The Augustin Center and The Sphere Packing Bound For Memoryless Channels

Barıs 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 \(\frac{1}{2}\) 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 \(n\). [4] Theorem 2]. In this manuscript, we derive analogous results for two families of memoryless channels. As we have done for the product channels in [2], 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 [4] §36]. In addition, Augustin established the connection between the exponent of Gallager’s inner bound for the cost constrained channels [4] Thm 8] and the sphere packing exponent [3] §35]. Our results surpass Augustin’s results in two ways:

1. Augustin assumes the cost function to be bounded [4] This hypothesis excludes certain important and interesting cases such as the Gaussian channels. Hence, Augustin’s analysis in [4] does not imply the sphere packing bounds derived by Shannon [5] 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. It is even possible to handle certain fading scenarios and additional per antenna power constraints.

2. The best asymptotic bound implied by Augustin’s non-asymptotic bound [3] Thm 36.6] is of the form \(P_{\text{err}}(n) \geq O(\frac{1}{\sqrt{n}}) e^{-\frac{1}{2}(\frac{1}{\sqrt{n}})} \sum_{t=1}^{n} O(\sqrt{\frac{1}{n^3}}) W_{n}(x_{t}, y_{t})\). In Theorem 1, we replace this by \(O(\frac{1}{\sqrt{n}})\) by \(O(\frac{1}{\sqrt{n}})\) 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. 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 \(X\), \(P(X)\) is the set of all probability mass functions that are non-zero only on finitely many members of \(X\); \(M^+(X)\) is the set of all non-zero mass functions with the same property. For any measurable space \((Y, \mathcal{Y})\), \(P(Y)\) is the set of all probability measures and \(M^+(Y)\) is set of all finite measures. For any \(\mu, \nu \in M^+(Y)\), \(\mu \leq \nu\) if \(\mu(E) \leq \nu(E)\) \(\forall E \in \mathcal{Y}\). Similarly, for any \(\mu, \nu \in \mathbb{R}^Y\), \(\mu \leq \nu\) if \(\mu(E) \leq \nu(E)\) \(\forall E \in \{1, \ldots, \ell\}\). For any \(\mu \in \mathbb{R}^Y\), \(\nu \in \mathbb{R}_+\) is the vector whose all entries are one. For any \(S \subseteq \mathbb{R}\) we denote the interior of \(S\) by \(\text{int} S\). For any set \(S\) in a vector space we denote the convex hull of \(S\) by \(\text{ch} S\).

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

A product channel is stationary iff all \(W_t\)’s are identical. If \(X \subseteq \bigotimes_{t \in \mathcal{T}} X_t\) then \(W\) is a memoryless channel.

1Shannon’s approximation error terms in $X_t$ 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.

2The issue here is not a matter of rescaling: certain conclusions of Augustin’s analysis are not correct when cost functions are not bounded.
An \((M, L)\) channel code on \(W : X \to \mathcal{P}(Y)\) is an ordered pair \((\Psi, \Theta)\) composed of an encoding function \(\Psi : M \to X\) and a decoding function \(\Theta : Y \to M\) where \(M = \{1, 2, \ldots, M\}\), \(M = \{L : L \in \mathbb{M}\) and \(|L| = L\), and \(\Theta\) is a measure as a function from the measurable space \((Y, \mathcal{Y})\).

Given an \((M, L)\) channel code \((\Psi, \Theta)\) on \(W : X \to \mathcal{P}(Y)\) the average error probability \(P^m\) and the conditional error probability \(P^m_{e}\) for \(m\) are given by

\[
P^m = \frac{1}{|M|} \sum_{m \in M} P^m_{e} = P^m(W(\Psi(m))) \text{ for } m \in M.
\]

A cost function \(p\) is a function from the input set to \(\mathbb{R}^+\) for some \(\ell \in \mathbb{Z}^+.\) We assume without loss of generality that

\[
\inf_{x \in X} p^\ell(x) = 0 \quad \forall \ell \in \{1, \ldots, \ell\}.
\]

Let \(I_p^\ell\) be the set of feasible cost constraints for \(\mathcal{P}(X)\):

\[
I_p^\ell \triangleq \{q \in \mathbb{R}^+_0 : \exists p \in \mathcal{P}(X) \text{ s.t. } \sum_m p(x)p(x) \leq q(x) \}.
\]

Then \(I_p^\ell\) is a convex set with non-empty interior. A cost function \(p\) for a product channel \(W\) is said to be additive iff there exists a \(p : X \to \mathbb{R}^+\) for each \(\ell \in \mathcal{T}\) such that

\[
\rho(x) = \sum_{\ell \in \mathcal{T}} p\ell(x) \quad \forall x \in X.
\]

An encoding function \(\Psi\), hence the corresponding cost, is said to satisfy the cost constraint \(q\) iff \(\forall m \in M \sum p(x)p(x) \leq q(x)\).

A code on a product channel \(W : \prod_{\ell \in \mathcal{T}} X_\ell \to \mathcal{P}(Y)\) is said to satisfy an empirical distribution constraint \(A \subset \mathcal{P}(\mathcal{X}_1)\) if the empirical distribution, i.e. type or composition, of \(\Psi(m)\) is in \(A\) for all \(m \in M\).

\section{B. Main Result}

\begin{assumption}
\{(\Psi_\ell, \rho_\ell, \varpi_\ell)\}_{\ell \in \mathcal{Z}_n} \text{ is an ordered sequence of channels with associated cost functions and cost constraints satisfying the following condition: } \exists \mu_0 \in \mathbb{Z}^+, K \in \mathbb{R}^+ \text{ s.t. } \max_{\ell \leq n} C_{\Psi_\ell, \rho_\ell, \varpi_\ell} \leq K \ln(n) \quad \text{and} \quad \varpi_{n} = \inf_{\ell \leq n} I_{\rho^\ell}^{\mu_0},
\end{assumption}

for all \(n \geq n_0\) where \(p_{[1,n]}(x_{[1,n]}) = \sum_{i=1}^{n} \rho_i(x_i)\).

\begin{theorem}
Let \(\{(\Psi_\ell, \rho_\ell, \varpi_\ell)\}_{\ell \in \mathcal{Z}_n}\) be a sequence satisfying \begin{assumption} \(a_{\alpha_0, \alpha_1}\) be orders satisfying \(0 < \alpha_0 < \alpha_1 < 1\) and \(\varepsilon \in \mathbb{R}^+\). Then for any sequence of codes \(\{(\Psi_\ell, \Theta_\ell)\}_{\ell \in \mathcal{Z}_n}\) on the product channels \(\{W_{[1,n]}\}_{\ell \in \mathcal{Z}_n}\) satisfying

\[
\forall m \in M, p_{[1,n]}(\Psi_\ell(m)) \leq q_\alpha \\
C_{\alpha_0, W_{[1,n]} : \Theta_\ell} + \varepsilon \ln 2^n \leq \ln \frac{1}{\varpi_\ell} \\
\exists \tau \in \mathbb{R}^+ \text{ and } \varepsilon_{n} \geq 0 \text{ such that } P^{\alpha'}_{e}(n) \geq n^{-\varepsilon_{n}} - B_{\rho}(\ln \frac{1}{\varpi_\ell}, W_{[1,n]} : \Theta_\ell) \quad \forall n \geq n_1
\]

where \(E_{\rho^\ell}(R, W, \mu) = \sup_{\delta \epsilon (0,1)} \frac{1}{\delta} \left( C_{\alpha_0, W_{[1,n]} : \Theta_\ell} + R \right)\).

\end{theorem}

\begin{lemma}
\end{lemma}

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

\footnote{Recall that for any encoder \(\Psi\) a deterministic MAP decoder obtains minimum \(P^m_{e}\) among all, possibly non-deterministic, decoders.}

\footnote{Augustin \[\] §33 has the following additional hypothesis: \(\forall x \in \mathcal{T} \rho^\ell(x) \leq 1.\)}
B. The Constrained Augustin Capacity and Center

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

\[
C_{\alpha, W, A} : = \sup_{p \in A} D_{\alpha}(W \| q | p).
\]

Using the definition of \( D_{\alpha}(p; W) \) we get

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

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 [14, §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 \( A \subset \mathcal{P}(\mathcal{X}) \)

(a) \( C_{\alpha, W, A} : (0, 1] \to [0, \infty) \) is increasing and continuous.
(b) \( C_{\alpha, W, A} : (0, 1) \to [0, \infty) \) is decreasing and continuous.
(c) \( \exists \alpha \in (0, 1) \) s.t. \( C_{\alpha, W, A} < \infty \) \( \forall \phi \in (0, 1) \).

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

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

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

\[
C_{\alpha, W, A} = \sup_{p \in A} D_{\alpha}(W \| q_{\alpha, W, A} | p).
\]

If \( \lim_{t \to 1} D_{\alpha}(p(t) | W) = C_{\alpha, W, A} < \infty \) for a \( \{p(t)\}_{t \in \mathbb{Z}^+} \subset A \) then \( \{p(t)\}_{t \in \mathbb{Z}^+} \) is a Cauchy sequence for the total variation metric on \( \mathcal{P}(\mathcal{Y}) \) and \( q_{\alpha, W, A} \) is its unique limit point.

Lemma [1] and Theorem [2] imply for all \( \alpha \in (0, 1], p \in A \) that

\[
C_{\alpha, W, A} = I_{\alpha}(p; W) \geq D_{\alpha}(q_{\alpha, W, A} | q | p).
\]

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 \( A \subset \mathcal{P}(\mathcal{X}) \) s.t. \( C_{\alpha, W, A} < \infty \), and \( q \in \mathcal{P}(\mathcal{Y}) \),

\[
\sup_{p \in A} D_{\alpha}(W \| q | p) \leq C_{\alpha, W, A} + D_{\alpha}(q_{\alpha, W, A} | q).
\]

Erven-Harremoes bound, the continuity of \( C_{\alpha, W, A} \) in \( \alpha \), and Pinsker’s inequality imply the continuity of \( q_{\alpha, W, A} \) in \( \alpha \) 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 \( A \subset \mathcal{P}(\mathcal{X}) \) s.t. \( C_{\eta, W, A} < \infty \), and \( \alpha, \phi \) satisfying \( 0 < \alpha < \phi \leq \eta \),

\[
C_{\phi, W, A} \geq C_{\alpha, W, A} + D_{\alpha}(q_{\alpha, W, A} | \phi | q_{\phi, W, A}).
\]

Furthermore, \( q_{\alpha, W, A} \) is continuous in \( \alpha \) 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 \( A_t \subset \mathcal{P}(\mathcal{X}_t) \) for each \( t \in \mathcal{T} \), and \( A = \bigcap_{t \in \mathcal{T}} A_t \subset \mathcal{P}(\mathcal{X}) \) for each \( t \in \mathcal{T} \),

\[
C_{\alpha, W, A} = \sum_{t \in \mathcal{T}} C_{\alpha, W, A_t}.
\]

Furthermore, if \( C_{\alpha, W, A} < \infty \) then \( q_{\alpha, W, A} = \bigcap_{t \in \mathcal{T}} q_{\alpha, W, A_t} \).

III. The Cost Constrained Augustin Capacity

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

\[
C_{\alpha, W, \varepsilon \phi} : = \sup_{p \in \mathcal{P}(\mathcal{X})(\mathcal{Y})} D_{\alpha}(W \| q | p) \quad \forall \varepsilon \in \mathcal{Y}.
\]

where \( \mathcal{A}(\alpha) : = \{ p \in \mathcal{P}(\mathcal{X}) : \sum_{\mathcal{Y}} p(x) \rho(x) \leq \varepsilon \} \). Note that Theorem [2] and Lemmas [3] and [4] hold for \( C_{\alpha, W, \varepsilon \phi} \) because \( \mathcal{A}(\alpha) \) is a convex set. We denote Augustin center by \( q_{\alpha, W, \varepsilon \phi} \).

Lemma 6. For any \( \alpha \in (0, 1], W : \mathcal{X} \to \mathcal{P}(\mathcal{Y}) \), \( \rho : \mathcal{X} \to \mathbb{R}_{\geq 0} \),

(a) \( C_{\alpha, W, \varepsilon \phi} : [0, \infty) \) is increasing and concave in \( \varepsilon \). It is either infinite \( \varepsilon \in \mathcal{P}(\mathcal{Y}) \) or finite and continuous on \( \mathcal{P}(\mathcal{Y}) \).

(b) If \( C_{\alpha, W, \varepsilon \phi} < \infty \) for an \( \varepsilon \in \mathcal{P}(\mathcal{Y}) \), then \( C_{\alpha, W, \varepsilon \phi} \) is continuous and compact.

The set of all such \( q_{\alpha, W, \varepsilon \phi} \)’s for an \( \alpha \) is convex and compact.

Lemma 7. For any \( \alpha \in (0, 1], \rho \) a product function \( W : \mathcal{X} \to \mathcal{P}(\mathcal{Y}) \), convex function \( \rho : \mathcal{X} \to \mathbb{R}_{\geq 0} \), \( \alpha \in \mathbb{R}_{\alpha} \), and \( \rho \in \mathcal{P}(\mathcal{X}) \),

the order \( \alpha \) Augustin-Legendre (A-L) information for prior \( \rho \) and Lagrange multiplier \( \lambda \) is

\[
I_{\alpha}^\rho(p; W) = I_{\alpha}(p; W) - \lambda \left( \sum_{x} p(x) \rho(x) \right).
\]

We call \( I_{\alpha}^\rho(p; W) \) A-L information because of the convex conjugate pair \( f_{\alpha, p} : \mathbb{R}_{\geq 0} \to (-\infty, \infty) \) and \( f_{\alpha, p}^* : \mathbb{R}_{\geq 0} \to \mathbb{R} \):

\[
f_{\alpha, p}(\varepsilon) = \begin{cases} -I_{\alpha}(p; W) & \varepsilon \geq E_{\rho[p]}[\varepsilon] \\ \infty & \varepsilon < E_{\rho[p]}[\varepsilon] \end{cases}
\]

Thus one can write \( C_{\alpha, W, \varepsilon \phi} \) in terms of \( I_{\alpha}^\rho(p; W) \) as

\[
C_{\alpha, W, \varepsilon \phi} = \sup_{p \in \mathcal{P}(\mathcal{X})} \inf_{\varepsilon \geq 0} I_{\alpha}^\rho(p; W) + \lambda \cdot \varepsilon.
\]

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

\[
I_{\alpha}^\rho(p; W) = D_{\alpha}(W \| q_{\alpha, W, \varepsilon \phi} | p) - \lambda \cdot E_{\rho[p]}[\varepsilon] \geq I_{\alpha}^\rho(p; W) + D_{\alpha}(q_{\alpha, W, \varepsilon \phi} \| q).
\]
For any \( \alpha \in (0, 1] \), \( W : X \to \mathcal{P}(Y) \), \( \rho : X \to \mathbb{R}_+^{(\ell)} \), and \( \lambda \in \mathbb{R}_+^{(\ell)} \), the A-L capacity \( C_{\alpha,W}^\lambda \) and the A-L radius \( S_{\alpha,W}^\lambda \) are given by

\[
C_{\alpha,W}^\lambda = \sup_{p \in \mathcal{P}(X)} I_{\alpha}^\lambda(p; W) \quad \text{s.t.} \quad S_{\alpha,W}^\lambda = \inf_{q \in \mathcal{P}(Y)} \sup_{x \in X} D_a(W(x)\|q) - \lambda \cdot \rho(x).
\]

Using the definition of \( I_{\alpha}^\lambda(p; W) \), we obtain:

\[
C_{\alpha,W}^\lambda = \sup_{p \in \mathcal{P}(X)} \inf_{q \in \mathcal{P}(Y)} D_a(W(x)\|q) - \lambda \cdot \rho(x).
\]

Lemma 8. For any \( \alpha \in (0, 1] \), \( W : X \to \mathcal{P}(Y) \), \( \rho : X \to \mathbb{R}_+^{(\ell)} \), \( \lambda \in \mathbb{R}_+^{(\ell)} \),

(a) \( C_{\alpha,W}^\lambda \) is convex, decreasing and lower semicontinuous in \( \lambda \) on \( \mathbb{R}_+^{(\ell)} \) and continuous in \( \lambda \) on \( \mathbb{R}_+^{(\ell)} \).

(b) \( C_{\alpha,W}^\lambda(p; W) \leq C_{\alpha,W}^\lambda(0; W) + \lambda \cdot \rho \) for all \( \rho \in \mathcal{P}(X) \).

(c) \( C_{\alpha,W}^\lambda(p; W) \leq \inf \{ I_{\alpha}^\lambda(p; W), S_{\alpha,W}^\lambda \} \).

(d) \( C_{\alpha,W}^\lambda(p; W) \leq \inf \{ I_{\alpha}^\lambda(p; W), S_{\alpha,W}^\lambda \} \).

Lemma 9. For any \( \alpha \in (0, 1] \), \( W : X \to \mathcal{P}(Y) \), \( \rho : X \to \mathbb{R}_+^{(\ell)} \), \( \lambda \in \mathbb{R}_+^{(\ell)} \), and \( \delta \in \mathbb{R}_+^{(\ell)} \),

(a) \( C_{\alpha,W}^\lambda(p; W) = \sup_{p \in \mathcal{P}(X)} \inf_{q \in \mathcal{P}(Y)} D_a(W(x)\|q) - \lambda \cdot \rho(x) \).

(b) \( C_{\alpha,W}^\lambda(p; W) = \sup_{p \in \mathcal{P}(X)} \inf_{q \in \mathcal{P}(Y)} D_a(W(x)\|q) - \lambda \cdot \rho(x) \).

(c) \( C_{\alpha,W}^\lambda(p; W) = \sup_{p \in \mathcal{P}(X)} \inf_{q \in \mathcal{P}(Y)} D_a(W(x)\|q) - \lambda \cdot \rho(x) \).

(d) \( C_{\alpha,W}^\lambda(p; W) = \sup_{p \in \mathcal{P}(X)} \inf_{q \in \mathcal{P}(Y)} D_a(W(x)\|q) - \lambda \cdot \rho(x) \).

Theorem 3. For any \( \alpha \in (0, 1] \), \( W : X \to \mathcal{P}(Y) \), \( \rho : X \to \mathbb{R}_+^{(\ell)} \), \( \lambda \in \mathbb{R}_+^{(\ell)} \),

\[
C_{\alpha,W}^\lambda = \sup_{p \in \mathcal{P}(X)} \sup_{\lambda \in \mathbb{R}_+^{(\ell)}} \left( \inf_{q \in \mathcal{P}(Y)} D_a(W(x)\|q) - \lambda \cdot \rho(x) \right).
\]

Lemma 11. For any \( w = \omega_1 \otimes \cdots \otimes \omega_n \), \( q = q_1 \otimes \cdots \otimes q_n \), \( \alpha \in (0, 1] \), \( \rho : X \to \mathbb{R}_+^{(\ell)} \),

\[
\mathbb{E}^p(\rho(W, A) \mid \mathcal{F}_n) = \mathbb{E}^p(\rho(W, A) \mid \mathcal{F}_n) \quad \text{for all} \quad \alpha \in (0, 1].
\]

Lemma 12. For any \( \alpha \in (0, 1] \), \( W : X \to \mathcal{P}(Y) \), \( \rho : X \to \mathbb{R}_+^{(\ell)} \), \( \lambda \in \mathbb{R}_+^{(\ell)} \),

\[
\mathbb{E}^p(\rho(W, A) \mid \mathcal{F}_n) = \mathbb{E}^p(\rho(W, A) \mid \mathcal{F}_n) \quad \text{for all} \quad \alpha \in (0, 1].
\]
Lemma 13. For any product channel $W$ for the index set \( \{1, \ldots, n\} \), cost function \( \rho(x) = \sum_{i \in \{1, \ldots, n\}} p_i(x_i) \) for \( p_i : \mathbb{X}_i \rightarrow \mathbb{R}^N \) and integers \( M, L \) satisfying

\[
\frac{M}{L} > \frac{a(v_0^m - (1-a_0)(1-a_1))}{\epsilon_1(1-a_2)}
\]

\[
\gamma \triangleq 3 \left( \frac{n}{\epsilon_1(1-a_2)} \right)^{\frac{1}{2}} \max \left( (C_2, W, t \phi) + \ln \frac{1}{\kappa} \right) \left( \frac{\ln \kappa}{\kappa} \right)^{\frac{1}{2}}
\]

for a \( \kappa \geq 3 \), an \( a_0 \in (0, 1) \), an \( \epsilon_1 \in (0, 1) \) and an \( \epsilon_2 \in (0, 1) \) satisfying

\[
(\frac{\epsilon_1}{\epsilon_2})(1-a_0)/(1-a_1) \geq 1,
\]

any \( (M, L) \) channel code \((\Psi, \Theta)\) on \( W \) satisfying \( \forall m \in \mathcal{M}(\rho(m)) \leq 0 \) satisfies

\[
P_\alpha^m \geq \left( \frac{\epsilon_1}{\epsilon_2}(1-a_0)/(1-a_1) \right)^{\frac{1}{2}} e^{-\frac{1}{2} \epsilon_2/2 \ln \kappa} \left( 1 \frac{\ln \kappa}{\kappa} \right)^{\frac{1}{2}}
\]

Proof Sketch. Since \( q \in \mathcal{F}_m^R, \forall \alpha \in (0, 1) \exists \lambda_\alpha \in \mathbb{R}^N \) and

\[
C_\alpha, W, \phi = \lambda_{\alpha, W} \phi + \lambda_{\alpha, W, \psi} \text{by Lemma 10.}
\]

Then \( q_{\alpha, W, \phi} = \lambda_{\alpha, W, \phi} \) by Lemma 10. Further, \( \lambda_{\alpha, W, \phi} = \sum_{m=0}^{\infty} \lambda_{\alpha, W, \phi} \) by Lemma 10.

Thus \( q_{W, \phi} = \sum_{m=0}^{\infty} \lambda_{\alpha, W, \phi} \).

Please refer to the reference for details.

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. Altug 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. Augsting. Error estimates for low rate codes. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 14(1):61–80, 1969.

[3] U. Augsting. Noisy Channels. 1976. (https://books.google.com.wpm.WM). 1966.

[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-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/1721.1/14295).

[7] T. van Erven and P. Harremoes. Rényi 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:1609.02424v2.

[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.