Simple Bounds for Lossless Source Coding in A Two-Hop Network

WeiHsin Gu
Department of Electrical Engineering
California Institute of Technology
Pasadena, California 91125 USA
Email: whgu@z.caltech.edu

Michelle Effros
Department of Electrical Engineering
California Institute of Technology
Pasadena, California 91125 USA
Email: effros@caltech.edu

***Abstract*— We describe new upper bounds on the lower convex hull of the lossless source coding region for a simple three-node network. While the given bound may not be tight, its advantage lies in its simplicity. Precisely, it allows us to easily calculate meaningful numerical bounds for all possible source distributions. This property contrasts markedly with earlier results where solutions characterized in terms of auxiliary random variables make it difficult to quantify results for a given distribution. The given result represents a first step in developing techniques for finding loose, calculable bounds for network source coding performance. Such techniques are likely to play a critical role in the advancement of this difficult field.**

I. Introduction

We seek concrete rate bounds on the lossless source coding region for the simple, two-hop network shown in Figure 1. Node 1 employs encoder α_1 to describe dependent sources X and Y at rate R_1. Node 2 uses decoder β_2 to map its received description to a reconstruction of Y; it also uses its encoder α_2 to map the received description at rate R_1 into a transmitted description at rate R_2. Node 3 employs decoder β_3 to build a reconstruction of X from its received rate-R_2 description and side information Z.

The lossless source coding region for the given network was partially characterized in [1], where the following upper bound on the lower convex hull was derived. A matching converse has not been proven to date.

Theorem 1 ([1, Corollary 1.2]): Consider i.i.d. samples of (X, Y, Z) from probability mass function $p(x, y, z)$. If nodes 1 and 2 output descriptions of rates R_1 and R_2 and the decoder combines these descriptions with side information Z, then we can recover X with arbitrarily low error probability if

$$\begin{aligned} R_X &\geq I(X, Y; U) + H(Y|U) + H(X|Z, U) \\ R_Y &\geq I(X, Y; U) + H(X|Z, U) \end{aligned}$$

for some joint probability mass function $p(x, y, z)p(u|x, y)$ where U is an auxiliary random variable with alphabet size $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{Y}| + 2$.

The given bound is intuitively reasonable. Auxiliary random variable U describes the information used by both decoders, and we describe the useful portion of its randomness at rate $I(X, Y; U)$. Rates $H(Y|U)$ and $H(X|Z, U)$ then suffice to describe the remaining uncertainty in both Y and X to decoders that receive U and (U, Z), respectively.

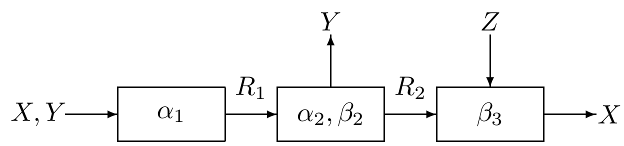

Fig. 1. A simple, two-hop network.

Despite its intuitive form, the given random bound is difficult to work with. Solution of the bound for any fixed $p(x, y, z)$ requires solution of a non-trivial optimization problem aimed at finding the family of auxiliary random variables U corresponding to the lower convex hull of the achievable rate region. The number of distinct probability mass functions $p(u|x, y)$ required to describe this lower convex hull is unknown. Further, the alphabet size for each distinct U may apparently be as large as $|\mathcal{X}||\mathcal{Y}| + 2$.

In this paper, we take a step back from the precise characterization of lossless source coding regions and focus instead on simple, calculable bounds. Using the two-hop network as an example, we break the given network down into a family of simpler source coding problems for which easily calculated bounds are already available. The resulting upper and lower bounds on the achievable rate region of the given network are simply characterized solely in terms of the source distributions.

II. Results

We begin our investigation of the two-hop network by simplifying the problem down to its bare essence. Since node 2 must lossless reconstruct Y without the help of side information, solving the problem at hand reduces to solving the network source coding problem shown in Figure 2. Here node 1 employs encoder α_1' to map observed sources X and Y to a binary description at rate R_1'. Node 2 uses encoder α_2' to map its received description at rate R_1' and its observed side information Y to a description at rate R_2'. Node 3 uses decoder β_3' to map its received description at rate-R_2' and its observed side information Z to build a reconstruction for source X. An immediate lower bound on the achievable rate region for this network is $R_1' \geq H(X|Y, Z)$ and $R_2 \geq H(X|Z)$. We find a simple achievable rate region by comparison with the network shown in Figure 3.

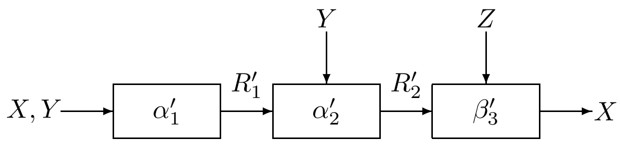

Fig. 2. Multihop source coding with side information.

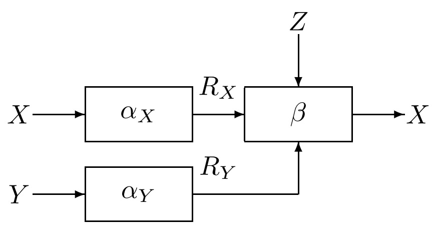

Fig. 3. Source coding with coded and uncoded side information at the decoder.

Figure 3 is a variation of the source coding with coded side information problem investigated by [2]. Here node 1 observes only X and transmits a description of that source at rate R_X. Node 2 likewise observes only Y and transmits a description of that source at rate R_Y. Node 3 receives both descriptions as well as side information Z and builds a reconstruction only of source Z. The given network differs from that of [2] only in its inclusion of uncoded side information in addition to coded side information. Theorem 2 fully characterizes the lossless source coding region for this network.

Theorem 2: Consider i.i.d. samples of (X, Y, Z) from probability mass function $p(x, y, z)$. If X and Y are independently encoded at rates R_X and R_Y and the decoder combines these descriptions with side information Z, then we can recover X with arbitrarily low error probability if and only if

$$\begin{aligned} R_X &\geq H(X|U, Z) \\ R_Y &\geq I(Y; U|Z) \end{aligned}$$

for some joint probability mass function $p(x, y, z)p(u|y)$ for some auxiliary random variable U with alphabet size $|\mathcal{U}| \leq |\mathcal{Y}| + 2$.

Proof: For the achievability result, fix a distribution $p(u|y)$. For each $x^n \in \mathcal{X}^n$, draw $\alpha_X(x^n)$ independently and uniformly over $\{1, \ldots, 2^{nR_X}\}$. Draw $U^n(1), U^n(2), \ldots, U^n(2^{nR'})$ i.i.d. from distribution $\prod_{i=1}^n p(u_i)$. Color each $s \in \{1, \ldots, 2^{nR'}\}$ a color $c(s)$ chosen independently and uniformly at random from colors $\{1, \ldots, 2^{nR_Y}\}$. For each $y^n \in \mathcal{Y}^n$, let $a(y^n)$ be the smallest index s such that $(y^n, U^n(s)) \in A_\epsilon^{*(n)}(Y, U)$, and let $\alpha_Y(y^n) = c(a(y^n))$. If y^n does not find a jointly typical U^n vector, then $a(y^n)$ is undefined and $\alpha_Y(y^n)$ is chosen arbitrarily on $\{1, \ldots, 2^{nR_Y}\}$.

For each $i_X \in \{1, \ldots, 2^{nR_X}\}$ and $i_Y \in \{1, \ldots, 2^{nR_Y}\}$, let $\sigma(i_Y, z^n)$ be the unique $s \in \{1, \ldots, 2^{nR_Y}\}$ for which $c(s) = i_Y$ and $(U^n(s), z^n) \in A_\epsilon^{*(n)}(U, Z)$, and let $\beta(i_X, i_Y, z^n)$ be the unique $x^n \in \mathcal{X}^n$ for which $\alpha_X(x^n) = i_X$ and $(x^n, z^n, U^n(\sigma(i_Y, z^n))) \in A_\epsilon^{*(n)}(X, Z, U)$. In either case, an error occurs if no unique solution exists.

The possible error events are

$$\begin{aligned} E_0 &= \{(X^n, Y^n, Z^n) \notin A_\epsilon^{*(n)}(X, Y, Z)\} \\ E_1 &= \{(Y^n, U^n(s)) \notin A_\epsilon^{*(n)}(Y, U) \forall s\} \\ E_2 &= \{(Z^n, U^n(a(Y^n))) \notin A_\epsilon^{*(n)}(Z, U)\} \\ E_3 &= \{\exists \hat{s} \neq a(Y^n) \text{ s.t. } c(\hat{s}) = c(a(Y^n)), \\ &\quad (Z^n, U^n(\hat{s})) \in A_\epsilon^{*(n)}(Z, U)\} \\ E_4 &= \{\exists \hat{x}^n \neq x^n \text{ s.t. } f_X(\hat{x}^n) = f_X(x^n), \\ &\quad (\hat{x}^n, Z^n, U^n(\sigma(i_Y, Z^n))) \in A_\epsilon^{*(n)}(X, Z, U)\}, \end{aligned}$$

giving an expected error probability

$$EP_e^{(n)} \leq E\left[\sum_{i=0}^{4} \Pr(E_i | E_0^c, \ldots, E_{i-1}^c)\right].$$

The first term can be made arbitrarily small by the law of large numbers. The second term becomes small provided $R' > I(Y : U)$ by [3, Lemma 13.6.2]; we set $R' = I(Y; U) + \epsilon$. The third term goes to zero by [3, Lemma 14.8.1] since $(X, Z) \to Y \to U$ forms a Markov chain. Again using [3, Lemma 13.6.2] which bounds $\Pr((Z^n, U^n(s)) \in A_\epsilon^{*(n)}(Z, U))$ from above and below for every s, We bound the fourth term as

$$\begin{aligned} & E \Pr(E_3 | E_0^c, E_1^c, E_2^c) \\ &\leq \sum_{\hat{s} \neq a(Y^n)} \Pr(c(\hat{s}) = c(a(Y^n))) \\ &\quad \cdot \Pr((Z^n, U^n(\hat{s})) \in A_\epsilon^{*(n)}(Z, U)) \\ &\leq 2^{n(I(Y;U)+\epsilon)} 2^{-nR_Y} 2^{-n(I(U;Z)-\epsilon)} \\ &= 2^{-n(R_Y - (I(U;Y) - I(U;Z) + 2\epsilon))}, \end{aligned}$$

which goes to zero provided that $R_Y > I(U; Y) - I(U; Z) + 2\epsilon$. We use the Markov condition to find $I(U; Y) - I(U; Z) = I(U; Y, Z) - I(U; Z) = I(U; Y|Z)$. Finally,

$$\begin{aligned} & E \Pr(E_4 | E_3^c, E_2^c, E_1^c, E_0^c) \\ &\leq \sum_{\hat{x}^n \in A_\epsilon^{*(n)}(X)} \Pr(\alpha_X(\hat{x}^n) = \alpha_X(x^n)) \\ &\quad \cdot \Pr((\hat{x}^n, Z^n, U^n(\sigma(i_Y, Z^n))) \in A_\epsilon^{*(n)}(X, Z, U)) \\ &\leq 2^{n(H(X)+\epsilon)} 2^{-nR_X} 2^{-n(I(X;Z,U)+3\epsilon)} \\ &= 2^{-n(R_X - (H(X|U,Z)+4\epsilon))}, \end{aligned}$$

which goes to zero provide $R_X > H(X|U, Z) + 4\epsilon$.

For the converse, let $S = \alpha_X(X^n)$ and $T = \alpha_Y(Y^n)$. Setting $U_i = (T, Y^{i-1}, Z^{i-1}, Z_{i+1}^n)$ gives

$$\begin{aligned} nR_Y &\geq H(T|Z^n) \geq I(Y^n; T|Z^n) \\ &= \sum_{i=1}^{n} I(Y_i; T, Y^{i-1}, Z^{i-1}, Z_{i+1}^n | Z_i) \\ &= \sum_{i=1}^{n} I(Y_i; U_i | Z_i) \end{aligned}$$

Similarly

$$\begin{aligned} nR_X &\geq H(S|T,Z^n) \\ &\geq I(X^n;S|T,Z^n) = H(X^n|T,Z^n) - H(X^n|S,T,Z^n) \\ &\geq \sum_{i=1}^n H(X_i|T,X^{i-1},Z^n) - n\epsilon_n \\ &\geq \sum_{i=1}^n H(X_i|U_i,X^{i-1},Z_i) - n\epsilon_n \\ &= \sum_{i=1}^n H(X_i|U_i,Z_i) - n\epsilon_n \end{aligned}$$

where the third inequality is Fano's inequality, the fourth inequality follows from the definition of U_i since conditioning reduces entropy, and the final line follows since $X_i \to (U_i, Z_i) \to X^{i-1}$ forms a Markov chain. □

While the rate region of Theorem 2 completely characterizes the full set of achievable rates in the given network, precise description of the given rate region for a fixed probability mass function $p(x, y, z)$ requires solution of a non-trivial optimization problem. Precisely, solving the region exactly requires finding the optimal family of auxiliary random variables for the given distribution. A partial solution to that problem for the original problem of Ahlswede and Körner appears in [4]. Generalizing that result to the case with both coded and uncoded side information gives the upper and lower bounds shown in Figure 4. The bounds match for some but not all probability mass functions, and the shaded portion indicates where the bounds may not match. The critical points $J(X;Y|Z)$ and $K(X;Y|Z)$ shown in that picture are found by applying the constructions of auxiliary random variable U used in the original coding with side information problem. We require some terminology to describe those values precisely.

Let $\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_K \subset \mathcal{Y}$ be the largest possible non-overlapping subsets of alphabet $\mathcal{Y}$ for which for every $x \in \mathcal{X}$, $p(x|y) = p(x|y')$ for all $y, y' \in \mathcal{Y}_K$. Adopting the terminology from [4], we call each $\mathcal{Y}_i$ a zero information component. If $\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_K \subset \mathcal{Y}$ are the zero information components for probability mass function $p(x, y, z)$, then we define $\mathcal{U}_1 = \{u_1, \ldots, u_K\}$ and define conditional probability mass function $p_1(u|y)$ as $p_1(u_k|y) = 1$ if $y \in \mathcal{Y}_k$ and $p_1(u_k|y) = 0$ otherwise. Then

$$H(X|U,Z)|_{p(x,y,z)p_1(u|y)} = H(X|Y,Z),$$

and we define

$$J(X;Y|Z) = I(Y;U|Z)|_{p(x,y,z)p_1(u|y)}.$$

The point $(R_X, R_Y) = (H(X|Y,Z), J(X;Y|Z))$ is achievable by Theorem 2.

From the given zero information components $\mathcal{Y}_1, \ldots, \mathcal{Y}_K$, we build larger disjoint components $\hat{\mathcal{Y}}_1, \ldots, \hat{\mathcal{Y}}_L$, where each $\hat{\mathcal{Y}}_i$ is a union of zero information components. Two zero information components $\mathcal{Y}_i$ and $\mathcal{Y}_j$ are contained in the same disjoint component if and only if there exists some $x \in \mathcal{X}$ such that $p(x|y_i)p(x|y_j) > 0$ for $y_i \in \mathcal{Y}_i$ and $y_j \in \mathcal{Y}_j$. If

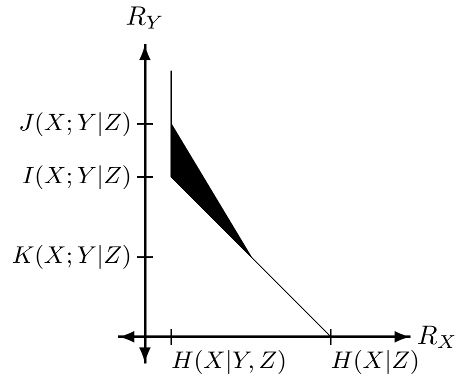

Fig. 4. Bounds on the achievable rate region for source coding with both coded and uncoded side information. The given picture shows upper and lower bounds on the achievable rate region. The shaded region shows where those bounds do not match.

$\hat{\mathcal{Y}}_1, \hat{\mathcal{Y}}_2, \ldots, \hat{\mathcal{Y}}_K \subset \mathcal{Y}$ are the disjoint components for probability mass function $p(x, y, z)$, then we define $\mathcal{U}_2 = \{u_1, \ldots, u_L\}$ and define conditional probability mass function $p_2(u|y)$ as $p_2(u_\ell|y) = 1$ if $y \in \hat{\mathcal{Y}}_\ell$ and $p_2(u_\ell|y) = 0$ otherwise. Then we define

$$K(X;Y|Z) = I(Y;U|Z)|_{p(x,y,z)p_2(u|y)}.$$

Note that

$$\begin{aligned} &[I(Y;U|Z) + H(X|U,Z)]_{p(x,y,z)p_2(u|y)} \\ &\quad= [H(U|Z) + H(X|U,Z) - H(U|Y,Z)]_{p(x,y,z)p_2(u|y)} \\ &\quad= [H(X|Z) + H(U|X,Z) - H(U|Y,Z)]_{p(x,y,z)p_2(u|y)} \\ &\quad= H(X|Z) \end{aligned}$$

since U is a deterministic function of X and U is a deterministic function of Y for the given distribution. Thus $(R_X, R_Y) = (H(X|Z) - K(X;Y|Z), K(X;Y|Z))$ is achievable by Theorem 2, and further the given point is on the lower convex hull of the lossless source coding region since $R_X + R_Y \geq H(X|Z)$ is by the Slepian-Wolf theorem.

Finally, the achievability of the point $(R_X, R_Y) = (H(X|Z), 0)$ is immediate from the Slepian-Wolf theorem. (Setting U equal to a constant in Theorem 2 achieves the same end.)

Together, the points $(H(X|Y,Z), J(X;Y|Z))$, $(H(X|Z) - K(X;Y|Z), K(X;Y|Z))$, and $(H(X|Z), 0)$ fully characterize our upper bound on the lower convex hull of the lossless source coding region. Our lower bound results from the earlier observation that $R_X + R_Y \geq H(X|Z)$ and the cut-set bound $R_X \geq H(X|Y,Z)$. The bounds are tight when $J(X;Y|Z) = K(X;Y|Z)$, which occurs when each disjoint component $\hat{\mathcal{Y}}_\ell$ is a zero information component.

To apply the achievable rate region from the coded and uncoded side information problem in the network of Figure 2, we define

$$\begin{aligned} \alpha_1'(X^n, Y^n) &= \alpha_X(X^n) \\ \alpha_2'(\alpha_1', Y^n) &= \alpha_X(X^n), \alpha_Y(Y^n) \\ \beta_3'(\alpha_1', \alpha_2', Z^n) &= \beta(\alpha_X(X^n), \alpha_Y(Y^n), Z^n) \end{aligned}$$

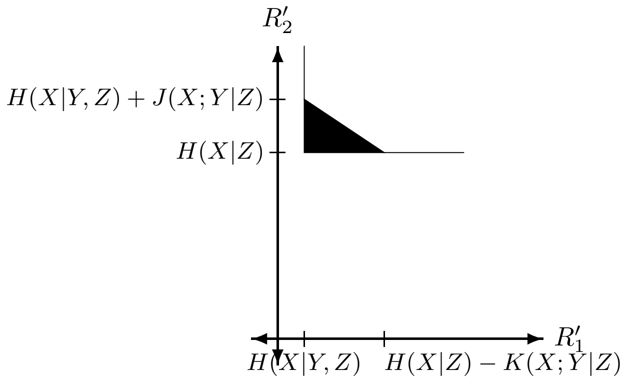

Fig. 5. Bounds on the achievable rate region for the single-source two-hop network of Figure 2. The given picture shows upper and lower bounds on the achievable rate region. The shaded region shows where those bounds do not match for some sources.

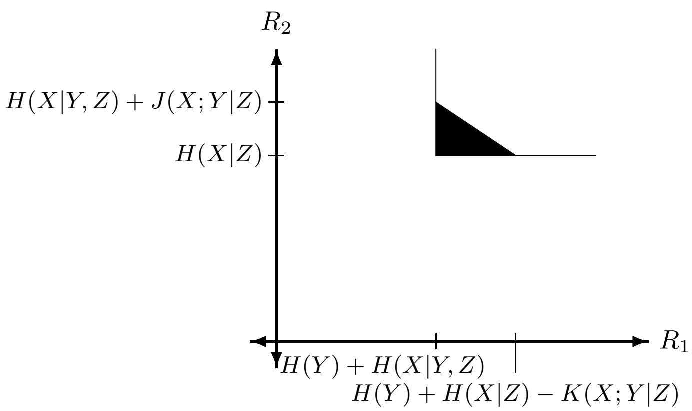

Fig. 6. Upper and lower bounds on the lower convex hull of the achievable rate region for source coding in the original two-hop network of Figure 1 using the distribution of Example 1. The shaded region shows where those bounds do not match for some sources.

That is, node 1 describes X at rate $H(X|U,Z)$. Node 2 appends to this a coded side information description of Y. Node 3 uses the decoder from the coded and uncoded side information problem to build a reconstruction of X^n from its received pair of descriptions. Combining this construction with our rate region bounds gives an upper bound on the lower convex hull of the achievable rate region, as shown in Figure 5. We achieve lower bounds on the rate region using a pair of simple cutset bounds: $R_1' \geq H(X|Y,Z)$ and $R_2' \geq H(X|Z)$.

Finally, we bound the rate region for our original two-hop network by adding to R_1' the cost of describing Y using a lossless entropy code. The resulting source coding bounds demonstrate that rate vector (R_1, R_2) is achievable if

$$\begin{aligned} R_1 &\geq H(Y)+H(X|U,Z)+I(Y;U|Z) \\ R_2 &\geq H(X|U,Z)+I(Y;U|Z) \end{aligned}$$

for some $p(u|y)$. The resulting non-parametric bounds appear in Figure 6.

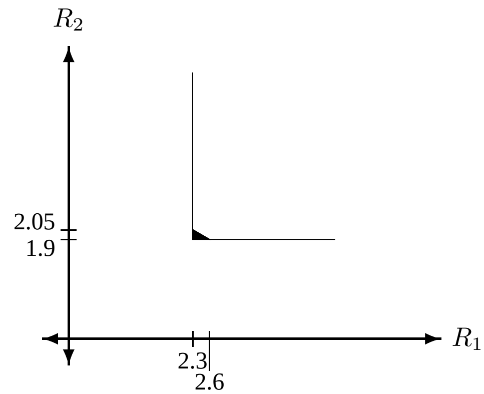

Fig. 7. Bounds on the achievable rate region for source coding in the original two-hop network of Figure 1. The given picture shows upper and lower bounds on the achievable rate region. The shaded region shows where those bounds do not match.

Example: Let $\mathcal{X} = \mathcal{Y} = \{1,2,3,4\}$, and suppose that

$$p(x|y) = \begin{cases} 0.1 & \text{if } (x,y) \in \{(1,1),(2,1),(3,4),(4,3)\} \\ 0.9 & \text{if } (x,y) \in \{(1,2),(2,2),(3,3),(4,4)\} \\ 0 & \text{otherwise.} \end{cases}$$

Further, let

$$p(z|x,y) = \begin{cases} 0.2 & \text{if } (x=y,\ z=1) \text{ or } (x \neq y,\ z=0) \\ 0.8 & \text{if } (x=y,\ z=0) \text{ or } (x \neq y,\ z=1) \end{cases}$$

III. Conclusions

We derive simple, calculable bounds on the achievable rate region for a two-hop network with side information. While the given region is not tight in general, it moves bounding the lossless achievable rate region from complex optimization to simple calculation. Since single-letter descriptions of lossless rate regions are elusive for many networks and solution of those bounds is often difficult even when single-letter solutions exist, we believe that systematic techniques for finding loose, calculable bounds will play a critical role for advancing the field of lossless network source coding. The result presented here are a first step in that direction.

Acknowledgment

This material is based upon work partially supported by NSF Grant No. CCR-0220039 and Caltech's Lee Center for Advanced Networking.

References

[1] W. Gu and M. Effros. Source coding for a relay network. In *Proceedings of the IEEE International Symposium on Information Theory*, Adelaide, Australia, September 2005.

[2] R. Ahlswede and J. Körner. Source coding with side information and a converse for degraded broadcast channels. *IEEE Transactions on Information Theory*, IT-21(6):629–637, November 1975.

[3] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 1991.

[4] D. Marco and M. Effros. A partial solution for lossless source coding with coded side information. In *Proceedings of the Information Theory Workshop*, Punta del Este, Uruguay, March 2006. IEEE. Invited paper. To appear.