

# On Source Coding with Coded Side Information for a Binary Source with Binary Side Information

WeiHsin Gu<sup>†</sup>, Ralf Koetter<sup>‡</sup>, Michelle Effros<sup>†</sup>, Tracey Ho<sup>†</sup>

<sup>†</sup>Department of Electrical Engineering  
California Institute of Technology  
Pasadena, California, CA 91125, USA  
Email: {wgu, effros, tho}@caltech.edu

<sup>‡</sup>Institute for Communications Engineering  
Technische Universitaet Muenchen  
D-80290 Muenchen, Germany  
Email: ralf.koetter@tum.de

**Abstract**— The lossless rate region for the coded side information problem is “solved,” but its solution is expressed in terms of an auxiliary random variable. As a result, finding the rate region for any fixed example requires an optimization over a family of allowed auxiliary random variables. While intuitive constructions are easy to come by and optimal solutions are known under some special conditions, proving the optimal solution is surprisingly difficult even for examples as basic as a binary source with binary side information. We derive the optimal auxiliary random variables and corresponding achievable rate regions for a family of problems where both the source and side information are binary. Our solution involves first tightening known bounds on the alphabet size of the auxiliary random variable and then optimizing the auxiliary random variable subject to this constraint. The technique used to tighten the bound on the alphabet size applies to a variety of problems beyond the one studied here.

## I. INTRODUCTION

Generalizing our understanding of the source coding problem from point-to-point communication systems to general networks remains a central underlying goal of source coding research. The problem of source coding with coded side information, perhaps one of the most basic components of network source coding systems, is an important stepping stone in this endeavor. The problem was introduced and solved by Ahlswede and Körner in [1]. Their achievable rate region describes the family of rate vectors  $(R_X, R_Y)$  such that independently describing source  $X$  at rate  $R_X$  and side information  $Y$  at rate  $R_Y$  allows the decoder to reconstruct  $X$  with asymptotically negligible error probability. (See Fig. 1.)

While the characterization given by Ahlswede and Körner is tight, it does not tell the full story. The given solution relies on an unknown auxiliary random variable. Thus numerically characterizing the achievable rate region for any joint distribution on  $(X, Y)$  requires an additional optimization over all admissible auxiliary random variables  $U$ .

Solution of the optimal auxiliary random variable is studied in [2]. The central results of that work are answers to two questions. What is the minimal achievable rate  $R_Y$  when  $R_X = H(X|Y)$ ? What is the maximal rate  $R_Y$  at which

$R_X + R_Y = H(X)$  is achievable? While the answers to these questions allow us to precisely characterize the achievable rate region in the special case where the answer to both questions is  $R_Y = I(X; Y)$  and to bound the achievable rate region more generally, it, too, fails to tell the full story. For example, when  $X$  and  $Y$  are uniformly distributed binary random variables related through a binary symmetric channel, the answer to the first question is  $R_Y = H(Y)$  and the answer to the second question is  $R_Y = 0$ , and the results of [2] tell us very little about the achievable rate region.

The remainder of this paper begins with background on the coded side information problem. We then provide solutions to a family of coded side information problems where both source and side information are binary. As a first step, we tighten the bound on the alphabet size of the auxiliary random variable from  $|\mathcal{Y}|+2$  to  $|\mathcal{Y}|$ . The technique used to improve this bound also applies to a variety of other problems, including those in [3], [4], [5], [6], [7]. We then derive the optimal  $U$  and corresponding optimal rate region for a variety of examples where  $X$  and  $Y$  are binary;  $|U| \leq 2$  in these examples by our first result. In Section IV, we prove that if the conditional distribution of  $X$  given  $Y$  is a binary symmetric channel, then  $U$  is optimal if and only if  $U$  and  $Y$  are related through a binary symmetric channel as well. (See Fig. 2.) In Section V, we show that if the conditional distribution of  $X$  given  $Y$  is a  $Z$ -channel, then  $U$  is optimal if and only if  $U$  and  $Y$  are related through a  $Z$ -channel as well. (See Fig. 3.) The result can be applied to bound the achievable region for general binary pairs  $(X, Y)$  using the concavity property derived in [8], and to fully describe the region for  $(X, Y)$  pairs whose joint distributions decompose into known irreducible components. (See [2].)

## II. BACKGROUND

Consider the coded side information problem shown in Fig. 1. Source  $X$  and side information  $Y$  are jointly distributed random variables on alphabets  $\mathcal{X}$  and  $\mathcal{Y}$  of sizes  $|\mathcal{X}|, |\mathcal{Y}| < \infty$ .

When  $(X, Y)$  are drawn i.i.d. according to a fixed joint probability mass function (pmf), every achievable point  $(R_X, R_Y)$  satisfies  $R_X \geq H(X|U)$  and  $R_Y \geq I(Y; U)$  for some random variable  $U$  satisfying Markov condition

$$X \rightarrow Y \rightarrow U, \quad (1)$$

<sup>0</sup>This material is based upon work partially supported by NSF Grant No. CCR-0325324 and Caltech’s Lee Center for Advanced Networking.

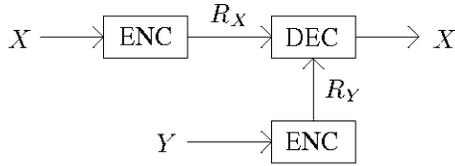


Fig. 1. The coded side information problem.

and alphabet bound  $|\mathcal{U}| \leq |\mathcal{Y}| + 2$  [1]. To compute the lower convex hull of this rate region, we minimize the Lagrangian functional  $H(X|U) + \mu I(Y;U)$  ( $\mu > 0$ ) under the natural constraints on  $p(y|u)$ . Unfortunately, this functional is neither convex nor concave in  $p(y|u)$  in general since  $H(X|U)$  is concave in  $p(y|u)$  while  $I(Y|U)$  is convex in  $p(y|u)$ . Therefore, the optimization is surprisingly difficult even in the case where both  $X$  and  $Y$  are binary random variables.

### III. ALPHABET SIZE OF $U$

While the following result treats the coded side information problem, the method used to prove it applies more widely.

**Theorem 1:** Alphabet size  $|\mathcal{U}| \leq |\mathcal{Y}|$  suffices to achieve any point  $(R_X, R_Y)$  on the lower boundary of the achievable rate region for the coded side-information problem.

*Proof:* The usual time-sharing argument implies that our achievable rate region is convex. Thus any point can be described by some auxiliary random variable  $U$  that minimizes  $H(X|U) + \mu I(Y;U)$  subject to constraint (1) for some  $\mu \geq 0$ . Since  $H(Y)$  is fixed,  $U$  minimizes  $H(X|U) + \mu I(Y;U)$  if and only if it minimizes  $H(X|U) - \mu H(Y|U)$ . Now fix an alphabet  $\mathcal{U}$ , and for each  $u \in \mathcal{U}$ , fix a conditional pmf  $\{p(y|u)\}_{y \in \mathcal{Y}}$ . We next show that no matter how large the original alphabet  $\mathcal{U}$  and no matter which conditional distribution  $\{p(y|u)\}_{y \in \mathcal{Y}}$  is used for each  $u \in \mathcal{U}$ , the optimal solution sets  $p(u) = 0$  for all but at most  $|\mathcal{Y}|$  values of  $u$ .

The given optimization problem is equivalent to choosing the pmf  $\{p(u)\}_{u \in \mathcal{U}}$  that minimizes

$$\sum_{u \in \mathcal{U}} [H(X|U = u) - \mu H(Y|U = u)] p(u)$$

$$\text{subject to } \begin{cases} \sum_{u \in \mathcal{U}} p(u) = 1. \\ p(y) = p(u)p(y|u) \quad \forall y \in \mathcal{Y} \\ p(u) \geq 0 \quad \forall u \in \mathcal{U}. \end{cases}$$

Fixing  $\{p(y|u)\}_{y \in \mathcal{Y}}$  for each  $u \in \mathcal{U}$  means that  $H(X|U = u)$ ,  $H(Y|U = u)$ , and  $p(y|u)$  are constants for each  $u$ , and the minimization is a linear program. Since the optimal point for any linear program can be achieved at a corner point, there exists an optimizing solution that satisfies  $|\mathcal{U}|$  constraints with equality. There is one constraint of the form  $\sum_{u \in \mathcal{U}} p(u) = 1$ , and there are  $|\mathcal{Y}| - 1$  independent constraints of the form  $p(y) = p(u)p(y|u)$  for some  $y \in \mathcal{Y}$ . Thus the remaining  $|\mathcal{U}| - (1 + |\mathcal{Y}| - 1)$  constraints that are satisfied with equality must take the form  $p(u) = 0$  for some  $u \in \mathcal{U}$ .  $\square$

### IV. BINARY SYMMETRICAL CASE

In this section, we consider the case where  $X$  and  $Y$  are related through a binary symmetric channel, namely,

$P_{X|Y}(1|0) = P_{X|Y}(0|1) = \epsilon$  for some  $0 < \epsilon < 1/2$  (the trivial cases  $\epsilon = 0$  and  $1/2$  are excluded). Theorem 2 is our central result for this case. For any  $x, \epsilon \in [0, 1]$ , we define  $H(x) := -x \log x - (1-x) \log(1-x)$  and  $H_\epsilon(x) := H(x_\epsilon)$ , where  $x_\epsilon := x(1-\epsilon) + (1-x)\epsilon = \epsilon + (1-2\epsilon)x$ .

**Theorem 2:** If  $(X, Y)$  is a binary pair with  $P_Y(0) = p$  and  $P_{X|Y}(1|0) = P_{X|Y}(0|1) = \epsilon$ , then the lower boundary of the achievable rate region is described by

$$\{(R_X, R_Y) : \begin{aligned} R_X &= H_\epsilon(\gamma), \quad R_Y = H(p) - H(\gamma), \\ \gamma &\in [0, \min\{p, 1-p\}]\}. \end{aligned}$$

This is achieved by binary auxiliary random variable  $U$  with  $U \rightarrow Y \rightarrow X$  if and only if  $P_{Y|U}(1|0) = P_{Y|U}(0|1) = \gamma$ .

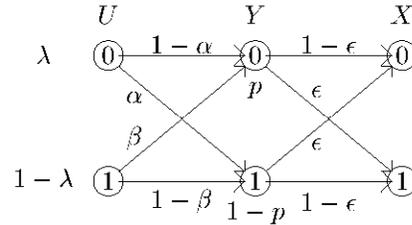


Fig. 2. The binary symmetric case.

To prove this theorem, let  $U$  be a binary random variable with  $U \rightarrow Y \rightarrow X$ ,  $\lambda = P_U(0)$ ,  $\alpha = P_{Y|U}(1|0)$ , and  $\beta = P_{Y|U}(0|1)$ . Since  $p = \lambda(1-\alpha) + (1-\lambda)\beta$ ,  $\lambda = (p-\beta)/(1-\alpha-\beta)$ . and

$$\begin{aligned} H(X|U) &= \frac{p-\beta}{1-\alpha-\beta} H_\epsilon(\alpha) + \frac{1-p-\alpha}{1-\alpha-\beta} H_\epsilon(\beta) \\ H(Y|U) &= \frac{p-\beta}{1-\alpha-\beta} H(\alpha) + \frac{1-p-\alpha}{1-\alpha-\beta} H(\beta). \end{aligned}$$

Therefore, finding an optimal  $U$  is equivalent to solving

$$\begin{aligned} \max & \left[ \frac{p-\beta}{1-\alpha-\beta} H(\alpha) + \frac{1-p-\alpha}{1-\alpha-\beta} H(\beta) \right] \\ \text{subject to} & \frac{p-\beta}{1-\alpha-\beta} H_\epsilon(\alpha) + \frac{1-p-\alpha}{1-\alpha-\beta} H_\epsilon(\beta) = K \end{aligned} \quad (2)$$

over  $0 \leq \alpha, \beta \leq 1$  for some  $K$ . Note that  $H(\epsilon) \leq K = \lambda H_\epsilon(\alpha) + (1-\lambda) H_\epsilon(\beta) \leq H_\epsilon(\lambda\alpha + (1-\lambda)\beta) = H_\epsilon(1-p)$ . We use the next theorem, which is proven in Appendix A, to solve this optimization problem.

**Theorem 3:** The following inequality holds for any  $0 < \alpha < \gamma < \beta \leq 1/2$  and  $0 < \epsilon < 1/2$

$$\frac{H(\beta) - H(\gamma)}{H(\beta) - H(\alpha)} < \frac{H_\epsilon(\beta) - H_\epsilon(\gamma)}{H_\epsilon(\beta) - H_\epsilon(\alpha)}.$$

Fix any  $K \in [H(\epsilon), H_\epsilon(p)]$  and let  $0 \leq \gamma < 1/2$  be chosen such that  $K = H_\epsilon(\gamma)$ . Then  $H_\epsilon(x)$  increasing in  $x \in [0, 1/2]$  implies that  $0 \leq \gamma \leq \min\{p, 1-p\}$ . We wish to show that

$$H(\gamma) > \lambda H(\alpha) + (1-\lambda) H(\beta)$$

for any distinct  $\alpha$  and  $\beta$  satisfying

$$K = \frac{p-\beta}{1-\alpha-\beta} H_\epsilon(\alpha) + \frac{1-p-\alpha}{1-\alpha-\beta} H_\epsilon(\beta).$$

Without loss of generality, we assume that  $0 \leq \alpha < \beta \leq 1/2$  (if not,  $\alpha$  or  $\beta$  can be replaced with  $1 - \alpha$  or  $1 - \beta$ , and the optimization problem is the same with an appropriate change of the formula of  $\lambda$  in terms of  $\alpha$  and  $\beta$ ). Then  $\gamma \in (\alpha, \beta)$  and  $H_\epsilon(\gamma) = \lambda H_\epsilon(\alpha) + (1 - \lambda)H_\epsilon(\beta)$ , giving  $\lambda = (H_\epsilon(\beta) - H_\epsilon(\gamma))/(H_\epsilon(\beta) - H_\epsilon(\alpha))$ . By Theorem 3,  $H(\gamma) > \lambda H(\alpha) + (1 - \lambda)H(\beta)$ , which implies that the symmetrical solution is optimal. On the other hand, for any  $K \in [H(\epsilon), H_\epsilon(p)]$ , there always exists some  $\gamma \in [0, \min\{p, 1 - p\}]$  such that  $K = H_\epsilon(\gamma)$ . Hence by letting  $\alpha = \beta = \gamma$ ,

$$H(\gamma) = \frac{p - \beta}{1 - \alpha - \beta}H(\alpha) + \frac{1 - p - \alpha}{1 - \alpha - \beta}H(\beta)$$

is maximized under the constraint

$$K = \frac{p - \beta}{1 - \alpha - \beta}H_\epsilon(\alpha) + \frac{1 - p - \alpha}{1 - \alpha - \beta}H_\epsilon(\beta).$$

This completes the proof.

### V. Z-CHANNEL CASE

We next consider  $X$  and  $Y$  related through a Z channel.

*Theorem 4:* If  $(X, Y)$  is a binary pair with  $P_Y(0) = p$ ,  $P_{X|Y}(1|0) = 0$ , and  $P_{X|Y}(0|1) = 1 - \epsilon$ , then the lower boundary of the achievable rate region is described by

$$\begin{aligned} \{(R_X, R_Y) : R_X &= \frac{1-p}{\gamma}H_\epsilon(\gamma), \\ R_Y &= H(p) - \frac{1-p}{\gamma}H(\gamma), \gamma \in [1-p, 1]\}. \end{aligned}$$

If  $U$  is binary with  $U \rightarrow Y \rightarrow X$ , then  $U$  achieves this bound if and only if  $P_{Y|U}(1|0) = 0$  and  $P_{Y|U}(1|1) = \gamma$ .

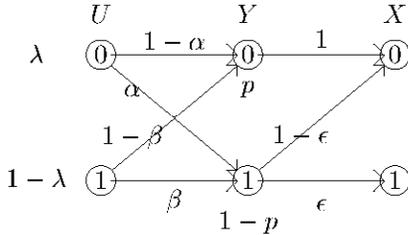


Fig. 3. The binary Z-channel case.

To prove this result, define  $\alpha$ ,  $\beta$ , and  $\lambda$  as before. (See Fig. 3.) Note that  $\lambda\alpha + (1 - \lambda)\beta = 1 - p$  implies  $\alpha \leq 1 - p \leq \beta$  or  $\alpha \geq 1 - p \geq \beta$ . Without loss of generality, we suppose  $\alpha \leq 1 - p \leq \beta$ . Here

$$\begin{aligned} H(X|U) &= \frac{\beta - (1 - p)}{\beta - \alpha}H(\epsilon\alpha) + \frac{(1 - p) - \alpha}{\beta - \alpha}H(\epsilon\beta) \\ H(Y|U) &= \frac{\beta - (1 - p)}{\beta - \alpha}H(\alpha) + \frac{(1 - p) - \alpha}{\beta - \alpha}H(\beta). \end{aligned}$$

Theorem 5 relates this quantities.

*Theorem 5:* Suppose  $0 < \alpha \leq 1 - p \leq \beta$ ,  $\alpha \neq \beta$ , and  $1 - p \leq \gamma \leq 1$ . Choose  $0 < \epsilon < 1$  such that

$$\frac{\beta - (1 - p)}{\beta - \alpha}H(\epsilon\alpha) + \frac{(1 - p) - \alpha}{\beta - \alpha}H(\epsilon\beta) = \frac{1 - p}{\gamma}H(\epsilon\gamma). \quad (3)$$

Then the following is true

$$\frac{\beta - (1 - p)}{\beta - \alpha}H(\alpha) + \frac{(1 - p) - \alpha}{\beta - \alpha}H(\beta) < \frac{1 - p}{\gamma}H(\gamma). \quad (4)$$

*Proof.* We first show that (3) implies that  $1 - p \leq \gamma < \beta$ . Consider any  $0 < \tau \leq 1$ , and some distinct  $x$  and  $y$  such that  $x \in [\alpha, 1 - p]$  and  $y \in [1 - p, \beta]$ , let

$$\theta_\tau(x, y) = \frac{y - (1 - p)}{y - x}H(\tau x) + \frac{(1 - p) - x}{y - x}H(\tau y).$$

Note that

$$\frac{d}{dx}\theta_\tau(x, y) = \frac{y - (1 - p)}{(y - x)^2}D(\tau y||\tau x) > 0$$

and that

$$\frac{d}{dy}\theta_\tau(x, y) = -\frac{(1 - p) - x}{(y - x)^2}D(\tau x||\tau y) < 0.$$

Hence for any  $\tau$ ,  $\theta_\tau(x, y)$  is strictly increasing in  $x$  (when  $y$  is fixed) and strictly decreasing in  $y$  (when  $x$  is fixed).

Rewriting (3) as  $\theta_\epsilon(\alpha, \beta) = \theta_\epsilon(0, \gamma)$ , we have immediately that

$$\theta_\epsilon(0, \beta) < \theta_\epsilon(\alpha, \beta) = \theta_\epsilon(0, \gamma),$$

which proves  $\beta > \gamma$ .

The condition (3) implies that

$$\theta_\epsilon(\alpha, \gamma) - \theta_\epsilon(\alpha, \beta) = \theta_\epsilon(\alpha, \gamma) - \theta_\epsilon(0, \gamma),$$

or equivalently,

$$\begin{aligned} &\int_\gamma^\beta \frac{(1 - p) - \alpha}{(y - \alpha)^2}D(\epsilon\alpha||\epsilon y)dy \\ &= \int_0^\alpha \frac{\gamma - (1 - p)}{(\gamma - x)^2}D(\epsilon\gamma||\epsilon x)dx, \end{aligned}$$

i.e.,

$$\frac{\int_\gamma^\beta \frac{(1 - p) - \alpha}{(y - \alpha)^2}D(\epsilon\alpha||\epsilon y)dy}{\int_0^\alpha \frac{\gamma - (1 - p)}{(\gamma - x)^2}D(\epsilon\gamma||\epsilon x)dx} = 1.$$

Now for any  $0 < x < \alpha < 1 - p < \gamma < y < \beta$ , by Lemma 5 (see Appendix B), we have

$$\frac{\frac{\gamma - (1 - p)}{(\gamma - x)^2}D(\gamma||x)}{D(\alpha||y)} < \frac{\frac{\gamma - (1 - p)}{(\gamma - x)^2}D(\epsilon\gamma||\epsilon x)}{D(\epsilon\alpha||\epsilon y)},$$

which implies that

$$\frac{\int_0^\alpha \frac{\gamma - (1 - p)}{(\gamma - x)^2}D(\gamma||x)dx}{D(\alpha||y)} < \frac{\int_0^\alpha \frac{\gamma - (1 - p)}{(\gamma - x)^2}D(\epsilon\gamma||\epsilon x)dx}{D(\epsilon\alpha||\epsilon y)}.$$

Hence for any  $\gamma < y < \beta$ ,

$$\frac{D(\alpha||y)}{\int_0^\alpha \frac{\gamma - (1 - p)}{(\gamma - x)^2}D(\gamma||x)dx} > \frac{D(\epsilon\alpha||\epsilon y)}{\int_0^\alpha \frac{\gamma - (1 - p)}{(\gamma - x)^2}D(\epsilon\gamma||\epsilon x)dx},$$

which leads to

$$\begin{aligned} \frac{\theta_1(\alpha, \gamma) - \theta_1(\alpha, \beta)}{\theta_1(\alpha, \gamma) - \theta_1(0, \gamma)} &= \frac{\int_\gamma^\beta \frac{(1 - p) - \alpha}{(y - \alpha)^2}D(\alpha||y)dy}{\int_0^\alpha \frac{\gamma - (1 - p)}{(\gamma - x)^2}D(\gamma||x)dx} \\ &> \frac{\int_\gamma^\beta \frac{(1 - p) - \alpha}{(y - \alpha)^2}D(\epsilon\alpha||\epsilon y)dy}{\int_0^\alpha \frac{\gamma - (1 - p)}{(\gamma - x)^2}D(\epsilon\gamma||\epsilon x)dx} = 1. \end{aligned}$$

Thus, (4) holds.  $\square$

Note that  $\frac{1-p}{\gamma}H(\epsilon\gamma)$  ranges from  $(1-p)H(\epsilon) = H(X|Y)$  (when  $\gamma = 1$ ) to  $H(\epsilon(1-p)) = H(X)$  (when  $\gamma = 1-p$ ). For any  $K \in [H(X|Y), H(X)]$ , let  $\gamma \in [1-p, 1]$  be such that

$$K = \frac{1-p}{\gamma}H(\epsilon\gamma).$$

Then by Theorem 5, for any distinct pair  $(\alpha, \beta)$  satisfying  $0 < \alpha \leq 1-p \leq \beta$  such that

$$K = \frac{\beta - (1-p)}{\beta - \alpha}H(\epsilon\alpha) + \frac{(1-p) - \alpha}{\beta - \alpha}H(\epsilon\beta),$$

$$\frac{\beta - (1-p)}{\beta - \alpha}H(\alpha) + \frac{(1-p) - \alpha}{\beta - \alpha}H(\beta) < \frac{1-p}{\gamma}.$$

Therefore, for every optimal solution,  $\alpha = 0$ . Conversely, for any pair  $(\alpha, \beta)$  with  $\alpha \leq 1-p \leq \beta$ , there exists  $\gamma \in [1-p, 1]$  such that (3) holds, and then (4) holds by Theorem 5.

## APPENDIX

### A. Proof of Theorem 3

We prove Theorem 3 by combining the following four lemmas.

*Lemma 1:* Let  $f$  and  $g$  be nonnegative differentiable functions on  $[a, b]$  for some  $a < b$ .<sup>1</sup> If  $f'(x)/g'(x)$  is (strictly) increasing, then

$$(a) \quad \frac{f(x) - f(0)}{g(x) - g(0)}$$

is (strictly) increasing in  $x$  for  $x$  such that  $g(x) \neq g(0)$ .

$$(b) \quad \frac{f(1) - f(x)}{g(1) - g(x)}$$

is (strictly) increasing in  $x$  for  $x$  such that  $g(x) \neq g(1)$ .

*Proof.* Without loss of generality, we assume that  $a = 0$  and  $b = 1$ . Given  $0 \leq t < 1$ , and  $0 < \Delta \leq 1 - t$ . For each  $n \in \mathbb{N}$ , consider the partition of  $[0, t]$

$$\left\{ \left[0, \frac{t}{n}\right], \left(\frac{t}{n}, \frac{2t}{n}\right], \dots, \left(\frac{(n-1)t}{n}, t\right] \right\},$$

and the partition of  $(t, t + \Delta]$

$$\left\{ \left(t, t + \frac{\Delta}{n}\right], \left(t + \frac{\Delta}{n}, t + \frac{2\Delta}{n}\right], \dots, \left(t + \frac{(n-1)\Delta}{n}, t + \Delta\right] \right\}.$$

Define the corresponding Riemann sums as follows

$$S_{f,n} = \sum_{k=1}^n f'\left(\frac{kt}{n}\right) \frac{t}{n}$$

$$S_{g,n} = \sum_{k=1}^n g'\left(\frac{kt}{n}\right) \frac{t}{n}$$

$$T_{f,n} = \sum_{k=1}^n f'\left(t + \frac{k\Delta}{n}\right) \frac{\Delta}{n}$$

$$T_{g,n} = \sum_{k=1}^n g'\left(t + \frac{k\Delta}{n}\right) \frac{\Delta}{n}.$$

<sup>1</sup>This lemma still works when  $a = -\infty$  or  $b = \infty$ .

$\square$  Then one has

$$\lim_{n \rightarrow \infty} S_{f,n} = f(t) - f(0)$$

$$\lim_{n \rightarrow \infty} S_{g,n} = g(t) - g(0)$$

$$\lim_{n \rightarrow \infty} T_{f,n} = f(t + \Delta) - f(t)$$

$$\lim_{n \rightarrow \infty} T_{g,n} = g(t + \Delta) - g(t),$$

and

$$\lim_{n \rightarrow \infty} \frac{S_{f,n}}{S_{g,n}} = \frac{f(t) - f(0)}{g(t) - g(0)}$$

$$\lim_{n \rightarrow \infty} \frac{T_{f,n}}{T_{g,n}} = \frac{f(t + \Delta) - f(t)}{g(t + \Delta) - g(t)}.$$

Now, since the function  $f'(x)/g'(x)$  is increasing,

$$\frac{S_{f,n}}{S_{g,n}} \leq \frac{S_{f,n} + T_{f,n}}{S_{g,n} + T_{g,n}},$$

and therefore

$$\frac{f(t) - f(0)}{g(t) - g(0)} \leq \frac{f(t + \Delta) - f(0)}{g(t + \Delta) - g(0)}.$$

A similar argument leads to (b).  $\square$

*Lemma 2:* The function

$$\frac{H'_\epsilon(x)}{H''(x)} = \frac{(1 - 2\epsilon)^2 x(1 - x)}{x_\epsilon(1 - x_\epsilon)}$$

is strictly increasing for  $x \in (0, 1/2)$ .

*Proof.* Notice that

$$\frac{d}{dx} \frac{x(1-x)}{x_\epsilon(1-x_\epsilon)} = \frac{(1-2x)x_\epsilon(1-x_\epsilon)}{(x_\epsilon(1-x_\epsilon))^2}$$

$$- \frac{(1-2\epsilon)(1-2x_\epsilon)x(1-x)}{(x_\epsilon(1-x_\epsilon))^2}.$$

Therefore, it suffices to show that

$$(1 - 2x)x_\epsilon(1 - x_\epsilon) - (1 - 2\epsilon)(1 - 2x_\epsilon)x(1 - x) \geq 0$$

for  $x \in (0, 1/2)$ . Fix  $x \in (0, 1/2)$ . Consider the function

$$\phi(\eta) = (1 - 2x)x_\eta(1 - x_\eta)$$

$$- (1 - 2\eta)(1 - 2x_\eta)x(1 - x)$$

for  $\eta \in [0, 1/2]$ .

Since  $x_\eta = x + (1 - 2x)\eta$ , the function  $x_\eta(1 - x_\eta)$  is strictly concave in  $\eta$ , and the function  $(1 - 2\eta)(1 - 2x_\eta)$  is convex in  $\eta$ . Thus  $\phi(\eta)$  is strictly concave in  $\eta$ . Now  $\phi(0) = 0$  and  $\phi(1/2) = (1 - 2x)/4 > 0$ , hence  $\phi(\epsilon) > 0$ .  $\square$

*Lemma 3:* (a) The function  $H'_\epsilon(x)/H''(x)$  is strictly increasing for  $x \in (0, 1/2)$ .

(b) The function

$$\frac{H_\epsilon(x) - H_\epsilon(\alpha)}{H(x) - H(\alpha)}$$

is strictly increasing for  $x \in (\alpha, \beta)$ .

(c) The function

$$\frac{H_\epsilon(\beta) - H_\epsilon(x)}{H(\beta) - H(x)}$$

is strictly increasing for  $x \in (\alpha, \beta)$ .

*Proof.*

(a) By Lemma 2,  $H'_\epsilon(x)/H''(x)$  is strictly increasing for  $x \in (0, 1/2)$ . Since

$$\frac{H'_\epsilon(x)}{H''(x)} = \frac{H'_\epsilon(1/2) - H'_\epsilon(x)}{H''(1/2) - H''(x)},$$

Lemma 1 (b) implies that  $H'_\epsilon(x)/H''(x)$  is strictly increasing for  $x \in (0, 1/2)$ .

(b)  $(H_\epsilon(x) - H_\epsilon(\alpha))/(H(x) - H(\alpha))$  is strictly increasing for  $x \in (\alpha, \beta)$  by (a) and Lemma 1 (a).

(c)  $(H_\epsilon(\beta) - H_\epsilon(x))/(H(\beta) - H(x))$  is strictly increasing for  $x \in (\alpha, \beta)$  by (a) and Lemma 1(b).

*Lemma 4:* For any  $x \in (\alpha, \beta)$ ,

$$\frac{H_\epsilon(\beta) - H_\epsilon(\alpha)}{H(\beta) - H(\alpha)} H(x) + \frac{H_\epsilon(\alpha)H(\beta) - H(\alpha)H_\epsilon(\beta)}{H(\beta) - H(\alpha)} > H_\epsilon(x).$$

*Proof.* By Lemma 3(b),

$$\begin{aligned} \frac{H_\epsilon(\beta) - H_\epsilon(\alpha)}{H(\beta) - H(\alpha)} &> \lim_{x \rightarrow \alpha^+} \frac{H_\epsilon(x) - H_\epsilon(\alpha)}{H(x) - H(\alpha)} \\ &= (1 - 2\epsilon) \frac{\ln(1 - \alpha_\epsilon) - \ln(\alpha_\epsilon)}{\ln(1 - \alpha) - \ln(\alpha)}. \end{aligned} \quad (\text{A-1})$$

Similarly, by Lemma 3(c),

$$\begin{aligned} \frac{H_\epsilon(\beta) - H_\epsilon(\alpha)}{H(\beta) - H(\alpha)} &< \lim_{x \rightarrow \beta^-} \frac{H_\epsilon(\beta) - H_\epsilon(x)}{H(\beta) - H(x)} \\ &= (1 - 2\epsilon) \frac{\ln(1 - \beta_\epsilon) - \ln(\beta_\epsilon)}{\ln(1 - \beta) - \ln(\beta)}. \end{aligned} \quad (\text{A-2})$$

Let

$$\begin{aligned} \phi(x) &= \frac{H_\epsilon(\beta) - H_\epsilon(\alpha)}{H(\beta) - H(\alpha)} H(x) - H_\epsilon(x) \\ &\quad + \frac{H_\epsilon(\alpha)H(\beta) - H(\alpha)H_\epsilon(\beta)}{H(\beta) - H(\alpha)} \end{aligned}$$

for  $x \in (\alpha, \beta)$ . (A-1) and (A-2) imply that  $\phi'(\alpha) > 0$  and  $\phi'(\beta) < 0$ . Now

$$\phi'(x) = \frac{H_\epsilon(\beta) - H_\epsilon(\alpha)}{H(\beta) - H(\alpha)} H'(x) - H'_\epsilon(x).$$

Since the function  $H'_\epsilon(x)/H''(x)$  is strictly increasing by Lemma 3(a), there is only one point  $s \in (\alpha, \beta)$  such that  $\phi'(s) = 0$ . This implies that  $\phi(x) > 0$  for  $x \in (\alpha, \beta)$ .  $\square$

### B. Proof of Lemma 5

*Lemma 5:* If  $0 < x < \alpha < \gamma < y < \beta$ , then for any  $\epsilon \in (0, 1)$ ,

$$\frac{D(\epsilon\alpha|\epsilon y)}{D(\epsilon\gamma|\epsilon x)} < \frac{D(\alpha|y)}{D(\gamma|x)}.$$

*Proof.* For any fixed  $s \neq t$ , consider  $D(\tau s|\tau t)$  as a function of  $\tau \in (-\infty, 1]$ . Then

$$\begin{aligned} D'(\tau s|\tau t) &= \frac{d}{d\tau} D(\tau s|\tau t) \\ &= s \ln \frac{s}{t} - s \ln \frac{1 - \tau s}{1 - \tau t} - s + \frac{1 - \tau s}{1 - \tau t} t \end{aligned}$$

and

$$D''(\tau s|\tau t) = \frac{(s - t)^2}{(1 - \tau s)(1 - \tau t)^2} > 0.$$

Thus the function

$$\frac{D'(\tau\alpha|\tau y)}{D'(\tau\gamma|\tau x)} = \frac{(\alpha - y)^2 (1 - \tau\gamma)(1 - \tau x)^2}{(\gamma - x)^2 (1 - \tau\alpha)(1 - \tau y)^2}$$

is positive and is strictly increasing in  $\tau$ .

$\square$  Hence by Lemma 1,

$$\frac{D'(\tau\alpha|\tau y)}{D'(\tau\gamma|\tau x)} = \frac{\int_{-\infty}^{\tau} D''(\eta\alpha|\eta y) d\eta}{\int_{-\infty}^{\tau} D''(\eta\gamma|\eta x) d\eta}$$

is strictly increasing in  $\tau$ .

Note that

$$\lim_{\tau \rightarrow -\infty} D'(\tau s|\tau t) = 0,$$

hence  $D'(\tau s|\tau t) > 0$  for all  $\tau \in (-\infty, 1]$ . Thus by Lemma 1 again, the function

$$\frac{D(\tau\alpha|\tau y)}{D(\tau\gamma|\tau x)} = \frac{\int_{-\infty}^{\tau} D'(\eta\alpha|\eta y) d\eta}{\int_{-\infty}^{\tau} D'(\eta\gamma|\eta x) d\eta}$$

is strictly increasing in  $\tau$ . In particular,

$$\frac{D(\epsilon\alpha|\epsilon y)}{D(\epsilon\gamma|\epsilon x)} < \frac{D(\alpha|y)}{D(\gamma|x)}.$$

$\square$

### REFERENCES

- [1] 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.
- [2] 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, 2006. IEEE.
- [3] A. El Gamal and T. M. Cover. Achievable rates for multiple descriptions. *IEEE Transactions on Information Theory*, IT-28(6):851–857, November 1982.
- [4] R. M. Gray and A. D. Wyner. Source coding for a simple network. *Bell System Technical Journal*, 53(9):1681–1721, November 1974.
- [5] C. Heegard and T. Berger. Rate distortion when side information may be absent. *IEEE Transactions on Information Theory*, IT-31:727–734, November 1985.
- [6] A. D. Wyner. On source coding with side information at the decoder. *IEEE Transactions on Information Theory*, IT-21(3):294–300, November 1975.
- [7] H. Yamamoto. Source coding theory for cascade and branching communication systems. *IEEE Transactions on Information Theory*, IT-27:299–308, May 1981.
- [8] W. Gu and M. Effros. On the concavity of lossless rate regions. In *Proceedings of the IEEE International Symposium on Information Theory*, Seattle, WA, 2006.