A new dual to the Gács-Körner common information defined via the Gray-Wyner system

Sudeep Kamath and Venkat Anantharam
EECS Department
University of California, Berkeley, CA, USA
sudeep,ananth@eecs.berkeley.edu

***Abstract*— We consider jointly distributed random variables X and Y. After describing the Gács-Körner common information between the random variables from the viewpoint of the capacity region of the Gray-Wyner system, we propose a new notion of common information between the random variables that is dual to the Gács-Körner common information from this viewpoint in a well-defined sense. We characterize this quantity explicitly in terms of two auxiliary quantities that are asymmetric in nature, and illustrate the operational significance of these new quantities by characterizing a corner point of the solution to a problem of source coding with side-information in terms of them. We also contrast this new concept of common information for a pair of random variables with the Wyner common information of the random variables, which is also a kind of dual to the Gács-Körner common information.**

I. Introduction

In information theory, mutual information is the most frequently used notion of common information between two jointly distributed random variables. Mutual information correctly captures the information given by one of the random variables about the other in a wide range of asymptotic formulations of important communication problems of interest. However, there are other notions of common information that arise naturally in certain problem formulations. Two such notions of common information between a pair of random variables, now well-known as the Gács-Körner common information and the Wyner common information, were defined respectively in [1] and [2].

Suppose we are given a pair of jointly distributed random variables (X, Y).

- The Gács-Körner common information $K(X;Y)$, roughly speaking, measures the amount of common randomness that can be separately extracted from either marginal of the two jointly distributed random variables, in an asymptotic formulation. This was defined and characterized in [1]. Witsenhausen strengthened the results of [1] in [3].
- The Wyner common information $C(X;Y)$ was defined in two alternate ways in [2]. It was shown that the two definitions are equivalent and the quantity was characterized in terms of an optimization problem involving an auxiliary random variable. One of the interpretations of this quantity is as the minimum rate at which external randomness must be supplied to physically separated agents, each responsible for one of the marginals, so that they are able to closely reproduce the given joint distribution, in an asymptotic formulation.

The descriptions above show that these two quantities can be roughly thought of as duals of each other in an operational sense. It is also well-known that $K(X;Y) \leq I(X;Y) \leq C(X;Y)$, where both the inequalities hold with equality if and only if any one of the inequalities holds with equality if and only if the joint distribution admits a decomposition of the form $X = (X', U), Y = (Y', U)$ with $X' \perp\!\!\!\perp Y'|U$.

There has been a lot of interesting work on problems of common information extending the above notions to more general scenarios. [4] develops a generalization of the Gács-Körner common information with applications to security. [5] and [6] look at the space of joint distributions that can be generated by nodes in a network under communication constraints. [7] considers the problem of simulating a channel given different rates of common randomness. [8] and [9] consider the problem of information-theoretically secure secret key agreement between nodes in a network in scenarios exploiting common randomness.

The Gray-Wyner source coding system was introduced in [10] and its capacity region was determined via an auxiliary random variable. The Gray-Wyner system is a two-source source coding problem, and there are three rates of interest: a common rate, and two private rates. Our contribution in this paper is as follows. We view the mutual information, the Gács-Körner common information, and the Wyner common information from the perspective of the capacity region of the Gray-Wyner system. Specifically, these quantities correspond to certain values of the common rate where the capacity region of the Gray-Wyner system bears a natural relationship to the elementary outer bound to this region defined by basic information-theoretic considerations. We present a formula expressing the Gács-Körner common information as the supremum of an information-theoretic

quantity over a set defined using one auxiliary random variable, that has a similar flavor to the well-known formula for the Wyner common information [2]. We then define a new concept of common information associated with a pair of jointly distributed random variables that is motivated by the relation between the capacity region of the Gray-Wyner system and its elementary outer bound, and, from this viewpoint, can be perceived as a dual to the Gács-Körner common information. We characterize this quantity in terms of two auxiliary asymmetric quantities, and show that these newly defined quantities have operational significance in the context of characterizing a corner point of the problem of source coding with side information that was defined in [11].

II. Setup and Definitions

Let (X, Y) be jointly distributed random variables taking values in finite sets $\mathcal{X}$ and $\mathcal{Y}$ respectively with the joint distribution $Q(x, y)$. [1] defines a quantity, nowadays called the Gács-Körner common information $K(X;Y)$ of (X, Y), and establishes it to be equal to the supremum of $H(V)$ over all random variables V, taking values in some finite set $\mathcal{V}$, that can be written as

$$V = f(X) = g(Y)$$

for some functions $f : \mathcal{X} \mapsto \mathcal{V}$ and $g : \mathcal{Y} \mapsto \mathcal{V}$. Since $H(V) = I(V;V) = I(f(X); g(Y)) \leq I(X;Y)$, we have $K(X;Y) \leq I(X;Y)$. An alternative approach is to define this quantity using the Gray-Wyner system as described below.

A. Gray-Wyner system

The Gray-Wyner system is a source coding system as shown in Figure 1. The problem formulation from [10] is as follows. Let $\{(X_i, Y_i)\}_{i=1}^{\infty}$ be a sequence of i.i.d. random variable pairs drawn from the set $\mathcal{X} \times \mathcal{Y}$ with the marginal distribution $(X, Y) \sim Q(x, y)$. A code with parameters (n, M_0, M_1, M_2) is comprised of an encoder, i.e. a mapping

$$f_E : \mathcal{X}^n \times \mathcal{Y}^n \mapsto [M_0] \times [M_1] \times [M_2],$$

where $[M]$ denotes the set $\{1, 2, \ldots, M\}$, and a decoder, i.e. a pair of mappings

$$f_D^{(\mathcal{X})} : [M_0] \times [M_1] \mapsto \mathcal{X}^n;$$

$$f_D^{(\mathcal{Y})} : [M_0] \times [M_2] \mapsto \mathcal{Y}^n.$$

Let $f_E(X^n, Y^n) = (S_0, S_1, S_2)$. We say that a triple of rates (R_0, R_1, R_2) is achievable if for arbitrary $\epsilon > 0$, there exists, for sufficiently large n, a code with parameters (n, M_0, M_1, M_2) with $\frac{1}{n} \log M_i \leq R_i + \epsilon$ for $i = 0, 1, 2$ and

$$\frac{1}{n}\mathbb{E}[d_H(X^n, f_D^{(\mathcal{X})}(S_0, S_1))] \leq \epsilon;$$

$$\frac{1}{n}\mathbb{E}[d_H(Y^n, f_D^{(\mathcal{Y})}(S_0, S_2))] \leq \epsilon,$$

where $d_H(\cdot, \cdot)$ is the Hamming distortion metric.

An elementary outer bound to the set of achievable rate triples in the Gray-Wyner system, coming from basic information-theoretic considerations, is given in Thm. 2 of [10]; the following inequalities must hold:

$$R_0 + R_1 \geq H(X), \quad (1)$$

$$R_0 + R_2 \geq H(Y), \quad (2)$$

$$R_0 + R_1 + R_2 \geq H(X, Y). \quad (3)$$

Let $\mathcal{L} := \{(R_0, R_1, R_2) : R_0, R_1, R_2 \geq 0, R_0 + R_1 \geq H(X), R_0 + R_2 \geq H(Y), R_0 + R_1 + R_2 \geq H(X, Y)\}$, and let $\mathcal{R}$ be the set of achievable rate triples. Clearly, $\mathcal{R} \subseteq \mathcal{L}$ and, in general, the inclusion is strict [10].

Let $\mathcal{P}$ be the family of probability distributions $p(x, y, w)$ with $x \in \mathcal{X}, y \in \mathcal{Y}, w \in \mathcal{W}$, where $\mathcal{W}$ is a finite set and which satisfy $\sum_{w \in \mathcal{W}} p(x, y, w) = Q(x, y)$, and $\mathcal{P}_b$ the subset of $\mathcal{P}$ where $\mathcal{W}$ has cardinality bounded as $|\mathcal{W}| \leq |\mathcal{X}|.|\mathcal{Y}| + 2$. For $p(x, y, w) \in \mathcal{P}$, define $\mathcal{R}^{(p)} := \{(R_0, R_1, R_2) : R_0 \geq I(X, Y; W), R_1 \geq H(X|W), R_2 \geq H(Y|W)$ for $(X, Y, W) \sim p(x, y, w)\}$, and define $\mathcal{R}_{\mathcal{P}} := \cup_{p \in \mathcal{P}} \mathcal{R}^{(p)}$ and $\mathcal{R}_{\mathcal{P}_b} := \cup_{p \in \mathcal{P}_b} \mathcal{R}^{(p)}$.

Theorem 2.1: $\mathcal{R} = \mathsf{cl}\,(\mathcal{R}_{\mathcal{P}}) = \mathcal{R}_{\mathcal{P}_b}$, where $\mathsf{cl}\,(\cdot)$ denotes closure of a set.

Proof : [10] establishes the first equality, where one does not have the cardinality bound on $\mathcal{W}$. The second equality can be obtained once one gets the cardinality bound by the Carathèodory-Fenchel theorem as follows.

Given (R_0, R_1, R_2) satisfying $R_0 \geq I(X, Y; W), R_1 \geq H(X|W)$ and $R_2 \geq H(Y|W)$ for some $(X, Y, W) \sim p(x, y, w)$ without a cardinality bound on $\mathcal{W}$ and where (X, Y) has the law of the source $Q(x, y)$, fix $p(x, y|w)$ and consider the equations

$$\sum_{w \in \mathcal{W}} p(w) p(x, y|w) = Q(x, y)$$

which give $|\mathcal{X}|.|\mathcal{Y}| - 1$ equality constraints since one of them is implied by the others, and the three inequalities

$$\sum_{w \in \mathcal{W}} p(w) H(p(x|w)) \leq H(X|W),$$

$$\sum_{w \in \mathcal{W}} p(w) H(p(y|w)) \leq H(Y|W),$$

$$\sum_{w \in \mathcal{W}} p(w) H(p(x, y|w)) \leq H(X, Y|W).$$

This system of linear equations and inequalities for $(p(w) : w \in \mathcal{W})$ is feasible in the simplex given by $p(w) \geq 0, \forall w \in \mathcal{W}$ and $\sum_{w \in \mathcal{W}} p(w) = 1$, so it must have a solution in this simplex with $(p(w), w \in \mathcal{W})$ being non-zero on a subset of size at most $(|\mathcal{X}|.|\mathcal{Y}| -$

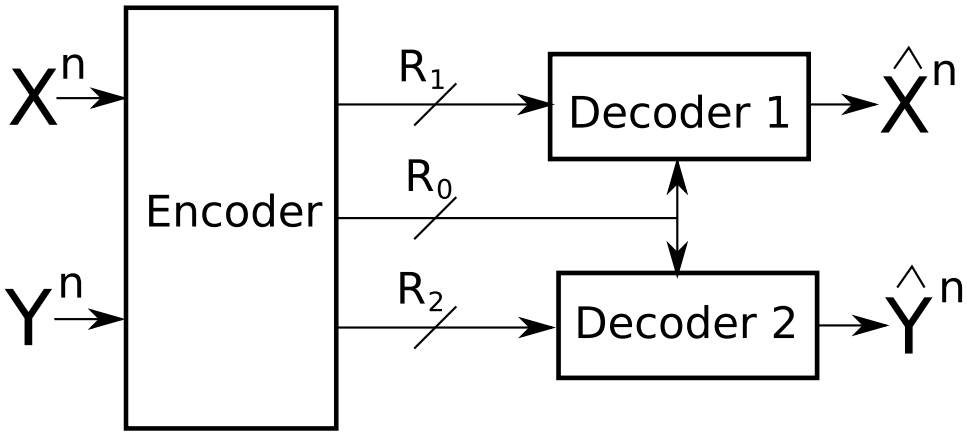

Fig. 1. The Gray-Wyner source coding system

$1)+3 = |\mathcal{X}|.|\mathcal{Y}|+2$, which proves the desired cardinality bound. ∎

The mutual information, the Gács-Körner common information, and the Wyner common information of a pair of random variables can be understood in terms of the relation between the region $\mathcal{R}$ in Theorem 2.1 and the elementary outer bound $\mathcal{L}$ to this region, as defined by the inequalities (1), (2), (3). To see this, we observe the following.

1) Mutual information $I(X;Y)$ is the value of R_0 at the intersection of the three planes constituting the boundary of the outer bound $\mathcal{L}$, i.e. the inequalities (1), (2), (3).
2) Gács-Körner common information $K(X;Y)$ is the supremum of R_0 over all achievable rate triples that satisfy inequalities (1) and (2) with equality. This is different from the original definition in [1]. It is proved in Theorem 3.1 below that the two notions are equivalent. Note that $K(X;Y) \leq I(X;Y)$.
3) Wyner common information $C(X;Y)$ is the infimum of R_0 over all achievable rate triples that satisfy the inequality (3) with equality. [2]

One may thus think of the Wyner common information as a dual to the Gács-Körner common information also in the above sense. However, we may also think of the Gács-Körner common information in a different way in the context of the relation between the region $\mathcal{R}$ and its elementary outer bound $\mathcal{L}$, which then suggests a different dual to the Gács-Körner common information.

4) Gács-Körner common information $K(X;Y)$ is the supremum of R_0 such that $R_0 \leq I(X;Y)$ and for which R_0, all non-negative rate triples allowed by inequalities (1), (2) and (3) are achievable. It is straightforward to see that this description is equivalent to the one given in item 2) above.
5) Define $U(X;Y)$ as the infimum of R_0 such that $R_0 \geq I(X;Y)$ and for which R_0, all non-negative rate triples allowed by inequalities (1), (2) and (3) are achievable.

We will provide a characterization of $U(X;Y)$ and discuss its connections with a problem of source coding with side information, introduced in Section IV.

III. Results and Proofs

First, we prove the equivalence of the two notions of the Gács-Körner common information, namely:

- $K(X;Y)$ is the supremum of $H(V)$ over all random variables V, taking values in some finite set $\mathcal{V}$, that can be written as
$$V = f(X) = g(Y);$$
- The alternative definition is $\tilde{K}(X;Y) = \sup\{R : R \leq I(X;Y), \{R_0 = R\} \cap \mathcal{L} \subseteq \mathcal{R}\}$.

Theorem 3.1: $K(X;Y) = \tilde{K}(X;Y)$.

Proof : First note that for $0 \leq R \leq I(X;Y)$, we have $\{R_0 = R\} \cap \mathcal{L} \subseteq \mathcal{R}$ if and only if $(R, H(X) - R, H(Y) - R) \in \mathcal{R}$.

Start with any $V = f(X) = g(Y)$. Setting $R = H(V), R_1 = H(X|V)$ and $R_2 = H(Y|V)$ we can verify that $R \leq I(X;Y), R_1 = H(X) - R$ and $R_2 = H(Y) - R$. With $W = V$, we get that $(R, H(X) - R, H(Y) - R) \in \mathcal{R}$. This gives $\tilde{K}(X;Y) \geq K(X;Y)$.

For the converse, suppose $(R, H(X) - R, H(Y) - R) \in \mathcal{R}$. From Theorem 2.1, $\exists W$ defined by $p(w|x,y)$, taking values in the finite set $\mathcal{W}$, so that

$$\begin{aligned} R &= I(X,Y;W) + \rho_0; \\ H(X) - R &= H(X|W) + \rho_1; \\ H(Y) - R &= H(Y|W) + \rho_2 \ , \end{aligned}$$

for $\rho_i \geq 0$ for $i = 0, 1, 2$.

Adding the first and second equality gives $\rho_0 + \rho_1 + I(Y;W|X) = 0$. Adding the first and the third gives $\rho_0 + \rho_2 + I(X;W|Y) = 0$. These yield $\rho_0 = \rho_1 =$

$\rho_2 = I(Y;W|X) = I(X;W|Y) = 0$. Thus, the joint distribution of (X,Y,W) must respect the Markov chains $W - X - Y$ and $X - Y - W$.

We may assume without loss of generality that $p(x) > 0$ for all $x \in \mathcal{X}$ and $p(y) > 0$ for all $y \in \mathcal{Y}$. Let

$$\mathcal{A} := \{(x,y) \in \mathcal{X} \times \mathcal{Y} \ : \ p(x,y) > 0\} \ .$$

Consider $(x,y) \in \mathcal{A}$. Then, for any $w \in \mathcal{W}$ such that $p(w|x,y) > 0$, we have

$$p(w,x,y) = p(w|x,y)Q(x,y) = p(w|x)Q(x,y)$$

from the Markov chain $W - X - Y$, and

$$p(w,x,y) = p(w|x,y)Q(x,y) = p(w|y)Q(x,y)$$

from the Markov chain $W - Y - X$. Let $\Phi^W_{(X,Y)}$ denote the function from $\mathcal{A}$ to $\mathbb{P}(\mathcal{W})$ (the simplex of probability distributions on $\mathcal{W}$), taking (x,y) to the conditional law of W given $\{X = x, Y = y\}$, i.e. to the law $(w \mapsto p(w|x,y))$. Similarly, let Φ^W_X denote the function from $\mathcal{X}$ to $\mathbb{P}(\mathcal{W})$ taking x to the conditional law of W given $\{X = x\}$ i.e. to the law $(w \mapsto p(w|x))$, and let Φ^W_Y denote the function from $\mathcal{Y}$ to $\mathbb{P}(\mathcal{W})$ taking y to the conditional law of W given $\{Y = y\}$ i.e. to the law $(w \mapsto p(w|y))$. The preceding two calculations tell us that for $(x,y) \in \mathcal{A}$ we have

$$\Phi^W_{(X,Y)}(x,y) = \Phi^W_X(x) = \Phi^W_Y(y) \ .$$

Next, consider the bipartite graph with left vertex set $\mathcal{X}$ and right vertex set $\mathcal{Y}$, where the edge (x,y) exists iff $(x,y) \in \mathcal{A}$. Consider the decomposition of this graph into connected components. By the above, on every connected component each edge maps to the same law on $\mathcal{W}$ under the map $\Phi^W_{(X,Y)}$. It follows that there must be a random variable V_{XY} that is constant on the edges of any connected component, taking values in a finite set $\mathcal{V}$ (one can think of this as just the index of the connected component) and another finite random variable U, which is independent of (X,Y), taking values in the finite set $\mathcal{U}$, such that W can be written as a deterministic function $W = t(U, V_{XY})$ (here U is just used to construct the law of W once one knows the index of the connected component). We will call V_{XY} the block index random variable of the pair (X,Y).

We now have

$$R_0 = I(X,Y;W) = I(X,Y;t(U,V_{XY})) \leq H(V_{XY}) \ ,$$

where the first equality is because $\rho_0 = 0$, and the last inequality is because U is independent of (X,Y). This gives

$$\tilde{K}(X,Y) \leq K(X,Y)$$

and concludes the proof. ∎

Corollary 3.2:

$$K(X;Y) = \sup_{\substack{W-X-Y \\ X-Y-W}} I(X,Y;W) \ .$$

Proof : From Theorem 3.1,

$$\begin{aligned} K(X;Y) &= \sup\{R : R \leq I(X;Y), \{R_0 = R\} \cap \mathcal{L} \subseteq \mathcal{R}\} \\ &= \sup\{R : (R, H(X) - R, H(Y) - R) \in \mathcal{R}\}. \end{aligned}$$

For any W with a joint distribution $p(x,y,w)$ satisfying $W - X - Y$ and $X - Y - W$, we have from Theorem 2.1 that $(I(X,Y;W), H(X) - I(X,Y;W), H(Y) - I(X,Y;W)) \in \mathcal{R}$ because $H(X|W) = H(X) - I(X,Y;W)$ and $H(Y|W) = H(Y) - I(X,Y;W)$. It is also straightforward to check that $I(X,Y;W) \leq I(X;Y)$. This gives $K(X;Y) \geq \sup_{W-X-Y, X-Y-W} I(X,Y;W)$.

For the converse, from Theorem 2.1, if $(R, H(X) - R, H(Y) - R) \in \mathcal{R}$, then $\exists p(x,y,w) \in \mathcal{P}$ such that $R \geq I(X,Y;W), H(X) - R \geq H(X|W), H(Y) - R \geq H(Y|W)$. Adding the first and second inequalities gives us $W - X - Y$ and equalities for both. Adding the first and third inequalities gives us $X - Y - W$ and equalities for both. So, $K(X;Y) \leq \sup_{W-X-Y, X-Y-W} I(X,Y;W)$. ∎

The Wyner common information was defined as in 3) and characterized in [2]. We now state this characterization and provide a proof for completeness.

Theorem 3.3: (Wyner [2])

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

Proof : If some triple $(R_0, R_1, R_2) \in \mathcal{R}$ satisfies $R_0 + R_1 + R_2 = H(X,Y)$, then by Theorem 2.1 we must have for some W that

$$\begin{aligned} R_0 &= I(X,Y;W) + \rho_0; \\ R_1 &= H(X|W) + \rho_1; \\ R_2 &= H(Y|W) + \rho_2, \end{aligned}$$

for $\rho_i \geq 0$ for $i = 0, 1, 2$.

Adding the three equalities gives $H(X,Y) = H(X,Y) + I(X;Y|W) + \rho_0 + \rho_1 + \rho_2$ which implies $X - W - Y$ and also that $\rho_0 = \rho_1 = \rho_2 = 0$ so that $R_0 = I(X,Y;W)$. So, $C(X;Y) \geq \inf_{X-W-Y} I(X,Y;W)$.

Conversely, for any W, $(I(X,Y;W), H(X|W), H(Y|W)) \in \mathcal{R}$ and if W happens to satisfy $X - W - Y$, then we have $I(X,Y;W) + H(X|W) + H(Y|W) = H(X,Y)$. Thus, $C(X;Y) \leq \inf_{X-W-Y} I(X,Y;W)$. ∎

Let us define the following.

Definition $G(Y \to X) := \inf\{R : (R, H(X|Y), H(Y) - R) \in \mathcal{R}\}$.

Definition $G(X \to Y) := \inf\{R : (R, H(X) - R, H(Y|X)) \in \mathcal{R}\}$.

$G(Y \to X)$ (respectively $G(X \to Y)$) is the infimum of R_0 over all achievable rate triples that satisfy (2) and (3) (respectively (1) and (3)) of the outer bound with equality.

We characterize the above defined quantities and $U(X;Y)$ in the following theorem.

Theorem 3.4:

$$G(Y \to X) = \inf_{\substack{X-Y-W \\ X-W-Y}} I(X,Y;W);$$

$$G(X \to Y) = \inf_{\substack{W-X-Y \\ X-W-Y}} I(X,Y;W);$$

$$U(X;Y) = \max\{G(Y \to X), G(X \to Y)\}.$$

Proof : We begin with $G(Y \to X)$. For any finite set $\mathcal{W}$, and $p(w|x,y)$ with $p(x,y,w) = Q(x,y)p(w|x,y) \in \mathcal{P}$, we have from Theorem 2.1 that $(I(X,Y;W), H(X|W), H(Y|W)) \in \mathcal{R}$. If W satisfies, in addition, the Markov conditions $X-Y-W$ and $X-W-Y$, we get $I(X,Y;W) + H(Y|W) = H(Y)$ and $I(X,Y;W) + H(X|W) + H(Y|W) = H(X,Y)$. This immediately gives $G(Y \to X) \leq \inf_{X-Y-W, X-W-Y} I(X,Y;W)$.

For the converse, suppose for some R, we have $(R, H(X|Y), H(Y) - R) \in \mathcal{R}$. From Theorem 2.1, it follows that $\exists p(x,y,w) \in \mathcal{P}$ such that

$$\begin{aligned} R &= I(X,Y;W) + \rho_0, \\ H(X|Y) &= H(X|W) + \rho_1, \\ H(Y) - R &= H(Y|W) + \rho_2. \end{aligned}$$

Adding the first and third equalities above gives $H(Y) = H(Y) + I(X;W|Y) + \rho_0 + \rho_2$. Adding all three gives $H(X,Y) = H(X,Y) + I(X;Y|W) + \rho_0 + \rho_1 + \rho_2$. It follows that $\rho_0 = \rho_1 = \rho_2 = I(X;W|Y) = I(X;Y|W) = 0$. Thus, $X-Y-W$ and $X-W-Y$ hold and so, $G(Y \to X) \geq \inf_{X-Y-W, X-W-Y} I(X,Y;W)$.

Similarly, we can prove the claimed formula for $G(X \to Y)$. It is easy to verify that $I(X;Y) \leq C(X;Y) \leq G(Y \to X) \leq H(Y)$. The first inequality is from the characterization of $C(X;Y)$ in Theorem 4.1, the second inequality is from the characterization in this theorem and because the infimum is over a smaller set of joint distributions, and the third inequality follows from the characterization in this theorem with $W = Y$. Similarly, $I(X;Y) \leq C(X;Y) \leq G(X \to Y) \leq H(X)$.

Now, we turn our attention to the main quantity $U(X;Y) = \inf\{R : R \geq I(X;Y), \{R_0 = R\} \cap \mathcal{L} \subseteq \mathcal{R}\}$. Assume without loss of generality that $H(X) \leq H(Y)$.

Note that for $R \geq I(X;Y)$,

$$\begin{aligned} & \{R_0 = R\} \cap \mathcal{L} \\ = \; & \{(R, R_1, R_2) : R_1 \geq (H(X) - R)^+, R_2 \geq (H(Y) - R)^+, \\ & \quad R_1 + R_2 \geq (H(X,Y) - R)^+\} \\ = \; & \begin{cases} \{(R, R_1, R_2) : R_1 \geq H(X) - R, R_2 \geq H(Y) - R, \\ \quad R_1 + R_2 \geq H(X,Y) - R\}, & \text{if } I(X;Y) \leq R < H(X) \\ \{(R, R_1, R_2) : R_1 \geq 0, R_2 \geq H(Y) - R, \\ \quad R_1 + R_2 \geq H(X,Y) - R\}, & \text{if } H(X) \leq R < H(Y) \\ \{(R, R_1, R_2) : R_1 \geq 0, R_2 \geq 0, \\ \quad R_1 + R_2 \geq H(X,Y) - R\}, & \text{if } H(Y) \leq R < H(X,Y) \\ \{(R, R_1, R_2) : R_1 \geq 0, R_2 \geq 0, \\ \quad R_1 + R_2 \geq 0\}, & \text{if } R \geq H(X,Y). \end{cases} \end{aligned}$$

Suppose that for some $R \geq I(X;Y)$ we have $\{R_0 = R\} \cap \mathcal{L} \subseteq \mathcal{R}$. If $R < H(X)$, then $(R, H(X) - R, H(Y|X))$ is achievable, so $R \geq G(X \to Y)$. If $R \geq H(X)$, then of course, $R \geq G(X \to Y)$. Thus, $U(X;Y) \geq G(X \to Y)$. Similarly, if $R < H(Y)$, then $(R, H(X|Y), H(Y) - R)$ is achievable, so $R \geq G(Y \to X)$. If $R \geq H(Y)$, then of course, $R \geq G(Y \to X)$. Thus, $U(X;Y) \geq G(X \to Y)$. This gives $U(X;Y) \geq \max\{G(Y \to X), G(X \to Y)\}$.

Now, choose an $R \geq \max\{G(Y \to X), G(X \to Y)\}$. We will show that $\{R_0 = R\} \cap \mathcal{L} \subseteq \mathcal{R}$. As argued above, $\{R_0 = R\} \cap \mathcal{L}$ as a subset of $\{R_0 = R\}$ is the polyhedral region of rate pairs that pointwise dominate a convex combination of two (possibly identical) distinguished rate pairs. We will show that the rate triples defined by these rate pairs, which we call corner points, are achievable.

This follows from the achievability of the points

- $(H(X,Y), 0, 0)$,
- $(H(Y), H(X|Y), 0)$,
- $(H(X), 0, H(Y|X))$,
- $(G(Y \to X), H(X|Y), H(Y) - G(Y \to X))$,
- $(G(X \to Y), H(X) - G(X \to Y), H(Y|X))$,

all of which can be easily verified. The details are below:

- Case I: $R \geq H(X,Y)$.
  Here, $\{R_0 = R\} \cap \mathcal{L} = \{(R, R_1, R_2) : R_1 \geq 0, R_2 \geq 0\}$. This is achievable since $(H(X,Y), 0, 0) \in \mathcal{R}$.
- Case II: $H(X,Y) > R \geq H(Y) \geq H(X)$.
  Here, $\{R_0 = R\} \cap \mathcal{L} = \{(R, R_1, R_2) : R_1 \geq 0, R_2 \geq 0, R_1 + R_2 \geq H(X,Y) - R\}$. The corner point $(R, 0, H(X,Y) - R)$ is achievable because it is a convex combination of the achievable points $(H(X,Y), 0, 0)$ and $(H(X), 0, H(Y|X))$. Similarly, the corner point $(R, H(X,Y) - R, 0)$ is achievable because it is a convex combination of achievable points $(H(X,Y), 0, 0)$ and $(H(Y), H(X|Y), 0)$.
- Case III: $H(Y) > R \geq H(X)$.
  Now, $\{R_0 = R\} \cap \mathcal{L} = \{(R, R_1, R_2) : R_1 \geq 0, R_2 \geq H(Y) - R, R_1 + R_2 \geq H(X,Y) - R\}$. The

corner point $(R, 0, H(X,Y)-R)$ is a convex combination of the achievable points $(H(X,Y), 0, 0)$ and $(H(X), 0, H(Y|X))$. The other corner point $(R, H(X|Y), H(Y)-R)$ is a convex combination of the achievable points $(H(Y), H(X|Y), 0)$ and $(G(Y \to X), H(X|Y), H(Y) - G(Y \to X))$.

- Case IV: $H(X) > R \geq I(X;Y)$.
  Now, $\{R_0 = R\} \cap \mathcal{L} = \{(R, R_1, R_2) : R_1 \geq H(X) - R, R_2 \geq H(Y) - R, R_1 + R_2 \geq H(X,Y) - R\}$. The corner point $(R, H(X|Y), H(Y)-R)$ is a convex combination of the achievable points $(H(Y), H(X|Y), 0)$ and $(G(Y \to X), H(X|Y), H(Y) - G(Y \to X))$. The other corner point $(R, H(X) - R, H(Y|X))$ is a convex combination of the achievable points $(H(X), 0, H(Y|X))$ and $(G(X \to Y), H(X) - G(X \to Y), H(Y|X))$.

Thus, we have $U(X;Y) = \max\{G(Y \to X), G(X \to Y)\}$. ∎

A. Explicit characterization of $G(Y \to X), G(X \to Y)$ and $U(X;Y)$

Assume without loss of generality that $p(y) > 0 \; \forall y \in \mathcal{Y}$. Let Φ_Y^X denote the function from $\mathcal{Y}$ to $\mathbb{P}(\mathcal{X})$ (the simplex of probability distributions on $\mathcal{X}$) taking y to the conditional law of X given y. We also think of it as a random variable. Let V_{XY} be the block index random variable of the joint distribution of $p(x,y)$, as defined in the proof of Theorem 3.1.

Some results of general interest are collected below:

Lemma 3.5:
1) If, for any random variable V, we have $H(V|X) = H(V|Y) = 0$, then $H(V|V_{XY}) = 0$.
2) V_{XY} has the largest entropy among all random variables V satisfying $H(V|X) = H(V|Y) = 0$.
3) If, for any random variable V, we have $H(V|Y) = 0$ and $X - V - Y$, then $H(\Phi_Y^X|V) = 0$.
4) Φ_Y^X has the smallest entropy among all random variables V satisfying $X - V - Y$ and $H(V|Y) = 0$.
5) For any random variable V satisfying $X - V - Y$ and $X - Y - V$, we have $H(\Phi_Y^X|V) = 0$, so Φ_Y^X has the smallest entropy among all random variables V satisfying $X - V - Y$ and $X - Y - V$. Thus, $G(Y \to X) = H(\Phi_Y^X)$.

Proof : The first four items are easy to prove. We provide a proof for the last item.

Let V taking values in $\mathcal{V}$ satisfy $X - V - Y$ and $X - Y - V$. Consistent with the earlier introduced notation, let Φ_V^X and $\Phi_{(V,Y)}^X$ denote the conditional law of X given V and (V,Y) respectively. Without loss of generality, assume $p(v) > 0 \; \forall \; v \in \mathcal{V}$.

Let us define

$$\mathcal{A} := \{(y,v) \in \mathcal{Y} \times \mathcal{V} \; : \; p(y,v) > 0\} \; .$$

$X - Y - V$ and $X - V - Y$ give $\Phi_V^X(v) = \Phi_Y^X(y) \; \forall (y,v) \in \mathcal{A}$, i.e. $\Phi_V^X = \Phi_Y^X$ as random variables. Thus, $H(\Phi_Y^X|V) = 0$.

So, we have $I(X,Y;V) = I(X,Y;V,\Phi_Y^X) \geq I(X,Y;\Phi_Y^X)$. Note that by definition, Φ_Y^X itself satisfies $X - Y - \Phi_Y^X$. It also satisfies $X - \Phi_Y^X - Y$ which can be checked by writing for all (x,y) with $p(x,y) > 0$,

$$\tau(y) := p(X = x|Y = y) = p(X = x|Y = y, \Phi_Y^X = \tau)$$

if $\Phi_Y^X(y) = \tau$, where the second equality holds because Φ_Y^X is a deterministic function of Y. But this equation verifies that the conditional law of X given Y and Φ_Y^X depends only on Φ_Y^X, i.e. $X - \Phi_Y^X - Y$, as claimed.

Thus, Φ_Y^X achieves the infimum in the definition of $G(Y \to X)$. Therefore, $G(Y \to X) = H(\Phi_Y^X)$. ∎

Similarly, we have $G(X \to Y) = H(\Phi_X^Y)$ and $U(X;Y)$ can be computed easily from the joint distribution of (X,Y) as $U(X;Y) = \max\{H(\Phi_Y^X), H(\Phi_X^Y)\}$.

Example : Let $\mathcal{X} = \{a,b,c\}$ and $\mathcal{Y} = \{\alpha, \beta, \gamma, \delta\}$. Consider the joint distribution on $\mathcal{X} \times \mathcal{Y}$ defined by the following

$$\begin{bmatrix} \frac{4}{37} & 0 & 0 & 0 \\ 0 & \frac{9}{37} & \frac{2}{37} & \frac{3}{37} \\ 0 & \frac{12}{37} & \frac{3}{37} & \frac{4}{37} \end{bmatrix}$$

where the rows are indexed by elements of $\mathcal{X}$ and the columns indexed by elements of $\mathcal{Y}$ in the specified order.

The Gács-Körner common information is given by $K(X;Y) = h\left(\frac{4}{37}, \frac{33}{37}\right)$.

$\Phi_Y^X(\alpha) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}, \Phi_Y^X(\beta) = \begin{bmatrix} 0 & \frac{3}{7} & \frac{4}{7} \end{bmatrix} = \Phi_Y^X(\delta), \Phi_Y^X(\gamma) = \begin{bmatrix} 0 & \frac{2}{5} & \frac{3}{5} \end{bmatrix}$. Thus, Φ_Y^X takes three distinct values with probabilities $4/37, 28/37$ and $5/37$, so $G(Y \to X) = H(\Phi_Y^X) = h\left(\frac{4}{37}, \frac{28}{37}, \frac{5}{37}\right) < H(Y)$.

Similarly, $\Phi_X^Y(a) = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}, \Phi_X^Y(b) = \begin{bmatrix} 0 & \frac{9}{14} & \frac{2}{14} & \frac{3}{14} \end{bmatrix}, \Phi_X^Y(c) = \begin{bmatrix} 0 & \frac{12}{19} & \frac{3}{19} & \frac{4}{19} \end{bmatrix}$. Here Φ_X^Y takes distinct values with probabilities $4/37, 14/37$ and $19/37$, so $G(X \to Y) = H(\Phi_X^Y) = h\left(\frac{4}{37}, \frac{14}{37}, \frac{19}{37}\right) = H(X)$.

Remark : For a *generic* joint probability distribution on the product of finite sets $\mathcal{X} \times \mathcal{Y}$, we have $K(X;Y) = 0$ and $G(Y \to X) = H(Y)$, $G(X \to Y) = H(X)$, $U(X;Y) = \max\{H(X), H(Y)\}$.

IV. Connection to a side-information problem

The side-information problem of interest is the following.

A source X is observed by encoder 1 and a correlated source Y is observed by encoder 2. The samples of the pair source are i.i.d. over time. Encoder i has a rate-limited link of capacity R_i to a decoder D, $i = 1, 2$. A rate pair (R_1, R_2) is achievable if the decoder can reconstruct X with asymptotically vanishing probability of error in the usual block-based source coding problem formulation. The closure of achievable rate pairs is the capacity region denoted by $\mathcal{R}_{\text{si}}$.

The capacity region for this problem, i.e. the closure of achievable rate pairs, is given below.

Theorem 4.1: (Wyner [11])

$\mathcal{R}_{\text{si}} = \cup_{p(w|y)}\{(R_1, R_2) : R_1 \geq H(X|W), R_2 \geq I(Y;W)\}$ where the union is taken over random variables W satisfying $X - Y - W$.

The cardinality of the auxiliary variable in the preceding theorem can be bounded as $|\mathcal{W}| \leq |\mathcal{Y}|+1$. We know that

- If R_2 is set to 0, then the minimum R_1 required is $H(X)$.
- If R_2 is set to $H(Y)$ or higher, then the minimum R_1 required is $H(X|Y)$.

But if R_1 is required to be the lowest possible, that is, $H(X|Y)$, do we need R_2 to be $H(Y)$ or can we make do with a smaller value of R_2? What is the minimum R_2 required? We answer this question in the following theorem.

Theorem 4.2:

$$\inf\{R : (H(X|Y), R) \in \mathcal{R}_{\text{si}}\} = G(Y \to X) = H(\Phi_Y^X)\,.$$

Proof :

If $(H(X|Y), R) \in \mathcal{R}_{\text{si}}$, then by Theorem 4.1, we must have some random variable W jointly distributed with (X, Y) satisfying $X - Y - W$, such that $H(X|Y) \geq H(X|W)$ and $R \geq I(Y;W)$. $X - Y - W$ implies $H(X|Y) \leq H(X|W)$ by the data processing inequality. This gives $H(X|W) = H(X|Y)$. So, $\inf\{R : (H(X|Y), R) \in \mathcal{R}_{\text{si}}\} \geq \inf_{X-Y-W, H(X|W)=H(X|Y)} I(Y;W)$.

Conversely, for any W satisfying $X - Y - W$ and $H(X|W) = H(X|Y)$, Theorem 4.1 asserts that $(H(X|W), I(Y;W))$ is an achievable rate pair. Therefore, $\inf\{R : (H(X|Y), R) \in \mathcal{R}_{\text{si}}\} \leq \inf_{X-Y-W, H(X|W)=H(X|Y)} I(Y;W)$.

This gives $\inf\{R : (H(X|Y), R) \in \mathcal{R}_{\text{si}}\} = \inf_{X-Y-W, H(X|W)=H(X|Y)} I(Y;W)$.

Now, note that $\{X - Y - W, X - W - Y\} \iff \{X - Y - W, H(X|W) = H(X|Y)\}$. To see this, first suppose $X - Y - W$ and $X - W - Y$. $X - Y - W$ gives $H(X|Y) \leq H(X|W)$ and $X - W - Y$ gives $H(X|Y) \geq H(X|W)$. Now, suppose $X - Y - W$ and $H(X|W) = H(X|Y)$. Then, $I(X;Y|W) = H(X|W) - H(X|W,Y) = H(X|Y) - H(X|W,Y) = I(X;W|Y) = 0$, so $X - W - Y$.

Thus, we have

$$\begin{aligned}
&\inf\{R : (H(X|Y), R) \in \mathcal{R}_{\text{si}}\} \\
&= \inf_{X-Y-W, H(X|W)=H(X|Y)} I(Y;W) \\
&= \inf_{X-Y-W, X-W-Y} I(Y;W) \\
&= \inf_{X-Y-W, X-W-Y} I(X,Y;W),
\end{aligned}$$

where the last step follows because $X - Y - W$ implies $I(Y;W) = I(X,Y;W)$. This gives us $\inf\{R : (H(X|Y), R) \in \mathcal{R}_{\text{si}}\} = G(Y \to X) = H(\Phi_Y^X)$ and concludes the proof. ∎

V. Acknowledgements

The research of the authors was supported by the ARO MURI grant W911NF-08-1-0233, "Tools for the Analysis and Design of Complex Multi-Scale Networks", by the NSF grants CCF-0635372 and CNS-0910702, by Marvell Semiconductor Inc., and by the U. C. Discovery program.

References

[1] P. Gács and J. Körner, "Common information is far less than mutual information", *Problems of Control and Information Theory*, vol. 2, no. 2, pp. 119–162, 1972.

[2] A.D. Wyner, "The common information of two dependent random variables", *IEEE Transactions On Information Theory*, vol. 21, no. 2, pp. 163–179, March 1975.

[3] H.S. Witsenhausen, "On sequences of pairs of dependent random variables", *SIAM Journal on Applied Mathematics*, vol. 28, no. 1, pp. 100–113, January 1975.

[4] V. Prabhakaran and M. Prabhakaran, "Assisted common information with applications to secure two-party computation", in *Proc. of IEEE ISIT*, Austin, Texas, June 2010.

[5] T. Cover and H. Permuter, "Capacity of coordinated actions", in *Proc. of IEEE ISIT*, Nice, France, 2007.

[6] P. Cuff, H. Permuter, and T. Cover, "Coordination capacity", *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4181–4206, September 2010.

[7] P. Cuff, "Communication requirements for generating correlated random variables", in *Proc. of IEEE ISIT*, Toronto, Canada, July 2008.

[8] A.A. Gohari and V. Anantharam, "Information-theoretic key agreement of multiple terminals - part I: Source model", *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3973–3996, August 2010.

[9] A.A. Gohari and V. Anantharam, "Information-theoretic key agreement of multiple terminals - part II: Channel model", *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3997–4010, August 2010.

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

[11] A.D. Wyner, "On source coding with side information at the decoder", *IEEE Transactions on Information Theory*, vol. 21, pp. 294–300, May 1975.