Graph Entropy Characterization of Relay-Assisted Zero-Error Source Coding with Side Information

Ofer Shayevitz
Information Theory & Applications Center
University of California, San Diego
La Jolla, CA 92093, USA
ofersha@ucsd.edu

Abstract

A sender knows X and a receiver knows a correlated Z and would like to learn X, without error. The sender can communicate with the receiver only via a relay, that knows a correlated Y. We study the expected number of bits per instance that need to be sent to and from the relay to that end, in the limit of multiple instances.

1 Introduction

We consider the problem of zero-error source coding with side-information over a simple network consisting of a sender, a relay, and a receiver. The sender has some information X. The receiver is in possession of a correlated Z, and would like to learn X, without error. The sender can communicate with the receiver only via a relay, that knows a correlated Y. We study the rates, i.e., the per-instance expected number of bits, that need to be sent to and from the relay to that end, in the limit of multiple instances.

The problem of point-to-point lossless source coding with receiver side information (no relay) under a vanishing block error probability criterion was originally studied in [1], where it was shown that the minimal asymptotical rate from sender to receiver is given by the conditional entropy $H(X|Z)$. The associated point-to-point zero-error source coding problem with receiver side information was later studied in [2]. In that work, the minimal asymptotical rate from sender to receiver was characterized as the limit of the (normalized) chromatic entropy of the AND product of the associated probabilistic confusability graph $(G_{X|Z}, X)$. This quantity was later interpreted as the corresponding complementary graph entropy [3]. A closed form expression for the complementary graph entropy is unknown; in fact, such an expression will yield in particular the zero-error capacity of a graph [4], a notorious open problem. In [2], the authors also considered a smaller family of transmission protocols which are robust to arbitrary side-information errors, in the sense of guaranteeing the reconstruction of the sender's information on each instance where the side-information is "correct", namely feasible with the sender's information. This setting was called "unrestricted inputs". It was shown that the minimal asymptotical rate in this case is exactly the (Körner) entropy $H(G_{X|Z}, X)$ of the confusability graph, which admits a single letter expression. This serves as an upper bound for the complementary graph entropy, whereas $H(X|Z)$ serves as a trivial lower bound. Both bounds can be arbitrarily loose.

The problem of relay-assisted source coding with side information when a vanishing block error probability is allowed, was considered in [5]. In that work, an inner (achievability) and outer (converse) bounds for the rate region were provided and shown to coincide in some cases, while in other cases a gap remains.

In this paper, we address the corresponding zero-error relay-assisted setting. We provide a graph coloring characterization of optimal protocols, and then limit our discussion to robust protocols in a way similar to the unrestricted inputs scenario in the point-to-point case. We provide inner and outer bounds for the region of achievable rates, by appropriately generalizing the notion of graph entropy. In some cases the bounds coincide, while in other a gap remains. One of our bounds raises a question regarding joint covering of random sequences, for which we provide a partial answer.

2 Preliminaries

We denote the set of all finite length binary strings by $\{0,1\}^*$. The length of a string $s \in \{0,1\}^*$ is denoted by $|s|$. A random variable (r.v.) X distributed over a finite alphabet $\mathcal{X}$ is associated with a probability mass function (p.m.f.) $p_X(x)$ over $\mathcal{X}$, where we omit the subscript and write $p(x)$ when there is no confusion. The (Shannon) entropy of X is denoted $H(X)$. The mutual information between two r.v.'s (X,Y) is denoted by $I(X;Y)$. Let U be a r.v. distributed over $2^{\mathcal{X}}$ (the set of all subsets of $\mathcal{X}$). We write $X \in U$ to denote that U contains X with probability one, i.e., that for any $x \in \mathcal{X}$ with $p(x) > 0$,

$$\sum_{u \ni x} p(u|x) = 1$$

For any $x^n \in \mathcal{X}^n$, let π_{x^n} be the p.m.f. over $\mathcal{X}$ that corresponds to the relative frequency of symbols in x^n. For $\varepsilon > 0$, define the (n,ε)-typical set associated with X to be[1]

$$\mathcal{T}_\varepsilon^n(X) \stackrel{\text{def}}{=} \{x^n \in \mathcal{X}^n : \forall x \in \mathcal{X}\,, |p(x) - \pi_{x^n}(x)| \leq \varepsilon p(x)\}$$

The joint typical set $\mathcal{T}_\varepsilon^n(X,Y)$ associated with a pair of r.v. X,Y is defined similarly.

For a graph G, we denote by $\Gamma(G)$ (resp. $\overline{\Gamma}(G)$) the set of all independent (resp. maximal independent) sets in G. Recall that a coloring of G is any function c over the vertex set such that $c^{-1}(\cdot)$ induces a partition of the vertex set into independent sets of G. We denote the n-fold OR product of G by $G^{\vee n}$.

A *probabilistic graph* consist of a graph G and a r.v. X distributed over the vertex set of the graph. The (Körner) *entropy* of a probabilistic graph (G,X) is defined to be (see [7, 8] for equivalent definitions and other properties)

$$H(G,X) \stackrel{\text{def}}{=} \min_{X \in U \in \Gamma(G)} I(X;U)$$

Namely, the minimum is taken over all conditional distributions $p(u|x)$ such that U, a random independent set of G, contains X with probability one. Note that without loss of generality, the minimum can be restricted to maximal independent sets.

Let $(X,Y) \sim p(x,y)$ be a pair of r.v.'s over a finite product alphabet $\mathcal{X} \times \mathcal{Y}$, and let S_X, S_Y denote the support sets of $p(x), p(y)$ respectively. For any $y \in \mathcal{Y}$, let $S_X(y) \stackrel{\text{def}}{=} \{x \in \mathcal{X} : p(x,y) > 0\}$ be the conditional support. We write $S_X(Y)$ for the associated r.v. distributed over $2^{\mathcal{X}}$, which always satisfies $X \in S_X(Y)$. The *confusability graph* $G_{X|Y}$ of X given Y has a vertex set S_X, where (x,x') are connected by an edge if and only if

[1] This definition of typically, also known as *robust typicality*, was originally introduced in [6]

there exists $y \in \mathcal{Y}$ such that both $p(x,y), p(x',y) > 0$. Clearly, $G_{X|Y}$ is a union over all $y \in \mathcal{Y}$ of $S_X(y)$-cliques. More generally, for any function $f(x,y)$, the f-confusability graph $G^f_{X|Y}$ of X given Y has (x,x') as an edge if and only if there exists $y \in \mathcal{Y}$ such that both $p(x,y), p(x',y) > 0$ and also $f(x,y) \neq f(x',y)$. The pairs $(G_{X|Y}, X)$ and $(G^f_{X|Y}, X)$ each constitute a probabilistic graph. The dependence on the random variable is implied hence omitted throughout. e.g., we write $H(G_{XY|Z})$ for $H(G_{XY|Z}, (X,Y))$.

3 Problem Statement

Let (X,Y,Z) be three discrete r.v.'s with a joint distribution $p(x,y,z)$ over a finite product alphabet $\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$. A *sender* knows X and a *receiver* knows Z and would like to learn X, *without error*. The sender can communicate with the receiver only via a *relay*, that knows Y. We study the expected number of bits that the sender must transmit to the relay, and the relay in turn to the receiver, to that end. We assume that both the relay and the receiver are able to tell when the message they receive ends, although this is not essential to our discussion. We will be especially interested in the corresponding multiple-instance setting with i.i.d. triplets $\{(X_k, Y_k, Z_k) \sim p(x,y,z)\}_{k=1}^n$.

A *(deterministic, zero-error) one-shot protocol* for (X,Y,Z) consists of a *sender mapping* $\phi_1 : \mathcal{X} \mapsto \{0,1\}^*$, and a *relay mapping* which is a partial mapping $\phi_2 : \{0,1\}^* \times \mathcal{Y} \mapsto \{0,1\}^*$ defined over the set $\{(\phi_1(x), y) \,:\, p(x,y) > 0\}$. The mappings satisfy:

(i) The ranges of ϕ_1 and ϕ_2 are prefix free sets.[2]

(ii) The pair $(z, \phi_2(\phi_1(x), y))$ uniquely determines x over the support of $p(x,y,z)$.

We continue somewhat informally. A *n-shot protocol* (ϕ_1, ϕ_2) for (X,Y,Z) is a one-shot protocol for (X^n, Y^n, Z^n). A n-shot protocol is said to be *robust* if for any two subsets $A_1, A_2 \subset \{1,\dots,n\}$, if the relay knows only $\{Y_k\}_{k\in A_1}$ and the receiver knows only $\{Z_k\}_{k\in A_2}$, then the receiver can reconstruct $\{X_k\}_{k\in A_1\cap A_2}$ without error[3]. This is the relay-assisted counterpart of the "unrestricted input" assumption in the point-to-point case [2].[4]

The rate pair (R_1, R_2) achieved by a n-shot protocol (ϕ_1, ϕ_2) is defined to be

$$R_1 \stackrel{\text{def}}{=} n^{-1}\mathbb{E}|\phi_1(X^n)|\,, \quad R_2 \stackrel{\text{def}}{=} n^{-1}\mathbb{E}|\phi_2(\phi_1(X^n), Y^n)|$$

We define the rate region $\mathcal{R}(X,Y,Z)$ associated with (X,Y,Z) to be the closure of the set of all rate pairs (R_1,R_2) achievable by some n-shot protocol. Similarly, we define the rate region $\mathscr{R}(X,Y,Z)$ to be the closure of the set of all rate pairs (R_1,R_2) achievable by some robust n-shot protocol.

Finally we comment that the prefix condition (i) can be relaxed by requiring that it holds only over edges of $G_{X|Y}$ and $G^x_{XY|Z}$ respectively, see [2]. This may save no more than $O(\log n)$ bits in each individual rate for a n-shot protocol, which is negligible for large n, hence this relaxation leaves both $\mathcal{R}(X,Y,Z)$ and $\mathscr{R}(X,Y,Z)$ unchanged.

[2]This can be relaxed, as discussed shortly. Note that ϕ_1, ϕ_2 are generally not one-to-one mappings.

[3]Here we implicitly assume that ϕ_2 can additionally depend on the set A_1.

[4]We require only robustness to arbitrary side information erasures, unlike "unrestricted inputs" which requires robustness to arbitrary errors. In the point-to-point case these requirements are equivalent, but in our setting robustness to errors at the relay would mean the relay could not perform multi-instance coding. Our setting does however support robustness to arbitrary errors in the receiver's side information.

4 Characterizing $\mathscr{R}(X, Y, Z)$

4.1 The point to point case (no relay)

Consider the case where $Y = const$. Clearly this corresponds to the case where the sender communicates X directly to the receiver that knows Z (the relay merely acts as a repeater forwarding the message). This setting was studied in [2], who proved that the optimal rate, for robust protocols, is given by $H(G_{X|Z})$. Therefore in this case

$$\mathscr{R}(X, Y, Z) = \big\{R_1, R_2 : R_1 \geq H(G_{X|Z}),\ R_2 \geq H(G_{X|Z})\big\}$$

4.2 Inner bounds

In this section we provide several inner bounds for the region $\mathscr{R}(X, Y, Z)$, which also trivially serve as inner bounds for the possibly larger region $\mathcal{R}(X, Y, Z)$.

4.2.1 The naive approach: point to point protocols

The following inner bound builds on the point-to-point results of [2].

Theorem 1. *Let*[5]

$$\begin{aligned}
\mathscr{R}_{0,1} &\stackrel{\text{def}}{=} \big\{(R_1, R_2)\ :\ R_1 \geq H(G_{X|Z}), R_2 \geq H(G_{X|Z})\big\} \\
\mathscr{R}_{0,2} &\stackrel{\text{def}}{=} \big\{(R_1, R_2)\ :\ R_1 \geq H(G_{X|Y}),\ R_2 \geq H(G^x_{XY|Z})\big\} \\
\mathscr{R}_{0,3} &\stackrel{\text{def}}{=} \big\{(R_1, R_2)\ :\ R_1 \geq H(G_{X|YZ}),\ R_2 \geq H(G_{X|YZ}) + H(G_{Y|Z})\big\}
\end{aligned}$$

Let $\mathscr{R}_0$ be the convex hull of $\mathscr{R}_{0,1} \cup \mathscr{R}_{0,2} \cup \mathscr{R}_{0,3}$. Then $\mathscr{R}_0 \subseteq \mathscr{R}$.

Proof. For the region $\mathscr{R}_{0,1}$, the sender describes X^n to the receiver using an expected $H(G_{X|Z})$ bits per instance, and the relay simply forwards the message. For the second region $\mathscr{R}_{0,1}$, the sender first describes X^n to the relay using an expected $H(G_{X|Y})$ bits per instance. The relay, after having reconstructed X^n, describes it to the receiver using an expected $H(G^x_{XY|Z})$ bits per instance, namely using Y^n as side information. Note that $H(G^x_{XY|Z}) \leq H(G_{X|Z})$, where the inequality can be strict, see Example 6. The region $\mathscr{R}_{0,3}$ is obtained as follows. The relay describes Y^n to the receiver using an expected $H(G_{Y|Z})$ bits per instance. The sender describes X^n to the receiver, who is now in possession of both Y^n, Z^n, using an expected $H(G_{X|YZ})$ bits per instance, and the relay forwards the message. The convex hull region $\mathscr{R}_0$ is obtained via simple time sharing. $\square$

4.2.2 Joint protocols

We now turn to discuss general protocols, making better use of the possible correlation between (X, Y, Z). The following observations are central to our discussion

Lemma 1. *(ϕ_1, ϕ_2) satisfies condition (ii) for a one-shot protocol, if and only if $\phi_1(x)$ is a coloring of $G_{X|YZ}$, and $\phi_2(\phi_1(x), y)$ is a coloring of $G^x_{XY|Z}$. Furthermore, for any ϕ_1 satisfying the former requirement, there exists ϕ_2 satisfying the latter.*

---

[5] By $G^x_{XY|Z}$ we mean $G^f_{XY|Z}$ with $f(x, y, z) = x$. Note that $G^x_{XY|Z}$ is obtained from $G_{XY|Z}$ by removing edges connecting vertices with the same x-coordinate.

Proof outline. Condition (ii) holds if and only if for any (x, y) and (x', y') connected in $G^x_{XY|Z}$, $\phi_2(\phi_1(x), y) \neq \phi_2(\phi_1(x'), y')$, i.e., $\phi_2(\phi_1(x), y)$ is a coloring of $G^x_{XY|Z}$. This is possible if and only if the relay can always identify an independent set of $G^x_{XY|Z}$, i.e, the product set $\left\{\phi_1^{-1}(\phi_1(x)) \cap S_X(y)\right\} \times \{y\}$ is an independent set of $G^x_{XY|Z}$, for any $p(x, y) > 0$. The latter holds if and only if $\phi_1(x)$ is a coloring of $G_{X|yZ}$ for any $y \in \mathcal{Y}$, or equivalently, if and only if $\phi_1(x)$ is a coloring of $\cup_y G_{X|yZ} = G_{X|YZ}$. □

Corollary 1. *Any rate pair (R_1, R_2) attainable via a one-shot protocol must satisfy*

$$H(c_1(X)) \leq R_1 \leq H(c_1(X)) + 1$$
$$H(c_2(X, Y)) \leq R_2 \leq H(c_2(X, Y)) + 1$$

for some colorings c_1 and c_2 of $G_{X|YZ}$ and $G^x_{XY|Z}$ respectively, satisfying

$$c_1^{-1}(c_1(x)) \times \{y\} \subseteq c_2^{-1}(c_2(x, y))$$

for all x, y with $p(x, y) > 0$.

Lemma 1 directly applies to n-shot protocols by considering the associated AND-product graphs. For robust n-shot protocols, it must hold *per-instance*, namely:

Corollary 2. *Lemma 1 holds for a robust n-shot protocol, by replacing $G_{X|YZ}$ and $G^x_{XY|Z}$ with their n-fold OR products $\left(G_{X|YZ}\right)^{\vee n}$ and $\left(G^x_{XY|Z}\right)^{\vee n}$, respectively.*

We have the following inner bound.

Theorem 2. *Let $\mathscr{R}_1$ be the closed convex hull of the union of all rate pairs satisfying*

$$R_1 \geq I(X; U), \quad R_2 \geq I(Y; V|U) + \min\{I(X; U), I(U; V)\} \tag{1}$$

Where $X \in U \in \Gamma(G_{X|YZ})$, $\{U \cap S_X(Y)\} \times \{Y\} \in V \in \Gamma(G^x_{XY|Z})$, and $U - X - Y$ and $X - (U, Y) - V$ are Markov chains. Then $\mathscr{R}_1 \subseteq \mathscr{R}$.

Proof. We write that an event A_n occurs *with high probability (w.h.p.)* to indicate that $\lim_{n\to\infty} \mathbb{P}(A_n) = 1$. We use the covering Lemma and the conditional typicality Lemma, see [9]. Randomly pick two mutually independent sets $\{U^n(m)\}_{m=1}^{2^{nr_1}}$ and $\{V^n(m)\}_{m=1}^{2^{nr_2}}$ of mutually independent random vectors, each distributed according to $\prod_{i=1}^n p_U(u_i)$ and $\prod_{i=1}^n p_V(v_i)$, respectively. Given X^n, the sender seeks an index m_1 such that $(X^n, U^n(m_1)) \in \mathcal{T}_\varepsilon^n(X, U)$. If no such index exists, the sender sends a binary representation of X^n with a leading zero, using at most $n \log |\mathcal{X}| + 2$ bits. If such an index m_1 exists, the sender sends a binary representation of m_1 with a leading one, using at most $nr_1 + 2$ bits. By the covering Lemma, the latter case holds w.h.p. if $r_1 > I(X; U)$. The expected sender rate is therefore $R_1 = r_1 + o(1)$ as desired.

The relay obtains either X^n or $U^n(m_1)$, and can distinguish by the leading bit. In the former (low probability) case, the relay simply forwards the message to the receiver, replacing the leading zero with two leading zeros. Consider the latter case. Since $U^n(m_1)$ is jointly typical with X^n, then knowing $U^n(m_1)$ means knowing an independent set of $G_{X|YZ}$ that contains X_k, for each $k \in \{1, \ldots, n\}$. Invoking Lemma 1 per instance, we see that $(U^n(m_1), Y^n)$ uniquely determine an independent set of $G^x_{XY|Z}$ containing X_k for each k, given by $\widetilde{V}_k \stackrel{\text{def}}{=} \{U_k \cap S_X(Y_k)\} \times \{Y_k\}$. Now, the relay seeks an index m_2 such that

$(U^n(m_1), Y^n, V^n((m_2)) \in \mathcal{T}_\varepsilon^n(U, Y, V)$. If no such index exists, the relay sends a binary representation of $\widetilde{V}^n$ with a leading 01, using at most $n(|\mathcal{X}| + \log|\mathcal{X}||\mathcal{Y}|) + 3$ bits. If such an index m_2 exists, the sender sends a binary representation of m_2 with two leading ones, using at most $nr_2 + 3$ bits. By the conditional typicality Lemma, $(U^n(m_1), X^n, Y^n)$ are jointly typical w.h.p. according to $U - X - Y$, and so by the covering Lemma, m_2 exists w.h.p. if $r_2 > I(Y, U; V) = I(Y; V|U) + I(U; V)$. The expected relay rate is therefore $R_2 = r_2 + o(1)$ which yields the expression pertaining to the first element in the minima. The receiver obtains either X^n itself (low probability), or $V^n(m_2)$ which is jointly typical with X^n, hence yields an independent set of $G^x_{XY|Z}$ containing X_k, for each k. Thus, $(V^n(m_2), Z^n)$ uniquely determine X^n. Since R_1, R_2 are achieved when averaged over our random ensemble, there exists a protocol achieving these rates. Note that the protocol is robust since the message describe per-instance independent sets: the relay can arbitrarily replace any erased Y_k, and the receiver just disregards the missing Z_k's.

To get $R_2 \geq I(Y; V|U) + I(X; U)$, the sender-relay protocol is the same. However here we independently pick for *each* $U^n(m_1)$ a set $\{V^n(m_1, m_2)\}_{m_2=1}^{2^{nr_2}}$, according to $\prod_{i=1}^n p_{V|U}(v_i|u_i(m_1))$. By the (conditional) covering Lemma, given m_1 we can find a jointly typical $V^n(m_1, m_2)$ w.h.p. if $r_2 > I(Y; V|U)$. However now, we need to convey $U^n(m_1)$ to the receiver as well, hence $R_2 = r_1 + r_2 + o(1)$. $\square$

Remark 1. Let us show that $\mathscr{R}_0 \subseteq \mathscr{R}_1$. The inclusion can be strict, see Example 5. To obtain $\mathscr{R}_{0,1}$, set U to achieve $H(G_{X|Z})$, and $V = U$. To obtain $\mathscr{R}_{0,2}$, set U to achieve $H(G_{X|Y})$. In this case X is a function of (U, Y), hence we can set V to achieve $H(G^x_{XY|Z})$ while satisfying the Markov chain $V - X - (U, Y)$, yielding $I(Y, U; V) = I(X; V) = H(G^x_{XY|Z})$. To obtain $\mathscr{R}_{0,3}$, set U to achieve $H(G_{X|YZ})$ and V' to achieve $H(G_{Y|Z})$ while satisfying the Markov chain $(X, U) - Y - V'$, and then define $V \stackrel{\text{def}}{=} U \times V'$.

Remark 2. Without loss of generality, V can be restricted to take values in the set of maximal independent sets $\overline{\Gamma}(G^x_{XY|Z})$. However, a similar restriction on U to take values in $\overline{\Gamma}(G_{X|YZ})$ may be suboptimal.

4.2.3 Digression: A covering problem

Let X, U, V be a triplet of r.v's. over a product alphabet $\mathcal{X} \times \mathcal{U} \times \mathcal{V}$. A set of pairs $S = \{(u^n(m), v^n(m) \in \mathcal{U}^n \times \mathcal{V}^n\}_{m=1}^M$ is said to be an (n, ε)-*cover* of X by (U, V) if

$$\mathbb{P}\left(\exists m, (X^n, u^n(m), v^n(m)) \in \mathcal{T}_\varepsilon^n(X, U, V)\right) \geq 1 - \varepsilon$$

A cover S is associated with a rate pair[6]

$$r_1(S) \stackrel{\text{def}}{=} n^{-1} \log \left|\{u^n(m)\}_{m=1}^M\right|, \quad r_2(S) \stackrel{\text{def}}{=} n^{-1} \log \left|\{v^n(m)\}_{m=1}^M\right|$$

A rate pair is called *covering* if for any $\varepsilon > 0$ there exists a (n, ε)-cover of X by (U, V) associated with it, for some large enough n. The *covering rate region* $\mathscr{C}(X|U, V)$ is defined to be the closure of the set of all covering rate pairs.

Problem 1. Determine $\mathscr{C}(X|U, V)$.

[6]Note that we count the number of distinct elements, hence $r_i(S) < n^{-1} \log M$ is possible.

While we do not know the solution to the problem above, we can derive bounds. Define:

$$\mathscr{C}_I^* \stackrel{\text{def}}{=} \{(R_1, R_2) \,:\, \min(R_1, R_2) \geq I(X;U,V)\}$$
$$\mathscr{C}_{II}^* \stackrel{\text{def}}{=} \{(R_1, R_2) \,:\, R_1 \geq I(X;U)\,, R_2 \geq I(X;V)\,, R_1 + R_2 \geq I(X;U) + I(X,U;V)\}$$

and let $\mathscr{C}^*(X|U,V)$ be the closed convex hull of $\mathscr{C}_I^* \cup \mathscr{C}_{II}^*$.

Theorem 3. $\mathscr{C}^*(X|U,V) \subseteq \mathscr{C}(X|U,V)$.

Proof outline. We pick a random cover in two different ways: 1) Jointly with $r_1 = r_2$ according to $\prod_{i=1}^n p_{UV}(u_i, v_i)$, and 2) independently for U, V according $\prod_{i=1}^n p_U(u_i)$ and $\prod_{i=1}^n p_V(v_i)$ respectively. Invoking the multivariate covering lemma [9], we show that for any $\varepsilon > 0$ these random covers are (n, ε)-covers w.h.p. under the constraints of regions $\mathscr{C}_I^*$ and $\mathscr{C}_{II}^*$ respectively. The convex hull is obtained by time-sharing the two strategies. □

The regions $\mathscr{C}_I^*$ and $\mathscr{C}_{II}^*$ do not contain one another in general, as we now exemplify.

Example 1. Suppose that $X - U - V$ forms a Markov chain, $I(X;U|V) > 0$, and $I(X;U) < I(U;V)$. It is easy to show this implies on the one hand that $I(X;V) < I(X;U,V)$ and hence $\mathscr{C}_{II}^* \not\subseteq \mathscr{C}_I^*$, and on the other hand $2I(X;U,V) < I(X;U) + I(X,U;V)$ hence $\mathscr{C}_I^* \not\subseteq \mathscr{C}_{II}^*$. A simple example where these conditions hold is $X \sim \mathrm{Bernoulli}(\frac{1}{2})$ and $U = X + Z_1, V = U + Z_2$ (mod-2 addition), where $Z_i \sim \mathrm{Bernoulli(p_i)}$, $p_2 < p_1 < \frac{1}{2}$, and X, Z_1, Z_2 are mutually independent.

4.2.4 When X knows Y

Let us now consider the special case where the sender knows that side information at the relay. Adapting Therrem 2 to this case, we obtain:

Corollary 3. *Suppose $Y = f(X)$. Then (1) in Theorem 2 reduces to*

$$R_1 \geq I(X;U)\,, R_2 \geq \min\{I(X;U,V), I(X,U;V)\} \tag{2}$$

However, in this case there exists another strategy which can sometimes do better.

Theorem 4. *Suppose $Y = f(X)$. Let $\mathscr{R}_2$ be the closed convex hull of the union of all the covering rate regions of the form $\mathscr{C}(X|U,V)$ where $X \in U \in \Gamma(G_{X|YZ})$ and $\{U \cap S_X(Y)\} \times \{Y\} \in V \in \Gamma(G^x_{XY|Z})$. Then $\mathscr{R}_2 \subseteq \mathscr{R}$.*

Proof outline. The protocol is similar to the first protocol in Theorem 2, with the distinction that here the sender can then simulate the relay hence can find both $U^n(m_1)$ and $V^n(m_2)$ in advance. This naturally puts forward the covering problem in the previous section as an achievable rate region. □

We now apply Theorem 3 to get an explicit bound. It is easy to see that the union over the $\mathscr{C}_I^*$ regions coincides with $\mathscr{R}_{0,1}$,[7] hence we only state the region obtained from $\mathscr{C}_{II}^*$.

Corollary 4. *Let $\mathscr{R}_2^*$ be the closed convex hull of the union of all rate pairs satisfying*

$$R_1 \geq I(X;U)\,, \quad R_2 \geq I(X;V)\,, \quad R_1 + R_2 \geq I(X;U) + I(X,U;V) \tag{3}$$

where $X \in U \in \Gamma(G_{X|YZ})$ and $\{U \cap S_X(Y)\} \times \{Y\} \in V \in \Gamma(G^x_{XY|Z})$. Then $\mathscr{R}_2^* \subseteq \mathscr{R}$.

[7] Loosely speaking, this holds since for $\mathscr{C}_I^*$, the U and V sequences have the same index, hence the relay need not use Y and can simply forward the message.

Remark 3. For any choice of U, V satisfying the conditions in Theorem 2, unless either $X - V - U$ or $U - X - V$ forms a Markov chain, the region (3) includes rate pairs outside the region (2). If also $I(X; U, V) \geq I(X, U; V)$, then the former strictly contains the latter. Furthermore, $\mathscr{R}_2^*$ can contain rate pairs outside $\mathscr{R}_1$, see Example 5. Note that Corollary 4 does not require the Markov condition $X - (U, Y) - V$ to hold.

4.3 An outer bound

Theorem 5. *The following outer bound holds:*

$$\mathscr{R} \subseteq \mathscr{R}_{\text{out}} \stackrel{\text{def}}{=} \{(R_1, R_2) \,:\, R_1 \geq H(G_{X|YZ}),\, R_2 \geq H(G^x_{XY|Z})\}$$

Proof. This is a consequence of Lemma 1, and can be proved along the lines of the converse in [2]. □

In the next Section we discuss simple cases where the inner and outer bounds coincide.

5 Examples

Example 2 (*Degraded receiver*)**.** Suppose that $G_{X|YZ} = G_{X|Y}$. A sufficient (but not necessary) condition to that end is that $X - Y - Z$ forms a Markov chain. Then $\mathscr{R}_{0,2} = \mathscr{R}_{\text{out}}$, yielding the exact rate region. One extreme case is where $X = f(Y)$ and $Z = g(Y)$. In this case the rate region is given by $\{R_1 \geq 0, R_2 \geq H(X|Z)\}$. This is also the optimal region under a vanishing error probability criterion. It can be derived from $\mathscr{R}_{0,2}$, or more directly by noticing that one can simply condition on Z at the relay. Note that if $G_{X|Z}$ is a full graph, then communicating X directly to the receiver requires a rate of $H(X)$ while when communicating through the relay a (possibly much lower) sum-rate of $H(X|Z)$ is sufficient. This should be contrasted with the vanishing error vanishing error probability case where the so-called *cutset bound* holds, i.e., the relay-receiver rate cannot be smaller than the optimal point-to-point sender-receiver rate [5].

Example 3 (*Degraded relay*)**.** Suppose $X - Z - Y$ forms a Markov chain and either $Z = f(X)$, or $S_Y(z) = S_Y$ for any $z \in S_Z$ (these conditions can be further refined). Then $G_{X|YZ} = G_{X|Z}$, and furthermore, if x, x' are connected (resp. not connected) in $G_{X|Z}$ then $(x, y), (x', y')$ are connected (resp. not connected) in $G^x_{XY|Z}$ for any feasible y, y'. It is readily verified this implies $H(G^x_{XY|Z}) = H(G_{X|Z})$, hence $\mathscr{R}_{0,1} = \mathscr{R}_{\text{out}}$, yielding the exact rate region.

Example 4. Suppose $Y - X - Z$ constitutes a Markov chain, and either $G_{X|Y} \subseteq G_{X|Z}$ or $G_{X|Z} \subseteq G_{X|Y}$. By Markovity, if x, x' are connected (resp. not connected) in $G_{X|Z}$ then $(x, y), (x', y')$ are connected (resp. not connected) in $G^x_{XY|Z}$ for any feasible y, y', hence $H(G^x_{XY|Z}) = H(G_{X|Z})$. Furthermore, Markovity implies that $G_{X|YZ} = G_{X|Y} \cap G_{X|Z}$. Therefore $\mathscr{R}_{0,1} \cup \mathscr{R}_{0,2} = \mathscr{R}_{\text{out}}$, yielding the exact rate region.

Example 5. Let X be uniformly distributed over a quaternary alphabet $\mathcal{X} = \{0, 1, 2, 3\}$. Let $Y = \min(X, 1)$ and $Z = \max(X, 2)$. The graph $G_{X|Y}$ has the edge set $\{(1, 2), (1, 3), (2, 3)\}$, and the graph $G_{X|Z}$ has the edge set $\{(0, 1), (0, 2), (1, 2)\}$.[8] The graph $G_{X|YZ}$ has a single edge $(1, 2)$. The graph $G_{Y|Z}$ is a full graph on two vertices. Finally, since Y is a function of X, then the $G^x_{XY|Z}$ is isomorphic to $G_{X|Z}$.

[8]Note that this setting satisfies the first condition in Example 4, but not the second.

Note that $\overline{\Gamma}(G_{X|Y}) = \{\{0,1\},\{0,2\},\{0,3\}\}$. It is readily verified that $H(G_{X|Y})$ is attained by setting $U = \{0, X\}$ given $X \neq 0$, and choosing U uniformly at random over $\overline{\Gamma}(G_{X|Y})$ given $X = 0$. This yields

$$H(G_{X|Y}) = I(X;U) = H(U) - H(U|X) = \log 3 - (1/4)\log 3 = (3/4)\log 3$$

Similarly, $H(G_{X|Z}) = H(G^x_{XY|Z}) = (3/4)\log 3$. Also clearly, $H(G_{Y|Z}) = H(Y) = H(1/4, 3/4) = 2 - (3/4)\log 3$. Now, $\overline{\Gamma}(G_{X|YZ}) = \{\{0,1,3\},\{0,2,3\}\}$, and it is readily verified that $H(G_{X|YZ})$ is attained by setting $U = \{0, X, 3\}$ given $X \in \{1,2\}$, and then choosing U uniformly at random over $\overline{\Gamma}(G_{X|YZ})$ given $X \in \{0,3\}$. This yields

$$H(G_{X|YZ}) = I(X;U) = H(U) - H(U|X) = 1 - 1/2 = 1/2$$

The naive inner bounds are thus given by

$$\mathscr{R}_{0,1} = \mathscr{R}_{0,2} = \{(R_1,R_2) : R_1 \geq (3/4)\log 3,\ R_2 \geq (3/4)\log 3\}$$
$$\mathscr{R}_{0,3} = \{(R_1,R_2) : R_1 \geq 1/2,\ R_2 \geq 5/2 - (3/4)\log 3\}$$

and the outer bound reads:

$$\mathscr{R} \subseteq \mathscr{R}_{\text{out}} = \{(R_1,R_2) : R_1 \geq 1/2,\ R_2 \geq (3/4)\log 3\}$$

Let us now consider the bound of Theorem 4. Set $U = \{0, X, 3\}$ given $X \in \{1,2\}$, then distributed over $\{\{0,1,3\},\{0,2,3\},\{0,3\}\}$ with probabilities $\{\frac{4}{9}, \frac{4}{9}, \frac{1}{9}\}$ respectively given $X = 0$, and chosen uniformly at random over the same set given $X = 3$. It is easy to check that

$$I(X;U) = (3/4)H\left(4/9, 4/9, 1/9\right) - (1/4)\log 3 = (5/4)\log 3 - 4/3$$

Furthermore, $U \cap S_X(Y)$ takes values in $\{\{0\},\{3\},\{1,3\},\{2,3\}\}$ with probabilities $\{\frac{1}{4}, \frac{1}{12}, \frac{1}{3}, \frac{1}{3}\}$. Set $V = (U \cap S_X(Y)) \times \{Y\}$ if $U \cap S_X(Y) \in \{\{1,3\},\{2,3\}\}$, and $V = \{0,3\} \times \{Y\}$ otherwise. This yields a V that achieves the graph entropy of $G^x_{XY|Z}$. i.e.,

$$I(X;V) = H(V) - H(V|X) = (3/4)\log 3$$

furthermore, $I(U;V|X) = H(V|X) = (1/4)\log 3$. Therefore the bound of Corollary 4 yields:

$$R_1 \geq (5/4)\log 3 - 4/3\,, \quad R_2 \geq (3/4)\log 3\,, \quad R_1 + R_2 \geq (9/4)\log 3 - 4/3$$

This region contains rate pairs strictly outside the convexified naive region $\mathscr{R}_0$, e.g., the rate pair $R_1 = (3/2)\log 3 - 4/3$, $R_2 = (3/4)\log 3$. In fact, it can be verified that this specific rate pair is also outside $\mathscr{R}_1$.

Example 6. Let $\mathcal{X} = \mathcal{Y} = \{0,1,2\}$ and $\mathcal{Z} = \{0,1,2,3\}$. Let $S_{XYZ} = \{000, 001, 010, 013, 110, 121, 122, 202, 203\}$, where $p(x,y,z) = 1/10$ for all triplets in S_{XYZ} with the exception of $p(110) = 1/5$. This results in X having a distribution with values $(2/5, 2/5, 1/5)$ over $\mathcal{X}$. $G_{X|Z}$ is a full graph hence $H(G_{X|Z}) = H(X) \approx 1.522$. Both $G_{X|Y}$ and $G_{Y|Z}$ have an edge set $\{(0,1),(0,2)\}$, hence $H(G_{X|Y}) = H(G_{Y|Z}) = H(2/5, 3/5) \approx 0.971$. $G_{X|YZ}$ has a single edge $(0,1)$, and $H(G_{X|YZ}) = 1 - (1/5)H(1/2,1/2) = 4/5$. This is achieved by letting $U = (X, 2)$ given $X \in \{0,1\}$, and choosing U uniformly over $\{(0,2),(1,2)\}$ given $X = 2$. The graph $G^x_{XY|Z}$ has a vertex set $\{00, 11, 01, 20, 12\}$, is a 5-cycle in that order, and is associated with a uniform distribution. It can be shown that its graph entropy is obtained by drawing U uniformly at random over the two possible maximal independent sets given each vertex, which yields $H(G^x_{XY|Z}) = \log(5/2) \approx 1.322$. Therefore, the naive inner bounds are

$$\begin{aligned}\mathscr{R}_{0,1} &= \{(R_1, R_2) : R_1 \geq 1.552,\ R_2 \geq 1.552\}\\ \mathscr{R}_{0,2} &= \{(R_1, R_2) : R_1 \geq 0.97,\quad R_2 \geq 1.322\}\\ \mathscr{R}_{0,3} &= \{(R_1, R_2) : R_1 \geq 0.8,\qquad R_2 \geq 1.771\}\end{aligned}$$

Consider now the bound in Theorem 2. Set U to achieve $H(G_{X|YZ})$ as above. Restrict V to take values in $\{(00,20),(11,20),(01,12)\}$. It is easy to check that V is deterministically determined from (U,Y), except when $U = (1,2)$ and $Y = 0$. In this latter case, choose V uniformly at random over $\{(00,20),(11,20)\}$. This results in V having a distribution with values $(7/20, 1/4, 2/5)$ over its alphabet, respectively. Hence $H(V) \approx 1.559$, and $H(V|U,Y) = 1/10$. Appealing to Theorem 2, we obtain the region:

$$\{(R_1, R_2) : R_1 \geq 0.8,\ R_2 \geq 1.459\}$$

Therefore, $\mathscr{R}_1$ strictly contains $\mathscr{R}_0$.

Acknowledgements

The vanishing error version of this problem was introduced to the author by Haim Permuter. Helpful discussions with Young-Han Kim, Amir Leshem, Alon Orlitsky and Haim Permuter, as well as useful comments by the reviewers, are greatly appreciated.

References

[1] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. on Inform. Theory*, vol. 19, no. 4, pp. 471 – 480, July 1973.

[2] N. Alon and A. Orlitsky, "Source coding and graph entropies," *IEEE Trans. on Inform. Theory*, vol. 42, no. 5, pp. 1329–1339, Sep 1996.

[3] P. Koulgi, E. Tuncel, S.L. Regunathan, and K. Rose, "On zero-error source coding with decoder side information," *IEEE Trans. on Inform. Theory*, vol. 49, no. 1, pp. 99 – 111, Jan. 2003.

[4] K. Marton, "On the shannon capacity of probabilistic graphs," *J. Comb. Theory, Ser. B*, vol. 57, no. 2, pp. 183–195, Mar. 1993.

[5] P. Ishwar and S. S. Pradhan, "A relay-assisted distributed source coding problem," in *Information Theory and Applications Workshop*, 2008, pp. 136–141.

[6] A. Orlitsky and J.R. Roche, "Coding for computing," *IEEE Trans. on Inform. Theory*, vol. 47, no. 3, pp. 903 –917, Mar. 2001.

[7] J. Körner, "Coding of an information source having ambiguous alphabet and the entropy of graphs," in *6th Prague Conf on Information Theory*, 1973, pp. 411–425.

[8] G. Simonyi, "Graph entropy: A survey," *in DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 20, pp. 399 – 441, 1995.

[9] A. El Gamal and Y-.H Kim, *Lecture Notes on Network Information Theory*, Stanford University and UCSD, 2009.