Exponential Strong Converse in Multi-user Problems

International Zurich Seminar on Information and Communication (IZS), March 6 – 8, 2024

Exponential Strong Converse in Multi-user Problems

Shun Watanabe
Tokyo University of Agriculture and Technology
Tokyo, Japan
email: shunwata@cc.tuat.ac.jp

***Abstract*—The exponential strong converse for a coding problem states that, if a coding rate (or a rate pair) is beyond the theoretical limit, the correct decoding probability converges to zero exponentially. The exponential strong converse theorem was initiated by Arimoto and by Dueck and Körner for the pont-to-point channel coding; even though tight exponents have been identified for single-user problems and simple multi-user problems, such as the Slepian-Wolf problem, tight exponents have been unsolved for multi-user problems. In this tutorial paper, we revisit the exponential strong converse theorems, and provide alternative proofs for single-user problems via manipulations of information quantities as in the weak converse argument (called "change-of-measure argument" in the literature). Then, we present the recently obtained result by Takeuchi and Watanabe providing the tight exponential strong converse for the source coding with coded side-information.**

I. Introduction

The strong converse for a coding theorem claims that the optimal asymptotic rate possible with vanishing probability cannot be improved by allowing a fixed error probability. The exponential strong converse further claims that, if a coding rate is beyond the asymptotic limit, the correct decoding probability converges to zero exponentially. Proving such a claim was initiated by Arimoto for the channel coding problem [2]; later, the strong converse exponent was studied by Dueck and Körner in [4]; see also [10] for the equivalence of the two exponents derived in [2] and [4]. Also, the strong converse exponent for the Slepian-Wolf problem was derived by Oohama and Han in [13].

Even though the tight strong converse exponent for point-to-point problems or simple multi-user problems, such as the Slepian-Wolf problem, have been identified, the strong converse exponent for multi-user problems have been unsolved until recently. A significant progress was made by Oohama in a series of paper including [11], [12]. More recently, the tight strong converse of the Wyner-Ahlswede-Körner (WAK) problem [1], [18] was derived in [14]; the converse part of [14] is based on a manipulations of information quantities as in the weak converse argument, called the "change-of-measure argument" in [15]. In this tutorial paper, we provide alternative proofs of the strong converse exponents for single-user problems by using the same methodology.

The change-of-measure argument was originally introduced by Gu and Effros in [6], [7] to prove strong converse for source coding problems where there exists a terminal that observes all the random variables involved; a particular example is the Gray-Wyner (GW) problem [5]. In the argument of [6], [7], we evaluate the performance of a given code not under the original source (or channel) but under another modified measure which depends on the code and under which the code is error free.[1] A type based modification of this argument was used in [17] to derive the second-order rate region of the GW problem. A difficulty of applying this argument to the so-called distributed coding problems, such as the WAK problem, is that the characterization of asymptotic limits involve auxiliary random variables and Markov chain constraints. This technical difficulty was circumvented in [16] for the WAK problem by relating the WAK problem to an extreme case of the GW problem. By using the idea of "soft Markov constraint" introduced by Oohama [11], the argument was further developed in [15] so that it can be applied to distributed coding problems; furthermore, the argument was also extended so that it can be applied to secrecy problems such as the secret key generation and the wiretap channel. More recently, a variation of the change-of-measure argument was further developed by Hamad, Wigger, and Sarkiss in [8] so that it can be applied to more involved multi-user networks in a concise manner; rather than adding a Markov constraint as a penalty term, they prove the Markov constraint in an asymptotic limit.

II. Preliminaries

We use the same notations as [3]. For instance, the entropy of random variable X is denoted as $H(X)$; the mutual information between X and Y is denoted as $I(X \wedge Y)$; and the KL-divergence between distributions P and Q is denoted as $D(P\|Q)$. The logarithm is base 2.

Let $X^n = (X_1, \ldots, X_n)$ be an independently identically distributed (i.i.d.) source on a finite alphabet $\mathcal{X}$. For a given set $\mathcal{C} \subset \mathcal{X}^n$, a key step of the change-of-measure argument is to construct a modified measure by conditioning:

$$P_{\tilde{X}^n}(x^n) := \frac{P_{X^n}(x^n)\mathbf{1}[x^n \in \mathcal{C}]}{P_{X^n}(\mathcal{C})},$$

where $\mathbf{1}[\cdot]$ is the indicator function. A key observation, which was used in Marton's proof of the blowing-up lemma [9], is that the modified measure is not too far from the original measure in the following sense:

$$D(P_{\tilde{X}^n}\|P_{X^n}) = \sum_{x^n \in \mathcal{C}} P_{\tilde{X}^n}(x^n) \log \frac{P_{\tilde{X}^n}(x^n)}{P_{X^n}(x^n)}$$

[1]In the original argument [6], [7], the modified measure is constructed by conditioning on typical sets in addition to the error free set; on the other hand, the argument in [15] only conditions on the error free set.

$$= \log \frac{1}{P_{X^n}(\mathcal{C})}.$$

The conditional measure $P_{\tilde{X}^n}$ is not i.i.d. in general. By using the sub-additivity and concavity of entropy, we can directly derive a single-letter upper bound on the joint entropy as

$$H(\tilde{X}^n) \le \sum_{j=1}^n H(\tilde{X}_j) \le nH(\tilde{X}_J),$$

where J is the random variable uniformly distributed on the index set $\{1,\ldots,n\}$. It is not possible to derive a single-letter lower bound on the joint entropy $H(\tilde{X}^n)$ directly; instead, we manipulate it with the divergence term:

$$\begin{aligned} H(\tilde{X}^n) + D(P_{\tilde{X}^n}\|P_{X^n}) &= \sum_{x^n} P_{\tilde{X}^n}(x^n) \log \frac{1}{P_{X^n}(x^n)} \\ &= \sum_{j=1}^n \sum_{x^n} P_{\tilde{X}^n}(x^n) \log \frac{1}{P_X(x_j)} \\ &= \sum_{j=1}^n \sum_{x} P_{\tilde{X}_j}(x) \log \frac{1}{P_X(x)} \\ &= n \sum_{x} P_{\tilde{X}_J}(x) \log \frac{1}{P_X(x)} \\ &= n\big[H(\tilde{X}_J) + D(P_{\tilde{X}_J}\|P_X)\big]. \quad (1) \end{aligned}$$

By the convexity of the KL-divergence, we can also derive a single-letter lower bound on the KL-divergence:

$$\begin{aligned} D(P_{\tilde{X}^n}\|P_{X^n}) &= \sum_{j=1}^n D(P_{\tilde{X}_j|\tilde{X}^{j-1}}\|P_X|P_{\tilde{X}^{j-1}}) \\ &\ge \sum_{j=1}^n D(P_{\tilde{X}_j}\|P_X) \\ &\ge nD(P_{\tilde{X}_J}\|P_X). \end{aligned}$$

The derivation of the strong converse exponent proceed by a judicious use of the above single-letter bounding manipulations.

III. Lossy Source Coding

In this section, we consider the lossy source coding. For a finite alphabet $\mathcal{X}$, let $X^n = (X_1,\ldots,X_n)$ be an independently identically distributed (i.i.d.) source with distribution $P_{X^n} = P_X^n$. For a finite reproduction alphabet $\mathcal{Y}$, we consider an encoder $\varphi : \mathcal{X}^n \to \mathcal{M}$ and a decoder $\psi : \mathcal{M} \to \mathcal{Y}^n$. For a distortion measure $d : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}_+$, let $d_n(x^n,y^n) = \sum_{j=1}^n d(x_j,y_j)$. For a distortion level $\Delta \ge 0$, we shall evaluate non-excess distortion probability:

$$p_c := \Pr\big(d_n(X^n,\psi(\varphi(X^n))) \le n\Delta\big).$$

For fixed Δ, a rate R is defined to be achievable if, for every $0 < \varepsilon < 1$ and for sufficiently large n, there exists a code (φ,ψ) such that the non-excess distortion probability satisfies $p_c \ge 1-\varepsilon$ and the coding rate satisfies $\frac{1}{n}\log|\mathcal{M}| \le R$. Then, the rate-distortion function $R(P_X,\Delta)$ is defined as the infimum of achievable rates. It is well known that the rate-distortion function is characterized as

$$R(P_X,\Delta) = \min_{\substack{P_{Y|X}: \\ \mathbb{E}[d(X,Y)]\le\Delta}} I(X\wedge Y).$$

We provide an alternative proof for the following exponential strong converse of the lossy source coding.

Proposition 1 For any code (φ,ψ) such that $\frac{1}{n}\log|\mathcal{M}| \le R$, the non-excess distortion probability satisfies

$$\frac{1}{n}\log(1/p_c) \ge \min_{P_{\tilde{X}}} \big[D(P_{\tilde{X}}\|P_X) + |R(P_{\tilde{X}},\Delta) - R|^+\big],$$

where $|a|^+ := \max[a,0]$.

Note that the exponent is positive if and only if $R < R(P_X,\Delta)$. It is known that the strong converse exponent in Proposition 1 is tight [3, Ex. 9.6].

Proof. Let

$$\mathcal{C} := \big\{x^n \in \mathcal{X}^n : d_n(x^n,\psi(\varphi(x^n))) \le n\Delta\big\},$$

and let

$$P_{\tilde{X}^n}(x^n) := \frac{P_{X^n}(x^n)\mathbf{1}[x^n\in\mathcal{C}]}{P_{X^n}(\mathcal{C})}.$$

Then, we have

$$D(P_{\tilde{X}^n}\|P_{X^n}) = \log(1/p_c)$$

and

$$\begin{aligned} \log(1/p_c) &= D(P_{\tilde{X}^n}\|P_{X^n}) \\ &\ge nD(P_{\tilde{X}_J}\|P_X). \end{aligned} \quad (2)$$

Note that the rate R can be lower bounded as

$$\begin{aligned} nR &\ge \log|\mathcal{M}| \\ &\ge H(\tilde{Y}^n) \\ &= I(\tilde{X}^n \wedge \tilde{Y}^n), \end{aligned}$$

where $\tilde{Y}^n = \psi(\varphi(\tilde{X}^n))$. Thus, we have

$$\begin{aligned} \log(1/p_c) &= D(P_{\tilde{X}^n}\|P_{X^n}) \\ &\ge D(P_{\tilde{X}^n}\|P_{X^n}) + I(\tilde{X}^n\wedge\tilde{Y}^n) - nR. \end{aligned}$$

Furthermore, by (1), we have

$$\begin{aligned} &D(P_{\tilde{X}^n}\|P_{X^n}) + I(\tilde{X}^n\wedge\tilde{Y}^n) \\ &= D(P_{\tilde{X}^n}\|P_{X^n}) + H(\tilde{X}^n) - H(\tilde{X}^n|\tilde{Y}^n) \\ &= nD(P_{\tilde{X}_J}\|P_X) + nH(\tilde{X}_J) - \sum_{j=1}^n H(\tilde{X}_j|\tilde{Y}^n,\tilde{X}_j^-) \\ &\ge nD(P_{\tilde{X}_J}\|P_X) + nH(\tilde{X}_J) - \sum_{j=1}^n H(\tilde{X}_j|\tilde{Y}_j) \\ &= nD(P_{\tilde{X}_J}\|P_X) + nH(\tilde{X}_J) - nH(\tilde{X}_J|\tilde{Y}_J,J) \\ &\ge nD(P_{\tilde{X}_J}\|P_X) + nH(\tilde{X}_J) - nH(\tilde{X}_J|\tilde{Y}_J) \\ &= nD(P_{\tilde{X}_J}\|P_X) + nI(\tilde{X}_J\wedge\tilde{Y}_J), \end{aligned}$$

where $\tilde{X}_j^- = (\tilde{X}_1, \ldots, \tilde{X}_{j-1})$. Also, since the support of the changed measure $P_{\tilde{X}^n}$ is $\mathcal{C}$, note that

$$\Delta \ge \mathbb{E}\left[\frac{1}{n} d_n(\tilde{X}^n, \tilde{Y}^n)\right] = \mathbb{E}[d(\tilde{X}_J, \tilde{Y}_J)].$$

Thus, we have

$$\begin{aligned}\frac{1}{n}\log(1/p_c) &\ge D(P_{\tilde{X}_J}\|P_X) + \big(I(\tilde{X}_J \wedge \tilde{Y}_J) - R\big) \\ &\ge D(P_{\tilde{X}_J}\|P_X) + \big(R(P_{\tilde{X}_J}, \Delta) - R\big). \quad (3)\end{aligned}$$

By combining (2) and (3), we have

$$\frac{1}{n}\log(1/p_c) \ge D(P_{\tilde{X}_J}\|P_X) + |R(P_{\tilde{X}_J}, \Delta) - R|^+.$$

Finally, by replacing $P_{\tilde{X}_J}$ with the minimum over $P_{\tilde{X}}$, we have the claim of the proposition. ■

IV. Channel Coding

In this section, we consider the channel coding. Let W^n be a discrete memoryless channel (DMC) from a finite input alphabet $\mathcal{X}$ to a finite output alphabet $\mathcal{Y}$. For a message set $\mathcal{M}$, a channel code consists of an encoder $\varphi : \mathcal{M} \to \mathcal{X}^n$ and a decoder $\psi : \mathcal{Y}^n \to \mathcal{M}$. Let

$$p_c := \sum_{m \in \mathcal{M}} \frac{1}{|\mathcal{M}|} W^n(\psi^{-1}(m)|\varphi(m))$$

be the average correct decoding probability. A rate R is defined to be achievable if, for every $0 < \varepsilon < 1$ and for sufficiently large n, there exists a code (φ, ψ) such that the average correct decoding probability satisfies $p_c \ge 1 - \varepsilon$ and the coding rate satisfies $\frac{1}{n}\log|\mathcal{M}| \ge R$. Then, the channel capacity $C(W)$ is defined as the supremum of achievable rates. It is well known that the channel capacity is characterized as

$$C(W) = \max_{P_X} I(X \wedge Y),$$

where the mutual information is evaluated with respect to (X, Y) induced by the input distribution P_X and the channel W.

We provide an alternative proof for the following exponential strong converse of the channel coding.

Proposition 2 For any code (φ, ψ) such that $\frac{1}{n}\log|\mathcal{M}| \ge R$, the average correct decoding probability satisfies

$$\frac{1}{n}\log(1/p_c) \ge \min_{P_{\tilde{X}\tilde{Y}}} \big[D(P_{\tilde{Y}|\tilde{X}}\|W|P_{\tilde{X}}) + |R - I(\tilde{X} \wedge \tilde{Y})|^+\big].$$

Note that the exponent is positive if and only if $R > C(W)$. It is known that the strong converse exponent in Proposition 2 is tight [4]. Furthermore, it also coincides with the strong converse exponent by Arimoto [2]; see [10] for the equivalence.

Proof. Let

$$\mathcal{C} := \big\{(m, x^n, y^n) : \psi(y^n) = m\big\}.$$

For

$$P_{MX^nY^n}(m, x^n, y^n) = \frac{1}{|\mathcal{M}|}\mathbf{1}[x^n = \varphi(m)] W^n(y^n|x^n),$$

let

$$\begin{aligned}&P_{\tilde{M}\tilde{X}^n\tilde{Y}^n}(m, x^n, y^n) \\ &:= \frac{P_{MX^nY^n}(m, x^n, y^n)\mathbf{1}[(m, x^n, y^n) \in \mathcal{C}]}{P_{MX^nY^n}(\mathcal{C})}.\end{aligned}$$

Then, we have

$$D(P_{\tilde{M}\tilde{X}^n\tilde{Y}^n}\|P_{MX^nY^n}) = \log(1/p_c).$$

By noting that $P_{Y^n|MX^n} = W^n$, we have

$$\begin{aligned}D(P_{\tilde{Y}^n|\tilde{M}\tilde{X}^n}\|W^n|P_{\tilde{M}\tilde{X}^n}) &= D(P_{\tilde{Y}^n|\tilde{M}\tilde{X}^n}\|P_{Y^n|MX^n}|P_{\tilde{M}\tilde{X}^n}) \\ &\le D(P_{\tilde{M}\tilde{X}^n\tilde{Y}^n}\|P_{MX^nY^n}) \\ &= \log(1/p_c).\end{aligned}$$

By the convexity of the KL-divergence, we also have

$$\begin{aligned}D(P_{\tilde{Y}^n|\tilde{X}^n}\|W^n|P_{\tilde{X}^n}) &\le D(P_{\tilde{Y}^n|\tilde{M}\tilde{X}^n}\|W^n|P_{\tilde{M}\tilde{X}^n}) \\ &\le \log(1/p_c). \quad (4)\end{aligned}$$

Furthermore, by the monotonicity of the KL-divergence, we also have

$$\begin{aligned}&D(P_{\tilde{M}}\|P_M) + D(P_{\tilde{Y}^n|\tilde{M}\tilde{X}^n}\|W^n|P_{\tilde{M}\tilde{X}^n}) \\ &\le D(P_{\tilde{M}\tilde{X}^n}\|P_{MX^n}) + D(P_{\tilde{Y}^n|\tilde{M}\tilde{X}^n}\|W^n|P_{\tilde{M}\tilde{X}^n}) \\ &= D(P_{\tilde{M}\tilde{X}^n\tilde{Y}^n}\|P_{MX^nY^n}) \\ &= \log(1/p_c). \quad (5)\end{aligned}$$

Now, by noting that P_M is uniform distribution on $\mathcal{M}$, we have

$$\begin{aligned}&nR \\ &\le \log|\mathcal{M}| \\ &= H(\tilde{M}) + D(P_{\tilde{M}}\|P_M) \\ &= I(\tilde{M} \wedge \tilde{Y}^n) + D(P_{\tilde{M}}\|P_M) \\ &\le I(\tilde{M} \wedge \tilde{Y}^n) + D(P_{\tilde{M}}\|P_M) \\ &\quad + \big[\log(1/p_c) - D(P_{\tilde{M}}\|P_M) - D(P_{\tilde{Y}^n|\tilde{M}\tilde{X}^n}\|W^n|P_{\tilde{M}\tilde{X}^n})\big] \\ &= I(\tilde{M} \wedge \tilde{Y}^n) - D(P_{\tilde{Y}^n|\tilde{M}\tilde{X}^n}\|W^n|P_{\tilde{M}\tilde{X}^n}) + \log(1/p_c) \\ &= I(\tilde{M}, \tilde{X}^n \wedge \tilde{Y}^n) - D(P_{\tilde{Y}^n|\tilde{M}\tilde{X}^n}\|W^n|P_{\tilde{M}\tilde{X}^n}) + \log(1/p_c) \\ &= I(\tilde{X}^n \wedge \tilde{Y}^n) + I(\tilde{M} \wedge \tilde{Y}^n|\tilde{X}^n) \\ &\quad - D(P_{\tilde{Y}^n|\tilde{M}\tilde{X}^n}\|W^n|P_{\tilde{M}\tilde{X}^n}) + \log(1/p_c) \\ &= I(\tilde{X}^n \wedge \tilde{Y}^n) + D(P_{\tilde{Y}^n|\tilde{M}\tilde{X}^n}\|P_{\tilde{Y}^n|\tilde{X}^n}|P_{\tilde{M}\tilde{X}^n}) \\ &\quad - D(P_{\tilde{Y}^n|\tilde{M}\tilde{X}^n}\|W^n|P_{\tilde{M}\tilde{X}^n}) + \log(1/p_c) \\ &= I(\tilde{X}^n \wedge \tilde{Y}^n) - D(P_{\tilde{Y}^n|\tilde{X}^n}\|W^n|P_{\tilde{X}^n}) + \log(1/p_c), \quad (6)\end{aligned}$$

where the second equality follows since $\tilde{M}$ can be decoded from $\tilde{Y}^n$ with 0 error probability, the second inequality follows from (5), and the forth equality follows since $\tilde{X}^n$ is a function of $\tilde{M}$.[2] Now, we conduct the single-letter procedure as follows:

$$\begin{aligned}
&I(\tilde{X}^n \wedge \tilde{Y}^n) - D(P_{\tilde{Y}^n|\tilde{X}^n}\|W^n|P_{\tilde{X}^n}) \\
&= H(\tilde{Y}^n) - H(\tilde{Y}^n|\tilde{X}^n) - D(P_{\tilde{Y}^n|\tilde{X}^n}\|W^n|P_{\tilde{X}^n}) \\
&= H(\tilde{Y}^n) - \sum_{x^n,y^n} P_{\tilde{X}^n\tilde{Y}^n}(x^n,y^n)\log\frac{1}{W^n(y^n|x^n)} \\
&= H(\tilde{Y}^n) - n\sum_{x,y} P_{\tilde{X}_J\tilde{Y}_J}(x,y)\log\frac{1}{W(y|x)} \\
&= H(\tilde{Y}^n) - nH(\tilde{Y}_J|\tilde{X}_J) - nD(P_{\tilde{Y}_J|\tilde{X}_J}\|W|P_{\tilde{X}_J}) \\
&\le nH(\tilde{Y}_J) - nH(\tilde{Y}_J|\tilde{X}_J) - nD(P_{\tilde{Y}_J|\tilde{X}_J}\|W|P_{\tilde{X}_J}) \\
&= nI(\tilde{X}_J \wedge \tilde{Y}_J) - nD(P_{\tilde{Y}_J|\tilde{X}_J}\|W|P_{\tilde{X}_J}). \qquad (7)
\end{aligned}$$

Thus, by combining (6) and (7), we have

$$\frac{1}{n}\log(1/p_c) \ge D(P_{\tilde{Y}_J|\tilde{X}_J}\|W|P_{\tilde{X}_J}) + \big(R - I(\tilde{X}_J \wedge \tilde{Y}_J)\big). \qquad (8)$$

Note also that

$$\begin{aligned}
D(P_{\tilde{Y}^n|\tilde{X}^n}\|W^n|P_{\tilde{X}^n}) &= \sum_{j=1}^n D(P_{\tilde{Y}_j|\tilde{X}^n\tilde{Y}_j^-}\|W|P_{\tilde{X}^n\tilde{Y}_j^-}) \\
&\ge \sum_{j=1}^n D(P_{\tilde{Y}_j|\tilde{X}_j}\|W|P_{\tilde{X}_j}) \\
&= nD(P_{\tilde{Y}_J|\tilde{X}_J J}\|W|P_{\tilde{X}_J J}) \\
&\ge nD(P_{\tilde{Y}_J|\tilde{X}_J}\|W|P_{\tilde{X}_J}). \qquad (9)
\end{aligned}$$

Thus, by combining (4) and (9), we have

$$\frac{1}{n}\log(1/p_c) \ge D(P_{\tilde{Y}_J|\tilde{X}_J}\|W|P_{\tilde{X}_J}). \qquad (10)$$

Thus, by combining (8) and (10), we have

$$\frac{1}{n}\log(1/p_c) \ge D(P_{\tilde{Y}_J|\tilde{X}_J}\|W|P_{\tilde{X}_J}) + |R - I(\tilde{X}_J \wedge \tilde{Y}_J)|^+.$$

Finally, by replacing $P_{\tilde{X}_J\tilde{Y}_J}$ with the minimum over $P_{\tilde{X}\tilde{Y}}$, we have the claim of the proposition. ■

V. Source Coding with Coded Side-information

In this section, we consider the source coding with coded side-information, also known as the Wyner-Ahlswede-Körner (WAK) problem [1], [18]. For finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, let (X^n, Y^n) be i.i.d. correlated source with distribution $P_{X^nY^n}$. A code consists of two encoders $\varphi_1 : \mathcal{X}^n \to \mathcal{M}_1$ and $\varphi_2 : \mathcal{Y}^n \to \mathcal{M}_2$, and a decoder $\psi : \mathcal{M}_1 \times \mathcal{M}_2 \to \mathcal{X}^n$. We shall evaluate the correct decoding probability:

$$p_c := \Pr\big(\psi(\varphi_1(X^n), \varphi_2(Y^n)) = X^n\big).$$

A rate pair (R_1, R_2) is defined to be achievable if, for every $0 < \varepsilon < 1$ and for sufficiently large n, there exists a code $(\varphi_1, \varphi_2, \psi)$ such that the correct decoding probability satisfies $p_c \ge 1 - \varepsilon$ and rate pair satisfies $\frac{1}{n}\log|\mathcal{M}_1| \le R_1$ and $\frac{1}{n}\log|\mathcal{M}_2| \le R_2$, respectively. Then, the achievable region $\mathcal{R}_{\mathtt{WAK}}(P_{XY})$ is defined as the closure of all achievable rate pairs. It is well known that the achievable region is characterized as

$$\begin{aligned}
\mathcal{R}_{\mathtt{WAK}}(P_{XY}) = \big\{(R_1,R_2) : \exists P_{U|Y} \in \mathcal{P}(\mathcal{U}|\mathcal{Y}) \text{ s.t.} \\
R_1 \ge H(X|U), R_2 \ge I(U \wedge Y)\big\}
\end{aligned}$$

where $\mathcal{P}(\mathcal{U}|\mathcal{Y})$ is the set of all channels from $\mathcal{Y}$ to an auxiliary alphabet $\mathcal{U}$ satisfying $|\mathcal{U}| \le |\mathcal{Y}| + 1$.

Note that the characterization of the achievable region involves an auxiliary random variable U that does not appear in the problem setting. Furthermore, U is generated only from Y via channel $P_{U|Y}$; in other words, U, Y, and X must satisfy the Markov chain. In many cases, difficulty of analyzing multi-user problems stem from the existence of auxiliary random variables and Markov chain constraints, and the WAK problem is the most basic problem involving such difficulties.

The following exponential strong converse of the WAK problem was obtained in [14].

Proposition 3 For any code $(\varphi_1, \varphi_2, \psi)$, the correct decoding probability satisfies

$$\begin{aligned}
&\frac{1}{n}\log(1/p_c) \\
&\ge \min_{P_{\tilde{U}\tilde{X}\tilde{Y}}} \big\{D(P_{\tilde{U}\tilde{X}\tilde{Y}}\|P_{\tilde{U}|\tilde{Y}}P_{XY}) + |I(\tilde{U} \wedge \tilde{Y}) - R_2|^+ : \\
&\qquad\qquad R_1 \ge H(\tilde{X}|\tilde{U})\big\},
\end{aligned}$$

where the minimization is taken over joint distributions on $\mathcal{U}\times\mathcal{X}\times\mathcal{Y}$ for an auxiliary alphabet satisfying $|\mathcal{U}| \le |\mathcal{X}||\mathcal{Y}|+2$.

For the proof, see [14]; furthermore, it can be proved that the bound in Proposition 3 is asymptotically tight.

In contrast to the characterization of the achievable region, the exponent in Proposition 3 does not involve the Markov chain constraint. In fact, we can decompose the divergence term as

$$D(P_{\tilde{U}\tilde{X}\tilde{Y}}\|P_{\tilde{U}|\tilde{Y}}P_{XY}) = D(P_{\tilde{X}\tilde{Y}}\|P_{XY}) + I(\tilde{U} \wedge \tilde{X}|\tilde{Y}).$$

Thus, in the analysis of the strong converse exponent, the Markov chain constraint is imposed as a (potentially non-zero) penalty term. The idea of introducing this kind of penalty term rather than the exact Markov chain constraint was proposed by Oohama in [11], which culminated in the tight strong exponent of the WAK problem in [14].

References

[1] R. Ahlswede and J. Körner, "Source coding with side information and a converse for the degraded broadcast channel," *IEEE Trans. Inform. Theory*, vol. 21, no. 6, pp. 629–637, November 1975.

[2] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. 19, no. 3, pp. 357–359, May 1973.

[3] I. Csiszár and J. Körner, *Information Theory, Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.

[2]Note that $\tilde{M}$ may not be a function of $\tilde{X}^n$ when the encoder is not one-to-one, and $I(\tilde{M} \wedge \tilde{Y}^n|\tilde{X}^n)$ may not be 0.

[4] G. Dueck and J. Körner, "Reliability function of a discrete memoryless channel at rates above capacity," *IEEE Trans. Inform. Theory*, vol. 25, no. 1, pp. 82–85, January 1979.

[5] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *Bell Labs. Technical Journal*, vol. 53, no. 9, pp. 1681–1721, November 1974.

[6] W. Gu and M. Effros, "A strong converse for a collection of network source coding problems," in *Proc. IEEE Int. Symp. Inf. Theory 2009*, 2009, pp. 2316–2320.

[7] ——, "A strong converse in source coding for super-source networks," in *Proc. IEEE Int. Symp. Inf. Theory 2011*, 2011, pp. 395–399.

[8] M. Hamad, M. A. Wigger, and M. Sarkiss, "Strong converse using typical changes of measures and asymptotic markov chains," 2023, arXiv:2301.06289.

[9] K. Marton, "A simple proof of the blowing-up lemma," *IEEE Trans. Inform. Theory*, vol. 32, no. 3, pp. 445–446, May 1986.

[10] Y. Oohama, "On two strong converse theorems for discrete memoryless channels," *IEICE Trans. Fundamentals*, vol. E98-A, no. 12, pp. 2471–2475, December 2015.

[11] ——, "Exponential strong converse for source coding with side information at the decoder," *Entropy*, vol. 20, no. 5, p. 352, April 2018.

[12] ——, "Exponential strong converse for one helper source coding problem," *Entropy*, vol. 21, no. 6, p. 567, June 2019.

[13] Y. Oohama and T. S. Han, "Universal coding for the Slepian-Wolf data compression system and the strong converse theorem," *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 1908–1919, November 1994.

[14] D. Takeuchi and S. Watanabe, "Tight exponential strong converse for source coding problem with encoded side information," in *Proc. IEEE Int. Symp. Inf. Theory 2023*, Taipei, Taiwan, 2023, pp. 1366–1371.

[15] H. Tyagi and S. Watanabe, "Strong converse using change of measure arguments," *IEEE Trans. Inform. Theory*, vol. 66, no. 2, pp. 689–703, February 2020.

[16] S. Watanabe, "A converse bound on Wyner-Ahlswede-Körner network via Gray-Wyner network," in *Proc. 2017 IEEE Information Theory Workshop (ITW)*, Kaohsiung, Taiwan, 2017.

[17] ——, "Second-order region for Gray-Wyner network," *IEEE Trans. Inform. Theory*, vol. 63, no. 2, pp. 1006–1018, February 2017.

[18] A. D. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 21, no. 3, pp. 294–300, May 1975.