/

Article / Book Information

Title	Non-Asymptotic Bounds and a General Formula for the Rate-Distortion Region of the Successive Refinement Problem
Authors	Tetsunao Matsuta, Tomohiko Uyematsu
Citation	IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Vol. E101-A, No. 12, pp. 2110-2124
Pub. date	2018, 12
URL	http://search.ieice.org/
Copyright	(c) 2018 Institute of Electronics, Information and Communication Engineers

Non-Asymptotic Bounds and a General Formula for the Rate-Distortion Region of the Successive Refinement Problem*

Tetsunao MATSUTA[†a], ***Member and* Tomohiko UYEMATSU**[†b]***, Fellow*

SUMMARY   In the successive refinement problem, a fixed-length sequence emitted from an information source is encoded into two codewords by two encoders in order to give two reconstructions of the sequence. One of two reconstructions is obtained by one of two codewords, and the other reconstruction is obtained by all two codewords. For this coding problem, we give non-asymptotic inner and outer bounds on pairs of numbers of codewords of two encoders such that each probability that a distortion exceeds a given distortion level is less than a given probability level. We also give a general formula for the rate-distortion region for general sources, where the rate-distortion region is the set of rate pairs of two encoders such that each maximum value of possible distortions is less than a given distortion level.
key words: general source, information spectrum, non-asymptotic bound, rate-distortion region, successive refinement

1. Introduction

The successive refinement problem is a fixed-length lossy source coding problem with many terminals (see Fig. 1). In this coding problem, a fixed-length sequence emitted from an information source is encoded into two codewords by two encoders in order to give two reconstructions of the sequence. One of two reconstructions is obtained by one of two codewords by using a decoder, and the other reconstruction is obtained by all two codewords by using the other decoder.

An important parameter of the successive refinement problem is a pair of *rates* of two encoders such that each distortion between the source sequence and a reconstruction is less than a given distortion level. The set of these pairs when the length (blocklength) of the source sequence is unlimited is called the *rate-distortion region*. Since a codeword is used in both decoders, we cannot always optimize rates like the case where each codeword is used for each reconstruction separately. However, there are some cases where we can achieve the optimum rates. Necessary and sufficient conditions for such cases were independently given by Koshelev [3], [4] and Equitz and Cover [5]. The complete characterization of the rate-distortion region for discrete stationary memoryless sources was given by Rimoldi [6]. Yamamoto [7] also gave the rate-distortion region as a special case of
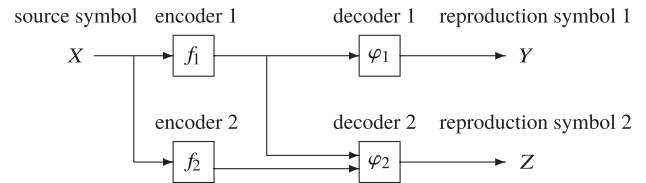

Fig. 1   Successive refinement problem.

a more general coding problem. Later, Effros [8] characterized the rate-distortion region for discrete stationary ergodic and non-ergodic sources.

Recently, the asymptotic analysis of the second-order rates to the blocklength becomes an active target of the study. Especially, for the successive refinement problem, No et al. [9] and Zhou et al. [10] gave a lot of results to the set of second-order rates for discrete and Gaussian stationary memoryless sources. No et al. [9] considered *separate* excess-distortion criteria such that a probability that a distortion exceeds a given distortion level is less than a given probability level separately for each reconstruction. On the other hand, Zhou et al. [10] considered the *joint* excess-distortion criterion such that a probability that either of distortions exceeds a given distortion level is less than a given probability level. Although they also gave several non-asymptotic bounds on the set of pairs of rates, they mainly focus on the asymptotic behavior of the set.

On the other hand, in this paper, we consider non-asymptotic bounds on pairs of rates in finite blocklengths. Especially, since a rate is easily calculated by a number of codewords, we focus on pairs of two numbers of codewords. Although we adopt separate excess-distortion criteria, our result can be easily applied to the joint excess-distortion criterion. We give inner and outer bounds on pairs of numbers of codewords. These bounds are characterized by using the *smooth max Rényi divergence* introduced by Warsi [11]. For the point-to-point lossy source coding problem, we also used the smooth max Rényi divergence to characterize the rate-distortion function which is the minimum rate when the blocklength is unlimited [12]. Proof techniques are similar to that of [12], but we employ several extended results for the successive refinement problem. The inner bound is derived by using an extended version of the previous lemma [12, Lemma 2]. We give this lemma as a special case of an extended version of the previous generalized covering lemma [13, Lemma 1]. The outer bound is derived by using an extended version of the previous converse bound

[12, Lemma 4].

In this paper, we also consider the rate-distortion region for general sources. In this case, we adopt the maximum-distortion criterion such that the maximum value of possible distortion is less than a given distortion level for each reconstruction. By using the spectral sup-mutual information rate (cf. [14]) and the non-asymptotic inner and outer bounds, we give a general formula for the rate-distortion region. We show that our rate-distortion region coincides with the region obtained by Rimoldi [6] when a source is discrete stationary memoryless. Furthermore, we consider a mixed source which is a mixture of two sources and show that the rate-distortion region is the intersection of those of two sources.

The rest of this paper is organized as follows. In Sect. 2, we provide some notations and the formal definition of the successive refinement problem. In Sect. 3, we give several lemmas for an inner bound on pairs of numbers of codewords and the rate-distortion region. These lemmas are extended versions of our previous results [12, Lemma 2] and [13, Lemma 1]. In Sect. 4, we give outer and inner bounds using the smooth max Rényi divergence on pairs of numbers of codewords. In Sect. 5, we give a general formula for the rate-distortion region. In this section, we consider the rate-distortion region for discrete stationary memoryless sources and mixed sources. In Sect. 6, we conclude the paper.

2. Preliminaries

Let $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{R}_{\geq 0}$ be sets of positive integers, real numbers, and non-negative real numbers, respectively.

Unless otherwise stated, we use the following notations. For a pair of integers $i \leq j$, the set of integers $\{i, i+1, \cdots, j\}$ is denoted by $[i : j]$. For finite or countably infinite sets $\mathcal{X}$ and $\mathcal{Y}$, the set of all probability distributions over $\mathcal{X}$ and $\mathcal{X} \times \mathcal{Y}$ are denoted by $\mathcal{P}_{\mathcal{X}}$ and $\mathcal{P}_{\mathcal{X}\mathcal{Y}}$, respectively. The set of all conditional probability distributions over $\mathcal{X}$ given $\mathcal{Y}$ is denoted by $\mathcal{P}_{\mathcal{X}|\mathcal{Y}}$. The probability distribution of a random variable (RV) X is denoted by the subscript notation P_X, and the conditional probability distribution for X given an RV Y is denoted by $P_{X|Y}$. The n-fold Cartesian product of a set $\mathcal{X}$ is denoted by $\mathcal{X}^n$ while an n-length sequence of symbols $(a_1, a_2, \cdots, a_n)$ is denoted by a^n. The sequence of RVs $\{X^n\}_{n=1}^{\infty}$ is denoted by the bold-face letter $\mathbf{X}$. Sequences of probability distributions $\{P_{X^n}\}_{n=1}^{\infty}$ and conditional probability distributions $\{P_{X^n|Y^n}\}_{n=1}^{\infty}$ are denoted by bold-face letters $\mathbf{P_X}$ and $\mathbf{P_{X|Y}}$, respectively.

For the successive refinement problem (Fig. 1), let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be finite or countably infinite sets, where $\mathcal{X}$ represents the source alphabet, and $\mathcal{Y}$ and $\mathcal{Z}$ represent two reconstruction alphabets. Let X over $\mathcal{X}$ be an RV which represents a single source symbol. Since the source can be characterized by X, we also refer to it as the source. When we consider $\mathcal{X}$ as an n-fold Cartesian product of a certain finite or countably infinite set, we can regard the source symbol X as an n-length source sequence. Thus, for the sake of brevity, we deal with the single source symbol unless otherwise stated.

Two encoders encoder 1 and encoder 2 are represented as functions $f_1 : \mathcal{X} \to [1 : M_1]$ and $f_2 : \mathcal{X} \to [1 : M_2]$, respectively, where M_1 and M_2 are positive integers which denote numbers of codewords. Two decoders decoder 1 and decoder 2 are represented as functions $\varphi_1 : [1 : M_1] \to \mathcal{Y}$ and $\varphi_2 : [1 : M_1] \times [1 : M_2] \to \mathcal{Z}$, respectively. We refer to a tuple of encoders and decoders $(f_1, f_2, \varphi_1, \varphi_2)$ as a *code*. In order to measure distortions between the source symbol and reconstruction symbols, we introduce distortion measures defined by functions $d_1 : \mathcal{X} \times \mathcal{Y} \to [0, +\infty)$ and $d_2 : \mathcal{X} \times \mathcal{Z} \to [0, +\infty)$.

We define two events of exceeding given distortion levels $D_1 \geq 0$ and $D_2 \geq 0$ as follows:

$$\mathcal{E}_1(D_1) \triangleq \{d_1(X, \varphi_1(f_1(X))) > D_1\},$$
$$\mathcal{E}_2(D_2) \triangleq \{d_2(X, \varphi_2(f_1(X), f_2(X))) > D_2\}.$$

Then, we define the achievability under the excess-distortion criterion.

Definition 1. For positive integers M_1, M_2, real numbers $D_1, D_2 \geq 0$, and $\epsilon_1, \epsilon_2 \in [0, 1]$, let $M = (M_1, M_2)$, $D = (D_1, D_2)$, and $\epsilon = (\epsilon_1, \epsilon_2)$. Then, for a source X, we say (M, D) is ϵ-*achievable* if and only if there exists a code $(f_1, f_2, \varphi_1, \varphi_2)$ such that numbers of codewords of encoder 1 and encoder 2 are M_1 and M_2, respectively, and

$$\Pr\{\mathcal{E}_i(D_i)\} \leq \epsilon_i, \quad \forall i \in \{1, 2\}.$$

In what follows, for constants $M_1, M_2, D_1, D_2, \epsilon_1$, and ϵ_2, we often use the above simple notations: $M = (M_1, M_2)$, $D = (D_1, D_2)$, and $\epsilon = (\epsilon_1, \epsilon_2)$. In this setting, we consider the set of all pairs (M_1, M_2) of numbers of codewords under the excess-distortion criterion. According to the ϵ-achievability, this set is defined as follows:

Definition 2. For a source X, real numbers $D_1, D_2 \geq 0$, and $\epsilon_1, \epsilon_2 \in [0, 1]$, we define

$$\mathcal{M}(D, \epsilon | X) \triangleq \{(M_1, M_2) \in \mathbb{N}^2 : (M, D) \text{ is } \epsilon\text{-achievable}\}.$$

Basically, this paper deals with a coding for a single source symbol. However, in Sect. 5, we deal with the coding for an n-length source sequence. Hence in that section, by abuse of notation, we regard the above sets $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ as n-fold Cartesian products $\mathcal{X}^n$, $\mathcal{Y}^n$, and $\mathcal{Z}^n$, respectively. We also regard source symbol X on $\mathcal{X}$ as an n-length source sequence X^n on $\mathcal{X}^n$. Then we call the sequence $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ of source sequences the *general source* that is not required to satisfy the consistency condition.

We use the superscript (n) for a code, distortion measures, and numbers of codewords (e.g., $(f_1^{(n)}, f_2^{(n)}, \varphi_1^{(n)}, \varphi_2^{(n)})$) to make clear that we are dealing with source sequences of length n. For a code, we define rates $R_1^{(n)}$ and $R_2^{(n)}$ as

$$R_i^{(n)} \triangleq \frac{1}{n} \log M_i^{(n)}, \quad \forall i \in \{1, 2\}.$$

Hereafter, log means the natural logarithm.

We introduce maximum distortions for a sequence of codes. To this end, we define the limit superior in probability [14].

Definition 3 (Limit superior in probability). For an arbitrary sequence $\mathbf{S} = \{S^n\}_{n=1}^{\infty}$ of real-valued RVs, we define the limit superior in probability by

$$\operatorname*{p-lim\,sup}_{n \to \infty} S_n \triangleq \inf \left\{ \alpha : \lim_{n \to \infty} \Pr \{S_n > \alpha\} = 0 \right\}.$$

Now we introduce the maximum distortions:

$$\operatorname*{p-lim\,sup}_{n \to \infty} d_1^{(n)}(X^n, \varphi_1^{(n)}(f_1^{(n)}(X^n))),$$

$$\operatorname*{p-lim\,sup}_{n \to \infty} d_2^{(n)}(X^n, \varphi_2^{(n)}(f_1^{(n)}(X^n), f_2^{(n)}(X^n))).$$

Then, we define the achievability under the maximum distortion criterion.

Definition 4. For real numbers $R_1, R_2 \geq 0$, let $R = (R_1, R_2)$. Then, for a general source $\mathbf{X}$, and real numbers $D_1, D_2 \geq 0$, we say a pair (R, D) is *fm-achievable* if and only if there exists a sequence $\{(f_1^{(n)}, f_2^{(n)}, \varphi_1^{(n)}, \varphi_2^{(n)})\}$ of codes satisfying

$$\operatorname*{p-lim\,sup}_{n \to \infty} d_1^{(n)}(X^n, \varphi_1^{(n)}(f_1^{(n)}(X^n))) \leq D_1,$$

$$\operatorname*{p-lim\,sup}_{n \to \infty} d_2^{(n)}(X^n, \varphi_2^{(n)}(f_1^{(n)}(X^n), f_2^{(n)}(X^n))) \leq D_2,$$

and

$$\limsup_{n \to \infty} R_i^{(n)} \leq R_i, \quad \forall i \in \{1, 2\}.$$

In what follows, for constants R_1 and R_2, we often use the above simple notation: $R = (R_1, R_2)$. In this setting, we consider the set of all rate pairs under the maximum distortion criterion. According to the fm-achievability, this set, usually called the rate-distortion region, is defined as follows:

Definition 5 (Rate-distortion region). For a general source $\mathbf{X}$ and real numbers $D_1, D_2 \geq 0$, we define

$$\mathcal{R}(D|\mathbf{X}) \triangleq \{(R_1, R_2) \in \mathbb{R}_{\geq 0}^2 : (R, D) \text{ is fm-achievable}\}.$$

Remark 1. We can show that the rate-distortion region $\mathcal{R}(D|\mathbf{X})$ is a closed set by the definition and using the diagonal line argument (cf. [14]).

We note that when we regard X as n-length sequence in the definition of $\mathcal{M}(D, \epsilon|X)$, it gives a non-asymptotic region of pairs of rates for a given finite blocklength.

3. Covering Lemma

In this section, we introduce some useful lemmas and corollaries for an inner bound on the set $\mathcal{M}(D, \epsilon|X)$ and $\mathcal{R}(D|\mathbf{X})$.

The next lemma is the most basic and important result in the sense that all subsequent results in this section are given by this lemma.

Lemma 1. Let $A \in \mathcal{A}$ be an arbitrary RV, and $\tilde{B} \in \mathcal{B}$ and $\tilde{C} \in C$ be RVs such that the pair $(\tilde{B}, \tilde{C})$ is independent of A. For an integer $M_1 \geq 1$, let $\tilde{B}_1, \tilde{B}_2, \cdots, \tilde{B}_{M_1}$ be RVs which are independent of each other and of A, and each distributed according to $P_{\tilde{B}}$. For an integer $i \in [1 : M_1]$ and $M_2 \geq 1$, let $\tilde{C}_{i,1}, \tilde{C}_{i,2}, \cdots, \tilde{C}_{i,M_2}$ be RVs which are independent of each other and of A, and each distributed according to $P_{\tilde{C}|\tilde{B}}(\cdot|\tilde{B}_i)$. Then, for any set $\mathcal{F} \subseteq \mathcal{A} \times \mathcal{B} \times C$, we have

$$\Pr \left\{ \bigcap_{i=1}^{M_1} \bigcap_{j=1}^{M_2} \{(A, \tilde{B}_i, \tilde{C}_{i,j}) \notin \mathcal{F}\} \right\}$$

$$= \mathrm{E} \left[\mathrm{E} \left[\mathrm{E} \left[\mathbf{1}\{(A, \tilde{B}, \tilde{C}) \notin \mathcal{F}\} \big| A, \tilde{B} \right]^{M_2} \big| A \right]^{M_1} \right], \quad (1)$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function, $\mathrm{E}[\cdot]$ denotes the expectation operator, and $\mathrm{E}[\cdot]^M$ denotes the M-th power of the expectation, i.e., $\mathrm{E}[\cdot]^M = (\mathrm{E}[\cdot])^M$.

Proof. We have

$$\Pr \left\{ \bigcap_{i=1}^{M_1} \bigcap_{j=1}^{M_2} \{(A, \tilde{B}_i, \tilde{C}_{i,j}) \notin \mathcal{F}\} \right\}$$

$$= \sum_{a \in \mathcal{A}} \sum_{(\tilde{b}_1 \cdots, \tilde{b}_{M_1}) \in \mathcal{B}^{M_1}} \sum_{(\tilde{c}_{1,1}, \cdots, \tilde{c}_{1,M_2}, c_{2,1} \cdots, \tilde{c}_{M_1, M_2}) \in C^{M_1 M_2}}$$

$$\times \left(P_A(a) \prod_{i=1}^{M_1} P_{\tilde{B}}(\tilde{b}_i) \prod_{j=1}^{M_2} P_{\tilde{C}|\tilde{B}}(\tilde{c}_{i,j}|\tilde{b}_i) \right)$$

$$\times \left(\prod_{i=1}^{M_1} \prod_{j=1}^{M_2} \mathbf{1}\{(a, \tilde{b}_i, \tilde{c}_{i,j}) \notin \mathcal{F}\} \right)$$

$$= \sum_{a \in \mathcal{A}} \sum_{(\tilde{b}_1 \cdots, \tilde{b}_{M_1}) \in \mathcal{B}^{M_1}} \sum_{(\tilde{c}_{1,1}, \cdots, \tilde{c}_{1,M_2}, c_{2,1} \cdots, \tilde{c}_{M_1, M_2}) \in C^{M_1 M_2}}$$

$$\times P_A(a) \prod_{i=1}^{M_1} P_{\tilde{B}}(\tilde{b}_i) \prod_{j=1}^{M_2} P_{\tilde{C}|\tilde{B}}(\tilde{c}_{i,j}|\tilde{b}_i)$$

$$\times \mathbf{1}\{(a, \tilde{b}_i, \tilde{c}_{i,j}) \notin \mathcal{F}\}$$

$$= \sum_{a \in \mathcal{A}} P_A(a) \prod_{i=1}^{M_1} \sum_{\tilde{b}_i \in \mathcal{B}} P_{\tilde{B}}(\tilde{b}_i) \prod_{j=1}^{M_2} \sum_{\tilde{c}_{i,j} \in C} P_{\tilde{C}|\tilde{B}}(\tilde{c}_{i,j}|\tilde{b}_i)$$

$$\times \mathbf{1}\{(a, \tilde{b}_i, \tilde{c}_{i,j}) \notin \mathcal{F}\}$$

$$= \sum_{a \in \mathcal{A}} P_A(a) \prod_{i=1}^{M_1} \sum_{\tilde{b} \in \mathcal{B}} P_{\tilde{B}}(\tilde{b}) \prod_{j=1}^{M_2} \sum_{\substack{\tilde{c} \in C: \\ (a,\tilde{b},\tilde{c}) \notin \mathcal{F}}} P_{\tilde{C}|\tilde{B}}(\tilde{c}|\tilde{b})$$

$$= \sum_{a \in \mathcal{A}} P_A(a) \left(\sum_{\tilde{b} \in \mathcal{B}} P_{\tilde{B}}(\tilde{b}) \left(\sum_{\substack{\tilde{c} \in C: \\ (a,\tilde{b},\tilde{c}) \notin \mathcal{F}}} P_{\tilde{C}|\tilde{B}}(\tilde{c}|\tilde{b}) \right)^{M_2} \right)^{M_1}.$$

By recalling that $(\tilde{B}, \tilde{C})$ is independent of A, this coincides with the right-hand side (RHS) of (1). □

This lemma implies an exact analysis of the error prob-

ability of covering a set $\mathcal{A}$ in terms of a given condition $\mathcal{F}$ by codewords $\{\tilde{B}_i\}$ and $\{\tilde{C}_{i,j}\}$ of random coding. Hence, this lemma can be regarded as an extended version of [15, Theorem 9].

Although the above lemma gives an exact analysis, it is difficult to use it for characterizing an inner bound on pairs of numbers of codewords and the rate-distortion region. Instead of it, we will use the next convenient lemma.

Lemma 2. Let $A \in \mathcal{A}$, $B \in \mathcal{B}$, and $C \in C$ be arbitrary RVs, and $\tilde{B} \in \mathcal{B}$ and $\tilde{C} \in C$ be RVs such that the pair $(\tilde{B}, \tilde{C})$ is independent of A. Let $\psi_1 : \mathcal{A} \times \mathcal{B} \to [0, 1]$ be a function and $\alpha_1 \in [0, 1]$ be a constant such that

$$P_A(a)P_{\tilde{B}}(b) \geq \alpha_1 \psi_1(a, b)P_{AB}(a, b),$$
$$\forall (a, b) \in \mathcal{A} \times \mathcal{B}. \tag{2}$$

Furthermore, let $\psi_2 : \mathcal{A} \times \mathcal{B} \times C \to [0, 1]$ be a function and $\alpha_2 \in [0, 1]$ be a constant such that

$$P_{AB}(a, b)P_{\tilde{C}|\tilde{B}}(c|b) \geq \alpha_2 \psi_2(a, b, c)P_{ABC}(a, b, c),$$
$$\forall (a, b, c) \in \mathcal{A} \times \mathcal{B} \times C. \tag{3}$$

Then, for any set $\mathcal{F} \subseteq \mathcal{A} \times \mathcal{B} \times C$, we have

$$E\left[E\left[E\left[\mathbf{1}\{(A, \tilde{B}, \tilde{C}) \notin \mathcal{F}\}\Big|A, \tilde{B}\right]^{M_2}\Big|A\right]^{M_1}\right]$$
$$\leq 1 - E[\psi_1(A, B)\psi_2(A, B, C)] + \Pr\{(A, B, C) \notin \mathcal{F}\}$$
$$+ e^{-\alpha_2 M_2 - \log \alpha_1} + e^{-\alpha_1 M_1}. \tag{4}$$

Proof. We have

$$E\left[E\left[E\left[\mathbf{1}\{(A, \tilde{B}, \tilde{C}) \notin \mathcal{F}\}\Big|A, \tilde{B}\right]^{M_2}\Big|A\right]^{M_1}\right]$$
$$= \sum_{\substack{a \in \mathcal{A}: \\ P_A(a) > 0}} P_A(a)\left(\sum_{b \in \mathcal{B}} P_{\tilde{B}}(b)\right.$$
$$\times \left.\left(1 - \sum_{\substack{c \in C: \\ (a,b,c) \in \mathcal{F}}} P_{\tilde{C}|\tilde{B}}(c|b)\right)^{M_2}\right)^{M_1}$$
$$\overset{(a)}{\leq} \sum_{\substack{a \in \mathcal{A}: \\ P_A(a) > 0}} P_A(a)\left(\sum_{b \in \mathcal{B}} P_{\tilde{B}}(b)\right.$$
$$\times \left.\left(1 - \alpha_2 \sum_{\substack{c \in C: \\ (a,b,c) \in \mathcal{F}, \\ P_{B|A}(b|a) > 0}} \psi_2(a, b, c)P_{C|AB}(c|a, b)\right)^{M_2}\right)^{M_1}$$
$$\overset{(b)}{\leq} \sum_{\substack{a \in \mathcal{A}: \\ P_A(a) > 0}} P_A(a)\left(\sum_{b \in \mathcal{B}} P_{\tilde{B}}(b)\left(1 - \sum_{\substack{c \in C: \\ (a,b,c) \in \mathcal{F}, \\ P_{B|A}(b|a) > 0}} \psi_2(a, b, c)\right.\right.$$

$$\times P_{C|AB}(c|a, b) + e^{-\alpha_2 M_2}\Big)\Big)^{M_1}$$
$$\overset{(c)}{\leq} \sum_{\substack{a \in \mathcal{A}: \\ P_A(a) > 0}} P_A(a)\left(1 - \sum_{\substack{b \in \mathcal{B}: \\ P_{B|A}(b|a) > 0}} \alpha_1 \psi_1(a, b)P_{B|A}(b|a)\right.$$
$$\times \left.\sum_{\substack{c \in C: \\ (a,b,c) \in \mathcal{F}}} \psi_2(a, b, c)P_{C|AB}(c|a, b) + e^{-\alpha_2 M_2}\right)^{M_1}$$
$$= \sum_{\substack{a \in \mathcal{A}: \\ P_A(a) > 0}} P_A(a)\left(1 - \alpha_1 \left(\sum_{\substack{(b,c) \in \mathcal{B} \times C: \\ (a,b,c) \in \mathcal{F}, \\ P_{B|A}(b|a) > 0}} \psi_1(a, b)\psi_2(a, b, c)\right.\right.$$
$$\times \left.\left. P_{BC|A}(b, c|a) - e^{-\alpha_2 M_2 - \log \alpha_1}\right)\right)^{M_1},$$

where (a) comes from (3), (b) follows since $(1 - xy)^M \leq 1 - y + e^{-xM}$ for $0 \leq x, y \leq 1$ and $M > 0$ (cf. [16, Lemma 10.5.3]), and (c) comes from (2). Since the probability is not greater than 1, we have

$$E\left[E\left[E\left[\mathbf{1}\{(A, \tilde{B}, \tilde{C}) \notin \mathcal{F}\}\Big|A, \tilde{B}\right]^{M_2}\Big|A\right]^{M_1}\right]$$
$$\leq \sum_{\substack{a \in \mathcal{A}: \\ P_A(a) > 0}} P_A(a)\left(1 - \alpha_1 \left|\sum_{\substack{(b,c) \in \mathcal{B} \times C: \\ (a,b,c) \in \mathcal{F}, \\ P_{B|A}(b|a) > 0}} \psi_1(a, b)\psi_2(a, b, c)\right.\right.$$
$$\times \left.\left. P_{BC|A}(b, c|a) - e^{-\alpha_2 M_2 - \log \alpha_1}\right|^+\right)^{M_1}$$
$$\overset{(a)}{\leq} \sum_{\substack{a \in \mathcal{A}: \\ P_A(a) > 0}} P_A(a)\left(1 - \left|\sum_{\substack{(b,c) \in \mathcal{B} \times C: \\ (a,b,c) \in \mathcal{F}, \\ P_{B|A}(b|a) > 0}} \psi_1(a, b)\psi_2(a, b, c)\right.\right.$$
$$\times P_{BC|A}(b, c|a) - e^{-\alpha_2 M_2 - \log \alpha_1}\Big|^+ + e^{-\alpha_1 M_1}\Big)$$
$$\leq 1 - \sum_{\substack{(a,b,c) \in \mathcal{A} \times \mathcal{B} \times C: \\ (a,b,c) \in \mathcal{F}, P_{AB}(a,b) > 0}} \psi_1(a, b)\psi_2(a, b, c)P_{ABC}(a, b, c)$$
$$+ e^{-\alpha_2 M_2 - \log \alpha_1} + e^{-\alpha_1 M_1}$$
$$\overset{(b)}{\leq} 1 - E[\psi_1(A, B)\psi_2(A, B, C)] + \Pr\{(A, B, C) \notin \mathcal{F}\}$$
$$+ e^{-\alpha_2 M_2 - \log \alpha_1} + e^{-\alpha_1 M_1},$$

where $|x|^+ = \max\{0, x\}$, (a) follows since $(1 - xy)^M \leq 1 - y + e^{-xM}$ for $0 \leq x, y \leq 1$ and $M > 0$, and (b) comes from the fact that $\psi_1(a, b)\psi_2(a, b, c) \leq 1$. $\square$

The importance of this lemma is to be able to change RVs from $(A, \tilde{B}, \tilde{C})$ to arbitrary correlated RVs (A, B, C). This makes it possible to characterize an inner bound on pairs of numbers of codewords and the rate-distortion region.

Lemma 2 can be regarded as an extended version of our previous lemma [13, Lemma 1] to multiple correlated RVs. Hence, like the previous lemma, by changing functions and constants, it gives many types of bounds such as the following two corollaries.

Corollary 1. For any set $\mathcal{F} \subseteq \mathcal{A} \times \mathcal{B} \times \mathcal{C}$, any real numbers $\gamma_1, \gamma_2 \in \mathbb{R}$, and any integers $M_1, M_2 \geq 1$ such that $M_1 \geq \exp(\gamma_1)$ and $M_2 \geq \exp(\gamma_2)$, we have

$$
\mathrm{E}\left[\mathrm{E}\left[\mathrm{E}\left[\mathbf{1}\{(A, \tilde{B}, \tilde{C}) \notin \mathcal{F}\}\big|A, \tilde{B}\right]^{M_2}\Big|A\right]^{M_1}\right]
$$

$$
\leq \Pr\left\{\log \frac{P_{B|A}(B|A)}{P_{\tilde{B}}(B)} > \log M_1 - \gamma_1\right.
$$

$$
\left. \text{or } \log \frac{P_{C|AB}(C|A, B)}{P_{\tilde{C}|\tilde{B}}(C|B)} > \log M_2 - \gamma_2\right\}
$$

$$
+ \Pr\{(A, B, C) \notin \mathcal{F}\} + e^{-\exp(\gamma_2) - \gamma_1 + \log M_1} + e^{-\exp(\gamma_1)}.
$$

Proof. Let $\alpha_1 = \frac{\exp(\gamma_1)}{M_1}$, $\alpha_2 = \frac{\exp(\gamma_2)}{M_2}$,

$$
\psi_1(a, b) = \mathbf{1}\left\{\log \frac{P_{B|A}(b|a)}{P_{\tilde{B}}(b)} \leq \log M_1 - \gamma_1\right\},
$$

$$
\psi_2(a, b, c) = \mathbf{1}\left\{\log \frac{P_{C|AB}(c|a, b)}{P_{\tilde{C}|\tilde{B}}(c|b)} \leq \log M_2 - \gamma_2\right\}.
$$

Then, we can easily check that these constants and functions satisfy (2) and (3). Plugging these functions and constants into (4), we have the desired bound. □

This corollary can be regarded as a bound in terms of the *information spectrum* (cf. [14]). To the best of our knowledge, this type of bound has not been reported so far (although, there are some converse bounds [10, Lemma 15] and [17, Theorem 3]).

On the other hand, the next corollary gives a bound in terms of the smooth max Rényi divergence $D_\infty^\delta(P\|Q)$ defined as

$$
D_\infty^\delta(P\|Q) \triangleq \inf_{\substack{\psi: \mathcal{A} \to [0,1]: \\ \sum_{a \in \mathcal{A}} \psi(a)P(a) \geq 1-\delta}} \left|\log \sup_{a \in \mathcal{A}} \frac{\psi(a)P(a)}{Q(a)}\right|^+,
$$

where $|x|^+ = \max\{0, x\}$.

Corollary 2. For any set $\mathcal{F} \subseteq \mathcal{A} \times \mathcal{B} \times \mathcal{C}$, any real numbers $\delta_1, \delta_2 \geq 0$, and any integers $M_1, M_2 \geq 1$, we have

$$
\mathrm{E}\left[\mathrm{E}\left[\mathrm{E}\left[\mathbf{1}\{(A, \tilde{B}, \tilde{C}) \notin \mathcal{F}\}\big|A, \tilde{B}\right]^{M_2}\Big|A\right]^{M_1}\right]
$$

$$
\leq \delta_1 + \delta_2 + \Pr\{(A, B, C) \notin \mathcal{F}\}
$$

$$
+ e^{-\exp(-D_\infty^{\delta_2}(P_{ABC}\|P_{AB} \times P_{\tilde{C}|\tilde{B}}))M_2 + D_\infty^{\delta_1}(P_{AB}\|P_A \times P_{\tilde{B}})}
$$

$$
+ e^{-\exp(-D_\infty^{\delta_1}(P_{AB}\|P_A \times P_{\tilde{B}}))M_1}.
$$

Proof. For an arbitrarily fixed $\epsilon > 0$, let ψ_1 and ψ_2 be functions such that $\mathrm{E}[\psi_1(A, B)] \geq 1 - \delta_1$, $\mathrm{E}[\psi_2(A, B, C)] \geq 1 - \delta_2$,

$$
\bar{D}_\infty^{\delta_1}(P_{AB}\|P_A \times P_{\tilde{B}}|\psi_1) \leq D_\infty^{\delta_1}(P_{AB}\|P_A \times P_{\tilde{B}}) + \epsilon, \tag{5}
$$

and

$$
\bar{D}_\infty^{\delta_2}(P_{ABC}\|P_{AB} \times P_{\tilde{C}|\tilde{B}}|\psi_2)
$$

$$
\leq D_\infty^{\delta_2}(P_{ABC}\|P_{AB} \times P_{\tilde{C}|\tilde{B}}) + \epsilon, \tag{6}
$$

where

$$
\bar{D}_\infty^{\delta}(P\|Q|\psi) = \left|\log \sup_{a \in \mathcal{A}} \frac{\psi(a)P(a)}{Q(a)}\right|^+.
$$

Then, we have

$$
\mathrm{E}[\psi_1(A, B)\psi_2(A, B, C)]
$$

$$
\overset{(a)}{\geq} \mathrm{E}[\psi_1(A, B)] + \mathrm{E}[\psi_2(A, B, C)] - 1
$$

$$
\geq 1 - \delta_1 - \delta_2, \tag{7}
$$

where (a) follows since $xy \geq x + y - 1$ for $x, y \in [0, 1]$.

On the other hand, let α_1 and α_2 be constants such that

$$
\alpha_1 = \exp\left(-\bar{D}_\infty^{\delta_1}(P_{AB}\|P_A \times P_{\tilde{B}}|\psi_1)\right),
$$

$$
\alpha_2 = \exp\left(-\bar{D}_\infty^{\delta_2}(P_{ABC}\|P_{AB} \times P_{\tilde{C}|\tilde{B}}|\psi_2)\right).
$$

Then, for any $(a, b, c) \in \mathcal{A} \times \mathcal{B} \times \mathcal{C}$, we have

$$
\alpha_1 \psi_1(a, b) P_{AB}(a, b)
$$

$$
\leq \left(\inf_{(a,b) \in \mathcal{A} \times \mathcal{B}} \frac{P_A(a)P_{\tilde{B}}(b)}{\psi_1(a, b)P_{AB}(a, b)}\right) \psi_1(a, b)P_{AB}(a, b)
$$

$$
\leq P_A(a)P_{\tilde{B}}(b),
$$

and

$$
\alpha_2 \psi_2(a, b, c) P_{ABC}(a, b, c)
$$

$$
\leq \left(\inf_{(a,b,c) \in \mathcal{A} \times \mathcal{B} \times \mathcal{C}} \frac{P_{AB}(a, b)P_{\tilde{C}|\tilde{B}}(c|b)}{\psi_2(a, b, c)P_{ABC}(a, b, c)}\right)
$$

$$
\times \psi_2(a, b, c)P_{ABC}(a, b, c)
$$

$$
\leq P_{AB}(a, b)P_{\tilde{C}|\tilde{B}}(c|b).
$$

Thus, ψ_1, ψ_2, α_1, and α_2 satisfy (2) and (3).

Plugging these functions and constants into (4), we have

$$
\mathrm{E}\left[\mathrm{E}\left[\mathrm{E}\left[\mathbf{1}\{(A, \tilde{B}, \tilde{C}) \notin \mathcal{F}\}\big|A, \tilde{B}\right]^{M_2}\Big|A\right]^{M_1}\right]
$$

$$
\leq \delta_1 + \delta_2 + \Pr\{(A, B, C) \notin \mathcal{F}\}
$$

$$
+ e^{-\exp(-D_\infty^{\delta_2}(P_{ABC}\|P_{\tilde{C}|\tilde{B}})-\epsilon)M_2}
$$

$$
\times e^{D_\infty^{\delta_1}(P_{AB}\|P_A \times P_{\tilde{B}})+\epsilon}
$$

$$
+ e^{-\exp(-D_\infty^{\delta_1}(P_{AB}\|P_A \times P_{\tilde{B}})-\epsilon)M_1},
$$

where we use inequalities (5), (6), and (7). Since $\epsilon > 0$ is arbitrary, this completes the proof. □

Remark 2. The original definition of the smooth max Rényi divergence (cf. [12]) is as follows:

$$D_\infty^\delta(P\|Q)^- \triangleq \inf_{\substack{\psi:\mathcal{A}\to[0,1]:\\ \sum_{a\in\mathcal{A}}\psi(a)P(a)\geq 1-\delta}} \log\sup_{\mathcal{B}\subseteq\mathcal{A}} \frac{\sum_{b\in\mathcal{B}}\psi(b)P(b)}{\sum_{b\in\mathcal{B}}Q(b)}.$$

Since for non-negative real valued functions $f(b)$ and $g(b)$, it holds that (cf. e.g. [16, Lemma 16.7.1])

$$\frac{\sum_{b\in\mathcal{B}}f(b)}{\sum_{b\in\mathcal{B}}g(b)} \leq \sup_{b\in\mathcal{B}}\frac{f(b)}{g(b)},$$

$D_\infty^\delta(P\|Q)^-$ can be simply defined as

$$D_\infty^\delta(P\|Q)^- \triangleq \inf_{\substack{\psi:\mathcal{A}\to[0,1]:\\ \sum_{a\in\mathcal{A}}\psi(a)P(a)\geq 1-\delta}} \log\sup_{a\in\mathcal{A}} \frac{\psi(a)P(a)}{Q(a)}.$$

In this definition, it may be a negative value depending on δ. Since this case is meaningless in this study, we adopt $D_\infty^\delta(P\|Q)$. Here we also note that

$$D_\infty^\delta(P\|Q) = \left|D_\infty^\delta(P\|Q)^-\right|^+. \tag{8}$$

4. Inner and Outer Bounds on the Set of Pairs of Numbers of Codewords

In this section, we give outer and inner bounds on $\mathcal{M}(D,\epsilon|X)$ by using the smooth max Rényi divergence.

First of all, we show a bound on the probability of the two events $\mathcal{E}_1(D_1)$ and $\mathcal{E}_2(D_2)$ for the successive refinement problem. In what follows, let $\mathcal{U}$ be an arbitrary finite or countably infinite set.

Theorem 1. For a source X, let $(\tilde{U},\tilde{Y},\tilde{Z}) \in \mathcal{U}\times\mathcal{Y}\times\mathcal{Z}$ be RVs such that $(\tilde{U},\tilde{Y},\tilde{Z})$ is independent of X. Then, for any real numbers $D_1, D_2 \geq 0$, there exists a code $(f_1,f_2,\varphi_1,\varphi_2)$ such that numbers of codewords of encoder 1 and encoder 2 are M_1 and M_2, respectively, and

$$\Pr\{\mathcal{E}_1(D_1)\cup\mathcal{E}_2(D_2)\}$$
$$\leq \mathrm{E}\left[\mathrm{E}\left[\mathrm{E}\left[\mathbf{1}\{(X,\tilde{U},\tilde{Y},\tilde{Z})\notin\mathcal{D}\}\Big|X,\tilde{U},\tilde{Y}\right]^{M_2}\Big|X\right]^{M_1}\right],$$

where

$$\mathcal{D} = \{(x,u,y,z)\in X\times\mathcal{U}\times\mathcal{Y}\times\mathcal{Z} :$$
$$d_1(x,y)\leq D_1, d_2(x,z)\leq D_2\}.$$

Proof. We generate $(\tilde{u}_1,\tilde{y}_1),(\tilde{u}_2,\tilde{y}_2),\cdots,(\tilde{u}_{M_1},\tilde{y}_{M_1})\in\mathcal{U}\times\mathcal{Y}$ independently subject to the probability distribution $P_{\tilde{U}\tilde{Y}}$, and define the set $C_1 \triangleq \{(\tilde{u}_1,\tilde{y}_1),(\tilde{u}_2,\tilde{y}_2),\cdots,(\tilde{u}_{M_1},\tilde{y}_{M_1})\}$. For any $i\in[1:M_1]$, we generate $\tilde{z}_{i,1},\tilde{z}_{i,2},\cdots,\tilde{z}_{i,M_2}\in\mathcal{Z}$ independently subject to the probability distribution $P_{\tilde{Z}|\tilde{U}\tilde{Y}}(\cdot|u_i,y_i)$, and define the set $C_{2,i} \triangleq \{\tilde{z}_{i,1},\tilde{z}_{i,2},\cdots,\tilde{z}_{i,M_2}\}$. We denote $\{C_{2,1},C_{2,2},\cdots,C_{2,M_1}\}$ as C_2. For a given set C_1, C_2 and a given symbol $x\in X$, we choose $i\in[1:M_1]$ and $j\in[1:M_2]$ such that

$$d_1(x,\tilde{y}_i)\leq D_1 \text{ and } d_2(x,\tilde{z}_{i,j})\leq D_2.$$

If there does not exist such pair, we set $(i,j)=(1,1)$. For

this pair, we define encoders f_1 and f_2 as

$$f_1(x)=i \text{ and } f_2(x)=j.$$

On the other hand, we define decoders φ_1 and φ_2 as

$$\varphi_1(i)=\tilde{y}_i \text{ and } \varphi_2(i,j)=\tilde{z}_{i,j}.$$

By taking the average over the random selection of C_1 and C_2, the average probability of $\Pr\{\mathcal{E}_1(D_1)\cup\mathcal{E}_2(D_2)\}$ is as follows:

$$\mathrm{E}[\Pr\{\mathcal{E}_1(D_1)\cup\mathcal{E}_2(D_2)\}]$$
$$= \Pr\left\{\bigcap_{i=1}^{M_1}\bigcap_{j=1}^{M_2}\{d_1(X,\tilde{Y}_i)>D_1 \text{ or } d_2(X,\tilde{Z}_{i,j})>D_2\}\right\}$$
$$= \Pr\left\{\bigcap_{i=1}^{M_1}\bigcap_{j=1}^{M_2}\{(X,\tilde{U}_i,\tilde{Y}_i,\tilde{Z}_{i,j})\notin\mathcal{D}\}\right\},$$

where $\{(\tilde{U}_i,\tilde{Y}_i,\tilde{Z}_{i,j})\}$ denote randomly selected sequences in C_1 and C_2. Now, by noting that $\tilde{Z}_{i,j}$ is generated for a given $(\tilde{U}_i,\tilde{Y}_i)$, the theorem follows from Lemma 1 by setting that $A=X$, $\tilde{B}=(\tilde{U},\tilde{Y})$, and $\tilde{C}=\tilde{Z}$. □

Remark 3. This proof is valid even without the RV $\tilde{U}$. This auxiliary RV is introduced merely for consistency with the outer bound. However, the following intuitive interpretation is possible: $\tilde{U}$ is partial information of $\tilde{Z}$ transmitted to two decoders. In an extreme case, if we set $\tilde{U}=\tilde{Z}$, M_2 has no effect on the bound of Theorem 1. Hence, we can make $M_2=1$. On the other hand, M_1 must be increased to satisfy a given probability level. If we set $\tilde{U}\neq\tilde{Z}$, M_1 may be decreased by increasing M_2. In other words, $\tilde{U}$ manages the balance of the numbers of codewords. This intuition may be useful to set the numbers of codewords in the actual code construction.

We use the next notation for the sake of simplicity.

Definition 6. For RVs (A,B,C), we define

$$I_\infty^\delta(A;B)\triangleq D_\infty^\delta(P_{AB}\|P_A\times P_B),$$
$$I_\infty^\delta(A;B|C)\triangleq D_\infty^\delta(P_{ABC}\|P_{AC}\times P_{B|C}).$$

We also define the following set of probability distributions for a given source X and constants D and ϵ.

$$\mathcal{P}(D,\epsilon|X)\triangleq \{P_{UYZ|X}\in\mathcal{P}_{\mathcal{U}\mathcal{Y}\mathcal{Z}|X} :$$
$$\Pr\{d_1(X,Y)>D_1\}\leq\epsilon_1,$$
$$\Pr\{d_2(X,Z)>D_2\}\leq\epsilon_2\}.$$

We note that $\mathcal{P}(D,\epsilon|X)$ depends on the set $\mathcal{U}$.

Now, by using the above theorem, we give an inner bound on $\mathcal{M}(D,\epsilon|X)$.

Theorem 2 (Inner bound). For a source X, real numbers $D_1, D_2\geq 0$ and $\epsilon_1,\epsilon_2>0$, and any set $\mathcal{U}$, we have

$$\mathcal{M}(D,\epsilon|X)\supseteq \bigcup_{(\delta,\beta,\gamma)\in\mathcal{S}(\epsilon)}\bigcup_{P_{UYZ|X}\in\mathcal{P}(D,\gamma|X)}\mathcal{M}_\mathrm{I}(\delta,\beta,P_{UYZ|X}),$$

where $\delta = (\delta_1, \delta_2)$, $\beta = (\beta_1, \beta_2)$, $\gamma = (\gamma_1, \gamma_2)$,

$$
\mathcal{S}(\epsilon) \triangleq \Big\{ (\delta, \beta, \gamma) \in (0, 1]^6 : \delta_1 + \delta_2 + \gamma_1 + \gamma_2
$$
$$
+ \beta_1 + \beta_2 \leq \min\{\epsilon_1, \epsilon_2\} \Big\}, \tag{9}
$$
$$
\mathcal{M}_{\mathrm{I}}(\delta, \beta, P_{UYZ|X}) \triangleq \Big\{ (M_1, M_2) \in \mathbb{N}^2 :
$$
$$
\log M_1 \geq I_\infty^{\delta_1}(X; U, Y) + \log \log \frac{1}{\beta_1},
$$
$$
\log M_2 \geq I_\infty^{\delta_2}(X; Z|U, Y)
$$
$$
+ \log \left(I_\infty^{\delta_1}(X; U, Y) + \log \frac{1}{\beta_2} \right) \Big\}, \tag{10}
$$

and (X, U, Y, Z) is a tuple of RVs with the probability distribution $P_X \times P_{UYZ|X}$.

Proof. We only have to show that (M, D) is ϵ-achievable for $(\delta, \beta, \gamma) \in \mathcal{S}(\epsilon)$, $P_{UYZ|X} \in \mathcal{P}(D, \gamma|X)$, and $M_1, M_2 \geq 1$ such that

$$
M_1 = \left\lceil \exp \left(I_\infty^{\delta_1}(X; U, Y) \right) \log \frac{1}{\beta_1} \right\rceil, \tag{11}
$$
$$
M_2 = \left\lceil \exp \left(I_\infty^{\delta_2}(X; Z|U, Y) \right) \left(I_\infty^{\delta_1}(X; U, Y) + \log \frac{1}{\beta_2} \right) \right\rceil. \tag{12}
$$

To this end, let $(\tilde{U}, \tilde{Y}, \tilde{Z}) \in \mathcal{U} \times \mathcal{Y} \times \mathcal{Z}$ be RVs that is independent of X and has the same marginal distribution as (U, Y, Z), i.e., $P_{\tilde{U}\tilde{Y}\tilde{Z}} = P_{UYZ}$. Then, according to Theorem 1, there exists a code $(f_1, f_2, \varphi_1, \varphi_2)$ such that numbers of codewords of encoder 1 and encoder 2 are M_1 and M_2, respectively, and

$$
\max\{\Pr\{\mathcal{E}_1(D_1)\}, \Pr\{\mathcal{E}_2(D_2)\}\}
$$
$$
\leq \mathrm{E} \left[\mathrm{E} \left[\mathrm{E} \left[\mathbf{1}\{(X, \tilde{U}, \tilde{Y}, \tilde{Z}) \notin \mathcal{D}\} \big| X, \tilde{U}, \tilde{Y} \right]^{M_2} \Big| X \right]^{M_1} \right]. \tag{13}
$$

On the other hand, according to Corollary 2, we have

$$
\mathrm{E} \left[\mathrm{E} \left[\mathrm{E} \left[\mathbf{1}\{(X, \tilde{U}, \tilde{Y}, \tilde{Z}) \notin \mathcal{D}\} | X, \tilde{U}, \tilde{Y} \right]^{M_2} \Big| X \right]^{M_1} \right]
$$
$$
\leq \delta_1 + \delta_2 + \Pr\{(X, U, Y, Z) \notin \mathcal{D}\}
$$
$$
+ e^{-\exp(-I_\infty^{\delta_2}(X; Z|U, Y)) M_2 + I_\infty^{\delta_1}(X; U, Y)}
$$
$$
+ e^{-\exp(-I_\infty^{\delta_1}(X; U, Y)) M_1}
$$
$$
\overset{(a)}{\leq} \delta_1 + \delta_2 + \gamma_1 + \gamma_2 + \beta_1 + \beta_2
$$
$$
\overset{(b)}{\leq} \min\{\epsilon_1, \epsilon_2\}, \tag{14}
$$

where (a) follows from (11), (12) and the fact that $P_{UYZ|X} \in \mathcal{P}(D, \gamma|X)$, and (b) comes from (9). This implies that (M, D) is ϵ-achievable. □

Remark 4. The proof is also valid if we do not restrict $(\tilde{U}, \tilde{Y}, \tilde{Z})$ to be the same distribution as (U, Y, Z). However, for the sake of simplicity, we consider the restricted case.

An outer bound on $\mathcal{M}(D, \epsilon|X)$ is given in the next theorem.

Theorem 3 (Outer bound)**.** For a source X, real numbers D_1, $D_2 \geq 0$ and $\epsilon_1, \epsilon_2 > 0$, and any set $\mathcal{U}$ such that $|\mathcal{U}| \geq |\mathcal{X}|$, we have

$$
\mathcal{M}(D, \epsilon|X) \subseteq \bigcup_{P_{UYZ|X} \in \mathcal{P}(D, \epsilon|X)} \bigcap_{\delta \in (0,1]^2} \mathcal{M}_{\mathrm{O}}(\delta, P_{UYZ|X}),
$$

where

$$
\mathcal{M}_{\mathrm{O}}(\delta, P_{UYZ|X}) \triangleq \Big\{ (M_1, M_2) \in \mathbb{N}^2 :
$$
$$
\log M_1 \geq I_\infty^{\delta_1}(X; U, Y) + \log \delta_1,
$$
$$
\log M_2 \geq I_\infty^{\delta_2}(X; Z|U, Y) + \log \delta_2 \Big\}.
$$

Remark 5. The outer bound can be further bounded as

$$
\bigcup_{P_{UYZ|X} \in \mathcal{P}(D, \epsilon|X)} \bigcap_{\delta \in (0,1]^2} \mathcal{M}_{\mathrm{O}}(\delta, P_{UYZ|X})
$$
$$
\subseteq \bigcup_{(\delta, \beta, \gamma) \in \mathcal{S}(\epsilon)} \bigcup_{P_{UYZ|X} \in \mathcal{P}(D, \epsilon|X)} \mathcal{M}_{\mathrm{O}}(\delta, P_{UYZ|X}),
$$

where we note that β and γ do not affect the bound. Although this bound is looser than that of Theorem 3, it may be easier to compare with the inner bound of Theorem 2. In fact, it can be immediately noticed that the difference between this loose bound and the inner bound comes from the difference between $\mathcal{P}(D, \gamma|X)$ and $\mathcal{P}(D, \epsilon|X)$, and the difference between $\mathcal{M}_{\mathrm{I}}(\delta, \beta, P_{UYZ|X})$ and $\mathcal{M}_{\mathrm{O}}(\delta, P_{UYZ|X})$.

Before proving the theorem, we show some necessary lemmas.

Lemma 3. Suppose that a pair of RVs (A, B) on $\mathcal{A} \times \mathcal{B}$ satisfies

$$
|\{a \in \mathcal{A} : P_{A|B}(a|b) > 0\}| \leq M,
$$
$$
\forall b \in \mathcal{B} \text{ s.t. } P_B(b) > 0, \tag{15}
$$

for some $M > 0$. Then, for any $\epsilon \in (0, 1]$, we have

$$
\Pr \left\{ P_{A|B}(A|B) \geq \frac{\epsilon}{M} \right\} > 1 - \epsilon.
$$

Proof. Since the lemma can be proved in a similar manner as [14, Lemma 2.6.2], we omit the proof. □

The next lemma is an extended version of [12, Lemma 4], which gives a bound on the size of the image of a function.

Lemma 4. For a function $g : \mathcal{A} \to \mathcal{B} \times C$ and $c \in C$, let $\|g\|_c$ denote the size of the image of g when one output is fixed to c, i.e.,

$$
\|g\|_c = |\{b \in \mathcal{B} : g(a) = (b, c), \exists a \in \mathcal{A}\}|.
$$

Then, for any $\delta \in (0, 1]$, any RV $A \in \mathcal{A}$, and $(B, C) = g(A)$, we have

$$\log \sup_{c \in C} \|g\|_c \geq I_\infty^\delta(A; B|C) + \log \delta.$$

Proof. Let $M = \sup_{c \in C} \|g\|_c$. Define a subset $\mathcal{D}_\delta \subseteq \mathcal{B} \times C$ and the function $\psi_o : \mathcal{A} \times \mathcal{B} \times C \to [0, 1]$ as

$$\mathcal{D}_\delta \triangleq \left\{ (b, c) \in \mathcal{B} \times C : P_{B|C}(b|c) \geq \frac{\delta}{M} \right\}, \quad (16)$$

$$\psi_o(a, b, c) \triangleq \mathbf{1}\{(a, b, c) \in \mathcal{A} \times \mathcal{D}_\delta\}. \quad (17)$$

Since $P_{BC}(b, c) = \sum_{a \in \mathcal{A}} P_A(a) \mathbf{1}\{(b, c) = g(a)\}$, $P_{B|C}(b|c) > 0$ for $c \in C$ such that $P_C(c) > 0$ if and only if there exists $a \in \mathcal{A}$ such that $(b, c) = g(a)$ and $P_A(a) > 0$. Thus, for $c \in C$ such that $P_C(c) > 0$, we have

$$
\begin{aligned}
&|\{b \in \mathcal{B} : P_{B|C}(b|c) > 0\} \\
&= |\{b \in \mathcal{B} : (b, c) = g(a), \exists a \in \mathcal{A} \text{ s.t. } P_A(a) > 0\} \\
&\leq \|g\|_c \\
&\leq M.
\end{aligned}
$$

Then, by using Lemma 3, it is easy to see that

$$
\begin{aligned}
&\sum_{(a,b,c) \in \mathcal{A} \times \mathcal{B} \times C} \psi_o(a, b, c) P_{ABC}(a, b, c) \\
&= \Pr\left\{ P_{B|C}(B|C) \geq \frac{\delta}{M} \right\} \\
&> 1 - \delta.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
&I_\infty^\delta(A; B|C) \\
&\leq \left| \log \sup_{(a,b,c) \in \mathcal{A} \times \mathcal{B} \times C} \frac{\psi_o(a, b, c) P_{ABC}(a, b, c)}{P_{AC}(a, c) P_{B|C}(b|c)} \right|^+ \\
&= \left| \log \sup_{(a,b,c) \in \mathcal{A} \times \mathcal{D}_\delta} \frac{P_{B|AC}(b|a, c)}{P_{B|C}(b|c)} \right|^+ \\
&\overset{(a)}{\leq} \left| \log \sup_{(a,b,c) \in \mathcal{A} \times \mathcal{D}_\delta} \frac{P_{B|AC}(b|a, c)}{\delta/M} \right|^+ \\
&\leq \log M - \log \delta,
\end{aligned}
$$

where (a) comes from the definition (16). This completes the proof. $\square$

Remark 6. For a triple of RVs (A, B, C) on $\mathcal{A} \times \mathcal{B} \times C$, let $M = |\mathcal{B}|$. Then, in the same way as the above proof, we can easily show that

$$I_\infty^\delta(A; B|C) \leq \log|\mathcal{B}| - \log \delta.$$

Now, we give the proof of Theorem 3.

Proof of Theorem 3. Let $\|f_1\|$ be the size of the image of an encoder f_1. Since $\|f_1\| \leq |\mathcal{X}|$ and $|\mathcal{X}| \leq |\mathcal{U}|$ by the assumption, there exists an injective function $\text{id} : [1 : \|f_1\|] \to \mathcal{U}$. For this function, let $\mathcal{U}_{\text{id}} \subseteq \mathcal{U}$ be the image of id and $\text{id}^{-1} : \mathcal{U}_{\text{id}} \to [1 : \|f_1\|]$ be the inverse function of id on $\mathcal{U}_{\text{id}}$.

Suppose that (M, D) is ϵ-achievable. Then, there exists

a code $(f_1, f_2, \varphi_1, \varphi_2)$ such that

$$
\begin{aligned}
\Pr\{d_1(X, \varphi_1(f_1(X))) > D_1\} &\leq \epsilon_1, \\
\Pr\{d_2(X, \varphi_2(f_1(X), f_2(X))) > D_2\} &\leq \epsilon_2.
\end{aligned}
$$

Thus, by setting $U = \text{id}(f_1(X))$, $Y = \varphi_1(f_1(X))$, and $Z = \varphi_2(f_1(X), f_2(X))$, we have

$$\Pr\{d_1(X, Y) > D_1\} \leq \epsilon_1, \quad (18)$$

$$\Pr\{d_2(X, Z) > D_2\} \leq \epsilon_2. \quad (19)$$

For a constant value c, let $g_1(x) = (\text{id}(f_1(x)), \varphi_1(f_1(x)), c)$, $A = X$, $B = (U, Y)$, and $C = c$. According to Lemma 4, for any $\delta_1 \in (0, 1]$, we have

$$
\begin{aligned}
\log \|g_1\|_c &\geq I_\infty^{\delta_1}(X; U, Y|C) + \log \delta_1 \\
&= I_\infty^{\delta_1}(X; U, Y) + \log \delta_1. \quad (20)
\end{aligned}
$$

On the other hand, we have

$$
\begin{aligned}
\|g_1\|_c &= |\{(u, y) \in \mathcal{U}_{\text{id}} \times \mathcal{Y} : g_1(x) = (u, y, c), \exists x \in \mathcal{X}\}| \\
&= \sum_{u \in \mathcal{U}_{\text{id}}} \sum_{y \in \mathcal{Y}} \mathbf{1}\{g_1(x) = (u, y, c), \exists x \in \mathcal{X}\} \\
&= \sum_{u \in \mathcal{U}_{\text{id}}} \mathbf{1}\{g_1(x) = (u, \varphi_1(\text{id}^{-1}(u)), c), \exists x \in \mathcal{X}\} \\
&\leq M_1, \quad (21)
\end{aligned}
$$

where the last inequality follows since the size of $\mathcal{U}_{\text{id}}$ is at most $\|f_1\|$ and $\|f_1\| \leq M_1$. Combining (20) and (21), we have

$$\log M_1 \geq I_\infty^{\delta_1}(X; U, Y) + \log \delta_1. \quad (22)$$

Let $g_2(x) = (\varphi_2(f_1(x), f_2(x)), \text{id}(f_1(x)), \varphi_1(f_1(x)))$, $A = X$, $B = Z$, and $C = (U, Y)$. Then, according to Lemma 4, for any $\delta_2 \in (0, 1]$, we have

$$\log \sup_{(u, y) \in \mathcal{U} \times \mathcal{Y}} \|g_2\|_{(u, y)} \geq I_\infty^{\delta_2}(X; Z|U, Y) + \log \delta_2. \quad (23)$$

On the other hand, for any $(u, y) \in \mathcal{U}_{\text{id}} \times \mathcal{Y}$, we have

$$
\begin{aligned}
\|g_2\|_{(u, y)} &= \sum_{z \in \mathcal{Z}} \mathbf{1}\{z = \varphi_2(\text{id}^{-1}(u), f_2(x)), \\
&\quad \text{id}^{-1}(u) = f_1(x), y = \varphi_1(\text{id}^{-1}(u)), \exists x \in \mathcal{X}\} \\
&\leq \sum_{z \in \mathcal{Z}} \mathbf{1}\{\exists x \in \mathcal{X}, z = \varphi_2(\text{id}^{-1}(u), f_2(x))\} \\
&\leq \sum_{z \in \mathcal{Z}} \mathbf{1}\{\exists j \in [1 : M_2], z = \varphi_2(\text{id}^{-1}(u), j)\} \\
&\leq \sum_{j \in [1:M_2]} \sum_{z \in \mathcal{Z}} \mathbf{1}\{z = \varphi_2(\text{id}^{-1}(u), j)\} \\
&= M_2. \quad (24)
\end{aligned}
$$

We note that for any $(u, y) \in \{\mathcal{U} \setminus \mathcal{U}_{\text{id}}\} \times \mathcal{Y}$, it holds that $\|g_2\|_{(u, y)} = 0$. Combining (23) and (24), we have

$$\log M_2 \geq I_\infty^{\delta_2}(X; Z|U, Y) + \log \delta_2. \qquad (25)$$

Since $\delta_1 \in (0, 1]$ and $\delta_2 \in (0, 1]$ are arbitrary, (22) and (25) imply that

$$(M_1, M_2) \in \bigcap_{\delta \in (0,1]^2} \mathcal{M}_O(\delta, P_{UYZ|X}).$$

Now, by recalling that (X, U, Y, Z) satisfy (18) and (19), for any ϵ-achievable pair (M, D), we have

$$(M_1, M_2) \in \bigcup_{P_{UYZ|X} \in \mathcal{P}(D, \epsilon|X)} \bigcap_{\delta \in (0,1]^2} \mathcal{M}_O(\delta, P_{UYZ|X}).$$

This completes the proof. $\qquad\square$

Remark 7. If we do not employ the RV U which has a role in fixing the RV $f_1(X)$ to a certain codeword, we cannot bound $\|g_2\|_{(u,y)}$ by M_2 in (24). Thus in this proof, introducing U is quite important.

Remark 8. In [1], we gave inner and outer bounds on $\mathcal{M}(D, \epsilon|X)$ by using the α-mutual information of order infinity [18], where the α-mutual information is a generalized version of the mutual information. In this paper, however, we use the smooth max Rényi divergence. This is because it is compatible with the information spectrum quantity which is well studied and useful to analyze rates of a code.

Finally, we discuss the difference between our inner and outer bounds of Theorems 2 and 3.

If cardinalities of sets $\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{U}$ are small, the outer bound and the inner bound may be given by computing their boundaries. Thus, the difference between these two bounds is actually evaluated. On the other hand, if cardinalities of the sets are large, it is difficult to compute their boundaries. However, we can evaluate the difference roughly.

Let $|\mathcal{U}| = |\mathcal{X}|$ for the sake of simplicity. Since the main interest is in the case where $\epsilon_1, \epsilon_2 > 0$ are sufficiently small, we assume that for a small real number $\rho \in (0, 1/2]$, $\epsilon_1 = \epsilon_2 = \rho$. Then, we can set that $\delta_i = \beta_i = \gamma_i \approx \epsilon_i$ $(i = 1, 2)$ in the inner bound. Thus, the inner bound can be approximated as

$$\bigcup_{(\delta, \beta, \gamma) \in \mathcal{S}(\epsilon)} \bigcup_{P_{UYZ|X} \in \mathcal{P}(D, \gamma|X)} \mathcal{M}_I(\delta, \beta, P_{UYZ|X})$$

$$\approx \bigcup_{P_{UYZ|X} \in \mathcal{P}(D, \epsilon|X)} \mathcal{M}_I(\epsilon, \epsilon, P_{UYZ|X}).$$

On the other hand, the outer bound can be bounded as

$$\bigcup_{P_{UYZ|X} \in \mathcal{P}(D, \epsilon|X)} \bigcap_{\delta \in (0,1]^2} \mathcal{M}_O(\delta, P_{UYZ|X})$$

$$\subseteq \bigcup_{P_{UYZ|X} \in \mathcal{P}(D, \epsilon|X)} \mathcal{M}_O(\epsilon, P_{UYZ|X}).$$

Thus, the difference between the outer bound and the inner bound can be evaluated by the difference between $\mathcal{M}_I(\epsilon, \epsilon, P_{UYZ|X})$ and $\mathcal{M}_O(\epsilon, P_{UYZ|X})$ for each fixed

$P_{UYZ|X} \in \mathcal{P}(D, \epsilon|X)$.

According to the definitions of $\mathcal{M}_I(\epsilon, \epsilon, P_{UYZ|X})$ and $\mathcal{M}_O(\epsilon, P_{UYZ|X})$, the difference comes from boundary pairs $(M_{I,1}, M_{I,2})$ of $\mathcal{M}_I(\epsilon, \epsilon, P_{UYZ|X})$ and $(M_{O,1}, M_{O,2})$ of $\mathcal{M}_O(\epsilon, P_{UYZ|X})$, where

$$M_{I,1} = \left\lceil \exp\left(I_\infty^\rho(X; U, Y) + \log\log\frac{1}{\rho}\right)\right\rceil,$$

$$M_{I,2} = \left\lceil \exp\left(I_\infty^\rho(X; Z|U, Y) + \log\left(I_\infty^\rho(X; U, Y) + \log\frac{1}{\rho}\right)\right)\right\rceil,$$

$$M_{O,1} = \left\lceil \exp\left(I_\infty^\rho(X; U, Y) + \log\rho\right)\right\rceil,$$

$$M_{O,2} = \left\lceil \exp\left(I_\infty^\rho(X; Z|U, Y) + \log\rho\right)\right\rceil.$$

Clearly, the difference between $M_{I,1}$ and $M_{O,1}$ can be evaluated by second terms of exponents, i.e, $\log\log\frac{1}{\rho} - \log\rho$. Similarly, the difference between $M_{I,2}$ and $M_{O,2}$ can be evaluated by second terms of exponents, i.e, $\log\left(I_\infty^\rho(X; U, Y) + \log\frac{1}{\rho}\right) - \log\rho$. Since $I_\infty^\rho(X; U, Y) \geq 0$, the differences are at most $\log\left(I_\infty^\rho(X; U, Y) + \log\frac{1}{\rho}\right) - \log\rho$.

Especially, for a finite set $\mathcal{X}$, since $I_\infty^\rho(X; U, Y) \leq \log|\mathcal{X}| - \log\rho$ (see Remark 6), the differences are at most $\log\left(\log|\mathcal{X}| + 2\log\frac{1}{\rho}\right) - \log\rho$. Furthermore, when we regard $\mathcal{X}$ as an n-fold Cartesian product $\mathcal{X}^n$, the differences of exponents are at most $\log\left(n\log|\mathcal{X}| + 2\log\frac{1}{\rho}\right) - \log\rho$. Dividing it by n, it obviously vanishes as n tends to infinity. This implies that the inner bound and the outer bound asymptotically coincide with each other in terms of *rate* (which is the exponent of the number of codewords divided by n) as n tends to infinity. In fact, in the next section, the rate-distortion region can be given by using our inner and outer bounds.

5. General Formula for the Rate-Distortion Region

In this section, we deal with the coding for an n-length source sequence and give a general formula for the rate-distortion region.

First of all, we introduce the spectral (conditional) sup-mutual information rate [14].

Definition 7. For a sequence $(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) = \{(X^n, Y^n, Z^n)\}_{n=1}^\infty$ of RVs, we define

$$\bar{I}(\mathbf{X}; \mathbf{Y}) \triangleq \text{p-}\limsup_{n\to\infty} \frac{1}{n}\log\frac{P_{Y^n|X^n}(Y^n|X^n)}{P_{Y^n}(Y^n)},$$

$$\bar{I}(\mathbf{X}; \mathbf{Y}|\mathbf{Z}) \triangleq \text{p-}\limsup_{n\to\infty} \frac{1}{n}\log\frac{P_{Y^n|X^nZ^n}(Y^n|X^n, Z^n)}{P_{Y^n|Z^n}(Y^n|Z^n)}.$$

The smooth max Rényi divergence is related to the spectral sup-mutual information rate as shown in the corollary of the next lemma.

Lemma 5. Consider two sequences $\mathbf{X}$ and $\mathbf{Y}$ of RVs over the same set. Then, we have

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\frac{1}{n}D_\infty^\delta(P_{X^n}\|P_{Y^n}) = \text{p-}\limsup_{n\to\infty}\frac{1}{n}\log\frac{P_{X^n}(X^n)}{P_{Y^n}(X^n)}.$$

Proof. For a sequence $\{a_n\}_{n=1}^{\infty}$ of real numbers, it holds that $\limsup_{n\to\infty} |a_n|^+ = |\limsup_{n\to\infty} a_n|^+$. Thus, according to (8), we have

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} D_\infty^\delta (P_{X^n} \| P_{Y^n})$$
$$= \left| \lim_{\delta\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} D_\infty^\delta (P_{X^n} \| P_{Y^n})^- \right|^+. \qquad (26)$$

According to [11, Lemma 3], it holds that

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} D_\infty^\delta (P_{X^n} \| P_{Y^n})^-$$
$$= \text{p-}\limsup_{n\to\infty} \frac{1}{n} \log \frac{P_{X^n}(X^n)}{P_{Y^n}(X^n)}. \qquad (27)$$

Furthermore, according to [14, Lemma 3.2.1], the RHS of (27) is non-negative. Thus, by combining (26) and (27), we have the lemma. □

Corollary 3. For a sequence $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ of RVs, we have

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} I_\infty^\delta (X^n; Y^n) = \overline{I}(\mathbf{X}; \mathbf{Y}),$$

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} I_\infty^\delta (X^n; Y^n | Z^n) = \overline{I}(\mathbf{X}; \mathbf{Y} | \mathbf{Z}).$$

Proof. Since this corollary immediately follows from Lemma 5 and Definition 7, we omit the proof. □

Let $\mathbf{P_{UYZ|X}}$ be a sequence of conditional probability distributions $P_{U^n Y^n Z^n | X^n} \in \mathcal{P}_{\mathcal{U}^n \mathcal{Y}^n \mathcal{Z}^n | \mathcal{X}^n}$. We define

$$\mathcal{P}_G(D|\mathbf{X}) \triangleq \{\mathbf{P_{UYZ|X}} : \overline{D}_1(\mathbf{X}, \mathbf{Y}) \le D_1,$$
$$\overline{D}_2(\mathbf{X}, \mathbf{Z}) \le D_2\},$$
$$\overline{D}_1(\mathbf{X}, \mathbf{Y}) \triangleq \text{p-}\limsup_{n\to\infty} d_1^{(n)}(X^n, Y^n),$$
$$\overline{D}_2(\mathbf{X}, \mathbf{Z}) \triangleq \text{p-}\limsup_{n\to\infty} d_2^{(n)}(X^n, Z^n),$$
$$\mathcal{R}_G(\mathbf{P_{UYZ|X}}|\mathbf{X}) \triangleq \{(R_1, R_2) \in \mathbb{R}^2 : R_1 \ge \overline{I}(\mathbf{X}; \mathbf{U}, \mathbf{Y}),$$
$$R_2 \ge \overline{I}(\mathbf{X}; \mathbf{Z} | \mathbf{U}, \mathbf{Y})\},$$

where $(\mathbf{X}, \mathbf{U}, \mathbf{Y}, \mathbf{Z})$ is a sequence of RVs (X^n, U^n, Y^n, Z^n) induced by $\mathbf{P_{UYZ|X}}$ and a general source $\mathbf{X}$. We note that $\mathcal{P}_G(D|\mathbf{X})$ depends on $\mathcal{U}$.

The main result of this section is the next theorem which gives a general formula for the rate-distortion region.

Theorem 4. For a general source $\mathbf{X}$, real numbers $D_1, D_2 \ge 0$, and any set $\mathcal{U}$ such that $|\mathcal{U}| \ge |\mathcal{X}|$, we have

$$\mathcal{R}(D|\mathbf{X}) = \bigcup_{\mathbf{P_{UYZ|X}} \in \mathcal{P}_G(D|\mathbf{X})} \mathcal{R}_G(\mathbf{P_{UYZ|X}}|\mathbf{X}). \qquad (28)$$

Remark 9. We can show that the RHS of (28) is a closed set by using the diagonal line argument (cf. [14, Remark 5.7.5]).

Remark 10. We are not sure whether a sequence $\mathbf{U}$ of auxiliary RVs is really necessary or not. It may be possible to

characterize the region without it.

The proof of this theorem is presented in the subsequent two sections. In these sections, for a code, we denote

$$\hat{Y}^n = \varphi_1^{(n)}(f_1^{(n)}(X^n)),$$
$$\hat{Z}^n = \varphi_2^{(n)}(f_1^{(n)}(X^n), f_2^{(n)}(X^n)).$$

5.1 Direct Part

In this section, we show

$$\mathcal{R}(D|\mathbf{X}) \supseteq \bigcup_{\mathbf{P_{UYZ|X}} \in \mathcal{P}_G(D|\mathbf{X})} \mathcal{R}_G(\mathbf{P_{UYZ|X}}|\mathbf{X}). \qquad (29)$$

Let $\mathbf{P_{UYZ|X}} \in \mathcal{P}_G(D|\mathbf{X})$ and suppose that $\overline{I}(\mathbf{X}; \mathbf{U}, \mathbf{Y}) < \infty$ and $\overline{I}(\mathbf{X}; \mathbf{Z} | \mathbf{U}, \mathbf{Y}) < \infty$. Then, for any $\epsilon, \epsilon_1, \epsilon_2 > 0$ such that $\epsilon_1 = \epsilon_2 = \epsilon$, any $(\delta, \beta, \gamma) \in \mathcal{S}(\epsilon)$ in (9), and every sufficiently large n, we have

$$\Pr\{d_1^{(n)}(X^n, Y^n) > D_1 + \epsilon\} \le \gamma_1,$$
$$\Pr\{d_2^{(n)}(X^n, Z^n) > D_2 + \epsilon\} \le \gamma_2.$$

Hence, according to Theorem 2, there exists a sequence of codes $\{(f_1^{(n)}, f_2^{(n)}, \varphi_1^{(n)}, \varphi_2^{(n)})\}$ such that for sufficiently large n,

$$\Pr\{d_1^{(n)}(X^n, \hat{Y}^n) > D_1 + \epsilon\} \le \epsilon,$$
$$\Pr\{d_2^{(n)}(X^n, \hat{Z}^n) > D_2 + \epsilon\} \le \epsilon,$$

and

$$\frac{1}{n} \log M_1^{(n)} = \frac{1}{n} \log \left[\exp\left(I_\infty^{\delta_1}(X^n; U^n, Y^n) \right) \log \frac{1}{\beta_1} \right],$$
$$\frac{1}{n} \log M_2^{(n)} = \frac{1}{n} \log \left[\exp\left(I_\infty^{\delta_2}(X^n; Z^n | U^n, Y^n) \right) \right.$$
$$\left. \times \left(I_\infty^{\delta_1}(X^n; U^n, Y^n) + \log \frac{1}{\beta_2} \right) \right].$$

Thus, we have

$$\limsup_{n\to\infty} R_1^{(n)} \le \limsup_{n\to\infty} \frac{1}{n} I_\infty^{\delta_1}(X^n; U^n, Y^n)$$
$$\overset{(a)}{\le} \overline{I}(\mathbf{X}; \mathbf{U}, \mathbf{Y}),$$

and

$$\limsup_{n\to\infty} R_2^{(n)}$$
$$\le \limsup_{n\to\infty} \frac{1}{n} \log \left(\exp\left(I_\infty^{\delta_2}(X^n; Z^n | U^n, Y^n) \right) \right.$$
$$\left. \times \left(I_\infty^{\delta_1}(X^n; U^n, Y^n) + \log \frac{1}{\beta_2} + 1 \right) \right)$$
$$\le \limsup_{n\to\infty} \frac{1}{n} \log \left(\exp\left(I_\infty^{\delta_2}(X^n; Z^n | U^n, Y^n) \right) \right.$$
$$\left. \times n \left(\limsup_{n\to\infty} \frac{1}{n} I_\infty^{\delta_1}(X^n; U^n, Y^n) + \delta_2 \right) \right)$$

$$\overset{(a)}{\leq} \limsup_{n\to\infty} \frac{1}{n} \log \left(\exp \left(I_\infty^{\delta_2}(X^n; Z^n | U^n, Y^n) \right) \right.$$
$$\left. \times n \left(\overline{I}(\mathbf{X}; \mathbf{U}, \mathbf{Y}) + \delta_2 \right) \right)$$
$$\overset{(b)}{=} \limsup_{n\to\infty} \frac{1}{n} I_\infty^{\delta_2}(X^n; Z^n | U^n, Y^n)$$
$$\overset{(a)}{\leq} \overline{I}(\mathbf{X}; \mathbf{Z} | \mathbf{U}, \mathbf{Y}),$$

where (a) comes from Corollary 3 and the fact that I_∞^δ is a non-increasing function of δ, and (b) follows since $\overline{I}(\mathbf{X}; \mathbf{U}, \mathbf{Y}) < \infty$.

Now, by using the usual diagonal line argument, we can construct a sequence of codes $\{(f_1^{(n)}, f_2^{(n)}, \varphi_1^{(n)}, \varphi_2^{(n)})\}$ such that

$$\limsup_{n\to\infty} R_1^{(n)} \leq \overline{I}(\mathbf{X}; \mathbf{U}, \mathbf{Y}),$$
$$\limsup_{n\to\infty} R_2^{(n)} \leq \overline{I}(\mathbf{X}; \mathbf{Z} | \mathbf{U}, \mathbf{Y}).$$

and for any $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr\{d_1^{(n)}(X^n, \hat{Y}^n) > D_1 + \epsilon\} = 0,$$
$$\lim_{n\to\infty} \Pr\{d_2^{(n)}(X^n, \hat{Z}^n) > D_2 + \epsilon\} = 0.$$

This implies that

$$\underset{n\to\infty}{\text{p-}\limsup} \, d_1^{(n)}(X^n, \hat{Y}^n) \leq D_1,$$
$$\underset{n\to\infty}{\text{p-}\limsup} \, d_2^{(n)}(X^n, \hat{Z}^n) \leq D_2.$$

Thus, for any $\mathbf{P_{UYZ|X}} \in \mathcal{P}_G(D|\mathbf{X})$ such that $\overline{I}(\mathbf{X}; \mathbf{U}, \mathbf{Y}) \in \mathbb{R}$ and $\overline{I}(\mathbf{X}; \mathbf{Z}|\mathbf{U}, \mathbf{Y}) \in \mathbb{R}$, we have

$$(\overline{I}(\mathbf{X}; \mathbf{U}, \mathbf{Y}), \overline{I}(\mathbf{X}; \mathbf{Z}|\mathbf{U}, \mathbf{Y})) \in \mathcal{R}(D|\mathbf{X}).$$

This implies (29).

5.2 Converse Part

In this section, we show that

$$\mathcal{R}(D|\mathbf{X}) \subseteq \bigcup_{\mathbf{P_{UYZ|X}} \in \mathcal{P}_G(D|\mathbf{X})} \mathcal{R}_G(\mathbf{P_{UYZ|X}}|\mathbf{X}). \quad (30)$$

Suppose that (R, D) is fm-achievable. Then, there exists a sequence of codes satisfying

$$\underset{n\to\infty}{\text{p-}\limsup} \, d_1^{(n)}(X^n, \hat{Y}^n) \leq D_1,$$
$$\underset{n\to\infty}{\text{p-}\limsup} \, d_2^{(n)}(X^n, \hat{Z}^n) \leq D_2,$$

and

$$\limsup_{n\to\infty} \frac{1}{n} \log M_i^{(n)} \leq R_i, \quad \forall i \in \{1, 2\}. \quad (31)$$

Thus, we have for any $\gamma > 0$

$$\lim_{n\to\infty} \Pr\{d_1^{(n)}(X^n, \hat{Y}^n) > D_1 + \gamma\} = 0,$$
$$\lim_{n\to\infty} \Pr\{d_2^{(n)}(X^n, \hat{Z}^n) > D_2 + \gamma\} = 0.$$

This implies that there exists a sequence $\{\gamma_n\}_{n=1}^\infty$ such that $\lim_{n\to\infty} \gamma_n = 0$ and

$$\Pr\{d_1^{(n)}(X^n, \hat{Y}^n) > D_1 + \gamma_n\} \leq \gamma_n,$$
$$\Pr\{d_2^{(n)}(X^n, \hat{Z}^n) > D_2 + \gamma_n\} \leq \gamma_n.$$

This means that $(M^{(n)}, D + \gamma_n)$ is γ_n-achievable. According to Theorem 3, there exists a sequence $\mathbf{P_{UYZ|X}} = \{P_{U^n Y^n Z^n | X^n}\}_{n=1}^\infty$ of conditional probability distributions such that $P_{U^n Y^n Z^n | X^n} \in \mathcal{P}(D + \gamma_n, \gamma_n | X^n)$ and for any $\delta \in (0, 1]^2$,

$$\limsup_{n\to\infty} \frac{1}{n} \log M_1^{(n)} \geq \limsup_{n\to\infty} \frac{1}{n} I_\infty^\delta(X^n; U^n, Y^n),$$
$$\limsup_{n\to\infty} \frac{1}{n} \log M_2^{(n)} \geq \limsup_{n\to\infty} \frac{1}{n} I_\infty^\delta(X^n; Z^n | U^n, Y^n).$$

Since this holds for any $\delta \in (0, 1]^2$, we have

$$\limsup_{n\to\infty} \frac{1}{n} \log M_1^{(n)} \geq \lim_{\delta\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} I_\infty^\delta(X^n; U^n, Y^n)$$
$$\overset{(a)}{=} \overline{I}(\mathbf{X}; \mathbf{U}, \mathbf{Y}), \quad (32)$$
$$\limsup_{n\to\infty} \frac{1}{n} \log M_2^{(n)} \geq \lim_{\delta\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} I_\infty^\delta(X^n; Z^n | U^n, Y^n)$$
$$\overset{(a)}{=} \overline{I}(\mathbf{X}; \mathbf{Z} | \mathbf{U}, \mathbf{Y}). \quad (33)$$

where (a) are comes from Corollary 3. On the other hand, since $P_{U^n Y^n Z^n | X^n} \in \mathcal{P}(D + \gamma_n, \gamma_n | X^n)$, $\mathbf{P_{UYZ|X}}$ must satisfy $\overline{D}_1(\mathbf{X}; \mathbf{Y}) \leq D_1$ and $\overline{D}_2(\mathbf{X}; \mathbf{Z}) \leq D_2$, i.e., $\mathbf{P_{UYZ|X}} \in \mathcal{P}_G(D|\mathbf{X})$.

By combining (31), (32), and (33), we can conclude that for any fm-achievable pair (R, D),

$$(R_1, R_2) \in \bigcup_{\mathbf{P_{UYZ|X}} \in \mathcal{P}_G(D|\mathbf{X})} \mathcal{R}_G(\mathbf{P_{UYZ|X}}|\mathbf{X}).$$

This implies (30).

5.3 Discrete Stationary Memoryless Sources

In this section, we show that the rate-distortion region given in Theorem 4 coincides with the region by Rimoldi [6] when a source $\mathbf{X}$ is a discrete stationary memoryless source.

Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be finite sets. Since $\mathbf{X} = \{X^n\}_{n=1}^\infty$ is a discrete stationary memoryless source, we assume that $X^n = (X_1, X_2, \cdots, X_n)$ is a sequence of independent copies of an RV X on $\mathcal{X}$. We also assume that distortion measures $d_1^{(n)}$ and $d_2^{(n)}$ are additive, i.e., for two functions $d_1 : \mathcal{X} \times \mathcal{Y} \to [0, +\infty)$ and $d_2 : \mathcal{X} \times \mathcal{Z} \to [0, +\infty)$, distortion measures are represented as

$$d_1^{(n)}(x^n, y^n) = \frac{1}{n} \sum_{i=1}^n d_1(x_i, y_i),$$

$$d_2^{(n)}(x^n, z^n) = \frac{1}{n} \sum_{i=1}^{n} d_2(x_i, z_i).$$

We define

$$\mathcal{P}_{\mathrm{M}}(D|X) \triangleq \{P_{YZ|X} \in \mathcal{P}_{\mathcal{Y}\mathcal{Z}|\mathcal{X}} : \mathrm{E}[d_1(X,Y)] \leq D_1,$$
$$\mathrm{E}[d_2(X,Z)] \leq D_2\},$$

and for $P_{YZ|X} \in \mathcal{P}_{\mathcal{Y}\mathcal{Z}|\mathcal{X}}$,

$$\mathcal{R}_{\mathrm{M}}(P_{YZ|X}|X) \triangleq \{(R_1, R_2) \in \mathbb{R}_{\geq 0}^2 : R_1 \geq I(X;Y),$$
$$R_1 + R_2 \geq I(X;Y,Z)\},$$

where (X, Y, Z) is the tuple of RVs induced by a conditional probability distribution $P_{YZ|X} \in \mathcal{P}_{\mathcal{Y}\mathcal{Z}|\mathcal{X}}$ and a given RV X. Then, we have the next theorem.

Theorem 5. For a discrete stationary memoryless source $\mathbf{X}$, additive distortion measures, and any set $\mathcal{U}$ such that $|\mathcal{U}| \geq |\mathcal{Z}| + 1$, we have

$$\bigcup_{\mathbf{P}_{\mathbf{UYZ|X}} \in \mathcal{P}_{\mathrm{G}}(D|\mathbf{X})} \mathcal{R}_{\mathrm{G}}(\mathbf{P}_{\mathbf{UYZ|X}}|\mathbf{X})$$
$$= \bigcup_{P_{YZ|X} \in \mathcal{P}_{\mathrm{M}}(D|X)} \mathcal{R}_{\mathrm{M}}(P_{YZ|X}|X). \qquad (34)$$

Remark 11. According to Theorem 4, the above theorem holds even if $|\mathcal{U}| \geq |\mathcal{Z}|$. However, we assume that $|\mathcal{U}| \geq |\mathcal{Z}| + 1$ in order to prove the above theorem without the help of Theorem 4.

Remark 12. The RHS of (34) can be written as

$$\{(R_1, R_2) \in \mathbb{R}_{\geq 0}^2 : R_1 \geq R_1(D_1), R_1 + R_2 \geq R_b(R_1, D_1, D_2)\}, \qquad (35)$$

where $R_1(D_1)$ is the rate-distortion function and $R_b(R_1, D_1, D_2)$ gives the boundary for a given R_1, which are defined as (see, e.g., [19, Corollary 1], [10, (22)])

$$R_1(D_1) \triangleq \min_{P_{Y|X} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}: \mathrm{E}[d_1(X,Y)] \leq D_1} I(X;Y),$$

$$R_b(R_1, D_1, D_2) \triangleq \min_{\substack{P_{YZ|X} \in \mathcal{P}_{\mathcal{Y}\mathcal{Z}|\mathcal{X}}: \\ \mathrm{E}[d_1(X,Y)] \leq D_1, \mathrm{E}[d_2(X,Z)] \leq D_2, \\ I(X;Y) \leq R_1}} I(X;Y,Z).$$

We note that $R_1(D_1)$ and $R_b(R_1, D_1, D_2)$ are convex and continuous functions of a triple (R_1, D_1, D_2) (see, e.g., [14, Remark 5.2.1] and [20, Lemma 4]).

We will prove the theorem by two parts separately.

Proof: The left-hand side (LHS) of (34) ⊆ The RHS of (34). We have

$$\bigcup_{\mathbf{P}_{\mathbf{UYZ|X}} \in \mathcal{P}_{\mathrm{G}}(D|\mathbf{X})} \mathcal{R}_{\mathrm{G}}(\mathbf{P}_{\mathbf{UYZ|X}}|\mathbf{X})$$
$$\subseteq \bigcup_{\mathbf{P}_{\mathbf{UYZ|X}} \in \mathcal{P}_{\mathrm{G}}(D|\mathbf{X})} \{(R_1, R_2) : R_1 \geq \overline{I}(\mathbf{X};\mathbf{U},\mathbf{Y}),$$
$$R_1 + R_2 \geq \overline{I}(\mathbf{X};\mathbf{U},\mathbf{Y}) + \overline{I}(\mathbf{X};\mathbf{Z}|\mathbf{U},\mathbf{Y})\}$$

$$\overset{(a)}{\subseteq} \bigcup_{\mathbf{P}_{\mathbf{UYZ|X}} \in \mathcal{P}_{\mathrm{G}}(D|\mathbf{X})} \{(R_1, R_2) : R_1 \geq \overline{I}(\mathbf{X};\mathbf{U},\mathbf{Y}),$$
$$R_1 + R_2 \geq \overline{I}(\mathbf{X};\mathbf{U},\mathbf{Y},\mathbf{Z})\}$$

$$\overset{(b)}{\subseteq} \bigcup_{\substack{\mathbf{P}_{\mathbf{YZ|X}}: \\ \overline{D}_1(\mathbf{X},\mathbf{Y}) \leq D_1, \overline{D}_2(\mathbf{X},\mathbf{Z}) \leq D_2}} \{(R_1, R_2) : R_1 \geq \overline{I}(\mathbf{X};\mathbf{Y}),$$
$$R_1 + R_2 \geq \overline{I}(\mathbf{X};\mathbf{Y},\mathbf{Z})\}, \qquad (36)$$

where (a) and (b) respectively come from the fact that

$$\overline{I}(\mathbf{X};\mathbf{U},\mathbf{Y}) + \overline{I}(\mathbf{X};\mathbf{Z}|\mathbf{U},\mathbf{Y}) \geq \overline{I}(\mathbf{X};\mathbf{U},\mathbf{Y},\mathbf{Z}),$$
$$\overline{I}(\mathbf{X};\mathbf{U},\mathbf{Y}) \geq \overline{I}(\mathbf{X};\mathbf{Y}).$$

For a sequence $(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) = \{(X^n, Y^n, Z^n)\}$ of RVs induced by $\mathbf{P}_{\mathbf{YZ|X}}$ and a given source $\mathbf{X}$, let $(\mathbf{X}, \bar{\mathbf{Y}}, \bar{\mathbf{Z}}) = \{(X^n, \bar{Y}^n, \bar{Z}^n)\}$ be a sequence of RVs such that $(X_1, \bar{Y}_1, \bar{Z}_1), (X_2, \bar{Y}_2, \bar{Z}_2), \cdots, (X_n, \bar{Y}_n, \bar{Z}_n)$ are independent and

$$P_{X_i \bar{Y}_i \bar{Z}_i}(x, y, z) = P_{X_i Y_i Z_i}(x, y, z),$$

where $P_{X_i Y_i Z_i}(x, y, z)$ is the i-th marginal distribution of (X^n, Y^n, Z^n). Then, according to [14, Lemma 5.8.1 and 5.8.2], we have $\overline{D}_1(\mathbf{X}, \mathbf{Y}) \geq \overline{D}_1(\mathbf{X}, \bar{\mathbf{Y}})$, $\overline{D}_2(\mathbf{X}, \mathbf{Z}) \geq \overline{D}_2(\mathbf{X}, \bar{\mathbf{Z}})$, $\overline{I}(\mathbf{X};\mathbf{Y}) \geq \overline{I}(\mathbf{X};\bar{\mathbf{Y}})$, and $\overline{I}(\mathbf{X};\mathbf{Y},\mathbf{Z}) \geq \overline{I}(\mathbf{X};\bar{\mathbf{Y}},\bar{\mathbf{Z}})$. Thus, by introducing the set $\mathcal{I}$ of probability distributions for independent RVs as

$$\mathcal{I} \triangleq \left\{ \mathbf{P}_{\mathbf{YZ|X}} = \{P_{Y^n Z^n|X^n}\} : P_{Y^n Z^n|X^n} = \prod_{i=1}^{n} P_{Y_i Z_i|X_i}, \right.$$
$$\left. \exists P_{Y_i Z_i|X_i} \in \mathcal{P}_{\mathcal{Y}\mathcal{Z}|\mathcal{X}}, i \in [1:n] \right\},$$

we have

$$\bigcup_{\substack{\mathbf{P}_{\mathbf{YZ|X}}: \\ \overline{D}_1(\mathbf{X},\mathbf{Y}) \leq D_1, \overline{D}_2(\mathbf{X},\mathbf{Z}) \leq D_2}} \{(R_1, R_2) : R_1 \geq \overline{I}(\mathbf{X};\mathbf{Y}),$$
$$R_1 + R_2 \geq \overline{I}(\mathbf{X};\mathbf{Y},\mathbf{Z})\}$$
$$\subseteq \bigcup_{\substack{\mathbf{P}_{\mathbf{YZ|X}} \in \mathcal{I}: \\ \overline{D}_1(\mathbf{X},\mathbf{Y}) \leq D_1, \overline{D}_2(\mathbf{X},\mathbf{Z}) \leq D_2}} \{(R_1, R_2) : R_1 \geq \overline{I}(\mathbf{X};\mathbf{Y}),$$
$$R_1 + R_2 \geq \overline{I}(\mathbf{X};\mathbf{Y},\mathbf{Z})\}. \qquad (37)$$

On the other hand, in the same way as [14, p.372], for any $\delta > 0, \gamma > 0$, and any $\mathbf{P}_{\mathbf{YZ|X}} \in \mathcal{I}$ such that $\overline{D}_1(\mathbf{X}, \mathbf{Y}) \leq D_1$ and $\overline{D}_2(\mathbf{X}, \mathbf{Z}) \leq D_2$, there exists $P_{YZ|X} \in \mathcal{P}_{\mathcal{Y}\mathcal{Z}|\mathcal{X}}$ such that

$$\overline{I}(\mathbf{X};\mathbf{Y}) \geq I(X;Y) - \delta,$$
$$\overline{I}(\mathbf{X};\mathbf{Y},\mathbf{Z}) \geq I(X;Y,Z) - \delta,$$

and

$$D_1 \geq \mathrm{E}[d_1(X,Y)] - \gamma,$$
$$D_2 \geq \mathrm{E}[d_2(X,Z)] - \gamma.$$

Thus, we have

$$\bigcup_{\substack{\mathbf{P_{YZ|X}} \in \mathcal{I}: \\ \overline{D}_1(\mathbf{X,Y}) \le D_1, \overline{D}_2(\mathbf{X,Z}) \le D_2}} \{(R_1, R_2) : R_1 \ge \overline{I}(\mathbf{X;Y}),$$

$$R_1 + R_2 \ge \overline{I}(\mathbf{X;Y,Z})\}$$

$$\subseteq \bigcap_{\gamma > 0} \bigcap_{\delta > 0} \bigcup_{P_{YZ|X} \in \mathcal{P}_M(D+\gamma|X)} \{(R_1, R_2) :$$

$$R_1 \ge I(X;Y) - \delta, R_1 + R_2 \ge I(X;Y,Z) - \delta\}$$

$$= \bigcap_{\gamma > 0} \bigcap_{\delta > 0} \{(R_1, R_2) : R_1 + \delta \ge R_1(D_1 + \gamma),$$

$$R_1 + R_2 + \delta \ge R_b(R_1 + \delta, D_1 + \gamma, D_2 + \gamma)\}, \quad (38)$$

where the last equality comes from (35).

Since $R_1(D_1)$ and $R_b(R_1, D_1, D_2)$ are convex and continuous functions of a triple (R_1, D_1, D_2) (see Remark 12), it holds that for any $\epsilon > 0$, there exist sufficiently small $\gamma_\epsilon > 0$ and $\delta_\epsilon > 0$ such that

$$\bigcap_{\delta > 0} \bigcap_{\gamma > 0} \{(R_1, R_2) : R_1 + \delta \ge R_1(D_1 + \gamma),$$

$$R_1 + R_2 + \delta \ge R_b(R_1 + \delta, D_1 + \gamma, D_2 + \gamma)\}$$

$$\subseteq \bigcap_{\delta_\epsilon > \delta > 0} \bigcap_{\gamma_\epsilon > \gamma > 0} \{(R_1, R_2) : R_1 + \delta \ge R_1(D_1 + \gamma),$$

$$R_1 + R_2 + \delta \ge R_b(R_1 + \delta, D_1 + \gamma, D_2 + \gamma)\}$$

$$\subseteq \bigcap_{\delta_\epsilon > \delta > 0} \bigcap_{\gamma_\epsilon > \gamma > 0} \{(R_1, R_2) : R_1 \ge R_1(D_1) - \delta - \epsilon,$$

$$R_1 + R_2 \ge R_b(R_1, D_1, D_2) - \delta - \epsilon\}$$

$$= \bigcap_{\delta_\epsilon > \delta > 0} \{(R_1, R_2) : R_1 \ge R_1(D_1) - \delta - \epsilon,$$

$$R_1 + R_2 \ge R_b(R_1, D_1, D_2) - \delta - \epsilon\}. \quad (39)$$

By combining (36), (37), (38), and (39), we have

$$\bigcup_{\mathbf{P_{UYZ|X}} \in \mathcal{P}_G(D|\mathbf{X})} \mathcal{R}_G(\mathbf{P_{UYZ|X}|X})$$

$$\subseteq \bigcap_{\epsilon > 0} \bigcap_{\delta_\epsilon > \delta > 0} \{(R_1, R_2) : R_1 \ge R_1(D_1) - \delta - \epsilon,$$

$$R_1 + R_2 \ge R_b(R_1, D_1, D_2) - \delta - \epsilon\}$$

$$= \{(R_1, R_2) : R_1 \ge R_1(D_1), R_1 + R_2 \ge R_b(R_1, D_1, D_2)\}.$$

According to Remark 12, this completes the proof. $\qquad \square$

Proof: The LHS of (34) $\supseteq$ *The RHS of* (34). Since $|\mathcal{U}| \ge |\mathcal{Z}| + 1$, there exists an injective function id $: \mathcal{Z} \to \mathcal{U}$. For this function, let $\mathcal{U}_{id} \subseteq \mathcal{U}$ be the image of id and $\mathrm{id}^{-1} : \mathcal{U}_{id} \to \mathcal{Z}$ be the inverse function of id on $\mathcal{U}_{id}$. Let $u^* \in \mathcal{U}$ be a symbol such that $u^* \notin \mathcal{U}_{id}$.

For any $P_{YZ|X} \in \mathcal{P}_M(D|X)$ and any $\alpha \in [0, 1]$, we define $P_{UYZ|X} \in \mathcal{P}_{\mathcal{U}\mathcal{Y}\mathcal{Z}|\mathcal{X}}$ as

$$P_{UYZ|X}(u, y, z|x) \triangleq \begin{cases} \alpha P_{YZ|X}(y, z|x) & \text{if } u = \mathrm{id}(z), \\ (1 - \alpha) P_{YZ|X}(y, z|x) & \text{if } u = u^*, \\ 0 & \text{otherwise.} \end{cases}$$

Since

$$P_{XUY}(x, u, y) = \begin{cases} \alpha P_{XYZ}(x, y, \mathrm{id}^{-1}(u)) & \text{if } u \in \mathcal{U}_{id}, \\ (1 - \alpha) P_{XY}(x, y) & \text{if } u = u^*, \\ 0 & \text{otherwise,} \end{cases}$$

$$P_{X|UY}(x|u, y) = \begin{cases} P_{X|YZ}(x|y, \mathrm{id}^{-1}(u)) & \text{if } u \in \mathcal{U}_{id}, \\ P_{X|Y}(x|y) & \text{if } u = u^*, \\ 0 & \text{otherwise,} \end{cases}$$

we have

$$I(X; U, Y) = H(X) - \alpha H(X|Y, Z) - (1 - \alpha) H(X|Y)$$

$$= \alpha I(X; Y, Z) + (1 - \alpha) I(X; Y).$$

Furthermore, since

$$P_{X|UYZ}(x|u, y, z) = \begin{cases} P_{X|YZ}(x|y, z) & \text{if } u = \mathrm{id}(z), \\ P_{X|YZ}(x|y, z) & \text{if } u = u^*, \\ 0 & \text{otherwise,} \end{cases}$$

we have

$$I(X; Z|U, Y) = H(X|U, Y) - H(X|U, Y, Z)$$

$$= \alpha H(X|Y, Z) + (1 - \alpha) H(X|Y) - H(X|Y, Z)$$

$$= (1 - \alpha) I(X; Z|Y).$$

Now by defining $\mathbf{P_{UYZ|X}}$ as

$$P_{U^n Y^n Z^n | X^n}(u^n, y^n, z^n | x^n) = \prod_{i=1}^{n} P_{UYZ|X}(u_i, y_i, z_i | x_i),$$

$(\mathbf{X, U, Y, Z})$ becomes a sequence of independent copies of RVs (X, U, Y, Z). Thus, we have

$$\overline{I}(\mathbf{X; U, Y}) = I(X; U, Y) = \alpha I(X; Y, Z) + (1 - \alpha) I(X; Y),$$

$$\overline{I}(\mathbf{X; Z|U, Y}) = I(X; Z|U, Y) = (1 - \alpha) I(X; Z|Y),$$

$$\overline{D}_1(\mathbf{X, Y}) = \mathrm{E}[d_1(X, Y)] \le D_1,$$

$$\overline{D}_2(\mathbf{X, Z}) = \mathrm{E}[d_2(X, Z)] \le D_2,$$

where we use the fact that for i.i.d. RVs $\{A_i\}_{i=1}^{\infty}$ taking values in a finite set, $\text{p-}\limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} A_i = \mathrm{E}[A_1]$. Hence, by noting that $\mathbf{P_{UYZ|X}} \in \mathcal{P}_G(D|\mathbf{X})$, we have for any $\alpha \in [0, 1]$,

$$(\alpha I(X; Y, Z) + (1 - \alpha) I(X; Y), (1 - \alpha) I(X; Z|Y))$$

$$\in \bigcup_{\mathbf{P_{UYZ|X}} \in \mathcal{P}_G(D|\mathbf{X})} \mathcal{R}_G(\mathbf{P_{UYZ|X}|X}). \quad (40)$$

This implies that for any $(R_1, R_2) \in \mathcal{R}_M(P_{YZ|X}|X)$, it holds that $(R_1, R_2) \in \bigcup_{\mathbf{P_{UYZ|X}} \in \mathcal{P}_G(D|\mathbf{X})} \mathcal{R}_G(\mathbf{P_{UYZ|X}|X})$. This is because for any $(R_1, R_2) \in \mathcal{R}_M(P_{YZ|X}|X)$ such that $I(X; Y, Z) \ge R_1$, there exists $\alpha \in [0, 1]$ such that

$$R_1 = \alpha I(X; Z|Y) + I(X; Y)$$

$$= \alpha I(X; Y, Z) + (1 - \alpha) I(X; Y).$$

Then, we have

$$R_2 \geq I(X;Y,Z) - R_1$$
$$= I(X;Y,Z) - \alpha I(X;Y,Z) - (1-\alpha)I(X;Y)$$
$$= (1-\alpha)I(X;Z|Y). \tag{41}$$

According to (40), such pair (R_1, R_2) is included in the region $\bigcup_{\mathbf{P}_{\mathbf{UYZ}|\mathbf{X}} \in \mathcal{P}_{\mathrm{G}}(D|\mathbf{X})} \mathcal{R}_{\mathrm{G}}(\mathbf{P}_{\mathbf{UYZ}|\mathbf{X}}|\mathbf{X})$. On the other hand, for any $(R_1, R_2) \in \mathcal{R}_{\mathrm{M}}(P_{YZ|X}|X)$ such that $R_1 > I(X;Y,Z)$, we have $R_2 \geq 0$. Since it holds that $(I(X;Y,Z),0) \in \bigcup_{\mathbf{P}_{\mathbf{UYZ}|\mathbf{X}} \in \mathcal{P}_{\mathrm{G}}(D|\mathbf{X})} \mathcal{R}_{\mathrm{G}}(\mathbf{P}_{\mathbf{UYZ}|\mathbf{X}}|\mathbf{X})$ due to (40), such pair (R_1, R_2) is also included in the region $\bigcup_{\mathbf{P}_{\mathbf{UYZ}|\mathbf{X}} \in \mathcal{P}_{\mathrm{G}}(D|\mathbf{X})} \mathcal{R}_{\mathrm{G}}(\mathbf{P}_{\mathbf{UYZ}|\mathbf{X}}|\mathbf{X})$.

Therefore, we have

$$\bigcup_{\mathbf{P}_{\mathbf{UYZ}|\mathbf{X}} \in \mathcal{P}_{\mathrm{G}}(D|\mathbf{X})} \mathcal{R}_{\mathrm{G}}(\mathbf{P}_{\mathbf{UYZ}|\mathbf{X}}|\mathbf{X}) \supseteq \mathcal{R}_{\mathrm{M}}(P_{YZ|X}|X).$$

Since this holds for any $P_{YZ|X} \in \mathcal{P}_{\mathrm{M}}(D|X)$, this completes the proof. □

Remark 13. Unlike the rate-distortion region by Rimoldi, our region includes a sequence $\mathbf{U}$ of auxiliary RVs. This comes from the fact that the time-sharing argument as in (41) cannot be employed because it holds that in general,

$$\bar{I}(\mathbf{X};\mathbf{Y},\mathbf{Z}) \neq \bar{I}(\mathbf{X};\mathbf{Y}) + \bar{I}(\mathbf{X};\mathbf{Z}|\mathbf{Y}).$$

5.4 Mixed Sources

In this section, we give the rate-distortion region for mixed sources.

The mixed source $\mathbf{X}$ is defined by $\mathbf{X}_1 = \{X_1^n\}_{n=1}^{\infty}$ and $\mathbf{X}_2 = \{X_2^n\}_{n=1}^{\infty}$ as

$$P_{X^n}(x^n) = \alpha_1 P_{X_1^n}(x^n) + \alpha_2 P_{X_2^n}(x^n),$$

where $\alpha_1, \alpha_2 \in [0,1]$ and $\alpha_1 + \alpha_2 = 1$.

The next lemma shows a fundamental property of the information spectrum of mixed sources.

Lemma 6. For sequences of RVs $(\mathbf{X}_1, \mathbf{Y}_1, \mathbf{Z}_1)$ and $(\mathbf{X}_2, \mathbf{Y}_2, \mathbf{Z}_2)$, let $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ be defined by

$$P_{X^n Y^n Z^n}(x^n, y^n, z^n) = \alpha_1 P_{X_1^n Y_1^n Z_1^n}(x^n, y^n, z^n)$$
$$+ \alpha_2 P_{X_2^n Y_2^n Z_2^n}(x^n, y^n, z^n).$$

Then, we have

$$\bar{I}(\mathbf{X};\mathbf{Y}) = \max\{\bar{I}(\mathbf{X}_1;\mathbf{Y}_1), \bar{I}(\mathbf{X}_2;\mathbf{Y}_2)\},$$
$$\bar{I}(\mathbf{X};\mathbf{Y}|\mathbf{Z}) = \max\{\bar{I}(\mathbf{X}_1;\mathbf{Y}_1|\mathbf{Z}_1), \bar{I}(\mathbf{X}_2;\mathbf{Y}_2|\mathbf{Z}_2)\}.$$

Proof. Since this lemma can be proved in the same way as [14, Lemma 7.9.1] by using [14, Lemma 1.4.2], we omit the details. □

The next theorem shows that the rate-distortion region for a mixed source is the intersection of those of two sources.

Theorem 6. For a mixed source $\mathbf{X}$ defined by $\mathbf{X}_1$ and $\mathbf{X}_2$, and any real numbers $D_1, D_2 \geq 0$, we have

$$\mathcal{R}(D|\mathbf{X}) = \mathcal{R}(D|\mathbf{X}_1) \cap \mathcal{R}(D|\mathbf{X}_2).$$

Proof. For $\mathbf{P}_{\mathbf{U}_1\mathbf{Y}_1\mathbf{Z}_1|\mathbf{X}_1}$ and $\mathbf{P}_{\mathbf{U}_2\mathbf{Y}_2\mathbf{Z}_2|\mathbf{X}_2}$, we define a mixture $\mathbf{P}_{\tilde{\mathbf{U}}\tilde{\mathbf{Y}}\tilde{\mathbf{Z}}|\mathbf{X}}$ of these two components by

$$P_{\tilde{U}^n \tilde{Y}^n \tilde{Z}^n | X^n}(u^n, y^n, z^n | x^n)$$
$$= \frac{\alpha_1 P_{X_1^n}(x^n) P_{U_1^n Y_1^n Z_1^n | X_1^n}(u^n, y^n, z^n | x^n)}{P_{X^n}(x^n)}$$
$$+ \frac{\alpha_2 P_{X_2^n}(x^n) P_{U_2^n Y_2^n Z_2^n | X_2^n}(u^n, y^n, z^n | x^n)}{P_{X^n}(x^n)}.$$

In order to prove the theorem, we give an equivalent expression of the rate-distortion region using $\mathbf{P}_{\mathbf{U}_1\mathbf{Y}_1\mathbf{Z}_1|\mathbf{X}_1}$, $\mathbf{P}_{\mathbf{U}_2\mathbf{Y}_2\mathbf{Z}_2|\mathbf{X}_2}$, and the mixture $\mathbf{P}_{\tilde{\mathbf{U}}\tilde{\mathbf{Y}}\tilde{\mathbf{Z}}|\mathbf{X}}$.

When $\mathbf{P}_{\mathbf{U}_1\mathbf{Y}_1\mathbf{Z}_1|\mathbf{X}_1} = \mathbf{P}_{\mathbf{U}_2\mathbf{Y}_2\mathbf{Z}_2|\mathbf{X}_2} = \mathbf{P}_{\mathbf{UYZ}|\mathbf{X}}$, we have $\mathbf{P}_{\tilde{\mathbf{U}}\tilde{\mathbf{Y}}\tilde{\mathbf{Z}}|\mathbf{X}} = \mathbf{P}_{\mathbf{UYZ}|\mathbf{X}}$ by the definition. This implies that for any $\mathbf{P}_{\mathbf{UYZ}|\mathbf{X}}$, there exist $\mathbf{P}_{\mathbf{U}_1\mathbf{Y}_1\mathbf{Z}_1|\mathbf{X}_1}$ and $\mathbf{P}_{\mathbf{U}_2\mathbf{Y}_2\mathbf{Z}_2|\mathbf{X}_2}$ such that $\mathbf{P}_{\mathbf{UYZ}|\mathbf{X}} = \mathbf{P}_{\tilde{\mathbf{U}}\tilde{\mathbf{Y}}\tilde{\mathbf{Z}}|\mathbf{X}}$. On the other hand, for any $\mathbf{P}_{\mathbf{U}_1\mathbf{Y}_1\mathbf{Z}_1|\mathbf{X}_1}$ and $\mathbf{P}_{\mathbf{U}_2\mathbf{Y}_2\mathbf{Z}_2|\mathbf{X}_2}$, there trivially exists $\mathbf{P}_{\mathbf{UYZ}|\mathbf{X}}$ such that $\mathbf{P}_{\mathbf{UYZ}|\mathbf{X}} = \mathbf{P}_{\tilde{\mathbf{U}}\tilde{\mathbf{Y}}\tilde{\mathbf{Z}}|\mathbf{X}}$. Thus, we have an equivalent expression:

$$\bigcup_{\mathbf{P}_{\mathbf{UYZ}|\mathbf{X}} \in \mathcal{P}_{\mathrm{G}}(D|\mathbf{X})} \mathcal{R}_{\mathrm{G}}(\mathbf{P}_{\mathbf{UYZ}|\mathbf{X}}|\mathbf{X})$$
$$= \bigcup_{\substack{\mathbf{P}_{\mathbf{U}_1\mathbf{Y}_1\mathbf{Z}_1|\mathbf{X}_1}, \mathbf{P}_{\mathbf{U}_2\mathbf{Y}_2\mathbf{Z}_2|\mathbf{X}_2}: \\ \bar{D}_1(\mathbf{X},\tilde{\mathbf{Y}}) \leq D_1, \bar{D}_2(\mathbf{X},\tilde{\mathbf{Z}}) \leq D_2}} \mathcal{R}_{\mathrm{G}}(\mathbf{P}_{\tilde{\mathbf{U}}\tilde{\mathbf{Y}}\tilde{\mathbf{Z}}|\mathbf{X}}|\mathbf{X}).$$

Thus, according to Theorem 4, we have

$$\mathcal{R}(D|\mathbf{X})$$
$$= \bigcup_{\substack{\mathbf{P}_{\mathbf{U}_1\mathbf{Y}_1\mathbf{Z}_1|\mathbf{X}_1}, \mathbf{P}_{\mathbf{U}_2\mathbf{Y}_2\mathbf{Z}_2|\mathbf{X}_2}: \\ \bar{D}_1(\mathbf{X},\tilde{\mathbf{Y}}) \leq D_1, \bar{D}_2(\mathbf{X},\tilde{\mathbf{Z}}) \leq D_2}} \mathcal{R}_{\mathrm{G}}(\mathbf{P}_{\tilde{\mathbf{U}}\tilde{\mathbf{Y}}\tilde{\mathbf{Z}}|\mathbf{X}}|\mathbf{X})$$
$$\overset{(a)}{=} \bigcup_{\substack{\mathbf{P}_{\mathbf{U}_1\mathbf{Y}_1\mathbf{Z}_1|\mathbf{X}_1}, \mathbf{P}_{\mathbf{U}_2\mathbf{Y}_2\mathbf{Z}_2|\mathbf{X}_2}: \\ \max\{\bar{D}_1(\mathbf{X}_1,\mathbf{Y}_1), \bar{D}_1(\mathbf{X}_2,\mathbf{Y}_2)\} \leq D_1, \\ \max\{\bar{D}_2(\mathbf{X}_1,\mathbf{Z}_1), \bar{D}_2(\mathbf{X}_2,\mathbf{Z}_2)\} \leq D_2}} \{(R_1, R_2):$$
$$R_1 \geq \max\{\bar{I}(\mathbf{X}_1;\mathbf{U}_1,\mathbf{Y}_1), \bar{I}(\mathbf{X}_2;\mathbf{U}_2,\mathbf{Y}_2)\}$$
$$R_2 \geq \max\{\bar{I}(\mathbf{X}_1;\mathbf{Z}_1|\mathbf{U}_1,\mathbf{Y}_1), \bar{I}(\mathbf{X}_2;\mathbf{Z}_2|\mathbf{U}_2,\mathbf{Y}_2)\}\}$$
$$= \bigcup_{\mathbf{P}_{\mathbf{U}_1\mathbf{Y}_1\mathbf{Z}_1|\mathbf{X}_1} \in \mathcal{P}_{\mathrm{G}}(D|\mathbf{X}_1)} \bigcup_{\mathbf{P}_{\mathbf{U}_2\mathbf{Y}_2\mathbf{Z}_2|\mathbf{X}_2} \in \mathcal{P}_{\mathrm{G}}(D|\mathbf{X}_2)}$$
$$\mathcal{R}_{\mathrm{G}}(\mathbf{P}_{\mathbf{U}_1\mathbf{Y}_1\mathbf{Z}_1|\mathbf{X}_1}|\mathbf{X}_1) \cap \mathcal{R}_{\mathrm{G}}(\mathbf{P}_{\mathbf{U}_2\mathbf{Y}_2\mathbf{Z}_2|\mathbf{X}_2}|\mathbf{X}_2)$$
$$= \mathcal{R}(D|\mathbf{X}_1) \cap \mathcal{R}(D|\mathbf{X}_2).$$

where (a) comes from [14, Lemma 1.4.2], Lemma 6 and the fact that

$$P_{X^n \tilde{U}^n \tilde{Y}^n \tilde{Z}^n}(x^n, u^n, y^n, z^n)$$
$$= \alpha_1 P_{X_1^n U_1^n Y_1^n Z_1^n}(x^n, u^n, y^n, z^n)$$
$$+ \alpha_2 P_{X_2^n U_2^n Y_2^n Z_2^n}(x^n, u^n, y^n, z^n).$$

This completes the proof. □

6. Conclusion

In this paper, we have dealt with the successive refinement

problem. We gave inner and outer bounds using the smooth max Rényi divergence on the set of pairs of numbers of codewords. These bounds are obtained by extended versions of our previous covering lemma and converse bound. By using these bounds, we also gave a general formula using the spectral sup-mutual information rate for the rate-distortion region. Further, we showed some special cases of our rate-distortion region for discrete stationary memoryless sources and mixed sources.

Acknowledgments

References

[1] T. Matsuta and T. Uyematsu, "Non-asymptotic bounds on numbers of codewords for the successive refinement problem," Proc. 38th Symp. on Inf. Theory and its Apps. (SITA2015), pp.43–48, Nov. 2015.

[2] T. Matsuta and T. Uyematsu, "A general formula of the achievable rate region for the successive refinement problem," Proc. IEICE Society Conference, p.37, Sept. 2016.

[3] V.N. Koshelev, "Estimation of mean error for a discrete successive-approximation scheme," Problemy Peredachi Informatsii, vol.17, no.3, pp.20–33, 1981.

[4] V.N. Koshelev, "Hierarchical coding of discrete sources," Problemy Peredachi Informatsii, vol.16, no.3, pp.31–49, 1980.

[5] W. Equitz and T. Cover, "Successive refinement of information," IEEE Trans. Inf. Theory, vol.37, no.2, pp.269–275, March 1991.

[6] B. Rimoldi, "Successive refinement of information: Characterization of the achievable rates," IEEE Trans. Inf. Theory, vol.40, no.1, pp.253–259, Jan. 1994.

[7] H. Yamamoto, "Source coding theory for a triangular communication system," IEEE Trans. Inf. Theory, vol.42, no.3, pp.848–853, May 1996.

[8] M. Effros, "Distortion-rate bounds for fixed- and variable-rate multiresolution source codes," IEEE Trans. Inf. Theory, vol.45, no.6, pp.1887–1910, Sept. 1999.

[9] A. No, A. Ingber, and T. Weissman, "Strong successive refinability and rate-distortion-complexity tradeoff," IEEE Trans. Inf. Theory, vol.62, no.6, pp.3618–3635, June 2016.

[10] L. Zhou, V.Y.F. Tan, and M. Motani, "Second-order and moderate deviations asymptotics for successive refinement," IEEE Trans. Inf. Theory, vol.63, no.5, pp.2896–2921, May 2017.

[11] N.A. Warsi, "One-shot bounds for various information theoretic problems using smooth min and max Rényi divergences," Proc. 2013 IEEE Inf. Theory Workshop, pp.1–5, Sept. 2013.

[12] T. Uyematsu and T. Matsuta, "Revisiting the rate-distortion theory using smooth max Rényi divergence," Proc. 2014 IEEE Inf. Theory Workshop, pp.202–206, Nov. 2014.

[13] T. Matsuta and T. Uyematsu, "New non-asymptotic bounds on numbers of codewords for the fixed-length lossy compression," IEICE Trans. Fundamentals, vol.E99-A, no.12, pp.2116–2129, Dec. 2016.

[14] T.S. Han, Information-Spectrum Methods in Information Theory, Springer, 2003.

[15] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," IEEE Trans. Inf. Theory, vol.58, no.6, pp.3309–3338, June 2012.

[16] T.M. Cover and J.A. Thomas, Elements of Information Theory, 2nd ed., Wiley, New York, 2006.

[17] V. Kostina and E. Tuncel, "The rate-distortion function for successive refinement of abstract sources," Proc. IEEE Int. Symp. on Inf. Theory, pp.1923–1927, June 2017.

[18] S. Verdú, "α-mutual information," 2015 Inf. Theory and Apps. Workshop, pp.1–6, Feb. 2015.

[19] A. Kanlis and P. Narayan, "Error exponents for successive refinement by partitioning," IEEE Trans. Inf. Theory, vol.42, no.1, pp.275–282, Jan. 1996.

[20] V. Kostina and E. Tuncel, "Successive refinement of abstract sources," arXiv preprint arXiv:1707.09567, July 2017.

Tetsunao Matsuta was born in Fukui, Japan, on March 17, 1985. He received the B.E. degree from Utsunomiya University in 2007. He received the M.E. degree, and the Dr.Eng. degree from Tokyo Institute of Technology, Japan, in 2009, 2012, respectively. Since 2012, he has been an assistant professor at Tokyo Institute of Technology, and currently he is in the Department of Information and Communications Engineering of Tokyo Institute of Technology. He received the Best Paper Award in 2016 from IEICE. His current research interests are in the area of multi-terminal information theory and non-asymptotic analysis of coding problems.

Tomohiko Uyematsu received the B.E., M.E. and Dr.Eng. degrees from Tokyo Institute of Technology in 1982, 1984 and 1988, respectively. From 1984 to 1992, he was with the Department of Electrical and Electronic Engineering of Tokyo Institute of Technology, first as research associate, next as lecturer, and lastly as associate professor. From 1992 to 1997, he was with School of Information Science of Japan Advanced Institute of Science and Technology as associate professor. Since 1997, he returned to Tokyo Institute of Technology as associate professor, and currently he is with the Department of Information and Communications Engineering as professor. In 1992 and 1996, he was a visiting researcher at the Supelec (Ecole superieure d'electricite), France and Delft University of Technology, Netherlands, respectively. He was an associate editor of IEEE Trans. Information Theory from 2010 to 2013. He received the Achievement Award in 2008, and the Best Paper Award in 1993, 1996, 2002, 2007, 2011, 2014 and 2016 both from IEICE. His current research interests are in the areas of information theory, especially Shannon theory and multi-terminal information theory. Dr. Uyematsu is a senior member of IEEE.