One-shot source coding with coded side information available at the decoder

Naqueeb Ahmad Warsi
Tata Institute of Fundamental Research
Homi Bhabha Road, Mumbai 400005
Email: naqueeb@tifr.res.in

***Abstract*—One-shot achievable rate region for source coding when coded side information is available at the decoder (source coding with a helper) is proposed. The achievable region proposed is in terms of conditional smooth max Rényi entropy and smooth max Rényi divergence. Asymptotically (in the limit of large block lengths) this region is quantified in terms of spectral-sup conditional entropy rate and spectral-sup mutual information rate. In particular, it coincides with the rate region derived in the limit of unlimited arbitrarily distributed copies of the sources.**

I. Introduction

The derivation of most of the fundamental results in information theory relies on the assumption that a random experiment is repeated identically and independently for large enough time. However, in practical scenarios both of these assumptions are not always justifiable. To overcome the limitations posed by these assumptions Renner et al. introduced the notion of *one-shot* information theory. One-shot information theory relies on the fact that a random experiment is available only *once*. Thus removing both the assumptions together.

The first one-shot bounds were given for the task of one-shot source coding [1]. These bounds were based on smooth Rényi entropies. The notion of smooth Rényi entropies was introduced for the very first time in the same work, i.e., in Ref. [1]. The elegance of the one-shot bounds obtained in Ref. [1] is that these bounds coincide with the Shannon entropy [2] of the information source in the limit of unlimited independent and identically distributed (i.i.d.) copies of the source. Furthermore, these bounds coincide with spectral sup-entropy rate as defined by Han and Vérdu in Ref. [3] in the limit of unlimited arbitrarily distributed copies of the source. One-shot bounds for distributed source coding were given by Sharma et al. in [4]. In Ref. [5] Wang et al. gave one-shot bounds for the channel coding problem in terms of smooth min Rényi divergence.

There has been a considerable work on the one-shot bounds for the quantum case under various scenarios (see for example Refs. [6], [7], [8], [9], [10], [11] and references therein).

In this work we give one-shot achievable rate region for source coding when coded state side information is available at the decoder. The achievable rate region derived for this problem is in terms of smooth max Rényi divergence and conditional smooth max Rényi entropy. The notion of smooth max Rényi divergence was introduced by Datta for the quantum case in [12]. We further show that the achievable region obtained asymptotically coincides with the rate region derived in [13] for the i.i.d case and with the rate region derived in [14] for the non-i.i.d. case.

The rest of this paper is organized as follows. In Section II we discuss the notations which we will be using throughout this paper. In Section III we give the definitions of smooth conditional Rényi entropy of order zero and smooth max Rényi divergence. In Section IV we state the problem of source coding with coded side information available at the decoder. We then propose a one-shot achievable rate region for this problem. We also prove a lemma pertaining to the asymptotic behavior of smooth max Rényi divergence. Although, this result is known in the quantum case we give a totally different proof. In particular, our proof involves more straight forward arguments. We then use this lemma to show that the proposed achievable rate region is asymptotically optimal.

II. Notation

In the discussions below we will be using X to represent a random variable. We will assume that all the random variables are discrete and have finite range. We represent a random sequence of length n by X^n and a particular realization of X^n by $\mathbf{x}$. Notation $\mathbf{X}$ will be used to represent an arbitrary sequence of random variables, i.e., $\mathbf{X} = \{X_n\}_{n=1}^{\infty}$. We use the notation $|\cdot|$ to represent the cardinality of a set. The set $\{\mathbf{x} : P_{X^n}(\mathbf{x}) > 0\}$ is denoted by $\text{Supp}(P_{X^n})$. We use the notation $X \to Y \to Z$ to denote the fact that random variables X, Y and Z form a Markov chain. We represent the set $\{x : 0 \leq x < \infty\}$ by $\mathbb{R}^+$. $\mathcal{X}\times\mathcal{Y}$ will represent the cartesian product of two sets. Similarly $(\mathcal{X}\times\mathcal{Y})^n$ will represent the n-th Cartesian product of the set $\mathcal{X}\times\mathcal{Y}$. The notation $\mathbb{N}$ is used to represent the set of natural numbers. Throughout this paper we will assume that log is to the base 2.

III. Smooth Rényi divergence of order infinity and conditional smooth Rényi entropy of order zero

Definition 1: (Max Rényi entropy [15]) Let $X \sim P_X$, with range $\mathcal{X}$. The max Rényi entropy of X is defined as

$$H_0(X) := \log|\text{Supp}(P_X)|.$$

Definition 2: (Conditional smooth max Rényi entropy [16]) Let $(X,Y) \sim P_{XY}$, with range $\mathcal{X}\times\mathcal{Y}$. For $\varepsilon \geq 0$, the conditional smooth max Rényi entropy of X given Y is defined as

$$H_0^{\varepsilon}(X|Y) := \min_{Q \in \mathcal{B}^{\varepsilon}(P_{XY})} \log \max_{y \in \mathcal{Y}} |\text{Supp}(Q(X|Y=y))|, \quad (1)$$

where $\mathcal{B}^{\varepsilon}(P_{XY}) = \{Q : \sum_{x,y \in \mathcal{X}\times\mathcal{Y}} Q(x,y) \geq 1 - \varepsilon, \forall\ (x,y) \in \mathcal{X} \times \mathcal{Y},\ P_{XY}(x,y) \geq Q(x,y) \geq 0\}$ and $Q(X=x|Y=y) := \frac{Q(x,y)}{P_Y(y)}$, for any $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. With the convention that $Q(X=x|Y=y) := 0$ if $P_Y(y) = 0$.

Definition 3: (Max Rényi divergence [15])
Let P and Q be two probability mass functions on the set $\mathcal{X}$ such that $\text{Supp}(P) \subseteq \text{Supp}(Q)$. The max Rényi divergence between P and Q is defined as

$$D_{\infty}(P||Q) := \log \max_{x:P(x)>0} \frac{P(x)}{Q(x)}. \quad (2)$$

Definition 4: (Smooth max Rényi divergence)
Let P and Q be two probability mass functions on the set $\mathcal{X}$ such that $\text{Supp}(P) \subseteq \text{Supp}(Q)$. The smooth max Rényi divergence between P and Q for $\varepsilon \in [0,1)$ is defined as

$$D_{\infty}^{\varepsilon}(P||Q) := \log \inf_{\phi \in \mathcal{B}^{\varepsilon}(P)} \max_{x:P(x)>0} \frac{\phi(x)}{Q(x)}, \quad (3)$$

where

$$\mathcal{B}^{\varepsilon}(P) = \Bigg\{\phi : 0 \leq \phi(x) \leq P(x), \forall x \in \mathcal{X} \text{ and } \sum_{x \in \mathcal{X}} \phi(x) \geq 1 - \varepsilon\Bigg\}.$$

Notice that $\mathcal{B}^{\varepsilon}(P)$ also contains functions which are not probability mass functions, therefore smooth max Rényi divergence can be negative. Also, $D_{\infty}^{\varepsilon}(P||Q)$ is a non-increasing function of ε, and for $\varepsilon = 0$ it reduces to max Rényi divergence.

IV. Source coding with coded state side information available at the decoder

Let $(X^n, Y^n) \sim P_{X^nY^n}$, with range $(\mathcal{X} \times \mathcal{Y})^n$, where

$$(X^n, Y^n) := [(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)].$$

The n-shot source coding with coded side information available at the decoder is formulated as follows. We first define two sets of integers

$$\mathcal{M}_n^{(1)} = \{1, \ldots, 2^{\ell^{\varepsilon}_{\text{d-enc}}(X^n)}\}, \quad (4)$$
$$\mathcal{M}_n^{(2)} = \{1, \ldots, 2^{\ell^{\varepsilon}_{\text{d-enc}}(Y^n)}\} \quad (5)$$

called the codes. Choose arbitrary mappings $e_n^{(1)} : \mathcal{X}^n \to \mathcal{M}_n^{(1)}$ (encoder1) and $e_n^{(2)} : \mathcal{Y}^n \to \mathcal{M}_n^{(2)}$ (encoder2). We call

$$\frac{\ell^{\varepsilon}_{\text{d}-\text{enc}}(X^n)}{n} = \frac{\log|\mathcal{M}_n^{(1)}|}{n},$$
$$\frac{\ell^{\varepsilon}_{\text{d}-\text{enc}}(Y^n)}{n} = \frac{\log|\mathcal{M}_n^{(2)}|}{n}$$

the coding rates of the encoder 1 and encoder 2, respectively. The decoder $d_n : \mathcal{M}_n^{(1)} \times \mathcal{M}_n^{(2)} \to \mathcal{X}^n$ receives two outputs $e_n^{(1)}(\mathbf{x})$ and $e_n^{(2)}(\mathbf{y})$ from the two encoders and tries to reconstruct the original source output $\mathbf{x}$. Thus the probability of error for this task is defined as

$$P_e^n := \Pr\{X^n \neq \hat{X}^n\},$$

where $\hat{X}^n = d_n(e_n^{(1)}(X^n), e_n^{(2)}(Y^n))$. Note here that the encoders $e_n^{(1)}$ and $e_n^{(2)}$ do not cooperate with each other. We call the triplet $(e_n^{(1)}, e_n^{(2)}, d_n)$ of two encoders and one decoder with the two codes in (4) and (5) and the error probability ε the $(n, 2^{\ell^{\varepsilon}_{\text{d-enc}}(X^n)}, 2^{\ell^{\varepsilon}_{\text{d-enc}}(Y^n)}, \varepsilon)$ n-shot code.

In this coding system we wish to minimize the two coding rates $\frac{\ell^{\varepsilon}_{\text{d-enc}}(X^n)}{n}$ and $\frac{\ell^{\varepsilon}_{\text{d-enc}}(Y^n)}{n}$ such that the probability of error is less than ε.

Definition 5: (One-shot ε-achievable rate pair)
A one-shot rate pair (R_1, R_2) is called ε-achievable if and only if there exists a $(1, 2^{\ell^{\varepsilon}_{\text{d-enc}}(X)}, 2^{\ell^{\varepsilon}_{\text{d-enc}}(Y)}, \varepsilon)$ one-shot code such that $P_e \leq \varepsilon$, $\ell^{\varepsilon}_{\text{d}-\text{enc}}(X) \leq R_1$ and $\ell^{\varepsilon}_{\text{d}-\text{enc}}(Y) \leq R_2$.

Definition 6: (Asymptotically achievable rate pair)
A rate pair (R_1, R_2) is asymptotically achievable if and only if there exists $(n, 2^{\ell^{\varepsilon}_{\text{d-enc}}(X^n)}, 2^{\ell^{\varepsilon}_{\text{d-enc}}(Y^n)}, \varepsilon)$ code such that $P_e^n \leq \varepsilon$,

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{\ell^{\varepsilon}_{\text{d}-\text{enc}}(X^n)}{n} \leq R_1$$

and

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{\ell^{\varepsilon}_{\text{d}-\text{enc}}(Y^n)}{n} \leq R_2.$$

Theorem 1: Let $(X, Y) \sim P_{XY}$, with range $\mathcal{X} \times \mathcal{Y}$, and let U satisfy $X \to Y \to U$. For the error $\varepsilon \in (0,1)$. The following one-shot rate region for source coding of X with a helper observing Y is achievable

$$\ell^{\varepsilon}_{\text{d}-\text{enc}}(X) \geq H_0^{\varepsilon_{11}}(X|U) - \log(\varepsilon - \varepsilon_1),$$
$$\ell^{\varepsilon}_{\text{d}-\text{enc}}(Y) \geq D_{\infty}^{\varepsilon_{11}}(P_{UY}||P_U \times P_Y) + \log[-\ln(\varepsilon_1 - \varepsilon_{11} - 2\varepsilon_{11}^{\frac{1}{2}})]$$

where $\varepsilon_1 < \varepsilon$ and ε_{11} is such that

$$\varepsilon_{11} + 2\varepsilon_{11}^{\frac{1}{2}} < \varepsilon_1 \text{ and } D_{\infty}^{\varepsilon_{11}}(P_{UY}||P_U \times P_Y) \geq 0. \quad (6)$$

Proof: Before giving the proof we would like to remark here that there always exists ε_{11} which satisfies the conditions mentioned in (6). This claim follows from the fact that $D_{\infty}^{\varepsilon_{11}}(P_{UY}||P_U \times P_Y)$ is a decreasing function of ε_{11}. The techniques used in the proof here are motivated by [13, Lemma 4.3] and [19, Lemma 4]. Fix a conditional probability mass function $P_{U|Y}$. Let $Q \in \mathcal{B}^{\varepsilon_{11}}(P_{UX})$ and $\phi \in \mathcal{B}^{\varepsilon_{11}}(P_{UY})$ be such that

$$H_0^{\varepsilon_{11}}(X|U) = \log \max_{u \in \mathcal{U}} |\text{Supp}(Q(X|U=u))| \quad (7)$$

and

$$D_{\infty}^{\varepsilon_{11}}(P_{UY}||P_U \times P_Y) = \log \max_{(u,y):P_{UY}(u,y)>0} \frac{\phi(u,y)}{P_U(u)P_Y(y)}, \quad (8)$$

where

$$\phi(U=u|Y=y) := \begin{cases} \frac{\phi(u,y)}{P_Y(y)} & \text{if } P_Y(y) > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

For details on the existence of Q which satisfies (7) see Ref. [16]. The existence of ϕ which satisfies (8) follows from the fact that the set $\mathcal{B}^{\varepsilon_{11}}(P)$ is compact, see Ref. [18, Lemma 1] for more details on this. For every $(u,y) \in \mathcal{U} \times \mathcal{Y}$, let

$$g(u,y) := \sum_{x \in \mathcal{X}} P_{X|Y}(x|y)\mathbf{I}(x,u), \quad (10)$$

where $\mathbf{I}(x,u)$ for every $(x,u) \in \mathcal{X} \times \mathcal{U}$ is defined as follows

$$\mathbf{I}(x,u) = \begin{cases} 1 & \text{if } (x,u) \notin \text{Supp}(Q), \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

Define the following set

$$\mathcal{F} := \left\{(u,y) \in \mathcal{U} \times \mathcal{Y} : g(u,y) \leq \varepsilon_{11}^{\frac{1}{2}}\right\}. \quad (12)$$

Random code generation: Randomly and independently assign an index $i \in [1 : 2^{\ell^{\varepsilon}_{\text{d-enc}}(X)}]$ to every realization $x \in \mathcal{X}$. The realizations with the same index i form a bin $\mathcal{B}(i)$. Randomly and independently generate $2^{\ell^{\varepsilon}_{\text{d-enc}}(Y)}$ realizations $u(k)$, $k \in [1 : 2^{\ell^{\varepsilon}_{\text{d-enc}}(Y)}]$, each according to P_U where $P_U = \sum_{y \in \mathcal{Y}} P_Y(y)P_{U|Y}(u|y)$.

Encoding: If the encoder 1 observes a realization $x \in \mathcal{B}(i)$, then the encoder 1 transmits i. For every realization $y \in \mathcal{Y}$ the encoder 2 finds an index k such that $(u(k), y) \in \mathcal{F}$. For the case when there are more than one such index, it sends the smallest one among them. If there is none, it then sends $k = 1$.

Decoding: The receiver finds the unique $x' \in \mathcal{B}(i)$ such that $(x', u(k)) \in \text{Supp}(Q)$. If there is none or more than one, then it declares an error.

Probability of error: We now calculate probability of error averaged over all code books. Let M_1 and M_2 be the random chosen indices for encoding X and Y. The error in the above mentioned encoding decoding strategy occurs if and only if one or more of the following error events occur

$$\begin{aligned} E_1 &= \left\{(U(m_2), Y) \notin \mathcal{F},\ \forall m_2 \in \left[1 : 2^{\ell^{\varepsilon}_{\text{d-enc}}(Y)}\right]\right\}, \\ E_2 &= \{(X, U(M_2)) \notin \text{Supp}(Q)\}, \\ E_3 &= \{\exists x' \in \mathcal{B}(m_1) : (x', U(M_2)) \in \text{Supp}(Q), x' \neq X\}. \end{aligned}$$

For more details on error events see [20, Theorem 10.2]. Notice that the typical sets $\mathcal{T}^{(n)}_{\varepsilon'}$ and $\mathcal{T}^{(n)}_{\varepsilon}$ used in the proof of Theorem 10.2 in [20] are replaced here by $\mathcal{F}$ and $\text{Supp}(Q)$. The probability of error is upper bounded as follows

$$\Pr\{E\} \leq \Pr\{E_1\} + \Pr\{E_1^c \cap E_2\} + \Pr\{E_3 | X \in \mathcal{B}(1)\}. \quad (13)$$

We now calculate $\Pr\{E_1\}$ as follows

$$\begin{aligned} &\Pr\{E_1\} \\ &= \sum_{y \in \mathcal{Y}} P_Y(y) \Pr\left\{(U(m_2), y) \notin \mathcal{F}, \forall m_2 \in \left[1 : 2^{\ell^{\varepsilon}_{\text{d-enc}}(Y)}\right]\right\} \\ &\stackrel{a}{=} \sum_{y \in \mathcal{Y}} P_Y(y) \left(1 - \sum_{u:(u,y) \in \mathcal{F}} P_U(u)\right)^{2^{\ell^{\varepsilon}_{\text{d-enc}}(Y)}} \\ &\stackrel{b}{\leq} \sum_{y \in \mathcal{Y}} P_Y(y) \Bigg(1 - 2^{-D^{\varepsilon_{11}}_{\infty}(P_{UY}||P_U \times P_Y)} \\ &\qquad \sum_{u:(u,y) \in \mathcal{F}} \phi(U=u|Y=y)\Bigg)^{2^{\ell^{\varepsilon}_{\text{d-enc}}(Y)}} \\ &\stackrel{c}{\leq} 1 - \sum_{(u,y) \in \mathcal{F}} \phi(u,y) + e^{-2^{\ell^{\varepsilon}_{\text{d-enc}}(Y)} 2^{-D^{\varepsilon_{11}}_{\infty}(P_{UY}||P_U \times P_Y)}} \\ &\stackrel{d}{\leq} \varepsilon_{11} + \Pr\{(U,Y) \notin \mathcal{F}\} + e^{-2^{\ell^{\varepsilon}_{\text{d-enc}}(Y)} 2^{-D^{\varepsilon_{11}}_{\infty}(P_{UY}||P_U \times P_Y)}}, \end{aligned}$$

where a follows because $U(1), \ldots, U(2^{\ell^{\varepsilon}_{\text{d-enc}}(Y)})$ are independent and subject to identical distribution P_U; b follows because for every $(u,y) \in \mathcal{U} \times \mathcal{Y}$

$$P_U(u) \geq 2^{-D^{\varepsilon_{11}}_{\infty}(P_{UY}||P_U \times P_Y)} \phi(U=u|Y=y), \quad (14)$$

where (14) follows from Definition 4, (8) and (9); c follows from the inequalities $(1-x)^y \leq e^{-xy}$ $(0 \leq x \leq 1, y \geq 0)$ and $e^{-xy} \leq 1 - y + x$ $(x \geq 0, 0 \leq y \leq 1)$ and (9); d is true because

$$1 - \varepsilon_{11} \leq \Pr\{(U,Y) \notin \mathcal{F}\} + \sum_{(u,y) \in \mathcal{F}} \phi(u,y), \quad (15)$$

where $\Pr\{(U,Y) \notin \mathcal{F}\} \leq \varepsilon_{11}^{\frac{1}{2}}$. See Ref. [18] for more details on (15) and about the calculation of $\Pr\{(U,Y) \notin \mathcal{F}\}$. The second term in (13) can be upper bounded by $\varepsilon_{11}^{\frac{1}{2}}$, i.e., $\Pr\{E_1^c \cap E_2\} \leq \varepsilon_{11}^{\frac{1}{2}}$. The details about the calculation of $\Pr\{E_1^c \cap E_2\}$ can be found in Ref. [18]. It now easily follows that

$$\begin{aligned} &\Pr\{E_1\} + \Pr\{E_1^c \cap E_2\} \\ &\quad \leq \varepsilon_{11} + 2\varepsilon_{11}^{\frac{1}{2}} + e^{-2^{\ell^{\varepsilon}_{\text{d-enc}}(Y)} 2^{-D^{\varepsilon_{11}}_{\infty}(P_{UY}||P_U \times P_Y)}}. \end{aligned}$$

Let

$$\varepsilon_1 \geq \varepsilon_{11} + 2\varepsilon_{11}^{\frac{1}{2}} + e^{-2^{\ell^{\varepsilon}_{\text{d-enc}}(Y)} 2^{-D^{\varepsilon_{11}}_{\infty}(P_{UY}||P_U \times P_Y)}}. \quad (16)$$

By taking log of both sides of (16) it now follows that

$$\begin{aligned} \ell^{\varepsilon}_{\text{d-enc}}(Y) \geq\ & D^{\varepsilon_{11}}_{\infty}(P_{UY}||P_U \times P_Y) \\ &+ \log[-\ln(\varepsilon_1 - \varepsilon_{11} - 2\varepsilon_{11}^{\frac{1}{2}})]. \end{aligned}$$

Finally, the third term in (13) can be upper bounded by $2^{-\ell^{\varepsilon}_{\text{d-enc}}(X)} \max_{u \in \mathcal{U}} |\text{Supp}(Q(X|U=u))|$, i.e.,

$$\begin{aligned} \Pr\{E_3\} &= \Pr\{E_3 | \mathcal{B}(1)\} \\ &\leq 2^{-\ell^{\varepsilon}_{\text{d-enc}}(X)} \max_{u \in \mathcal{U}} |\text{Supp}(Q(X|U=u))|. \quad (17) \end{aligned}$$

For more details on (17) see Ref. [18]. Thus from (16) and (17) it follows that

$$\Pr\{E\} \leq \varepsilon_1 + 2^{-\ell^{\varepsilon}_{\text{d-enc}}(X)} \max_{u\in\mathcal{U}} |\text{Supp}(Q(X|U=u))|.$$

Let

$$\varepsilon_1 + 2^{-\ell^{\varepsilon}_{\text{d-enc}}(X)} \max_{u\in\mathcal{U}} |\text{Supp}(Q(X|U=u))| \leq \varepsilon. \tag{18}$$

Taking $\log$ of both sides of (18) and using (7) we get

$$\ell^{\varepsilon}_{\text{d-enc}}(X) \geq H_0^{\varepsilon_{11}}(X|U) - \log(\varepsilon-\varepsilon_1).$$

This completes the proof. ∎

It would have been nice if the bounds derived in Theorem 1 had no $\log$ terms. However, this is true with other one-shot bounds derived in the literature see for example Refs. [1], [5], [16]. The $\log$ terms appear because of the use of random coding arguments.

We now mention two lemmas which will help us to prove the asymptotic optimality of the achievable region obtained in Theorem 1.

Lemma 1: (Datta and Renner [6]) Let $(\mathbf{X},\mathbf{U}) = \{(X_n,U_n)\}_{n=1}^{\infty}$ be an arbitrary random sequence taking values over the set $\{(\mathcal{X}\times\mathcal{U})^n\}_{n=1}^{\infty}$, where $(\mathcal{X}\times\mathcal{U})^n$ is the n-th Cartesian product of $\mathcal{X}\times\mathcal{U}$. Then

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{H_0^{\varepsilon}(X^n|U^n)}{n} = \overline{H}(\mathbf{X}|\mathbf{U}),$$

where $\overline{H}(\mathbf{X}|\mathbf{U})$ is called the spectral-sup conditional entropy rate of $\mathbf{X}$ given $\mathbf{U}$ [17]. In particular, if $(\mathbf{X},\mathbf{U}) = \{(X_n,U_n)\}_{n=1}^{\infty}$ is a random sequence of independent and identically distributed random pairs distributed according to P_{XY} then

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{H_0^{\varepsilon}(X^n|U^n)}{n} = H(X|U).$$

Lemma 2: Let $\mathbf{P} = \{P_n\}_{n=1}^{\infty}$ and $\mathbf{Q} = \{Q_n\}_{n=1}^{\infty}$ be an arbitrary sequences of probability mass functions defined on the set $\{\mathcal{X}^n\}_{n=1}^{\infty}$, where $\mathcal{X}^n$ is the n-th cartesian product of the set $\mathcal{X}$ and $|\mathcal{X}| < \infty$. Assume that for every $n\in\mathbb{N}$, $\text{Supp}(P_n)\subseteq\text{Supp}(Q_n)$. Then

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}D_{\infty}^{\varepsilon}(P_n||Q_n) = \bar{I}(\mathbf{P};\mathbf{Q}), \tag{19}$$

where

$$\bar{I}(\mathbf{P};\mathbf{Q}) := \inf\left\{\alpha \Big| \lim_{n\to\infty}\Pr\left\{\frac{1}{n}\log\frac{P_n}{Q_n}\leq\alpha\right\} = 1\right\}. \tag{20}$$

The probability in the R.H.S. of (20) is calculated with respect to P_n and $\bar{I}(\mathbf{P};\mathbf{Q})$ is called the spectral sup-mutual information rate between $\mathbf{P}$ and $\mathbf{Q}$ [17]. In particular, if $\mathbf{P} = \{P^{\times n}\}_{n=1}^{\infty}$ and $\mathbf{Q} = \{Q^{\times n}\}_{n=1}^{\infty}$, where $P^{\times n}$ and $Q^{\times n}$ represent the product distributions of P and Q on $\mathcal{X}^n$. Then

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}D_{\infty}^{\varepsilon}(P_n||Q_n) = D(P||Q). \tag{21}$$

Proof: See Appendix. ∎

The following lemma is an immediate consequence of Definition 6, Lemma 1 and Lemma 2.

Lemma 3: The normalized asymptotic limits of the bounds obtained in Theorem 1 are given by

$$R_1 \geq \overline{H}(\mathbf{X}|\mathbf{U}),$$
$$R_2 \geq \bar{I}(\mathbf{Y};\mathbf{U}).$$

Note that the above region is similar to the region obtained in [14, Theorem 2].

Appendix

We give here the proof for Lemma 2. We will first prove

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}D_{\infty}^{\varepsilon}(P_n||Q_n) \leq \bar{I}(\mathbf{P};\mathbf{Q}).$$

Consider any $\lambda > \bar{I}(\mathbf{P};\mathbf{Q})$. Let us define the following set

$$\mathcal{A}_n(\lambda) := \left\{\mathbf{x} : \frac{1}{n}\log\frac{P_n(\mathbf{x})}{Q_n(\mathbf{x})}\leq\lambda\right\}. \tag{22}$$

Let $\phi_n : \mathcal{X}^n\to[0,1]$, $n\in\mathbb{N}$, such that

$$\phi_n(\mathbf{x}) = \begin{cases} P_n(\mathbf{x}) & \text{if } \mathbf{x}\in\mathcal{A}_n(\lambda), \\ 0 & \text{otherwise.}\end{cases} \tag{23}$$

From (20) it easily follows that

$$\lim_{n\to\infty}\Pr\{\mathcal{A}_n(\lambda)\} = 1. \tag{24}$$

Thus from our construction of ϕ_n, (23), it follows that

$$\lim_{n\to\infty}\sum_{\mathbf{x}\in\mathcal{X}^n}\phi_n(\mathbf{x}) = \lim_{n\to\infty}\Pr\{\mathcal{A}_n(\lambda)\} = 1. \tag{25}$$

Notice the following inequalities

$$\begin{aligned}\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}D_{\infty}^{\varepsilon}(P_n||Q_n) &\overset{a}{\leq} \limsup_{n\to\infty}D_{\infty}(\phi_n||Q_n) \\ &= \limsup_{n\to\infty}\frac{1}{n}\log\max_{\mathbf{x}\in\mathcal{X}^n}\frac{\phi_n(\mathbf{x})}{Q_n(\mathbf{x})} \\ &\overset{b}{\leq} \lambda,\end{aligned}$$

where a follows from Definition 4 and (25); b follows from (22) and (23).

We now prove the other direction, i.e.,

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}D_{\infty}^{\varepsilon}(P_n||Q_n) \geq \bar{I}(\mathbf{P};\mathbf{Q}).$$

Consider any $\gamma < \bar{I}(\mathbf{P};\mathbf{Q})$. For every $n\in\mathbb{N}$, let us define the following the set

$$\mathcal{D}_n(\gamma) := \left\{\mathbf{x} : \frac{1}{n}\log\frac{P_n(\mathbf{x})}{Q_n(\mathbf{x})}\geq\gamma\right\}. \tag{26}$$

From (20) it follows that there exists $\eta\in(0,1]$, such that

$$\limsup_{n\to\infty}\Pr\{\mathcal{D}_n(\gamma)\} = \eta. \tag{27}$$

Since $\Pr\{\mathcal{D}_n(\gamma)\}+\Pr\{\mathcal{D}_n^c(\gamma)\} = 1$, for every $n\in\mathbb{N}$, we have $\liminf_{n\to\infty}\Pr\{\mathcal{D}_n^c(\gamma)\} = 1-\eta$. For $\varepsilon\in(0,\eta)$, consider a sequence of positive functions $\{\phi_n\}_{n=1}^{\infty}$, such that for every $n \in \mathbb{N}$

$$\begin{aligned}&\phi_n : \mathcal{X}^n \to [0,1], \phi_n(\mathbf{x}) \leq P_n(\mathbf{x}), \forall \mathbf{x} \in \mathcal{X}^n \\ &\text{and} \sum_{\mathbf{x}\in\mathcal{X}^n} \phi_n(\mathbf{x}) \geq 1-\varepsilon. \end{aligned} \tag{28}$$

For every $n \in \mathbb{N}$, we have

$$\begin{aligned}1-\varepsilon &\leq \sum_{\mathbf{x}\in\mathcal{D}_n^c(\gamma)} \phi_n(\mathbf{x}) + \sum_{\mathbf{x}\in\mathcal{D}_n(\gamma)} \phi_n(\mathbf{x}) \\ &\leq \Pr\{\mathcal{D}_n^c(\gamma)\} + \sum_{\mathbf{x}\in\mathcal{D}_n(\gamma)} \phi_n(\mathbf{x}).\end{aligned}$$

By rearranging the terms in the above equation we get

$$1-\varepsilon-\Pr\{\mathcal{D}_n^c(\gamma)\} \leq \sum_{\mathbf{x}\in\mathcal{D}_n(\gamma)} \phi_n(\mathbf{x}). \tag{29}$$

Taking $\limsup$ on both sides of (29), we have

$$\begin{aligned}\limsup_{n\to\infty} \sum_{\mathbf{x}\in\mathcal{D}_n(\gamma)} \phi_n(\mathbf{x}) &\geq 1-\varepsilon-\liminf_{n\to\infty} \Pr\{\mathcal{D}_n^c(\gamma)\} \\ &\geq \eta-\varepsilon. \end{aligned}\tag{30}$$

Now notice the following set of inequalities

$$\begin{aligned}1 &\geq \sum_{\mathbf{x}\in\mathcal{D}_n(\gamma)} P_n(\mathbf{x}) \\ &\overset{a}{\geq} \sum_{\mathbf{x}\in\mathcal{D}_n(\gamma)} 2^{n\gamma} Q_n(\mathbf{x}) \\ &\overset{b}{\geq} 2^{\left(n\gamma - \max_{\mathbf{x}\in\mathcal{X}^n} \log \frac{\phi_n(\mathbf{x})}{Q_n(\mathbf{x})}\right)} \sum_{\mathbf{x}\in\mathcal{D}_n(\gamma)} \phi_n(\mathbf{x}) \end{aligned}\tag{31}$$

where a follows from (26); b follows from the fact that for every $\mathbf{x} \in \mathcal{D}^n(\gamma)$,

$$\frac{\phi_n(\mathbf{x})}{Q_n(\mathbf{x})} \leq \max_{\mathbf{x}\in\mathcal{D}^n(\gamma)} \frac{\phi_n(\mathbf{x})}{Q_n(\mathbf{x})} \leq \max_{\mathbf{x}\in\mathcal{X}^n} \frac{\phi_n(\mathbf{x})}{Q_n(\mathbf{x})}.$$

By taking $\log$ of both sides of (31) and rearranging the terms we get

$$\max_{\mathbf{x}\in\mathcal{X}^n} \frac{1}{n} \log \frac{\phi_n(\mathbf{x})}{Q_n(\mathbf{x})} \geq \gamma + \frac{1}{n} \log \sum_{\mathbf{x}\in\mathcal{D}_n(\gamma)} \phi_n(\mathbf{x}).$$

Taking $\limsup$ on both sides of the above equation we have

$$\begin{aligned}\limsup_{n\to\infty} \max_{\mathbf{x}\in\mathcal{X}^n} \frac{1}{n} \log \frac{\phi_n(\mathbf{x})}{Q_n(\mathbf{x})} &\geq \gamma + \limsup_{n\to\infty} \frac{1}{n} \log \sum_{\mathbf{x}\in\mathcal{D}_n(\gamma)} \phi_n(\mathbf{x}) \\ &\geq \gamma \end{aligned}\tag{32}$$

where (32) follows from (30). Notice that (32) is true for every ϕ_n satisfying (28). Thus

$$\limsup_{n\to\infty} \frac{1}{n} D_{\infty}^{\varepsilon}(P_n || Q_n) \geq \gamma. \tag{33}$$

Since (33) is true for every $\varepsilon \in (0, \eta)$, the result will hold true for $\varepsilon \downarrow 0$.

(21) easily follows from the law of large numbers and (19). This completes the proof.

Conclusion and Acknowledgements

We proved that smooth max divergence and smooth max conditional Rényi entropy can be used to obtain one-shot achievable rate region for source coding when coded side information is available at the decoder. Furthermore, we showed that asymptotically this region coincides with the rate region as derived by Wyner in [13, Theorem 2.1] for the i.i.d. case and with the region derived by Miyake and Kanaya in [14, Theorem 2] for the non-i.i.d. case.

The author gratefully acknowledges the helpful discussions with Mohit Garg and Sarat Moka. The author also thanks Jaikumar Rahdhakrishnan for his comments on the previous version of this manuscript.

References

[1] R. Renner and S. Wolf, "Smooth Rényi entropy and applications," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, (Chicago, IL, USA), June 2004.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2nd ed., 2006.

[3] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, pp. 752 –772, May 1993.

[4] N. Sharma and N. A. Warsi, "One-shot Slepian-Wolf." arXiv:1112.1687, Jan. 2012.

[5] L. Wang, R. Colbeck, and R. Renner, "Simple channel coding bounds," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, (Seoul, Korea), June 2009.

[6] N. Datta and R. Renner, "Smooth Rényi entropies and the quantum information spectrum," *IEEE Trans. Inf. Theory*, vol. 55, pp. 2807–2815, 2009.

[7] R. König, R. Renner, and C. Schaffner, "The operational meaning of min- and max-entropy," *IEEE Trans. Inf. Theory*, vol. 55, pp. 4337–4347, Sept. 2009.

[8] F. Dupuis, P. Hayden, and K. Li, "A father protocol for quantum broadcast channels," *IEEE Trans. Inf. Theory*, vol. 56, pp. 2946–2956, June 2010.

[9] M. Berta, M. Christandl, and R. Renner, "The quantum reverse Shannon theorem based on one-shot information theory," *Commun. Math. Phys.*, vol. 306, pp. 579–615, Sept. 2011.

[10] N. Datta and M.-H. Hsieh, "The apex of the family tree of protocols: optimal rates and resource inequalities," *New J. Phys.*, vol. 13, p. 093042, Sept. 2011.

[11] J. M. Renes and R. Renner, "Noisy channel coding via privacy amplification and information reconciliation," *IEEE Trans. Inf. Theory*, vol. 57, pp. 7377–7385, Nov. 2011.

[12] N. Datta, "Min- and max-relative entropies and a new entanglement monotone," *IEEE Trans. Inf. Theory*, vol. 55, pp. 2816 –2826, June 2009.

[13] A. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 21, pp. 94 – 300, May 1975.

[14] S. Miyake and F. Kanaya, "Coding theorems on correlated general sources," *IEICE Trans. Fundamentals*, vol. E78-A(9), pp. 1063–1070, Sept. 1995.

[15] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math Stat. Prob.*, pp. 547–561, 1960.

[16] R. Renner and S. Wolf, "Simple and tight bounds for information reconciliation and privacy amplification," in *Advances in Cryptology—ASIACRYPT 2005, Lecture Notes in Computer Science*, pp. 199–216, Springer-Verlag, 2005.

[17] T. S. Han, *Information-Spectrum Methods in Information Theory*. Berlin, Germany: Springer-Verlag, 2003.

[18] N. A. Warsi, "One-shot source coding with coded state side information available at the decoder." arxiv: 1303.2579, Mar. 2013.

[19] S. Kuzuoka, "A simple technique for bounding the redundancy of source coding with side information," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, (Cambridge, MA, USA), pp. 910–914, July 2012.

[20] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge, U.K: Cambridge University Press, 2012.