2012 IEEE International Symposium on Information Theory Proceedings

Lossy Common Information of Two Dependent Random Variables

Kumar Viswanatha, Emrah Akyol and Kenneth Rose
ECE Department, University of California - Santa Barbara
{kumar,eakyol,rose}@ece.ucsb.edu

***Abstract*—The two most prevalent notions of common information are due to Wyner and Gács-Körner and both the notions can be stated as two different characteristic points in the *lossless* Gray-Wyner region. Although these quantities can be easily evaluated for random variables with infinite entropy (eg. continuous random variables), the operational significance underlying their definition is applicable only to the lossless framework. The primary objective of this paper is to generalize these two notions of common information to the lossy Gray-Wyner network, which extends the theoretical intuition underlying their definitions for general sources and distortion measures. We begin with the lossy generalization of Wyner's common information, defined as the minimum rate on the shared branch of the Gray-Wyner network at minimum sum rate when the two decoders reconstruct the sources subject to individual distortion constraints. We derive a complete single letter information theoretic characterization for this quantity and use it to compute the common information of symmetric bivariate Gaussian random variables. We then derive similar results to generalize Gács-Körner's definition to the lossy framework. These two characterizations allow us to carry the practical insight underlying the two notions of common information to general sources and distortion measures.**

***Index Terms*—Wyner's common information, Gács and Körner's common information, Lossy Gray-Wyner network**

I. Introduction

The quest for a meaningful and useful notion of common information (CI) of two random variables (denoted by X and Y) has been actively pursued by researchers in information theory for over three decades. An early seminal approach to quantify CI is due to Gács and Körner [1] (denoted here by $C_{GK}(X,Y)$), who defined it as the maximum amount of information relevant to both random variables, one can extract from the knowledge of either one of them. Their result was of considerable theoretical interest, but also fundamentally negative in nature. They showed that $C_{GK}(X,Y)$ is typically much smaller than the mutual information and only depends on the zeros of the joint distribution. Unsatisfied with the negative implication, Wyner proposed an alternative idea of CI [2] (denoted here by $C_W(X,Y)$) inspired by earlier work in multi-terminal source coding [3]. He defined the CI as the minimum rate on the shared branch of the lossless Gray-Wyner network (described in section I-A and Fig. 1), when the sum rate is constrained to be the joint entropy. His conception was to look at the minimum amount of shared information that must be sent to both decoders, while restricting the overall transmission rate to the minimum, $H(X,Y)$. He obtained the single letter characterization for $C_W(X,Y)$ as,

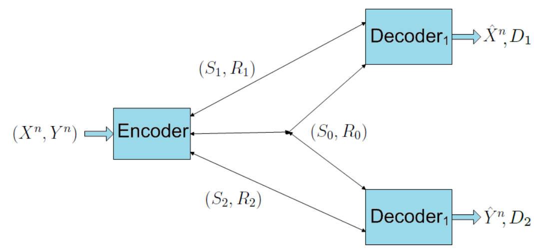

Figure 1. The Gray-Wyner network

$$C_W(X,Y) = \inf I(X,Y;U) \tag{1}$$

where the infimum is over all random variables, U, such that $X \leftrightarrow U \leftrightarrow Y$ form a Markov chain in that order. We note that although $C_{GK}(X,Y)$ and $C_W(X,Y)$ were defined from theoretical standpoints, they play critical roles in understanding the performance limits in several practical networking and database applications, see for eg. [4]. We further note in passing that several other definitions of CI, with applications in different fields, have appeared in the literature [6], [7], but are less relevant to us here.

Although the quantity in (1) can be evaluated for random variables with infinite entropy (eg. continuous random variables), for such random variables it lacks the underlying theoretical interpretation of Wyner's notion of CI as one of the distinctive operating points in the Gray-Wyner region and thereby lacks the fundamental intuition behind the definition. This largely compromises its practical significance and calls for a useful generalization which can be easily extended to infinite entropy distributions. Our primary step is to characterize a lossy coding extension of Wyner's CI (denoted by $C_W(X,Y;D_1,D_2)$), defined as the minimum rate on the shared branch of the Gray-Wyner network at minimum sum rate when the sources are decoded at respective distortions of D_1 and D_2. Note that the minimum sum rate at distortions D_1 and D_2 is given by the Shannon's rate distortion function, hereafter denoted by $R_{X,Y}(D_1,D_2)$. In this paper, our main objective is to derive an information theoretic characterization for $C_W(X,Y;D_1,D_2)$ for general sources and distortion measures. We note that although there is no prior work on characterizing $C_W(X,Y;D_1,D_2)$, in a recent work [8], Xu et

The work was supported by the NSF under grants CCF-0728986, CCF - 1016861 and CCF-1118075

978-1-4673-2579-0/12/$31.00 ©2012 IEEE

al. defined the asymptotic quantity $C_W(X,Y;D_1,D_2)$[1] and showed that there exists a region of distortions around the origin where $C_W(X,Y;D_1,D_2)$ is equal to the Wyner's single letter characterization in (1).

We note that much of the focus in the beginning of this paper in section II will be towards characterizing Wyner's definition of CI in the lossy Gray-Wyner setting. However, similar results and proof techniques are used in section III to extend Gács and Körner's definition to the lossy framework. We further note that there have been other physical interpretations of both the notions of common information besides the Gray-Wyner network, see for example [1], [2]. We are not yet sure of the relations between such interpretations and the lossy generalizations we consider in this paper. We will be investigating these connections are part of our future work.

A. Gray-Wyner Network [3]

Let (X,Y) be any two dependent random variables taking values in the alphabets $\mathcal{X}$ and $\mathcal{Y}$ respectively. Let $\hat{\mathcal{X}}$ and $\hat{\mathcal{Y}}$ be two given reconstruction alphabets for random variables X and Y respectively. We denote the set $\{1,2\ldots M\}$ by I_M for any positive integer M, n iid samples of a random variable by X^n and the corresponding alphabet by $\mathcal{X}^n$. In what follows, for any pair of random variables X and Y, $R_X(\cdot)$, $R_Y(\cdot)$ and $R_{X,Y}(\cdot,\cdot)$ denote the respective rate distortion functions.

A rate-distortion tuple (R_0,R_1,R_2,D_1,D_2) is said to be achievable for the Gray-Wyner network if for all $\epsilon>0$, there exists encoder mappings $f_E:\mathcal{X}^n\times\mathcal{Y}^n\to I_{M_0}\times I_{M_1}\times I_{M_2}$ and decoder mappings $f_D^{(X)}:I_{M_0}\times I_{M_1}\to\hat{\mathcal{X}}^n$, $f_D^{(Y)}:I_{M_0}\times I_{M_2}\to\hat{\mathcal{Y}}^n$ such that the following hold:

$$M_i \leq 2^{nR_i}+\epsilon \;\; i\in\{0,1,2\}$$
$$\Delta_X\leq D_1+\epsilon \qquad \Delta_Y\leq D_2+\epsilon \tag{2}$$

where $\hat{X}^n=f_D^{(X)}(S_0,S_1)$, $\hat{Y}^n=f_D^{(Y)}(S_0,S_2)$ and $\Delta_X=\frac{1}{n}\sum_{i=1}^{n}d_X(X_i,\hat{X}_i)$, $\Delta_Y=\frac{1}{n}\sum_{i=1}^{n}d_Y(Y_i,\hat{Y}_i)$ for some well defined single letter distortion measures $d_X(\cdot,\cdot)$ and $d_Y(\cdot,\cdot)$. The convex closure over all such achievable rate-distortion tuples is called the achievable region for the Gray-Wyner network. The set of all achievable rate tuples for any given distortion D_1 and D_2 is denoted here by $\mathcal{R}_{GW}(D_1,D_2)$.

Gray and Wyner [3] gave the following complete characterization for $\mathcal{R}_{GW}(D_1,D_2)$. Let $(U,\hat{X},\hat{Y})$ be any random variables jointly distributed with (X,Y) and taking values over alphabets $\mathcal{U},\hat{\mathcal{X}}$ and $\hat{\mathcal{Y}}$ respectively, for some arbitrary $\mathcal{U}$. Let the joint density be $P(X,Y,U,\hat{X},\hat{Y})$. All rate-distortion tuples (R_0,R_1,R_2,D_1,D_2) satisfying the following conditions are achievable:

$$R_0\geq I(X,Y;U)$$
$$R_1\geq I(X;\hat{X}|U),\;\; R_2\geq I(Y;\hat{Y}|U) \tag{3}$$

if $E(d_X(X,\hat{X}))\leq D_1$ and $E(d_Y(Y,\hat{Y}))\leq D_2$. The closure of the achievable rate distortion tuples over all such joint densities is the complete rate distortion region for the Gray-Wyner network.

[1]They denote $C_W(X,Y;D_1,D_2)$ by $C_3(D_1,D_2)$

II. Wyner's Lossy Common Information

A. Definition

We next define Wyner's CI generalized to the lossy framework denoted by $C_W(X,Y;D_1,D_2)$. It is defined as the infimum over all shared rates R_0, such that $(R_0,R_1,R_2)\in\{\mathcal{R}_{GW}(D_1,D_2)\cap\text{Pangloss plane}\}$, where for any distortion pair (D_1,D_2) the plane $R_0+R_1+R_2=R_{X,Y}(D_1,D_2)$ is defined as the Pangloss plane. We note that the above operational definition of $C_W(X,Y;D_1,D_2)$, has already appeared recently in [8]. However, a complete single letter information theoretic characterization of $C_W(X,Y;D_1,D_2)$ has never been considered in any earlier work. The primary objective of section II-B is to characterize $C_W(X,Y;D_1,D_2)$ for general sources and distortion measures. It is important to note that Wyner gave the complete single letter characterization of $C_W(X,Y;0,0)$ when X and Y have finite joint entropy. His result is stated formally as:

$$C_W(X,Y)=C_W(X,Y;0,0)=\inf I(X,Y;U) \tag{4}$$

where the infimum is over all U satisfying $X\leftrightarrow U\leftrightarrow Y$.

B. Single Letter Characterization of $C_W(X,Y;D_1,D_2)$

In the following theorem, we characterize $C_W(X,Y;D_1,D_2)$. We denote the set of all channels which achieve $R_{X,Y}(D_1,D_2)$ by $\mathcal{P}^{X,Y}_{D_1,D_2}$, i.e.,

$$\inf I(X,Y;\hat{X},\hat{Y}) = I(X,Y;X^*,Y^*) \tag{5}$$

$\forall P(X^*,Y^*|X,Y)\in\mathcal{P}^{X,Y}_{D_1,D_2}$, where the infimum is over channels such that $E(d_X(X,\hat{X}))\leq D_1, E(d_X(X,\hat{X}))\leq D_2$. Hereafter we assume that, for every distortion pair (D_1,D_2), there exists at least one channel $P(X^*,Y^*|X,Y)$ such that $I(X,Y;X^*,Y^*)=R_{X,Y}(D_1,D_2)$. We note that, our results can be easily extended to all other 'well behaved' continuous joint densities using standard techniques from probability measures.

Theorem 1. *A single letter characterization of* $C_W(X,Y;D_1,D_2)$ *is given by:*

$$C_W(X,Y;D_1,D_2)=\inf I(X,Y;U) \tag{6}$$

where the infimum is over all joint densities (X,Y,X^*,Y^*,U) *such that the following Markov conditions hold:*

$$X^*\leftrightarrow U\leftrightarrow Y^* \tag{7}$$
$$(X,Y)\leftrightarrow (X^*,Y^*)\leftrightarrow U \tag{8}$$

and where $P(X^*,Y^*|X,Y)\in\mathcal{P}^{X,Y}_{D_1,D_2}$ *is any joint distribution which achieves the rate distortion function at* (D_1,D_2).

Remark 1. If we set $\hat{\mathcal{X}}=\mathcal{X}$, $\hat{\mathcal{Y}}=\mathcal{Y}$ and consider the Hamming distortion measure, at $(D_1,D_2)=(0,0)$, it is easy to show that Wyner's common information is obtained as a special case, i.e., $C_W(X,Y;0,0)=C_W(X,Y)$.

Proof: We note that, although there are arguably simpler methods to prove this theorem, we choose the following approach as it uses only the Gray-Wyner theorem without recourse to any supplementary results. We also assume that there exists a unique channel $P^*(X^*, Y^*|X, Y)$ which achieves $R_{X,Y}(D_1, D_2)$. The proof of the theorem when there are multiple channels in $\mathcal{P}^{X,Y}_{D_1,D_2}$ follows directly.

Our objective is to show that every point in the intersection of $\mathcal{R}_{GW}(D_1, D_2)$ and the Pangloss plane has $R_0 = I(X, Y; U)$ for some U jointly distributed with (X, Y, X^*, Y^*) and satisfying conditions (7) and (8). We first prove that every point in the intersection of the Pangloss plane and $\mathcal{R}_{GW}(D_1, D_2)$ is achieved by a joint density satisfying (7) and (8). Towards showing this, we begin with an alternate characterization of $\mathcal{R}_{GW}(D_1, D_2)$ (which is also complete) due to Venkataramani et.al in [10][2]. Let $(U, \hat{X}, \hat{Y})$ be any random variables jointly distributed with (X, Y) such that $E(d_X(X, \hat{X})) \leq D_1$ and $E(d_Y(Y, \hat{Y})) \leq D_2$. Then any rate tuple (R_0, R_1, R_2) satisfying the following conditions belongs to $\mathcal{R}_{GW}(D_1, D_2)$:

$$\begin{aligned} R_0 &\geq I(X, Y; U) \\ R_1 + R_0 &\geq I(X, Y; U, \hat{X}) \\ R_2 + R_0 &\geq I(X, Y; U, \hat{Y}) \\ R_0 + R_1 + R_2 &\geq I(X, Y; U, \hat{X}, \hat{Y}) + I(\hat{X}; \hat{Y}|U) \quad (9) \end{aligned}$$

It is very easy to show that the above characterization is equivalent to (3). As the above characterization is complete, this implies that, if a rate-distortion tuple $(R_0, R_1, R_2, D_1, D_2)$ is achievable for the Gray-Wyner network, then we can always find random variables $(U, \hat{X}, \hat{Y})$ such that $E(d_X(X, \hat{X})) \leq D_1$, $E(d_Y(Y, \hat{Y})) \leq D_2$ and satisfying (9). We are further interested in characterizing the points in $\mathcal{R}_{GW}(D_1, D_2)$ which also lie on the Pangloss plane, i.e., $R_0 + R_1 + R_2 = R_{X,Y}(D_1, D_2)$. Therefore, for any rate tuple (R_0, R_1, R_2) on the Pangloss plane in $\mathcal{R}_{GW}(D_1, D_2)$, we have the following series of inequalities:

$$\begin{aligned} R_{X,Y}(D_1, D_2) &= R_0 + R_1 + R_2 \\ &\geq I(X, Y; U, \hat{X}, \hat{Y}) - H(\hat{X}, \hat{Y}|U) \\ &\quad + H(\hat{X}|U) + H(\hat{Y}|U) \\ &\geq I(X, Y; \hat{X}, \hat{Y}) - H(\hat{X}, \hat{Y}|U) \\ &\quad + H(\hat{X}|U) + H(\hat{Y}|U) \\ &\geq^{(a)} R_{X,Y}(D_1, D_2) - H(\hat{X}, \hat{Y}|U) \\ &\quad + H(\hat{X}|U) + H(\hat{Y}|U) \\ &\geq R_{X,Y}(D_1, D_2) \quad (10) \end{aligned}$$

where (a) follows because $(\hat{X}, \hat{Y})$ satisfy the distortion constraints. As the LHS and RHS of the above series of inequalities are the same, all the inequalities must be equalities and hence we have:

$$\begin{aligned} H(\hat{X}|U) + H(\hat{Y}|U) &= H(\hat{X}, \hat{Y}|U) \\ I(X, Y; U, \hat{X}, \hat{Y}) &= I(X, Y; \hat{X}, \hat{Y}) \\ I(X, Y; \hat{X}, \hat{Y}) &= R_{X,Y}(D_1, D_2) \quad (11) \end{aligned}$$

From our assumption, there is a unique channel, $P^*(X^*, Y^*|X, Y)$ which achieves $I(X, Y; \hat{X}, \hat{Y}) = R_{X,Y}(D_1, D_2)$. It therefore follows that every point in $\mathcal{R}_{GW}(D_1, D_2)$ that lies on the Pangloss plane satisfies (9) for some joint density satisfying (7) and (8).

What remains is to show that any joint density (X, Y, X^*, Y^*, U) satisfying (7) and (8) leads to a subregion of $\mathcal{R}_{GW}(D_1, D_2)$ which has at least one point on the Pangloss plane with $R_0 = I(X, Y; U)$. Formally, denote by $\mathcal{R}(U)$, the region (9) achieved by a joint density (X, Y, X^*, Y^*, U) satisfying (7) and (8). Then we have to show that $\exists (R_0, R_1, R_2) \in \mathcal{R}(U)$ such that:

$$\begin{aligned} R_0 + R_1 + R_2 &= R_{X,Y}(D_1, D_2) \\ R_0 &= I(X, Y; U) \quad (12) \end{aligned}$$

Consider the point, $(R_0, R_1, R_2) = (I(X, Y; U), I(X, Y; X^*|U), I(X, Y; Y^*|U, X^*))$ for any joint density (X, Y, X^*, Y^*, U) satisfying (7) and (8). Clearly the point satisfies the first two conditions in (9). Next, we note that:

$$\begin{aligned} R_0 + R_2 &= I(X, Y; U) + I(X, Y; Y^*|U, X^*) \\ &\geq^{(b)} I(X, Y; U) + I(X, Y; Y^*|U) \\ &= I(X, Y; Y^*, U) \quad (13) \end{aligned}$$

$$\begin{aligned} R_0 + R_1 + R_2 &=^{(c)} I(X, Y; X^*, Y^*, U) \quad (14) \\ &= I(X, Y; X^*, Y^*) = R_{X,Y}(D_1, D_2) \end{aligned}$$

where (b) and (c) follow from the fact that the joint density satisfies (7) and (8). Hence, we have shown the existence of one point in $\mathcal{R}(U)$ satisfying (12) for every joint density (X, Y, X^*, Y^*, U) satisfying (7) and (8) proving the theorem. ■

We note that, in general, it is hard to establish convexity/monotonicity of $C_W(X, Y; D_1, D_2)$ with respect to (D_1, D_2). This makes it hard to establish conclusive inequality relations between $C_W(X, Y; D_1, D_2)$ and $C_W(X, Y)$ for all distortions. However, in the following lemma, we establish sufficient conditions on (D_1, D_2) for $C_W(X, Y; D_1, D_2) \lesseqgtr C_W(X, Y)$.

Lemma 1. *For any pair of random variables (X, Y),*

- (i) $C_W(X, Y; D_1, D_2) \leq C_W(X, Y)$ at (D_1, D_2) if $\exists (\tilde{D}_1, \tilde{D}_2)$ such that $\tilde{D}_1 \leq D_1$, $\tilde{D}_2 \leq D_2$ and $R_{X,Y}(\tilde{D}_1, \tilde{D}_2) = C_W(X, Y)$
- (ii) $C_W(X, Y; D_1, D_2) \geq C_W(X, Y)$ if Shannon lower bound for $R_{XY}(D_1, D_2)$ is tight at (D_1, D_2)

Proof: The proof of (i) is rather straightforward and hence we choose to omit it. Towards proving (ii), it is easy to show using standard techniques [11], [9] that the conditional

[2] We note that the theorem can be proved even using the original Gray-Wyner's characterization. However, if we begin with that characterization, we would require the random variables to satisfy two more Markov conditions beyond (7) and (8). These Markov chains can in fact be shown to be redundant using Kuhn-Tucker conditions. We choose this alternate approach to avoid recourse to such supplementary results.

distribution $P(X^*, Y^*|X, Y)$ which achieves $R_{X,Y}(D_1, D_2)$ when Shannon lower bound is tight has independent backward channels, i.e.:

$$P_{X,Y|X^*,Y^*}(x, y|x^*, y^*) = Q_{X|X^*}(x|x^*)Q_{Y|Y^*}(y|y^*) \quad (15)$$

Let us consider any U which satisfies $(X^* \leftrightarrow U \leftrightarrow Y^*)$ and $(X, Y) \leftrightarrow (X^*, Y^*) \leftrightarrow U$. It is easy to verify that any such joint density also satisfies $X \leftrightarrow U \leftrightarrow Y$. As the infimum for $C_W(X, Y)$ is taken over a larger set of joint densities, we have $C_W(X, Y; D_1, D_2) \geq C_W(X, Y)$. ■

The above lemma highlights the anomalous behavior of $C_W(X, Y; D_1, D_2)$ with respect to the distortions. Determining the conditions for equality in Lemma 1.(ii) is an interesting problem in its own right. For the symmetric setting, i.e., $D_1 = D_2 = D$, it was shown in [8] using a completely different approach leveraging prior results from conditional rate distortion theory (see for eg. [9]) that $C_W(X, Y; D, D) = C_W(X, Y)$ iff $D \leq R^{-1}_{X,Y}(C_W(X, Y))$, where $R^{-1}_{X,Y}(\cdot)$ denotes the distortion-rate function. We will further explore the underlying connections between these results as part of our future work.

C. Bivariate Gaussian Example

Let X and Y be jointly Gaussian random variables with zero mean, unit variance and a correlation coefficient of ρ. We focus on the symmetric distortion scenario, i.e. $D_1 = D_2 = D$, under mean squared error distortion metric. The joint rate distortion function is given by [11]:

$$R_{X,Y}(D, D) = \begin{cases} \frac{1}{2}\log\left(\frac{1-\rho^2}{D^2}\right) & \text{if } 0 < D \leq 1-\rho \\ \frac{1}{2}\log\left(\frac{1+\rho}{2D-1+\rho}\right) & \text{if } 1-\rho \leq D \leq 1 \end{cases} \quad (16)$$

We first consider the range $0 < D \leq 1 - \rho$. The RD-optimal random encoder is such that $P(X|X^*)$ and $P(Y|Y^*)$ are two independent zero mean Gaussian channels with variance D. It is easy to verify that the optimal reproduction distribution (for (X^*, Y^*)) is jointly Gaussian with zero mean. The covariance matrix for (X^*, Y^*) is:

$$\Sigma_{X^*Y^*} = \begin{bmatrix} 1-D & \frac{\rho}{1-D} \\ \frac{\rho}{1-D} & 1-D \end{bmatrix} \quad (17)$$

At these distortions, the Shannon lower bound is tight and hence from Lemma 1, $C_W(X, Y; D, D) \geq C_W(X, Y)$. Further, it was shown in [8] that $C_W(X, Y) = C_W(X, Y; 0, 0) = \frac{1}{2}\log\frac{1+\rho}{1-\rho}$ and the infimum achieving U^* is a standard Gaussian random variable jointly distributed with (X, Y) as:

$$\begin{aligned} X &= \sqrt{\rho}U^* + \sqrt{1-\rho}N_1 \\ Y &= \sqrt{\rho}U^* + \sqrt{1-\rho}N_2 \end{aligned} \quad (18)$$

where N_1 and N_2 are independent standard Gaussian random variables. We can generate (X^*, Y^*) by passing U^* through independent Gaussian channels as follows:

$$\begin{aligned} X^* &= \sqrt{\frac{\rho}{1-D}}U^* + \sqrt{\frac{(1-D)^2-\rho}{1-D}}\tilde{N}_1 \\ Y^* &= \sqrt{\frac{\rho}{1-D}}U^* + \sqrt{\frac{(1-D)^2-\rho}{1-D}}\tilde{N}_2 \end{aligned} \quad (19)$$

where $\tilde{N}_1$ and $\tilde{N}_2$ are independent standard Gaussian random variables independent of both N_1 and N_2. Therefore there exists a joint density over (X, Y, X^*, Y^*, U^*) satisfying $X^* \leftrightarrow U^* \leftrightarrow Y^*$ and $(X, Y) \leftrightarrow (X^*, Y^*) \leftrightarrow U^*$. This shows that $C_W(X, Y; D, D) \leq C_W(X, Y)$. Therefore in the range $0 < D \leq 1-\rho$, we have $C_W(X, Y; D, D) = C_W(X, Y)$. We note that this specific result was already deduced in [8].

However $C_W(X, Y; D, D)$ in the range $1-\rho \leq D \leq 1$, has never been considered to date. Note that the Shannon lower bound for $R_{X,Y}(D, D)$ is not tight in this range. However, the RD-optimal conditional distribution $P(X^*, Y^*|X, Y)$ in this distortion range is such that $X^* = Y^*$. Therefore the only U which satisfies $(X^* \leftrightarrow U \leftrightarrow Y^*)$ is $U = X^* = Y^*$. Therefore from Theorem 1, we conclude that $C_W(X, Y; D, D) = R_{X,Y}(D, D)$ for $1 - \rho \leq D \leq 1$. Of course, for $D > 1$, $C_W(X, Y; D, D) = 0$. Hence we have completely characterized $C_W(X, Y; D, D)$ for (X, Y) jointly Gaussian for all symmetric distortions $D_1 = D_2 = D > 0$.

III. Gács-Körner's CI

Gács and Körner [1] defined CI of X and Y as the maximum rate of the codeword that can be generated individually at two encoders observing X^n and Y^n separately. Formally:

$$C_{GK}(X, Y) = \sup \frac{1}{n}H(f_1(X^n)) \quad (20)$$

where sup is taken over all f_1 and f_2 such that $P(f_1(X^n) \neq f_2(Y^n)) \to 0$. They showed that $C_{GK}(X, Y)$ is equal to the entropy of the random variable which defines the ergodic decomposition of the stochastic matrix of conditional probabilities $P(X = x|Y = y)$, i.e., if J is a random variable such that $J = j$ iff $x \in \mathcal{X}_j \Leftrightarrow y \in \mathcal{Y}_j$, where $\mathcal{X} \times \mathcal{Y} = \bigcup_j \mathcal{X}_j \times \mathcal{Y}_j$, then $C_{GK}(X, Y) = H(J)$.

Gács-Körner's original definition of CI was naturally unrelated to the Gray-Wyner network, which it predates. However, an equivalent and insightful characterization of $C_{GK}(X, Y)$ was given by Ahlswede and Körner [5] in terms of $\mathcal{R}_{GW}(0, 0)$ as follows:

$$C_{GK}(X, Y) = \max R_0 : (R_0, R_1, R_2) \in \mathcal{R}_{GW}(0, 0) \quad (21)$$

subject to,

$$R_0 + R_1 = H(X) \qquad R_0 + R_2 = H(Y) \quad (22)$$

Specifically, Ahlswede and Körner showed that:

$$H(J) = C_{GK}(X, Y) = \sup I(X, Y; U) \quad (23)$$

subject to,

$$Y \leftrightarrow X \leftrightarrow U \text{ and } X \leftrightarrow Y \leftrightarrow U \quad (24)$$

Though the original definition of Gács-Körner's CI does not have a direct lossy interpretation, the equivalent definition given by Ahlswede and Körner in terms of the lossless Gray-Wyner region can be easily extended to the lossy setting similar to Wyner's CI. These generalizations provide theoretical insight into the performance limits of practical databases for fusion storage of correlated sources as described in [4].

We define the lossy generalization of Gács-Körner's CI at (D_1, D_2), denoted by $C_{GK}(X, Y; D_1, D_2)$ as.

$$\sup R_0 : (R_0, R_1, R_2) \in \mathcal{R}_{GW}(D_1, D_2) \tag{25}$$

subject to,

$$R_0 + R_1 = R_X(D_1) \qquad R_0 + R_2 = R_Y(D_2) \tag{26}$$

We provide an information theoretic characterization for $C_{GK}(X, Y; D_1, D_2)$ in the following theorem. We again assume that there exists channels which achieve $R_X(D_1)$ and $R_Y(D_2)$ respectively, noting that the results can be easily extended to more general source densities. We denote the set of all channels which achieve $R_X(D_1)$ and $R_Y(D_2)$ by $\mathcal{P}_{D_1}^X$ and $\mathcal{P}_{D_1}^X$ respectively.

Theorem 2. *A single letter characterization of $C_{GK}(X, Y; D_1, D_2)$ is given by:*

$$C_{GK}(X, Y; D_1, D_2) = \sup I(X, Y; U) \tag{27}$$

where the supremum is over all joint densities $(X, Y, \tilde{X}, \tilde{Y}, U)$ such that the following Markov conditions hold:

$$\begin{aligned} Y \leftrightarrow X \leftrightarrow U \qquad X \leftrightarrow Y \leftrightarrow U \\ X \leftrightarrow \tilde{X} \leftrightarrow U \qquad Y \leftrightarrow \tilde{Y} \leftrightarrow U \end{aligned} \tag{28}$$

where $P(\tilde{X}|X) \in \mathcal{P}_{D_1}^X$ and $P(\tilde{Y}|Y) \in P_{D_2}^Y$.

Proof: The proof follows in very similar lines to the proof of Theorem 1. The original Gray-Wyner's characterization is in fact sufficient in this case. Again we first assume that there are unique channels $P(\tilde{X}|X)$ and $P(\tilde{Y}|Y)$ which achieve $R_X(D_1)$ and $R_Y(D_2)$ respectively. The proof extends directly to the case of multiple RD optimal channels.

We are interested in characterizing the points in $\mathcal{R}_{GW}(D_1, D_2)$ which lie on both the planes $R_0 + R_1 = R_X(D_1)$ and $R_0 + R_2 = R_Y(D_2)$. Therefore we have the following series of inequalities:

$$\begin{aligned} R_X(D_1) &= R_0 + R_1 \\ &\geq I(X, Y; U) + I(X; \hat{X}|U) \\ &= I(X; \hat{X}, U) + I(Y; U|X) \\ &\geq I(X; \hat{X}) \geq R_X(D_1) \end{aligned}$$

Writing similar inequality relations for Y and following the same arguments as in Theorem 1, it follows that for all joint densities satisfying (28) and for which $P(\tilde{X}|X) \in \mathcal{P}_{D_1}^X$ and $P(\tilde{Y}|Y) \in \mathcal{P}_{D_2}^Y$, there exists at least one point in $\mathcal{R}_{GW}(D_1, D_2)$ which satisfies both $R_0 + R_1 = R_X(D_1)$ and $R_0 + R_2 = R_Y(D_2)$ and for which $R_0 = I(X, Y; U)$. This proves the theorem. ■

Corollary 1. $C_{GK}(X, Y; D_1, D_2) \leq C_{GK}(X, Y)$

Proof: This corollary follows directly from Theorem 2 as conditions in (24) are a subset of the conditions in (28). ■

It is easy to show that if the random variables (X, Y) are jointly Gaussian with a correlation coefficient $\rho < 1$, then $C_{GK}(X, Y) = 0$. Hence from Corollary 1, it follows that, for jointly Gaussian random variables with correlation coefficient strictly less than 1, $C_{GK}(X, Y; D_1, D_2) = 0 \; \forall D_1, D_2$ under any distortion metric. It is well known that $C_{GK}(X, Y)$ is typically very small (usually zero) and depends only on the zeros of the joint distribution. In the general setting, as $C_{GK}(X, Y; D_1, D_2) \leq C_{GK}(X, Y)$, it would seem that Theorem 2 has very limited practical significance. However in a separate paper, currently under preparation, we show that $C_{GK}(X, Y; D_1, D_2)$ plays a central role in scalable coding of sources that are not successively refinable.

IV. Conclusion

In this paper we derived single letter information theoretic characterizations for the lossy generalizations of the two most prevalent notions of common information due to Wyner and Gács-Körner. These generalizations allow us to extend the theoretical interpretation underlying their original definitions to sources with infinite entropy (eg. continuous random variables). We use these information theoretic characterizations to derive the common information of symmetric bivariate Gaussian random variables.

References

[1] P. Gács and J. Körner, "Common information is far less than mutual information", Problems of Control and Information Theory, vol. 2, pp. 149-162, 1973.

[2] A. Wyner, "The common information of two dependent random variables", IEEE Trans. on Information Theory, vol. 21, pp. 163 – 179, Mar 1975.

[3] R. Gray and A. Wyner, "Source coding for a simple network", Bell systems technical report, Dec 1974.

[4] K. Viswanatha, E. Akyol, K. Rose, "An optimal transmit-receive rate tradeoff in Gray-Wyner network and its relation to common information," in Proc. IEEE Information Theory Workshop (ITW), pp.105-109, Oct 2011.

[5] R. Ahlswede and J. Körner, "On common information and related characteristics of correlated information sources", preprint presented at the 7th Prague Conference on Information Theory, 1974.

[6] H. Yamamoto, "Coding theorems for Shannon's cipher system with correlated source outputs and common information," IEEE Trans. on Information Theory, vol. 40, no. 1, pp. 85-95, Jan 1994.

[7] S. Kamath and V. Anantharam, "A new dual to the Gács - Körner common information defined via the Gray-Wyner system," in Proc. 48th Allerton Conf. on Communication, Control, and Computing, pp. 1340–1346, 2010.

[8] G. Xu, W. Liu, B. Chen; , "Wyners common information for continuous random variables - A lossy source coding interpretation," in Proc. 45th Annual Conference on Information Sciences and Systems (CISS), pp.1-6, Mar 2011.

[9] R. Gray, "A new class of lower bounds to information rates of stationary sources via conditional rate-distortion functions," IEEE Trans. on Information Theory, vol. 19, no. 4, pp. 480- 489, Jul 1973.

[10] R. Venkataramani, G. Kramer, V.K. Goyal, "Multiple description coding with many channels," IEEE Trans. on Information Theory, vol. 49, no. 9, pp. 2106- 2114, Sept 2003.

[11] T. Berger, "Rate distortion theory", Englewood, New Jersey, Prentice-Hall, 1971.