Remote Source Coding with Two-Sided Information

Basak Guler Ebrahim MolavianJazi Aylin Yener

Wireless Communications and Networking Laboratory
Department of Electrical Engineering
The Pennsylvania State University, University Park, PA 16802
basak@psu.edu *eum20@psu.edu* *yener@engr.psu.edu*

***Abstract*—This paper studies the impact of side information on the lossy compression of a remote source, one which is indirectly accessed by the encoder. In particular, we identify the conditions under which sharing side information between the encoder and the decoder may be superior or inferior to having two-sided, i.e., correlated but not identical, side information. As a special case, we characterize the optimal rate-distortion function for a direct binary source with two-sided information by proposing an achievable scheme and proving a converse. This example suggests a hierarchy on the impact of side information, in that the performance is mainly determined by how well the decoder learns about the source and then by how well the encoder learns about the decoder's observation.**

I. Introduction

We study the lossy compression of a remote source with two-sided side information. The encoder has access only to a noisy observation of the source. The encoder and the decoder have either common or individual side information correlated with the source. We focus on identifying whether common or correlated side information is beneficial from the source coding perspective, and investigate the conditions on the indirect observation of the source as well as the encoder's and decoder's side information under which one system outperforms the other. As a special case, we characterize the rate-distortion function for a binary source with two-sided information and elaborate on the hierarchy between encoder and decoder side information in lossy source coding.

Related Work: Shannon introduced the rate-distortion function as the fundamental metric for (direct) source coding with a fidelity criterion [1]. Wyner and Ziv showed that the rate-distortion function with only decoder side information [2] is inferior to the conditional rate-distortion of the system with shared information between the encoder and the decoder [3]. Lossy source coding with two-sided side information is studied in [4]. The rate-distortion function for the case with decoder side information as well as an additional shared information is numerically evaluated in [5]. For a remote source with only decoder side information, which we call a *remote Wyner-Ziv* system, the rate-distortion function is characterized in [6].

II. System Model and Preliminaries

In the remainder of the paper, X represents a random variable, and x is its realization. $\mathcal{X}$ denotes a set, and $|\mathcal{X}|$ is its cardinality. $\mathbb{E}[X]$ is the expected value of X.

Consider a discrete memoryless remote (indirect) source with two-sided side information as in Fig. 1. $\tilde{X}$ is the encoder's

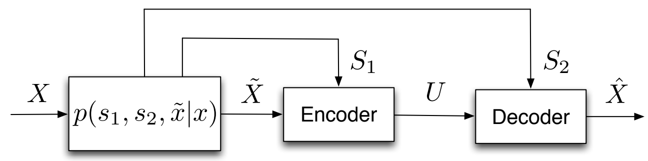

Fig. 1. Remote source coding with two-sided information.

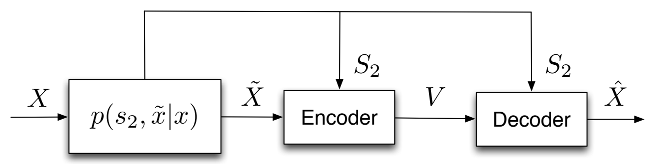

Fig. 2. Remote source coding with shared information.

observation of the source X. The direct source is a special case with $X = \tilde{X}$. The encoder and the decoder have access to side information S_1 and S_2, respectively. The random variables X, $\tilde{X}$, S_1, and S_2 are described by the joint distribution $p(x, \tilde{x}, s_1, s_2)$ over the corresponding alphabets $\mathcal{X}$, $\tilde{\mathcal{X}}$, $\mathcal{S}_1$, $\mathcal{S}_2$. The n-letter encoding and decoding schemes are defined analogously to [4]. In the resulting single letter expression, a lossy version $\hat{X}$ of X is recovered, with a fidelity criterion $\mathbb{E}[d(X, \hat{X})] \leq D$, where $d(x, \hat{x})$ is the distortion measure.

The remote source with two-sided side information is a special case of the remote Wyner-Ziv system in [6], by setting $\tilde{X}S_1$ as the encoder's observation of the source and S_2 as the decoder side information. Therefore, the rate-distortion function for the remote two-sided system is

$$R_{\text{TS}}(D) = \min_{p(u|\tilde{x}, s_1), \hat{x}=g(u,s_2)} I(U; \tilde{X}S_1|S_2), \qquad (1)$$

such that $U - \tilde{X}S_1 - XS_2$ forms a Markov chain and $\mathbb{E}[d(X, g(U, S_2))] \leq D$. For the cardinality of the auxiliary variable, it suffices to consider $|\mathcal{U}| \leq |\tilde{\mathcal{X}}||\mathcal{S}_1| + 1$.

Next, consider a remote source with shared information as in Fig. 2, in which the decoder side information is shared with the encoder. Its rate distortion function follows from (1) by letting $S_1 = S_2$ and noting that $I(V; \tilde{X}, S_2|S_2) = I(V; \tilde{X}|S_2)$ and that $V - \tilde{X}S_2 - XS_2$ holds if and only if $V - \tilde{X}S_2 - X$. Therefore the rate-distortion function for the remote system with shared-information is

$$R_{\text{SH}}(D) = \min_{\tilde{p}(v|\tilde{x}, s_2), \hat{x}=f(v,s_2)} I(V; \tilde{X}|S_2), \qquad (2)$$

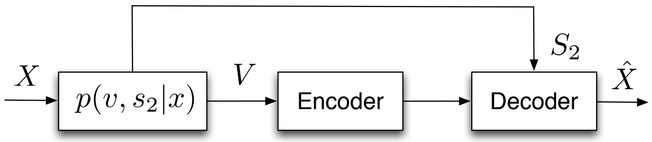

Fig. 3. Remote source with decoder only side information (remote Wyner-Ziv).

such that $V - \tilde{X}S_2 - X$ forms a Markov chain and $\mathbb{E}[d(X, f(V, S_2))] \leq D$.

Finally, we present a technical result on reducing two-sided systems to "Wyner-Ziv" ones.

Proposition 1. *Given $\tilde{X} - S_1 - XS_2$ or $S_1 - \tilde{X} - XS_2$, the remote source system with two-sided side information (Fig. 1) is equivalent to a remote source system with decoder only side information in Fig. 3 with $V = S_1$ or $V = \tilde{X}$, respectively.*

Proof. Let $\tilde{X} - S_1 - XS_2$ hold for the system in Fig. 1. First,

$$I(U; \tilde{X}S_1|S_2) \geq I(U; S_1|S_2). \quad (3)$$

Moreover, $\tilde{X} - S_1 - XS_2$ and $U - \tilde{X}S_1 - XS_2$ hold if and only if $U\tilde{X} - S_1 - XS_2$, as a result of non-negativity of mutual information and the chain rule $I(U\tilde{X}; XS_2|S_1) = I(\tilde{X}; XS_2|S_1) + I(U; XS_2|\tilde{X}S_1)$. Since the Markov chain $U\tilde{X} - S_1 - XS_2$ implies $U - S_1 - XS_2$, then (3) states that the search space can be limited to the distributions that satisfy $U - S_1 - XS_2$ with no rate increase. Hence, the remote system with two-sided information can be reduced to a remote source with decoder side information in Fig. 3 with $V = S_1$. The case when $S_1 - \tilde{X} - XS_2$ follows the same steps with the roles of $\tilde{X}$ and S_1 exchanged. □

III. Cases Where Shared Information is Better

In this section, we study the conditions under which two-sided information can never outperform shared information.

Theorem 1. *For remote source coding systems in Figs. 1 and 2, whenever $S_1 - \tilde{X}S_2 - X$, it holds for all D that*

$$R_{SH}(D) \leq R_{TS}(D). \quad (4)$$

Proof. We show that any optimal strategy for the two-sided case leads to a valid strategy for the shared case, but with a larger rate. Without loss of generality, let U in (1) and V in (2) be from a set with cardinality $\max(|\mathcal{U}|, |\mathcal{V}|)$. Let the minimum in (1) be achieved by $p^*(u|\tilde{x}, s_1)$ and $g^*(u, s_2)$. Then,

$$\mathbb{E}[d(X, g^*(U, S_2))] = \sum_{u,x,\tilde{x},s_1,s_2} d(x, g^*(u, s_2))p^*(u|\tilde{x}, s_1)p(x, \tilde{x}, s_1, s_2) \leq D. \quad (5)$$

For the shared information system, consider the distribution

$$\bar{p}(u|\tilde{x}, s_2) = \sum_{s_1} p^*(u|\tilde{x}, s_1)p(s_1|\tilde{x}, s_2), \quad (6)$$

due to $U - \tilde{X}S_1 - S_2$ since $U - \tilde{X}S_1 - XS_2$, and the function

$$f(u, s_2) = g^*(u, s_2), \quad (7)$$

and use $S_1 - \tilde{X}S_2 - X$ to determine its expected distortion as

$$\begin{aligned}\mathbb{E}[d(X, f(U, S_2))] &= \sum_{u,x,\tilde{x},s_2} d(x, f(u, s_2))\bar{p}(u|\tilde{x}, s_2)p(x, s_2, \tilde{x}) \\ &= \sum_{u,x,\tilde{x},s_2} d(x, g^*(u, s_2)) \sum_{s_1} p^*(u|\tilde{x}, s_1)p(s_1|\tilde{x}, s_2)p(x, s_2, \tilde{x}) \\ &= \sum_{u,x,\tilde{x},s_1,s_2} d(x, g^*(u, s_2))p^*(u|\tilde{x}, s_1)p(s_1|\tilde{x}, s_2, x)p(x, s_2, \tilde{x}) \quad (8)\end{aligned}$$

which is $\leq D$ due to (5). Hence, $\bar{p}(u|\tilde{x}, s_2)$ and $f(u, s_2)$ in (6) and (7) are valid assignments for the shared information system achieving a rate $I(U; \tilde{X}|S_2)$ characterized by $\bar{p}(u|\tilde{x}, s_2)p(x, s_2, \tilde{x})$ in (2), whereas the minimum compression rate for the system with two-sided information according to (1) is

$$R_{\text{TS}}(D) = I(U; \tilde{X}S_1|S_2) \geq I(U; \tilde{X}|S_2) \quad (9)$$

in which $I(U; \tilde{X}|S_2)$ in (9) is characterized by $p^*(u|\tilde{x}, s_2)p(x, s_2, \tilde{x})$ where

$$p^*(u|\tilde{x}, s_2) = \sum_{s_1} p^*(u|\tilde{x}, s_1)p(s_1|\tilde{x}, s_2) = \bar{p}(u|\tilde{x}, s_2), \quad (10)$$

due to (6). We have from (9) and (10) that, for every optimal distribution and reconstruction function corresponding to a given distortion D for the two-sided information system, there exists a feasible assignment for the shared information system whose rate is no greater than that of the two-sided system. □

Remark 1. *Encoder side information is redundant whenever it is conditionally independent of the source given the observation of the source and decoder side information. In this case, it is better to share the decoder side information, no matter how noisy it may be.*

Remark 2. *Theorem 1 can alternatively be obtained by providing S_2 to the encoder in Fig. 1, and using the fact that $R_{TS}(D) \geq \min_U I(U; \tilde{X}, S_1, S_2|S_2) \geq \min_U I(U; \tilde{X}|S_2)$ over all U and g satisfying $U - (\tilde{X}, S_1, S_2) - (X, S_2)$ and $d(X, g(U, S_2)) \leq D$. Since $X - (\tilde{X}, S_1, S_2) - U$ and $X - (\tilde{X}, S_2) - S_1$ imply $X - (\tilde{X}, S_2) - U$, we have $R_{SH}(D) \leq R_{TS}(D)$.*

Corollary 1. *A direct source with two-sided information can never outperform the direct source with shared information.*

Proof. For $\tilde{X} = X$, the condition in Theorem 1 becomes $S_1 - XS_2 - X$ which always holds. □

Therefore, for a direct source, sharing even noisy decoder side information is better than having individual side information at the encoder and the decoder, no matter how good the encoder side information might be.

Corollary 2. *If $S_1 - X - S_2$, the direct source with two-sided information is equal to a Wyner-Ziv system with only decoder side information S_2.*

Proof. For a direct source $\tilde{X} = X$, so we can invoke Prop. 1 with $S_1 - \tilde{X} - XS_2$ and $\tilde{X} = X$, since $S_1 - X - S_2$ is equivalent to $S_1 - X - XS_2$. □

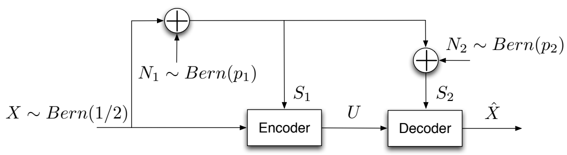

Fig. 4. Binary source with two-sided information, $0 \leq p_1, p_2 \leq 0.5$.

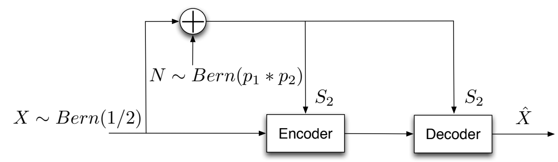

Fig. 5. Binary source with shared information (decoder information is shared).

Hence, encoder side information is totally useless for a direct source if it is does not convey any information about the decoder's observation.

IV. Cases Where Two-sided Information is Better

In this section, we study the conditions under which shared information can never outperform two-sided information.

Theorem 2. *For the remote source coding systems in Figs. 1 and 2, if we have $\tilde{X} - S_1 - XS_2$, then it holds for all D that*

$$R_{SH}(D) \geq R_{TS}(D). \tag{11}$$

Proof. We show that any attainable distortion for the shared information system is attainable for the two-sided information system with the same (zero) rate. Since $\tilde{X} - S_1 - XS_2$, applying Prop. 1 to Fig. 1 reduces it to Fig. 3 with $V = S_1$, which we refer to as System 1. Applying Prop. 1 again to Fig. 2 by letting $S_1 = S_2$ reduces it to Fig. 3 with $V = S_2$ and $p(v, s_2|x) = p(s_2|x)$, which we refer to as System 2. In this case, no additional information is provided to the encoder about X that is not already known to the decoder, hence the only feasible distortion levels are the ones that allow the decoder to predict X from S_2 only, which requires zero rate. As both systems are remote Wyner-Ziv type, the rate distortion function of System 1 is [2]

$$R_1(D) = \min_{p(u|s_1), p(\hat{x}|u, s_2)} I(U; S_1|S_2) \tag{12}$$

such that $U - S_1 - XS_2$ and

$$\mathbb{E}[d(X, \hat{X})] = \sum_{x,u,s_1,s_2,\hat{x}} d(x, \hat{x}) p(\hat{x}|u, s_2) p(u|s_1) p(x, s_1, s_2) \leq D.$$

The rate distortion function of System 2 is

$$R_2(D) = \min_{\substack{\bar{p}(u|s_2), \bar{p}(\hat{x}|u, s_2) \\ U - S_2 - X}} I(U; S_2|S_2) = 0, \tag{13}$$

where

$$\mathbb{E}[d(X, \hat{X})] = \sum_{x,u,s_2,\hat{x}} d(x, \hat{x}) \bar{p}(\hat{x}|u, s_2) \bar{p}(u|s_2) p(x, s_2) \leq D, \tag{14}$$

since $U - S_2 - XS_2$ is equivalent to $U - S_2 - X$. Hence the rate is zero for system 2 whenever D is attainable. Suppose that a given D is attainable for System 2. Then

$$\mathbb{E}[d(X, \hat{X})] = \sum_{x,\hat{x},s_2} d(x, \hat{x}) \bar{p}^*(\hat{x}|s_2) p(x, s_2) \leq D, \tag{15}$$

where $\bar{p}^*(\hat{x}|s_2) = \sum_u \bar{p}(\hat{x}|u, s_2) \bar{p}(u|s_2)$. Next, assume that $p(\hat{x}|u, s_2) = \bar{p}^*(\hat{x}|s_2)$ for all $u \in \mathcal{U}$ and $p(u|s_1) = p^*(u)$ $\forall s_1 \in \mathcal{S}_1$ for some $p^*(u)$ in System 1, and determine its expected distortion as

$$\begin{aligned}\mathbb{E}[d(X, \hat{X})] &= \sum_{x,u,s_1,s_2,\hat{x}} d(x, \hat{x}) p(\hat{x}|u, s_2) p(u|s_1) p(x, s_1, s_2) \\ &= \sum_{x,u,s_1,s_2,\hat{x}} d(x, \hat{x}) \bar{p}^*(\hat{x}|s_2) p^*(u) p(x, s_1, s_2) \\ &= \sum_{x,s_2,\hat{x}} d(x, \hat{x}) \bar{p}^*(\hat{x}|s_2) p(x, s_2) \leq D, \end{aligned} \tag{16}$$

where (16) follows from (15). Hence, D is also attainable for System 1 with a rate given by

$$I(U; S_1|S_2) = I(U; S_1) - I(U; S_2) \leq I(U; S_1) = 0, \tag{17}$$

since $U - S_1 - XS_2$ and U is independent of S_1. Hence, System 1 always performs at least as good as System 2. In other words, shared information cannot outperform two-sided information. □

V. Binary Direct Source with Two-sided Information

In this section, we study a binary direct source with two-sided information shown in Fig. 4, and characterize its rate-distortion function using an optimal *ternary* alphabet for the auxiliary random variable. Our achievability and converse analysis is inspired by [7]. In the following, we frequently use the notations $\bar{p} = 1 - p$ and $p * q = (1 - p)q + p(1 - q)$, for scalars $0 \leq p, q \leq 1$, as well as the binary entropy function $h(p) = -p \log p - (1 - p) \log(1 - p)$.

Theorem 3. *The rate distortion function for the binary direct source with two-sided side information is given by*

$$R_{TS}(D) = \min_{\substack{0 \leq \alpha, \beta, \theta \leq 1 \\ 0 \leq \gamma \leq 0.5 \\ 0 \leq d \leq p_1 * p_2}} R(\alpha, \beta, \theta, d, \gamma) \tag{18}$$

where

$$\begin{aligned} R(\alpha, \beta, \theta, d, \gamma) := h(p_1) - \theta h(d) + \theta h\big(p_2 * (\alpha(1 - d) + \beta d)\big) \\ - \theta(1 - d)h(\alpha) - \theta d h(\beta) - (1 - \theta)h(\gamma) \end{aligned} \tag{19}$$

subject to the constraints

$$\theta d + (1 - \theta)(p_2 * \gamma) \leq D, \tag{20}$$

$$\theta\beta d + \theta(1 - \alpha)(1 - d) + (1 - \theta)\gamma = p_1, \tag{21}$$

$$\gamma = 0.5 \text{ whenever } \theta = 1. \tag{22}$$

Proof. (Achievability) Consider the distribution $p(u, x, s_1, s_2) = p(u, x, s_1)p(s_2|s_1)$, using $p(u, x, s_1)$ given in Table I and $U - XS_1 - S_2$ to calculate

$$R(\alpha, \beta, \theta, d, \gamma) = I(XS_1; U|S_2) \tag{23}$$

$$= H(XS_1) - H(S_2) - H(X|U) + H(S_2|U) - H(S_1|XU) \tag{24}$$

TABLE I
JOINT DISTRIBUTION $p(u,x,s_1)$.

	$S_1=0$		$S_1=1$	
	$X=0$	$X=1$	$X=0$	$X=1$
$U=0$	$\theta\alpha\bar{d}/2$	$\theta\beta d/2$	$\theta\bar{\alpha}\bar{d}/2$	$\theta\bar{\beta}d/2$
$U=1$	$\theta\bar{\beta}d/2$	$\theta\bar{\alpha}\bar{d}/2$	$\theta\beta d/2$	$\theta\alpha\bar{d}/2$
$U=2$	$\bar{\theta}\bar{\gamma}/2$	$\bar{\theta}\gamma/2$	$\bar{\theta}\gamma/2$	$\bar{\theta}\bar{\gamma}/2$

from which (19) immediately follows. Consider an assignment

$$\hat{X}=\begin{cases}U & \text{if} \quad U=0,1\\ S_2 & \text{if} \quad U=2\end{cases} \tag{25}$$

so that the expected distortion is

$$\mathbb{E}[d(X,\hat{X})]=\theta d+(1-\theta)(p_2*\gamma)\leq D. \tag{26}$$

Lastly, the condition $p(x,s_1)=p_1/2$ if $x\neq s_1$ leads to (21), which ensures a valid marginal distribution for (X,S_1).

(Converse) We prove an equivalent expression to (19):

$$R(\alpha,\beta,\theta,d,\gamma)=I(XS_1;U|S_2)$$
$$=H(XS_1)-H(S_2)-H(S_1|U)+H(S_2|U)-H(X|S_1U) \tag{27}$$
$$=h(p_1)-(1-\theta)h(\gamma)+\theta h\big(p_2*(\alpha\bar{d}+\beta d)\big)-\theta h(\alpha\bar{d}+\beta d)$$
$$-\theta(\alpha\bar{d}+\beta d)h\Big(\frac{\beta d}{\alpha\bar{d}+\beta d}\Big)-\theta(1-(\alpha\bar{d}+\beta d))h\Big(\frac{\bar{\beta}d}{1-(\alpha\bar{d}+\beta d)}\Big). \tag{28}$$

Consider all tuples (X,S_1,S_2,U) satisfying $U-XS_1-S_2$ and $\mathbb{E}[d(X,f(U,S_2))]\leq D$ for some optimal $f(u,s_2)=\hat{x}$ and achieving the minimum of (27). Define the set

$$\mathcal{A}=\{u:f(u,0)=f(u,1)=g(u)\}, \tag{29}$$

for some function $g(u)$, and its complement $\mathcal{A}^c=\{u: f(u,0)\neq f(u,1)\}$. Denote

$$\theta:=p(U\in A),\quad 0\leq\theta\leq 1, \tag{30}$$

$$d_u:=p(X\neq g(u)|U=u)\qquad \text{for } u\in\mathcal{A}, \tag{31}$$

$$d:=\sum_{u\in A}\frac{p_u}{\theta}d_u, \tag{32}$$

where $p_u=p(U=u)$, so that $\sum_{u\in A}p_u d_u=\theta d$, and $\sum_{u\in A}p_u(1-d_u)=\theta(1-d)$. Note that, the attained distortion within set $\mathcal{A}$ must be lower than that achieved by solely using S_2, therefore $d_u\leq p_1*p_2$ for all $u\in\mathcal{A}$ and thus $d\leq p_1*p_2$. Next, for each $u\in\mathcal{A}$, define

$$\alpha_u:=p(S_1=g(u)|X=g(u),U=u), \tag{33}$$
$$\beta_u:=p(S_1=g(u)|X\neq g(u),U=u). \tag{34}$$

and

$$\alpha:=\sum_{u\in\mathcal{A}}\frac{p_u(1-d_u)}{\theta(1-d)}\alpha_u,\qquad 0\leq\alpha\leq 1, \tag{35}$$

$$\beta:=\sum_{u\in\mathcal{A}}\frac{p_u d_u}{\theta d}\beta_u,\qquad 0\leq\beta\leq 1. \tag{36}$$

Now, it can be determined within the set $\mathcal{A}$ that,

$$\sum_{u\in\mathcal{A}}p_uH(S_2|U=u)-\sum_{u\in\mathcal{A}}p_uH(S_1|U=u)$$
$$=\sum_{u\in\mathcal{A}}p_u\Big(h(p_2*p(S_1=g(u)|U=u))-h(p(S_1=g(u)|U=u))\Big) \tag{37}$$
$$\geq\theta\Big(h\big(\sum_{u\in\mathcal{A}}\frac{p_u}{\theta}p(S_1=g(u)|U=u)\big)$$
$$-h\Big(p_2*\big(\sum_{u\in\mathcal{A}}\frac{p_u}{\theta}p(S_1=g(u)|U=u)\big)\Big)\Big) \tag{38}$$
$$=\theta\Big(h(\alpha(1-d)+\beta d)-h(p_2*(\alpha(1-d)+\beta d))\Big) \tag{39}$$

where (37) follows from $U-S_1-S_2$, (38) from the convexity of $h(p_2*\delta)-h(\delta)$ for all $0\leq\delta\leq 1$ [2], and (39) from $p(S_1=g(u)|U=u)=\alpha_u(1-d_u)+\beta_u d_u$. Moreover,

$$\sum_{u\in\mathcal{A}}p_uH(X|S_1,U=u)$$
$$=\sum_{u\in\mathcal{A}}p_up(S_1=g(u)|U=u)h(p(X\neq g(u)|S_1=g(u),U=u))$$
$$+\sum_{u\in\mathcal{A}}p_up(S_1\neq g(u)|U=u)h(p(X\neq g(u)|S_1\neq g(u),U=u))$$
$$\leq(\theta(1-d)\alpha+\theta d\beta)h\Big(\sum_{u\in\mathcal{A}}\frac{p_u\beta_u d_u}{\theta(1-d)\alpha+\theta d\beta}\Big)$$
$$+(\theta-(\theta(1-d)\alpha+\theta d\beta))h\Big(\sum_{u\in\mathcal{A}}\frac{p_u(1-\beta_u)d_u}{\theta-(\theta(1-d)\alpha+\theta d\beta)}\Big) \tag{40}$$
$$=\theta(\alpha(1-d)+\beta d)h\Big(\frac{\beta d}{(1-d)\alpha+d\beta}\Big)$$
$$+\theta(1-((1-d)\alpha+d\beta))h\Big(\frac{(1-\beta)d}{1-((1-d)\alpha+d\beta)}\Big) \tag{41}$$

where (40) follows from (31) - (36), Jensen's inequality, and that $\sum_{u\in\mathcal{A}}p_u(\alpha_u(1-d_u)+\beta_u d_u)=\theta(1-d)\alpha+\theta d\beta$, whereas (41) is implied by (35) and (36).

If $u\in\mathcal{A}^c$, define $\gamma_u:=p(X\neq S_1|U=u)$ and

$$\gamma:=\sum_{u\in\mathcal{A}^c}\frac{p_u}{1-\theta}\gamma_u. \tag{42}$$

Lemma 1. *For any optimal U in the binary two-sided system, $0\leq\gamma_u\leq 0.5$ for all $u\in\mathcal{A}^c$, where $\mathcal{A}$ is defined in* (29).

Proof. Assume an optimal $p(u,x,s_1,s_2)$, $f(u,s_2)$ is given, for which $\gamma_u=p(X\neq S_1|U=u)>0.5$ for some $u\in\mathcal{A}^c$. Partition $\mathcal{A}^c$ into disjoint subsets $\mathcal{A}^c=\mathcal{A}_1\cup\mathcal{A}_2\cup\mathcal{A}_3\cup\mathcal{A}_4$,

$$\mathcal{A}_1=\{u:u\in\mathcal{A}^c,\gamma_u>0.5,f(u,s_2)=s_2,\ \forall s_2\} \tag{43}$$
$$\mathcal{A}_2=\{u:u\in\mathcal{A}^c,\gamma_u\leq 0.5,f(u,s_2)=s_2,\ \forall s_2\} \tag{44}$$
$$\mathcal{A}_3=\{u:u\in\mathcal{A}^c,\gamma_u>0.5,f(u,s_2)=\bar{s}_2,\ \forall s_2\} \tag{45}$$
$$\mathcal{A}_4=\{u:u\in\mathcal{A}^c,\gamma_u\leq 0.5,f(u,s_2)=\bar{s}_2,\ \forall s_2\} \tag{46}$$

Consider a new $p'(u,x,s_1,s_2)$ and $f'(u,s_2)$ such that for all $s_1\in\mathcal{S}_1$ and $u\in\mathcal{A}^c$ we let $p'_u=p_u$ and

$$p'(S_1=s_1|U=u)=p(S_1=s_1|U=u), \tag{47}$$

if $u\in\mathcal{A}_1$ we let $f'(u,s_2)=s_2$ and

$$p'(X\neq s_1|S_1=s_1,U=u)=1-p(X\neq s_1|S_1=s_1,U=u), \tag{48}$$

if $u\in\mathcal{A}_2$ we let $f'(u,s_2)=s_2$ and

$$p'(X\neq s_1|S_1=s_1,U=u)=p(X\neq s_1|S_1=s_1,U=u), \tag{49}$$

if $u \in \mathcal{A}_3$ we let $f'(u, s_2) = s_2$ and

$$p'(X \neq s_1|S_1 = s_1, U = u) = 1 - p(X \neq s_1|S_1 = s_1, U = u), \quad (50)$$

if $u \in \mathcal{A}_4$ we let $f'(u, s_2) = \bar{s}_2$ and

$$p'(X \neq s_1|S_1 = s_1, U = u) = p(X \neq s_1|S_1 = s_1, U = u). \quad (51)$$

It can be calculated that the total contribution of $p'(u, x, s_1, s_2)$ to the expected distortion is no greater than that of $p(u, x, s_1, s_2)$, whereas the contributions of the two distributions to the rate are equal. Hence, $p(u, x, s_1, s_2)$ and $f(u, s_2)$ can be replaced with $p'(u, x, s_1, s_2)$ and function $f'(u, s_2)$ without loss of optimality, for which $\gamma_u \leq 0.5$ for all $u \in \mathcal{A}^c$ for $p'(u, x, s_1, s_2)$. Moreover, $\gamma \leq 0.5$ follows from (42). □

Using Lemma 1, we observe within set $\mathcal{A}^c$ that,

$$\sum_{u \in \mathcal{A}^c} p_u H(X|S_1, U = u) = \sum_{u \in \mathcal{A}^c} p_u h(p(S_1 \neq X|U = u)) \quad (52)$$

$$\leq (1 - \theta) h\Big(\sum_{u \in \mathcal{A}^c} \frac{p_u \gamma_u}{1 - \theta}\Big) = (1 - \theta) h(\gamma), \quad (53)$$

and that,

$$\sum_{u \in \mathcal{A}^c} p_u H(S_2|U = u) - \sum_{u \in \mathcal{A}^c} p_u H(S_1|U = u) \quad (54)$$

$$= \sum_{u \in \mathcal{A}^c} p_u H(p_2 * p(S_1|U = u)) - \sum_{u \in \mathcal{A}^c} p_u H(p(S_1|U = u)) \geq 0, \quad (55)$$

from $U - S_1 - S_2$, implied by $U - XS_1 - S_2$ and $X - S_1 - S_2$. Then, combining (27), (39), (41), (53), and (55), we obtain[1]

$$I(XS_1; U|S_2) \geq \text{RHS of (28)} = \text{RHS of (19)}.$$

We next show (20) as follows:

$$D \geq E(d(X \neq \hat{X})) \quad (56)$$

$$= p(X \neq \hat{X}, U \in \mathcal{A}) + p(X \neq \hat{X}, U \in A^c) \quad (57)$$

$$\geq \sum_{u \in \mathcal{A}} p_u d_u + (1 - \theta)\Big(p_2 * \Big(\sum_{u \in \mathcal{A}^c} \frac{p_u \gamma_u}{(1 - \theta)}\Big)\Big) \quad (58)$$

$$= \theta d + (1 - \theta)(p_2 * \gamma), \quad (59)$$

where (58) holds because if $u \in \mathcal{A}^c$ and $f(u, s_2) = s_2$, then

$$p(X \neq \hat{X}|U = u) = p(X \neq S_2|U = u) = p_2 * \gamma_u, \quad (60)$$

whereas if $u \in \mathcal{A}^c$ and $f(u, s_2) = \bar{s}_2$, then

$$p(X \neq \hat{X}|U = u) = 1 - (p_2 * \gamma_u) \geq p_2 * \gamma_u, \quad (61)$$

since $p_2 * \gamma_u \leq 0.5$ due to Lemma 1. Lastly,

$$p_1 = p(X \neq S_1) \quad (62)$$

$$= \sum_{u \in \mathcal{A}^c} p_u \gamma_u + \sum_{u \in \mathcal{A}} p_u (1 - d_u)(1 - \alpha_u) + \sum_{u \in \mathcal{A}} p_u d_u \beta_u \quad (63)$$

$$= (1 - \theta)\gamma + \theta(1 - d)(1 - \alpha) + \theta d \beta \quad (64)$$

which establishes (21) and completes the proof of converse. □

Remark 3. For $p_1 = 0$, Fig. 4 reduces to a Wyner-Ziv source with a rate distortion function $R_{\mathrm{WZ}}(D)$, in which the source and decoder side information forms a doubly symmetric binary source (DSBS) with parameter p_2. Then, $R_{\mathrm{WZ}}(D) = R_{\mathrm{TS}}(D)$ with $\alpha = 1$, $\beta = 0$, $\gamma = 0$ in (19). For $p_2 = 0$, Fig. 4 reduces to a shared information system with parameter p_1, and $R_{\mathrm{SH}}(D) = R_{\mathrm{TS}}(D)$ with $\theta = 1$, $\alpha = \beta$, $\bar{\beta} * D = p_1$, and $d = D$ in (19).

[1]RHS = right hand side

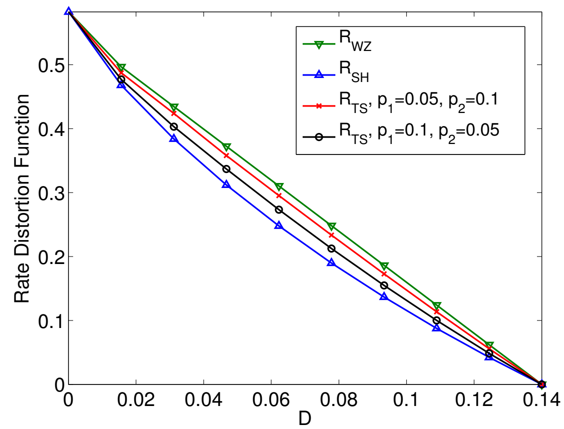

Fig. 6. Comparison of the rate distortion functions $R_{\mathrm{WZ}}(D)$, $R_{\mathrm{SH}}(D)$, and $R_{\mathrm{TS}}(D)$, for the Wyner-Ziv, shared information, and two-sided information systems, respectively, for a binary source with $p_1 * p_2 = 0.14$.

Remark 4. Numerical evaluations suggest that inequality in (20) can be replaced with an equality.

We compare (18) first with a Wyner-Ziv upper bound such that the source and decoder side information forms a DSBS with parameter $p_1 * p_2$, and then a conditional rate distortion lower bound obtained by Fig. 5, such that the source and shared information forms a DSBS with parameter $p_1 * p_2$. We also present our two-sided results in Fig. 6 for two sets of p_1 and p_2 for which $p_1 * p_2$ are equal. Fig. 6 indicates an inherent hierarchy between the encoder and the decoder side information, such that having a less noisy side information at the decoder is more essential than at the encoder.

VI. Conclusion

We have considered the rate distortion problem for a remote source with side information, and studied the conditions under which two-sided or shared information is useful. As a special case, we have completely characterized the direct binary source with two-sided side information.

References

[1] C. E. Shannon, “Coding theorems for a discrete source with a fidelity criterion,” *IRE Nat. Conv. Rec*, vol. 4, no. 142-163, p. 1, 1959.

[2] A. D. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Trans. on Inf. Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[3] R. M. Gray, *Conditional rate-distortion theory*. Inf. Sys. Laboratory, Stanford Electronics Laboratories, 1972.

[4] T. M. Cover and M. Chiang, “Duality between channel capacity and rate distortion with two-sided state information,” *IEEE Trans. on Inf. Theory*, vol. 48, no. 6, pp. 1629–1638, 2002.

[5] M. Fleming and M. Effros, “On rate-distortion with mixed types of side information,” *IEEE Trans. on Inf. Theory*, vol. 52, no. 4, pp. 1698–1705, 2006.

[6] H. Yamamoto and K. Itoh, “Source coding theory for multiterminal communication systems with a remote source,” *IEICE Trans.*, vol. 63, no. 10, pp. 700–706, 1980.

[7] C. Tian and S. N. Diggavi, “On multistage successive refinement for wyner–ziv source coding with degraded side informations,” *IEEE Trans. on Inf. Theory*, vol. 53, no. 8, pp. 2946–2960, 2007.