The Augustin Center and The Sphere Packing Bound For Memoryless Channels

Barış Nakiboğlu
nakib@alum.mit.edu

***Abstract*—For any channel with a convex constraint set and finite Augustin capacity, existence of a unique Augustin center and associated Erven-Harremoes bound are established. Augustin-Legendre capacity, center, and radius are introduced and proved to be equal to the corresponding Renyi-Gallager entities. Sphere packing bounds with polynomial prefactors are derived for codes on two families of channels: (possibly non-stationary) memoryless channels with multiple additive cost constraints and stationary memoryless channels with convex constraints on the empirical distribution of the input codewords.**

I. Introduction

Augustin [2], [3] derived the sphere packing bound for the product channels without assuming the stationarity. Assuming that order ½ Renyi capacity of the component channels are $O(\ln n)$, we have derived the sphere packing bound for product channels with a prefactor that is polynomial in the block length n, [12, Theorem 2]. In this manuscript, we derive analogous results for two families of memoryless channels. As we have done for the product channels in [12], we first derive a non-asymptotic outer bound for codes on a given memoryless channel, then we derive our asymptotic result using this bound.

In [3, Chapter VII], Augustin pursued an analysis similar to ours and derived the sphere packing bound for memoryless channels with cost constraints [3, §36]. In addition, Augustin established the connection between the exponent of Gallager's inner bound for the cost constrained channels [8, Thm 8] and the sphere packing exponent [3, §35]. Our results surpass Augustin's results in two ways:

- Augustin assumes the cost function to be bounded.[1] This hypothesis excludes certain important and interesting cases such as the Gaussian channels. Hence, Augustin's analysis in [3] does not imply the sphere packing bounds derived by Shannon [15] and Ebert [6]. We don't assume the cost function to be bounded. Thus, Theorem 1 establishes the sphere packing bound for a wider class of channels including the Gaussian channels with multiple antennas.[2] It is even possible to handle certain fading scenarios and additional per antenna power constraints.
- The best asymptotic bound implied by Augustin's non-asymptotic bound [3, Thm 36.6] is of the form $P_{\mathbf{e}}^{av(n)} \geq O(\frac{1}{e^{\sqrt{n}}})e^{-E_{sp}\left(\ln \frac{M_n}{L_n} - O(\sqrt{n}), W_{[1,n]}, \varrho_n\right)}$. In Theorem 1 we replace $O(\frac{1}{e^{\sqrt{n}}})$ by $O(\frac{1}{n^{\tau}})$ by $O(\sqrt{n})$ to 0.

For stationary memoryless channels with finite input sets, the sphere packing bound is well-known [4, Ch. 10], [5]. For such a channel, one first chooses the most populous constant composition sub-code and then derives the sphere packing bound for the code using the sphere packing bound for the constant composition sub-code.[3] This technique, however, fails when the input set of the channel is infinite. We show that a sphere packing bound similar to Theorem 1 holds for codes on stationary memoryless channels with convex constraints on the empirical distribution of the input codewords.

In the rest of this section, we describe our model and notation and state our main asymptotic result. In Section II, we introduce and analyze Augustin information, mean, capacity, and center as purely measure theoretic concepts. The role of these concepts in our analysis is analogous to the role of corresponding Renyi concepts in [11], [12]. In Section III, we investigate the cost constrained Augustin capacity more closely and introduce the concepts of Augustin-Legendre information and Renyi-Gallager information, together with the associated means, capacities, centers, and radii. Our main aim in Section III is to express the cost constrained Augustin capacity and center in terms of Augustin-Legendre capacity and center. In Section IV, we derive non-asymptotic outer bounds for codes on two families memoryless channels.

A. Model and Notation

For any set $\mathcal{X}$, $\mathcal{P}(\mathcal{X})$ is the set of all probability mass functions that are non-zero only on finitely many members of $\mathcal{X}$; $\mathcal{M}^{+}(\mathcal{X})$ is the set of all non-zero mass functions with the same property. For any measurable space $(\mathcal{Y},\mathscr{Y})$, $\mathcal{P}(\mathscr{Y})$ is the set of all probability measures and $\mathcal{M}^{+}(\mathscr{Y})$ is set of all finite measures. For any $\mu, q \in \mathcal{M}^{+}(\mathscr{Y})$, $\mu \leq q$ iff $\mu(\mathcal{E}) \leq q(\mathcal{E})$ $\forall \mathcal{E} \in \mathscr{Y}$. Similarly, for any $\mu, q \in \Re^{\ell}$, $\mu \leq q$ iff $\mu^{\imath} \leq q^{\imath}$ $\forall \imath \in \{1,\ldots,\ell\}$. For any $\mu, q \in \Re^{\ell}$, $\mu \cdot q \triangleq \sum_{\jmath=1}^{\ell} \mu^{\jmath} q^{\jmath}$. For any $\ell \in \mathbb{Z}_{+}$, $\mathbb{1} \in \Re^{\ell}$ is the vector whose all entries are one. For any $\mathcal{S} \subset \Re^{\ell}$ we denote the interior of $\mathcal{S}$ by $\mathtt{int}\mathcal{S}$. For any set $\mathcal{S}$ in a vector space we denote the convex hull of $\mathcal{S}$ by $\mathtt{ch}\mathcal{S}$.

A *channel* W is a function from *the input set* $\mathcal{X}$ to the set of all probability measures on *the output space* $(\mathcal{Y},\mathscr{Y})$. A channel $W : \mathcal{X} \to \mathcal{P}(\mathscr{Y})$ is a *product channel* for a finite index set $\mathcal{T}$ iff there exist channels $W_t : \mathcal{X}_t \to \mathcal{P}(\mathscr{Y}_t)$ for all $t \in \mathcal{T}$ satisfying $W(x) = \prod_{t\in\mathcal{T}}^{\otimes} W_t(x_t)$ for all $x \in \mathcal{X}$ where

$$\mathcal{X} = \prod\nolimits_{t\in\mathcal{T}}^{\otimes} \mathcal{X}_t \qquad \mathcal{Y} = \prod\nolimits_{t\in\mathcal{T}}^{\times} \mathcal{Y}_t \qquad \mathscr{Y} = \prod\nolimits_{t\in\mathcal{T}}^{\otimes} \mathscr{Y}_t.$$

A product channel is *stationary* iff all W_t's are identical. If $\mathcal{X} \subset \prod_{t\in\mathcal{T}}^{\otimes} \mathcal{X}_t$ then W is a *memoryless channel*.

[1]The issue here is not a matter of rescaling: certain conclusions of Augustin's analysis are not correct when cost functions are not bounded.

[2]Shannon's approximation error terms in [15] are considerably better than ours. But his derivation relies heavily on the geometry of the output space. Our derivation, on the other hand, is oblivious towards it.

[3]Haroutunian [9] was the first one to give a complete proof of the sphere packing bound for constant composition codes. Recently, Altug and Wagner [1] sharpened the prefactor of the bound for channels with finite output sets.

An (M,L) *channel code* on $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$ is an ordered pair (Ψ,Θ) composed of an *encoding function* $\Psi:\mathcal{M}\to\mathcal{X}$ and a *decoding function*[4] $\Theta:\mathcal{Y}\to\widehat{\mathcal{M}}$ where $\mathcal{M}\triangleq\{1,2,\ldots,M\}$, $\widehat{\mathcal{M}}\triangleq\{\mathcal{L}:\mathcal{L}\subset\mathcal{M}$ and $|\mathcal{L}|=L\}$, and Θ is a measurable as a function from the measurable space $(\mathcal{Y},\mathcal{Y})$.

Given an (M,L) channel code (Ψ,Θ) on $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$ *the average error probability* $P_{\mathbf{e}}^{av}$ and *the conditional error probability* $P_{\mathbf{e}}^{m}$ for $m\in\mathcal{M}$ are given by

$$P_{\mathbf{e}}^{av}\triangleq\tfrac{1}{M}\sum\nolimits_{m\in\mathcal{M}}P_{\mathbf{e}}^{m}\qquad\qquad P_{\mathbf{e}}^{m}\triangleq W(\Psi(m))(\{m\notin\widehat{m}\}).$$

A *cost function* ρ is a function from the input set to $\Re_{\geq0}^{\ell}$ for some $\ell\in\mathbb{Z}_{+}$. We assume without loss of generality that[5]

$$\inf\nolimits_{x\in\mathcal{X}}\rho^{\imath}(x)=0\qquad\qquad\forall\imath\in\{1,\ldots,\ell\}.$$

Let Γ_{ρ} be the set of feasible cost constraints for $\mathcal{P}(\mathcal{X})$:

$$\Gamma_{\rho}\triangleq\{\varrho\in\Re_{\geq0}^{\ell}:\exists p\in\mathcal{P}(\mathcal{X})\text{ s.t. }\sum\nolimits_{x}p(x)\rho(x)\leq\varrho\}.$$

Then Γ_{ρ} is a convex set with non-empty interior. A cost function ρ for a product channel W is said to be *additive* iff there exists a $\rho_t:\mathcal{X}_t\to\Re_{\geq0}^{\ell}$ for each $t\in\mathcal{T}$ such that

$$\rho(x)=\sum\nolimits_{t\in\mathcal{T}}\rho_t(x_t)\qquad\qquad\forall x\in\mathcal{X}.$$

An encoding function Ψ, hence the corresponding code, is said to satisfy the cost constraint ϱ iff $\vee_{m\in\mathcal{M}}\rho(\Psi(m))\leq\varrho$. A code on a product channel $W:\prod_{t\in\mathcal{T}}^{\otimes}\mathcal{X}_t\to\mathcal{P}(\mathcal{Y})$ is said to satisfy an empirical distribution constraint $\mathcal{A}\subset\mathcal{P}(\mathcal{X}_1)$ iff the empirical distribution, i.e. type or composition, of $\Psi(m)$ is in $\mathcal{A}$ for all $m\in\mathcal{M}$.

B. Main Result

Assumption 1. $\{(W_t,\rho_t,\varrho_t)\}_{t\in\mathbb{Z}_+}$ is an ordered sequence of channels with associated cost functions and cost constraints satisfying the following condition: $\exists n_0\in\mathbb{Z}_+, K\in\Re_+$ s.t.

$$\max\nolimits_{t:t\leq n}C_{\frac{1}{2},W_t,\varrho_n}\leq K\ln(n)\quad\text{ and }\quad\varrho_n\in\mathtt{int}\Gamma_{\rho_{[1,n]}}$$

for all $\forall n\geq n_0$ where $\rho_{[1,n]}(x_{[1,n]})=\sum_{t=1}^{n}\rho_t(x_t)$.

Theorem 1. *Let $\{(W_t,\rho_t,\varrho_t)\}_{t\in\mathbb{Z}_+}$ be a sequence satisfying Assumption 1, α_0,α_1 be orders satisfying $0<\alpha_0<\alpha_1<1$ and $\varepsilon\in\Re_{\geq0}$. Then for any sequence of codes $\{(\Psi_t,\Theta_t)\}_{t\in\mathbb{Z}_+}$ on the product channels $\{W_{[1,n]}\}_{n\in\mathbb{Z}_+}$ satisfying*

$$\vee_{m\in\mathcal{M}_n}\rho_{[1,n]}(\Psi_t(m))\leq\varrho_n\qquad\qquad\forall n\in\mathbb{Z}_+$$
$$C_{\alpha_0,W_{[1,n]},\varrho_n}+\varepsilon\ln^2 n\leq\ln\tfrac{M_n}{L_n}\leq C_{\alpha_1,W_{[1,n]},\varrho_n}\qquad\forall n\geq n_0$$

there exists a $\tau\in\Re_+$ and an $n_1\geq n_0$ such that

$$P_{\mathbf{e}}^{av(n)}\geq n^{-\tau}e^{-E_{sp}\left(\ln\frac{M_n}{L_n},W_{[1,n]},\varrho_n\right)}\qquad\qquad\forall n\geq n_1$$

where $E_{sp}(R,W,\varrho)=\sup_{\alpha\in(0,1)}\frac{1-\alpha}{\alpha}\left(C_{\alpha,W,\varrho}-R\right)$.

Theorem 1 follows from Lemma 12 and Lemma 13, through an analysis similar to the one in [12, §III-E]. An asymptotic result similar to Theorem 1 for codes on stationary memoryless channels with convex empirical distribution constraints can be proved using Lemma 12 and the bound given in equation (10).

[4]Recall that for any encoder Ψ a deterministic MAP decoder obtains minimum $P_{\mathbf{e}}^{av}$ among all, possibly non-deterministic, decoders.

[5]Augustin [3, §33] has the following additional hypothesis: $\vee_{x\in\mathcal{X}}\rho(x)\leq\mathbb{1}$.

II. The Augustin Information and Capacity

$\forall\alpha\in\Re_+$, $w,q\in\mathcal{M}^{+}(\mathcal{Y})$, *the order α Renyi divergence* is

$$D_{\alpha}(w\|\,q)\triangleq\begin{cases}\frac{1}{\alpha-1}\ln\int(\frac{\mathrm{d}w}{\mathrm{d}\nu})^{\alpha}(\frac{\mathrm{d}q}{\mathrm{d}\nu})^{1-\alpha}\nu(\mathrm{d}y) & \alpha\neq1\\ \int\frac{\mathrm{d}w}{\mathrm{d}\nu}\left[\ln\frac{\mathrm{d}w}{\mathrm{d}\nu}-\ln\frac{\mathrm{d}q}{\mathrm{d}\nu}\right]\nu(\mathrm{d}y) & \alpha=1\end{cases}$$

where ν is any measure s.t. $w\prec\nu, q\prec\nu$. If $D_{\alpha}(w\|q)<\infty$ then *the order α tilted probability measure* $v_{\alpha}^{w,q}$ is

$$\tfrac{\mathrm{d}v_{\alpha}^{w,q}}{\mathrm{d}\nu}\triangleq e^{(1-\alpha)D_{\alpha}(w\|\,q)}(\tfrac{\mathrm{d}w}{\mathrm{d}\nu})^{\alpha}(\tfrac{\mathrm{d}q}{\mathrm{d}\nu})^{1-\alpha}.$$

A. The Augustin Information and Mean

Definition 1. For any $\alpha\in\Re_+$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, and $p\in\mathcal{P}(\mathcal{X})$ *the order α Augustin information for the prior p* is

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

where $D_{\alpha}(W\|\,q|\,p)\triangleq\sum_{x\in\mathcal{X}}p(x)D_{\alpha}(W(x)\|\,q)$.

Whenever it exists, the uniqueness of $q_{\alpha,p}\in\mathcal{P}(\mathcal{Y})$ satisfying $I_{\alpha}(p;W)=D_{\alpha}(W\|\,q_{\alpha,p}|\,p)$ follows from the strict convexity of $D_{\alpha}(w\|\,q)$ in q, i.e. [7, Thm 12]. Such a $q_{\alpha,p}$ is called *the order α Augustin mean for the prior p*. If $|\mathcal{Y}|<\infty$ then $\mathcal{P}(\mathcal{Y})$ is compact and the existence of $q_{\alpha,p}$ follows from the lower semicontinuity of $D_{\alpha}(w\|\,q)$ in q, i.e [11, Lem 7], and the extreme value theorem [10, Ch3§12.2].

Lemma 1 asserts the existence of a unique $q_{\alpha,p}$ for arbitrary channels and describes $q_{\alpha,p}$ via the identities it has to satisfy. Part (a) is well known; part (b) is due to[6] Augustin [3, 34.2]. A generalization of Lemma 1 for all $\alpha\in\Re_+$ is proved in [13].

Definition 2. For any $\alpha\in\Re_+$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, and $p\in\mathcal{P}(\mathcal{X})$,

- $\mathtt{T}_{\alpha,p}(\cdot):\{q\in\mathcal{M}^{+}(\mathcal{Y}):D_{\alpha}(W\|\,q|\,p)<\infty\}\to\mathcal{P}(\mathcal{Y})$ is
$$\mathtt{T}_{\alpha,p}(q)\triangleq\sum\nolimits_{x}p(x)v_{\alpha}^{W(x),q}.$$
Furthermore, $\mathtt{T}_{\alpha,p}^{\imath+1}(q)\triangleq\mathtt{T}_{\alpha,p}(\mathtt{T}_{\alpha,p}^{\imath}(q))$ for $\imath\in\mathbb{Z}_+$.
- $\mu_{\alpha,p}\in\mathcal{M}^{+}(\mathcal{Y})$ and $q_{\alpha,p}^{g}\in\mathcal{P}(\mathcal{Y})$ are given by
$$\tfrac{\mathrm{d}\mu_{\alpha,p}}{\mathrm{d}\nu}\triangleq\left[\sum\nolimits_{x}p(x)\left(\tfrac{\mathrm{d}W(x)}{\mathrm{d}\nu}\right)^{\alpha}\right]^{\frac{1}{\alpha}}\qquad\qquad q_{\alpha,p}^{g}\triangleq\tfrac{\mu_{\alpha,p}}{\|\mu_{\alpha,p}\|}$$
where ν is any measure for which $(\sum_{x}p(x)W(x))\prec\nu$.

Lemma 1. *For any $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$ and $p\in\mathcal{P}(\mathcal{X})$,*

(a) $I_1(p;W)=D_1(W\|\,q_{1,p}|\,p)$ *for* $q_{1,p}\triangleq\sum_{x}p(x)W(x)$.

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

(b) $\forall\alpha\in(0,1)\exists!q_{\alpha,p}$ *s.t.* $I_{\alpha}(p;W)=D_{\alpha}(W\|q_{\alpha,p}|\,p)$. $q_{\alpha,p}\sim q_{1,p}$,

$$D_{\alpha}(W\|\,q|\,p)-I_{\alpha}(p;W)\geq D_{\alpha}(q_{\alpha,p}\|\,q)\qquad\forall q\in\mathcal{P}(\mathcal{Y})\quad(2)$$
$$\mathtt{T}_{\alpha,p}(q_{\alpha,p})=q_{\alpha,p}\quad(3)$$
$$\lim_{\jmath\to\infty}\left\|q_{\alpha,p}-\mathtt{T}_{\alpha,p}^{\jmath}(q_{\alpha,p}^{g})\right\|=0.\quad(4)$$

Furthermore, if a $q\in\mathcal{P}(\mathcal{Y})$ satisfying $q_{1,p}\prec q$ is a fixed point of $\mathtt{T}_{\alpha,p}(\cdot)$ then $q=q_{\alpha,p}$.

(c) If $\alpha\in(0,1]$, W is a product channel for a finite index set $\mathcal{T}$, and p is of the form $\prod_{t\in\mathcal{T}}^{\otimes}p_t$ for $p_t\in\mathcal{P}(\mathcal{X}_t)$ then

$$q_{\alpha,p}=\prod\nolimits_{t\in\mathcal{T}}^{\otimes}q_{\alpha,p_t}\qquad I_{\alpha}(p;W)=\sum\nolimits_{t\in\mathcal{T}}I_{\alpha}(p_t;W_t).\quad(5)$$

[6][3, 34.2] claims eq. (4) for $q_{1,p}^{g}$ instead of $q_{\alpha,p}^{g}$. We could not confirm the correctness of Augustin's proof of [3, 34.2], see [13].

B. The Constrained Augustin Capacity and Center

Definition 3. For any $\alpha\in\Re_+$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, and $\mathcal{A}\subset\mathcal{P}(\mathcal{X})$, *the order α Augustin capacity of W for constraint set $\mathcal{A}$* is

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

Using the definition of $I_\alpha(p;W)$ we get

$$C_{\alpha,W,\mathcal{A}}=\sup\nolimits_{p\in\mathcal{A}}\inf\nolimits_{q\in\mathcal{P}(\mathcal{Y})} D_\alpha(W\|\,q|\,p)\,.$$

Proofs of the propositions presented in this subsection can be found in [13]. They are very similar to the proofs of the corresponding claims in [11, §III, §IV, §F] for Renyi capacity; we invoke Lemma 1 instead of [11, Lem 10].

Lemma 2. *. For any $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$ and $\mathcal{A}\subset\mathcal{P}(\mathcal{X})$*

(a) $C_{\alpha,W,\mathcal{A}}:(0,1]\to[0,\infty]$ is increasing and continuous

(b) $\frac{1-\alpha}{\alpha}C_{\alpha,W,\mathcal{A}}:(0,1)\to[0,\infty]$ is decreasing and continuous

(c) $\exists\alpha\in(0,1)$ s.t. $C_{\alpha,W,\mathcal{A}}<\infty$ iff $C_{\phi,W,\mathcal{A}}<\infty$ $\forall\phi\in(0,1)$.

Theorem 2. *$\forall\alpha\in(0,1]$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, and convex $\mathcal{A}\subset\mathcal{P}(\mathcal{X})$,*

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

If $C_{\alpha,W,\mathcal{A}}<\infty$ then $\exists!q_{\alpha,W,\mathcal{A}}\in\mathcal{P}(\mathcal{Y})$, called the order α Augustin center of W for the constraint set $\mathcal{A}$, such that

$$C_{\alpha,W,\mathcal{A}}=\sup\nolimits_{p\in\mathcal{A}} D_\alpha(W\|\,q_{\alpha,W,\mathcal{A}}|\,p)\,.$$

If $\lim_{\imath\to\infty}I_\alpha\big(p^{(\imath)};W\big)=C_{\alpha,W,\mathcal{A}}<\infty$ for a $\{p^{(\imath)}\}_{\imath\in\mathbb{Z}_+}\subset\mathcal{A}$ then $\{q_{\alpha,p^{(\imath)}}\}_{\imath\in\mathbb{Z}_+}$ is a Cauchy sequence for the total variation metric on $\mathcal{P}(\mathcal{Y})$ and $q_{\alpha,W,\mathcal{A}}$ is its unique limit point.

Lemma 1 and Theorem 2 imply for all $\alpha\in(0,1]$, $p\in\mathcal{A}$ that

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

Using Lemma 1 and Theorem 2 we can prove the following Erven-Harremoes bound for Augustin capacity.

Lemma 3. *For any $\alpha\in(0,1]$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, and convex $\mathcal{A}\subset\mathcal{P}(\mathcal{X})$ s.t. $C_{\alpha,W,\mathcal{A}}<\infty$, and $q\in\mathcal{P}(\mathcal{Y})$*

$$\sup\nolimits_{p\in\mathcal{A}} D_\alpha(W\|\,q|\,p)\geq C_{\alpha,W,\mathcal{A}}+D_\alpha(q_{\alpha,W,\mathcal{A}}\|\,q)\,.$$

Erven-Harremoes bound, the continuity of $C_{\alpha,W,\mathcal{A}}$ in α, and Pinsker's inequality imply the continuity of $q_{\alpha,W,\mathcal{A}}$ in α for the total variation topology on $\mathcal{P}(\mathcal{Y})$.

Lemma 4. *For any $\eta\in(0,1]$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, convex $\mathcal{A}\subset\mathcal{P}(\mathcal{X})$ s.t. $C_{\eta,W,\mathcal{A}}<\infty$, and α, ϕ satisfying $0<\alpha<\phi\leq\eta$,*

$$C_{\phi,W,\mathcal{A}}-C_{\alpha,W,\mathcal{A}}\geq D_\alpha(q_{\alpha,W,\mathcal{A}}\|\,q_{\phi,W,\mathcal{A}})\,.$$

Furthermore, $q_{\alpha,W,\mathcal{A}}:(0,\eta]\to\mathcal{P}(\mathcal{Y})$ is continuous in α for the total variation topology on $\mathcal{P}(\mathcal{Y})$.

Lemma 5. *For any $\alpha\in(0,1]$, product channel W for a finite index set $\mathcal{T}$, convex sets $\mathcal{A}_t\subset\mathcal{P}(\mathcal{X}_t)$ for each $t\in\mathcal{T}$, and $\mathcal{A}=\mathtt{ch}\{\prod_{t\in\mathcal{T}}^{\otimes}p_t:p_t\in\mathcal{P}(\mathcal{X}_t)\ \forall t\in\mathcal{T}\}$*

$$C_{\alpha,W,\mathcal{A}}=\sum\nolimits_{t\in\mathcal{T}} C_{\alpha,W_t,\mathcal{A}_t}.$$

Furthermore, if $C_{\alpha,W,\mathcal{A}}<\infty$ then $q_{\alpha,W,\mathcal{A}}=\prod_{t\in\mathcal{T}}^{\otimes}q_{\alpha,W_t,\mathcal{A}_t}$.

III. The Cost Constrained Augustin Capacity

With a slight abuse of notation we define the cost constrained Augustin capacity as

$$C_{\alpha,W,\varrho}\triangleq\sup\nolimits_{p\in\mathcal{A}(\varrho)} I_\alpha(p;W)\qquad\forall\varrho\in\Gamma_\rho$$

where $\mathcal{A}(\varrho)\triangleq\{p\in\mathcal{P}(\mathcal{X}):\sum_x p(x)\rho(x)\leq\varrho\}$. Note that Theorem 2 and Lemmas 3 and 4 hold for $C_{\alpha,W,\varrho}$ because $\mathcal{A}(\varrho)$ is a convex set. We denote Augustin center by $q_{\alpha,W,\varrho}$.

Lemma 6. *For any $\alpha\in(0,1]$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, $\rho:\mathcal{X}\to\Re^{\ell}_{\geq0}$,*

(a) $C_{\alpha,W,\varrho}:\Gamma_\rho\to[0,\infty]$ is increasing and concave in ϱ. It is either infinite $\forall\varrho\in\mathtt{int}\Gamma_\rho$ or finite and continuous on $\mathtt{int}\Gamma_\rho$.

(b) If $C_{\alpha,W,\varrho}<\infty$ for a $\varrho\in\mathtt{int}\Gamma_\rho$ then $\exists\lambda_{\alpha,W,\varrho}\in\Re^{\ell}_{\geq0}$ s.t.

$$C_{\alpha,W,\varrho}+\lambda_{\alpha,W,\varrho}\cdot(\tilde{\varrho}-\varrho)\geq C_{\alpha,W,\tilde{\varrho}}\qquad\forall\tilde{\varrho}\in\Gamma_\rho.$$

The set of all such $\lambda_{\alpha,W,\varrho}$'s for an α is convex and compact.

Lemma 7. *For any $\alpha\in(0,1]$, product channel W for a finite index set $\mathcal{T}$, additive cost function $\rho:\mathcal{X}\to\Re^{\ell}_{\geq0}$ satisfying $\rho(x)=\sum_{t\in\mathcal{T}}\rho_t(x_t)$ for some $\rho_t:\mathcal{X}_t\to\Re^{\ell}_{\geq0}$ and $\varrho\in\Gamma_\rho$*

$$C_{\alpha,W,\varrho}=\sup\left\{\sum\nolimits_{t\in\mathcal{T}} C_{\alpha,W_t,\varrho_t}:\sum\nolimits_{t\in\mathcal{T}}\varrho_t\leq\varrho,\ \varrho_t\in\Gamma_{\rho_t}\right\}$$

If $\exists\{\varrho_t\}_{t\in\mathcal{T}}$ s.t. $C_{\alpha,W,\varrho}=\sum_{t\in\mathcal{T}}C_{\alpha,W_t,\varrho_t}$ and $C_{\alpha,W,\varrho}<\infty$ then $q_{\alpha,W,\varrho}=\prod_{t\in\mathcal{T}}^{\otimes}q_{\alpha,W_t,\varrho_t}$.

Since Augustin capacity is concave in the cost constraint by Lemma 6-(a), $C_{\alpha,W,\varrho}=\sum_{t\in\mathcal{T}}C_{\alpha,W_t,\frac{\varrho}{|\mathcal{T}|}}$ whenever W is stationary and $\rho_t=\rho_1$ for all $t\in\mathcal{T}$. Alternatively, if Γ_{ρ_t}'s are closed and $C_{\alpha,W_t,\varrho}$'s are upper semicontinuous functions of ϱ on Γ_{ρ_t}'s then we can use the extreme value theorem for the upper semicontinuous functions to establish the existence of a $\{\varrho_t\}_{t\in\mathcal{T}}$ s.t. $C_{\alpha,W,\varrho}=\sum_{t\in\mathcal{T}}C_{\alpha,W_t,\varrho_t}$. However, such an existence assertion does not hold in general.

A. The A-L Information, Capacity, Center, and Radius

This subsection is a generalization of parts of [4, Ch. 8], which is confined to $|\mathcal{X}|\vee|\mathcal{Y}|<\infty$, $\alpha=1$, and $\ell=1$ case.

For any $\alpha\in\Re_+$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, cost function $\rho:\mathcal{X}\to\Re^{\ell}_{\geq0}$, $\lambda\in\Re^{\ell}_{\geq0}$, and $p\in\mathcal{P}(\mathcal{X})$ *the order α Augustin-Legendre (A-L) information for prior p and Lagrange multiplier λ* is

$$I_\alpha^\lambda(p;W)\triangleq I_\alpha(p;W)-\lambda\cdot\left(\sum\nolimits_x p(x)\rho(x)\right).$$

We call $I_\alpha^\lambda(p;W)$ A-L information because of the convex conjugate pair $f_{\alpha,p}:\Re^{\ell}_{\geq0}\to(-\infty,\infty]$ and $f^*_{\alpha,p}:\Re^{\ell}_{\leq0}\to\Re$:

$$f_{\alpha,p}(\varrho)\triangleq\begin{cases}-I_\alpha(p;W) & \varrho\geq\mathbf{E}_p[\rho]\\ \infty & \text{else}\end{cases}\quad=\sup_{\xi\leq0}\xi\cdot\varrho-f^*_{\alpha,p}(\xi)$$

$$f^*_{\alpha,p}(\xi)\triangleq\sup\nolimits_{\varrho\geq0}\xi\cdot\varrho-f_{\alpha,p}(\varrho)\qquad=\xi\cdot\mathbf{E}_p[\rho]+I_\alpha(p;W)$$

Thus one can write $C_{\alpha,W,\varrho}$ in terms of $I_\alpha^\lambda(p;W)$ as

$$C_{\alpha,W,\varrho}=\sup\nolimits_{p\in\mathcal{P}(\mathcal{X})}\inf\nolimits_{\lambda\geq0} I_\alpha^\lambda(p;W)+\lambda\cdot\varrho.$$

$I_\alpha^\lambda(p;W)$ is convex, decreasing and continuous in λ. Furthermore, by Lemma 1 for $\alpha\in(0,1]$ we have:

$$I_\alpha^\lambda(p;W)=D_\alpha(W\|\,q_{\alpha,p}|\,p)-\lambda\cdot\mathbf{E}_p[\rho]$$

$$D_\alpha(W\|\,q|\,p)-\lambda\cdot\mathbf{E}_p[\rho]\geq I_\alpha^\lambda(p;W)+D_\alpha(q_{\alpha,p}\|\,q)\,.$$

For any $\alpha\in(0,1]$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, $\rho:\mathcal{X}\to\Re_{\geq 0}^{\ell}$, and $\lambda\in\Re_{\geq 0}^{\ell}$, *the A-L capacity* $C_{\alpha,W}^{\lambda}$ and *the A-L radius* $S_{\alpha,W}^{\lambda}$ are given by

$$C_{\alpha,W}^{\lambda}\triangleq\sup\nolimits_{p\in\mathcal{P}(\mathcal{X})} I_{\alpha}^{\lambda}(p;W)$$
$$S_{\alpha,W}^{\lambda}\triangleq\inf\nolimits_{q\in\mathcal{P}(\mathcal{Y})}\sup\nolimits_{x\in\mathcal{X}} D_{\alpha}(W(x)\|\, q)-\lambda\cdot\rho(x).$$

Using the definition of $I_{\alpha}^{\lambda}(p;W)$, $I_{\alpha}(p;W)$ and $S_{\alpha,W}^{\lambda}$ we get

$$C_{\alpha,W}^{\lambda}=\sup\nolimits_{p\in\mathcal{P}(\mathcal{X})}\inf\nolimits_{q\in\mathcal{P}(\mathcal{Y})} D_{\alpha}(W\|\, q|\, p)-\lambda\cdot\mathbf{E}_{p}[\rho]$$
$$S_{\alpha,W}^{\lambda}=\inf\nolimits_{q\in\mathcal{P}(\mathcal{Y})}\sup\nolimits_{p\in\mathcal{P}(\mathcal{X})} D_{\alpha}(W\|\, q|\, p)-\lambda\cdot\mathbf{E}_{p}[\rho]\,.$$

Lemma 8. *For any $\alpha\in(0,1]$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, $\rho:\mathcal{X}\to\Re_{\geq 0}^{\ell}$,*

(a) $C_{\alpha,W}^{\lambda}$ is convex, decreasing and lower semicontinuous in λ on $\Re_{\geq 0}^{\ell}$ and continuous in λ on $\{\lambda:\exists\epsilon>0\, s.t.\, C_{\alpha,W}^{\lambda-\epsilon\mathbb{1}}<\infty\}$.
(b) $C_{\alpha,W,\varrho}\leq\inf_{\lambda\geq 0} C_{\alpha,W}^{\lambda}+\lambda\cdot\varrho$ for all $\varrho\in\Gamma_{\rho}$.
(c) $C_{\alpha,W,\varrho}=\inf_{\lambda\geq 0} C_{\alpha,W}^{\lambda}+\lambda\varrho$ if either $|\mathcal{X}|<\infty$ or $\varrho\in\mathtt{int}\Gamma_{\rho}$.
(d) If $\exists\varrho\in\mathtt{int}\Gamma_{\rho}$ s.t. $C_{\alpha,W,\varrho}<\infty$ then $\forall\varrho\in\mathtt{int}\Gamma_{\rho}$ $\exists\lambda\in\Re_{\geq 0}^{\ell}$ s.t. $C_{\alpha,W,\varrho}=C_{\alpha,W}^{\lambda}+\lambda\cdot\varrho$.
(e) If $C_{\alpha,W,\varrho}=C_{\alpha,W}^{\lambda}+\lambda\cdot\varrho<\infty$ for a $(\varrho,\lambda)\in\Gamma_{\rho}\times\Re_{\geq 0}^{\ell}$, and $\lim_{\imath\to\infty} I_{\alpha}(p^{(\imath)};W)=C_{\alpha,W,\varrho}$ for a $\{p^{(\imath)}\}_{\imath\in\mathbb{Z}_{+}}\subset\mathcal{A}(\varrho)$ then $\lim_{\imath\to\infty} I_{\alpha}^{\lambda}(p^{(\imath)};W)=C_{\alpha,W}^{\lambda}$.

If $\exists\lambda\in\Re_{\geq 0}$ s.t. $C_{\alpha,W}^{\lambda}<\infty$ then $C_{\alpha,W,\varrho}<\infty$ $\forall\varrho\in\Gamma_{\rho}$ by Lemma 8-(a). However, the converse claim is not true. There are cases for which $C_{\alpha,W,\varrho}$ is finite for all $\varrho\in\Gamma_{\rho}$, yet $C_{\alpha,W}^{\lambda}$ is infinite for λ small enough.[7] The equality given in (c) might not hold if $\varrho\in\Gamma_{\rho}\setminus\mathtt{int}\Gamma_{\rho}$ and $|\mathcal{X}|=\infty$.

Theorem 3. *$\forall\alpha\in(0,1]$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, $\rho:\mathcal{X}\to\Re_{\geq 0}^{\ell}$, $\lambda\in\Re_{\geq 0}^{\ell}$,*

$$C_{\alpha,W}^{\lambda}=S_{\alpha,W}^{\lambda}.$$

If $C_{\alpha,W}^{\lambda}<\infty$ then $\exists!q_{\alpha,W}^{\lambda}\in\mathcal{P}(\mathcal{Y})$, called the order α A-L center of W for the Lagrange multiplier λ, such that

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

If $\lim_{\imath\to\infty} I_{\alpha}^{\lambda}(p^{(\imath)};W)=C_{\alpha,W}^{\lambda}<\infty$ for a $\{p^{(\imath)}\}_{\imath\in\mathbb{Z}_{+}}\subset\mathcal{P}(\mathcal{X})$ then corresponding $\{q_{\alpha,p^{(\imath)}}\}_{\imath\in\mathbb{Z}_{+}}$ is a Cauchy sequence for the total variation metric on $\mathcal{P}(\mathcal{Y})$ and $q_{\alpha,W}^{\lambda}$ is its unique limit point.

Lemma 9. *If $\alpha\in(0,1]$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, $\rho:\mathcal{X}\to\Re_{\geq 0}^{\ell}$, $\varrho\in\Gamma_{\rho}$ s.t. $C_{\alpha,W,\varrho}<\infty$ and $\lambda\in\Re_{\geq 0}^{\ell}$ s.t. $C_{\alpha,W,\varrho}=C_{\alpha,W}^{\lambda}+\lambda\cdot\varrho$ then $q_{\alpha,W,\varrho}=q_{\alpha,W}^{\lambda}$.*

Lemma 10. *$\forall\alpha\in(0,1]$, product channel W for finite index set $\mathcal{T}$, and ρ satisfying $\rho(x)=\sum_{t\in\mathcal{T}}\rho_{t}(x_{t})$ for some $\rho_{t}:\mathcal{X}_{t}\to\Re_{\geq 0}^{\ell}$,*

$$C_{\alpha,W}^{\lambda}=\sum\nolimits_{t\in\mathcal{T}} C_{\alpha,W_{t}}^{\lambda}\qquad\forall\lambda\in\Re_{\geq 0}^{\ell}.$$

If $C_{\alpha,W}^{\lambda}<\infty$ then $q_{\alpha,W}^{\lambda}=\prod_{t\in\mathcal{T}}^{\otimes} q_{\alpha,W_{t}}^{\lambda}$.

Recall that the product structure assertion for $q_{\alpha,W,\varrho}$ in Lemma 7, was qualified by the existence of a $\{\varrho_{t}\}_{t\in\mathcal{T}}$ satisfying $\sum_{t\in\mathcal{T}} C_{\alpha,W_{t},\varrho_{t}}=C_{\alpha,W,\varrho}<\infty$. In Lemma 10, on the other hand, the product structure assertion for $q_{\alpha,W}^{\lambda}$ is qualified only by $C_{\alpha,W}^{\lambda}<\infty$.

[7] In [3, §33§35], Augustin considers bounded ρ's of the form $\rho:\mathcal{X}\to[0,1]^{\ell}$. In that case, it is easy to see that if $\exists\varrho\in\mathtt{int}\Gamma_{\rho}$ s.t. $C_{\alpha,W,\varrho}<\infty$ then $\sup_{\varrho\in\Gamma_{\rho}} C_{\alpha,W,\varrho}=C_{\alpha,W,\mathbb{1}}<\infty$ and $C_{\alpha,W}^{\lambda}<\infty$ for all $\lambda\in\Re_{\geq 0}^{\ell}$.

B. The R-G Information, Mean, Capacity, and Center

For any $\alpha\in\Re_{+}\setminus\{1\}$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, cost function $\rho:\mathcal{X}\to\Re_{\geq 0}^{\ell}$, $\lambda\in\Re_{\geq 0}^{\ell}$, and $p\in\mathcal{P}(\mathcal{X})$ *the order α Renyi-Gallager (R-G) information for prior p and Lagrange multiplier λ* is

$$I_{\alpha}^{g\lambda}(p;W)\triangleq\inf\nolimits_{q\in\mathcal{P}(\mathcal{Y})} D_{\alpha}\left(p\circ W e^{\frac{1-\alpha}{\alpha}\lambda\cdot\rho}\middle\|\, p\otimes q\right).$$

The order α R-G capacity for Lagrange multiplier λ is

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

Using the definition of $I_{\alpha}^{g\lambda}(p;W)$ and $C_{\alpha,W}^{g\lambda}$ we get

$$C_{\alpha,W}^{g\lambda}=\sup\nolimits_{p\in\mathcal{P}(\mathcal{X})}\inf\nolimits_{q\in\mathcal{P}(\mathcal{Y})} D_{\alpha}\left(p\circ W e^{\frac{1-\alpha}{\alpha}\lambda\cdot\rho}\middle\|\, p\otimes q\right).$$

Using the concavity of log function and Jensen's inequality one can show that $I_{\alpha}^{\lambda}(p;W)\geq I_{\alpha}^{g\lambda}(p;W)$ for $\alpha\in(0,1)$ and $I_{\alpha}^{\lambda}(p;W)\leq I_{\alpha}^{g\lambda}(p;W)$ for $\alpha\in(1,\infty)$. On the other hand, one can show by substitution that $\forall q\in\mathcal{P}(\mathcal{Y})$ and $\alpha\in\Re_{+}\setminus\{1\}$,

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

where $q_{\alpha,p}^{g\lambda}$ is *the R-G mean* given in terms of $\mu_{\alpha,p}^{\lambda}$ as follows,

$$q_{\alpha,p}^{g\lambda}\triangleq\tfrac{\mu_{\alpha,p}^{\lambda}}{\|\mu_{\alpha,p}^{\lambda}\|}\qquad\tfrac{\mathrm{d}\mu_{\alpha,p}^{\lambda}}{\mathrm{d}\nu}\triangleq\left[\sum\nolimits_{x} p(x)e^{(1-\alpha)\lambda\cdot\rho(x)}\left(\tfrac{\mathrm{d}W(x)}{\mathrm{d}\nu}\right)^{\alpha}\right]^{\frac{1}{\alpha}}.$$

For $\lambda=0\mathbb{1}$, R-G information and mean are equal to the corresponding Renyi information and mean analyzed in [11]. Following a similar analysis one can show that a minimax theorem similar to [11, Thm 1] holds for R-G quantities:

$$C_{\alpha,W}^{g\lambda}=\inf\nolimits_{q\in\mathcal{P}(\mathcal{Y})}\sup\nolimits_{p\in\mathcal{P}(\mathcal{X})} D_{\alpha}\left(p\circ W e^{\frac{1-\alpha}{\alpha}\lambda\cdot\rho}\middle\|\, p\otimes q\right)$$
$$=\inf\nolimits_{q\in\mathcal{P}(\mathcal{Y})}\sup\nolimits_{x\in\mathcal{X}} D_{\alpha}(W(x)\|\, q)-\lambda\cdot\rho(x).$$

Then $C_{\alpha,W}^{g\lambda}=C_{\alpha,W}^{\lambda}$ $\forall\lambda\in\Re_{\geq 0}^{\ell},\alpha\in(0,1)$ by Theorem 3.

IV. Sphere Packing Bounds

Lemma 11. *For any $w=w_{1}\otimes\cdots\otimes w_{n}$, $q=q_{1}\otimes\cdots\otimes q_{n}$, $\kappa\geq 3$, $\alpha\in(0,1)$, if $q(\mathcal{E})\leq(1/\sqrt{16n})e^{-D_{1}(v_{\alpha}^{w,q}\|q)-\alpha 3g_{\kappa}}$ for $\mathcal{E}\in\mathcal{Y}$ and $g_{\kappa}\triangleq\left(\sum_{t=1}^{n}\mathbf{E}_{v_{\alpha}^{w,q}}\left[\left|\ln\frac{\mathrm{d}w_{t\sim}}{\mathrm{d}q_{t}}-\mathbf{E}_{v_{\alpha}^{w,q}}\left[\ln\frac{\mathrm{d}w_{t\sim}}{\mathrm{d}q_{t}}\right]\right|^{\kappa}\right]\right)^{\frac{1}{\kappa}}$ then $w(\mathcal{Y}\setminus\mathcal{E})\geq(1/\sqrt{16n})e^{-D_{1}(v_{\alpha}^{w,q}\|w)-(1-\alpha)3g_{\kappa}}$.*

Lemma 11 is in the spirit of [16, Thm 5], but instead of Chebyshev ineq, it relies on Berry-Essen Thm via [12, Lem9].

Our sphere packing bounds are expressed in terms of the averaged Augustin capacity and[8] averaged sphere packing exponent: for all $\epsilon\in(0,1)$ and $R\in\Re_{+}$:

$$\widetilde{C}_{\alpha,W,\mathcal{A}}^{\epsilon}\triangleq\tfrac{1}{\epsilon}\int_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)}\left[1\vee\left(\tfrac{\alpha}{1-\alpha}\tfrac{1-\phi}{\phi}\right)\right]C_{\phi,W,\mathcal{A}}\mathrm{d}\phi$$
$$\widetilde{E}_{sp}^{\epsilon}(R,W,\mathcal{A})\triangleq\sup\nolimits_{\alpha\in(0,1)}\tfrac{1-\alpha}{\alpha}\left(\widetilde{C}_{\alpha,W,\mathcal{A}}^{\epsilon}-R\right).$$

Lemma 12. *For any $\alpha\in(0,1]$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, $\mathcal{A}\subset\mathcal{P}(\mathcal{X})$ s.t. $C_{1/2,W,\mathcal{A}}\in\Re_{+}$, $\phi\in(0,1)$, $R\in[C_{\phi,W,\mathcal{A}},C_{1,W,\mathcal{A}})$ and $\epsilon\in(0,\phi)$. Then $0\leq\widetilde{E}_{sp}^{\epsilon}(R,W,\mathcal{A})-E_{sp}(R,W,\mathcal{A})\leq\frac{\epsilon}{\phi-\epsilon}\frac{R}{\phi}$.*

Proof of Lemma 12 is identical to that of [12, Lem 11].

[8] Note $\widetilde{C}_{\alpha,W,\varrho}^{\epsilon}=\widetilde{C}_{\alpha,W,\mathcal{A}(\varrho)}^{\epsilon}$ and $\widetilde{E}_{sp}^{\epsilon}(R,W,\varrho)=\widetilde{E}_{sp}^{\epsilon}(R,W,\mathcal{A}(\varrho))$.

Lemma 13. *For any product channel W for the index set $\{1,\ldots,n\}$, cost function ρ satisfying $\rho(x)=\sum_{t\in\mathcal{T}}\rho_t(x_t)$ for $\rho_t:\mathcal{X}_t\to\Re^{\ell}_{\geq 0}$, $\varrho\in\mathtt{int}\Gamma_\rho$, and integers M, L satisfying*

$$\tfrac{M}{L}>\tfrac{8e^2(1-\alpha_0)(1-\epsilon_1)\epsilon_2 n^{2.5}}{\epsilon_1(1-\epsilon_2)}e^{\widetilde{C}^{\epsilon_1}_{\alpha_0,W,\varrho}+\frac{\gamma}{1-\alpha_0}}$$

$$\gamma\triangleq 3\sqrt[\kappa]{3}\left(\sum\nolimits_{t=1}^{n}\left((C_{1/2,W_t,\varrho}+\ln\tfrac{1}{\epsilon_2})\vee\kappa\right)^{\kappa}\right)^{\frac{1}{\kappa}}$$

for a $\kappa\geq 3$, an $\alpha_0\in(0,1)$, an $\epsilon_1\in(0,1)$ and an $\epsilon_2\in(0,1)$ satisfying $\frac{(n-1)(1-\alpha_0)(1-\epsilon_1)}{\epsilon_1}\geq 1$, any (M,L) channel code (Ψ,Θ) on W satisfying $\vee_{m\in\mathcal{M}}\rho(\Psi(m))\leq\varrho$ satisfies

$$P_{\mathbf{e}}^{av}\geq\left(\tfrac{\epsilon_1 e^{-2\gamma}}{8e^2(1-\alpha_0)(1-\epsilon_1)n^{1.5}}\right)^{\frac{1}{\alpha_0}}e^{-\widetilde{E}^{\epsilon_1}_{sp}\left(\ln\frac{M}{L},W,\varrho\right)}.$$

Proof Sketch. Since $\varrho\in\mathtt{int}\Gamma_\rho$, $\forall\alpha\in(0,1)\exists\lambda_{\alpha,W,\varrho}\in\Re^{\ell}_{\geq 0}$ s.t. $C_{\alpha,W,\varrho}=C^{\lambda_{\alpha,W,\varrho}}_{\alpha,W}+\lambda_{\alpha,W,\varrho}\varrho$ by Lem.8-(d). Then $q_{\alpha,W,\varrho}=q^{\lambda_{\alpha,W,\varrho}}_{\alpha,W}$ by Lemma 9. Furthermore, $q^{\lambda_{\alpha,W,\varrho}}_{\alpha,W}=\prod_t^{\otimes}q^{\lambda_{\alpha,W,\varrho}}_{\alpha,W_t}$ by Lemma 10. Then $q^{\lambda_{\alpha,W,\varrho}}_{\alpha,W_t}:(0,1)\to\mathcal{P}(\mathcal{Y}_t)$ is continuous in α for the total variation topology on $\mathcal{P}(\mathcal{Y}_t)$ because $q_{\alpha,W,\varrho}$ is by Lemma 4. Then $q^{\lambda_{\cdot,\varrho}}_{\cdot,W_t}$ is a transition probability from $((0,1),\mathcal{B}((0,1)))$ to $(\mathcal{Y}_t,\mathcal{Y}_t)$. We define $q^{\epsilon}_{\alpha,W_t}$ as the $\mathcal{Y}_t$ marginal of the probability measure $u_{\alpha,\epsilon}\circ q^{\lambda_{\cdot,\varrho}}_{\cdot,W_t}$ where $u_{\alpha,\epsilon}$ is the uniform probability distribution on $(\alpha-\epsilon\alpha,\alpha+\epsilon(1-\alpha))$:

$$q^{\epsilon}_{\alpha,W_t}=\tfrac{1}{\epsilon}\int_{\alpha-\alpha\epsilon}^{\alpha+(1-\alpha)\epsilon}q^{\lambda_{\phi,W,\varrho}}_{\phi,W_t}\mathrm{d}\phi. \tag{6}$$

Let $\Psi_t(m)$ be the $\mathcal{Y}_t$ marginal of $\Psi(m)$ and $q_{\alpha,t}$, q_α, v^m_α be

$$q_{\alpha,t}\triangleq(1-\epsilon_2)q^{\epsilon_1}_{\alpha,W_t}+\epsilon_2 q_{\frac{1}{2},W_t,\varrho}\qquad q_\alpha\triangleq\prod\nolimits_t^{\otimes}q_{\alpha,t}\qquad v^m_\alpha\triangleq v^{\Psi(m),q_\alpha}_\alpha.$$

By [11, Lem 9-(b,d)], Lemma 10 and $\ln\tau\leq\tau-1$ we have

$$D_\alpha(\Psi(m)\|\,q_\alpha)\leq\tfrac{n\epsilon_2}{1-\epsilon_2}+\int_{\alpha(1-\epsilon_1)}^{\alpha+(1-\alpha)\epsilon_1}\frac{D_\alpha\left(\Psi(m)\|\,q^{\lambda_{\phi,W,\varrho}}_{\phi,W}\right)}{\epsilon_1}\mathrm{d}\phi.$$

Using Lemma 9, [11, Lem 9-(a)], [7, Prop 2], Theorem 2, $\rho(\Psi(m))\leq\varrho$ and the definition of $\widetilde{C}^{\epsilon}_{\alpha,W,\mathcal{A}}$ we get

$$D_\alpha(\Psi(m)\|\,q_\alpha)\leq\tfrac{n\epsilon_2}{1-\epsilon_2}+\widetilde{C}^{\epsilon_1}_{\alpha,W,\varrho}. \tag{7}$$

Let $(\Psi_t(m))_\sim$ be the component of $\Psi_t(m)$ that is absolutely continuous in $q_{\alpha,t}$. Furthermore, let $\xi^m_{\alpha,t}$ and ξ^m_α be

$$\xi^m_{\alpha,t}\triangleq\ln\tfrac{\mathrm{d}(\Psi_t(m))_\sim}{\mathrm{d}q_{\alpha,t}}-\mathbf{E}_{v^m_\alpha}\left[\ln\tfrac{\mathrm{d}(\Psi_t(m))_\sim}{\mathrm{d}q_{\alpha,t}}\right]\qquad\xi^m_\alpha\triangleq\sum\nolimits_{t=1}^{n}\xi^m_{\alpha,t}.$$

Then using [12, Lem 6], [11, Lem 9-(b,a)], [7, Prop 2], and Theorem 2 we get

$$\mathbf{E}_{v^m_\alpha}\left[\left|\xi^m_{\alpha,t}\right|^\kappa\right]^{\frac{1}{\kappa}}\leq 3^{\frac{1}{\kappa}}\tfrac{(C_{1/2,W_t,\varrho}-\ln\epsilon_2)\vee\kappa}{\alpha(1-\alpha)}\qquad\forall\kappa\in\Re_+,\alpha\in(0,1).$$

Then using the definition of γ we get

$$\left[\sum\nolimits_{t=1}^{n}\mathbf{E}_{v^m_\alpha}\left[\left|\xi^m_{\alpha,t}\right|^\kappa\right]\right]^{\frac{1}{\kappa}}\leq\tfrac{\gamma}{3\alpha(1-\alpha)}. \tag{8}$$

On the other hand, $\forall m\in\mathcal{M},\alpha\in(0,1)$ by the definition of v^m_α

$$D_1(v^m_\alpha\|\,q_\alpha)=D_\alpha(\Psi(m)\|\,q_\alpha)-\tfrac{\alpha}{1-\alpha}D_1(v^m_\alpha\|\,\Psi(m)). \tag{9}$$

Thus we can bound $D_1(v^m_\alpha\|\,q_\alpha)$ using the non-negativity of the Renyi divergence, i.e. [7, Thm 8], and equation (7) as $D_1(v^m_\alpha\|\,q_\alpha)\leq\frac{n\epsilon_2}{1-\epsilon_2}+\widetilde{C}^{\epsilon_1}_{\alpha,W,\varrho}$. Hence,

$$\lim_{\alpha\to\alpha_0}D_1(v^m_\alpha\|\,q_\alpha)+\tfrac{\gamma}{3(1-\alpha)}<\ln\tfrac{M}{L}+\ln\tfrac{\epsilon_1}{8e^2(1-\alpha_0)(1-\epsilon_1)n^{1.5}}$$

$$\lim_{\alpha\to 1}D_1(v^m_\alpha\|\,q_\alpha)+\tfrac{\gamma}{3(1-\alpha)}=\infty.$$

$D_1(v^m_\alpha\|\,q_\alpha)$ is continuous in α by [12, Lem 7], then by the intermediate value theorem [14, 4.23] $\forall m\in\mathcal{M}\exists\alpha_m\in(\alpha_0,1)$ s.t.

$$D_1(v^m_\alpha\|\,q_\alpha)+\tfrac{\gamma}{3(1-\alpha)}|_{\alpha=\alpha_m}=\ln\tfrac{M}{L}+\ln\tfrac{\epsilon_1}{8e^2(1-\alpha_0)(1-\epsilon_1)n^{1.5}}.$$

Lemma 13 follows from Lemma 11 through a pigeon hole argument similar to the one invoked in [12, eq (68)-(69)]. ☐

If W is stationary and memoryless Lemma 13 can be proved $\forall\varrho\in\Gamma_\rho$ by setting $q^{\epsilon}_{\alpha,W_t}=\int u_{\alpha,\epsilon}\circ q_{\phi,W_t,\frac{\varrho}{n}}\mathrm{d}\phi$. Furthermore, bound given in (10) can be obtained for codes satisfying a convex empirical distribution constraint $\mathcal{A}\subset\mathcal{P}(\mathcal{X}_1)$ by setting $q^{\epsilon}_{\alpha,W_t}=\int u_{\alpha,\epsilon}\circ q_{\phi,W_t\mathcal{A}}\mathrm{d}\phi$ and $q_{\alpha,t}=(1-\epsilon_2)q^{\epsilon_1}_{\alpha,W_t}+\epsilon_2 q_{\frac{1}{2},W_t,\mathcal{B}_{\mathcal{A}}}$ where $\mathcal{B}_{\mathcal{A}}\triangleq\mathcal{P}(\{x\in\mathcal{X}_1:\exists p\in\mathcal{A}\text{ s.t. }p(x)>0\})$.

$$\tilde{\gamma}=3\sqrt[\kappa]{3n}\left((C_{1/2,W_1,\mathcal{B}(\mathcal{A})}+\ln\tfrac{1}{\epsilon_2})\vee\kappa\right)$$

$$P_{\mathbf{e}}^{av}\geq\left(\tfrac{\epsilon_1 e^{-2\tilde{\gamma}}}{8e^2(1-\alpha_0)(1-\epsilon_1)n^{1.5}}\right)^{\frac{1}{\alpha_0}}e^{-n\widetilde{E}^{\epsilon_1}_{sp}\left(\frac{1}{n}\ln\frac{M}{L},W_1,\mathcal{A}\right)}. \tag{10}$$

ACKNOWLEDGMENT

Author would like to thank Fatma Nakiboğlu and Mehmet Nakiboğlu for their hospitality; this work simply would not have been possible without it.

REFERENCES

[1] Y. Altuğ and A. B. Wagner. Refinement of the sphere-packing bound: Asymmetric channels. *IEEE Trans. on Information Theory*, 60(3):1592–1614, March 2014.

[2] U. Augustin. Error estimates for low rate codes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 14(1):61–88, 1969.

[3] U. Augustin. *Noisy Channels*. 1978. (https://goo.gl/8wupWM).

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

[5] M. Dalai. Classical and classical-quantum sphere packing bounds: Rényi vs Kullback and Leibler. In *International Zurich Seminar on Communications*, pages 198–202, 2016.

[6] P. M. Ebert. *Error Bounds For Parallel Communication Channels*. 1966. (https://hdl.handle.net/1721.1/4295).

[7] T. van Erven and P. Harremoes. Renyi divergence and Kullback-Leibler divergence. *IEEE Trans. on Information Theory*, 60(7):3797–3820, July 2014.

[8] R. G. Gallager. A simple derivation of the coding theorem and some applications. *IEEE Trans. on Information Theory*, 11(1):3–18, Jan. 1965.

[9] E. A. Haroutunian. Estimates of the error probability exponent for a semicontinuous memoryless channel. *Problems of Information Transmission*, 4(4):37–48, 1968.

[10] A. N. Kolmogorov and S. V. Fomin. *Introductory Real Analysis*. Dover, 1975.

[11] B. Nakiboğlu. Augustin's method - part I: the Renyi center. *arXiv:*1608.02424v2.

[12] B. Nakiboğlu. Augustin's method - part II: the sphere packing bound. *arXiv:*1611.06924v1.

[13] B. Nakiboğlu. The Augustin center and the sphere packing bound for memoryless channels. *(in preparation)*.

[14] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, 1976.

[15] C. E. Shannon. Probability of error for optimal codes in a Gaussian channel. *The Bell System Technical Journal*, 38(3):611–656, May 1959.

[16] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp. Lower bounds to error probability for coding on discrete memoryless channels. I. *Information and Control*, 10(1):65–103, 1967.