4-11-2011

Wyner’s Common Information for Continuous Random Variables - A Lossy Source Coding Interpretation

Ge Xu
Syracuse University

Wei Liu
Syracuse University

Biao Chen
Syracuse University, Department of Electrical Engineering and Computer Science, bchien@ecs.syr.edu

Department of Electrical Engineering and Computer Science

Technical Report

SYR-EECS-2011-06 April 11, 2011

Wyner's Common Information for Continuous Random Variables - A Lossy Source Coding Interpretation

Ge Xu gexu@syr.edu

Wei Liu wliu28@syr.edu

Biao Chen bichen@syr.edu

ABSTRACT: Wyner's common information can be easily generalized for continuous random variables. We provide an operational meaning for such generalization using the Gray-Wyner network with *lossy* source coding. Specifically, a Gray-Wyner network consists of one encoder and two decoders. A sequence of independent copies of a pair of random variables $(X, Y) \sim p(x, y)$ is encoded into three messages, one of them is a common input to both decoders. The two decoders attempt to reconstruct the two sequences respectively subject to individual distortion constraints. We show that Wyner's common information equals the smallest common message rate when the total rate is arbitrarily close to the rate-distortion function with joint decoding. A surprising observation is that such equality holds independent of the values of distortion constraints as long as the distortions are less than certain thresholds. An interpretation for such thresholds is given for the symmetric case.

KEYWORDS: Wyner's common information, lossy source coding, continuous random variables, Gaussian random variables

Syracuse University - Department of EECS,
4-206 CST, Syracuse, NY 13244
(P) 315.443.2652 (F) 315.443.2583
http://eecs.syr.edu

Wyner's Common Information For Continuous Random Variables - A Lossy Source Coding Interpretation

Ge Xu
Department of EECS
Syracuse University
Email: gexu@syr.edu

Wei Liu
Department of EECS
Syracuse University
Email: wliu28@syr.edu

Biao Chen
Department of EECS
Syracuse University
Email: bichen@syr.edu

***Abstract*— Wyner's common information can be easily generalized for continuous random variables. We provide an operational meaning for such generalization using the Gray-Wyner network with *lossy* source coding. Specifically, a Gray-Wyner network consists of one encoder and two decoders. A sequence of independent copies of a pair of random variables $(X, Y) \sim p(x, y)$ is encoded into three messages, one of them is a common input to both decoders. The two decoders attempt to reconstruct the two sequences respectively subject to individual distortion constraints. We show that Wyner's common information equals the smallest common message rate when the total rate is arbitrarily close to the rate-distortion function with joint decoding. A surprising observation is that such equality holds independent of the values of distortion constraints as long as the distortions are less than certain thresholds. An interpretation for such thresholds is given for the symmetric case.**

I. Introduction

Classical notions that characterize the amount of common information contained in a pair of dependent random variables $(X, Y) \sim p(x, y)$ include Shannon's mutual information and Gács and Körner's common randomness [1]. Each of these notions carries clearly defined operational meaning. Wyner provided yet another definition to quantify the common information for a pair of discrete random variables with finite alphabet [3]:

$$C(X; Y) = \inf_{X-W-Y} I(X, Y; W). \tag{1}$$

Here, the infimum is taken over all auxiliary random variables W such that X, W, and Y forms a Markov chain, i.e., X and Y are conditionally independent given W.

Wyner's common information also carries its own operational meanings, two of them were given in Wyner's original paper [3]. The first approach is shown in Fig. 1. The encoder observes a pair of sequences (X^n, Y^n) and outputs three messages W_0, W_1, and W_2 where

$$W_i \in \{1, \cdots, 2^{nR_i}\}$$

for $i = 0, 1, 2$. Decoder 1 recontructs X^n from messages (W_0, W_1), and decoder 2 reconstructs Y^n from (W_0, W_2). For any given ϵ, define C_1 to be the minimum R_0 for the system in Fig. 1 such that the total rate $\sum_{i=0}^{2} R_i < H(X, Y) + \epsilon$ and the probabilities of error at both decoders are bounded by ϵ.

Wyner's second approach is shown in Fig. 2. In this model, a common message W uniformly distributed on $\mathcal{W} = \{1, \cdots, 2^{nR_0}\}$ is passed to two separate random number generators, whose outputs are generated independently according to distributions $p_1(X|W)$ and $p_2(Y|W)$. The output sequences of the two processors, denoted as $\tilde{X}^n$ and $\tilde{Y}^n$ respectively, have joint distribution

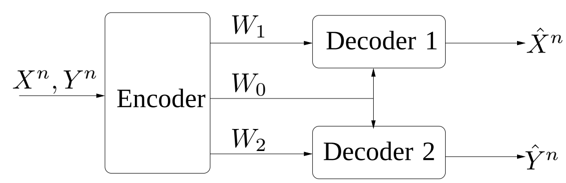

Fig. 1. Gray-Wyner source coding network.

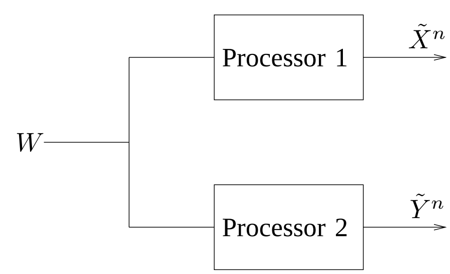

Fig. 2. Distribution Approximation.

$$p(\tilde{X}^n, \tilde{Y}^n) = \sum_{w \in \mathcal{W}} \frac{1}{|\mathcal{W}|} p_1(X^n|W) p_2(Y^n|W). \tag{2}$$

Let C_2 be the smallest R_0 such that the distribution in (2) is arbitrarily close to $Q(X^n, Y^n)$, in the sense that

$$D_n(Q; P) = \frac{1}{n} \sum_{x^n \in \mathcal{X}^n, y^n \in \mathcal{Y}^n} P(x^n, y^n) \log \frac{P(x^n, y^n)}{Q(x^n, y^n)} \tag{3}$$

can be made arbitrarily small.

Wyner proved that

$$C(X, Y) = C_1 = C_2, \tag{4}$$

lending clear practical interpretations to $C(X, Y)$ defined in equation (1). However, these interpretations only apply to random variables with finite alphabet sets. Indeed, Wyner's common information was originally defined only for a pair of finite-alphabet discrete random variables. Its generalization to multiple dependent variables was first mentioned in [6] and further developed in [7]. The definition in (1) also applies to random variables with continuous alphabet. However, it is not clear what physical interpretation such quantity carries for continuous random variables.

We provide such an interpretation using the rate distortion result for the Gray-Wyner network as described in Fig. 1. That is, instead of requiring the sources to be reproduced losslessly at the two decoders, we allow certain distortions subject to given distortion constraints [9]. It turns out that Wyner's common information is precisely the smallest common message rate for a certain range of distortion constraints when the total rate is arbitrarily close to the rate distortion function with joint decoding. A surprising result is that, as Wyner's common information is only a function of the joint distribution, this smallest common rate remains constant even if the distortion constraints vary, as long as they are less than certain thresholds.

The rest of the paper is organized as follows. Section II gives the problem formulation and the main results. The proofs are given in Appendix. In section III, two examples, the doubly symmetric binary source studied in [3] and the bivariate Gaussian source, are given. We conclude in Section IV.

II. Problem formulation and main result

Let $\{(X_k, Y_k)\}_{k=1}^{\infty}$ be independent copies of a pair of dependent random variables $(X, Y) \sim Q(x, y)$ which take values in some arbitrary (finite, countable, or continuous) spaces $\mathcal{X} \times \mathcal{Y}$. Here, we use $Q(x, y)$ to denote the joint distribution of (X, Y), i.e., probability mass function if (X, Y) are discrete and probability density function if (X, Y) are continuous. Thus the joint distribution of length n vectors (X^n, Y^n) is

$$Q^n(x^n, y^n) = \prod_{i=1}^{n} Q(x_i, y_i). \tag{5}$$

The common information of the pair (X, Y) is a functional of Q and is defined as

$$C(X, Y) \triangleq \inf I(X, Y; W), \tag{6}$$

where the infimum is taken over all random variable triples X, Y, W satisfying

- (C1) The marginal distribution for X, Y is $Q(x, y)$,
- (C2) X and Y are conditionally independent given W.

Let us consider the lossy source coding problem described in Fig. 1. The encoder observes a pair of sequences (X^n, Y^n), and map them to three messages W_0, W_1, W_2 with

$$W_i \in \{1, \cdots, 2^{nR_i}\},$$

for $i = 0, 1, 2$. Let $d_1(x, \hat{x})$ and $d_2(x, \hat{y})$ be bounded single letter distortion functions defined on $\mathcal{X} \times \mathcal{X}$ and $\mathcal{Y} \times \mathcal{Y}$ respectively. Decoder 1 reproduces X^n from (W_0, W_1) subject to an average distortion constraint Δ_1; decoder 2 reproduces Y^n from (W_0, W_2) subject to an average distortion constraint Δ_2. We now give a precise definition of the quantity $C_3(\Delta_1, \Delta_2)$, which is the smallest common rate R_0 such that the total rate meets the rate distortion bound with joint decoding.

Definition 1: An (n, M_0, M_1, M_2) rate distortion code consists of the following:

- One encoder mapping f_E
$$f_E \quad : \quad \mathcal{X}^n \times \mathcal{Y}^n \rightarrow I_{M_0} \times I_{M_1} \times I_{M_2}, \tag{7}$$
where $I_{M_i} = \{0, 1, 2, \cdots, M_i - 1\}$ for $i = 0, 1, 2$.
- Two decoder mappings $f_D^{(X)}, f_D^{(Y)}$
$$f_D^{(X)} \quad : \quad I_{M_0} \times I_{M_1} \rightarrow \mathcal{X}^n, \tag{8}$$
$$f_D^{(Y)} \quad : \quad I_{M_0} \times I_{M_2} \rightarrow \mathcal{Y}^n. \tag{9}$$

Let $f_E(X^n, Y^n) = (W_0, W_1, W_2)$, $1 \leq W_i \leq M_i$ and

$$\hat{X}^n = f_D^{(X)}(W_0, W_1), \tag{10}$$
$$\hat{Y}^n = f_D^{(Y)}(W_0, W_2). \tag{11}$$

Denote by (Δ_X, Δ_Y) the average distortion between encoder inputs and decoder outputs:

$$\Delta_X = Ed_1(X^n, \hat{X}^n), \tag{12}$$
$$\Delta_Y = Ed_2(Y^n, \hat{Y}^n), \tag{13}$$

where

$$d_1(x^n, \hat{x}^n) = \frac{1}{n} \sum_{k=1}^{n} d_1(x_k, \hat{x}_k), \tag{14}$$

$$d_2(y^n, \hat{y}^n) = \frac{1}{n} \sum_{k=1}^{n} d_2(y_k, \hat{y}_k). \tag{15}$$

An (n, M_0, M_1, M_2) code with distortion (Δ_X, Δ_Y) is referred to as an $(n, M_0, M_1, M_2, \Delta_X, \Delta_Y)$ rate distortion code.

Definition 2: For any $\Delta_1, \Delta_2 \geq 0$, a number R_0 is said to be (Δ_1, Δ_2)-achievable if for any $\epsilon > 0$ we can find n sufficiently large such that there exists a $(n, M_0, M_1, M_2, \Delta_X, \Delta_Y)$ rate distortion code with

$$M_0 \quad \leq \quad 2^{nR_0}, \tag{16}$$
$$\sum_{i=0}^{2} \frac{1}{n} \log M_i \quad \leq \quad R_{XY}(\Delta_1, \Delta_2) + \epsilon, \tag{17}$$
$$\Delta_X \leq \Delta_1 + \epsilon \quad , \quad \Delta_Y \leq \Delta_2 + \epsilon. \tag{18}$$

where $R_{XY}(\Delta_1, \Delta_2)$ is the rate distortion function for (X, Y) with joint encoding and decoding, i.e.,

$$R_{XY}(\Delta_1, \Delta_2) = \min I(X, Y; \hat{X}, \hat{Y}), \tag{19}$$

where the minimum is taken over all the test channels $q_t(\hat{x}, \hat{y}|x, y)$ such that $Ed_1(X, \hat{X}) \leq \Delta_1$, $Ed_2(Y, \hat{Y}) \leq \Delta_2$.

Definition 3: $C_3(\Delta_1, \Delta_2)$ is defined as the infimum of all R_0 that is (Δ_1, Δ_2)-achievable.

We now state the main results.

Theorem 1: The common information $C(X; Y) = C_3(\Delta_1, \Delta_2)$ in some neighborhood of the origin $\{(\Delta_1, \Delta_2) : 0 \leq \Delta_1, \Delta_2 \leq \gamma\}$ provided that

$$Q(x, y) > 0 \quad \text{all } x \in \mathcal{X}, y \in \mathcal{Y}, \tag{20}$$

and d_1, d_2 satisfy

$$d_1(x, \hat{x}) \quad > \quad d_1(x, x) = 0, x \neq \hat{x}, \tag{21}$$
$$d_2(y, \hat{y}) \quad > \quad d_2(y, y) = 0, y \neq \hat{y}. \tag{22}$$

A proof of Theorem 1 is given in Appendix A.

The condition on d_1 and d_2 set forth in the theorem amounts to requiring the distortion function be normal, as defined in [10].

If (X, Y) are discrete random variables with finite alphabet, and $d_1 = d_2 = d_H$ are the Hamming distortion, defined as

$$d_H(u, \hat{u}) = \begin{cases} 0, & u = \hat{u} \\ 1, & u \neq \hat{u}, \end{cases} \tag{23}$$

then for $\Delta_1 = \Delta_2 = 0$, $C_3(\Delta_1, \Delta_2) = C_1 = C(X;Y)$. Therefore, approach 1 in [3] is a special case of Theorem 1.

Theorem 2: For the symmetric case $\Delta_1 = \Delta_2 = \Delta$, $C_3(\Delta) = C(X,Y)$ if and only if $\Delta \leq R_{XY}^{-1}(C(X,Y))$, where $R_{XY}^{-1}(\cdot)$ denotes the inverse function of $R_{XY}(\Delta, \Delta)$, i.e, the distortion rate function.

A proof of Theorem 2 is given in Appendix B.

III. Examples

A. Doubly symmetric binary source (DSBS)

Consider a DSBS as in [3], [9]. That is, a binary source where $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ and for $x, y = 0, 1$,

$$Q(x, y) = \frac{1}{2}(1 - a_0)\delta_{x,y} + \frac{1}{2}a_0(1 - \delta_{x,y}), \tag{24}$$

$0 \leq a_0 \leq \frac{1}{2}$ and $\delta_{x,y}$ is an indicator function of $x = y$. X can be considered as an unbiased binary input to a binary symmetric channel (BSC) with crossover probability a_0 and Y as the corresponding output, or vice versa.

It is shown in [3] that for the DSBS

$$C(X;Y) = 1 + h(a_0) - 2h(a_1), \tag{25}$$

where $h(a_0)$ is the binary entropy function for $0 \leq a_0 \leq 1$ and $a_1 = \frac{1}{2} - \frac{1}{2}(1 - 2a_0)^{\frac{1}{2}}$.

For a DSBS with Hamming distortion $d_1 = d_2 = d_H$ and symmetric distortion constraint $\Delta_1 = \Delta_2 = \Delta$, the joint rate distortion function [12] is given by (26), where $L(x) = -x \log x$. It can be seen that

$$R_{XY}(a_1, a_1) = 1 + h(a_0) - 2h(a_1) = C(X, Y).$$

Therefore, by Theorem 2 we have $\gamma = a_1$. $C_3(\Delta, \Delta) = 1 + h(a_0) - 2h(a_1)$ for any $0 \leq \Delta \leq a_1$.

Remark: $C_3(\Delta, \Delta)$ for any $0 \leq \Delta \leq a_1$ is achieved by $R_0 = R_{XY}(a_1, a_1) = 1 + h(a_0) - 2h(a_1)$, $R_1 = R_{X|\tilde{X}\tilde{Y}}(\Delta)$, and $R_2 = R_{Y|\tilde{X}\tilde{Y}}(\Delta)$, where $(\tilde{X}, \tilde{Y})$ are the random variables achieving $R_{XY}(a_1, a_1)$. The test channels are

$$\begin{aligned} Pr\{X = x|\tilde{x}\tilde{y}\} &= (1 - a_1)\delta_{x,\tilde{x}} + a_1(1 - \delta_{x,\tilde{x}}), & (27) \\ Pr\{Y = y|\tilde{x}\tilde{y}\} &= (1 - a_1)\delta_{y,\tilde{y}} + a_1(1 - \delta_{y,\tilde{y}}). & (28) \end{aligned}$$

Hence, $R_1 = R_2 = h(a_1) - h(\Delta)$.

B. Gaussian source

In this section we consider the case when X, Y are bivariate Gaussian with zero mean and covariance matrix

$$K = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \tag{29}$$

Proposition 1: For the Gaussian random variable (X, Y) described above, the common information is

$$C(X;Y) = \frac{1}{2} \log \frac{1 + \rho}{1 - \rho}. \tag{30}$$

The proof is given in Appendix C. Proposition 1 can be extended to multivariate Gaussian distributions.

Corollary 1: For N joint Gaussian random variables $X_1, X_2, \cdots, X_N$ with covariance matrix

$$K_N = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \cdot & \cdot & \cdots & \cdot \\ \rho & \rho & \cdots & 1 \end{bmatrix}, \tag{31}$$

the common information is

$$C(X_1, X_2, \cdots, X_N) = \frac{1}{2} \log(1 + \frac{N\rho}{1 - \rho}). \tag{32}$$

Proposition 2: For bivariate Gaussian random variables X, Y with zero mean and covariance matrix in (29) and squared error distortion $d_1(u, \hat{u}) = d_2(u, \hat{u}) = (u - \hat{u})^2$, we have

$$C_3(\Delta, \Delta) = C(X;Y). \tag{33}$$

for any $\Delta \leq 1 - \rho$.

Proof: The joint rate distortion function for Gaussian random variables with symmetric squared error distortion [12] is

$$R_{XY}(\beta, \beta) = \begin{cases} \frac{1}{2} \log \frac{1 - \rho^2}{\beta^2} & 0 \leq \beta \leq 1 - \rho \\ \frac{1}{2} \log \frac{1 + \rho}{2\beta - (1 - \rho)} & 1 - \rho \leq \beta \leq 1 \\ 0 & \beta \geq 1 \end{cases}. \tag{34}$$

Thus we have

$$R_{XY}(1 - \rho, 1 - \rho) = \frac{1}{2} \log \frac{1 + \rho}{1 - \rho} = C(X, Y). \tag{35}$$

By Theorem 2, $\gamma = 1 - \rho$. This means that $C_3(\Delta, \Delta) = C(X;Y)$ for any $\Delta \leq 1 - \rho$.

Remark: $C_3(\Delta, \Delta)$ for any $0 \leq \Delta \leq 1 - \rho$ is achieved by $R_0 = R_{XY}(1 - \rho, 1 - \rho)$, $R_1 = R_{X|\tilde{X}\tilde{Y}}(\Delta) = \frac{1}{2} \log \frac{1 - \rho}{\Delta}$, $R_2 = R_{Y|\tilde{X}\tilde{Y}}(\Delta) = \frac{1}{2} \log \frac{1 - \rho}{\Delta}$, where $(\tilde{X}, \tilde{Y})$ are the random variables achieving $R_{XY}(1 - \rho, 1 - \rho)$.

IV. Conclusion

In this paper, we generalized Wyner's common information to that of continuous random variables and provided a lossy source coding interpretation using the Gray-Wyner network. A surprising observation is that the the minimum common rate for lossy source coding is invariant to the distortion constraint as long as it is less than a certain threshold.

$$R_{XY}(\beta,\beta)=\begin{cases} 1+h(a_0)-2h(\beta) & \text{if } 0\le\beta\le a_1 \\ L(1-a_0)-\frac{1}{2}\{L(2\beta-a_0)+L[2(1-\beta)-a_0]\} & a_1\le\beta\le\frac{1}{2} \end{cases} \quad (26)$$

APPENDIX

A. Proof of Theorem 1

We first introduce the following two lemmas. The first one is given by Gray [8].

Lemma 1: Given a two-dimensional source X, Y and a compound distortion measure, we have the following inequalities

$$\begin{aligned} R_{XY}(\Delta_1,\Delta_2) &\geq R_{X|Y}(\Delta_1)+R_Y(\Delta_2), & (36)\\ R_{X|Y}(\Delta_1) &\geq R_X(\Delta_1)-I(X;Y), & (37)\end{aligned}$$

and equalities hold in some neighborhood of the origin $\{(\Delta_1,\Delta_2): 0\le\Delta_1,\Delta_2\le\gamma\}$, provided that

$$Q(x,y)>0 \ \text{ all } x\in\mathcal{X}, y\in\mathcal{Y}, \quad (38)$$

and d_1, d_2 satisfy

$$\begin{aligned} d_1(x,\hat{x}) &> d_1(x,x)=0, x\neq\hat{x}, & (39)\\ d_2(y,\hat{y}) &> d_2(y,y)=0, y\neq\hat{y}. & (40)\end{aligned}$$

Here $R_{X|Y}(\Delta)$ is the conditional rate distortion function which is defined as

$$R_{X|Y}(\Delta)=\min I(X;\hat{X}|Y), \quad (41)$$

where the minimum is taken with respect to all test channels $q_t(\hat{x}|x,y)$ such that $Ed(X,\hat{X})\le\Delta$.

The second lemma is given by Gray and Wyner [9].

Lemma 2: For the lossy source coding problem described in the previous section, for $\Delta_1,\Delta_2\ge 0$, the rate distortion region is given by

$$\begin{aligned}\mathcal{R}(\Delta_1,\Delta_2)=\{&(R_0,R_1,R_2): R_0\ge I(X,Y;W),\\ &R_1\ge R_{X|W}(\Delta_1), R_2\ge R_{Y|W}(\Delta_2)\}, \quad (42)\end{aligned}$$

for some distributions $p(w|x,y)Q(x,y)$.

Note that Lemma 2 is valid for both the discrete case and the continuous case. Although Gray and Wyner only treated the discrete case in [9], the result can be generalized to the continuous case [11].

We now prove Theorem 1.

1) Achievability: For a given $Q(x,y)>0$ $x\in\mathcal{X}, y\in\mathcal{Y}$, let $C(X;Y)=I(XY;W)$ where (X,Y,W) satisfies $X-W-Y$ and $\sum_w p(x,y,w)=Q(x,y)$, i.e., W is the auxiliary variable that achieves $C(X,Y)$. Let (Δ_1,Δ_2) be in the range $\{0\le\Delta_1,\Delta_2\le\gamma\}$ where γ is chosen such that the following equalities hold

$$\begin{aligned} R_X(\Delta_1) &= R_{X|W}(\Delta_1)+I(X;W), & (43)\\ R_Y(\Delta_2) &= R_{Y|W}(\Delta_2)+I(Y;W), & (44)\\ R_{XY}(\Delta_1,\Delta_2) &= R_X(\Delta_1)+R_Y(\Delta_2)-I(X;Y). & (45)\end{aligned}$$

We now prove that $C_3(\Delta_1,\Delta_2)\le C(X,Y)$ in the range $\{0\le\Delta_1,\Delta_2\le\gamma\}$. For any $R_0>C(X,Y)$ and $\epsilon>0$ let

$$\epsilon_1=\min(\epsilon/3, R_0-C(X;Y)). \quad (46)$$

Since $\epsilon_1>0$, we know from Lemma 2 that there exists a code $(n,M_0,M_1,M_2,\Delta_X,\Delta_Y)$ with $\Delta_X\le\Delta_1+\epsilon_1$, $\Delta_Y\le\Delta_2+\epsilon_1$ and

$$\begin{aligned} \frac{1}{n}\log M_0 &\le I(X,Y;W)+\epsilon_1 \\ &= C(X;Y)+\epsilon_1\le R_0, & (47)\\ \frac{1}{n}\log M_1 &\le R_{X|W}(\Delta_1)+\epsilon_1, & (48)\\ \frac{1}{n}\log M_2 &\le R_{Y|W}(\Delta_2)+\epsilon_1. & (49)\end{aligned}$$

From (47-49), we have that

$$\begin{aligned} &\frac{1}{n}\sum_{i=0}^{2}\log M_i \\ &\le I(X,Y;W)+R_{X|W}(\Delta_1)+R_{Y|W}(\Delta_2)+3\epsilon_1, & (50)\\ &= I(X;W)+R_{X|W}(\Delta_1)+I(Y;W)+R_{Y|W}(\Delta_2)\\ &\quad -I(X;Y)+3\epsilon_1, & (51)\\ &= R_X(\Delta_1)+R_Y(\Delta_2)-I(X;Y)+3\epsilon_1, & (52)\\ &\le R_{XY}(\Delta_1,\Delta_2)+\epsilon. & (53)\end{aligned}$$

where (51) follows from the chain rule and the Markov Chain $X-W-Y$, (52) and (53) follow from (43-46).

This proves that the code satisfies (16)-(18), i.e., R_0 is (Δ_1,Δ_2)-achievable. This completes the proof of $C_3(\Delta_1,\Delta_2)\le C(X;Y)$.

2) Converse: Let Δ_1,Δ_2 be in the region $\{0\le\Delta_1,\Delta_2\le\gamma\}$ such that

$$R_X(\Delta_1)+R_Y(\Delta_2)-I(X;Y)=R_{XY}(\Delta_1,\Delta_2). \quad (54)$$

Let R_0 be (Δ_1,Δ_2)-achievable. We will show that $R_0\ge C(X;Y)$. The proof follows similar procedures as the proof of Theorem 5.1 in [3].

Since R_0 is (Δ_1,Δ_2)-achievable, there exists an $(n,M_0,M_1,M_2,\Delta_X,\Delta_Y)$ code satisfying (16)-(18). Let $f_E(X^n,Y^n)=(W_0,W_1,W_2)$, we have that

$$\begin{aligned} R_0 &\ge \frac{1}{n}\log M_0\ge\frac{1}{n}H(W_0), & (55)\\ &\ge \frac{1}{n}I(X^n,Y^n;W_0), & (56)\\ &= \frac{1}{n}H(X^n,Y^n)-\frac{1}{n}H(X^n,Y^n|W_0), & (57)\\ &= H(X,Y)-\frac{1}{n}\sum_{k=1}^{n}H(X_k,Y_k|X^{k-1},Y^{k-1},W_0) & (58)\\ &\ge H(X,Y)-\frac{1}{n}\sum_{k=1}^{n}\Gamma_1(\delta^{(k)}), & (59)\end{aligned}$$

$$\geq H(X,Y)-\Gamma_1(\frac{1}{n}\sum_{k=1}^{n}\delta^{(k)}), \tag{60}$$

where (59) comes from the definition of $\Gamma_1(\cdot)$ (c.f. Corollary 4.5, [3]) and the definition of $\delta^{(k)}$, where

$$\delta^{(k)}=I(X_k;Y_k|X^{k-1},Y^{k-1},W_0).$$

Inequality (60) follows from the concavity of $\Gamma_1(\delta)$.

Therefore, since $C(X;Y)=H(X,Y)-\Gamma_1(0)$ (c.f. equation (4.4) in [3]) and (60), to establish $R_0\geq C(X;Y)$ we only need to prove that, for arbitrary $\epsilon>0$,

$$\frac{1}{n}\sum_{k=1}^{n}\delta^{(k)}\leq v(\epsilon), \tag{61}$$

$$\lim_{\epsilon\to 0}v(\epsilon)=0. \tag{62}$$

From (57), we have that

$$\begin{aligned}\frac{1}{n}\log M_0 &\geq \frac{1}{n}H(X^n,Y^n)-\frac{1}{n}H(X^n,Y^n|W_0), &(63)\\ &= \frac{1}{n}H(X^n,Y^n)+\frac{1}{n}I(X^n;Y^n|W_0)\\ &\quad -\frac{1}{n}H(X^n|W_0)-\frac{1}{n}H(Y^n|W_0). &(64)\end{aligned}$$

Consider again the $(n,M_0,M_1,M_2,\Delta_X,\Delta_Y)$ code that satisfies (16)-(18) for arbitrary $\epsilon>0$. Set $\hat{X}^n=f_D^{(X)}(W_0,W_1)$ and $\hat{Y}^n=f_D^{(Y)}(W_0,W_2)$, we have

$$\begin{aligned}\frac{1}{n}\log M_1 &\geq \frac{1}{n}H(W_1), &(65)\\ &\geq \frac{1}{n}H(W_1|W_0), &(66)\\ &\geq \frac{1}{n}I(X^n;W_1|W_0), &(67)\\ &\geq \frac{1}{n}I(X^n;\hat{X}^n|W_0). &(68)\end{aligned}$$

where inequality (68) follows from the Markov chain $X^n-W_0,W_1-\hat{X}^n$. Similarly, we have

$$\frac{1}{n}\log M_2\geq\frac{1}{n}I(Y^n;\hat{Y}^n|W_0). \tag{69}$$

Adding (64), (68) and (69), we obtain

$$\begin{aligned}&\sum_{i=0}^{2}\frac{1}{n}\log M_i\\ \geq\ &\frac{1}{n}(H(X^n,Y^n)+I(X^n;Y^n|W_0)-H(X^n|W_0)\\ &-H(Y^n|W_0)+I(X^n;\hat{X}^n|W_0)+I(Y^n;\hat{Y}^n|W_0)) &(70)\\ =\ &\frac{1}{n}(I(X^n;W_0)+I(Y^n;W_0)-I(X^n;Y^n)\\ &+I(X^n;Y^n|W_0)+I(X^n;\hat{X}^n|W_0)+I(Y^n;\hat{Y}^n|W_0)), &(71)\\ =\ &\frac{1}{n}(I(X^n;\hat{X}^n,W_0)+I(Y^n;\hat{Y}^n,W_0)-I(X^n;Y^n)\\ &+I(X^n;Y^n|W_0)), &(72)\\ \geq\ &\frac{1}{n}(I(X^n;\hat{X}^n)+I(Y^n;\hat{Y}^n)-nI(X;Y)\\ &+I(X^n;Y^n|W_0)), &(73)\\ \geq\ &R_X(\Delta_1)+R_Y(\Delta_2)-I(X;Y)+\frac{1}{n}I(X^n;Y^n|W_0), &(74)\\ =\ &R_{XY}(\Delta_1,\Delta_2)+\frac{1}{n}I(X^n;Y^n|W_0). &(75)\end{aligned}$$

where (71), (72) follow from the chain rule, (73) follows from the fact that conditioning does not increase entropy, (74) follows from the definition of rate distortion function and (75) is from (54).

On the other hand, the code satisfies (17), so we have

$$\sum_{i=0}^{2}\frac{1}{n}\log M_i\leq R_{XY}(\Delta_1,\Delta_2)+\epsilon. \tag{76}$$

Combining (75) and (76) we will have that

$$\frac{1}{n}I(X^n;Y^n|W_0)\leq\epsilon. \tag{77}$$

Also, it is easy to check that the following inequality is true.

$$\frac{1}{n}I(X^n;Y^n|W_0)\geq\frac{1}{n}\sum_{k=1}^{n}\delta^{(k)}. \tag{78}$$

Combining (77) and (78), we obtain

$$\frac{1}{n}\sum_{k=1}^{n}\delta^{(k)}\leq\epsilon, \tag{79}$$

which completes the proof.

B. *Proof of Theorem 2*

Before proving Theorem 2, we first introduce two lemmas.

Lemma 3: For any Δ_1,Δ_2,

$$C_3(\Delta_1,\Delta_2)\leq R_{XY}(\Delta_1,\Delta_2). \tag{80}$$

Proof: The lemma follows from the fact that $R_0=R_{XY}(\Delta_1,\Delta_2)$ is (Δ_1,Δ_2)-achievable.

Lemma 4: Let $\tau=R_{XY}^{-1}(C(X,Y))$, $\Delta\leq\tau$, if R_0 is Δ-achievable, then there exists a W such that $X-W-Y$, $R_0\geq I(X,Y;W)$, and

$$I(X,Y;W)+R_{X|W}(\Delta)+R_{Y|W}(\Delta)=R_{XY}(\Delta,\Delta). \tag{81}$$

Proof: For $\Delta\leq\tau$, if R_0 is Δ-achievable, we have that for any $\epsilon>0$, there exists a code $(n,M_0,M_1,M_2,\Delta,\Delta)$ that satisfies (16)-(18). Let $R_i'=\frac{1}{n}\log M_i$ for $i=0,1,2$, we have that

$$\sum_{i=0}^{2}R_i'\leq R_{XY}(\Delta,\Delta)+\epsilon. \tag{82}$$

From the definition of rate distortion region [9], we know that $(R_0',R_1'-\epsilon/2,R_2'-\epsilon/2)$ is in the rate distortion region $\mathcal{R}$. By Lemma 2, there exists a W jointly distributed with X,Y as $p(w|x,y)Q(x,y)$ and satisfies

$$\begin{aligned}R_0'+R_1'+R_2'-\epsilon &\geq I(X,Y;W)+R_{X|W}(\Delta)+R_{Y|W}(\Delta), &(83)\\ &\geq I(X,Y;W)+R_{XY|W}(\Delta,\Delta), &(84)\\ &\geq R_{XY}(\Delta,\Delta), &(85)\end{aligned}$$

where inequalities (84) and (85) are from Them 3.1 in [8]. The equality in (84) holds only when X is conditionally independent of Y given W and equality in (85) holds only when $0 \le \Delta \le \gamma$. For $\Delta = \tau$, combined with (82), we have that $I(X,Y;W) = R_{XY}(\tau,\tau)$. Hence for any $\Delta \le \tau$,

$$I(X,Y;W) + R_{X|W}(\Delta) + R_{Y|W}(\Delta) = R_{XY}(\Delta,\Delta). \quad (86)$$

This completes the proof.

We now prove Theorem 2.

First we show that for any Δ such that $C_3(\Delta) = C(X,Y)$, we have $\Delta \le \tau$. From Lemma 3, $R_{XY}(\Delta,\Delta) \ge C(X,Y)$. $R_{XY}(\Delta,\Delta)$ is a non increasing function of Δ, therefore, $\Delta \le R_{XY}^{-1}(C(X,Y)) = \tau$.

Next we will show that for any distortion $\Delta \le \tau$, $C_3(\Delta) = C(X,Y)$.

For any R_0 that is Δ-achievable, from Lemma 4, there exists a W such that $X - W - Y$ and $R_0 \ge I(X,Y;W)$. Hence, $R_0 \ge I(X,Y;W) \ge C(X,Y)$, which implies $C_3(\Delta) \ge C(X,Y)$.

From Lemma 3, $C_3(\tau) \le C(X,Y)$. Hence, $C_3(\tau) = C(X,Y)$. Thus, any rate $R_0 > C(X,Y)$ is τ-achievable. By Lemma 4, we have

$$C(X,Y) + R_{X|W}(\tau) + R_{Y|W}(\tau) = R_{XY}(\tau,\tau),$$

where W is the random variable such that $C(X,Y) = I(X,Y;W)$. Thus, by Lemma 1, for any $\Delta \le \tau$,

$$C(X,Y) + R_{X|W}(\Delta) + R_{Y|W}(\Delta) = R_{XY}(\Delta,\Delta).$$

Then use the same proof as the achievability part of Theorem 1, we can prove that when the distortion $\Delta \le \tau$, any rate $R_0 > C(X,Y)$ is Δ-achievable. Hence, $C_3(\Delta) \le C(X,Y)$, completing the proof.

C. Proof of Proposition 1

Let W, N_1 and N_2 be standard Gaussian random variables independent of each other and express X, Y as

$$X = \sqrt{\rho}W + \sqrt{1-\rho}N_1, \quad (87)$$
$$Y = \sqrt{\rho}W + \sqrt{1-\rho}N_2. \quad (88)$$

It is easy to verify that conditions (C1) and (C2) are satisfied. Straightforward calculation yields $I(X,Y;W) = \frac{1}{2}\log\frac{1+\rho}{1-\rho}$.

The proof is thus complete if one can prove $I(X,Y;W) > \frac{1}{2}\log\frac{1+\rho}{1-\rho}$ for all W satisfying the conditions (C1) and (C2).

Let $P_{X,W,Y}$ be any joint distribution satisfying the conditions (C1) and (C2) and let K denote the corresponding covariance matrix. Let $\tilde{P}_{X,W,Y}$ be joint Gaussian satisfying the conditions (C1) and (C2) with zero mean and the same covariance matrix K. From the fact that conditional differential entropy is maximized under Gaussian distribution for a given covariance matrix [13], we have

$$h(X,Y|W) \le h_{\tilde{P}}(X,Y|W). \quad (89)$$

Therefore $I(X,Y;W) \ge I_{\tilde{P}}(X,Y;W)$. Hence we only need to consider (X,W,Y) that are jointly Gaussian distributed.

Without loss of generality, let W be a Gaussian random variable with zero mean and variance σ^2, and

$$X = \rho_1 W + \sqrt{1-\rho_1^2\sigma^2}N_1, \quad (90)$$
$$Y = \rho_2 W + \sqrt{1-\rho_2^2\sigma^2}N_2, \quad (91)$$

where N_1 and N_2 are standard Gaussian random variables and W, N_1, N_2 are mutually independent with each other.

Since $EXY = \rho$, we have

$$\rho = \rho_1\rho_2\sigma^2, \quad (92)$$

and due to the Markov chain $X - W - Y$, we have $H(X|W) = H(X|W,Y)$, i.e.,

$$1 - \rho_1^2 = \frac{1 + 2\rho\rho_1\rho_2 - \rho^2 - \rho_1^2 - \rho_2^2}{1-\rho_2^2}. \quad (93)$$

Combining (92) and (93), we get $\sigma^2 = 1$. Therefore, we can lower bound $I(X,Y;W)$ by

$$\begin{aligned} I(X,Y;W) &= h(X,Y) - h(X|W) - h(Y|W), && (94)\\ &= \frac{1}{2}\log\frac{1-\rho^2}{(1-\rho_1^2)(1-\rho_2^2)}, && (95)\\ &= \frac{1}{2}\log\frac{1-\rho^2}{1+\rho^2-\rho_1^2-\rho_2^2}, && (96)\\ &\ge \frac{1}{2}\log\frac{1-\rho^2}{1+\rho^2-2\rho}, && (97)\\ &= \frac{1}{2}\log\frac{1+\rho}{1-\rho}, && (98)\end{aligned}$$

where we use the facts that $\rho_1\rho_2 = \rho$ and $\rho_1^2 + \rho_2^2 \ge 2\rho_1\rho_2$.

References

[1] P. Gács and J. Körner, "Common information is much less than mutual information," *Problems Contr. Inform. Theory*, vol. 2, pp. 149-162, 1973.

[2] R. Ahlswede and J. Körner, "On common information and related characteristics of correlated information sources," *Proc. of the 7th Prague Conference of Information Theory*, 1974.

[3] A. D. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 163-179, March, 1975.

[4] H. S. Witsenhausen, "On sequences of pairs of dependent random varibles," *SIAM J. Appl. Math.*, vol. 28, pp. 100-113, Jan. 1975.

[5] H. S. Witsenhausen, "Values and bounds for the common information of two discrete random variables," *SIAM J. Appl. Math.*, vol. 31, no. 2, pp. 313-333, 1976.

[6] P. Cuff, H.Permuter and T. M. Cover, "Coordination capacity," *IEEE Trans. Inf. Theory*, vol.56, pp. 4181-4206, Sep. 2010.

[7] W. Liu, G. Xu and B. Chen, "The common information of N dependent random variables," *Proc. Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sep. 2010.

[8] R. M. Gray, "A new class of lower bounds to information rates of stationary sources via conditional rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. 19, pp.480-489, Jul. 1973.

[9] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *Bell Syst. Tech. J.*, vol. 58, pp. 1681-1721, Nov. 1974.

[10] R. W. Yeung, *Information Theory and Network Coding*, Springer, 2008.

[11] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder-II: General sources," *Information and control*, vol. 38, pp. 60-80, 1978.

[12] T. Berger, *Rate Distortion Theory*, Englewood Cliffs, N.J.: Prentice-Hall, 1971.

[13] J. A. Thomas, "Feedback can at most double Gaussian multiple access channel capacity." *IEEE Trans. Inf. Theory*,vol. IT-33, no. 5, pp.711-716, Sep. 1987.