On the Existence of the Augustin Mean

Hao-Chung Cheng
Department of Electrical Engineering and Graduate Institute of Communication Engineering, Department of Mathematics, 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*—The existence of a unique Augustin mean and its invariance under the Augustin operator are established for arbitrary input distributions with finite Augustin information for channels with countably generated output σ-algebras. The existence is established by representing the conditional Rényi divergence as a lower semicontinuous and convex functional in an appropriately chosen uniformly convex space and then invoking the Banach–Saks property in conjunction with the lower semicontinuity and the convexity. A new family of operators is proposed to establish the invariance of the Augustin mean under the Augustin operator for orders greater than one. Some members of this new family strictly decrease the conditional Rényi divergence, when applied to the second argument of the divergence, unless the second argument is a fixed point of the Augustin operator.**

I. Introduction

In sixties and seventies, Shannon's fundamental result has been strengthened for memoryless channels in terms of three exponent functions:

(i) For codes operating at rates below the Shannon capacity, the exponential decay rate of the error probability with the block length is bounded from below by the random coding exponent [1]–[5] and from above by the sphere packing exponent [4]–[7].

(ii) For codes operating at rates above the Shannon capacity, the exponential rate that the correct transmission (decoding) probability vanishes with the block length is equal to the strong converse exponent, [8]–[10].

These exponent functions have been characterized in terms of Gallager's functions [11], auxiliary channels [12], [13], and Augustin information measures [5]. To obtain the right exponent functions for cost constrained codes in terms of Gallager's functions, one has to apply the Lagrange multipliers method in a somewhat non-standard way described in [1]–[3]. The corresponding modification works for convex composition constraints, as well; see [5], [14]. This non-standard application of the Lagrange multipliers method to Gallager's function has recently been shown to be equivalent to the standard application of the Lagrange multipliers method to the Augustin information measures in [15, §5]. However, the Lagrange multipliers method is unnecessary to express the exponent functions in terms of Augustin information measures, either for composition constrained codes or for cost constrained codes. The right exponent functions are obtained by imposing the same constraints to the domain of the supremum defining Augustin capacity in terms of Augustin information [5], [15]–[25]. Such characterizations permit relatively simple derivations of tight polynomial prefactors under certain symmetry hypothesis [23], [24].

Both the Augustin information and the Rényi information (i.e. a scaled and reparametrized version of Gallager's function [26]), can be seen as generalizations of the mutual information. However, unlike the mutual information and the Rényi information, the Augustin information does not have a closed form expression. The order α Augustin information for the input distribution p is defined as

$$I_\alpha(p;W) \triangleq \inf_{q\in\mathcal{P}(\mathcal{Y})} D_\alpha(W\|\, q|\, p)\,, \tag{1}$$

where $\mathcal{P}(\mathcal{Y})$ is the set of all probability measures on the output space. For the case when the output set is a finite set (e.g. when W is a discrete memoryless channel as in [17], [27]), the compactness of $\mathcal{P}(\mathcal{Y})$, the lower semicontinuity of Rényi divergence in its second argument [28, Thm 15], and the extreme value theorem imply the existence of an order α Augustin mean $q_{\alpha,p}\in\mathcal{P}(\mathcal{Y})$ satisfying

$$I_\alpha(p;W) = D_\alpha(W\|\, q_{\alpha,p}|\, p)\,. \tag{2}$$

The Augustin mean $q_{\alpha,p}$ is unique because of the strict convexity of the Rényi divergence in its second argument described in [28, Thm 12]. Other properties of the Augustin mean and information established in [5], [15] can be derived independently, once the existence of a unique Augustin mean is established.

For channels whose output space is an arbitrary measurable space $(\mathcal{Y},\mathcal{Y})$, we no longer have the compactness of $\mathcal{P}(\mathcal{Y})$ and establishing the existence of the Augustin mean becomes a more delicate issue. It has been established for the case when p is a probability mass function with a finite support set for arbitrary channels in [5], [15]. In addition, the closed form expression for the Augustin mean has been derived for certain special cases: for Gaussian input distributions on scalar or vector Gaussian channels in [15] and for Augustin capacity achieving input distribution on additive exponential noise channels with a mean constraint in [25]. But a general

H.-C. Cheng is supported by the Young Scholar Fellowship (Einstein Program) of the Ministry of Science and Technology in Taiwan (R.O.C.) under grant number MOST 109-2636-E-002-001 & 110-2636-E-002-009, and is supported by the Yushan Young Scholar Program of the Ministry of Education in Taiwan (R.O.C.) under grant number NTU-109V0904 & NTU-110V0904.

B. Nakiboğlu is supported by the Science Academy, Turkey, under The Science Academy's Young Scientist Award Program (BAGEP), and by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under Grant 119E053.

978-1-6654-0312-2/21/$31.00 ©2021 IEEE

existence result for the Augustin mean has not been proved yet; see Remark 4 of §IV for a discussion regarding [25].

In this paper, we prove, under finite Augustin information hypothesis, the existence of a unique Augustin mean, its invariance under the Augustin operator, and its equivalence to the $q_{\widetilde{p}}$ defined in (31), which is absolute continuous in the output distribution q_p generated by the input distribution p. Our presentation will be as follows: In §II, we introduce our model and notation and prove that the infimum defining the Augustin information in (1) can be taken over the probability measures that are absolutely continuous in q_p, rather than the whole $\mathcal{P}(\mathcal{Y})$. In §III, we first use Radon–Nikodym theorem to express this optimization in $L^\tau(q_p)$ for some $\tau>1$, with the help of a functional corresponding to the conditional Rényi divergence. Then we show that this functional inherits the convexity and the norm lower semicontinuity from the conditional Rényi divergence and use them together with the Banach–Saks property to establish the existence of a unique Augustin mean. In §IV, we propose a new family of operators related to the Augustin operator, establish a new monotonicity property for the conditional Rényi divergence, see Lemma 6, and use it to establish the invariance of the Augustin mean under the Augustin operator. In §V, we briefly discuss the novelty of our approach in comparison to the previous analysis methods. Proofs of Lemmas 2, 3, 5, and equation (33) of Lemma 7 are presented in [29] because of page restrictions.

II. Preliminaries

For any measurable space $(\mathcal{Y},\mathcal{Y})$, we denote the set of all probability measures on $(\mathcal{Y},\mathcal{Y})$ by $\mathcal{P}(\mathcal{Y})$. With a slight abuse of notation we denote the set of all probability measures that are absolutely continuous with respect to a finite measure q by $\mathcal{P}(q)$. For finite measures, we use $\mathcal{M}^+(\cdot)$ instead of $\mathcal{P}(\cdot)$. We use $\|\cdot\|$ for the total variation norm and corresponding metric.

Definition 1. For any $\alpha\in(0,\infty]$, $w\in\mathcal{P}(\mathcal{Y})$, and $q\in\mathcal{M}^+(\mathcal{Y})$ *the order α Rényi divergence between w and q* is

$$D_\alpha(w\|\, q)\triangleq\begin{cases}\frac{1}{\alpha-1}\ln\int(\frac{\mathrm{d}w}{\mathrm{d}\nu})^\alpha(\frac{\mathrm{d}q}{\mathrm{d}\nu})^{1-\alpha}\nu(\mathrm{d}y) & \alpha\in\mathbb{R}_+\setminus\{1\}\\ \int\frac{\mathrm{d}w}{\mathrm{d}\nu}\left[\ln\frac{\mathrm{d}w}{\mathrm{d}\nu}-\ln\frac{\mathrm{d}q}{\mathrm{d}\nu}\right]\nu(\mathrm{d}y) & \alpha=1\\ \ln\operatorname{ess\,sup}_\nu\frac{\mathrm{d}w}{\mathrm{d}\nu}/\frac{\mathrm{d}q}{\mathrm{d}\nu} & \alpha=\infty\end{cases}$$

where ν is any measure satisfying $w\prec\nu$ and $q\prec\nu$.

If $q\in\mathcal{P}(\mathcal{Y})$, then $D_\alpha(w\|\, q)$ is positive unless $w=q$ by [28, Thm. 8] and the following Pinsker's inequality holds by [28, Thms. 3 and 31],

$$D_\alpha(w\|\, q)\geq\tfrac{1\wedge\alpha}{2}\|w-q\|^2 \qquad \forall q,w\in\mathcal{P}(\mathcal{Y}). \tag{3}$$

We denote the set of all transition probabilities[1] from $(\mathcal{X},\mathcal{X})$ to $(\mathcal{Y},\mathcal{Y})$ by $\mathcal{P}(\mathcal{Y}|\mathcal{X})$ and model the channel W as a transition probability in $\mathcal{P}(\mathcal{Y}|\mathcal{X})$. Thus [30, Thm. 10.7.2] ensures the existence of a joint distribution $p\circledast W$ on $\mathcal{X}\otimes\mathcal{Y}$ for any input distribution p in $\mathcal{P}(\mathcal{X})$. We call the $\mathcal{Y}$-marginal of $p\circledast W$ the output distribution induced by p and denote it by q_p.

$$q_p(\mathcal{E})\triangleq p\circledast W(\mathcal{X}\times\mathcal{E}) \qquad \forall\mathcal{E}\in\mathcal{Y}. \tag{4}$$

[1] See [26, Definition 9], [30, Definition 10.7.1] for the formal definition.

Applying [30, Thm. 10.7.2] for $f(x,y)=\mathbb{1}_{\{y\in\mathcal{E}\}}$ we get

$$q_p(\mathcal{E})=\int_{\mathcal{X}}W(\mathcal{E}|x)p(\mathrm{d}x) \qquad \forall\mathcal{E}\in\mathcal{Y}. \tag{5}$$

With a slight abuse of notation, for a $W\in\mathcal{P}(\mathcal{Y}|\mathcal{X})$ and $x\in\mathcal{X}$, we denote the probability measure $W(\cdot|x)\in\mathcal{P}(\mathcal{Y})$ by $W(x)$, whenever it is possible to do so without any ambiguity.

Definition 2. For any $\alpha\in(0,\infty]$, countably generated σ-algebra $\mathcal{Y}$ of subsets of $\mathcal{Y}$, $W\in\mathcal{P}(\mathcal{Y}|\mathcal{X})$, $q\in\mathcal{M}^+(\mathcal{Y})$, and $p\in\mathcal{P}(\mathcal{X})$ *the order α conditional Rényi divergence for the input distribution p* is

$$D_\alpha(W\|\, q|\, p)\triangleq\int D_\alpha(W(x)\|\, q)\, p(\mathrm{d}x). \tag{6}$$

We assume $\mathcal{Y}$ to be countably generated, so as to ensure the $\mathcal{X}$-measurablity of the integrand in (6) by[2] [15, Lemma 37].

For $\alpha=1$ case, one can confirm by substitution that the conditional Rényi divergence can be expressed in terms of the joint distribution $p\circledast W$ induced by $p\in\mathcal{P}(\mathcal{X})$ as follows

$$D_1(W\|\, q|\, p)=D_1(p\circledast W\|\, p\otimes q) \qquad \forall q\in\mathcal{P}(\mathcal{Y}), \tag{7}$$

where $p\otimes q$ is the product measure. Furthermore, (5) and (7) can be used to confirm by substitution that

$$D_1(W\|\, q|\, p)=D_1(W\|\, q_p|\, p)+D_1(q_p\|\, q) \qquad \forall q\in\mathcal{P}(\mathcal{Y}). \tag{8}$$

Definition 3. For any $\alpha\in(0,\infty]$, countably generated σ-algebra $\mathcal{Y}$, $W\in\mathcal{P}(\mathcal{Y}|\mathcal{X})$, and $p\in\mathcal{P}(\mathcal{X})$ *the order α Augustin information for the input distribution p* is given by (1).

For $\alpha=1$ case, (8) provides us a closed form expression of the Augustin information by (3): $I_1(p;W)=D_1(W\|\, q_p|\, p)$. For other orders, however, a general closed form expression does not exist either for the Augustin information or for the probability measure that achieves the infimum given in (1), called the Augustin mean. Nevertheless q_p, can be used to restrict the domain of the optimization problem defining Augustin information as follows.

Lemma 1. *For any $\alpha\in(0,\infty]$, countably generated σ-algebra $\mathcal{Y}$, $W\in\mathcal{P}(\mathcal{Y}|\mathcal{X})$, and $p\in\mathcal{P}(\mathcal{X})$,*

$$I_\alpha(p;W)=\inf\nolimits_{q\in\mathcal{P}(q_p)}D_\alpha(W\|\, q|\, p)\,. \tag{9}$$

Proof. Any $q\in\mathcal{P}(\mathcal{Y})$ can be written as the sum of absolutely continuous and singular components with respect to q_p by the Lebesgue decomposition theorem [30, Thm. 3.2.3], i.e. there exist $q_\sim\prec q_p$ and $q_\perp\perp q_p$ such that $q=q_\sim+q_\perp$. Hence, there exists an $\mathcal{E}\in\mathcal{Y}$ satisfying $q_p(\mathcal{E})=0$ and $q_\perp(\mathcal{Y}\setminus\mathcal{E})=0$ because $q_\perp\perp q_p$. Then $W(\mathcal{E}|x)=0$ p-a.s. by (5) and consequently

$$D_\alpha(W(x)\|\, q)=D_\alpha(W(x)\|\, q_\sim) \qquad p\text{-a.s.}$$

Thus $\|q_\sim\|>0$ for all q satisfying $D_\alpha(W\|\, q|\, p)<\infty$ and

$$D_\alpha(W\|\, q|\, p)=D_\alpha\Big(W\|\,\tfrac{q_\sim}{\|q_\sim\|}\Big|\, p\Big)-\ln\|q_\sim\| \tag{10}$$

for all $q\in\mathcal{P}(\mathcal{Y})$ satisfying $D_\alpha(W\|\, q|\, p)<\infty$. Then we can replace $\mathcal{P}(\mathcal{Y})$ with $\mathcal{P}(q_p)$ in (1), without changing the value of the infimum because $-\ln\|q_\sim\|\geq 0$ and $\frac{q_\sim}{\|q_\sim\|}\in\mathcal{P}(q_p)$. ☐

[2] [15, Lemma 37] establishes $\mathcal{X}$-measurability for $q\in\mathcal{P}(\mathcal{Y})$ and $\alpha\in\mathbb{R}_+$ case, but a similar proof works for $q\in\mathcal{M}^+(\mathcal{Y})$ and $\alpha\in(0,\infty]$ case.

III. Existence of a Unique Augustin Mean

The uniform convexity[3] of L^{τ} for $\tau>1$, plays a central role in our proof of the existence of a unique Augustin mean for input distributions with finite Augustin information. Let us first recall the definition of the τ-norm. For any $\tau\geq 1$ and q_p-measurable function $f:\mathcal{Y}\to\mathbb{R}$, the τ-norm of f is

$$\|f\|_{\tau}\triangleq\left(\int|f(y)|^{\tau}q_p(\mathrm{d}y)\right)^{1/\tau}. \tag{11}$$

The set of all finite τ-norm functions $L^{\tau}(q_p)$ form a complete normed vector space, i.e. Banach space, under the pointwise addition and the scalar multiplication by [30, Thm. 4.1.3]

$$L^{\tau}(q_p)\triangleq\{f:\|f\|_{\tau}<\infty\}. \tag{12}$$

As a result of Radon–Nikdoym theorem [30, Thm. 3.2.2], we know that elements of $\mathcal{P}(q_p)$ can be represented via their Radon–Nikodym derivatives with respect to q_p, which will be non-negative functions of unit norm in $L^1(q_p)$. By taking pointwise τ^{th} root of these Radon–Nikodym derivatives, we can obtain analogous representations in $L^{\tau}(q_p)$ for any positive τ. Motivated by these observations we define the following subsets of $L^{\tau}(q_p)$:

$$\mathcal{B}^{\tau}(q_p)\triangleq\{f\in L^{\tau}(q_p):f(y)\geq 0\ q_p\text{-a.s.}\}, \tag{13}$$

$$\mathcal{B}^{\tau}_1(q_p)\triangleq\{f\in\mathcal{B}^{\tau}(q_p):\|f\|_{\tau}=1\}, \tag{14}$$

$$\mathcal{B}^{\tau}_{\leq 1}(q_p)\triangleq\{f\in\mathcal{B}^{\tau}(q_p):\|f\|_{\tau}\leq 1\}. \tag{15}$$

Let $\omega_{\tau}(\cdot):\mathcal{B}^{\tau}(q_p)\to\mathcal{M}^{+}(q_p)$ be the function defined through the following relation

$$\omega_{\tau}(f)(\mathcal{E})\triangleq\int_{\mathcal{E}}[f(y)]^{\tau}q_p(\mathrm{d}y)\qquad\forall f\in\mathcal{B}^{\tau}(q_p),\mathcal{E}\in\mathcal{Y}. \tag{16}$$

Using the conditional Rényi divergence and $\omega_1(\cdot)$, we can define the functional $D_{\alpha}(W\|\,\omega_1(\cdot)|\,p)$ on $\mathcal{B}^1(q_p)$, which inherits the convexity and norm lower semicontinuity from the Rényi divergence by the linearity and continuity of $\omega_1(\cdot)$. Lemmas 2 and 3 demonstrate that for an appropriately chosen $\tau>1$, the functional $D_{\alpha}(W\|\,\omega_{\tau}(\cdot)|\,p)$ on $\mathcal{B}^{\tau}(q_p)$ inherits the convexity and norm lower semicontinuity, as well. These observations are important because, unlike $L^1(q_p)$, $L^{\tau}(q_p)$ is uniformly convex for any $\tau>1$, and thus it has the Banach–Saks property.

Definition 4. Let $\mathfrak{D}_{\alpha}(\cdot):\mathcal{B}^{\tau_{\alpha}}(q_p)\to(-\infty,\infty]$ be

$$\mathfrak{D}_{\alpha}(f)\triangleq D_{\alpha}(W\|\,\omega_{\tau_{\alpha}}(f)|\,p) \tag{17}$$

for all $f\in\mathcal{B}^{\tau_{\alpha}}(q_p)$ and $\alpha\in(0,\infty]$, where

$$\tau_{\alpha}\triangleq\begin{cases}2 & \alpha\in[0.5,\infty]\\ \frac{1}{1-\alpha} & \alpha\in(0,0.5)\end{cases}. \tag{18}$$

Lemma 2. *For all $\alpha\in(0,\infty]$, functional $\mathfrak{D}_{\alpha}(\cdot)$, defined in* (17)*, is convex on $\mathcal{B}^{\tau_{\alpha}}(q_p)$.*

Lemma 3. *For all $\alpha\in(0,\infty]$, functional $\mathfrak{D}_{\alpha}(\cdot)$, defined in* (17)*, is norm lower semicontinuous on $\mathcal{B}^{\tau_{\alpha}}(q_p)$.*

[3]Usually, p rather than τ is used to name the norm and the associated Banach space. We deviate from the convention to reserve the symbol p for the input distributions.

Proofs of Lemmas 2 and 3 are presented in [29, Appendix A and Appendix B].

Lemma 4. *For all $\alpha\in(0,\infty]$, there exists an $f_{\alpha}\in\mathcal{B}^{\tau_{\alpha}}(q_p)$ satisfying $\|f_{\alpha}\|_{\tau_{\alpha}}=1$ and*

$$\mathfrak{D}_{\alpha}(f_{\alpha})=I_{\alpha}(p;W). \tag{19}$$

Proof. Note that $\omega_{\tau}(\gamma f)=\gamma^{\tau}\omega_{\tau}(f)$ for all $\tau\geq 1$ and $\gamma\geq 0$ by (16). Thus

$$\mathfrak{D}_{\alpha}\left(f/\|f\|_{\tau_{\alpha}}\right)=\mathfrak{D}_{\alpha}(f)+\ln\|f\|_{\tau_{\alpha}}^{\tau_{\alpha}}. \tag{20}$$

for all $f\in\mathcal{B}^{\tau_{\alpha}}(q_p)$ by (17). Consequently,

$$\inf\nolimits_{f\in\mathcal{B}^{\tau_{\alpha}}_{\leq 1}(q_p)}\mathfrak{D}_{\alpha}(f)=\inf\nolimits_{f\in\mathcal{B}^{\tau_{\alpha}}_{1}(q_p)}\mathfrak{D}_{\alpha}(f).$$

Hence the definition of $\mathfrak{D}_{\alpha}(\cdot)$, the Radon–Nikdoym theorem [30, Thm 3.2.2], and Lemma 1 imply

$$\inf\nolimits_{f\in\mathcal{B}^{\tau_{\alpha}}_{\leq 1}(q_p)}\mathfrak{D}_{\alpha}(f)=I_{\alpha}(p;W). \tag{21}$$

Thus there exists a sequence $\{f_n\}\subset\mathcal{B}^{\tau_{\alpha}}_{\leq 1}(q_p)$ satisfying[4]

$$\mathfrak{D}_{\alpha}(f_n)\downarrow I_{\alpha}(p;W), \tag{22}$$

$$\sum\nolimits_{n\in\mathbb{Z}_{+}}[\mathfrak{D}_{\alpha}(f_n)-I_{\alpha}(p;W)]<\infty. \tag{23}$$

$L^{\tau_{\alpha}}(q_p)$ has the Banach–Saks property for $\tau_{\alpha}\in(1,2]$ by [30, Cor. 4.7.17], because it is uniformly convex by [30, Thm. 4.7.15]. Thus for the norm bounded sequence $\{f_n\}$, there exist a subsequence $\{f_{n_k}\}$ and an $f_{\alpha}\in L^{\tau_{\alpha}}(q_p)$ such that

$$\lim\nolimits_{k\to\infty}\left\|\tfrac{f_{n_1}+\cdots+f_{n_k}}{k}-f_{\alpha}\right\|_{\tau_{\alpha}}=0. \tag{24}$$

Furthermore, $f_{\alpha}\in\mathcal{B}^{\tau_{\alpha}}_{\leq 1}(q_p)$ because $\mathcal{B}^{\tau_{\alpha}}_{\leq 1}(q_p)$ is closed and $\frac{f_{n_1}+\cdots+f_{n_k}}{k}\in\mathcal{B}^{\tau_{\alpha}}_{\leq 1}(q_p)$ for all k by the non-negativity of f_n's and the triangle inequality of $\|\cdot\|_{\tau_{\alpha}}$.
The norm lower semicontinuity of $\mathfrak{D}_{\alpha}(\cdot)$ established in Lemma 3, $f_{\alpha}\in\mathcal{B}^{\tau_{\alpha}}_{\leq 1}(q_p)$, and (24) imply

$$\mathfrak{D}_{\alpha}(f_{\alpha})\leq\liminf\nolimits_{k\to\infty}\mathfrak{D}_{\alpha}\left(\tfrac{f_{n_1}+\cdots+f_{n_k}}{k}\right). \tag{25}$$

On the other hand, the convexity of $\mathfrak{D}_{\alpha}(\cdot)$ established in Lemma 2 implies

$$\mathfrak{D}_{\alpha}\left(\tfrac{f_{n_1}+\cdots+f_{n_k}}{k}\right)\leq\tfrac{\mathfrak{D}_{\alpha}(f_{n_1})+\cdots+\mathfrak{D}_{\alpha}(f_{n_k})}{k}. \tag{26}$$

$\mathfrak{D}_{\alpha}(f_{\alpha})\leq I_{\alpha}(p;W)$ by (22), (23), (25) and (26). Hence, (19) follows from (21) and the fact that $f_{\alpha}\in\mathcal{B}^{\tau_{\alpha}}_{\leq 1}(q_p)$. Furthermore, $\|f_{\alpha}\|_{\tau_{\alpha}}=1$ as a result of (19), (20), and (21). ☐

For finite orders, Lemma 5, expresses Lemma 4 in terms of probability measures and strengthens it with uniqueness assertion for the finite Augustin information case.

Lemma 5. *For any $\alpha\in\mathbb{R}_{+}$, channel $W\in\mathcal{P}(\mathcal{Y}|\mathcal{X})$ with a countably generated output σ-algebra $\mathcal{Y}$, and input distribution $p\in\mathcal{P}(\mathcal{X})$ with a finite* order α Augustin information*, there exists a unique $q_{\alpha,p}\in\mathcal{P}(\mathcal{Y})$ satisfying*

$$I_{\alpha}(p;W)=D_{\alpha}(W\|\,q_{\alpha,p}|\,p), \tag{27}$$

called the order α Augustin mean for the input distribution p. Furthermore, $q_{\alpha,p}$ is absolutely continuous in q_p, i.e. $q_{\alpha,p}\prec q_p$.

Proof of Lemma 5 is presented in [29, Appendix C].

[4]For example let f_n be such that $\mathfrak{D}_{\alpha}(f_n)\leq I_{\alpha}(p;W)+(1/n)^2$.

IV. Fixed Point Properties of Augustin Mean

The existence of a unique Augustin mean $q_{\alpha,p}$ and its absolute continuity in q_p are important observations. But they do not provide an easy way to decide whether $q_{\alpha,p}=q$ for a $q\prec q_p$ or not. For input distributions that are probability mass functions with finite support set, this issue was addressed by characterizing $q_{\alpha,p}$ as the only fixed point of the Augustin operator that is equivalent to q_p, see[5] [5, Lemma 34.2], [15, Lemma 13]. Our main goal in this section is to establish an analogous characterization of the Augustin mean $q_{\alpha,p}$ for a general input distribution p merely by assuming that $I_\alpha(p;W)$ is finite, see Lemma 7. Let $\mathcal{Q}_{\alpha,p}$, $\mathcal{X}^q_{\alpha,p}$, and $\mathcal{X}^q_{\alpha,p}$ be

$$\mathcal{Q}_{\alpha,p} \triangleq \{q\in\mathcal{P}(\mathcal{Y}) : D_\alpha(W\| q| p)<\infty\},$$
$$\mathcal{X}^q_{\alpha,p} \triangleq \{x : D_\alpha(W(x)\| q) < \infty\},$$
$$\mathcal{X}^q_{\alpha,p} \triangleq \{\mathcal{E}\cap\mathcal{X}^q_{\alpha,p} : \mathcal{E}\in\mathcal{X}\}.$$

Definition 5. For any $\alpha\in\mathbb{R}_+$, countably generated σ-algebra $\mathcal{Y}$ of subsets of $\mathcal{Y}$, $W\in\mathcal{P}(\mathcal{Y}|\mathcal{X})$, $q\in\mathcal{P}(\mathcal{Y})$, and $x\in\mathcal{X}^q_{\alpha,p}$

$$\tfrac{\mathrm{d}W^q_\alpha(x)}{\mathrm{d}\nu} \triangleq e^{(1-\alpha)D_\alpha(W(x)\|q)}\left(\tfrac{\mathrm{d}W(x)}{\mathrm{d}\nu}\right)^{\alpha}\left(\tfrac{\mathrm{d}q}{\mathrm{d}\nu}\right)^{1-\alpha}. \tag{28}$$

Then W^q_α defines a transition probability called *the order α tilted channel* $W^q_\alpha\in\mathcal{P}(\mathcal{Y}|\mathcal{X}^q_{\alpha,p})$.

Remark 1. *If $q\in\mathcal{Q}_{\alpha,p}$, then $p(\mathcal{X}^q_{\alpha,p})=1$. Hence, for input distributions that are absolutely continuous in p, the fact that W^q_α is an element of $\mathcal{P}(\mathcal{Y}|\mathcal{X}^q_{\alpha,p})$ rather than $\mathcal{P}(\mathcal{Y}|\mathcal{X})$ is inconsequential.*

Definition 6. Under the hypothesis of Lemma 5, *the Augustin operator* $\mathtt{T}_{\alpha,p}(\cdot):\mathcal{Q}_{\alpha,p}\to\mathcal{P}(\mathcal{Y})$ is defined as

$$\mathtt{T}_{\alpha,p}(q)(\mathcal{E})\triangleq\mathbf{E}_p[W^q_\alpha(\mathcal{E}|\mathsf{X})]\qquad\forall\mathcal{E}\in\mathcal{Y},\ q\in\mathcal{Q}_{\alpha,p}. \tag{29}$$

Furthermore, for any $\beta\in\mathbb{R}_+$ satisfying $D_\beta(\mathtt{T}_{\alpha,p}(q)\|q)<\infty$, *the tilted Augustin operator* $\mathtt{T}^\beta_{\alpha,p}(q)$ is defined as

$$\tfrac{\mathrm{d}\mathtt{T}^\beta_{\alpha,p}(q)}{\mathrm{d}\nu} \triangleq e^{(1-\beta)D_\beta(\mathtt{T}_{\alpha,p}(q)\|q)}\left(\tfrac{\mathrm{d}\mathtt{T}_{\alpha,p}(q)}{\mathrm{d}\nu}\right)^{\beta}\left(\tfrac{\mathrm{d}q}{\mathrm{d}\nu}\right)^{1-\beta}. \tag{30}$$

The Augustin operator has been used before either implicitly [7], [16], [32] or explicitly [5], [15], [25], [31]. However, to the best of our knowledge, the tilted Augustin operator is first defined and analyzed in the present work.

Lemma 6. *Under the hypothesis of Lemma 5, if either $\alpha\in(0,1)$ and $\beta\in(0,1]$, or $\alpha\in(1,\infty)$ and $\beta\in(0,1\wedge\frac{1}{\alpha-1})$, then for any $q\in\mathcal{Q}_{\alpha,p}$ we have*

$$\begin{aligned}D_\alpha(W\|q|p)-D_\alpha\big(W\|\mathtt{T}^\beta_{\alpha,p}(q)\big|p\big)&\\ \geq \beta D_{1-\beta|\alpha-1|^+}(\mathtt{T}_{\alpha,p}(q)\|q)+(1-\beta)D_\beta(\mathtt{T}_{\alpha,p}(q)\|q)&\\ \geq \tfrac{\beta(2-\beta(\alpha\vee 1))}{2}\|\mathtt{T}_{\alpha,p}(q)-q\|^2.&\end{aligned}$$

A particular case of Lemma 6 for $\alpha\in(0,1)$ and $\beta=1$ was proved in [5, p. 236] and [15, (B.4)], and was used to show that the Augustin mean is a fixed point of the Augustin operator[6] in [5, Lemma 34.2] and [15, Lemma 13 (c)] for $\alpha\in(0,1)$. Lemma 6 allows us to invoke this simpler argument for establishing the fixed point property for $\alpha\in(1,\infty)$ case.

Proof.

$$\begin{aligned}
&D_\alpha(W\|q|p)-D_\alpha\big(W\|\mathtt{T}^\beta_{\alpha,p}(q)\big|p\big)\\
&=\tfrac{1}{1-\alpha}\mathbf{E}_p\left[\ln\int\left(\tfrac{\mathrm{d}\mathtt{T}^\beta_{\alpha,p}(q)}{\mathrm{d}q}\right)^{1-\alpha}W^q_\alpha(\mathrm{d}y|\mathsf{X})\right]\\
&\overset{(a)}{\geq}\begin{cases}\tfrac{1}{1-\alpha}\mathbf{E}_p\left[\int\ln\left(\tfrac{\mathrm{d}\mathtt{T}^\beta_{\alpha,p}(q)}{\mathrm{d}q}\right)^{1-\alpha}W^q_\alpha(\mathrm{d}y|\mathsf{X})\right] & \text{if } \alpha<1\\ \tfrac{1}{1-\alpha}\ln\mathbf{E}_p\left[\int\left(\tfrac{\mathrm{d}\mathtt{T}^\beta_{\alpha,p}(q)}{\mathrm{d}q}\right)^{1-\alpha}W^q_\alpha(\mathrm{d}y|\mathsf{X})\right] & \text{if } \alpha>1\end{cases}\\
&\overset{(b)}{=}\begin{cases}\int\tfrac{\mathrm{d}\mathtt{T}_{\alpha,p}(q)}{\mathrm{d}q}\ln\left(\tfrac{\mathrm{d}\mathtt{T}^\beta_{\alpha,p}(q)}{\mathrm{d}q}\right)q(\mathrm{d}y) & \text{if } \alpha<1\\ \tfrac{1}{1-\alpha}\ln\int\left(\tfrac{\mathrm{d}\mathtt{T}^\beta_{\alpha,p}(q)}{\mathrm{d}q}\right)^{1-\alpha}\tfrac{\mathrm{d}\mathtt{T}_{\alpha,p}(q)}{\mathrm{d}q}q(\mathrm{d}y) & \text{if } \alpha>1\end{cases}\\
&\overset{(c)}{=}\begin{cases}\beta D_1(\mathtt{T}_{\alpha,p}(q)\|q)+(1-\beta)D_\beta(\mathtt{T}_{\alpha,p}(q)\|q) & \text{if } \alpha<1\\ \beta D_{1+\beta(1-\alpha)}(\mathtt{T}_{\alpha,p}(q)\|q)+(1-\beta)D_\beta(\mathtt{T}_{\alpha,p}(q)\|q) & \text{if } \alpha>1\end{cases}
\end{aligned}$$

where (a) follows from Jensen's inequality and the concavity of natural logarithm function, (b) follows from (29) and Fubini's theorem [30, Thm. 3.4.4], (c) follows (30). The second inequality of the lemma follows from (3). ☐

For most, but not all, cases of interest $W(x)\prec q_p$ p-a.s., e.g. see [29, Example 1]. To avoid introducing "$W(x)\prec q_p$ p-a.s." as a separate hypothesis, we define $q_{\widetilde{p}}$ as follows

$$\tfrac{\mathrm{d}q_{\widetilde{p}}}{\mathrm{d}q_p} \triangleq \mathbf{E}_p\left[\tfrac{\mathrm{d}W_\sim(\mathsf{X})}{\mathrm{d}q_p}\right], \tag{31}$$

where $W_\sim(x)$ is the q_p-absolutely continuous part of $W(x)$. Note that $W_\sim(x)=W(x)$ p-a.s. and thus $q_{\widetilde{p}}=q_p$ whenever $W(x)\prec q_p$ p-a.s. and thus whenever $I_1(p;W)<\infty$.

Lemma 7. *Under the hypothesis of Lemma 5, $q_{\alpha,p}\sim q_{\widetilde{p}}$,*

$$\mathtt{T}_{\alpha,p}(q_{\alpha,p})=q_{\alpha,p}, \tag{32}$$

$$D_{1\vee\alpha}(q_{\alpha,p}\|q)\geq D_\alpha(W\|q|p)-I_\alpha(p;W)\geq D_{1\wedge\alpha}(q_{\alpha,p}\|q), \tag{33}$$

for all $q\in\mathcal{P}(\mathcal{Y})$. Furthermore, if $q_{\widetilde{p}}\prec q$ and $\mathtt{T}_{\alpha,p}(q)=q$ for a $q\in\mathcal{P}(\mathcal{Y})$, then $q_{\alpha,p}=q$.

Remark 2. *For $\alpha\in(1,\infty)$, $D_\alpha(W\|q|p)$ is finite and $\mathtt{T}_{\alpha,p}(q)$ is defined only for q's satisfying $q_p\prec q$; furthermore finite $I_\alpha(p;W)$ hypothesis of Lemma 5 implies $q_{\widetilde{p}}=q_p$. Thus $q_{\widetilde{p}}\prec q$ hypothesis can be omitted for $\alpha\in(1,\infty)$. For $\alpha\in(0,1)$, however, $q_{\widetilde{p}}\prec q$ hypothesis cannot be dropped; see [15, footnote 11], and [20, (15)], [21, Thm. IV.14] for classical-quantum channels, and a related problem in [33, Lem. 5].*

Proof. For $\alpha=1$ case lemma follows from (3) and (8) for $q_{1,p}=q_p$. For other orders, first apply Lemma 6 for $\beta=1\wedge{}^1\!/\!_\alpha$

$$D_\alpha(W\|q|p)-D_\alpha\Big(W\|\mathtt{T}^{1\wedge 1/\alpha}_{\alpha,p}(q)\Big|p\Big)\geq\tfrac{1\wedge\alpha}{2\alpha}\|\mathtt{T}_{\alpha,p}(q)-q\|^2.$$

Then (32) follows from by (1) and (27).

For $\alpha\in(0,1)$, $W^q_\alpha(x)\prec q$ whenever $W^q_\alpha(x)$ is defined. Thus using (29), (32), and $q_{\alpha,p}\prec q_p$ established in Lemma 5, we can obtain the following identity for all $\alpha\in(0,1)$,

[5] This is the case even for certain quantum models [31, Proposition 4].

[6] Although we will not rely on it, it is worth mentioning that $\mathtt{T}^\beta_{\alpha,p}(q)=q$ holds either for all positive real β's or for none.

$$\begin{aligned}\tfrac{\mathrm{d}q_{\alpha,p}}{\mathrm{d}q_p} &= \left(\mathbf{E}_p\left[\left(\tfrac{\mathrm{d}W_{\sim}(\mathsf{X})}{\mathrm{d}q_p}\right)^{\alpha} e^{(1-\alpha)D_{\alpha}(W(\mathsf{X})\|\, q_{\alpha,p})}\right]\right)^{1/\alpha}, \qquad (34)\\ &\geq \left(\mathbf{E}_p\left[\left(\tfrac{\mathrm{d}W_{\sim}(\mathsf{X})}{\mathrm{d}q_p}\right)^{\alpha}\right]\right)^{1/\alpha},\end{aligned}$$

where the inequality follows from $D_{\alpha}(W(\mathsf{X})\|\, q_{\alpha,p}) \geq 0$. Thus $q_{\widetilde{p}} \prec q_{\alpha,p}$ by (31) because $\mathbf{E}[\mathsf{Z}^{\alpha}] > 0$ iff $\mathbf{E}[\mathsf{Z}] > 0$ for any non-negative random variable Z. Furthermore, for any $\alpha \in (0,1)$

$$\begin{aligned}D_{\alpha}(W\|\, q|\, p) &\overset{(a)}{=} D_{\alpha}(W\|\, q_{\sim}|\, p)\\ &\overset{(b)}{=} \tfrac{1}{\alpha-1}\mathbf{E}_p\left[\ln\!\int\!\left(\tfrac{\mathrm{d}W_{\sim}(\mathsf{X})}{\mathrm{d}q_p}\right)^{\alpha}\left(\tfrac{\mathrm{d}q_{\sim}}{\mathrm{d}q_p}\right)^{1-\alpha} q_p(\mathrm{d}y)\right]\\ &\overset{(c)}{=} \tfrac{1}{\alpha-1}\mathbf{E}_p\left[\ln\!\int\!\left(\tfrac{\mathrm{d}W_{\sim}(\mathsf{X})}{\mathrm{d}q_p}\right)^{\alpha}\left(\tfrac{\mathrm{d}q_{ac}}{\mathrm{d}q_p}\right)^{1-\alpha} q_p(\mathrm{d}y)\right]\\ &\overset{(d)}{=} D_{\alpha}(W\|\, q_{ac}|\, p), \qquad (35)\end{aligned}$$

where $q_{\sim}$ is q_p-absolutely continuous part of q, q_{ac} is $q_{\widetilde{p}}$-absolutely continuous part of both $q_{\sim}$ and q, (a) follows from (10), (b) follows from the definition of Rényi divergence for $\alpha \in (0,1)$ and $q_{\sim} \prec q_p$, (c) follows from (31) because as a result only the $q_{\widetilde{p}}$-absolutely continuous part of $q_{\sim}$ contributes to the integral p-a.s., (d) follows from the definition of Rényi divergence of $\alpha \in (0,1)$ and $q_{ac} \prec q_p$. Note that (35) implies $q_{\alpha,p} \prec q_{\widetilde{p}}$ and hence $q_{\alpha,p} \sim q_{\widetilde{p}}$ for $\alpha \in (0,1)$ because we have already established $q_{\widetilde{p}} \prec q_{\alpha,p}$.

For $\alpha \in (1,\infty)$, $D_{\alpha}(W\|\, q_{\alpha,p}|\, p) < \infty$ implies $W(x) \prec q_{\alpha,p}$ p-a.s. and $q_{\widetilde{p}} = q_p$. Thus $q_p \prec q_{\alpha,p}$ by (5) and consequently $q_p \sim q_{\alpha,p}$ by Lemma 5. Thus $q_{\alpha,p} \sim q_{\widetilde{p}}$, for $\alpha \in (1,\infty)$.

Let $s \in \mathcal{P}(\mathcal{Y})$ satisfy $\mathtt{T}_{\alpha,p}(s) = s$ and $q_{\widetilde{p}} \prec s$, and q_{ac} be $q_{\widetilde{p}}$-absolutely continuous part of a $q \in \mathcal{P}(\mathcal{Y})$. For $\alpha > 1$, finite $I_{\alpha}(p;W)$ hypothesis implies $q_{\widetilde{p}} = q_p$. Then invoking (35) for $\alpha \in (0,1)$ and (10) for $\alpha \in (1,\infty)$ we get

$$\begin{aligned}&D_{\alpha}(W\|\, q|\, p) - D_{\alpha}(W\|\, s|\, p)\\ &\quad= \tfrac{1}{\alpha-1}\mathbf{E}_p\left[\ln\int\left(\tfrac{\mathrm{d}q_{ac}}{\mathrm{d}s}\right)^{1-\alpha} W_{\alpha}^{s}(\mathrm{d}y|\mathsf{X})\right]\\ &\quad\overset{(a)}{\geq} \begin{cases}\tfrac{1}{\alpha-1}\ln\mathbf{E}_p\left[\int\left(\tfrac{\mathrm{d}q_{ac}}{\mathrm{d}s}\right)^{1-\alpha} W_{\alpha}^{s}(\mathrm{d}y|\mathsf{X})\right] & \text{if } \alpha < 1\\ \tfrac{1}{\alpha-1}\mathbf{E}_p\left[\int\ln\left(\tfrac{\mathrm{d}q_{ac}}{\mathrm{d}s}\right)^{1-\alpha} W_{\alpha}^{s}(\mathrm{d}y|\mathsf{X})\right] & \text{if } \alpha > 1\end{cases}\\ &\quad\overset{(b)}{=} \begin{cases}\tfrac{1}{\alpha-1}\ln\int\tfrac{\mathrm{d}\mathtt{T}_{\alpha,p}(s)}{\mathrm{d}s}\left(\tfrac{\mathrm{d}q_{ac}}{\mathrm{d}s}\right)^{1-\alpha} s(\mathrm{d}y) & \text{if } \alpha < 1\\ -\int\tfrac{\mathrm{d}\mathtt{T}_{\alpha,p}(s)}{\mathrm{d}s}\ln\left(\tfrac{\mathrm{d}q_{ac}}{\mathrm{d}s}\right) s(\mathrm{d}y) & \text{if } \alpha > 1\end{cases}\\ &\quad\overset{(c)}{\geq} D_{1\wedge\alpha}(s\|\, q) \qquad (36)\end{aligned}$$

where (a) follows from Jensen's inequality and the concavity of natural logarithm function, (b) follows from (29) and Fubini's theorem [30, Thm. 3.4.4], (c) follows from $\mathtt{T}_{\alpha,p}(s) = s$, [15, Lemma 1], and $q_{ac} \leq q$. Thus $D_{\alpha}(W\|\, q|\, p) > D_{\alpha}(W\|\, s|\, p)$ for all $q \in \mathcal{P}(\mathcal{Y}) \setminus \{s\}$ by (3) and $s = q_{\alpha,p}$ by Lemma 5, for any $s \in \mathcal{P}(\mathcal{Y})$ satisfying both $\mathtt{T}_{\alpha,p}(s) = s$ and $q_{\widetilde{p}} \prec s$. Proof of (33) is presented in [29, Appendix D]. ☐

Remark 3. *The identity* (34) *holds not only for* $\alpha \in (0,1)$ *but for any* $\alpha \in \mathbb{R}_+$ *satisfying* $I_{\alpha}(p;W) < \infty$. *For* $\alpha \in (1,\infty)$ *case, if* $I_{\alpha}(p;W) < \infty$ *then* (34) *follows from* (29), (32), $q_{\alpha,p} \prec q_p$, *and the fact that* $W(x) \prec q_{\alpha,p}$ *p-a.s. and it can be written as*

$$\tfrac{\mathrm{d}q_{\alpha,p}}{\mathrm{d}q_p} = \left(\mathbf{E}_p\left[\left(\tfrac{\mathrm{d}W(\mathsf{X})}{\mathrm{d}q_p}\right)^{\alpha} e^{(1-\alpha)D_{\alpha}(W(\mathsf{X})\|\, q_{\alpha,p})}\right]\right)^{1/\alpha}. \qquad (37)$$

For $\alpha \in (0,1)$ *case,* (37) *holds whenever* $W(x) \prec q_p$ *p-a.s., e.g. when* p *is a probability mass function as in [15, (38)].*

Remark 4. *In [25], the channel* W *is assumed to satisfy* $W(x) \prec q_p$ *p-a.s. for all* p*, which is a reasonable but not completely general assumption. Ref. [25] defines the Augustin mean, which it calls* $\langle\alpha\rangle$*-response to* p*, as the element of* $\mathcal{P}(q_p)$ *satisfying* (37)*; see [25, (92)]. The existence of a unique element of* $\mathcal{P}(q_p)$ *satisfying* (37)*, however, is not a definition, but an assertion that requires a proof. Furthermore, the proof of Lemma 13-(c)-i and 13-(d)-i in [15], had previously shown for any probability mass function* p *with a finite support set that when a* $q \in \mathcal{P}(\mathcal{Y})$ *satisfying both* $\mathtt{T}_{\alpha,p}(q) = q$ *and* $q_p \prec q$ *exists, it has to be the Augustin mean, and these arguments are valid as they are for general input distributions* p*, as well.*

V. Discussions

Augustin information was defined for arbitrary channels with countably generated output σ-algebras in [15, §5.4]. The existence of a unique Agustin center was confirmed both for the unconstrained and cost constrained cases for channels with countably separated input σ-algebras, provided that Augustin capacity is finite; see [15, Thms 4 and 5]. However, the existence of a unique Augustin mean was not proved for general input distributions on these channels in [15].

The technical challenge arises from the lack of closed form expression for the minimizer in (1). If the output set is finite, then the probability simplex is compact; thus the lower semicontinuity and the extreme value theorem implies the existence of a minimizer. When the output space is an arbitrary measurable space, the existence of the minimizer has only been proved for input distributions with finite support set, [5], [15], [31]. In these proofs, finite support of the input distribution is used to reach an intermediary problem with compactness. Thus previous proofs of the existence of the Augustin mean relied on some form of compactness directly.

The novelty of our approach is the use of Banach–Saks property and convexity in lieu of compactness. We lift the optimization in (1) from the set of all probability measures to an L^{τ} space for a $\tau > 1$ because the space of probability measure $\mathcal{P}$ does not have the Banach–Saks property.[7] Despite the change in the underlying vector space structure, the new functional $\mathfrak{D}_{\alpha}(\cdot)$ inherits both the convexity and norm lower semicontinuity from the Rényi divergence, for an appropriately chosen τ. Use of the Banach–Saks theorem in conjunction with the (quasi-)convexity and the norm lower semicontinuity of the objective function to prove the existence of its minimizer seems to be a novel approach more generally in the context of information theoretic optimization problems.

[7] One might think of working in L^1 instead of L^{τ} and invoking the Komlós theorem [34, Theorem 1.a], [30, 4.7.24 Theorem]—every norm bounded sequence in L^1 contains a subsequence whose Cesàro mean converges almost everywhere. However, this fact alone does not guarantee the setwise convergence that is crucial to the application of lower semicontinuity of the Rényi divergence in its second argument.

References

[1] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Transactions on Information Theory*, vol. 11, no. 1, pp. 3–18, Jan. 1965.

[2] P. M. Ebert, "Error bounds for parallel communication channels," Technical Report 448, Research Laboratory of Electronics at Massachusetts Institute of Technology, Cambridge, MA, 1966, (http://hdl.handle.net/1721.1/4295).

[3] J. S. Richters, "Communication over fading dispersive channels," Technical Report 464, Research Laboratory of Electronics at Massachusetts Institute of Technology, Cambridge, MA, 1967, (http://hdl.handle.net/1721.1/4279).

[4] U. Augustin, "Error estimates for low rate codes," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 14, no. 1, pp. 61–88, 1969.

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

[6] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Information and Control*, vol. 10, no. 1, pp. 65–103, 1967.

[7] 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.

[8] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels (corresp.)," *IEEE Transactions on Information Theory*, vol. 19, no. 3, pp. 357–359, May 1973.

[9] J. K. Omura, "A lower bounding method for channel and source coding probabilities," *Information and Control*, vol. 27, no. 2, pp. 148 – 177, 1975.

[10] G. Dueck and J. Korner, "Reliability function of a discrete memoryless channel at rates above capacity (corresp.)," *IEEE Transactions on Information Theory*, vol. 25, no. 1, pp. 82–85, Jan 1979.

[11] R. G. Gallager, *Information theory and reliable communication*. New York, NY: John Wiley & Sons, Inc., 1968.

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

[13] I. Csiszár, "The method of types [information theory]," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2505–2523, Oct 1998.

[14] J. Scarlett, A. Martinez, and A. G. i. Fabregas, "Mismatched decoding: Error exponents, second-order rates and saddlepoint approximations," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2647–2666, May 2014.

[15] 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]).

[16] G. S. Poltyrev, "Random coding bounds for discrete memoryless channels," *Problems of Information Transmission*, vol. 18, no. 1, pp. 9–21, 1982.

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

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

[19] 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.

[20] 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]).

[21] 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].

[22] 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]).

[23] ——, "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]).

[24] 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]).

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

[26] 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]).

[27] S. Verdú, "α-mutual information," in *2015 Information Theory and Applications Workshop*, 2015.

[28] 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.

[29] H.-C. Cheng and B. Nakiboğlu, "On the existence of the Augustin mean," *arXiv:2109.00443v1 [cs.IT]*, 2021.

[30] V. I. Bogachev, *Measure Theory*. Berlin Heidelberg: Springer-Verlag, 2007.

[31] H.-C. Cheng, L. Gao, and M. Hsieh, "Properties of scaled noncommutative Rényi and Augustin information," in *2019 IEEE International Symposium on Information Theory (ISIT)*, July 2019, pp. 2219–2223, (arXiv:1811.04218 [quant-ph]).

[32] R. M. Fano, *Transmission of Information: A Statistical Theory of Communications*. Cambridge, MA: M.I.T. Press, 1961.

[33] M. Hayashi and M. Tomamichel, "Correlation detection and an operational interpretation of the rényi mutual information," *Journal of Mathematical Physics*, vol. 57, no. 10, p. 102201, Oct 2016.

[34] J. Komlós, "A generalization of a problem of Steinhaus," *Acta Mathematica Academiae Scientiarum Hungarica*, vol. 18, no. 1-2, pp. 117–229, 1967.