ISSN 0032-9460, Problems of Information Transmission, 2019, Vol. 55, No. 4, pp. 299–342. © Pleiades Publishing, Inc., 2019.
Russian Text © The Author(s), 2019, published in Problemy Peredachi Informatsii, 2019, Vol. 55, No. 4, pp. 3–51.

INFORMATION THEORY

The Augustin Capacity and Center

B. Nakiboğlu

Middle East Technical University, Ankara, Turkey
e-mail: bnakib@metu.edu.tr

Received March 22, 2018; revised August 20, 2019; accepted November 12, 2019

Abstract—For any channel, the existence of a unique Augustin mean is established for any positive order and probability mass function on the input set. The Augustin mean is shown to be the unique fixed point of an operator defined in terms of the order and the input distribution. The Augustin information is shown to be continuously differentiable in the order. For any channel and convex constraint set with finite Augustin capacity, the existence of a unique Augustin center and the associated van Erven–Harremoës bound are established. The Augustin–Legendre (A-L) information, capacity, center, and radius are introduced, and the latter three are proved to be equal to the corresponding Rényi–Gallager quantities. The equality of the A-L capacity to the A-L radius for arbitrary channels and the existence of a unique A-L center for channels with finite A-L capacity are established. For all interior points of the feasible set of cost constraints, the cost constrained Augustin capacity and center are expressed in terms of the A-L capacity and center. Certain shift-invariant families of probabilities and certain Gaussian channels are analyzed as examples.

Key words: Rényi divergence, Rényi information, Augustin information, Augustin mean, Augustin center, Augustin capacity, cost constrained capacity and center, Augustin–Legendre information measures, Rényi–Gallager information measures.

DOI: 10.1134/S003294601904001X

1. INTRODUCTION

The mutual information, which is sometimes called the Shannon information, is a pivotal quantity in the analysis of various information transmission problems. It is defined without referring to an optimization problem, but it satisfies the following two identities given in terms of the Kullback–Leibler divergence:

$$I(p;W) = \inf_{q\in\mathcal{P}(\mathcal{Y})} D(p \circledast W \,\|\, p\otimes q) \tag{1}$$

$$= \inf_{q\in\mathcal{P}(\mathcal{Y})} \sum_x p(x) D(W(x) \,\|\, q), \tag{2}$$

where $\mathcal{P}(\mathcal{Y})$ is the set of all probability measures on the output space $(\mathscr{Y},\mathcal{Y})$, p is a probability mass function that is positive only on a finite subset of the input set $\mathscr{X}$, and W is a function of the form $W\colon \mathscr{X}\to\mathcal{P}(\mathcal{Y})$. Either of the expressions on the right-hand side can be taken as the definition of the mutual information. One can define the order α Rényi information via these expressions by replacing the Kullback–Leibler divergence with the order α Rényi divergence. Since the order one Rényi divergence is the Kullback–Leibler divergence, the order one Rényi information is equal to the mutual information for both definitions. For other orders, however, these two definitions are not equivalent to the definition of the mutual information or to one another, as was pointed out by Csiszár [2]. The generalization associated with the expression in (1) is called the order α Rényi information and denoted by $I_\alpha^{\mathrm{g}}(p;W)$. The generalization associated with the expression in (2) is called the order α Augustin information and denoted by $I_\alpha(p;W)$. Following the convention for the constrained Shannon capacity, the order α Augustin capacity for the constraint set $\mathscr{A}$ is defined as $\sup_{p\in\mathscr{A}} I_\alpha(p;W)$.

For constant composition codes on memoryless classical-quantum channels, the Augustin information for orders less than one arises in the expression for the sphere packing exponent and the Augustin information for orders greater than one arises in the expression for the strong converse exponent, as was recently pointed out by Dalai [3] and by Mosonyi and Ogawa [4], respectively. For the constant composition codes on the discrete stationary product channels, these observations were made implicitly by Csiszár and Körner in [5, p. 172] and by Csiszár in [2]. For the cost constrained codes on (possibly nonstationary) product channels with additive cost functions, the cost constrained Augustin capacity plays an analogous role in the expressions for the sphere packing exponent and the strong converse exponent. The observations about the sphere packing exponent were also reported by Augustin in [6, Remark 36.7(i) and Section 36] for quite general channel models. Therefore, Augustin's information measures do have operational significance, at the very least for the channel coding problem. Our main aim in the current manuscript, however, is to analyze the Augustin information and capacity as measure theoretic concepts. Throughout the manuscript, we will refrain from referring to the channel coding problem or the operational significance of Augustin's information measures, because we believe that the Augustin information and capacity can and should be understood as measure theoretic concepts first. The operational significance of the Augustin information and capacity can be established afterward using information theoretic techniques together with the results of the measure theoretic analysis, as we do in [7].

All of the previous works on the Augustin information or capacity, except Augustin's [6], assume the output set $\mathscr{Y}$ of the channel W to be finite [2, 3, 8–11]. This, however, is a major drawback, because the finite output set assumption is violated by certain analytically interesting models that are also important because of their prominence in engineering applications, such as the Gaussian and Poisson channel models. We pursue our analysis on a more general model and assume[1] the output space $(\mathscr{Y}, \mathcal{Y})$ to be a measurable space composed of an output set $\mathscr{Y}$ and a σ-algebra of its subsets $\mathcal{Y}$. Our analysis of the Augustin information and capacity in this general framework is built around two fundamental concepts: the Augustin mean and the Augustin center.

Recall that the mutual information is defined as $I(p;W) \triangleq \sum_x p(x)D(W(x)\,\|\,q_{1,p})$, where $q_{1,p} = \sum_x p(x)W(x)$. Hence, the infimum in (2) is achieved by $q_{1,p}$. Furthermore, one can confirm by substitution that

$$\sum_x p(x)D(W(x)\,\|\,q) = I(p;W) + D(q_{1,p}\,\|\,q), \quad \forall q \in \mathcal{P}(\mathcal{Y}).$$

Thus, $q_{1,p}$ is the only probability measure achieving the infimum in (2), because the Kullback–Leibler divergence is positive for distinct probability measures. A similar relation holds for other orders, as well: for any α in $\mathbb{R}_+$ there exists a unique probability measure $q_{\alpha,p}$ satisfying $I_\alpha(p;W) = \sum_x p(x)D_\alpha(W(x)\,\|\,q_{\alpha,p})$. We call the probability measure $q_{\alpha,p}$ the *order α Augustin mean*. In [6, Lemma 34.2], Augustin established the existence of a unique $q_{\alpha,p}$ for α's in $(0,1]$ and derived certain important characteristics of $q_{\alpha,p}$, which are corner stones of the analysis of the Augustin information and capacity. We establish analogous relations for orders greater than one in Section 3; see Lemma 13(d).

In [12], Kemperman proved the equality of the (unconstrained) Shannon capacity to the Shannon radius[2] for any channel of the form $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$ and the existence of a unique Shannon center

[1] We have additional hypotheses in Section 5.4, but those assumptions are satisfied by essentially all models of interest, as well.

[2] The Shannon radius is defined as $\inf_{q\in\mathcal{P}(\mathcal{Y})} \sup_{x\in\mathscr{X}} D(W(x)\,\|\,q)$.

for channels with finite Shannon capacity. Using ideas that are already present in Kemperman's proof, one can establish a similar result for the constrained Shannon capacity provided that the constrained set is convex; see [13, Theorem 2]: For any channel W of the form $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$ and convex constraint set $\mathscr{A}$,

$$\sup_{p\in\mathscr{A}} I(p;W) = \inf_{q\in\mathcal{P}(\mathcal{Y})} \sup_{p\in\mathscr{A}} \sum_x p(x) D(W(x)\,\|\,q). \tag{3}$$

Considering (2), one can interpret (3) as a minimax theorem. Furthermore, if the Shannon capacity for the constraint set $\mathscr{A}$ is finite, then there exists a unique probability measure $q_{1,W,\mathscr{A}}$, called the Shannon center for the constraint set $\mathscr{A}$, such that

$$\sup_{p\in\mathscr{A}} I(p;W) = \sup_{p\in\mathscr{A}} \sum_x p(x) D(W(x)\,\|\,q_{1,W,\mathscr{A}}).$$

The name center is reminiscent of the name of the corresponding quantity in the unconstrained case, which is discussed in [12]. Augustin proved an analogous result for $I_\alpha(p;W)$ assuming α to be an order in $(0,1]$ and $\mathscr{A}$ to be a constraint set determined by cost constraints; see [6, Lemma 34.7]. We prove an analogous proposition for $I_\alpha(p;W)$ for any α in $\mathbb{R}_+$ and convex constraint set $\mathscr{A}$ in Section 4; see Theorem 1. We call the corresponding probability measure $q_{\alpha,W,\mathscr{A}}$ the *order α Augustin center for the constraint set $\mathscr{A}$*.

Constraint sets determined by cost constraints are frequently encountered while employing the Augustin capacity to analyze channel coding problems. One can apply the convex conjugation techniques to provide an alternative characterization of the cost constrained Augustin capacity and center. Augustin did so in [6, Section 35], relying on a quantity that was previously employed in discrete channels by Gallager [14, pp. 13–15; 15, Section 7.3] and in various Gaussian channel models[3] [14, pp. 15 and 16; 15, Sections 7.4 and 7.5; 16; 17]. We call this quantity the Rényi–Gallager information and analyze it in Section 5.3. Compared to the application of convex conjugation techniques to the cost constrained Shannon capacity provided by Csiszár and Körner in [5, ch. 8], Augustin's analysis in [6, Section 35], relying on the Rényi–Gallager information, is rather convoluted. In Section 5.2, we adhere to a more standard approach and provide an analysis, which can be seen as a generalization of [5, ch. 8], relying on a new quantity, which we call the Augustin–Legendre information. We show the equivalence of these two approaches using minimax theorems similar to the one described above for the constrained Augustin capacity.

Some of the most important observations presented in this paper have already been derived previously in [6, Sections 33–35; 10; 18; 19]. In order to delineate our main contributions in the context of these works, we provide a tally in Section 1.3. Before doing that, we describe our notational conventions in Section 1.1 and our model in Section 1.2.

1.1. Notational Conventions

The inner product of any two vectors μ and q in $\mathbb{R}^\ell$, i.e., $\sum\limits_{i=1}^{\ell} \mu^i q^i$, is denoted by $\mu\cdot q$. The ℓ-dimensional vector whose all entries are one is denoted by $\mathbb{1}$ for any $\ell\in\mathbb{Z}_+$, the dimension ℓ will be clear from the context. We denote the closure, interior, and convex hull of a set $\mathscr{S}$ by $\operatorname{cl}\mathscr{S}$, $\operatorname{int}\mathscr{S}$, and $\operatorname{ch}\mathscr{S}$, respectively; the relevant topology or vector space structure will be evident from the context.

For any set $\mathscr{Y}$, we denote the set of all subsets of $\mathscr{Y}$ (i.e., the power set of $\mathscr{Y}$) by $2^{\mathscr{Y}}$, the set of all probability measures on finite subsets of $\mathscr{Y}$ by $\mathscr{P}(\mathscr{Y})$, and the set of all nonzero finite measures

[3] Augustin assumed neither a specific noise model nor the finiteness of the output set. Nevertheless, Gaussian channels are not subsumed by Augustin's model in [6, Section 35], because Augustin assumed a bounded cost function.

with the same property by $\mathscr{M}^{+}(\mathscr{Y})$. For any p in $\mathscr{M}^{+}(\mathscr{Y})$, we call the set of all y's satisfying $p(y)>0$ the support of p and denote it by $\operatorname{supp} p$.

On a measurable space $(\mathscr{Y},\mathcal{Y})$, we denote the set of all finite signed measures by $\mathcal{M}(\mathcal{Y})$, the set of all finite measures by $\mathcal{M}_0^{+}(\mathcal{Y})$, the set of all nonzero finite measures by $\mathcal{M}^{+}(\mathcal{Y})$, and the set of all probability measures by $\mathcal{P}(\mathcal{Y})$. Let μ and q be two measures on the measurable space $(\mathscr{Y},\mathcal{Y})$. Then μ is absolutely continuous with respect to q, i.e., $\mu \prec q$, whenever $\mu(\mathscr{E})=0$ for any $\mathscr{E}\in\mathcal{Y}$ such that $q(\mathscr{E})=0$; μ and q are equivalent, i.e., $\mu \sim q$, whenever $\mu \prec q$ and $q \prec \mu$; μ and q are singular, i.e., $\mu \perp q$, whenever $\exists \mathscr{E}\in\mathcal{Y}$ such that $\mu(\mathscr{E})=q(\mathscr{Y}\setminus\mathscr{E})=0$. Furthermore, a set of measures $\mathscr{W}$ on $(\mathscr{Y},\mathcal{Y})$ is absolutely continuous with respect to q, i.e., $\mathscr{W}\prec q$, whenever $w \prec q$ for all $w\in\mathscr{W}$ and uniformly absolutely continuous with respect to q, i.e., $\mathscr{W}\prec^{\text{uni}} q$, whenever for every $\varepsilon>0$ there exists a $\delta>0$ such that $w(\mathscr{E})<\varepsilon$ for all $w\in\mathscr{W}$ provided that $q(\mathscr{E})<\delta$.

We denote the integral of a measurable function f with respect to the measure μ by $\int f\mu(dy)$ or $\int f(y)\mu(dy)$. If the integral is on the real line and if it is with respect to the Lebesgue measure, we denote it by $\int f dy$ or $\int f(y)dy$, as well. If μ is a probability measure, then we also call the integral of f with respect to μ the expectation of f or the expected value of f and denote it by $\mathbf{E}_{\mu}[f]$ or $\mathbf{E}_{\mu}[f(y)]$.

Our notation will be overloaded for certain symbols; however, the relations represented by these symbols will be clear from the context. We use $\hbar(\cdot)$ to denote both the Shannon entropy and the binary entropy: $\hbar(p)\triangleq\sum_y p(y)\ln\frac{1}{p(y)}$ for all $p\in\mathscr{P}(\mathscr{Y})$ and $\hbar(z)\triangleq z\ln\frac{1}{z}+(1-z)\ln\frac{1}{1-z}$ for all $z\in[0,1]$. We denote the product of topologies [20, p. 38], σ-algebras [20, p. 118], and measures [20, Theorem 4.4.4] by $\otimes$. We denote the Cartesian product of sets [20, p. 38] by $\times$. We use the shorthand notation $\mathscr{X}_1^n$ for the Cartesian product of sets $\mathscr{X}_1,\ldots,\mathscr{X}_n$ and $\mathcal{Y}_1^n$ for the product of the σ-algebras $\mathcal{Y}_1,\ldots,\mathcal{Y}_n$. We use $|\cdot|$ to denote the absolute value of real numbers and the size of sets. The sign $\leq$ stands for the usual less than or equal to relation for real numbers and the corresponding pointwise inequality for functions and vectors. For two measures μ and q on the measurable space $(\mathscr{Y},\mathcal{Y})$, $\mu\leq q$ whenever $\mu(\mathscr{E})\leq q(\mathscr{E})$ for all $\mathscr{E}\in\mathcal{Y}$.

For $a,b\in\mathbb{R}$, $a\wedge b$ is the minimum of a and b. For $f\colon\mathscr{Y}\to\mathbb{R}$ and $g\colon\mathscr{Y}\to\mathbb{R}$, the function $f\wedge g$ is the pointwise minimum of f and g. For $\mu,q\in\mathcal{M}(\mathcal{Y})$, $\mu\wedge q$ is the unique measure satisfying $\frac{d\mu\wedge q}{d\nu}=\frac{d\mu}{d\nu}\wedge\frac{dq}{d\nu}$ ν-a.e. for any ν satisfying $\mu\prec\nu$ and $q\prec\nu$. For a collection $\mathscr{F}$ of real valued functions $\bigwedge_{f\in\mathscr{F}} f$ is the pointwise infimum of f's in $\mathscr{F}$, which is an extended real valued function. For a collection of measures $\mathscr{U}\subset\mathcal{M}(\mathcal{Y})$ satisfying $w\leq u$ for all $u\in\mathscr{U}$ for some $w\in\mathcal{P}(\mathcal{Y})$, $\bigwedge_{u\in\mathscr{U}} u$ is the infimum of $\mathscr{U}$ with respect to the partial order $\leq$. There exists a unique infimum measure by [21, Theorem 4.7.5]. We use the symbol $\vee$ analogously to $\wedge$ but we represent maxima and suprema with it, rather than minima and infima.

1.2. Channel Model

A *channel* W is a function from the *input set* $\mathscr{X}$ to the set of all probability measures on the *output space* $(\mathscr{Y},\mathcal{Y})$:

$$W\colon\mathscr{X}\to\mathcal{P}(\mathcal{Y}). \tag{4}$$

$\mathscr{Y}$ is called the *output set*, and $\mathcal{Y}$ is called the *σ-algebra of the output events*. We denote the set of all channels from the input set $\mathscr{X}$ to the output space $(\mathscr{Y},\mathcal{Y})$ by $\mathcal{P}(\mathcal{Y}\,|\,\mathscr{X})$. For any $p\in\mathscr{P}(\mathscr{X})$ and $W\in\mathcal{P}(\mathcal{Y}\,|\,\mathscr{X})$, the probability measure whose marginal on $\mathscr{X}$ is p and whose conditional distribution given x is $W(x)$ is denoted by $p\circledast W$. Until Section 5.4, we confine our discussion to the input distributions in $\mathscr{P}(\mathscr{X})$ and avoid the subtleties related to measurability. The more general case of input distributions in $\mathcal{P}(\mathcal{X})$ is considered[4] in Section 5.4.

A channel W is called a *discrete channel* if both $\mathscr{X}$ and $\mathcal{Y}$ are finite sets. For any $n \in \mathbb{Z}_+$ and channels $W_t\colon \mathscr{X}_t \to \mathcal{P}(\mathcal{Y}_t)$ for $t \in \{1, \ldots, n\}$, the *length n product channel* $W_{[1,n]}\colon \mathscr{X}_1^n \to \mathcal{P}(\mathcal{Y}_1^n)$ is defined via the following relation:

$$W_{[1,n]}(x_1^n) = \bigotimes_{t=1}^{n} W_t(x_t), \quad \forall x_1^n \in \mathscr{X}_1^n.$$

A product channel is *stationary* whenever $W_t = W$ for all $t \in \{1, \ldots, n\}$ for some $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$.

For any $\ell \in \mathbb{Z}_+$, an ℓ-dimensional *cost function* ρ is a function from the input set to $\mathbb{R}^\ell$ that is bounded from below, i.e., that is of the form $\rho\colon \mathscr{X} \to \mathbb{R}^\ell_{\ge z}$ for some $z \in \mathbb{R}$. We assume without loss of generality that[5]

$$\inf_{x \in \mathscr{X}} \rho^i(x) \ge 0, \quad \forall i \in \{1, \ldots, \ell\}.$$

We denote the set of all cost constraints that can be satisfied by some member of $\mathscr{X}$ by Γ_ρ^{ex} and the set of all cost constraints that can be satisfied by some member of $\mathscr{P}(\mathscr{X})$ by Γ_ρ:

$$\Gamma_\rho^{\text{ex}} \triangleq \{\varrho \in \mathbb{R}^\ell_{\ge 0} : \exists x \in \mathscr{X} \text{ such that } \rho(x) \le \varrho\} \tag{5}$$

$$\Gamma_\rho \triangleq \{\varrho \in \mathbb{R}^\ell_{\ge 0} : \exists p \in \mathscr{P}(\mathscr{X}) \text{ such that } \sum_x p(x)\rho(x) \le \varrho\}. \tag{6}$$

Then both Γ_ρ^{ex} and Γ_ρ have nonempty interiors and Γ_ρ is the convex hull of Γ_ρ^{ex}, i.e., $\Gamma_\rho = \operatorname{ch} \Gamma_\rho^{\text{ex}}$.

A cost function on a product channel is said to be additive whenever it can be written as the sum of cost functions defined on the component channels. Given $W_t\colon \mathscr{X}_t \to \mathcal{P}(\mathcal{Y}_t)$ and $\rho_t\colon \mathscr{X}_t \to \mathbb{R}^\ell_{\ge 0}$ for $t \in \{1, \ldots, n\}$, we denote the resulting additive cost function on $\mathscr{X}_1^n$ for the channel $W_{[1,n]}$ by $\rho_{[1,n]}$, i.e.,

$$\rho_{[1,n]}(x_1^n) = \sum_{t=1}^{n} \rho_t(x_t), \quad \forall x_1^n \in \mathscr{X}_1^n.$$

1.3. Previous Work and Main Contributions

The following is a list of our contributions that are important for a thorough understanding of the Augustin information measures and related results that have been reported before.

I. For all α in $(0, 1)$, [6, Lemma 34.2] of Augustin asserts the existence of a unique probability measure $q_{\alpha,p}$ satisfying $I_\alpha(p; W) = D_\alpha(W \,\|\, q_{\alpha,p} \,|\, p)$ and characterizes $q_{\alpha,p}$ in terms of the operator[6] $\mathrm{T}_{\alpha,p}(\cdot)$ as follows:

- $\mathrm{T}_{\alpha,p}(q_{\alpha,p}) = q_{\alpha,p}$ and $q_{\alpha,p} \sim q_{1,p}$.
- If $q_{1,p} \prec q$ and $\mathrm{T}_{\alpha,p}(q) = q$, then $q_{\alpha,p} = q$.
- $\lim_{j \to \infty} \|q_{\alpha,p} - \mathrm{T}^j_{\alpha,p}(q_{1,p})\| = 0$.

[4] The structure described in (4) is not sufficient on its own to ensure the existence of a unique $p \circledast W$ with the desired properties for all p in $\mathcal{P}(\mathcal{X})$. The existence of such a unique $p \circledast W$ is guaranteed for all p in $\mathcal{P}(\mathcal{X})$ if W is a transition probability from $(\mathscr{X}, \mathcal{X})$ to $(\mathscr{Y}, \mathcal{Y})$, i.e., a member of $\mathcal{P}(\mathcal{Y} \,|\, \mathcal{X})$ rather than $\mathcal{P}(\mathcal{Y} \,|\, \mathscr{X})$.

[5] Augustin [6, Section 33] has an additional hypothesis, $\bigvee_{x \in \mathscr{X}} \rho(x) \le \mathbb{1}$. This hypothesis, however, excludes certain important cases, such as the Gaussian channels.

[6] The operator $\mathrm{T}_{\alpha,p}(\cdot)$, defined in (28), is determined uniquely by α and p and well-defined for all q with finite $D_\alpha(W \,\|\, q \,|\, p)$.

- $D_\alpha(W \,\|\, q \,|\, p) \geq I_\alpha(p; W) + D_\alpha(q_{\alpha,p} \,\|\, q)$ for[7] all $q \in \mathcal{P}(\mathcal{Y})$.

We cannot verify the correctness of the proof of [6, Lemma 34.2]; we discuss our reservations in [22, Appendix C]. Lemma 13(c) is proved[8] relying on the ideas employed in Augustin's proof of [6, Lemma 34.2]. Lemma 13(c) implies all assertions of [6, Lemma 34.2] except for $\lim_{j\to\infty} \|q_{\alpha,p} - \mathrm{T}_{\alpha,p}^{j}(q_{1,p})\| = 0$; Lemma 13(c) establishes $\lim_{j\to\infty} \|q_{\alpha,p} - \mathrm{T}_{\alpha,p}^{j}(q_{\alpha,p}^{\mathrm{g}})\| = 0$ instead—see (37) and [22, Remark 6]. Unlike [6, Lemma 34.2], Lemma 13(c) also bounds $D_\alpha(W \,\|\, q \,|\, p)$ from above. This bound is new to the best of our knowledge. The following inequality summarizes the upper and lower bounds on $D_\alpha(W \,\|\, q \,|\, p)$ established in Lemma 13(c),(d):

$$D_{1\vee\alpha}(q_{\alpha,p} \,\|\, q) \geq D_\alpha(W \,\|\, q \,|\, p) - I_\alpha(p; W) \geq D_{1\wedge\alpha}(q_{\alpha,p} \,\|\, q), \quad \forall q \in \mathcal{P}(\mathcal{Y}). \tag{7}$$

For the finite $\mathscr{Y}$ case, the existence of a q in $\mathscr{P}(\mathscr{Y})$ satisfying both $q \sim q_{1,p}$ and $\mathrm{T}_{\alpha,p}(q) = q$ has been discussed by other authors. We make a brief digression to point out the discussion of the aforementioned existence result in these works.

- While deriving the sphere packing bound for the constant composition codes on discrete stationary product channels, Fano implicitly asserts the existence of a fixed point that is equivalent to $q_{1,p}$ for each α in $(0, 1)$; see [23, Section 9.2, equation (9.24) and p. 292]. Fano, however, does not explain why such a fixed point must exist and does not elaborate on its uniqueness or on its relation to $q_{\alpha,p}$ in [23, Section 9.2].
- While establishing the equivalence of his expression for the sphere packing exponent in the finite $\mathscr{Y}$ case to the one provided by Fano in [23], Haroutunian proved the existence of a fixed point that is equivalent to $q_{1,p}$ for each α in $(0, 1)$; see [18, equations (16)–(19)].
- While discussing the random coding bounds for discrete stationary product channels, Poltyrev makes an observation that is equivalent to asserting for each α in $[1/2, 1)$ the existence of a fixed point that is equivalent to $q_{1,p}$; see [19, equations (3.15) and (3.16) and Theorem 3.2]. Poltyrev, however, does not formulate his observations as a fixed point property.

In our understanding, the main conceptual contribution of [6, Lemma 34.2] is the characterization of the Augustin mean as a fixed point of $\mathrm{T}_{\alpha,p}(\cdot)$ that is equivalent to $q_{1,p}$. Bounds such as the one given in (7) follow from this observation via Jensen's inequality.

II. For $\alpha \in (1, \infty)$, Lemma 13(d) establishes the existence of a unique Augustin mean $q_{\alpha,p}$ and proves that it satisfies (7) as well as the following two assertions:

- $\mathrm{T}_{\alpha,p}(q_{\alpha,p}) = q_{\alpha,p}$ and $q_{\alpha,p} \sim q_{1,p}$.
- If $\mathrm{T}_{\alpha,p}(q) = q$, then $q_{\alpha,p} = q$.

Lemma 13(d) is new to the best of our knowledge. For the $\alpha \in (1, \infty)$ case, neither the characterization of $q_{\alpha,p}$ in terms of $\mathrm{T}_{\alpha,p}(\cdot)$, nor the inequalities given in (7) have been reported before, even for the finite $\mathscr{Y}$ case.

III. $I_\alpha(p; W)$ is a continuously differentiable function of α from $\mathbb{R}_+$ to $[0, \hbar(p)]$ by Lemma 17(e).

IV. The following minimax identity is established in Theorem 1 for any convex constraint set $\mathscr{A}$:

$$\sup_{p \in \mathscr{A}} \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(W \,\|\, q \,|\, p) = \inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{p \in \mathscr{A}} D_\alpha(W \,\|\, q \,|\, p).$$

Theorem 1 establishes the existence of a unique Augustin center, $q_{\alpha,W,\mathscr{A}}$, for any convex $\mathscr{A}$ with finite Augustin capacity and the convergence of $\{q_{\alpha,p^{(i)}}\}_{i\in\mathbb{Z}_+}$ to $q_{\alpha,W,\mathscr{A}}$ in total variation topology

[7] To be precise, [6, Lemma 34.2] asserts the inequality $D_\alpha(W \,\|\, q \,|\, p) \geq I_\alpha(p; W) + \frac{\alpha}{2}\|q_{\alpha,p} - q\|^2$ rather than the one given above. But Augustin proves the inequality given above first and then uses Pinsker's inequality to establish the one given in [6, Lemma 34.2].

[8] One can prove Lemma 13(c) using the ideas employed in the proof of Lemma 13(d), as well.

for any $\{p^{(i)}\}_{i\in\mathbb{Z}_+} \subset \mathscr{A}$ satisfying $\lim_{i\to\infty} I_\alpha(p^{(i)};W) = C_{\alpha,W,\mathscr{A}}$. Augustin proved this result only for α's in $(0,1]$ and the constraint sets determined by cost constraints; see [6, Lemma 34.7]. For the $\mathscr{A} = \mathscr{P}(\mathscr{X})$ case, similar results were proved by Csiszár [2, Proposition 1] assuming both $\mathscr{X}$ and $\mathscr{Y}$ to be finite sets and by van Erven and Harremoës [8, Theorem 34] assuming $\mathscr{Y}$ to be a finite set.

V. The following bound in terms of the Augustin capacity and center established in Lemma 21 is new to the best of our knowledge:

$$\sup_{p\in\mathscr{A}} D_\alpha(W\,\|\,q\,|\,p) \geq C_{\alpha,W,\mathscr{A}} + D_{\alpha\wedge 1}(q_{\alpha,W,\mathscr{A}}\,\|\,q), \quad \forall q \in \mathcal{P}(\mathcal{Y}).$$

A similar bound has been conjectured by van Erven and Harremoës in [8]. For the Rényi capacity and center, we have proved that conjecture and extended it to the constrained case elsewhere; see [13, Lemmas 19 and 25].

VI. The Augustin–Legendre information $I_\alpha^\lambda(p;W)$, defined as $I_\alpha(p;W) - \lambda\cdot\mathbf{E}_p[\rho]$, as well as the resulting capacity, center, and radius are new concepts that have not been studied before, except for the $\alpha = 1$ case. Thus, formally speaking, all of the propositions in Section 5.2 are new. The analysis presented in Section 5.2 is a standard application of the convex conjugation techniques to characterize the cost constrained Augustin capacity and center. A similar analysis for the $\alpha = 1$ case is provided by Csiszár and Körner in [5, ch. 8] for discrete channels with a single cost constraint. The most important conclusions of the analysis presented in Section 5.2 are the following:

- $C_{\alpha,W}^\lambda$, defined as $\sup_{p\in\mathscr{P}(\mathscr{X})} I_\alpha^\lambda(p;W)$, satisfies $C_{\alpha,W}^\lambda = \sup_{\varrho\geq 0} C_{\alpha,W,\varrho} - \lambda\cdot\varrho$ for all $\lambda \in \mathbb{R}_{\geq 0}^{\ell}$ by (76).
- $C_{\alpha,W,\varrho} = \inf_{\lambda\geq 0} C_{\alpha,W}^\lambda + \lambda\cdot\varrho$ for all $\varrho \in \operatorname{int}\Gamma_\rho$ and the set of λ's achieving this infimum form a nonempty convex compact set whenever $C_{\alpha,W,\varrho}$ is finite by Lemma 29.
- $C_{\alpha,W}^\lambda = S_{\alpha,W}^\lambda$ where $S_{\alpha,W}^\lambda$ is defined as $\inf_{q\in\mathcal{P}(\mathcal{Y})} \sup_{x\in\mathscr{X}} D_\alpha(W(x)\,\|\,q) - \lambda\cdot\rho(x)$ by Theorem 2.
- If $C_{\alpha,W}^\lambda < \infty$, then there exists a unique A-L center $q_{\alpha,W}^\lambda$ satisfying

  $$C_{\alpha,W}^\lambda = \sup_{x\in\mathscr{X}} D_\alpha(W(x)\,\|\,q_{\alpha,W}^\lambda) - \lambda\cdot\rho(x).$$

  Furthermore, $\lim_{i\to\infty} \|q_{\alpha,p} - q_{\alpha,W}^\lambda\| = 0$ for all $\{p^{(i)}\}_{i\in\mathbb{Z}_+} \subset \mathscr{P}(\mathscr{X})$ satisfying $\lim_{i\to\infty} I_\alpha^\lambda(p^{(i)};W) = C_{\alpha,W}^\lambda$ by Theorem 2.
- If $C_{\alpha,W,\varrho} = C_{\alpha,W}^\lambda + \lambda\cdot\varrho < \infty$ for a $\lambda \in \mathbb{R}_{\geq 0}^{\ell}$, then $q_{\alpha,W,\varrho} = q_{\alpha,W}^\lambda$ by Lemma 31.
- If $W_{[1,n]}$ is a product channel with an additive cost function, then $C_{\alpha,W_{[1,n]}}^\lambda = \sum_{t=1}^{n} C_{\alpha,W_t}^\lambda$ for all $\lambda \in \mathbb{R}_{\geq 0}^{\ell}$ and $\alpha \in \mathbb{R}_+$, and whenever either of them exists, $q_{\alpha,W_{[1,n]}}^\lambda$ is equal to $\bigotimes_{t=1}^{n} q_{\alpha,W_t}^\lambda$ by Lemma 32.

VII. The Rényi–Gallager information $I_\alpha^{\mathrm{g}\lambda}(p;W)$ is a generalization of the Rényi information $I_\alpha^{\mathrm{g}}(p;W)$ with a Lagrange multiplier, because $I_\alpha^{\mathrm{g}0}(p;W) = I_\alpha^{\mathrm{g}}(p;W)$. This quantity was first employed by Gallager in [14] by a different parametrization and scaling; later considered in [24, Section IV; 6; 16; 17; 25–27] with various parametrizations, scalings, and names. We chose the scaling and the parametrization so as to be compatible with the ones for Augustin–Legendre information. The most important conclusions of our analysis in Section 5.3 are the following:

- $C_{\alpha,W}^{\mathrm{g}\lambda} = S_{\alpha,W}^\lambda$ by Theorem 3, where $C_{\alpha,W}^{\mathrm{g}\lambda}$ is defined as $\sup_{p\in\mathscr{P}(\mathscr{X})} I_\alpha^{\mathrm{g}\lambda}(p;W)$.
- If $C_{\alpha,W}^\lambda < \infty$ and $\lim_{i\to\infty} I_\alpha^{\mathrm{g}}(p^{(i)};W) = C_{\alpha,W}^\lambda$, then $\lim_{i\to\infty} \|q_{\alpha,p}^{\mathrm{g}\lambda} - q_{\alpha,W}^\lambda\| = 0$ by Theorem 3.
- $\sup_{x\in\mathscr{X}} D_\alpha(W(x)\,\|\,q) - \lambda\cdot\rho(x) \geq C_{\alpha,W}^\lambda + D_\alpha(q_{\alpha,W}^\lambda\,\|\,q)$ for all $q \in \mathcal{P}(\mathcal{Y})$ by Lemma 35.

Lemma 35 is new to the best of our knowledge. For the case where both $\alpha \in (0,1)$ and $\bigvee_{x\in\mathscr{X}} \rho(x) \leq \mathbb{1}$, Theorem 3 is implied by [6, Lemma 35.2].

While pursuing a similar analysis in [6, Section 35], Augustin assumed the cost function to be bounded. This hypothesis, however, excludes certain important and interesting cases such as the Gaussian channels. The issue here is not a matter of rescaling: certain conclusions of Augustin's analysis, e.g., [6, Lemma 35.3(a)], are not correct when the cost function is unbounded. We do not assume the cost function to be bounded. Thus, our model subsumes not only Augustin's model in [6, Section 35] but also other previously considered models, which were either discrete [14, pp. 13–15; 15, Section 7.3; 24, Section IV; 26; 27] or Gaussian [14, pp. 15 and 16; 15, Sections 7.4 and 7.5; 16; 17; 25].

VIII. For channels with uncountable input sets, the Shannon information and capacity is often defined via the probability measures on the input space $(\mathscr{X},\mathcal{X})$ rather than the probability mass functions on the input set $\mathscr{X}$. In Section 5.4, we discuss how and under which conditions one can make such a generalization for Augustin's information measures. The most important conclusions of our analysis are the following:

- If W is a transition probability $(\mathscr{X},\mathcal{X})$ to $(\mathscr{Y},\mathcal{Y})$—i.e., $W \in \mathcal{P}(\mathcal{Y}\,|\,\mathcal{X})$—and $\mathcal{Y}$ is countably generated, then
  - $I_\alpha(p;W)$ is well defined for all $\alpha \in \mathbb{R}_+$ and $p \in \mathcal{P}(\mathcal{X})$ by (112), (113), and Lemma 37.
  - $I_\alpha^\lambda(p;W)$ is well defined for all $\alpha \in \mathbb{R}_+$, $p \in \mathcal{P}(\mathcal{X})$, and $\lambda \in \mathbb{R}_{\geq 0}^\ell$ by (114) provided that ρ is $\mathcal{X}$-measurable.
- If $W \in \mathcal{P}(\mathcal{Y}\,|\,\mathcal{X})$, $\mathcal{X}$ is countably separated, $\mathcal{Y}$ is countably generated, and ρ is $\mathcal{X}$-measurable, then
  - $C_{\alpha,W}^\lambda = \sup_{p\in\mathcal{A}^\lambda} I_\alpha^\lambda(p;W)$ for all λ in $\mathbb{R}_{\geq 0}^\ell$ by Theorem 4 where $\mathcal{A}^\lambda$ is defined as $\{p \in \mathcal{P}(\mathcal{X}) : \lambda \cdot \mathbf{E}_p[\rho] < \infty\}$.
  - If $C_{\alpha,W}^\lambda < \infty$ for a λ in $\mathbb{R}_{\geq 0}^\ell$, then $C_{\alpha,W}^\lambda = \sup_{p\in\mathcal{A}^\lambda} D_\alpha(W\,\|\,q_{\alpha,W}^\lambda\,|\,p) - \lambda \cdot \mathbf{E}_p[\rho]$ by Theorem 4.
  - $C_{\alpha,W,\varrho} = \sup_{p\in\mathcal{A}(\varrho)} I_\alpha(p;W)$ for all ϱ in $\operatorname{int}\Gamma_\rho$ by Theorem 5 where $\mathcal{A}(\varrho)$ is defined as $\{p \in \mathcal{P}(\mathcal{X}) : \mathbf{E}_p[\rho] \leq \varrho\}$.
  - If $C_{\alpha,W,\varrho} < \infty$ for a ϱ in $\operatorname{int}\Gamma_\rho$, then $C_{\alpha,W,\varrho} = \sup_{p\in\mathcal{A}(\varrho)} D_\alpha(W\,\|\,q_{\alpha,W,\varrho}\,|\,p)$ by Theorem 5.

  Thus, the A-L capacity and center as well as the cost constrained Augustin capacity and center defined via probability mass functions are equal to the corresponding quantities that might be defined via probability measures on $(\mathscr{X},\mathcal{X})$, provided that $\mathcal{X}$ is countably separated and $\mathcal{Y}$ is countably generated.

2. PRELIMINARIES

2.1. The Rényi Divergence

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

$$D_\alpha(w\,\|\,q) \triangleq \begin{cases} \dfrac{1}{\alpha-1} \ln \displaystyle\int \left(\frac{dw}{d\nu}\right)^\alpha \left(\frac{dq}{d\nu}\right)^{1-\alpha} \nu(dy) & \text{for } \alpha \neq 1, \\ \displaystyle\int \frac{dw}{d\nu}\left[\ln\frac{dw}{d\nu} - \ln\frac{dq}{d\nu}\right]\nu(dy) & \text{for } \alpha = 1, \end{cases} \tag{8}$$

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

Customarily, the Rényi divergence is defined for pairs of probability measures—see [8, 28] for example—rather than pairs of nonzero finite measures. We adopt this slightly more general definition, because it allows us to use the Rényi divergence to express certain observations more succinctly; see Lemma 1 in the following and Section 5.3. For pairs of probability measures Definition 1 is equivalent to usual definition employed in [8] by [8, Theorem 5].

Lemma 1 [13, Lemma 8]. *Let α be a positive real number and w, q, v be nonzero finite measures on $(\mathscr{Y}, \mathcal{Y})$.*

- *If $v \leq q$, then $D_\alpha(w \,\|\, q) \leq D_\alpha(w \,\|\, v)$.*
- *If $q = \gamma v$ for some $\gamma \in \mathbb{R}_+$ and either w is a probability measure or $\alpha \neq 1$, then $D_\alpha(w \,\|\, q) = D_\alpha(w \,\|\, v) - \ln \gamma$.*

If both arguments of the Rényi divergence are probability measures, then it is positive unless the arguments are equal to one another by Lemma 2.

Lemma 2 [8, Theorems 3 and 31]. *For any $\alpha \in \mathbb{R}_+$ and probability measures w and q on $(\mathscr{Y}, \mathcal{Y})$,*

$$D_\alpha(w \,\|\, q) \geq \frac{1 \wedge \alpha}{2} \|w - q\|^2.$$

For orders in $(0, 1]$, this inequality is called Pinsker's inequality [29, 30]. For orders in $(0, 1)$ it is possible to bound the Rényi divergence from above in terms of the total variation distance. For the $\alpha = 1/2$ case, [31, equation (21), p. 364] asserts

$$D_{1/2}(w \,\|\, q) \leq 2 \ln \frac{2}{2 - \|w - q\|}. \tag{9}$$

As a function of its arguments, the order α Rényi divergence is continuous for the total variation topology provided that $\alpha \in (0, 1)$. For arbitrary orders we only have lower semicontinuity, but that holds even for the topology of setwise convergence.

Lemma 3 [8, Theorem 15]. *For any $\alpha \in \mathbb{R}_+$, $D_\alpha(w \,\|\, q)$ is a lower semicontinuous function of the pair of probability measures (w, q) in the topology of setwise convergence.*

Lemma 4 [8, Theorem 17]. *For any $\alpha \in (0, 1)$, $D_\alpha(w \,\|\, q)$ is a uniformly continuous function of the pair of probability measures (w, q) in the total variation topology.*

The Rényi divergence is convex in its second argument for all positive orders, jointly convex in its arguments for positive orders that are not greater than one, and jointly quasi-convex in its arguments for all positive orders.

Lemma 5 [8, Theorem 12]. *For all $\alpha \in \mathbb{R}_+$, $w, q_0, q_1 \in \mathcal{P}(\mathcal{Y})$, $\beta \in (0, 1)$, and ν satisfying $(q_0 + q_1) \prec \nu$,*

$$D_\alpha(w \,\|\, \beta q_1 + (1 - \beta) q_0) \leq \beta D_\alpha(w \,\|\, q_1) + (1 - \beta) D_\alpha(w \,\|\, q_0).$$

Furthermore, the equality holds if and only if $\frac{dq_1}{d\nu} = \frac{dq_0}{d\nu}$ w-almost surely.

Lemma 6 [8, Theorem 11]. *For all $\alpha \in (0, 1]$, $w_0, w_1, q_0, q_1 \in \mathcal{P}(\mathcal{Y})$, $\beta \in (0, 1)$, and ν satisfying $(w_0 + w_1 + q_0 + q_1) \prec \nu$,*

$$D_\alpha(\beta w_1 + (1 - \beta) w_0 \,\|\, \beta q_1 + (1 - \beta) q_0) \leq \beta D_\alpha(w_1 \,\|\, q_1) + (1 - \beta) D_\alpha(w_0 \,\|\, q_0). \tag{10}$$

Furthermore, for $\alpha = 1$ the equality holds if and only if $\frac{dw_0}{d\nu} \frac{dq_1}{d\nu} = \frac{dw_1}{d\nu} \frac{dq_0}{d\nu}$, and for $\alpha \in (0, 1)$ the equality holds if and only if $\frac{dw_0}{d\nu} \frac{dq_1}{d\nu} = \frac{dw_1}{d\nu} \frac{dq_0}{d\nu}$ and $D_\alpha(w_1 \,\|\, q_1) = D_\alpha(w_0 \,\|\, q_0)$.

Lemma 7 [8, Theorem 13]. *For all $\alpha \in \mathbb{R}_+$, $w_0, w_1, q_0, q_1 \in \mathcal{P}(\mathcal{Y})$, and $\beta \in (0, 1)$,*

$$D_\alpha(\beta w_1 + (1 - \beta) w_0 \,\|\, \beta q_1 + (1 - \beta) q_0) \leq D_\alpha(w_1 \,\|\, q_1) \vee D_\alpha(w_0 \,\|\, q_0).$$

Lemma 8 [8, Theorems 3 and 7]. *For all $w, q \in \mathcal{P}(\mathcal{Y})$, $D_\alpha(w \,\|\, q)$ is a nondecreasing and lower semicontinuous function of α on $\mathbb{R}_+$ that is continuous on $(0, (1 \vee \chi_{w,q})]$ where $\chi_{w,q} \triangleq \sup\{\alpha : D_\alpha(w \,\|\, q) < \infty\}$.*

Since $D_\alpha(w \,\|\, q) = \frac{\alpha}{1-\alpha} D_{1-\alpha}(q \,\|\, w)$ for all $\alpha \in (0,1)$, Lemma 8 and (9) imply

$$D_\alpha(w \,\|\, q) \le \begin{cases} D_{1/2}(w \,\|\, q) & \text{if } \alpha \in (0, 1/2], \\ \frac{\alpha}{1-\alpha} D_{1/2}(w \,\|\, q) & \text{if } \alpha \in (1/2, 1) \end{cases}$$
$$\le \frac{2}{1-\alpha} \ln \frac{2}{2 - \|w - q\|}, \quad \forall \alpha \in (0,1). \tag{11}$$

For a slightly tighter bound, see [31, equation (24), p. 365].

If $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{Y}$, then for any w and q in $\mathcal{P}(\mathcal{Y})$ the identities $w|_{\mathcal{G}}(\mathscr{E}) = w(\mathscr{E})$ for all $\mathscr{E} \in \mathcal{G}$ and $q|_{\mathcal{G}}(\mathscr{E}) = q(\mathscr{E})$ for all $\mathscr{E} \in \mathcal{G}$ uniquely define probability measures $w|_{\mathcal{G}}$ and $q|_{\mathcal{G}}$ on $(\mathscr{Y}, \mathcal{G})$. We denote $D_\alpha(w|_{\mathcal{G}} \,\|\, q|_{\mathcal{G}})$ by $D_\alpha^{\mathcal{G}}(w \,\|\, q)$.

Lemma 9 [8, Theorem 21]. *Let $\mathcal{Y}_1 \subset \mathcal{Y}_2 \subset \ldots \subset \mathcal{Y}$ be an increasing family of σ-algebras, and let $\mathcal{Y}_\infty = \sigma\Big(\bigcup_{i=1}^{\infty} \mathcal{Y}_i\Big)$ be the smallest σ-algebra containing them. Then for any order $\alpha \in \mathbb{R}_+$,*

$$\lim_{i \to \infty} D_\alpha^{\mathcal{Y}_i}(w \,\|\, q) = D_\alpha^{\mathcal{Y}_\infty}(w \,\|\, q).$$

2.2. Tilted Probability Measure

Definition 2. For any $\alpha \in \mathbb{R}_+$ and $w, q \in \mathcal{P}(\mathcal{Y})$ satisfying $D_\alpha(w \,\|\, q) < \infty$, the *order α tilted probability measure w_α^q* is

$$\frac{dw_\alpha^q}{d\nu} \triangleq e^{(1-\alpha)D_\alpha(w \| q)} \left(\frac{dw}{d\nu}\right)^\alpha \left(\frac{dq}{d\nu}\right)^{1-\alpha}.$$

Note that $w_1^q = w$ for any q satisfying $D_1(w \,\|\, q) < \infty$. For other orders one can confirm the following identity by substitution: if $D_\alpha(w \,\|\, q) < \infty$, then for any $v \in \mathcal{P}(\mathcal{Y})$ satisfying both $D_1(v \,\|\, w) < \infty$ and $D_1(v \,\|\, q) < \infty$ also satisfies

$$\frac{1}{1-\alpha} D_1(v \,\|\, w_\alpha^q) + D_\alpha(w \,\|\, q) = \frac{\alpha}{1-\alpha} D_1(v \,\|\, w) + D_1(v \,\|\, q).$$

This identity is used to derive the following variational characterization of the Rényi divergence for orders other than one.

Lemma 10 [8, Theorem 30]. *For any $w, q \in \mathcal{P}(\mathcal{Y})$*

$$D_\alpha(w \,\|\, q) = \begin{cases} \inf_{v \in \mathcal{P}(\mathcal{Y})} \frac{\alpha}{1-\alpha} D_1(v \,\|\, w) + D_1(v \,\|\, q) & \text{for } \alpha \in (0,1), \\ \sup_{v \in \mathcal{P}(\mathcal{Y})} \frac{\alpha}{1-\alpha} D_1(v \,\|\, w) + D_1(v \,\|\, q) & \text{for } \alpha \in (1,\infty), \end{cases}$$

where $\frac{\alpha}{1-\alpha} D_1(v \,\|\, w) + D_1(v \,\|\, q)$ stands for $-\infty$ when $\alpha \in (1,\infty)$ and $D_1(v \,\|\, w) = D_1(v \,\|\, q) = \infty$. Furthermore, if $D_\alpha(w \,\|\, q)$ is finite and either $\alpha \in (0,1)$ or $D_1(w_\alpha^q \,\|\, w) < \infty$, then

$$D_\alpha(w \,\|\, q) = \frac{\alpha}{1-\alpha} D_1(w_\alpha^q \,\|\, w) + D_1(w_\alpha^q \,\|\, q). \tag{12}$$

We have observed in Lemma 8 that $D_\alpha(w \,\|\, q)$ is continuous in α on the closure of the interval where it is finite. Lemma 11, in the following, establishes the analyticity of $D_\alpha(w \,\|\, q)$ in α on the interior of the interval where $D_\alpha(w\,\|\,q)$ is finite. Lemma 11 also establishes the analyticity—and hence the finiteness—of $D_1(w_\alpha^q\,\|\,w)$ and $D_1(w_\alpha^q\,\|\,q)$ on the same interval. This allows us to assert the validity of (12) on the same interval:

$$D_\alpha(w\,\|\,q) = \frac{\alpha}{1-\alpha} D_1(w_\alpha^q\,\|\,w) + D_1(w_\alpha^q\,\|\,q), \quad \forall \alpha \in (0, \chi_{w,q}).$$

Lemma 11. *For any $w, q \in \mathcal{P}(\mathcal{Y})$ satisfying $\chi_{w,q} > 0$, for $\chi_{w,q} \triangleq \sup\{\alpha : D_\alpha(w\,\|\,q) < \infty\}$, $D_\alpha(w\,\|\,q)$, $D_1(w_\alpha^q\,\|\,w)$, and $D_1(w_\alpha^q\,\|\,q)$ are analytic functions of α on $(0, \chi_{w,q})$. Furthermore,*

$$\left.\frac{\partial^\kappa D_\alpha(w\,\|\,q)}{\partial \alpha^\kappa}\right|_{\alpha=\phi} = \begin{cases} \kappa! \sum_{t=0}^{\kappa} \frac{(-1)^{\kappa-t}}{(\phi-1)^{\kappa-t+1}} G_{w,q}^t(\phi) & \text{for } \phi \neq 1, \\ \kappa!\, G_{w,q}^{\kappa+1}(1) & \text{for } \phi = 1, \end{cases} \tag{13}$$

where $G_{w,q}^t(\phi)$ is defined in terms of the set $\mathscr{J}_t$ as follows:

$$\mathscr{J}_t \triangleq \{(j_1, j_2, \ldots, j_t) : j_i \in \mathbb{Z}_{\geq 0}\ \forall i \text{ and } 1j_1 + 2j_2 + \ldots + tj_t = t\}, \tag{14}$$

$$G_{w,q}^t(\phi) \triangleq \begin{cases} (\phi-1) D_\phi(w\,\|\,q) & \text{for } t = 0, \\ \sum_{\mathscr{J}_t} \frac{-(j_1+j_2+\ldots+j_t-1)!}{j_1!\,j_2!\ldots j_t!} \prod_{i=1}^{t} \left(\frac{(-1)}{i!} \mathbf{E}_{w_\phi^q}\left[\left(\ln\frac{dw}{d\nu} - \ln\frac{dq}{d\nu}\right)^i\right]\right)^{j_i} & \text{for } t \in \mathbb{Z}_+. \end{cases} \tag{15}$$

Lemma 11 is new to the best of our knowledge; it is proved in [22, Appendix A] using standard results on the continuity and differentiability of parametric integrals and Faà di Bruno formula for derivatives of compositions of smooth functions.

Note that $\mathscr{J}_1 = \{(1)\}$, $\mathscr{J}_2 = \{(2,0),(0,1)\}$, and $\mathscr{J}_3 = \{(3,0,0),(1,1,0),(0,0,1)\}$. Thus, one can confirm using (15) by substitution that

$$\begin{aligned} G_{w,q}^1(\phi) &= \mathbf{E}_{w_\phi^q}[\xi] \\ G_{w,q}^2(\phi) &= \frac{1}{2}(\mathbf{E}_{w_\phi^q}[\xi^2] - \mathbf{E}_{w_\phi^q}[\xi]^2) \\ &= \frac{1}{2}\mathbf{E}_{w_\phi^q}\left[(\xi - \mathbf{E}_{w_\phi^q}[\xi])^2\right] \\ G_{w,q}^3(\phi) &= \frac{1}{3!}\mathbf{E}_{w_\phi^q}[\xi^3] - \frac{1}{2}\mathbf{E}_{w_\phi^q}[\xi^2]\,\mathbf{E}_{w_\phi^q}[\xi] + \frac{1}{3}\mathbf{E}_{w_\phi^q}[\xi]^3 \\ &= \frac{1}{3!}\mathbf{E}_{w_\phi^q}\left[(\xi - \mathbf{E}_{w_\phi^q}[\xi])^3\right] \end{aligned}$$

where $\xi = \ln\frac{dw}{d\nu} - \ln\frac{dq}{d\nu}$. If we substitute these expressions for $G_{w,q}^1(\phi)$, $G_{w,q}^2(\phi)$, and $G_{w,q}^3(\phi)$ in (13) and use the identity $\xi = \frac{1}{\phi-1}\left(\ln\frac{dw_\phi^q}{dw} + G_{w,q}^0(\phi)\right)$ which holds w_ϕ^q-almost surely for $\phi \in (0, \chi_{w,q})\setminus\{1\}$, we get the following more succinct expressions for the first two derivatives of $D_\alpha(w\,\|\,q)$ with respect to α:

$$\left.\frac{\partial}{\partial\alpha} D_\alpha(w\,\|\,q)\right|_{\alpha=\phi} = \begin{cases} \frac{1}{(\phi-1)^2} D_1(w_\phi^q\,\|\,w) & \text{for } \phi \neq 1, \\ \frac{1}{2}\mathbf{E}_w\left[\left(\ln\frac{dw}{dq} - D_1(w\,\|\,q)\right)^2\right] & \text{for } \phi = 1, \end{cases} \tag{16}$$

$$\left.\frac{\partial^2}{\partial\alpha^2} D_\alpha(w\,\|\,q)\right|_{\alpha=\phi} = \begin{cases} \frac{1}{(\phi-1)^3}\left(\mathbf{E}_{w_\phi^q}\left[\left(\ln\frac{dw_\phi^q}{dw}\right)^2\right] - 2D_1(w_\phi^q\,\|\,w) - \left[D_1(w_\phi^q\,\|\,w)\right]^2\right) & \text{for } \phi \neq 1, \\ \frac{1}{3}\mathbf{E}_w\left[\left(\ln\frac{dw}{dq} - D_1(w\,\|\,q)\right)^3\right] & \text{for } \phi = 1. \end{cases} \tag{17}$$

Analyticity of $D_\alpha(w\,\|\,q)$ on $(0,\chi_{w,q})$ implies that for any $\phi\in(0,\chi_{w,q})$ there exists an open interval containing ϕ on which $D_\alpha(w\,\|\,q)$ is equal to the power series determined by the derivatives of $D_\alpha(w\,\|\,q)$ at $\alpha=\phi$. If we have a finite collection of pairs of probability measures $\{(w_i,q_i)\}_{i\in\mathscr{I}}$, then for any ϕ that is in $(0,\chi_{w_i,q_i})$ for all $i\in\mathscr{I}$ there exists an open interval containing ϕ on which each $D_\alpha(w_i\,\|\,q_i)$ is equal to the power series determined by the derivatives of $D_\alpha(w_i\,\|\,q_i)$ at $\alpha=\phi$. When the collection of pairs of probability measures is infinite, then there might not be an open interval containing ϕ that is contained in all $(0,\chi_{w_i,q_i})$'s. Lemma 12, in the following, asserts the existence of such an interval when $D_\beta(w_i\,\|\,q_i)$ is uniformly bounded for a $\beta>\phi$ for all $i\in\mathscr{I}$. In addition, Lemma 12 asserts uniform approximation error terms, over all $i\in\mathscr{I}$, for the power series on that interval.

Lemma 12. *For any $\gamma,\phi,\beta\in\mathbb{R}_+$ satisfying $\phi\in(0,\beta)$ and $w,q\in\mathcal{P}(\mathcal{Y})$ satisfying $D_\beta(w\,\|\,q)\le\gamma$,*

$$\left|\frac{\partial^\kappa D_\alpha(w\,\|\,q)}{\partial\alpha^\kappa}\bigg|_{\alpha=\phi}\right|\le\begin{cases}\kappa!\,\tau^{\kappa+1}\kappa & \text{for } \phi\neq1,\\ \kappa!\,\tau^{\kappa+1} & \text{for } \phi=1,\end{cases}\tag{18}$$

$$\left|D_\eta(w\,\|\,q)-\sum_{i=0}^{\kappa-1}\frac{(\eta-\phi)^i}{i!}\,\frac{\partial^i D_\alpha(w\,\|\,q)}{\partial\alpha^i}\bigg|_{\alpha=\phi}\right|$$
$$\le\begin{cases}\dfrac{\tau^{\kappa+1}|\eta-\phi|^\kappa}{1-|\eta-\phi|\tau}\left[\kappa-1+\dfrac{1}{1-|\eta-\phi|\tau}\right] & \text{for } \phi\neq1,\\ \dfrac{\tau^{\kappa+1}|\eta-\phi|^\kappa}{1-|\eta-\phi|\tau} & \text{for } \phi=1,\end{cases}\qquad\forall\eta:\ |\eta-\phi|\le\frac{1}{\tau}\tag{19}$$

where

$$\tau\triangleq\begin{cases}\dfrac{1}{|\phi-1|}\vee\left[\dfrac{1+e^{(1\vee\beta)\gamma}}{\phi\wedge(\beta-\phi)}+\gamma\right] & \text{for } \phi\neq1,\\ \dfrac{1+e^{\beta\gamma}}{1\wedge(\beta-1)} & \text{for } \phi=1.\end{cases}\tag{20}$$

Lemma 12 is new to the best of our knowledge; it is proved in [22, Appendix A] using (13) together with the elementary properties of the real analytic functions and power series.

2.3. The Conditional Rényi Divergence and Tilted Channel

The conditional Rényi divergence and the tilted channel allows us to write certain frequently used expressions more succinctly.

Definition 3. For any $\alpha\in\mathbb{R}_+$, $W\colon\mathscr{X}\to\mathcal{P}(\mathcal{Y})$, $Q\colon\mathscr{X}\to\mathcal{P}(\mathcal{Y})$, and $p\in\mathscr{P}(\mathscr{X})$ the *order α conditional Rényi divergence for the input distribution p* is

$$D_\alpha(W\,\|\,Q\,|\,p)\triangleq\sum_{x\in\mathscr{X}}p(x)D_\alpha(W(x)\,\|\,Q(x)).\tag{21}$$

If $\exists q\in\mathcal{P}(\mathcal{Y})$ such that $Q(x)=q$ for all $x\in\mathscr{X}$, then we denote $D_\alpha(W\,\|\,Q\,|\,p)$ by $D_\alpha(W\,\|\,q\,|\,p)$.

Remark 1. In [11, 32], $D_\alpha(W\,\|\,Q\,|\,p)$ stands for $D_\alpha(p\circledast W\,\|\,p\circledast Q)$. For the $\alpha=1$ case, the convention used in [11, 32] is equivalent to ours; for the $\alpha\neq1$ case, however, it is not. If either $\alpha=1$ or $D_\alpha(W(x)\,\|\,Q(x))$ has the same value for all x's with positive $p(x)$, then $D_\alpha(p\circledast W\,\|\,p\circledast Q)=\sum_x p(x)D_\alpha(W(x)\,\|\,Q(x))$; otherwise, $D_\alpha(p\circledast W\,\|\,p\circledast Q)<\sum_x p(x)D_\alpha(W(x)\,\|\,Q(x))$ for $\alpha\in(0,1)$ and $D_\alpha(p\circledast W\,\|\,p\circledast Q)>\sum_x p(x)D_\alpha(W(x)\,\|\,Q(x))$ for $\alpha\in(1,\infty)$. The inequalities follow from Jensen's inequality and the strict concavity of the natural logarithm function.

Definition 4. For any $\alpha \in \mathbb{R}_+$, $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$ and $Q\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$, the *order α tilted channel* W_α^Q is a function from $\{x:\ D_\alpha(W(x)\,\|\,Q(x)) < \infty\}$ to $\mathcal{P}(\mathcal{Y})$ given by

$$\frac{dW_\alpha^Q(x)}{d\nu} \triangleq e^{(1-\alpha)D_\alpha(W(x)\|Q(x))}\left[\frac{dW(x)}{d\nu}\right]^\alpha \left[\frac{dQ(x)}{d\nu}\right]^{1-\alpha}. \tag{22}$$

If $\exists q \in \mathcal{P}(\mathcal{Y})$ such that $Q(x) = q$ for all $x \in \mathscr{X}$, then we denote W_α^Q by W_α^q.

3. THE AUGUSTIN INFORMATION

The main aim of this section is to introduce the concepts of Augustin information and mean. We define the order α Augustin information for the input distribution p and establish the existence of a unique Augustin mean for any input distribution p and positive finite order α in Section 3.1. After that we analyze the Augustin information, first as a function of the input distribution for a given order in Section 3.2 and then as a function of the order for a given input distribution in Section 3.3. We conclude our discussion by comparing the Augustin information with the Rényi information and characterizing each quantity in terms of the other in Section 3.4. Some of the most important observations about the Augustin information and mean were first reported by Augustin in [6, Section 34] for orders not exceeding one. This is why we suggest naming these concepts after him. Proof of the lemmas presented in this section are presented in [22, Appendix B].

3.1. Existence of a Unique Augustin Mean

Definition 5. For any $\alpha \in \mathbb{R}_+$, $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$, and $p \in \mathscr{P}(\mathscr{X})$ the *order α Augustin information for the input distribution* p is

$$I_\alpha(p;W) \triangleq \inf_{q\in\mathcal{P}(\mathcal{Y})} D_\alpha(W\,\|\,q\,|\,p). \tag{23}$$

One can confirm by substitution that

$$D_1(W\,\|\,q\,|\,p) = D_1(W\,\|\,q_{1,p}\,|\,p) + D_1(q_{1,p}\,\|\,q), \quad \forall q \in \mathcal{P}(\mathcal{Y}), \tag{24}$$

where

$$q_{1,p} \triangleq \sum_x p(x)W(x). \tag{25}$$

Then Lemma 2 and (23) imply

$$I_1(p;W) = D_1(W\,\|\,q_{1,p}\,|\,p).$$

Thus, the order one Augustin information has a closed form expression, which is equal to the mutual information. For other orders, however, Augustin information does not have a closed form expression. Nonetheless, Lemma 13, presented in the following, establishes the existence of a unique probability measure $q_{\alpha,p}$ satisfying $I_\alpha(p;W) = D_\alpha(W\,\|\,q_{\alpha,p}\,|\,p)$ for[9] any positive order α and input distribution p. Furthermore, parts (c) and (d) of Lemma 13 present an alternative characterization of $q_{\alpha,p}$ by showing that $q_{\alpha,p}$ is the unique fixed point of the operator $\mathrm{T}_{\alpha,p}(\cdot)$ satisfying $q_{1,p} \prec q_{\alpha,p}$. Lemma 13(e) provides an alternative characterization of the Augustin information for orders other than one.[10]

[9] This is rather easy to prove when $\mathscr{Y}$ is a finite set. The uniqueness of $q_{\alpha,p}$ follows from the strict convexity of the Rényi divergence in its second argument described in Lemma 5. If $\mathscr{Y}$ is finite, then $\mathcal{P}(\mathcal{Y})$ is compact and the existence of $q_{\alpha,p}$ follows from the lower semicontinuity of the Rényi divergence in its second argument—which follows from Lemma 3—and the extreme value theorem for the lower semicontinuous functions [33, ch. 3, Section 12.2]. For channels with arbitrary output spaces, however, $\mathcal{P}(\mathcal{Y})$ is not compact; thus, we cannot invoke the extreme value theorem to establish the existence of $q_{\alpha,p}$.

[10] This alternative characterization is employed to prove the equivalence of two definitions of the sphere packing exponent and the strong converse exponent.

Definition 6. Let α be a positive real number and W be a channel of the form $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$.

- For any $p \in \mathscr{M}^{+}(\mathscr{X})$, the *order α mean measure for the input distribution p* is given by

$$\frac{d\mu_{\alpha,p}}{d\nu} \triangleq \left[\sum_{x} p(x) \left(\frac{dW(x)}{d\nu}\right)^{\alpha}\right]^{\frac{1}{\alpha}} \tag{26}$$

where ν is any measure for which $\left(\sum_{x} p(x)W(x)\right) \prec \nu$.

- For any $p \in \mathscr{P}(\mathscr{X})$, the *order α Rényi mean for the input distribution p* is given by

$$q^{\mathrm{g}}_{\alpha,p} \triangleq \frac{\mu_{\alpha,p}}{\|\mu_{\alpha,p}\|}. \tag{27}$$

- For any $p \in \mathscr{P}(\mathscr{X})$, the *order α Augustin operator for the input distribution p*, $\mathrm{T}_{\alpha,p}(\cdot)\colon \mathscr{Q}_{\alpha,p} \to \mathcal{P}(\mathcal{Y})$, is given by

$$\mathrm{T}_{\alpha,p}(q) \triangleq \sum_{x} p(x) W^{q}_{\alpha}(x), \quad \forall q \in \mathscr{Q}_{\alpha,p}, \tag{28}$$

where $\mathscr{Q}_{\alpha,p} \triangleq \{q \in \mathcal{P}(\mathcal{Y}) : D_{\alpha}(W \,\|\, q \,|\, p) < \infty\}$ and the tilted channel W^{q}_{α} is defined in (22). Furthermore, $\mathrm{T}^{0}_{\alpha,p}(q) = q$ and $\mathrm{T}^{i+1}_{\alpha,p}(q) \triangleq \mathrm{T}_{\alpha,p}(\mathrm{T}^{i}_{\alpha,p}(q))$ for any nonnegative integer i.

Lemma 13. *Let W be a channel of the form $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$ and p be an input distribution in $\mathscr{P}(\mathscr{X})$.*

(a) $I_{\alpha}(p;W) \le D_{\alpha}(W \,\|\, q_{1,p} \,|\, p) \le \hbar(p) < \infty$ *for all* $\alpha \in \mathbb{R}_{+}$ *where* $q_{1,p}$ *is defined in* (25).

(b) $I_{1}(p;W) = D_{1}(W \,\|\, q_{1,p} \,|\, p)$. *Furthermore,*

$$D_{1}(W \,\|\, q \,|\, p) - I_{1}(p;W) = D_{1}(q_{1,p} \,\|\, q), \quad \forall q \in \mathcal{P}(\mathcal{Y}). \tag{29}$$

(c) *If* $\alpha \in (0,1)$, *then* $\exists!\, q_{\alpha,p}$ *such that* $I_{\alpha}(p;W) = D_{\alpha}(W \,\|\, q_{\alpha,p} \,|\, p)$. *Furthermore,*

$$\mathrm{T}_{\alpha,p}(q_{\alpha,p}) = q_{\alpha,p}, \tag{30}$$

$$\lim_{j\to\infty} \|q_{\alpha,p} - \mathrm{T}^{j}_{\alpha,p}(q^{\mathrm{g}}_{\alpha,p})\| = 0, \tag{31}$$

$$D_{1}(q_{\alpha,p} \,\|\, q) \ge D_{\alpha}(W \,\|\, q \,|\, p) - I_{\alpha}(p;W) \ge D_{\alpha}(q_{\alpha,p} \,\|\, q), \quad \forall q \in \mathcal{P}(\mathcal{Y}), \tag{32}$$

and $q_{\alpha,p} \sim q_{1,p}$. *In addition,*[11] *if* $q_{1,p} \prec q$ *and* $\mathrm{T}_{\alpha,p}(q) = q$, *then* $q_{\alpha,p} = q$.

(d) *If* $\alpha \in (1,\infty)$, *then* $\exists!\, q_{\alpha,p}$ *such that* $I_{\alpha}(p;W) = D_{\alpha}(W \,\|\, q_{\alpha,p} \,|\, p)$. *Furthermore,*

$$\mathrm{T}_{\alpha,p}(q_{\alpha,p}) = q_{\alpha,p}, \tag{33}$$

$$D_{\alpha}(q_{\alpha,p} \,\|\, q) \ge D_{\alpha}(W \,\|\, q \,|\, p) - I_{\alpha}(p;W) \ge D_{1}(q_{\alpha,p} \,\|\, q), \quad \forall q \in \mathcal{P}(\mathcal{Y}), \tag{34}$$

and $q_{\alpha,p} \sim q_{1,p}$. *In addition, if* $\mathrm{T}_{\alpha,p}(q) = q$, *then* $q_{\alpha,p} = q$.

(e) *If* $\alpha \in \mathbb{R}_{+} \setminus \{1\}$, *then*

$$I_{\alpha}(p;W) = \frac{\alpha}{1-\alpha} D_{1}(W^{q_{\alpha,p}}_{\alpha} \,\|\, W \,|\, p) + I_{1}(p; W^{q_{\alpha,p}}_{\alpha}) \tag{35}$$

$$= \begin{cases} \inf\limits_{V \in \mathcal{P}(\mathcal{Y}|\mathscr{X})} \dfrac{\alpha}{1-\alpha} D_{1}(V \,\|\, W \,|\, p) + I_{1}(p;V), & \alpha \in (0,1), \\ \sup\limits_{V \in \mathcal{P}(\mathcal{Y}|\mathscr{X})} \dfrac{\alpha}{1-\alpha} D_{1}(V \,\|\, W \,|\, p) + I_{1}(p;V), & \alpha \in (1,\infty), \end{cases} \tag{36}$$

$$= \frac{\alpha}{1-\alpha} \inf_{V \in \mathcal{P}(\mathcal{Y}|\mathscr{X})} \left(D_{1}(V \,\|\, W \,|\, p) + \frac{1-\alpha}{\alpha} I_{1}(p;V) \right).$$

[11] Note that $\mathrm{T}_{\alpha,p}(q) = q$, on its own, does not imply $q_{\alpha,p} = q$ for α's in $(0,1)$. Consider for example a binary symmetric channel and let q be the probability measure that puts all its probability to one of the output letters. Then $\mathrm{T}_{\alpha,p}(q) = q$, but $q_{\alpha,p} \neq q$, for all $p \in \mathscr{P}(\mathscr{X})$ and $\alpha \in (0,1)$.

The convergence described in (31) holds not just for the Rényi mean $q^{\mathrm{g}}_{\alpha,p}$ but also for certain other probability measures, as well. [22, Remark 6 in Appendix B] describes how one can establish the following more general convergence result for any $\alpha \in (0,1)$ and $p \in \mathscr{P}(\mathscr{X})$:

$$\lim_{j\to\infty} \|q_{\alpha,p} - \mathrm{T}^{j}_{\alpha,p}(q)\| = 0 \quad \text{if} \quad q \sim q_{1,p} \quad \text{and} \quad \operatorname{ess\,sup}_{q_{1,p}} \left|\ln \frac{dq}{dq_{1,p}}\right| < \infty. \tag{37}$$

Part (a) is proved using Lemma 1; $I_\alpha(p;W) \leq \hbar(p)$ was proved by Csiszár through a different argument in [2, equation (24)]. Part (b), which is well known, is proved by substitution. Part (c) is due to[12] Augustin [6, Lemma 34.2]. Part (d) is new to the best our knowledge. Part (e) was proved for the finite $\mathscr{Y}$ case by Csiszár [2, equations (A24) and (A27)].

Definition 7. For any $\alpha \in \mathbb{R}_+$, $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$, $p \in \mathscr{P}(\mathscr{X})$ the unique probability measure $q_{\alpha,p}$ on $(\mathscr{Y},\mathcal{Y})$ satisfying $I_\alpha(p;W) = D_\alpha(W\,\|\,q_{\alpha,p}\,|\,p)$ is called the *order α Augustin mean for the input distribution p.*

Lemma 2 and equations (29), (32), and (34) imply the following bound, which is analogous to [34, Theorem 3.1] of Csiszár:

$$\sqrt{2\frac{D_\alpha(W\,\|\,q\,|\,p) - I_\alpha(p;W)}{\alpha \wedge 1}} \geq \|q_{\alpha,p} - q\|, \quad \forall q \in \mathcal{P}(\mathcal{Y}), \quad \forall \alpha \in \mathbb{R}_+.$$

The Augustin information and mean have closed form expressions only for $\alpha = 1$; for other orders they do not have closed form expressions. However, the fixed point property $\mathrm{T}_{\alpha,p}(q_{\alpha,p}) = q_{\alpha,p}$ established in Lemma 13(c),(d) and the definition of $\mathrm{T}_{\alpha,p}(\cdot)$ given in (28) imply the following identity for the Augustin mean:

$$\frac{dq_{\alpha,p}}{d\nu} = \left[\sum_x p(x) \left(\frac{dW(x)}{d\nu}\right)^{\alpha} e^{(1-\alpha)D_\alpha(W(x)\|q_{\alpha,p})}\right]^{\frac{1}{\alpha}}, \quad \forall \nu:\ q_{1,p} \prec \nu. \tag{38}$$

In Section 3.3, we use this identity in lieu of a closed form expression while analyzing $I_\alpha(p;W)$ and $q_{\alpha,p}$ as a function of α.

Lemma 14. *For any length n product channel $W_{[1,n]}\colon \mathscr{X}_1^n \to \mathcal{P}(\mathcal{Y}_1^n)$ and input distribution $p \in \mathscr{P}(\mathscr{X}_1^n)$ we have*

$$I_\alpha(p;W_{[1,n]}) \leq \sum_{t=1}^{n} I_\alpha(p_t;W_t) \tag{39}$$

for all $\alpha \in \mathbb{R}_+$, where $p_t \in \mathscr{P}(\mathscr{X}_t)$ is the marginal of p on $\mathscr{X}_t$. Furthermore, the inequality in (39) *is an equality for an $\alpha \in \mathbb{R}_+$ if and only if $q_{\alpha,p}$ satisfies*

$$q_{\alpha,p} = \bigotimes_{t=1}^{n} q_{\alpha,p_t}. \tag{40}$$

If $p = \bigotimes_{t=1}^{n} p_t$, then (40) *holds for all $\alpha \in \mathbb{R}_+$ and consequently* (39) *holds as an equality for all $\alpha \in \mathbb{R}_+$.*

3.2. Augustin Information as a Function of the Input Distribution

The order α Augustin information for the input distribution p is defined as the infimum of a family of conditional Rényi divergences, which are linear in p. Then the Augustin information is

[12] To be precise, [6, Lemma 34.2] does not include the assertion $D_1(q_{\alpha,p}\,\|\,q) \geq D_\alpha(W\,\|\,q\,|\,p) - I_\alpha(p;W)$ and claims (31) for $q_{1,p}$ instead of $q^{\mathrm{g}}_{\alpha,p}$. We cannot verify the correctness of Augustin's proof of [6, Lemma 34.2]; see [22, Appendix C] for a more detailed discussion.

concave in p, because pointwise infimum of a family of concave functions is concave. Lemma 15 strengthens this observation using Lemma 13.

Lemma 15. *For any $\alpha \in \mathbb{R}_+$ and $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$, $I_\alpha(p;W)$ is a concave function of p satisfying*

$$I_\alpha(p_\beta;W) \geq \beta I_\alpha(p_1;W) + (1-\beta)I_\alpha(p_0;W) + \beta D_{\alpha\wedge 1}(q_{\alpha,p_1} \,\|\, q_{\alpha,p_\beta}) + (1-\beta)D_{\alpha\wedge 1}(q_{\alpha,p_0} \,\|\, q_{\alpha,p_\beta}), \tag{41}$$

$$I_\alpha(p_\beta;W) \leq \beta I_\alpha(p_1;W) + (1-\beta)I_\alpha(p_0;W) + \beta D_{\alpha\vee 1}(q_{\alpha,p_1} \,\|\, q_{\alpha,p_\beta}) + (1-\beta)D_{\alpha\vee 1}(q_{\alpha,p_0} \,\|\, q_{\alpha,p_\beta}), \tag{42}$$

$$I_\alpha(p_\beta;W) \leq \beta I_\alpha(p_1;W) + (1-\beta)I_\alpha(p_0;W) + \hbar(\beta) - D_{\alpha\wedge 1}(q_{\alpha,p_\beta} \,\|\, \beta q_{\alpha,p_1} + (1-\beta)q_{\alpha,p_0}), \tag{43}$$

where $p_\beta = \beta p_1 + (1-\beta)p_0$ for all $p_0, p_1 \in \mathscr{P}(\mathscr{X})$ and $\beta \in [0,1]$.

Lemma 15 implies that for any positive order α and channel W, the order α Augustin information $I_\alpha(p;W)$ is a continuous function of the input distribution p if and only if $\sup_{p\in\mathscr{P}(\mathscr{X})} I_\alpha(p;W)$ is finite.[13] Furthermore, if $\sup_{p\in\mathscr{P}(\mathscr{X})} I_\eta(p;W)$ is finite for an $\eta \in \mathbb{R}_+$, then $\{I_\alpha(p;W)\}_{\alpha\in(0,\eta]}$ is uniformly equicontinuous in p on $\mathscr{P}(\mathscr{X})$.

In order to see why the finiteness of $\sup_{p\in\mathscr{P}(\mathscr{X})} I_\alpha(p;W)$ is necessary for the continuity, note that the nonnegativity of the Rényi divergence for probability measures and (41) imply that

$$\begin{aligned} I_\alpha(p_\beta;W) - I_\alpha(p_0;W) &\geq \beta(I_\alpha(p_1;W) - I_\alpha(p_0;W)) + \beta D_{\alpha\wedge 1}(q_{\alpha,p_1} \,\|\, q_{\alpha,p_\beta}) + (1-\beta)D_{\alpha\wedge 1}(q_{\alpha,p_0} \,\|\, q_{\alpha,p_\beta}) \\ &\geq \beta(I_\alpha(p_1;W) - I_\alpha(p_0;W)). \end{aligned}$$

On the other hand $\|p_\beta - p_0\| \leq 2\beta$. Thus, if there exists a $\{p_i\}_{i\in\mathbb{Z}_+} \subset \mathscr{P}(\mathscr{X})$ such that $\lim_{i\uparrow\infty} I_\alpha(p_i;W) = \infty$, then $I_\alpha(p;W)$ is discontinuous at every p in $\mathscr{P}(\mathscr{X})$.

The converse statement, i.e., the sufficiency, can be established together with the equicontinuity. For any $p_0, p_1 \in \mathscr{P}(\mathscr{X})$ such that $p_0 \neq p_1$ let $s_\wedge$, s_1, and s_0 be

$$s_\wedge = \frac{p_1 \wedge p_0}{\|p_1 \wedge p_0\|}, \qquad s_1 = \frac{p_1 - p_1 \wedge p_0}{1 - \|p_1 \wedge p_0\|}, \qquad s_0 = \frac{p_0 - p_1 \wedge p_0}{1 - \|p_1 \wedge p_0\|}.$$

Then $s_\wedge, s_1, s_0 \in \mathscr{P}(\mathscr{X})$ and $s_1 \perp s_0$. On the other hand $\|p_1 - p_0\| = 2 - 2\|p_1 \wedge p_0\|$. Therefore,

$$\begin{aligned} p_1 &= \left(\frac{2 - \|p_1 - p_0\|}{2}\right) s_\wedge + \frac{\|p_1 - p_0\|}{2} s_1, \\ p_0 &= \left(\frac{2 - \|p_1 - p_0\|}{2}\right) s_\wedge + \frac{\|p_1 - p_0\|}{2} s_0. \end{aligned}$$

Thus, as a result of Lemmas 2 and 15 we have

$$\begin{aligned} I_\alpha(p_0;W) - I_\alpha(p_1;W) &\leq \hbar\left(\frac{\|p_1 - p_0\|}{2}\right) + \frac{\|p_1 - p_0\|}{2}(I_\alpha(s_0;W) - I_\alpha(s_1;W)) \\ &\leq \hbar\left(\frac{\|p_1 - p_0\|}{2}\right) + \frac{\|p_1 - p_0\|}{2} I_\alpha(s_0;W), \quad \forall p_1, p_0 \in \mathscr{P}(\mathscr{X}),\ \alpha \in \mathbb{R}_+. \end{aligned} \tag{44}$$

[13] The Rényi information, discussed in Section 3.4, has already shown to satisfy analogous relations; see [13, Lemma 16(d),(e)]. The only substantial subtlety is that for orders in $(0,1)$ the Rényi information is a continuous function of p even when the corresponding capacity expression is infinite, because the Rényi information is quasi-concave rather than concave in p for orders in $(0,1)$; see [13, Lemma 6(a)].

Thus,

$$|I_\alpha(p_0;W)-I_\alpha(p_1;W)| \le \hbar\left(\frac{\|p_1-p_0\|}{2}\right)+\frac{\|p_1-p_0\|}{2}\sup_{p\in\mathscr{P}(\mathscr{X})} I_\eta(p;W), \quad \forall p_1,p_0\in\mathscr{P}(\mathscr{X}),\ \alpha\in(0,\eta].$$

3.3. Augustin Information as a Function of the Order

The main goal of this subsection is to characterize the behavior of the Augustin information as a function of the order for a given input distribution. Lemma 16 presents preliminary observations that facilitate the analysis of Augustin information as a function of the order; results of this analysis are presented in Lemma 17.

Lemma 16. *For any channel W of the form $W\colon \mathscr{X}\to\mathcal{P}(\mathcal{Y})$ and input distribution $p\in\mathscr{P}(\mathscr{X})$,*

(a) $D_\alpha(W(x)\,\|\,q_{\alpha,p})\le \ln\dfrac{1}{p(x)}$.

(b) $[p(x)]^{\frac{1}{\alpha\wedge 1}}W(x)\le q_{\alpha,p}$.

(c) $\left|\ln\dfrac{dq_{\alpha,p}}{dq_{1,p}}\right|\le\dfrac{|\alpha-1|}{\alpha}\ln\dfrac{1}{\min\limits_{x\colon p(x)>0} p(x)}$.

Bounds given in Lemma 16 follow from (38) via elementary manipulations.

Lemma 17. *For any channel W of the form $W\colon \mathscr{X}\to\mathcal{P}(\mathcal{Y})$ and input distribution $p\in\mathscr{P}(\mathscr{X})$,*

(a) *Either $(\alpha-1)I_\alpha(p;W)$ is a strictly convex function of α from $\mathbb{R}_+$ to $[-\hbar(p),\infty)$ or $I_\alpha(p;W)=\sum\limits_x p(x)\ln\gamma(x)$ for some $\gamma\colon\mathscr{X}\to[1,\infty)$ satisfying $\dfrac{dW(x)}{dq_{1,p}}=\gamma(x)$ $W(x)$-a.s. for all $x\in\operatorname{supp}p$ and $q_{\alpha,p}=q_{1,p}$ for all $\alpha\in\mathbb{R}_+$.*

(b) *$\dfrac{1-\alpha}{\alpha}I_\alpha(p;W)$ is a nonincreasing and continuous function of α from $\mathbb{R}_+$ to $\mathbb{R}$.*

(c) *$I_\alpha(p;W)$ is a nondecreasing and continuous function of α from $\mathbb{R}_+$ to $[0,\hbar(p)]$.*

(d) *$\left\{\ln\dfrac{dq_{\alpha,p}}{dq_{1,p}}\right\}_{y\in\mathscr{Y}}$ is an equicontinuous family of functions of α on $\mathbb{R}_+$.*

(e) *$I_\alpha(p;W)$ is a continuously differentiable function of α from $\mathbb{R}_+$ to $[0,\hbar(p)]$ such that*

$$\left.\frac{\partial}{\partial\alpha}I_\alpha(p;W)\right|_{\alpha=\phi}=\left.\frac{\partial}{\partial\alpha}D_\alpha(W\,\|\,q_{\phi,p}\,|\,p)\right|_{\alpha=\phi} \tag{45}$$

$$=\begin{cases}\dfrac{1}{(\phi-1)^2}D_1\big(W_\phi^{q_{\phi,p}}\,\|\,W\,|\,p\big) & \text{for } \phi\ne 1,\\ \displaystyle\sum_x\frac{p(x)}{2}\mathbf{E}_{W(x)}\left[\left(\ln\frac{dW(x)}{dq_{1,p}}-D_1(W(x)\,\|\,q_{1,p})\right)^2\right] & \text{for } \phi=1.\end{cases} \tag{46}$$

(f) *If $(\alpha-1)I_\alpha(p;W)$ is strictly convex in α, then $I_1(p;W_\alpha^{q_{\alpha,p}})$, i.e., $D_1(W_\alpha^{q_{\alpha,p}}\,\|\,q_{\alpha,p}\,|\,p)$, is a monotonically increasing continuous function of α on $\mathbb{R}_+$; otherwise, $I_1(p;W_\alpha^{q_{\alpha,p}})=\sum\limits_x p(x)\ln\gamma(x)$ (i.e., $D_1(W_\alpha^{q_{\alpha,p}}\,\|\,q_{\alpha,p}\,|\,p)=\sum\limits_x p(x)\ln\gamma(x)$) for some $\gamma\colon\mathscr{X}\to[1,\infty)$ satisfying $\dfrac{dW(x)}{dq_{1,p}}=\gamma(x)$ $W(x)$-a.s. for all $x\in\operatorname{supp}p$ and $q_{\alpha,p}=q_{1,p}$ for all $\alpha\in\mathbb{R}_+$.*

(g) $\lim\limits_{\alpha\downarrow 0} I_1(p;W_\alpha^{q_{\alpha,p}})=\lim\limits_{\alpha\downarrow 0} I_\alpha(p;W)$.

The (strict) convexity of $(\alpha-1)I_\alpha(p;W)$ in α on $\mathbb{R}_+$ is equivalent to the (strict) concavity of the function $sI_{\frac{1}{1+s}}(p;W)$ in s on $(-1,\infty)$; for a proof, see the proof of part (f). The concavity of the function $sI_{\frac{1}{1+s}}(p;W)$ in s on $(-1,\infty)$ and parts (b) and (c) of Lemma 17 have been reported by Augustin in [6, Lemma 34.3] for orders between zero and one. Parts (a) and (d)–(g) of Lemma 17 are new to the best of our knowledge. Lemma 17 is primarily about the Augustin information as a function of the order for a given input distribution. Part (d), i.e., the equicontinuity of $\left\{\ln \frac{dq_{\alpha,p}}{dq_{1,p}}\right\}_{y\in\mathscr{Y}}$ as a family of functions of the order α, is derived as a necessary tool for establishing the continuity of the derivative of the Augustin information, i.e., part (e). Note that Lemma 16(c) has already established this equicontinuity at $\alpha = 1$.

3.4. Augustin Information vs Rényi Information

The Augustin information is not the only information that has been defined in terms of the Rényi divergence; there are others. The Rényi information, defined first by Gallager[14] [14] and then by Sibson [35], is arguably the most prominent one among them because of its operational significance established by Gallager [14].

Definition 8. For any $\alpha \in \mathbb{R}_+$, $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$, and $p \in \mathscr{P}(\mathscr{X})$ the *order α Rényi information for the input distribution p* is

$$I^{\mathrm{g}}_{\alpha}(p;W) \triangleq \inf_{q\in\mathcal{P}(\mathcal{Y})} D_{\alpha}(p \circledast W \,\|\, p \otimes q). \tag{47}$$

As was noted by Sibson [35], one can confirm by substitution that

$$D_{\alpha}(p \circledast W \,\|\, p \otimes q) = D_{\alpha}(p \circledast W \,\|\, p \otimes q^{\mathrm{g}}_{\alpha,p}) + D_{\alpha}(q^{\mathrm{g}}_{\alpha,p} \,\|\, q), \quad \forall p \in \mathscr{P}(\mathscr{X}),\ q \in \mathcal{P}(\mathcal{Y}),\ \alpha \in \mathbb{R}_+$$

where $q^{\mathrm{g}}_{\alpha,p}$ is the Rényi mean defined in (27). Then using Lemma 2 we can conclude that

$$I^{\mathrm{g}}_{\alpha}(p;W) = D_{\alpha}(p \circledast W \,\|\, p \otimes q^{\mathrm{g}}_{\alpha,p}), \quad \forall p \in \mathscr{P}(\mathscr{X}),\ \alpha \in \mathbb{R}_+, \tag{48}$$

$$D_{\alpha}(p \circledast W \,\|\, p \otimes q) = I^{\mathrm{g}}_{\alpha}(p;W) + D_{\alpha}(q^{\mathrm{g}}_{\alpha,p} \,\|\, q), \quad \forall p \in \mathscr{P}(\mathscr{X}),\ q \in \mathcal{P}(\mathcal{Y}),\ \alpha \in \mathbb{R}_+. \tag{49}$$

For orders other than one the closed form expression given in (48) is equal to the following expression, which is sometimes taken as the definition of the Rényi information:

$$I^{\mathrm{g}}_{\alpha}(p;W) = \frac{\alpha}{\alpha-1} \ln \|\mu_{\alpha,p}\|, \quad \alpha \in \mathbb{R}_+ \setminus \{1\}.$$

Note that unlike the order α Augustin mean, the order α Rényi mean has a closed form expression for orders other than one, as well. Furthermore, the inequalities given in equations (29), (32), and (34) of Lemma 13 are replaced by the equality given in (49). A discussion of the Rényi information similar to the one we have presented in this section for the Augustin information can be found in [13].

The order one Rényi information is equal to the order one Augustin information for all input distributions. For other orders such an equality does not hold for arbitrary input distributions. However, it is possible to characterize the Augustin information and the Rényi information in terms of one another through appropriate variational forms. Characterizing the Augustin information in a variational form in terms of the Rényi information is especially useful, because the Augustin information does not have a closed form expression whereas the Rényi information does. This characterization also implies another variational characterization of the Augustin information.

Lemma 18. *Let W be a channel of the form $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$ and p be an input distribution in $\mathscr{P}(\mathscr{X})$.*

[14] Gallager uses a different parametrization and confines his discussion to the $\alpha \in (0,1)$ case.

(a) *Let* $u_{\alpha,p} \in \mathscr{P}(\mathscr{X})$ *be* $u_{\alpha,p}(x) = \dfrac{p(x)e^{(1-\alpha)D_\alpha(W(x)\,\|\,q_{\alpha,p})}}{\sum_{\widetilde{x}} p(\widetilde{x})e^{(1-\alpha)D_\alpha(W(\widetilde{x})\,\|\,q_{\alpha,p})}}$ *for all* x; *then*

$$I_\alpha(p;W) = I_\alpha^{\mathrm{g}}(u_{\alpha,p};W) + \frac{1}{\alpha-1}D_1(p\,\|\,u_{\alpha,p}) \tag{50}$$

$$= \begin{cases} \sup\limits_{u\in\mathscr{P}(\mathscr{X})} I_\alpha^{\mathrm{g}}(u;W) + \dfrac{1}{\alpha-1}D_1(p\,\|\,u) & \text{for } \alpha\in(0,1), \\ \inf\limits_{u\in\mathscr{P}(\mathscr{X})} I_\alpha^{\mathrm{g}}(u;W) + \dfrac{1}{\alpha-1}D_1(p\,\|\,u) & \text{for } \alpha\in(1,\infty). \end{cases} \tag{51}$$

(b) *Let* $a_{\alpha,p} \in \mathscr{P}(\mathscr{X})$ *be* $a_{\alpha,p}(x) = \dfrac{p(x)e^{(\alpha-1)D_\alpha\left(W(x)\,\|\,q^{\mathrm{g}}_{\alpha,p}\right)}}{\sum_{\widetilde{x}} p(\widetilde{x})e^{(\alpha-1)D_\alpha(W(\widetilde{x})\,\|\,q^{\mathrm{g}}_{\alpha,p})}}$ *for all* x; *then*

$$I_\alpha^{\mathrm{g}}(p;W) = I_\alpha(a_{\alpha,p};W) - \frac{1}{\alpha-1}D_1(a_{\alpha,p}\,\|\,p) \tag{52}$$

$$= \begin{cases} \inf\limits_{a\in\mathscr{P}(\mathscr{X})} I_\alpha(a;W) - \dfrac{1}{\alpha-1}D_1(a\,\|\,p) & \text{for } \alpha\in(0,1), \\ \sup\limits_{a\in\mathscr{P}(\mathscr{X})} I_\alpha(a;W) - \dfrac{1}{\alpha-1}D_1(a\,\|\,p) & \text{for } \alpha\in(1,\infty). \end{cases} \tag{53}$$

(c) *Let* $f_{\alpha,p}\colon \mathscr{X}\to\mathbb{R}$ *be* $f_{\alpha,p}(x) = [D_\alpha(W(x)\,\|\,q_{\alpha,p}) - I_\alpha(p;W)]\mathbb{1}_{p(x)>0}$ *for all* x; *then*

$$I_\alpha(p;W) = \frac{\alpha}{\alpha-1}\ln \mathbf{E}_\nu\left[\left(\sum_x p(x)e^{(1-\alpha)f_{\alpha,p}(x)}\left[\frac{dW(x)}{d\nu}\right]^\alpha\right)^{1/\alpha}\right] \tag{54}$$

$$= \frac{\alpha}{\alpha-1}\ln \inf_{f\colon \mathbf{E}_p[f]=0} \mathbf{E}_\nu\left[\left(\sum_x p(x)e^{(1-\alpha)f(x)}\left[\frac{dW(x)}{d\nu}\right]^\alpha\right)^{1/\alpha}\right]. \tag{55}$$

Lemma 18(a) was first proved by Poltyrev, [19, Theorem 3.4], in a slightly different form for the $\alpha\in[1/2,1)$ case assuming that $\mathscr{Y}$ is finite. Equation (53) of Lemma 18(b) was first proved by Shayevitz, [10, Theorem 1], for the finite $\mathscr{Y}$ case. Shayevitz, however, neither gave the expression for the optimal $a_{\alpha,p}$, nor asserted its existence in [10]. Lemma 18(c) was first proved by Augustin, [6, Lemma 35.7], for orders less than one.[15]

The following inequalities are implied by both $u=p$ point in the variational characterization given in Lemma 18(a) and $a=p$ point in the variational characterization given in Lemma 18(b):

$$I_\alpha(p;W) \ge I_\alpha^{\mathrm{g}}(p;W) \quad \text{for } \alpha\in(0,1], \tag{56}$$

$$I_\alpha(p;W) \le I_\alpha^{\mathrm{g}}(p;W) \quad \text{for } \alpha\in[1,\infty). \tag{57}$$

These inequalities can also be obtained using Jensen's inequality and the concavity of the natural logarithm function.

4. THE AUGUSTIN CAPACITY

In the previous section we have defined and analyzed the Augustin information and mean; our main aim in this section is doing the same for the Augustin capacity and center. In Section 4.1, we establish the existence of a unique Augustin center for all convex constraint sets with finite Augustin capacity and investigate the implications of the existence of an Augustin center for a given order and constraint set. In Section 4.2, we analyze the Augustin capacity and center as a function of the

[15] [6, Lemma 35.7(d)] is implied by the stronger inequalities established using (32) and Lemma 18(c).

order for a given constraint set. In Section 4.3, we bound the Augustin capacity of the convex hull of a collection of constraint sets on a given channel in terms of the Augustin capacities of individual constraint sets and determine the Augustin capacity of products of constraint sets on the product channels. Proofs of the propositions presented in this section can be found in [22, Appendix D].

Augustin provided a presentation similar to the current section in [6, Sections 33 and 34] and derived many of the key results—such as the existence of unique Augustin center and its continuity as a function of order; see [6, Lemmas 34.6–34.8]—for orders not exceeding one. Augustin, however, defines capacity and center only for the subsets of $\mathscr{P}(\mathscr{X})$ defined through cost constraints. We investigate that important special case more closely in Section 5.

4.1. Existence of a Unique Augustin Center

Definition 9. For any $\alpha \in \mathbb{R}_+$, $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$, and $\mathscr{A} \subset \mathscr{P}(\mathscr{X})$, the *order α Augustin capacity of W for constraint set $\mathscr{A}$* is

$$C_{\alpha,W,\mathscr{A}} \triangleq \sup_{p\in\mathscr{A}} I_\alpha(p;W).$$

When the constraint set $\mathscr{A}$ is the whole $\mathscr{P}(\mathscr{X})$, we denote the order α Augustin capacity by $C_{\alpha,W}$, i.e., $C_{\alpha,W} \triangleq C_{\alpha,W,\mathscr{P}(\mathscr{X})}$.

Using the definition of the Augustin information $I_\alpha(p;W)$ given in (23) we get the following expression for $C_{\alpha,W,\mathscr{A}}$:

$$C_{\alpha,W,\mathscr{A}} = \sup_{p\in\mathscr{A}} \inf_{q\in\mathcal{P}(\mathcal{Y})} D_\alpha(W \,\|\, q \,|\, p). \tag{58}$$

Theorem 1 in the following demonstrates that at least for convex $\mathscr{A}$'s one can exchange the order of the supremum and infimum without changing the value in the above expression.

Theorem 1. *For any order $\alpha \in \mathbb{R}_+$, channel W of the form $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$, and convex constraint set $\mathscr{A} \subset \mathscr{P}(\mathscr{X})$,*

$$\sup_{p\in\mathscr{A}} \inf_{q\in\mathcal{P}(\mathcal{Y})} D_\alpha(W \,\|\, q \,|\, p) = \inf_{q\in\mathcal{P}(\mathcal{Y})} \sup_{p\in\mathscr{A}} D_\alpha(W \,\|\, q \,|\, p). \tag{59}$$

If the expression on the left-hand side of (59) is finite, i.e., if $C_{\alpha,W,\mathscr{A}} \in \mathbb{R}_{\geq 0}$, then $\exists!\, q_{\alpha,W,\mathscr{A}} \in \mathcal{P}(\mathcal{Y})$, called the order α Augustin center of W for the constraint set $\mathscr{A}$, satisfying

$$C_{\alpha,W,\mathscr{A}} = \sup_{p\in\mathscr{A}} D_\alpha(W \,\|\, q_{\alpha,W,\mathscr{A}} \,|\, p). \tag{60}$$

Furthermore, for every sequence of input distributions $\{p^{(i)}\}_{i\in\mathbb{Z}_+} \subset \mathscr{A}$ satisfying $\lim_{i\to\infty} I_\alpha(p^{(i)};W) = C_{\alpha,W,\mathscr{A}}$, the corresponding sequence of order α Augustin means $\{q_{\alpha,p^{(i)}}\}_{i\in\mathbb{Z}_+}$ is a Cauchy sequence for the total variation metric on $\mathcal{P}(\mathcal{Y})$ and $q_{\alpha,W,\mathscr{A}}$ is the unique limit point of that Cauchy sequence.

In order to prove Theorem 1, we follow the program put forward by Kemperman [12] for establishing a similar result for the $\alpha = 1$ and $\mathscr{A} = \mathscr{P}(\mathscr{X})$ case. We first state and prove Theorem 1 assuming that the input set is finite. Then we generalize the result to the case with arbitrary input sets. In the case where $\mathscr{X}$ is a finite set, we can also assert the existence of an optimal input distribution for which the Augustin information is equal to the Augustin capacity.

Lemma 19. *For any order $\alpha \in \mathbb{R}_+$, channel W of the form $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$ with a finite input set $\mathscr{X}$, and closed convex constraint set $\mathscr{A} \subset \mathscr{P}(\mathscr{X})$, there exists $\widetilde{p} \in \mathscr{A}$ such that $I_\alpha(\widetilde{p};W) = C_{\alpha,W,\mathscr{A}}$ and $\exists!\, q_{\alpha,W,\mathscr{A}} \in \mathcal{P}(\mathcal{Y})$ satisfying*

$$D_\alpha(W \,\|\, q_{\alpha,W,\mathscr{A}} \,|\, p) \leq C_{\alpha,W,\mathscr{A}}, \quad \forall p \in \mathscr{A}. \tag{61}$$

Furthermore, $q_{\alpha,\widetilde{p}} = q_{\alpha,W,\mathscr{A}}$ for all $\widetilde{p} \in \mathscr{A}$ such that $I_\alpha(\widetilde{p};W) = C_{\alpha,W,\mathscr{A}}$.

If $\mathscr{A}$ is $\mathscr{P}(\mathscr{X})$, then the expression on the right-hand side of (60) is equal to the Rényi radius $S_{\alpha,W}$ defined in the following. Thus, Theorem 1 implies $C_{\alpha,W}=S_{\alpha,W}$.

Definition 10. For any $\alpha\in\mathbb{R}_{+}$ and $W\colon\mathscr{X}\to\mathcal{P}(\mathcal{Y})$, the *order α Rényi radius of W* is

$$S_{\alpha,W}\triangleq\inf_{q\in\mathcal{P}(\mathcal{Y})}\sup_{x\in\mathscr{X}}D_{\alpha}(W(x)\,\|\,q).$$

Theorem 1 asserts the existence of a unique order α Augustin center for convex constraint sets with finite Augustin capacity. However, a probability measure $q_{\alpha,W,\mathscr{A}}$ satisfying (60), i.e., an order α Augustin center, can in principle exist even for nonconvex constraint sets.

Definition 11. A constraint set $\mathscr{A}$ for the channel $W\colon\mathscr{X}\to\mathcal{P}(\mathcal{Y})$ has an order α Augustin center if and only if $\exists q\in\mathcal{P}(\mathcal{Y})$ such that

$$\sup_{p\in\mathscr{A}}D_{\alpha}(W\,\|\,q\,|\,p)=C_{\alpha,W,\mathscr{A}}. \tag{62}$$

If $C_{\alpha,W,\mathscr{A}}$ is infinite, then all probability measures on the output space satisfy (62) as a result of (58) and the max-min inequality. Thus, for constraint sets with infinite order α Augustin capacity, all probability measures on the output space are order α Augustin centers. On the other hand, some constraint sets do not have any order α Augustin center. Consider for example p_1 and p_2 satisfying $q_{\alpha,p_1}\neq q_{\alpha,p_2}$ and $I_{\alpha}(p_1;W)=I_{\alpha}(p_2;W)$. Then (62) is not satisfied by any probability measure for $\mathscr{A}=\{p_1,p_2\}$ and $\mathscr{A}$ does not have an order α Augustin center. Lemma 20 asserts that if Augustin center exists for a constraint set with finite Augustin capacity, then the Augustin center is unique.

Lemma 20. *Let $\mathscr{A}\subset\mathscr{P}(\mathscr{X})$ be a constraint set satisfying $C_{\alpha,W,\mathscr{A}}\in\mathbb{R}_{\geq 0}$, and $q_{\alpha,W,\mathscr{A}}$ be a probability measure satisfying* (62). *Then for every $\{p^{(i)}\}_{i\in\mathbb{Z}_{+}}\subset\mathscr{A}$ satisfying $\lim\limits_{i\to\infty}I_{\alpha}(p^{(i)};W)=C_{\alpha,W,\mathscr{A}}$ the sequence of order α Augustin means $\{q_{\alpha,p^{(i)}}\}_{i\in\mathbb{Z}_{+}}$ is a Cauchy sequence with the limit point $q_{\alpha,W,\mathscr{A}}$ and the order α Augustin center $q_{\alpha,W,\mathscr{A}}$ is unique.*

For any $\mathscr{A}$ that has an order α Augustin center and a finite $C_{\alpha,W,\mathscr{A}}$, Lemmas 13(b)–(d) and 20 imply that

$$C_{\alpha,W,\mathscr{A}}-I_{\alpha}(p;W)\geq D_{\alpha\wedge 1}(q_{\alpha,p}\,\|\,q_{\alpha,W,\mathscr{A}}),\quad\forall p\in\mathscr{A}.$$

Lemmas 13(b)–(d) and 20 can also be used establish a lower bound on $\sup\limits_{p\in\mathscr{A}}D_{\alpha}(W\,\|\,q\,|\,p)$ in terms of the Augustin capacity and center.

Lemma 21. *For any constraint set $\mathscr{A}$ that has an order α Augustin center and a finite $C_{\alpha,W,\mathscr{A}}$ we have*

$$\sup_{p\in\mathscr{A}}D_{\alpha}(W\,\|\,q\,|\,p)\geq C_{\alpha,W,\mathscr{A}}+D_{\alpha\wedge 1}(q_{\alpha,W,\mathscr{A}}\,\|\,q),\quad\forall q\in\mathcal{P}(\mathcal{Y}). \tag{63}$$

Note that the form of the lower bound given in (63) is, in a sense, analogous to the ones given in (29), (32), and (34). The bound given in (63) is a van Erven–Harremoës bound[16] for $\alpha\in(0,1]$, but it is not a van Erven–Harremoës bound for $\alpha\in(1,\infty)$, because we have a $D_{1}(q_{\alpha,W,\mathscr{A}}\,\|\,q)$ term rather than a $D_{\alpha}(q_{\alpha,W,\mathscr{A}}\,\|\,q)$ term for $\alpha\in(1,\infty)$.

[16] In [8] van Erven and Harremoës have conjectured that the inequality $\sup\limits_{x\in\mathscr{X}}D_{\alpha}(W(x)\,\|\,q)\geq C_{\alpha,W}+D_{\alpha}(q_{\alpha,W}\,\|\,q)$ holds for all $q\in\mathcal{P}(\mathcal{Y})$. Van Erven and Harremoës have also proved the bound for the case where $\alpha=\infty$, assuming that $\mathscr{Y}$ is countable [8, Theorem 37]. We have confirmed the van Erven–Harremoës conjecture in [13, Lemma 19] and generalized it to the convex constrained case for the Rényi capacity and center in [13, Lemma 25]. See Section 4.4 for a brief discussion of the Rényi capacity and center; a more comprehensive discussion can be found in [13].

For orders other than one, using Csiszár's form for the Augustin information given in (36) and the definition of the Augustin capacity, we obtain the following expressions:

$$C_{\alpha,W,\mathscr{A}} = \begin{cases} \sup\limits_{p\in\mathscr{A}} \inf\limits_{V\in\mathcal{P}(\mathcal{Y}|\mathscr{X})} \frac{\alpha}{1-\alpha} D_1(V \,\|\, W \,|\, p) + I_1(p;V) & \text{for } \alpha\in(0,1), \\ \sup\limits_{p\in\mathscr{A}} \sup\limits_{V\in\mathcal{P}(\mathcal{Y}|\mathscr{X})} \frac{\alpha}{1-\alpha} D_1(V \,\|\, W \,|\, p) + I_1(p;V) & \text{for } \alpha\in(1,\infty). \end{cases} \tag{64}$$

Then

$$C_{\alpha,W,\mathscr{A}} = \sup_{V\in\mathcal{P}(\mathcal{Y}|\mathscr{X})} \sup_{p\in\mathscr{A}} \frac{\alpha}{1-\alpha} D_1(V \,\|\, W \,|\, p) + I_1(p;V), \quad \forall\alpha\in(1,\infty).$$

For $\alpha\in(0,1)$, if the constraint set $\mathscr{A}$ has an order α Augustin center, e.g., when $\mathscr{A}$ is convex, then one can exchange the order of the supremum and the infimum and replace the infimum with a minimum whenever the Augustin capacity is finite by Lemma 22, given in the following.

Lemma 22. *For any $\alpha\in(0,1)$, if the constraint set $\mathscr{A}$ for the channel $W\colon \mathscr{X}\to\mathcal{P}(\mathcal{Y})$ has an order α Augustin center, then*

$$C_{\alpha,W,\mathscr{A}} = \inf_{V\in\mathcal{P}(\mathcal{Y}|\mathscr{X})} \sup_{p\in\mathscr{A}} \frac{\alpha}{1-\alpha} D_1(V \,\|\, W \,|\, p) + I_1(p;V). \tag{65}$$

If $C_{\alpha,W,\mathscr{A}}$ is finite, then $W_\alpha^{q_{\alpha,W,\mathscr{A}}}$ satisfies

$$C_{\alpha,W,\mathscr{A}} = \sup_{p\in\mathscr{A}} \frac{\alpha}{1-\alpha} D_1(W_\alpha^{q_{\alpha,W,\mathscr{A}}} \,\|\, W \,|\, p) + I_1(p;W_\alpha^{q_{\alpha,W,\mathscr{A}}}). \tag{66}$$

Lemma 22 is proved using Csiszár's form for the Augustin information, given in Lemma 13(e), and Lemma 20. In [36], Blahut proved a similar result assuming both $\mathscr{X}$ and $\mathscr{Y}$ are finite sets and $\mathscr{A}=\mathscr{P}(\mathscr{X})$. Even under those assumptions Blahut's result [36, Theorem 16] imply (65) and (66) for all orders in $(0,1)$ only when $C_{\alpha,W}$ is a differentiable function of the order α. Blahut was motivated by the expression for the sphere packing exponent; consequently, [36, Theorem 16] is stated in terms of an optimal input distribution at a given rate $R\in(C_{0,W},C_{1,W})$ and the corresponding optimal order $\alpha^*(R)$.

4.2. Augustin Capacity and Center as a Function of the Order

Lemma 23. *For any channel W of the form $W\colon\mathscr{X}\to\mathcal{P}(\mathcal{Y})$ and constraint set $\mathscr{A}\subset\mathscr{P}(\mathscr{X})$,*

(a) *$C_{\alpha,W,\mathscr{A}}$ is a nondecreasing and lower semicontinuous function of α on $\mathbb{R}_+$;*

(b) *$\frac{1-\alpha}{\alpha}C_{\alpha,W,\mathscr{A}}$ is a nonincreasing and continuous function of α on*[17] *$(0,1)$;*

(c) *$(\alpha-1)C_{\alpha,W,\mathscr{A}}$ is a convex function of α on $(1,\infty)$;*

(d) *$C_{\alpha,W,\mathscr{A}}$ is nondecreasing and continuous in α on $(0,1]$ and $(1,\chi_{W,\mathscr{A}}]$, where $\chi_{W,\mathscr{A}} \triangleq \sup\{\phi : C_{\phi,W,\mathscr{A}}\in\mathbb{R}_{\geq0}\}$;*

(e) *If $\sup\limits_{p\in\mathscr{A}} I_\phi^{\mathrm{g}}(p;W)\in\mathbb{R}_{\geq0}$ for a $\phi>1$, then $C_{\alpha,W,\mathscr{A}}$ is nondecreasing and continuous in α on $(0,(1\vee\chi_{W,\mathscr{A}})]$.*

The continuity results presented in parts (d) and (e) are somewhat unsatisfactory. One would like to either establish the continuity of $C_{\alpha,W,\mathscr{A}}$ from the right at $\alpha=1$ whenever $C_{\phi,W,\mathscr{A}}$ is finite for a $\phi>1$ or provide a channel W and a constraint set $\mathscr{A}$ for which $C_{\phi,W,\mathscr{A}}$ is finite for a $\phi>1$ and $\lim\limits_{\alpha\downarrow1} C_{\alpha,W,\mathscr{A}} > C_{1,W,\mathscr{A}}$. We could not do either. Instead we establish the continuity of $C_{\alpha,W,\mathscr{A}}$ from the right at $\alpha=1$ assuming that $\sup\limits_{p\in\mathscr{A}} I_\phi^{\mathrm{g}}(p;W)$ is finite for a $\phi>1$.

[17] We exclude the $\alpha=1$ case, because we do not want to assume $C_{1,W,\mathscr{A}}$ to be finite.

Since $C_{\phi,W}=S_{\phi,W}$ by Theorem 1 and $I_{\phi}^{\mathrm{g}}(p;W)\leq S_{\phi,W}$ for all $p\in\mathscr{P}(\mathscr{X})$ by (47), $\sup_{p\in\mathscr{A}} I_{\phi}^{\mathrm{g}}(p;W)$ is finite for all $\mathscr{A}\subset\mathscr{P}(\mathscr{X})$ whenever $C_{\phi,W}$ is finite. Thus, $C_{\alpha,W,\mathscr{A}}$ is nondecreasing and continuous in α on $(0,\chi_{W,\mathscr{A}}]$ for all $\mathscr{A}\subset\mathscr{P}(\mathscr{X})$, provided that $C_{\phi,W}$ is finite for a $\phi>1$.

Lemma 21 allows us to use the continuity of $C_{\alpha,W,\mathscr{A}}$ in α and Lemma 2 to establish the continuity of $q_{\alpha,W,\mathscr{A}}$ in α for the total variation topology on $\mathcal{P}(\mathcal{Y})$.

Lemma 24. *For any $\eta\in\mathbb{R}_{+}$, $W\colon\mathscr{X}\to\mathcal{P}(\mathcal{Y})$, and convex $\mathscr{A}\subset\mathscr{P}(\mathscr{X})$ such that $C_{\eta,W,\mathscr{A}}\in\mathbb{R}_{+}$,*

$$D_{\alpha\wedge 1}(q_{\alpha,W,\mathscr{A}}\,\|\,q_{\phi,W,\mathscr{A}})\leq C_{\phi,W,\mathscr{A}}-C_{\alpha,W,\mathscr{A}},\quad \forall\alpha,\phi \text{ such that } 0<\alpha<\phi\leq\eta. \tag{67}$$

Consequently, if $C_{\alpha,W,\mathscr{A}}$ is continuous in α on $\mathcal{I}$ for some $\mathcal{I}\subset(0,\eta]$, then $q_{\alpha,W,\mathscr{A}}\colon\mathcal{I}\to\mathcal{P}(\mathcal{Y})$ is continuous in α on $\mathcal{I}$ for the total variation topology on $\mathcal{P}(\mathcal{Y})$.

4.3. Convex Hulls of Constraints and Product Constraints

In the following we consider two kinds of frequently encountered constraint sets that are described in terms of simpler constraint sets. Lemma 25 considers convex hull of a family of constraint sets and bounds the Augustin capacity for the convex hull in terms of the Augustin capacities of the individual constraint sets. Lemma 26 considers a product channel for the constraint set that is the product of convex hulls of the constraint sets on the component channels that have Augustin centers and shows that Augustin capacity has an additive form and Augustin center has a product form.

Lemma 25. *Let α be a positive real, W be a channel of the form $W\colon\mathscr{X}\to\mathcal{P}(\mathcal{Y})$, and $\mathscr{A}^{(i)}$ be a constraint set that has an order α Augustin center and a finite $C_{\alpha,W,\mathscr{A}^{(i)}}$ for all $i\in\mathscr{T}$. Then*

$$\sup_{i\in\mathscr{T}} C_{\alpha,W,\mathscr{A}^{(i)}}\leq C_{\alpha,W,\mathscr{A}}\leq\ln\sum_{i\in\mathscr{T}} e^{C_{\alpha,W,\mathscr{A}^{(i)}}}$$

where $\mathscr{A}$ is the convex hull of the union, i.e., $\mathscr{A}=\mathrm{ch}\Big(\bigcup_{i\in\mathscr{T}}\mathscr{A}^{(i)}\Big)$. Furthermore,

- $C_{\alpha,W,\mathscr{A}^{(i)}}=C_{\alpha,W,\mathscr{A}}<\infty\Leftrightarrow\sup_{p\in\mathscr{A}} D_{\alpha}(W\,\|\,q_{\alpha,W,\mathscr{A}^{(i)}}\,|\,p)\leq C_{\alpha,W,\mathscr{A}^{(i)}}\Rightarrow q_{\alpha,W,\mathscr{A}}=q_{\alpha,W,\mathscr{A}^{(i)}}$;
- $C_{\alpha,W,\mathscr{A}}=\ln\sum_{i\in\mathscr{T}} e^{C_{\alpha,W,\mathscr{A}^{(i)}}}<\infty\Leftrightarrow q_{\alpha,W,\mathscr{A}^{(i)}}\perp q_{\alpha,W,\mathscr{A}^{(j)}}$ $\forall i\neq j$ *and* $|\mathscr{T}|<\infty\Rightarrow q_{\alpha,W,\mathscr{A}}=\sum_{i\in\mathscr{T}}\frac{e^{C_{\alpha,W,\mathscr{A}^{(i)}}}}{e^{C_{\alpha,W,\mathscr{A}}}}q_{\alpha,W,\mathscr{A}^{(i)}}$.

Note that if $\mathscr{A}^{(i)}$ is convex and $C_{\alpha,W,\mathscr{A}^{(i)}}$ is finite, then $\mathscr{A}^{(i)}$ has a unique order α Augustin center by Theorem 1.

Lemma 26. *For any $\alpha\in\mathbb{R}_{+}$, length n product channel $W_{[1,n]}\colon\mathscr{X}_1^n\to\mathcal{P}(\mathcal{Y}_1^n)$, and constraint sets $\mathscr{A}_t\subset\mathscr{P}(\mathscr{X}_t)$ that have order α Augustin centers*

$$C_{\alpha,W_{[1,n]},\mathscr{A}}=C_{\alpha,W_{[1,n]},\mathscr{A}_1^n}=\sum_{t=1}^{n} C_{\alpha,W_t,\mathscr{A}_t}$$

where $\mathscr{A}=\{p\in\mathscr{P}(\mathscr{X}_1^n)\colon\ p_t\in\mathrm{ch}\,\mathscr{A}_t\ \forall t\in\{1,\ldots,n\}\}$, i.e., a $p\in\mathscr{P}(\mathscr{X}_1^n)$ is in $\mathscr{A}$ if and only if for all $t\in\{1,\ldots,n\}$ its $\mathscr{X}_t$ marginal p_t is in the convex hull of $\mathscr{A}_t$. Furthermore, if $C_{\alpha,W_t,\mathscr{A}_t}$ is finite for all $t\in\{1,\ldots,n\}$, then $q_{\alpha,W_{[1,n]},\mathscr{A}}=q_{\alpha,W_{[1,n]},\mathscr{A}_1^n}=\bigotimes_{t=1}^{n} q_{\alpha,W_t,\mathscr{A}_t}$.

Remark 2. Note that the convex hull of any subset of $\mathscr{A}$ is a subset of $\mathscr{A}$, because $\mathscr{A}$ is convex by definition. In particular, $\mathscr{A}_1^n\subset\mathrm{ch}\,\mathscr{A}_1^n\subset\mathscr{A}$. Then $C_{\alpha,W_{[1,n]},\mathrm{ch}\,\mathscr{A}_1^n}=\sum_{t=1}^{n} C_{\alpha,W_t,\mathscr{A}_t}$ by Lemma 26. Furthermore, if $C_{\alpha,W_t,\mathscr{A}_t}$ is finite for all $t\in\{1,\ldots,n\}$, then $q_{\alpha,W_{[1,n]},\mathrm{ch}\,\mathscr{A}_1^n}=\bigotimes_{t=1}^{n} q_{\alpha,W_t,\mathscr{A}_t}$ by Lemma 25.

Remark 3. The constraint set $\mathscr{A}_1^n$ described in Lemma 26 may not be convex, yet $\mathscr{A}_1^n$ is guaranteed to have an order α Augustin center.

4.4. Augustin Capacity vs Rényi Capacity

Using the Rényi information instead of the Augustin information, one can define the Rényi capacity, as follows.

Definition 12. For any $\alpha \in \mathbb{R}_+$, $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$, and $\mathscr{A} \subset \mathscr{P}(\mathscr{X})$ the *order α Rényi capacity of W for constraint set $\mathscr{A}$* is

$$C^{\mathrm{g}}_{\alpha,W,\mathscr{A}} \triangleq \sup_{p\in\mathscr{A}} I^{\mathrm{g}}_{\alpha}(p;W).$$

When the constraint set $\mathscr{A}$ is the whole $\mathscr{P}(\mathscr{X})$, we denote the order α Rényi capacity by $C^{\mathrm{g}}_{\alpha,W}$, i.e., $C^{\mathrm{g}}_{\alpha,W} \triangleq C^{\mathrm{g}}_{\alpha,W,\mathscr{P}(\mathscr{X})}$.

Since $I_1(p;W) = I^{\mathrm{g}}_1(p;W)$, $C^{\mathrm{g}}_{1,W,\mathscr{A}} = C_{1,W,\mathscr{A}}$ by definition. We cannot say the same for other orders; by (56) and (57) we have

$$\begin{aligned} C^{\mathrm{g}}_{\alpha,W,\mathscr{A}} &\le C_{\alpha,W,\mathscr{A}} \quad \text{for } \alpha \in (0,1],\\ C^{\mathrm{g}}_{\alpha,W,\mathscr{A}} &\ge C_{\alpha,W,\mathscr{A}} \quad \text{for } \alpha \in [1,\infty). \end{aligned}$$

As a result of definitions of the Rényi information and capacity, we have

$$C^{\mathrm{g}}_{\alpha,W,\mathscr{A}} = \sup_{p\in\mathscr{A}} \inf_{q\in\mathcal{P}(\mathcal{Y})} D_{\alpha}(p \circledast W \,\|\, p\otimes q).$$

The Rényi capacity satisfies a minimax theorem, [13, Theorem 2], similar to Theorem 1: For any convex constraint set $\mathscr{A} \subset \mathscr{P}(\mathscr{X})$,

$$\sup_{p\in\mathscr{A}} \inf_{q\in\mathcal{P}(\mathcal{Y})} D_{\alpha}(p \circledast W \,\|\, p\otimes q) = \inf_{q\in\mathcal{P}(\mathcal{Y})} \sup_{p\in\mathscr{A}} D_{\alpha}(p \circledast W \,\|\, p\otimes q).$$

If $C^{\mathrm{g}}_{\alpha,W,\mathscr{A}}$ is finite, then $\exists!\, q^{\mathrm{g}}_{\alpha,W,\mathscr{A}} \in \mathcal{P}(\mathcal{Y})$, the *order α Rényi center of W for the constraint set $\mathscr{A}$*, satisfying

$$C^{\mathrm{g}}_{\alpha,W,\mathscr{A}} = \sup_{p\in\mathscr{A}} D_{\alpha}(p \circledast W \,\|\, p\otimes q^{\mathrm{g}}_{\alpha,W,\mathscr{A}}).$$

Consequently, the Rényi capacity equals to the Rényi radius provided that $\mathscr{A} = \mathscr{P}(\mathscr{X})$. Hence, $C^{\mathrm{g}}_{\alpha,W} = C_{\alpha,W}$ and $q^{\mathrm{g}}_{\alpha,W} = q_{\alpha,W}$ by Theorem 1. The other observations presented in this section have their counter parts for the Rényi capacity and center; compare, for example, Lemma 21 and [13, Lemma 25].

5. THE COST CONSTRAINED PROBLEM

In the previous section, we have defined the Augustin capacity for arbitrary constraint sets and proved the existence of a unique Augustin center for any convex constraint set with finite Augustin capacity. The convex constraint sets of interest are often defined via the cost constraints; the main aim of this section is to investigate this important special case more closely. In Section 5.1 we investigate the immediate consequences of the definition of the cost constrained Augustin capacity and ramifications of the analysis presented in the previous section. In Section 5.2 we define and analyze the Augustin–Legendre (A-L) information, capacity, radius, and center. The discussion in Section 5.2 is a generalization of certain parts of the analysis presented by Csiszár and Körner in [5, ch. 8] for the supremum of the mutual information for discrete channels with single cost constraint, i.e., the $\alpha = 1$, $|\mathscr{X}| < \infty$, $|\mathscr{Y}| < \infty$, $\ell = 1$ case. In Section 5.3 we define and analyze the Rényi–Gallager (R-G) information, mean, capacity, radius, and center. The most important conclusion of our analysis in Section 5.3 is the equality of the A-L capacity and center to the R-G capacity and center. In Section 5.4, we demonstrate how the results presented in Sections 5.1–5.3 can be used to determine the Augustin capacity and center of a transition probability with cost constraints. Proofs of the propositions presented in Sections 5.1–5.3 can be found in [22, Appendix E].

Augustin presented a discussion of the cost constrained capacity $C_{\alpha,W,\varrho}$ in [6, Section 34] for the case where the cost function ρ is a bounded function of the form $\rho\colon \mathscr{X} \to [0,1]^{\ell}$ and the order α is in $(0,1]$. In [6, Section 35], Augustin also analyzed quantities closely related to the R-G information and capacity. The quantities analyzed by Augustin in [6, Section 35] have first appeared in Gallager's error exponents analysis for cost constrained channels [14, Section 6; 15, Sections 7.3–7.5]. Unlike Augustin, Gallager did not assume ρ to be bounded; but Gallager confined his analysis to the case where there is a single cost constraint, i.e., the $\ell = 1$ case, and refrained from defining the R-G capacity as a quantity that is of interest on its own right. Other authors studying cost constrained problems, [24, Section IV; 25–27], have considered the R-G information and capacity, as well. Yet to the best of our knowledge for orders other than one the A-L information measures, which are obtained through a more direct application of convex conjugation, have not been studied before.

5.1. The Cost Constrained Augustin Capacity and Center

We denote the set of all probability mass functions satisfying a cost constraint ϱ by $\mathscr{A}(\varrho)$, i.e.,

$$\mathscr{A}(\varrho) \triangleq \{p \in \mathscr{P}(\mathscr{X}) : \mathbf{E}_p[\rho] \leq \varrho\}.$$

$\mathscr{A}(\varrho) \neq \varnothing$ if and only if $\varrho \in \Gamma_\rho$ where Γ_ρ is defined in (6) as the set of all feasible cost constraints for the cost function ρ. $\mathscr{A}(\varrho)$ is nondecreasing in ϱ, i.e., $\varrho_1 \leq \varrho_2$ implies $\mathscr{A}(\varrho_1) \subset \mathscr{A}(\varrho_2)$. We define the order α Augustin capacity of W for the cost constraint ϱ as

$$C_{\alpha,W,\varrho} \triangleq \begin{cases} \sup\limits_{p \in \mathscr{A}(\varrho)} I_\alpha(p;W) & \text{if } \varrho \in \Gamma_\rho, \\ -\infty & \text{if } \varrho \in \mathbb{R}^{\ell}_{\geq 0} \setminus \Gamma_\rho, \end{cases} \qquad \forall \alpha \in \mathbb{R}_+. \tag{68}$$

We defined $C_{\alpha,W,\varrho}$ for ϱ's that are not feasible in order to be able to use standard results without modifications. Since $\mathscr{A}(\varrho)$ is a convex set, Theorem 1 holds for $\mathscr{A}(\varrho)$. We denote[18] the order α Augustin center of W for the cost constraint ϱ by $q_{\alpha,W,\varrho}$.

For a given order α, the Augustin capacity $C_{\alpha,W,\varrho}$ is a concave function of the cost constraint ϱ. Hence, if it is finite at an interior point of Γ_ρ, then it is a continuous function of the cost constraint ϱ that lies below its tangent planes drawn at interior points of Γ_ρ. Lemma 27 below summarizes these observations.

Lemma 27. *Let W be a channel of the form $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$ with the cost function ρ of the form $\rho\colon \mathscr{X} \to \mathbb{R}^{\ell}_{\geq 0}$.*

(a) *For any $\alpha \in \mathbb{R}_+$, $C_{\alpha,W,\varrho}$ is a nondecreasing and concave function of ϱ on $\mathbb{R}^{\ell}_{\geq 0}$, which is either infinite on every point in $\operatorname{int}\Gamma_\rho$ or finite and continuous on $\operatorname{int}\Gamma_\rho$.*

(b) *If $C_{\alpha,W,\varrho}$ is finite on $\operatorname{int}\Gamma_\rho$ for an $\alpha \in \mathbb{R}_+$, then for every $\varrho \in \operatorname{int}\Gamma_\rho$ there exists a $\lambda_{\alpha,W,\varrho} \in \mathbb{R}^{\ell}_{\geq 0}$ such that*

$$C_{\alpha,W,\varrho} + \lambda_{\alpha,W,\varrho} \cdot (\widetilde{\varrho} - \varrho) \geq C_{\alpha,W,\widetilde{\varrho}}, \quad \forall \widetilde{\varrho} \in \mathbb{R}^{\ell}_{\geq 0}. \tag{69}$$

Furthermore, the set of all such $\lambda_{\alpha,W,\varrho}$ is convex and compact.

(c) *Either $C_{\alpha,W,\varrho} = \infty$ for all $(\alpha,\varrho) \in (0,1) \times \operatorname{int}\Gamma_\rho$ or $C_{\alpha,W,\varrho}$ and $q_{\alpha,W,\varrho}$ are continuous in (α,ϱ) on $(0,1) \times \operatorname{int}\Gamma_\rho$ for the total variation topology on $\mathcal{P}(\mathcal{Y})$.*

[18] This slight abuse of notation—which can be avoided by using $C_{\alpha,W,\mathscr{A}(\varrho)}$ and $q_{\alpha,W,\mathscr{A}(\varrho)}$ instead of $C_{\alpha,W,\varrho}$ and $q_{\alpha,W,\varrho}$—provides brevity without leading to any notational ambiguity.

If the cost function for a product channel is additive, then the cost constrained Augustin capacity of the product channel is equal to the supremum of the sum of the cost constrained Augustin capacities of the component channels over all feasible cost allocations. Furthermore, if there exists an optimal cost allocation, then the Augustin center of the product channel is a product measure. Lemma 28, given below, states these observations formally.

Lemma 28. *For any length n product channel* $W_{[1,n]}\colon \mathscr{X}_1^n \to \mathcal{P}(\mathcal{Y}_1^n)$ *and additive cost function* $\rho_{[1,n]}\colon \mathscr{X}_1^n \to \mathbb{R}_{\geq 0}^{\ell}$ *we have*[19]

$$C_{\alpha,W_{[1,n]},\varrho} = \sup\left\{\sum_{t=1}^{n} C_{\alpha,W_t,\varrho_t} : \sum_{t=1}^{n} \varrho_t \leq \varrho,\ \varrho_t \in \mathbb{R}_{\geq 0}^{\ell}\right\}, \quad \forall \varrho \in \mathbb{R}_{\geq 0}^{\ell}, \alpha \in \mathbb{R}_{+}. \tag{70}$$

If $C_{\alpha,W_{[1,n]},\varrho} \in \mathbb{R}_{\geq 0}$ *for an* $\alpha \in \mathbb{R}_+$ *and* $\exists(\varrho_1,\ldots,\varrho_n)$ *such that* $C_{\alpha,W_{[1,n]},\varrho} = \sum_{t=1}^{n} C_{\alpha,W_t,\varrho_t}$, *then* $q_{\alpha,W_{[1,n]},\varrho} = \bigotimes_{t=1}^{n} q_{\alpha,W_t,\varrho_t}$.

Since the Augustin capacity is concave in the cost constraint by Lemma 27(a), $C_{\alpha,W_{[1,n]},\varrho} = \sum_{t=1}^{n} C_{\alpha,W_t,\varrho/n}$ whenever $W_{[1,n]}$ is stationary and $\rho_t = \rho_1$ for all $t \in \{1,\ldots,n\}$. Alternatively, if Γ_{ρ_t}'s are closed and $C_{\alpha,W_t,\varrho}$'s are upper semicontinuous functions of ϱ on Γ_{ρ_t}'s, then we can use the extreme value theorem[20] for the upper semicontinuous functions to establish the existence of a $(\varrho_1,\ldots,\varrho_n)$ satisfying both $C_{\alpha,W_{[1,n]},\varrho} = \sum_{t=1}^{n} C_{\alpha,W_t,\varrho_t}$ and $\sum_{t=1}^{n} \varrho_t \leq \varrho$. However, such an existence assertion does not hold in general; see Example 3.

5.2. The Augustin–Legendre Information Measures

The cost constrained Augustin capacity $C_{\alpha,W,\varrho}$ and center $q_{\alpha,W,\varrho}$ can be characterized using convex conjugation, as well. In this part of the paper, we introduce and analyze the concepts of the Augustin–Legendre information, capacity, center, and radius in order to obtain a more complete understanding of this characterization. The current method seems to us to be the standard application of the convex conjugation technique to characterize the cost constrained Augustin capacity. Yet, it is not the customary method. Starting with the seminal work of Gallager [14], a more ad hoc method based on the Rényi information became the customary way to apply Lagrange multipliers techniques to characterize the Augustin capacity; see [6, Section 35; 25; 26]. We discuss that approach in Section 5.3. Theorem 2 presented in the following and Theorem 3 presented in Section 5.3 establish the equivalence of these two approaches by establishing the equality of the Augustin–Legendre capacity and center to the Rényi–Gallager capacity and center.

Definition 13. For any $\alpha \in \mathbb{R}_+$, channel W of the form $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$ with a cost function $\rho\colon \mathscr{X} \to \mathbb{R}_{\geq 0}^{\ell}$, $p \in \mathscr{P}(\mathscr{X})$, and $\lambda \in \mathbb{R}_{\geq 0}^{\ell}$, the *order* α *Augustin–Legendre information for the input distribution* p *and the Lagrange multiplier* λ is

$$I_\alpha^\lambda(p;W) \triangleq I_\alpha(p;W) - \lambda \cdot \mathbf{E}_p[\rho]. \tag{71}$$

Note that as an immediate consequence of the definition of the A-L information we have

$$\inf_{\lambda \geq 0} I_\alpha^\lambda(p;W) + \lambda \cdot \varrho = \xi_{\alpha,p}(\varrho) \tag{72}$$

[19] If $C_{\alpha,W_t,\varrho_t} = -\infty$ for any $t \in \{1,\ldots,n\}$, then $\sum_{t=1}^{n} C_{\alpha,W_t,\varrho_t}$ stands for $-\infty$; even if one or more of other C_{α,W_t,ϱ_t}'s are equal to ∞.

[20] Consider the function $f(\varrho_1,\ldots,\varrho_n)$ which is equal to $\sum_{t=1}^{n} C_{\alpha,W_t,\varrho_t}$ if $\sum_{t=1}^{n} \varrho_t \leq \varrho$ and $\varrho_t \in \Gamma_{\rho_t}$ for all $t \in \{1,\ldots,n\}$ and to $-\infty$ otherwise. We choose a large enough but bounded set using the vector ϱ to obtain a compact set for the supremum.

where $\xi_{\alpha,p}(\cdot)\colon \mathbb{R}^{\ell}_{\geq 0} \to [-\infty,\infty)$ is defined as

$$\xi_{\alpha,p}(\varrho) \triangleq \begin{cases} I_{\alpha}(p;W) & \text{if } \varrho \geq \mathbf{E}_p[\rho], \\ -\infty & \text{otherwise.} \end{cases} \tag{73}$$

Then the Augustin–Legendre information $I^{\lambda}_{\alpha}(p;W)$ can also be expressed as

$$I^{\lambda}_{\alpha}(p;W) = \sup_{\varrho \geq 0} \xi_{\alpha,p}(\varrho) - \lambda \cdot \varrho. \tag{74}$$

Remark 4. Note that if $f\colon \mathbb{R}^{\ell}_{\geq 0} \to (-\infty,\infty]$ and $f^{*}\colon (-\infty,0]^{\ell} \to \mathbb{R}$ are defined as $f(\varrho) \triangleq -\xi_{\alpha,p}(\varrho)$ and $f^{*}(\gamma) \triangleq I^{-\gamma}_{\alpha}(p;W)$, then f^{*} is the convex conjugate, i.e., Legendre transform, of the convex function f. This is why we call $I^{\lambda}_{\alpha}(p;W)$ the Augustin–Legendre information.

Definition 14. For any $\alpha \in \mathbb{R}_{+}$, channel W of the form $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$ with a cost function $\rho\colon \mathscr{X} \to \mathbb{R}^{\ell}_{\geq 0}$, and $\lambda \in \mathbb{R}^{\ell}_{\geq 0}$ the *order α Augustin–Legendre (A-L) capacity for the Lagrange multiplier λ* is

$$C^{\lambda}_{\alpha,W} \triangleq \sup_{p \in \mathscr{P}(\mathscr{X})} I^{\lambda}_{\alpha}(p;W). \tag{75}$$

Then as a result of (73) and (74) we have

$$C^{\lambda}_{\alpha,W} = \sup_{\varrho \geq 0} C_{\alpha,W,\varrho} - \lambda \cdot \varrho, \quad \forall \lambda \in \mathbb{R}^{\ell}_{\geq 0}. \tag{76}$$

Hence, using (72) and the max-min inequality we can conclude that

$$C_{\alpha,W,\varrho} \leq \inf_{\lambda \geq 0} C^{\lambda}_{\alpha,W} + \lambda \cdot \varrho, \quad \forall \varrho \in \mathbb{R}^{\ell}_{\geq 0}. \tag{77}$$

Then $C_{\alpha,W,\varrho} < \infty$ for all $\varrho \in \mathbb{R}^{\ell}_{\geq 0}$ provided that $C^{\lambda}_{\alpha,W} < \infty$ for a $\lambda \in \mathbb{R}_{\geq 0}$. But $C^{\lambda}_{\alpha,W} = \infty$ might hold for λ small enough even when $C_{\alpha,W,\varrho} < \infty$ for all $\varrho \in \mathbb{R}^{\ell}_{\geq 0}$; see Example 1.

Remark 5. In [6, Sections 33–35], Augustin considered the case where the cost function ρ is a bounded function of the form $\rho\colon \mathscr{X} \to [0,1]^{\ell}$. In that case $C^{\lambda}_{\alpha,W} < \infty$ for all $\lambda \in \mathbb{R}^{\ell}_{\geq 0}$ provided that $C_{\alpha,W,\varrho} < \infty$ for a $\varrho \in \operatorname{int}\Gamma_{\rho}$, because $C_{\alpha,W,\mathbb{1}} < \infty$ by Lemma 27(b) and $C_{\alpha,W,\mathbb{1}} = C_{\alpha,W}$ and $C^{\lambda}_{\alpha,W} \leq C^{0}_{\alpha,W} = C_{\alpha,W}$ for all $\lambda \in \mathbb{R}^{\ell}_{\geq 0}$ by definition.

The inequality given in (77) is an equality for many cases of interest as demonstrated by the following lemma. However, the inequality given in (77) is not an equality in general; see Example 2.

Lemma 29. *Let $\alpha \in \mathbb{R}_{+}$ and W be a channel of the form $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$ with a cost function $\rho\colon \mathscr{X} \to \mathbb{R}^{\ell}_{\geq 0}$. Then*

(a) $C^{\lambda}_{\alpha,W}$ *is convex, nonincreasing, and lower semicontinuous in λ on $\mathbb{R}^{\ell}_{\geq 0}$ and continuous in λ on $\{\lambda : \exists \varepsilon > 0$ such that $C^{\lambda-\varepsilon\mathbb{1}}_{\alpha,W} < \infty\}$.*

(b) *If $\mathscr{X}$ is a finite set, then $C_{\alpha,W,\varrho} = \inf_{\lambda \geq 0} C^{\lambda}_{\alpha,W} + \lambda \cdot \varrho$.*

(c) *If $\varrho \in \operatorname{int}\Gamma_{\rho}$, then $C_{\alpha,W,\varrho} = \inf_{\lambda \geq 0} C^{\lambda}_{\alpha,W} + \lambda \cdot \varrho$. If in addition $C_{\alpha,W,\varrho} < \infty$, then there exists a nonempty convex, compact set of $\lambda_{\alpha,W,\varrho}$'s satisfying both (69) and $C_{\alpha,W,\varrho} = C^{\lambda_{\alpha,W,\varrho}}_{\alpha,W} + \lambda_{\alpha,W,\varrho} \cdot \varrho$.*

(d) *If $C_{\alpha,W,\varrho}$ is finite and $C_{\alpha,W,\varrho} = C^{\lambda}_{\alpha,W} + \lambda \cdot \varrho$ for some $\varrho \in \Gamma_{\rho}$ and $\lambda \in \mathbb{R}^{\ell}_{\geq 0}$, then we have $\lim_{i \to \infty} I^{\lambda}_{\alpha}(p^{(i)};W) = C^{\lambda}_{\alpha,W}$ for all $\{p^{(i)}\}_{i \in \mathbb{Z}_{+}} \in \mathscr{A}(\varrho)$ such that $\lim_{i \to \infty} I_{\alpha}(p^{(i)};W) = C_{\alpha,W,\varrho}$.*

Using the definitions of $I_{\alpha}(p;W)$, $I^{\lambda}_{\alpha}(p;W)$, and $C^{\lambda}_{\alpha,W}$ given in (23), (71), and (75), we get the following expression for $C^{\lambda}_{\alpha,W}$:

$$C^{\lambda}_{\alpha,W} = \sup_{p \in \mathscr{P}(\mathscr{X})} \inf_{q \in \mathcal{P}(\mathcal{Y})} D_{\alpha}(W \,\|\, q \,|\, p) - \lambda \cdot \mathbf{E}_p[\rho]. \tag{78}$$

The A-L capacity satisfies a minimax theorem similar to the one satisfied by the Augustin capacity, which allows us to assert the existence of a unique A-L center whenever the A-L capacity is finite.

Theorem 2. *For any $\alpha \in \mathbb{R}_+$, channel $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$ with a cost function $\rho\colon \mathscr{X} \to \mathbb{R}^{\ell}_{\geq 0}$, and Lagrange multiplier $\lambda \in \mathbb{R}^{\ell}_{\geq 0}$,*

$$\sup_{p \in \mathscr{P}(\mathscr{X})} \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(W \,\|\, q \,|\, p) - \lambda \cdot \mathbf{E}_p[\rho] = \inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{p \in \mathscr{P}(\mathscr{X})} D_\alpha(W \,\|\, q \,|\, p) - \lambda \cdot \mathbf{E}_p[\rho] \tag{79}$$

$$= \inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{x \in \mathscr{X}} D_\alpha(W(x) \,\|\, q) - \lambda \cdot \rho(x). \tag{80}$$

If the expression on the left-hand side of (79) is finite, i.e., if $C^{\lambda}_{\alpha,W} < \infty$, then $\exists!\, q^{\lambda}_{\alpha,W} \in \mathcal{P}(\mathcal{Y})$, called the order α Augustin–Legendre center of W for the Lagrange multiplier λ, satisfying

$$C^{\lambda}_{\alpha,W} = \sup_{p \in \mathscr{P}(\mathscr{X})} D_\alpha(W \,\|\, q^{\lambda}_{\alpha,W} \,|\, p) - \lambda \cdot \mathbf{E}_p[\rho] \tag{81}$$

$$= \sup_{x \in \mathscr{X}} D_\alpha(W(x) \,\|\, q^{\lambda}_{\alpha,W}) - \lambda \cdot \rho(x). \tag{82}$$

Furthermore, for every sequence of input distributions $\{p^{(i)}\}_{i \in \mathbb{Z}_+} \subset \mathscr{P}(\mathscr{X})$ with $\lim_{i \to \infty} I^{\lambda}_{\alpha}(p^{(i)}; W) = C^{\lambda}_{\alpha,W}$, the corresponding sequence of order α Augustin means $\{q_{\alpha,p^{(i)}}\}_{i \in \mathbb{Z}_+}$ is a Cauchy sequence for the total variation metric on $\mathcal{P}(\mathcal{Y})$ and $q^{\lambda}_{\alpha,W}$ is the unique limit point of that Cauchy sequence.

Note that Theorem 2 for $\lambda = 0$ is nothing but Theorem 1 for $\mathscr{A} = \mathscr{P}(\mathscr{X})$. The proof of Theorem 2 is very similar to that of Theorem 1, as well; it employs Lemma 30, presented in the following, instead of Lemma 19. Note that, Lemma 30 for $\lambda = 0$ is nothing but Lemma 19 for $\mathscr{A} = \mathscr{P}(\mathscr{X})$, as well.

Lemma 30. *For any $\alpha \in \mathbb{R}_+$, channel $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$ with a cost function $\rho\colon \mathscr{X} \to \mathbb{R}^{\ell}_{\geq 0}$ for a finite input set $\mathscr{X}$, and Lagrange multiplier $\lambda \in \mathbb{R}^{\ell}_{\geq 0}$, there exists a $\widetilde{p} \in \mathscr{P}(\mathscr{X})$ such that $I^{\lambda}_{\alpha}(\widetilde{p}; W) = C^{\lambda}_{\alpha,W}$ and $\exists!\, q^{\lambda}_{\alpha,W} \in \mathcal{P}(\mathcal{Y})$ satisfying*

$$D_\alpha(W \,\|\, q^{\lambda}_{\alpha,W} \,|\, p) - \lambda \cdot \mathbf{E}_p[\rho] \leq C^{\lambda}_{\alpha,W}, \quad \forall p \in \mathscr{P}(\mathscr{X}). \tag{83}$$

Furthermore, $q_{\alpha,\widetilde{p}} = q^{\lambda}_{\alpha,W}$ for all $\widetilde{p} \in \mathscr{P}(\mathscr{X})$ such that $I^{\lambda}_{\alpha}(\widetilde{p}; W) = C^{\lambda}_{\alpha,W}$.

Note that the expression on the left-hand side of equation (79) is nothing but the A-L capacity. Thus, Theorem 2 is establishes the equality of the A-L capacity to the A-L radius defined in the following.

Definition 15. For any $\alpha \in \mathbb{R}_+$, channel $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$ with a cost function $\rho\colon \mathscr{X} \to \mathbb{R}^{\ell}_{\geq 0}$, and $\lambda \in \mathbb{R}^{\ell}_{\geq 0}$, the *order α Augustin–Legendre radius of W for the Lagrange multiplier λ* is

$$S^{\lambda}_{\alpha,W} \triangleq \inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{x \in \mathscr{X}} D_\alpha(W(x) \,\|\, q) - \lambda \cdot \rho(x). \tag{84}$$

If $C^{\lambda}_{\alpha,W}$ is finite, then Lemma 13(b)–(d), Theorem 2, and the definition of $I^{\lambda}_{\alpha}(p; W)$ given in (71) imply that

$$C^{\lambda}_{\alpha,W} - I^{\lambda}_{\alpha}(p; W) \geq D_{\alpha \wedge 1}(q_{\alpha,p} \,\|\, q^{\lambda}_{\alpha,W}), \quad \forall p \in \mathscr{P}(\mathscr{X}).$$

Using Lemma 13 and Theorem 2, one can also establish a bound similar to the one given in Lemma 21. However, we will not do so here, because one can obtain a slightly stronger results using the characterization of the A-L capacity and center via R-G capacity and center presented in Section 5.3; see Lemma 35 and the ensuing discussion.

As a result of Lemma 29(c), we know that if $C_{\alpha,W,\varrho}$ is finite for a $\varrho \in \operatorname{int}\Gamma_\rho$, then there exists at least one $\lambda_{\alpha,W,\varrho}$ for which $C_{\alpha,W,\varrho} = C_{\alpha,W}^{\lambda_{\alpha,W,\varrho}} + \lambda_{\alpha,W,\varrho} \cdot \varrho$ holds. Lemma 31, given in the following, asserts that for any such Lagrange multiplier the corresponding order α A-L center should be equal to the order α Augustin center for the cost constraint ϱ. Thus, if there are multiple $\lambda_{\alpha,W,\varrho}$'s satisfying $C_{\alpha,W,\varrho} = C_{\alpha,W}^{\lambda_{\alpha,W,\varrho}} + \lambda_{\alpha,W,\varrho} \cdot \varrho$, then they all have the same order α A-L center.

Lemma 31. *For any $\alpha \in \mathbb{R}_+$, channel $W \colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$ with a cost function $\rho \colon \mathscr{X} \to \mathbb{R}^\ell_{\geq 0}$, and a cost constraint $\varrho \in \Gamma_\rho$ such that $C_{\alpha,W,\varrho} < \infty$, if $C_{\alpha,W,\varrho} = C^\lambda_{\alpha,W} + \lambda \cdot \varrho$ for a $\lambda \in \mathbb{R}^\ell_{\geq 0}$, then $q_{\alpha,W,\varrho} = q^\lambda_{\alpha,W}$.*

For product constraints on product channels, the Augustin capacity has an additive form and the Augustin center has a multiplicative form—whenever it exists—by Lemma 26. The cost constraints for additive cost functions, however, are not product constraints. In order to calculate the cost constrained Augustin capacity for product channels with additive cost functions, we need to optimize over the feasible allocations of the cost over the component channels by Lemma 28. In addition, we can express the cost constrained Augustin center of the product channel as the product of the cost constrained Augustin centers of the components channels—using Lemma 28—only when there exists a feasible allocation of the cost that achieves the optimum value. For the A-L capacity and center, on the other hand, we have a considerably neater picture: For product channels with additive cost functions the A-L capacity is additive and the A-L center is multiplicative, whenever it exists.

Lemma 32. *For any length n product channel $W_{[1,n]} \colon \mathscr{X}_1^n \to \mathcal{P}(\mathcal{Y}_1^n)$ and additive cost function $\rho_{[1,n]} \colon \mathscr{X}_1^n \to \mathbb{R}^\ell_{\geq 0}$ we have*

$$C^\lambda_{\alpha,W_{[1,n]}} = \sum_{t=1}^{n} C^\lambda_{\alpha,W_t}, \quad \forall \lambda \in \mathbb{R}^\ell_{\geq 0},\ \alpha \in \mathbb{R}_+. \tag{85}$$

Furthermore, if $C^\lambda_{\alpha,W_{[1,n]}} < \infty$, then $q^\lambda_{\alpha,W_{[1,n]}} = \bigotimes_{t=1}^{n} q^\lambda_{\alpha,W_t}$.

The additivity of the cost function $\rho_{[1,n]}$ implies for any p in $\mathscr{P}(\mathscr{X}_1^n)$

$$\mathbf{E}_p[\rho_{[1,n]}] = \sum_{t=1}^{n} \mathbf{E}_{p_t}[\rho_t]$$

where $p_t \in \mathscr{P}(\mathscr{X}_t)$ is the $\mathscr{X}_t$ marginal of p. Thus, Lemma 14 and the definition of the A-L information imply

$$\begin{aligned} I^\lambda_\alpha(p; W_{[1,n]}) &\leq I^\lambda_\alpha(p_1 \otimes \ldots \otimes p_n; W_{[1,n]}) \\ &= \sum_{t=1}^{n} I^\lambda_\alpha(p_t; W_t). \end{aligned} \tag{86}$$

Lemma 32 is proved using (86) together with Theorem 2.

5.3. The Rényi–Gallager Information Measures

In Section 5.2, we have characterized the cost constrained Augustin capacity and center in terms of the A-L capacity and center. The A-L capacity is defined as the supremum of the A-L information. Gallager—implicitly—proposed another information with a Lagrange multiplier in [14, (103) and (116)]. Augustin characterized the cost constrained Augustin capacity in terms of the supremum of this information, assuming that the cost function is bounded, in [6, Lemmas 35.4(b) and 35.8(b)]. We call this supremum the R-G capacity. The main aim of this subsection is establishing the equality of the A-L capacity and center to the R-G capacity and center. We will also derive a van Erven–Harremoës bound for the A-L capacity and center and use it to derive the continuity of the A-L center as a function of the Lagrange multiplier λ.

Definition 16. For any $\alpha \in \mathbb{R}_+$, channel $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$ with a cost function $\rho\colon \mathscr{X} \to \mathbb{R}^{\ell}_{\geq 0}$, $p \in \mathscr{P}(\mathscr{X})$, and $\lambda \in \mathbb{R}^{\ell}_{\geq 0}$, the *order α Rényi–Gallager (R-G) information for the input distribution p and the Lagrange multiplier λ* is

$$I^{\mathrm{g}\lambda}_{\alpha}(p;W) \triangleq \begin{cases} \inf\limits_{q \in \mathcal{P}(\mathcal{Y})} D_{\alpha}(p \circledast W e^{\frac{1-\alpha}{\alpha}\lambda\cdot\rho} \,\|\, p \otimes q) & \text{for } \alpha \in \mathbb{R}_+ \setminus \{1\}, \\ \inf\limits_{q \in \mathcal{P}(\mathcal{Y})} D_1(p \circledast W \,\|\, p \otimes q) - \lambda \cdot \mathbf{E}_p[\rho] & \text{for } \alpha = 1. \end{cases} \tag{87}$$

If λ is a vector of zeros, then the R-G information is the Rényi information. Similar to the Rényi information, the R-G information has a closed form expression, described in terms of the probability measure achieving the infimum in its definition.

Definition 17. For any $\alpha \in \mathbb{R}_+$, channel $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$ with a cost function $\rho\colon \mathscr{X} \to \mathbb{R}^{\ell}_{\geq 0}$, $p \in \mathscr{P}(\mathscr{X})$, and $\lambda \in \mathbb{R}^{\ell}_{\geq 0}$, the *order α mean measure for the input distribution p and the Lagrange multiplier λ* is

$$\frac{d\mu^{\lambda}_{\alpha,p}}{d\nu} \triangleq \left[\sum_{x} p(x) e^{(1-\alpha)\lambda\cdot\rho(x)} \left(\frac{dW(x)}{d\nu}\right)^{\alpha}\right]^{\frac{1}{\alpha}}. \tag{88}$$

The *order α Rényi–Gallager (R-G) mean for the input distribution p and the Lagrange multiplier λ* is

$$q^{\mathrm{g}\lambda}_{\alpha,p} \triangleq \frac{\mu^{\lambda}_{\alpha,p}}{\|\mu^{\lambda}_{\alpha,p}\|}. \tag{89}$$

Both $\mu^{\lambda}_{\alpha,p}$ and $q^{\mathrm{g}\lambda}_{\alpha,p}$ depend on the Lagrange multiplier λ for $\alpha \in \mathbb{R}_+ \setminus \{1\}$. Furthermore, one can confirm by substitution that

$$D_{\alpha}(p \circledast W e^{\frac{1-\alpha}{\alpha}\lambda\cdot\rho} \,\|\, p \otimes q) = D_{\alpha}(p \circledast W e^{\frac{1-\alpha}{\alpha}\lambda\cdot\rho} \,\|\, p \otimes q^{\mathrm{g}\lambda}_{\alpha,p}) + D_{\alpha}(q^{\mathrm{g}\lambda}_{\alpha,p} \,\|\, q), \quad \alpha \in \mathbb{R}_+ \setminus \{1\}. \tag{90}$$

Then, as a result of Lemma 2, we have

$$I^{\mathrm{g}\lambda}_{\alpha}(p;W) = D_{\alpha}(p \circledast W e^{\frac{1-\alpha}{\alpha}\lambda\cdot\rho} \,\|\, p \otimes q^{\mathrm{g}\lambda}_{\alpha,p}) \tag{91}$$

$$= \frac{\alpha}{\alpha-1} \ln \|\mu^{\lambda}_{\alpha,p}\| \quad \text{for } \alpha \in \mathbb{R}_+ \setminus \{1\}. \tag{92}$$

Neither $\mu^{\lambda}_{1,p}$, nor $q^{\mathrm{g}\lambda}_{1,p}$ depends on the Lagrange multiplier λ. In addition, one can confirm by substitution that

$$D_1(p \circledast W \,\|\, p \otimes q) - \lambda \cdot \mathbf{E}_p[\rho] = D_1(p \circledast W \,\|\, p \otimes q^{\mathrm{g}\lambda}_{1,p}) - \lambda \cdot \mathbf{E}_p[\rho] + D_1(q^{\mathrm{g}\lambda}_{1,p} \,\|\, q). \tag{93}$$

Then as a result of Lemma 2, we have

$$I^{\mathrm{g}\lambda}_{1}(p;W) = D_1(p \circledast W \,\|\, p \otimes q^{\mathrm{g}\lambda}_{1,p}) - \lambda \cdot \mathbf{E}_p[\rho]. \tag{94}$$

Using the definitions of the A-L information and the R-G information given in (71) and (87) together with Jensen's inequality and the concavity of the natural logarithm function we get

$$I^{\lambda}_{\alpha}(p;W) \geq I^{\mathrm{g}\lambda}_{\alpha}(p;W) \quad \text{for } \alpha \in (0,1],$$
$$I^{\lambda}_{\alpha}(p;W) \leq I^{\mathrm{g}\lambda}_{\alpha}(p;W) \quad \text{for } \alpha \in [1,\infty).$$

It is possible to strengthen these relations by expressing the A-L information and the R-G information in terms of one another as follows.

Lemma 33. *Let W be a channel of the form $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$ with a cost function $\rho\colon \mathscr{X} \to \mathbb{R}^{\ell}_{\geq 0}$, p be an input distribution in $\mathscr{P}(\mathscr{X})$, and λ be a Lagrange multiplier in $\mathbb{R}^{\ell}_{\geq 0}$.*

(a) *Let* $u^{\lambda}_{\alpha,p} \in \mathscr{P}(\mathscr{X})$ *be* $u^{\lambda}_{\alpha,p}(x) = \frac{p(x)e^{(1-\alpha)D_{\alpha}(W(x)\| q_{\alpha,p})+(\alpha-1)\lambda\cdot\rho(x)}}{\sum_{\widetilde{x}} p(\widetilde{x})e^{(1-\alpha)D_{\alpha}(W(\widetilde{x})\| q_{\alpha,p})+(\alpha-1)\lambda\cdot\rho(\widetilde{x})}}$ *for all* x; *then*

$$I^{\lambda}_{\alpha}(p;W) = I^{\mathrm{g}\lambda}_{\alpha}(u_{\alpha,p};W) + \frac{1}{\alpha-1}D_1(p\,\|\,u_{\alpha,p}) \tag{95}$$

$$= \begin{cases} \sup\limits_{u\in\mathscr{P}(\mathscr{X})} I^{\mathrm{g}\lambda}_{\alpha}(u;W) + \dfrac{1}{\alpha-1}D_1(p\,\|\,u) & \text{for } \alpha\in(0,1),\\ \inf\limits_{u\in\mathscr{P}(\mathscr{X})} I^{\mathrm{g}\lambda}_{\alpha}(u;W) + \dfrac{1}{\alpha-1}D_1(p\,\|\,u) & \text{for } \alpha\in(1,\infty). \end{cases} \tag{96}$$

(b) *Let* $a^{\lambda}_{\alpha,p} \in \mathscr{P}(\mathscr{X})$ *be* $a^{\lambda}_{\alpha,p}(x) = \frac{p(x)e^{(\alpha-1)D_{\alpha}\left(W(x)\| q^{\mathrm{g}\lambda}_{\alpha,p}\right)+(1-\alpha)\lambda\cdot\rho(x)}}{\sum_{\widetilde{x}} p(\widetilde{x})e^{(\alpha-1)D_{\alpha}(W(\widetilde{x})\| q^{\mathrm{g}\lambda}_{\alpha,p})+(1-\alpha)\lambda\cdot\rho(\widetilde{x})}}$ *for all* x; *then*

$$I^{\mathrm{g}\lambda}_{\alpha}(p;W) = I^{\lambda}_{\alpha}(a^{\lambda}_{\alpha,p};W) - \frac{1}{\alpha-1}D_1(a^{\lambda}_{\alpha,p}\,\|\,p) \tag{97}$$

$$= \begin{cases} \inf\limits_{a\in\mathscr{P}(\mathscr{X})} I^{\lambda}_{\alpha}(a;W) - \dfrac{1}{\alpha-1}D_1(a\,\|\,p) & \text{for } \alpha\in(0,1),\\ \sup\limits_{a\in\mathscr{P}(\mathscr{X})} I^{\lambda}_{\alpha}(a;W) - \dfrac{1}{\alpha-1}D_1(a\,\|\,p) & \text{for } \alpha\in(1,\infty). \end{cases} \tag{98}$$

(c) *Let* $f^{\lambda}_{\alpha,p}\colon \mathscr{X}\to\mathbb{R}$ *be* $f^{\lambda}_{\alpha,p}(x) = [D_{\alpha}(W(x)\,\|\,q_{\alpha,p}) - \lambda\cdot\rho(x) - I^{\lambda}_{\alpha}(p;W)]\mathbb{1}_{p(x)>0}$ *for all* x; *then*

$$I^{\lambda}_{\alpha}(p;W) = \frac{\alpha}{\alpha-1}\ln \mathbf{E}_{\nu}\left[\left(\sum_x p(x)e^{(1-\alpha)(f^{\lambda}_{\alpha,p}(x)+\lambda\cdot\rho(x))}\left[\frac{dW(x)}{d\nu}\right]^{\alpha}\right)^{1/\alpha}\right] \tag{99}$$

$$= \frac{\alpha}{\alpha-1}\ln \inf_{f:\mathbf{E}_p[f]=0} \mathbf{E}_{\nu}\left[\left(\sum_x p(x)e^{(1-\alpha)(f(x)+\lambda\cdot\rho(x))}\left[\frac{dW(x)}{d\nu}\right]^{\alpha}\right)^{1/\alpha}\right]. \tag{100}$$

Lemma 33 for $\lambda = 0$ is Lemma 18, which was previously discussed by Poltyrev [19], Shayevitz [10], and Augustin [6].

Definition 18. For any $\alpha\in\mathbb{R}_+$, channel $W\colon\mathscr{X}\to\mathcal{P}(\mathcal{Y})$ with a cost function $\rho\colon\mathscr{X}\to\mathbb{R}^{\ell}_{\geq 0}$, and $\lambda\in\mathbb{R}^{\ell}_{\geq 0}$, the *order* α *Rényi–Gallager (R-G) capacity for the Lagrange multiplier* λ is

$$C^{\mathrm{g}\lambda}_{\alpha,W} \triangleq \sup_{p\in\mathscr{P}(\mathscr{X})} I^{\mathrm{g}\lambda}_{\alpha}(p;W).$$

Using the definition of $I^{\mathrm{g}\lambda}_{\alpha}(p;W)$, given in (87), we get the following expression for $C^{\mathrm{g}\lambda}_{\alpha,W}$:

$$C^{\mathrm{g}\lambda}_{\alpha,W} = \begin{cases} \sup\limits_{p\in\mathscr{P}(\mathscr{X})}\inf\limits_{q\in\mathcal{P}(\mathcal{Y})} D_{\alpha}(p\circledast We^{\frac{1-\alpha}{\alpha}\lambda\cdot\rho}\,\|\,p\otimes q) & \text{for } \alpha\in\mathbb{R}_+\setminus\{1\},\\ \sup\limits_{p\in\mathscr{P}(\mathscr{X})}\inf\limits_{q\in\mathcal{P}(\mathcal{Y})} D_{\alpha}(p\circledast W\,\|\,p\otimes q) - \lambda\cdot\mathbf{E}_p[\rho] & \text{for } \alpha = 1. \end{cases} \tag{101}$$

The R-G capacity satisfies a minimax theorem similar to the one satisfied by the A-L capacity, i.e., Theorem 2. Since both the statement and the proof of the minimax theorems are identical for the order one A-L capacity and the order one R-G capacity, we state the minimax theorem for the R-G capacity only for finite positive orders other than one.

Theorem 3. *For any* $\alpha\in\mathbb{R}_+\setminus\{1\}$, *channel* $W\colon\mathscr{X}\to\mathcal{P}(\mathcal{Y})$ *with a cost function* $\rho\colon\mathscr{X}\to\mathbb{R}^{\ell}_{\geq 0}$, *and Lagrange multiplier* $\lambda\in\mathbb{R}^{\ell}_{\geq 0}$,

$$\sup_{p\in\mathscr{P}(\mathscr{X})}\inf_{q\in\mathcal{P}(\mathcal{Y})} D_{\alpha}(p\circledast We^{\frac{1-\alpha}{\alpha}\lambda\cdot\rho}\,\|\,p\otimes q) = \inf_{q\in\mathcal{P}(\mathcal{Y})}\sup_{p\in\mathscr{P}(\mathscr{X})} D_{\alpha}(p\circledast We^{\frac{1-\alpha}{\alpha}\lambda\cdot\rho}\,\|\,p\otimes q) \tag{102}$$

$$= \inf_{q\in\mathcal{P}(\mathcal{Y})}\sup_{x\in\mathscr{X}} D_{\alpha}(W(x)\,\|\,q) - \lambda\cdot\rho(x). \tag{103}$$

If the expression on the left-hand side of (102) *is finite, i.e., if* $C^{g\lambda}_{\alpha,W} < \infty$, *then* $\exists!\, q^{g\lambda}_{\alpha,W} \in \mathcal{P}(\mathcal{Y})$, *called the order* α *Rényi–Gallager center of* W *for the Lagrange multiplier* λ, *satisfying*

$$C^{g\lambda}_{\alpha,W} = \sup_{p \in \mathscr{P}(\mathscr{X})} D_\alpha(p \circledast W e^{\frac{1-\alpha}{\alpha}\lambda\cdot\rho} \,\|\, p \otimes q^{g\lambda}_{\alpha,W}) \tag{104}$$

$$= \sup_{x \in \mathscr{X}} D_\alpha(W(x) \,\|\, q^{g\lambda}_{\alpha,W}) - \lambda \cdot \rho(x). \tag{105}$$

Furthermore, for every sequence of input distributions $\{p^{(i)}\}_{i\in\mathbb{Z}_+} \subset \mathscr{P}(\mathscr{X})$ *with* $\lim_{i\to\infty} I^{g\lambda}_\alpha(p^{(i)}; W) = C^{g\lambda}_{\alpha,W}$, *the corresponding sequence of the order* α *Rényi–Gallager means* $\{q^{g\lambda}_{\alpha,p^{(i)}}\}_{i\in\mathbb{Z}_+}$ *is a Cauchy sequence for the total variation metric on* $\mathcal{P}(\mathcal{Y})$ *and* $q^{g\lambda}_{\alpha,W}$ *is the unique limit point of that Cauchy sequence.*

The proof of Theorem 3 is very similar to the proofs of Theorem 1 and Theorem 2. It relies on Lemma 34, given in the following, instead of Lemma 19 or Lemma 30.

Lemma 34. *For any* $\alpha \in \mathbb{R}_+ \setminus \{1\}$, *channel* $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$ *with cost function* $\rho\colon \mathscr{X} \to \mathbb{R}^\ell_{\geq 0}$ *for a finite input set* $\mathscr{X}$, *and Lagrange multiplier* $\lambda \in \mathbb{R}^\ell_{\geq 0}$, *there exists a* $\widetilde{p} \in \mathscr{P}(\mathscr{X})$ *such that* $I^\lambda_\alpha(\widetilde{p}; W) = C^\lambda_{\alpha,W}$ *and* $\exists!\, q^\lambda_{\alpha,W} \in \mathcal{P}(\mathcal{Y})$ *satisfying*

$$D_\alpha(p \circledast W e^{\frac{1-\alpha}{\alpha}\lambda\cdot\rho} \,\|\, p \otimes q^{g\lambda}_{\alpha,W}) \leq C^{g\lambda}_{\alpha,W}, \quad \forall p \in \mathscr{P}(\mathscr{X}). \tag{106}$$

Furthermore, $q^{g\lambda}_{\alpha,\widetilde{p}} = q^{g\lambda}_{\alpha,W}$ *for all* $\widetilde{p} \in \mathscr{P}(\mathscr{X})$ *such that* $I^{g\lambda}_\alpha(\widetilde{p}; W) = C^{g\lambda}_{\alpha,W}$.

The expression on the left-hand side of (102) is the R-G capacity, whereas the expression in (103) is the A-L radius defined in (84). Thus, Theorems 2 and 3 imply that

$$C^\lambda_{\alpha,W} = S^\lambda_{\alpha,W} = C^{g\lambda}_{\alpha,W}, \quad \forall \alpha \in \mathbb{R}_+,\ \lambda \in \mathbb{R}^\ell_{\geq 0}. \tag{107}$$

Furthermore, whenever $C^\lambda_{\alpha,W}$ is finite the unique A-L center described in (82) is equal to the unique R-G center described in (105) by Theorems 2 and 3, as well:

$$q^\lambda_{\alpha,W} = q^{g\lambda}_{\alpha,W}, \quad \forall \alpha \in \mathbb{R}_+,\ \lambda \in \mathbb{R}^\ell_{\geq 0} \text{ such that } C^\lambda_{\alpha,W} < \infty. \tag{108}$$

In order to avoid using multiple names for the same quantity, we will state our propositions in terms of the A-L capacity and center in the rest of the paper.

If $C^\lambda_{\alpha,W}$ is finite, then (90), (91), and Theorem 3 for $\alpha \in \mathbb{R}_+ \setminus \{1\}$ and (93), (94), and Theorem 2 for $\alpha = 1$ imply that

$$C^\lambda_{\alpha,W} - I^{g\lambda}_\alpha(p; W) \geq D_\alpha(q^{g\lambda}_{\alpha,p} \,\|\, q^\lambda_{\alpha,W}), \quad \forall p \in \mathscr{P}(\mathscr{X}).$$

Using the same observations, we can prove a van Erven–Harremoës bound for the A-L capacity, as well.

Lemma 35. *For any* $\alpha \in \mathbb{R}_+$, *channel* $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$ *with a cost function* $\rho\colon \mathscr{X} \to \mathbb{R}^\ell_{\geq 0}$, *and Lagrange multiplier* $\lambda \in \mathbb{R}^\ell_{\geq 0}$ *satisfying* $C^\lambda_{\alpha,W} < \infty$,

$$\sup_{x \in \mathscr{X}} D_\alpha(W(x) \,\|\, q) - \lambda \cdot \rho(x) \geq C^\lambda_{\alpha,W} + D_\alpha(q^\lambda_{\alpha,W} \,\|\, q), \quad \forall q \in \mathcal{P}(\mathcal{Y}). \tag{109}$$

One can prove a similar, but weaker, result using Lemma 13 and Theorem 2. The right most term of the resulting bound is $D_{\alpha\wedge 1}(q^\lambda_{\alpha,W} \,\|\, q)$ rather than $D_\alpha(q^\lambda_{\alpha,W} \,\|\, q)$.

Lemma 35 and the continuity of the A-L capacity $C^\lambda_{\alpha,W}$ as a function of λ, established in Lemma 29(a), imply the continuity of the A-L center $q^\lambda_{\alpha,W}$ in λ for the total variation topology on $\mathcal{P}(\mathcal{Y})$ via Lemma 2.

Lemma 36. *For any $\alpha \in \mathbb{R}_+$, channel $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$ with a cost function $\rho\colon \mathscr{X} \to \mathbb{R}^{\ell}_{\geq 0}$, and Lagrange multiplier $\lambda_0 \in \mathbb{R}^{\ell}_{\geq 0}$ satisfying $C^{\lambda_0}_{\alpha,W} < \infty$,*

$$D_\alpha(q^{\lambda_2}_{\alpha,W} \,\|\, q^{\lambda_1}_{\alpha,W}) \leq C^{\lambda_1}_{\alpha,W} - C^{\lambda_2}_{\alpha,W}, \quad \forall \lambda_1, \lambda_2 \in \mathbb{R}^{\ell}_{\geq 0} \text{ such that } \lambda_0 \leq \lambda_1 \leq \lambda_2. \tag{110}$$

Furthermore, $q^{\lambda}_{\alpha,W}$ is continuous in λ on $\{\lambda : \exists \varepsilon > 0$ such that $C^{\lambda-\varepsilon\mathbb{1}}_{\alpha,W} < \infty\}$ for the total variation topology on $\mathcal{P}(\mathcal{Y})$.

5.4. Information Measures for Transition Probabilities

We have defined the conditional Rényi divergence, the Augustin information, the A-L information, and the R-G information, only for input distributions in $\mathscr{P}(\mathscr{X})$, i.e., for probability mass functions that are zero in all but finite number of elements of $\mathscr{X}$. In many practically relevant and analytically interesting models, however, the input set $\mathscr{X}$ is an uncountably infinite set equipped with a σ-algebra $\mathcal{X}$. The Gaussian channels—possibly with multiple input and output antennas and fading—and the Poisson channels are among the most prominent examples of such models. For such models, it is often desirable to extend the definitions of the Augustin information and the A-L information from $\mathscr{P}(\mathscr{X})$ to $\mathcal{P}(\mathcal{X})$. For instance, in the additive Gaussian channels described in Examples 4 and 5, the equality $I_\alpha(p; W) = C_{\alpha,W,\varrho}$ is not satisfied by any probability mass function p satisfying the cost constraint; but it is satisfied by the zero mean Gaussian distribution with variance ϱ.

In the following, we will first show that if $\mathcal{Y}$ is a countably generated σ-algebra, then one can generalize the definitions of the conditional Rényi divergence, the Augustin information, and the A-L information from $\mathscr{P}(\mathscr{X})$ to $\mathcal{P}(\mathcal{X})$ provided that W and Q are not only functions from $\mathscr{X}$ to $\mathcal{P}(\mathcal{Y})$, but also transition probabilities from $(\mathscr{X}, \mathcal{X})$ to $(\mathscr{Y}, \mathcal{Y})$. After that we will show that if in addition $\mathcal{X}$ is countably separated, then the supremum of A-L information $I^{\lambda}_{\alpha}(p; W)$ over $\mathcal{P}(\mathcal{X})$ is equal to the A-L radius $S^{\lambda}_{\alpha,W}$; see Theorem 4. This will imply that the cost constrained Augustin capacity $C_{\alpha,W,\varrho}$—defined in (68)—is equal to the supremum of the Augustin information $I_\alpha(p; W)$ over members of $\mathcal{P}(\mathcal{X})$ satisfying $\mathbf{E}_p[\rho] \leq \varrho$, as well, at least for the cost constraints that are in the interior of the set of all feasible constraints; see Theorem 5.

Let us first recall the definition of transition probability. We adopt the definition provided by Bogachev [21, Definition 10.7.1] with a minor modification: we use $W(\mathscr{E} \,|\, x)$ instead of $W(x \,|\, \mathscr{E})$.

Definition 19. Let $(\mathscr{X}, \mathcal{X})$ and $(\mathscr{Y}, \mathcal{Y})$ be measurable spaces. Then a function $W\colon \mathcal{Y} \times \mathscr{X} \to [0, 1]$ is called a transition probability (a stochastic kernel / a Markov kernel) from $(\mathscr{X}, \mathcal{X})$ to $(\mathscr{Y}, \mathcal{Y})$ if it satisfies the following two conditions:

(i) For all $x \in \mathscr{X}$, the function $W(\cdot \,|\, x)\colon \mathcal{Y} \to [0, 1]$ is a probability measure on $(\mathscr{Y}, \mathcal{Y})$.
(ii) For all $\mathscr{E} \in \mathcal{Y}$, the function $W(\mathscr{E} \,|\, \cdot)\colon \mathscr{X} \to [0, 1]$ is an $(\mathcal{X}, \mathcal{B}([0, 1]))$-measurable function.

We denote the set of all transition probabilities from $(\mathscr{X}, \mathcal{X})$ to $(\mathscr{Y}, \mathcal{Y})$ by $\mathcal{P}(\mathcal{Y} \,|\, \mathcal{X})$ with the tacit understanding that $\mathscr{X}$ and $\mathscr{Y}$ will be clear from the context. If W satisfies (i), then $W\colon \mathscr{X} \to \mathcal{P}(\mathcal{Y})$ is a channel, i.e., W is a member of $\mathcal{P}(\mathcal{Y} \,|\, \mathscr{X})$, even if W does not satisfy (ii). Hence, $\mathcal{P}(\mathcal{Y} \,|\, \mathcal{X}) \subset \mathcal{P}(\mathcal{Y} \,|\, \mathscr{X})$. Inspired by this observation, we denote the probability measure $W(\cdot \,|\, x)$ by $W(x)$ whenever it is notationally convenient and unambiguous.

In order to extend the definition of the conditional Rényi divergence from $\mathscr{P}(\mathscr{X})$ to $\mathcal{P}(\mathcal{X})$, we ensure the $\mathcal{X}$-measurability of $D_\alpha(W(x) \,\|\, Q(x))$ on $\mathscr{X}$ and replace the sum in (21) with an integral. If $(\mathscr{X}, \tau)$ is a topological space and $\mathcal{X}$ is the associated Borel σ-algebra, then one can establish the measurability by first establishing the continuity. Such a continuity result holds if both $\frac{dW(x)}{d\nu}$ and $\frac{dQ(x)}{d\nu}$ are continuous in x for ν-almost every y for some probability measure ν for which $(W(x) + Q(x)) \prec \nu$ for all $x \in \mathscr{X}$. At times this hypothesis on W and Q might not be easy to confirm. If, on the other hand, W and Q are transition probabilities from $(\mathscr{X},\mathcal{X})$ to $(\mathscr{Y},\mathcal{Y})$ for a countably generated $\mathcal{Y}$, then the desired measurability follows from the elementary properties of the measurable functions and Lemma 9, as we demonstrate in the following.

Lemma 37. *For any $\alpha\in\mathbb{R}_+$, countable generated σ-algebra $\mathcal{Y}$ of subsets of $\mathscr{Y}$, and $W,Q\in\mathcal{P}(\mathcal{Y}\,|\,\mathcal{X})$ the function $D_\alpha(W(\cdot)\,\|\,Q(\cdot))\colon\mathscr{X}\to[0,\infty]$ is $\mathcal{X}$-measurable.*

Proof. There exists $\{\mathscr{E}_i\}_{i\in\mathbb{Z}_+}\subset\mathcal{Y}$ such that $\mathcal{Y}=\sigma(\{\mathscr{E}_i:\ i\in\mathbb{Z}_+\})$, because $\mathcal{Y}$ is countably generated σ-algebra. Let $\mathcal{Y}_i$ be

$$\mathcal{Y}_i\triangleq\sigma(\{\mathscr{E}_1,\ldots,\mathscr{E}_i\}),\quad i\in\mathbb{Z}_+.$$

Then $\mathcal{Y}_1\subset\mathcal{Y}_2\subset\ldots\subset\mathcal{Y}$, $\mathcal{Y}=\sigma\Big(\bigcup\limits_{i=1}^{\infty}\mathcal{Y}_i\Big)$, and Lemma 9 implies that

$$D_\alpha(W(x)\,\|\,Q(x))=\lim_{i\to\infty}D_\alpha^{\mathcal{Y}_i}(W(x)\,\|\,Q(x)),\quad\forall x\in\mathscr{X}.\tag{111}$$

On the other hand $\mathcal{Y}_i$ is finite set for all $i\in\mathbb{Z}_+$. Thus, for all $i\in\mathbb{Z}_+$ there exists a $\mathcal{Y}_i$-measurable finite partition $\mathcal{E}_i$ of $\mathscr{Y}$. Thus, as a result of the definition of the Rényi divergence given in (8) we have

$$D_\alpha^{\mathcal{Y}_i}(W(x)\,\|\,Q(x))=\begin{cases}\dfrac{1}{\alpha-1}\ln\displaystyle\sum_{\mathscr{E}\in\mathcal{E}_i}(W(\mathscr{E}\,|\,x))^{\alpha}\,(Q(\mathscr{E}\,|\,x))^{1-\alpha} & \text{for }\alpha\in\mathbb{R}_+\setminus\{1\},\\ \displaystyle\sum_{\mathscr{E}\in\mathcal{E}_i}W(\mathscr{E}\,|\,x)\ln\frac{W(\mathscr{E}\,|\,x)}{Q(\mathscr{E}\,|\,x)} & \text{for }\alpha=1.\end{cases}$$

Then $D_\alpha^{\mathcal{Y}_i}(W(x)\,\|\,Q(x))$ is an $\mathcal{X}$-measurable function for any $i\in\mathbb{Z}_+$ by [21, Theorem 2.1.5(i)-(iv) and Remark 2.1.6], because $W(\mathscr{E}\,|\,x)$ and $Q(\mathscr{E}\,|\,x)$ are $\mathcal{X}$-measurable for all $\mathscr{E}\in\mathcal{E}_i$ by the hypothesis of the lemma. Then $D_\alpha(W(x)\,\|\,Q(x))$ is $\mathcal{X}$-measurable as a result of (111) by [21, Theorem 2.1.5(v) and Remark 2.1.6]. △

Definition 20. For any $\alpha\in\mathbb{R}_+$, countable generated σ-algebra $\mathcal{Y}$ of subsets of $\mathscr{Y}$, $W\in\mathcal{P}(\mathcal{Y}\,|\,\mathcal{X})$, and $p\in\mathcal{P}(\mathcal{X})$ the *order α conditional Rényi divergence for the input distribution p* is

$$D_\alpha(W\,\|\,Q\,|\,p)\triangleq\int D_\alpha(W(x)\,\|\,Q(x))p(dx).\tag{112}$$

If $\exists q\in\mathcal{P}(\mathcal{Y})$ such that $Q(x)=q$ for p-a.s., then we denote $D_\alpha(W\,\|\,Q\,|\,p)$ by $D_\alpha(W\,\|\,q\,|\,p)$.

Then one can define the Augustin information and the A-L information for all p in $\mathcal{P}(\mathcal{X})$, provided that W is in $\mathcal{P}(\mathcal{Y}\,|\,\mathcal{X})$ for a countably generated $\mathcal{Y}$ and ρ is an $\mathcal{X}$-measurable function.

Definition 21. For any $\alpha\in\mathbb{R}_+$, countable generated σ-algebra $\mathcal{Y}$ of subsets of $\mathscr{Y}$, $W\in\mathcal{P}(\mathcal{Y}\,|\,\mathcal{X})$, and $p\in\mathcal{P}(\mathcal{X})$ the *order α Augustin information for the input distribution p* is

$$I_\alpha(p;W)\triangleq\inf_{q\in\mathcal{P}(\mathcal{Y})}D_\alpha(W\,\|\,q\,|\,p).\tag{113}$$

Furthermore, for any $\mathcal{X}$-measurable cost function $\rho\colon\mathscr{X}\to\mathbb{R}^{\ell}_{\ge0}$ and $\lambda\in\mathbb{R}^{\ell}_{\ge0}$ the *order α Augustin–Legendre information for the input distribution p and the Lagrange multiplier λ* is defined as

$$I_\alpha^{\lambda}(p;W)\triangleq I_\alpha(p;W)-\lambda\cdot\mathbf{E}_p[\rho]\tag{114}$$

with the understanding that if $\lambda\cdot\mathbf{E}_p[\rho]=\infty$, then $I_\alpha^{\lambda}(p;W)=-\infty$.

Although we have included the $\lambda\cdot\mathbf{E}_p[\rho]=\infty$ case in the formal definition of the A-L information, we will only be interested in p's for which $\lambda\cdot\mathbf{E}_p[\rho]$ is finite. We define $\mathcal{A}^{\lambda}$ to be the set of all such p's:

$$\mathcal{A}^{\lambda}\triangleq\{p\in\mathcal{P}(\mathcal{X}):\ \lambda\cdot\mathbf{E}_p[\rho]<\infty\}.\tag{115}$$

For an arbitrary σ-algebra $\mathcal{X}$, the singletons (i.e., sets with only one element) are not necessarily measurable sets; thus, $\mathscr{P}(\mathscr{X})$ is not necessarily a subset of $\mathcal{A}^{\lambda}$. If $\mathcal{X}$ is countably separated, then the singletons are in $\mathcal{X}$ by [21, Theorem 6.5.7], $\mathscr{P}(\mathscr{X}) \subset \mathcal{A}^{\lambda}$ and $\sup_{p\in\mathcal{A}^{\lambda}} I_{\alpha}^{\lambda}(p;W) \geq C_{\alpha,W}^{\lambda}$. The reverse inequality follows from Theorem 2 and we have $\sup_{p\in\mathcal{A}^{\lambda}} I_{\alpha}^{\lambda}(p;W) = C_{\alpha,W}^{\lambda}$. Theorem 4 states these observations formally together with the ones about the A-L center through a minimax theorem.

Theorem 4. *Let $\mathcal{X}$ be a countably separated σ-algebra, $\mathcal{Y}$ a countably generated σ-algebra, W a transition probability from $(\mathscr{X},\mathcal{X})$ to $(\mathscr{Y},\mathcal{Y})$, $\rho\colon \mathscr{X}\to\mathbb{R}_{\geq 0}^{\ell}$ an $\mathcal{X}$-measurable cost function, and $\alpha\in\mathbb{R}_{+}$. Then for all $\lambda\in\mathbb{R}_{\geq 0}^{\ell}$ we have*

$$\sup_{p\in\mathcal{A}^{\lambda}} \inf_{q\in\mathcal{P}(\mathcal{Y})} D_{\alpha}(W\,\|\,q\,|\,p) - \lambda\cdot\mathbf{E}_{p}[\rho] = \inf_{q\in\mathcal{P}(\mathcal{Y})} \sup_{p\in\mathcal{A}^{\lambda}} D_{\alpha}(W\,\|\,q\,|\,p) - \lambda\cdot\mathbf{E}_{p}[\rho] \tag{116}$$

$$= \inf_{q\in\mathcal{P}(\mathcal{Y})} \sup_{x\in\mathscr{X}} D_{\alpha}(W(x)\,\|\,q) - \lambda\cdot\rho(x) \tag{117}$$

$$= C_{\alpha,W}^{\lambda} \tag{118}$$

where $\mathcal{A}^{\lambda}$ is defined in (115). *If $C_{\alpha,W}^{\lambda}$ is finite, then $\exists!\, q_{\alpha,W}^{\lambda}\in\mathcal{P}(\mathcal{Y})$, called the order α Augustin–Legendre center of W for the Lagrange multiplier λ, satisfying*

$$C_{\alpha,W}^{\lambda} = \sup_{p\in\mathcal{A}^{\lambda}} D_{\alpha}(W\,\|\,q_{\alpha,W}^{\lambda}\,|\,p) - \lambda\cdot\mathbf{E}_{p}[\rho] \tag{119}$$

$$= \sup_{x\in\mathscr{X}} D_{\alpha}(W(x)\,\|\,q_{\alpha,W}^{\lambda}) - \lambda\cdot\rho(x). \tag{120}$$

Proof. Since $\mathscr{P}(\mathscr{X})\subset\mathcal{A}^{\lambda}$, the max-min inequality implies

$$\begin{aligned}\sup_{p\in\mathscr{P}(\mathscr{X})} \inf_{q\in\mathcal{P}(\mathcal{Y})} D_{\alpha}(W\,\|\,q\,|\,p) - \lambda\cdot\mathbf{E}_{p}[\rho] &\leq \sup_{p\in\mathcal{A}^{\lambda}} \inf_{q\in\mathcal{P}(\mathcal{Y})} D_{\alpha}(W\,\|\,q\,|\,p) - \lambda\cdot\mathbf{E}_{p}[\rho]\\ &\leq \inf_{q\in\mathcal{P}(\mathcal{Y})} \sup_{p\in\mathcal{A}^{\lambda}} D_{\alpha}(W\,\|\,q\,|\,p) - \lambda\cdot\mathbf{E}_{p}[\rho]\\ &= \inf_{q\in\mathcal{P}(\mathcal{Y})} \sup_{x\in\mathscr{X}} D_{\alpha}(W(x)\,\|\,q) - \lambda\cdot\rho(x).\end{aligned}$$

Thus, (116) and (117) hold as a result of (79) and (80) of Theorem 2 and (118) follows by (80) of Theorem 2 and (78).

If $C_{\alpha,W}^{\lambda}$ is finite, then as a result of Theorem 2 there exist a unique $q_{\alpha,W}^{\lambda}\in\mathscr{P}(\mathcal{Y})$ satisfying

$$\sup_{x\in\mathscr{X}} D_{\alpha}(W(x)\,\|\,q_{\alpha,W}^{\lambda}) - \lambda\cdot\rho(x) = C_{\alpha,W}^{\lambda}.$$

Then (119) and (120) hold, because $\sup_{p\in\mathcal{A}^{\lambda}} D_{\alpha}(W\,\|\,q\,|\,p) - \lambda\cdot\mathbf{E}_{p}[\rho] = \sup_{x\in\mathscr{X}} D_{\alpha}(W(x)\,\|\,q) - \lambda\cdot\rho(x)$ for any $q\in\mathscr{P}(\mathcal{Y})$. △

Let $\mathcal{A}(\varrho)$ be the subset $\mathcal{P}(\mathcal{X})$ composed of the probability measures satisfying the cost constraint ϱ,

$$\mathcal{A}(\varrho) \triangleq \{p\in\mathcal{P}(\mathcal{X}):\ \mathbf{E}_{p}[\rho]\leq\varrho\}.$$

Then $\mathscr{A}(\varrho)\subset\mathcal{A}(\varrho)$ and $\sup_{p\in\mathcal{A}(\varrho)} I_{\alpha}(p;W)\geq C_{\alpha,W,\varrho}$ whenever $\mathscr{X}$ is countably separated. For the cost constraints in $\operatorname{int}\Gamma_{\rho}$ reverse inequality holds as a result of Lemma 29(c) and Theorem 4 and we have $\sup_{p\in\mathcal{A}(\varrho)} I_{\alpha}(p;W) = C_{\alpha,W,\varrho}$. Theorem 5 states these observations formally together with the ones about the Augustin center through a minimax theorem.

Theorem 5. *Let $\mathcal{X}$ be a countably separated σ-algebra, $\mathcal{Y}$ be a countably generated σ-algebra, W be a transition probability from $(\mathscr{X},\mathcal{X})$ to $(\mathscr{Y},\mathcal{Y})$, $\rho\colon\mathscr{X}\to\mathbb{R}_{\geq 0}^{\ell}$ be an $\mathcal{X}$-measurable cost* *function, and* $\alpha \in \mathbb{R}_+$. *For any* $\varrho \in \operatorname{int}\Gamma_\rho$ *we have*

$$\sup_{p\in\mathcal{A}(\varrho)} \inf_{q\in\mathcal{P}(\mathcal{Y})} D_\alpha(W\,\|\,q\,|\,p) = \inf_{q\in\mathcal{P}(\mathcal{Y})} \sup_{p\in\mathcal{A}(\varrho)} D_\alpha(W\,\|\,q\,|\,p) \tag{121}$$

$$= C_{\alpha,W,\varrho}, \tag{122}$$

where $C_{\alpha,W,\varrho}$ *is defined in* (68). *If* $C_{\alpha,W,\varrho} \in \mathbb{R}_{\geq 0}$, *then* $\exists!\, q_{\alpha,W,\varrho} \in \mathcal{P}(\mathcal{Y})$, *called the order* α *Augustin center of* W *for the cost constraint* ϱ, *satisfying*

$$C_{\alpha,W,\varrho} = \sup_{p\in\mathcal{A}(\varrho)} D_\alpha(W\,\|\,q_{\alpha,W,\varrho}\,|\,p) \tag{123}$$

$$= \sup_{p\in\mathscr{A}(\varrho)} D_\alpha(W\,\|\,q_{\alpha,W,\varrho}\,|\,p). \tag{124}$$

Furthermore, $q_{\alpha,W,\varrho} = q^\lambda_{\alpha,W}$ *for all* $\lambda \in \mathbb{R}^\ell_{\geq 0}$ *satisfying* $C_{\alpha,W,\varrho} = C^\lambda_{\alpha,W} + \lambda\cdot\varrho$.

Proof. Since $\mathscr{A}(\varrho) \subset \mathcal{A}(\varrho)$, the max-min inequality implies

$$\sup_{p\in\mathscr{A}(\varrho)} \inf_{q\in\mathcal{P}(\mathcal{Y})} D_\alpha(W\,\|\,q\,|\,p) \leq \sup_{p\in\mathcal{A}(\varrho)} \inf_{q\in\mathcal{P}(\mathcal{Y})} D_\alpha(W\,\|\,q\,|\,p) \leq \inf_{q\in\mathcal{P}(\mathcal{Y})} \sup_{p\in\mathcal{A}(\varrho)} D_\alpha(W\,\|\,q\,|\,p).$$

Thus, both (121) and (122) hold whenever $C_{\alpha,W,\varrho} = \infty$ by (58). On the other hand, as a result of Theorem 4 for any λ with finite $C^\lambda_{\alpha,W}$ there exists a unique $q^\lambda_{\alpha,W}$ satisfying (120). Thus, we have

$$\begin{aligned}\inf_{q\in\mathcal{P}(\mathcal{Y})} \sup_{p\in\mathcal{A}(\varrho)} D_\alpha(W\,\|\,q\,|\,p) &\leq \sup_{p\in\mathcal{A}(\varrho)} D_\alpha(W\,\|\,q^\lambda_{\alpha,W}\,|\,p)\\ &\leq \sup_{p\in\mathcal{A}(\varrho)} D_\alpha(W\,\|\,q^\lambda_{\alpha,W}\,|\,p) - \lambda\cdot\mathbf{E}_p[\rho] + \lambda\cdot\varrho\\ &\leq C^\lambda_{\alpha,W} + \lambda\cdot\varrho.\end{aligned}$$

Furthermore, if $C_{\alpha,W,\varrho} \in \mathbb{R}$, then there exists at least one $\lambda \in \mathbb{R}^\ell_{\geq 0}$ satisfying $C_{\alpha,W,\varrho} = C^\lambda_{\alpha,W} + \lambda\cdot\varrho$ by Lemma 29(c). Then (121) and (122) hold when $C_{\alpha,W,\varrho} \in \mathbb{R}$ and (123) holds for $q_{\alpha,W,\varrho} = q^\lambda_{\alpha,W}$ provided that $C_{\alpha,W,\varrho} = C^\lambda_{\alpha,W} + \lambda\cdot\varrho$. On the other hand, $q_{\alpha,W,\varrho}$ is a probability measure satisfying (124) by Theorem 1 and $q_{\alpha,W,\varrho} = q^\lambda_{\alpha,W}$ for all λ satisfying $C_{\alpha,W,\varrho} = C^\lambda_{\alpha,W} + \lambda\cdot\varrho$ by Lemma 31. △

The countable separability of $\mathcal{X}$ and countable generatedness of $\mathcal{Y}$ are fairly mild assumptions satisfied by most transition probabilities considered in practice. Hence, Theorems 4 and 5 provide further justification for studying the relatively simple case of probability mass functions, first.

The existence of an input distribution p satisfying both $\mathbf{E}_p[\rho] \leq \varrho$ and $I_\alpha(p;W) = C_{\alpha,W,\varrho}$ is immaterial to the existence of a unique $q_{\alpha,W,\varrho}$ or its characterization through $q^\lambda_{\alpha,W}$ for λ's satisfying $C_{\alpha,W,\varrho} = C^\lambda_{\alpha,W} + \lambda\cdot\varrho$ by Lemma 29(c),(d) and Theorem 5. Although one can prove the existence of such a p for certain special cases, such an input distribution does not exist in general. Thus, we believe, it is preferable to separate the issue of the existence of an optimal input distribution from the discussion of $C_{\alpha,W,\varrho}$ and $q_{\alpha,W,\varrho}$ and their characterization via $C^\lambda_{\alpha,W}$ and $q^\lambda_{\alpha,W}$. That, however, is not the standard practice, [37, Theorem 1].

We have assumed $\mathcal{Y}$ to be countably generated in order to ensure that the conditional Rényi divergence used in (113) is well-defined. In order to define the Rényi information, however, we do not need to assume $\mathcal{Y}$ to be countably generated; the transition probability structure is sufficient. Recall that if $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$, then for any $p \in \mathcal{P}(\mathcal{X})$ there exists a unique probability measure $p \circledast W$ on $(\mathscr{X}\times\mathscr{Y}, \mathcal{X}\otimes\mathcal{Y})$ such that

$$p \circledast W(\mathscr{E}_x \times \mathscr{E}_y) = \int_{\mathscr{E}_x} W(\mathscr{E}_y\,|\,x)p(dx), \quad \forall \mathscr{E}_x \in \mathcal{X},\ \mathscr{E}_y \in \mathcal{Y},$$

by [21, Theorem 10.7.2]. Thus, $I_{\alpha}^{\mathrm{g}}(p;W)$ is well defined for any $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ and $p \in \mathcal{P}(\mathcal{X})$.

Unfortunately, the situation is not nearly as simple for the R-G information. In order to define the R-G information using a similar approach one first shows that $We^{\frac{1-\alpha}{\alpha}\lambda\cdot\rho}$ is a transition kernel—rather than a transition probability (i.e., Markov kernel)—and then proceeds with establishing the existence a unique measure $p \circledast We^{\frac{1-\alpha}{\alpha}\lambda\cdot\rho}$ for all p in $\mathcal{P}(\mathcal{Y})$. For orders greater than one, resulting measure is a sub-probability measure and one can use (87) as the definition of the R-G information. For orders between zero and one, on the other hand, $p \circledast We^{\frac{1-\alpha}{\alpha}\lambda\cdot\rho}$ is a σ-finite measure for all p's in $\mathcal{P}(\mathcal{X})$, but it is not necessarily a finite measure for all p's in $\mathcal{P}(\mathcal{X})$. Thus, for orders between zero and one, one can use (87) as the definition of the R-G information, only after extending the definition of the Rényi divergence to σ-finite measures.

6. EXAMPLES

In this section, we will first demonstrate certain subtleties that we have pointed out in the earlier sections. After that we will study Gaussian channels and obtain closed form expressions for their Augustin capacity and center.

6.1. Shift Invariant Families

Example 1 (channel with an affine capacity). Let the channel $W\colon \mathbb{R}_{\geq 0} \to \mathcal{P}(\mathcal{B}([0,1)))$ and the associated cost function $\rho\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be

$$\frac{dW(x)}{d\nu} = f_{\lfloor x \rfloor}(y - x - \lfloor y - x \rfloor),$$
$$\rho(x) = \lfloor x \rfloor$$

where ν is the Lebesgue measure on $[0,1)$ and f_i's are given by

$$f_i(y) = e^{i+1}\mathbb{1}_{y \in [0, e^{-i-1})}, \quad \forall i \in \mathbb{Z}_{\geq 0}.$$

Let u_i be uniform distribution on $[i, i+1)$; then one can confirm by substitution that $\mathrm{T}_{\alpha,u_i}(u_0) = u_0$. Then, using Jensen's inequality together with the fixed point property, we get[21]

$$D_{\alpha}(W \,\|\, q \,|\, u_i) \geq D_{\alpha}(W \,\|\, u_0 \,|\, u_i) + D_{\alpha\wedge 1}(u_0 \,\|\, q).$$

Thus, u_0 is the unique order α Augustin mean for the input distribution u_i, i.e., $q_{\alpha,u_i} = u_0$, and $I_{\alpha}(u_i;W) = D_{\alpha}(W \,\|\, u_0 \,|\, u_i)$—and hence $I_{\alpha}(u_i;W) = i+1$—for all $i \in \mathbb{Z}_{+}$ and $\alpha \in \mathbb{R}_{+}$. Then using $\mathbf{E}_{u_i}[\rho] = i$, we can conclude that $C_{\alpha,W,\varrho} \geq (\varrho+1)$ not only for $\varrho \in \mathbb{Z}_{\geq 0}$ but also for $\varrho \in \mathbb{R}_{\geq 0}$, because $C_{\alpha,W,\varrho}$ is concave in ϱ by Lemma 27(a). One the other hand, one can confirm by substitution that

$$D_{\alpha}(W \,\|\, u_0 \,|\, p) = \mathbf{E}_p[\rho] + 1. \tag{125}$$

Thus, $I_{\alpha}(p;W) \leq (\varrho + 1)$ for any p satisfying the cost constraint ϱ. Hence,

$$C_{\alpha,W,\varrho} = \varrho + 1,$$
$$q_{\alpha,W,\varrho} = u_0.$$

Then as a result of (76) we have

$$C_{\alpha,W}^{\lambda} = \begin{cases} \infty & \text{for } \lambda \in [0,1), \\ 1 & \text{for } \lambda \in [1,\infty). \end{cases}$$

Then using (125) and Theorem 4, we can conclude that $q_{\alpha,W}^{\lambda} = u_0$ for all $\lambda \in [1,\infty)$.

[21] See the derivation of (32) and (34) of Lemma 13(c),(d) given in [22, Appendix B].

Example 2 (channel with a non-upper semicontinuous capacity). Let $W\colon \mathbb{R} \to \mathcal{P}(\mathcal{B}([0,1)))$ and the associated cost function $\rho\colon \mathbb{R} \to \mathbb{R}_{\geq 0}$ be

$$\frac{dW(x)}{d\nu} = f_{\lfloor x \rfloor}(y - x - \lfloor y - x \rfloor),$$
$$\rho(x) = \begin{cases} \lfloor x \rfloor & \text{for } x \geq 0, \\ 2^{\lfloor x \rfloor} & \text{for } x < 0, \end{cases}$$

where ν is the Lebesgue measure on $[0,1)$ and $f_i\colon [0,1) \to \mathbb{R}_{\geq 0}$ are given by

$$f_i(y) = \begin{cases} 2^{i+1}\mathbb{1}_{y \in [0, 2^{-i-1})} & \text{for } i > 0, \\ 3/2\mathbb{1}_{y \in [0, 2/3)} & \text{for } i = 0, \\ 2\mathbb{1}_{y \in [0, 1/2)} & \text{for } i < 0. \end{cases}$$

Following an analysis similar to the one described above, we can conclude that

$$C_{\alpha,W,\varrho} = \begin{cases} (\varrho + 1)\ln 2 & \text{for } \varrho > 0, \\ \ln 3/2 & \text{for } \varrho = 0, \end{cases}$$
$$C^{\lambda}_{\alpha,W} = \begin{cases} \infty & \text{for } \lambda \in [0, \ln 2), \\ \ln 2 & \text{for } \lambda \in [\ln 2, \infty). \end{cases}$$

Hence, $C_{\alpha,W,\varrho} \neq \inf\limits_{\lambda \geq 0} C^{\lambda}_{\alpha,W} + \lambda \cdot \varrho$ for $\varrho = 0$.

Example 3 (product channel without an optimal cost allocation). Let W_1 and W_2 be the channels described in Examples 1 and 2 and ρ_1 and ρ_2 be the associated cost functions. Let $W_{[1,2]}$ be the product of these two channels with the additive cost function $\varrho_{[1,2]}$, i.e.,

$$W_{[1,2]}(x_1, x_2) = W_1(x_1) \otimes W_2(x_2),$$
$$\rho_{[1,2]}(x_1, x_2) = \rho_1(x_1) + \rho_2(x_2).$$

Then Lemma 28 implies

$$C_{\alpha,W_{[1,2]},\varrho} = \begin{cases} \varrho + 1 + \ln 2 & \text{for } \varrho > 0, \\ 1 + \ln \dfrac{3}{2} & \text{for } \varrho = 0. \end{cases}$$

Note that for positive values of ϱ there does not exist any (ϱ_1, ϱ_2) pair satisfying both $C_{\alpha,W_{[1,2]},\varrho} = C_{\alpha,W_1,\varrho_1} + C_{\alpha,W_2,\varrho_2}$ and the cost constraint $\varrho_1 + \varrho_2 \leq \varrho$ at the same time.

6.2. Gaussian Channels

In the following, we denote the zero mean Gaussian probability measure on $\mathcal{B}(\mathbb{R})$ with variance σ^2 by φ_{σ^2}. With a slight abuse of notation, we denote the corresponding probability density function by the same symbol:

$$\varphi_{\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}, \quad \forall x \in \mathbb{R}.$$

We use the Gaussian channels and the corresponding transition probabilities interchangeably; they have the same cost constrained Augustin capacity and center by Theorems 4 and 5.

Example 4 (scalar Gaussian channel). Let W be the scalar Gaussian channel with noise variance σ^2 and the associated cost function ρ be the quadratic one, i.e.,

$$W(\mathscr{E}\,|\,x) = \int_{\mathscr{E}} \varphi_{\sigma^2}(y-x)dy, \quad \forall \mathscr{E} \in \mathcal{B}(\mathbb{R}),$$
$$\rho(x) = x^2, \quad \forall x \in \mathbb{R}.$$

The Augustin capacity and center of this channel are given by the following expressions:

$$C_{\alpha,W,\varrho} = \begin{cases} \dfrac{\alpha\varrho}{2(\alpha\theta_{\alpha,\sigma,\varrho}+(1-\alpha)\sigma^2)} + \dfrac{1}{\alpha-1}\ln\dfrac{(\theta_{\alpha,\sigma,\varrho})^{\alpha/2}\sigma^{(1-\alpha)}}{\sqrt{\alpha\theta_{\alpha,\sigma,\varrho}+(1-\alpha)\sigma^2}} & \text{for } \alpha \in \mathbb{R}_+ \setminus \{1\}, \\ \dfrac{1}{2}\ln\left(1+\dfrac{\varrho}{\sigma^2}\right) & \text{for } \alpha = 1, \end{cases} \tag{126}$$

$$q_{\alpha,W,\varrho} = \varphi_{\theta_{\alpha,\sigma,\varrho}}, \tag{127}$$

$$\theta_{\alpha,\sigma,\varrho} \triangleq \sigma^2 + \frac{\varrho}{2} - \frac{\sigma^2}{2\alpha} + \sqrt{(\frac{\varrho}{2}-\frac{\sigma^2}{2\alpha})^2 + \varrho\sigma^2}. \tag{128}$$

Furthermore, $C_{\alpha,W,\varrho}$ is continuously differentiable in ϱ and its derivative is a continuous, decreasing, and bijective function of ϱ from $\mathbb{R}_+$ to $[0, \alpha/2\sigma^2)$ given by

$$\frac{d}{d\varrho}C_{\alpha,W,\varrho} = \frac{\alpha}{2(\alpha\theta_{\alpha,\sigma,\varrho}+(1-\alpha)\sigma^2)} \tag{129}$$
$$= \frac{\alpha}{\alpha\varrho+\sigma^2+\sqrt{(\alpha\varrho-\sigma^2)^2+4\varrho\alpha^2\sigma^2}}. \tag{130}$$

In order to prove these, we first demonstrate that the Augustin mean for the zero mean Gaussian distribution with variance ϱ is the zero mean Gaussian distribution with variance $\theta_{\alpha,\sigma,\varrho}$, i.e., $q_{\alpha,\varphi_\varrho} = \varphi_{\theta_{\alpha,\sigma,\varrho}}$. This will imply $I_\alpha(\varphi_\varrho; W) = D_\alpha(W\,\|\,\varphi_{\theta_{\alpha,\sigma,\varrho}}\,|\,\varphi_\varrho)$. $D_\alpha(W\,\|\,\varphi_{\theta_{\alpha,\sigma,\varrho}}\,|\,\varphi_\varrho)$ is equal to the expression on the right-hand side of (126). In order to establish (126) and (127), we demonstrate that this value is the greatest value for the Augustin information among all input distributions satisfying the cost constraint ϱ. Consequently, we have $C_{\alpha,W,\varrho} = I_\alpha(\varphi_\varrho; W)$ and $q_{\alpha,W,\varrho} = q_{\alpha,\varphi_\varrho}$. Then we confirm (129) using an identity, i.e., (133), obtained while establishing $q_{\alpha,\varphi_\varrho} = \varphi_{\theta_{\alpha,\sigma,\varrho}}$.

One can confirm by substitution that

$$D_\alpha(W(x)\,\|\,\varphi_\theta) = \begin{cases} \dfrac{\alpha x^2}{2(\alpha\theta+(1-\alpha)\sigma^2)} + \dfrac{1}{\alpha-1}\ln\dfrac{\theta^{\alpha/2}\sigma^{(1-\alpha)}}{\sqrt{\alpha\theta+(1-\alpha)\sigma^2}} & \text{for } \alpha \in \mathbb{R}_+ \setminus \{1\}, \\ \dfrac{\sigma^2+x^2-\theta}{2\theta} + \dfrac{1}{2}\ln\dfrac{\theta}{\sigma^2} & \text{for } \alpha = 1. \end{cases} \tag{131}$$

Then the order α tilted channel $W_\alpha^{\varphi_\theta}$, defined in (22), is a Gaussian channel as well:

$$W_\alpha^{\varphi_\theta}(\mathscr{E}\,|\,x) = \int_{\mathscr{E}} \varphi_{\frac{\sigma^2\theta}{\alpha\theta+(1-\alpha)\sigma^2}}\left(y - \frac{\alpha\theta}{\alpha\theta+(1-\alpha)\sigma^2}x\right)dy.$$

Then $\mathrm{T}_{\alpha,p}(q)$ is a zero mean Gaussian probability measure whenever both p and q are so. In particular,

$$\mathrm{T}_{\alpha,\varphi_\varrho}(\varphi_\theta) = \varphi_{(\frac{\alpha\theta}{\alpha\theta+(1-\alpha)\sigma^2})^2\varrho + \frac{\sigma^2\theta}{\alpha\theta+(1-\alpha)\sigma^2}}. \tag{132}$$

Consequently, if φ_θ is a fixed point of $\mathrm{T}_{\alpha,\varphi_\varrho}(\cdot)$, then θ satisfies the following equality:

$$\theta\left[\theta^2 - \theta\Big(\varrho + \Big(2-\frac{1}{\alpha}\Big)\sigma^2\Big) + \Big(1-\frac{1}{\alpha}\Big)\sigma^4\right] = 0. \tag{133}$$

$\theta_{\alpha,\sigma,\varrho}$, defined in (128), is the only root of the equality given in (133) that is greater than σ^2 for α's in $\mathbb{R}_+$; it is the only positive root for α's in $(0,1)$, as well. Furthermore, using (132) one can confirm that $\mathrm{T}_{\alpha,\varphi_\varrho}(\varphi_{\theta^2_{\alpha,\sigma,\varrho}}) = \varphi_{\theta_{\alpha,\sigma,\varrho}}$, i.e., $\varphi_{\theta_{\alpha,\sigma,\varrho}}$ is a fixed point of $\mathrm{T}_{\alpha,\varphi_\varrho}(\cdot)$. Then, using Jensen's inequality together with this fixed point property, we get[22]

$$D_\alpha(W \,\|\, q \,|\, \varphi_\varrho) \geq D_\alpha(W \,\|\, \varphi_{\theta_{\alpha,\sigma,\varrho}} \,|\, \varphi_\varrho) + D_{1\wedge\alpha}(\varphi_{\theta_{\alpha,\sigma,\varrho}} \,\|\, q), \quad \forall q \in \mathcal{P}(\mathcal{B}(\mathbb{R})).$$

Thus, $\varphi_{\theta_{\alpha,\sigma,\varrho}}$ is the order α Augustin mean for the input distribution φ_ϱ, i.e., $q_{\alpha,\varphi_\varrho} = \varphi_{\theta_{\alpha,\sigma,\varrho}}$ and $I_\alpha(\varphi_\varrho; W) = D_\alpha(W \,\|\, \varphi_{\theta_{\alpha,\sigma,\varrho}} \,|\, \varphi_\varrho)$. On the other hand, (131) implies

$$D_\alpha(W \,\|\, \varphi_{\theta_{\alpha,\sigma,\varrho}} \,|\, p) = \frac{\alpha(\mathbf{E}_p[\rho] - \varrho)}{2(\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2)} + I_\alpha(\varphi_\varrho; W), \quad \forall p \in \mathcal{P}(\mathcal{B}(\mathbb{R})). \tag{134}$$

Then $I_\alpha(p; W) \leq I_\alpha(\varphi_\varrho; W)$ for all p satisfying $\mathbf{E}_p[\rho] \leq \varrho$. Consequently, $C_{\alpha,W,\varrho} = I_\alpha(\varphi_\varrho; W)$ and $q_{\alpha,W,\varrho} = q_{\alpha,\varphi_\varrho}$.

For the $\alpha = 1$ case, (129) is evident. In order to establish (129) for the $\alpha \in \mathbb{R}_+ \setminus \{1\}$ case, note that

$$\begin{aligned}
\frac{d}{d\varrho} C_{\alpha,W,\varrho} &= \frac{\alpha}{2(\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2)} \\
&\quad + \left[\frac{-\alpha^2\varrho}{2(\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2)^2} + \frac{\alpha(\theta_{\alpha,\sigma,\varrho} - \sigma^2)}{2(\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2)\theta_{\alpha,\sigma,\varrho}}\right] \frac{d}{d\varrho}\theta_{\alpha,\sigma,\varrho} \\
&= \frac{\alpha}{2(\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2)} + \frac{\alpha^2}{2(\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2)^2\theta_{\alpha,\sigma,\varrho}} \\
&\quad \times \left[\theta^2_{\alpha,\sigma,\varrho} - \theta_{\alpha,\sigma,\varrho}\Big(\varrho + \Big(2 - \frac{1}{\alpha}\Big)\sigma^2\Big) + \Big(1 - \frac{1}{\alpha}\Big)\sigma^4\right] \frac{d}{d\varrho}\theta_{\alpha,\sigma,\varrho}.
\end{aligned}$$

Then (129) holds for $\alpha \in \mathbb{R}_+ \setminus \{1\}$, because $\theta_{\alpha,\sigma,\varrho}$ is a root of the equality in (133).

The A-L capacity and center of this channel are given by the following expressions:

$$C^\lambda_{\alpha,W} = \begin{cases} \left(\frac{\alpha}{\alpha-1}\ln\sqrt{\frac{1}{\alpha} + \frac{\alpha-1}{\alpha}\frac{2\sigma^2\lambda}{\alpha}} - \ln\sqrt{\frac{2\sigma^2\lambda}{\alpha}}\right)\mathbb{1}_{\lambda\in(0,\frac{\alpha}{2\sigma^2})} & \text{for } \alpha \in \mathbb{R}_+ \setminus \{1\}, \\ \left(\sigma^2\lambda - \ln\sqrt{2e\sigma^2\lambda}\right)\mathbb{1}_{\lambda\in(0,\frac{1}{2\sigma^2})} & \text{for } \alpha = 1, \end{cases} \tag{135}$$

$$q^\lambda_{\alpha,W} = \varphi_{\theta^\lambda_{\alpha,\sigma}}, \tag{136}$$

$$\theta^\lambda_{\alpha,\sigma} \triangleq \sigma^2 + \left|\frac{1}{2\lambda} - \frac{\sigma^2}{\alpha}\right|^+. \tag{137}$$

Then $C^\lambda_{\alpha,W}$ is a continuously differentiable function of λ and its derivative is a continuous, increasing, and bijective function of λ from $\mathbb{R}_+$ to $(-\infty, 0]$ given by

$$\frac{d}{d\lambda}C^\lambda_{\alpha,W} = -\frac{\alpha - 2\sigma^2\lambda}{2\lambda(\alpha + (\alpha-1)2\sigma^2\lambda)}\mathbb{1}_{\lambda\leq\frac{\alpha}{2\sigma^2}}. \tag{138}$$

The expressions for the A-L capacity and center given in (135) and (136) are derived using the expressions for Augustin capacity and center, (76), (129)–(131), and Lemma 31.

- If $\lambda \in (0, \alpha/2\sigma^2)$, then there exists a unique $\varrho^\lambda_{\alpha,W}$ satisfying $\frac{d}{d\varrho}C_{\alpha,W,\varrho}\Big|_{\varrho=\varrho^\lambda_{\alpha,W}} = \lambda$ by (130). Furthermore, $\varrho^\lambda_{\alpha,W}$ satisfies $C^\lambda_{\alpha,W} = C_{\alpha,W,\varrho^\lambda_{\alpha,W}} - \lambda\varrho^\lambda_{\alpha,W}$ by (76), because $\frac{d}{d\varrho}C_{\alpha,W,\varrho}$ is decreasing in ϱ.

[22] Derivation of this inequality is analogous to the derivation of (32) and (34) of Lemma 13(c),(d), presented in [22, Appendix B].

Then (135) follows from (126) and (129). On the other hand, $q^{\lambda}_{\alpha,W}=q_{\alpha,W,\varrho^{\lambda}_{\alpha,W}}$ by Lemma 31, because $C_{\alpha,W,\varrho^{\lambda}_{\alpha,W}}=C^{\lambda}_{\alpha,W}+\lambda\varrho^{\lambda}_{\alpha,W}$. Then (136) follows from (127)–(129) and (137). In addition one can confirm that $\varrho^{\lambda}_{\alpha,W}=-\frac{d}{d\lambda}C^{\lambda}_{\alpha,W}$ by solving $\frac{d}{d\varrho}C_{\alpha,W,\varrho}\Big|_{\varrho=\varrho^{\lambda}_{\alpha,W}}=\lambda$ explicitly for $\varrho^{\lambda}_{\alpha,W}$. We, however, do not need to obtain that explicit solution to confirm (135) and (136).

- If $\lambda\in[\alpha/2\sigma^2,\infty)$, then $D_{\alpha}(W\,\|\,\varphi_{\sigma^2}\,|\,p)-\lambda\,\mathbf{E}_p[\varrho]\leq 0$ by (131). On the other hand, $C^{\lambda}_{\alpha,W}\geq 0$, because A-L information is zero for the probability measure that puts all its probability mass to $x=0$. Hence, $C^{\lambda}_{\alpha,W}=0$ and $q^{\lambda}_{\alpha,W}=\varphi_{\sigma^2}$. Thus, both (135) and (136) hold.

Example 5 (parallel Gaussian channels). Let $W_{[1,n]}$ be the product of scalar Gaussian channels W_i with noise variance σ_i for $i\in\{1,\dots,n\}$ and $\rho_{[1,n]}$ be the additive cost function, i.e.,

$$W_{[1,n]}(\mathscr{E}\,|\,x_1^n)=\int\limits_{\mathscr{E}}\left[\prod_{i=1}^{n}\varphi_{\sigma_i^2}(y_i-x_i)\right]dy_1^n,\quad \forall\mathscr{E}\in\mathcal{B}(\mathbb{R}^n),$$

$$\rho_{[1,n]}(x_1^n)=\sum_{i=1}^{n}x_i^2,\quad \forall x_1^n\in\mathbb{R}^n.$$

As a result of Lemma 28, the cost constrained Augustin capacity of $W_{[1,n]}$ satisfies

$$C_{\alpha,W_{[1,n]},\varrho}=\sup_{\varrho_1,\dots,\varrho_n:\,\sum\limits_i\varrho_i\leq\varrho}C_{\alpha,W_i,\varrho_i}.$$

Since C_{α,W_i,ϱ_i}'s are continuous, strictly concave, and increasing in ϱ_i the supremum is achieved at a unique $(\varrho_{\alpha,1},\dots,\varrho_{\alpha,n})$. Then $q_{\alpha,W_{[1,n]},\varrho}=q_{\alpha,W_1,\varrho_{\alpha,1}}\otimes\dots\otimes q_{\alpha,W_n,\varrho_{\alpha,n}}$ by Lemma 28. Furthermore, since C_{α,W_i,ϱ_i}'s are continuously differentiable in ϱ_i, the unique point $(\varrho_{\alpha,1},\dots,\varrho_{\alpha,n})$ can be determined via the derivative test: $\frac{d}{d\varrho_i}C_{\alpha,W_i,\varrho_i}\Big|_{\varrho_i=\varrho_{\alpha,i}}=\lambda_{\alpha}$ for all i's with a positive $\varrho_{\alpha,i}$ and $\frac{d}{d\varrho_i}C_{\alpha,W_i,\varrho_i}\Big|_{\varrho_i=\varrho_{\alpha,i}}\leq\lambda_{\alpha}$ for all i's with a zero $\varrho_{\alpha,i}$ for some $\lambda_{\alpha}\in\mathbb{R}_+$. Thus, using (130), we can conclude that the optimal cost allocation, i.e., $(\varrho_{\alpha,1},\dots,\varrho_{\alpha,n})$, satisfies

$$\varrho_{\alpha,i}=\frac{|\alpha-2\sigma_i^2\lambda_{\alpha}|^{+}}{2\lambda_{\alpha}(\alpha+2(\alpha-1)\sigma_i^2\lambda_{\alpha})}\tag{139}$$

for some λ_{α} that is uniquely determined by constraint $\sum\limits_{i=1}^{n}\varrho_{\alpha,i}=\varrho$, because the expression on the right-hand side of (139) is nonincreasing in λ_{α} for each i. Consequently,

$$C_{\alpha,W_{[1,n]},\varrho}=\sum_{i=1}^{n}C_{\alpha,W_i,\varrho_{\alpha,i}}\tag{140}$$

$$q_{\alpha,W_{[1,n]},\varrho}=\bigotimes_{i=1}^{n}\varphi_{\theta_{\alpha,\sigma_i,\varrho_{\alpha,i}}}\tag{141}$$

where $\theta_{\alpha,\sigma,\varrho}$ is defined in (128). Using the constraints for the optimality of a cost allocation we obtained via the derivative test, i.e., $\frac{d}{d\varrho_i}C_{\alpha,W_i,\varrho_i}\Big|_{\varrho_i=\varrho_{\alpha,i}}=\lambda_{\alpha}$ for all i's with a positive $\varrho_{\alpha,i}$ and $\frac{d}{d\varrho_i}C_{\alpha,W_i,\varrho_i}\Big|_{\varrho_i=\varrho_{\alpha,i}}\leq\lambda_{\alpha}$ for all i's with a zero $\varrho_{\alpha,i}$, together with (129)—instead of (130)—we obtain the following alternative characterization of $\theta_{\alpha,\sigma_i,\varrho_{\alpha,i}}$ in terms of σ_i and λ_{α} that does not depend on $\varrho_{\alpha,i}$'s explicitly:

$$\theta_{\alpha,\sigma_i,\varrho_{\alpha,i}}=\sigma_i^2+\left|\frac{1}{2\lambda_{\alpha}}-\frac{\sigma_i^2}{\alpha}\right|^{+}.\tag{142}$$

The A-L capacity and center of $W_{[1,n]}$ can be written in terms of the corresponding quantities for the component channels using Lemma 32 as follows:

$$C^{\lambda}_{\alpha,W_{[1,n]}} = \sum_{i=1}^{n} C^{\lambda}_{\alpha,W_i}, \qquad q^{\lambda}_{\alpha,W_{[1,n]}} = \bigotimes_{i=1}^{n} q^{\lambda}_{\alpha,W_i}.$$

The cost constrained Augustin capacity and center and A-L capacity and center of vector Gaussian channels with multiple input and output antennas can be analyzed with a similar approach with the help of singular value decomposition.

7. DISCUSSION

Similar to the Rényi information, the Augustin information is a generalization of the mutual information defined in terms of the Rényi divergence. Unlike the order α Rényi information, however, the order α Augustin information does not have a closed form expression, except for the order one case. This makes it harder to prove certain properties of the Augustin information such as its continuous differentiability as a function of the order α, the existence of a unique order α Augustin mean $q_{\alpha,p}$, or the bounds given in (7). However, once these fundamental properties of the Augustin information are established, the analysis of the Augustin capacity is rather straightforward and very similar to the analogous analysis for the Rényi capacity, presented in [13].

Previously, the convex conjugation techniques have been applied to the calculation of the cost constrained Augustin capacity through the quantity $I^{\mathrm{g}\lambda}_{\alpha}(p;W)$, which we have called the R-G information. Although such an approach can successfully characterize the cost constrained Augustin capacity via the R-G capacity, it is nonstandard and somewhat convoluted. A more standard approach, based on the concept of A-L information $I^{\lambda}_{\alpha}(p;W)$, is presented in Section 5.2. The A-L information has not been used or studied before to the best of our knowledge; nevertheless the resulting capacity is identical to the one associated with the R-G information. The optimality of the approach based on the R-G information seems more intuitive, in the light of this observation.

Our analysis of the Augustin information and capacity was primarily motivated by their operational significance in the channel coding problem, [6]. We investigate that operational significance more closely and derive sphere packing bounds with polynomial prefactors for two families of memoryless channels —composition constrained and cost constrained—in [7]. Broadly speaking, the derivation of the sphere packing bound for memoryless channels in [7] is similar to the derivation of the sphere packing bound for product channels in [38], except for the use of the Augustin capacity and center instead of the Rényi capacity and center.

ACKNOWLEDGMENT

The author would like to thank Fatma Nakiboğlu and Mehmet Nakiboğlu for their hospitality; this work would not have been possible without it. The author would like to thank Marco Dalai for informing him about Fano's implicit assertion of the fixed point property in [23] and Gonzalo Vazquez-Vilar for pointing out Poltyrev's paper [19] on the random coding bound. Author would also like to thank the reviewer for his meticulous report, which allowed the author to correct a number of inaccurate and/or imprecise statements in the original manuscript.

ADDITIONAL INFORMATION

Results of this research were presented in part at the 2017 IEEE International Symposium on Information Theory [1]. Detailed proofs of statements formulated in Sections 2.2, 3, 4, and 5.1–5.3 can be found in [22].

REFERENCES

1. Nakiboğlu, B., The Augustin Center and the Sphere Packing Bound for Memoryless Channels, *Proc. 2017 IEEE Int. Sympos. on Information Theory (ISIT'2017), Aachen, Germany, June 25–30, 2017*, pp. 1401–1405.

2. Csiszár, I., Generalized Cutoff Rates and Rényi's Information Measures, *IEEE Trans. Inform. Theory*, 1995, vol. 41, no. 1, pp. 26–34.

3. Dalai, M., Some Remarks on Classical and Classical-Quantum Sphere Packing Bounds: Rényi vs. Kullback–Leibler, *Entropy*, 2017, vol. 19, no. 7, pp. 355 (11 pp.).

4. Mosonyi, M. and Ogawa, T., Divergence Radii and the Strong Converse Exponent of Classical-Quantum Channel Coding with Constant Compositions, `arXiv:1811.10599v4 [quant-ph]`, 2018.

5. Csiszár, I. and Körner, J., *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Cambridge, UK: Cambridge Univ. Press, 2011, 2nd ed. First edition translated under the title *Teoriya informatsii: teoremy kodirovaniya dlya diskretnykh sistem bez pamyati*, Moscow: Mir, 1985.

6. Augustin, U., Noisy Channels, *Habilitation Thesis*, Universität Erlangen-Nürnberg, 1978. Avaliable at `http://bit.ly/2ID8h7m`.

7. Nakiboğlu, B., The Sphere Packing Bound for Memoryless Channels, `arXiv:1804.06372 [cs.IT]`, 2018.

8. van Erven, T. and Harremoës, P., Rényi Divergence and Kullback–Leibler Divergence, *IEEE Trans. Inform. Theory*, 2014, vol. 60, no. 7, pp. 3797–3820.

9. Shayevitz, O., A Note on a Characterization of Rényi Measures and Its Relation to Composite Hypothesis Testing, `arXiv:1012.4401v2 [cs.IT]`, 2010.

10. Shayevitz, O., On Rényi Measures and Hypothesis Testing, in *Proc. 2010 IEEE Int. Sympos. on Information Theory (ISIT'2010), Austin, Texas, USA, June 13–18, 2010*, pp. 894–898.

11. Verdú, S., α-Mutual Information, in *Proc. 2015 Information Theory and Applications Workshop (ITA'2015), San Diego, CA, USA, Feb. 1–6, 2015*, pp. 1–6. Available at `http://ita.ucsd.edu/workshop/15/files/paper/paper_374.pdf`.

12. Kemperman, J.H.B., On the Shannon Capacity of an Arbitrary Channel, *Indag. Math.*, 1974, vol. 77, no. 2, pp. 101–115.

13. Nakiboğlu, B., The Rényi Capacity and Center, *IEEE Trans. Inform. Theory*, 2019, vol. 65, no. 2, pp. 841–860.

14. Gallager, R.G., A Simple Derivation of the Coding Theorem and Some Applications, *IEEE Trans. Inform. Theory*, 1965, vol. 11, no. 1, pp. 3–18.

15. Gallager, R.G., *Information Theory and Reliable Communication*, New York: Wiley, 1968.

16. Ebert, P.M., Error Bounds For Parallel Communication Channels, *Tech. Rep. of Research Lab. of Electronics, MIT*, Cambridge, MA, USA, 1966, no. 448. Available at `https://dspace.mit.edu/handle/1721.1/4295`.

17. Richters, J.S., Communication over Fading Dispersive Channels, *Tech. Rep. of Research Lab. of Electronics, MIT*, Cambridge, MA, USA, 1967, no. 464. Available at `https://dspace.mit.edu/handle/1721.1/4279`.

18. Haroutunian, E.A., Bounds for the Exponent of the Probability of Error for a Semicontinuous Memoryless Channel, *Probl. Peredachi Inf.*, 1968, vol. 4, no. 4, pp. 37–48 [*Probl. Inf. Transm.* (Engl. Transl.), 1968, vol. 4, no. 4, pp. 29–39].

19. Poltyrev, G.Sh., Random Coding Bounds for Discrete Memoryless Channels, *Probl. Peredachi Inf.*, 1982, vol. 18, no. 1, pp. 12–26 [*Probl. Inf. Transm.* (Engl. Transl.), 1982, vol. 18, no. 1, pp. 9–21].

20. Dudley, R.M., *Real Analysis and Probability*, Cambridge: Cambridge Univ. Press, 2002.

21. Bogachev, V.I., *Measure Theory*, Berlin: Springer, 2007.

22. Nakiboğlu, B., The Augustin Capacity and Center, arXiv:1803.07937 [cs.IT], 2018.

23. Fano, R.M., *Transmission of Information: A Statistical Theory of Communications*, New York: M.I.T. Press, 1961.

24. Arimoto, S., Computation of Random Coding Exponent Functions, *IEEE Trans. Inform. Theory*, 1976, vol. 22, no. 6, pp. 665–671.

25. Oohama, Y., The Optimal Exponent Function for the Additive White Gaussian Noise Channel at Rates above the Capacity, in *Proc. 2017 IEEE Int. Sympos. on Information Theory (ISIT'2017), Aachen, Germany, June 25–30, 2017*, pp. 1053–1057.

26. Oohama, Y., Exponent Function for Stationary Memoryless Channels with Input Cost at Rates above the Capacity, arXiv:1701.06545v3 [cs.IT], 2017.

27. Vazquez-Vilar, G., Martinez, A., and Guillén i Fàbregas, A., A Derivation of the Cost-Constrained Sphere-Packing Exponent, in *Proc. 2015 IEEE Int. Sympos. on Information Theory (ISIT'2015), Hong Kong, China, June 14–19, 2015*, pp. 929–933.

28. Rényi, A., On Measures of Entropy and Information, *Proc. 4th Berkeley Sympos. on Mathematical Statistics and Probability, Berkely, CA, USA, June 20 – July 30, 1960*, Neyman, J., Ed., Berkely: Univ. of California Press, 1961, vol. 1: Contributions to the Theory of Statistics, pp. 547–561.

29. Csiszár, I., Information-type Measures of Difference of Probability Distributions and Indirect Observations, *Studia Sci. Math. Hungar.*, 1967, vol. 2, no. 3–4, pp. 299–318.

30. Gilardoni, G.L., On Pinsker's and Vajda's Type Inequalities for Csiszár's f-Divergences, *IEEE Trans. Inform. Theory*, 2010, vol. 56, no. 11, pp. 5377–5386.

31. Shiryaev, A.N., *Probability*, New York: Springer, 1995.

32. Polyanskiy, Y. and Verdú, S., Arimoto Channel Coding Converse and Rényi Divergence, in *Proc. 48th Annual Allerton Conf. on Communication, Control, and Computation, Sept. 29 – Oct. 1, 2010, Allerton, IL, USA*, pp. 1327–1333.

33. Kolmogorov, A.N. and Fomin, S.V., *Elementy teorii funktsii i funktsional'nogo analiza* (Basics of Function Theory and Functional Analysis), Moscow: Nauka, 1968. Translated under the title *Introductory Real Analysis*, New York: Dover, 1975.

34. Csiszár, I., A Class of Measures of Informativity of Observation Channels, *Period. Math. Hungar.*, 1972, vol. 2, no. 1–4, pp. 191–213.

35. Sibson, R., Information Radius, *Z. Wahrsch. Verw. Gebiete*, 1969, vol. 14, no. 2, pp. 149–160.

36. Blahut, R.E., Hypothesis Testing and Information Theory, *IEEE Trans. Inform. Theory*, 1974, vol. 20, no. 4, pp. 405–417.

37. Kostina, V. and Verdú, S., Channels with Cost Constraints: Strong Converse and Dispersion, *IEEE Trans. Inform. Theory*, 2015, vol. 61, no. 5, pp. 2415–2429.

38. Nakiboğlu, B., The Sphere Packing Bound via Augustin's Method, *IEEE Trans. Inform. Theory*, 2019, vol. 65, no. 2, pp. 816–840.