Extremal Relations Between Shannon Entropy and $\ell_\alpha$-Norm

Yuta Sakai and Ken-ichi Iwata
Department of Information Science, University of Fukui,
3-9-1 Bunkyo, Fukui, Fukui, 910-8507, Japan,
E-mail: {ji140117,k-iwata}@u-fukui.ac.jp

Abstract
The paper examines relationships between the Shannon entropy and the $\ell_\alpha$-norm for $n$-ary probability vectors, $n \geq 2$. More precisely, we investigate the tight bounds of the $\ell_\alpha$-norm with a fixed Shannon entropy, and vice versa. As applications of the results, we derive the tight bounds between the Shannon entropy and several information measures which are determined by the $\ell_\alpha$-norm. Moreover, we apply these results to uniformly focusing channels. Then, we show the tight bounds of Gallager’s $E_0$ functions with a fixed mutual information under a uniform input distribution.

I. INTRODUCTION

Information measures of random variables are used in several fields. The Shannon entropy [1] is one of the famous measures of uncertainty for a given random variable. On the studies of information measures, inequalities for information measures are commonly used in many applications. As an instance, Fano’s inequality [2] gives the tight upper bound of the conditional Shannon entropy with a fixed error probability. Then, note that the tight means the existence of the distribution which attains the equality of the bound. Later, the reverse of Fano’s inequality, i.e., the tight lower bound of the conditional Shannon entropy with a fixed error probability, are established [3]–[5]. On the other hand, Harremoës and Topsøe [8] derived the exact range between the Shannon entropy and the index of coincidence (or the Simpson index) for all $n$-ary probability vectors, $n \geq 3$. In the above studies, note that the error probability and the index of coincidence are closely related to $\ell_\infty$-norm and $\ell_2$-norm, respectively. Similarly, several axiomatic definitions of the entropies [9]–[14] are also related to the $\ell_\alpha$-norm. Furthermore, the $\ell_\alpha$-norm are also related to some diversity indices, such as the index of coincidence.

In this study, we examine extremal relations between the Shannon entropy and the $\ell_\alpha$-norm for $n$-ary probability vectors, $n \geq 2$. More precisely, we establish the tight bounds of $\ell_\alpha$-norm with a fixed Shannon entropy in Theorem 1. Similarly, we also derive the tight bounds of the Shannon entropy with a fixed $\ell_\alpha$-norm in Theorem 2. Directly extending Theorem 1 to Corollary 1, we can obtain the tight bounds of several information measures which are determined by the $\ell_\alpha$-norm with a fixed Shannon entropy, as shown in Table I. In particular, we illustrate the exact feasible regions between the Shannon entropy and the Rényi entropy in Fig. 2 by using (295) and (296). In Section III-B we consider applications of Corollary 1 for a particular class of discrete memoryless channels, defined in Definition 2 which is called uniformly focusing [15] or uniform from the output [16].
II. Preliminaries

A. n-ary probability vectors and its information measures

Let the set of all \( n \)-ary probability vectors be denoted by
\[
P_n \triangleq \left\{ (p_1, p_2, \ldots, p_n) \in \mathbb{R}^n \mid p_j \geq 0 \text{ and } \sum_{i=1}^{n} p_i = 1 \right\} \tag{1}
\]
for an integer \( n \geq 2 \). For \( p = (p_1, p_2, \ldots, p_n) \in P_n \), let
\[
p[1] \geq p[2] \geq \cdots \geq p[n] \tag{2}
\]
denote the components of \( p \) in decreasing order, and let
\[
p_\downarrow \triangleq (p[1], p[2], \ldots, p[n]) \tag{3}
\]
denote the decreasing rearrangement\(^1\) of \( p \). In particular, we define the following two \( n \)-ary probability vectors: (i) an \( n \)-ary deterministic distribution
\[
d_n \triangleq (d_1, d_2, \ldots, d_n) \in P_n \tag{4}
\]
is defined by \( d_1 = 1 \) and \( d_i = 0 \) for \( i \in \{2, 3, \ldots, n\} \) and (ii) the \( n \)-ary equiprobable distribution
\[
u_n \triangleq (u_1, u_2, \ldots, u_n) \in P_n \tag{5}
\]
is defined by \( u_i = \frac{1}{n} \) for \( i \in \{1, 2, \ldots, n\} \).

For an \( n \)-ary random variable \( X \sim p \in P_n \), we define the Shannon entropy [1] as
\[
H(X) = H(p) \triangleq -\sum_{i=1}^{n} p_i \ln p_i, \tag{6}
\]
where \( \ln \) denotes the natural logarithm and assume that \( 0 \ln 0 = 0 \). Moreover, we define the \( \ell_\alpha \)-norm of \( p \in P_n \) as
\[
\|p\|_\alpha \triangleq \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}} \tag{7}
\]
for \( \alpha \in (0, \infty) \). Note that \( \lim_{\alpha \to \infty} \|p\|_\alpha = \|p\|_\infty \triangleq \max\{p_1, p_2, \ldots, p_n\} \) for \( p \in P_n \). On the works of extending Shannon entropy, the \( \ell_\alpha \)-norm is appear in the several information measures. As an instance, Rényi [9] generalized the Shannon entropy axiomatically to the Rényi entropy of order \( \alpha \in (0, 1) \cup (1, \infty) \), defined as
\[
H_\alpha(X) = H_\alpha(p) \triangleq \frac{\alpha}{1 - \alpha} \ln \|p\|_\alpha \tag{8}
\]
for \( X \sim p \in P_n \). Note that it is usually defined that \( H_1(X) \triangleq H(X) \) since \( \lim_{\alpha \to 1} H_\alpha(X) = H(X) \) by L'Hôpital's rule. In other axiomatic definitions of entropies [10]–[14], we can also define them by using the \( \ell_\alpha \)-norm, as with [8].

\(^1\)This rearrangement is denoted by reference to the notation of [7].
In this study, we analyze relations between $H(p)$ and $\|p\|_\alpha$ to examine relationships between the Shannon entropy and several information measures. Note that $H(p)$ and $\|p\|_\alpha$ are invariant for any permutation of the indices of $p \in P_n$; that is,

$$H(p) = H(p_\downarrow) \quad \text{and} \quad \|p\|_\alpha = \|p_\downarrow\|_\alpha$$

(9)

for any $p \in P_n$. Hence, we only consider $p_\downarrow$ for $p \in P_n$ in the analyses of the study. Since $\|p\|_1 = 1$ for any $p \in P_n$, we have no interest in the case $\alpha = 1$; hence, we omit the case $\alpha = 1$ in this study. Furthermore, since

$$H(p) = \ln n \iff \|p\|_\alpha = n^{\frac{1}{\alpha} - 1} \iff p = u_n,$$

(10)

$$H(p) = 0 \iff \|p\|_\alpha = 1 \iff p_\downarrow = d_n,$$

(11)

the cases $p = u_n$ and $p_\downarrow = d_n$ are trivial; thus, we also omit these cases in the analyses of this study.

B. Properties of two distributions $v_n(\cdot)$ and $w_n(\cdot)$

For a fixed $n \geq 2$, let the $n$-ary distribution $v_n(p) \triangleq (v_1(p), v_2(p), \ldots, v_n(p)) \in P_n$ be defined by

$$v_i(p) = \begin{cases} 1 - (n - 1)p & \text{if } i = 1, \\ p & \text{otherwise} \end{cases}$$

(12)

for $p \in [0, \frac{1}{n-1}]$, and let the $n$-ary distribution $w_n(p) \triangleq (w_1(p), w_2(p), \ldots, w_n(p)) \in P_n$ be defined by

$$w_i(p) = \begin{cases} p & \text{if } 1 \leq i \leq \lfloor p^{-1} \rfloor, \\ 1 - \lfloor p^{-1} \rfloor p & \text{if } i = \lfloor p^{-1} \rfloor + 1, \\ 0 & \text{otherwise} \end{cases}$$

(13)

for $p \in [\frac{1}{n}, 1]$, where $\lfloor \cdot \rfloor$ denotes the floor function. Note that $v_n(p)_\downarrow = w_n(p)$ for $p \in [\frac{1}{n}, \frac{1}{n-1}]$. In this subsection, we examine the properties of the Shannon entropies and the $\ell_\alpha$-norms for $v_n(\cdot)$ and $w_n(\cdot)$. For simplicity, we define

$$H_{v_n}(p) \triangleq H(v_n(p))$$

(14)

$$= -(1 - (n - 1)p) \ln(1 - (n - 1)p) - (n - 1)p \ln p,$$

(15)

$$H_{w_n}(p) \triangleq H(w_n(p))$$

(16)

$$= -\lfloor p^{-1} \rfloor p \ln p - (1 - \lfloor p^{-1} \rfloor p) \ln(1 - \lfloor p^{-1} \rfloor p).$$

(17)

Then, we first show the monotonicity of $H_{v_n}(p)$ with respect to $p \in [0, \frac{1}{n}]$ in the following lemma.

**Lemma 1.** $H_{v_n}(p)$ is strictly increasing for $p \in [0, \frac{1}{n}]$.

2The definition of $w_n(\cdot)$ is similar to the definition of \cite{6} Eq. (26).
Proof of Lemma 1. It is easy to see that
\[ H_{v_n}(p) = - \sum_{i=1}^{n} v_i(p) \ln v_i(p) \]  
(18)
\[ = -v_1(p) \ln v_1(p) - \sum_{i=2}^{n} v_i(p) \ln v_i(p) \]  
(19)
\[ = -(1 - (n - 1)p) \ln(1 - (n - 1)p) - \sum_{i=2}^{n} v_i(p) \ln v_i(p) \]  
(20)
\[ = -(1 - (n - 1)p) \ln(1 - (n - 1)p) - (n - 1)p \ln p. \]  
(21)

Then, the first-order derivative of \( H_{v_n}(p) \) with respect to \( p \) is
\[ \frac{\partial H_{v_n}(p)}{\partial p} = \frac{\partial}{\partial p} \left( -(n - 1)p \ln p - (1 - (n - 1)p) \ln(1 - (n - 1)p) \right) \]  
(22)
\[ = -(n - 1) \left( \frac{d}{dp} (p \ln p) \right) - \left( \frac{\partial}{\partial p} ((1 - (n - 1)p) \ln(1 - (n - 1)p)) \right) \]  
(23)
\[ = -(n - 1) \left( \ln p + 1 \right) + (n - 1) \left( \ln(1 - (n - 1)p) + 1 \right) \]  
(24)
\[ = (n - 1) \left( \ln(1 - (n - 1)p) - \ln p \right) \]  
(25)
\[ = (n - 1) \ln \frac{1 - (n - 1)p}{p}. \]  
(26)

Since \( 1 - (n - 1)p > p > 0 \) for \( p \in (0, \frac{1}{n}) \), it follows from (26) that
\[ \frac{\partial H_{v_n}(p)}{\partial p} > 0 \]  
(27)
for \( p \in (0, \frac{1}{n}) \). Note that \( H_{v_n}(p) \) is continuous for \( p \in [0, \frac{1}{n}] \) since \( \lim_{p \to \frac{1}{n}} H_{v_n}(p) = H_{v_n}(\frac{1}{n}) = \ln n \) and \( \lim_{p \to 0^+} H_{v_n}(p) = H_{v_n}(0) = 0 \) by the assumption \( 0 \ln 0 = 0 \). Therefore, \( H_{v_n}(p) \) is strictly increasing for \( p \in [0, \frac{1}{n}] \).

Lemma 1 implies the existence of the inverse function of \( H_{v_n}(p) \) for \( p \in [0, \frac{1}{n}] \). We second show the monotonicity of \( H_{w_n}(p) \) with respect to \( p \in [\frac{1}{n}, 1] \) as follows:

**Lemma 2.** \( H_{w_n}(p) \) is strictly decreasing for \( p \in [\frac{1}{n}, 1] \).

Proof of Lemma 2. For an integer \( m \in [2, n] \), assume that \( p \in [\frac{1}{m}, \frac{1}{m-1}] \). Then, note that \( \lfloor p^{-1} \rfloor = m \). It is easy to see that
\[ H_{w_n}(p) = - \sum_{i=1}^{n} w_i(p) \ln w_i(p) \]  
(28)
\[ = - \sum_{i=1}^{m} w_i(p) \ln w_i(p) - w_{m+1}(p) \ln w_{m+1}(p) - \sum_{j=m+2}^{n} w_j(p) \ln w_j(p) \]  
(29)
\[ \overset{(a)}{=} - \sum_{i=1}^{m} w_i(p) \ln w_i(p) - w_{m+1}(p) \ln w_{m+1}(p) \]  
(30)
\[ = -m \ln p - w_{m+1}(p) \ln w_{m+1}(p) \]  
(31)
\[ = -m \ln p - (1 - m \ln p) \ln(1 - m \ln p), \]  
(32)
where (a) follows by the assumption \(0 \ln 0 = 0\). Then, the first-order derivative of \(H_{w_n}(p)\) with respect to \(p\) is

\[
\frac{\partial H_{w_n}(p)}{\partial p} = \frac{\partial}{\partial p} \left( -m p \ln p - (1 - m p) \ln(1 - m p) \right)
\]

(33)

\[
= -m \left( \frac{d}{dp} (p \ln p) \right) - \left( \frac{\partial}{\partial p} ((1 - m p) \ln(1 - m p)) \right)
\]

(34)

\[
= -m \left( \ln p + 1 \right) + m \left( \ln(1 - m p) + 1 \right)
\]

(35)

\[
= m \left( \ln(1 - m p) - \ln p \right)
\]

(36)

\[
= m \ln \frac{1 - m p}{p}.
\]

(37)

Since \(p > 1 - m p > 0\) for \(p \in \left( \frac{1}{m}, \frac{1}{m-1} \right)\), it follows from (37) that

\[
\frac{\partial H_{w_n}(p)}{\partial p} < 0
\]

(38)

for \(p \in \left( \frac{1}{m}, \frac{1}{m-1} \right)\). On the other hand, we observe that

\[
\lim_{p \to \left( \frac{1}{m} \right)^-} H_{w_n}(p) = \lim_{p \to \left( \frac{1}{m} \right)^-} \left( -|p^{-1}|p \ln p - (1 - |p^{-1}p|) \ln(1 - |p^{-1}|p) \right)
\]

(39)

\[
= \lim_{p \to \left( \frac{1}{m} \right)^-} \left( -m p \ln p - (1 - m p) \ln(1 - m p) \right)
\]

(40)

\[
= \ln m - \lim_{p \to \left( \frac{1}{m} \right)^-} \left( (1 - m p) \ln(1 - m p) \right)
\]

(41)

\[
= \ln m - \lim_{x \to 0^+} \left( x \ln x \right)
\]

(42)

\[
= \ln m
\]

(43)

for an integer \(m \in [1, n - 1]\) and

\[
\lim_{p \to \left( \frac{1}{m} \right)^+} H_{w_n}(p) = \lim_{p \to \left( \frac{1}{m} \right)^+} \left( -|p^{-1}|p \ln p - (1 - |p^{-1}|p) \ln(1 - |p^{-1}|p) \right)
\]

(44)

\[
= \lim_{p \to \left( \frac{1}{m} \right)^+} \left( -m \ln p - (1 - m p) \ln(1 - m p) \right)
\]

(45)

\[
= \left( 1 - \frac{1}{m} \right) \ln m - \lim_{p \to \left( \frac{1}{m} \right)^+} \left( (1 - m p) \ln(1 - m p) \right)
\]

(46)

\[
= \left( 1 - \frac{1}{m} \right) \ln m - \left( -\frac{1}{m} \ln m \right)
\]

(47)

\[
= \ln m
\]

(48)

for an integer \(m \in [2, n]\). Note that \(H_{w_n}(\frac{1}{m}) = \ln m\) from (43) and the assumption \(0 \ln 0 = 0\). Hence, for any integer \(m \in [2, n - 1]\), we get that

\[
\lim_{p \to \left( \frac{1}{m} \right)^+} H_{w_n}(p) = H_{w_n}(\frac{1}{m}) = \ln m
\]

(49)

\[
\lim_{p \to \frac{1}{m}} H_{w_n}(p) = H_{w_n}(\frac{1}{m}) = \ln m,
\]

(50)

\[
\lim_{p \to 1^-} H_{w_n}(p) = H_{w_n}(1) = 0,
\]

(51)

which imply that \(H_{w_n}(p)\) is continuous for \(p \in \left[ \frac{1}{n}, 1 \right]\). Therefore, \(H_{w_n}(p)\) is strictly decreasing for \(p \in \left[ \frac{1}{n}, 1 \right]\). □
As with Lemma 1, Lemma 2 also implies the existence of the inverse function of $H_{\alpha_n}(p)$ for $p \in \left[\frac{1}{n}, 1\right]$. Since $H_{\alpha_n}(0) = 0$, $H_{\alpha_n}(\frac{1}{n}) = \ln n$, $H_{\alpha_n}(\frac{1}{n}) = \ln n$, and $H_{\alpha_n}(1) = 0$, we can denote the inverse functions of $H_{\alpha_n}(p)$ and $H_{\alpha_n}(p)$ with respect to $p$ as follows: We denote by $H_{\alpha_n}^{-1} : \left[0, \ln n\right] \rightarrow \left[0, \frac{1}{n}\right]$ the inverse function of $H_{\alpha_n}(p)$ for $p \in \left[0, \frac{1}{n}\right]$. Moreover, we also denote by $H_{\alpha_n}^{-1} : \left[0, \ln n\right] \rightarrow \left[\frac{1}{n}, 1\right]$ the inverse function of $H_{\alpha_n}(p)$ for $p \in \left[\frac{1}{n}, 1\right]$.

Now, we provide the monotonicity of $\|v_n(p)\|_{\alpha}$ with respect to $H_{\alpha_n}(p)$ in the following lemma.

**Lemma 3.** For any fixed $n \geq 2$ and any fixed $\alpha \in (-\infty, 0) \cup (0, 1) \cup (1, \infty)$, if $p \in \left[0, \frac{1}{n}\right]$, the following monotonicity hold:

(i) if $\alpha > 1$, then $\|v_n(p)\|_{\alpha}$ is strictly decreasing for $H_{\alpha_n}(p) \in \left[0, \ln n\right]$ and

(ii) if $\alpha < 1$, then $\|v_n(p)\|_{\alpha}$ is strictly increasing for $H_{\alpha_n}(p) \in \left[0, \ln n\right]$.

**Proof of Lemma 3.** The proof of Lemma 3 is given in a similar manner with [20 Appendix I]. By the chain rule of the derivation and the inverse function theorem, we have

\[
\frac{\partial \|v_n(p)\|_{\alpha}}{\partial H_{\alpha_n}(p)} = \left(\frac{\partial \|v_n(p)\|_{\alpha}}{\partial p}\right) \cdot \left(\frac{\partial p}{\partial H_{\alpha_n}(p)}\right): \quad (52)
\]

Direct calculation shows

\[
\frac{\partial \|v_n(p)\|_{\alpha}}{\partial p} = \frac{\partial}{\partial p} \left((n-1)p^{\alpha} + (1 - (n-1)p)^{\alpha}\right)^{\frac{1}{\alpha}} = \frac{1}{\alpha} \left((n-1)p^{\alpha} + (1 - (n-1)p)^{\alpha}\right)^{\frac{1}{\alpha} - 1} \left(\frac{\partial}{\partial p} \left((n-1)p^{\alpha} + (1 - (n-1)p)^{\alpha}\right)\right) \quad (54)
\]

\[
= \frac{1}{\alpha} \left((n-1)p^{\alpha} + (1 - (n-1)p)^{\alpha}\right)^{\frac{1}{\alpha} - 1} \left(\alpha(n-1) \left(p^{\alpha-1} - (1 - (n-1)p)^{\alpha-1}\right)\right) \quad (55)
\]

\[
= (n-1) \left((n-1)p^{\alpha} + (1 - (n-1)p)^{\alpha}\right)^{\frac{1}{\alpha} - 1} \left(p^{\alpha-1} - (1 - (n-1)p)^{\alpha-1}\right) \quad (56)
\]

Substituting (56) and (57) into (53), we obtain

\[
\frac{\partial \|v_n(p)\|_{\alpha}}{\partial H_{\alpha_n}(p)}
= (n-1) \left((n-1)p^{\alpha} + (1 - (n-1)p)^{\alpha}\right)^{\frac{1}{\alpha} - 1} \left(p^{\alpha-1} - (1 - (n-1)p)^{\alpha-1}\right) \left(\frac{1}{(n-1) \ln \frac{1-(n-1)p}{p}}\right) \quad (58)
\]

\[
= \left((n-1)p^{\alpha} + (1 - (n-1)p)^{\alpha}\right)^{\frac{1}{\alpha} - 1} \left(p^{\alpha-1} - (1 - (n-1)p)^{\alpha-1}\right) \left(\frac{1}{\ln \frac{1-(n-1)p}{p}}\right) \quad (59)
\]

We now define the sign function as

\[
\text{sgn}(x) \triangleq \begin{cases} 
1 & \text{if } x > 0, \\
0 & \text{if } x = 0, \\
-1 & \text{if } x < 0.
\end{cases}
\]

January 29, 2016
Lemma 4. For any fixed $\parallel p \parallel < 1 - (n - 1)p$ for $p \in (0, \frac{1}{n})$, we observe that

$$\text{sgn} \left( \left( (n - 1)p^\alpha + (1 - (n - 1)p)^\alpha \right)^{\frac{1}{\alpha} - 1} \right) = 1,$$

(61) for $p \in (0, \frac{1}{n})$ and $\alpha \in (-\infty, 0) \cup (0, +\infty)$; and therefore, we have

$$\text{sgn} \left( \frac{\partial \| v_n(p) \|_\alpha}{\partial H_{w_n}(p)} \right)$$

(65)

$$\text{sgn} \left( \left( (n - 1)p^\alpha + (1 - (n - 1)p)^\alpha \right)^{\frac{1}{\alpha} - 1} \left( p^{\alpha - 1} - (1 - (n - 1)p)^{\alpha - 1} \right) \frac{1}{\ln \left( \frac{1}{1 - (n - 1)p} \right)} \right) = 1$$

(64)

$$= \text{sgn} \left( \left( (n - 1)p^\alpha + (1 - (n - 1)p)^\alpha \right)^{\frac{1}{\alpha} - 1} \right) \cdot \text{sgn} \left( p^{\alpha - 1} - (1 - (n - 1)p)^{\alpha - 1} \right) \cdot \text{sgn} \left( \frac{1}{\ln \left( \frac{1}{1 - (n - 1)p} \right)} \right)$$

(65)

$$= \begin{cases} 1 & \text{if } \alpha < 1, \\ 0 & \text{if } \alpha = 1, \\ -1 & \text{if } \alpha > 1, \end{cases}$$

(66)

for $p \in (0, \frac{1}{n})$ and $\alpha \in (-\infty, 0) \cup (0, +\infty)$, which implies Lemma 3.

It follows from Lemmas 1 and 3 that, for each $\alpha \in (0, 1) \cup (1, \infty)$, $\| v_n(p) \|_\alpha$ is bijective for $p \in [0, \frac{1}{n}]$. Similarly, we also show the monotonicity of $\| w_n(p) \|_\alpha$ with respect to $H_{w_n}(p)$ in the following lemma.

Lemma 4. For any fixed $n \geq 2$ and any fixed $\alpha \in (0, 1) \cup (1, \infty)$, if $p \in [\frac{1}{n}, 1]$, the following monotonicity hold:

(i) if $\alpha > 1$, then $\| w_n(p) \|_\alpha$ is strictly decreasing for $H_{w_n}(p) \in [0, \ln n]$ and

(ii) if $\alpha < 1$, then $\| w_n(p) \|_\alpha$ is strictly increasing for $H_{w_n}(p) \in [0, \ln n]$.

Proof of Lemma 2: Since $w_n(p) = v_n(p)$ for $p \in [\frac{1}{n}, \frac{1}{n - 1}]$, we can obtain immediately from (59) that

$$\frac{\partial \| w_n(p) \|_\alpha}{\partial H_{w_n}(p)} = \left( (n - 1)p^\alpha + (1 - (n - 1)p)^\alpha \right)^{\frac{1}{\alpha} - 1} \left( p^{\alpha - 1} - (1 - (n - 1)p)^{\alpha - 1} \right) \frac{1}{\ln \left( \frac{1}{1 - (n - 1)p} \right)}$$

(67)

for $p \in (\frac{1}{n}, \frac{1}{n - 1})$. Since $0 < 1 - (n - 1)p < p \in (\frac{1}{n}, \frac{1}{n - 1})$, we observe that

$$\text{sgn} \left( \left( (n - 1)p^\alpha + (1 - (n - 1)p)^\alpha \right)^{\frac{1}{\alpha} - 1} \right) = 1,$$

(68)

$$\text{sgn} \left( p^{\alpha - 1} - (1 - (n - 1)p)^{\alpha - 1} \right) = \begin{cases} 1 & \text{if } \alpha > 1, \\ 0 & \text{if } \alpha = 1, \\ -1 & \text{if } \alpha < 1, \end{cases}$$

(69)
Theorems 1 and 2. Then, we can identify the exact feasible region of

\[
\text{sgn}\left(\frac{1}{\ln \frac{1}{1-(n-1)p}}\right) = -1
\]  

(70)

for \( p \in \left(\frac{1}{n}, \frac{1}{n-1}\right) \) and \( \alpha \in (-\infty, 0) \cup (0, +\infty) \); and therefore, we have

\[
\text{sgn}\left(\frac{\partial\|w_n(p)\|_\alpha}{\partial H_{w_n}(p)}\right) = \begin{cases} 
1 & \text{if } \alpha < 1, \\
0 & \text{if } \alpha = 1, \\
-1 & \text{if } \alpha > 1,
\end{cases}
\]  

(71)

for \( p \in \left(\frac{1}{n}, \frac{1}{n-1}\right) \) and \( \alpha \in (-\infty, 0) \cup (0, +\infty) \), as with (66). Hence, for \( \alpha \in (-\infty, 0) \cup (0, +\infty) \), we have that

- if \( \alpha > 1 \), then \( \|w_n(p)\|_\alpha \) is strictly decreasing for \( H_{w_n}(p) \in [\ln(n-1), \ln n] \) and
- if \( \alpha < 1 \), then \( \|w_n(p)\|_\alpha \) is strictly increasing for \( H_{w_n}(p) \in [\ln(n-1), \ln n] \).

Finally, since \( H_{w_m}(p) = H_{w_n}(p) \) and \( \|w_m(p)\|_\alpha = \|w_n(p)\|_\alpha \) for any integer \( m \in [2, n-1] \), any \( p \in \left[\frac{1}{m}, \frac{1}{m-1}\right] \), and any \( \alpha \in (0, 1) \cup (1, +\infty) \), we can obtain that

- if \( \alpha > 1 \), then \( \|w_n(p)\|_\alpha \) is strictly decreasing for \( H_{w_n}(p) \in [\ln(m-1), \ln m] \) and
- if \( \alpha < 1 \), then \( \|w_n(p)\|_\alpha \) is strictly increasing for \( H_{w_n}(p) \in [\ln(m-1), \ln m] \)

for any integer \( m \in [2, n] \) and any \( \alpha \in (0, 1) \cup (1, +\infty) \). This completes the proof of Lemma 4.

It also follows from Lemmas 2 and 4 that, for each \( \alpha \in (0, 1) \cup (1, +\infty) \), \( \|w_n(p)\|_\alpha \) is also bijective for \( p \in \left[\frac{1}{n}, 1\right] \).

III. RESULTS

In Section III-A, we examine the extremal relations between the Shannon entropy and the \( \ell_\alpha \)-norm, as shown in Theorems 1 and 2. Then, we can identify the exact feasible region of

\[
\mathcal{R}_n(\alpha) \triangleq \{(H(p), \|p\|_\alpha) \mid p \in \mathcal{P}_n\}
\]  

(72)

for any \( n \geq 2 \) and any \( \alpha \in (0, 1) \cup (1, +\infty) \). Extending Theorems 1 and 2 to Corollary 1, we can obtain the tight bounds between the Shannon entropy and several information measures which are determined by the \( \ell_\alpha \)-norm, as shown in Table I. In Section III-B, we apply the results of Section III-A to uniformly focusing channels of Definition 2.

A. Bounds on Shannon entropy and \( \ell_\alpha \)-norm

Let the \( \alpha \)-logarithm function \([19]\) be denoted by

\[
\ln_\alpha x \triangleq \frac{x^{1-\alpha} - 1}{1-\alpha}
\]  

(73)

for \( \alpha \neq 1 \) and \( x > 0 \); besides, since \( \lim_{\alpha \to 1} \ln_\alpha x = \ln x \) by L’Hôpital’s rule, it is defined that \( \ln_1 x \triangleq \ln x \). For the \( \alpha \)-logarithm function, we can see the following lemma.

Lemma 5. For \( \alpha < \beta \) and \( 1 \leq x \leq y \) \((y \neq 1)\), we observe that

\[
\frac{\ln_\alpha x}{\ln_\alpha y} \leq \frac{\ln_\beta x}{\ln_\beta y}
\]  

(74)

with equality if and only if \( x \in \{1, y\} \).
Proof of Lemma 5. For $1 \leq x \leq y$ ($y \neq 1$), we consider the monotonicity of $\frac{\ln x}{\ln y}$ with respect to $\alpha$. Direct calculation shows

$$\frac{\partial}{\partial \alpha} \left( \frac{\ln x}{\ln y} \right) = \frac{\partial}{\partial \alpha} \left( x^{1-\alpha} - 1 \right) \left( \frac{1}{y^{1-\alpha} - 1} \right) + (x^{1-\alpha} - 1) \left( \frac{\partial}{\partial \alpha} \left( \frac{1}{y^{1-\alpha} - 1} \right) \right) \quad (75)$$

Then, we can see that

$$\text{sgn} \left( \frac{\partial}{\partial \alpha} \left( \frac{\ln x}{\ln y} \right) \right) = \text{sgn} \left( -\frac{1}{(y^{1-\alpha} - 1)^2} \left( x^{1-\alpha}(\ln x)(y^{1-\alpha} - 1) - y^{1-\alpha}(\ln y)(x^{1-\alpha} - 1) \right) \right) \quad (81)$$

where

- the equality (a) follows from the fact that

$$\text{sgn} \left( -\frac{1}{(y^{1-\alpha} - 1)^2} \right) = -1 \quad (89)$$

for $y > 0$ ($y \neq 1$) and $\alpha \in (-\infty, 1) \cup (1, +\infty)$,

- the equality (b) follows from the fact that $x^{1-\alpha}, y^{1-\alpha} > 0$ for $\alpha \in (-\infty, +\infty)$ and $x, y > 0$, and

- the equality (c) follows by the change of variables: $a = a(x, \alpha) \triangleq x^{\alpha-1}$ and $b = b(y, \alpha) \triangleq y^{\alpha-1}$.

Then, it can be easily seen that

$$\text{sgn} \left( \frac{1}{\alpha - 1} \right) = \begin{cases} 1 & \text{if } \alpha > 1, \\ -1 & \text{if } \alpha < 1. \end{cases} \quad (90)$$
Thus, to check the sign of \( \frac{\partial}{\partial \alpha} \left( \frac{\ln \alpha}{\ln a} \right) \), we now examine the function \((b - 1) \ln a - (a - 1) \ln b\). We readily see that
\[
\left. \left( (b - 1) \ln a - (a - 1) \ln b \right) \right|_{a=1} = \left. \left( (b - 1) \ln a - (a - 1) \ln b \right) \right|_{a=b} = 0
\] (91)
for \(b > 0\). We calculate the second order derivative of \((b - 1) \ln a - (a - 1) \ln b\) with respect to \(a\) as follows:
\[
\frac{\partial^2}{\partial a^2} \left( (b - 1) \ln a - (a - 1) \ln b \right) = \frac{\partial}{\partial a} \left( \frac{\partial}{\partial a} \left( (b - 1) \ln a - (a - 1) \ln b \right) \right)
\] (92)
\[
= \frac{\partial}{\partial a} \left( \frac{b - 1}{a} - \ln b \right)
\] (93)
\[
= \left( b - 1 \right) \left( 1 - \frac{\ln a}{a} \right)
\] (94)
\[
= \frac{-b - 1}{a^2}.
\] (95)
Hence, we observe that
\[
\text{sgn} \left( \frac{\partial^2}{\partial a^2} \left( (b - 1) \ln a - (a - 1) \ln b \right) \right) = \text{sgn} \left( -\frac{b - 1}{a^2} \right)
\] (96)
for \(a > 0\), which implies that
- if \(b > 1\), then \((b - 1) \ln a - (a - 1) \ln b\) is strictly concave in \(a > 0\) and
- if \(0 < b < 1\), then \((b - 1) \ln a - (a - 1) \ln b\) is strictly convex in \(a > 0\).

Therefore, it follows from (91) that
- if \(b > 1\), then
\[
\text{sgn} \left( (b - 1) \ln a - (a - 1) \ln b \right) = \begin{cases} 1 & \text{if } 1 < a < b, \\ 0 & \text{if } a = 1 \text{ or } a = b, \\ -1 & \text{if } 0 < a < 1 \text{ or } a > b \end{cases}
\] (97)
and
- if \(0 < b < 1\), then
\[
\text{sgn} \left( (b - 1) \ln a - (a - 1) \ln b \right) = \begin{cases} 1 & \text{if } 0 < a < b \text{ or } a > 1, \\ 0 & \text{if } a = b \text{ or } a = 1, \\ -1 & \text{if } b < a < 1. \end{cases}
\] (98)

Since \(a = x^{\alpha-1}\) and \(b = y^{\alpha-1}\), note that
- if \(\alpha > 1\), then \(1 \leq a \leq b (b \neq 1)\) for \(1 \leq x \leq y (y \neq 1)\) and
- if \(\alpha < 1\), then \(0 < b \leq a \leq 1 (b \neq 1)\) for \(1 \leq x \leq y (y \neq 1)\).
Hence, we obtain

\[
\text{sgn}\left( (y^{1-\alpha} - 1) \ln x^{1-\alpha} - (x^{1-\alpha} - 1) \ln y^{1-\alpha} \right) = \begin{cases} 
1 & \text{if } 1 < x < y \text{ and } \alpha > 1, \\
0 & \text{if } x = 1 \text{ or } x = y \text{ or } \alpha = 1, \\
-1 & \text{if } 1 < y < x \text{ and } \alpha < 1 
\end{cases}
\]

(101)

for \( \alpha \in (-\infty, +\infty) \) and \( 1 \leq x \leq y \) \((y \neq 1)\). Concluding the above analyses, we have

\[
\text{sgn}\left( \frac{\partial}{\partial \alpha} \left( \frac{\ln \alpha}{\ln \alpha} \right) \right) \cdot \text{sgn}\left( (y^{1-\alpha} - 1) \ln x^{1-\alpha} - (x^{1-\alpha} - 1) \ln y^{1-\alpha} \right)
\]

\[
= \begin{cases} 
1 & \text{if } 1 < x < y, \\
0 & \text{if } x = 1 \text{ or } x = y 
\end{cases}
\]

(102)

for \( \alpha \in (-\infty, 1) \cup (1, \infty) \), where the last equality follows from (90) and (101). Note that

\[
\lim_{\alpha \to 1} \frac{\ln \alpha}{\ln \alpha} = \frac{\ln x}{\ln y}
\]

(104)

for \( x, y > 0 \) \((y \neq 1)\), which implies that \( \frac{\ln \alpha}{\ln \alpha} \) is continuous at \( \alpha = 1 \). Therefore, we have that, if \( 1 < x < y \), then \( \frac{\ln \alpha}{\ln \alpha} \) is strictly increasing for \( \alpha \in (-\infty, +\infty) \), which implies Lemma 6.

The following two lemmas have important roles in the proving Theorem 1.

**Lemma 6.** For any \( n \geq 2 \) and any \( p \in \mathcal{P}_n \), there exists \( p \in \left[0, \frac{1}{2}\right]\) such that \( H_{v_n}(p) = H(p) \) and \( \|v_n(p)\|_\alpha \geq \|p\|_\alpha \) for all \( \alpha \in (0, \infty) \).

**Proof of Lemma 6.** If \( n = 2 \), then it can be easily seen that \( p_\downarrow = v_2(p) \) for any \( p \in \mathcal{P}_2 \) and some \( p \in \left[0, \frac{1}{2}\right] \); therefore, the lemma obviously holds when \( n = 2 \). Moreover, since

\[
H(p) = \ln n \quad \iff \quad p = u_n = v_n(\frac{1}{n}),
\]

(105)

\[
H(p) = 0 \quad \iff \quad p_\downarrow = d_n = v_n(0)
\]

(106)

the lemma obviously holds if \( H(p) \in \{0, \ln n\} \). Thus, we omit the cases \( n = 2 \) and \( H(p) \in \{0, \ln n\} \) in the analyses and consider \( p \in \mathcal{P}_n \) for \( H(p) \in (0, \ln n) \). For a fixed \( n \geq 3 \) and a constant \( A \in (0, \ln n) \), we assume for \( p \in \mathcal{P}_n \) that

\[
H(p) = A
\]

(107)

For that \( p \), let \( k \in \{2, 3, \ldots, n-1\} \) be the index such that \( p_{[k-1]} > p_{[k+1]} = p_{[n]} \); namely, the index \( k \) is chosen to satisfy the following inequalities:

\[
p_{[1]} \geq p_{[2]} \geq \cdots \geq p_{[k-1]} \geq p_{[k]} \geq p_{[k+1]} \geq \cdots = p_{[n]} \quad (p_{[k-1]} > p_{[k+1]}).
\]

(108)
Since $p_1 + p_2 + \cdots + p_n = 1$, we observe that

$$\sum_{i=1}^{n} p_i = 1 \quad (109)$$

$$\implies \quad \frac{d}{dp_{[k]}} \left( \sum_{i=1}^{n} p_i \right) = \frac{d}{dp_{[k]}} (1) \quad (110)$$

$$\iff \quad \frac{d}{dp_{[k]}} \left( \sum_{i=1}^{n} p_{[i]} \right) = 0 \quad (111)$$

$$\iff \quad \frac{dp_{[k]}}{dp_{[k]}} + \sum_{i=1;i \neq k}^{n} \frac{dp_{[i]}}{dp_{[k]}} = 0 \quad (112)$$

$$\iff \quad 1 + \sum_{i=1;i \neq k}^{n} \frac{dp_{[i]}}{dp_{[k]}} = 0 \quad (113)$$

$$\iff \quad \sum_{i=1;i \neq k}^{n} \frac{dp_{[i]}}{dp_{[k]}} = -1. \quad (114)$$

In this proof, we further assume that

$$\frac{dp_{[i]}}{dp_{[k]}} = 0 \quad (115)$$

for $i \in \{2, 3, \ldots, k-1\}$ and

$$\frac{dp_{[j]}}{dp_{[k]}} = \frac{dp_{[n]}}{dp_{[k]}} \quad (116)$$

for $j \in \{k+1, k+2, \ldots, n-1\}$. By constraints (115) and (116), we get

$$\sum_{i=1}^{n} p_i = 1 \quad (117)$$

$$\iff \quad \sum_{i=1;i \neq k}^{n} \frac{dp_{[i]}}{dp_{[k]}} = -1 \quad (118)$$

$$\iff \quad \sum_{i=1}^{k-1} \frac{dp_{[i]}}{dp_{[k]}} + \sum_{j=k+1}^{n} \frac{dp_{[j]}}{dp_{[k]}} = -1 \quad (119)$$

$$\iff \quad \sum_{i=1}^{n} \frac{dp_{[i]}}{dp_{[k]}} = -1 \quad (120)$$

$$\iff \quad \frac{dp_{[1]}}{dp_{[k]}} + (n-k) \frac{dp_{[n]}}{dp_{[k]}} = -1 \quad (121)$$

$$\iff \quad \frac{dp_{[1]}}{dp_{[k]}} = -1 - (n-k) \frac{dp_{[n]}}{dp_{[k]}}, \quad (122)$$

Moreover, since $H(p) = A$, we observe that

$$- \sum_{i=1}^{n} p_i \ln p_i = A \quad (123)$$

$$\implies \quad \frac{d}{dp_{[k]}} \left( - \sum_{i=1}^{n} p_i \ln p_i \right) = \frac{d}{dp_{[k]}} (A) \quad (124)$$

$$\iff \quad \frac{d}{dp_{[k]}} \left( - \sum_{i=1}^{n} p_i \ln p_{[i]} \right) = 0 \quad (125)$$
\[\sum_{i=1}^{n} \frac{d}{dp[k]} (p[i] \ln p[i]) = 0 \quad (126)\]

\[\sum_{i=1: i \neq k}^{n} \frac{d}{dp[k]} (p[i] \ln p[i]) = 0 \quad (127)\]

\[- \frac{d}{dp[k]} (\ln p[k]) - \sum_{i=1: i \neq k}^{n} \frac{d}{dp[k]} (p[i] \ln p[i]) = 0 \quad (128)\]

\[\sum_{i=1: i \neq k}^{n} \frac{d}{dp[k]} (p[i] \ln p[i]) = \ln p[k] + 1 \quad (129)\]

\[\sum_{i=1: i \neq k}^{n} \frac{d}{dp[i]} (p[i] \ln p[i]) = \ln p[k] + 1 \quad (130)\]

\[- \sum_{i=1: i \neq k}^{n} \left( \frac{d}{dp[i]} \right) (\ln p[i] + 1) = \ln p[k] + 1 \quad (131)\]

\[- \sum_{i=1}^{k-1} \left( \frac{d}{dp[i]} \right) (\ln p[i] + 1) - \sum_{i=k+1}^{n} \left( \frac{d}{dp[i]} \right) (\ln p[i] + 1) = \ln p[k] + 1 \quad (132)\]

\[- \sum_{i=1}^{k-1} \left( \frac{d}{dp[i]} \right) (\ln p[i] + 1) - \sum_{i=k+1}^{n} \left( \frac{d}{dp[i]} \right) (\ln p[i] + 1) = \ln p[k] + 1 \quad (133)\]

\[- \sum_{i=1}^{k-1} \left( \frac{d}{dp[i]} \right) (\ln p[i] + 1) - \sum_{i=k+1}^{n} \left( \frac{d}{dp[i]} \right) (\ln p[i] + 1) = \ln p[k] + 1 \quad (134)\]

\[- \sum_{i=1}^{k-1} \left( \frac{d}{dp[i]} \right) (\ln p[i] + 1) - \sum_{i=k+1}^{n} \left( \frac{d}{dp[i]} \right) (\ln p[i] + 1) = \ln p[k] + 1 \quad (135)\]

\[- \sum_{i=1}^{k-1} \left( \frac{d}{dp[i]} \right) (\ln p[i] + 1) - \sum_{i=k+1}^{n} \left( \frac{d}{dp[i]} \right) (\ln p[i] + 1) = \ln p[k] + 1 \quad (136)\]

\[\ln p[1] + (n - k) \frac{d}{dp[k]} (\ln p[1] + 1) = \ln p[k] + 1 \quad (137)\]

\[\ln p[1] + (n - k) \frac{d}{dp[k]} (\ln p[1] + 1) = \ln p[k] + 1 \quad (138)\]

\[- (1 - (n - k) \frac{d}{dp[k]} (\ln p[1] + 1) = \ln p[k] - \ln p[1] \quad (139)\]

\[- (n - k) \frac{d}{dp[k]} (\ln p[1] - \ln p[n]) = \ln p[k] - \ln p[1] \quad (140)\]

\[- \frac{d}{dp[k]} \frac{\ln p[k] - \ln p[1]}{\ln p[1] - \ln p[n]} = \frac{1}{n - k} \left( \frac{\ln p[1] - \ln p[k]}{\ln p[1] - \ln p[n]} \right) \quad (141)\]

where the equivalence (a) follows by the chain rule. We now check the sign of the right-hand side of (141). If

\[1 > p[1] > p[k] \geq p[n] > 0,\]

then

\[0 < \frac{\ln p[1] - \ln p[k]}{\ln p[1] - \ln p[n]} < 1 \quad (142)\]

since \(0 > \ln p[1] > \ln p[k] > \ln p[n]\); therefore, we get from (141) that

\[- \frac{1}{n - k} < \frac{d}{dp[k]} \frac{\ln p[n]}{\ln p[k]} < 0 \quad (143)\]
for \(1 > p[1] > p[k] > p[n] > 0\). Note that \(n - k \geq 1\). Moreover, if \(1 > p[1] = p[k] > p[n] > 0\), then

\[
\frac{dp[n]}{dp[k]} = -\frac{1}{n - k} \left( \frac{\ln p[1] - \ln p[k]}{\ln p[1] - \ln p[n]} \right)
\]

(144)

\[
= -\frac{1}{n - k} \left( \frac{0}{\ln p[1] - \ln p[n]} \right)
\]

(145)

\[
= 0.
\]

(146)

Furthermore, if \(1 > p[1] > p[k] = p[n] > 0\), then

\[
\frac{dp[n]}{dp[k]} = -\frac{1}{n - k} \left( \frac{\ln p[1] - \ln p[k]}{\ln p[1] - \ln p[n]} \right)
\]

(147)

\[
= -\frac{1}{n - k}.
\]

(148)

Combining (143), (146), and (148), we get under the constraints (107), (108), (115), and (116) that

\[
\text{sgn} \left( \frac{dp[n]}{dp[k]} \right) = \begin{cases} 0 & \text{if } p[1] = p[k], \\ -1 & \text{otherwise} \end{cases}
\]

(149)

for \(1 > p[1] \geq p[k] \geq p[n] > 0\) \((p[1] > p[n])\). Note for the constraint (107) that

\[
\lim_{(p[k+1], p[k+2], \ldots, p[n]) \to (0^+, 0^+, \ldots, 0^+)} H(p[1], p[2], \ldots, p[n]) = H(p[1], p[2], \ldots, p[k], 0, 0, \ldots, 0)
\]

(150)

since \(\lim_{x \to 0^+} x \ln x = 0 \ln 0\) by the assumption \(0 \ln 0 = 0\). Thus, it follows from (149) that, for all \(j \in \{k + 1, k + 2, \ldots, n\}\), \(p[j]\) is strictly decreasing for \(p[k]\) under the constraints (107), (108), (115), and (116). Similarly, we check the sign of the right-hand side of (122):

\[
\frac{dp[1]}{dp[k]} = -1 - (n - k) \frac{dp[n]}{dp[k]}
\]

(151)

By (143), (146), and (148), we can see that

\[
-1 \leq \frac{dp[1]}{dp[k]} < 0
\]

(152)

for \(1 > p[1] \geq p[k] > p[n] > 0\) and

\[
\frac{dp[1]}{dp[k]} = 0
\]

(153)

for \(1 > p[1] > p[k] = p[n] > 0\); therefore, we also get under the constraints (107), (108), (115), and (116) that

\[
\text{sgn} \left( \frac{dp[1]}{dp[k]} \right) = \begin{cases} 0 & \text{if } p[k] = p[n], \\ -1 & \text{otherwise} \end{cases}
\]

(154)

for \(1 > p[1] \geq p[k] \geq p[n] > 0\) \((p[1] > p[n])\). As with (149), it follows from (154) that \(p[1]\) is strictly decreasing for \(p[k]\) under the constraints (107), (108), (115), and (116).
On the other hand, for a fixed $\alpha \in (-\infty, 1) \cup (1, +\infty)$, we have

$$\frac{d\|p\|_\alpha}{dp_{[k]}} = \frac{d}{dp_{[k]}} \left( \sum_{i=1}^{n} p_i^\alpha \right)$$

(155)

$$= \frac{1}{\alpha} \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}-1} \left( \frac{d}{dp_{[k]}} \sum_{i=1}^{n} p_i^\alpha \right)$$

(156)

$$= \frac{1}{\alpha} \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}-1} \left( \frac{d}{dp_{[k]}} \sum_{i=1}^{n} p_i^\alpha \right)$$

(157)

$$= \frac{1}{\alpha} \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}-1} \left( \sum_{i=1}^{n} \frac{d}{dp_{[k]}} (p_i^\alpha) \right)$$

(158)

$$= \frac{1}{\alpha} \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}-1} \left( \frac{d}{dp_{[k]}} (p_i^\alpha) + \sum_{i=1;i\neq k}^{n} \frac{d}{dp_{[k]}} (p_i^\alpha) \right)$$

(159)

$$= \frac{1}{\alpha} \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}-1} \left( \alpha p_i^{\alpha-1} + \sum_{i=1;i\neq k}^{n} \frac{d}{dp_{[k]}} (p_i^\alpha) \right)$$

(160)

$$= \frac{1}{\alpha} \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}-1} \left( p_{[k]}^{\alpha-1} + \sum_{i=1;i\neq k}^{n} \left( \frac{d}{dp_{[k]}} (p_i^\alpha) \right) \right)$$

(161)

$$= \frac{1}{\alpha} \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}-1} \left( \alpha p_i^{\alpha-1} + \sum_{i=1;i\neq k}^{n} \left( \frac{d}{dp_{[k]}} (p_i^\alpha) \right) \right)$$

(162)

$$= \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}-1} \left( p_{[k]}^{\alpha-1} + \sum_{i=1;i\neq k}^{n} \left( \frac{d}{dp_{[k]}} (p_i^\alpha) \right) \right)$$

(163)

$$= \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}-1} \left( p_{[k]}^{\alpha-1} + \sum_{i=1}^{k-1} \left( \frac{d}{dp_{[k]}} (p_i^\alpha) \right) \right) + \sum_{j=k+1}^{n} \left( \frac{d}{dp_{[k]}} (p_i^\alpha) \right)$$

(164)

$$= \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}-1} \left( p_{[k]}^{\alpha-1} + \sum_{i=1}^{k-1} \left( \frac{d}{dp_{[k]}} (p_i^\alpha) \right) \right) + \sum_{j=k+1}^{n} \left( \frac{d}{dp_{[k]}} (p_i^\alpha) \right)$$

(165)

$$= \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}-1} \left( p_{[k]}^{\alpha-1} + \sum_{i=1}^{k-1} \left( \frac{d}{dp_{[k]}} (p_i^\alpha) \right) \right) + \sum_{j=k+1}^{n} \left( \frac{d}{dp_{[k]}} (p_i^\alpha) \right)$$

(166)

$$= \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}-1} \left( p_{[k]}^{\alpha-1} + \sum_{i=1}^{k-1} \left( \frac{d}{dp_{[k]}} (p_i^\alpha) \right) \right) + \sum_{j=k+1}^{n} \left( \frac{d}{dp_{[k]}} (p_i^\alpha) \right)$$

(167)

$$= \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}-1} \left( p_{[k]}^{\alpha-1} + \sum_{i=1}^{k-1} \left( \frac{d}{dp_{[k]}} (p_i^\alpha) \right) \right) + \sum_{j=k+1}^{n} \left( \frac{d}{dp_{[k]}} (p_i^\alpha) \right)$$

(168)

$$= \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}-1} \left( p_{[k]}^{\alpha-1} + (n-k) \left( \frac{d}{dp_{[k]}} (p_i^\alpha) \right) \right)$$

(169)

$$= \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}-1} \left( p_{[k]}^{\alpha-1} + (n-k) \left( \frac{d}{dp_{[k]}} (p_i^\alpha) \right) \right)$$

(170)
\[
\sum_{i=1}^{n} p_i^\alpha \left( (p_{[k]}^\alpha - p_{[1]}^\alpha) + (n - k) \left( -\frac{1}{n - k} \left( \ln p_{[1]} - \ln p_{[k]} \right) \right) (p_{[n]}^\alpha - p_{[1]}^\alpha) \right)
\]

(171)

\[
= \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}} \left( p_{[k]}^\alpha - p_{[1]}^\alpha \right) + (n - k) \left( -\frac{1}{n - k} \left( \ln p_{[1]} - \ln p_{[k]} \right) \right) (p_{[n]}^\alpha - p_{[1]}^\alpha)
\]

(172)

\[
= \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}} \left( p_{[n]}^\alpha - p_{[1]}^\alpha \right) \left( (\ln p_{[1]} - \ln p_{[k]}) (p_{[n]}^\alpha - p_{[1]}^\alpha) \right) - \ln p_{[1]} \ln p_{[n]}
\]

(173)

\[
= \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}} \left( p_{[n]}^\alpha - p_{[1]}^\alpha \right) \left( (\ln p_{[1]} - \ln p_{[k]}) (p_{[n]}^\alpha - p_{[1]}^\alpha) \right) - \ln p_{[1]} \ln p_{[n]}
\]

(174)

\[
= \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}} \left( p_{[n]}^\alpha - p_{[1]}^\alpha \right) \left( \ln p_{[1]} - \ln p_{[k]} \right) + \ln p_{[1]} - \ln p_{[n]}
\]

(175)

\[
= \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}} \left( p_{[n]}^\alpha - p_{[1]}^\alpha \right) \left( \ln p_{[1]} - \ln p_{[k]} \right) + \ln p_{[1]} - \ln p_{[n]}
\]

(176)

\[
= \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}} \left( p_{[n]}^\alpha - p_{[1]}^\alpha \right) \left( \ln p_{[1]} - \ln p_{[k]} \right) + \ln p_{[1]} - \ln p_{[n]}
\]

(177)

Hence, we can see that

\[
\text{sgn} \left( \frac{d||p||_\alpha}{dp_{[k]}} \right) = \text{sgn} \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}} \left( p_{[n]}^\alpha - p_{[1]}^\alpha \right) \left( \frac{\ln p_{[1]} - \ln p_{[k]}}{\ln p_{[1]} - \ln p_{[n]}} \right)
\]

(178)

\[
= \text{sgn} \left( \sum_{i=1}^{n} p_i^\alpha \right)^{\frac{1}{\alpha}} \cdot \text{sgn} \left( p_{[n]}^\alpha - p_{[1]}^\alpha \right) \cdot \text{sgn} \left( \frac{\ln p_{[1]} - \ln p_{[k]}}{\ln p_{[1]} - \ln p_{[n]}} \right)
\]

(179)

\[
= \text{sgn} \left( p_{[n]}^\alpha - p_{[1]}^\alpha \right) \cdot \text{sgn} \left( \frac{\ln p_{[1]} - \ln p_{[k]}}{\ln p_{[1]} - \ln p_{[n]}} \right)
\]

(180)

for \( \alpha \in (-\infty, 0) \cup (0, +\infty) \). Since \( p \neq u_n \), i.e., \( p_{[1]} > p_{[n]} \), we readily see that

\[
\text{sgn} \left( p_{[n]}^\alpha - p_{[1]}^\alpha \right) = \begin{cases} 
1 & \text{if } \alpha < 1, \\
0 & \text{if } \alpha = 1, \\
-1 & \text{if } \alpha > 1.
\end{cases}
\]

(181)

Moreover, for \( 1 \leq \frac{p_{[1]}}{p_{[k]}} \leq \frac{p_{[1]}}{p_{[n]}} \neq 1 \), we observe from Lemma 5 that

\[
\text{sgn} \left( \frac{\ln p_{[1]} - \ln p_{[k]}}{\ln p_{[1]} - \ln p_{[n]}} \right) = \begin{cases} 
1 & \text{if } \alpha > 1 \text{ and } p_{[1]} > p_{[k]} > p_{[n]}, \\
0 & \text{if } \alpha = 1 \text{ or } p_{[1]} = p_{[k]} \text{ or } p_{[k]} = p_{[n]}, \\
-1 & \text{if } \alpha < 1 \text{ and } p_{[1]} > p_{[k]} > p_{[n]}.
\end{cases}
\]

(182)
Therefore, under the constraints (107), (108), (115), and (116), we have
\[
\text{sgn} \left( \frac{d\|p\|_\alpha}{dp[k]} \right) = \text{sgn} \left( \frac{p[n]}{p[k]} - 1 \right) \cdot \text{sgn} \left( \ln \frac{p[n]}{p[k]} - \ln \frac{p[n]}{p[n]} \right)
\]
\[
= \begin{cases} 0 & \text{if } \alpha = 1 \text{ or } p[1] = p[k] \text{ or } p[k] = p[n], \\ -1 & \text{if } \alpha \neq 1 \text{ and } p[1] > p[k] > p[n] \end{cases}
\]
for \( \alpha \in (-\infty, 0) \cup (0, +\infty) \), where the last equality follows from (181) and (182). Hence, we have that \( \|p\|_\alpha \) with a fixed \( \alpha \in (-\infty, 0) \cup (0, 1) \cup (1, +\infty) \) is strictly decreasing for \( p[k] \) under the constraints (107), (108), (115), and (116).

Using the above results, we now prove this lemma. If \( p[k] = p[k+1] \), then we reset the index \( k \in \{3, 4, \ldots, n-1\} \) to \( k-1 \); namely, we now choose the index \( k \in \{2, 3, \ldots, n-1\} \) to satisfy the following inequalities:
\[
p[1] \geq p[2] \geq \cdots \geq p[k-1] \geq p[k] > p[k+1] = p[k+2] = \cdots = p[n] \geq 0.
\]
(185)

Then, we consider to decrease \( p[k] \) under the constraints of (107), (108), (115), and (116). It follows from (154) that \( p[1] \) is strictly increased by according to decreasing \( p[k] \). Hence, if \( p[k] \) is decreased, then the condition \( p[1] > p[2] \) must be held. Similarly, it follows from (116) and (149) that, for all \( j \in \{k+1, k+2, \ldots, n\} \), \( p[j] \) is also strictly increased by according to decreasing \( p[k] \). Hence, if \( p[k] \) is decreased, then the condition \( p[k+1] = p[k+2] = \cdots = p[n] > 0 \) must be held. Let \( q = (q_1, q_2, \ldots, q_n) \) denote the probability vector that made from \( p \) by continuing the above operation until to satisfy \( p[k] = p[k+1] \) under the conditions of (107), (108), (115), (116), and (185). Namely, the probability vector \( q \) satisfies the following inequalities:
\[
q[1] > q[2] \geq q[3] \geq \cdots \geq q[k-1] > q[k] = q[k+1] = \cdots = q[n] > 0.
\]
(186)

Since \( q \) is made from \( p \) under the constraint (107), note that
\[
H(p) = H(q).
\]
(187)

Moreover, it follows from (184) that \( \|p\|_\alpha \) with a fixed \( \alpha \in (0, 1) \cup (1, +\infty) \) is also strictly increased by according to decreasing \( p[k] \); that is, we observe that
\[
\|p\|_\alpha \leq \|q\|_\alpha
\]
(188)
for \( \alpha \in (0, 1) \cup (1, +\infty) \). Repeating these operation until to satisfy \( k = 2 \) and \( p[k] = p[n] \), we have that
\[
H(p) = H_{w_n}(p),
\]
(189)
\[
\|p\|_\alpha \leq \|w_n(p)\|_\alpha
\]
(190)
for all \( \alpha \in (0, 1) \cup (1, +\infty) \) and some \( p \in [0, \frac{1}{n}] \). That completes the proof of Lemma 6.

**Lemma 7.** For any \( n \geq 2 \) and any \( p \in \mathcal{P}_n \), there exists \( p \in [\frac{1}{n}, 1] \) such that \( H_{w_n}(p) = H(p) \) and \( \|w_n(p)\|_\alpha \leq \|p\|_\alpha \) for all \( \alpha \in (0, \infty) \).
Proof of Lemma 7. This proof is similar to the proof of Lemma 6. If \( n = 2 \), then it can be easily seen that \( p_i = w_2(p) \) for any \( p \in P_2 \) and some \( p \in \left[ \frac{1}{2}, 1 \right] \); therefore, the lemma obviously holds when \( n = 2 \). Moreover, since

\[
H(p) = \ln n \iff p = u_n = w_n(\frac{1}{n}), \quad (191)
\]

\[
H(p) = 0 \iff p_i = d_n = w_n(1), \quad (192)
\]

the lemma obviously holds if \( H(p) \in \{0, \ln n\} \). Furthermore, if \( p_i = w_n(\frac{1}{m}) \) for an integer \( 2 \leq m \leq n - 1 \), then the lemma also obviously holds. Thus, we omit the cases \( n = 2, H(p) \in \{0, \ln n\} \), and \( p_i = w_n(\frac{1}{m}) \) in the analyses. For a fixed \( n \geq 3 \) and a constant \( A \in (0, \ln n) \), we assume for \( p \in P_n \) that

\[
H(p) = A. \quad (193)
\]

For that \( p \), let \( k, l \in \{2, 3, \ldots, n\} \ (k < l) \) be the indices such that \( p_{[1]} = p_{[k-1]} > p_{[k+1]} \) and \( p_{[l]} > p_{[l+1]} = 0 \); namely, the indices \( k, l \) are chosen to satisfy the following inequalities:

\[
p_{[1]} = \cdots = p_{[k-1]} \geq p_{[k]} \geq p_{[k+1]} \geq \cdots \geq p_{[l-1]} \geq p_{[l]} > p_{[l+1]} = \cdots = p_{[n]} = 0 \quad (p_{[k-1]} > p_{[k+1]}). \quad (194)
\]

Since \( p_1 + p_2 + \cdots + p_n = 1 \), we observe as with (114) that

\[
\sum_{i=1}^{n} p_i = 1 \implies \sum_{i=1; \, i \neq k}^{n} \frac{dp_{[i]}}{dp_{[k]}} = -1. \quad (195)
\]

In this proof, we further assume that

\[
\frac{dp_{[i]}}{dp_{[k]}} = \frac{dp_{[i]}}{dp_{[k]}} \quad (196)
\]

for \( i \in \{2, 3, \ldots, k - 1\} \),

\[
\frac{dp_{[j]}}{dp_{[k]}} = 1 \quad (197)
\]

for \( j \in \{k + 1, k + 2, \ldots, l - 1\} \), and

\[
\frac{dp_{[m]}}{dp_{[k]}} = 0 \quad (198)
\]

for \( m \in \{l + 1, l + 2, \ldots, n\} \). Note that (197) implies that, for all \( j \in \{k + 1, k + 2, \ldots, l - 1\} \), the increase/decrease rate of \( p_{[j]} \) is equivalent to the increase/decrease rate of \( p_{[k]} \). By constraints (196), (197), and (198), we get

\[
\sum_{i=1}^{n} p_i = 1 \quad (199)
\]

\[
\sum_{i=1; \, i \neq k}^{n} \frac{dp_{[i]}}{dp_{[k]}} = -1 \quad (200)
\]

\[
(k - 1) \frac{dp_{[1]}}{dp_{[k]}} + \sum_{j=k+1}^{l-1} \frac{dp_{[j]}}{dp_{[k]}} + \sum_{m=l+1}^{n} \frac{dp_{[m]}}{dp_{[k]}} = -1 \quad (201)
\]

\[
(k - 1) \frac{dp_{[1]}}{dp_{[k]}} + \sum_{j=k+1}^{l-1} \frac{dp_{[j]}}{dp_{[k]}} + \sum_{m=l+1}^{n} \frac{dp_{[m]}}{dp_{[k]}} = -1 \quad (202)
\]

\[
(k - 1) \frac{dp_{[1]}}{dp_{[k]}} + (l - k - 1) + \sum_{m=l+1}^{n} \frac{dp_{[m]}}{dp_{[k]}} = -1 \quad (203)
\]
\[ (k - 1) \frac{dp_{[1]}}{dp_{[k]}} + (l - k - 1) + \frac{dp_{[l]}}{dp_{[k]}} = -1 \]  
\[ (k - 1) \frac{dp_{[1]}}{dp_{[k]}} + \frac{dp_{[l]}}{dp_{[k]}} = -(l - k) \]  
\[ (k - 1) \frac{dp_{[1]}}{dp_{[k]}} = -(l - k) - \frac{dp_{[l]}}{dp_{[k]}} \]  
\[ \frac{dp_{[1]}}{dp_{[k]}} = \frac{-1}{k - 1} \left( (l - k) + \frac{dp_{[l]}}{dp_{[k]}} \right), \]  
where note in (207) that \( k \geq 2 \). Moreover, since \( H(p) = A \), we observe that 
\[ - \sum_{i=1}^{n} p_i \ln p_i = A \]  
\[ -(\ln p_{[1]} + 1) \sum_{i=1}^{k-1} \left( \frac{dp_{[i]}}{dp_{[k]}} \right) (\ln p_{[i]} + 1) - \sum_{j=k+1}^{l-1} \left( \frac{dp_{[j]}}{dp_{[k]}} \right) (\ln p_{[j]} + 1) = \ln p_{[k]} + 1 \]  
(a) \[ -(\ln p_{[1]} + 1) \sum_{i=1}^{k-1} \frac{dp_{[i]}}{dp_{[k]}} (\ln p_{[i]} + 1) - \sum_{j=k+1}^{l-1} \frac{dp_{[j]}}{dp_{[k]}} (\ln p_{[j]} + 1) = \ln p_{[k]} + 1 \]  
\[ -(k - 1)(\ln p_{[1]} + 1) \left( \frac{dp_{[1]}}{dp_{[k]}} \right) - \sum_{j=k+1}^{l-1} (\ln p_{[j]} + 1) = \ln p_{[k]} + 1 \]  
\[ (\ln p_{[1]} + 1) \left( (l - k) + \frac{dp_{[l]}}{dp_{[k]}} \right) - \sum_{j=k+1}^{l-1} (\ln p_{[j]} + 1) = \ln p_{[k]} + 1 \]  
where (a) follows from the fact that \( \left( \frac{dp_{[m]}}{dp_{[k]}} \right) (\ln p_{[m]} + 1) = 0 \) for \( m \in \{l + 1, l + 2, \ldots, n\} \) since \( \frac{dp_{[m]}}{dp_{[k]}} = 0 \) (see Eq. (198)), \( p_{[m]} = 0 \) (see Eq. (194)), and \( 0 \ln 0 = 0 \). Hence, under the constraints (193), (194), (196), (197), and (198), we observe that 
\[ \frac{dp_{[l]}}{dp_{[k]}} = - \frac{\sum_{j=k}^{l-1} (\ln p_{[1]} - \ln p_{[j]})}{\ln p_{[1]} - \ln p_{[l]}}. \]
We now check the sign of the right-hand side of (220). Note that
\[-(l - k) \left( \frac{\ln p(1) - \ln p(l-1)}{\ln p(1) - \ln p[l]} \right) \leq \frac{dp[l]}{dp[k]} \leq -(l - k) \left( \frac{\ln p(1) - \ln p[l]}{\ln p(1) - \ln p[l]} \right)\]  
(221)
since \(\ln p[k] \geq \ln p[j] \geq \ln p[l-1]\) for all \(j \in \{k, k+1, \ldots, l-1\}\). If \(1 > p[1] > p[k] \geq p[l] > 0\), then
\[
\frac{\ln p(1) - \ln p[l]}{\ln p(1) - \ln p[l]} > 0
\]
(222)
since \(0 \ln p[l] > \ln p[l]\); therefore, we get for the upper bound of (221) that
\[-(l - k) \left( \frac{\ln p(1) - \ln p[l]}{\ln p(1) - \ln p[l]} \right) < 0\]  
(223)
for \(1 > p[1] > p[k] \geq p[l] > 0\), where note that \(l - k \geq 1\). Moreover, if \(1 > p[1] = p[k] > p[l] > 0\), then
\[-(l - k) \left( \frac{\ln p(1) - \ln p[l]}{\ln p(1) - \ln p[l]} \right) = -(l - k) \left( \frac{0}{\ln p(1) - \ln p[l]} \right)\]  
(224)
\[= 0.\]  
(225).
Combining (223) and (225), we see that the upper bound of (221) is always nonpositive for \(1 > p[1] \geq p[k] \geq p[l] > 0\) \((p[1] > p[l])\); that is, we observe under the constraints (193), (194), (196), (197), and (198) that
\[
\text{sgn} \left( \frac{dp[l]}{dp[k]} \right) \leq \text{sgn} \left( -(l - k) \left( \frac{\ln p(1) - \ln p[l]}{\ln p(1) - \ln p[l]} \right) \right)\]  
(226)
\[= \left\{ \begin{array}{ll} 
0 & \text{if } p[1] = p[k], \\
-1 & \text{otherwise} 
\end{array} \right.\]  
(227)
for \(1 > p[1] \geq p[k] \geq p[l] > 0\) \((p[1] > p[l])\). Note for the constraint (193) that
\[
\lim_{p[l] \to 0^+} H(p[1], p[2], \ldots, p[l-1], p[l], 0, 0, \ldots, 0) = H(p[1], p[2], \ldots, p[l-1], 0, 0, \ldots, 0)\]  
(228)
since \(\lim_{x \to 0^+} x \ln x = 0 \ln 0 = 0\) by the assumption \(0 \ln 0 = 0\). Thus, it follows from (227) that \(p[l]\) is strictly decreasing for \(p[k]\) under the constraints (193), (194), (196), (197), and (198). Similarly, we check the sign of the right-hand side of (207). Substituting the lower bound of (221) into the right-hand side of (207), we observe that
\[
\frac{dp[1]}{dp[k]} \leq \frac{l - k}{k - 1} \left( 1 - \frac{\ln p(1) - \ln p[l-1]}{\ln p(1) - \ln p[l]} \right).\]  
(229)
If \(1 > p[1] \geq p[l-1] > p[l] > 0\), then
\[
\frac{\ln p(1) - \ln p[l-1]}{\ln p(1) - \ln p[l]} < 1
\]
(230)
since \(0 \ln p[l] > \ln p[l]\); therefore, we get for the upper bound of (229) that
\[
\frac{l - k}{k - 1} \left( 1 - \frac{\ln p(1) - \ln p[l-1]}{\ln p(1) - \ln p[l]} \right) < 0
\]
(231)
for \(1 > p[1] \geq p[l-1] > p[l] > 0\), where note that \(\frac{l - k}{k - 1} > 0\). Moreover, if \(1 > p[1] = p[l-1] > p[l] > 0\), then
\[
\frac{l - k}{k - 1} \left( 1 - \frac{\ln p(1) - \ln p[l-1]}{\ln p(1) - \ln p[l]} \right) = \frac{l - k}{k - 1} \left( 1 - \frac{\ln p(1) - \ln p[l]}{\ln p(1) - \ln p[l]} \right)\]  
(232)
\[= \frac{l - k}{k - 1} (1 - 1)\]  
(233)
\[= 0.\]  
(234)
It follows from (231) and (234) that the upper bound of (229) is always nonpositive for \( 1 > p_{[1]} \geq p_{[l-1]} \geq p_{[l]} > 0 \) \((p_{[1]} > p_{[l]}):\) that is, we observe under the constraints \((193), (194), (196), (197),\) and \((198)\) that

\[
\text{sgn} \left( \frac{dp_{[l]}}{dp_{[k]}} \right) \leq \text{sgn} \left( -\frac{l - k}{k - 1} \left( 1 - \frac{\ln p_{[1]} - \ln p_{[l-1]}}{\ln p_{[1]} - \ln p_{[l]}} \right) \right) \quad \text{(235)}
\]

\[
= \begin{cases} 
0 & \text{if } p_{[l-1]} = p_{[l]}, \\
-1 & \text{otherwise}
\end{cases} \quad \text{(236)}
\]

for \( 1 > p_{[1]} \geq p_{[l-1]} \geq p_{[l]} > 0 \) \((p_{[1]} > p_{[l]}).\) As with \((227),\) it follows from \((236)\) that, for all \( i \in \{1, 2, \ldots, k-1\},\)

\( p_{[i]} \) is strictly decreasing for \( p_{[k]} \) under the constraints \((193), (194), (196), (197),\) and \((198).\)

On the other hand, for a fixed \( \alpha \in (0, 1) \cup (1, +\infty),\) we have

\[
\frac{d}{dp_{[k]}} \left( \sum_{i=1}^{n} p_{i}^{\alpha} \right)^{\frac{1}{\alpha} - 1} \left( \alpha p_{[k]}^{\alpha - 1} + \sum_{i: i \neq k}^{n} \frac{d}{dp_{[i]}} (p_{i}^{\alpha}) \right) \quad \text{(237)}
\]

\[
= \frac{1}{\alpha} \sum_{i=1}^{n} p_{i}^{\alpha} \left( \alpha p_{[k]}^{\alpha - 1} + \sum_{i: i \neq k}^{l} \frac{d}{dp_{[i]}} (p_{i}^{\alpha}) \right) \quad \text{(238)}
\]

\[
\overset{(a)}{=} \frac{1}{\alpha} \left( \sum_{i=1}^{n} p_{i}^{\alpha} \right)^{\frac{1}{\alpha} - 1} \left( \alpha p_{[k]}^{\alpha - 1} + \sum_{i: i \neq k}^{l} \frac{d}{dp_{[i]}} (p_{i}^{\alpha}) \right) \quad \text{(239)}
\]

\[
= \frac{1}{\alpha} \left( \sum_{i=1}^{n} p_{i}^{\alpha} \right)^{\frac{1}{\alpha} - 1} \left( \alpha p_{[k]}^{\alpha - 1} + \sum_{i: i \neq k}^{l} \left( \frac{dp_{[i]}}{dp_{[k]}} \right) \left( \frac{1}{p_{i}^{\alpha}} \right) \right) \quad \text{(240)}
\]

\[
= \left( \sum_{i=1}^{n} p_{i}^{\alpha} \right)^{\frac{1}{\alpha} - 1} \left( p_{[k]}^{\alpha - 1} + \sum_{i: i \neq k}^{l} \left( \frac{dp_{[i]}}{dp_{[k]}} \right) \left( p_{i}^{\alpha - 1} \right) \right) \quad \text{(241)}
\]

\[
\overset{(237)}{=} \left( \sum_{i=1}^{n} p_{i}^{\alpha} \right)^{\frac{1}{\alpha} - 1} \left( \left( \frac{dp_{[i]}}{dp_{[k]}} \right) \left( p_{i}^{\alpha - 1} \right) \right) \quad \text{(242)}
\]

\[
\overset{(244)}{=} \left( \sum_{i=1}^{n} p_{i}^{\alpha} \right)^{\frac{1}{\alpha} - 1} \left( p_{[k]}^{\alpha - 1} + \sum_{i=1}^{k-1} \left( \frac{dp_{[i]}}{dp_{[k]}} \right) \left( p_{i}^{\alpha - 1} \right) + \sum_{j=k+1}^{l-1} \left( \frac{dp_{[j]}}{dp_{[k]}} \right) \left( p_{j}^{\alpha - 1} \right) \right) \quad \text{(243)}
\]

\[
\overset{(246)}{=} \left( \sum_{i=1}^{n} p_{i}^{\alpha} \right)^{\frac{1}{\alpha} - 1} \left( \left( \frac{dp_{[i]}}{dp_{[k]}} \right) \left( p_{i}^{\alpha - 1} \right) \right) \quad \text{(244)}
\]

\[
\overset{(248)}{=} \left( \sum_{i=1}^{n} p_{i}^{\alpha} \right)^{\frac{1}{\alpha} - 1} \left( l - k \right) \left( \frac{dp_{[i]}}{dp_{[k]}} \right) \left( p_{i}^{\alpha - 1} \right) \quad \text{(245)}
\]

\[
\overset{(247)}{=} \left( \sum_{i=1}^{n} p_{i}^{\alpha} \right)^{\frac{1}{\alpha} - 1} \left( \left( \frac{dp_{[i]}}{dp_{[k]}} \right) \left( p_{i}^{\alpha - 1} \right) \right) \quad \text{(246)}
\]

\[
= \left( \sum_{i=1}^{n} p_{i}^{\alpha} \right)^{\frac{1}{\alpha} - 1} \left( \left( \frac{dp_{[i]}}{dp_{[k]}} \right) \left( p_{i}^{\alpha - 1} \right) \right) \quad \text{(247)}
\]
where (a) holds since the constraint (198) implies that \( p_{[\alpha]} \) is constant for \( \alpha \). Hence, we can see that

\[
\text{sgn} \left( \frac{d\|p\|_{\alpha}}{dp_{[\alpha]}} \right) = \text{sgn} \left( \sum_{i=1}^{n} p_{i}^{\alpha - 1} \left( p_{[\alpha]}^{\alpha - 1} - p_{[\alpha]}^{-1} \right) \sum_{j=k}^{l-1} \left( \frac{\ln p_{[\alpha]}^{\alpha - 1}}{\ln p_{[\alpha]}^{-1}} - \frac{\ln p_{[\alpha]}^{-1}}{\ln p_{[\alpha]}^{\alpha - 1}} \right) \right)
\]

\[
= \text{sgn} \left( \sum_{i=1}^{n} p_{i}^{\alpha - 1} \left( p_{[\alpha]}^{\alpha - 1} - p_{[\alpha]}^{-1} \right) \sum_{j=k}^{l-1} \left( \frac{\ln p_{[\alpha]}^{\alpha - 1}}{\ln p_{[\alpha]}^{-1}} - \frac{\ln p_{[\alpha]}^{-1}}{\ln p_{[\alpha]}^{\alpha - 1}} \right) \right)
\]

\[
= \text{sgn} \left( p_{[\alpha]}^{\alpha - 1} - p_{[\alpha]}^{-1} \right) \cdot \text{sgn} \left( \sum_{j=k}^{l-1} \left( \frac{\ln p_{[\alpha]}^{\alpha - 1}}{\ln p_{[\alpha]}^{-1}} - \frac{\ln p_{[\alpha]}^{-1}}{\ln p_{[\alpha]}^{\alpha - 1}} \right) \right)
\]

for \( \alpha \in (0, 1) \cup (1, \infty) \). As with (181), we readily see that

\[
\text{sgn} \left( p_{[\alpha]}^{\alpha - 1} - p_{[\alpha]}^{-1} \right) = \begin{cases} 
1 & \text{if } \alpha < 1, \\
0 & \text{if } \alpha = 1, \\
-1 & \text{if } \alpha > 1
\end{cases}
\]

for \( p_{[\alpha]} > p_{[\alpha]}^{-1} > 0 \). Moreover, since \( 1 \leq \frac{p_{[\alpha]}}{p_{[\alpha]}} \leq \frac{p_{[\alpha]}}{p_{[\alpha]}} (\frac{p_{[\alpha]}}{p_{[\alpha]}} \neq 1) \) for \( j \in \{k, k+1, \ldots, l-1\} \), we observe from Lemma 5 that

\[
\text{sgn} \left( \frac{\ln p_{[\alpha]}^{\alpha - 1}}{\ln p_{[\alpha]}^{-1}} - \frac{\ln p_{[\alpha]}^{-1}}{\ln p_{[\alpha]}^{\alpha - 1}} \right) = \begin{cases} 
1 & \text{if } \alpha > 1 \text{ and } p_{[\alpha]} > p_{[\alpha]^{-1}} > 0, \\
0 & \text{if } \alpha = 1 \text{ or } p_{[\alpha]} = p_{[\alpha]^{-1}} \text{ or } p_{[\alpha]} = p_{[\alpha]^{-1}}, \\
-1 & \text{if } \alpha < 1 \text{ and } p_{[\alpha]} > p_{[\alpha]^{-1}} > 0.
\end{cases}
\]
for $j \in \{k, k+1, \ldots, l-1\}$; and therefore, we have

$$
\text{sgn} \left( \sum_{j=k}^{l-1} \left( \frac{\ln_{\alpha} \frac{p_{[1]}^{(j)}}{p_{[j]}} - \ln_{\alpha} \frac{p_{[1]}^{(j)}}{p_{[j]}}} \right) \right) = \begin{cases} 
1 & \text{if } \alpha > 1 \text{ and } (p_{[1]}^{(j)} > p_{[k]} \geq p_{[l]} \text{ or } p_{[1]}^{(j)} \geq p_{[k]} > p_{[l]}), \\
0 & \text{if } \alpha = 1 \text{ or } (p_{[1]}^{(j)} = p_{[k]} \text{ and } p_{[k]}^{(j)} = p_{[l]} \text{ or } p_{[1]} = p_{[l]}), \\
-1 & \text{if } \alpha < 1 \text{ and } (p_{[1]}^{(j)} > p_{[k]} \geq p_{[l]} \text{ or } p_{[1]} \geq p_{[k]} > p_{[l]}),
\end{cases}
$$

(260)

for $p \in \mathcal{P}_n$ under the constraint (194). Therefore, under the constraints (193), (194), (196), (197), and (198), we obtain

$$
\text{sgn} \left( \frac{\partial \|p\|_{\alpha}}{\partial p_{[k]}} \right) \geq 0
\quad \text{sgn} \left( p_{[k]}^{\alpha-1} - p_{[1]}^{\alpha-1} \right) \cdot \text{sgn} \left( \sum_{j=k}^{l-1} \left( \frac{\ln_{\alpha} \frac{p_{[1]}^{(j)}}{p_{[j]}} - \ln_{\alpha} \frac{p_{[1]}^{(j)}}{p_{[j]}}} \right) \right)
$$

(261)

for $\alpha \in (0, 1) \cup (1, +\infty)$, where the last equality follows from (258) and (260). Hence, we have that $\|p\|_{\alpha}$ with a fixed $\alpha \in (0, 1) \cup (1, +\infty)$ is strictly decreasing for $p_{[k]}$ under the constraints (193), (194), (196), (197), and (198).

Using the above results, we now prove this lemma. Note that, if $p_{[k-1]} = p_{[k]}$ and $k = l - 1$, then $p_* = w_n(p)$ for some $p \in [\frac{1}{n}, 1]$. If $p_{[k-1]} = p_{[k]}$ and $k < l - 1$, then we reset the index $k \in \{2, 3, \ldots, n-2\}$ to $k+1$; namely, we now choose the indices $k, l \in \{2, 3, \ldots, n\}$ ($k < l$) to satisfy the following inequalities:

$$
p_{[1]} = p_{[2]} = \cdots = p_{[k-1]} > p_{[k]} \geq p_{[k+1]} \geq \cdots \geq p_{[l-1]} \geq p_{[l]} > p_{[l+1]} = p_{[l+2]} = \cdots = p_{[n]} = 0.
$$

(263)

Then, we consider to increase $p_{[k]}$ under the constraints of (193), (194), (196), (197), and (198). Note that the constraint (197) implies that, for all $j \in \{k+1, k+2, \ldots, l-1\}$, $p_{[j]}$ is strictly increased with the same speed of increasing $p_{[k]}$. It follows from (196) and (260) that, for all $i \in \{1, 2, \ldots, k - 1\}$, $p_{[i]}$ is strictly decreased by according to increasing $p_{[k]}$. Hence, if $p_{[k]}$ is decreased, then there is a possibility that $p_{[1]} = \cdots = p_{[k-1]} = p_{[k]}$. Similarly, it follows from (227) that $p_{[l]}$ is also strictly decreased by according to increasing $p_{[k]}$. Hence, if $p_{[k]}$ is decreased, then there is a possibility that $p_{[l]} = p_{[l+1]} = \cdots = p_{[n]} = 0$. Let $q = (q_1, q_2, \ldots, q_n)$ denotes the probability vector that made from $p$ by continuing the above operation until to satisfy $p_{[1]} = p_{[k]}$ or $p_{[l]} = 0$ under the conditions of (193), (196), (197), and (198). Namely, the probability vector $q$ satisfies either

$$
q_{[1]} = q_{[2]} = \cdots = q_{[k-1]} = q_{[k]} = q_{[k+1]} \geq q_{[l-1]} > q_{[l]} \geq q_{[l+1]} = q_{[l+2]} = \cdots = q_{[n]} = 0
$$

(264)

or

$$
q_{[1]} = q_{[2]} = \cdots = q_{[k-1]} \geq q_{[k]} \geq q_{[k+1]} \geq \cdots \geq q_{[l-1]} > q_{[l]} = q_{[l+1]} = q_{[l+2]} = \cdots = q_{[n]} = 0.
$$

(265)

Note that there is a possibility that both of (264) and (265) hold; that is,

$$
q_{[1]} = q_{[2]} = \cdots = q_{[k-1]} = q_{[k]} \geq q_{[k+1]} \geq \cdots \geq q_{[l-1]} > q_{[l]} = q_{[l+1]} = q_{[l+2]} = \cdots = q_{[n]} = 0
$$

(266)

holds. Since $q$ is made under the constraint (193), note that

$$
H(p) = H(q).
$$

(267)
Moreover, it follows from (262) that \( \|p\|_\alpha \) with a fixed \( \alpha \in (0, 1) \cup (1, +\infty) \) is also strictly decreased by according to increasing \( p_{[k]} \); therefore, we observe that
\[
\|p\|_\alpha \geq \|q\|_\alpha \tag{268}
\]
for \( \alpha \in (0, 1) \cup (1, +\infty) \). Repeating these operation until to satisfy \( k = l - 1 \) and \( p_{[l]} = p_{[l]} > p_{[l]} \geq p_{[l-1]} = p_{[n]} = 0 \), we have that
\[
H(p) = H_{w_n}(p), \tag{269}
\]
\[
\|p\|_\alpha \geq \|w_n(p)\|_\alpha \tag{270}
\]
for all \( \alpha \in (0, 1) \cup (1, +\infty) \) and some \( p \in [\frac{1}{n}, 1] \). That completes the proof of Lemma 7. □

Lemmas 6 and 7 are derived by using Lemma 5. Lemmas 6 and 7 imply that the distributions \( v_n(\cdot) \) and \( w_n(\cdot) \) have extremal properties in the sense of a relation between the Shannon entropy and the \( \ell_\alpha \)-norm. Then, we can derive tight bounds of \( \ell_\alpha \)-norms with a fixed Shannon entropy as follows:

**Theorem 1.** Let \( \tilde{v}_n(p) \triangleq v_n(H_{v_n}^{-1}(H(p))) \) and \( \tilde{w}_n(p) \triangleq w_n(H_{w_n}^{-1}(H(p))) \) for \( p \in \mathcal{P}_n \). Then, we observe that
\[
\|w_n(p)\|_\alpha \leq \|p\|_\alpha \leq \|\tilde{v}_n(p)\|_\alpha \tag{271}
\]
for any \( n \geq 2 \), any \( p \in \mathcal{P}_n \), and any \( \alpha \in (0, \infty) \).

**Proof of Theorem 7.** It follows from Lemmas 6 and 7 that, for any \( n \geq 2 \) and any \( p \in \mathcal{P}_n \), there exist \( p \in [0, \frac{1}{n}] \) and \( p' \in [\frac{1}{n}, 1] \) such that
\[
H_{w_n}(p') = H(p) = H_{v_n}(p), \tag{272}
\]
\[
\|w_n(p')\|_\alpha \leq \|p\|_\alpha \leq \|v_n(p)\|_\alpha \tag{273}
\]
for all \( \alpha \in (0, +\infty) \). Then, we now consider \( q, q' \in \mathcal{P}_n \) such that
\[
H(q') = H_{w_n}(p') = H_{v_n}(p) = H(q), \tag{274}
\]
\[
\|q'\|_\alpha \leq \|w_n(p')\|_\alpha \leq \|v_n(p)\|_\alpha \leq \|q\|_\alpha \tag{275}
\]
for \( \alpha \in (0, +\infty) \). It also follows from Lemmas 6 and 7 that there exist \( q \in [0, \frac{1}{n}] \) and \( q' \in [\frac{1}{n}, 1] \) such that
\[
H_{w_n}(q') = H(q') = H(q) = H_{v_n}(q), \tag{276}
\]
\[
\|w_n(q')\|_\alpha \leq \|q'\|_\alpha \leq \|q\|_\alpha \leq \|v_n(q)\|_\alpha \tag{277}
\]
for \( \alpha \in (0, +\infty) \). Note from (274) and (276) that
\[
H_{v_n}(p) = H_{v_n}(q), \tag{278}
\]
\[
H_{w_n}(p') = H_{w_n}(q'). \tag{279}
\]
Note that it follows from Lemmas 1 and 2 that $H_{v_n}(p)$ and $H_{w_n}(p')$ are both bijective functions of $p \in [0, \frac{1}{n}]$ and $p' \in [\frac{1}{n}, 1]$, respectively. Therefore, we get

$$p = q, \quad (280)$$

$$p' = q', \quad (281)$$

which imply that, for $q$ and $q'$, the following equalities must be held:

$$\|v_n(p)\|_\alpha = \|q\|_\alpha = \|v_n(q)\|_\alpha, \quad (282)$$

$$\|w_n(p')\|_\alpha = \|q'\|_\alpha = \|w_n(q')\|_\alpha. \quad (283)$$

That completes the proof of Theorem 1. ■

Note that the distributions $\bar{v}_n(p)$ and $\bar{w}_n(p)$ denote $v_n(p)$ and $w_n(q)$, respectively, such that $H_{v_n}(p) = H_{w_n}(q) = H(p)$ for a given $p \in \mathcal{P}_n$. Theorem 1 shows that, among all $n$-ary probability vectors with a fixed Shannon entropy, the distributions $v_n(\cdot)$ and $w_n(\cdot)$ take the maximum and the minimum $\ell_\alpha$-norm, respectively. Thus, the bounds (271) of Theorem 1 are tight in the sense of the existences of the distributions $v_n(\cdot)$ and $w_n(\cdot)$ which attain both equalities of the bounds (271). In other words, Theorem 1 implies that the boundaries of $\mathcal{R}_n(\alpha)$ defined in (72), can be attained by $v_n(\cdot)$ and $w_n(\cdot)$. We illustrate the graphs of the boundaries of $\mathcal{R}_n(\alpha)$ in Fig. 1. Note that $\|\bar{v}_2(p)\|_\alpha = \|\bar{w}_2(p)\|_\alpha$ for any $p \in \mathcal{P}_2$ and any $\alpha \in (0, \infty)$ since $v_2(p) = w_2(1 - p)$ for $p \in [0, \frac{1}{2}]$. Therefore, Theorem 1 becomes meaningful for $n \geq 3$.

On the other hand, the following theorem shows that, among all $n$-ary probability vectors with a fixed $\ell_\alpha$-norm, the distributions $v_n(\cdot)$ and $w_n(\cdot)$ also take the extreme values of the Shannon entropy.

**Theorem 2.** Let $p \in [0, \frac{1}{n}]$ and $p' \in [\frac{1}{n}, 1]$ be chosen to satisfy

$$\|v_n(p)\|_\alpha = \|p\|_\alpha = \|w_n(p')\|_\alpha \quad (284)$$

for a fixed $\alpha \in (0, 1) \cup (1, \infty)$. Then, we observe that

$$0 < \alpha < 1 \implies H_{v_n}(p) \leq H(p) \leq H_{w_n}(p'), \quad (285)$$

$$\alpha > 1 \implies H_{w_n}(p') \leq H(p) \leq H_{v_n}(p) \quad (286)$$

for any $n \geq 2$ and any $p \in \mathcal{P}_n$.

**Proof of Theorem 2.** From Theorem 1 for a fixed $n \geq 2$, we consider $p \in \mathcal{P}_n$, $p \in [0, \frac{1}{n}]$, and $p' \in [\frac{1}{n}, 1]$ such that

$$H_{w_n}(p') = H(p) = H_{v_n}(p), \quad (287)$$

$$\|w_n(p')\|_\alpha \leq \|p\|_\alpha \leq \|v_n(p)\|_\alpha \quad (288)$$

for $\alpha \in (0, 1) \cup (1, +\infty)$. Note that $p$ and $p'$ are uniquely determined for a given $p \in \mathcal{P}_n$. It follows from Lemmas 1 and 2 that $H_{v_n}(p) \in [0, \ln n]$ and $H_{w_n}(p') \in [0, \ln n]$ are strictly increasing for $p \in [0, \frac{1}{n}]$ and strictly decreasing for $p' \in [\frac{1}{n}, 1]$, respectively. Moreover, it follows from Lemmas 3 and 4 that, if $\alpha \in (0, 1)$, then $\|v_n(p)\|_\alpha$ and
Let $\ell$-information measures, which are related to distributions $v$.

Theorems 1 and 2 show that extremality between the Shannon entropy and the $\ell$-norm can be attained by the distributions $v_n(\cdot)$ and $w_n(\cdot)$.

\[\|w_n(p')\|_\alpha\] are strictly increasing for $p \in [0, \frac{1}{n}]$ and strictly decreasing for $p' \in [\frac{1}{n}, 1]$, respectively. Therefore, decreasing both $p \in [0, \frac{1}{n}]$ and $p' \in [\frac{1}{n}, 1]$, we can obtain $q \in [0, \frac{1}{n}]$ and $q' \in [\frac{1}{n}, 1]$ such that

\[H_{w_n}(q') \geq H(p) \geq H_{v_n}(q),\]

\[\|w_n(q')\|_\alpha = \|p\|_\alpha = \|v_n(q)\|_\alpha\] (289)

for a fixed $\alpha \in (0, 1)$.

On the other hand, it follows from Lemmas 3 and 4 that, if $\alpha \in (1, +\infty)$, then $\|v_n(p)\|_\alpha$ and $\|w_n(p')\|_\alpha$ are strictly decreasing for $p \in [0, \frac{1}{n}]$ and strictly increasing for $p \in [\frac{1}{n}, 1]$, respectively. Therefore, increasing both $p \in [0, \frac{1}{n}]$ and $p' \in [\frac{1}{n}, 1]$, we can obtain $q \in [0, \frac{1}{n}]$ and $q' \in [\frac{1}{n}, 1]$ such that

\[H_{w_n}(q') \leq H(p) \leq H_{v_n}(q),\]

\[\|w_n(q')\|_\alpha = \|p\|_\alpha = \|v_n(q)\|_\alpha\] (291)

for a fixed $\alpha \in (1, +\infty)$.

Finally, we note that the strict monotonicity of Lemmas 3 and 4 prove the uniquenesses of the values $q \in [0, \frac{1}{n}]$ and $q' \in [\frac{1}{n}, 1]$. In fact, it follows from Lemmas 1, 2, and 4 that, for a fixed $n \geq 2$ and a fixed $\alpha \in (0, 1) \cup (1, +\infty)$, $\|v_n(p)\|_\alpha$ and $\|w_n(p')\|_\alpha$ are both bijective function of $p \in [0, \frac{1}{n}]$ and $p' \in [\frac{1}{n}, 1]$, respectively. That completes the proof of Theorem 2.

In Theorem 2, note that the values $p \in [0, \frac{1}{n}]$ and $p' \in [\frac{1}{n}, 1]$ are uniquely determined by the value of $\|p\|_\alpha$.

Theorems 1 and 2 show that extremality between the Shannon entropy and the $\ell_\alpha$-norm can be attained by the distributions $v_n(\cdot)$ and $w_n(\cdot)$.

Following a same manner with [20, Theorem 2], we extend the bounds of Theorem 1 from the $\ell_\alpha$-norm to several information measures, which are related to $\ell_\alpha$-norm, as follows:

**Corollary 1.** Let $f(\cdot)$ be a strictly monotonic function. Then, we observe that: (i) if $f(\cdot)$ is strictly increasing, then

\[f(\|w_n(p)\|_\alpha) \leq f(\|p\|_\alpha) \leq f(\|v_n(p)\|_\alpha)\] (293)
and (ii) if $f(\cdot)$ is strictly decreasing, then
\[
 f(\|\tilde{v}_n(p)\|_\alpha) \leq f(\|p\|_\alpha) \leq f(\|\tilde{w}_n(p)\|_\alpha)
\] (294)
for any $n \geq 2$, any $p \in \mathcal{P}_n$, and any $\alpha \in (0, \infty)$.

Proof of Corollary 1 Since any strictly increasing function $f(\cdot)$ satisfies $f(x) < f(y)$ for $x < y$, it is easy to see that (293) from (271) of Theorem 1. Similarly, since any strictly decreasing function $f(\cdot)$ satisfies $f(x) > f(y)$ for $x < y$, it is also easy to see that (294) from (271) of Theorem 1.

Therefore, we can obtain tight bounds of several information measures, which are determined by $\ell_\alpha$-norm, with a fixed Shannon entropy. As an instance, we introduce the application of Corollary 1 to the Rényi entropy as follows: Let $f_\alpha(x) = \frac{\alpha}{\alpha - 1} \ln x$. Then, we readily see that $H_\alpha(p) = f_\alpha(\|p\|_\alpha)$. It can be easily seen that $f_\alpha(x)$ is strictly increasing for $x \geq 0$ when $\alpha \in (0, 1)$ and strictly decreasing for $x \geq 0$ when $\alpha \in (1, \infty)$. Hence, it follows from Corollary 1 that
\[
 0 < \alpha < 1 \implies H_\alpha(\tilde{w}_n(p)) \leq H_\alpha(p) \leq H_\alpha(\tilde{v}_n(p))
\] (295)
\[
 \alpha > 1 \implies H_\alpha(\tilde{v}_n(p)) \leq H_\alpha(p) \leq H_\alpha(\tilde{w}_n(p))
\] (296)
for any $n \geq 2$ and any $p \in \mathcal{P}_n$. Moreover, if $p \in [0, \frac{1}{2}]$ and $p' \in [\frac{1}{2}, 1]$ are chosen to satisfy $H_\alpha(p) = H_\alpha(\tilde{v}_n(p)) = H_\alpha(\tilde{w}_n(p'))$ for a fixed $\alpha \in (0, 1) \cup (1, \infty)$, then (285) and (286) hold for any $n \geq 2$ and any $p \in \mathcal{P}_n$. These bounds between the Shannon entropy and the Rényi entropy imply the boundary of the region $\mathcal{R}_n^{\text{Rényi}}(\alpha) \triangleq \{(H(p), H_\alpha(p)) \mid p \in \mathcal{P}_n\}$ for any $n \geq 2$ and any $\alpha \in (0, 1) \cup (1, \infty)$. We illustrate the boundaries of $\mathcal{R}_n^{\text{Rényi}}(\alpha)$ in Fig. 2. Similarly, we can apply Corollary 1 to several entropies as shown in Table 1 and we illustrate these exact feasible regions in Figs. 3-5.

Remark 1. Harremoës and Topsøe [8] showed that the exact region of $\Delta_n = \{(H(p), IC(p)) \mid p \in \mathcal{P}_n\}$ for $n \geq 3$, where $IC(p) \triangleq \|p\|^2_2$ denotes the index of coincidence. Then, we can see that Corollary 1 contains its result by $f(x) = x^2$.

B. Applications for uniformly focusing channels

In this subsection, we consider applications of Corollary 1 for a particular class of discrete memoryless channels (DMCs), i.e., uniformly focusing channels [15]. Let the Rényi divergence [9] of order $\alpha \in (0, 1) \cup (1, \infty)$ is denoted

| Table 1: Applications of Corollary 1 |
|--------------------------------------|
| Entropies                      | function $f_\alpha(\cdot)$ | monotonicity $(0 < t < 1)$ | monotonicity $(t > 1)$ |
|----------------------------------|-----------------------------|-----------------------------|-------------------------|
| Rényi entropy $H_\alpha(p) = f_\alpha(\|p\|_\alpha)$ | $f_\alpha(x) = \frac{x}{1-x} \ln x$ | strictly increasing for $x > 0$ | strictly decreasing for $x > 0$ |
| Tsallis entropy $S_q(p) = f_q(\|p\|_q)$ | $f_q(x) = \frac{1}{q} (x^q - 1)$ | strictly increasing for $x > 0$ | strictly decreasing for $x > 0$ |
| Entropy of type-β $H_\beta(p) = f_\beta(\|p\|_\beta)$ | $f_\beta(x) = \frac{1}{\beta} (x^\beta - 1)$ | strictly increasing for $x > 0$ | strictly decreasing for $x > 0$ |
| $\gamma$-entropy $H_\gamma(p) = f_\gamma(\|p\|_{1/\gamma})$ | $f_\gamma(x) = \frac{1}{1-\gamma} (1-x)$ | strictly decreasing for $x > 0$ | strictly decreasing for $x > 0$ |
| The R-norm information $H_R(p) = f_R(\|p\|_R)$ | $f_R(x) = \frac{1}{R-1} (1-x)$ | strictly decreasing for $x > 0$ | strictly decreasing for $x > 0$ |

January 29, 2016
Fig. 2. Plots of the boundaries of $R_n^{\text{Renyi}}(\alpha)$ with $n = 6$. If $0 < \alpha < 1$, then the upper- and lower-boundaries correspond to distributions $v_n(.)$ and $w_n(.)$, respectively. If $\alpha > 1$, then these correspondences are reversed.

Fig. 3. Plots of the boundaries of $\{(H(p),S_q(p)) \mid p \in \mathcal{P}_n\}$ with $n = 6$. If $0 < \alpha < 1$, then the upper- and lower-boundaries correspond to distributions $v_n(.)$ and $w_n(.)$, respectively. If $\alpha > 1$, then these correspondences are reversed.

Fig. 4. Plots of the boundaries of $\{(H(p),H_\beta(p)) \mid p \in \mathcal{P}_n\}$ with $n = 6$. If $0 < \alpha < 1$, then the upper- and lower-boundaries correspond to distributions $v_n(.)$ and $w_n(.)$, respectively. If $\alpha > 1$, then these correspondences are reversed.

by

$$D_\alpha(p \parallel q) \triangleq \frac{1}{\alpha - 1} \ln \sum_{i=1}^n p_i^\alpha q_i^{1-\alpha},$$

(297)
(a) The case $\gamma = \frac{1}{2}$.

(b) The case $\gamma = 2$.

Fig. 5. Plots of the boundaries of $\{(H(p), H\gamma(p)) \mid p \in P_n\}$ with $n = 6$. If $0 < \alpha < 1$, then the upper- and lower-boundaries correspond to distributions $v_n(\cdot)$ and $w_n(\cdot)$, respectively. If $\alpha > 1$, then these correspondences are reversed.

(a) The case $R = \frac{1}{2}$.

(b) The case $R = 2$.

Fig. 6. Plots of the boundaries of $\{(H(p), H_R(p)) \mid p \in P_n\}$ with $n = 6$. If $0 < \alpha < 1$, then the upper- and lower-boundaries correspond to distributions $v_n(\cdot)$ and $w_n(\cdot)$, respectively. If $\alpha > 1$, then these correspondences are reversed.

for $p, q \in P_n$. Since $\lim_{\alpha \to 1} D_\alpha(p \parallel q) = D(p \parallel q)$ by L’Hôpital’s rule, we write $D_1(p \parallel q) \triangleq D(p \parallel q)$, where

$$D(p \parallel q) \triangleq \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i}$$

(298)

denotes the relative entropy. Since

$$D_\alpha(p \parallel u_n) = \ln n - H_\alpha(p)$$

(299)

for $\alpha \in (0, \infty)$, we can obtain Corollary 2 from (295) and (296).

Corollary 2. If $0 < \alpha < 1$, then

$$D_\alpha(\bar{v}_n(p) \parallel u_n) \leq D_\alpha(p \parallel u_n) \leq D_\alpha(\bar{w}_n(p) \parallel u_n)$$

(300)

for any $n \geq 2$ and any $p \in P_n$. Moreover, if $\alpha > 1$, then

$$D_\alpha(\bar{w}_n(p) \parallel u_n) \leq D_\alpha(p \parallel u_n) \leq D_\alpha(\bar{v}_n(p) \parallel u_n)$$

(301)

for any $n \geq 2$ and any $p \in P_n$. 

January 29, 2016 DRAFT
D\alpha (p \parallel u_n) \ [\text{nats}] = D(p \parallel u_n) + n (1 - \alpha) H(p)

(a) The case \alpha = \frac{1}{2}.

(b) The case \alpha = 2.

Fig. 7. Plots of the boundaries of \{(D(p \parallel u_n), D_{\alpha}(p \parallel u_n)) \mid p \in \mathcal{P}_n\} with n = 6. If 0 < \alpha < 1, then the upper- and lower-boundaries correspond to distributions \(w_n(\cdot)\) and \(v_n(\cdot)\), respectively. If \alpha > 1, then these correspondences are reversed.

Since \(D(p \parallel u_n) = \ln n - H(p)\), we note that Corollary 2 shows the tight bounds of Rényi divergence from a uniform distribution with a fixed relative entropy from a uniform distribution. Namely, Corollary 2 implies the boundary of

\[\{(D(p \parallel u_n), D_{\alpha}(p \parallel u_n)) \mid p \in \mathcal{P}_n\}\]  

for any \(n \geq 2\) and any \(\alpha \in (0, 1) \cup (1, \infty)\). We illustrate boundaries of its region in Fig. 7.

We now define DMCs as follows: Let the discrete random variables \(X \in \mathcal{X}\) and \(Y \in \mathcal{Y}\) denote the input and output of a DMC, respectively, where \(\mathcal{X}\) and \(\mathcal{Y}\) denote the finite input and output alphabets, respectively. Let \(P_{Y|X}(y \mid x)\) denote the transition probability of a DMC \((X, Y)\) for \((x, y) \in \mathcal{X} \times \mathcal{Y}\). Then, we define the following three classes of DMCs.

**Definition 1.** A channel \((X, Y)\) is said to be uniformly dispersive [15] or uniform from the input [16] if there exists a permutation \(\pi_x : \mathcal{Y} \to \mathcal{Y}\) for each \(x \in \mathcal{X}\) such that \(P_{Y|X}(x \mid \pi_x(y)) = P_{Y|X}(x' \mid \pi_{x'}(y))\) for all \((x, x', y) \in \mathcal{X}^2 \times \mathcal{Y}\).

**Definition 2.** A channel \((X, Y)\) is said to be uniformly focusing [15] or uniform from the output [16] if there exists a permutation \(\pi_y : \mathcal{X} \to \mathcal{X}\) for each \(y \in \mathcal{Y}\) such that \(P_{Y|X}(\pi_y(x) \mid y) = P_{Y|X}(\pi_{y'}(x) \mid y')\) for all \((x, y, y') \in \mathcal{X} \times \mathcal{Y}^2\).

**Definition 3.** A channel is said to be strongly symmetric [15] or doubly uniform [16] if it is both uniformly dispersive and uniformly focusing.

For a uniformly dispersive channel \((X, Y)\), it is known that

\[H(Y \mid X) = H(Y \mid X = x)\]  

January 29, 2016 DRAFT
for any $x \in \mathcal{X}$ (see [16, Eq. (5.18)] or [15, Lemma 4.1]), where the conditional Shannon entropy $H(X | Y)$ of $(X, Y) \sim P_{X|Y}P_Y$ is defined by
\[
H(X | Y) \triangleq \mathbb{E}[H(P_{X|Y}(\cdot | Y))] \tag{304}
\]
and $\mathbb{E}[\cdot]$ denotes the expected value of the random variable. Moreover, let the conditional Rényi entropy of order $\alpha \in (0, 1) \cup (1, \infty)$ be denoted by
\[
H_\alpha(X | Y) \triangleq \frac{\alpha}{1 - \alpha} \ln \mathbb{E}[P_{X|Y}^\alpha(\cdot | Y)] \tag{305}
\]
for $(X, Y) \sim P_{X|Y}P_Y$. By convention, we write $H_1(X | Y) \triangleq H(X | Y)$. As with uniformly focusing channels, for uniformly focusing channels, we can provide the following lemma.

**Lemma 8.** If a channel $(X, Y)$ is uniformly focusing and the input $X$ follows a uniform distribution, then
\[
H_\alpha(X | Y) = H_\alpha(X | Y = y) \tag{306}
\]
for any $y \in \mathcal{Y}$ and any $\alpha \in (0, \infty)$.

**Proof of Lemma 8.** Consider a uniformly focusing channel $(X, Y)$. Assume that the input $X$ follows a uniform distribution, i.e., $P_X(x) = \frac{1}{|\mathcal{X}|}$ for all $x \in \mathcal{X}$. Note from [16, p. 127] or [15, Vol. I, Lemma 4.2] that, if the input $X$ follows a uniform distribution, then the output $Y$ also follows a uniform distribution, i.e., $P_Y(y) = \frac{1}{|\mathcal{Y}|}$ for all $y \in \mathcal{Y}$. Then, since the a posteriori probability of $(X, Y)$ is written as
\[
P_{X|Y}(x | y) = \frac{P_X(x)P_Y(y | x)}{P_Y(y)} \tag{307}
\]
for $(x, y) \in \mathcal{X} \times \mathcal{Y}$ by Bayes’ rule and the fraction $\frac{P_X(x)}{P_Y(y)}$ is constant for $(x, y) \in \mathcal{X} \times \mathcal{Y}$, it follows from Definition 2 that there exists a permutation $\pi_y : \mathcal{X} \rightarrow \mathcal{X}$ for each $y \in \mathcal{Y}$ such that
\[
P_{X|Y}(\pi_y(x) | y) = P_{X|Y}(\pi_y'(x) | y') \tag{308}
\]
for all $(x, y, y') \in \mathcal{X} \times \mathcal{Y}^2$. Hence, we get
\[
H(X | Y) = \sum_{y \in \mathcal{Y}} P_Y(y)H(X | Y = y) \tag{309}
\]
\[
= \sum_{y \in \mathcal{Y}} P_Y(y) \left( -\sum_{x \in \mathcal{X}} P_{X|Y}(x | y) \ln P_{X|Y}(x | y) \right) \tag{310}
\]
\[
= \sum_{y \in \mathcal{Y}} P_Y(y) \left( -\sum_{x \in \mathcal{X}} P_{X|Y}(\pi_y(x) | y) \ln P_{X|Y}(\pi_y(x) | y) \right) \tag{311}
\]
\[
= \sum_{y \in \mathcal{Y}} P_Y(y) \left( -\sum_{x \in \mathcal{X}} P_{X|Y}(\pi_y'(x) | y') \ln P_{X|Y}(\pi_y'(x) | y') \right) \tag{312}
\]
\[
= -\sum_{x \in \mathcal{X}} P_{X|Y}(\pi_y'(x) | y') \ln P_{X|Y}(\pi_y'(x) | y') \tag{313}
\]
\[
= -\sum_{x \in \mathcal{X}} P_{X|Y}(x | y') \ln P_{X|Y}(x | y') \tag{314}
\]
\[
= H(X | Y = y') \tag{315}
\]
for any \( y' \in \mathcal{Y} \). Similarly, we also get

\[
E[\|P_{X|Y}(\cdot \mid Y)\|_\alpha] = \sum_{y \in \mathcal{Y}} P_Y(y) \|P_{X|Y}(\cdot \mid y)\|_\alpha \\
= \sum_{y \in \mathcal{Y}} P_Y(y) \left( \sum_{x \in \mathcal{X}} P_{X|Y}(x \mid y)^\alpha \right)^{\frac{1}{\alpha}}
\]

(316)

or \( \geq \) if the input \( X \) follows a uniform distribution, as with (295) and (296). For a channel \( (X,Y) \), let the mutual information of order \( \alpha \in (0, \infty) \) [17] between \( X \) and \( Y \) be denoted by

\[
I_\alpha(X;Y) \triangleq H_\alpha(X) - H_\alpha(X \mid Y)
\]

(323)

for \( \alpha \in (0, \infty) \). Note that \( I_1(X;Y) \triangleq I(X;Y) \) denotes the (ordinary) mutual information between \( X \) and \( Y \). In this paragraph, we assume that a channel \( (X,Y) \) is uniformly focusing and the input \( X \) follows a uniform distribution. Since \( H_\alpha(u_n) = \ln n \) for \( \alpha \in (0, \infty) \), it follows from Lemma 8 that

\[
I_\alpha(X;Y) = \ln |\mathcal{X}| - H_\alpha(X \mid Y = y)
\]

(324)

\[
D_\alpha(P_{X|Y}(\cdot \mid y) \parallel u_{|\mathcal{X}|})
\]

(325)

for any \( y \in \mathcal{Y} \) and any \( \alpha \in (0, \infty) \), where \( |\cdot| \) denotes the cardinality of the finite set. Therefore, it follows that the tight bounds of \( I_\alpha(X;Y) \) with a fixed \( I(X;Y) \) are equivalent to the bounds of Corollary 2 under the hypotheses.

Furthermore, we consider Gallager’s \( E_0 \) function [18] of a channel \( (X,Y) \), defined by

\[
E_0(\rho, X, Y) = E_0(\rho, P_X, P_{Y \mid X})
\]

(326)

\[
\triangleq -\ln \sum_{y \in \mathcal{Y}} \left( \sum_{x \in \mathcal{X}} P_X(x) P_{Y \mid X}(y \mid x) \right)^{\frac{1}{1+\rho}}
\]

(327)

for \( \rho \in (-1, \infty) \). Then, we can obtain the following theorem.
Theorem 3. For a uniformly focusing channel \((X,Y)\), let

\[
E_0^{(v_n)}(\rho, X, Y) \triangleq \rho D_{\frac{1}{1+\rho}}(\hat{v}_n(X \mid Y) \parallel u_n),
\]

\[
E_0^{(w_n)}(\rho, X, Y) \triangleq \rho D_{\frac{1}{1+\rho}}(\hat{w}_n(X \mid Y) \parallel u_n),
\]

where \(\hat{v}_n(X \mid Y) \triangleq v_n(H_{w_n}^{-1}(H(X \mid Y)))\), \(\hat{w}_n(X \mid Y) \triangleq w_n(H_{w_n}^{-1}(H(X \mid Y)))\), and \(n = |X|\). If the input \(X\) follows a uniform distribution, then we observe that

\[
E_0^{(v_n)}(\rho, X, Y) \leq E_0(\rho, X, Y) \leq E_0^{(w_n)}(\rho, X, Y)
\]

for any \(\rho \in (-1, \infty)\).

Proof of Theorem 3: We can see from [17, Eq. (16)] that

\[
\frac{E_0(\rho, P_{X^\alpha}, P_{Y|X})}{\rho} = I_{\frac{1}{1+\rho}}(X;Y),
\]

where

\[
P_{X^\alpha}(x) \triangleq \frac{P_X(x)^\alpha}{\sum_{x' \in X} P_X(x')^\alpha}
\]

denotes the tilted distribution. We can see from [331] that the \(E_0\) function is closely related to the mutual information of order \(\alpha\). Note that, if the distribution \(P_X\) is a uniform distribution, then its tilted distribution \(P_{X^\alpha}\) is also a uniform distribution for any \(\alpha \in (0, \infty)\). Thus, if a channel \((X,Y)\) is uniformly focusing and the input \(X\) follows a uniform distribution, then it follows from (325) and (331) that

\[
E_0(\rho, X,Y) = \rho D_{\frac{1}{1+\rho}}(P_{X|Y}(- \mid y) \parallel u_{|X|})
\]

for any \(\rho \in (-1, \infty)\) and any \(y \in Y\). Hence, noting the relations

\[
-1 < \rho < 0 \iff 1 < \alpha < \infty,
\]

\[
0 < \rho < \infty \iff 0 < \alpha < 1,
\]

the \(E_0\) function can also be evaluate as with Corollary 2.

Note that the distributions \(\hat{v}_n(X \mid Y)\) and \(\hat{w}_n(X \mid Y)\) denote \(v_n(p)\) and \(w_n(q)\), respectively, such that \(H_{w_n}(p) = H_{w_n}(q) = H(X \mid Y)\) for a given channel \((X,Y)\). Since \(I(X;Y) = \ln |X| - H(X \mid Y)\) under a uniform input distribution, Theorem 3 shows bounds of the \(E_0\) function with a fixed mutual information. Note that, since (328) and (329) are defined by the \(v_n(-)\) and \(w_n(-)\), respectively, there exist two strongly symmetric channels which attain each equality of the bounds (330). Namely, Theorem 3 provides tight bounds (330). We illustrate graphical representations of Theorem 3 in Fig. 8, as with Figs. 1 and 2. Theorem 3 is a generalization of [22 Theorem 2] from ternary-input strongly symmetric channels to \(n\)-ary input uniformly focusing channels under a uniform input distribution.
Finally, we consider the hypothesis of a uniform input distribution. If a channel \((X,Y)\) is symmetric\(^3\), then the mutual information of order \(\alpha\) is maximized by a uniform input distribution\(^4\) for \(\alpha \in (0, \infty)\). Therefore, since a strongly symmetric channel is symmetric, the hypothesis is optimal if the channel \((X,Y)\) is strongly symmetric.

IV. CONCLUSION

In this study, we established the tight bounds of the \(\ell_\alpha\)-norm with a fixed Shannon entropy in Theorem 1 and vice versa in Theorem 2. Previously, the tight bounds of the Shannon entropy with a fixed error probability were derived \(^2\)–\(^6\), \(^21\). Since the error probability is closely related to the \(\ell_\infty\)-norm, this study is a generalization of previous studies \(^2\)–\(^6\), \(^21\). Note that the set of all \(n\)-ary probability vectors, which are sorted in decreasing order, with a fixed \(\ell_\alpha\)-norm is convex set. The previous works \(^2\)–\(^6\), \(^21\) used the concavity of the Shannon entropy in probability vectors to examine the Shannon entropy with a fixed \(\ell_\alpha\)-norm. However, since \(\|p\|_\alpha\) is strictly concave in \(p \in \mathcal{P}_n\) when \(\alpha \in (0,1)\) and is strictly convex in \(p \in \mathcal{P}_n\) when \(\alpha \in (1,\infty)\), the concavity of the Shannon entropy in probability vectors turns out to be hard-to-use when the \(\ell_\alpha\)-norm is fixed. In this study, we derived Theorems 1 and 2 by using elementary calculus without using the concavity of the Shannon entropy.

ACKNOWLEDGMENT

This study was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (C) 26420352.

REFERENCES

[1] C. E. Shannon, “A mathematical theory of communication,” Bell Syst. Tech. J., vol. 27, pp. 379–423 and 623–656, July and Oct. 1948.
[2] R. M. Fano, “Class notes for Transmission of Information,” Course 6.574, MIT, Cambridge, MA, 1952.
[3] V. A. Kovalevsky, “The problem of character recognition from the point of view of mathematical statistics,” Character Readers and Pattern Recognition. New York: Spartan, pp. 3–30, 1968. (Russian edition in 1965).

\(^3\)Symmetric channels are defined in \(^{18}, \text{p. 94}\).

\(^4\)This fact can be verified by using, e.g., \(^{23}, \text{Theorem 7.2}\).
[4] D. L. Tebbe and S. J. Dwyer III, “Uncertainty and probability of error,” *IEEE Trans. Inf. Theory*, vol. 14, no. 3, pp. 516–518, May 1968.
[5] M. Feder and N. Merhav, “Relations between entropy and error probability,” *IEEE Trans. Inf. Theory*, vol. 40, no. 1, pp. 259–266, Jan. 1994.
[6] S.-W. Ho and S. Verdú, “On the interplay between conditional entropy and error probability,” *IEEE Trans. Inf. Theory* vol. 56, no. 12, pp. 5930–5942, Dec. 2010.
[7] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic, 1979.
[8] P. Harremoës and F. Topsøe, “Inequalities between entropy and index of coincidence derived from information diagrams,” *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2944–2960, Nov. 2001.
[9] A. Rényi, “On measures of information and entropy,” *Proc. 4th Berkeley Symp. Math. Statist. Prob.*, Berkeley, Calif., vol. 1, Univ. of Calif. Press, pp. 547–561, 1961.
[10] C. Tsallis, “Possible generalization of Boltzmann-Gibbs statistics,” *J. Statist. Phys.*, vol. 52, no. 1–2, pp. 479–487, 1988.
[11] J. Havrda and F. Charvát, “Quantification method of classification processes. Concept of structural α-entropy,” *Kybernetika*, vol. 3, no. 1, pp. 30–35, 1967.
[12] Z. Daróczy, “Generalized information functions,” *Inf. Control*, vol. 16, no. 1, pp. 36–51, Mar. 1970.
[13] M. Behara and J. S. Chawla, “Generalized γ-entropy,” *Entropy and Ergodic Theory: Selecta Statistica Canadiana*, vol. 2, pp. 15–38, 1974.
[14] D. E. Boekee and J. C. A. Van der Lubbe, “The R-norm information measure,” *Inf. Control*, vol. 45, no. 2, pp. 136–155, May 1980.
[15] J. L. Massey, *Applied digital information theory I and II*. Lecture notes, Signal and Information Processing Laboratory, ETH Zurich, 1995–1996. [Online]. Available at [http://www.isiweb.ee.ethz.ch/archive/massey_scr/](http://www.isiweb.ee.ethz.ch/archive/massey_scr/)
[16] R. M. Fano, *Transmission of Information: A Statistical Theory of Communications*. New York: M.I.T. Press, 1961.
[17] S. Arimoto, “Information measures and capacity of order α for discrete memoryless channels,” in *Topics in Information Theory, 2nd Colloq. Math. Soc. J. Bolyai*, Keszthely, Hungary, vol. 16, pp. 41–52, 1977.
[18] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[19] C. Tsallis, “What are the numbers that experiments provide?” *Quimica Nova*, vol. 17, no. 6, pp. 468–471, 1994.
[20] A. Guillén i Fàbregas, I. Land, and A. Martinez, “Extremes of error exponents,” *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2201–2207, Apr. 2013.
[21] M. Ben-Bassat, “f-entropies, probability of error, and feature selection,” *Inf. Control*, vol. 39, no. 3, pp. 227–242, Dec. 1978.
[22] Y. Sakai and K. Iwata, “Feasible regions of symmetric capacity and Gallager’s $E_0$ function for ternary-input discrete memoryless channels,” *Proc. IEEE Int. Symp. Inf. Theory* (ISIT’2015), Hong Kong, pp. 81–85, June 2015.
[23] F. Jelinek *Probabilistic Information Theory: Discrete and Memoryless Models*. New York: McGraw-Hill, 1968.
[24] Y. Sakai and K. Iwata, “Relations between conditional Shannon entropy and expectation of $\ell_\gamma$-norm,” submitted to *IEEE Int. Symp. Inf. Theory*, (ISIT’2016) Barcelona, Spain, July 2016.