e. for some $\tau\in\mathbb{R}$, where $f_{\beta}=\beta f_{1}+(1-\beta)f_{0}$. For any $\alpha\in\mathbb{R}_{+}$, the usual definition of the order α Rényi divergence, e.g., [20], can be extended as follows for any probability measure $w\in\mathcal{P}(\mathcal{Y})$ and finite measure $q\in\mathcal{M}^{+}(\mathcal{Y})$

$$D_{\alpha}(w\|q)=\begin{cases}\frac{1}{\alpha-1}\ln\int\left(\frac{\mathrm{d}w}{\mathrm{d}\nu}\right)^{\alpha}\left(\frac{\mathrm{d}q}{\mathrm{d}\nu}\right)^{1-\alpha}\mathrm{d}\nu & \alpha\neq1\\ \int\frac{\mathrm{d}w}{\mathrm{d}\nu}\left(\ln\frac{\mathrm{d}w}{\mathrm{d}\nu}-\ln\frac{\mathrm{d}q}{\mathrm{d}\nu}\right)\mathrm{d}\nu & \alpha=1\end{cases}. \tag{10}$$

Definition 3. For any $\alpha\in\mathbb{R}_{+}$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, $p\in\mathcal{M}^{+}(\mathcal{X})$ with a finite support set and $n\in\mathbb{Z}_{+}$, the power mean operator $\mathtt{M}_{\alpha,p}(\cdot):\mathcal{F}_{p}\to\mathcal{F}_{p}$ is defined as

$$\mathtt{M}_{\alpha,p}(f):=D_{\alpha}\left(W\middle\|\mu_{\alpha,p}^{f}\right) \tag{11}$$

Note that the definition of $\mathtt{M}_{\alpha,p}(f)$, (3), and (10) imply

$$\mathtt{M}_{\alpha,\gamma p}(f+\tau)=\mathtt{M}_{\alpha,p}(f)+\tfrac{\alpha-1}{\alpha}\tau-\tfrac{1}{\alpha}\ln\gamma, \tag{12}$$

for all $\gamma\in\mathbb{R}_{+}$ and $\tau\in\mathbb{R}$.

Theorem 1. *For any $\alpha\in\mathbb{R}_{+}$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, $p\in\mathcal{M}^{+}(\mathcal{X})$ with a finite support set, the power mean operator $\mathtt{M}_{\alpha,p}(\cdot)$ has a unique fixed point $f_{\alpha,p}$, called the order α Augustin dual for the input distribution p, satisfying*

$$f_{\alpha,p}=\mathtt{M}_{\alpha,p}(f_{\alpha,p}). \tag{13}$$

The Augustin dual $f_{\alpha,p}$ satisfies

$$\left\|\mu_{\alpha,p}^{f_{\alpha,p}}\right\|_{1}=\langle p,\mathbb{1}\rangle. \tag{14}$$

$f_{\alpha,p}$ is a maximizer of $A_{\alpha}(p,\cdot)$, i.e., $A_{\alpha}(p)=A_{\alpha}(p,f_{\alpha,p})$, hence

$$A_{\alpha}(p)=\tfrac{1-\alpha}{\alpha}\langle p,f_{\alpha,p}\rangle-\langle p,\mathbb{1}\rangle\ln\langle p,\mathbb{1}\rangle. \tag{15}$$

Furthermore, maximizers of $A_{\alpha}(p,\cdot)$ are characterized by the condition $f=f_{\alpha,p}+\tau$ holds p-.a.e. for some $\tau\in\mathbb{R}$, i.e,

$$A_{\alpha}(p)=A_{\alpha}(p,f) \quad\Leftrightarrow\quad \exists\tau\in\mathbb{R}:f=f_{\alpha,p}+\tau\ p\text{-a.e.} \tag{16}$$

Proof of Theorem 1. First confirm by substitution that

$$\mu_{\alpha,p}^{f}=\mu_{\alpha,u_{\alpha,p}^{f}} \quad\text{and}\quad D_{1}\left(\tfrac{p}{\langle p,\mathbb{1}\rangle}\middle\|u_{\alpha,p}^{f}\right)=\tfrac{(\alpha-1)\langle p,f\rangle}{\langle p,\mathbb{1}\rangle}-\ln\langle p,\mathbb{1}\rangle$$

where $u_{\alpha,p}^{f}:=pe^{(1-\alpha)f}$. Thus

$$A_{\alpha}(p,f)=-\langle p,\mathbb{1}\rangle\frac{D_{1}\left(\frac{p}{\langle p,\mathbb{1}\rangle}\middle\|u_{\alpha,p}^{f}\right)+\ln\langle p,\mathbb{1}\rangle}{\alpha}-\langle p,\mathbb{1}\rangle\ln\left\|\mu_{\alpha,u_{\alpha,p}^{f}}\right\|_{1}$$

Furthermore, $f_{\alpha,p}^{u}:=\frac{1}{1-\alpha}\ln\left(\frac{u}{p}\right)$ satisfies $u_{\alpha,p}^{f}|_{f=f_{\alpha,p}^{u}}=u$. Hence there is a one-to-one correspondence between functions defined on the support of p and the mass functions that have the same support with p. Thus,

$$A_{\alpha}(p)=\tfrac{-\langle p,\mathbb{1}\rangle}{\alpha}\inf_{u\in\mathcal{M}^{+}(\mathcal{X}):u\sim p}D_{1}\left(\tfrac{p}{\langle p,\mathbb{1}\rangle}\middle\|u\right)+\ln\left(\langle p,\mathbb{1}\rangle\|\mu_{\alpha,u}\|_{1}^{\alpha}\right),$$

where $u\sim p$ stands for the equivalence of p and u, i.e., the support of u being equal to the support of p in this case. On the other hand, the function minimized on the right hand side is invariant under scaling of u by a positive constant γ by (3) and (10). Hence, for $\mathcal{U}_{p}=\{u\in\mathcal{P}(\mathcal{X}):u\sim p\}$, we have

$$A_{\alpha}(p)=\tfrac{-\langle p,\mathbb{1}\rangle}{\alpha}\inf\nolimits_{u\in\mathcal{U}_{p}}D_{1}\left(\tfrac{p}{\langle p,\mathbb{1}\rangle}\middle\|u\right)+\ln\left(\langle p,\mathbb{1}\rangle\|\mu_{\alpha,u}\|_{1}^{\alpha}\right).$$

Note that $\|\mu_{\alpha,u}\|_{1}$ is continuous in u by [19, Lemma 4-(d), Lemma 16-(c)] and the triangle inequality. On the other hand, $D_{1}\left(\frac{p}{\langle p,\mathbb{1}\rangle}\middle\|u\right)$ is continuous in u on $\mathcal{P}(\mathtt{supp}(p))$, by [20, Theorem 18]. Thus the infimum value will not change, if we replace $\mathcal{U}_{p}$ with its closure $\mathtt{cl}(\mathcal{U}_{p})=\mathcal{P}(\mathtt{supp}(p))$, i.e.,

$$A_{\alpha}(p)=\tfrac{-\langle p,\mathbb{1}\rangle}{\alpha}\inf\nolimits_{\mathtt{cl}(\mathcal{U}_{p})}D_{1}\left(\tfrac{p}{\langle p,\mathbb{1}\rangle}\middle\|u\right)+\ln\left(\langle p,\mathbb{1}\rangle\|\mu_{\alpha,u}\|_{1}^{\alpha}\right).$$

Since the set of all probability mass functions on a finite set is compact, using the extreme value theorem we can assert that there exists a u_* achieving the infimum. Furthermore, p is absolutely continuous in u_* because otherwise $D_1\left(\frac{p}{\langle p,\mathbb{1}\rangle}\middle\|u\right)$ will be infinite. Thus $A_\alpha(p;f_{\alpha,p}^{u_*}) = A_\alpha(p)$ holds for the corresponding $f_{\alpha,p}^{u_*}$.

On the other hand,

$$\tfrac{\partial}{\partial f(x)}A_\alpha(p,f)=\left(1-\langle p,\mathbb{1}\rangle\tfrac{e^{(1-\alpha)[f(x)-\mathtt{M}_{\alpha,p}(f)(x)]}}{\|\mu_{\alpha,p}^{f}\|_1}\right)\tfrac{1-\alpha}{\alpha}p(x).$$

Thus the optimality condition $\frac{\partial}{\partial f}A_\alpha(p,f)\Big|_{f=f_{\alpha,p}}=0$ implies that $f_{\alpha,p}^{u_*}-\mathtt{M}_{\alpha,p}(f_{\alpha,p}^{u_*})=\tau$ holds p-a.e. for some $\tau\in\mathbb{R}$. Then (13), (14), (15), and (16) holds for $f_{\alpha,p}=f_{\alpha,p}^{u_*}-\alpha\tau$ by (2) and (12).

To prove the uniqueness of the Augustin dual $f_{\alpha,p}$, let us assume that there exist two distinct functions, f_1 and f_0, satisfying (13). Then f_1 and f_0 will both be maximizers of $A_\alpha(p;\cdot)$ because they will both satisfy the optimality condition. Consequently, $f_1=f_0+\tau$ should p-a.e. for some $\tau\in\mathbb{R}$ by the equality condition for (9). Then only one of f_1 and f_0 can satisfy (13) as a result of (12). ☐

The Augustin mean is customarily defined as the minimizer in the definition of Augustin information given in (1), see [1], [13]. In the following, we define Augustin mean as a function of the Augustin dual, which is the unique fixed point of the power mean operator by Theorem 1. Theorem 2 given in the following establish the equivalence of these two definitions.

Definition 4. For any $\alpha\in\mathbb{R}_+$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, $p\in\mathcal{M}^+(\mathcal{X})$ with a finite support set, the Augustin mean $q_{\alpha,p}\in\mathcal{M}^+(\mathcal{Y})$ and the Augustin channel $W_{\alpha,p}:\mathcal{X}_{\alpha,p}\to\mathcal{P}(\mathcal{Y})$ are defined as

$$q_{\alpha,p}:=\mu_{\alpha,p}^{f_{\alpha,p}}, \tag{17}$$

$$\tfrac{\mathrm{d}W_{\alpha,p}}{\mathrm{d}\nu}:=\left(\tfrac{\mathrm{d}W}{\mathrm{d}\nu}\right)^{\alpha}\left(\tfrac{\mathrm{d}q_{\alpha,p}}{\mathrm{d}\nu}\right)^{(1-\alpha)}e^{(1-\alpha)\mathtt{M}_{\alpha,p}(f_{\alpha,p})}, \tag{18}$$

where $\mathcal{X}_{\alpha,p}:=\{x\in\mathcal{X}:D_\alpha(W(x)\|q_{\alpha,p})<\infty\}$ and ν is any σ-finite reference measure.

Note that as a result of (2), we have

$$\left\langle p,\tfrac{\mathrm{d}W_{\alpha,p}}{\mathrm{d}\nu}\right\rangle=\left(\tfrac{\mathrm{d}\mu_{\alpha,p}^{\mathtt{M}_{\alpha,p}(f_{\alpha,p})}}{\mathrm{d}\nu}\right)^{\alpha}\left(\tfrac{\mathrm{d}q_{\alpha,p}}{\mathrm{d}\nu}\right)^{(1-\alpha)}\qquad \nu\text{-a.e.}. \tag{19}$$

Thus as a result of (13) and (17), we have

$$\tfrac{\mathrm{d}q_{\alpha,p}}{\mathrm{d}\nu}=\left\langle p,\tfrac{\mathrm{d}W_{\alpha,p}}{\mathrm{d}\nu}\right\rangle \qquad \nu\text{-a.e.}. \tag{20}$$

On the other hand $f_{\alpha,p}=D_\alpha(W\|q_{\alpha,p})$ by (11), (13), and (17). Thus as a result of Theorem 1, we have

$$A_\alpha(p)=\tfrac{1-\alpha}{\alpha}D_\alpha(W\|q_{\alpha,p}|p)-\langle p,\mathbb{1}\rangle\ln\langle p,\mathbb{1}\rangle, \tag{21}$$

where $D_\alpha(W\|q|p)$ stands for $\sum_x p(x)D_\alpha(W(x)\|q)$.

Theorem 2. *For any $\alpha\in\mathbb{R}_+\setminus\{1\}$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, $p\in\mathcal{P}(\mathcal{X})$ with a finite support set,*

$$I_\alpha(p;W)=D_\alpha(W\|q_{\alpha,p}|p) \tag{22}$$

$$D_{1\vee\alpha}(q_{\alpha,p}\|q)\geq D_\alpha(W\|q|p)-I_\alpha(p;W)\geq D_{1\wedge\alpha}(q_{\alpha,p}\|q) \tag{23}$$

Proof. Note that (15) and (21) imply

$$D_\alpha(W\|q|p)-D_\alpha(W\|q_{\alpha,p}|p)=\langle p,D_\alpha(W\|q)-f_{\alpha,p}\rangle.$$

Using (10), (13), and (18), we get

$$\langle p,D_\alpha(W\|q)-f_{\alpha,p}\rangle=\tfrac{1}{\alpha-1}\left\langle p,\ln\mathbf{E}_{W_{\alpha,p}}\left[\left(\tfrac{\mathrm{d}q}{\mathrm{d}q_{\alpha,p}}\right)^{1-\alpha}\right]\right\rangle.$$

Using Jensen's inequality to move $\ln(\cdot)$ inside expectation for $\alpha\in(1,\infty)$ case and the summation over x inside $\ln(\cdot)$ for $\alpha\in(0,1)$ case and invoking (20), we get

$$D_\alpha(W\|q|p)-D_\alpha(W\|q_{\alpha,p}|p)\geq D_{1\wedge\alpha}(q_{\alpha,p}\|q).$$

Thus (22) follows from (1) because $D_{1\wedge\alpha}(q_{\alpha,p}\|q)>0$ when $q_{\alpha,p}\neq q$ by [20, Theorem 8]. To establish the upper bound in (23), we switch the operations done for $\alpha>1$ and $\alpha<1$ cases above. ☐

We can also bound from below the decrease in $A_\alpha(p,f)$ due to using a suboptimal f as follows.

Lemma 1. *For any $\alpha\in\mathbb{R}_+\setminus\{1\}$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, $p\in\mathcal{P}(\mathcal{X})$ with a finite support set, there exists $\delta>0$ and $\gamma>0$ s.t.*

$$A_\alpha(p)-A_\alpha(p,f)\geq\begin{cases}\gamma\|f-f_{\alpha,p}\|_{\xi_p}^2 & \text{if } \|f-f_{\alpha,p}\|_{\xi_p}\leq\delta\\ \gamma\delta\|f-f_{\alpha,p}\|_{\xi_p} & \text{if } \|f-f_{\alpha,p}\|_{\xi_p}>\delta\end{cases}$$
$$=\gamma\|f-f_{\alpha,p}\|_{\xi_p}\left(\delta\wedge\|f-f_{\alpha,p}\|_{\xi_p}\right)$$

for all $f\in\mathcal{F}_p$, satisfying $\langle p,f\rangle=\langle p,f_{\alpha,p}\rangle$, where only non-zero entries of ξ_p are ones at the diagonal entries for x's with positive $p(x)$, i.e., $[\xi_p]_{x,z}=\mathbb{1}_{\{x=z\}}\mathbb{1}_{\{x\in\mathtt{supp}(p)\}}$.

Proof. The value $A_\alpha(p,f)$ depends only on the value of f on the support of p. Thus in the following we interpret f and $f_{\alpha,p}$ as functions on $\mathtt{supp}(p)$ and ξ_p as a $\mathtt{supp}(p)$-by-$\mathtt{supp}(p)$ matrix.

$$\nabla A_\alpha(p,f)\big|_{f=f_{\alpha,p}}=0,$$
$$\nabla^T\nabla A_\alpha(p,f)\big|_{f=f_{\alpha,p}}=\tfrac{(1-\alpha)^2}{\alpha^2}\left(\tfrac{pp^T}{\langle p,\mathbb{1}\rangle}-\xi_{\alpha,p}\right)$$

where $\xi_{\alpha,p}:=\alpha\mathrm{diag}(p)+(1-\alpha)\mathrm{diag}(p)\Lambda_{\alpha,p}\mathrm{diag}(p)$ and

$$[\Lambda_{\alpha,p}]_{x,z}=\int\tfrac{\mathrm{d}W_{\alpha,p}(x)}{\mathrm{d}q_{\alpha,p}}\tfrac{\mathrm{d}W_{\alpha,p}(z)}{\mathrm{d}q_{\alpha,p}}\mathrm{d}q_{\alpha,p}.$$

$\xi_{\alpha,p}$ is positive definite. For $\alpha\in(0,1)$ case, this is evident because $\xi_{\alpha,p}$ is the sum of a positive definite matrix and a positive semi-definite matrix. For $\alpha\in(1,\infty)$ case, we first observe that considering $\langle p,\mathbb{1}\rangle=1$ case suffices because $\xi_{\alpha,\gamma p}=\gamma\xi_{\alpha,p}$ for all $\gamma>0$. Then the proof follows from expressing $f^T\xi_{\alpha,p}f$ as the weighted sum of conditional variance of f and expectation of the square of the conditional expectation of f.

Consider the Taylor series expansion of $A_\alpha(p,f)$ around the point $f=f_{\alpha,p}$. Since $\xi_{\alpha,p}$ is positive definite, for small enough neighborhoods of $f_{\alpha,p}$ the error term in the second order Taylor expansion will contribute less than the half of the second order term. Thus

$$A_\alpha(p)-A_\alpha(p,f)\geq\tfrac{(1-\alpha)^2}{4\alpha^2}\|f-f_{\alpha,p}\|_{\xi_{\alpha,p}}^2$$

for all f satisfying $\|f-f_{\alpha,p}\|\leq\delta$ for some $\delta>0$.

For f for which $\|f-f_{\alpha,p}\|>\delta$, first note that as a result of (9) and (13) for any $\beta\in(0,1)$, we have

$$A_\alpha(p)-A_\alpha(p,f)\geq\tfrac{1}{\beta}(A_\alpha(p)-A_\alpha(p,(1-\beta)f_{\alpha,p}+\beta f)),$$
$$=\tfrac{1}{\beta}(A_\alpha(p)-A_\alpha(p,f_{\alpha,p}+(f-f_{\alpha,p})\beta)).$$

If set $\beta=\frac{\delta}{\|f-f_{\alpha,p}\|}$ desired result for $\|f-f_{\alpha,p}\|>\delta$ case follows from the case when $\|f-f_{\alpha,p}\|\leq\delta$. ☐

For $\alpha<1$ case, to calculate $q_{\alpha,p}$ and $I_\alpha(p;W)$, Augustin applied an operator repeatedly to an appropriately chosen initial output distribution, say $q_{1,p}$, see [1, Lemma 34.2]. The resulting sequence output distributions $\{q_n\}_{n\in\mathbb{Z}_+}$ satisfy $\lim_{n\to\infty}D_\alpha(W\|q_n|p)=I_\alpha(p;W)$ by the lower semicontinuity of the Rényi divergence because $\{q_n\}_{n\in\mathbb{Z}_+}$ converges to $q_{\alpha,p}$, as a result of [13, B.4] and Pinsker's inequality. The bound given in [18, Lemma 6] can replace [13, B.4] in the above argument for the case when $\alpha>1$. Lemma 2 given in the following can be used in a similar argument to prove $\lim_{n\to\infty}A_\alpha(p,f_n)=A_\alpha(p)$ where $f_{n+1}=\mathtt{M}_{\alpha,p}(f_n)$ for $n\in\mathbb{Z}_+$ and $f_0=f$. Recently, other iterative methods for calculating $I_\alpha(p;W)$ have been studied in [21] and [22]. We define $q^f_{\alpha,p}$ as

$$\tfrac{\mathrm{d}q^f_{\alpha,p}}{\mathrm{d}\nu}:=\left\langle p,\left(\tfrac{\mathrm{d}W}{\mathrm{d}\nu}\right)^\alpha\left(\tfrac{\mathrm{d}\mu^f_{\alpha,p}}{\mathrm{d}\nu}\right)^{1-\alpha}e^{(1-\alpha)D_\alpha\left(W\middle\|\mu^f_{\alpha,p}\right)}\right\rangle. \quad (24)$$

Then $q^f_{\alpha,p}\in\mathcal{P}(\mathcal{Y})$ and $q^f_{\alpha,p}$ satisfies the following identity as result of definitions of power mean and power mean operator,

$$\tfrac{\mathrm{d}q^f_{\alpha,p}}{\mathrm{d}\nu}=\left(\tfrac{\mathrm{d}\mu^{\mathtt{M}_{\alpha,p}(f)}_{\alpha,p}}{\mathrm{d}\nu}\right)^\alpha\left(\tfrac{\mathrm{d}\mu^f_{\alpha,p}}{\mathrm{d}\nu}\right)^{1-\alpha}. \quad (25)$$

Lemma 2. *For any $\alpha>1$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, $p\in\mathcal{P}(\mathcal{X})$ with a finite support set,*

$$A_\alpha(p,\mathtt{M}_{\alpha,p}(f))\geq A_\alpha(p,f)+\tfrac{\alpha-1}{\alpha}D_{\frac{1}{\alpha}}\left(q^f_{\alpha,p}\middle\|\tfrac{\mu^f_{\alpha,p}}{\|\mu^f_{\alpha,p}\|_1}\right)\geq A_\alpha(p,f),$$

both inequalities are strict unless $f=f_{\alpha,p}+\tau$ for some $\tau\in\mathbb{R}$.

Proof.

$$\begin{aligned}
&A_\alpha(p,\mathtt{M}_{\alpha,p}(f))-A_\alpha(p,f)\\
&=\tfrac{-1}{\alpha}\left\langle p,\ln\int\left(\tfrac{\mathrm{d}W}{\mathrm{d}\nu}\right)^\alpha e^{(1-\alpha)f}\left(\tfrac{\mathrm{d}\mu^f_{\alpha,p}}{\mathrm{d}\nu}\right)^{1-\alpha}\mathrm{d}\nu\right\rangle+\ln\tfrac{\|\mu^f_{\alpha,p}\|_1}{\left\|\mu^{\mathtt{M}_{\alpha,p}(f)}_{\alpha,p}\right\|_1}\\
&\overset{(a)}{\geq}\tfrac{-1}{\alpha}\ln\int\left\langle p,\left(\tfrac{\mathrm{d}W}{\mathrm{d}\nu}\right)^\alpha e^{(1-\alpha)f}\right\rangle\left(\tfrac{\mathrm{d}\mu^f_{\alpha,p}}{\mathrm{d}\nu}\right)^{1-\alpha}\mathrm{d}\nu+\ln\tfrac{\|\mu^f_{\alpha,p}\|_1}{\left\|\mu^{\mathtt{M}_{\alpha,p}(f)}_{\alpha,p}\right\|_1}\\
&\overset{(b)}{=}\left(\tfrac{\alpha-1}{\alpha}\right)\left(\ln\left\|\mu^f_{\alpha,p}\right\|_1+\tfrac{\alpha}{1-\alpha}\ln\left\|\mu^{\mathtt{M}_{\alpha,p}(f)}_{\alpha,p}\right\|_1\right)\\
&\overset{(c)}{=}\left(\tfrac{\alpha-1}{\alpha}\right)D_{\frac{1}{\alpha}}\left(q^f_{\alpha,p}\middle\|(\|\mu^f_{\alpha,p}\|_1)^{-1}\mu^f_{\alpha,p}\right)\\
&\overset{(d)}{\geq}0
\end{aligned}$$

where (a) follows from Jensen's inequality and concavity of the natural logarithm function and the inequality is strict unless there exits a $\widetilde{\tau}$ satisfying $f=\mathtt{M}_{\alpha,p}(f)+\widetilde{\tau}$ p-a.e., (b) follows from (2), (c) follows from (10) and (25), and (d) follows from [20, Theorem 8] and the inequity is strict unless $\mu^{\mathtt{M}_{\alpha,p}(f)}_{\alpha,p}=(\|\mu^f_{\alpha,p}\|_1)^{-1/\alpha}\mu^f_{\alpha,p}$ by (25), i.e., unless there exits a $\widetilde{\tau}$ satisfying $f=\mathtt{M}_{\alpha,p}(f)+\widetilde{\tau}$ p-a.e. by (4).

On the other hand, if $f=\mathtt{M}_{\alpha,p}(f)+\widetilde{\tau}$ holds p-a.e., then $f-\alpha\widetilde{\tau}$ is a fixed point of $\mathtt{M}_{\alpha,p}(\cdot)$ by (12); and $f=f_{\alpha,p}+\alpha\widetilde{\tau}$ holds p-a.e. because $f_{\alpha,p}$ is the unique fixed point of $\mathtt{M}_{\alpha,p}(\cdot)$ by Theorem 1. Thus both the inequality in (b) and the inequality in (d) are strict unless there exists a $\tilde{\tau}$ satisfying $f=\mathtt{M}_{\alpha,p}(f)+\widetilde{\tau}$. ☐

III. Quantum Setting

Let $\mathcal{H}$ be a separable (possibly infinite-dimensional) Hilbert space, and $\mathcal{S}(\mathcal{H})$ be the set of density operators (i.e., positive semi-definite operators with unit trace) on $\mathcal{H}$. For a bounded linear operator T on $\mathcal{H}$, the trace norm of T is defined as

$$\|T\|_1:=\mathrm{Tr}\left[\sqrt{T^*T}\right], \quad (26)$$

where T^* is the adjoint of T.

A classical-quantum channel $W:\mathcal{X}\to\mathcal{S}(\mathcal{H})$ maps each letter of the input alphabet $\mathcal{X}$ to a density operator on the output Hilbert space $\mathcal{H}$.

Definition 5. For any order $\alpha\in\mathbb{R}_+$, classical-quantum channel $W:\mathcal{X}\to\mathcal{S}(\mathcal{H})$, mass function $p\in\mathcal{M}^+(\mathcal{X})$ with a finite support set, and function $f\in\mathcal{F}_p$, the (adjusted) power mean $\mu^f_{\alpha,p}$ is

$$\mu^f_{\alpha,p}:=\left(\sum\nolimits_{x\in\mathcal{X}}p(x)e^{(1-\alpha)f(x)}W(x)^\alpha\right)^{1/\alpha}. \quad (27)$$

By substitution, one can confirm that (3) continues to hold for $W:\mathcal{X}\to\mathcal{S}(\mathcal{H})$ case, as well.

Lemma 3. *For any $\alpha\in\mathbb{R}_+$, $W:\mathcal{X}\to\mathcal{S}(\mathcal{H})$, and $p\in\mathcal{M}^+(\mathcal{X})$ with a finite support set, $\left\|\mu^f_{\alpha,p}\right\|_1$ is log-convex in f, i.e.,*

$$\ln\left\|\mu^{f_\beta}_{\alpha,p}\right\|_1\leq\beta\ln\left\|\mu^{f_1}_{\alpha,p}\right\|_1+(1-\beta)\ln\left\|\mu^{f_0}_{\alpha,p}\right\|_1, \quad (28)$$

for all $\beta\in(0,1)$ with $f_\beta=\beta f_1+(1-\beta)f_0$. Furthermore, the inequality is strict unless $f_1=f_0+\tau$ holds p-a.e.for some $\tau\in\mathbb{R}$.

Proof. First recall [14, Prop. 3]: For any $\alpha\in\mathbb{R}_+$, positive semi-definite operator A, bounded operator Z, the function $g(t):=\left\|(Z^*A^tZ)^{1/\alpha}\right\|_1$ is log-convex in t on $(-\infty,\infty)$.

Let A and B be the following the block operator and column operator, respectively

$$A=\bigoplus_{x\in\mathtt{supp}(p)}Ie^{(1-\alpha)[f_1(x)-f_0(x)]},$$

$$Z=\begin{bmatrix}\sqrt{p(1)W(1)^\alpha e^{(1-\alpha)f_0(1)}}\\ \sqrt{p(2)W(2)^\alpha e^{(1-\alpha)f_0(2)}}\\ \vdots\end{bmatrix},$$

where I denotes the identity operator on $\mathcal{H}$. Then for all β in $[0,1]$ we have $g(\beta)=\left\|\mu^{f_\beta}_{\alpha,p}\right\|_1$, where $f_\beta=\beta f_1+(1-\beta)f_0$. Then the log-convexity of $g(t)$ in t implies (28).

Next, we prove that the log-convexity of $g(t)$ is saturated iff the operator A is proportional to identity operator on the support of Z. This then implies that (28) is strict unless $f_1=f_0+\tau$ holds p-a.e. for some $\tau\in\mathbb{R}$.

Suppose $A=\gamma I$ for some $\gamma\in\mathbb{R}_+$. Then, for $\beta\in(0,1)$,

$$g(\beta t_1+(1-\beta)t_0)=\gamma^{\frac{\beta t_1+(1-\beta)t_0}{\alpha}}\left\|(Z^*Z)^r\right\|_1$$
$$=\gamma^{\frac{\beta t_1}{\alpha}}\left\|(Z^*Z)^{1/\alpha}\right\|_1^{\beta}\gamma^{\frac{(1-\beta)t_0}{\alpha}}\left\|(Z^*Z)^{1/\alpha}\right\|_1^{1-\beta}$$
$$=g(t_1)^{\beta}\cdot g(t_0)^{1-\beta},$$

proving the sufficient condition.

To show the necessary condition of the above equality for A and Z, it is sufficient to consider the case that Z commutes with A and that Z is a projection onto some subspace. Let $X:=(Z^*A^{t_0}Z)^{1/\alpha}$ and $Y:=(Z^*A^{t_1}Z)^{1/\alpha}$. Using the facts that Z is a projection, $X=Z^*A^{\frac{t_0}{\alpha}}Z$, and $Y=Z^*A^{\frac{t_1}{\alpha}}Z$, we have

$$g(\beta t_1+(1-\beta)t_0)=\left\|X^{\beta}\cdot Y^{1-\beta}\right\|_1.$$

On the other hand,

$$g(t_1)^{\beta}g(t_0)^{1-\beta}=\|X\|_1^{\beta}\|Y\|_1^{1-\beta}\quad=\left\|X^{\beta}\right\|_{\frac{1}{\beta}}\left\|Y^{1-\beta}\right\|_{\frac{1}{1-\beta}}.$$

Note that X commutes with Y. By Hölder's inequality, the equality $g(\beta t_1+(1-\beta)t_0)=g(t_1)^{\beta}g(t_0)^{1-\beta}$ holds only if $X\propto Y$, ensuring that $A^{\frac{t_0}{\alpha}}\propto A^{\frac{t_1}{\alpha}}$ on the support of Z. This is possible only if $A\propto I$ on the support of Z, concluding the proof. ☐

Definition 2 for the (adjusted) Gallager's function $A_\alpha(p,f)$ and the maximal (adjusted) Gallager's function $A_\alpha(p)$ naturally extend to the case of classical-quantum channels. The only difference is that now $\|\cdot\|_1$ stands for the trace norm defined in (26), rather than the L^1 norm.

The log-convexity of the trace norm of the power mean $\mu_{\alpha,p}^{f}$ in f, i.e., Lemma 3, implies the concavity of $A_\alpha(p,f)$ in f as before; thus (9) holds for classical-quantum channels, as well.

Definition 6. For any order $\alpha\in(0,1)$, any density operators ρ and σ on a separable Hilbert space $\mathcal{H}$, the Petz–Rényi divergence (see [23]) is defined as

$$D_\alpha(\rho\|\sigma):=\tfrac{1}{\alpha-1}\log\mathrm{Tr}\left[\rho^{\alpha}\sigma^{1-\alpha}\right] \tag{29}$$

for non-orthogonal ρ and σ with $\alpha\in(0,1)$. Otherwise, it is defined to be infinite.

Definition 3 of the power mean operator can be extended to classical-quantum channels, using Petz-Rényi divergence [23]. Furthermore, (12) continues to hold, because (3) holds for classical-quantum channels.

Theorem 1, continues to hold for classical-quantum channels for $\alpha\in(0,1)$. Even the proof of Theorem 1 directly extends to classical-quantum channels, using the observations we have made extended from the classical channels. The only ingredient of the proof of Theorem 1 that we have not established for the classical-quantum channels is the continuity of $\|\mu_{\alpha,p}\|_1$ in p for the case when $p\in\mathcal{P}(\mathcal{X})$ for a finite set $\mathcal{X}$, which was established in [14, Proposition 4-(c)].

For $\alpha\in(0,1)$, $W:\mathcal{X}\to\mathcal{S}(\mathcal{H})$, and $p\in\mathcal{P}(\mathcal{X})$, order α Quantum Augustin information is defined as

$$I_\alpha(p;W):=\inf\nolimits_{\sigma\in\mathcal{S}(\mathcal{H})}D_\alpha(W\|\sigma|p). \tag{30}$$

Customarily, the Augustin mean $\sigma_{\alpha,p}$ is defined as the minimizer of the infimum on the right-hand side, see [3], [9], [14]. However, we define Augustin mean $\sigma_{\alpha,p}$ as

$$\sigma_{\alpha,p}:=\mu_{\alpha,p}^{f_{\alpha,p}}. \tag{31}$$

Theorem 3, given below, implies the equivalence of these two definitions. Furthermore, (13), (27) and (31) imply

$$\sigma_{\alpha,p}=\left(\sum\nolimits_{x\in\mathcal{X}}p(x)e^{(1-\alpha)D_\alpha(W(x)\|\sigma_{\alpha,p})}W(x)^{\alpha}\right)^{1/\alpha}. \tag{32}$$

(32), holds for infinite-dimensional separable Hilbert spaces and it is another way to present the fixed point property given in (20), see [13, (38)]. Previously, the fixed-point property of $\sigma_{\alpha,p}$ has only been proved for channels with finite-dimensional Hilbert spaces $\mathcal{H}$ at the output, first by [3, Proposition 2-(b)] and then by [10, Theorem IV.14].

Theorem 3. *For any $\alpha\in(0,1)$, $W:\mathcal{X}\to\mathcal{S}(\mathcal{H})$, $p\in\mathcal{P}(\mathcal{X})$ with a finite support set, and a density operator σ on $\mathcal{H}$,*

$$I_\alpha(p;W)=D_\alpha(W\|\sigma_{\alpha,p}|p), \tag{33}$$
$$D_\alpha(W\|\sigma|p)-I_\alpha(p;W)\geq D_{1\wedge\alpha}(\sigma_{\alpha,p}\|\sigma). \tag{34}$$

Proof. For any $x\in\mathcal{X}$ satisfying $D_\alpha(W(x)\|\sigma_{\alpha,p})<\infty$, let $W_{\alpha,p}(x)\in\mathcal{S}(\mathcal{H})$ be

$$W_{\alpha,p}(x):=\sigma_{\alpha,p}^{\frac{1-\alpha}{2}}W(x)^{\alpha}\sigma_{\alpha,p}^{\frac{1-\alpha}{2}}e^{(1-\alpha)\mathtt{M}_{\alpha,p}(f_{\alpha,p})}. \tag{35}$$

Since $f_{\alpha,p}=D_\alpha(W\|\sigma_{\alpha,p})$ by (11), (13), (31), we have

$$D_\alpha(W\|\sigma|p)-D_\alpha(W\|\sigma_{\alpha,p}|p)=\langle p,D_\alpha(W\|q)-f_{\alpha,p}\rangle$$
$$=\tfrac{1}{\alpha-1}\Big\langle p,\ln\mathrm{Tr}\Big[W_{\alpha,p}\tfrac{\sigma^{1-\alpha}}{\sigma_{\alpha,p}^{1-\alpha}}\Big]\Big\rangle.$$

Here, we denote a noncommutative quotient for selfadjoint operator T and positive semi-definite operator M with $\mathtt{supp}(T)\subseteq\mathtt{supp}(M)$ as $\frac{T}{M}:=M^{-\frac{1}{2}}TM^{-\frac{1}{2}}$.

On the other hand for any $\alpha\in(0,1)$, Jensen's inequality and concavity of the logarithmic function imply

$$\tfrac{1}{\alpha-1}\Big\langle p,\ln\mathrm{Tr}\Big[W_{\alpha,p}\tfrac{\sigma^{1-\alpha}}{\sigma_{\alpha,p}^{1-\alpha}}\Big]\Big\rangle\geq\tfrac{1}{\alpha-1}\ln\Big\langle p,\mathrm{Tr}\Big[W_{\alpha,p}\tfrac{\sigma^{1-\alpha}}{\sigma_{\alpha,p}^{1-\alpha}}\Big]\Big\rangle$$
$$=\tfrac{1}{\alpha-1}\ln\Big\langle p,\mathrm{Tr}\Big[e^{(1-\alpha)f_{\alpha,p}}W^{\alpha}\sigma^{1-\alpha}\Big]\Big\rangle$$
$$\overset{(a)}{=}\tfrac{1}{\alpha-1}\ln\mathrm{Tr}\left[\sigma_{\alpha,p}^{\alpha}\sigma^{1-\alpha}\right]$$
$$=D_\alpha(\sigma_{\alpha,p}\|\sigma),$$

where (a) follows from (32). Thus (33) follows (30) because $D_\alpha(\sigma_{\alpha,p}\|\sigma)>0$ when $\sigma_{\alpha,p}\neq\sigma$ by [23], [24]. ☐

Lemmas 1 and 2 continue to hold for classical-quantum channels, as well. Their proofs, however, are more nuanced.

IV. Discussion

The Augustin information and mean have been characterized in terms of an optimization of a function defined on the input set before, see [1, Lemma 35.7] and [13, Lemma 18]. The primary novelty of our approach is the characterization of the Augustin dual $f_{\alpha,p}$ as the unique fixed point of the Augustin operator, i.e., Theorem 1 in general and (13) in particular.

For the classical-quantum channels, the natural extension of the definition of power mean given in (27) works with Petz–Rényi divergence. For Petz–Rényi divergence operationally relevant orders are the ones in $(0,1)$, see [25]. Thus we have restricted our discussion for the classical-quantum channels to Petz–Rényi divergence of orders in $(0,1)$ only.

REFERENCES

[1] U. Augustin, "Noisy channels," Habilitation Thesis, Universität Erlangen-Nürnberg, 1978, (http://bit.ly/3bsWDgG).

[2] B. Nakiboğlu, "The Augustin center and the sphere packing bound for memoryless channels," in *2017 IEEE International Symposium on Information Theory (ISIT)*, Aachen, Germany, June 2017, pp. 1401–1405.

[3] H.-C. Cheng, M. H. Hsieh, and M. Tomamichel, "Quantum sphere-packing bounds with polynomial prefactors," *IEEE Transactions on Information Theory*, vol. 65, no. 5, pp. 2872–2898, May 2019, (arXiv:1704.05703 [quant-ph]).

[4] B. Nakiboğlu, "The Sphere Packing Bound for Memoryless Channels," *Problems of Information Transmission*, vol. 56, no. 3, pp. 201–244, July 2020, (arXiv:1804.06372 [cs.IT]).

[5] ——, "A Simple Derivation of the Refined SPB for the Constant Composition Codes," in *2019 IEEE International Symposium on Information Theory (ISIT)*, Paris, France, July 2019, pp. 2659–2663.

[6] ——, "A simple derivation of the refined sphere packing bound under certain symmetry hypotheses," *Turkish Journal Of Mathematics*, vol. 44, no. 3, pp. 919–948, 2020, (arXiv:1904.12780 [cs.IT]).

[7] H.-C. Cheng and B. Nakiboğlu, "Refined strong converse for the constant composition codes," in *2020 IEEE International Symposium on Information Theory (ISIT)*, June 2020, pp. 2149–2154, (arXiv:2002.11414 [cs.IT]).

[8] S. Verdú, "Error exponents and α-mutual information," *Entropy*, vol. 23, no. 2, 2021.

[9] M. Dalai and A. Winter, "Constant compositions in the sphere packing bound for classical-quantum channels," *IEEE Transactions on Information Theory*, vol. 63, no. 9, pp. 5603–5617, Sept 2017.

[10] M. Mosonyi and T. Ogawa, "Divergence radii and the strong converse exponent of classical-quantum channel coding with constant compositions," *IEEE Transactions on Information Theory*, vol. 67, no. 3, pp. 1668–1698, 2021, arXiv:1811.10599v6 [cs.IT].

[11] I. Csiszár, "Generalized cutoff rates and Rényi's information measures," *IEEE Transactions on Information Theory*, vol. 41, no. 1, pp. 26–34, Jan 1995.

[12] M. Dalai, "Some remarks on classical and classical-quantum sphere packing bounds: Rényi vs. Kullback-Leibler," *Entropy*, vol. 19, no. 7, p. 355, 2017.

[13] B. Nakiboğlu, "The Augustin Capacity and Center," *Problems of Information Transmission*, vol. 55, no. 4, pp. 299–342, October 2019, (arXiv:1803.07937 [cs.IT]).

[14] H.-C. Cheng, L. Gao, and M.-H. Hsieh, "Properties of noncommutative Rényi and Augustin information," *Communications in Mathematical Physics*, feb 2022.

[15] E. A. Haroutunian, "Bounds for the exponent of the probability of error for a semicontinuous memoryless channel," *Problems of Information Transmission*, vol. 4, no. 4, pp. 29–39, 1968.

[16] E. A. Haroutunian, M. E. Haroutunian, and A. N. Harutyunyan, "Reliability criteria in information theory and in statistical hypothesis testing," *Foundations and Trends in Communications and Information Theory*, vol. 4, no. 2-3, pp. 97–263, 2007.

[17] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge, UK: Cambridge University Press, 2011.

[18] H.-C. Cheng and B. Nakiboğlu, "On the existence of the Augustin mean," in *2021 IEEE Information Theory Workshop (ITW)*, October 2021, pp. 1–6, (arXiv:2109.00443v1 [cs.IT]).

[19] B. Nakiboğlu, "The Rényi Capacity and Center," *IEEE Transactions on Information Theory*, vol. 65, no. 2, pp. 841–860, Feb 2019, (arXiv:1608.02424 [cs.IT]).

[20] T. v. Erven and P. Harremoës, "Rényi divergence and Kullback–Leibler divergence," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 3797–3820, July 2014.

[21] G.-R. Wang, T. C.-E., H.-C. Cheng, and Y.-H. Li, "Computing Augustin information via hybrid geodesically convex optimization," *arXiv:2402.02731 [cs.IT]*, 2024.

[22] A. Kamatsuka, K. Kazama, and T. Yoshida, "Algorithms for computing the Augustin–Csiszar mutual information and Lapidoth–Pfister mutual information," *2404.10950 [cs.IT]*, 2024.

[23] D. Petz, "Quasi-entropies for finite quantum systems," *Reports on Mathematical Physics*, vol. 23, no. 1, pp. 57–65, Feb 1986.

[24] F. Hiai, *Quantum f-Divergences in von Neumann Algebras*. Springer Singapore, 2021.

[25] K. M. R. Audenaert, M. Nussbaum, A. Szkoła, and F. Verstraete, "Asymptotic error rates in quantum hypothesis testing," *Communications in Mathematical Physics*, vol. 279, no. 1, pp. 251–283, Feb 2008.