Lossy Source Coding with Gaussian or Erased Side-Information

EPFL Technical Report LICOS-REPORT-2009-001

Etienne Perron, Suhas Diggavi, Emre Telatar
EPFL, Lausanne, Switzerland
Email: {etienne.perron,suhas.diggavi,emre.telatar}@epfl.ch

***Abstract*—In this paper we find properties that are shared between two seemingly unrelated lossy source coding setups with side-information. The first setup is when the source and side-information are jointly Gaussian and the distortion measure is quadratic. The second setup is when the side-information is an erased version of the source. We begin with the observation that in both these cases the Wyner-Ziv and conditional rate-distortion functions are equal. We further find that there is a continuum of optimal strategies for the conditional rate distortion problem in both these setups. Next, we consider the case when there are two decoders with access to different side-information sources. For the case when the encoder has access to the side-information we establish bounds on the rate-distortion function and a sufficient condition for tightness. Under this condition, we find a characterization of the rate-distortion function for physically degraded side-information. This characterization holds for both the Gaussian and erasure setups.**

I. Introduction

Lossy source coding with side-information available at the encoder and the decoder is a well studied problem with applications in the distribution of correlated pieces of data, for instance in video coding. When there is only one decoder, the problem is simply a conditional version of the rate-distortion problem. In [8], Wyner and Ziv solved the case when only the decoder has access to the side-information. They showed that for Gaussian sources and quadratic distortion, the rate-distortion tradeoff of their problem is the same as for the conditional rate-distortion problem. In [1], Heegard and Berger considered the Wyner-Ziv problem for several decoders, and solved it for the case when the side-information sources are stochastically degraded. In [2], Kaspi provided the optimal tradeoff for the same problem, but with only one decoder having access to side-information. He also solved a version of this particular problem where the encoder knows the side-information. In [5], we considered the problem of lossy source coding for two decoders, each of which has access to a side-information source, and where the encoder has full knowledge of the side-information. The encoder sends the same message to both decoders. This is a generalization of the problem studied by Kaspi and we refer here to this problem as "source coding for two informed decoders". In [5], we solved the Gaussian version of this problem for physically degraded side-information.

In this paper, we extend the existing results in several ways. First, we introduce discrete sources with "erased" side-information, *i.e.*, where the side-information source is the output of an erasure channel whose input is the data source. We show that with this type of side-information, discrete memoryless sources have an intimate connection, in terms of properties, to Gaussian sources with Gaussian side-information. We first show that the Wyner-Ziv and conditional rate-distortion functions are equal for discrete sources with erased side-information[1]. For all other results, we focus on the "binary-erasure" special case, where the data source is binary. We show that for the conditional rate-distortion problem, there is a continuum of optimal strategies for both the Gaussian and the binary-erasure case. This fact turns out to be useful in source coding for two informed decoders. For this problem, we give upper and lower bounds on the rate-distortion function, as well as a sufficient condition for equality of these bounds. Using this condition, the Gaussian result for physically degraded side-information in [5] is easy to prove. We also show an analogous result for the binary-erasure case, which shows another connection between the two cases. As an auxiliary result, we compute the binary-erasure rate-distortion function for the Kaspi problem.

The paper is organized as follows. In Section III we state and discuss the main results. Section IV contains a few additional results about the Kaspi problem. These results are somewhat auxiliary, but important because they are used in the proofs of the main results. Section V contains these proofs. Because of space limitations, we omit some of the proofs and for some results, we provide only a rough proof outline. The detailed proofs can be found in several technical reports available online ([3],[6],[4]).

II. Terminology

Definition 1: A *Gaussian* source coding problem is a setup with a source X and (one or) two side-information sources Y and Z, where (X, Y, Z) are real, jointly Gaussian random variables. All reconstruction alphabets are real and we use the quadratic distortion measure $d(x, \hat{x}) = (x - \hat{x})^2$ (for all reconstructions).

[1] This result was found independently and simultaneously in [7] and by us in [4].

Definition 2: A *binary-erasure* source coding problem is a setup with a source X and (one or) two side-information sources Y and Z, where X is a Bernoulli-$\frac{1}{2}$ random variable, and Y and Z are the outputs of two binary erasure channels (BECs) whose input is X. The two BECs may be correlated. All reconstruction alphabets are binary and the distortion measure is the Hamming distance $d(x, \hat{x}) = x \oplus \hat{x}$ (for all reconstructions), where $\oplus$ denotes modulo-2 addition over the binary field.

Definition 3: For a given source coding problem, the rate-distortion function expresses the smallest rate for which the rate distortion pair (or triple) is achievable for the given distortion(s).

III. Main Results

The five theorems contained in this section constitute the main results of this paper. All the results hold in an analogous way for both the binary-erasure and the Gaussian case, hence demonstrating the connection between the two cases. To the best of our knowledge, all the results in this section are new, with the exception of Theorem 1 which was found independently and simultaneously in [7] and in [4]. Also, the Gaussian part of Theorem 5 was already published in [5].

A. One Decoder

To convince the reader how similar the binary-erasure and the Gaussian setups are, we first state two important results for the case when there is only one decoder. It is well-known that in the Gaussian case, the rate-distortion function for a single decoder with side-information is

$$R_{X|Y}(D) = R^{\mathrm{WZ}}(D) = \frac{1}{2}\log\frac{\mathrm{Var}\,(X|Y)}{D},$$

no matter whether Y is available at the encoder or not. We find that for the erasure case, a similar property holds:

Theorem 1: When X is a source taking values in a discrete set $\mathcal{X}$ and Y is an erased version of X, *i.e.*,

$$Y = \begin{cases} X & \text{w.p. } 1-p \\ \epsilon & \text{w.p. } p, \end{cases}$$

then we have the following: Provided that the reconstruction alphabet is $\mathcal{X}$ and the distortion measure $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+$ is such that $d(x, x) = 0$ for all $x \in \mathcal{X}$, the rate-distortion function for a single decoder with side-information is

$$R_{X|Y}(D) = R^{\mathrm{WZ}}(D) = pR_X(\frac{D}{p}),$$

no matter whether Y is available at the encoder or not.
Here, $R_X(\cdot)$ is the rate-distortion function of the source X under the distortion measure d, when the decoder has no side-information. To the best of our knowledge, there is no other case in which the conditional and Wyner-Ziv rate-distortion functions match for a discrete source. The proof of Theorem 1 can be found in the appendix of a technical report [4]. Note that Theorem 1 holds for arbitrary, discrete sources. In the rest of this paper, we focus on binary-erasure problems, where the source X takes values in $\{0, 1\}$. In this case, Theorem 1 becomes:

Corollary 1: For the binary-erasure source coding problem with one decoder and only one side-information Y, we have

$$R_{X|Y}(D) = R^{\mathrm{WZ}}(D) = p(1 - h(\frac{D}{p})),$$

where $h(\cdot)$ is the binary entropy function.

When the encoder knows the side-information, one can either use the conditional rate-distortion scheme or the Wyner-Ziv scheme, which is also applicable because the encoder can ignore the side-information. Our second result for one decoder shows that there are actually infinitely many different schemes who are all optimal in this case. The conditional rate-distortion function can be written as $R_{X|Y}(D) = \min_{p_{W|X,Y}} \big(I(X, Y; W) - I(W; Y)\big)$, where W takes values in the reconstruction alphabet, and $p_{W|X,Y}$ is such that the distortion requirement can be satisfied by some estimator $\hat{X} = g(W, Y)$.

Theorem 2: For the Gaussian case and the binary-erasure case, $R_{X|Y}(D)$ can be achieved by a continuum of auxiliary random variables W (taking values in the reconstruction alphabet $\mathbb{R}$ and $\{0, 1\}$), with $I(W; Y)$ varying in $[0, \infty]$ and $[0, 1-p]$ for the Gaussian and the binary-erasure case, respectively.
The proof of Theorem 2 can be found in Section V.
Remark 1: When $I(W; Y) = 0$, then W corresponds to a lossy description of the "innovation", *i.e.*, of the quantity $X - f(Y)$, where $f(Y)$ is the best estimator of X from Y.
Remark 2: The continuum contains one choice of W for which $W \multimap X \multimap Y$ is a Markov chain. This choice of W corresponds to the Wyner-Ziv scheme [8].

B. Two Informed Decoders

Consider now the rate-distortion setup depicted in Figure 1. The source X is to be described at distortion D_1 and D_2, respectively to two decoders who have access to different side-information sources Y and Z, respectively. Both Y and Z are also available at the encoder. We refer to this problem as "source coding for two informed decoders", and we denote the corresponding rate-distortion function by $R(D_1, D_2)$. In the following, we present an upper and a lower bound on $R(D_1, D_2)$, a sufficient condition for equality of the bounds, and we show that this condition holds when the side-information is physically degraded. All of these results are valid for both the binary-erasure and the Gaussian case.

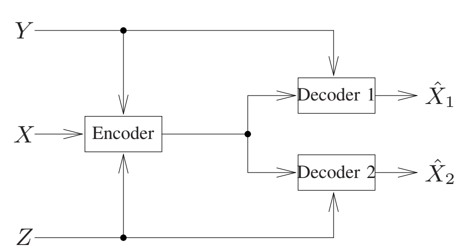

Fig. 1. Source coding for two informed decoders.

In [5], we presented an achievable rate-distortion region for the problem of source coding for two informed decoders when (X, Y, Z) is a general discrete memoryless multisource (Theorem 1 of that publication). For the Gaussian and binary-erasure versions of the problem, we simplify that region by dropping the auxiliary random variables U and V. The result is the following upper bound on the rate-distortion function, stated as a corollary to Theorem 1 in [5]:

Corollary 2: In the Gaussian and the binary-erasure case, the rate-distortion function for two informed decoders $R(D_1, D_2)$ is at least as small as

$$R^+(D_1, D_2) = \min_{p_{W|XYZ} \in \mathcal{A}(D_1,D_2)} \max\{I(X, Y; W|Z), I(X, Z; W|Y)\} \quad (1)$$

where $\mathcal{A}(D_1, D_2)$ is the set of all conditional distributions of random variables W such that

- W is Gaussian or binary symmetric, respectively, and jointly distributed with (X, Y, Z),
- $\exists\ \hat{X}_1(W, Y)$ such that $\mathrm{E}\left[d(\hat{X}_1, X)\right] \leq D_1$,
- $\exists\ \hat{X}_2(W, Z)$ such that $\mathrm{E}\left[d(\hat{X}_2, X)\right] \leq D_2$.

Next, we provide a lower bound on $R(D_1, D_2)$.

Theorem 3: In the Gaussian and the binary-erasure case, the rate-distortion function for two informed decoders $R(D_1, D_2)$ is lower bounded by

$$R^-(D_1, D_2) = \max\Big\{ \min_{p_{W|XYZ} \in \mathcal{A}(D_1,D_2)} I(X, Y; W|Z), \min_{p_{W|XYZ} \in \mathcal{A}(D_1,D_2)} I(X, Z; W|Y)\Big\}, \quad (2)$$

where $\mathcal{A}(D_1, D_2)$ is defined as in Corollary 2.

This lower bound is obtained by assuming that a genie makes one of the side-information sources available to the decoder who does not know it, followed by additional steps. In [5], we developed a similar lower bound for the Gaussian case. The proof of Theorem 3 can be found in Section V.

We find that in some cases, $R^-(D_1, D_2)$ and $R^+(D_1, D_2)$ match, and in the next result, we provide a sufficient condition for this equality.

Definition 4: Let $\mathcal{S}^* \subseteq \mathcal{A}(D_1, D_2)$ be the set of all optimizers of the first minimization in (2), *i.e.*, $p_{W^*|XYZ} \in \mathcal{S}^*$ if and only if

$$I(X, Y; W^*|Z) = \min_{p_{W|XYZ} \in \mathcal{A}(D_1,D_2)} I(X, Y; W|Z). \quad (3)$$

Similarly, let $\mathcal{S}^{**}$ be the set of all optimizers of the second minimization in (2).

Theorem 4: When $I(W^*; Z) \leq I(W^*; Y)$ for some $p_{W^*|XYZ} \in \mathcal{S}^*$ or $I(W^{**}; Z) \geq I(W^{**}; Y)$ for some $p_{W^{**}|XYZ} \in \mathcal{S}^{**}$, then $R^+(D_1, D_2) = R^-(D_1, D_2)$.

The proof of Theorem 4 can be found in Section V.

This theorem is particularly useful to find an exact characterization of the rate-distortion function for physically degraded side-information.

Definition 5: In source coding for two informed decoders, we say that the side-information is *physically degraded* if (X, Y, Z) forms a Markov chain in either of the orders $X \multimap Y \multimap Z$ or $X \multimap Z \multimap Y$.

Theorem 5: In source coding for two informed encoders with physically degraded side-information, the rate-distortion function is $R(D_1, D_2) = R^+(D_1, D_2)$ for both the Gaussian and the binary-erasure case.

This theorem generalizes a Gaussian result that was presented in [5]. A rather lengthy proof of that Gaussian result was provided in [6]. The novelty here is that the same result also holds for the binary-erasure case, and that Theorem 4 can be used to obtain a relatively simple proof for both the Gaussian and the binary-erasure case. The proof can be found in Section V.

Application of Theorem 2: In [6], we noted that if (X, Y) has the same statistics as (X, Z), then a Wyner-Ziv code, optimized for the decoder with the more strict distortion requirement, is optimal. Theorem 2 provides us with an interesting extension of this idea: Pick one of the two decoders, say Decoder 1, and implement a scheme that achieves the conditional rate-distortion function for Decoder 1, $R_{X|Y}(D_1)$. Different random variables W out of the continuum described in Theorem 2 are more or less useful for Decoder 2, *i.e.*, for estimating X from (W, Z). To find the best out of all the strategies that have rate $R_{X|Y}(D_1)$, one should find the W from the continuum that is most useful for Decoder 2 and that satisfies $I(W; Y) \leq I(W; Z)$. This last condition is required to ensure that although the encoder uses a binning scheme for Decoder 1, Decoder 2 is also able to identify the correct member of the bin described by the message. A more detailed study of this technique is left as future work.

Bound on the Gap: If we compare the upper and lower bounds given in Corollary 2 and Theorem 3, respectively, we notice that the only difference between upper and lower bound is that the minimization and the maximization are inverted. This fact can be used to bound the gap between the two bounds. For instance, assume that $I(X, Y; W^*|Z) \geq I(X, Z; W^{**}|Y)$, where W^* and W^{**} are as in Definition 4. Then,

$$\begin{aligned} R^+(D_1, D_2) &- R^-(D_1, D_2) \\ &\leq \max\{I(X, Y; W^*|Z), I(X, Z; W^*|Y)\} \\ &\quad - I(X, Y; W^*|Z) \\ &= \big[I(X, Z; W^*|Y) - I(X, Y; W^*|Z)\big]^+ \\ &= \big[I(W^*; Z) - I(W^*; Y)\big]^+. \end{aligned}$$

IV. Auxiliary Results

In [2], Kaspi found the rate-distortion function to a setup where a source X is encoded for two decoders. One decoder has access to a side-information source Y, while the other decoder is uninformed. The encoder knows Y. We call this problem the Kaspi problem. The Kaspi problem is a special case of source coding for two informed decoders, in which one of the side-information sources is constant (Figure 1 with Z = constant). The results in this section could be viewed as a special case of Theorem 5. However, we present them as independent results, and we will use them in Section V to prove our main results, including Theorem 5.

In [5], we computed the rate-distortion function for the Gaussian Kaspi problem. We repeat that result here:

Theorem 6: (Theorem 2 in [5].) Let $X \sim \mathcal{N}(0, \sigma_X^2)$, and $Y = X + N$, with $N \sim \mathcal{N}(0, \sigma_N^2)$ independent of X. In the Gaussian Kaspi problem defined by these sources, the rate-distortion function $R_{\text{Kaspi}}(D_1, D_2)$ is given by the following four regimes:
1) if $D_1 \geq \frac{\sigma_X^2 \sigma_N^2}{\sigma_X^2 + \sigma_N^2}$ and $D_2 \geq \sigma_X^2$, then $R_{\text{Kaspi}}(D_1, D_2) = 0$,
2) if $D_1 \geq \frac{D_2 \sigma_N^2}{D_2 + \sigma_N^2}$ and $D_2 < \sigma_X^2$, then
$R_{\text{Kaspi}}(D_1, D_2) = \frac{1}{2} \log \frac{\sigma_X^2}{D_2}$,
3) if $D_1 < \frac{\sigma_X^2 \sigma_N^2}{\sigma_X^2 + \sigma_N^2}$ and $D_2 \geq D_1 + \frac{\sigma_X^4}{\sigma_X^2 + \sigma_N^2}$, then
$R_{\text{Kaspi}}(D_1, D_2) = \frac{1}{2} \log \frac{\sigma_X^2 \sigma_N^2}{D_1(\sigma_X^2 + \sigma_N^2)}$,
4) otherwise, $R_{\text{Kaspi}}(D_1, D_2) = \frac{1}{2} \log \frac{\sigma_X^2}{D_2(1-\rho_0^2)}$,
where $\rho_0 = \frac{\sigma_X D_1}{\sigma_N (\sigma_X^2 - D_1) D_2} \Phi$ and

$$\Phi = \sqrt{D_2(\sigma_X^2 - D_2)} - \sqrt{(\sigma_X^2 \frac{D_2}{D_1}(\sigma_N^2 - D_1) - \sigma_N^2 D_2)(\frac{D_2}{D_1} - 1)}.$$

The detailed proof of this theorem can be found in [3].

It turns out that the binary-erasure Kaspi problem is very similar to its Gaussian counterpart. In particular, we have the following result:

Theorem 7: In the binary-erasure Kaspi problem, the rate-distortion function $R_{\text{Kaspi}}(D_1, D_2)$ is given by the following four regimes (where p is the erasure probability of the BEC):
1) if $D_1 \geq \frac{p}{2}$ and $D_2 \geq \frac{1}{2}$, then $R_{\text{Kaspi}}(D_1, D_2) = 0$,
2) if $D_1 \geq p D_2$ and $D_2 < \frac{1}{2}$, then
$R_{\text{Kaspi}}(D_1, D_2) = 1 - h(D_2)$,
3) if $D_1 < \frac{p}{2}$ and $D_2 \geq D_1 + \frac{1-p}{2}$, then
$R_{\text{Kaspi}}(D_1, D_2) = p\big(1 - h(\frac{D_1}{p})\big)$,
4) otherwise,

$$R_{\text{Kaspi}}(D_1, D_2) = 1 - h(\frac{D_2 - D_1}{1-p}) + p\Big(h(\frac{D_2 - D_1}{1-p}) - h(\frac{D_1}{p})\Big).$$

The detailed proof of this theorem can be found in [4].

Corollary 3: In both the Gaussian case and the binary-erasure case, the rate-distortion function of the Kaspi problem is

$$R_{\text{Kaspi}}(D_1, D_2) = \min_{p_{W|XY}} I(X, Y; W),$$

where the indicated minimization is over all auxiliary random variables W such that

- W is Gaussian or binary symmetric, respectively, and jointly distributed with (X, Y),
- $\exists$ a function $\hat{X}_1(W, Y)$ such that $\mathrm{E}\left[d(\hat{X}_1, X)\right] \leq D_1$,
- $\exists$ a function $\hat{X}_2(W)$ such that $\mathrm{E}\left[d(\hat{X}_2, X)\right] \leq D_2$.

The proof of this corollary is given in the appendix.

V. Proofs of the Main Results

A. Sketch of the proof of Theorem 2

In the Gaussian case, W can be parametrized by 3 real parameters. Since scaling of W by a constant changes neither the rate nor the distortion, 2 relevant paremeters remain. However, since $I(X; W|Y) = \frac{1}{2} \log \frac{\mathrm{Var}(X|Y)}{\mathrm{Var}(X|W,Y)}$, we see that any choice of W that satisfies $\mathrm{Var}\,(X|W, Y) = D$ is optimal. After imposing this constraint, one of the parameters that specify W is still free, hence the continuum. In the binary-erasure case, W can be parametrized by 2 crossover-probabilities connecting X and W (for the two cases when Y is an erasure or not). It is clear that when Y is not an erasure, *i.e.*, when $Y = X$, the best estimate of X given (W, Y) should be Y. Hence, one of the two parameters has no importance in computing the expected distortion achieved by the optimal scheme. In addition, the same parameter plays no role in the rate expression $I(X; W|Y)$, either. Hence the continuum.

B. Proof of Theorem 3

If a genie makes Z available to Decoder 1, we obtain a new setup that we call the "conditional Kaspi setup". In this new setup, Decoder 1 is more powerful than before, and hence, the rate-distortion function of the conditional Kaspi setup gives a lower bound on $R(D_1, D_2)$.

Note that in the conditional Kaspi setup, all three entities have access to Z. Hence, the problem is the same as the Kaspi problem for sources $(\tilde{X}, \tilde{Y})$ that correspond to "(X, Y) conditioned on Z".

From Corollary 3, we conclude that the rate-distortion function of the conditional Kaspi problem can be written as

$$R_Z^-(D_1, D_2) = \min_{p_{W|XYZ} \in \mathcal{B}(D_1, D_2)} I(X, Y; W|Z), \qquad (4)$$

where $\mathcal{B}(D_1, D_2)$ is the set of all conditional distributions of random variables W such that

- W is Gaussian or binary symmetric, respectively, and jointly distributed with (X, Y, Z),
- $\exists$ a function $\hat{X}_1(W, Y, Z)$ such that $\mathrm{E}\left[d(\hat{X}_1, X)\right] \leq D_1$,
- $\exists$ a function $\hat{X}_2(W, Z)$ such that $\mathrm{E}\left[d(\hat{X}_2, X)\right] \leq D_2$.

We can apply the exact same reasoning also for the case when the genie makes Y available to Decoder 2. In this case, we obtain a different lower bound on $R(D_1, D_2)$, namely

$$R_Y^-(D_1, D_2) = \min_{p_{W|XYZ} \in \mathcal{C}(D_1, D_2)} I(X, Z; W|Y),$$

where $\mathcal{C}(D_1, D_2)$ is the symmetric counterpart of $\mathcal{B}(D_1, D_2)$.

The following lemma provides a simplification of the lower bounds given so far.

Lemma 1: In $R_Z^-(D_1, D_2)$ and in $R_Y^-(D_1, D_2)$, the sets $\mathcal{B}(D_1, D_2)$ and $\mathcal{C}(D_1, D_2)$ can both be replaced by $\mathcal{A}(D_1, D_2)$, without changing the outcome of the optimizations.

The proof of Lemma 1 can be found in the appendix. By combining the two bounds $R_Z^-(D_1, D_2)$ and $R_Y^-(D_1, D_2)$, the claim of the theorem follows.

C. Proof of Theorem 4

From (2), we know that

$$R_Z^-(D_1, D_2) = \min_{\mathcal{A}(D_1, D_2)} \left[I(X, Y, Z; W) - I(Z; W)\right]. \qquad (5)$$

is a lower bound on $R(D_1, D_2)$. The upper bound (1), on the other hand, can be written as

$$R^+(D_1, D_2) = \min_{p_{W|XYZ} \in \mathcal{A}(D_1,D_2)} \Big[I(X,Y,Z;W) - \min\{I(W;Z), I(W;Y)\} \Big]. \quad (6)$$

Let W^* be a random variable whose distribution is a minimizer of (5), and assume that $I(W^*;Z) \le I(W^*;Y)$ Then,

$$\begin{aligned} R^-(D_1, D_2) &\ge R_Z^-(D_1, D_2) \\ &= I(X,Y,Z;W^*) - I(Z;W^*) \\ &= I(X,Y,Z;W^*) - \min\{I(W^*;Z), I(W^*;Y)\} \\ &\ge R^+(D_1, D_2) \end{aligned}$$

where we used (6) in the last inequality. From Corollary 2 and Theorem 3, it is clear that $R^+(D_1, D_2) \ge R^-(D_1, D_2)$. Hence, $R^-(D_1, D_2) = R^+(D_1, D_2)$.

When $I(W^{**};Z) \ge I(W^{**};Y)$, where W^{**} is as defined in Definition 4, an analogous argument holds, with $R_Z^-(D_1, D_2)$ replaced by $R_Y^-(D_1, D_2)$.

D. Proof of Theorem 5

We only prove the result for the Markov chain $X -\!\!\circ\!\!- Y -\!\!\circ\!\!- Z$. The proof for the other Markov chain is analogous.

Gaussian case: W.l.o.g., we can assume that $X \sim \mathcal{N}(0, \sigma_X^2)$, $Y = X + N_1$ and $Z = Y + N_2$, where $N_i \sim \mathcal{N}(0, \sigma_i^2)$, $i = 1, 2$ are Gaussian random variables, independent of X and of each other. We parametrize W as

$$W = aX + bN_1 + cN_2 + \xi,$$

where $\xi \sim \mathcal{N}(0, 1)$, and ξ is independent of everything else. Using Lagrange multipliers, we find that (3) is optimized by a W^* for which $c = 0$. It follows that $W^* -\!\!\circ\!\!- (X,Y) -\!\!\circ\!\!- Z$. This, together with Lemma 2 below and Theorem 4, lets us conclude the Gaussian part of the theorem. Details regarding the computation of the Lagrange equations can be found in [6].

Binary-erasure case: Let X be a Bernoulli random variable with mean $\frac{1}{2}$. Let Y be the output of a BEC with erasure probability p_1, when X is the input, and let Z be the output of a BEC with erasure probability p_2, when Y is the input. Any Bernoulli-$\frac{1}{2}$ random variable W jointly distributed with (X,Y,Z) can be expressed using three parameters (q, r, s):

$$\begin{aligned} q &= \mathbf{P}(W \neq X | Y \neq \epsilon, Z \neq \epsilon) \\ r &= \mathbf{P}(W \neq X | Y \neq \epsilon, Z = \epsilon) \\ s &= \mathbf{P}(W \neq X | Y = \epsilon, Z = \epsilon). \end{aligned}$$

In other words, depending on the values of (Y,Z), X is connected to W through one of three virtual BSC's that have crossover probabilities q, r and s, respectively. Using the parameters (q, r, s), the objective function in (3) can be written as

$$\begin{aligned} I(X,Y;W|Z) &= I(X,Y,Z;W) - I(W;Z) \\ &= H(W|Z) - H(W|X,Y,Z) \\ &= p_1 + (1-p_1)p_2 \\ &\quad - (1-p_1)p_2 h(r) - p_1 h(s), \end{aligned} \quad (7)$$

where the last equality follows after the terms containing $h(q)$ have cancelled out. The best estimate of X given (W,Y) is simply $\hat{X} = Y$ whenever $Y \neq \epsilon$ and $\hat{X} = W$ otherwise. Hence, the distortion constraint for Decoder 1 is

$$\mathbf{P}(\hat{X}_1 \neq X) = p_1 s \le D_1. \quad (8)$$

Likewise, for Decoder 2, we require

$$\mathbf{P}(\hat{X}_2 \neq X) = p_1 s + (1-p_1)p_2 r \le D_2. \quad (9)$$

The parameter q figures in none of (7), (8) and (9). Hence, q can be freely chosen in the optimization. In particular, one optimal solution to (3) is a W^* such that $q = r^*$, where (r^*, s^*) are the optimizers of (7) subject to the distortion constraints (8) and (9). For that particular choice of q, we have $W^* -\!\!\circ\!\!- (X,Y) -\!\!\circ\!\!- Z$. This, together with Lemma 2 and Theorem 4, concludes the binary-erasure part of the theorem.

Lemma 2: Let (W,X,Y,Z) be arbitrary random variables such that the two Markov chains $W -\!\!\circ\!\!- (X,Y) -\!\!\circ\!\!- Z$ and $X -\!\!\circ\!\!- Y -\!\!\circ\!\!- Z$ are satisfied. Then, $I(W;Z) \le I(W;Y)$.

The proof of this lemma is given in the appendix.

VI. Conclusion

The binary-erasure setting is promising as a tool to analyze problems in source coding with side-information. All the results that we provide for the Gaussian case hold in a analogous way for the binary-erasure case. In addition, the binary-erasure case is often more easy to analyze. This is due to the fact that in the binary-erasure case, the side-information is either perfect or completely useless. The results here suggest that there may be more connections between the Gaussian and erasure setups and it is likely that one may find further analogies between them. Such connections may provide insights into previously unresolved questions.

For the problem of source coding for two informed decoders, we intend to provide a more complete discussion on how to apply Theorem 2 in future work.

References

[1] C. Heegard and T. Berger, *Rate distortion when side information may be absent*, IEEE Trans. Inform. Theory, 31 (1985), pp. 727–734.

[2] A. Kaspi, *Rate-distortion function when side-information may be present at the decoder*, IEEE Trans. Inform. Theory, 40 (1994), pp. 2031–2034.

[3] E. Perron, S. Diggavi, and E. Telatar, *The Kaspi rate-distortion problem with encoder side-information: Gaussian case*, Tech. Rep. LICOS-REPORT-2005-001, École Polytechnique Fédérale de Lausanne, Switzerland, Nov. 2005. http://licos.epfl.ch/index.php?p=pubs&l=en.

[4] ———, *The Kaspi rate-distortion problem with encoder side-information: Binary erasure case*, Tech. Rep. LICOS-REPORT-2006-004, École Polytechnique Fédérale de Lausanne, Switzerland, Nov. 2006. http://licos.epfl.ch/index.php?p=pubs&l=en.

[5] ———, *On the role of side-information in source coding for multiple decoders*, in Proc. of the IEEE Int. Symposium on Inform. Theory, Seattle, USA, 2006.

[6] ———, *On the role of side-information in source coding for multiple decoders*, Tech. Rep. LICOS-REPORT-2006-003, École Polytechnique Fédérale de Lausanne, Switzerland, Aug. 2006. http://licos.epfl.ch/index.php?p=pubs&l=en.

[7] S. Verdu and T. Weissman, *The information lost in erasures*, IEEE Trans. Inform. Theory, 54 (2008), pp. 5030–5058.

[8] A. Wyner and J. Ziv, *The rate-distortion function for source coding with side information at the decoder*, IEEE Trans. Inform. Theory, 22 (1976).

VII. Appendix

A. Proof of Corollary 3

For the Gaussian case, we have:

1) if $D_1 \geq \frac{\sigma_X^2 \sigma_N^2}{\sigma_X^2+\sigma_N^2}$ and $D_2 \geq \sigma_X^2$, then $W = 0$,

2) if $D_1 \geq \frac{D_2 \sigma_N^2}{D_2+\sigma_N^2}$ and $D_2 < \sigma_X^2$, then $W = \sqrt{\frac{\sigma_X^2 - D_2}{\sigma_X^2 D_2}} X + \xi$,

3) if $D_1 < \frac{\sigma_X^2 \sigma_N^2}{\sigma_X^2+\sigma_N^2}$ and $D_2 \geq D_1 + \frac{\sigma_X^4}{\sigma_X^2+\sigma_N^2}$, then
$W = \sqrt{\frac{1}{D_1} - \frac{(\sigma_X^2+\sigma_N^2)}{\sigma_X^2 \sigma_N^2}} \Big(- \frac{\sigma_N^2}{\sigma_X^2+\sigma_N^2} X + \frac{\sigma_X^2}{\sigma_X^2+\sigma_N^2} N \Big) + \xi$,

4) otherwise, $W = \frac{1}{\sqrt{1-\rho_0^2}} \Big(\sqrt{\frac{\sigma_X^2 - D_2}{\sigma_X^2 D_2}} X + \frac{\rho_0}{\sigma_N} N + \xi$.

Here, ξ is independent of (X, Y) and $\xi \sim \mathcal{N}(0,1)$.

For the binary-erasure case, we have:

1) if $D_1 \geq \frac{p}{2}$ and $D_2 \geq \frac{1}{2}$, then $W = \begin{cases} \bar{X} & \text{w.p. } \frac{1}{2} \\ X & \text{w.p. } \frac{1}{2}, \end{cases}$

2) if $D_1 \geq p D_2$ and $D_2 < \frac{1}{2}$, then

$$W = \begin{cases} \bar{X} & \text{w.p. } D_2 \\ X & \text{w.p. } 1 - D_2, \end{cases}$$

3) if $D_1 < \frac{p}{2}$ and $D_2 \geq D_1 + \frac{1-p}{2}$, then

$$W = \begin{cases} 0 & \text{w.p. } \frac{1}{2} & \text{if } Y \neq \epsilon \\ 1 & \text{w.p. } \frac{1}{2} & \text{if } Y \neq \epsilon \\ \bar{X} & \text{w.p. } \frac{D_1}{p} & \text{if } Y = \epsilon \\ X & \text{w.p. } 1 - \frac{D_1}{p} & \text{if } Y = \epsilon. \end{cases}$$

4) otherwise,

$$W = \begin{cases} \bar{X} & \text{w.p. } \frac{D_2 - D_1}{1-p} & \text{if } Y \neq \epsilon \\ X & \text{w.p. } 1 - \frac{D_2 - D_1}{1-p} & \text{if } Y \neq \epsilon \\ \bar{X} & \text{w.p. } \frac{D_1}{p} & \text{if } Y = \epsilon \\ X & \text{w.p. } 1 - \frac{D_1}{p} & \text{if } Y = \epsilon. \end{cases}$$

Above, $\bar{x}$ stands for $x \oplus 1$. It can be verified that by plugging the above choices of W into the expression $I(X, Y; W)$, one obtains the rate-distortion trade-off given in Theorems 6 and 7. By inspecting the proof of Theorem 6 in [3] and the proof of Theorem 7 in [4], one can verify that the above choices of W also satisfy the distortion requirements.

B. Proof of Lemma 1

We only prove the result for $R_Z^-(D_1, D_2)$; the proof for $R_Y^-(D_1, D_2)$ follows by the symmetry of the setup. Let

$$r_A(D_1, D_2) \triangleq \min_{p_{W|XYZ} \in \mathcal{A}(D_1, D_2)} I(X, Y; W|Z) \tag{10}$$

and

$$r_B(D_1, D_2) \triangleq \min_{p_{W|XYZ} \in \mathcal{B}(D_1, D_2)} I(X, Y; W|Z). \tag{11}$$

The aim is to show that $r_A(D_1, D_2) = r_B(D_1, D_2)$. Note that if for a given W, there exists a function $\hat{X}_1(W, Y)$ such that $\mathrm{E}\left[d(\hat{X}_1, X)\right] \leq D_1$, then there exists also a function $\hat{X}_1(W, Y, Z)$ with that same property. Hence, $\mathcal{A}(D_1, D_2) \subseteq \mathcal{B}(D_1, D_2)$ and therefore,

$$r_B(D_1, D_2) \leq r_A(D_1, D_2).$$

It remains to show that $r_A(D_1, D_2) \leq r_B(D_1, D_2)$.

Gaussian case:
Let $\breve{W}$ be the minimizer of (11). Since $(\breve{W}, Y, Z)$ are jointly Gaussian, the best estimate of X given $(\breve{W}, Y, Z)$ (computed at Decoder 1) is a linear combination $\hat{X}_1 = a\breve{W} + bY + cZ$. Define $\tilde{W} = a\breve{W} + cZ$. Using Y and $\tilde{W}$, Decoder 1 can produce the same estimate $\hat{X}_1$, and therefore,

$$\text{Var}\left(X|\tilde{W}, Y\right) = \text{Var}\left(X|\breve{W}, Y, Z\right) \leq D_1.$$

In addition,

$$\begin{aligned} \text{Var}\left(X|\tilde{W}, Z\right) &= \text{Var}\left(X|a\breve{W} + cZ, Z\right) \\ &= \text{Var}\left(X|\breve{W}, Z\right) \leq D_2. \end{aligned}$$

Hence, $\tilde{W} \in \mathcal{A}(D_1, D_2)$. In addition,

$$\begin{aligned} I(X, Y; \tilde{W}|Z) &= H(X, Y|Z) - H(X, Y|Z, a\breve{W} + cZ) \\ &= H(X, Y|Z) - H(X, Y|Z, \breve{W}) \\ &= I(X, Y; \breve{W}|Z). \end{aligned}$$

Hence,

$$\begin{aligned} r_A(D_1, D_2) &\leq I(X, Y; \tilde{W}|Z) \\ &= I(X, Y; \breve{W}|Z) \\ &= r_B(D_1, D_2). \end{aligned}$$

Binary-erasure case:
Again, let $\breve{W}$ be the minimizer of (11). The optimal reconstruction functions should be

$$\hat{X}_1 = \begin{cases} \breve{W} & \text{if } Y = Z = \epsilon \\ Z & \text{if } Y = \epsilon, Z \neq \epsilon \\ Y & \text{if } Y \neq \epsilon \end{cases}$$

and

$$\hat{X}_2 = \begin{cases} \breve{W} & \text{if } Z = \epsilon \\ Z & \text{if } Z \neq \epsilon. \end{cases}$$

This is due to the fact that when the side-information sources are erasures, the optimal estimate of X given $\breve{W}$ is $\breve{W}$ itself. Assume that $\breve{W}$ is such that

$$\begin{aligned} \mathbf{P}(\breve{W} \neq X) &= q \text{ if } Y = \epsilon,\ Z = \epsilon \\ \mathbf{P}(\breve{W} \neq X) &= r \text{ if } Y \neq \epsilon,\ Z = \epsilon \\ \mathbf{P}(\breve{W} \neq X) &= s \text{ if } Y \neq \epsilon,\ Z \neq \epsilon \\ \mathbf{P}(\breve{W} \neq X) &= t \text{ if } Y = \epsilon,\ Z \neq \epsilon. \end{aligned}$$

Then,

$$\begin{aligned}I(X,Y;\breve{W}|Z) =& H(\breve{W}|Z) - H(\breve{W}|X,Y,Z)\\ =&\mathbf{P}(Z=\epsilon)h(\frac{1}{2})\\ &+\mathbf{P}(Z\neq\epsilon)h((1-a)s+at)\\ &-\mathbf{P}(Y=\epsilon,Z=\epsilon)h(q)\\ &-\mathbf{P}(Y\neq\epsilon,Z=\epsilon)h(r)\\ &-\mathbf{P}(Z\neq\epsilon)(1-a)h(s)\\ &-\mathbf{P}(Z\neq\epsilon)ah(t),\end{aligned}$$

where we defined $a \triangleq \mathbf{P}(Y=\epsilon|Z\neq\epsilon)$. Note that the terms that depend on s and t can be written as

$$\mathbf{P}(Z\neq\epsilon)\Big(h((1-a)s+at)-(1-a)h(s)-ah(t)\Big). \tag{12}$$

By Jensen's inequality, since $h(\cdot)$ is a concave function, the above expression is non-negative. It is also smaller than one because $h(\cdot)\leq 1$. Define a new auxiliary random variable $\tilde{W}$ such that

$$\begin{aligned}\mathbf{P}(\tilde{W}\neq X) =& q \text{ if } Y=\epsilon,\ Z=\epsilon\\ \mathbf{P}(\tilde{W}\neq X) =& r \text{ if } Y\neq\epsilon,\ Z=\epsilon\\ \mathbf{P}(\tilde{W}\neq X) =& \tilde{s} \text{ if } Y\neq\epsilon,\ Z\neq\epsilon\\ \mathbf{P}(\tilde{W}\neq X) =& 0 \text{ if } Y=\epsilon,\ Z\neq\epsilon,\end{aligned}$$

for some $\tilde{s}$ to be defined. The terms in $I(X,Y;\tilde{W}|Z)$ that depend on $\tilde{s}$ can be written as

$$\mathbf{P}(Z\neq\epsilon)\Big(h\big((1-a)\tilde{s}\big)-(1-a)h(\tilde{s})\Big). \tag{13}$$

As $\tilde{s}$ varies in $[0,1]$, (13) can take any value in $[0,h(1-a)]$. Hence, one can always find a value of $\tilde{s}$ for which (13) is smaller than (12), and hence

$$I(X,Y;\tilde{W}|Z)\leq I(X,Y;\breve{W}|Z).$$

In addition, define the new reconstruction functions

$$\tilde{X}_1=\begin{cases}\tilde{W} & \text{if } Y=\epsilon\\ Y & \text{if } Y\neq\epsilon\end{cases}$$

and

$$\tilde{X}_2=\begin{cases}\tilde{W} & \text{if } Z=\epsilon\\ Z & \text{if } Z\neq\epsilon.\end{cases}$$

One can verify that $\tilde{X}_i=\hat{X}_i$, $i=1,2$. Because of this and because we use only $(\tilde{W},Y)$ to compute $\tilde{X}_1$, we conclude that $\tilde{W}\in\mathcal{A}(D_1,D_2)$. Hence,

$$\begin{aligned}r_A(D_1,D_2)\leq& I(X,Y;\tilde{W}|Z)\\ \leq& I(X,Y;\breve{W}|Z)\\ =& r_B(D_1,D_2).\end{aligned}$$

C. *Proof of Lemma 2*

From the Markov chain $W \mathrel{-\!\!\circ\!\!-} (X,Y) \mathrel{-\!\!\circ\!\!-} Z$, we obtain

$$\begin{aligned}I(W;X,Y,Z) =& I(W;X,Y)\\ I(W;Z)+I(W;X,Y|Z) =& I(W;Y)+I(W;X|Y).\end{aligned}$$

Hence, to prove Lemma 2, it suffices to show that when $X \mathrel{-\!\!\circ\!\!-} Y \mathrel{-\!\!\circ\!\!-} Z$ is a Markov chain, then

$$I(W;X,Y|Z)-I(W;X|Y)\geq 0.$$

Indeed, we have

$$\begin{aligned}&I(W;X,Y|Z)-I(W;X|Y)\\ &=I(W,Z;X,Y)-I(Z;X,Y)-I(W;X|Y)\\ &=I(W,Z;Y)+I(W,Z;X|Y)\\ &\quad -I(Z;X,Y)-I(W;X|Y)\\ &=I(W,Z;Y)+I(Z;X|W,Y)-I(Z;X,Y)\\ &=I(Z;Y)+I(W;Y|Z)\\ &\quad +I(Z;X|W,Y)-I(Z;X,Y)\\ &=I(W;Y|Z)+I(Z;X|W,Y)\\ &\geq 0,\end{aligned}$$

where the last simplification follows because $X \mathrel{-\!\!\circ\!\!-} Y \mathrel{-\!\!\circ\!\!-} Z$ implies that $I(Z;Y)=I(Z;X,Y)$.