

The Augustin Capacity and Center

Bariş Nakiboğlu
bnakib@metu.edu.tr

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.

CONTENTS

1	Introduction	2
1.1	Notational Conventions	3
1.2	Channel Model	4
1.3	Previous Work and Main Contributions	4
2	Preliminaries	6
2.1	The Rényi Divergence	6
2.2	Tilted Probability Measure	7
2.3	The Conditional Rényi Divergence and Tilted Channel	9
3	The Augustin Information	9
3.1	Existence of a Unique Augustin Mean	9
3.2	Augustin Information as a Function of the Input Distribution	11
3.3	Augustin Information as a Function of the Order	12
3.4	Augustin Information vs Rényi Information	13
4	The Augustin Capacity	14
4.1	Existence of a Unique Augustin Center	14
4.2	Augustin Capacity and Center as a Function of the Order	16
4.3	Convex Hulls of Constraints and Product Constraints	16
4.4	Augustin Capacity vs Rényi Capacity	17
5	The Cost Constrained Problem	17
5.1	The Cost Constrained Augustin Capacity and Center	17
5.2	The Augustin-Legendre Information Measures	18
5.3	The Rényi-Gallager Information Measures	20
5.4	Information Measures for Transition Probabilities	23
6	Examples	26
6.1	Shift Invariant Families	26
6.2	Gaussian Channels	27
7	Discussion	29
	Appendix	30
A	Proofs of Lemmas on the Analyticity of the Rényi Divergence	30
B	Proofs of Lemmas on the Augustin Information	32
C	Augustin's Proof of Lemma 13-(c)	43
D	Proofs of Lemmas on the Augustin Capacity	45
E	Proofs of Lemmas on the Cost Constrained Problem	50
	References	59

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

$$\begin{aligned} I(p; W) &= \inf_{q \in \mathcal{P}(\mathcal{Y})} D(p \otimes W \| p \otimes q) & (1) \\ &= \inf_{q \in \mathcal{P}(\mathcal{Y})} \sum_x p(x) D(W(x) \| q) & (2) \end{aligned}$$

where $\mathcal{P}(\mathcal{Y})$ is the set of all probability measures on the output space $(\mathcal{Y}, \mathcal{Y})$, p is a probability mass function that is positive only on a finite subset of the input set \mathcal{X} , and W is a function of the form $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$. 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 pointed out by Csiszár [2]. The generalization associated with the expression in (1) is called the order α Rényi information and denoted by $I_\alpha^g(p; W)$. 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 \mathcal{A} is defined as $\sup_{p \in \mathcal{A}} I_\alpha(p; W)$.

For constant composition codes on the 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 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 non-stationary) 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 §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 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 \mathcal{Y} of the channel W to be a finite set [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 assume¹ the output space $(\mathcal{Y}, \mathcal{Y})$ to be a measurable space composed of an output set \mathcal{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 $q_{\alpha,p}$ that are the corner stones of the analysis of the Augustin information and capacity. We establish analogous relations for orders greater than one in §3, see Lemma 13-(d).

In [12], Kemperman proved the equality of the (unconstrained) Shannon capacity to the Shannon radius² for any channel of the form $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ and the existence a unique Shannon center 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, Thm. 2]: For any channel W of the form $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ and convex constraint set \mathcal{A} ,

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

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

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

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

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

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 \mathcal{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 \mathcal{A} in §4, see Theorem 1. We call the corresponding probability measure $q_{\alpha,W,\mathcal{A}}$ *the order α Augustin center for the constraint set \mathcal{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, §35], relying on a quantity that was previously employed in discrete channels by Gallager [14, pp. 13-15], [15, §7.3] and in various Gaussian channel models³ by Gallager [14, pp. 15,16], [15, §§7.4,7.5], Ebert [16], and Richters [17]. We call this quantity the Rényi-Gallager information and analyze it in §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, §35] relying on the Rényi-Gallager information is rather convoluted. In §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 we present in this paper have already been derived previously in [6, §§33-35], [10], [18], [19]. In order to delineate our main contributions in the context of these works, we provide a tally in §1.3. Before doing that, we describe our notational conventions in §1.1 and our model in §1.2.

1.1. Notational Conventions

The inner product of any two vectors μ and q in \mathbb{R}^ℓ , i.e. $\sum_{i=1}^\ell \mu^i q^i$, is denoted by $\mu \cdot q$. The ℓ dimensional vector whose all entries are one is denoted by $\mathbf{1}$ 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 \mathcal{S} by $\text{cl}\mathcal{S}$, $\text{int}\mathcal{S}$, and $\text{ch}\mathcal{S}$, respectively; the relevant topology or vector space structure will be evident from the context.

For any set \mathcal{Y} , we denote the set of all subsets of \mathcal{Y} —i.e. the power set of \mathcal{Y} —by $\mathcal{2}^{\mathcal{Y}}$, the set of all probability measures on finite subsets of \mathcal{Y} by $\mathcal{P}(\mathcal{Y})$, and the set of all non-zero finite measures with the same property by $\mathcal{M}^+(\mathcal{Y})$. For any p in $\mathcal{M}^+(\mathcal{Y})$, we call the set of all y 's satisfying $p(y) > 0$ the support of p and denote it by $\text{supp}(p)$.

On a measurable space $(\mathcal{Y}, \mathcal{Y})$, 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 non-zero 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 $(\mathcal{Y}, \mathcal{Y})$. Then μ is absolutely continuous with respect to q , i.e. $\mu \prec q$, iff $\mu(\mathcal{E}) = 0$ for any $\mathcal{E} \in \mathcal{Y}$ such that $q(\mathcal{E}) = 0$; μ and q are equivalent, i.e. $\mu \sim q$, iff $\mu \prec q$ and $q \prec \mu$; μ and q are singular, i.e. $\mu \perp q$, iff $\exists \mathcal{E} \in \mathcal{Y}$ such that $\mu(\mathcal{E}) = q(\mathcal{Y} \setminus \mathcal{E}) = 0$. Furthermore, a set of measures \mathcal{W} on $(\mathcal{Y}, \mathcal{Y})$ is absolutely continuous with respect to q , i.e. $\mathcal{W} \prec q$, iff $w \prec q$ for all $w \in \mathcal{W}$ and uniformly absolutely continuous with respect to q , i.e. $\mathcal{W} \prec^{uni} q$, iff for every $\epsilon > 0$ there exists a $\delta > 0$ such that $w(\mathcal{E}) < \epsilon$ for all $w \in \mathcal{W}$ provided that $q(\mathcal{E}) < \delta$.

We denote the integral of a measurable function f with respect to the measure μ by $\int f \mu(d\mathcal{Y})$ or $\int f(y) \mu(d\mathcal{Y})$. If the integral is on the real line and if it is with respect to the Lebesgue measure, we denote it by $\int f d\mathcal{Y}$ or $\int f(y) d\mathcal{Y}$, 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 $\mathbf{E}_\mu[f(\mathcal{Y})]$.

Our notation will be overloaded for certain symbols; however, the relations represented by these symbols will be clear from the context. We use $\tilde{h}(\cdot)$ to denote both the Shannon entropy and the binary entropy: $\tilde{h}(p) \triangleq \sum_y p(y) \ln \frac{1}{p(y)}$ for all $p \in \mathcal{P}(\mathcal{Y})$ and $\tilde{h}(z) \triangleq z \ln \frac{1}{z} + (1-z) \ln \frac{1}{1-z}$ for all $z \in [0, 1]$. We denote the product of topologies [20, p. 38], σ -algebras [20, p. 118], and measures [20, Thm. 4.4.4] by \otimes . We denote the Cartesian product of sets [20, p. 38] by \times . We use the short hand \mathcal{X}_1^n for the Cartesian product of sets $\mathcal{X}_1, \dots, \mathcal{X}_n$ and \mathcal{Y}_1^n for the product of the σ -algebras $\mathcal{Y}_1, \dots, \mathcal{Y}_n$. 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 point-wise inequity for functions and vectors. For two measures μ and q on the measurable space $(\mathcal{Y}, \mathcal{Y})$, $\mu \leq q$ iff $\mu(\mathcal{E}) \leq q(\mathcal{E})$ for all $\mathcal{E} \in \mathcal{Y}$.

For $a, b \in \mathbb{R}$, $a \wedge b$ is the minimum of a and b . For $f : \mathcal{Y} \rightarrow \mathbb{R}$ and $g : \mathcal{Y} \rightarrow \mathbb{R}$, 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 \mathcal{F} of real valued functions $\wedge_{f \in \mathcal{F}} f$ is the pointwise infimum of f 's in \mathcal{F} , which is an extended real valued function. For a collection of measures $\mathcal{U} \subset \mathcal{M}(\mathcal{Y})$ satisfying $w \leq u$ for all $u \in \mathcal{U}$ for some $w \in \mathcal{P}(\mathcal{Y})$, $\wedge_{u \in \mathcal{U}} u$

³Augustin 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, §35] because Augustin assumed a bounded cost function.

is the infimum of \mathcal{U} with respect to the partial order \leq . There exists a unique infimum measure by [21, Thm. 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 \mathcal{X} to the set of all probability measures on the output space $(\mathcal{Y}, \mathcal{Y})$:

$$W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y}). \quad (4)$$

\mathcal{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 \mathcal{X} to the output space $(\mathcal{Y}, \mathcal{Y})$ by $\mathcal{P}(\mathcal{Y}|\mathcal{X})$. For any $p \in \mathcal{P}(\mathcal{X})$ and $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$, the probability measure whose marginal on \mathcal{X} is p and whose conditional distribution given x is $W(x)$ is denoted by $p \otimes W$. Until §5.4, we confine our discussion to the input distributions in $\mathcal{P}(\mathcal{X})$ and avoid the subtleties related to measurability. The more general case of input distributions in $\mathcal{P}(\mathcal{X})$ is considered⁴ in §5.4.

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

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

A product channel is *stationary* iff $W_t = W$ for all $t \in \{1, \dots, n\}$ for some $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$.

For any $\ell \in \mathbb{Z}_+$, an ℓ dimensional *cost function* ρ is a function from the input set to \mathbb{R}^ℓ that is bounded from below, i.e. that is of the form $\rho : \mathcal{X} \rightarrow \mathbb{R}_{\geq z}^\ell$ for some $z \in \mathbb{R}$. We assume without loss of generality that⁵

$$\inf_{x \in \mathcal{X}} \rho^i(x) \geq 0 \quad \forall i \in \{1, \dots, \ell\}.$$

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

$$\Gamma_\rho^{ex} \triangleq \{\varrho \in \mathbb{R}_{\geq 0}^\ell : \exists x \in \mathcal{X} \text{ s.t. } \rho(x) \leq \varrho\} \quad (5)$$

$$\Gamma_\rho \triangleq \{\varrho \in \mathbb{R}_{\geq 0}^\ell : \exists p \in \mathcal{P}(\mathcal{X}) \text{ s.t. } \sum_x p(x) \rho(x) \leq \varrho\}. \quad (6)$$

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

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

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

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 operator⁶ $T_{\alpha,p}(\cdot)$ as follows:

- $T_{\alpha,p}(q_{\alpha,p}) = q_{\alpha,p}$ and $q_{\alpha,p} \sim q_{1,p}$.
- If $q_{1,p} \prec q$ and $T_{\alpha,p}(q) = q$, then $q_{\alpha,p} = q$.
- $\lim_{j \rightarrow \infty} \|q_{\alpha,p} - T_{\alpha,p}^j(q_{1,p})\| = 0$.
- $D_\alpha(W \| q | p) \geq I_\alpha(p; W) + D_\alpha(q_{\alpha,p} \| q)$ for⁷ all $q \in \mathcal{P}(\mathcal{Y})$.

We can not verify the correctness of the proof of [6, Lemma 34.2]; we discuss our reservations in Appendix C. Lemma 13-(c) is proved⁸ 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 $\lim_{j \rightarrow \infty} \|q_{\alpha,p} - T_{\alpha,p}^j(q_{1,p})\| = 0$; Lemma 13-(c) establishes $\lim_{j \rightarrow \infty} \|q_{\alpha,p} - T_{\alpha,p}^j(q_{\alpha,p}^g)\| = 0$ instead —see (37) and Remark 6. Unlike [6, Lemma 34.2], Lemma 13-(c) also bounds

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

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

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

⁷To 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].

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

$D_\alpha(W\|q|p)$ 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):

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

For finite \mathcal{Y} case, the existence of a q in $\mathcal{P}(\mathcal{Y})$ satisfying both $q \sim q_{1,p}$ and $T_{\alpha,p}(q) = q$ 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 [22, §9.2, (9.24) & 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 [22, §9.2].
- While establishing the equivalence of his expression for the sphere packing exponent in finite \mathcal{Y} case to the one provided by Fano in [22], Haroutunian proved the existence of a fixed point that is equivalent to $q_{1,p}$ for each α in $(0, 1)$, see [18, (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, (3.15), (3.16) and Thm. 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 $T_{\alpha,p}(\cdot)$ 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:

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

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

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

Theorem 1 establishes the existence of a unique Augustin center, $q_{\alpha,W,\mathcal{A}}$, for any convex \mathcal{A} with finite Augustin capacity and the convergence of $\{q_{\alpha,p^{(i)}}\}_{i \in \mathbb{Z}_+}$ to $q_{\alpha,W,\mathcal{A}}$ in total variation topology for any $\{p^{(i)}\}_{i \in \mathbb{Z}_+} \subset \mathcal{A}$ satisfying $\lim_{i \rightarrow \infty} I_\alpha(p^{(i)};W) = C_{\alpha,W,\mathcal{A}}$. 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 $\mathcal{A} = \mathcal{P}(\mathcal{X})$ case similar results were proved by Csiszár [2, Proposition 1] assuming both \mathcal{X} and \mathcal{Y} are finite sets and by van Erven and Harremoës [8, Thm. 34] assuming \mathcal{Y} is 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

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

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 & 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 $\alpha = 1$ case. Thus, formally speaking, all of the propositions in §5.2 are new. The analysis presented in §5.2 is a standard application of the convex conjugation techniques to characterize the cost constrained Augustin capacity and center. A similar analysis for $\alpha = 1$ 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 §5.2 are the followings:

- $C_{\alpha,W}^\lambda$, defined as $\sup_{p \in \mathcal{P}(\mathcal{X})} I_\alpha^\lambda(p;W)$, 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 $\varrho \in \text{int}I_\rho$ and the set of λ 's achieving this infimum form a non-empty 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 $\inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{x \in \mathcal{X}} D_\alpha(W(x)\|q) - \lambda \cdot \rho(x)$ by Theorem 2.
- If $C_{\alpha,W}^\lambda < \infty$, then there exists a unique A-L center $q_{\alpha,W}^\lambda$ satisfying $C_{\alpha,W}^\lambda = \sup_{x \in \mathcal{X}} D_\alpha(W(x)\|q_{\alpha,W}^\lambda) - \lambda \cdot \rho(x)$. Furthermore, $\lim_{i \rightarrow \infty} \|q_{\alpha,p} - q_{\alpha,W}^\lambda\| = 0$ for all $\{p^{(i)}\}_{i \in \mathbb{Z}_+} \subset \mathcal{P}(\mathcal{X})$ satisfying $\lim_{i \rightarrow \infty} I_\alpha^\lambda(p^{(i)};W) = C_{\alpha,W}^\lambda$ 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 $\lambda \in \mathbb{R}_{\geq 0}^\ell$, $\alpha \in \mathbb{R}_+$ 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 $I_{\alpha}^{g\lambda}(p; W)$ is a generalization of the Rényi information $I_{\alpha}^g(p; W)$ with a Lagrange multiplier because $I_{\alpha}^{g0}(p; W) = I_{\alpha}^g(p; W)$. This quantity was first employed by Gallager in [14] by a different parametrization and scaling; later considered by Arimoto [23, §IV], Augustin [6], Ebert [16], Richters [17], Oohama [24], [25], and Vazquez-Vilar, Martinez, and Fàbregas [26] 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 §5.3 are the followings:

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

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

While pursuing a similar analysis in [6, §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, §35] but also other previously considered models, which were either discrete [14, pp. 13-15], [15, §7.3], [23, §IV], [25], [26] or Gaussian [14, pp. 15,16], [15, §§7.4,7.5], [16], [17], [24].

VIII. For channels with uncountable input sets the Shannon information and capacity is often defined via the probability measures on the input space $(\mathcal{X}, \mathcal{X})$, rather than the probability mass functions on the input set \mathcal{X} . In §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 followings:

- If W is a transition probability $(\mathcal{X}, \mathcal{X})$ to $(\mathcal{Y}, \mathcal{Y})$ —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}^{λ} 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 $\text{int}\Gamma_{\rho}$ 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 $\text{int}\Gamma_{\rho}$, 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 $(\mathcal{X}, \mathcal{X})$, 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

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

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] and [27] for example— rather than pairs of non-zero 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 §5.3. For pairs of probability measures Definition 1 is equivalent to usual definition employed in [8] by [8, Thm. 5].

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

- *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, Thm. 3, Thm. 31]). *For any $\alpha \in \mathbb{R}_+$, probability measure w and q on $(\mathcal{Y}, \mathcal{Y})$*

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

For orders in $(0, 1]$ this inequality is called the Pinsker's inequality, [28], [29]. For orders in $(0, 1)$ it is possible to bound the Rényi divergence from above in terms of the total variation distance. For $\alpha = 1/2$ case [30, eq. (21), p. 364] asserts

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

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, Thm. 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, Thm. 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, Thm. 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 iff $\frac{dq_1}{d\nu} = \frac{dq_0}{d\nu}$ w -almost surely.

Lemma 6 ([8, Thm. 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 $(w_0 + w_1 + q_0 + q_1) \prec \nu$,*

$$D_\alpha(\beta w_1 + (1 - \beta) w_0 \| \beta q_1 + (1 - \beta) q_0) \leq \beta D_\alpha(w_1 \| q_1) + (1 - \beta) D_\alpha(w_0 \| q_0). \quad (10)$$

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

Lemma 7 ([8, Thm. 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, Thm. 3, Thm. 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

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

For a slightly tighter bound, see [30, eq. (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 $w|_{\mathcal{G}}(\mathcal{E}) = w(\mathcal{E})$ for all $\mathcal{E} \in \mathcal{G}$ and $q|_{\mathcal{G}}(\mathcal{E}) = q(\mathcal{E})$ for all $\mathcal{E} \in \mathcal{G}$ uniquely define probability measures $w|_{\mathcal{G}}$ and $q|_{\mathcal{G}}$ on $(\mathcal{Y}, \mathcal{G})$. We denote $D_\alpha(w|_{\mathcal{G}} \| q|_{\mathcal{G}})$ by $D_\alpha^{\mathcal{G}}(w \| q)$.

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

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

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, Thm. 30]). *For any $w, q \in \mathcal{P}(\mathcal{Y})$*

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

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

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

We have observed in Lemma 8 that $D_\alpha(w\|q)$ is continuous in α on the closure of the interval that it is finite. Lemma 11, in the following, establishes the analyticity of $D_\alpha(w\|q)$ in α on the interior of the interval that $D_\alpha(w\|q)$ 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,

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

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

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

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

Lemma 11 is new to the best of our knowledge; it is proved in 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 $\mathcal{J}_1 = \{(1)\}$, $\mathcal{J}_2 = \{(2,0), (0,1)\}$, and $\mathcal{J}_3 = \{(3,0,0), (1,1,0), (0,0,1)\}$. Thus one can confirm using (15) by substitution that

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

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 $\xi = \frac{1}{\phi-1} \left(\ln \frac{dw_\phi^q}{d\nu} + G_{w,q}^0(\phi) \right)$ 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 :

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

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

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 $\{(w_i, q_i)\}_{i \in \mathcal{J}}$, then for any ϕ that is in $(0, \chi_{w_i, q_i})$ for all $i \in \mathcal{J}$ 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 $i \in \mathcal{J}$. In addition, Lemma 12 asserts uniform approximation error terms, over all $i \in \mathcal{J}$, 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 $D_\beta(w \| q) \leq \gamma$,

$$\left| \frac{\partial^\kappa D_\alpha(w \| q)}{\partial \alpha^\kappa} \Big|_{\alpha=\phi} \right| \leq \begin{cases} \kappa! \tau^{\kappa+1} \kappa & \phi \neq 1 \\ \kappa! \tau^{\kappa+1} & \phi = 1 \end{cases} \quad (18)$$

$$\left| D_\eta(w \| q) - \sum_{i=0}^{\kappa-1} \frac{(\eta-\phi)^i}{i!} \frac{\partial^i D_\alpha(w \| q)}{\partial \alpha^i} \Big|_{\alpha=\phi} \right| \leq \begin{cases} \frac{\tau^{\kappa+1} |\eta-\phi|^\kappa}{1-|\eta-\phi|^\tau} \left[\kappa - 1 + \frac{1}{1-|\eta-\phi|^\tau} \right] & \phi \neq 1 \\ \frac{\tau^{\kappa+1} |\eta-\phi|^\kappa}{1-|\eta-\phi|^\tau} & \phi = 1 \end{cases} \quad \forall \eta : |\eta - \phi| \leq \frac{1}{\tau} \quad (19)$$

where

$$\tau \triangleq \begin{cases} \frac{1}{|\phi-1|} \vee \left[\frac{1+e^{(1 \vee \beta)\gamma}}{\phi^{\wedge(\beta-\phi)}} + \gamma \right] & \phi \neq 1 \\ \frac{1+e^{\beta\gamma}}{1 \wedge (\beta-1)} & \phi = 1 \end{cases}. \quad (20)$$

Lemma 12 is new to the best of our knowledge; it is proved in 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 $\alpha \in \mathbb{R}_+$, $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$, $Q : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$, and $p \in \mathcal{P}(\mathcal{X})$ the order α conditional Rényi divergence for the input distribution p is

$$D_\alpha(W \| Q | p) \triangleq \sum_{x \in \mathcal{X}} p(x) D_\alpha(W(x) \| Q(x)). \quad (21)$$

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

Remark 1. In [11] and [31], $D_\alpha(W \| Q | p)$ stands for $D_\alpha(p \otimes W \| p \otimes Q)$. For $\alpha = 1$ case the convention used in [11] and [31] is equivalent to ours; for $\alpha \neq 1$ 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 $D_\alpha(p \otimes W \| p \otimes Q) = \sum_x p(x) D_\alpha(W(x) \| Q(x))$, else $D_\alpha(p \otimes W \| p \otimes Q) < \sum_x p(x) D_\alpha(W(x) \| Q(x))$ for $\alpha \in (0, 1)$ and $D_\alpha(p \otimes W \| p \otimes Q) > \sum_x p(x) D_\alpha(W(x) \| Q(x))$ for $\alpha \in (1, \infty)$. The inequalities follow from the Jensen's inequality and the strict concavity of the natural logarithm function.

Definition 4. For any $\alpha \in \mathbb{R}_+$, $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ and $Q : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$, 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

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

If $\exists q \in \mathcal{P}(\mathcal{Y})$ such that $Q(x) = q$ for all $x \in \mathcal{X}$, 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 §3.1. After that we analyze the Augustin information, first as a function of the input distribution for a given order in §3.2 and then as a function of the order for a given input distribution in §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 §3.4. Some of the most important observations about the Augustin information and mean were first reported by Augustin in [6, §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 Appendix B.

3.1. Existence of a Unique Augustin Mean

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

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

One can confirm by substitution that

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

where

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

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 $I_\alpha(p; W) = D_\alpha(W \| q_{\alpha,p} | p)$ for⁹ 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 $T_{\alpha,p}(\cdot)$ 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

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

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

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

where ν is any measure for which $(\sum_x p(x) W(x)) \prec \nu$.

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

$$q_{\alpha,p}^g \triangleq \frac{\mu_{\alpha,p}}{\|\mu_{\alpha,p}\|}. \quad (27)$$

- For any $p \in \mathcal{P}(\mathcal{X})$, the order α Augustin operator for the input distribution p , i.e. $T_{\alpha,p}(\cdot) : \mathcal{Q}_{\alpha,p} \rightarrow \mathcal{P}(\mathcal{Y})$, is given by

$$T_{\alpha,p}(q) \triangleq \sum_x p(x) W_\alpha^q(x) \quad \forall q \in \mathcal{Q}_{\alpha,p} \quad (28)$$

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

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

(a) $I_\alpha(p; W) \leq D_\alpha(W \| q_{1,p} | p) \leq \hbar(p) < \infty$ 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,

$$D_1(W \| q | p) - I_1(p; W) = D_1(q_{1,p} \| q) \quad \forall q \in \mathcal{P}(\mathcal{Y}). \quad (29)$$

(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,

$$T_{\alpha,p}(q_{\alpha,p}) = q_{\alpha,p}, \quad (30)$$

$$\lim_{j \rightarrow \infty} \|q_{\alpha,p} - T_{\alpha,p}^j(q_{\alpha,p}^g)\| = 0, \quad (31)$$

$$D_1(q_{\alpha,p} \| q) \geq D_\alpha(W \| q | p) - I_\alpha(p; W) \geq D_\alpha(q_{\alpha,p} \| q) \quad \forall q \in \mathcal{P}(\mathcal{Y}), \quad (32)$$

and $q_{\alpha,p} \sim q_{1,p}$. In addition,¹¹ if $q_{1,p} \prec q$ and $T_{\alpha,p}(q) = q$, 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,

$$T_{\alpha,p}(q_{\alpha,p}) = q_{\alpha,p}, \quad (33)$$

$$D_\alpha(q_{\alpha,p} \| q) \geq D_\alpha(W \| q | p) - I_\alpha(p; W) \geq D_1(q_{\alpha,p} \| q) \quad \forall q \in \mathcal{P}(\mathcal{Y}), \quad (34)$$

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

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

$$I_\alpha(p; W) = \frac{\alpha}{1-\alpha} D_1(W_{\alpha}^{q_{\alpha,p}} \| W | p) + I_1(p; W_{\alpha}^{q_{\alpha,p}}) \quad (35)$$

$$= \begin{cases} \inf_{V \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} \frac{\alpha}{1-\alpha} D_1(V \| W | p) + I_1(p; V) & \alpha \in (0, 1) \\ \sup_{V \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} \frac{\alpha}{1-\alpha} D_1(V \| W | p) + I_1(p; V) & \alpha \in (1, \infty) \end{cases} \quad (36)$$

$$= \frac{\alpha}{1-\alpha} \inf_{V \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} (D_1(V \| W | p) + \frac{1-\alpha}{\alpha} I_1(p; V)).$$

⁹This is rather easy to prove when \mathcal{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 \mathcal{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 [32, Ch3§12.2]. For channels with arbitrary output spaces, however, $\mathcal{P}(\mathcal{Y})$ is not compact; thus we can not invoke the extreme value theorem to establish the existence of $q_{\alpha,p}$.

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

¹¹Note that $T_{\alpha,p}(q) = q$, 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 of its probability to one of the output letters. Then $T_{\alpha,p}(q) = q$, but $q_{\alpha,p} \neq q$, for all $p \in \mathcal{P}(\mathcal{X})$ and $\alpha \in (0, 1)$.

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

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

Part (a) is proved using Lemma 1; $I_\alpha(p; W) \leq \hbar(p)$ was proved by Csiszár through a different argument in [2, (24)]. Part (b), which is well known, is proved by substitution. Part (c) is due to¹² Augustin [6, Lemma 34.2]. Part (d) is new to the best of our knowledge. Part (e) was proved for the finite \mathcal{Y} case by Csiszár [2, (A24), (A27)].

Definition 7. For any $\alpha \in \mathbb{R}_+$, $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$, $p \in \mathcal{P}(\mathcal{X})$ the unique probability measure $q_{\alpha,p}$ on $(\mathcal{Y}, \mathcal{Y})$ 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 (29), (32), (34), imply the following bound, which is analogous to [33, Thm. 3.1] of Csiszár:

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

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 $T_{\alpha,p}(q_{\alpha,p}) = q_{\alpha,p}$ established in Lemma 13-(c,d) and the definition of $T_{\alpha,p}(\cdot)$ given in (28) imply the following identity for the Augustin mean:

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

In §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 $W_{[1,n]} : \mathcal{X}_1^n \rightarrow \mathcal{P}(\mathcal{Y}_1^n)$ and input distribution $p \in \mathcal{P}(\mathcal{X}_1^n)$ we have

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

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

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

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 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 $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$, $I_\alpha(p; W)$ is a concave function of p satisfying

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

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

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

where $p_\beta = \beta p_1 + (1 - \beta) p_0$ for all $p_0, p_1 \in \mathcal{P}(\mathcal{X})$ 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 iff $\sup_{p \in \mathcal{P}(\mathcal{X})} I_\alpha(p; W)$ is finite.¹³ Furthermore, if $\sup_{p \in \mathcal{P}(\mathcal{X})} I_\eta(p; W)$ is finite for an $\eta \in \mathbb{R}_+$ then $\{I_\alpha(p; W)\}_{\alpha \in (0, \eta]}$ is uniformly equicontinuous in p on $\mathcal{P}(\mathcal{X})$.

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

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

¹²To 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 $q_{\alpha,p}^g$. We cannot verify the correctness of Augustin's proof of [6, Lemma 34.2], see Appendix C for a more detailed discussion.

¹³The Rényi information, discussed in §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)].

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

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

$$\begin{aligned} s_\wedge &= \frac{p_1 \wedge p_0}{\|p_1 \wedge p_0\|}, \\ s_1 &= \frac{p_1 - p_1 \wedge p_0}{1 - \|p_1 \wedge p_0\|}, \\ s_0 &= \frac{p_0 - p_1 \wedge p_0}{1 - \|p_1 \wedge p_0\|}. \end{aligned}$$

Then $s_\wedge, s_1, s_0 \in \mathcal{P}(\mathcal{X})$ 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

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

Thus

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

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 $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ and input distribution $p \in \mathcal{P}(\mathcal{X})$,

- (a) $D_\alpha(W(x) \| q_{\alpha,p}) \leq \ln \frac{1}{p(x)}$,
- (b) $[p(x)]^{\frac{1}{\alpha-1}} W(x) \leq q_{\alpha,p}$,
- (c) $\left| \ln \frac{dq_{\alpha,p}}{dq_{1,p}} \right| \leq \frac{|\alpha-1|}{\alpha} \ln \frac{1}{\min_{x:p(x)>0} p(x)}$.

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

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

- (a) Either $(\alpha - 1)I_\alpha(p; W)$ is a strictly convex function of α from \mathbb{R}_+ to $[-\hbar(p), \infty)$ or $I_\alpha(p; W) = \sum_x p(x) \ln \gamma(x)$ for some $\gamma : \mathcal{X} \rightarrow [1, \infty)$ satisfying $\frac{dW(x)}{dq_{1,p}} = \gamma(x) W(x)$ -a.s. for all $x \in \text{supp}(p)$ and $q_{\alpha,p} = q_{1,p}$ for all $\alpha \in \mathbb{R}_+$.
- (b) $\frac{1-\alpha}{\alpha} I_\alpha(p; W)$ 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) $\{\ln \frac{dq_{\alpha,p}}{dq_{1,p}}\}_{y \in \mathcal{Y}}$ 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

$$\frac{\partial}{\partial \alpha} I_\alpha(p; W) \Big|_{\alpha=\phi} = \frac{\partial}{\partial \alpha} D_\alpha(W \| q_{\phi,p} | p) \Big|_{\alpha=\phi} \quad (45)$$

$$= \begin{cases} \frac{1}{(\phi-1)^2} D_1 \left(W_\phi^{q_{\phi,p}} \Big\| W \Big| p \right) & \phi \neq 1 \\ \sum_x \frac{p(x)}{2} \mathbf{E}_{W(x)} \left[\left(\ln \frac{dW(x)}{dq_{1,p}} - D_1(W(x) \| q_{1,p}) \right)^2 \right] & \phi = 1 \end{cases}. \quad (46)$$

- (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}_+ ; else $I_1(p; W_\alpha^{q_{\alpha,p}}) = \sum_x p(x) \ln \gamma(x)$ —i.e. $D_1(W_\alpha^{q_{\alpha,p}} \| q_{\alpha,p} | p) = \sum_x p(x) \ln \gamma(x)$ —for some $\gamma : \mathcal{X} \rightarrow [1, \infty)$ satisfying $\frac{dW(x)}{dq_{1,p}} = \gamma(x) W(x)$ -a.s. for all $x \in \text{supp}(p)$ and $q_{\alpha,p} = q_{1,p}$ for all $\alpha \in \mathbb{R}_+$.
- (g) $\lim_{\alpha \downarrow 0} I_1(p; W_\alpha^{q_{\alpha,p}}) = \lim_{\alpha \downarrow 0} I_\alpha(p; W)$.

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 $(-1, \infty)$, see the proof of part (f) for a proof. 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), (d), (e), (f), and (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 $\{\ln \frac{dq_{\alpha,p}}{dq_{1,p}}\}_{y \in \mathcal{Y}}$ 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 Gallager¹⁴ [14] and then by Sibson [34], is arguably the most prominent one among them because of its operational significance established by Gallager [14].

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

$$I_\alpha^g(p; W) \triangleq \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(p \otimes W \| p \otimes q). \quad (47)$$

As noted by Sibson [34], one can confirm by substitution that

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

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

$$I_\alpha^g(p; W) = D_\alpha(p \otimes W \| p \otimes q_{\alpha,p}^g) \quad \forall p \in \mathcal{P}(\mathcal{X}), \alpha \in \mathbb{R}_+, \quad (48)$$

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

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,

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

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), (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 $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ and p be an input distribution in $\mathcal{P}(\mathcal{X})$.

(a) Let $u_{\alpha,p} \in \mathcal{P}(\mathcal{X})$ be $u_{\alpha,p}(x) = \frac{p(x)e^{(1-\alpha)D_\alpha(W(x)\|q_{\alpha,p}^g)}}{\sum_{\tilde{x}} p(\tilde{x})e^{(1-\alpha)D_\alpha(W(\tilde{x})\|q_{\alpha,p}^g)}}$ for all x ; then

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

$$= \begin{cases} \sup_{u \in \mathcal{P}(\mathcal{X})} I_\alpha^g(u; W) + \frac{1}{\alpha-1} D_1(p \| u) & \alpha \in (0, 1) \\ \inf_{u \in \mathcal{P}(\mathcal{X})} I_\alpha^g(u; W) + \frac{1}{\alpha-1} D_1(p \| u) & \alpha \in (1, \infty) \end{cases}. \quad (51)$$

(b) Let $a_{\alpha,p} \in \mathcal{P}(\mathcal{X})$ be $a_{\alpha,p}(x) = \frac{p(x)e^{(\alpha-1)D_\alpha(W(x)\|q_{\alpha,p}^g)}}{\sum_{\tilde{x}} p(\tilde{x})e^{(\alpha-1)D_\alpha(W(\tilde{x})\|q_{\alpha,p}^g)}}$ for all x ; then

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

$$= \begin{cases} \inf_{a \in \mathcal{P}(\mathcal{X})} I_\alpha(a; W) - \frac{1}{\alpha-1} D_1(a \| p) & \alpha \in (0, 1) \\ \sup_{a \in \mathcal{P}(\mathcal{X})} I_\alpha(a; W) - \frac{1}{\alpha-1} D_1(a \| p) & \alpha \in (1, \infty) \end{cases}. \quad (53)$$

(c) Let $f_{\alpha,p} : \mathcal{X} \rightarrow \mathbb{R}$ 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

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

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

Lemma 18-(a) was first proved by Poltyrev, [19, Thm. 3.4], in a slightly different form for $\alpha \in [1/2, 1)$ case assuming that \mathcal{Y} is finite. Equation (53) of Lemma 18-(b) was first proved by Shayevitz, [10, Thm. 1], for finite \mathcal{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

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

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

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

$$I_\alpha(p; W) \geq I_\alpha^g(p; W) \quad \alpha \in (0, 1] \quad (56)$$

$$I_\alpha(p; W) \leq I_\alpha^g(p; W) \quad \alpha \in [1, \infty) \quad (57)$$

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 §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 §4.2, we analyze the Augustin capacity and center as a function of the order for a given constraint set. In §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 Appendix D.

Augustin provided a presentation similar to the current section in [6, §§33,34] 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, 34.7, 34.8]— for orders not exceeding one. Augustin, however, defines capacity and center only for the subsets of $\mathcal{P}(\mathcal{X})$ defined through cost constraints. We investigate that important special case more closely in §5.

4.1. Existence of a Unique Augustin Center

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

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

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

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

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

Theorem 1 in the following demonstrates that at least for convex \mathcal{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 $W: \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$, and convex constraint set $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$

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

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

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

Furthermore, for every sequence of input distributions $\{p^{(i)}\}_{i \in \mathbb{Z}_+} \subset \mathcal{A}$ satisfying $\lim_{i \rightarrow \infty} I_\alpha(p^{(i)}; W) = C_{\alpha, W, \mathcal{A}}$, 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 $q_{\alpha, W, \mathcal{A}}$ 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 $\alpha = 1$ and $\mathcal{A} = \mathcal{P}(\mathcal{X})$ 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 when \mathcal{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 $W: \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ with a finite input set \mathcal{X} , and closed convex constraint set $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$, there exists $\tilde{p} \in \mathcal{A}$ such that $I_\alpha(\tilde{p}; W) = C_{\alpha, W, \mathcal{A}}$ and $\exists! q_{\alpha, W, \mathcal{A}} \in \mathcal{P}(\mathcal{Y})$ satisfying

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

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

If \mathcal{A} is $\mathcal{P}(\mathcal{X})$, 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 $W: \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$, the order α Rényi radius of W is

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

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

Definition 11. A constraint set \mathcal{A} for the channel $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ has an order α Augustin center iff $\exists q \in \mathcal{P}(\mathcal{Y})$ such that

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

If $C_{\alpha,W,\mathcal{A}}$ 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 $\mathcal{A} = \{p_1, p_2\}$ and \mathcal{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 $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$ be a constraint set satisfying $C_{\alpha,W,\mathcal{A}} \in \mathbb{R}_{\geq 0}$, and $q_{\alpha,W,\mathcal{A}}$ be a probability measure satisfying (62). Then for every $\{p^{(i)}\}_{i \in \mathbb{Z}_+} \subset \mathcal{A}$ satisfying $\lim_{i \rightarrow \infty} I_\alpha(p^{(i)}; W) = C_{\alpha,W,\mathcal{A}}$ the sequence of order α Augustin means $\{q_{\alpha,p^{(i)}}\}_{i \in \mathbb{Z}_+}$ is a Cauchy sequence with the limit point $q_{\alpha,W,\mathcal{A}}$ and the order α Augustin center $q_{\alpha,W,\mathcal{A}}$ is unique.

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

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

Lemma 13-(b,c,d) and Lemma 20 can also be used establish a lower bound on $\sup_{p \in \mathcal{A}} D_\alpha(W \| q | p)$ in terms of the Augustin capacity and center.

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

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

Note that the form of the lower bound given in (63) is, in a sense, analogous to the ones given in (29), (32), (34). The bound given in (63) is a van Erven-Harremoës bound¹⁶ for $\alpha \in (0, 1]$, but it is not a van Erven-Harremoës bound for $\alpha \in (1, \infty)$ because we have a $D_1(q_{\alpha,W,\mathcal{A}} \| q)$ term rather than a $D_\alpha(q_{\alpha,W,\mathcal{A}} \| q)$ term for $\alpha \in (1, \infty)$.

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:

$$C_{\alpha,W,\mathcal{A}} = \begin{cases} \sup_{p \in \mathcal{A}} \inf_{V \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} \frac{\alpha}{1-\alpha} D_1(V \| W | p) + I_1(p; V) & \alpha \in (0, 1) \\ \sup_{p \in \mathcal{A}} \sup_{V \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} \frac{\alpha}{1-\alpha} D_1(V \| W | p) + I_1(p; V) & \alpha \in (1, \infty) \end{cases}. \quad (64)$$

Then

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

For $\alpha \in (0, 1)$, if the constraint set \mathcal{A} has an order α Augustin center, e.g. when \mathcal{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 \mathcal{A} for the channel $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ has an order α Augustin center, then

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

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

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

Lemma 22 is proved using Csiszár's form for the Augustin information, given in Lemma 13-(e), and Lemma 20. In [35], Blahut proved a similar result assuming both \mathcal{X} and \mathcal{Y} are finite sets and $\mathcal{A} = \mathcal{P}(\mathcal{X})$. Even under those assumptions Blahut's result [35, Thm. 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, [35, Thm. 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 $\alpha^*(R)$.

¹⁶In [8] van Erven and Harremoës have conjectured that the inequality $\sup_{x \in \mathcal{X}} D_\alpha(W(x) \| q) \geq C_{\alpha,W} + D_\alpha(q_{\alpha,W} \| q)$ holds for all $q \in \mathcal{P}(\mathcal{Y})$. Van Erven and Harremoës have also proved the bound for the case when $\alpha = \infty$, assuming that \mathcal{Y} is countable [8, Thm. 37]. We have confirmed 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 §4.4 for a brief discussion of the Rényi capacity and center; a more comprehensive discussion can be found in [13].

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

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

- (a) $C_{\alpha, W, \mathcal{A}}$ is a nondecreasing and lower semicontinuous function of α on \mathbb{R}_+ .
- (b) $\frac{1-\alpha}{\alpha} C_{\alpha, W, \mathcal{A}}$ is a nonincreasing and continuous function of α on¹⁷ $(0, 1)$.
- (c) $(\alpha - 1)C_{\alpha, W, \mathcal{A}}$ is a convex function of α on $(1, \infty)$.
- (d) $C_{\alpha, W, \mathcal{A}}$ is nondecreasing and continuous in α on $(0, 1]$ and $(1, \chi_{W, \mathcal{A}}]$ where $\chi_{W, \mathcal{A}} \triangleq \sup\{\phi : C_{\phi, W, \mathcal{A}} \in \mathbb{R}_{\geq 0}\}$.
- (e) If $\sup_{p \in \mathcal{A}} I_\phi^g(p; W) \in \mathbb{R}_{\geq 0}$ for a $\phi > 1$, then $C_{\alpha, W, \mathcal{A}}$ is nondecreasing and continuous in α on $(0, (1 \vee \chi_{W, \mathcal{A}})]$.

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

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

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

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

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

Consequently, if $C_{\alpha, W, \mathcal{A}}$ is continuous in α on \mathcal{I} for some $\mathcal{I} \subset (0, \eta]$, then $q_{\alpha, W, \mathcal{A}} : \mathcal{I} \rightarrow \mathcal{P}(\mathcal{Y})$ 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 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 $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$, and $\mathcal{A}^{(i)}$ be a constraint set that has an order α Augustin center and a finite $C_{\alpha, W, \mathcal{A}^{(i)}}$ for all $i \in \mathcal{J}$. Then

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

where \mathcal{A} is the convex hull of the union, i.e. $\mathcal{A} = \text{ch}(\cup_{i \in \mathcal{J}} \mathcal{A}^{(i)})$. Furthermore,

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

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

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

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

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

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

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

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

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 $\alpha \in \mathbb{R}_+$, $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$, and $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$ the order α Rényi capacity of W for constraint set \mathcal{A} is

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

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

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

$$\begin{aligned} C_{\alpha,W,\mathcal{A}}^g &\leq C_{\alpha,W,\mathcal{A}} & \alpha \in (0, 1], \\ C_{\alpha,W,\mathcal{A}}^g &\geq C_{\alpha,W,\mathcal{A}} & \alpha \in [1, \infty). \end{aligned}$$

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

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

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

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

If $C_{\alpha,W,\mathcal{A}}^g$ is finite, then $\exists! q_{\alpha,W,\mathcal{A}}^g \in \mathcal{P}(\mathcal{Y})$, the order α Rényi center W for the constraint set \mathcal{A} , satisfying

$$C_{\alpha,W,\mathcal{A}}^g = \sup_{p \in \mathcal{A}} D_\alpha(p \otimes W \| p \otimes q_{\alpha,W,\mathcal{A}}^g).$$

Consequently, the Rényi capacity equals to the Rényi radius provided that $\mathcal{A} = \mathcal{P}(\mathcal{X})$. Hence $C_{\alpha,W}^g = C_{\alpha,W}$ and $q_{\alpha,W}^g = q_{\alpha,W}$ 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 §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 §5.2 we define and analyze the Augustin-Legendre (A-L) information, capacity, radius, and center. The discussion in §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. $\alpha = 1$, $|\mathcal{X}| < \infty$, $|\mathcal{Y}| < \infty$, $\ell = 1$ case. In §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 §5.3 is the equality of the A-L capacity and center to the R-G capacity and center. In §5.4, we demonstrate how the results presented in §5.1, §5.2, and §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 §5.1, §5.2, and §5.3 can be found in Appendix E.

Augustin presented a discussion of the cost constrained capacity $C_{\alpha,W,\varrho}$ in [6, §34] for the case when the cost function ρ is a bounded function of the form $\rho : \mathcal{X} \rightarrow [0, 1]^\ell$ and the order α is in $(0, 1]$. In [6, §35], Augustin also analyzed quantities closely related to the R-G information and capacity. The quantities analyzed by Augustin in [6, §35] have first appeared in Gallager's error exponents analysis for cost constrained channels [14, §6], [15, §7.3, §7.4, §7.5]. Unlike Augustin, Gallager did not assume ρ to be bounded; but Gallager confined his analysis to the case when there is a single cost constraint, i.e. $\ell = 1$ 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, [23, §IV], [24]–[26], 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 $\mathcal{A}(\varrho)$, i.e.

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

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

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

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 $\mathcal{A}(\varrho)$ is a convex set, Theorem 1 holds for $\mathcal{A}(\varrho)$. We denote¹⁸ 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, in the following, summarizes these observations.

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

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

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

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

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

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

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 in the following, states these observations formally.

Lemma 28. *For any length n product channel $W_{[1,n]} : \mathcal{X}_1^n \rightarrow \mathcal{P}(\mathcal{Y}_1^n)$ and additive cost function $\rho_{[1,n]} : \mathcal{X}_1^n \rightarrow \mathbb{R}_{\geq 0}^\ell$ we have¹⁹*

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

If $C_{\alpha,W_{[1,n]},\varrho} \in \mathbb{R}_{\geq 0}$ for an $\alpha \in \mathbb{R}_+$ and $\exists(\varrho_1, \dots, \varrho_n)$ 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), $C_{\alpha,W_{[1,n]},\varrho} = \sum_{t=1}^n C_{\alpha,W_t,\frac{\varrho}{n}}$ whenever $W_{[1,n]}$ is stationary and $\rho_t = \rho_1$ for all $t \in \{1, \dots, n\}$. 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 theorem²⁰ for the upper semicontinuous functions to establish the existence of a $(\varrho_1, \dots, \varrho_n)$ 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, §35], [24], [25]. We discuss that approach in §5.3. Theorem 2 presented in the following and Theorem 3 presented in §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 $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ with a cost function $\rho : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}^\ell$, $p \in \mathcal{P}(\mathcal{X})$, 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

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

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

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

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

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

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

¹⁹If $C_{\alpha,W_t,\varrho_t} = -\infty$ for any $t \in \{1, \dots, n\}$, 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 ∞ .

²⁰Consider the function $f(\varrho_1, \dots, \varrho_n)$ 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 $t \in \{1, \dots, n\}$ and which is equal to $-\infty$ otherwise. We choose a large enough but bounded set using the vector ϱ to obtain a compact set for the supremum.

Then the Augustin-Legendre information $I_\alpha^\lambda(p; W)$ can also be expressed as

$$I_\alpha^\lambda(p; W) = \sup_{\varrho \geq 0} \xi_{\alpha, p}(\varrho) - \lambda \cdot \varrho \quad (74)$$

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

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

$$C_{\alpha, W}^\lambda \triangleq \sup_{p \in \mathcal{P}(\mathcal{X})} I_\alpha^\lambda(p; W). \quad (75)$$

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

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

Hence, using the max-min inequality we can conclude that

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

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

Remark 5. In [6, §33-§35], Augustin considered the case when the cost function ρ is a bounded function of the form $\rho : \mathcal{X} \rightarrow [0, 1]^\ell$. In that case $C_{\alpha, W}^\lambda < \infty$ for all $\lambda \in \mathbb{R}_{\geq 0}^\ell$ provided that $C_{\alpha, W, \varrho} < \infty$ for a $\varrho \in \text{int}\Gamma_\rho$ because $C_{\alpha, W, \mathbf{1}} < \infty$ by Lemma 27-(b) and $C_{\alpha, W, \mathbf{1}} = C_{\alpha, W}$ and $C_{\alpha, W}^\lambda \leq C_{\alpha, W}^0 = C_{\alpha, W}$ for all $\lambda \in \mathbb{R}_{\geq 0}^\ell$ 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 $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ with a cost function $\rho : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}^\ell$. Then

- (a) $C_{\alpha, W}^\lambda$ is convex, nonincreasing, and lower semicontinuous in λ on $\mathbb{R}_{\geq 0}^\ell$ and continuous in λ on $\{\lambda : \exists \epsilon > 0 \text{ s.t. } C_{\alpha, W}^{\lambda - \epsilon \mathbf{1}} < \infty\}$.
- (b) If \mathcal{X} is a finite set, then $C_{\alpha, W, \varrho} = \inf_{\lambda \geq 0} C_{\alpha, W}^\lambda + \lambda \cdot \varrho$.
- (c) If $\varrho \in \text{int}\Gamma_\rho$, then $C_{\alpha, W, \varrho} = \inf_{\lambda \geq 0} C_{\alpha, W}^\lambda + \lambda \cdot \varrho$. If in addition $C_{\alpha, W, \varrho} < \infty$, then there exists a non-empty convex, compact set of $\lambda_{\alpha, W, \varrho}$'s satisfying both (69) and $C_{\alpha, W, \varrho} = C_{\alpha, W}^{\lambda_{\alpha, W, \varrho}} + \lambda_{\alpha, W, \varrho} \cdot \varrho$.
- (d) If $C_{\alpha, W, \varrho}$ is finite and $C_{\alpha, W, \varrho} = C_{\alpha, W}^\lambda + \lambda \cdot \varrho$ for some $\varrho \in \Gamma_\rho$ and $\lambda \in \mathbb{R}_{\geq 0}^\ell$, then $\lim_{i \rightarrow \infty} I_\alpha^\lambda(p^{(i)}; W) = C_{\alpha, W}^\lambda$ for all $\{p^{(i)}\}_{i \in \mathbb{Z}_+} \in \mathcal{A}(\varrho)$ s.t. $\lim_{i \rightarrow \infty} I_\alpha(p^{(i)}; W) = C_{\alpha, W, \varrho}$.

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

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

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 $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ with a cost function $\rho : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}^\ell$, and Lagrange multiplier $\lambda \in \mathbb{R}_{\geq 0}^\ell$

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

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

If the expression on the left hand side of (79) is finite, i.e. if $C_{\alpha, W}^\lambda < \infty$, 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

$$C_{\alpha, W}^\lambda = \sup_{p \in \mathcal{P}(\mathcal{X})} D_\alpha(W \| q_{\alpha, W}^\lambda | p) - \lambda \cdot \mathbf{E}_p[\rho] \quad (81)$$

$$= \sup_{x \in \mathcal{X}} D_\alpha(W(x) \| q_{\alpha, W}^\lambda) - \lambda \cdot \rho(x). \quad (82)$$

Furthermore, for every sequence of input distributions $\{p^{(i)}\}_{i \in \mathbb{Z}_+} \subset \mathcal{P}(\mathcal{X})$ such that $\lim_{i \rightarrow \infty} I_\alpha^\lambda(p^{(i)}; W) = C_{\alpha, W}^\lambda$, 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 $q_{\alpha, W}^\lambda$ is the unique limit point of that Cauchy sequence.

Note that Theorem 2 for $\lambda = 0$ is nothing but Theorem 1 for $\mathcal{A} = \mathcal{P}(\mathcal{X})$. 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 $\mathcal{A} = \mathcal{P}(\mathcal{X})$, as well.

Lemma 30. For any $\alpha \in \mathbb{R}_+$, channel $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ with a cost function $\rho : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}^\ell$ for a finite input set \mathcal{X} , and Lagrange multiplier $\lambda \in \mathbb{R}_{\geq 0}^\ell$, there exists a $\tilde{p} \in \mathcal{P}(\mathcal{X})$ such that $I_\alpha^\lambda(\tilde{p}; W) = C_{\alpha, W}^\lambda$ and $\exists! q_{\alpha, W}^\lambda \in \mathcal{P}(\mathcal{Y})$ satisfying

$$D_\alpha(W \| q_{\alpha, W}^\lambda | p) - \lambda \cdot \mathbf{E}_p[\rho] \leq C_{\alpha, W}^\lambda \quad \forall p \in \mathcal{P}(\mathcal{X}). \quad (83)$$

Furthermore, $q_{\alpha, \tilde{p}} = q_{\alpha, W}^\lambda$ for all $\tilde{p} \in \mathcal{P}(\mathcal{X})$ such that $I_\alpha^\lambda(\tilde{p}; W) = C_{\alpha, W}^\lambda$.

Note that the expression on the left hand side of equation (79) is nothing but the A-L capacity. Thus Theorem 2 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 $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ with a cost function $\rho : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}^\ell$, and $\lambda \in \mathbb{R}_{\geq 0}^\ell$, the order α Augustin-Legendre radius of W for the Lagrange multiplier λ is

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

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

$$C_{\alpha, W}^\lambda - I_\alpha^\lambda(p; W) \geq D_{\alpha \wedge 1}(q_{\alpha, p} \| q_{\alpha, W}^\lambda) \quad \forall p \in \mathcal{P}(\mathcal{X}).$$

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 §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 $\varrho \in \text{int} \Gamma_\rho$, 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 $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ with a cost function $\rho : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}^\ell$, and a cost constraint $\varrho \in \Gamma_\rho$ such that $C_{\alpha, W, \varrho} < \infty$, if $C_{\alpha, W, \varrho} = C_{\alpha, W}^\lambda + \lambda \cdot \varrho$ for a $\lambda \in \mathbb{R}_{\geq 0}^\ell$, then $q_{\alpha, W, \varrho} = q_{\alpha, W}^\lambda$.

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 $W_{[1, n]} : \mathcal{X}_1^n \rightarrow \mathcal{P}(\mathcal{Y}_{[1, n]}^n)$ and additive cost function $\rho_{[1, n]} : \mathcal{X}_1^n \rightarrow \mathbb{R}_{\geq 0}^\ell$ we have

$$C_{\alpha, W_{[1, n]}}^\lambda = \sum_{t=1}^n C_{\alpha, W_t}^\lambda \quad \forall \lambda \in \mathbb{R}_{\geq 0}^\ell, \alpha \in \mathbb{R}_+. \quad (85)$$

Furthermore, if $C_{\alpha, W_{[1, n]}}^\lambda < \infty$, then $q_{\alpha, W_{[1, n]}}^\lambda = \bigotimes_{t=1}^n q_{\alpha, W_t}^\lambda$.

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

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

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

$$\begin{aligned} I_\alpha^\lambda(p; W_{[1, n]}) &\leq I_\alpha^\lambda(p_1 \otimes \cdots \otimes p_n; W_{[1, n]}) \\ &= \sum_{t=1}^n I_\alpha^\lambda(p_t; W_t). \end{aligned} \quad (86)$$

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

5.3. The Rényi-Gallager Information Measures

In §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 $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ with a cost function $\rho : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}^\ell$, $p \in \mathcal{P}(\mathcal{X})$, and $\lambda \in \mathbb{R}_{\geq 0}^\ell$ the order α Rényi-Gallager (R-G) information for the input distribution p and the Lagrange multiplier λ is

$$I_\alpha^{g\lambda}(p; W) \triangleq \begin{cases} \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha \left(p \otimes W e^{\frac{1-\alpha}{\alpha} \lambda \cdot \rho} \middle\| p \otimes q \right) & \alpha \in \mathbb{R}_+ \setminus \{1\} \\ \inf_{q \in \mathcal{P}(\mathcal{Y})} D_1(p \otimes W \middle\| p \otimes q) - \lambda \cdot \mathbf{E}_p[\rho] & \alpha = 1 \end{cases}. \quad (87)$$

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 $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ with a cost function $\rho : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}^\ell$, $p \in \mathcal{P}(\mathcal{X})$, and $\lambda \in \mathbb{R}_{\geq 0}^\ell$, the order α mean measure for the input distribution p and the Lagrange multiplier λ is

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

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

$$q_{\alpha,p}^{g\lambda} \triangleq \frac{\mu_{\alpha,p}^\lambda}{\|\mu_{\alpha,p}^\lambda\|}. \quad (89)$$

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

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

Then as a result of Lemma 2, we have

$$I_\alpha^{g\lambda}(p; W) = D_\alpha \left(p \otimes W e^{\frac{1-\alpha}{\alpha} \lambda \cdot \rho} \middle\| p \otimes q_{\alpha,p}^{g\lambda} \right) \quad (91)$$

$$= \frac{\alpha}{\alpha-1} \ln \|\mu_{\alpha,p}^\lambda\| \quad \alpha \in \mathbb{R}_+ \setminus \{1\}. \quad (92)$$

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

$$D_1(p \otimes W \middle\| p \otimes q) - \lambda \cdot \mathbf{E}_p[\rho] = D_1 \left(p \otimes W \middle\| p \otimes q_{1,p}^{g\lambda} \right) - \lambda \cdot \mathbf{E}_p[\rho] + D_1 \left(q_{1,p}^{g\lambda} \middle\| q \right). \quad (93)$$

Then as a result of Lemma 2, we have

$$I_1^{g\lambda}(p; W) = D_1 \left(p \otimes W \middle\| p \otimes q_{1,p}^{g\lambda} \right) - \lambda \cdot \mathbf{E}_p[\rho]. \quad (94)$$

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

$$\begin{aligned} I_\alpha^\lambda(p; W) &\geq I_\alpha^{g\lambda}(p; W) & \alpha \in (0, 1) \\ I_\alpha^\lambda(p; W) &\leq I_\alpha^{g\lambda}(p; W) & \alpha \in [1, \infty). \end{aligned}$$

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 $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ with a cost function $\rho : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}^\ell$, p be an input distribution in $\mathcal{P}(\mathcal{X})$ and λ be a Lagrange multiplier in $\mathbb{R}_{\geq 0}^\ell$.

(a) Let $u_{\alpha,p}^\lambda \in \mathcal{P}(\mathcal{X})$ be $u_{\alpha,p}^\lambda(x) = \frac{p(x) e^{(1-\alpha)D_\alpha(W(x) \middle\| q_{\alpha,p}) + (\alpha-1)\lambda \cdot \rho(x)}}{\sum_{\tilde{x}} p(\tilde{x}) e^{(1-\alpha)D_\alpha(W(\tilde{x}) \middle\| q_{\alpha,p}) + (\alpha-1)\lambda \cdot \rho(x)}}$ for all x ; then

$$I_\alpha^\lambda(p; W) = I_\alpha^{g\lambda}(u_{\alpha,p}^\lambda; W) + \frac{1}{\alpha-1} D_1(p \middle\| u_{\alpha,p}^\lambda) \quad (95)$$

$$= \begin{cases} \sup_{u \in \mathcal{P}(\mathcal{X})} I_\alpha^{g\lambda}(u; W) + \frac{1}{\alpha-1} D_1(p \middle\| u) & \alpha \in (0, 1) \\ \inf_{u \in \mathcal{P}(\mathcal{X})} I_\alpha^{g\lambda}(u; W) + \frac{1}{\alpha-1} D_1(p \middle\| u) & \alpha \in (1, \infty) \end{cases}. \quad (96)$$

(b) Let $a_{\alpha,p}^\lambda \in \mathcal{P}(\mathcal{X})$ be $a_{\alpha,p}^\lambda(x) = \frac{p(x) e^{(\alpha-1)D_\alpha(W(x) \middle\| q_{\alpha,p}^{g\lambda}) + (1-\alpha)\lambda \cdot \rho(x)}}{\sum_{\tilde{x}} p(\tilde{x}) e^{(\alpha-1)D_\alpha(W(\tilde{x}) \middle\| q_{\alpha,p}^{g\lambda}) + (1-\alpha)\lambda \cdot \rho(x)}}$ for all x ; then

$$I_\alpha^{g\lambda}(p; W) = I_\alpha^\lambda(a_{\alpha,p}^\lambda; W) - \frac{1}{\alpha-1} D_1(a_{\alpha,p}^\lambda \middle\| p) \quad (97)$$

$$= \begin{cases} \inf_{a \in \mathcal{P}(\mathcal{X})} I_\alpha^\lambda(a; W) - \frac{1}{\alpha-1} D_1(a \middle\| p) & \alpha \in (0, 1) \\ \sup_{a \in \mathcal{P}(\mathcal{X})} I_\alpha^\lambda(a; W) - \frac{1}{\alpha-1} D_1(a \middle\| p) & \alpha \in (1, \infty) \end{cases}. \quad (98)$$

(c) Let $f_{\alpha,p}^\lambda : \mathcal{X} \rightarrow \mathbb{R}$ be $f_{\alpha,p}^\lambda(x) = [D_\alpha(W(x) \| q_{\alpha,p}) - \lambda \cdot \rho(x) - I_\alpha^\lambda(p; W)] \mathbb{1}_{\{p(x) > 0\}}$ for all x ; then

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

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

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 $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ with a cost function $\rho : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}^\ell$, and $\lambda \in \mathbb{R}_{\geq 0}^\ell$, the order α Rényi-Gallager (R-G) capacity for the Lagrange multiplier λ is

$$C_{\alpha,W}^{g\lambda} \triangleq \sup_{p \in \mathcal{P}(\mathcal{X})} I_\alpha^{g\lambda}(p; W).$$

Using the definition of $I_\alpha^{g\lambda}(p; W)$, given in (87), we get the following expression for $C_{\alpha,W}^{g\lambda}$.

$$C_{\alpha,W}^{g\lambda} = \begin{cases} \sup_{p \in \mathcal{P}(\mathcal{X})} \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha \left(p \otimes W e^{\frac{1-\alpha}{\alpha} \lambda \cdot \rho} \middle\| p \otimes q \right) & \alpha \in \mathbb{R}_+ \setminus \{1\} \\ \sup_{p \in \mathcal{P}(\mathcal{X})} \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(p \otimes W \| p \otimes q) - \lambda \cdot \mathbf{E}_p[\rho] & \alpha = 1 \end{cases} \quad (101)$$

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 $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ with a cost function $\rho : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}^\ell$, and Lagrange multiplier $\lambda \in \mathbb{R}_{\geq 0}^\ell$

$$\sup_{p \in \mathcal{P}(\mathcal{X})} \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha \left(p \otimes W e^{\frac{1-\alpha}{\alpha} \lambda \cdot \rho} \middle\| p \otimes q \right) = \inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{p \in \mathcal{P}(\mathcal{X})} D_\alpha \left(p \otimes W e^{\frac{1-\alpha}{\alpha} \lambda \cdot \rho} \middle\| p \otimes q \right) \quad (102)$$

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

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

$$C_{\alpha,W}^{g\lambda} = \sup_{p \in \mathcal{P}(\mathcal{X})} D_\alpha \left(p \otimes W e^{\frac{1-\alpha}{\alpha} \lambda \cdot \rho} \middle\| p \otimes q_{\alpha,W}^{g\lambda} \right) \quad (104)$$

$$= \sup_{x \in \mathcal{X}} D_\alpha \left(W(x) \| q_{\alpha,W}^{g\lambda} \right) - \lambda \cdot \rho(x). \quad (105)$$

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

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 $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ with cost function $\rho : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}^\ell$ for a finite input set \mathcal{X} , and Lagrange multiplier $\lambda \in \mathbb{R}_{\geq 0}^\ell$, there exists a $\tilde{p} \in \mathcal{P}(\mathcal{X})$ such that $I_\alpha^\lambda(\tilde{p}; W) = C_{\alpha,W}^\lambda$ and $\exists! q_{\alpha,W}^\lambda \in \mathcal{P}(\mathcal{Y})$ satisfying

$$D_\alpha \left(p \otimes W e^{\frac{1-\alpha}{\alpha} \lambda \cdot \rho} \middle\| p \otimes q_{\alpha,W}^\lambda \right) \leq C_{\alpha,W}^{g\lambda} \quad \forall p \in \mathcal{P}(\mathcal{X}). \quad (106)$$

Furthermore, $q_{\alpha,\tilde{p}}^{g\lambda} = q_{\alpha,W}^{g\lambda}$ for all $\tilde{p} \in \mathcal{P}(\mathcal{X})$ such that $I_\alpha^{g\lambda}(\tilde{p}; W) = C_{\alpha,W}^{g\lambda}$.

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

$$C_{\alpha,W}^\lambda = S_{\alpha,W}^\lambda = C_{\alpha,W}^{g\lambda} \quad \forall \alpha \in \mathbb{R}_+, \lambda \in \mathbb{R}_{\geq 0}^\ell. \quad (107)$$

Furthermore, whenever $C_{\alpha,W}^\lambda$ 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.

$$q_{\alpha,W}^\lambda = q_{\alpha,W}^{g\lambda} \quad \forall \alpha \in \mathbb{R}_+, \lambda \in \mathbb{R}_{\geq 0}^\ell \text{ s.t. } C_{\alpha,W}^\lambda < \infty. \quad (108)$$

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 $C_{\alpha,W}^\lambda$ 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

$$C_{\alpha,W}^\lambda - I_\alpha^{g\lambda}(p; W) \geq D_\alpha(q_{\alpha,p}^{g\lambda} \| q_{\alpha,W}^\lambda) \quad \forall p \in \mathcal{P}(\mathcal{X}).$$

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 $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ with a cost function $\rho : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}^\ell$, and Lagrange multiplier $\lambda \in \mathbb{R}_{\geq 0}^\ell$ satisfying $C_{\alpha, W}^\lambda < \infty$

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

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

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

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

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

Furthermore $q_{\alpha, W}^\lambda$ is continuous in λ on $\{\lambda : \exists \epsilon > 0 \text{ s.t. } C_{\alpha, W}^{\lambda - \epsilon \mathbf{1}} < \infty\}$ 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 $\mathcal{P}(\mathcal{X})$, i.e. for probability mass functions that are zero in all but finite number of elements of \mathcal{X} . In many practically relevant and analytically interesting models, however, the input set \mathcal{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 $\mathcal{P}(\mathcal{X})$ to $\mathcal{P}(\mathcal{X})$. For instance, in the additive Gaussian channels described in Examples 4 and 5, the equality $I_\alpha(p; W) = C_{\alpha, W, \rho}$ 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 $\mathcal{P}(\mathcal{X})$ to $\mathcal{P}(\mathcal{X})$ provided that W and Q are not only functions from \mathcal{X} to $\mathcal{P}(\mathcal{Y})$, but also transition probabilities from $(\mathcal{X}, \mathcal{X})$ to $(\mathcal{Y}, \mathcal{Y})$. After that we will show that if in addition \mathcal{X} is countably separated, then the supremum of A-L information $I_\alpha^\lambda(p; W)$ over $\mathcal{P}(\mathcal{X})$ is equal to the A-L radius $S_{\alpha, W}^\lambda$, see Theorem 4. This will imply that the cost constrained Augustin capacity $C_{\alpha, W, \rho}$ —defined in (68)—is equal to the supremum of the Augustin information $I_\alpha(p; W)$ over members of $\mathcal{P}(\mathcal{X})$ satisfying $\mathbf{E}_p[\rho] \leq \rho$, 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, 10.7.1] with a minor modification: we use $W(\mathcal{E}|x)$ instead of $W(x|\mathcal{E})$.

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

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

We denote the set of all transition probabilities from $(\mathcal{X}, \mathcal{X})$ to $(\mathcal{Y}, \mathcal{Y})$ by $\mathcal{P}(\mathcal{Y}|\mathcal{X})$ with the tacit understanding that \mathcal{X} and \mathcal{Y} will be clear from the context. If W satisfies (i), then $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ is a channel, i.e. W is a member of $\mathcal{P}(\mathcal{Y}|\mathcal{X})$, even if W does not satisfy (ii). Hence $\mathcal{P}(\mathcal{Y}|\mathcal{X}) \subset \mathcal{P}(\mathcal{Y}|\mathcal{X})$. 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 $\mathcal{P}(\mathcal{X})$ to $\mathcal{P}(\mathcal{X})$, we ensure the \mathcal{X} -measurability of $D_\alpha(W(x) \| Q(x))$ on \mathcal{X} and replace the sum in (21) with an integral. If (\mathcal{X}, τ) 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 $(W(x) + Q(x)) \ll \nu$ for all $x \in \mathcal{X}$. 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 $(\mathcal{X}, \mathcal{X})$ to $(\mathcal{Y}, \mathcal{Y})$ 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 \mathcal{Y} , and $W, Q \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ the function $D_\alpha(W(\cdot) \| Q(\cdot)) : \mathcal{X} \rightarrow [0, \infty]$ is \mathcal{X} -measurable.

Proof of Lemma 37. There exists $\{\mathcal{E}_i\}_{i \in \mathbb{Z}_+} \subset \mathcal{Y}$ such that $\mathcal{Y} = \sigma(\{\mathcal{E}_i : i \in \mathbb{Z}_+\})$ because \mathcal{Y} is countably generated σ -algebra. Let \mathcal{Y}_i be

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

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

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

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 \mathcal{Y} . Thus as a result of the definition of the Rényi divergence given in (8) we have

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

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

Definition 20. For any $\alpha \in \mathbb{R}_+$, countable generated σ -algebra \mathcal{Y} of subsets of \mathcal{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

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

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 a \mathcal{X} -measurable function.

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

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

Furthermore, for any \mathcal{X} -measurable cost function $\rho : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}^{\ell}$ 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 defined as

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

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

Although we have included $\lambda \cdot \mathbf{E}_p[\rho] = \infty$ 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}^{λ} to be the set of all such p 's:

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

For an arbitrary σ -algebra \mathcal{X} , the singletons (i.e. sets with only one element) are not necessarily measurable sets; thus $\mathcal{P}(\mathcal{X})$ is not necessarily a subset of \mathcal{A}^{λ} . If \mathcal{X} is countably separated, then the singletons are in \mathcal{X} by [21, Thm. 6.5.7], $\mathcal{P}(\mathcal{X}) \subset \mathcal{A}^{\lambda}$ 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} be a countably generated σ -algebra, W be a transition probability from $(\mathcal{X}, \mathcal{X})$ to $(\mathcal{Y}, \mathcal{Y})$, $\rho : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}^{\ell}$ be a \mathcal{X} -measurable cost function, and $\alpha \in \mathbb{R}_+$. Then for all $\lambda \in \mathbb{R}_{\geq 0}^{\ell}$ we have

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

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

$$= C_{\alpha, W}^{\lambda} \quad (118)$$

where \mathcal{A}^{λ} 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

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

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

Proof of Theorem 4. Since $\mathcal{P}(\mathcal{X}) \subset \mathcal{A}^{\lambda}$, the max-min inequality implies

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

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 $q_{\alpha, W}^\lambda \in \mathcal{P}(\mathcal{Y})$ satisfying

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

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

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 $\mathcal{A}(\varrho) \subset \mathcal{A}(\varrho)$ and $\sup_{p \in \mathcal{A}(\varrho)} I_\alpha(p; W) \geq C_{\alpha, W, \varrho}$ whenever \mathcal{X} is countably separated. For the cost constraints in $\text{int}\Gamma_\rho$ 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 $(\mathcal{X}, \mathcal{X})$ to $(\mathcal{Y}, \mathcal{Y})$, $\rho : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}^\ell$ be a \mathcal{X} -measurable cost function, and $\alpha \in \mathbb{R}_+$. For any $\varrho \in \text{int}\Gamma_\rho$ we have*

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

$$= C_{\alpha, W, \varrho} \quad (122)$$

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

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

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

Furthermore, $q_{\alpha, W, \varrho} = q_{\alpha, W}^\lambda$ for all $\lambda \in \mathbb{R}_{\geq 0}^\ell$ satisfying $C_{\alpha, W, \varrho} = C_{\alpha, W}^\lambda + \lambda \cdot \varrho$.

Proof of Theorem 5. Since $\mathcal{A}(\varrho) \subset \mathcal{A}(\varrho)$, the max-min inequality implies

$$\begin{aligned} \sup_{p \in \mathcal{A}(\varrho)} \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(W \| q | p) &\leq \sup_{p \in \mathcal{A}(\varrho)} \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(W \| q | p) \\ &\leq \inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{p \in \mathcal{A}(\varrho)} D_\alpha(W \| q | p). \end{aligned}$$

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 $C_{\alpha, W}^\lambda$ there exists a unique $q_{\alpha, W}^\lambda$ satisfying (120). Thus we have,

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

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

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 $q_{\alpha, W}^\lambda$ for λ 's satisfying $C_{\alpha, W, \varrho} = C_{\alpha, W}^\lambda + \lambda \cdot \varrho$ 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 $C_{\alpha, W}^\lambda$ and $q_{\alpha, W}^\lambda$. That, however, is not the standard practice, [36, Thm. 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 $p \otimes W$ on $(\mathcal{X} \times \mathcal{Y}, \mathcal{X} \otimes \mathcal{Y})$ such that

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

by [21, Thm. 10.7.2.]. Thus $I_\alpha^g(p; W)$ 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 $W e^{\frac{1-\alpha}{\alpha}\lambda \cdot \rho}$ is a transition kernel—rather than a transition probability (i.e. Markov kernel)—and then proceeds with establishing the existence a unique measure $p \otimes W e^{\frac{1-\alpha}{\alpha}\lambda \cdot \rho}$ 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, $p \otimes W e^{\frac{1-\alpha}{\alpha}\lambda \cdot \rho}$ 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 (A Channel with an Affine Capacity). Let the channel $W : \mathbb{R}_{\geq 0} \rightarrow \mathcal{P}(\mathcal{B}([0, 1]))$ and the associated cost function $\rho : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ be

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

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

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

Let u_i be uniform distribution on $[i, i+1)$; then one can confirm by substitution that $T_{\alpha, u_i}(u_0) = u_0$. Then using the Jensen's inequality together with the fixed point property we get²¹

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

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). On the other hand, one can confirm by substitution that

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

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

$$\begin{aligned} C_{\alpha, W, \varrho} &= \varrho + 1, \\ q_{\alpha, W, \varrho} &= u_0. \end{aligned}$$

Then as a result of (76) we have

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

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

Example 2 (A Channel with a Non-Upper Semicontinuous Capacity). Let the channel $W : \mathbb{R} \rightarrow \mathcal{P}(\mathcal{B}([0, 1]))$ and the associated cost function $\rho : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ be

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

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

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

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

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

$$C_{\alpha, W, \varrho} = \begin{cases} (\varrho + 1) \ln 2 & \varrho > 0 \\ \ln 3/2 & \varrho = 0 \end{cases},$$

$$C_{\alpha, W}^\lambda = \begin{cases} \infty & \lambda \in [0, \ln 2) \\ \ln 2 & \lambda \in [\ln 2, \infty) \end{cases}.$$

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

Example 3 (A 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

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

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.

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

$$\rho(x) = x^2 \quad \forall x \in \mathbb{R}.$$

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

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

$$q_{\alpha, W, \varrho} = \varphi_{\theta_{\alpha, \sigma, \varrho}}, \quad (127)$$

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

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

$$\frac{d}{d\varrho} C_{\alpha, W, \varrho} = \frac{\alpha}{2(\alpha \theta_{\alpha, \sigma, \varrho} + (1-\alpha)\sigma^2)} \quad (129)$$

$$= \frac{\alpha}{\alpha \varrho + \sigma^2 + \sqrt{(\alpha \varrho - \sigma^2)^2 + 4\varrho \alpha^2 \sigma^2}}. \quad (130)$$

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

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

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

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

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

$$T_{\alpha,\varphi_\theta}(\varphi_\theta) = \varphi \left(\frac{\alpha\theta}{(\alpha\theta+(1-\alpha)\sigma^2)^2\varrho} + \frac{\sigma^2\theta}{\alpha\theta+(1-\alpha)\sigma^2} \right). \quad (132)$$

Consequently, if φ_θ is a fixed point of $T_{\alpha,\varphi_\theta}(\cdot)$, then θ satisfies the following equality

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

$\theta_{\alpha,\sigma,\varrho}$, 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 $T_{\alpha,\varphi_\theta}(\varphi_{\theta_{\alpha,\sigma,\varrho}}^2) = \varphi_{\theta_{\alpha,\sigma,\varrho}}$, i.e. $\varphi_{\theta_{\alpha,\sigma,\varrho}}$ is a fixed point of $T_{\alpha,\varphi_\theta}(\cdot)$. Then using the Jensen's inequality together with this fixed point property we get²²

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

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

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

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

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

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

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:

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

$$q_{\alpha,W}^\lambda = \varphi_{\theta_{\alpha,\sigma}^\lambda} \quad (136)$$

$$\theta_{\alpha,\sigma}^\lambda \triangleq \sigma^2 + \left| \frac{1}{2\lambda} - \frac{\sigma^2}{\alpha} \right|^+ \quad (137)$$

Then $C_{\alpha,W}^\lambda$ 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

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

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), (130), (131), and Lemma 31.

- If $\lambda \in (0, \alpha/2\sigma^2)$, then there exists a unique $\varrho_{\alpha,W}^\lambda$ satisfying $\frac{d}{d\varrho} C_{\alpha,W,\varrho}|_{\varrho=\varrho_{\alpha,W}^\lambda} = \lambda$ by (130). Furthermore, $\varrho_{\alpha,W}^\lambda$ satisfies $C_{\alpha,W}^\lambda = C_{\alpha,W,\varrho_{\alpha,W}^\lambda} - \lambda\varrho_{\alpha,W}^\lambda$ by (76) because $\frac{d}{d\varrho} C_{\alpha,W,\varrho}$ is decreasing in ϱ . Then (135) follows from (126) and (129). On the other hand $q_{\alpha,W}^\lambda = q_{\alpha,W,\varrho_{\alpha,W}^\lambda}$ by Lemma 31 because $C_{\alpha,W,\varrho_{\alpha,W}^\lambda} = C_{\alpha,W}^\lambda + \lambda\varrho_{\alpha,W}^\lambda$. Then (136) follows from (127), (128), (129), and (137). In addition one can confirm that $\varrho_{\alpha,W}^\lambda = -\frac{d}{d\lambda} C_{\alpha,W}^\lambda$ by solving $\frac{d}{d\varrho} C_{\alpha,W,\varrho}|_{\varrho=\varrho_{\alpha,W}^\lambda} = \lambda$ explicitly for $\varrho_{\alpha,W}^\lambda$. We, however, do not need to obtain that explicit solution to confirm (135) and (136).
- If $\lambda \in [\alpha/2\sigma^2, \infty)$, then $D_\alpha(W||\varphi_{\sigma^2}|p) - \lambda\mathbf{E}_p[\rho] \leq 0$ by (131). On the other hand, $C_{\alpha,W}^\lambda \geq 0$ because A-L information is zero for the probability measure that puts all its probability mass to $x = 0$. Hence $C_{\alpha,W}^\lambda = 0$ and $q_{\alpha,W}^\lambda = \varphi_{\sigma^2}$. Thus, both (135) and (136) hold.

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

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.

$$\begin{aligned} W_{[1,n]}(\mathcal{E}|x_1^n) &= \int_{\mathcal{E}} \left[\prod_{i=1}^n \varphi_{\sigma_i^2}(y_i - x_i) \right] dy_1^n & \forall \mathcal{E} \in \mathcal{B}(\mathbb{R}^n), \\ \rho_{[1,n]}(x_1^n) &= \sum_{i=1}^n x_i^2 & \forall x_1^n \in \mathbb{R}^n. \end{aligned}$$

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

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

Since $C_{\alpha, W_i, \varrho_i}$'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 $C_{\alpha, W_i, \varrho_i}$'s are continuously differentiable in ϱ_i , the unique point $(\varrho_{\alpha,1}, \dots, \varrho_{\alpha,n})$ can be determined via the derivative test: $\frac{d}{d\varrho_i} C_{\alpha, W_i, \varrho_i} |_{\varrho_i = \varrho_{\alpha,i}} = \lambda_{\alpha}$ for all i 's with a positive $\varrho_{\alpha,i}$ and $\frac{d}{d\varrho_i} C_{\alpha, W_i, \varrho_i} |_{\varrho_i = \varrho_{\alpha,i}} \leq \lambda_{\alpha}$ 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

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

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

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

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

where $\theta_{\alpha, \sigma_i, \varrho_{\alpha,i}}$ is defined in (128). Using the constraints for the optimality of a cost allocation we obtained via the derivative test, i.e. $\frac{d}{d\varrho_i} C_{\alpha, W_i, \varrho_i} |_{\varrho_i = \varrho_{\alpha,i}} = \lambda_{\alpha}$ for all i 's with a positive $\varrho_{\alpha,i}$ and $\frac{d}{d\varrho_i} C_{\alpha, W_i, \varrho_i} |_{\varrho_i = \varrho_{\alpha,i}} \leq \lambda_{\alpha}$ 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

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

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:

$$\begin{aligned} C_{\alpha, W_{[1,n]}}^{\lambda} &= \sum_{i=1}^n C_{\alpha, W_i}^{\lambda}, \\ q_{\alpha, W_{[1,n]}}^{\lambda} &= \bigotimes_{i=1}^n q_{\alpha, W_i}^{\lambda}. \end{aligned}$$

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 $I_{\alpha}^{\lambda}(p; W)$, 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 non-standard and somewhat convoluted. A more standard approach, based on the concept of A-L information $I_{\alpha}^{\lambda}(p; W)$, is presented in §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 [37], except for the use of the Augustin capacity and center instead of the Rényi capacity and center.

APPENDIX

A. Proofs of Lemmas on the Analyticity of the Rényi Divergence

Proof of Lemma 11. Let $g(\alpha)$ and $f(\alpha, y)$ be

$$g(\alpha) \triangleq \int \left(\frac{dw}{d\nu}\right)^\alpha \left(\frac{dq}{d\nu}\right)^{1-\alpha} \nu(dy), \quad (\text{A.1})$$

$$f(\alpha, y) \triangleq \left(\frac{dw}{d\nu}\right)^\alpha \left(\frac{dq}{d\nu}\right)^{1-\alpha} \quad (\text{A.2})$$

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

$$D_\alpha(w \| q) = \frac{1}{\alpha-1} \ln g(\alpha) \quad \alpha \in \mathbb{R}_+ \setminus \{1\}. \quad (\text{A.3})$$

Furthermore $g(\alpha)$ does not depend on the choice of ν , but $f(\alpha, y)$ does.

$$\frac{\partial^\kappa}{\partial \alpha^\kappa} f(\alpha, y) = \left(\ln \frac{dw}{d\nu} - \ln \frac{dq}{d\nu} \right)^\kappa f(\alpha, y) \quad \forall \kappa \in \mathbb{Z}_{\geq 0}. \quad (\text{A.4})$$

Then using the inequality $z \ln z \geq -1/e$ we get

$$\left| \frac{\partial^\kappa}{\partial \alpha^\kappa} f(\alpha, y) \right| \leq \left(\frac{\kappa}{\alpha e}\right)^\kappa \frac{dq}{d\nu} \mathbb{1}_{\left\{\frac{dw}{d\nu} \leq \frac{dq}{d\nu}\right\}} + \left(\frac{\kappa}{\phi - \alpha}\right)^\kappa f(\phi, y) \mathbb{1}_{\left\{\frac{dw}{d\nu} > \frac{dq}{d\nu}\right\}} \quad \forall \kappa \in \mathbb{Z}_{\geq 0}, \phi \in (\alpha, \infty).$$

Invoking the Stirling's approximation for the factorial function, i.e. $\sqrt{2\pi\kappa}(\kappa/e)^\kappa \leq \kappa! \leq e\sqrt{\kappa}(\kappa/e)^\kappa$, we get

$$\left| \frac{\partial^\kappa}{\partial \alpha^\kappa} f(\alpha, y) \right| \leq \frac{\kappa!}{\sqrt{2\pi\kappa}} \left(\frac{1}{\alpha^\kappa} \frac{dq}{d\nu} \mathbb{1}_{\left\{\frac{dw}{d\nu} \leq \frac{dq}{d\nu}\right\}} + \frac{f(\phi, y)}{(\phi - \alpha)^\kappa} \mathbb{1}_{\left\{\frac{dw}{d\nu} > \frac{dq}{d\nu}\right\}} \right) \quad \forall \kappa \in \mathbb{Z}_{\geq 0}, \phi \in (\alpha, \infty). \quad (\text{A.5})$$

On the other hand $\int f(\phi, y) \nu(dy) = e^{(\phi-1)D_\phi(w \| q)}$ and for all α in $(0, \chi_{w,q})$ there exists a ϕ in $(\alpha, \chi_{w,q})$ with finite $D_\phi(w \| q)$. Then as a result of [21, Corollary 2.8.7-(ii)], $g(\alpha)$ is an infinitely differentiable function of α on $(0, \chi_{w,q})$ such that

$$\frac{\partial^\kappa}{\partial \alpha^\kappa} g(\alpha) = \int \left[\frac{\partial^\kappa}{\partial \alpha^\kappa} f(\alpha, y) \right] \nu(dy) \quad \forall \kappa \in \mathbb{Z}_{\geq 0}. \quad (\text{A.6})$$

Consequently, if $\chi_{w,q} > 1$, then

$$D_1(w \| q) = \frac{\partial}{\partial \alpha} \ln g(\alpha) \Big|_{\alpha=1}. \quad (\text{A.7})$$

Using (A.5) and (A.6) we get

$$\left| \frac{\partial^\kappa}{\partial \alpha^\kappa} g(\alpha) \right| \leq \frac{\kappa!}{\sqrt{2\pi\kappa}} \left(\frac{1}{\alpha^\kappa} + \frac{g(\phi)}{(\phi - \alpha)^\kappa} \right) \quad \forall \kappa \in \mathbb{Z}_+, \phi \in (\alpha, \chi_{w,q}). \quad (\text{A.8})$$

Thus $g(\alpha)$ is not only infinitely differentiable but also analytic in α on $(0, \chi_{w,q})$ by [38, Proposition 1.2.12]. On the other hand $g(\alpha) \in \mathbb{R}_+$ for all $\alpha \in (0, \chi_{w,q})$ because $g(\alpha) = e^{(\alpha-1)D_\alpha(w \| q)}$ by (A.3) and $D_\alpha(w \| q) \in \mathbb{R}_{\geq 0}$ by Lemmas 2 and 8 and the definition of $\chi_{w,q}$. Thus $\ln g(\alpha)$ is analytic in α on $(0, \chi_{w,q})$ because composition of analytic functions is analytic by [38, Proposition 1.4.2]. Then $D_\alpha(w \| q)$ is analytic in α on $(0, \chi_{w,q}) \setminus \{1\}$ because the quotient of analytic functions is analytic at points with open neighborhoods on which the function in the denominator is non-zero by [38, Proposition 1.1.12].

Now we proceed with establishing the analyticity of $D_\alpha(w \| q)$ at $\alpha = 1$ for $\chi_{w,q} > 1$ case. Since $\ln g(\alpha)$ is analytic in α on $(0, \chi_{w,q})$ we can write $\ln g(\alpha)$ as a convergent power series around any point in $(0, \chi_{w,q})$ for some neighborhood. Thus, there exists a $\delta > 0$ for which the following two identities hold for all $\eta \in (1 - \delta, 1 + \delta)$

$$\begin{aligned} \sum_{i=0}^{\infty} \frac{|\eta-1|^i}{i!} \left| \frac{\partial^i}{\partial \alpha^i} \ln g(\alpha) \Big|_{\alpha=1} \right| &< \infty, \\ \sum_{i=0}^{\infty} \frac{(\eta-1)^i}{i!} \frac{\partial^i}{\partial \alpha^i} \ln g(\alpha) \Big|_{\alpha=1} &= \ln g(\eta). \end{aligned}$$

Then using $\ln g(1) = 0$ together with (A.3) and (A.7) we get

$$D_\eta(w \| q) = D_1(w \| q) + \sum_{i=2}^{\infty} \frac{(\eta-1)^{i-1}}{i!} \frac{\partial^i}{\partial \alpha^i} \ln g(\alpha) \Big|_{\alpha=1} \quad \forall \eta \in (1 - \delta, 1 + \delta). \quad (\text{A.9})$$

Then $D_\eta(w \| q)$ is analytic on $(1 - \delta, 1 + \delta)$ by [38, Corollary 1.2.4] because it is equal to a function defined by a convergent power series.

The convergent power series given in (A.9) determines the derivatives of $D_\alpha(w \| q)$ at $\alpha = 1$ by [38, Corollary 1.1.16]:

$$\frac{\partial^\kappa}{\partial \alpha^\kappa} D_\alpha(w \| q) \Big|_{\alpha=1} = \frac{1}{\kappa+1} \frac{\partial^{\kappa+1}}{\partial \alpha^{\kappa+1}} \ln g(\alpha) \Big|_{\alpha=1} \quad \kappa \in \mathbb{Z}_+. \quad (\text{A.10})$$

Using (A.3) together with the elementary rules of differentiation we can express the derivatives of $D_\alpha(w\|q)$ in terms of the derivatives of $\ln g(\alpha)$ for other orders in $(0, \chi_{w,q})$, as well

$$\begin{aligned} \frac{\partial^\kappa}{\partial \alpha^\kappa} D_\alpha(w\|q) \Big|_{\alpha=\phi} &= \sum_{t=0}^{\kappa} \frac{\kappa!}{t!(\kappa-t)!} \left(\frac{\partial^{\kappa-t}}{\partial \alpha^{\kappa-t}} \frac{1}{\alpha-1} \Big|_{\alpha=\phi} \right) \left(\frac{\partial^t}{\partial \alpha^t} \ln g(\alpha) \Big|_{\alpha=\phi} \right) \\ &= \sum_{t=0}^{\kappa} \frac{\kappa!}{t!} \frac{(-1)^{\kappa-t}}{(\phi-1)^{\kappa-t+1}} \frac{\partial^t}{\partial \alpha^t} \ln g(\alpha) \Big|_{\alpha=\phi} \quad \kappa \in \mathbb{Z}_+, \phi \in (0, \chi_{w,q}) \setminus \{1\}. \end{aligned} \quad (\text{A.11})$$

On the other hand by Faà di Bruno formula for derivatives of the composition of smooth functions [38, Thm. 1.3.2] we have

$$\begin{aligned} \frac{\partial^t}{\partial \alpha^t} \ln g(\alpha) &= \sum_{j_t} \frac{t!}{j_1! j_2! \dots j_t!} \left(\frac{\partial^{j_1+j_2+\dots+j_t}}{\partial \tau^{j_1+j_2+\dots+j_t}} \ln \tau \Big|_{\tau=g(\alpha)} \right) \left(\frac{1}{1!} \frac{\partial^{j_1}}{\partial \alpha^{j_1}} g(\alpha) \right)^{j_1} \left(\frac{1}{2!} \frac{\partial^{j_2}}{\partial \alpha^{j_2}} g(\alpha) \right)^{j_2} \dots \left(\frac{1}{t!} \frac{\partial^{j_t}}{\partial \alpha^{j_t}} g(\alpha) \right)^{j_t} \\ &= \sum_{j_t} \frac{t!}{j_1! j_2! \dots j_t!} \frac{(-1)^{(j_1+j_2+\dots+j_t-1)!}}{(-g(\alpha))^{j_1+j_2+\dots+j_t}} \left(\frac{1}{1!} \frac{\partial^{j_1}}{\partial \alpha^{j_1}} g(\alpha) \right)^{j_1} \left(\frac{1}{2!} \frac{\partial^{j_2}}{\partial \alpha^{j_2}} g(\alpha) \right)^{j_2} \dots \left(\frac{1}{t!} \frac{\partial^{j_t}}{\partial \alpha^{j_t}} g(\alpha) \right)^{j_t} \quad \forall t \in \mathbb{Z}_+. \end{aligned}$$

Then using (A.1), (A.2), (A.4), and (A.6) we get

$$\frac{\partial^t}{\partial \alpha^t} \ln g(\alpha) = t! \sum_{j_t} \frac{(-1)^{(j_1+j_2+\dots+j_t-1)!}}{j_1! j_2! \dots j_t!} \prod_{i=1}^t \left(\frac{(-1)}{i!} \mathbf{E}_{w_\alpha^q} \left[\left(\ln \frac{dw}{d\nu} - \ln \frac{dq}{d\nu} \right)^i \right] \right)^{j_i} \quad \forall t \in \mathbb{Z}_+. \quad (\text{A.12})$$

The expression given in (13) for κ^{th} derivative of $D_\alpha(w\|q)$ with respect to α follows from the identity $\ln g(1) = 0$ and equations (A.3), (A.10), (A.11), and (A.12).

In order to prove the analyticity of $D_1(w_\alpha^q\|w)$ and $D_1(w_\alpha^q\|q)$, first note that as a result of (16), which follows from (13), we have

$$D_1(w_\phi^q\|w) = (\phi-1)^2 \frac{\partial}{\partial \alpha} D_\alpha(w\|q) \Big|_{\alpha=\phi} \quad \forall \phi \in (0, \chi_{w,q}). \quad (\text{A.13})$$

Since $D_\alpha(w\|q)$ is analytic in α on $(0, \chi_{w,q})$, so is $\frac{\partial}{\partial \alpha} D_\alpha(w\|q)$. Hence, $D_1(w_\alpha^q\|w)$ is analytic in α on $(0, \chi_{w,q})$. Since $D_1(w_\phi^q\|w)$ is analytic in ϕ on $(0, \chi_{w,q})$, it is finite on $(0, \chi_{w,q})$. Thus (12) holds for all α in $(0, \chi_{w,q})$ and (A.13) implies

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

Thus $D_1(w_\alpha^q\|q)$ is an analytic function of α on $(0, \chi_{w,q})$, as well. \square

Proof of Lemma 12. As results of (A.2), (A.4), (A.5), and Definition 2 we have

$$\mathbf{E}_{w_\alpha^q} \left[\left| \ln \frac{dw}{d\nu} - \ln \frac{dq}{d\nu} \right|^i \right] \leq \frac{e^{(1-\alpha)D_\alpha(w\|q)} i!}{\sqrt{2\pi i}} \left(\frac{1}{\alpha^i} + \frac{e^{(\beta-1)D_\beta(w\|q)}}{(\beta-\alpha)^i} \right) \quad \forall i \in \mathbb{Z}_+.$$

Then using $D_\beta(w\|q) \leq \gamma$ together with Lemma 8 we get

$$\mathbf{E}_{w_\alpha^q} \left[\left| \ln \frac{dw}{d\nu} - \ln \frac{dq}{d\nu} \right|^i \right] \leq \frac{i!}{\sqrt{i}} \frac{e^{(1\vee\beta)\gamma}}{(\alpha \wedge (\beta-\alpha))^i} \quad \forall i \in \mathbb{Z}_+.$$

Then using (13) and (15) we get,

$$\left| \frac{\partial^\kappa D_\alpha(w\|q)}{\partial \alpha^\kappa} \Big|_{\alpha=\phi} \right| \leq \begin{cases} \kappa! \frac{\gamma}{|\phi-1|^\kappa} + \kappa! \sum_{t=1}^{\kappa} \frac{1}{|\phi-1|^{\kappa-t+1}} \frac{1}{(\phi \wedge (\beta-\phi))^t} \sum_{j_t} \frac{(j_1+j_2+\dots+j_t-1)!}{j_1! j_2! \dots j_t!} (e^{(1\vee\beta)\gamma})^{j_1+j_2+\dots+j_t} & \phi \neq 1 \\ \kappa! \frac{1}{(1 \wedge (\beta-1))^{\kappa+1}} \sum_{j_{\kappa+1}} \frac{(j_1+j_2+\dots+j_{\kappa+1}-1)!}{j_1! j_2! \dots j_{\kappa+1}!} (e^{(1\vee\beta)\gamma})^{j_1+j_2+\dots+j_{\kappa+1}} & \phi = 1 \end{cases}.$$

On the other hand $\sum_{j_t} \frac{(j_1+j_2+\dots+j_t)!}{j_1! j_2! \dots j_t!} \xi^{j_1+j_2+\dots+j_t} = \xi(1+\xi)^{t-1}$ by [38, Thm. 1.4.1]. Thus we get the following inequality, which implies (18) for the τ defined in (20).

$$\left| \frac{\partial^\kappa D_\alpha(w\|q)}{\partial \alpha^\kappa} \Big|_{\alpha=\phi} \right| \leq \begin{cases} \kappa! \sum_{t=1}^{\kappa} \frac{1}{|\phi-1|^{\kappa-t+1}} \left(\frac{1+e^{(1\vee\beta)\gamma}}{\phi \wedge (\beta-\phi)} + \gamma \mathbf{1}_{\{t=1\}} \right)^t & \phi \neq 1 \\ \kappa! \left(\frac{1+e^{\beta\gamma}}{1 \wedge (\beta-1)} \right)^{\kappa+1} & \phi = 1 \end{cases}.$$

As a result of [38, Corollaries 1.2.4 and 1.2.5] the following equality holds on the open interval in which the power series on the right hand side is convergent,

$$D_\eta(w\|q) = \sum_{j=0}^{\infty} \frac{(\eta-\phi)^j}{j!} \frac{\partial^j}{\partial \alpha^j} D_\alpha(w\|q) \Big|_{\alpha=\phi}. \quad (\text{A.14})$$

Note that as a result of (18) we have

$$\limsup_{\kappa \rightarrow \infty} \sqrt[\kappa]{\frac{1}{\kappa!} \left| \frac{\partial^\kappa}{\partial \alpha^\kappa} D_\alpha(w\|q) \Big|_{\alpha=\phi} \right|} \leq \tau.$$

Thus radius of convergence of the power series on the right hand side of (A.14) is at least $\frac{1}{\tau}$ by [38, Lemma 1.1.8], i.e. by Hadamard formula. Thus for all $\eta \in (\phi - \frac{1}{\tau}, \phi + \frac{1}{\tau})$ using (18) and (A.14) we get

$$\left| D_\eta(w \| q) - \sum_{i=0}^{\kappa-1} \frac{(\eta-\phi)^i}{i!} \frac{\partial^i D_\alpha(w \| q)}{\partial \alpha^i} \Big|_{\alpha=\phi} \right| \leq \sum_{i=\kappa}^{\infty} |\eta - \phi|^i \tau^{i+1} (\mathbb{1}_{\{\phi=1\}} + \nu \mathbb{1}_{\{\phi \neq 1\}}).$$

Using identities $\sum_{i=0}^{\infty} z^i = \frac{1}{1-z}$ and $\sum_{i=0}^{\infty} (i+1)z^i = \frac{1}{(1-z)^2}$ for $|z| < 1$ we get,

$$\left| D_\eta(w \| q) - \sum_{i=0}^{\kappa-1} \frac{(\eta-\phi)^i}{i!} \frac{\partial^i D_\alpha(w \| q)}{\partial \alpha^i} \Big|_{\alpha=\phi} \right| \leq \frac{\tau^{\kappa+1} |\eta-\phi|^\kappa}{1-|\eta-\phi|^\tau} \left[\mathbb{1}_{\{\phi=1\}} + \left(\kappa - 1 + \frac{1}{1-|\eta-\phi|^\tau} \right) \mathbb{1}_{\{\phi \neq 1\}} \right].$$

□

B. Proofs of Lemmas on the Augustin Information

Proof of Lemma 13.

(13-a) $I_\alpha(p; W) \leq D_\alpha(W \| q | p)$ for all $q \in \mathcal{P}(\mathcal{Y})$ by definition. On the other hand, $D_\alpha(W(x) \| q_{1,p}) \leq -\ln p(x)$ for all x with positive $p(x)$ by Lemma 1 because $p(x)W(x) \leq q_{1,p}$. Hence, $I_\alpha(p; W) \leq -\sum_x p(x) \ln p(x)$.

(13-b) Note that as a result of Lemma 2 and (24),

$$D_1(W \| q | p) \geq D_1(W \| q_{1,p} | p) + \frac{1}{2} \|q_{1,p} - q\|^2 \quad \forall q \in \mathcal{P}(\mathcal{Y}).$$

Then $q_{1,p}$ is the unique probability measure satisfying $I_1(p; W) = D_1(W \| q_{1,p} | p)$. Then (29) follows from (24).

(13-c) Let \mathfrak{S} and ς be

$$\begin{aligned} \varsigma &\triangleq \min_{x:p(x)>0} p(x), \\ \mathfrak{S} &\triangleq \left\{ s \in \mathcal{M}^+(\mathcal{X}) : \varsigma \mathbb{1}_{\{p(x)>0\}} \leq s(x) \leq \left(e^{\frac{1-\alpha}{\varsigma} D_\alpha(W \| q_{\alpha,p}^g | p)} \right) \mathbb{1}_{\{p(x)>0\}} \quad \forall x \in \mathcal{X} \right\}. \end{aligned}$$

The statements proved in (c-i), (c-iv), (c-vi), and (c-vii) collectively imply part (c).

(c-i) If $q_{1,p} \prec u$ and $T_{\alpha,p}(u) = u$, then $D_\alpha(W \| u | p) = I_\alpha(p; W)$, (30) and (32) hold for $q_{\alpha,p} = u$, and $q_{\alpha,p}$ is unique: Note that $T_{\alpha,p}(u) = u$ and $q_{1,p} \prec u$ imply

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

Then one can confirm by substitution that

$$D_\alpha(u \| q) = \frac{1}{\alpha-1} \ln \sum_x p(x) e^{(\alpha-1)(D_\alpha(W(x) \| q) - D_\alpha(W(x) \| u))} \quad \forall q \in \mathcal{P}(\mathcal{Y}).$$

Then Jensen's inequality and convexity of the exponential function imply

$$D_\alpha(u \| q) \leq D_\alpha(W \| q | p) - D_\alpha(W \| u | p) \quad \forall q \in \mathcal{P}(\mathcal{Y}). \quad (\text{B.1})$$

Then u is the unique probability measure satisfying $I_\alpha(p; W) = D_\alpha(W \| u | p)$ by Lemma 2. Consequently (30) and the lower bound given in (32) hold.

In order to establish the upper bound given in (32) for $q \in \mathcal{Q}_{\alpha,p}$, first note that $W(x) \prec u$ for all x with a positive $p(x)$ because $q_{1,p} \prec u$. Thus for all x with positive $p(x)$ we have

$$\begin{aligned} D_\alpha(W(x) \| q) - D_\alpha(W(x) \| u) &= \frac{1}{\alpha-1} \left[\ln \int \left(\frac{dW(x)}{du} \right)^\alpha \left(\frac{dq_\sim}{du} \right)^{1-\alpha} u(dy) - (\alpha-1) D_\alpha(W(x) \| u) \right] \\ &= \frac{1}{\alpha-1} \ln \int \left(\frac{dq_\sim}{du} \right)^{1-\alpha} \left(\frac{dW(x)}{du} \right)^\alpha e^{(1-\alpha)D_\alpha(W(x) \| u)} u(dy) \\ &= \frac{1}{\alpha-1} \ln \int \left(\frac{dq_\sim}{du} \right)^{1-\alpha} \frac{dW_\alpha^u(x)}{du} u(dy) \\ &= \frac{1}{\alpha-1} \ln \int \left(\frac{dq_\sim}{du} \right)^{1-\alpha} W_\alpha^u(x) (dy) \quad \forall q \in \mathcal{Q}_{\alpha,p} \end{aligned}$$

where q_\sim is the component of q that is absolutely continuous in u . Consequently,

$$D_\alpha(W \| q | p) - D_\alpha(W \| u | p) = \frac{1}{\alpha-1} \sum_x p(x) \ln \int \left(\frac{dq_\sim}{du} \right)^{1-\alpha} W_\alpha^u(x) (dy) \quad \forall q \in \mathcal{Q}_{\alpha,p}. \quad (\text{B.2})$$

On the other hand using the Jensen's inequality and concavity of the natural logarithm function we get

$$\begin{aligned} \frac{1}{\alpha-1} \sum_x p(x) \ln \int \left(\frac{dq_\sim}{du} \right)^{1-\alpha} W_\alpha^u(x) (dy) &\leq \frac{1}{\alpha-1} \sum_x p(x) \int \left[\ln \left(\frac{dq_\sim}{du} \right)^{1-\alpha} \right] W_\alpha^u(x) (dy) \\ &= \int \left[\ln \left(\frac{du}{dq_\sim} \right) \right] T_{\alpha,p}(u) (dy). \quad (\text{B.3}) \end{aligned}$$

Since $T_{\alpha,p}(u) = u$ by the hypothesis, using (B.2) and (B.3) we get

$$D_1(u||q) \geq D_\alpha(W||q|p) - D_\alpha(W||u|p) \quad \forall q \in \mathcal{Q}_{\alpha,p}.$$

In order to establish the upper bound given in (32) for $q \notin \mathcal{Q}_{\alpha,p}$ we need to make the following additional observation. If $q \notin \mathcal{Q}_{\alpha,p}$, then there exists an x for which $p(x) > 0$ and $W(x) \perp q$ because $D_\alpha(W(x)||q) = \infty$ implies $W(x) \perp q$ by (11). As a result there exists an event $\mathcal{E} \in \mathcal{Y}$ such that $u(\mathcal{E}) > 0$ and $q(\mathcal{E}) = 0$ because $W(x) \prec_{q_{1,p}}$ and $q_{1,p} \prec u$. Consequently $D_1(u||q) = \infty$ and the upper bound in equation (32) holds for $q \notin \mathcal{Q}_{\alpha,p}$, as well.

(c-ii) $D_\alpha(W||q|p) - D_\alpha(W||T_{\alpha,p}(q)|p) \geq D_1(T_{\alpha,p}(q)||q)$ for all $q \in \mathcal{Q}_{\alpha,p}$. Note that $T_{\alpha,p}(q) \prec q$ for all $q \in \mathcal{Q}_{\alpha,p}$ by definition. Then

$$\begin{aligned} D_\alpha(W||q|p) - D_\alpha(W||T_{\alpha,p}(q)|p) &= \frac{1}{1-\alpha} \sum_x p(x) \ln \int \left(\frac{dT_{\alpha,p}(q)}{dq} \right)^{1-\alpha} W_\alpha^q(x) (dy) \\ &\geq \frac{1}{1-\alpha} \sum_x p(x) \int \left[\ln \left(\frac{dT_{\alpha,p}(q)}{dq} \right)^{1-\alpha} \right] W_\alpha^q(x) (dy) \\ &= D_1(T_{\alpha,p}(q)||q). \end{aligned} \quad (\text{B.4})$$

The inequality follows from the Jensen's inequality and the concavity of the natural logarithm function.

(c-iii) $\{T_{\alpha,p}^\iota(q_{\alpha,p}^g)\}_{\iota \in \mathbb{Z}_+}$ is totally bounded for total variation metric on $\mathcal{M}^+(\mathcal{Y})$: For any $q \in \mathcal{Q}_{\alpha,p}$, as a result of definitions of $T_{\alpha,p}(\cdot)$ and $\mu_{\alpha,p}$ we have

$$\frac{dT_{\alpha,p}(q)}{d\nu} = \left(\frac{d\mu_{\alpha,s}}{d\nu} \right)^\alpha \left(\frac{dq}{d\nu} \right)^{1-\alpha}$$

where $s(x) = p(x)e^{(1-\alpha)D_\alpha(W(x)||q)}$. Furthermore, if $D_\alpha(W||q|p) \leq D_\alpha(W||q_{\alpha,p}^g|p)$, then $s \in \mathcal{S}$. In addition $q_{\alpha,p}^g$ is equal to $\mu_{\alpha,s}$ for an $s \in \mathcal{S}$. In particular

$$q_{\alpha,p}^g = \mu_{\alpha,s_0}$$

where $s_0 = \|\mu_{\alpha,p}\|^{-\alpha} p$. One can confirm by substitution that $\|\mu_{\alpha,p}\|^{-\alpha} = e^{(1-\alpha)D_\alpha(p \otimes W || p \otimes q_{\alpha,p}^g)}$. Furthermore, $D_\alpha(p \otimes W || p \otimes q_{\alpha,p}^g) \geq 0$ by Lemma 2 and $D_\alpha(p \otimes W || p \otimes q_{\alpha,p}^g) \leq D_\alpha(W || q_{\alpha,p}^g | p)$ by the Jensen's inequality and the concavity of the natural logarithm function. Thus $s_0 \in \mathcal{S}$.

On the other hand, $D_\alpha(W || T_{\alpha,p}^\iota(q_{\alpha,p}^g) | p) \leq D_\alpha(W || q_{\alpha,p}^g | p)$ for all $\iota \geq \mathbb{Z}_+$. Thus we can write $T_{\alpha,p}^\iota(q_{\alpha,p}^g)$ in terms of the elements of $\mu_{\alpha,\mathcal{S}}$ as follows:

$$\frac{dT_{\alpha,p}^\iota(q_{\alpha,p}^g)}{d\nu} = \left(\frac{d\mu_{\alpha,s_0}}{d\nu} \right)^{(1-\alpha)^\iota} \prod_{j=1}^{\iota} \left(\frac{d\mu_{\alpha,s_j}}{d\nu} \right)^{\alpha(1-\alpha)^{\iota-j}}$$

where $s_j(x) = p(x)e^{(1-\alpha)D_\alpha(W(x)||T_{\alpha,p}^{\iota-1}(q_{\alpha,p}^g))}$.

In order to prove that $\{T_{\alpha,p}^\iota(q_{\alpha,p}^g)\}_{\iota \in \mathbb{Z}_+}$ is totally bounded, we prove that a superset of it, i.e. \mathcal{B} defined in the following, is totally bounded.

$$\mathcal{B}_\iota \triangleq \left\{ b \in \mathcal{M}^+(\mathcal{Y}) : \frac{db}{d\nu} = \left(\frac{d\mu_{\alpha,s_0}}{d\nu} \right)^{(1-\alpha)^\iota} \prod_{j=1}^{\iota} \left(\frac{d\mu_{\alpha,s_j}}{d\nu} \right)^{\alpha(1-\alpha)^{\iota-j}} \text{ for some } s_j \in \mathcal{S} \right\} \quad (\text{B.5})$$

$$\mathcal{B} \triangleq \cup_{\iota \in \mathbb{Z}_+} \mathcal{B}_\iota. \quad (\text{B.6})$$

Let us denote the number of x 's with $p(x) > 0$ by κ . Then \mathcal{S} is isometric to a cube in²³ \mathbb{R}^κ . We divide each side of the cube into n equal length intervals. Thus \mathcal{S} is composed of n^κ sub-cubes. Furthermore, $\mu_{\alpha,s} \leq \mu_{\alpha,\tilde{s}}$ whenever $s \leq \tilde{s}$ by definition. Thus, for any $s \in \mathcal{S}$ we have

$$\mu_{\alpha,\lfloor s \rfloor_n} \leq \mu_{\alpha,s} \leq \left[1 + \frac{e^{\frac{1-\alpha}{\varsigma} D_\alpha(W || q_{\alpha,p}^g | p) - \varsigma}}{\varsigma n} \right]^{\frac{1}{\alpha}} \mu_{\alpha,\lfloor s \rfloor_n}$$

where $\lfloor s \rfloor_n$ is the corner point that satisfies $\lfloor s \rfloor_n \leq \tilde{s}$ for all \tilde{s} in the sub-cube for the sub-cube that s is in.

In order to approximate members of \mathcal{B}_ι one can use the preceding discretization on each s_j given in definition \mathcal{B}_ι . Thus we have $n^{(\iota+1)\kappa}$ point set $\mathcal{K}_{\iota,n}$ such that:

$$\forall b \in \mathcal{B}_\iota \exists \mu \in \mathcal{K}_{\iota,n} \text{ such that } \mu \leq b \leq \left[1 + \frac{e^{\frac{1-\alpha}{\varsigma} D_\alpha(W || q_{\alpha,p}^g | p) - \varsigma}}{\varsigma n} \right]^{\frac{1}{\alpha}} \mu.$$

²³We assume \mathbb{R}^κ has the metric $d : \mathbb{R}^\kappa \times \mathbb{R}^\kappa \rightarrow \mathbb{R}_{\geq 0}$ given by $d(z, \tilde{z}) = \sum_{i=1}^{\kappa} |z_i - \tilde{z}_i|$ for all $z, \tilde{z} \in \mathbb{R}^\kappa$.

One can use the points of $\mathcal{K}_{i,n}$ to approximate the points in $\cup_{t>i} \mathcal{B}_t$, as well. We apply the approximation with the sub-cubes described above for the last i components of b , i.e. for i s_j 's with the largest indices. The remaining component of μ is set to the minimum element of²⁴ $\mu_{\alpha,S}$. Then

$$\forall b \in \cup_{t>i} \mathcal{B}_t \exists \mu \in \mathcal{K}_{i,n} \text{ such that } \mu \leq b \leq \left[1 + \frac{e^{\frac{1-\alpha}{\varsigma} D_\alpha(W \| q_{\alpha,p}^g | p)} - \varsigma}{\varsigma n} \right]^{\frac{1-(1-\alpha)^i}{\alpha}} \left[\frac{e^{\frac{1-\alpha}{\varsigma} D_\alpha(W \| q_{\alpha,p}^g | p)}}{\varsigma} \right]^{\frac{(1-\alpha)^i}{\alpha}} \mu.$$

Let \mathcal{K}_n be $\mathcal{K}_n = \cup_{j \in \{0, \dots, n\}} \mathcal{K}_{j,n}$. Then

$$\forall b \in \mathcal{B} \exists \mu \in \mathcal{K}_n \text{ such that } \|b - \mu\| \leq \left(\left[1 + \frac{e^{\frac{1-\alpha}{\varsigma} D_\alpha(W \| q_{\alpha,p}^g | p)} - \varsigma}{\varsigma n} \right]^{\frac{1}{\alpha}} \left[\frac{e^{\frac{1-\alpha}{\varsigma} D_\alpha(W \| q_{\alpha,p}^g | p)}}{\varsigma} \right]^{\frac{(1-\alpha)^n}{\alpha}} - 1 \right) \sup_{s \in \mathcal{S}} \|\mu_{\alpha,s}\|.$$

Note that $\sup_{s \in \mathcal{S}} \|\mu_{\alpha,s}\|$ is finite and its coefficient converges to zero as n diverges. Furthermore, \mathcal{K}_n is a finite set for any n . Thus \mathcal{B} is totally bounded. As a result every subset of \mathcal{B} , and hence $\{\mathbb{T}_{\alpha,p}^i(q_{\alpha,p}^g)\}_{i \in \mathbb{Z}_+}$, is totally bounded.

- (c-iv) $\{\mathbb{T}_{\alpha,p}^i(q_{\alpha,p}^g)\}_{i \in \mathbb{Z}_+}$ has a subsequence $\{\mathbb{T}_{\alpha,p}^{i(j)}(q_{\alpha,p}^g)\}_{j \in \mathbb{Z}_+}$ satisfying $\lim_{j \rightarrow \infty} \|\mathbb{T}_{\alpha,p}^{i(j)}(q_{\alpha,p}^g) - u\| = 0$ for a $u \sim q_{1,p}$. The existence of a limit point u and convergent subsequence follow from the compactness of the completion of $\{\mathbb{T}_{\alpha,p}^i(q_{\alpha,p}^g)\}_{i \in \mathbb{Z}_+}$. The completion is compact by [39, Thm. 45.1] because $\{\mathbb{T}_{\alpha,p}^i(q_{\alpha,p}^g)\}_{i \in \mathbb{Z}_+}$ is totally bounded. Note that $\mathbb{T}_{\alpha,p}^i(q_{\alpha,p}^g) \prec q_{1,p}$ because $q_{\alpha,p}^g \sim q_{1,p}$. Then $u \prec q_{1,p}$ because any probability measure that is not absolute continuous in $q_{1,p}$ is outside the closure of $\{\mathbb{T}_{\alpha,p}^i(q_{\alpha,p}^g)\}_{i \in \mathbb{Z}_+}$. On the other hand, $\mu_{\alpha,p} \leq q_{\alpha,p}^g$ by definition because $\|\mu_{\alpha,p}\| \leq 1$. Furthermore, for any $q \in \mathcal{Q}_{\alpha,p}$ we have

$$\begin{aligned} \frac{d\mathbb{T}_{\alpha,p}(q)}{d\nu} &\geq \sum_x p(x) \left(\frac{dW(x)}{d\nu} \right)^\alpha \left(\frac{dq}{d\nu} \right)^{1-\alpha} \\ &= \left(\frac{d\mu_{\alpha,p}}{d\nu} \right)^\alpha \left(\frac{dq}{d\nu} \right)^{1-\alpha}. \end{aligned}$$

Hence, if $\mu_{\alpha,p} \leq q$, then $\mu_{\alpha,p} \leq \mathbb{T}_{\alpha,p}(q)$. Consequently, $\mu_{\alpha,p} \leq \mathbb{T}_{\alpha,p}^i(q_{\alpha,p}^g)$ for all $i \in \mathbb{Z}_+$. Hence $\mu_{\alpha,p} \leq u$, because otherwise u can not be in the closure of $\{\mathbb{T}_{\alpha,p}^i(q_{\alpha,p}^g)\}_{i \in \mathbb{Z}_+}$. Then $q_{1,p} \prec u$ because $q_{1,p} \sim \mu_{\alpha,p}$.

- (c-v) $\mathbb{T}_{\alpha,p}(\cdot) : \mathcal{Q}_{\alpha,p} \rightarrow \mathcal{P}(\mathcal{Y})$ is continuous if both $\mathcal{Q}_{\alpha,p}$ and $\mathcal{P}(\mathcal{Y})$ have the total variation topology: First, note that $(z+t)^{1-\alpha} - z^{1-\alpha}$ is a monotonically decreasing function of z on $\mathbb{R}_{\geq 0}$ for fixed $t \in \mathbb{R}_{\geq 0}$ and $\alpha \in (0, 1)$. Then for any x with positive $p(x)$ and $q_1, q_2 \in \mathcal{Q}_{\alpha,p}$ as a result of Holder's inequality we have

$$\begin{aligned} \int \left| \left(\frac{dW(x)}{d\nu} \right)^\alpha \left(\frac{dq_1}{d\nu} \right)^{1-\alpha} - \left(\frac{dW(x)}{d\nu} \right)^\alpha \left(\frac{dq_2}{d\nu} \right)^{1-\alpha} \right| \nu(dy) &\leq \int \left(\frac{dW(x)}{d\nu} \right)^\alpha \left| \frac{dq_1}{d\nu} - \frac{dq_2}{d\nu} \right|^{1-\alpha} \nu(dy) \\ &\leq \|q_1 - q_2\|^{1-\alpha}. \end{aligned}$$

Hence $e^{(\alpha-1)D_\alpha(W(x) \| q)} W_\alpha^q(x)$ is a continuous function of q from $\mathcal{Q}_{\alpha,p}$ to $\mathcal{M}^+(\mathcal{Y})$ for the total variation topology. Then $W_\alpha^q(x)$ is a continuous function of q for the total variation topology, as well, because $D_\alpha(W(x) \| q)$ is continuous in q for the total variation topology by Lemma 4. Thus $\mathbb{T}_{\alpha,p}(\cdot) : \mathcal{Q}_{\alpha,p} \rightarrow \mathcal{P}(\mathcal{Y})$ is continuous.

- (c-vi) The limit point of the convergent subsequence $\{\mathbb{T}_{\alpha,p}^{i(j)}(q_{\alpha,p}^g)\}_{j \in \mathbb{Z}_+}$ is a fixed point of $\mathbb{T}_{\alpha,p}(\cdot)$, i.e. $\mathbb{T}_{\alpha,p}(u) = u$: Using the non-negativity of the Rényi divergence for probability measures and (B.4) we get

$$\begin{aligned} D_\alpha(W \| q_{\alpha,p}^g | p) &\geq \sum_{i \in \mathbb{Z}_{\geq 0}} D_\alpha(W \| \mathbb{T}_{\alpha,p}^i(q_{\alpha,p}^g) | p) - D_\alpha(W \| \mathbb{T}_{\alpha,p}(\mathbb{T}_{\alpha,p}^i(q_{\alpha,p}^g)) | p) \\ &\geq \sum_{i \in \mathbb{Z}_{\geq 0}} D_1(\mathbb{T}_{\alpha,p}(\mathbb{T}_{\alpha,p}^i(q_{\alpha,p}^g)) \| \mathbb{T}_{\alpha,p}^i(q_{\alpha,p}^g)). \end{aligned}$$

Then $\lim_{i \rightarrow \infty} D_1(\mathbb{T}_{\alpha,p}(\mathbb{T}_{\alpha,p}^i(q_{\alpha,p}^g)) \| \mathbb{T}_{\alpha,p}^i(q_{\alpha,p}^g)) = 0$. Hence $\lim_{j \rightarrow \infty} D_1(\mathbb{T}_{\alpha,p}(\mathbb{T}_{\alpha,p}^{i(j)}(q_{\alpha,p}^g)) \| \mathbb{T}_{\alpha,p}^{i(j)}(q_{\alpha,p}^g)) = 0$. On the other hand, $D_1(\mathbb{T}_{\alpha,p}(q) \| q)$ is lower semicontinuous in q for the total variation topology because the Rényi divergence is lower semicontinuous in its arguments for the topology of setwise convergence—and hence to the total variation topology—by Lemma 3 and $\mathbb{T}_{\alpha,p}(\cdot)$ is continuous in the total variation topology. Then $D_1(\mathbb{T}_{\alpha,p}(u) \| u) = 0$ because $\mathbb{T}_{\alpha,p}^{i(j)}(q_{\alpha,p}^g)$ converges to u in total variation topology as j diverges. Thus $\mathbb{T}_{\alpha,p}(u) = u$ as a result of Lemma 2.

- (c-vii) $q_{\alpha,p}$ satisfies (31): Recall that $D_\alpha(w \| q)$ is continuous in q for the total variation topology by Lemma 4. Furthermore, $\lim_{j \rightarrow \infty} \|\mathbb{T}_{\alpha,p}^{i(j)}(q_{\alpha,p}^g) - q_{\alpha,p}\| = 0$, and $D_\alpha(W \| q_{\alpha,p} | p) = I_\alpha(p; W)$. Then

$$\lim_{j \rightarrow \infty} D_\alpha(W \| \mathbb{T}_{\alpha,p}^{i(j)}(q_{\alpha,p}^g) | p) = I_\alpha(p; W).$$

²⁴Such a minimum element might not exist for an arbitrary set of measures, but for the image of \mathcal{S} it exists: the minimum element is the image of the minimum point of \mathcal{S} .

On the other hand $D_\alpha(W \| \mathbb{T}_{\alpha,p}^i(q_{\alpha,p}^g) | p) \geq D_\alpha(W \| \mathbb{T}_{\alpha,p}^{i+t}(q_{\alpha,p}^g) | p) \geq I_\alpha(p; W)$ for all $t \in \mathbb{Z}_+$ by (B.4) and the definition of the Augustin information. Thus

$$\lim_{i \rightarrow \infty} D_\alpha(W \| \mathbb{T}_{\alpha,p}^i(q_{\alpha,p}^g) | p) = I_\alpha(p; W).$$

Then as a result of (32), which is implied by the assertions we have already established, we have

$$\lim_{i \rightarrow \infty} D_\alpha(q_{\alpha,p} \| \mathbb{T}_{\alpha,p}^i(q_{\alpha,p}^g)) = 0.$$

Then $\lim_{i \rightarrow \infty} \|q_{\alpha,p} - \mathbb{T}_{\alpha,p}^i(q_{\alpha,p}^g)\| = 0$ as a result of Lemma 2.

Remark 6. For any q satisfying $q \sim \mu_{\alpha,p}$ with a finite $\text{ess sup}_{\mu_{\alpha,p}} \left| \ln \frac{dq}{d\mu_{\alpha,p}} \right|$, we can define the sets \mathcal{S} and \mathcal{B}_i as follows

$$\begin{aligned} \mathcal{S} &\triangleq \left\{ s \in \mathcal{M}^+(\mathcal{X}) : s \mathbb{1}_{\{p(x)>0\}} \leq s(x) \leq \left(e^{\frac{1-\alpha}{\zeta} D_\alpha(W \| q | p)} \right) \mathbb{1}_{\{p(x)>0\}} \quad \forall x \in \mathcal{X} \right\}, \\ \mathcal{B}_i &\triangleq \left\{ b \in \mathcal{M}^+(\mathcal{Y}) : \frac{db}{d\nu} = (e^{\gamma \frac{dq}{d\nu}})^{(1-\alpha)^i} \prod_{j=1}^i \left(\frac{d\mu_{\alpha,s_j}}{d\nu} \right)^{\alpha(1-\alpha)^{i-j}} \text{ for some } \gamma \in \{-\Gamma, 0, \Gamma\} \text{ and } s_j \in \mathcal{S} \right\}, \end{aligned}$$

where $\Gamma = \frac{(1-\alpha)D_\alpha(W \| q | p)}{\zeta} - \frac{\ln \zeta}{\alpha} + \text{ess sup}_{\mu_{\alpha,p}} \left| \ln \frac{dq}{d\mu_{\alpha,p}} \right|$. Then one can confirm that $e^{-\Gamma} q \leq \mu_{\alpha,s} \leq e^\Gamma q$ for all $s \in \mathcal{S}$. Using this property, we can repeat the rest of the analysis with appropriate modifications to establish the following:

$$\lim_{j \rightarrow \infty} \|q_{\alpha,p} - \mathbb{T}_{\alpha,p}^j(q)\| = 0 \quad \text{if } q \sim \mu_{\alpha,p} \text{ and } \text{ess sup}_{\mu_{\alpha,p}} \left| \ln \frac{dq}{d\mu_{\alpha,p}} \right| < \infty, \quad (\text{B.7})$$

On the other hand, $q_{1,p} \sim \mu_{\alpha,p}$ by [13, Lemma 1-(a)] and $\left| \ln \frac{dq_{1,p}}{d\mu_{\alpha,p}} \right| \leq \frac{|\alpha-1|}{\alpha} \ln \frac{1}{\zeta}$ holds $q_{1,p}$ -a.s. by [13, Lemma 2-(a)]. Thus $q_{1,p}$ satisfies the condition given in (B.7) and the convergence described in (B.7) is equivalent to the one in (37).

(13-d) Let the function $f(\cdot)$ and the set of channels \mathcal{U} be

$$\begin{aligned} f(V) &\triangleq \frac{\alpha}{1-\alpha} D_1(V \| W | p) + I_1(p; V) & \forall V \in \mathcal{U}, \\ \mathcal{U} &\triangleq \{ V \in \mathcal{P}(\mathcal{Y} | \text{supp}(p)) : D_1(V \| W | p) < \infty \}. \end{aligned}$$

The statements proved in (d-i), (d-iv), (d-v), and (d-vi) collectively imply part (d).

(d-i) If $\mathbb{T}_{\alpha,p}(u) = u$, then $D_\alpha(W \| u | p) = I_\alpha(p; W)$, (33) and (34) hold for $q_{\alpha,p} = u$, $q_{\alpha,p}$ is unique and $q_{\alpha,p} \sim q_{1,p}$:

$$D_\alpha(W \| q | p) - D_\alpha(W \| u | p) = \frac{1}{\alpha-1} \sum_x p(x) \ln \int \left(\frac{dq_{\sim}}{du} \right)^{1-\alpha} W_\alpha^u(x) (dy) \quad \forall q \in \mathcal{Q}_{\alpha,p} \quad (\text{B.8})$$

where q_{\sim} is the component of q that is absolutely continuous in u .

On the other hand using the Jensen's inequality and concavity of the natural logarithm function we get

$$\begin{aligned} \frac{1}{\alpha-1} \sum_x p(x) \ln \int \left(\frac{dq_{\sim}}{du} \right)^{1-\alpha} W_\alpha^u(x) (dy) &\geq \frac{1}{\alpha-1} \sum_x p(x) \int \left[\ln \left(\frac{dq_{\sim}}{du} \right)^{1-\alpha} \right] W_\alpha^u(x) (dy) \\ &= \int \left[\ln \left(\frac{du}{dq_{\sim}} \right) \right] \mathbb{T}_{\alpha,p}(u) (dy). \end{aligned} \quad (\text{B.9})$$

Since $\mathbb{T}_{\alpha,p}(u) = u$ by the hypothesis, using (B.8) and (B.9) we get

$$D_\alpha(W \| q | p) - D_\alpha(W \| u | p) \geq D_1(u \| q) \quad \forall q \in \mathcal{Q}_{\alpha,p}. \quad (\text{B.10})$$

$D_1(u \| q) > 0$ for all $q \in \mathcal{P}(\mathcal{Y}) \setminus \{u\}$ by Lemma 2 and $D_\alpha(W \| q | p) = \infty$ for $q \notin \mathcal{Q}_{\alpha,p}$ by definition. Then u is the unique probability measure satisfying $I_\alpha(p; W) = D_\alpha(W \| u | p)$ and (33) holds. In addition $q_{1,p} \prec u$ because otherwise $D_\alpha(W \| u | p)$ would have been infinite. Furthermore, $u \prec q_{1,p}$ because $D_1(u \| q_{1,p})$ is finite by (B.10) and part (a).

The lower bound given in (34) holds for $q \in \mathcal{Q}_{\alpha,p}$ by (B.10) and for $q \notin \mathcal{Q}_{\alpha,p}$ by definition. In order to establish the upper bound given in (32), note that $\mathbb{T}_{\alpha,p}(u) = u$ implies

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

Then one can confirm by substitution that

$$D_\alpha(u \| q) = \frac{1}{\alpha-1} \ln \sum_x p(x) e^{(\alpha-1)(D_\alpha(W(x) \| q) - D_\alpha(W(x) \| u))} \quad \forall q \in \mathcal{P}(\mathcal{Y}).$$

Then Jensen's inequality and convexity of the exponential function imply

$$D_\alpha(u \| q) \geq D_\alpha(W \| q | p) - D_\alpha(W \| u | p) \quad \forall q \in \mathcal{P}(\mathcal{Y}).$$

(d-ii) $f(\cdot) : \mathcal{U} \rightarrow \mathbb{R}$ is concave and upper semicontinuous on \mathcal{U} for the topology of setwise convergence:²⁵ Using the definition of the tilted channel given in (22), the identity given in (29), and the joint convexity of the order one Rényi divergence in its arguments, i.e. Lemma 6, we can write $f(V)$ as the sum of three finite terms as follows for all $V \in \mathcal{U}$:

$$f(V) = \frac{1}{1-\alpha} D_1(V \| W_\alpha^{q_{1,p}} | p) - D_1\left(\sum_x p(x) V(x) \middle| q_{1,p}\right) + D_\alpha(W \| q_{1,p} | p). \quad (\text{B.11})$$

Then $f(\cdot)$ is a concave because the order one Rényi divergence is convex in its first argument by Lemma 6. Similarly, $f(\cdot)$ is upper semicontinuity for the topology of setwise convergence, because Rényi divergence is lower semicontinuous in its first argument for the topology of setwise convergence by Lemma 3.

(d-iii) $\mathcal{U}' \triangleq \{V \in \mathcal{U} : \max_{x:p(x)>0} p(x) D_1(V(x) \| W(x)) \leq \frac{\alpha-1}{\alpha} \hbar(p)\}$ is compact for the topology of setwise convergence: For any $v \in \mathcal{P}(\mathcal{Y})$ and $w \in \mathcal{P}(\mathcal{Y})$, the identity $z \ln z \geq -1/e$ implies that

$$\int \left| \frac{dv}{dw} \ln \frac{dv}{dw} \right|^+ w(dy) \leq D_1(v \| w) + 1/e.$$

Then for any $\gamma \in \mathbb{R}_+$ and $w \in \mathcal{P}(\mathcal{Y})$, the set of Radon-Nikodym derivatives $\{\frac{dv}{dw}\}_{v:D_1(v\|w)\leq\gamma}$ is uniformly integrable because it satisfies the necessary and sufficient condition for the uniform integrability given by de la Vallée Poussin [21, Thm. 4.5.9]. Hence, $\{v \in \mathcal{P}(\mathcal{Y}) : D_1(v \| w) \leq \gamma\} \prec^{uni} w$. Then²⁶ $\{v \in \mathcal{P}(\mathcal{Y}) : D_1(v \| w) \leq \gamma\}$ has compact closure in the topology of setwise convergence by [21, Thm. 4.7.25]. On the other hand the set $\{v \in \mathcal{P}(\mathcal{Y}) : D_1(v \| w) \leq \gamma\}$ is closed, i.e. it is equal to its closure, because Rényi divergence is lower semicontinuous in its arguments for the topology of setwise convergence by Lemma 3. Hence $\{v \in \mathcal{P}(\mathcal{Y}) : D_1(v \| w) \leq \gamma\}$ is compact in the topology of setwise convergence for any $\gamma \in \mathbb{R}_+$ and $w \in \mathcal{P}(\mathcal{Y})$. Then \mathcal{U}' is compact in the topology of setwise convergence because product of finite number of compact sets is compact by [39, Thm. 26.7].

(d-iv) $\exists U_* \in \mathcal{U}'$ s.t. $f(U_*) = \sup_{V \in \mathcal{U}} f(V)$: Note that $W \in \mathcal{U}$ and $f(W) = I_1(p; W)$. Furthermore, $I_1(p; W) \geq 0$ by Lemma 2 and part (b). On the other hand, if $p(x) D_1(V(x) \| W(x)) > \frac{\alpha-1}{\alpha} \hbar(p)$ for an x , then $f(V) < 0$ because $D_1(V(x) \| W(x)) \geq 0$ by Lemma 2 and $I_1(p; V) \leq \hbar(p)$ by part (a). Thus,

$$\sup_{V \in \mathcal{U}} f(V) = \sup_{V \in \mathcal{U}'} f(V).$$

On the other hand, $\exists U_*$ such that $f(U_*) = \sup_{V \in \mathcal{U}'} f(V)$ by the extreme value theorem for the upper semicontinuous functions [32, Ch3§12.2] because \mathcal{U}' is compact and $f(\cdot)$ is upper semicontinuous for the topology of setwise convergence.

(d-v) $f(U_*) = D_\alpha(W \| u_* | p)$ where $u_* \triangleq \sum_x p(x) U_*(x)$: As a result of Lemma 10 we have

$$D_\alpha(W \| u_* | p) = \sup_{V \in \mathcal{U}} \frac{\alpha}{1-\alpha} D_1(V \| W | p) + D_1(V \| u_* | p). \quad (\text{B.12})$$

On the other hand $\frac{\alpha}{1-\alpha} D_1(U_* \| W | p) + D_1(U_* \| u_* | p) = f(U_*)$ because $I_1(p; U_*) = D_1(U_* \| u_* | p)$ by part (b). Then $D_\alpha(W \| u_* | p) \geq f(U_*)$ is evident by (B.12).

In order to prove $D_\alpha(W \| u_* | p) \leq f(U_*)$, let us consider a $V \in \mathcal{U}$ and define $V^{(\iota)}$ and $q^{(\iota)}$ for each $\iota \in \mathbb{Z}_+$ as

$$\begin{aligned} V^{(\iota)} &\triangleq \frac{\iota-1}{\iota} U_* + \frac{1}{\iota} V, \\ q^{(\iota)} &\triangleq \frac{\iota-1}{\iota} u_* + \frac{1}{\iota} \sum_x p(x) V(x). \end{aligned}$$

As a result of the decomposition given in (B.11) we have

$$f(V^{(\iota)}) = \frac{1}{1-\alpha} D_1\left(V^{(\iota)} \middle| W_\alpha^{q_{1,p}} \middle| p\right) - D_1\left(q^{(\iota)} \middle| q_{1,p}\right) + D_\alpha(W \| q_{1,p} | p).$$

Then using the Jensen's inequality and convexity of the order one Rényi divergence in its first argument established in Lemma 6 we get

$$\begin{aligned} f(V^{(\iota)}) &\geq \frac{1}{1-\alpha} \left[\frac{\iota-1}{\iota} D_1(U_* \| W_\alpha^{q_{1,p}} | p) + \frac{1}{\iota} D_1(V \| W_\alpha^{q_{1,p}} | p) \right] - D_1\left(q^{(\iota)} \middle| q_{1,p}\right) + D_\alpha(W \| q_{1,p} | p) \\ &= \frac{\iota-1}{\iota} [f(U_*) + D_1(u_* \| q_{1,p})] + \frac{1}{\iota} \left[\frac{\alpha}{1-\alpha} D_1(V \| W | p) + D_1(V \| q_{1,p} | p) \right] - D_1\left(q^{(\iota)} \middle| q_{1,p}\right) \\ &= \frac{\iota-1}{\iota} \left[f(U_*) + D_1(u_* \| q^{(\iota)}) \right] + \frac{1}{\iota} \left[\frac{\alpha}{1-\alpha} D_1(V \| W | p) + D_1(V \| q^{(\iota)} | p) \right]. \end{aligned}$$

Then using $f(U_*) = \sup_{V \in \mathcal{U}} f(V) \geq f(V^{(\iota)})$ and $D_1(u_* \| q^{(\iota)}) \geq 0$ we get

$$f(U_*) \geq \frac{\alpha}{1-\alpha} D_1(V \| W | p) + D_1(V \| q^{(\iota)} | p) \quad \forall \iota \in \mathbb{Z}_+.$$

²⁵The set \mathcal{U} is a subset of the Cartesian product of a finite number of copies of $\mathcal{P}(\mathcal{Y})$. What we mean by the topology of setwise convergence on \mathcal{U} is the product topology obtained by assuming topology of setwise convergence on each component of the Cartesian product. We employ this terminology in the rest of the proof without explicitly mentioning it.

²⁶Note that $\{v \in \mathcal{P}(\mathcal{Y}) : D_1(v \| w) \leq \gamma\}$ is bounded in variation norm by definition.

On the other hand, $\lim_{i \rightarrow \infty} D_1(V \| q^{(i)} | p) \geq D_1(V \| u_* | p)$ because $\lim_{i \rightarrow \infty} \|q^{(i)} - u_*\| = 0$ by construction and the Rényi divergence is lower semicontinuous in its second argument for the topology of setwise convergence by Lemma 3. Then

$$f(U_*) \geq \frac{\alpha}{1-\alpha} D_1(V \| W | p) + D_1(V \| u_* | p) \quad \forall V \in \mathcal{U}.$$

Hence, $f(U_*) \geq D_\alpha(W \| u_* | p)$ by (B.12).

(d-vi) $T_{\alpha,p}(u_*) = u_*$ and $U_*(x) = W_\alpha^{u_*}(x)$ for all x such that $p(x) > 0$: Note that $D_\alpha(W \| u_* | p) < \infty$ by part (d-v) because $f(U_*) \leq \tilde{h}(p) < \infty$ by definition. Consequently, we can define the tilted probability measure $W_\alpha^{u_*}(x)$ for each x such that $p(x) > 0$. Using the definitions of $f(\cdot)$ and the tilted channel W_α^q together with the identity $I_1(p; U_*) = D_1(U_* \| u_* | p)$, which follows from part (b), we get

$$f(U_*) = D_\alpha(W \| u_* | p) + \frac{1}{1-\alpha} D_1(U_* \| W_\alpha^{u_*} | p).$$

Since $f(U_*) = D_\alpha(W \| u_* | p)$ by part (d-v) we get $D_1(U_* \| W_\alpha^{u_*} | p) = 0$. Hence $U_*(x) = W_\alpha^{u_*}(x)$ for all x such that $p(x) > 0$ by Lemma 2. As a result $T_{\alpha,p}(u_*) = u_*$ because $T_{\alpha,p}(u_*) = \sum_x p(x) W_\alpha^{u_*}(x)$ and $u_* = \sum_x p(x) U_*(x)$ by definition.

(13-e) We prove the statement for $\alpha \in (0, 1)$ and $\alpha \in (1, \infty)$ cases separately,

- Recall that $\sum_x p(x) W_\alpha^{q_{\alpha,p}}(x) = q_{\alpha,p}$ by parts (c). Then as a result of (29) we have

$$I_1(p; W_\alpha^{q_{\alpha,p}}) = D_1(W_\alpha^{q_{\alpha,p}} \| q_{\alpha,p} | p). \quad (\text{B.13})$$

Then (35) follows from Lemma 10 and $I_\alpha(p; W) = D_\alpha(W \| q_{\alpha,p} | p)$ for $\alpha \in (0, 1)$.

On the other hand as a result the definition of the Augustin information, and Lemma 10 we have

$$\begin{aligned} I_\alpha(p; W) &= \inf_{q \in \mathcal{P}(\mathcal{Y})} \inf_{V \in \mathcal{P}(\mathcal{Y}|X)} \frac{\alpha}{1-\alpha} D_1(V \| W | p) + D_1(V \| q | p) \\ &= \inf_{V \in \mathcal{P}(\mathcal{Y}|X)} \inf_{q \in \mathcal{P}(\mathcal{Y})} \frac{\alpha}{1-\alpha} D_1(V \| W | p) + D_1(V \| q | p). \end{aligned}$$

Then (36) follows from the definition of the order one Augustin information.

- Note that for $\alpha \in (1, \infty)$ identity given in (35) is nothing but $f(W_\alpha^{q_{\alpha,p}}) = I_\alpha(p; W)$ which is already established in the proof of part (d). Similarly (36) is equivalent to $\sup_{V \in \mathcal{P}(\mathcal{Y}|X)} f(V) = I_\alpha(p; W)$ which is established in the proof of part (d). □

Proof of Lemma 14. The following identity can be confirmed by substitution

$$D_\alpha\left(W_{[1,n]} \| \bigotimes_{t=1}^n q_{\alpha,p_t} | p\right) = \sum_{t=1}^n D_\alpha(W_t \| q_{\alpha,p_t} | p_t).$$

Then using (29), (32), (34) we get

$$\sum_{t=1}^n I_\alpha(p_t; W_t) - D_{1 \wedge \alpha}\left(q_{\alpha,p} \| \bigotimes_{t=1}^n q_{\alpha,p_t}\right) \geq I_\alpha(p; W_{[1,n]}) \geq \sum_{t=1}^n I_\alpha(p_t; W_t) - D_{\alpha \vee 1}\left(q_{\alpha,p} \| \bigotimes_{t=1}^n q_{\alpha,p_t}\right).$$

Thus (39) holds for all $p \in \mathcal{P}(\mathcal{X}_1^n)$ because Rényi divergence between probability measures is non-negative. Furthermore, (39) holds as an equality iff for an $\alpha \in \mathbb{R}_+$ iff $q_{\alpha,p}$ satisfies (40) because the Rényi divergence between distinct probability measures is positive.

If $p = \bigotimes_{t=1}^n p_t$, then one can confirm (40) for $\alpha = 1$ case by substitution. In addition for any $\alpha \in \mathbb{R}_+ \setminus \{1\}$ one can show by substitution that the probability measure $q = \bigotimes_{t=1}^n q_{\alpha,p_t}$ is a fixed point of $T_{\alpha,p}(\cdot)$, i.e. $T_{\alpha,p}(q) = q$. Furthermore, $q_{1,p} \prec q$ because $q_{1,p_t} \prec q_{\alpha,p_t}$ for each $t \in \{1, \dots, n\}$. Thus for $\alpha \in \mathbb{R}_+ \setminus \{1\}$ the identity in (40) follows from Lemma 13-(c,d). □

Proof of Lemma 15. Note that $D_\alpha(W \| q | p)$ is linear and hence concave in p for any q by definition. Then $I_\alpha(p; W)$ is concave in p because pointwise infimum of a family of concave functions is concave. Furthermore by Lemma 13-(b,c,d), $\exists! q_{\alpha,p_\beta}$ such that $D_\alpha(W \| q_{\alpha,p_\beta} | p_\beta) = I_\alpha(p_\beta; W)$. In addition,

$$D_\alpha(W \| q_{\alpha,p_\beta} | p_\beta) = \beta D_\alpha(W \| q_{\alpha,p_\beta} | p_1) + (1 - \beta) D_\alpha(W \| q_{\alpha,p_\beta} | p_0).$$

Then equation (41) and (42) are obtained by bounding $D_\alpha(W \| q_{\alpha,p_\beta} | p_1)$ and $D_\alpha(W \| q_{\alpha,p_\beta} | p_0)$ using Lemma 13-(b,c,d).

On the other hand, Lemma 1 implies

$$\begin{aligned} D_\alpha(W \| \beta q_{\alpha,p_1} + (1 - \beta) q_{\alpha,p_0} | p_\beta) &= \beta D_\alpha(W \| \beta q_{\alpha,p_1} + (1 - \beta) q_{\alpha,p_0} | p_1) + (1 - \beta) D_\alpha(W \| \beta q_{\alpha,p_1} + (1 - \beta) q_{\alpha,p_0} | p_0) \\ &\leq \beta D_\alpha(W \| q_{\alpha,p_1} | p_1) - \beta \ln \beta + (1 - \beta) D_\alpha(W \| q_{\alpha,p_0} | p_0) - (1 - \beta) \ln(1 - \beta) \\ &= \beta I_\alpha(p_1; W) + (1 - \beta) I_\alpha(p_0; W) + \tilde{h}(\beta). \end{aligned}$$

Then (43) follows from the lower bound on $D_\alpha(W \| q | p)$ given in Lemma 13-(b,c,d). □

Proof of Lemma 16. Note that as result of (38) we have,

$$[p(x)]^{\frac{1}{\alpha}} e^{\frac{1-\alpha}{\alpha} D_{\alpha}(W(x) \| q_{\alpha,p})} W(x) \leq q_{\alpha,p} \quad \forall x \in \mathcal{X}. \quad (\text{B.14})$$

(16-a) Using Lemma 1 and (B.14) we get

$$\begin{aligned} D_{\alpha}(W(x) \| q_{\alpha,p}) &\leq D_{\alpha}\left(W(x) \| [p(x)]^{\frac{1}{\alpha}} e^{\frac{1-\alpha}{\alpha} D_{\alpha}(W(x) \| q_{\alpha,p})} W(x)\right) \\ &= \frac{1}{\alpha} \ln \frac{1}{p(x)} + \frac{\alpha-1}{\alpha} D_{\alpha}(W(x) \| q_{\alpha,p}). \end{aligned}$$

(16-b) We employ (B.14) together with $D_{\alpha}(W(x) \| q_{\alpha,p}) > 0$ for $\alpha \in (0, 1]$ case and together with part (a) for $\alpha \in (1, \infty)$ case.

(16-c) We prove $\alpha \in (0, 1)$ case and $\alpha \in (1, \infty)$ case separately.

- $\alpha \in (0, 1)$: First use Jensen's inequality, i.e. $\mathbf{E}[\xi^{\alpha}] \leq \mathbf{E}[\xi]^{\alpha}$, in (38); then invoke $D_{\alpha}(W(x) \| q_{\alpha,p}) \leq \ln \frac{1}{p(x)}$:

$$\begin{aligned} \frac{dq_{\alpha,p}}{dq_{1,p}} &\leq \sum_x p(x) \frac{dW(x)}{dq_{1,p}} e^{\frac{1-\alpha}{\alpha} D_{\alpha}(W(x) \| q_{\alpha,p})} \\ &\leq (\min_{x:p(x)>0} p(x))^{\frac{\alpha-1}{\alpha}}. \end{aligned}$$

Recall that if $\xi(x) \geq 0$ for all x , then $\sum_x [\xi(x)]^{\alpha} \geq [\sum_x \xi(x)]^{\alpha}$. Using $D_{\alpha}(W(x) \| q_{\alpha,p}) \geq 0$ we get

$$\begin{aligned} \frac{dq_{\alpha,p}}{dq_{1,p}} &\geq \sum_x (p(x))^{\frac{1}{\alpha}} \frac{dW(x)}{dq_{1,p}} e^{(\frac{1-\alpha}{\alpha}) D_{\alpha}(W(x) \| q_{\alpha,p})} \\ &\geq (\min_{x:p(x)>0} p(x))^{\frac{1-\alpha}{\alpha}}. \end{aligned}$$

- $\alpha \in (1, \infty)$: First use Jensen's inequality, i.e. $\mathbf{E}[\xi^{\alpha}] \geq \mathbf{E}[\xi]^{\alpha}$, in (38), then invoke $D_{\alpha}(W(x) \| q_{\alpha,p}) \leq \ln \frac{1}{p(x)}$:

$$\begin{aligned} \frac{dq_{\alpha,p}}{dq_{1,p}} &\geq \sum_x p(x) \frac{dW(x)}{dq_{1,p}} e^{\frac{1-\alpha}{\alpha} D_{\alpha}(W(x) \| q_{\alpha,p})} \\ &\geq (\min_{x:p(x)>0} p(x))^{\frac{\alpha-1}{\alpha}}. \end{aligned}$$

Recall that if $\xi(x) \geq 0$ for all x , then $\sum_x [\xi(x)]^{\alpha} \leq [\sum_x \xi(x)]^{\alpha}$. Using $D_{\alpha}(W(x) \| q_{\alpha,p}) \geq 0$ we get

$$\begin{aligned} \frac{dq_{\alpha,p}}{dq_{1,p}} &\leq \sum_x (p(x))^{\frac{1}{\alpha}} \frac{dW(x)}{dq_{1,p}} e^{(\frac{1-\alpha}{\alpha}) D_{\alpha}(W(x) \| q_{\alpha,p})} \\ &\leq (\min_{x:p(x)>0} p(x))^{\frac{1-\alpha}{\alpha}}. \end{aligned}$$

□

Proof of Lemma 17.

(17-a) For brevity, let us denote $(\alpha - 1)I_{\alpha}(p; W)$ by $g(\alpha)$ in this part of the proof. We first prove the dichotomy about $g(\cdot)$ on $(0, 1)$ and on $(1, \infty)$. Then we extend this dichotomy to \mathbb{R}_+ assuming that $g(\cdot)$ is convex on \mathbb{R}_+ . After that we establish the assumed convexity of $g(\cdot)$ on \mathbb{R}_+ .

Let $\alpha_{\beta} = \beta\alpha_1 + (1 - \beta)\alpha_0$ for any $\alpha_0, \alpha_1 \in (0, 1)$ and $\beta \in (0, 1)$. Then for any $\alpha_0, \alpha_1 \in (0, 1)$ and $\beta \in (0, 1)$ we have

$$\begin{aligned} \beta g(\alpha_1) + (1 - \beta)g(\alpha_0) &\geq \beta(\alpha_1 - 1)D_{\alpha_1}(W \| q_{\alpha_{\beta},p} | p) + (1 - \beta)(\alpha_0 - 1)D_{\alpha_0}(W \| q_{\alpha_{\beta},p} | p) \\ &= \sum_x p(x) \ln \left(\mathbf{E}_{q_{\alpha_{\beta},p}} \left[\left(\frac{dW(x)}{dq_{\alpha_{\beta},p}} \right)^{\alpha_1} \right] \right)^{\beta} \left(\mathbf{E}_{q_{\alpha_{\beta},p}} \left[\left(\frac{dW(x)}{dq_{\alpha_{\beta},p}} \right)^{\alpha_0} \right] \right)^{1-\beta} \\ &\geq \sum_x p(x) \ln \mathbf{E}_{q_{\alpha_{\beta},p}} \left[\left(\frac{dW(x)}{dq_{\alpha_{\beta},p}} \right)^{\alpha_{\beta}} \right] \\ &= g(\alpha_{\beta}) \end{aligned}$$

where the first inequality follows from the definition of the Augustin information and the second inequality follows from the Hölder's inequality. Furthermore, the first inequality is an equality iff $q_{\alpha_0,p} = q_{\alpha_{\beta},p} = q_{\alpha_1,p}$ by Lemma 13-(c) and the second inequality is an equality iff $\frac{dW(x)}{dq_{\alpha_{\beta},p}} = \gamma(x)$ holds $W(x)$ -a.s. for all $x \in \text{supp}(p)$. On the other hand, if $\frac{dW(x)}{dq_{\alpha,p}} = \gamma(x)$ holds $W(x)$ -a.s., then $W_{\alpha}^{q_{\alpha,p}}(x) = W(x)$. Consequently, if $\frac{dW(x)}{dq_{\alpha,p}} = \gamma(x)$ holds $W(x)$ -a.s. for all $x \in \text{supp}(p)$ then $q_{\alpha,p} = q_{1,p}$ by Lemma 13-(c) because $T_{\alpha,p}(q_{\alpha,p}) = q_{1,p}$. Thus either $g(\cdot)$ is strictly convex on $(0, 1)$ or $\frac{dW(x)}{dq_{1,p}} = \gamma(x)$ $W(x)$ -a.s. for all $x \in \text{supp}(p)$ and $I_{\alpha}(p; W) = \sum_x p(x) \ln \gamma(x)$ for all $\alpha \in \mathbb{R}_+$.

Let $\alpha_{\beta} = \beta\alpha_1 + (1 - \beta)\alpha_0$ and $\frac{d\mu}{dq_{1,p}} = \left(\frac{dq_{\alpha_1,p}}{dq_{1,p}} \right)^{\frac{(\alpha_1-1)\beta}{\alpha_{\beta}-1}} \left(\frac{dq_{\alpha_0,p}}{dq_{1,p}} \right)^{\frac{(\alpha_0-1)(1-\beta)}{\alpha_{\beta}-1}}$ for any $\alpha_0, \alpha_1 \in (1, \infty)$ and $\beta \in (0, 1)$. Then

$$\begin{aligned} \beta g(\alpha_1) + (1 - \beta)g(\alpha_0) &= \sum_x p(x) \ln \left(\mathbf{E}_{q_{1,p}} \left[\left(\frac{dW(x)}{dq_{1,p}} \right)^{\alpha_1} \left(\frac{dq_{\alpha_1,p}}{dq_{1,p}} \right)^{1-\alpha_1} \right] \right)^{\beta} \left(\mathbf{E}_{q_{1,p}} \left[\left(\frac{dW(x)}{dq_{1,p}} \right)^{\alpha_0} \left(\frac{dq_{\alpha_0,p}}{dq_{1,p}} \right)^{1-\alpha_0} \right] \right)^{1-\beta} \\ &\geq \sum_x p(x) \ln \mathbf{E}_{q_{1,p}} \left[\left(\frac{dW(x)}{dq_{1,p}} \right)^{\alpha_{\beta}} \left(\frac{dq_{\alpha_1,p}}{dq_{1,p}} \right)^{\beta(1-\alpha_1)} \left(\frac{dq_{\alpha_0,p}}{dq_{1,p}} \right)^{(1-\beta)(1-\alpha_0)} \right] \\ &= (\alpha_{\beta} - 1)D_{\alpha_{\beta}}\left(W \| \frac{\mu}{\|\mu\|} | p\right) - (\alpha_{\beta} - 1) \ln \|\mu\| \\ &\geq g(\alpha_{\beta}) \end{aligned}$$

where the first inequality follows from the Hölder's inequality and the second inequality follows from the definition of Augustin information and the fact that $\|\mu\| \leq 1$, which is consequence of the Hölder's inequality. Furthermore, the first inequality is an equality iff $\frac{dW(x)}{dq_{1,p}} \left(\frac{dq_{\alpha_1,p}}{dq_{1,p}} \right)^{\frac{1-\alpha_1}{\alpha_1-\alpha_0}} \left(\frac{dq_{\alpha_0,p}}{dq_{1,p}} \right)^{\frac{\alpha_0-1}{\alpha_1-\alpha_0}} = \gamma(x)$ holds $W(x)$ -a.s. for all $x \in \text{supp}(p)$ and the second inequality is an equality iff $\mu = q_{\alpha_\beta,p}$ by Lemma 13-(d) because $\|\mu\| \leq 1$ by the Hölder's inequality. On the other hand, $\|\mu\| = 1$ iff $q_{\alpha_0,p} = q_{\alpha_1,p}$ as a result of the Hölder's inequality. Thus the second inequality is an equality iff $q_{\alpha_0,p} = q_{\alpha_1,p} = q_{\alpha_\beta,p}$. Hence both inequalities are equalities, i.e. $\beta g(\alpha_1) + (1-\beta)g(\alpha_0) = g(\alpha_\beta)$, iff $q_{\alpha_0,p} = q_{\alpha_\beta,p} = q_{\alpha_1,p}$ and $\frac{dW(x)}{dq_{\alpha_\beta,p}} = \gamma(x)$ holds $W(x)$ -a.s. for all $x \in \text{supp}(p)$. Following a reasoning similar to the one for $\alpha_0, \alpha_1 \in (0, 1)$ case and invoking Lemma 13-(d) instead of Lemma 13-(c), we conclude that either $g(\cdot)$ is strictly convex on $(1, \infty)$ or $\frac{dW(x)}{dq_{1,p}} = \gamma(x)$ $W(x)$ -a.s. for all $x \in \text{supp}(p)$ and $I_\alpha(p; W) = \sum_x p(x) \ln \gamma(x)$ for all $\alpha \in \mathbb{R}_+$. We assume the convexity of $g(\cdot)$ on \mathbb{R}_+ , in order to establish the strict convexity of $g(\cdot)$ on \mathbb{R}_+ using the strict convexity on $(0, 1)$ and $(1, \infty)$. Note that if $\alpha_0 \in (0, 1]$, $\alpha_1 \in (1, \infty)$ and $\alpha_\beta \in (1, \infty)$, then there exists an $\epsilon \in (0, \beta)$ such that $\alpha_{\beta-\epsilon} \in (1, \infty)$. Thus

$$\begin{aligned} \beta g(\alpha_1) + (1-\beta)g(\alpha_0) &= \frac{\epsilon}{1-\beta+\epsilon}g(\alpha_1) + \frac{1-\beta}{1-\beta+\epsilon} [(\beta-\epsilon)g(\alpha_1) + (1-\beta+\epsilon)g(\alpha_0)] \\ &\geq \frac{\epsilon}{1-\beta+\epsilon}g(\alpha_1) + \frac{1-\beta}{1-\beta+\epsilon}g(\alpha_{\beta-\epsilon}) \\ &> g\left(\frac{\epsilon}{1-\beta+\epsilon}\alpha_1 + \frac{1-\beta}{1-\beta+\epsilon}\alpha_{\beta-\epsilon}\right) \\ &= g(\alpha_\beta). \end{aligned}$$

Similar manipulations can be used to prove the strict inequality for $\alpha_0 \in (0, 1)$, $\alpha_1 \in [1, \infty)$, $\alpha_\beta \in (0, 1)$ case and $\alpha_0 \in (0, 1)$, $\alpha_1 \in (1, \infty)$, $\alpha_\beta = 1$ case.

Now we are left with establishing the convexity of $g(\cdot)$ on \mathbb{R}_+ that we have assumed. Invoking Lemma 13-(e) for $\alpha \in \mathbb{R}_+ \setminus \{1\}$ case and using $D_1(W \| W | p) = 0$ for $\alpha = 1$ case we get

$$g(\alpha) = \sup_{V \in \mathcal{P}(\mathcal{Y}|X)} (\alpha - 1)I_1(p; V) - \alpha D_1(V \| W | p).$$

Then $g(\alpha)$ is convex in α because pointwise supremum of a family of linear/convex functions is convex.

On the other hand, using $V = W$ we can deduce that, $g(\alpha) \geq (\alpha - 1)I_1(p; W)$. and $I_1(p; W) \in [0, \hbar(p)]$ by Lemma 13-(a). Thus $g(\alpha) \geq -\hbar(p)$.

(17-b) Since $(\alpha - 1)I_\alpha(p; W)$ is finite and convex in α on \mathbb{R}_+ , it is continuous on \mathbb{R}_+ by [20, Thm. 6.3.3]. Then $\frac{1-\alpha}{\alpha}I_\alpha(p; W)$ is continuous in α on \mathbb{R}_+ , as well. Furthermore,

$$\frac{1-\alpha}{\alpha}I_\alpha(p; W) = \inf_{V \in \mathcal{P}(\mathcal{Y}|X)} \frac{1-\alpha}{\alpha}I_1(p; V) + D_1(V \| W | p)$$

by Lemma 13-(a,e) and $D_1(W \| W | p) = 0$. Then $\frac{1-\alpha}{\alpha}I_\alpha(p; W)$ is nonincreasing in α because infimum of a family of nonincreasing functions is nonincreasing. Note that $\frac{1-\alpha}{\alpha}I_1(p; V) + D_1(V \| W | p)$ is nonincreasing in α because $I_1(p; V)$ is nonnegative.

(17-c) $I_\alpha(p; W)$ is nondecreasing in α because the pointwise infimum of a family of nondecreasing functions is nondecreasing and the Rényi divergence is nondecreasing in its order by Lemma 8.

Since $(\alpha - 1)I_\alpha(p; W)$ is finite and convex in α on \mathbb{R}_+ , it is continuous on \mathbb{R}_+ by [20, Thm. 6.3.3]. Then $I_\alpha(p; W)$ is continuous on $(0, 1)$ and $(1, \infty)$. In order to extend the continuity to \mathbb{R}_+ we need to prove that $I_\alpha(p; W)$ is continuous at $\alpha = 1$. Note that as a result of the definition of the Augustin information we have $I_\alpha(p; W) \leq D_\alpha(W \| q_{1,p} | p)$ for all $\alpha \in \mathbb{R}_+$. Since $I_\alpha(p; W)$ is nondecreasing in α we have

$$I_1(p; W) \leq I_\alpha(p; W) \leq D_\alpha(W \| q_{1,p} | p) \quad \forall \alpha \in (1, \infty). \quad (\text{B.15})$$

Since $q_{\alpha,p} \leq (\min_{x:p(x)>0} p(x))^{-\frac{1-\alpha}{\alpha}} q_{1,p}$ by Lemma 16-(c), using Lemma 1 we get

$$D_\alpha(W \| q_{\alpha,p} | p) \geq D_\alpha(W \| q_{1,p} | p) + \frac{1-\alpha}{\alpha} \ln(\min_{x:p(x)>0} p(x)) \quad \forall \alpha \in (0, 1).$$

Recall that $D_\alpha(W \| q_{\alpha,p} | p) = I_\alpha(p; W)$ by Lemma 13-(c) and $I_\alpha(p; W)$ is nondecreasing in α . Thus we have

$$D_\alpha(W \| q_{1,p} | p) + \frac{1-\alpha}{\alpha} \ln(\min_{x:p(x)>0} p(x)) \leq I_\alpha(p; W) \leq I_1(p; W) \quad \forall \alpha \in (0, 1). \quad (\text{B.16})$$

On the other hand, $D_\phi(W \| q_{1,p} | p) \leq \hbar(p) < \infty$ for any $\phi \in \mathbb{R}_+$ by Lemma 13-(a). Then $D_\alpha(W \| q_{1,p} | p)$ is continuous in α by Lemma 8. Furthermore, $D_1(W \| q_{1,p} | p) = I_1(p; W)$ by Lemma 13-(b). Then

$$\lim_{\alpha \rightarrow 1} D_\alpha(W \| q_{1,p} | p) = I_1(p; W).$$

Then the continuity of $I_\alpha(p; W)$ at $\alpha = 1$ follows from (B.15) and (B.16).

(17-d) Let $\tau_x(\alpha)$ be $\tau_x(\alpha) \triangleq \frac{\alpha-1}{\alpha} D_\alpha(W(x) \| q_{\alpha,p})$. Then we can rewrite (38) as follows:

$$\ln \frac{dq_{\eta,p}}{dq_{1,p}} = \frac{1}{\eta} \ln \sum_x p(x) \left(\left(\frac{dW(x)}{dq_{1,p}} \right)^\phi e^{(1-\phi)D_\phi(W(x) \| q_{\phi,p})} \right)^{\frac{\eta}{\phi}} e^{\eta(\tau_x(\phi) - \tau_x(\eta))}.$$

Let us assume without loss of generality that $\phi > \eta$. Then using the Jensen's inequality we get

$$\ln \frac{dq_{\eta,p}}{dq_{1,p}} \leq \ln \frac{dq_{\phi,p}}{dq_{1,p}} + \max_{x:p(x)>0} (\tau_x(\phi) - \tau_x(\eta)). \quad (\text{B.17})$$

On the other hand, using the fact that $\sum_x [\xi(x)]^{\frac{\eta}{\phi}} \geq [\sum_x \xi(x)]^{\frac{\eta}{\phi}}$ for non-negative $\xi(x)$ we get

$$\ln \frac{dq_{\eta,p}}{dq_{1,p}} \geq \ln \frac{dq_{\phi,p}}{dq_{1,p}} + (1 - \frac{\eta}{\phi}) \ln(\min_{x:p(x)>0} p(x)) + \min_{x:p(x)>0} (\tau_x(\phi) - \tau_x(\eta)). \quad (\text{B.18})$$

If $\{\tau_x(\alpha)\}_{x:p(x)>0}$ is equicontinuous in α , then $\{\ln \frac{dq_{\alpha,p}}{dq_{1,p}}\}_{y \in \mathcal{Y}}$ is equicontinuous in α by (B.17) and (B.18). On the other hand, there are only finitely many x 's with positive $p(x)$. Thus $\{\tau_x(\alpha)\}_{x:p(x)>0}$ is equicontinuous if each $\tau_x(\alpha)$ is continuous. We are left with establishing the continuity of $\tau_x(\alpha)$.

Let $g_x(\cdot) : [\eta, \phi] \rightarrow \mathbb{R}_+$, $f_x(\cdot, \cdot) : [\eta, \phi] \times \mathcal{Y} \rightarrow \mathbb{R}_{\geq 0}$, and $s : [\eta, \phi] \rightarrow \mathcal{P}(\mathcal{Y})$ be

$$\begin{aligned} g_x(\alpha) &\triangleq \int f_x(\alpha, y) \nu(dy), \\ f_x(\alpha, y) &\triangleq \left(\frac{dW(x)}{d\nu} \right)^\alpha \left(\frac{ds_\alpha}{d\nu} \right)^{1-\alpha}, \\ s_\alpha &\triangleq \frac{\phi-\alpha}{\phi-\eta} q_{\eta,p} + \frac{\alpha-\eta}{\phi-\eta} q_{\phi,p}. \end{aligned}$$

Then $f_x(\alpha, y)$ is differentiable in α and its derivative can be bounded using Lemma 16-(b) and the identity $\tau \ln \frac{1}{\tau} \leq \frac{1}{e}$:

$$\begin{aligned} \frac{\partial}{\partial \alpha} f_x(\alpha, y) &= \left(\frac{dW(x)}{d\nu} \right)^\alpha \left(\frac{ds_\alpha}{d\nu} \right)^{1-\alpha} \ln \frac{dW(x)}{ds_\alpha} + \left(\frac{dW(x)}{d\nu} \right)^\alpha \left(\frac{ds_\alpha}{d\nu} \right)^{-\alpha} \frac{1-\alpha}{\phi-\eta} \left[\frac{dq_{\phi,p}}{d\nu} - \frac{dq_{\eta,p}}{d\nu} \right] \\ \left| \frac{\partial}{\partial \alpha} f_x(\alpha, y) \Big|_{\alpha=\beta} \right| &\leq \frac{ds_\beta}{d\nu} \left[\frac{1}{\beta e} + \left(\frac{1}{p(x)^{\frac{1}{\eta\lambda_1}}} \right)^\beta \ln \frac{1}{p(x)^{\frac{1}{\eta\lambda_1}}} \right] + \left(\frac{1}{p(x)^{\frac{1}{\eta\lambda_1}}} \right)^\beta \frac{|1-\beta|}{\phi-\eta} \left[\frac{dq_{\phi,p}}{d\nu} + \frac{dq_{\eta,p}}{d\nu} \right] \\ &\leq \left(\frac{dq_{\phi,p}}{d\nu} + \frac{dq_{\eta,p}}{d\nu} \right) \left[\frac{1}{\eta e} + \left(\frac{1}{p(x)^{\frac{1}{\eta\lambda_1}}} \right)^\phi \left(\ln \frac{1}{p(x)^{\frac{1}{\eta\lambda_1}}} + \frac{1+\phi}{\phi-\eta} \right) \right] \quad \forall \beta \in [\eta, \phi]. \end{aligned}$$

The expression on the right hand side is ν -integrable. Thus as a result of [21, Corollary 2.8.7] we have

$$\frac{\partial}{\partial \alpha} g_x(\alpha) = \int \frac{\partial}{\partial \alpha} f_x(\alpha, y) \nu(dy).$$

Furthermore, $\frac{\partial}{\partial \alpha} g_x(\alpha)$ is continuous by [21, Corollary 2.8.7] because $\frac{\partial}{\partial \alpha} f_x(\alpha, y)$ is continuous in α . Then $\frac{\ln g_x(\alpha)}{\alpha}$ is a continuous function on $[\eta, \phi]$ that is continuously differentiable on (η, ϕ) . Then, as a result of mean value theorem [40, Thm. 5.10] we have

$$\left| \frac{\ln g_x(\phi)}{\phi} - \frac{\ln g_x(\eta)}{\eta} \right| \leq (\phi - \eta) \sup_{\beta \in (\eta, \phi)} \left| \frac{\partial}{\partial \alpha} \frac{\ln g_x(\alpha)}{\alpha} \Big|_{\alpha=\beta} \right|. \quad (\text{B.19})$$

Using Lemma 16-(b) and the identity $\tau \ln \frac{1}{\tau} \leq \frac{1}{e}$ we get

$$\begin{aligned} \left| \frac{\partial}{\partial \alpha} \frac{\ln g_x(\alpha)}{\alpha} \Big|_{\alpha=\beta} \right| &= \left| \frac{-\ln g_x(\beta)}{\beta^2} + \frac{1}{\beta g_x(\beta)} \int \frac{\partial}{\partial \alpha} f_x(\alpha, y) \Big|_{\alpha=\beta} \nu(dy) \right| \\ &\leq \left| \frac{\ln g_x(\beta)}{\beta^2} \right| + \frac{1}{\alpha g_x(\beta)} \left[\frac{1}{\beta e} + \left(\frac{1}{p(x)^{\frac{1}{\eta\lambda_1}}} \right)^\beta \left(\ln \frac{1}{p(x)^{\frac{1}{\eta\lambda_1}}} + \frac{|1-\beta|}{\phi-\eta} \|q_{\phi,p} - q_{\eta,p}\| \right) \right]. \end{aligned} \quad (\text{B.20})$$

We bound $\ln g_x(\beta)$ using the definition of $g_x(\beta)$ together with Lemmas 16-(b) and 1:

$$\begin{aligned} |\ln g_x(\beta)| &= |\beta - 1| D_\beta \left(W(x) \parallel \frac{\phi-\beta}{\phi-\eta} q_{\eta,p} + \frac{\beta-\eta}{\phi-\eta} q_{\phi,p} \right) \\ &\leq |\beta - 1| D_\beta \left(W(x) \parallel \left[\frac{\phi-\beta}{\phi-\eta} [p(x)]^{\frac{1}{1-\eta}} + \frac{\beta-\eta}{\phi-\eta} [p(x)]^{\frac{1}{1-\phi}} \right] W(x) \right) \\ &\leq (\beta \vee 1) D_\beta \left(W(x) \parallel [p(x)]^{\frac{1}{1-\eta}} W(x) \right) \\ &\leq \frac{\phi \vee 1}{\eta \wedge 1} \ln \frac{1}{p(x)} \quad \forall \beta \in [\eta, \phi]. \end{aligned} \quad (\text{B.21})$$

The Augustin information is nondecreasing in its order by part (c). Thus Lemmas 2 and 13 imply that

$$\|q_{\phi,p} - q_{\eta,p}\| \leq \sqrt{\frac{2}{\eta \wedge 1} (I_\phi(p; W) - I_\eta(p; W))}. \quad (\text{B.22})$$

On the other hand, one can confirm by substitution that

$$\tau_x(\phi) - \tau_x(\eta) = \frac{\ln g_x(\phi)}{\phi} - \frac{\ln g_x(\eta)}{\eta}. \quad (\text{B.23})$$

Then the continuity of $\tau_x(\alpha)$ in α is implied by (B.19), (B.20), (B.21), (B.22), (B.23) and the continuity of the Augustin information in the order established in part (c).

(17-e) $D_\eta(W \| q_{\phi,p} | p) \geq I_\eta(p; W)$ for any $\phi \in \mathbb{R}_+$ and $\eta \in \mathbb{R}_+$ by the definition of the Augustin information. Then the differentiability of $D_\alpha(W \| q | p)$ in α established in Lemma 11 implies that

$$\lim_{\eta \downarrow \phi} \frac{I_\eta(p; W) - I_\phi(p; W)}{\eta - \phi} \leq \lim_{\eta \downarrow \phi} \frac{D_\eta(W \| q_{\phi,p} | p) - D_\phi(W \| q_{\phi,p} | p)}{\eta - \phi} = \frac{\partial}{\partial \alpha} D_\alpha(W \| q_{\phi,p} | p) \Big|_{\alpha=\phi} \quad (\text{B.24})$$

$$\lim_{\eta \uparrow \phi} \frac{I_\eta(p; W) - I_\phi(p; W)}{\eta - \phi} \geq \lim_{\eta \downarrow \phi} \frac{D_\eta(W \| q_{\phi,p} | p) - D_\phi(W \| q_{\phi,p} | p)}{\eta - \phi} = \frac{\partial}{\partial \alpha} D_\alpha(W \| q_{\phi,p} | p) \Big|_{\alpha=\phi}. \quad (\text{B.25})$$

Similarly,²⁷ $D_\phi(W \| q_{\eta,p} | p) \geq I_\phi(p; W)$ for any $\phi \in \mathbb{R}_+$ and $\eta \in \mathbb{R}_+$ by the definition of the Augustin information. Hence,

$$\lim_{\eta \downarrow \phi} \frac{I_\eta(p; W) - I_\phi(p; W)}{\eta - \phi} \geq \lim_{\eta \downarrow \phi} \frac{D_\eta(W \| q_{\eta,p} | p) - D_\phi(W \| q_{\eta,p} | p)}{\eta - \phi} \quad (\text{B.26})$$

$$\lim_{\eta \uparrow \phi} \frac{I_\eta(p; W) - I_\phi(p; W)}{\eta - \phi} \leq \lim_{\eta \uparrow \phi} \frac{D_\eta(W \| q_{\eta,p} | p) - D_\phi(W \| q_{\eta,p} | p)}{\eta - \phi}. \quad (\text{B.27})$$

For any $\delta \in (0, \phi)$ by Lemma 16-(b) and Lemma 1 we have

$$D_\alpha(W(x) \| q_{\eta,p}) \leq \frac{1}{(\phi - \delta) \wedge 1} \ln \frac{1}{p(x)} \quad \forall \eta : |\eta - \phi| < \delta, \alpha \in \mathbb{R}_+. \quad (\text{B.28})$$

Then as a result of Lemma 12, there exists²⁸ a $K_{\phi,p} > 0$ such that for η close enough to ϕ we have

$$\left| D_\eta(W \| q_{\eta,p} | p) - D_\phi(W \| q_{\eta,p} | p) - (\eta - \phi) \frac{\partial}{\partial \alpha} D_\alpha(W \| q_{\eta,p} | p) \Big|_{\alpha=\phi} \right| \leq K_{\phi,p} |\eta - \phi|^2. \quad (\text{B.29})$$

We show in the following that $\frac{\partial}{\partial \alpha} D_\alpha(W \| q_{\eta,p} | p) \Big|_{\alpha=\phi}$ is a continuous function of η , i.e.

$$\lim_{\eta \rightarrow \beta} \frac{\partial}{\partial \alpha} D_\alpha(W \| q_{\eta,p} | p) \Big|_{\alpha=\phi} = \frac{\partial}{\partial \alpha} D_\alpha(W \| q_{\beta,p} | p) \Big|_{\alpha=\phi} \quad \forall \phi, \beta \in \mathbb{R}_+. \quad (\text{B.30})$$

Using (B.29) and (B.30) we get

$$\lim_{\eta \rightarrow \phi} \frac{D_\eta(W \| q_{\eta,p} | p) - D_\phi(W \| q_{\eta,p} | p)}{\eta - \phi} = \frac{\partial}{\partial \alpha} D_\alpha(W \| q_{\phi,p} | p) \Big|_{\alpha=\phi}. \quad (\text{B.31})$$

Differentiability of the Augustin information and (45) follow from (B.24), (B.25), (B.26), (B.27), and (B.31).

In order to establish the continuity of $\frac{\partial}{\partial \alpha} D_\alpha(W \| q_{\eta,p} | p) \Big|_{\alpha=\phi}$ in η , i.e. (B.30), let us first recall that the expression for the derivative of the Rényi divergence given in (16):

$$\frac{\partial}{\partial \alpha} D_\alpha(W \| q_{\eta,p} | p) \Big|_{\alpha=\phi} = \begin{cases} \frac{1}{(\phi-1)^2} \sum_x p(x) \int \frac{dW_\phi^{q_{\eta,p}}(x)}{dq_{1,p}} \left(\ln \frac{dW_\phi^{q_{\eta,p}}(x)}{dW(x)} \right) q_{1,p}(dy) & \phi \neq 1 \\ \sum_x \frac{p(x)}{2} \left(\int \frac{dW(x)}{dq_{1,p}} \left(\ln \frac{dW(x)}{dq_{1,p}} \right)^2 q_{1,p}(dy) - [D_1(W(x) \| q_{\eta,p})]^2 \right) & \phi = 1 \end{cases}.$$

Recall that,

$$D_\phi(W(x) \| q_{\eta,p}) = \begin{cases} \frac{1}{\phi-1} \ln \int \left(\frac{dW(x)}{dq_{1,p}} \right)^\phi \left(\frac{dq_{\eta,p}}{dq_{1,p}} \right)^{1-\phi} q_{1,p}(dy) & \phi \neq 1 \\ \int \frac{dW(x)}{dq_{1,p}} \left(\ln \frac{dW(x)}{dq_{1,p}} - \ln \frac{dq_{\eta,p}}{dq_{1,p}} \right) q_{1,p}(dy) & \phi = 1 \end{cases}.$$

Then $D_\phi(W(x) \| q_{\eta,p})$ is continuous in η for any $\phi \in \mathbb{R}_+$ by [21, Corollary 2.8.7-(i)] because $\{\ln \frac{dq_{\eta,p}}{dq_{1,p}}\}_{y \in \mathcal{Y}}$ is equicontinuous function of η by part (d) and $\left| \ln \frac{dq_{\eta,p}}{dq_{1,p}} \right| \leq \frac{|\eta-1|}{\eta} \ln \frac{1}{\min_{x:p(x)>0} p(x)}$ by Lemma 16-(c). On the other hand,

$$\ln \frac{dW_\phi^{q_{\eta,p}}(x)}{dq_{1,p}} = \phi \ln \frac{dW(x)}{dq_{1,p}} + (1 - \phi) \ln \frac{dq_{\eta,p}}{dq_{1,p}} + (1 - \phi) D_\phi(W(x) \| q_{\eta,p}).$$

Thus $\{\frac{dW_\phi^{q_{\eta,p}}(x)}{dq_{1,p}}\}_{y \in \mathcal{Y}}$ is equicontinuous in η because $\{\ln \frac{dq_{\eta,p}}{dq_{1,p}}\}_{y \in \mathcal{Y}}$ is equicontinuous and $D_\phi(W(x) \| q_{\eta,p})$ is continuous in η . Furthermore, using Lemma 1, Lemma 16-(b), and the identity $\tau \ln \frac{1}{\tau} \leq \frac{1}{e}$ we obtain the following bounds

$$\frac{dW_\phi^{q_{\eta,p}}(x)}{dq_{1,p}} \left| \ln \frac{dW_\phi^{q_{\eta,p}}(x)}{dW(x)} \right| \leq \frac{dW(x)}{dq_{1,p}} \left(\mathbb{1}_{\{\frac{dW_\phi^{q_{\eta,p}}(x)}{dW(x)} \leq 1\}} \frac{1}{e} + \mathbb{1}_{\{\frac{dW_\phi^{q_{\eta,p}}(x)}{dW(x)} > 1\}} \left[\frac{1}{p(x)} \right]^{\frac{\phi-1}{\eta \wedge 1}} \ln \left[\frac{1}{p(x)} \right]^{\frac{\phi-1}{\eta \wedge 1}} \right) \quad \text{if } \phi \in [1, \infty), \quad (\text{B.32})$$

$$\frac{dW_\phi^{q_{\eta,p}}(x)}{dq_{1,p}} \left| \ln \frac{dW_\phi^{q_{\eta,p}}(x)}{dW(x)} \right| \leq \frac{dW(x)}{dq_{1,p}} \mathbb{1}_{\{\frac{dW_\phi^{q_{\eta,p}}(x)}{dW(x)} \leq 1\}} \frac{1}{e} + \frac{dq_{\eta,p}}{dq_{1,p}} \mathbb{1}_{\{\frac{dW_\phi^{q_{\eta,p}}(x)}{dW(x)} > 1\}} \frac{1-\phi}{\phi e} \left[\frac{1}{p(x)} \right]^{\frac{1}{\eta \wedge 1}} \quad \text{if } \phi \in (0, 1). \quad (\text{B.33})$$

²⁷For $\phi \in (1, \infty)$, we can also use (36) of Lemma 13-(e) to establish reverse inequalities for (B.24) and (B.25).

²⁸Note that, this is not just the Taylor expansion of $D_\alpha(w \| q)$ around $\alpha = \phi$ for a given (w, q) pair. Lemma 11 allows us to apply the Taylor expansion for a family of (w, q) pairs around $\alpha = \phi$ simultaneously if we can bound $D_\beta(w \| q)$ uniformly for all (w, q) 's for some $\beta > \phi$.

Using $\left| \ln \frac{dq_{\eta,p}}{dq_{1,p}} \right| \leq \frac{|\eta-1|}{\eta} \ln \frac{1}{\min_{x:p(x)>0} p(x)}$ —i.e. Lemma 16-(c)— in (B.33), we get

$$\frac{dW_{\phi}^{q_{\eta,p}}(x)}{dq_{1,p}} \left| \ln \frac{dW_{\phi}^{q_{\eta,p}}(x)}{dW(x)} \right| \leq \frac{dW(x)}{dq_{1,p}} \frac{1}{e} + \left[\min_{x:p(x)>0} p(x) \right]^{\frac{-|\eta-1|}{\eta}} \frac{1}{\phi e} \left[\frac{1}{p(x)} \right]^{\frac{1}{\eta \wedge 1}} \quad \text{if } \phi \in (0, 1). \quad (\text{B.34})$$

Using (B.32) and (B.34), we get the following bound for all $\phi \in \mathbb{R}_+$ and $\eta \in [a, b]$

$$\frac{dW_{\phi}^{q_{\eta,p}}(x)}{dq_{1,p}} \left| \ln \frac{dW_{\phi}^{q_{\eta,p}}(x)}{dW(x)} \right| \leq \frac{dW(x)}{dq_{1,p}} \left(\frac{1}{e} + \left[\frac{1}{p(x)} \right]^{\frac{\phi}{a \wedge 1}} \ln \left[\frac{1}{p(x)} \right]^{\frac{\phi}{a \wedge 1}} \right) + \left[\min_{x:p(x)>0} p(x) \right]^{-1-\frac{1}{a}} \frac{1}{\phi e} \left[\frac{1}{p(x)} \right]^{\frac{1}{\eta \wedge 1}}.$$

Then $\left\{ \frac{dW_{\phi}^{q_{\eta,p}}(x)}{dq_{1,p}} \left| \ln \frac{dW_{\phi}^{q_{\eta,p}}(x)}{dW(x)} \right| \right\}_{\eta \in [a,b]}$ is bounded from above by a $q_{1,p}$ -integrable function for any closed interval

$[a, b] \subset \mathbb{R}_+$. Thus $D_1 \left(W_{\phi}^{q_{\eta,p}} \left\| \left\| W \right\| p \right) \right)$ is a continuous function of η by [21, Corollary 2.8.7-(i)] for all $\phi \in \mathbb{R}_+$. Then $\frac{\partial}{\partial \alpha} D_{\alpha}(W \| q_{\eta,p} | p) \Big|_{\alpha=\phi}$ is continuous in η for $\phi \in \mathbb{R}_+ \setminus 1$. The continuity of $\frac{\partial}{\partial \alpha} D_{\alpha}(W \| q_{\eta,p} | p) \Big|_{\alpha=1}$ in η follows from the continuity of $D_1(W(x) \| q_{\eta,p})$ and [21, Corollary 2.8.7-(i)] via the following bound, which can be established using the identity $\tau(\ln \tau)^2 \mathbb{1}_{\{\tau \in (0,1]\}} \leq \frac{4}{e^2}$ and Lemma 16-(b),

$$\frac{dW(x)}{dq_{1,p}} \left(\ln \frac{dW(x)}{dq_{\eta,p}} \right)^2 \leq \frac{dq_{\eta,p}}{dq_{1,p}} \frac{4}{e^2} + \frac{dW(x)}{dq_{1,p}} \left(\frac{\ln p(x)}{\eta \wedge 1} \right)^2. \quad (\text{B.35})$$

Now we are left with establishing the continuity of the derivative of the Augustin information. Since $\{\ln \frac{dq_{\alpha,p}}{dq_{1,p}}\}_{y \in \mathcal{Y}}$ is equicontinuous in α by part (d), for any $\epsilon > 0$ there exists a δ such that

$$e^{-\epsilon} q_{\phi,p} \leq q_{\eta,p} \leq e^{\epsilon} q_{\phi,p} \quad \forall \eta : |\eta - \phi| < \delta.$$

On the other hand $(p(x))^{\frac{1}{\phi \wedge 1}} W(x) \leq q_{\phi,p}$ by Lemma 16-(b). Then as a result of Lemma 1

$$D_{\alpha}(W(x) \| q_{\eta,p}) \leq \frac{1}{\phi \wedge 1} \ln \frac{1}{p(x)} + \epsilon \quad \forall \eta : |\phi - \eta| < \delta, \quad \forall \alpha \in \mathbb{R}_+.$$

Hence, Lemma 12 implies the existence of a $\tau \in \mathbb{R}_+$ that does not depend on η such that

$$\left| \frac{\partial^{\kappa} D_{\alpha}(W \| q_{\eta,p} | p)}{\partial \alpha^{\kappa}} \Big|_{\alpha=\phi} \right| \leq \kappa! \tau^{\kappa+1} \kappa \quad \forall \eta : |\phi - \eta| < \delta. \quad (\text{B.36})$$

Then $\limsup_{\kappa \rightarrow \infty} \left| \frac{1}{\kappa!} \frac{\partial^{\kappa}}{\partial \alpha^{\kappa}} \left(\frac{\partial}{\partial \alpha} D_{\alpha}(W \| q_{\eta,p} | p) \right) \Big|_{\alpha=\phi} \right|^{1/\kappa} \leq \tau$. Thus the radius of convergence of the Taylor's expansion of $\frac{\partial D_{\alpha}(W \| q_{\eta,p} | p)}{\partial \alpha}$ around $\alpha = \phi$ is at least $\frac{1}{\tau}$ for all $\eta \in [\phi - \delta, \phi + \delta]$ by Hadamard's formula [38, Lemma 1.1.8]. Furthermore, we can use (B.36) to bound higher order derivatives:

$$\left| \frac{\partial D_{\alpha}(W \| q_{\eta,p} | p)}{\partial \alpha} \Big|_{\alpha=\beta} - \frac{\partial D_{\alpha}(W \| q_{\eta,p} | p)}{\partial \alpha} \Big|_{\alpha=\phi} \right| \leq \sum_{i=1}^{\infty} \frac{|\beta - \phi|^i}{i!} \left| \frac{\partial^{i+1} D_{\alpha}(W \| q_{\eta,p} | p)}{\partial \alpha^{i+1}} \Big|_{\alpha=\phi} \right| \leq \tau^2 \sum_{i=1}^{\infty} (i^2 + 2i + 1) |\beta - \phi|^i \tau^i \quad \forall \eta : |\eta - \phi| \leq \delta, \quad \forall \beta : |\beta - \phi| < \frac{1}{\tau}.$$

Using identities $\sum_{i=1}^{\infty} \xi^i \leq \sum_{i=1}^{\infty} i \xi^{i-1} \leq \sum_{i=1}^{\infty} i^2 \xi^{i-2}$ for $\xi \geq 0$ and $\sum_{i=1}^{\infty} i^2 \xi^i = \frac{(1+\xi)\xi}{(1-\xi)^3}$ for $|\xi| < 1$ we get,

$$\left| \frac{\partial D_{\alpha}(W \| q_{\eta,p} | p)}{\partial \alpha} \Big|_{\alpha=\beta} - \frac{\partial D_{\alpha}(W \| q_{\eta,p} | p)}{\partial \alpha} \Big|_{\alpha=\phi} \right| \leq 4\tau^3 |\beta - \phi| \frac{1+\tau|\beta-\phi|}{(1-\tau|\beta-\phi|)^3} \quad \forall \eta : |\eta - \phi| \leq \delta, \quad \forall \beta : |\beta - \phi| < \frac{1}{\tau}.$$

Then using (45) we get

$$\left| \frac{\partial}{\partial \alpha} I_{\alpha}(p; W) \Big|_{\alpha=\eta} - \frac{\partial D_{\alpha}(W \| q_{\eta,p} | p)}{\partial \alpha} \Big|_{\alpha=\phi} \right| \leq 4\tau^3 |\eta - \phi| \frac{1+\tau|\eta-\phi|}{(1-\tau|\eta-\phi|)^3} \quad \forall \eta : |\eta - \phi| \leq \delta \wedge \frac{1}{\tau}.$$

Hence, $\lim_{\eta \rightarrow \phi} \frac{\partial}{\partial \alpha} I_{\alpha}(p; W) \Big|_{\alpha=\eta} = \lim_{\eta \rightarrow \phi} \frac{\partial D_{\alpha}(W \| q_{\eta,p} | p)}{\partial \alpha} \Big|_{\alpha=\phi}$, if the latter limit exists. However, we have already established the existence of that limit in order to calculate the derivative of the Augustin information: it is equal to $\frac{\partial}{\partial \alpha} I_{\alpha}(p; W) \Big|_{\alpha=\phi}$. Thus the Augustin information is continuously differentiable in the order.

(17-f) Let us start with analyzing the case when $(\alpha - 1)I_{\alpha}(p; W)$ is strictly convex in α . The chain rule for derivatives implies

$$\frac{\partial}{\partial s} s I_{\frac{1}{1+s}}(p; W) = I_{\frac{1}{1+s}}(p; W) + s \frac{(-1)}{(1+s)^2} \frac{\partial}{\partial \alpha} I_{\alpha}(p; W) \Big|_{\alpha=\frac{1}{1+s}}.$$

Using (46), (B.35), and the fact that $\mathbf{E}_{W(x)} \left[\left(\ln \frac{dW(x)}{dq_{1,p}} - D_1(W(x) \| q_{1,p}) \right)^2 \right] \leq \mathbf{E}_{W(x)} \left[\left(\ln \frac{dW(x)}{dq_{1,p}} \right)^2 \right]$, we get

$$\frac{\partial}{\partial s} s I_{\frac{1}{1+s}}(p; W) \Big|_{s=0} = \begin{cases} I_1(p; W) & s = 0 \\ I_{\frac{1}{1+s}}(p; W) - \frac{1}{s} D_1 \left(W_{\frac{1}{1+s}}^{q_{\frac{1}{1+s},p}} \left\| \left\| W \right\| p \right) \right) & s \in (-1, 0) \cup (0, \infty) \end{cases}.$$

Then as a result of (35), we can assert the following expression for all $s \in (-1, \infty)$

$$\frac{\partial}{\partial s} sI_{\frac{1}{1+s}}(p; W) = I_1\left(p; W^{\frac{q_{\frac{1}{1+s}, p}}{\frac{1}{1+s}}}\right). \quad (\text{B.37})$$

Then the continuous differentiability of $I_\alpha(p; W)$ in α on \mathbb{R}_+ , established in part (e), implies the continuity of $I_1(p; W_\alpha^{q_{\alpha, p}})$ in α on \mathbb{R}_+ .

In order to prove that $I_1(p; W_\alpha^{q_{\alpha, p}})$ is monotonically increasing in α on \mathbb{R}_+ , note that the strict convexity of $(\alpha - 1)I_\alpha(p; W)$ in α on \mathbb{R}_+ is equivalent to the strict concavity of $sI_{\frac{1}{1+s}}(p; W)$ in s on $(-1, \infty)$ because the inequality

$$(\alpha_\beta - 1)I_{\alpha_\beta}(p; W) < \beta(\alpha_1 - 1)I_{\alpha_1}(p; W) + (1 - \beta)(\alpha_0 - 1)I_{\alpha_0}(p; W)$$

holds for $\alpha_0, \alpha_1 \in \mathbb{R}_+$, $\beta \in (0, 1)$, and $\alpha_\beta = \beta\alpha_1 + (1 - \beta)\alpha_0$ iff the inequality

$$s_\mu I_{\frac{1}{1+s_\mu}}(p; W) > \mu s_1 I_{\frac{1}{1+s_1}}(p; W) + (1 - \mu)s_0 I_{\frac{1}{1+s_0}}(p; W).$$

holds for $s_0 = \frac{1-\alpha_0}{\alpha_0}$, $s_1 = \frac{1-\alpha_1}{\alpha_1}$, $s_\mu = \mu s_1 + (1 - \mu)s_0$ and $\mu = \frac{\beta\alpha_1}{\alpha_\beta}$.

On the other hand for any strictly concave function $f(\cdot)$ and s_1, s_2, s_3, s_4 satisfying $s_1 < s_2 < s_3 < s_4$ we have²⁹

$$\frac{f(s_2) - f(s_1)}{s_2 - s_1} > \frac{f(s_4) - f(s_3)}{s_4 - s_3}.$$

Thus $I_1\left(p; W^{\frac{q_{\frac{1}{1+s}, p}}{\frac{1}{1+s}}}\right)$ is a decreasing function of s on $(-1, \infty)$ by (B.37) and the definition of the derivative because $sI_{\frac{1}{1+s}}(p; W)$ is strictly concave in s . Hence $I_1(p; W_\alpha^{q_{\alpha, p}})$ is an increasing function of α on \mathbb{R}_+ .

If $(\alpha - 1)I_\alpha(p; W)$ is not strictly convex in α then there exists a $\gamma : \mathcal{X} \rightarrow [1, \infty)$ satisfying $\frac{dW(x)}{dq_{1, p}} = \gamma(x)$ $W(x)$ -a.s. for all $x \in \text{supp}(p)$ and $q_{\alpha, p} = q_{1, p}$ for all $\alpha \in \mathbb{R}_+$ by part (a). Thus $W_\alpha^{q_{\alpha, p}}(x) = W(x)$ and $\frac{dW_\alpha^{q_{\alpha, p}}(x)}{dq_{\alpha, p}} = \gamma(x)$ for all $x \in \text{supp}(p)$. Consequently $I_1(p; W_\alpha^{q_{\alpha, p}}) = \sum_x p(x) \ln \gamma(x)$.

(17-g) Let us define $I_0(p; W)$ to be $\lim_{\alpha \downarrow 0} I_\alpha(p; W)$, such a limit exists because $I_\alpha(p; W)$ is non-decreasing function of α on \mathbb{R}_+ . Then $(\alpha - 1)I_\alpha(p; W)$ is convex in α on $[0, \infty)$, as well. Thus for any $\alpha \in \mathbb{R}_+$ and $\epsilon \in \mathbb{R}_+$

$$\frac{(\alpha - 1)I_\alpha(p; W) + I_0(p; W)}{\alpha} \leq \frac{(\alpha + \epsilon - 1)I_{\alpha + \epsilon}(p; W) - (\alpha - 1)I_\alpha(p; W)}{\epsilon}$$

by [20, Proposition 6.3.2]. Taking the limits as $\epsilon \downarrow 0$ and invoking (46) we get

$$\frac{(\alpha - 1)I_\alpha(p; W) + I_0(p; W)}{\alpha} \leq I_\alpha(p; W) + \frac{1}{\alpha - 1} D_1(W_\alpha^{q_{\alpha, p}} \| W | p).$$

Thus (35) implies

$$I_0(p; W) \leq I_1(p; W_\alpha^{q_{\alpha, p}}).$$

On the other hand for all $\alpha \in (0, 1)$ the non-negativity of the Rényi divergence and (35) implies

$$I_1(p; W_\alpha^{q_{\alpha, p}}) \leq I_\alpha(p; W).$$

Hence $\lim_{\alpha \downarrow 0} I_1(p; W_\alpha^{q_{\alpha, p}}) = I_0(p; W)$, i.e. $\lim_{\alpha \downarrow 0} I_1(p; W_\alpha^{q_{\alpha, p}}) = \lim_{\alpha \downarrow 0} I_\alpha(p; W)$. □

Proof of Lemma 18. Lemma 18 is nothing but Lemma 33 for the case when λ is a vector of zeros. Thus we do not present a separate proof for Lemma 18, see the proof of Lemma 33. □

C. Augustin's Proof of Lemma 13-(c)

We have employed the relative compactness in the total variation topology for proving Lemma 13-(c) because we wanted to assert $q_{\alpha, p} \sim q_{1, p}$, the convergence described in (31), and the inequality given in (32). Establishing the existence of a unique Augustin mean together with the fixed point property described in (30) is considerably easier. It can be done using the concept of relative compactness in the topology of setwise convergence, as demonstrated by Augustin in [6, §34]. Augustin claims to establish other assertions of Lemma 13-(c), as well. In the following, we discuss why we think there are caveats in Augustin's argument in [6, §34].

²⁹Note that $f(s_2) > \frac{s_3 - s_2}{s_3 - s_1} f(s_1) + \frac{s_2 - s_1}{s_3 - s_1} f(s_3)$ implies $\frac{f(s_2) - f(s_1)}{s_2 - s_1} > \frac{f(s_3) - f(s_2)}{s_3 - s_2}$.

Let us first establish the existence of a unique Augustin mean. First, we establish (B.4) as we have done in the current proof. Then we consider the set $\mathcal{Q}' \triangleq \{q \in \mathcal{P}(\mathcal{Y}) : D_\alpha(W \| q | p) < D_\alpha(W \| q_{\alpha,p}^g | p)\}$. Note that $I_\alpha(p; W) = \inf_{q \in \mathcal{T}_{\alpha,p}(\mathcal{Q}')} D_\alpha(W \| q | p)$ because of the definition of \mathcal{Q}' and (B.4). Furthermore for all $q \in \mathcal{Q}'$ and $\mathcal{E} \in \mathcal{Y}$,

$$\begin{aligned}
T_{\alpha,p}(q)(\mathcal{E}) &= \sum_x p(x) e^{(1-\alpha)D_\alpha(W(x) \| q)} \int_{\mathcal{E}} \left(\frac{dW(x)}{d\nu}\right)^\alpha \left(\frac{dq}{d\nu}\right)^{1-\alpha} \nu(dy) \\
&\leq e^{\frac{(1-\alpha)D_\alpha(W(x) \| q_{\alpha,p}^g | p)}{\min_{x:p(x)>0}}} \sum_x p(x) \int_{\mathcal{E}} \left(\frac{dW(x)}{d\nu}\right)^\alpha \left(\frac{dq}{d\nu}\right)^{1-\alpha} \nu(dy) && \text{by the definition of } \mathcal{Q}' \text{ and Lemma 2} \\
&= e^{\frac{(1-\alpha)D_\alpha(W(x) \| q_{\alpha,p}^g | p)}{\min_{x:p(x)>0}}} \int_{\mathcal{E}} \left(\frac{d\mu_{\alpha,p}}{d\nu}\right)^\alpha \left(\frac{dq}{d\nu}\right)^{1-\alpha} \nu(dy) && \text{by (26)} \\
&\leq e^{\frac{(1-\alpha)D_\alpha(W(x) \| q_{\alpha,p}^g | p)}{\min_{x:p(x)>0}}} [\mu_{\alpha,p}(\mathcal{E})]^\alpha [q(\mathcal{E})]^{1-\alpha} && \text{by Holder's inequality} \\
&\leq e^{\frac{(1-\alpha)D_\alpha(W(x) \| q_{\alpha,p}^g | p)}{\min_{x:p(x)>0}}} [\mu_{\alpha,p}(\mathcal{E})]^\alpha && \text{because } q(\mathcal{E}) \leq 1
\end{aligned}$$

Thus $\mathcal{T}_{\alpha,p}(\mathcal{Q}') \prec^{uni} q_{\alpha,p}^g$ and $\mathcal{T}_{\alpha,p}(\mathcal{Q}')$ has compact closure in the topology of setwise convergence by a version of Dunford-Pettis theorem [21, 4.7.25]. On the other hand, $D_\alpha(W \| q | p)$ is lower-semicontinuous in q for the topology of setwise convergence because $D_\alpha(W(x) \| q)$ is, by Lemma 3. Then there exists a $q_{\alpha,p}$ in the closure of $\mathcal{T}_{\alpha,p}(\mathcal{Q}')$ for the topology of setwise convergence such that $D_\alpha(W \| q_{\alpha,p} | p) = \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(W \| q | p)$ by the extreme value theorem for lower semicontinuous functions [32, Ch3§12.2]. The uniqueness of $q_{\alpha,p}$ follows from the strict convexity of the Rényi divergence in its second argument described in Lemma 5.

This construction establishes certain additional properties of the Augustin mean, as well. Note that $q_{\alpha,p} \prec q_{1,p}$ because $q_{\alpha,p}$ is in the closure of $\mathcal{T}_{\alpha,p}(\mathcal{Q}')$ for the topology of setwise convergence. In addition, $\mathcal{T}_{\alpha,p}(q_{\alpha,p}) = q_{\alpha,p}$ because of Lemma 2 and (B.4). Furthermore, any q satisfying $\mathcal{T}_{\alpha,p}(q) = q$ and $q_{1,p} \prec q$ is equal to $q_{\alpha,p}$ because of the argument presented in step (c-i) of the proof of Lemma 13-(c). These observations, with minor differences, exist in Augustin's proof of [6, Lemma 34.2].

Above discussion establishes Lemma 13-(c) except for the following three assertions:

- (i) $q_{1,p} \prec q_{\alpha,p}$,
- (ii) the identity given in (31),
- (iii) the inequality given in (32).

Note that $\mathcal{T}_{\alpha,p}(\mathcal{Q}') \prec^{uni} q_{\alpha,p}^g$ and $\{\mathcal{T}_{\alpha,p}^i(q_{\alpha,p}^g)\}_{i \in \mathbb{Z}_+} \subset \mathcal{T}_{\alpha,p}(\mathcal{Q}')$. Then by [21, Thm. 4.7.25], $\{\mathcal{T}_{\alpha,p}^i(q_{\alpha,p}^g)\}_{i \in \mathbb{Z}_+}$ has a subsequence $\{\mathcal{T}_{\alpha,p}^{i(j)}(q_{\alpha,p}^g)\}_{j \in \mathbb{Z}_+}$ converging to a $q \in \text{cl}(\{\mathcal{T}_{\alpha,p}^i(q_{\alpha,p}^g)\}_{i \in \mathbb{Z}_+})$ where both the convergence and the closure are for the topology of setwise convergence. Furthermore, $q \sim q_{1,p}$ because of the arguments used in step (c-iv) of the proof of Lemma 13-(c). There are two ways one can prove remaining assertions of Lemma 13-(c) without using the totally boundedness of $\{\mathcal{T}_{\alpha,p}^i(q_{\alpha,p}^g)\}_{i \in \mathbb{Z}_+}$ established in step (c-iii) of the proof of Lemma 13-(c):

- If one can show that $\frac{d\mathcal{T}_{\alpha,p}^{i(j)}(q_{\alpha,p}^g)}{dq_{1,p}}$ converges to $\frac{dq}{dq_{1,p}}$ in measure $q_{1,p}$, then because of the Lebesgue-Vitali convergence theorem [21, 4.5.4] one would have $\lim_{j \rightarrow \infty} \left\| \mathcal{T}_{\alpha,p}^{i(j)}(q_{\alpha,p}^g) - q \right\| = 0$, established step (c-iv). Thus one can skip steps (c-iii) and (c-iv) and proceed with the step (c-v) of the proof.
- If one can show that the limit point q of the subsequence $\frac{d\mathcal{T}_{\alpha,p}^{i(j)}(q_{\alpha,p}^g)}{dq_{1,p}}$ is a fixed point of $\mathcal{T}_{\alpha,p}(\cdot)$, then one would have a statement equivalent to step (c-vi). Thus one can skip steps (c-iii), (c-iv), and (c-vi) and proceed with the step (c-vii) after deriving (c-v).

We proved Lemma 13-(c) using the concept of totally boundedness because we could not find an easy way to establish either the convergence in measure property or the fixed point property mentioned in the preceding discussion. However, we do know that both properties hold. The convergence in measure holds because of the only if part of the Lebesgue-Vitali convergence theorem [21, 4.5.4]. The fixed point property holds because $\{\mathcal{T}_{\alpha,p}^i(q_{\alpha,p}^g)\}_{i \in \mathbb{Z}_+}$ has a unique limit point both in total variation topology and in topology of setwise convergence by (31).

While proving [6, Lemma 34.2], after establishing the weak convergence of $\{\mathcal{T}_{\alpha,p}^{i(j)}(q_{1,p})\}_{j \in \mathbb{Z}_+}$ to q , Augustin asserts that $\lim_{j \rightarrow \infty} \left\| \mathcal{T}_{\alpha,p}^{i(j)}(q_{1,p}) - q \right\| = 0$. This is the first one of our two major reservations for Augustin's proof of [6, Lemma 34.2]. Note that convergence in the topology of setwise convergence and weak convergence are one and the same thing for sequences of measures by [21, Corollary 4.7.26]. But convergence in the topology of setwise convergence does not imply convergence in total variation topology.³⁰ Thus we don't know how one can justify such an assertion.

In order to prove [6, Lemma 34.2], Augustin establishes the totally boundedness of $\{\mathcal{T}_{\alpha,p}^i(q_{1,p})\}_{i \in \mathbb{Z}_+}$ for the total variation metric. In that proof Augustin asserts that $\mathcal{T}_{\alpha,p}^i(q_{1,p})$ is in \mathcal{B}_j , defined in equation (B.5), for some $j \in \mathbb{Z}_+$. We don't know whether such an assertion is correct or not. But we know that $\mathcal{T}_{\alpha,p}^i(q_{\alpha,p}^g)$ is in \mathcal{B}_i . Thus one can fix this problem easily.

³⁰Consider, for example, the sequence of measure on the unit interval whose Radon-Nikodym derivatives with respect to the Lebesgue measure is given by $\{(1 + \cos(\pi j z))\}_{j \in \mathbb{Z}_+}$. This set of probability measures converges to the Lebesgue measure on every measurable set, but not in total variation.

A more important problem stems from Augustin's obliviousness about the infiniteness of the set of positive integers. Either in his discussion or in his equations there is no evidence suggesting that he makes a distinction of cases for approximating $\{T_{\alpha,p}^t(q_{\alpha,p}^g)\}_{t \leq \iota}$ and $\{T_{\alpha,p}^t(q_{\alpha,p}^g)\}_{t > \iota}$. This is our other major reservation for Augustin's proof of [6, Lemma 34.2].

D. Proofs of Lemmas on the Augustin Capacity

Proof of Lemma 19.

(i) $\forall \alpha \in \mathbb{R}_+ \exists \tilde{p} \in \mathcal{P}(\mathcal{X})$ s.t. $I_\alpha(\tilde{p}; W) = C_{\alpha,W,\mathcal{A}}$: Since $\tilde{h}(p) \leq \ln |\mathcal{X}|$ for all $p \in \mathcal{P}(\mathcal{X})$, (44) and Lemma 13-(a) imply that

$$|I_\alpha(p_2; W) - I_\alpha(p_1; W)| \leq \tilde{h}\left(\frac{\|p_1 - p_2\|}{2}\right) + \frac{\|p_1 - p_2\|}{2} \ln |\mathcal{X}|. \quad (\text{D.1})$$

Hence, $I_\alpha(p; W)$ is continuous in p on $\mathcal{P}(\mathcal{X})$. On the other hand, $\mathcal{P}(\mathcal{X})$ is compact because \mathcal{X} is a finite set. Then \mathcal{A} is compact because any closed subset of a compact set is compact, [39, Thm. 26.2]. Then there exists a $\tilde{p} \in \mathcal{A}$ such that $I_\alpha(\tilde{p}; W) = \sup_{p \in \mathcal{A}} I_\alpha(p; W)$ by the extreme value theorem,³¹ [39, 27.4].

(ii) If $\alpha \in \mathbb{R}_+$ and $I_\alpha(\tilde{p}; W) = C_{\alpha,W,\mathcal{A}}$, then $D_\alpha(W \| q_{\alpha,\tilde{p}} | p) \leq C_{\alpha,W,\mathcal{A}}$ for all $p \in \mathcal{A}$: Let p be any member of \mathcal{A} and $p^{(\iota)}$ be $\frac{\iota-1}{\iota}\tilde{p} + \frac{1}{\iota}p$ for $\iota \in \mathbb{Z}_+$. Then $p^{(\iota)} \in \mathcal{A}$ because \mathcal{A} is convex. Furthermore, by Lemma 13-(b,c,d) we have

$$\begin{aligned} I_\alpha(p^{(\iota)}; W) &= \frac{\iota-1}{\iota} D_\alpha\left(W \| q_{\alpha,p^{(\iota)}} \middle| \tilde{p}\right) + \frac{1}{\iota} D_\alpha\left(W \| q_{\alpha,p^{(\iota)}} \middle| p\right) \\ &\geq \frac{\iota-1}{\iota} \left[I_\alpha(\tilde{p}; W) + D_{\alpha \wedge 1}\left(q_{\alpha,\tilde{p}} \| q_{\alpha,p^{(\iota)}}\right) \right] + \frac{1}{\iota} D_\alpha\left(W \| q_{\alpha,p^{(\iota)}} \middle| p\right) \quad \forall \iota \in \mathbb{Z}_+. \end{aligned}$$

Using $I_\alpha(p^{(\iota)}; W) \leq C_{\alpha,W,\mathcal{A}}$, $I_\alpha(\tilde{p}; W) = C_{\alpha,W,\mathcal{A}}$, and $D_{\alpha \wedge 1}\left(q_{\alpha,\tilde{p}} \| q_{\alpha,p^{(\iota)}}\right) \geq 0$ we get

$$C_{\alpha,W,\mathcal{A}} \geq D_\alpha\left(W \| q_{\alpha,p^{(\iota)}} \middle| p\right) \quad \forall \iota \in \mathbb{Z}_+. \quad (\text{D.2})$$

On the other hand, using $I_\alpha(p^{(\iota)}; W) \leq C_{\alpha,W,\mathcal{A}}$, $I_\alpha(\tilde{p}; W) = C_{\alpha,W,\mathcal{A}}$, and $D_\alpha\left(W \| q_{\alpha,p^{(\iota)}} \middle| p\right) \geq 0$ we get

$$\frac{1}{\iota} C_{\alpha,W,\mathcal{A}} \geq \frac{\iota-1}{\iota} D_{\alpha \wedge 1}\left(q_{\alpha,\tilde{p}} \| q_{\alpha,p^{(\iota)}}\right) \quad \forall \iota \in \mathbb{Z}_+.$$

Then using Lemma 2 we get

$$\sqrt{\frac{2}{\alpha \wedge 1} \frac{1}{\iota-1} C_{\alpha,W,\mathcal{A}}} \geq \|q_{\alpha,\tilde{p}} - q_{\alpha,p^{(\iota)}}\| \quad \forall \iota \in \mathbb{Z}_+.$$

Thus $q_{\alpha,p^{(\iota)}}$ converges to $q_{\alpha,\tilde{p}}$ in the total variation topology and hence in the topology of setwise convergence. Since the Rényi divergence is lower semicontinuous in the topology of setwise convergence by Lemma 3, we have

$$\liminf_{\iota \rightarrow \infty} D_\alpha\left(W \| q_{\alpha,p^{(\iota)}} \middle| p\right) \geq D_\alpha\left(W \| q_{\alpha,\tilde{p}} \middle| p\right) \quad (\text{D.3})$$

Then the inequality $D_\alpha(W \| q_{\alpha,\tilde{p}} | p) \leq C_{\alpha,W,\mathcal{A}}$ follows from (D.2) and (D.3).

(iii) If $\alpha \in \mathbb{R}_+$, then $\exists! q_{\alpha,W,\mathcal{A}} \in \mathcal{P}(\mathcal{Y})$ satisfying (61) such that $q_{\alpha,p} = q_{\alpha,W,\mathcal{A}}$ for all $p \in \mathcal{A}$ satisfying $I_\alpha(p; W) = C_{\alpha,W,\mathcal{A}}$: If $I_\alpha(p; W) = C_{\alpha,W,\mathcal{A}}$ for a $p \in \mathcal{A}$, then Lemma 13-(b,c,d) and Lemma 2 imply

$$D_\alpha(W \| q_{\alpha,\tilde{p}} | p) \geq C_{\alpha,W,\mathcal{A}} + \frac{\alpha \wedge 1}{2} \|q_{\alpha,p} - q_{\alpha,\tilde{p}}\|^2. \quad (\text{D.4})$$

Since we have already established that $D_\alpha(W \| q_{\alpha,\tilde{p}} | p) \leq C_{\alpha,W,\mathcal{A}}$ for all $p \in \mathcal{A}$, (D.4) implies that $q_{\alpha,p} = q_{\alpha,\tilde{p}}$ for any $p \in \mathcal{A}$ satisfying $I_\alpha(p; W) = C_{\alpha,W,\mathcal{A}}$. □

Proof of Theorem 1. The right hand side of (59) is an upper bound on the left hand side because of the max-min inequality. Furthermore, the left hand side of (59) is equal to $C_{\alpha,W,\mathcal{A}}$ by (58). Thus when $C_{\alpha,W,\mathcal{A}}$ is infinite, (59) holds trivially. When $C_{\alpha,W,\mathcal{A}}$ is finite, (59) follows from (60) and the max-min inequality. Thus we can assume $C_{\alpha,W,\mathcal{A}}$ to be finite and prove the claims about $q_{\alpha,W,\mathcal{A}}$ in order to prove the theorem.

(i) If $C_{\alpha,W,\mathcal{A}} \in \mathbb{R}_{\geq 0}$ and $\lim_{\iota \rightarrow \infty} I_\alpha(p^{(\iota)}; W) = C_{\alpha,W,\mathcal{A}}$ for a $\{p^{(\iota)}\}_{\iota \in \mathbb{Z}_+} \subset \mathcal{A}$, then $\{q_{\alpha,p^{(\iota)}}\}_{\iota \in \mathbb{Z}_+}$ is a Cauchy sequence in $\mathcal{P}(\mathcal{Y})$ for the total variation metric: For any sequence of members of \mathcal{A} satisfying $\lim_{\iota \rightarrow \infty} I_\alpha(p^{(\iota)}; W) = C_{\alpha,W,\mathcal{A}}$, let $\{\mathcal{A}^{(\iota)}\}_{\iota \in \mathbb{Z}_+}$ be a nested sequence of closed, convex, subsets of \mathcal{A} defined as follows,

$$\mathcal{A}^{(\iota)} \triangleq \text{ch}\left(\bigcup_{j=1}^{\iota} \{p^{(j)}\}\right).$$

³¹We do not need to establish the continuity of $I_\alpha(p; W)$ in p ; the upper semicontinuity is sufficient as a result of [32, Ch3§12.2]. Note that $I_\alpha(p; W)$ is upper semicontinuous in p because it is the infimum of a family of linear functions.

Furthermore, each $\mathcal{A}^{(\iota)}$ can be interpreted as a constraint set for a $W^{(\iota)}$ with a finite input set $\mathcal{X}^{(\iota)}$ defined as follows,

$$\mathcal{X}^{(\iota)} \triangleq \{x \in \mathcal{X} : \exists j \in \{1, \dots, \iota\} \text{ such that } p^{(j)}(x) > 0\}.$$

With a slight abuse of notation we use the symbol $\mathcal{A}^{(\iota)}$ not only for a subset of $\mathcal{P}(\mathcal{X})$ but also for the corresponding subset of $\mathcal{P}(\mathcal{X}^{(\iota)})$. For any $\iota \in \mathbb{Z}_+$, there exists a unique $q_{\alpha, W^{(\iota)}, \mathcal{A}^{(\iota)}}$ satisfying inequality (61) by Lemma 19. Furthermore, $\mathcal{A}^{(j)} \subset \mathcal{A}^{(\iota)}$ for any $\iota, j \in \mathbb{Z}_+$ such that $j \leq \iota$. In order to bound $\|q_{\alpha, p^{(j)}} - q_{\alpha, p^{(\iota)}}\|$ for positive integers $j < \iota$, we use the triangle inequality for $q_{\alpha, p^{(j)}}$, $q_{\alpha, p^{(\iota)}}$, and $q_{\alpha, W^{(\iota)}, \mathcal{A}^{(\iota)}}$

$$\|q_{\alpha, p^{(j)}} - q_{\alpha, p^{(\iota)}}\| \leq \|q_{\alpha, p^{(j)}} - q_{\alpha, W^{(\iota)}, \mathcal{A}^{(\iota)}}\| + \|q_{\alpha, p^{(\iota)}} - q_{\alpha, W^{(\iota)}, \mathcal{A}^{(\iota)}}\|. \quad (\text{D.5})$$

Let us proceed with bounding $\|q_{\alpha, p^{(j)}} - q_{\alpha, W^{(\iota)}, \mathcal{A}^{(\iota)}}\|$ and $\|q_{\alpha, p^{(\iota)}} - q_{\alpha, W^{(\iota)}, \mathcal{A}^{(\iota)}}\|$ from above.

$$\begin{aligned} \|q_{\alpha, p^{(j)}} - q_{\alpha, W^{(\iota)}, \mathcal{A}^{(\iota)}}\| &\stackrel{(a)}{\leq} \sqrt{\frac{2}{\alpha \wedge 1} D_{\alpha \wedge 1}(q_{\alpha, p^{(j)}} \| q_{\alpha, W^{(\iota)}, \mathcal{A}^{(\iota)}})} \\ &\stackrel{(b)}{\leq} \sqrt{\frac{2}{\alpha \wedge 1} \sqrt{D_{\alpha}(W \| q_{\alpha, W^{(\iota)}, \mathcal{A}^{(\iota)}} | p^{(j)}) - I_{\alpha}(p^{(j)}; W^{(\iota)})}} \\ &\stackrel{(c)}{\leq} \sqrt{\frac{2}{\alpha \wedge 1} \sqrt{C_{\alpha, W^{(\iota)}, \mathcal{A}^{(\iota)}} - I_{\alpha}(p^{(j)}; W^{(\iota)})}} \\ &\stackrel{(d)}{\leq} \sqrt{\frac{2}{\alpha \wedge 1} \sqrt{C_{\alpha, W, \mathcal{A}} - I_{\alpha}(p^{(j)}; W)}} \end{aligned}$$

where (a) follows from Lemma 2, (b) follows from Lemma 13-(b,c,d), (c) follows from Lemma 19 because $p^{(j)} \in \mathcal{A}^{(\iota)}$, and (d) follows from the identities $C_{\alpha, W^{(\iota)}, \mathcal{A}^{(\iota)}} = C_{\alpha, W, \mathcal{A}^{(\iota)}} \leq C_{\alpha, W, \mathcal{A}}$ and $I_{\alpha}(p^{(j)}; W^{(\iota)}) = I_{\alpha}(p^{(j)}; W)$. We can obtain a similar bound on $\|q_{\alpha, p^{(\iota)}} - q_{\alpha, W^{(\iota)}, \mathcal{A}^{(\iota)}}\|$. Then $\{q_{\alpha, p^{(\iota)}}\}$ is a Cauchy sequence by (D.5) because $\lim_{j \rightarrow \infty} I_{\alpha}(p^{(j)}; W) = C_{\alpha, W, \mathcal{A}}$.

- (ii) If $C_{\alpha, W, \mathcal{A}} \in \mathbb{R}_{\geq 0}$, then $\exists! q_{\alpha, W, \mathcal{A}} \in \mathcal{P}(\mathcal{Y})$ satisfying $\lim_{\iota \rightarrow \infty} \|q_{\alpha, W, \mathcal{A}} - q_{\alpha, p^{(\iota)}}\| = 0$ for all $\{p^{(\iota)}\}_{\iota \in \mathbb{Z}_+} \subset \mathcal{A}$ such that $\lim_{\iota \rightarrow \infty} I_{\alpha}(p^{(\iota)}; W) = C_{\alpha, W, \mathcal{A}}$. Note that $\mathcal{M}(\mathcal{Y})$ is a complete metric space for the total variation metric, i.e. every Cauchy sequence has a unique limit point in $\mathcal{M}(\mathcal{Y})$, because $\mathcal{M}(\mathcal{Y})$ is a Banach space for the total variation topology [21, Thm. 4.6.1]. Then $\{q_{\alpha, p^{(\iota)}}\}_{\iota \in \mathbb{Z}_+}$ has a unique limit point q_{α, p^*} in $\mathcal{M}(\mathcal{Y})$. Since $\mathcal{P}(\mathcal{Y})$ is a closed set for the total variation topology and $\cup_{\iota \in \mathbb{Z}_+} q_{\alpha, p^{(\iota)}} \subset \mathcal{P}(\mathcal{Y})$, then $q_{\alpha, p^*} \in \mathcal{P}(\mathcal{Y})$ by [39, Thm. 2.1.3].

We have established the existence of a unique limit point q_{α, p^*} , for any $\{p^{(\iota)}\}_{\iota \in \mathbb{Z}_+} \subset \mathcal{A}$ satisfying $\lim_{\iota \rightarrow \infty} I_{\alpha}(p^{(\iota)}; W) = C_{\alpha, W, \mathcal{A}}$. This, however, implies $\lim_{\iota \rightarrow \infty} \|q_{\alpha, \tilde{p}^{(\iota)}} - q_{\alpha, p^*}\| = 0$ for any $\{\tilde{p}^{(\iota)}\}_{\iota \in \mathbb{Z}_+} \subset \mathcal{A}$ satisfying $\lim_{\iota \rightarrow \infty} I_{\alpha}(\tilde{p}^{(\iota)}; W) = C_{\alpha, W, \mathcal{A}}$ because we can interleave the elements of $\{p^{(\iota)}\}_{\iota \in \mathbb{Z}_+}$ and $\{\tilde{p}^{(\iota)}\}_{\iota \in \mathbb{Z}_+}$ to obtain a new sequence $\{\hat{p}^{(\iota)}\}_{\iota \in \mathbb{Z}_+} \subset \mathcal{A}$ satisfying $\lim_{\iota \rightarrow \infty} I_{\alpha}(\hat{p}^{(\iota)}; W) = C_{\alpha, W, \mathcal{A}}$ for which $\{q_{\alpha, \hat{p}^{(\iota)}}\}$ is a Cauchy sequence. Then $q_{\alpha, W, \mathcal{A}} = q_{\alpha, p^*}$.

- (iii) $q_{\alpha, W, \mathcal{A}}$ satisfies the equality given in (60): For any $p \in \mathcal{A}$, let us consider any sequence $\{p^{(\iota)}\}_{\iota \in \mathbb{Z}_+} \subset \mathcal{A}$ satisfying $p^{(1)} = p$ and $\lim_{\iota \rightarrow \infty} I_{\alpha}(p^{(\iota)}; W) = C_{\alpha, W, \mathcal{A}}$. Then $p \in \mathcal{A}^{(\iota)}$ for all $\iota \in \mathbb{Z}_+$. Using Lemma 19 we get

$$D_{\alpha}(W \| q_{\alpha, W^{(\iota)}, \mathcal{A}^{(\iota)}} | p) \leq C_{\alpha, W^{(\iota)}, \mathcal{A}^{(\iota)}} \quad \forall \iota \in \mathbb{Z}_+. \quad (\text{D.6})$$

Since $W^{(\iota)}$ has a finite input set, $\exists \tilde{p}^{(\iota)} \in \mathcal{A}^{(\iota)}$ satisfying $I_{\alpha}(\tilde{p}^{(\iota)}; W^{(\iota)}) = C_{\alpha, W^{(\iota)}, \mathcal{A}^{(\iota)}}$ and $q_{\alpha, \tilde{p}^{(\iota)}} = q_{\alpha, W^{(\iota)}, \mathcal{A}^{(\iota)}}$ by Lemma 19. Then $I_{\alpha}(\tilde{p}^{(\iota)}; W^{(\iota)}) \geq I_{\alpha}(p^{(\iota)}; W^{(\iota)})$ and consequently $\lim_{\iota \rightarrow \infty} I_{\alpha}(\tilde{p}^{(\iota)}; W) = C_{\alpha, W, \mathcal{A}}$. We have already established that for such a sequence $q_{\alpha, \tilde{p}^{(\iota)}} \rightarrow q_{\alpha, W, \mathcal{A}}$ in the total variation topology, and hence in the topology of setwise convergence. Then the lower semicontinuity of the Rényi divergence in its arguments for the topology of setwise convergence, i.e. Lemma 3, the identity $C_{\alpha, W^{(\iota)}, \mathcal{A}^{(\iota)}} = C_{\alpha, W, \mathcal{A}^{(\iota)}} \leq C_{\alpha, W, \mathcal{A}}$, and (D.6) imply

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

On the other hand $D_{\alpha}(W \| q_{\alpha, W, \mathcal{A}} | p) \geq I_{\alpha}(p; W)$ and $\sup_{p \in \mathcal{A}} I_{\alpha}(p; W) = C_{\alpha, W, \mathcal{A}}$ by definition. Thus (60) holds. \square

Proof of Lemma 20. Lemma 13-(b,c,d) and the hypothesis given in (62) imply

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

Then as a result of Lemma 2,

$$\sqrt{\frac{2(C_{\alpha, W, \mathcal{A}} - I_{\alpha}(p; W))}{\alpha \wedge 1}} \geq \|q_{\alpha, p} - q_{\alpha, W, \mathcal{A}}\| \quad \forall p \in \mathcal{A}.$$

Thus $\{q_{\alpha, p^{(\iota)}}\}_{\iota \in \mathbb{Z}_+}$ is a Cauchy sequence with the limit point $q_{\alpha, W, \mathcal{A}}$ for any sequence of input distributions $\{p^{(\iota)}\}_{\iota \in \mathbb{Z}_+} \subset \mathcal{A}$ satisfying $\lim_{\iota \rightarrow \infty} I_{\alpha}(p^{(\iota)}; W) = C_{\alpha, W, \mathcal{A}}$. \square

Proof of Lemma 21. As a result of Lemma 13-(b,c,d) we have

$$\begin{aligned} \sup_{\tilde{p} \in \mathcal{A}} D_\alpha(W \| q | \tilde{p}) &\geq D_\alpha(W \| q | p) && \forall p \in \mathcal{A} \\ &\geq I_\alpha(p; W) + D_{\alpha \wedge 1}(q_{\alpha, p} \| q) && \forall p \in \mathcal{A}. \end{aligned} \quad (\text{D.7})$$

Let $\{p^{(i)}\}_{i \in \mathbb{Z}_+} \subset \mathcal{A}$ be a sequence such that $\lim_{i \rightarrow \infty} I_\alpha(p^{(i)}; W) = C_{\alpha, W, \mathcal{A}}$. Then $\{q_{\alpha, p^{(i)}}\} \rightarrow q_{\alpha, W, \mathcal{A}}$ in total variation topology and hence in the topology of set wise convergence by Lemma 20. On the other hand, the Rényi divergence is lower semicontinuous in its arguments for the topology of setwise convergence by Lemma 3. Then

$$\liminf_{i \rightarrow \infty} \left[I_\alpha(p^{(i)}; W) + D_{\alpha \wedge 1}(q_{\alpha, p^{(i)}} \| q) \right] \geq C_{\alpha, W, \mathcal{A}} + D_{\alpha \wedge 1}(q_{\alpha, W, \mathcal{A}} \| q). \quad (\text{D.8})$$

(63) follows from (D.7) and (D.8). \square

Proof of Lemma 22. Note that as a result of (64) and the max-min inequality we have

$$C_{\alpha, W, \mathcal{A}} \leq \inf_{V \in \mathcal{P}(\mathcal{Y}|X)} \sup_{p \in \mathcal{A}} \frac{\alpha}{1-\alpha} D_1(V \| W | p) + I_1(p; V). \quad (\text{D.9})$$

Hence, (65) holds trivially whenever $C_{\alpha, W, \mathcal{A}} = \infty$ and (66) implies (65) whenever $C_{\alpha, W, \mathcal{A}} \in \mathbb{R}_{\geq 0}$.

In order to establish (66) assuming $C_{\alpha, W, \mathcal{A}} \in \mathbb{R}_{\geq 0}$, first note that whenever $C_{\alpha, W, \mathcal{A}} \in \mathbb{R}_{\geq 0}$ there exists a unique $q_{\alpha, W, \mathcal{A}}$ satisfying (62) by Lemma 20. Then as a result of Definitions 1, 2, 3, 4 and Lemma 10 we have

$$D_\alpha(W \| q_{\alpha, W, \mathcal{A}} | p) = \frac{\alpha}{1-\alpha} D_1(W_\alpha^{q_{\alpha, W, \mathcal{A}}} \| W | p) + D_1(W_\alpha^{q_{\alpha, W, \mathcal{A}}} \| q_{\alpha, W, \mathcal{A}} | p).$$

Then using Lemma 13-(b) and Lemma 2, we get

$$\begin{aligned} D_\alpha(W \| q_{\alpha, W, \mathcal{A}} | p) &= \frac{\alpha}{1-\alpha} D_1(W_\alpha^{q_{\alpha, W, \mathcal{A}}} \| W | p) + I_1(p; W_\alpha^{q_{\alpha, W, \mathcal{A}}}) + D_1\left(\sum_x p(x) W_\alpha^{q_{\alpha, W, \mathcal{A}}}(x) \| q_{\alpha, W, \mathcal{A}}\right) \\ &\geq \frac{\alpha}{1-\alpha} D_1(W_\alpha^{q_{\alpha, W, \mathcal{A}}} \| W | p) + I_1(p; W_\alpha^{q_{\alpha, W, \mathcal{A}}}) \end{aligned} \quad \forall p \in \mathcal{A}.$$

Thus (62) and Lemma 20 implies that

$$C_{\alpha, W, \mathcal{A}} \geq \sup_{p \in \mathcal{A}} \frac{\alpha}{1-\alpha} D_1(W_\alpha^{q_{\alpha, W, \mathcal{A}}} \| W | p) + I_1(p; W_\alpha^{q_{\alpha, W, \mathcal{A}}}) \quad (\text{D.10})$$

$$\geq \inf_V \sup_{p \in \mathcal{A}} \frac{\alpha}{1-\alpha} D_1(V \| W | p) + I_1(p; V). \quad (\text{D.11})$$

Note that (65) follows from (D.9) and (D.11). On the other hand, using the Csiszár's form for the Augustin information, given in (36), we get

$$\begin{aligned} \frac{\alpha}{1-\alpha} D_1(W_\alpha^{q_{\alpha, W, \mathcal{A}}} \| W | p) + I_1(p; W_\alpha^{q_{\alpha, W, \mathcal{A}}}) &\geq \inf_V \frac{\alpha}{1-\alpha} D_1(V \| W | p) + I_1(p; V) \\ &= I_\alpha(p; W) \end{aligned} \quad \forall p \in \mathcal{P}(X).$$

Then (66) follows from the definition of $C_{\alpha, W, \mathcal{A}}$ and (D.10). \square

Proof of Lemma 23.

(23-a) $C_{\alpha, W, \mathcal{A}}$ is nondecreasing and lower semicontinuous because it is the pointwise supremum of $I_\alpha(p; W)$ for $p \in \mathcal{A}$ and $I_\alpha(p; W)$ is nondecreasing and continuous in α by Lemma 17-(c).

(23-b) $\frac{1-\alpha}{\alpha} I_\alpha(p; W)$ is nonincreasing and continuous in α on \mathbb{R}_+ for all $p \in \mathcal{P}(X)$ by Lemma 17-(b). Furthermore,

$$\frac{1-\alpha}{\alpha} C_{\alpha, W, \mathcal{A}} = \sup_{p \in \mathcal{A}} \frac{1-\alpha}{\alpha} I_\alpha(p; W) \quad \forall \alpha \in (0, 1).$$

Then $\frac{1-\alpha}{\alpha} C_{\alpha, W, \mathcal{A}}$ is nonincreasing and lower semicontinuous in α on $(0, 1)$ because the pointwise supremum of a family of nonincreasing (lower semicontinuous) functions is nonincreasing (lower semicontinuous). Thus $\frac{1-\alpha}{\alpha} C_{\alpha, W, \mathcal{A}}$ and $C_{\alpha, W, \mathcal{A}}$ are both continuous from the right on $(0, 1)$. On the other hand $C_{\alpha, W, \mathcal{A}}$ and $\frac{1-\alpha}{\alpha} C_{\alpha, W, \mathcal{A}}$ are both continuous from the left on $(0, 1)$ because $C_{\alpha, W, \mathcal{A}}$ is nondecreasing and lower semicontinuous on $(0, 1)$ by part (a). Consequently, $C_{\alpha, W, \mathcal{A}}$ and $\frac{1-\alpha}{\alpha} C_{\alpha, W, \mathcal{A}}$ are both continuous on $(0, 1)$.

(23-c) $(\alpha - 1)I_\alpha(p; W)$ is convex in α on \mathbb{R}_+ by Lemma 17-(a). Furthermore,

$$(\alpha - 1)C_{\alpha, W, \mathcal{A}} = \sup_{p \in \mathcal{A}} (\alpha - 1)I_\alpha(p; W) \quad \forall \alpha \in (1, \infty).$$

Then $(\alpha - 1)C_{\alpha, W, \mathcal{A}}$ is convex in α because the pointwise supremum of a family of convex functions is convex.

(23-d) $C_{\alpha, W, \mathcal{A}}$ is continuous in α on $(0, 1)$ by part (b). Furthermore, $C_{\alpha, W, \mathcal{A}}$ is continuous from the left because it is nondecreasing and lower semicontinuous. Thus $C_{\alpha, W, \mathcal{A}}$ is continuous in α on $(0, 1]$. If $\chi_{W, \mathcal{A}} = 1$ we are done.

If $\chi_{W, \mathcal{A}} > 1$, then $(\alpha - 1)C_{\alpha, W, \mathcal{A}}$ is finite and convex in α on $[1, \chi_{W, \mathcal{A}}]$ by part (c) and the definition of $\chi_{W, \mathcal{A}}$. Then $(\alpha - 1)C_{\alpha, W, \mathcal{A}}$ is continuous in α on $(1, \chi_{W, \mathcal{A}})$ by [20, Thm. 6.3.3]. The continuity of $(\alpha - 1)C_{\alpha, W, \mathcal{A}}$ on $(1, \chi_{W, \mathcal{A}})$ implies the continuity of $C_{\alpha, W, \mathcal{A}}$ on $(1, \chi_{W, \mathcal{A}})$. Furthermore, $C_{\alpha, W, \mathcal{A}}$ is continuous from the left because $C_{\alpha, W, \mathcal{A}}$ is nondecreasing and lower semicontinuous. Hence, $C_{\alpha, W, \mathcal{A}}$ is continuous in α on $(1, \chi_{W, \mathcal{A}}]$, as well.

(23-e) As a result of part (d), we only need to prove the continuity of $C_{\alpha,W,\mathcal{A}}$ from the right at $\alpha = 1$ when $\chi_{W,\mathcal{A}} > 1$. As a result of [13, (30)] we have

$$I_\alpha^g(p; W) \leq I_1^g(p; W) + \frac{8(\alpha-1)}{\epsilon^2 e^2} e^{\frac{\eta-1}{\eta} I_\eta^g(p; W)} \quad \forall \epsilon \in (0, \frac{\eta-1}{\eta}), \alpha \in [1, (1-\epsilon)\eta].$$

On the other hand $I_1(p; W) = I_1^g(p; W)$ and $I_\alpha(p; W) \leq I_\alpha^g(p; W)$ for $\alpha > 1$. Then,

$$C_{\alpha,W,\mathcal{A}} \leq C_{1,W,\mathcal{A}} + \frac{8(\alpha-1)}{\epsilon^2 e^2} e^{\frac{\eta-1}{\eta} \sup_{p \in \mathcal{A}} I_\eta^g(p; W)} \quad \forall \epsilon \in (0, \frac{\eta-1}{\eta}), \alpha \in [1, (1-\epsilon)\eta].$$

Thus $C_{\alpha,W,\mathcal{A}}$ is continuous at $\alpha = 1$ from the right because $C_{\alpha,W,\mathcal{A}} \geq C_{1,W,\mathcal{A}}$. \square

Proof of Lemma 24. Note that $C_{\alpha,W,\mathcal{A}}$ is finite for all $\alpha \in (0, \eta]$ by Lemma 23. Then there exists a unique order α Augustin center, $q_{\alpha,W,\mathcal{A}}$, for all $\alpha \in (0, \eta]$ by Theorem 1. We apply Lemma 21 for $q = q_{\alpha,W,\mathcal{A}}$ to get

$$\sup_{p \in \mathcal{A}} D_\alpha(W \| q_{\alpha,W,\mathcal{A}} | p) \geq C_{\alpha,W,\mathcal{A}} + D_{\alpha \wedge 1}(q_{\alpha,W,\mathcal{A}} \| q_{\alpha,W,\mathcal{A}}). \quad (\text{D.12})$$

Note that $D_\alpha(W \| q_{\alpha,W,\mathcal{A}} | p)$ is nondecreasing in α for all $p \in \mathcal{A}$, because $D_\alpha(W(x) \| q_{\alpha,W,\mathcal{A}})$ is, by Lemma 8. Then,

$$D_\phi(W \| q_{\alpha,W,\mathcal{A}} | p) \geq D_\alpha(W \| q_{\alpha,W,\mathcal{A}} | p) \quad \forall p \in \mathcal{A}, \phi \in [\alpha, \eta]. \quad (\text{D.13})$$

On the other hand, by (60) of Theorem 1 we have

$$\sup_{p \in \mathcal{A}} D_\phi(W \| q_{\alpha,W,\mathcal{A}} | p) = C_{\phi,W,\mathcal{A}} \quad \forall \phi \in (0, \eta]. \quad (\text{D.14})$$

(67) follows from (D.12), (D.13), and (D.14).

Using Lemma 2 together with (67) we get

$$\|q_{\phi,W,\mathcal{A}} - q_{\alpha,W,\mathcal{A}}\| \leq \sqrt{\frac{2}{\alpha \wedge 1} (C_{\phi,W,\mathcal{A}} - C_{\alpha,W,\mathcal{A}})} \quad \forall \alpha, \phi \text{ such that } 0 < \alpha < \phi \leq \eta. \quad (\text{D.15})$$

Then the continuity $q_{\alpha,W,\mathcal{A}}$ in α for the total variation topology on $\mathcal{P}(\mathcal{Y})$ follows from the continuity $C_{\alpha,W,\mathcal{A}}$ in α on \mathcal{I} . \square

Proof of Lemma 25. We analyze the upper bound on $C_{\alpha,W,\mathcal{A}}$ and the lower bound on $C_{\alpha,W,\mathcal{A}}$ separately.

- $\sup_{\iota \in \mathcal{T}} C_{\alpha,W,\mathcal{A}^{(\iota)}} \leq C_{\alpha,W,\mathcal{A}}$: Note that $C_{\alpha,W,\mathcal{A}^{(\iota)}} \leq C_{\alpha,W,\mathcal{A}}$ by definition because $\mathcal{A}^{(\iota)} \subset \mathcal{A}$. Thus $C_{\alpha,W,\mathcal{A}}$ is bounded from below by $\sup_{\iota \in \mathcal{T}} C_{\alpha,W,\mathcal{A}^{(\iota)}}$, as well.
 - If $C_{\alpha,W,\mathcal{A}^{(\iota)}} = C_{\alpha,W,\mathcal{A}} < \infty$, then $q_{\alpha,W,\mathcal{A}} = q_{\alpha,W,\mathcal{A}^{(\iota)}}$ because using Theorem 1 for (a), Lemma 21 for (b) and Lemma 2 for (c) we get

$$\begin{aligned} C_{\alpha,W,\mathcal{A}} &\stackrel{(a)}{\geq} \sup_{p \in \mathcal{A}^{(\iota)}} D_\alpha(W \| q_{\alpha,W,\mathcal{A}} | p) \\ &\stackrel{(b)}{\geq} C_{\alpha,W,\mathcal{A}^{(\iota)}} + D_{\alpha \wedge 1}(q_{\alpha,W,\mathcal{A}^{(\iota)}} \| q_{\alpha,W,\mathcal{A}}) \\ &\stackrel{(c)}{\geq} C_{\alpha,W,\mathcal{A}^{(\iota)}} + \frac{\alpha \wedge 1}{2} \|q_{\alpha,W,\mathcal{A}^{(\iota)}} - q_{\alpha,W,\mathcal{A}}\|^2. \end{aligned}$$

If $C_{\alpha,W,\mathcal{A}^{(\iota)}} = C_{\alpha,W,\mathcal{A}}$ and $q_{\alpha,W,\mathcal{A}^{(\iota)}} = q_{\alpha,W,\mathcal{A}}$, then $\sup_{p \in \mathcal{A}} D_\alpha(W \| q_{\alpha,W,\mathcal{A}^{(\iota)}} | p) \leq C_{\alpha,W,\mathcal{A}^{(\iota)}}$ by Theorem 1.

- If $\sup_{p \in \mathcal{A}} D_\alpha(W \| q_{\alpha,W,\mathcal{A}^{(\iota)}} | p) \leq C_{\alpha,W,\mathcal{A}^{(\iota)}}$, then $C_{\alpha,W,\mathcal{A}} \leq C_{\alpha,W,\mathcal{A}^{(\iota)}}$ by (58) and Theorem 1. On the other hand, $C_{\alpha,W,\mathcal{A}} \geq C_{\alpha,W,\mathcal{A}^{(\iota)}}$ because $\mathcal{A}^{(\iota)} \subset \mathcal{A}$. Hence, $C_{\alpha,W,\mathcal{A}} = C_{\alpha,W,\mathcal{A}^{(\iota)}} < \infty$.

- $C_{\alpha,W,\mathcal{A}} \leq \ln \sum_{\iota \in \mathcal{T}} e^{C_{\alpha,W,\mathcal{A}^{(\iota)}}}$: If \mathcal{T} is an infinite set, then the inequality holds trivially because $\ln \sum_{\iota \in \mathcal{T}} e^{C_{\alpha,W,\mathcal{A}^{(\iota)}}} = \infty$. Thus we assume \mathcal{T} to be a finite set for the rest of the proof. Let $\mu = \bigvee_{\iota \in \mathcal{T}} e^{C_{\alpha,W,\mathcal{A}^{(\iota)}}} q_{\alpha,W,\mathcal{A}^{(\iota)}}$. Then as a result of Lemma 1 we have

$$\begin{aligned} \sup_{p \in \mathcal{A}^{(\iota)}} D_\alpha\left(W \| \frac{\mu}{\|\mu\|} \middle| p\right) &= \sup_{p \in \mathcal{A}^{(\iota)}} D_\alpha(W \| \mu | p) + \ln \|\mu\| \\ &\leq \sup_{p \in \mathcal{A}^{(\iota)}} D_\alpha(W \| q_{\alpha,W,\mathcal{A}^{(\iota)}} | p) - C_{\alpha,W,\mathcal{A}^{(\iota)}} + \ln \|\mu\| \quad \forall \iota \in \mathcal{T}. \end{aligned}$$

Since $\sup_{p \in \mathcal{A}^{(\iota)}} D_\alpha(W \| q_{\alpha,W,\mathcal{A}^{(\iota)}} | p) = C_{\alpha,W,\mathcal{A}^{(\iota)}}$ by hypothesis, we have

$$\sup_{p \in \mathcal{A}^{(\iota)}} D_\alpha\left(W \| \frac{\mu}{\|\mu\|} \middle| p\right) \leq \ln \|\mu\| \quad \forall \iota \in \mathcal{T}.$$

Then using (58) and Theorem 1 we get

$$\begin{aligned} C_{\alpha,W,\mathcal{A}} &\leq \sup_{p \in \mathcal{A}} D_\alpha\left(W \| \frac{\mu}{\|\mu\|} \middle| p\right) \\ &= \sup_{\iota \in \mathcal{T}} \sup_{p \in \mathcal{A}^{(\iota)}} D_\alpha\left(W \| \frac{\mu}{\|\mu\|} \middle| p\right) \\ &\leq \ln \|\mu\| \\ &\leq \ln \sum_{\iota \in \mathcal{T}} e^{C_{\alpha,W,\mathcal{A}^{(\iota)}}}. \end{aligned}$$

- If $q_{\alpha, W, \mathcal{A}^{(i)}}$ and $q_{\alpha, W, \mathcal{A}^{(j)}}$ are not singular for some $i \neq j$, then $\|\mu\| < \sum_{i \in \mathcal{J}} e^{C_{\alpha, W, \mathcal{A}^{(i)}}}$. Thus $C_{\alpha, W, \mathcal{A}} < \ln \sum_{i \in \mathcal{J}} e^{C_{\alpha, W, \mathcal{A}^{(i)}}}$. Consequently, if $C_{\alpha, W, \mathcal{A}} = \ln \sum_{i \in \mathcal{J}} e^{C_{\alpha, W, \mathcal{A}^{(i)}}}$, then $q_{\alpha, W, \mathcal{A}^{(i)}} \perp q_{\alpha, W, \mathcal{A}^{(j)}}$ for all $i \neq j$.
- If $q_{\alpha, W, \mathcal{A}^{(i)}} \perp q_{\alpha, W, \mathcal{A}^{(j)}}$ for all $i \neq j$, then any $s \in \mathcal{P}(\mathcal{Y})$ can be written as $s = \sum_{i=1}^{|\mathcal{J}|+1} s_i$ where s_i 's are finite measures such that $s_i \prec_{q_{\alpha, W, \mathcal{A}^{(i)}}}$ for $i \in \mathcal{J}$ and $s_{|\mathcal{J}|+1} \perp (\sum_{i \in \mathcal{J}} q_{\alpha, W, \mathcal{A}^{(i)}})$ by the Lebesgue decomposition theorem [20, 5.5.3]. Using Lemmas 1, 2, and 21 we get

$$\begin{aligned} \sup_{p \in \mathcal{A}^{(i)}} D_{\alpha}(W \| s | p) &\geq C_{\alpha, W, \mathcal{A}^{(i)}} + D_{\alpha \wedge 1}(q_{\alpha, W, \mathcal{A}^{(i)}} \| s) \\ &= C_{\alpha, W, \mathcal{A}^{(i)}} + D_{\alpha \wedge 1}\left(q_{\alpha, W, \mathcal{A}^{(i)}} \left\| \frac{s_i}{\|s_i\|} \right.\right) - \ln \|s_i\| \\ &\geq C_{\alpha, W, \mathcal{A}^{(i)}} - \ln \|s_i\|. \end{aligned}$$

$\sup_{p \in \mathcal{A}} D_{\alpha}(W \| s | p) = \max_{i \in \mathcal{J}} \sup_{p \in \mathcal{A}^{(i)}} D_{\alpha}(W \| s | p)$ because $D_{\alpha}(W \| s | p)$ is linear in p and $\mathcal{A} = \text{ch}(\cup_{i \in \mathcal{J}} \mathcal{A}^{(i)})$. Then using $\sum_{i=1}^{|\mathcal{J}|} \|s_i\| \leq \|s\| = 1$ we get,

$$\begin{aligned} \sup_{p \in \mathcal{A}} D_{\alpha}(W \| s | p) &\geq \max_{i \in \mathcal{J}} \ln \frac{e^{C_{\alpha, W, \mathcal{A}^{(i)}}}}{\|s_i\|} \\ &\geq \ln \sum_{i \in \mathcal{J}} e^{C_{\alpha, W, \mathcal{A}^{(i)}}} \quad \forall s \in \mathcal{P}(\mathcal{Y}). \end{aligned}$$

Then $C_{\alpha, W, \mathcal{A}} \geq \ln \sum_{i \in \mathcal{J}} e^{C_{\alpha, W, \mathcal{A}^{(i)}}}$ by (58) and Theorem 1. Since we have already established the reverse inequality, we have $C_{\alpha, W, \mathcal{A}} = \ln \sum_{i \in \mathcal{J}} e^{C_{\alpha, W, \mathcal{A}^{(i)}}}$.

We have proved that if $q_{\alpha, W, \mathcal{A}^{(i)}} \perp q_{\alpha, W, \mathcal{A}^{(j)}}$ for all $i \neq j$, then $C_{\alpha, W, \mathcal{A}} = \sum_{i \in \mathcal{J}} e^{C_{\alpha, W, \mathcal{A}^{(i)}}}$. One can confirm by substitution that $\sup_{p \in \mathcal{A}^{(i)}} D_{\alpha}(W \| s | p) \leq C_{\alpha, W, \mathcal{A}}$ for all $i \in \mathcal{J}$ for $s = \sum_{i \in \mathcal{J}} e^{-C_{\alpha, W, \mathcal{A}} + C_{\alpha, W, \mathcal{A}^{(i)}}} q_{\alpha, W, \mathcal{A}^{(i)}}$. On the other hand, $\sup_{p \in \mathcal{A}} D_{\alpha}(W \| s | p) = \max_{i \in \mathcal{J}} \sup_{p \in \mathcal{A}^{(i)}} D_{\alpha}(W \| s | p)$ because $D_{\alpha}(W \| s | p)$ is linear in p . Then s is the unique order α Augustin center by Theorem 1. □

Proof of Lemma 26. Let α be any fixed positive real order. Then as a result of Lemma 14 we have

$$C_{\alpha, W_{[1, n]}, \mathcal{A}_1^n} = \sum_{t=1}^n C_{\alpha, W_t, \mathcal{A}_t}. \quad (\text{D.16})$$

On the other hand, $C_{\alpha, W_{[1, n]}, \mathcal{A}_1^n} \leq C_{\alpha, W_{[1, n]}, \mathcal{A}}$ because $\mathcal{A}_1^n \subset \mathcal{A}$. Then

$$\sum_{t=1}^n C_{\alpha, W_t, \mathcal{A}_t} \leq C_{\alpha, W_{[1, n]}, \mathcal{A}}. \quad (\text{D.17})$$

We proceed to prove $C_{\alpha, W_{[1, n]}, \mathcal{A}} \leq \sum_{t=1}^n C_{\alpha, W_t, \mathcal{A}_t}$. If there exists a $t \in \{1, \dots, n\}$ such that $C_{\alpha, W_t, \mathcal{A}_t} = \infty$, then the inequality holds trivially. Else $C_{\alpha, W_t, \mathcal{A}_t}$ is finite for all $t \in \{1, \dots, n\}$ and by Lemma 20 there exists a unique $q_{\alpha, W_t, \mathcal{A}_t}$ for each $t \in \{1, \dots, n\}$ such that

$$D_{\alpha}(W_t \| q_{\alpha, W_t, \mathcal{A}_t} | \tilde{p}_t) \leq C_{\alpha, W_t, \mathcal{A}_t} \quad \forall \tilde{p}_t \in \mathcal{A}_t.$$

Since the conditional Rényi divergence $D_{\alpha}(W_t \| q_{\alpha, W_t, \mathcal{A}_t} | \tilde{p}_t)$ is linear in the input distribution \tilde{p}_t , this implies

$$D_{\alpha}(W_t \| q_{\alpha, W_t, \mathcal{A}_t} | \tilde{p}_t) \leq C_{\alpha, W_t, \mathcal{A}_t} \quad \forall \tilde{p}_t \in \text{ch} \mathcal{A}_t. \quad (\text{D.18})$$

Let q be $q = \bigotimes_{t=1}^n q_{\alpha, W_t, \mathcal{A}_t}$. Then as a result of Tonelli-Fubini theorem [20, 4.4.5] we have

$$D_{\alpha}(W_{[1, n]}(x_1^n) \| q) = \sum_{t=1}^n D_{\alpha}(W_t(x_t) \| q_{\alpha, W_t, \mathcal{A}_t}) \quad \forall x_1^n \in \mathcal{X}_1^n.$$

Hence,

$$D_{\alpha}(W_{[1, n]} \| q | p) = \sum_{t=1}^n D_{\alpha}(W_t \| q_{\alpha, W_t, \mathcal{A}_t} | p_t) \quad \forall p \in \mathcal{P}(\mathcal{X}_1^n),$$

where $p_t \in \mathcal{P}(\mathcal{X}_t)$ is the \mathcal{X}_t marginal of p for each $t \in \{1, \dots, n\}$. Note that $p_t \in \text{ch} \mathcal{A}_t$ for all $t \in \{1, \dots, n\}$ by the definition constraint set \mathcal{A} . Thus (D.18) implies

$$D_{\alpha}(W_{[1, n]} \| q | p) \leq \sum_{t=1}^n C_{\alpha, W_t, \mathcal{A}_t} \quad \forall p \in \mathcal{A}. \quad (\text{D.19})$$

On the other hand $D_{\alpha}(W_{[1, n]} \| q | p) \geq I_{\alpha}(p; W_{[1, n]})$ by definition. Thus (D.17) and (D.19) imply $C_{\alpha, W_{[1, n]}, \mathcal{A}} = \sum_{t=1}^n C_{\alpha, W_t, \mathcal{A}_t}$ and $q_{\alpha, W_{[1, n]}, \mathcal{A}} = q$. Then $q_{\alpha, W_{[1, n]}, \mathcal{A}_1^n} = q$ by Lemma 25, as well, because $\mathcal{A}_1^n \subset \mathcal{A}$ and $C_{\alpha, W_{[1, n]}, \mathcal{A}_1^n} = C_{\alpha, W_{[1, n]}, \mathcal{A}}$. □

E. Proofs of Lemmas on the Cost Constrained Problem

Proof of Lemma 27.

- (27-a) If $\varrho_1 \leq \varrho_2$, then $C_{\alpha,W,\varrho_1} \leq C_{\alpha,W,\varrho_2}$ because $\mathcal{A}(\varrho_1) \subset \mathcal{A}(\varrho_2)$. Thus $C_{\alpha,W,\varrho}$ is nondecreasing in ϱ . Let $\varrho_\beta = \beta\varrho_1 + (1-\beta)\varrho_0$; then $(\beta p_1 + (1-\beta)p_0) \in \mathcal{A}(\varrho_\beta)$ for any $p_1 \in \mathcal{A}(\varrho_1)$ and $p_0 \in \mathcal{A}(\varrho_0)$. Hence, using the concavity of the Augustin information in its input distribution established in Lemma 15 we get

$$\begin{aligned} C_{\alpha,W,\varrho_\beta} &\geq \sup_{p_1 \in \mathcal{A}(\varrho_1), p_0 \in \mathcal{A}(\varrho_0)} I_\alpha(\beta p_1 + (1-\beta)p_0; W) \\ &\geq \sup_{p_1 \in \mathcal{A}(\varrho_1), p_0 \in \mathcal{A}(\varrho_0)} \beta I_\alpha(p_1; W) + (1-\beta)I_\alpha(p_0; W) \\ &= \beta C_{\alpha,W,\varrho_1} + (1-\beta)C_{\alpha,W,\varrho_0}. \end{aligned}$$

Thus $C_{\alpha,W,\varrho}$ is concave in ϱ .

Now let us proceed by proving that if $C_{\alpha,W,\varrho_0} = \infty$ for a $\varrho_0 \in \text{int}\Gamma_\rho$; then $C_{\alpha,W,\varrho} = \infty$ for all $\varrho \in \text{int}\Gamma_\rho$. Note that any point ϱ in $\text{int}\Gamma_\rho$ can be written as $\varrho = \beta\varrho_1 + (1-\beta)\varrho_0$ for some $\beta \in (0,1)$ and $\varrho_1 \in \text{int}\Gamma_\rho$ because $\text{int}\Gamma_\rho$ is a convex open subset \mathbb{R}^ℓ . Then $C_{\alpha,W,\varrho} = \infty$ follows from the concavity of $C_{\alpha,W,\varrho}$.

If $C_{\alpha,W,\varrho}$ is finite on $\text{int}\Gamma_\rho$, then $C_{\alpha,W,\varrho}$ is continuous on $\text{int}\Gamma_\rho$ by [20, Thm. 6.3.4] because $\text{int}\Gamma_\rho$ is a convex open subset \mathbb{R}^ℓ and $(-C_{\alpha,W,\varrho})$ is a convex function of ϱ on $\text{int}\Gamma_\rho$.

- (27-b) Let us extend the definition of $C_{\alpha,W,\varrho}$ from $\mathbb{R}_{\geq 0}^\ell$ to \mathbb{R}^ℓ by setting $C_{\alpha,W,\varrho}$ to $-\infty$ for all $\varrho \in \mathbb{R}^\ell \setminus \mathbb{R}_{\geq 0}^\ell$. Then $(-C_{\alpha,W,\varrho})$ is a proper convex function, i.e. $(-C_{\alpha,W,\varrho}): \mathbb{R}^\ell \rightarrow (-\infty, \infty]$ is a convex function and $\exists \varrho$ such that $(-C_{\alpha,W,\varrho}) < \infty$. Furthermore, $\text{int}\Gamma_\rho$ is also the interior of the effective domain of the extended function. Hence the sub-differential $\partial(-C_{\alpha,W,\varrho})$ is non-empty and compact by [41, Proposition 4.4.2]. Then (69) follows from the fact that the epigraph of a convex function lies above the tangent planes drawn at any point. The non-negativity of the components of $\lambda_{\alpha,W,\varrho}$ follows from the monotonicity of $C_{\alpha,W,\varrho}$ in ϱ .

- (27-c) If $C_{\tilde{\alpha},W,\tilde{\varrho}} = \infty$ for a $\tilde{\alpha} \in (0,1)$ and $\tilde{\varrho} \in \text{int}\Gamma_\rho$, then $C_{\alpha,W,\tilde{\varrho}} = \infty$ for all $\alpha \in (0,1)$ by Lemma 23-(a,b). Therefore, $C_{\alpha,W,\varrho} = \infty$ for all $\alpha \in (0,1)$ and $\varrho \in \text{int}\Gamma_\rho$ by part (a).

In order to prove the continuity when $C_{\alpha,W,\varrho}$ is finite, note that as a result of the triangle inequality we have

$$|C_{\alpha_1,W,\varrho_1} - C_{\alpha_2,W,\varrho_2}| \leq |C_{\alpha_1,W,\varrho_1} - C_{\alpha_1,W,\varrho_2}| + |C_{\alpha_1,W,\varrho_2} - C_{\alpha_2,W,\varrho_2}|.$$

The first term converges to zero as $\varrho_2 \rightarrow \varrho_1$ as a result of the continuity of the Augustin capacity in the constraint established in part (a). The second term converges to zero as $\alpha_2 \rightarrow \alpha_1$ because of (E.1) established in the following. Thus $C_{\alpha,W,\varrho}$ is continuous in the pair (α, ϱ) .

Using the monotonicity of $C_{\alpha,W,\varrho}$ and $\frac{1-\alpha}{\alpha}C_{\alpha,W,\varrho}$ established in Lemma 23-(a) and Lemma 23-(b) we get

$$|C_{\alpha_1,W,\varrho_2} - C_{\alpha_2,W,\varrho_2}| \leq \frac{|\alpha_2 - \alpha_1|}{(\alpha_1 \wedge \alpha_2)(1 - \alpha_1 \vee \alpha_2)} C_{\alpha_1,W,\varrho_2}.$$

Thus using (69) to bound C_{α_1,W,ϱ_2} we get

$$|C_{\alpha_1,W,\varrho_2} - C_{\alpha_2,W,\varrho_2}| \leq \frac{|\alpha_2 - \alpha_1|}{(\alpha_1 \wedge \alpha_2)(1 - \alpha_1 \vee \alpha_2)} (C_{\alpha_1,W,\varrho_1} + |\lambda_{\alpha_1,W,\varrho_1} \cdot (\varrho_2 - \varrho_1)|^+). \quad (\text{E.1})$$

In order to prove the continuity of the Augustin center, note that by the triangle inequality we have

$$\|q_{\alpha_1,W,\varrho_1} - q_{\alpha_2,W,\varrho_2}\| \leq \|q_{\alpha_1,W,\varrho_1} - q_{\alpha_1,W,\varrho_2}\| + \|q_{\alpha_1,W,\varrho_2} - q_{\alpha_2,W,\varrho_2}\|. \quad (\text{E.2})$$

Using first Lemmas 2 and 24, and then (E.1) we get

$$\begin{aligned} \|q_{\alpha_1,W,\varrho_2} - q_{\alpha_2,W,\varrho_2}\| &\leq \sqrt{\frac{2|C_{\alpha_1,W,\varrho_2} - C_{\alpha_2,W,\varrho_2}|}{\alpha_1 \wedge \alpha_2}} \\ &\leq \sqrt{\frac{2|\alpha_2 - \alpha_1|}{(\alpha_1 \wedge \alpha_2)^2(1 - \alpha_1 \vee \alpha_2)}} (C_{\alpha_1,W,\varrho_1} + |\lambda_{\alpha_1,W,\varrho_1} \cdot (\varrho_2 - \varrho_1)|^+). \end{aligned} \quad (\text{E.3})$$

In order to bound $\|q_{\alpha_1,W,\varrho_1} - q_{\alpha_1,W,\varrho_2}\|$, we use triangle inequality once more

$$\|q_{\alpha_1,W,\varrho_1} - q_{\alpha_1,W,\varrho_2}\| \leq \|q_{\alpha_1,W,\varrho_1} - q_{\alpha_1,W,\varrho_\vee}\| + \|q_{\alpha_1,W,\varrho_\vee} - q_{\alpha_1,W,\varrho_2}\|$$

where $\varrho_\vee = \varrho_1 \vee \varrho_2$, i.e. $\varrho_\vee^i = \varrho_1^i \vee \varrho_2^i$ for each $i \in \{1, \dots, \ell\}$.

On the other hand by Lemma 21 we have

$$\sup_{p \in \mathcal{A}(\varrho_1)} D_{\alpha_1}(W \| q_{\alpha_1,W,\varrho_\vee} | p) \geq C_{\alpha_1,W,\varrho_1} + D_{1 \wedge \alpha_1}(q_{\alpha_1,W,\varrho_1} \| q_{\alpha_1,W,\varrho_\vee}).$$

Since $\mathcal{A}(\varrho_1) \subset \mathcal{A}(\varrho_\vee)$, using Theorem 1 and Lemma 2 we get,

$$C_{\alpha_1,W,\varrho_\vee} - C_{\alpha_1,W,\varrho_1} \geq \frac{\alpha_1}{2} \|q_{\alpha_1,W,\varrho_1} - q_{\alpha_1,W,\varrho_\vee}\|^2.$$

Repeating the same analysis for $\|q_{\alpha_1, W, \varrho_2} - q_{\alpha_1, W, \varrho_1}\|$ and bounding $C_{\alpha_1, W, \varrho_1}$ using (69) we get

$$\|q_{\alpha_1, W, \varrho_1} - q_{\alpha_1, W, \varrho_2}\| \leq \sqrt{\frac{2}{\alpha_1}} \left(\sqrt{C_{\alpha_1, W, \varrho_1} - C_{\alpha_1, W, \varrho_2} + \lambda_{\alpha_1, W, \varrho_1} \cdot (\varrho_1 - \varrho_2)} + \sqrt{\lambda_{\alpha_1, W, \varrho_1} \cdot (\varrho_1 - \varrho_2)} \right) \quad (\text{E.4})$$

The continuity of $q_{\alpha, W, \varrho}$ for the total variation topology on $\mathcal{P}(\mathcal{Y})$ follows from (E.2), (E.3), (E.4) and the continuity of Augustin capacity as a function of the constraint established in part (a). \square

Proof of Lemma 28. Let $\mathcal{B}(\varrho)$ be

$$\mathcal{B}(\varrho) \triangleq \left\{ (\varrho_1, \dots, \varrho_n) : \sum_{t=1}^n \varrho_t \leq \varrho, \varrho_t \in \Gamma_{\rho_t} \right\}.$$

Note that if $\mathcal{B}(\varrho) = \emptyset$, then $\varrho \notin \Gamma_{\rho_{[1, n]}}$ and $C_{\alpha, W_{[1, n]}, \varrho} = -\infty$. On the other hand, $\sum_{t=1}^n C_{\alpha, W_t, \varrho_t}$ is minus infinity for $(\varrho_1, \dots, \varrho_n)$'s that are outside $\mathcal{B}(\varrho)$ by the convention stated in the lemma. Thus (70) holds for $\varrho \in \mathbb{R}_{\geq 0}^\ell \setminus \Gamma_{\rho_{[1, n]}}$ case and the constraints $\varrho_t \in \mathbb{R}_{\geq 0}^\ell$ can be replaced by $\varrho_t \in \Gamma_{\rho_t}$ for $\varrho \in \Gamma_{\rho_{[1, n]}}$ case in (70).

Furthermore, as a result of Lemma 26 for any $(\varrho_1, \dots, \varrho_n) \in \mathcal{B}(\varrho)$ we have

$$C_{\alpha, W_{[1, n]}, \times_{t=1}^n \mathcal{A}_t(\varrho_t)} = \sum_{t=1}^n C_{\alpha, W_t, \varrho_t}.$$

On the other hand $(\times_{t=1}^n \mathcal{A}_t(\varrho_t)) \subset \mathcal{A}(\varrho)$ for any $(\varrho_1, \dots, \varrho_n) \in \mathcal{B}(\varrho)$. Thus as a result of Lemma 25 we have

$$C_{\alpha, W_{[1, n]}, \varrho} \geq \sup \left\{ \sum_{t=1}^n C_{\alpha, W_t, \varrho_t} : \sum_{t=1}^n \varrho_t \leq \varrho, \varrho_t \in \Gamma_{\rho_t} \right\}.$$

For deriving the reverse inequality, first recall that Lemma 14 implies $I_\alpha(p; W_{[1, n]}) \leq \sum_{t=1}^n I_\alpha(p_t; W_t)$ for all $p \in \mathcal{P}(\mathcal{X}_1^n)$ where $p_t \in \mathcal{P}(\mathcal{X}_t)$ is the \mathcal{X}_t marginal of p . On the other hand, $\mathbf{E}_p[\rho_{[1, n]}] = \sum_{t=1}^n \mathbf{E}_{p_t}[\rho_t]$ and $\mathbf{E}_{p_t}[\rho_t] \in \Gamma_{\rho_t}$. Hence,

$$\begin{aligned} \sup_{p: \mathbf{E}_p[\rho_{[1, n]}] \leq \varrho} I_\alpha(p; W_{[1, n]}) &\leq \sup_{p_1, \dots, p_n: \sum_{t=1}^n \mathbf{E}_{p_t}[\rho_t] \leq \varrho} \sum_{t=1}^n I_\alpha(p_t; W_t) \\ &= \sup \left\{ \sum_{t=1}^n C_{\alpha, W_t, \varrho_t} : \sum_{t=1}^n \varrho_t \leq \varrho, \varrho_t \in \Gamma_{\rho_t} \right\}. \end{aligned}$$

Thus (70) holds. In addition, $\mathcal{A}(\varrho)$ can be interpreted as the union of $(\times_{t=1}^n \mathcal{A}_t(\varrho_t))$ and $\mathcal{A}(\varrho) \setminus (\times_{t=1}^n \mathcal{A}_t(\varrho_t))$. Therefore, if there exists a $(\varrho_1, \dots, \varrho_n)$ such that $C_{\alpha, W_{[1, n]}, \varrho} = \sum_{t=1}^n C_{\alpha, W_t, \varrho_t}$ and $C_{\alpha, W_{[1, n]}, \varrho} < \infty$, then $q_{\alpha, W_{[1, n]}, \varrho} = \bigotimes_{t=1}^n q_{\alpha, W_t, \varrho_t}$ because $C_{\alpha, W_{[1, n]}, \varrho} = C_{\alpha, W_{[1, n]}, \times_{t=1}^n \mathcal{A}_t(\varrho_t)}$ and $C_{\alpha, W_{[1, n]}, \varrho} < \infty$ imply $q_{\alpha, W_{[1, n]}, \varrho} = q_{\alpha, W_{[1, n]}, \times_{t=1}^n \mathcal{A}_t(\varrho_t)}$ by Lemma 25 and $q_{\alpha, W_{[1, n]}, \times_{t=1}^n \mathcal{A}_t(\varrho_t)} = \bigotimes_{t=1}^n q_{\alpha, W_t, \varrho_t}$ by Lemma 26. \square

Proof of Lemma 29.

(a) $C_{\alpha, W}^\lambda$ is convex, nonincreasing, and lower semicontinuous in λ because $C_{\alpha, W}^\lambda$ is the pointwise supremum of such functions as a result of (76).

Since $C_{\alpha, W}^\lambda$ is convex it is continuous on the interior of $\{\lambda \in \mathbb{R}_{\geq 0}^\ell : C_{\alpha, W}^\lambda < \infty\}$ by [20, Thm. 6.3.4]. The interior of $\{\lambda \in \mathbb{R}_{\geq 0}^\ell : C_{\alpha, W}^\lambda < \infty\}$ is $\{\lambda \in \mathbb{R}_{\geq 0}^\ell : \exists \epsilon > 0 \text{ s.t. } C_{\alpha, W}^{\lambda - \epsilon \mathbf{1}} < \infty\}$ because $C_{\alpha, W}^\lambda$ is nonincreasing in λ .

(b) Note that $C_{\alpha, W, \varrho} = \sup_{p \in \mathcal{P}(\mathcal{X})} \xi_{\alpha, p}(\varrho)$ as a result of (68) and (73). Then as a result of (72), we have

$$C_{\alpha, W, \varrho} = \sup_{p \in \mathcal{P}(\mathcal{X})} \inf_{\lambda \geq 0} I_\alpha^\lambda(p; W) + \lambda \cdot \varrho \quad \forall \varrho \in \mathbb{R}_{\geq 0}^\ell. \quad (\text{E.5})$$

If \mathcal{X} is finite, then $\mathcal{P}(\mathcal{X})$ is compact. Furthermore, using (D.1) together with triangle inequality we get

$$|I_\alpha^\lambda(p_2; W) - I_\alpha^\lambda(p_1; W)| \leq \hbar \left(\frac{\|p_1 - p_2\|}{2} \right) + \frac{\|p_1 - p_2\|}{2} \ln |\mathcal{X}| + \frac{\|p_1 - p_2\|}{2} \max_{x \in \mathcal{X}} \lambda \cdot \rho(x). \quad (\text{E.6})$$

Then $I_\alpha^\lambda(p; W) + \lambda \cdot \varrho$ is continuous in p on $\mathcal{P}(\mathcal{X})$. On the other hand, $I_\alpha^\lambda(p; W) + \lambda \cdot \varrho$ is concave in p by Lemma 15 and convex and continuous in λ . Thus we can change the order of the infimum and supremum in (E.5) —using the Sion's minimax theorem, [42], [43]— and $C_{\alpha, W, \varrho} = \inf_{\lambda \geq 0} C_{\alpha, W}^\lambda + \lambda \cdot \varrho$ by (75).

(c) If $\varrho \in \text{int} \Gamma_\rho$ and $C_{\alpha, W, \varrho}$ is infinite, then $C_{\alpha, W, \varrho} = \inf_{\lambda \geq 0} C_{\alpha, W}^\lambda + \lambda \cdot \varrho$ follows from (77) trivially.

If $\varrho \in \text{int} \Gamma_\rho$ and $C_{\alpha, W, \varrho}$ is finite, then there exists a non-empty, convex, and compact set of $\lambda_{\alpha, W, \varrho}$'s satisfying (69) by Lemma 27-(b). Furthermore, (76) implies for any $\lambda_{\alpha, W, \varrho}$ satisfying (69) the following identity

$$\begin{aligned} C_{\alpha, W}^{\lambda_{\alpha, W, \varrho}} &= \sup_{\tilde{\varrho} \geq 0} C_{\alpha, W, \tilde{\varrho}} - \lambda_{\alpha, W, \varrho} \cdot \tilde{\varrho} \\ &= C_{\alpha, W, \varrho} - \lambda_{\alpha, W, \varrho} \cdot \varrho. \end{aligned}$$

Then $C_{\alpha, W, \varrho} = \inf_{\lambda \geq 0} C_{\alpha, W}^\lambda + \lambda \cdot \varrho$ by (77).

(d) Note that $I_\alpha^\lambda(p; W) \leq C_{\alpha, W}^\lambda$ by definition. Hence $\limsup_{\iota \rightarrow \infty} I_\alpha^\lambda(p^{(\iota)}; W) \leq C_{\alpha, W}^\lambda$. Furthermore,

$$I_\alpha^\lambda(p; W) \geq I_\alpha(p; W) - \lambda \cdot \varrho \quad \forall p \in \mathcal{A}(\varrho).$$

Thus $\liminf_{\iota \rightarrow \infty} I_\alpha^\lambda(p^{(\iota)}; W) \geq C_{\alpha, W, \varrho} - \lambda \cdot \varrho = C_{\alpha, W}^\lambda$, as well. Thus $\lim_{\iota \rightarrow \infty} I_\alpha^\lambda(p^{(\iota)}; W) = C_{\alpha, W}^\lambda$. \square

Proof of Lemma 30.

- (i) $\forall \alpha \in \mathbb{R}_+ \exists \tilde{p} \in \mathcal{P}(\mathcal{X})$ such that $I_\alpha^\lambda(\tilde{p}; W) = C_{\alpha, W}^\lambda$: Note that $I_\alpha^\lambda(p; W)$ is continuous in p on $\mathcal{P}(\mathcal{X})$ by (E.6). On the other hand, $\mathcal{P}(\mathcal{X})$ is compact because \mathcal{X} is a finite set. Then there exists a $\tilde{p} \in \mathcal{P}(\mathcal{X})$ such that $I_\alpha^\lambda(\tilde{p}; W) = \sup_{p \in \mathcal{P}(\mathcal{X})} I_\alpha^\lambda(p; W)$ by the extreme value theorem, [39, 27.4].
- (ii) If $\alpha \in \mathbb{R}_+$ and $I_\alpha^\lambda(\tilde{p}; W) = C_{\alpha, W}^\lambda$, then $D_\alpha(W \| q_{\alpha, \tilde{p}} | p) - \lambda \cdot \mathbf{E}_p[\rho] \leq C_{\alpha, W}^\lambda$ for all $p \in \mathcal{P}(\mathcal{X})$: Let p be any member of $\mathcal{P}(\mathcal{X})$ and $p^{(\iota)}$ be $\frac{\iota-1}{\iota} \tilde{p} + \frac{1}{\iota} p$ for $\iota \in \mathbb{Z}_+$. Then by Lemma 13

$$\begin{aligned} I_\alpha^\lambda(p^{(\iota)}; W) &= \frac{\iota-1}{\iota} \left[D_\alpha(W \| q_{\alpha, p^{(\iota)}} | \tilde{p}) - \lambda \cdot \mathbf{E}_{\tilde{p}}[\rho] \right] + \frac{1}{\iota} \left[D_\alpha(W \| q_{\alpha, p^{(\iota)}} | p) - \lambda \cdot \mathbf{E}_p[\rho] \right] \\ &\geq \frac{\iota-1}{\iota} \left[I_\alpha^\lambda(\tilde{p}; W) + D_{\alpha \wedge 1}(q_{\alpha, \tilde{p}} \| q_{\alpha, p^{(\iota)}}) \right] + \frac{1}{\iota} \left[D_\alpha(W \| q_{\alpha, p^{(\iota)}} | p) - \lambda \cdot \mathbf{E}_p[\rho] \right] \quad \forall \iota \in \mathbb{Z}_+. \end{aligned}$$

Then using $I_\alpha^\lambda(p^{(\iota)}; W) \leq C_{\alpha, W}^\lambda$, $I_\alpha^\lambda(\tilde{p}; W) = C_{\alpha, W}^\lambda$, and $D_{\alpha \wedge 1}(q_{\alpha, \tilde{p}} \| q_{\alpha, p^{(\iota)}}) \geq 0$ we get

$$C_{\alpha, W}^\lambda \geq D_\alpha(W \| q_{\alpha, p^{(\iota)}} | p) - \lambda \cdot \mathbf{E}_p[\rho]. \quad (\text{E.7})$$

On the other hand, using $I_\alpha^\lambda(p^{(\iota)}; W) \leq C_{\alpha, W}^\lambda$, $I_\alpha^\lambda(\tilde{p}; W) = C_{\alpha, W}^\lambda$ and $D_\alpha(W \| q_{\alpha, p^{(\iota)}} | p) \geq 0$ we get

$$\frac{C_{\alpha, W}^\lambda + \lambda \cdot \mathbf{E}_p[\rho]}{\iota} \geq \frac{\iota-1}{\iota} D_{\alpha \wedge 1}(q_{\alpha, \tilde{p}} \| q_{\alpha, p^{(\iota)}}) \quad \forall \iota \in \mathbb{Z}_+.$$

Then using, Lemma 2 we get

$$\sqrt{\frac{2}{\alpha \wedge 1} \frac{C_{\alpha, W}^\lambda + \lambda \cdot \mathbf{E}_p[\rho]}{\iota-1}} \geq \|q_{\alpha, \tilde{p}} - q_{\alpha, p^{(\iota)}}\| \quad \forall \iota \in \mathbb{Z}_+.$$

Thus $q_{\alpha, p^{(\iota)}}$ converges to $q_{\alpha, \tilde{p}}$ in the total variation topology and hence in the topology of setwise convergence. Since the Rényi divergence is lower semicontinuous in the topology of setwise convergence by Lemma 3, we have

$$\liminf_{\iota \rightarrow \infty} D_\alpha(W \| q_{\alpha, p^{(\iota)}} | p) \geq D_\alpha(W \| q_{\alpha, \tilde{p}} | p). \quad (\text{E.8})$$

Then the inequality $D_\alpha(W \| q_{\alpha, \tilde{p}} | p) - \lambda \cdot \mathbf{E}_p[\rho] \leq C_{\alpha, W}^\lambda$ follows from (E.7) and (E.8).

- (iii) If $\alpha \in \mathbb{R}_+$, then $\exists! q_{\alpha, W}^\lambda \in \mathcal{P}(\mathcal{Y})$ satisfying (83) such that $q_{\alpha, p} = q_{\alpha, W}^\lambda$ for all $p \in \mathcal{P}(\mathcal{X})$ satisfying $I_\alpha^\lambda(p; W) = C_{\alpha, W}^\lambda$: If $I_\alpha^\lambda(p; W) = C_{\alpha, W}^\lambda$ for a $p \in \mathcal{P}(\mathcal{X})$, then Lemma 13-(b,c,d) and Lemma 2 imply

$$D_\alpha(W \| q_{\alpha, \tilde{p}} | p) - \lambda \cdot \mathbf{E}_p[\rho] \geq C_{\alpha, W}^\lambda + \frac{\alpha \wedge 1}{2} \|q_{\alpha, p} - q_{\alpha, \tilde{p}}\|^2. \quad (\text{E.9})$$

Since we have already established that $D_\alpha(W \| q_{\alpha, \tilde{p}} | p) - \lambda \cdot \mathbf{E}_p[\rho] \leq C_{\alpha, W}^\lambda$ for any $p \in \mathcal{P}(\mathcal{X})$, (E.9) implies that $q_{\alpha, p} = q_{\alpha, \tilde{p}}$ for any $p \in \mathcal{P}(\mathcal{X})$ satisfying $I_\alpha^\lambda(p; W) = C_{\alpha, W}^\lambda$. \square

Proof of Theorem 2. First note that (79) implies (80) and (81) implies (82). Furthermore, the left hand side of (79) is equal to $C_{\alpha, W}^\lambda$ by (78). Thus when $C_{\alpha, W}^\lambda$ is infinite, (79) holds trivially by the max-min inequality. When $C_{\alpha, W}^\lambda$ is finite, (79) follows from (81) and the max-min inequality. Thus we can assume $C_{\alpha, W}^\lambda$ to be finite and prove the claims about $q_{\alpha, W}^\lambda$, in order to prove the theorem.

- (i) If $C_{\alpha, W}^\lambda < \infty$ and $\lim_{\iota \rightarrow \infty} I_\alpha^\lambda(p^{(\iota)}; W) = C_{\alpha, W}^\lambda$, then $\{q_{\alpha, p^{(\iota)}}\}_{\iota \in \mathbb{Z}_+}$ is a Cauchy sequence in $\mathcal{P}(\mathcal{Y})$ for the total variation metric: For any sequence $\{p^{(\iota)}\}_{\iota \in \mathbb{Z}_+} \subset \mathcal{P}(\mathcal{X})$ satisfying $\lim_{\iota \rightarrow \infty} I_\alpha^\lambda(p^{(\iota)}; W) = C_{\alpha, W}^\lambda$, let us consider a sequence of channels $\{W^{(\iota)}\}_{\iota \in \mathbb{Z}_+}$ whose input sets $\{\mathcal{X}^{(\iota)}\}_{\iota \in \mathbb{Z}_+}$ form a nested sequence of finite subsets of \mathcal{X} defined as follows,

$$\mathcal{X}^{(\iota)} \triangleq \{x \in \mathcal{X} : \exists j \in \{1, \dots, \iota\} \text{ such that } p^{(j)}(x) > 0\}.$$

Then for any $\iota \in \mathbb{Z}_+$, there exists a unique $q_{\alpha, W^{(\iota)}}^\lambda$ satisfying (83) by Lemma 30. Furthermore, $\mathcal{P}(\mathcal{X}^{(j)}) \subset \mathcal{P}(\mathcal{X}^{(\iota)})$ for any $\iota, j \in \mathbb{Z}_+$ such that $j \leq \iota$. In order to bound $\|q_{\alpha, p^{(j)}} - q_{\alpha, p^{(\iota)}}\|$ for positive integers $j < \iota$ we use the triangle inequality for $q_{\alpha, p^{(j)}}$, $q_{\alpha, p^{(\iota)}}$, and $q_{\alpha, W^{(\iota)}}^\lambda$

$$\|q_{\alpha, p^{(j)}} - q_{\alpha, p^{(\iota)}}\| \leq \|q_{\alpha, p^{(j)}} - q_{\alpha, W^{(\iota)}}^\lambda\| + \|q_{\alpha, p^{(\iota)}} - q_{\alpha, W^{(\iota)}}^\lambda\|. \quad (\text{E.10})$$

Let us proceed with bounding $\|q_{\alpha,p^{(j)}} - q_{\alpha,W^{(j)}}^\lambda\|$ and $\|q_{\alpha,p^{(i)}} - q_{\alpha,W^{(i)}}^\lambda\|$ from above.

$$\begin{aligned} \|q_{\alpha,p^{(j)}} - q_{\alpha,W^{(j)}}^\lambda\| &\stackrel{(a)}{\leq} \sqrt{\frac{2}{\alpha\lambda 1} D_{\alpha\wedge 1}(q_{\alpha,p^{(j)}} \| q_{\alpha,W^{(j)}}^\lambda)} \\ &\stackrel{(b)}{\leq} \sqrt{\frac{2}{\alpha\lambda 1} \sqrt{D_\alpha(W \| q_{\alpha,W^{(j)}}^\lambda | p^{(j)}) - I_\alpha(p^{(j)}; W^{(j)})}} \\ &\stackrel{(c)}{\leq} \sqrt{\frac{2}{\alpha\lambda 1} \sqrt{C_{\alpha,W^{(j)}}^\lambda - I_\alpha(p^{(j)}; W^{(j)})}} \\ &\stackrel{(d)}{\leq} \sqrt{\frac{2}{\alpha\lambda 1} \sqrt{C_{\alpha,W}^\lambda - I_\alpha(p^{(j)}; W)}} \end{aligned}$$

where (a) follows from Lemma 2, (b) follows from Lemma 13-(b,c,d), (c) follows from Lemma 30 because $p^{(j)} \in \mathcal{P}(\mathcal{X}^{(j)})$, and (d) follows from the identities $I_\alpha^\lambda(p^{(j)}; W^{(j)}) = I_\alpha^\lambda(p^{(j)}; W)$ and $C_{\alpha,W^{(j)}}^\lambda \leq C_{\alpha,W}^\lambda$. We can obtain a similar bound on $\|q_{\alpha,p^{(i)}} - q_{\alpha,W^{(i)}}^\lambda\|$. Then $\{q_{\alpha,p^{(i)}}\}$ is a Cauchy sequence as a result of (E.10) because $\lim_{j \rightarrow \infty} I_\alpha^\lambda(p^{(j)}; W) = C_{\alpha,W}^\lambda$.

(ii) If $C_{\alpha,W}^\lambda < \infty$, then $\exists! q_{\alpha,W}^\lambda \in \mathcal{P}(\mathcal{Y})$ satisfying $\lim_{i \rightarrow \infty} \|q_{\alpha,W}^\lambda - q_{\alpha,p^{(i)}}\| = 0$ for all $\{p^{(i)}\}_{i \in \mathbb{Z}_+} \subset \mathcal{P}(\mathcal{X})$ such that $\lim_{i \rightarrow \infty} I_\alpha^\lambda(p^{(i)}; W) = C_{\alpha,W}^\lambda$: Note that $\mathcal{M}(\mathcal{Y})$ is a complete metric space for the total variation metric because $\mathcal{M}(\mathcal{Y})$ is a Banach space for the total variation topology [21, Thm. 4.6.1]. Then $\{q_{\alpha,p^{(i)}}\}_{i \in \mathbb{Z}_+}$ has a unique limit point q_{α,p^*} in $\mathcal{M}(\mathcal{Y})$. Since $\mathcal{P}(\mathcal{Y})$ is a closed set for the total variation topology and $\cup_{i \in \mathbb{Z}_+} q_{\alpha,p^{(i)}} \subset \mathcal{P}(\mathcal{Y})$, then $q_{\alpha,p^*} \in \mathcal{P}(\mathcal{Y})$, by [39, Thm. 2.1.3].

We have established the existence of a unique limit point q_{α,p^*} , for any sequence $\{p^{(i)}\}_{i \in \mathbb{Z}_+} \subset \mathcal{P}(\mathcal{X})$ satisfying $\lim_{i \rightarrow \infty} I_\alpha^\lambda(p^{(i)}; W) = C_{\alpha,W}^\lambda$. This, however, implies $\lim_{i \rightarrow \infty} \|q_{\alpha,\tilde{p}^{(i)}} - q_{\alpha,p^*}\| = 0$ for any $\{\tilde{p}^{(i)}\}_{i \in \mathbb{Z}_+}$ satisfying $\lim_{i \rightarrow \infty} I_\alpha^\lambda(\tilde{p}^{(i)}; W) = C_{\alpha,W}^\lambda$ because we can interleave the elements of $\{p^{(i)}\}_{i \in \mathbb{Z}_+}$ and $\{\tilde{p}^{(i)}\}_{i \in \mathbb{Z}_+}$ to obtain a new sequence $\{\hat{p}^{(i)}\}_{i \in \mathbb{Z}_+}$ satisfying $\lim_{i \rightarrow \infty} I_\alpha^\lambda(\hat{p}^{(i)}; W) = C_{\alpha,W}^\lambda$ for which $\{q_{\alpha,\hat{p}^{(i)}}\}$ is a Cauchy sequence. Then $q_{\alpha,W}^\lambda = q_{\alpha,p^*}$.

(iii) $q_{\alpha,W}^\lambda$ satisfies the equality given in (81): For any $p \in \mathcal{P}(\mathcal{X})$, let us consider any sequence $\{p^{(i)}\}_{i \in \mathbb{Z}_+}$ satisfying $p^{(1)} = p$ and $\lim_{i \rightarrow \infty} I_\alpha^\lambda(p^{(i)}; W) = C_{\alpha,W}^\lambda$. Then $p \in \mathcal{P}(\mathcal{X}^{(i)})$ for all $i \in \mathbb{Z}_+$. Using Lemma 30 we get

$$D_\alpha(W \| q_{\alpha,W^{(i)}} | p) - \lambda \cdot \mathbf{E}_p[\rho] \leq C_{\alpha,W^{(i)}}^\lambda \quad \forall i \in \mathbb{Z}_+. \quad (\text{E.11})$$

Since $\mathcal{X}^{(i)}$ is a finite set, $\exists \tilde{p}^{(i)} \in \mathcal{P}(\mathcal{X}^{(i)})$ satisfying $I_\alpha^\lambda(\tilde{p}^{(i)}; W^{(i)}) = C_{\alpha,W^{(i)}}^\lambda$ and $q_{\alpha,\tilde{p}^{(i)}} = q_{\alpha,W^{(i)}}^\lambda$ by Lemma 30. Then $I_\alpha^\lambda(\tilde{p}^{(i)}; W^{(i)}) \geq I_\alpha^\lambda(p^{(i)}; W^{(i)})$ and consequently $\lim_{i \rightarrow \infty} I_\alpha^\lambda(\tilde{p}^{(i)}; W) = C_{\alpha,W}^\lambda$. We have already established that for such a sequence $q_{\alpha,\tilde{p}^{(i)}} \rightarrow q_{\alpha,W}^\lambda$ in the total variation topology, and hence in the topology of setwise convergence. Then the lower semicontinuity of the Rényi divergence in its arguments for the topology of setwise convergence, i.e. Lemma 3, the identity $C_{\alpha,W^{(i)}}^\lambda \leq C_{\alpha,W}^\lambda$, and (E.11) imply that

$$D_\alpha(W \| q_{\alpha,W}^\lambda | p) - \lambda \cdot \mathbf{E}_p[\rho] \leq C_{\alpha,W}^\lambda \quad \forall p \in \mathcal{P}(\mathcal{X}).$$

On the other hand $D_\alpha(W \| q_{\alpha,W}^\lambda | p) - \lambda \cdot \mathbf{E}_p[\rho] \geq I_\alpha^\lambda(p; W)$ and $C_{\alpha,W}^\lambda = \sup_{p \in \mathcal{P}(\mathcal{X})} I_\alpha^\lambda(p; W)$ by the definitions of $I_\alpha(p; W)$, $I_\alpha^\lambda(p; W)$, and $C_{\alpha,W}^\lambda$. Thus (81) holds. \square

Proof of Lemma 31. Let $\{p^{(i)}\}_{i \in \mathbb{Z}_+} \subset \mathcal{A}(\varrho)$ be such that $\lim_{i \rightarrow \infty} I_\alpha(p^{(i)}; W) = C_{\alpha,W,\varrho}$. Then $\{q_{\alpha,p^{(i)}}\}_{i \in \mathbb{Z}_+}$ is a Cauchy sequence with the limit point $q_{\alpha,W,\varrho}$ by Theorem 1. On the other hand, $\lim_{i \rightarrow \infty} I_\alpha^\lambda(p^{(i)}; W) = C_{\alpha,W}^\lambda$ by Lemma 29-(d). Then $\{q_{\alpha,p^{(i)}}\}_{i \in \mathbb{Z}_+}$ is a Cauchy sequence with the limit point $q_{\alpha,W}^\lambda$ by Theorem 2. Hence $q_{\alpha,W,\varrho} = q_{\alpha,W}^\lambda$. \square

Proof of Lemma 32. As a result of (86) we have

$$\sup_{p \in \mathcal{P}(\mathcal{X}_1^n)} I_\alpha^\lambda(p; W_{[1,n]}) = \sup_{p_1 \in \mathcal{P}(\mathcal{X}_1), \dots, p_n \in \mathcal{P}(\mathcal{X}_n)} \sum_{t=1}^n I_\alpha^\lambda(p_t; W_t).$$

Thus (85) holds. In order to establish $q_{\alpha,W_{[1,n]}}^\lambda = \otimes_{t=1}^n q_{\alpha,W_t}^\lambda$, one can confirm by substitution that $\otimes_{t=1}^n q_{\alpha,W_t}^\lambda$ satisfies (82). \square

Proof of Lemma 33.

(a) Note that as a result of Lemma 13-(c,d) and the definition of $I_\alpha^\lambda(p; W)$ given in (71) we have

$$D_1(p \| u_{\alpha,p}^\lambda) = (\alpha - 1) I_\alpha^\lambda(p; W) + \ln \sum_{\tilde{x}} p(\tilde{x}) e^{(1-\alpha) D_\alpha(W(\tilde{x}) \| q_{\alpha,p}) + (\alpha-1)\lambda \cdot \rho(x)}. \quad (\text{E.12})$$

On the other hand as a result of (38), (88), and (92)

$$\begin{aligned} I_\alpha^\lambda(u_{\alpha,p}^\lambda; W) &= \frac{\alpha}{\alpha-1} \ln \int \left(\sum_x u_{\alpha,p}^\lambda(x) e^{(1-\alpha)\lambda \cdot \rho(x)} \left(\frac{dW(x)}{d\nu} \right)^\alpha \right)^{1/\alpha} \nu(dy) \\ &= \frac{\alpha}{\alpha-1} \ln \int \frac{d q_{\alpha,p}}{d\nu} \nu(dy) - \frac{1}{\alpha-1} \ln \sum_{\tilde{x}} p(\tilde{x}) e^{(1-\alpha) D_\alpha(W(\tilde{x}) \| q_{\alpha,p}) + (\alpha-1)\lambda \cdot \rho(x)}. \end{aligned}$$

Then (95) follows from (E.12)

- In order to prove (96) for $\alpha \in (0, 1)$ case, we prove the following inequality

$$I_{\alpha}^{g\lambda}(u; W) + \frac{1}{\alpha-1} D_1(p \| u) \leq I_{\alpha}^{\lambda}(p; W) \quad \forall u \in \mathcal{P}(\mathcal{X}).$$

Preceding inequality together with (95) imply (96) for $\alpha \in (0, 1)$. Note that the inequality holds trivially when $p \not\prec u$ because $D_1(p \| u)$ is infinite in that case. Thus we are left with $p \prec u$ case. On the other hand, any $u \in \mathcal{P}(\mathcal{X})$ can be written as $u = u_{ac} + u_s$ where $u_{ac} \prec p$ and $u_s \perp p$. Then

$$\begin{aligned} I_{\alpha}^{g\lambda}(u; W) &\stackrel{(i)}{\leq} D_{\alpha}\left(u \otimes W e^{\frac{1-\alpha}{\alpha} \lambda \cdot \rho} \middle\| u \otimes q_{\alpha, p}\right) \\ &\stackrel{(ii)}{\leq} \frac{1}{\alpha-1} \ln \left[\sum_x u_{ac}(x) e^{(\alpha-1) D_{\alpha}(W(x) \| q_{\alpha, p}) + (1-\alpha) \lambda \cdot \rho(x)} \right] \\ &\stackrel{(iii)}{=} \frac{1}{\alpha-1} \ln \left[\sum_x p(x) \frac{u_{ac}(x)}{p(x)} e^{(\alpha-1) D_{\alpha}(W(x) \| q_{\alpha, p}) + (1-\alpha) \lambda \cdot \rho(x)} \right] \\ &\stackrel{(iv)}{\leq} \frac{1}{\alpha-1} \left[\sum_x p(x) \ln \frac{u_{ac}(x)}{p(x)} e^{(\alpha-1) D_{\alpha}(W(x) \| q_{\alpha, p}) + (1-\alpha) \lambda \cdot \rho(x)} \right] \\ &\stackrel{(v)}{=} I_{\alpha}^{\lambda}(p; W) - \frac{1}{\alpha-1} D_1(p \| u_{ac}) \\ &\stackrel{(vi)}{=} I_{\alpha}^{\lambda}(p; W) - \frac{1}{\alpha-1} D_1(p \| u). \end{aligned}$$

where (i) follows from (87), (ii) follows from (8) and the monotonicity of the natural logarithm function, (iii) follows from $u_{ac} \sim p$ which holds because $p \prec u$, (iv) follows from the Jensen's inequality and the concavity of the natural logarithm function, (v) follows from Lemma 13-(c) and the definition of $I_{\alpha}^{\lambda}(p; W)$ given in (71).

- In order to prove (96) for $\alpha \in (1, \infty)$ case, we prove the following inequality

$$I_{\alpha}^{g\lambda}(u; W) + \frac{1}{\alpha-1} D_1(p \| u) \geq I_{\alpha}^{\lambda}(p; W) \quad \forall u \in \mathcal{P}(\mathcal{X}).$$

Preceding inequality together with (95) imply (96) for $\alpha \in (1, \infty)$. Note that the inequality holds trivially when $p \not\prec u$ because $D_1(p \| u)$ is infinite in that case. Thus we are left with $p \prec u$ case. On the other hand, any $u \in \mathcal{P}(\mathcal{X})$ can be written as $u = u_{ac} + u_s$ where $u_{ac} \prec p$ and $u_s \perp p$. Then

$$\begin{aligned} I_{\alpha}^{g\lambda}(u; W) &\stackrel{(i)}{=} D_{\alpha}\left(u \otimes W e^{\frac{1-\alpha}{\alpha} \lambda \cdot \rho} \middle\| u \otimes q_{\alpha, u}^{g\lambda}\right) \\ &\stackrel{(ii)}{\geq} \frac{1}{\alpha-1} \ln \left[\sum_x u_{ac}(x) e^{(\alpha-1) D_{\alpha}(W(x) \| q_{\alpha, u}^{g\lambda}) + (1-\alpha) \lambda \cdot \rho(x)} \right] \\ &\stackrel{(iii)}{=} \frac{1}{\alpha-1} \ln \left[\sum_x p(x) \frac{u_{ac}(x)}{p(x)} e^{(\alpha-1) D_{\alpha}(W(x) \| q_{\alpha, u}^{g\lambda}) + (1-\alpha) \lambda \cdot \rho(x)} \right] \\ &\stackrel{(iv)}{\geq} \frac{1}{\alpha-1} \left[\sum_x p(x) \ln \frac{u_{ac}(x)}{p(x)} e^{(\alpha-1) D_{\alpha}(W(x) \| q_{\alpha, u}^{g\lambda}) + (1-\alpha) \lambda \cdot \rho(x)} \right] \\ &\stackrel{(v)}{\geq} I_{\alpha}(p; W) + D_1(q_{\alpha, p} \| q_{\alpha, u}^{g\lambda}) - \lambda \cdot \mathbf{E}_p[\rho] - \frac{1}{\alpha-1} D_1(p \| u_{ac}) \\ &\stackrel{(vi)}{=} I_{\alpha}^{\lambda}(p; W) - \frac{1}{\alpha-1} D_1(p \| u). \end{aligned}$$

where (i) follows from (91), (ii) follows from (8) and the monotonicity of the natural logarithm function, (iii) follows from $u_{ac} \sim p$ which holds because $p \prec u$, (iv) follows from the Jensen's inequality and the concavity of the natural logarithm function, (v) follows from Lemma 13-(d), (vi) follows from Lemma 2 and the definition of $I_{\alpha}^{\lambda}(p; W)$ given in (71).

- (b) Note that the order α R-G mean for the input distribution p and the Lagrange multiplier λ is a fixed point of the order α Augustin operator for the input distribution $a_{\alpha, p}^{\lambda}$, i.e.

$$\begin{aligned} \frac{d\mathbf{T}_{\alpha, a_{\alpha, p}^{\lambda}}(q_{\alpha, p}^{g\lambda})}{d\nu} &= \sum_x a_{\alpha, p}^{\lambda}(x) \left(\frac{dW(x)}{d\nu} \right)^{\alpha} \left(\frac{dq_{\alpha, p}^{g\lambda}}{d\nu} \right)^{1-\alpha} e^{(1-\alpha) D_{\alpha}(W(x) \| q_{\alpha, p}^{g\lambda})} \\ &= \frac{1}{e^{(\alpha-1) I_{\alpha}^{g\lambda}(p; W)}} \sum_x p(x) e^{(1-\alpha) \lambda \cdot \rho(x)} \left(\frac{dW(x)}{d\nu} \right)^{\alpha} \left(\frac{dq_{\alpha, p}^{g\lambda}}{d\nu} \right)^{1-\alpha} \\ &= \frac{dq_{\alpha, p}^{g\lambda}}{d\nu}. \end{aligned}$$

Consequently $I_\alpha(a_{\alpha,p}^\lambda; W) = D_\alpha(W \| q_{\alpha,p}^{g\lambda} | a_{\alpha,p}^\lambda)$ by Lemma 13-(c,d). Then

$$\begin{aligned} D_1(a_{\alpha,p}^\lambda \| p) &= \sum_x a_{\alpha,p}^\lambda(x) \ln \frac{p(x) e^{(\alpha-1)D_\alpha(W(x) \| q_{\alpha,p}^{g\lambda}) + (1-\alpha)\lambda \cdot \rho(x)}}{\sum_{\bar{x}} p(\bar{x}) e^{(\alpha-1)D_\alpha(W(\bar{x}) \| q_{\alpha,p}^{g\lambda}) + (1-\alpha)\lambda \cdot \rho(\bar{x})}} \frac{1}{p(x)} \\ &= (\alpha-1)I_\alpha^\lambda(a_{\alpha,p}^\lambda; W) - \ln \sum_{\bar{x}} p(\bar{x}) e^{(\alpha-1)D_\alpha(W(\bar{x}) \| q_{\alpha,p}^{g\lambda}) + (1-\alpha)\lambda \cdot \rho(\bar{x})} \\ &= (\alpha-1)[I_\alpha^\lambda(a_{\alpha,p}^\lambda; W) - I_\alpha^\lambda(p; W)]. \end{aligned}$$

Thus (97) holds.

- In order to prove (98) for $\alpha \in (0, 1)$ case, we prove the following inequality

$$I_\alpha^\lambda(a; W) - \frac{1}{\alpha-1}D_1(a \| p) \geq I_\alpha^\lambda(p; W) \quad \forall a \in \mathcal{P}(\mathcal{X}).$$

Preceding inequality together with (97) imply (98) for $\alpha \in (0, 1)$. Note that the inequality holds trivially when $a \not\prec p$ because $D_1(a \| p)$ is infinite in that case. Thus we are left with $a \prec p$ case. On the other hand, for any $a \in \mathcal{P}(\mathcal{X})$, p can be written as $p = p_{ac} + p_s$ where $p_{ac} \prec a$ and $p_s \perp a$. Then

$$\begin{aligned} I_\alpha^\lambda(a; W) - \frac{1}{\alpha-1}D_1(a \| p) &\stackrel{(i)}{=} D_\alpha(W \| q_{\alpha,a} | a) - \lambda \cdot \mathbf{E}_a[\rho] - \frac{1}{\alpha-1}D_1(a \| p_{ac}) \\ &\stackrel{(ii)}{=} \frac{1}{\alpha-1} \sum_x a(x) \ln \left[\frac{p_{ac}(x)}{a(x)} e^{(\alpha-1)D_\alpha(W(x) \| q_{\alpha,a}) + (1-\alpha)\lambda \cdot \rho(x)} \right] \\ &\stackrel{(iii)}{\geq} \frac{1}{\alpha-1} \ln \sum_x p_{ac}(x) e^{(\alpha-1)D_\alpha(W(x) \| q_{\alpha,a}) + (1-\alpha)\lambda \cdot \rho(x)} \\ &\stackrel{(iv)}{\geq} \frac{1}{\alpha-1} \ln \sum_x p(x) e^{(\alpha-1)D_\alpha(W(x) \| q_{\alpha,a}) + (1-\alpha)\lambda \cdot \rho(x)} \\ &\stackrel{(v)}{\geq} I_\alpha^\lambda(p; W). \end{aligned}$$

where (i) follows from (8), (71), and Lemma 13-(c), (ii) follows from $p_{ac} \sim a$ which holds because $a \prec p$, (iii) follows from the Jensen's inequality and the concavity of the natural logarithm function, (iv) follows from the monotonicity of the natural logarithm function, (v) follows from (8) and (87).

- In order to prove (98) for $\alpha \in (1, \infty)$ case, we prove the following inequality

$$I_\alpha^\lambda(a; W) - \frac{1}{\alpha-1}D_1(a \| p) \leq I_\alpha^\lambda(p; W) \quad \forall a \in \mathcal{P}(\mathcal{X}).$$

Preceding inequality together with (97) imply (98) for $\alpha \in (1, \infty)$. Note that the inequality holds trivially when $a \not\prec p$ because $D_1(a \| p)$ is infinite in that case. Thus we are left with $a \prec p$ case. On the other hand, for any $a \in \mathcal{P}(\mathcal{X})$, p can be written as $p = p_{ac} + p_s$ where $p_{ac} \prec a$ and $p_s \perp a$. Then

$$\begin{aligned} I_\alpha^\lambda(a; W) - \frac{1}{\alpha-1}D_1(a \| p) &\stackrel{(i)}{\leq} D_\alpha(W \| q_{\alpha,p}^{g\lambda} | a) - \lambda \cdot \mathbf{E}_a[\rho] - \frac{1}{\alpha-1}D_1(a \| p_{ac}) \\ &\stackrel{(ii)}{=} \frac{1}{\alpha-1} \sum_x a(x) \ln \left[\frac{p_{ac}(x)}{a(x)} e^{(\alpha-1)D_\alpha(W(x) \| q_{\alpha,p}^{g\lambda}) + (1-\alpha)\lambda \cdot \rho(x)} \right] \\ &\stackrel{(iii)}{\leq} \frac{1}{\alpha-1} \ln \sum_x p_{ac}(x) e^{(\alpha-1)D_\alpha(W(x) \| q_{\alpha,p}^{g\lambda}) + (1-\alpha)\lambda \cdot \rho(x)} \\ &\stackrel{(iv)}{\leq} \frac{1}{\alpha-1} \ln \sum_x p(x) e^{(\alpha-1)D_\alpha(W(x) \| q_{\alpha,p}^{g\lambda}) + (1-\alpha)\lambda \cdot \rho(x)} \\ &\stackrel{(v)}{=} I_\alpha^\lambda(p; W). \end{aligned}$$

where (i) follows from (8), (23), and (71), (ii) follows from $p_{ac} \sim a$ which holds because $a \prec p$, (iii) follows from the Jensen's inequality and the concavity of the natural logarithm function, (iv) follows from the monotonicity of the natural logarithm function, (v) follows from (8) and (91).

(c) (99) follows from (38) by substitution. On the other hand, (92) and (96) imply

$$\frac{\alpha-1}{\alpha}I_\alpha^\lambda(p; W) \leq \ln \|\mu_{\alpha,u}^\lambda\| + \frac{D_1(p \| u)}{\alpha} \quad \forall u \in \mathcal{P}(\mathcal{X}). \quad (\text{E.13})$$

For any f satisfying $f : \mathbf{E}_p[f] = 0$, let $u_f \in \mathcal{P}(\mathcal{X})$ be $u_f(x) \triangleq \frac{p(x)e^{(1-\alpha)f(x)}}{\sum_z p(z)e^{(1-\alpha)f(z)}}$ for all $x \in \mathcal{X}$. Thus as a result of (E.13) and (88) we have

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

Then (100) follows from (99). \square

Proof of Lemma 34.

(i) $\exists \tilde{p} \in \mathcal{P}(\mathcal{X})$ such that $I_{\alpha}^{g\lambda}(\tilde{p}; W) = C_{\alpha, W}^{g\lambda}$: Note that $\mathcal{P}(\mathcal{X})$ is compact because \mathcal{X} is a finite set. If $I_{\alpha}^{g\lambda}(p; W)$ is continuous in p , then the existence of \tilde{p} follows from the extreme value theorem, [39, 27.4]. Thus we are left with establishing the continuity of $I_{\alpha}^{g\lambda}(p; W)$ in p .

Note that for any p_1 and p_0 there exist probability mass functions s_1, s_0 , and s_{\wedge} satisfying $s_0 \perp s_1, p_1 = (1 - \delta)s_{\wedge} + \delta s_1$, and $p_0 = (1 - \delta)s_{\wedge} + \delta s_0$ where $\delta = \frac{\|p_1 - p_0\|}{2}$. Then applying first (90) and (91) we get

$$I_{\alpha}^{g\lambda}(p_1; W) = \frac{1}{\alpha-1} \ln \left[(1 - \delta)e^{(\alpha-1)[I_{\alpha}^{g\lambda}(s_{\wedge}; W) + D_{\alpha}(q_{\alpha, s_{\wedge}}^{\lambda} \| q_{\alpha, p_1}^{\lambda})]} + \delta e^{(\alpha-1)[I_{\alpha}^{g\lambda}(s_1; W) + D_{\alpha}(q_{\alpha, s_1}^{\lambda} \| q_{\alpha, p_1}^{\lambda})]} \right]. \quad (\text{E.14})$$

Note that $I_{\alpha}^{g\lambda}(p; W) \leq I_{\alpha}^g(p; W)$ and $I_{\alpha}^g(p; W) \leq D_{\alpha}(p \otimes W \| p \otimes q_{1, u})$ for all $p \in \mathcal{P}(\mathcal{X})$ by definition where u is the uniform distribution on \mathcal{X} . Furthermore, $D_{\alpha}(p \otimes W \| p \otimes q_{1, u}) = \frac{1}{\alpha-1} \ln \sum_x p(x) e^{(\alpha-1)D_{\alpha}(W(x) \| q_{1, u})} \leq \ln |\mathcal{X}|$ for all $p \in \mathcal{P}(\mathcal{X})$ by Lemma 1. Thus $I_{\alpha}^{g\lambda}(s_{\wedge}; W) \leq \ln |\mathcal{X}|$ and using Lemma 2 to bound the expression in (E.14) we get

$$I_{\alpha}^{g\lambda}(p_1; W) \geq I_{\alpha}^{g\lambda}(s_{\wedge}; W) + \frac{1}{\alpha-1} \ln \left[(1 - \delta) + \delta e^{(1-\alpha) \ln |\mathcal{X}|} \right]. \quad (\text{E.15})$$

On the other hand $(1 - \delta)^{\frac{1}{\alpha}} \mu_{\alpha, s_{\wedge}}^{\lambda} \leq \mu_{\alpha, p_1}^{\lambda}$ and $\delta^{\frac{1}{\alpha}} \mu_{\alpha, s_1}^{\lambda} \leq \mu_{\alpha, p_1}^{\lambda}$ by (88). Then using (89) and Lemma 1 we get

$$\begin{aligned} D_{\alpha}(q_{\alpha, s_{\wedge}}^{\lambda} \| q_{\alpha, p_1}^{\lambda}) &\leq \frac{1}{\alpha} \ln \frac{1}{1-\delta} - \frac{\alpha-1}{\alpha} (I_{\alpha}^{g\lambda}(s_{\wedge}; W) - I_{\alpha}^{g\lambda}(p_1; W)), \\ D_{\alpha}(q_{\alpha, s_1}^{\lambda} \| q_{\alpha, p_1}^{\lambda}) &\leq \frac{1}{\alpha} \ln \frac{1}{\delta} - \frac{\alpha-1}{\alpha} (I_{\alpha}^{g\lambda}(s_1; W) - I_{\alpha}^{g\lambda}(p_1; W)). \end{aligned}$$

Since $I_{\alpha}^{g\lambda}(s_1; W) \leq I_{\alpha}^g(s_1; W) \leq \ln |\mathcal{X}|$ using (E.14) and we get

$$I_{\alpha}^{g\lambda}(p_1; W) \leq I_{\alpha}^{g\lambda}(s_{\wedge}; W) + \frac{\alpha}{\alpha-1} \ln \left[(1 - \delta)^{\frac{1}{\alpha}} + \delta^{\frac{1}{\alpha}} e^{\frac{\alpha-1}{\alpha} \ln |\mathcal{X}|} \right]. \quad (\text{E.16})$$

Using (E.15) and (E.16) we get

$$|I_{\alpha}^{g\lambda}(p_1; W) - I_{\alpha}^{g\lambda}(p_2; W)| \leq \frac{\alpha}{\alpha-1} \ln \left[(1 - \delta)^{\frac{1}{\alpha}} + \delta^{\frac{1}{\alpha}} e^{\frac{\alpha-1}{\alpha} \ln |\mathcal{X}|} \right] - \frac{1}{\alpha-1} \ln \left[(1 - \delta) + \delta e^{(1-\alpha) \ln |\mathcal{X}|} \right].$$

Then $I_{\alpha}^{g\lambda}(p; W)$ is continuous in p .

(ii) If $I_{\alpha}^{g\lambda}(\tilde{p}; W) = C_{\alpha, W}^{g\lambda}$, then $D_{\alpha}(p \otimes W e^{\frac{1-\alpha}{\alpha} \lambda \cdot \rho} \| p \otimes q_{\alpha, \tilde{p}}^{g\lambda}) \leq C_{\alpha, W}^{g\lambda}$ for all $p \in \mathcal{P}(\mathcal{X})$: Let $\tilde{p} \in \mathcal{P}(\mathcal{X})$ be such that $I_{\alpha}^{g\lambda}(\tilde{p}; W) = C_{\alpha, W}^{g\lambda}$, p be any member of $\mathcal{P}(\mathcal{X})$ and $p^{(\iota)}$ be $\frac{\iota-1}{\iota} \tilde{p} + \frac{1}{\iota} p$ for $\iota \in \mathbb{Z}_+$. Then

$$I_{\alpha}^{g\lambda}(p^{(\iota)}; W) = \frac{1}{\alpha-1} \ln \left[\frac{\iota-1}{\iota} e^{(\alpha-1) \left(I_{\alpha}^{g\lambda}(\tilde{p}; W) + D_{\alpha}(q_{\alpha, \tilde{p}}^{g\lambda} \| q_{\alpha, p^{(\iota)}}^{g\lambda}) \right)} + \frac{1}{\iota} e^{(\alpha-1) \left(I_{\alpha}^{g\lambda}(p; W) + D_{\alpha}(q_{\alpha, p}^{g\lambda} \| q_{\alpha, p^{(\iota)}}^{g\lambda}) \right)} \right].$$

Then using $I_{\alpha}^{g\lambda}(p^{(\iota)}; W) \leq C_{\alpha, W}^{g\lambda}$, $I_{\alpha}^{g\lambda}(\tilde{p}; W) = C_{\alpha, W}^{g\lambda}$, and $D_{\alpha}(q_{\alpha, \tilde{p}}^{g\lambda} \| q_{\alpha, p^{(\iota)}}^{g\lambda}) \geq 0$ we get

$$I_{\alpha}^{g\lambda}(p; W) + D_{\alpha}(q_{\alpha, p}^{g\lambda} \| q_{\alpha, p^{(\iota)}}^{g\lambda}) \leq C_{\alpha, W}^{g\lambda} \quad \forall \iota \in \mathbb{Z}_+. \quad (\text{E.17})$$

On the other hand using $I_{\alpha}^{g\lambda}(p^{(\iota)}; W) \leq C_{\alpha, W}^{g\lambda}$, $I_{\alpha}^{g\lambda}(\tilde{p}; W) = C_{\alpha, W}^{g\lambda}$, $I_{\alpha}^{g\lambda}(p; W) \geq 0$, and $D_{\alpha}(q_{\alpha, p}^{g\lambda} \| q_{\alpha, p^{(\iota)}}^{g\lambda}) \geq 0$, we get

$$D_{\alpha}(q_{\alpha, \tilde{p}}^{g\lambda} \| q_{\alpha, p^{(\iota)}}^{g\lambda}) \leq \frac{1}{\alpha-1} \ln \frac{\iota - e^{(1-\alpha)C_{\alpha, W}^{g\lambda}}}{\iota-1} \quad \forall \iota \in \mathbb{Z}_+.$$

Thus Lemma 2 implies

$$\limsup_{\iota \rightarrow \infty} \left\| q_{\alpha, \tilde{p}}^{g\lambda} - q_{\alpha, p^{(\iota)}}^{g\lambda} \right\| \leq 0.$$

Then $q_{\alpha, p^{(\iota)}}^{g\lambda}$ converges to $q_{\alpha, \tilde{p}}^{g\lambda}$ in the total variation topology and hence in the topology of setwise convergence. Since the Rényi divergence is lower semicontinuous in the topology of setwise convergence by Lemma 3, we have

$$D_{\alpha}(q_{\alpha, p}^{g\lambda} \| q_{\alpha, \tilde{p}}^{g\lambda}) \leq \liminf_{\iota \rightarrow \infty} D_{\alpha}(q_{\alpha, p}^{g\lambda} \| q_{\alpha, p^{(\iota)}}^{g\lambda}). \quad (\text{E.18})$$

Equations (90), (91), (E.17), (E.18) imply that $D_{\alpha}(p \otimes W e^{\frac{1-\alpha}{\alpha} \lambda \cdot \rho} \| p \otimes q_{\alpha, \tilde{p}}^{g\lambda}) \leq C_{\alpha, W}^{g\lambda}$ for all $p \in \mathcal{P}(\mathcal{X})$.

(iii) $\exists! q_{\alpha, W}^{g\lambda} \in \mathcal{P}(\mathcal{Y})$ satisfying (106) such that $q_{\alpha, p}^{g\lambda} = q_{\alpha, W}^{g\lambda}$ for all p with $I_{\alpha}^{g\lambda}(p; W) = C_{\alpha, W}^{g\lambda}$: If $I_{\alpha}^{g\lambda}(p; W) = C_{\alpha, W}^{g\lambda}$ for a $p \in \mathcal{P}(\mathcal{X})$, then as a result of (90), (91), and Lemma 2 we have

$$D_{\alpha}(p \otimes W e^{\frac{1-\alpha}{\alpha} \lambda \cdot \rho} \| p \otimes q_{\alpha, \tilde{p}}^{g\lambda}) \geq C_{\alpha, W}^{g\lambda} + \frac{\alpha \wedge 1}{2} \left\| q_{\alpha, p}^{g\lambda} - q_{\alpha, \tilde{p}}^{g\lambda} \right\|^2. \quad (\text{E.19})$$

Since we have already established that $D_{\alpha}(p \otimes W e^{\frac{1-\alpha}{\alpha} \lambda \cdot \rho} \| p \otimes q_{\alpha, \tilde{p}}^{g\lambda}) \leq C_{\alpha, W}^{g\lambda}$ for any $p \in \mathcal{P}(\mathcal{X})$, (E.19) implies that $q_{\alpha, p}^{g\lambda} = q_{\alpha, \tilde{p}}^{g\lambda}$ for any $p \in \mathcal{P}(\mathcal{X})$ satisfying $I_{\alpha}^{g\lambda}(p; W) = C_{\alpha, W}^{g\lambda}$.

□

Proof of Theorem 3. Note that (102) implies (103) and (104) implies (105). Furthermore, the left hand side of (102) is equal to $C_{\alpha, W}^{g\lambda}$ by (101). Thus when $C_{\alpha, W}^{g\lambda}$ is infinite, (102) holds trivially by the max-min inequality. When $C_{\alpha, W}^{g\lambda}$ is finite, (102) follows from (104) and the max-min inequality. Thus we can assume $C_{\alpha, W}^{g\lambda}$ to be finite and prove the claims about $q_{\alpha, W}^{g\lambda}$, in order to prove the theorem.

- (i) If $C_{\alpha, W}^{g\lambda} < \infty$ and $\lim_{i \rightarrow \infty} I_{\alpha}^{g\lambda}(p^{(i)}; W) = C_{\alpha, W}^{g\lambda}$, then $\{q_{\alpha, p^{(i)}}^{g\lambda}\}_{i \in \mathbb{Z}_+}$ is a Cauchy sequence in $\mathcal{P}(\mathcal{Y})$ for the total variation metric: For any sequence $\{p^{(i)}\}_{i \in \mathbb{Z}_+} \subset \mathcal{P}(\mathcal{X})$ satisfying $\lim_{i \rightarrow \infty} I_{\alpha}^{g\lambda}(p^{(i)}; W) = C_{\alpha, W}^{g\lambda}$, let us consider a sequence of channels $\{W^{(i)}\}_{i \in \mathbb{Z}_+}$ whose input sets $\{\mathcal{X}^{(i)}\}_{i \in \mathbb{Z}_+}$ form a nested sequence of finite subsets of \mathcal{X} defined as follows,

$$\mathcal{X}^{(i)} \triangleq \{x \in \mathcal{X} : \exists j \in \{1, \dots, i\} \text{ such that } p^{(j)}(x) > 0\}.$$

Then for any $i \in \mathbb{Z}_+$, there exists a unique $q_{\alpha, W^{(i)}}^{g\lambda}$ satisfying (106) by Lemma 34. Furthermore, $\mathcal{P}(\mathcal{X}^{(j)}) \subset \mathcal{P}(\mathcal{X}^{(i)})$ for any $i, j \in \mathbb{Z}_+$ such that $j \leq i$. In order to bound $\|q_{\alpha, p^{(j)}}^{g\lambda} - q_{\alpha, p^{(i)}}^{g\lambda}\|$ for positive integers $j < i$, we use the triangle inequality for $q_{\alpha, p^{(j)}}^{g\lambda}$, $q_{\alpha, p^{(i)}}^{g\lambda}$ and $q_{\alpha, W^{(i)}}^{g\lambda}$

$$\|q_{\alpha, p^{(j)}}^{g\lambda} - q_{\alpha, p^{(i)}}^{g\lambda}\| \leq \|q_{\alpha, p^{(j)}}^{g\lambda} - q_{\alpha, W^{(i)}}^{g\lambda}\| + \|q_{\alpha, p^{(i)}}^{g\lambda} - q_{\alpha, W^{(i)}}^{g\lambda}\|. \quad (\text{E.20})$$

Let us proceed with bounding $\|q_{\alpha, p^{(j)}}^{g\lambda} - q_{\alpha, W^{(i)}}^{g\lambda}\|$ and $\|q_{\alpha, p^{(i)}}^{g\lambda} - q_{\alpha, W^{(i)}}^{g\lambda}\|$.

$$\begin{aligned} \|q_{\alpha, p^{(j)}}^{g\lambda} - q_{\alpha, W^{(i)}}^{g\lambda}\| &\stackrel{(a)}{\leq} \sqrt{\frac{2}{\alpha\lambda 1} D_{\alpha}\left(q_{\alpha, p^{(j)}}^{g\lambda} \parallel q_{\alpha, W^{(i)}}^{g\lambda}\right)} \\ &\stackrel{(b)}{\leq} \sqrt{\frac{2}{\alpha\lambda 1}} \sqrt{D_{\alpha}\left(p^{(j)} \otimes W^{(i)} e^{\frac{1-\alpha}{\alpha}\lambda \cdot \rho} \parallel p^{(j)} \otimes q_{\alpha, W^{(i)}}^{g\lambda}\right) - I_{\alpha}^{g\lambda}(p^{(j)}; W^{(i)})} \\ &\stackrel{(c)}{\leq} \sqrt{\frac{2}{\alpha\lambda 1}} \sqrt{C_{\alpha, W^{(i)}}^{g\lambda} - I_{\alpha}^{g\lambda}(p^{(j)}; W^{(i)})} \\ &\stackrel{(d)}{\leq} \sqrt{\frac{2}{\alpha\lambda 1}} \sqrt{C_{\alpha, W}^{g\lambda} - I_{\alpha}^{g\lambda}(p^{(j)}; W)} \end{aligned}$$

where (a) follows from Lemma 2, (b) follows from (90) and (91), (c) follows Lemma 34 because $p^{(j)} \in \mathcal{P}(\mathcal{X}^{(i)})$, and (d) follows from $I_{\alpha}^{g\lambda}(p^{(j)}; W^{(i)}) = I_{\alpha}^{g\lambda}(p^{(j)}; W)$ and $C_{\alpha, W^{(i)}}^{g\lambda} \leq C_{\alpha, W}^{g\lambda}$. We can obtain a similar bound on $\|q_{\alpha, p^{(i)}}^{g\lambda} - q_{\alpha, W^{(i)}}^{g\lambda}\|$. Then $\{q_{\alpha, p^{(i)}}^{g\lambda}\}$ is a Cauchy sequence as a result of (E.20) because $\lim_{i \rightarrow \infty} I_{\alpha}^{g\lambda}(p^{(i)}; W) = C_{\alpha, W}^{g\lambda}$.

- (ii) If $C_{\alpha, W}^{g\lambda} < \infty$, then $\exists! q_{\alpha, W}^{g\lambda} \in \mathcal{P}(\mathcal{Y})$ satisfying $\lim_{i \rightarrow \infty} \|q_{\alpha, W}^{g\lambda} - q_{\alpha, p^{(i)}}^{g\lambda}\| = 0$ for all $\{p^{(i)}\}_{i \in \mathbb{Z}_+} \subset \mathcal{P}(\mathcal{X})$ such that $\lim_{i \rightarrow \infty} I_{\alpha}^{g\lambda}(p^{(i)}; W) = C_{\alpha, W}^{g\lambda}$: Note that $\mathcal{M}(\mathcal{Y})$ is a complete metric space for the total variation metric. Then $\{q_{\alpha, p^{(i)}}^{g\lambda}\}_{i \in \mathbb{Z}_+}$ has a unique limit point $q_{\alpha, W}^{g\lambda}$ in $\mathcal{M}(\mathcal{Y})$. Since $\mathcal{P}(\mathcal{Y})$ is a closed set for the total variation topology and $\cup_{i \in \mathbb{Z}_+} q_{\alpha, p^{(i)}}^{g\lambda} \subset \mathcal{P}(\mathcal{Y})$, then $q_{\alpha, W}^{g\lambda} \in \mathcal{P}(\mathcal{Y})$, by [39, Thm. 2.1.3].

We have established the existence of a unique limit point $q_{\alpha, W}^{g\lambda}$, for any sequence $\{p^{(i)}\}_{i \in \mathbb{Z}_+} \subset \mathcal{P}(\mathcal{X})$ satisfying $\lim_{i \rightarrow \infty} I_{\alpha}^{g\lambda}(p^{(i)}; W) = C_{\alpha, W}^{g\lambda}$. This, however, implies $\lim_{i \rightarrow \infty} \|q_{\alpha, \tilde{p}^{(i)}}^{g\lambda} - q_{\alpha, p^*}^{g\lambda}\| = 0$ for any $\{\tilde{p}^{(i)}\}_{i \in \mathbb{Z}_+}$ satisfying $\lim_{i \rightarrow \infty} I_{\alpha}^{g\lambda}(\tilde{p}^{(i)}; W) = C_{\alpha, W}^{g\lambda}$ because we can interleave the elements of $\{p^{(i)}\}_{i \in \mathbb{Z}_+}$ and $\{\tilde{p}^{(i)}\}_{i \in \mathbb{Z}_+}$ to obtain a new sequence $\{\hat{p}^{(i)}\}_{i \in \mathbb{Z}_+}$ satisfying $\lim_{i \rightarrow \infty} I_{\alpha}^{g\lambda}(\hat{p}^{(i)}; W) = C_{\alpha, W}^{g\lambda}$ for which $\{q_{\alpha, \hat{p}^{(i)}}^{g\lambda}\}$ is a Cauchy sequence. Then $q_{\alpha, W}^{g\lambda} = q_{\alpha, p^*}^{g\lambda}$.

- (iii) $q_{\alpha, W}^{g\lambda}$ satisfies the equality given in (104): For any $p \in \mathcal{P}(\mathcal{X})$, let us consider any sequence $\{p^{(i)}\}_{i \in \mathbb{Z}_+}$ satisfying $p^{(1)} = p$ and $\lim_{i \rightarrow \infty} I_{\alpha}^{g\lambda}(p^{(i)}; W) = C_{\alpha, W}^{g\lambda}$. Then $p \in \mathcal{P}(\mathcal{X}^{(i)})$ for all $i \in \mathbb{Z}_+$. Using Lemma 34 we get

$$D_{\alpha}\left(p \otimes W e^{\frac{1-\alpha}{\alpha}\lambda \cdot \rho} \parallel p \otimes q_{\alpha, W^{(i)}}^{g\lambda}\right) \leq C_{\alpha, W^{(i)}}^{g\lambda} \quad \forall i \in \mathbb{Z}_+. \quad (\text{E.21})$$

Since $\mathcal{X}^{(i)}$ is a finite set, $\exists \tilde{p}^{(i)} \in \mathcal{P}(\mathcal{X}^{(i)})$ satisfying $I_{\alpha}^{g\lambda}(\tilde{p}^{(i)}; W^{(i)}) = C_{\alpha, W^{(i)}}^{g\lambda}$ and $q_{\alpha, \tilde{p}^{(i)}}^{g\lambda} = q_{\alpha, W^{(i)}}^{g\lambda}$ by Lemma 34. Then $I_{\alpha}^{g\lambda}(\tilde{p}^{(i)}; W^{(i)}) \geq I_{\alpha}^{g\lambda}(p; W^{(i)})$ and consequently $\lim_{i \rightarrow \infty} I_{\alpha}^{g\lambda}(\tilde{p}^{(i)}; W) = C_{\alpha, W}^{g\lambda}$. We have already established that for such a sequence $q_{\alpha, \tilde{p}^{(i)}}^{g\lambda} \rightarrow q_{\alpha, W}^{g\lambda}$ in the total variation topology, and hence in the topology of setwise convergence. Then the lower semicontinuity of the Rényi divergence (i.e. Lemma 3) and the identity $C_{\alpha, W^{(i)}}^{g\lambda} \leq C_{\alpha, W}^{g\lambda}$ imply that

$$D_{\alpha}\left(p \otimes W e^{\frac{1-\alpha}{\alpha}\lambda \cdot \rho} \parallel p \otimes q_{\alpha, W}^{g\lambda}\right) \leq C_{\alpha, W}^{g\lambda} \quad \forall p \in \mathcal{P}(\mathcal{X}).$$

On the other hand $D_{\alpha}\left(p \otimes W e^{\frac{1-\alpha}{\alpha}\lambda \cdot \rho} \parallel p \otimes q_{\alpha, W}^{g\lambda}\right) \geq I_{\alpha}^{g\lambda}(p; W)$ and $C_{\alpha, W}^{g\lambda} = \sup_{p \in \mathcal{P}(\mathcal{X})} I_{\alpha}^{g\lambda}(p; W)$ by definitions of $I_{\alpha}^{g\lambda}(p; W)$ and $C_{\alpha, W}^{g\lambda}$. Thus (104) holds.

□

Proof of Lemma 35. Let us first consider the case $\alpha \in \mathbb{R}_+ \setminus \{1\}$. As a result of (90) and (91) we have,

$$\begin{aligned} \sup_{x \in \mathcal{X}} D_\alpha(W(x) \| q) - \lambda \cdot \rho(x) &= \sup_{x \in \mathcal{X}} D_\alpha\left(W(x) e^{\frac{1-\alpha}{\alpha} \lambda \cdot \rho(x)} \middle\| q\right) \\ &\geq D_\alpha\left(p \otimes W e^{\frac{1-\alpha}{\alpha} \lambda \cdot \rho} \middle\| p \otimes q\right) \\ &\geq I_\alpha^\lambda(p; W) + D_\alpha(q_{\alpha, p}^{g^\lambda} \| q) \quad \forall p \in \mathcal{P}(\mathcal{X}). \end{aligned} \quad (\text{E.22})$$

Let $\{p^{(\iota)}\}_{\iota \in \mathbb{Z}_+}$ be a sequence of elements of $\mathcal{P}(\mathcal{X})$ such that $\lim_{\iota \rightarrow \infty} I_\alpha^\lambda(p^{(\iota)}; W) = C_{\alpha, W}^{g^\lambda}$. Then the sequence $\{q_{\alpha, p^{(\iota)}}^{g^\lambda}\}_{\iota \in \mathbb{Z}_+}$ is a Cauchy sequence with the unique limit point $q_{\alpha, W}^{g^\lambda}$ by Theorem 3. Since $\{q_{\alpha, p^{(\iota)}}^{g^\lambda}\} \rightarrow q_{\alpha, W}^{g^\lambda}$ in total variation topology, same convergence holds in the topology of setwise convergence. On the other hand, the order α Rényi divergence is lower semicontinuous for the topology of setwise convergence by Lemma 3. Thus we have

$$\liminf_{\iota \rightarrow \infty} \left[I_\alpha^\lambda(p^{(\iota)}; W) + D_\alpha(q_{\alpha, p^{(\iota)}}^{g^\lambda} \| q) \right] \geq C_{\alpha, W}^{g^\lambda} + D_\alpha(q_{\alpha, W}^{g^\lambda} \| q). \quad (\text{E.23})$$

(E.22) and (E.23) imply (109) for $\alpha \in \mathbb{R}_+ \setminus \{1\}$ because $C_{\alpha, W}^{g^\lambda} = C_{\alpha, W}^\lambda$ by (107) and $q_{\alpha, W}^{g^\lambda} = q_{\alpha, W}^\lambda$ by (108).

For $\alpha = 1$ case, as a result of Lemma 13-(b) and the definition of A-L information given in (71) we have,

$$\sup_{x \in \mathcal{X}} D_1(W(x) \| q) - \lambda \cdot \rho(x) \geq I_1^\lambda(p; W) + D_1(q_{1, p} \| q) \quad \forall p \in \mathcal{P}(\mathcal{X}). \quad (\text{E.24})$$

Repeating the argument leading to (E.23) and invoking Theorem 2, rather than Theorem 3, we get

$$\liminf_{\iota \rightarrow \infty} \left[I_1^\lambda(p^{(\iota)}; W) + D_1(q_{1, p^{(\iota)}} \| q) \right] \geq C_{1, W}^\lambda + D_1(q_{1, W}^\lambda \| q). \quad (\text{E.25})$$

(E.24) and (E.25) imply (109) for $\alpha = 1$ case. □

Proof of Lemma 36. Since $C_{\alpha, W}^\lambda$ is nonincreasing in λ by Lemma 29-(a), $C_{\alpha, W}^{\lambda_2} \leq C_{\alpha, W}^{\lambda_1} \leq C_{\alpha, W}^{\lambda_0} < \infty$. We apply Lemma 35 for $\lambda = \lambda_2$ and $q = q_{\alpha, W}^{\lambda_1}$ and use the fact that $0 \leq \rho(x)$ for all $x \in \mathcal{X}$ to obtain

$$\begin{aligned} D_\alpha(q_{\alpha, W}^{\lambda_2} \| q_{\alpha, W}^{\lambda_1}) + C_{\alpha, W}^{\lambda_2} &\leq \sup_{x \in \mathcal{X}} D_\alpha(W(x) \| q_{\alpha, W}^{\lambda_1}) - \lambda_2 \cdot \rho(x) \\ &\leq \sup_{x \in \mathcal{X}} D_\alpha(W(x) \| q_{\alpha, W}^{\lambda_1}) - \lambda_1 \cdot \rho(x). \end{aligned}$$

Then (110) follows from (82) of Theorem 2.

For any two point λ_1 and λ_2 in $\{\lambda : \exists \epsilon > 0 \text{ s.t. } C_{\alpha, W}^{\lambda - \epsilon \mathbb{1}} < \infty\}$, not necessarily satisfying $\lambda_1 \leq \lambda_2$, let λ_\vee be $\lambda_1 \vee \lambda_2$, i.e. $\lambda_\vee^\iota = \lambda_1^\iota \vee \lambda_2^\iota$ for all $\iota \in \{1, \dots, \ell\}$. Then as a result of the triangle inequality we have

$$\left\| q_{\alpha, W}^{\lambda_1} - q_{\alpha, W}^{\lambda_2} \right\| \leq \left\| q_{\alpha, W}^{\lambda_1} - q_{\alpha, W}^{\lambda_\vee} \right\| + \left\| q_{\alpha, W}^{\lambda_\vee} - q_{\alpha, W}^{\lambda_2} \right\|. \quad (\text{E.26})$$

On the other hand, as a result of Lemma 2 and (110) we have,

$$\left\| q_{\alpha, W}^{\lambda_1} - q_{\alpha, W}^{\lambda_\vee} \right\| = \sqrt{\frac{2}{\alpha \wedge 1}} \sqrt{C_{\alpha, W}^{\lambda_1} - C_{\alpha, W}^{\lambda_\vee}}, \quad (\text{E.27})$$

$$\left\| q_{\alpha, W}^{\lambda_\vee} - q_{\alpha, W}^{\lambda_2} \right\| = \sqrt{\frac{2}{\alpha \wedge 1}} \sqrt{C_{\alpha, W}^{\lambda_2} - C_{\alpha, W}^{\lambda_\vee}}. \quad (\text{E.28})$$

Then continuity of $q_{\alpha, W}^\lambda$ in λ on $\{\lambda : \exists \epsilon > 0 \text{ s.t. } C_{\alpha, W}^{\lambda - \epsilon \mathbb{1}} < \infty\}$ for the total variation topology on $\mathcal{P}(\mathcal{Y})$ follows from (E.26), (E.27), (E.28), and the continuity of $C_{\alpha, W}^\lambda$ in λ on $\{\lambda : \exists \epsilon > 0 \text{ s.t. } C_{\alpha, W}^{\lambda - \epsilon \mathbb{1}} < \infty\}$ established in Lemma 29-(a). □

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 [22] 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.

REFERENCES

- [1] B. Nakiboğlu. The Augustin center and the sphere packing bound for memoryless channels. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 1401–1405, Aachen, Germany, June 2017.
- [2] I. Csiszár. Generalized cutoff rates and Rényi’s information measures. *IEEE Transactions on Information Theory*, 41(1):26–34, Jan 1995.
- [3] M. Dalai. Some remarks on classical and classical-quantum sphere packing bounds: Rényi vs. Kullback-Leibler. *Entropy*, 19(7):355, 2017.
- [4] Mosonyi and Ogawa. Divergence radii and the strong converse exponent of classical-quantum channel coding with constant compositions. *arXiv:1811.10599v4 [cs.IT]*, 2018.
- [5] Imre Csiszár and János Körner. *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, Cambridge, UK, 2011.
- [6] Udo Augustin. *Noisy Channels*. Habilitation thesis, Universität Erlangen-Nürnberg, 1978. (<http://bit.ly/2ID8h7m>).
- [7] B. Nakiboğlu. The sphere packing bound for memoryless channels. *arXiv:1804.06372 [cs.IT]*, 2018.
- [8] T. van Erven and P. Harremoës. Rényi divergence and Kullback-Leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, July 2014.
- [9] O. Shayevitz. A note on a characterization of Rényi measures and its relation to composite hypothesis testing. *arXiv:1012.4401v2 [cs.IT]*, 2010.
- [10] O. Shayevitz. On Rényi measures and hypothesis testing. In *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pages 894–898, Austin, Texas, July 2011.
- [11] S. Verdú. α -mutual information. In *2015 Information Theory and Applications Workshop*, 2015.
- [12] J. H. B. Kemperman. On the Shannon capacity of an arbitrary channel. *Indagationes Mathematicae (Proceedings)*, 77(2):101–115, 1974.
- [13] B. Nakiboğlu. The Rényi capacity and center. *IEEE Transactions on Information Theory*, 65(2):841–860, Feb 2019. ([arXiv:1608.02424 \[cs.IT\]](https://arxiv.org/abs/1608.02424)).
- [14] R. G. Gallager. A simple derivation of the coding theorem and some applications. *IEEE Transactions on Information Theory*, 11(1):3–18, Jan. 1965.
- [15] Robert G. Gallager. *Information theory and reliable communication*. John Wiley & Sons, Inc., New York, NY, 1968.
- [16] Paul Michael Ebert. *Error Bounds For Parallel Communication Channels*. Technical report 448, Research Laboratory of Electronics at Massachusetts Institute of Technology, Cambridge, MA, 1966. (<http://hdl.handle.net/1721.1/4295>).
- [17] John Stephen Richters. *Communication over fading dispersive channels*. Technical report 464, Research Laboratory of Electronics at Massachusetts Institute of Technology, Cambridge, MA, 1967. (<http://hdl.handle.net/1721.1/4279>).
- [18] E. A. Haroutunian. Estimates of the error probability exponent for a semicontinuous memoryless channel. *Problems of Information Transmission*, 4(4):37–48, 1968.
- [19] G. Sh. Poltyrev. Random coding bounds for discrete memoryless channels. *Problems of Information Transmission*, 18(1):9–21, 1982.
- [20] Richard M. Dudley. *Real analysis and probability*, volume 74. Cambridge University Press, New York, NY, 2002.
- [21] Vladimir I. Bogachev. *Measure Theory*. Springer-Verlag, Berlin Heidelberg, 2007.
- [22] Roberto M. Fano. *Transmission of Information: A Statistical Theory of Communications*. M.I.T. Press, Cambridge, MA, 1961.
- [23] S. Arimoto. Computation of random coding exponent functions. *IEEE Transactions on Information Theory*, 22(6):665–671, Nov 1976.
- [24] Y. Oohama. The optimal exponent function for the additive white gaussian noise channel at rates above the capacity. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 1053–1057, Aachen, Germany, June 2017.
- [25] Y. Oohama. Exponent function for stationary memoryless channels with input cost at rates above the capacity. *arXiv:1701.06545v3 [cs.IT]*, 2017.
- [26] G. Vazquez-Vilar, A. Martinez, and A. G. i Fabregas. A derivation of the cost-constrained sphere-packing exponent. In *2015 IEEE International Symposium on Information Theory*, pages 929–933, Hong Kong, P.R.C., June 2015.
- [27] A. Rényi. On measures of entropy and information. In *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, pages 547–561, 1961.
- [28] I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*, 2(3-4):299–318, 1967.
- [29] G. L. Gilardoni. On Pinsker’s and Vajda’s type inequalities for Csiszár’s f -divergences. *IEEE Transactions on Information Theory*, 56(11):5377–5386, Nov 2010.
- [30] Albert N. Shiryaev. *Probability*. Springer-Verlag, New York, NY, 1995.
- [31] Y. Polyanskiy and S. Verdú. Arimoto channel coding converse and Rényi divergence. In *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, pages 1327–1333, Oct 2010.
- [32] A. N. Kolmogorov and S. V. Fomin. *Introductory Real Analysis*. Dover Books on Mathematics. Dover Publications, New York, NY, 1975.
- [33] I. Csiszár. A class of measures of informativity of observation channels. *Periodica Mathematica Hungarica*, 2(1-4):191–213, 1972.
- [34] R. Sibson. Information radius. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 14(2):149–160, 1969.
- [35] R. E. Blahut. Hypothesis testing and information theory. *IEEE Transactions on Information Theory*, 20(4):405–417, Jul 1974.
- [36] V. Kostina and S. Verdú. Channels with cost constraints: Strong converse and dispersion. *IEEE Transactions on Information Theory*, 61(5):2415–2429, May 2015.
- [37] B. Nakiboğlu. The sphere packing bound via Augustin’s method. *IEEE Transactions on Information Theory*, 65(2):816–840, Feb 2019. ([arXiv:1611.06924 \[cs.IT\]](https://arxiv.org/abs/1611.06924)).
- [38] S.G. Krantz and H.R. Parks. *A Primer of Real Analytic Functions*. Birkhäuser, Boston, 2002.
- [39] James R. Munkres. *Topology*. Prentice Hall Inc., Upper Saddle River, NJ 07458, 2000.
- [40] Walter Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, New York, NY, 1976.
- [41] Dimitri P. Bertsekas, Angelia Nedić, and Asuman E. Ozdaglar. *Convex Analysis and Optimization*. Athena Scientific, Belmont, Massachusetts, 2003.
- [42] H. Komiya. Elementary proof for Sion’s minimax theorem. *Kodai Mathematical Journal*, 11(1):5–7, 1988.
- [43] M. Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958.