Vector Gaussian Rate-Distortion with Variable Side Information

Sinem Unal and Aaron B. Wagner

Abstract

We consider rate-distortion with two decoders, each with distinct side information. This problem is well understood when the side information at the decoders satisfies a certain degradedness condition. We consider cases in which this degradedness condition is violated but the source and the side information consist of jointly Gaussian vectors. We provide a hierarchy of four lower bounds on the optimal rate. These bounds are then used to determine the optimal rate for several classes of instances.

I. INTRODUCTION

We consider a rate-distortion problem with multiple decoders, each with potentially different side information, as shown in Fig. 1. This problem, which is sometimes called the Heegard-Berger problem, is one of the most basic network information theory problems that is not well understood, and it can arise in a couple of ways. First, the side information may represent reconstructions of the source due to prior transmissions. If the decoders have received different transmissions in the past, either due to channel loss or because they are not always listening to the transmitter, then their side information will be different. Second, it can be viewed as an instance of a “compound Wyner-Ziv” problem, in which there is in reality only one decoder, but the encoder is uncertain about its side information. The side information associated with different decoders in the problem could then represent the transmitter’s view of the set of possible side information configurations at the decoder. The transmitter then seeks to construct a message that would work reasonably well for any of these possible side information configurations.

The problem is well understood when the problem is degraded, i.e., the side information at one of the decoders is stochastically degraded with respect to the other’s [1]. Recently, the following class of instances, schemes for which are called index coding, has received particular attention [2], [3]. The source at each time step is a vector of independent and identically distributed (i.i.d.) bits, each side information variable is a subset of these bits, and the goal of each decoder is to losslessly reproduce a subset of the bits that are not contained in its side information. Treating the source as an i.i.d. vector of uniform bits is appropriate if the source is first compressed by an optimal rate-distortion encoder. Thus index coding implicitly assumes a separation-based architecture in which lossy compression is performed first and then the broadcasting with side information is performed at the bit level. Ideally, one would consider both types of coding jointly. In previous work [4], we studied the index coding problem using tools from network information theory, in contrast to most past work on index coding which used techniques from network coding and graph theory. One of the advantages of using network information-theoretic tools, which was not pursued in the previous work, is that it allows one to consider the problems of lossy compression and coding for side information together, by allowing for a richer class of source models and distortion constraints. Our goal in this paper is to study systems that involve both lossy compression and coding for side information.

We shall focus on the case in which the source and the side information at the decoders are all jointly Gaussian vectors. This class of instances is important in applications, since vector Gaussian sources are natural stepping stones on the path from discrete memoryless sources to more sophisticated models of multimedia. The vector Gaussian setup can also be motivated theoretically since, like index coding, it is one of the simplest classes of instances that are not degraded in general. We shall focus on the case of two decoders; unlike index coding, for vector Gaussian problems even the two-decoder case is nontrivial.

We provide a hierarchy of four lower bounds on the optimal rate. For three separate special cases, we show that at least one of the lower bounds matches the best-known achievable rate [1], [5], thereby determining the optimal rate. The four lower bounds are all obtained using variations on the following argument. Since the rate-distortion function is known when the side information is degraded [1], a natural approach to proving lower bounds is to enhance the side information of one encoder or the other in order to make the problem degraded. The optimal rate for the newly-obtained instance is thus known and provides a lower bound on the optimal rate for the original instance. This idea can be applied several ways, leading to lower bounds of varying strength and usability. The weakest of these bounds is quite weak but also quite simple. The strongest, on the other hand, is quite strong but also difficult to apply. The intermediate bounds attempt to provide the best attributes of both.

We consider three different distortion constraints, all phrased as constraints on the error covariance matrices, averaged over the block, at the two decoders. The first stipulates an upper bound on the mean square error of the reproduction of each component of the source; this can be viewed as constraints on the diagonal elements of the time-average error covariance
matrix. The second requires that the average error covariance matrix itself must be dominated, in a positive definite sense, by a given scaled identity matrix. In the final case, we require the trace of the average error covariance matrix to be upper bounded by a constant. For each of the three distortion measures, we solve a class of instances using the lower bounds developed in the paper. The necessary achievability arguments are standard, although our analysis does provide insight into how the auxiliary random variables therein should be chosen. Specifically, we show how to divide the signal space into “regions,” in which the side information at one decoder is “stronger” than that of the other. We then show that it is optimal for certain auxiliary random variables to live in certain of these regions.

The balance of the paper is organized as follows. The next two sections provide the problem formulation and the four lower bounds, respectively. Section IV contains the statements of our optimality results for all three cases described above. The achievability analysis for these problems is presented in Section V. Section VI shows how the lower bounds can be used to prove the converse half of the optimality results. Section VII contains a brief epilogue describing a conjectured difference among the lower bounds.

II. Problem Definition

Let X, Y_1, Y_2 be correlated vector Gaussian sources of size k \times 1, k_1 \times 1 and k_2 \times 1 respectively where X is the source to be compressed at the encoder and Y_1 and Y_2 comprise the side information at Decoder 1 and Decoder 2, respectively. We assume that the conditional covariance matrix of X given Y_i, K_{X|Y_i}, i \in \{1, 2\} is invertible. Both Decoder 1 and 2 wish to reconstruct X subject to given distortion constraints. The objective is to characterize the rate distortion function for this setting. The following definitions are used to formulate the problem.

Definition 1. \( \Gamma_i, i \in \{1, 2\} \) is defined as a mapping from the set of all \( k \times k \) positive semi-definite (PSD) matrices to the set of \( k_0 \times k_0 \) PSD matrices such that
1) \( \Gamma_i(\cdot) \) is linear,
2) \( A \preceq B \) implies that \( \Gamma_i(A) \preceq \Gamma_i(B) \).

Definition 2. An \((n, M, D_1, D_2)\) code where \( D_1 \) and \( D_2 \) are positive definite matrices, is composed of
- an encoding function
  \[ f : \mathbb{R}^{kn} \to \{1, \ldots, M\} \]
- and decoding functions
  \[ g_1 : \{1, \ldots, M\} \times \mathbb{R}^{k_1n} \to \mathbb{R}^{kn} \]
  \[ g_2 : \{1, \ldots, M\} \times \mathbb{R}^{k_2n} \to \mathbb{R}^{kn} \]
satisfying the distortion constraints
  \[ E \left[ \frac{1}{n} \sum_{k=1}^{n} \Gamma_i \left( (X_k - \hat{X}_{ik})(X_k - \hat{X}_{ik})^T \right) \right] \preceq D_i, \ i \in \{1, 2\} \]
where \( \hat{X}_{1n} = g_1(f(X^n), Y_{1n}^n) \), and \( \hat{X}_{2n} = g_2(f(X^n), Y_{2n}^n) \). We call \( n \) the block length and \( M \) the message size of the code.

Definition 3. A rate \( R \) is \((D_1, D_2)\)-achievable if for every \( \epsilon > 0 \), there exists an \((n, M, D_1 + \epsilon I, D_2 + \epsilon I)\) code such that \( n^{-1} \log M \leq R + \epsilon \).

1We use bold letters to denote vectors.
2Unless otherwise is stated, we assume that all Gaussian random variables are zero mean.
3\( A \preceq B \) means that \( B - A \) is a positive semidefinite matrix.
Definition 4. The rate-distortion function is defined as

\[ R(D) = \inf \{ R : R \text{ is } D\text{-achievable} \}, \]

where \( D = (D_1, D_2) \).

We shall prove our lower bounds for arbitrary distortion measures \( \Gamma \) satisfying the requirements of Definition 4. We conclude this section by introducing the following notations used in rest of the paper.

Notation 1. Let \( X \) be a \( k \times 1 \) vector where \( k = l_1 + l_2 \). Then \( (X)_{l_1} \) denotes the \( l_1 \times 1 \) vector consisting of the first \( l_1 \times 1 \) components of \( X \) and \( [X]_{l_2} \) denotes the remaining part of \( X \).

Notation 2. Let \( E \) be a \( p \times p \) matrix. Then \((E)_{i,j}\) denotes the element of \( E \) which is in the \( i^{th} \) row and \( j^{th} \) column of \( E \).

Notation 3. Let \( E \) and \( F \) be \( p \times p \) and \( r \times r \) matrices where \( p \geq l_1 \) and \( r \geq l_2 \). Then \((E)_{l_1} \) denotes the upper-left \( l_1 \times l_1 \) submatrix of \( E \) and \([F]_{l_2} \) denotes the lower-right \( l_2 \times l_2 \) submatrix of \( F \).

Notation 4. Let \( E \) and \( F \) be \( p \times p \) and \( r \times r \) matrices where \( p \geq l_1 \) and \( r \geq l_2 \). Then \((E)_{\text{diag}} \) denotes the \( p \times p \) diagonal matrix whose diagonal elements are the same as that of \( E \). Also, \((E)_{l_1,\text{diag}} \) denotes the \( l_1 \times l_1 \) diagonal matrix whose diagonal elements are the same as that of upper-left \( l_1 \times l_1 \) submatrix of \( E \) and \([F]_{l_2,\text{diag}} \) denotes the \( l_2 \times l_2 \) diagonal matrix whose diagonal elements are the same as that of lower-right \( l_2 \times l_2 \) submatrix of \( F \).

Notation 5. Let \( E \) and \( F \) be \( p \times p \) diagonal matrices. Then \( \min\{E, F\} \) denotes the \( p \times p \) diagonal matrix whose each diagonal entry is the minimum of corresponding diagonal entries of \( E \) and \( F \).

Notation 6. Let \( (X, Y, Z) \) be a random vector. Then \( X \perp Y | Z \) denotes that \( X \) and \( Y \) are independent given \( Z \), \( X \leftrightarrow Y \leftrightarrow Z \) denotes that \( X, Y \) and \( Z \) forms a Markov chain, and \( K_X \) denotes the covariance matrix of \( X \).

III. LOWER BOUNDS

We turn to lower bounds on the optimal rate. We shall provide four such bounds. In order of strongest (largest) to weakest (smallest), these are

1) The Minimax bound (MLB);
2) The Maximin bound (MLB);
3) The Enhanced-Enhancement bound (Enhanced-ELB or E\textsuperscript{2}LB);
4) The Enhancement bound (ELB).

Although the Maximin bound, the Enhanced-Enhancement bound, and the Enhancement bound are never larger than the Minimax bound, they are useful in that they are simpler to work with in some respects. We begin with the simplest, and weakest, of the bounds. This bound is folklore, and it turns out to be quite weak indeed.

A. Enhancement Lower Bound

If the side information at the decoders is degraded, meaning that we can find a joint distribution of \( (X, Y_1, Y_2) \) such that

\[ X \leftrightarrow Y_{\sigma(1)} \leftrightarrow Y_{\sigma(2)} \tag{1} \]

for some permutation \( \sigma(\cdot) \), then the rate distortion function is known \cite{1, 5}. Hence a natural way to obtain a lower bound to \( R(D) \) is to create degraded problems by providing extra side information to one decoder or the other. We call this lower bound enhancement lower bound, abbreviated as ELB, due to its similarity to the converse results for broadcast channels \cite{6}. Proposition 1 states this lower bound.

Proposition 1. The rate distortion function \( R(D) \) is lower bounded by

\[ R_{\text{ELB}}(D) = \max \{ \sup_{S_G} \inf_{\bar{C}_{1}(D)} R_{l_01} , \sup_{S_G} \inf_{\bar{C}_{2}(D)} R_{l_02} \}, \tag{2} \]

where

\[ R_{l_01} = I(\bar{X}; W, U | Y_1) + I(\bar{X}; V | W, U, Y), \tag{3} \]

\[ R_{l_02} = I(\bar{X}; W, V | Y_2) + I(\bar{X}; W, U, V, Y), \tag{4} \]

\( S_G = \{ Y \text{ jointly Gaussian with } (X, Y_1, Y_2) | X \leftrightarrow Y \leftrightarrow (Y_1, Y_2) \} \), and

\( \bar{C}_{1}(D) \) is the set of \( (W, U, V) \) such that

\( (W, U, V) \leftrightarrow X \leftrightarrow (Y, Y_1, Y_2) \)

\( (K_{X|W,U,Y_1} \leq D_1, \Gamma_2 (K_{X|W,U,V,Y}) \leq D_2) \)

\( \bar{C}_{2}(D) \) is the set of \( (W, U, V) \) such that

\( (W, U, V) \leftrightarrow X \leftrightarrow (Y, Y_1, Y_2) \)

\( (K_{X|W,U,Y} \leq D_1, \Gamma_2 (K_{X|W,U,V,Y}) \leq D_2) \).
The ELB is quite weak. Consider, for example, what is arguably the simplest nontrivial instance of the problem: the source $X$ is bivariate, $K_X$, $K_{X|Y_1}$, and $K_{X|Y_2}$ are all diagonal, and the reconstructions at decoders are subject to component-wise MSE distortion constraints. This is essentially the parallel scalar Gaussian version of the problem. If the overall problem is degraded then the ELB is of course tight. But if one of the two components is degraded in one direction and the other component is degraded in the other, then Watanabe [7] has shown that the ELB is not tight, at least for the natural choice of $Y$ that has

$$K_{X|Y} = \min(K_{X|Y_1}, K_{X|Y_2}).$$

Comparing the ELB against the achievable bound in Theorem 5 to follow, one sees several potential sources of looseness. We shall see that the culprit is that the distortion constraints in this way allows less informative ($W$, $U$, $V$) to be feasible, because one can make use of the enhanced side information $Y$ for estimation purposes. We shall make this intuition precise by showing that the Maximin and Enhanced-Enhancement lower bound, which differ from the ELB only in the distortion constraints, are tight for this problem. For reasons of expeditiousness, we shall state and prove the Minimax lower bound first, and then weaken it to obtain the Maximin and Enhanced-Enhancement lower bound.

**B. Minimax Lower Bound**

Theorem 1. The rate distortion function, $R(D)$, is lower bounded by

$$R_{\text{MLB}}(D) = \sup_{S} \inf_{C_l(D)} \max \{R_{l01}, R_{l02}\}$$

(5)

where $R_{l01}$ and $R_{l02}$ are as in (3) and (4), and

$$S = \{Y|X \leftrightarrow Y \leftrightarrow (Y_1, Y_2)\}$$

$$C_l(D): \text{the set of } (W, U, V) \text{ such that}$$

$$(W, U, V) \leftrightarrow X \leftrightarrow (Y_1, Y_2, Y)$$

$$\Gamma_1 \left( K_{X|W,U,Y_1} \right) \leq D_1$$

$$\Gamma_2 \left( K_{X|W,U,Y_2} \right) \leq D_2$$

here. Weakening the constraints in this way allows less informative ($W$, $U$, $V$) to be feasible, because one can make use of the enhanced side information $Y$ for estimation purposes. We shall make this intuition precise by showing that the Maximin and Enhanced-Enhancement lower bound, which differ from the ELB only in the distortion constraints, are tight for this problem. For reasons of expeditiousness, we shall state and prove the Minimax lower bound first, and then weaken it to obtain the Maximin and Enhanced-Enhancement lower bound.

**Proof of Theorem 1:** By definition, for any $D$–achievable rate, $R$, and for all $\epsilon > 0$, we can find a $(n, 2^{n(R+\epsilon)}, D+\epsilon(I, I))$ code. Let $\epsilon > 0$ be given and $J$ denote the output of the encoder. Also let $Y$ be an auxiliary source in $S$. Then, we can write

$$n(R + \epsilon) \geq H(J)$$

$$\geq I(X^n, Y^n_1, Y^n_2; J)$$

$$\overset{a}{=} I(Y^n_1, J) + I(Y^n_2; J|Y^n_1) + I(X^n; J|Y^n_1, Y^n)$$

$$\overset{b}{=} I(Y^n_1, J|Y^n_1) + I(X^n; J|Y^n_1, Y^n)$$

$$\overset{b}{=} \sum_{i=1}^{n} \left[ I(Y_i; J, Y_{1i}|Y_{1i}) + I(X_i; J, Y_{1i}, Y_{1i}|Y_{1i}, Y_i) \right]$$

(6)

where $Y_{1i}$ denotes all $Y^n_1$ except $Y_{1i}$ and $a$ is due to the chain rule, and $b$ is due to the chain rule and that conditioning reduces entropy. Then if we apply the chain rule to the last term above, the right hand side of (6) equals

$$\sum_{i=1}^{n} \left[ I(Y_i; J, Y_{1i}|Y_{1i}) + I(X_i; J, Y_{1i}, Y_{1i}|Y_{1i}, Y_i) \right]$$

(7)

$$= \sum_{i=1}^{n} \left[ I(X_i; Y_i; J, Y_{1i}|Y_{1i}) + I(X_i; Y_{1i}; J, Y_{1i}, Y_{1i}, Y_i) \right]$$

$$\geq \sum_{i=1}^{n} \left[ I(X_i; Y_{1i}|Y_{1i}) + I(X_i; Y_{1i}; J, Y_{1i}, Y_{1i}, Y_i) \right].$$

(8)
Also, since \( X \leftrightarrow Y \leftrightarrow (Y_1, Y_2) \) the right hand side of (8) is equal to
\[
\sum_{i=1}^{n} \left[ I(X_i; J, Y_{1i}|Y_i) + I(X_i; Y_{2i}|Y_i) \right]
\geq \sum_{i=1}^{n} \left[ I(X_i; J, Y_{1i}|Y_i) + I(X_i; Y_{2i}|Y_i) \right]
= \sum_{i=1}^{n} \left[ I(X_i; W'_i, U'_i|Y_i) + I(X_i; V'_i|W'_i, U'_i, Y_i) \right]
\]
where \( W'_i = J, U'_i = Y_{1i} \) and \( V'_i = Y_{2i} \). Note that \( (W'_i, U'_i, V'_i) \leftrightarrow X_i \leftrightarrow (Y_{1i}, Y_{2i}, Y_i) \) for all \( i \in [n] \). Let \( T \) be a random variable uniformly distributed on \([n]\) and independent of the source, side information and all \((W'_i, U'_i, V'_i), i \in [n] \). Then we can write the right hand side of (9) as
\[
\sum_{i=1}^{n} \left[ I(X_i; W'_i, U'_i|Y_i, T = i) + I(X_i; V'_i|W'_i, U'_i, Y_i, T = i) \right]
= n \left[ I(X; W', U', T|Y_1) + I(X; V', T|W', U', T, Y) \right]
= nR_{lo2}, \text{by denoting } (W', T), (U', T), (V', T) \text{ as } W, U, V \text{ respectively. (10)}
\]
If we swap the role of \( Y_1 \) and \( Y_2 \) and apply the same procedure above, we can get
\[
R + \epsilon \geq I(X; W, V|Y_2) + I(X; U|W, V, Y)
= R_{lo2}.
\]
Note that since \((W'_i, U'_i, V'_i) \leftrightarrow X_i \leftrightarrow (Y_{1i}, Y_{2i}, Y_i) \) for all \( i \in [n] \), we have \((W, U, V) \leftrightarrow X \leftrightarrow (Y_1, Y_2, Y) \). Moreover since \((W'_i, U'_i, Y_1) = (J, Y_1^j) \) and \((W'_i, V'_i, Y_2) = (J, Y_2^j) \), given \((W'_i, U'_i, Y_1) \) Decoder 1 can reconstruct the source, \( X_i \), subject to its distortion constraint. Similarly, Decoder 2 can reconstruct the source, \( X_i \), given \((W'_i, V'_i, Y_2) \). Hence, \((W, U, V) \in C_l(D + \epsilon(I, I)) \) and we have
\[
R(D) \geq \inf_{C_l(D + \epsilon(I, I))} \max \{R_{lo1}, R_{lo2}\} - \epsilon.
\]
Let \( R_{lo1}'(D + \epsilon(I, I), Y) \) denote the right hand side of (12). Note that (12) holds for any \( Y \in S \), where \( S \) as in Theorem 1. Hence we can write
\[
R(D) \geq \sup_S R_{lo1}'(D + \epsilon(I, I), Y) - \epsilon.
\]
Note that from Lemma 9 in Appendix C, \( R_{lo1}'(D, Y) \) is convex in \( D \). Since \( 0 < D_i, i \in \{1, 2\} \) we can find \( \delta(D_1, D_2) > 0 \) such that \( 0 < D_i - \delta(D_1, D_2)I \) for \( i \in \{1, 2\} \). Hence \( R_{lo1}'(D + \epsilon(I, I), Y) \) is also convex in \( \gamma \), where \( \gamma \geq -\delta(D_1, D_2) \). Note that \( \sup_S R_{lo1}'(D + \epsilon(I, I), Y) \) is also convex since supremum of convex functions is convex. Then, we can conclude that \( \sup_S R_{lo1}'(D + \epsilon(I, I), Y) \) is continuous at \( \epsilon = 0 \) since a convex function on an open set is continuous. Lastly, since \( \epsilon \) was arbitrary, letting \( \epsilon \rightarrow 0 \) gives the result.

It is worth noting how one can prove a bound similar to mLB for non-Gaussian sources and general additive distortion constraints. Although the mLB is quite powerful, it can be difficult to apply. In particular, it is not clear that it is sufficient to consider \((W, U, V) \) that are jointly Gaussian with \((X, Y_1, Y_2) \). Similarly, when considering the analogous form of this bound for discrete memoryless sources, it is not clear how to obtain cardinality bounds on the auxiliary random variables \((W, U, V) \). As such, it is not clear how to compute this bound in general. We shall therefore consider a slightly weakened form of the bound that is easier to apply. It turns out that simply swapping the min and the max in the objective and adding that \( Y \) is jointly Gaussian with \((X, Y_1, Y_2) \) to \( S \) yields a bound that is significantly more tractable.

### C. Maximin Lower Bound

The next proposition gives the Maximin lower bound, abbreviated as mLB.

**Proposition 2.** The rate distortion function, \( R(D) \), is lower bounded by
\[
R_{\text{MLB}}(D) = \max \{ \sup_{S} \inf_{C_{1}(D)} R_{lo1}, \sup_{S} \inf_{C_{2}(D)} R_{lo2} \},
\]
(14)
where $R_{101}$ and $R_{102}$ are as in (3) and (4) respectively, $S_G$ as in Proposition 7 and

\[ C_{11}(D) : \text{the set of } (W, U, V) \text{ such that } \]
\[ (W, U, V) \leftrightarrow X \leftrightarrow (Y_1, Y_2, Y) \]
\[ \Gamma_1(K_{X|W,U,Y_1}) \leq D_1, \quad \Gamma_2(K_{X|W,U,Y_1}^{-1} - K)^{-1} \leq D_2 \]

\[ C_{12}(D) : \text{the set of } (W, U, V) \text{ such that } \]
\[ (W, U, V) \leftrightarrow X \leftrightarrow (Y_1, Y_2, Y) \]
\[ \Gamma_1(K_{X|W,U,Y_1}) \leq D_1, \quad \Gamma_2(K_{X|W,U,Y_1}^{-1} - K)^{-1} \leq D_2. \]

\[ R_{EELB}(D) = \max \{ \sup_{S_G} \inf_{C_{11}(D)} R_{101}, \sup_{S_G} \inf_{C_{12}(D)} R_{102} \}, \] \hspace{1cm} (15)

where $R_{101}$ and $R_{102}$ are as in (3) and (4) respectively, $S_G$ as in Proposition 7 and

\[ \tilde{K} = K_{X|Y}^{-1} - K_{X|Y_1}^{-1}, \quad \tilde{K} = K_{X|Y}^{-1} - K_{X|Y_1}. \]

Proposition 3. The rate distortion function, $R(D)$, is lower bounded by

\[ R_{EELB}(D) = \max \{ \sup_{S_G} \inf_{C_{11}(D)} R_{101}, \sup_{S_G} \inf_{C_{12}(D)} R_{102} \}, \] \hspace{1cm} (15)

where $R_{101}$ and $R_{102}$ are as in (3) and (4) respectively, $S_G$ as in Proposition 7 and

\[ \tilde{K} = K_{X|Y}^{-1} - K_{X|Y_1}^{-1}, \quad \tilde{K} = K_{X|Y}^{-1} - K_{X|Y_1}. \]

\[ \Gamma_1(K_{X|W,U,Y_1}) \leq D_1, \quad \Gamma_2(K_{X|W,U,Y_1}^{-1} - K)^{-1} \leq D_2. \]

Prop: Note that only difference between MLB and Enhanced-ELB is the optimization sets over which the infima are taken. Hence it is enough to show that \( C_{1i}(D) \subseteq \tilde{C}_{1i}(D) \) for \( i \in \{1, 2\} \). Let \( (W, U, V) \in C_{11}(D) \). Then \( (W, U, V) \) satisfy the Markov chain condition \( (W, U, V) \leftrightarrow X \leftrightarrow (Y_1, Y_2, Y) \) and we have \( \Gamma_1(K_{X|W,U,Y_1}) \leq D_1 \). Also, the inequalities \( K_{X|W,U,V_1,Y_2} \leq K_{X|W,U,Y_1,Y_2} \) and \( K_{X|W,U,V_1,Y_2}^{-1} \leq K_{X|W,U,V,Y_1}^{-1} - \tilde{K} \) imply, by the Gaussian “variance-drop” lemma (Lemma 7 in Appendix A), that \( \Gamma_2(K_{X|W,U,Y_1}^{-1} - K)^{-1} \leq D_2. \) Hence \( (W, U, V) \) is also in \( \tilde{C}_{11}(D) \), giving \( C_{11}(D) \subseteq \tilde{C}_{11}(D) \). We can apply similar procedure to get \( C_{12}(D) \subseteq \tilde{C}_{12}(D) \), which concludes the proof.

E. Properties of the Lower Bounds

It is evident from the proofs in this section that the four lower bounds can be ordered as follows

\[ R_{ELB}(D) \leq R_{EELB}(D) \leq R_{MLB}(D) \leq R_{MB}(D). \]

We shall show that Gaussian auxiliary random variables are optimal for MLB, Enhanced-ELB, and ELB, and that the MLB and Enhanced-ELB are in fact equal. We begin by showing that Gaussian auxiliary random variables are optimal for the ELB and Enhanced-ELB.

Lemma 1. One may add the constraint that \( (W, U, V) \) is jointly Gaussian with \( (X, Y_1, Y_2, Y) \) to the optimization problem in the ELB in (2) and the Enhanced-ELB in (15) without affecting the optimal value.
**Proof:** See Appendix B. 

**Proposition 4.** The Maximin bound and Enhanced-Enhancement bound in Proposition 2 and 3 respectively, coincide:

\[ R_{MLB}(D) = R_{E^2LB}(D). \]

**Proof:** It suffices to show that

\[ R_{MLB}(D) \leq R_{E^2LB}(D) \]

By Lemma 1 \((\mathbf{W}, \mathbf{U}, \mathbf{V}) \) in \( \tilde{C}_{11}(D) \) or \( \tilde{C}_{12}(D) \) can be restricted to vector Gaussian random variables without loss of optimality. Furthermore, any \( \mathbf{U} \in \tilde{C}_{11} \) can be lumped into \( \mathbf{W} \in \tilde{C}_{11}(D) \), i.e. \( \mathbf{U} \) is deterministic, without loss of optimality since \( \mathbf{W} \) and \( \mathbf{U} \) always appear together both in the objective and the conditions. The same argument holds when we swap the roles of \( \mathbf{U} \) and \( \mathbf{V} \) in \( \tilde{C}_{12}(D) \). Hence, with those additional conditions we can write the optimizing sets, \( \tilde{C}_{11}(D) \) and \( \tilde{C}_{12}(D) \), as

\[ \begin{align*}
\tilde{C}_{11}(D) : & \quad (\mathbf{W}, \mathbf{U}, \mathbf{V}) \leftrightarrow (\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}) \\
& \quad (\mathbf{W}, \mathbf{U}, \mathbf{V}, \mathbf{X}, \mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}) \text{ jointly Gaussian, } \mathbf{U} = \emptyset \\
& \quad \mathbf{K}_{\mathbf{X}|\mathbf{W}, \mathbf{Y}_1} \preceq \mathbf{D}_1, \mathbf{K}_{\mathbf{X}|\mathbf{W}, \mathbf{Y}_2} \preceq \mathbf{D}_2
\end{align*} \]

Then any such \((\mathbf{W}, \mathbf{U}, \mathbf{V}) \in \tilde{C}_{11}(D)\) (or \((\mathbf{W}, \mathbf{U}, \mathbf{V}) \in \tilde{C}_{12}(D)\)) is also in \( C_{11}(D) \) (or \( C_{12}(D) \)). Hence, \( R_{MLB}(D) \leq R_{E^2LB}(D) \).

It follows from the two previous results that Gaussian auxiliary random variables are optimal for the MLB. To see this, let \( R_{E^2LB}(D) \) denote the Enhanced-ELB with the auxiliary random variables constrained to be jointly Gaussian with \((\mathbf{X}, \mathbf{Y}_1, \mathbf{Y}_2)\). Define \( R_{MLB}^G(D) \) likewise. Then we have

\[ R_{MLB}^G(D) \geq R_{MLB}(D) \]

\[ \overset{a}{=} R_{E^2LB}(D) \]

\[ \overset{b}{=} R_{E^2LB}^G(D) \]

\[ \overset{c}{=} R_{MLB}(D), \]

where \( a \) follows from Proposition 4, \( b \) follows from Lemma 1, and \( c \) is straightforward to verify.

We now proceed to state our optimality results.

**IV. OPTIMALITY RESULTS**

We shall determine the optimal rate for the following choices of \( \Gamma_1, \Gamma_2, D_1, \) and \( D_2 \):

1) **Mean square error (MSE):** \( \Gamma_1 \) and \( \Gamma_2 \) are chosen as

\[ \Gamma_i(K) = (K)_{diag} \quad i \in \{1, 2\} \]

and \( D_1 \) and \( D_2 \) are diagonal matrices satisfying

\[ D_1 \preceq \mathbf{K}_{\mathbf{X}|\mathbf{Y}_1} \quad \text{and} \quad D_2 \preceq \mathbf{K}_{\mathbf{X}|\mathbf{Y}_2}. \]

2) **Error covariance matrix:** \( \Gamma_1 \) and \( \Gamma_2 \) are chosen as

\[ \Gamma_i(K) = K \quad i \in \{1, 2\} \]

and \( D_1 \) and \( D_2 \) are scaled identity matrices satisfying

\[ D_1 \preceq \mathbf{K}_{\mathbf{X}|\mathbf{Y}_1} \quad \text{and} \quad D_2 \preceq \mathbf{K}_{\mathbf{X}|\mathbf{Y}_2}. \]

Note that scaled identity matrix constraints on the error covariance matrix enable us to bound the MSE of the reconstruction vector uniformly from all directions.

3) **Trace of the error covariance matrix:** \( \Gamma_1 \) and \( \Gamma_2 \) are chosen as

\[ \Gamma_i(K) = \text{Tr}(K) \quad i \in \{1, 2\} \]

and \( D_1 \) and \( D_2 \) are scalars satisfying

\[ D_1 I \preceq \mathbf{K}_{\mathbf{X}|\mathbf{Y}_1} \quad \text{and} \quad D_2 I \preceq \mathbf{K}_{\mathbf{X}|\mathbf{Y}_2}. \]

Most of the prior work on the Heegard-Berger problem assumes some sort of degradedness structure between the source and the side information at the two decoders (e.g. [1], [7], [9]). Watanabe [7], in particular, assumes that the source and the side information all consist of two components, and the first components of all three variables are independent of the second.
components of all three variables. The two components are “mismatched degraded,” i.e., each component is individually degraded, but the two components are degraded in opposite order. Although we do not assume any degradedness structure, we shall reduce our problems to one that resembles Watanabe’s. Specifically, we shall decompose the signal space into “regions,” one of which is such that the side information at Decoder 1 is “stronger” than that of Decoder 2 and one such that the reverse is true. Many such candidate decompositions are possible; we shall use the following one.

Recall that we assume that $K_{X|Y_i}$, $i \in \{1, 2\}$ are invertible matrices. Now consider the matrix $K_{X|Y_1}^{-1} - K_{X|Y_2}^{-1}$. Since it is symmetric we can find an orthogonal matrix $Q_1$ such that $Q_1 (K_{X|Y_2}^{-1} - K_{X|Y_1}^{-1}) Q_1^T$ is diagonal. Furthermore, we can find another orthogonal matrix $Q_2$ such that $Q_2 Q_1 (K_{X|Y_2}^{-1} - K_{X|Y_1}^{-1}) Q_1^T Q_2^T$ is of the form

$$K = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

(23)

where $A \succeq 0$ is an $l_1 \times l_1$ diagonal matrix, $B < 0$ is an $l_2 \times l_2$ diagonal matrix and $l_1 + l_2 = k$.

Let $Q = Q_2 Q_1$. Note that $Q D Q^T = D$ when $D$ is a scaled identity matrix and distortion measure in (21) is invariant under $(X, \tilde{X}_i) \rightarrow (QX, Q\tilde{X}_i)$.

Note that MSE distortion measure is not invariant under $(X, \tilde{X}_i) \rightarrow (QX, Q\tilde{X}_i)$. Then for MSE and any $i$, such that it is not invariant under $(X, \tilde{X}_i) \rightarrow (QX, Q\tilde{X}_i)$, we restrict our attention to the source $X$ and side information $Y_i$ such that

$$K_{X|Y_1}^{-1} - K_{X|Y_2}^{-1} = K.$$  

(24)

Therefore, the rate-distortion problems where $QX$ is the source, $Y_1$ is side information at Decoder $i$ subject to the distortion constraints $D_i$, $i \in \{1, 2\}$ are equivalent to the problems that we defined at the beginning. For the rest of the paper, we assume that $QX$ is the source and we relabel $QX$ as $X$ for the ease of notation, $Y_1$ and $Y_2$ are side information and $D_1$ and $D_2$ distortion constraints for Decoder 1 and 2, respectively, as shown in Figure. Note that we have not entirely reduced the problem to that of Watanabe because the components of $X$ may be dependent.

From now on we use the abbreviation RDSI for the problem of finding the rate distortion function where reconstructions at decoders are subjected to error covariance distortion constraints that are scaled identity matrices as in (20) and denote the corresponding rate distortion function as $R^SC(D)$, where $D = (D_1, D_2)$. Also RDT and RDMSE denote the rate distortion problems where decoders have distortion constraints as in (22) on the error covariance matrices and (13) componentwise MSE constraints, respectively. The corresponding rate distortion functions for RDT and RDMSE are denoted by $R^{T\text{r}}(D)$ and $R^{\text{MSE}}(D)$, respectively.

**Remark 1.** Since $(K_{X|Y_2}^{-1})_{l_1} \succeq (K_{X|Y_1}^{-1})_{l_1}$, we say that $Y_2$ is “stronger” than $Y_1$ in the “region” involving the upper-left part of the inverse covariance matrices. Similarly, $Y_1$ is “stronger” than $Y_2$ in the lower-right part of the inverse covariance matrices since $(K_{X|Y_1}^{-1})_{l_2} \preceq (K_{X|Y_2}^{-1})_{l_2}$.

Now we are ready to state our optimality results.

**Theorem 2.** Let $K_{X|Y_i}$, $i \in \{1, 2\}$ be diagonal matrices. Then the rate distortion function of RDMSE, $R^{\text{MSE}}(D)$, can be written as

$$R^{\text{MSE}}(D) = \max \{R_1^{\text{MSE}}(D), R_2^{\text{MSE}}(D)\},$$

where

$$R_1^{\text{MSE}}(D) = \frac{1}{2} \log \frac{|K_{X|Y_1}|}{|D_1|_{l_1}} + \frac{1}{2} \log \frac{|D_1|_{l_1}}{|\tilde{D}_1|_{l_1}}$$

$$R_2^{\text{MSE}}(D) = \frac{1}{2} \log \frac{|K_{X|Y_2}|}{|D_2|_{l_2}} + \frac{1}{2} \log \frac{|D_2|_{l_2}}{|\tilde{D}_2|_{l_2}},$$

(25)

and

$$\tilde{D}_1 = (D_1^{-1} + K)^{-1}, \quad \tilde{D}_2 = (D_2^{-1} - K)^{-1}.$$

To prove Theorem 2 first we find an upper bound based on the achievable scheme in (5) in Section V and then we utilize the Enhanced-ELB bound in the previous section, which turns out to match the upper bound.

**Remark 2.** Theorem 2 subsumes the Gaussian version of Watanabe’s result by allowing for $X$ to have dimension exceeding two. Watanabe points out that the rate-distortion for his problem, and thus for ours, does not in general equal the sum of the individual rate-distortion functions across the components of $X$, even though they are independent, independent given either side information vector, and subject to separate distortion constraints. Thus, even in this case, it is necessary to code across the different components of $X$.

The distortion constraints in (20), (22), and (13) also imply that $K_{X|Y_1}$ and $K_{X|Y_2}$ are positive definite matrices.

Note that $D_1$ and $D_2$ are positive definite since $D_1^{-1} \succeq K_{X|Y_1}^{-1} > 0$, $D_2^{-1} \succeq K_{X|Y_2}^{-1} > 0$, and $K_{X|Y_2}^{-1} = K_{X|Y_1}^{-1} + K.$
Theorem 3. The rate-distortion function for RDSI, \( R^{sc}(D) \), can be expressed as
\[
R^{sc}(D) = \max \{ R_1^{sc}(D), R_2^{sc}(D) \},
\]
where
\[
R_1^{sc}(D) = \frac{1}{2} \log \frac{|K_{X|Y_1}|}{|D_1|_{t_1}} + \frac{1}{2} \log \frac{|(D_1)_{t_1}|}{\min\{|D_1|_{t_1}, (D_2)_{t_1}\}},
\]
\[
R_2^{sc}(D) = \frac{1}{2} \log \frac{|K_{X|Y_2}|}{|D_2|_{t_2}} + \frac{1}{2} \log \frac{|(D_2)_{t_2}|}{\min\{|D_1|_{t_2}, (D_2)_{t_2}\}},
\]
and \( \tilde{D}_1 = (D_1^{-1} + K)^{-1}, \tilde{D}_2 = (D_2^{-1} - K)^{-1} \).
For the direct part of the proof of Theorem 3 we utilize the achievable scheme in Section V. For the converse result presented in Section VI we use the Enhanced-ELB bound.

Theorem 4. The rate distortion function for RDTr, \( R^{Tr}(D) \), can be characterized as
\[
R^{Tr}(D) = \min_{C^{Tr}(D)} \max \{ R_1^{Tr}(D), R_2^{Tr}(D) \}
\]
where
\[
R_1^{Tr}(D) = \frac{1}{2} \log \frac{|K_{X|Y_1}|}{I + A(K_{X|W,Y_1})_{t_1}} + \frac{1}{2} \log \frac{1}{|(K_{X|W,Y_2})_{t_1}|},
\]
\[
R_2^{Tr}(D) = \frac{1}{2} \log \frac{|K_{X|Y_2}|}{I - B(K_{X|W,Y_2})_{t_2}} + \frac{1}{2} \log \frac{1}{|(K_{X|W,Y_2})_{t_2}|},
\]
and \( C^{Tr}(D) \) denotes
\[(W, U, V) \text{ jointly Gaussian with (X, Y_1, Y_2)} \]
\[(W, U, V) \leftrightarrow X \leftrightarrow (Y_1, Y_2) \]
\[U \perp (X)_{t_1} | (W, Y_1), V \perp [X]_{t_2} | (W, Y_2) \]
\[K_{X|W,Y_1}, K_{X|W,Y_2}, K_{X|W,U,Y_1}, K_{X|W,V,Y_2}, \text{ diagonal} \]
\[\text{Tr}(K_{X|W,Y_1}) + \text{Tr}(K_{X|W,U,Y_1}) \leq D_1 \]
\[\text{Tr}(K_{X|W,Y_2}) + \text{Tr}(K_{X|W,V,Y_2}) \leq D_2 \]

Remark 3. Let \( W \) be jointly Gaussian with \( (X, Y_1, Y_2) \) such that \( W \leftrightarrow X \leftrightarrow (Y_1, Y_2) \). Due to (24), \( K_{X|W,Y_1} \) is a diagonal matrix if and only if \( K_{X|W,Y_2} \) is diagonal.

Similar to the proof of Theorem 3 and 2 we begin with proving the direct part using the same achievable scheme for RDSI by changing the distortion measure. For the converse part; however, we utilize the MLB bound, which is stronger than the Enhanced-ELB bound in general.

V. Achievable Scheme

Heegard and Berger [1] give an achievable scheme for a more general version of our problem. For more than two decoders, the Heegard and Berger result was corrected by Timo et al. [5], but we shall only consider the two-decoder version here. Particularizing the Heegard-Berger result to our problem implies the following.

Theorem 5 (cf. [1], [5]). The rate distortion function, \( R(D) \), is upper bounded by
\[
R_{ach}(D) = \inf_{C_{ach}(D)} \max \{ I(X; W, U|Y_1) + I(X; V|W, Y_2), I(X; W, V|Y_2) + I(X; U|W, Y_1) \}
\]
where
\[
C_{ach}(D) : \text{ set of } (W, U, V) \text{ such that}
\]
\[(W, U, V) \text{ jointly Gaussian with } (X, Y_1, Y_2) \]
\[(W, U, V) \leftrightarrow X \leftrightarrow (Y_1, Y_2) \]
\[\Gamma_1 (K_{X|W,U,Y_1}) \leq D_1, \Gamma_2 (K_{X|W,V,Y_2}) \leq D_2, \]
and \( \Gamma_i \) can be equal to one of the mappings in (17), (19), and (21) and the corresponding distortion constraints are as in (18), (20), and (22) respectively.

Note that \( \tilde{D}_1 \) and \( \tilde{D}_2 \) are positive definite due to similar reasoning as in Theorem 3.
Here \( W \) can be viewed as a common message to both decoders, and \( U \) and \( V \) are private messages for Decoder 1 and 2 respectively. The encoder first creates \( W \) via vector quantization with a given Gaussian test channel and then generates \( U \) and \( V \) with respect to the source and \( W \). Then \( W \) is sent to both decoders and \( U \) and \( V \) are sent to Decoder 1 and Decoder 2, respectively. At the Decoder side, Decoder 1 decodes \( W \) and \( U \) by using its side information \( Y_1 \). Similarly, Decoder 2 decodes \( W \) and \( V \) using \( Y_2 \).

Heegard and Berger do not require \((W, U, V)\) to be jointly Gaussian with \((X, Y_1, Y_2)\), but we shall only apply Theorem 5 with \((W, U, V)\) of this form, so we have added it as a constraint in the statement of the result. Note that when \((W, U, V)\) are jointly Gaussian with \((X, Y_1, Y_2)\), we can write \( R_{ach}(D) \) as

\[
R_{ach}(D) = \inf_{C_u(D)} \max \{R_1, R_2\}
\]

where

\[
R_1 = \frac{1}{2} \log \frac{|K_{X|Y_1}|}{|K_{X|W,U,Y_1}|}, \quad R_2 = \frac{1}{2} \log \frac{|K_{X|Y_2}|}{|K_{X|W,V,Y_2}|}.
\]

To get an explicit expression for the upper bounds we need to specify the auxiliary random variables more explicitly. The next three propositions give an explicit upper bound on the \( R^{MSE}(D) \), \( R^{Sc}(D) \), and properties of \((W, U, V)\) in the optimizing set \( C_u(D) \) for trace distortion constraints.

**Proposition 5.** \( R^{MSE}(D) \) is upper bounded by

\[
R_u^{MSE}(D) = \max \{R_1^{MSE}(D), R_2^{MSE}(D)\}
\]

where \( R_1^{MSE}(D) \) and \( R_2^{MSE}(D) \) are as in (25) and (26) respectively.

**Proof:** We start the proof by showing that

\[
G = \begin{pmatrix} \hat{D}_1 & 0 \\ 0 & [D_2]_{l_2} \end{pmatrix},
\]

where \( \hat{D}_1 \) as in Theorem 2 is dominated by \( K_{X|Y_2} \). Since \( K_{X|Y_2} \) and \( G \) are diagonal matrices and \( D_2 \preceq K_{X|Y_2} \), it is enough to show that \( \hat{D}_1 \preceq (K_{X|Y_2})_{l_1} \). Note that \( \hat{D}_1 = (D_1^{-1} + K)^{-1} \preceq K_{X|Y_2} \) since \( D_1 \preceq K_{X|Y_1} \). Thus, \( (\hat{D}_1)_{l_1} \preceq (K_{X|Y_2})_{l_1} \) and \( G \preceq K_{X|Y_2} \). Then we can select \( W \) such that it is jointly Gaussian with \( X \) and \( K_{X|W,Y_2} = G \). This implies

\[
K_{X|W,Y_1} = (K_{X|W,Y_2} - K)^{-1} = (G^{-1} - K)^{-1} = \begin{pmatrix} (D_1)_{l_1} & 0 \\ 0 & [D_2]_{l_2} \end{pmatrix},
\]

where \( \hat{D}_2 \) as in Theorem 2.

Lastly, we select \( U \) and \( V \) jointly Gaussian with \( X \) and \( W \) such that

\[
K_{X|W,V,Y_2} = \begin{pmatrix} \min\{(\hat{D}_1)_{l_1}, (D_2)_{l_1}\} & 0 \\ 0 & [D_2]_{l_2} \end{pmatrix},
\]

\[
K_{X|W,U,Y_1} = \begin{pmatrix} (D_1)_{l_1} & 0 \\ 0 & \min\{(D_1)_{l_1}, [D_2]_{l_2}\} \end{pmatrix},
\]

satisfy the distortion constraints. Evaluating \( R_1 \) and \( R_2 \) for this choice of \((W, U, V)\) gives us \( R^{MSE}(D) \) and \( R^{Sc}(D) \).

From the selection of the “common” and “private” messages, we can make the following observation. The “common” message is used to hit the distortion constraint of each decoder with equality over the region in which it is “weaker.” We shall apply this strategy in all three problems, in fact. Note that each decoder may undershoot its distortion constraint over the region in which it is “stronger” depending on \( D_1, D_2 \) and \( K \). Now we provide the following proposition which gives an explicit upper bound on \( R^{Sc}(D) \).

**Proposition 6.** \( R^{Sc}(D) \) is upper bounded by

\[
R_u^{Sc}(D) = \max \{R_1^{Sc}(D), R_2^{Sc}(D)\}
\]

where \( R_1^{Sc}(D) \) and \( R_2^{Sc}(D) \) are as in (27) and (28) respectively.
Proof: We follow similar approach in the proof of Proposition 5. We take a particular feasible choice of $(W, U, V)$ in $R_{ach}(D)$ to get an explicit upper bound on the rate-distortion function, $R^{Sc}(D)$. We would like to choose $W$ jointly Gaussian with $X$ so that $K_{X|W,Y_2}$ is equal to

$$G = \begin{pmatrix} (\bar{D}_1)_{l_1} & 0 \\ 0 & [D_2]_{l_2} \end{pmatrix}.$$  

This is possible if and only if $G$ is dominated by $K_{X|Y_2}$. To see that this is the case, note that $K^{-1}_{X|Y_2} \preceq D_1^{-1}$ so we have $K^{-1}_{X|Y_2} + K \preceq D_1^{-1} + K$, where $K$ is in $\mathcal{D}$. This implies that $K^{-1}_{X|Y_2} \preceq \bar{D}_1^{-1}$ since $D_1^{-1} + K = \bar{D}_1^{-1}$.

Now since $D_1$ and $D_2$ are scaled identity matrices, we must have either $D_1 \preceq D_2$ or $D_1 \succeq D_2$. We shall show that we have $K^{-1}_{X|Y_2} \succeq G^{-1}$ in both cases.

**Case 1:** $D_1 \preceq D_2$.

Note that $(\bar{D}_1^{-1})_{l_1} \succeq (D_1^{-1})_{l_1} \succeq (D_2^{-1})_{l_1}$. Then

$$G^{-1} - D_2^{-1} = \begin{pmatrix} (\bar{D}_1^{-1})_{l_1} - (D_2^{-1})_{l_1} & 0 \\ 0 & 0 \end{pmatrix} \succeq 0.$$ 

So $G^{-1} \succeq D_2^{-1} \preceq K^{-1}_{X|Y_2}$.

**Case 2:** $D_1 \succeq D_2$.

Note that $|\bar{D}_1^{-1}|_{l_2} \succeq |D_1^{-1}|_{l_2} \succeq |D_2^{-1}|_{l_2}$. Then

$$G^{-1} - \bar{D}_1^{-1} = \begin{pmatrix} 0 & 0 \\ 0 & |D_2^{-1}|_{l_2} - |\bar{D}_1^{-1}|_{l_2} \end{pmatrix} \succeq 0.$$ 

So $G^{-1} \succeq \bar{D}_1^{-1} \preceq K^{-1}_{X|Y_2}$. This shows that $K_{X|Y_2} \succeq G$ as desired. Hence we can select $K_{X|W,Y_2} = G$.

Now for any $W$ that is jointly Gaussian with $X$ and has the specified $K_{X|W,Y_2}$, we will have

$$K_{X|W,Y_1} = (K^{-1}_{X|W,Y_2} - K)^{-1} = \left( \begin{pmatrix} (\bar{D}_1^{-1})_{l_1} & 0 \\ 0 & [D_2]_{l_2} \end{pmatrix} - \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \right)^{-1}$$

$$= \begin{pmatrix} (D_1)_{l_1} & 0 \\ 0 & [D_2]_{l_2} \end{pmatrix}.$$ 

Then select $U$ and $V$ jointly Gaussian with $X$ and $W$ so that

$$K_{X|W,V,Y_2} = \begin{pmatrix} \min\{(D_1)_{l_1}, (D_2)_{l_2}\} & 0 \\ 0 & [D_2]_{l_2} \end{pmatrix},$$

$$K_{X|W,U,Y_1} = \begin{pmatrix} (D_1)_{l_1} & 0 \\ 0 & \min\{[D_1]_{l_2}, [D_2]_{l_2}\} \end{pmatrix}.$$ 

Note that $K_{X|W,U,Y_1} \preceq D_1$ and $K_{X|W,V,Y_2} \preceq D_2$ as required. Evaluating $R_1$ and $R_2$ for this choice of $(W, U, V)$ gives us $R^{Sc}_1(D)$ and $R^{Sc}_2(D)$.

As in the achievable scheme for RDMSE in Proposition 5, each decoder hits its own distortion constraint with equality on the region where it is “weaker” while each may undershoot its distortion constraint where it is “stronger” depending on $D_1$, $D_2$ and $K$. Finally, we provide the following proposition giving additional constraints on the optimizers in the optimization set $C_u(D)$ when we have trace distortion constraints.

**Proposition 7.** $R^{Tr}(D)$ is upper bounded by

$$R^{Tr}_u(D) = \min_{C^{Tr}(D)} \max\{R^{Tr}_1(D), R^{Tr}_2(D)\}$$

where $R^{Tr}_1(D)$, $R^{Tr}_2(D)$ and $C^{Tr}(D)$ as in Theorem 4

**Proof:** Notice that we can include the conditions

$$U \perp (X)_{l_1}|(W, Y_1), \quad V \perp [X]_{l_2}|(W, Y_2) \quad (41)$$

$$K_{X|W,Y_2}, K_{X|W,Y_2}, K_{X|W,U,Y_1}, K_{X|W,V,Y_2} \text{ diagonal} \quad (42)$$

to $C_u(D)$ of $R_{ach}(D)$, which gives the result. ■
VI. CONVERSE RESULTS

A. Converse for RDMSE and RDSI

It turns out that the Enhancement-ELB is sufficient for the RDMSE and RDSI problems, so we will use that bound. We shall select $Y$ in the Enhancement-ELB with the properties stated in the following lemma.

Lemma 2. Let the joint distribution of the source and side information pairs $(X, Y_i)$, $i \in \{1, 2\}$ be given. We can find a random vector, $Y$, jointly Gaussian with $(X, Y_1, Y_2)$ such that

$$X \leftrightarrow Y \leftrightarrow (Y_1, Y_2)$$

and

$$K^{-1}_{X|Y} = K^{-1}_{X|Y_1} + \tilde{K}$$

$$= K^{-1}_{X|Y_2} + \tilde{K}$$

where $\tilde{K} = \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}$ and $\tilde{K} = \begin{pmatrix} 0 & 0 \\ 0 & -B \end{pmatrix}$.

Proof: Observe that if $(X, Y, Y_1)$ can be coupled so that $X \leftrightarrow Y \leftrightarrow Y_1$ holds for $i \in \{1, 2\}$ and $(X, Y)$ has the same distribution under both couplings then it is possible to couple all four variables such that $X \leftrightarrow Y \leftrightarrow (Y_1, Y_2)$ holds.

Next note that the matrix $K^{-1}_{X|Y} - K^{-1}_{X|Y_1} + \tilde{K} = K^{-1}_{X|Y_2} - K^{-1}_{X|Y_1} + \tilde{K}$ is positive semidefinite. Thus, we can find a matrix $M$ such that $M^T M - K^{-1}_{X|Y} - K^{-1}_{X|Y_1} + \tilde{K} = -K^{-1}_{X|Y_2} + \tilde{K}$. Then, let $N$ be a Gaussian random vector, independent of $X$, with covariance matrix $K_N = I$ and let $Y = MX + N$. Then, $K^{-1}_{X|Y} = K^{-1}_{X|Y_1} + M^T M = K^{-1}_{X|Y_1} + \tilde{K} = K^{-1}_{X|Y_2} + \tilde{K}$. Since we have $K|_X \leq K|_{X|Y_1}$, $i \in \{1, 2\}$, we can couple $(X, Y, Y_1)$ so that $X \leftrightarrow Y \leftrightarrow Y_1$.

Let $Y$ be selected as in Lemma 2. By Lemma 3 we can add the condition that $(W, U, V)$ is jointly Gaussian with the source and side information at decoders to optimization sets $C_{l1}$ and $C_{l2}$ in the Enhanced-ELB. Then we can write $R_{lo1}$ in (15) as

$$R_{lo1} = \frac{1}{2} \log \frac{|K_{X|W,Y_1}|}{|K_{X|W,Y_2}|} \frac{|K_{X|W,U,Y_1}|}{|K_{X|W,U,Y_2}|}.$$

Likewise, $R_{lo2}$ in (15) can be written as

$$R_{lo2} = \frac{1}{2} \log \frac{|K_{X|W,Y_2}|}{|K_{X|W,Y_1}|} \frac{|K_{X|W,V,Y}|}{|K_{X|W,V,Y_1}|}.$$

We can further write,

$$R_{lo1} = \frac{1}{2} \log \frac{|K_{X|Y_1}|}{|K_{X|W,U,Y_1}|} \frac{|K^{-1}_{X|W,U,Y_1}|}{|K^{-1}_{X|W,U,Y_1}|}$$

$$= \frac{1}{2} \log \frac{|K_{X|Y_1}|}{|K_{X|W,Y_1}|} \frac{1}{|K_{X|W,Y_1}|}$$

$$\geq \frac{1}{2} \log \frac{|K_{X|Y_1}|}{\prod_{i=1}^{t+1} (1 + (K)_{ii}(K|_{X,W,U,Y_1})_{ii})} \frac{1}{\prod_{i=1}^{t+1} (K|_{X,W,V,Y})_{ii}}.$$

Now we focus on RDMSE where $K_{X|Y_i}, i \in \{1, 2\}$ are diagonal matrices and $D_i, i \in \{1, 2\}$ are as in (18). Since $(W, U, V)$ is jointly Gaussian with $(X, Y_1, Y_2)$, we can write $K^{-1}_{X|W,U,Y_1} = K^{-1}_{X|W,U,V} - \tilde{K}$, where $\tilde{K}$ as in Lemma 2. Then we can write $(K_{X|W,U,Y_1})_{diag} \leq D_1$ and $((K_{X|W,U,V} - \tilde{K})^{-1})_{diag} \leq D_2$, the constraints at $C_{l1}$, as $(K_{X|W,U,Y_1})_{diag} \leq D_1$ and $(K_{X|W,U,V,Y_2})_{diag} \leq D_2$.

The following lemma will be useful for matching the distortion constraints in the achievable scheme and the Enhanced-ELB.

Lemma 3. Let $A \geq 0$ be an $m \times m$ diagonal matrix, $M > 0$ be an $m \times m$ matrix and $M_{diag}$ denote $(M)_{diag}$. Then $[(M_{diag})^{-1} + A]^{-1} \preceq [(M^{-1} + A^{-1})_{diag}]$.

Proof: See Appendix D.

From $(K_{X|W,U,Y_1})_{diag} \leq D_1$ and $(K_{X|W,U,V,Y_2})_{diag} \leq D_2$, the constraints in $C_{l1}$, and by Lemma 3 we can get

$$(K_{X|W,U,Y})_{diag} \preceq (D_1^{-1} + \tilde{K})^{-1}$$

$$(K_{X|W,U,V,Y})_{diag} \preceq (D_2^{-1} + \tilde{K})^{-1}.$$
which implies

$$(K_{X|W,u,v,y})_{diag} \leq \min((D_1^{-1} + \tilde{K})^{-1}, (D_2^{-1} + \tilde{K})^{-1}).$$

Let $\tilde{D}_1$ and $\tilde{D}_2$ be as in Theorem 2. Note that $((D_1^{-1} + \tilde{K})^{-1})_{i_1} = (\tilde{D}_1)_{i_1}$ and $[(D_2^{-1} + \tilde{K})^{-1}]_{i_2} = [\tilde{D}_2]_{i_2}$. Then the right hand side of (46) is lower bounded by

$$1/2 \log \frac{|K_{X|Y_1}|}{|I + A(D_1)_{i_1}| \min((\tilde{D}_1)_{i_1}, (D_2)_{i_1}) \min([D_1]_{i_2}, [D_2]_{i_2})}.$$

Since

$$1/2 \log \frac{|K_{X|Y_1}|}{|I + A(D_1)_{i_1}|} = 1/2 \log \frac{|K_{X|Y_1}| \cdot |(\tilde{D}_1)_{i_1}|}{|(D_1)_{i_1}|},$$

we have $R_{lo1} \geq R_{1}^{MSE}(D)$. If we follow a similar procedure for $R_{lo2}$, we obtain

$$R_{lo2} = 1/2 \log \frac{|K_{X|Y_2}|}{|I + K K_{X|W,v,y} Y_2| |K_{X|W,u,v,Y}|}$$

$$\geq 1/2 \log \frac{|K_{X|Y_2}|}{|I + B[D_2]_{i_2}| \min((\tilde{D}_1)_{i_1}, (D_2)_{i_1}) \min([D_1]_{i_2}, [D_2]_{i_2})}.$$

Since $1/2 \log \frac{|K_{X|Y_2}|}{|I - B[D_2]_{i_2}|} = 1/2 \log \frac{|K_{X|Y_2}| \cdot |[D_2]_{i_2}|}{|[D_2]_{i_2}|}$, we have $R_{lo2} \geq R_{2}^{MSE}(D)$. Hence together with Proposition, this proves Theorem 2.

Note that for RDSI we can lower bound the right hand side of (46) by

$$\frac{1}{2} \log \frac{|K_{X|Y_1}|}{|I + A(D_1)_{i_1}| \min((\tilde{D}_1)_{i_1}, (D_2)_{i_1}) \min([D_1]_{i_2}, [D_2]_{i_2})},$$

where $D_i, i \in \{1, 2\}$. Then we have $R_{lo1} \geq R_{1}^{R}(D)$. If we follow a similar procedure for $R_{lo2}$, we obtain

$$R_{lo2} = 1/2 \log \frac{|K_{X|Y_2}|}{|I + K K_{X|W,v,y} Y_2| |K_{X|W,u,v,Y}|}$$

$$\geq 1/2 \log \frac{|K_{X|Y_2}|}{|I + B[D_2]_{i_2}| \min((\tilde{D}_1)_{i_1}, (D_2)_{i_1}) \min([D_1]_{i_2}, [D_2]_{i_2})}.$$

Since $1/2 \log \frac{|K_{X|Y_2}|}{|I - B[D_2]_{i_2}|} = 1/2 \log \frac{|K_{X|Y_2}| \cdot |[D_2]_{i_2}|}{|[D_2]_{i_2}|}$, we have $R_{lo2} \geq R_{2}^{R}(D)$. Hence together with Proposition, this proves Theorem 3.

B. Converse for RDTR

For RDTR, we utilize the MLB. Similar to the converse of RDMSE and RDSI, let $Y$ in MLB be selected as in Lemma 2. Then, by Lemma 8 in Appendix A we can create a $\tilde{Y}_i, i \in \{1, 2\}$ so that $(X, Y, Y_1, \tilde{Y}_1)$ is jointly Gaussian, $\tilde{Y}_i \leftrightarrow X \leftrightarrow Y_i$ and $E[X|Y_1, \tilde{Y}_1] = E[X|Y_1, Y]$ almost surely. Since $\tilde{Y}_i \leftrightarrow X \leftrightarrow Y_i$, we can write

$$\tilde{Y}_i = A_{\tilde{Y}_i} X + N_{\tilde{Y}_i}, i \in \{1, 2\},$$

where $N_{\tilde{Y}_i}$ is independent of $X$ and $Y_i$.

Then,

$$K_{X|\tilde{Y}_i,Y_i}^{-1} = K_{X|Y_i}^{-1} + A_{\tilde{Y}_i}^T K_{N_{\tilde{Y}_i}}^{-1} A_{\tilde{Y}_i},$$

(47)
Also, since \( E[X|Y_1, \hat{Y}_1] = E[X|Y_1, Y] \) almost surely \( K_{X|Y_1}^{-1} = K_{X,Y_1,Y_1}^{-1} = K_{X|Y_1,Y_1}^{-1} \). Then, from (47), \( K_{X|Y}^{-1} - K_{X|Y_1}^{-1} = \hat{K} \) and \( K_{X|Y}^{-1} - K_{X|Y_2}^{-1} = \hat{K} \),

\[
\begin{align*}
A_{Y_1}^T K_{N_{Y_1}}^{-1} A_{Y_1} &= \hat{K}, \\
A_{Y_2}^T K_{N_{Y_2}}^{-1} A_{Y_2} &= \hat{K}.
\end{align*}
\] (48)

(49)

Now, we consider any feasible variable satisfying the constraints in the optimization of \( R_{lo}(D) \) in Theorem 1. We can rewrite \( R_{lo} \) in (3) as

\[
R_{lo} = I(X; W, U|Y_1) + I(X; V|W, U, Y)
\]

\[
= h(X|Y_1) - h(X|W, U, Y_1) + h(X|W, U, Y) - h(X|W, U, Y)
\]

\[
= h(X|Y_1) - h(X|W, U, Y_1) + h(X|W, U, Y, Y) - h(X|W, U, V, Y)
\]

\[
\geq \frac{1}{2} \log \frac{|K_{X,Y_1}|}{|K_{Y_1,X,Y_1}|} = \frac{|K_{N_{Y_1}}|}{|K_{N_{Y_1}} + A_{Y_1} K_{X,W,U,Y_1} A_{Y_1}^T|}.
\] (50)

with equality if \((W, U, V)\) is Gaussian achieving the given covariance matrices. Now, let us focus on the ratio \( \frac{|K_{Y_1,X,Y_1}|}{|K_{Y_1,W,U,Y_1}|} \).

Since \( \hat{Y}_1 \leftrightarrow X \leftrightarrow Y_1 \) we can write

\[
\frac{|K_{Y_1,X,Y_1}|}{|K_{Y_1,W,U,Y_1}|} = \frac{|K_{N_{Y_1}}|}{|K_{N_{Y_1}} + A_{Y_1} K_{X,W,U,Y_1} A_{Y_1}^T|}.
\]

Since \( K_{N_{Y_1}} \) is positive definite we can write it as \( S_{Y_1} S_{Y_1} \) where \( S_{Y_1} \) is an invertible matrix. Then we can write,

\[
\frac{|K_{Y_1,X,Y_1}|}{|K_{Y_1,W,U,Y_1}|} = \frac{1}{|I + S_{Y_1}^{-1} A_{Y_1} K_{X,W,U,Y_1} A_{Y_1}^T S_{Y_1}^{-1}|} = \frac{1}{|I + K_{X,W,U,Y_1} A_{Y_1}^T S_{Y_1}^{-1} S_{Y_1}^{-1} A_{Y_1}^T|},
\]

by Sylvester’s determinant identity

\[
= \frac{1}{|I + K_{X,W,U,Y_1} A_{Y_1}^T K_{N_{Y_1}}^{-1} A_{Y_1}|} = \frac{1}{|I + K_{X,W,U,Y_1} \left( \begin{array}{cc} A & 0 \\ 0 & 0 \end{array} \right) |}
\]

where the last equality is due to (48). Then we can write (50) as

\[
R_{lo} \geq \frac{1}{2} \log \frac{|K_{X,Y_1}|}{|I + K_{X,W,U,Y_1} \left( \begin{array}{cc} A & 0 \\ 0 & 0 \end{array} \right) |} = \frac{1}{2} \log \frac{|K_{X,Y_1}|}{|K_{X,W,U,Y}|} = \frac{1}{2} \log \frac{|K_{X,Y_1}|}{|I + (K_{X|W,U,Y_1})_t A|} = \frac{1}{2} \log \frac{|K_{X,Y_1}|}{|I + (K_{X|W,U,Y_1})_t A| K_{X,W,U,Y}|} \geq \frac{1}{2} \log \frac{|K_{X,Y_1}|}{|I + \prod_{i=1}^{l_1} (1 + (K_{X|W,U,Y_1})_{ii}(A))|} = \frac{1}{2} \log \frac{|K_{X,Y_1}|}{\prod_{i=1}^{l_1} (1 + (K_{X|W,U,Y_1})_{ii}(A))} \prod_{i=1}^{l_1} (K_{X|W,U,Y_1})_{ii},
\]

by Hadamard inequality,
Lemma 5. One can add the constraint that \( \hat{R}_{i0} \) bound to \( R(\mathbf{d}) \). There exists a feasible constraint to the optimization set, \( C \) becomes (51) and (52) does not increase \( X^U \). Let \( K_{X|W,U,Y_1} \) be jointly Gaussian with \( X, Y, Y_1, Y_2 \). Furthermore, such \( (W_G, U_G, V_G) \) do not increase \( R_{i0} \) and \( \hat{R}_{i0} \). Proposition 8. The rate distortion function of RDTR, \( R(\mathbf{d}) \), is lower bounded by

\[
\min_{C(\mathbf{d})} \max \{ R^T_1(\mathbf{d}), R^T_2(\mathbf{d}) \}
\]

where \( C(\mathbf{d}) \), \( R_1^T(\mathbf{d}) \) and \( R_2^T(\mathbf{d}) \) are as in Theorem 4.

The proof follows from the next four lemmas. At each lemma, we show that without loss of optimality we can add a constraint to the optimization set, \( C_i(\mathbf{d}) \) of Theorem 4 for the trace constraints. With those additional constraints \( C_i(\mathbf{d}) \) becomes \( C(\mathbf{d}) \) and \( \hat{R}_{i0} = R^T_i(\mathbf{d}) \) for \( i \in \{1, 2\} \).

Lemma 4. There exists a feasible \((W_G, U_G, V_G)\) for \( R_{i0}(\mathbf{d}) \) such that \((W_G, U_G, V_G)\) are jointly Gaussian with \( (X, Y, Y_1, Y_2) \). Furthermore, such \((W_G, U_G, V_G)\) do not increase \( R_{i0} \) and \( \hat{R}_{i0} \).

Proof: Let \((W_G, U_G, V_G)\) be jointly Gaussian with \((X, Y, Y_1, Y_2)\) and \((W_G, U_G, V_G) \leftrightarrow X \leftrightarrow (Y, Y_1, Y_2)\) such that

\[
K_{X|W,G \cup U_1} = K_{X|W,U,Y_1}, \\
K_{X|W,G \cup U_2} = K_{X|W,V,Y_2}.
\]

By Lemma 4 we have \( K_{X|W, U} \leq K_{X|W, U_G} \) and \( K_{X|W, V} \leq K_{X|W, V_G} \). This implies \( \min \{ (K_{X|W, U}^i), (K_{X|W, V}^i) \} \) is lower than or equal to \( \min \{ (K_{X|W, U_G}^i), (K_{X|W, V_G}^i) \} \) for all \( i \in [k] \). Hence, we can conclude that \((W_G, U_G, V_G)\) is feasible for \( R_{i0}(\mathbf{d}) \) and replacing the \((W, U, V)\) with \((W_G, U_G, V_G)\) on (51) and (52) does not increase \( \hat{R}_{i0} \) and \( \hat{R}_{i02} \).

Then by Lemma 4 we can write

\[
R_{i0}(\mathbf{d}) \geq \hat{R}_{i0}(\mathbf{d})
\]

where

\[
\hat{R}_{i0}(\mathbf{d}) = \inf_{C_i(\mathbf{d})} \max \{ \hat{R}_{i01}, \hat{R}_{i02} \}
\]

and \( \hat{C}_i(\mathbf{d}) = \{ (W, U, V) \in C_i(\mathbf{d}) \mid (W, U, V) \text{ jointly Gaussian with } (X, Y, Y_1, Y_2) \} \).

The following lemmas show that without loss of optimality we can add the conditions \( U \perp (X)_i | (W, Y_1), V \perp \mid X \} \mid (W, Y_2), \) and \( K_{X|W,Y_1}, K_{X|W,U,Y_1}, K_{X|W,V,Y_2} \) are diagonal matrices to \( \hat{C}_i(\mathbf{d}) \).

Lemma 5. One can add the constraint that \( K_{X|W, U,Y_1}, K_{X|W, V,Y_2} \) are diagonal matrices to \( \hat{C}_i(\mathbf{d}) \) without increasing the optimal value, \( \hat{R}_{i0}(\mathbf{d}) \).

Proof: Note that for each feasible \((W, U, V)\) in \( \hat{C}_i(\mathbf{d}) \), we can find a \((W', U', V')\) jointly Gaussian with \((X, Y, Y_1, Y_2)\) and \((W', U', V') \leftrightarrow X \leftrightarrow (Y_1, Y_2, Y)\) such that

\[
K_{X|W, U,Y_1} = (K_{X|W, U,Y_1})_{\text{diag}} \\
K_{X|W, V,Y_2} = (K_{X|W, V,Y_2})_{\text{diag}}
\]

since \( K_{X|W, U,Y_1} \leq D_i \leq K_{X|Y_i} \) for \( i \in \{1, 2\} \). Also notice that \((W', U', V')\) satisfies the corresponding distortion constraints. Lastly we need to check that \((K_{X|W', U', Y}^i)_{\text{diag}} \geq (K_{X|W, U,Y}^i)_{\text{diag}} \) and \((K_{X|W', V, Y}^i)_{\text{diag}} \geq (K_{X|W, V,Y}^i)_{\text{diag}} \).
Since
\[ K_{X|W',U',Y} = \left[ K_{X|W',U',Y_1}^{-1} + \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} \right]^{-1} = \left( (K_{X|W,U,Y_1})_{\text{diag}}^{-1} + \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} \right)^{-1} \]

from Lemma 3, we have \( K_{X|W',U',Y} \preceq (K_{X|W,U,Y})_{\text{diag}} \) and similarly \( K_{X|W',V',Y} \preceq (K_{X|W,V,Y})_{\text{diag}} \). Hence, without loss of optimality, we can add the condition that \( K_{X|W,U,Y_1}, K_{X|W,V,Y_2} \) are diagonal matrices to \( \tilde{C}_I(D) \).

By Lemma 5, we can write
\[ \tilde{R}_{lo}(D) = \inf_{\tilde{C}_I(D)} \max \{ \tilde{R}_{lo1}, \tilde{R}_{lo2} \} \]

(55)

where \( \tilde{C}_I(D) = \{(W, U, V) \in \tilde{C}_I(D) | K_{X|W,U,Y_1}, K_{X|W,V,Y_2} \text{ are diagonal}\} \).

**Lemma 6.** One may add the constraints

\[
U \perp (X)_{l_1}|W, Y_1, \\
V \perp [X]_{l_2}|W, Y_2, \\
K_{X|W,G,Y_1}, K_{X|W,G,Y_2} \text{ are diagonal matrices}
\]

(56) (57) (58)

to the optimization set \( \tilde{C}_I(D) \) without increasing the optimal value, \( \tilde{R}_{lo}(D) \).

**Proof:** Let \((W, U, V)\) be feasible for \( \tilde{R}_{lo}(D) \), i.e. \((W, U, V) \in \tilde{C}_I(D)\). From these, we shall construct \((\tilde{W}, \tilde{U}, \tilde{V})\) that are feasible for \( \tilde{R}_{lo}(D) \) and also satisfy the conditions in (56), (57), (58) and for which the objective is only lower.

First suppose that \( d_2 \leq d_1 \). Then note that

\[
\begin{pmatrix}
(K_{X|W,U,Y_1})_{l_1}^{-1} & 0 \\
0 & [K_{X|W,V,Y_2})_{l_2}^{-1} - B
\end{pmatrix} \succeq \begin{pmatrix}
d_1^{-1}I & 0 \\
0 & d_2^{-1}I - B
\end{pmatrix} \geq \begin{pmatrix}
d_1^{-1}I \\
0
\end{pmatrix} \geq K_{X|Y_1}^{-1}.
\]

Then we may choose \( \tilde{W} \) such that

\[
K_{X|\tilde{W},Y_1}^{-1} = \begin{pmatrix}
(K_{X|W,U,Y_1})_{l_1}^{-1} & 0 \\
0 & [K_{X|W,V,Y_2})_{l_2}^{-1} - B
\end{pmatrix} \]

(59)

in which case we have

\[
K_{X|\tilde{W},Y_2}^{-1} = K_{X|\tilde{W},Y_1}^{-1} + K = \begin{pmatrix}
(K_{X|W,U,Y_1})_{l_1}^{-1} + A & 0 \\
0 & [K_{X|W,V,Y_2})_{l_2}^{-1}
\end{pmatrix} \]

(60)

Likewise, if \( d_1 < d_2 \), we have

\[
\begin{pmatrix}
(K_{X|W,U,Y_1})_{l_1}^{-1} + A & 0 \\
0 & [K_{X|W,V,Y_2})_{l_2}^{-1}
\end{pmatrix} \succeq \begin{pmatrix}
d_1^{-1}I + A & 0 \\
0 & d_2^{-1}I
\end{pmatrix} \geq d_2^{-1}I \geq K_{X|Y_2}^{-1}.
\]

Hence we may choose \( \tilde{W} \) such that

\[
K_{X|\tilde{W},Y_2}^{-1} = \begin{pmatrix}
(K_{X|W,U,Y_1})_{l_1}^{-1} + A & 0 \\
0 & [K_{X|W,V,Y_2})_{l_2}^{-1}
\end{pmatrix} \]

in which case

\[
K_{X|\tilde{W},Y_1}^{-1} = K_{X|\tilde{W},Y_2}^{-1} - K = \begin{pmatrix}
(K_{X|W,U,Y_1})_{l_1}^{-1} & 0 \\
0 & [K_{X|W,V,Y_2})_{l_2}^{-1} - B
\end{pmatrix} \]

Thus either way, we may choose \( \tilde{W} \) such that (59) and (60) hold, and so \( K_{X|\tilde{W},Y_1} \) and \( K_{X|\tilde{W},Y_2} \) are both diagonal.
Next we choose $\tilde{U}$ and $\tilde{V}$ such that $(\tilde{W}, \tilde{U}, \tilde{V}) \leftrightarrow X \leftrightarrow (Y, Y_1, Y_2)$ and

$$K_{X|\tilde{W},\tilde{U},Y_1} = \begin{pmatrix} (K_{X|\tilde{W},Y_1})_{t_1} & 0 \\ 0 & \min[[K_{X|\tilde{W},Y_1}]_{t_2}, [K_{X|\tilde{W},U,Y_1}]_{t_2}] \end{pmatrix}$$

$$= \begin{pmatrix} (K_{X|\tilde{W},U,Y_1})_{t_1} & 0 \\ 0 & \min[[K_{X|\tilde{W},Y_1}]_{t_2}, [K_{X|\tilde{W},U,Y_1}]_{t_2}] \end{pmatrix}$$

(61)

and

$$K_{X|\tilde{W},\tilde{V},Y_2} = \begin{pmatrix} \min\{(K_{X|\tilde{W},Y_2})_{t_1}, (K_{X|\tilde{W},V,Y_2})_{t_1}\} & 0 \\ 0 & [K_{X|\tilde{W},Y_2}]_{t_2} \end{pmatrix}$$

$$= \begin{pmatrix} \min\{(K_{X|\tilde{W},V,Y_2})_{t_1}, (K_{X|\tilde{W},V,Y_2})_{t_1}\} & 0 \\ 0 & [K_{X|\tilde{W},V,Y_2}]_{t_2} \end{pmatrix}$$

(62)

Evidently we have $\tilde{U} \perp (X)_{t_1} | \tilde{W}, Y_1$ and $\tilde{V} \perp [X]_{t_2} | \tilde{W}, Y_2,$ and $(\tilde{W}, \tilde{U}, \tilde{V})$ satisfy the distortion constraints.

Finally, from (59) we have

$$K_{X|\tilde{W},Y}^{-1} = K_{X|\tilde{W},Y_1}^{-1} + \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} (K_{X|\tilde{W},U,Y_1})_{t_1} + A & 0 \\ 0 & [K_{X|\tilde{W},V,Y_2}]_{t_2} - B \end{pmatrix}$$

$$= \begin{pmatrix} (K_{X|\tilde{W},U,Y_1})_{t_1} & 0 \\ 0 & [K_{X|\tilde{W},V,Y_2}]_{t_2} \end{pmatrix}$$

(63)

Similarly,

$$K_{X|\tilde{W},\tilde{U},Y}^{-1} = K_{X|\tilde{W},\tilde{U},Y_1}^{-1} + \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}$$

Substituting (61) into this equation gives,

$$K_{X|\tilde{W},\tilde{U},Y} = \begin{pmatrix} (K_{X|\tilde{W},U,Y})_{t_1} & 0 \\ 0 & \min[[K_{X|\tilde{W},U,Y}]_{t_2}, [K_{X|\tilde{W},V,Y}]_{t_2}] \end{pmatrix}$$

(64)

Likewise,

$$K_{X|\tilde{W},\tilde{V},Y} = \begin{pmatrix} \min\{(K_{X|\tilde{W},U,Y})_{t_1}, (K_{X|\tilde{W},V,Y})_{t_1}\} & 0 \\ 0 & [K_{X|\tilde{W},V,Y}]_{t_2} \end{pmatrix}$$

(65)

From (61), (62), (64) and (65), we see that the objective for $(\tilde{W}, \tilde{U}, \tilde{V})$ is equal to the objective for $(W, U, V)$. By Lemma 6, we can conclude that $R_{\ell_0}(D)$ is equal to $R_{\ell_0}^{R_{D}}(D)$.

VI. CONCLUDING REMARKS

Recall that we used the Enhanced-Enhancement lower bound to prove the converse for the RD-MSE the RDSI problems, while for the RDT problem we used the MLB. It appears that the other lower bounds are in fact insufficient for the RDT problem.

Conjecture 1. There exists an instance of the RDT problem such that the Minimax lower bound is strictly greater than the Maximin lower bound (and hence the Enhanced-Enhancement lower bound and the ELB).

To support this conjecture, one can apply the same arguments in the proof of Proposition 8 to the each minimization in the MLB separately. This way we obtain a lower bound, which is the same as in (53) except that the minimization and maximization are swapped. Consider the case where the vectors $X, Y_1, Y_2$ are bivariate Gaussian random vectors such that

$$K_{X|Y_1} = \begin{pmatrix} \frac{4}{5} & 0 \\ 0 & \frac{4}{5} \end{pmatrix}, \quad K_{X|Y_2} = \begin{pmatrix} \frac{4}{17} & 0 \\ 0 & \frac{4}{17} \end{pmatrix}$$

and the distortion constraints are $d_1 = d_2 = 0.15$. When we use CVX, a package for solving convex programs [10], [11], and the sgp function of Octave [12] to solve for the minimum rate using Theorem 4 we get a solution of 1.7808784 while we get 1.7802127 from both solvers when we swap the min and max in (53). Thus it appears that there are instances for which the added strength provided by the MLB is necessary.
VIII. ACKNOWLEDGMENT

This work was supported by Intel, Cisco, and Verizon under the Video-Aware Wireless Networks (VAWN) program and by the National Science Foundation under grants CCF-1117128 and CCF-1218578.

APPENDIX A

The aim of this appendix is to prove the following lemma.

Lemma 7 (Gaussian Variance-Drop Lemma). Let \((W, W_G, X, \tilde{Z}, Z)\) be random vectors such that \((W_G, X, \tilde{Z}, Z)\) is jointly Gaussian, \((W, W_G) \leftrightarrow X \leftrightarrow \tilde{Z} \leftrightarrow Z\) and \(K_X|W, Z > 0\). If \(K_X|W_G, Z = K_X|W, Z\) then \(K_X|W, Z \leq K_X|W_G, Z\). Also, if \(K_X|W, \tilde{Z} = K_X|W_G, \tilde{Z}\) then \(K_X|W_G, Z \leq K_X|W, Z\).

This lemma can be interpreted as follows. We view \(X\) as an underlying source of interest and \(W, W_G, \tilde{Z}\), and \(Z\) as “noisy observations” of \(X\). All except possibly \(W\) are jointly Gaussian. If \((W, Z)\) and \((W_G, Z)\) are equally-good observations, in terms of their error covariance matrix, then \((W, \tilde{Z})\) can only be better than \((W_G, Z)\). That is, replacing \(Z\) with \(\tilde{Z}\) results in a “variance drop,” and this drop is smallest in the Gaussian case.

To prove this result we will make use of the following technical lemma.

Lemma 8. Let \((X, \tilde{Z}, Z)\) be jointly Gaussian random vectors such that \(X \leftrightarrow \tilde{Z} \leftrightarrow Z\) and \(K_X|Z > 0\). We can form a \(\tilde{Z}\) such that \((X, \tilde{Z}, Z, \tilde{Z})\) is jointly Gaussian, \(Z \leftrightarrow X \leftrightarrow \tilde{Z}\), and \(E[X|Z, \tilde{Z}] = E[X|Z, \tilde{Z}]\) almost surely.

Proof: Given such \((X, \tilde{Z}, Z)\), we can create a \(\tilde{Z}\) such that \(\tilde{Z} = Z + \tilde{N}\) where \(\tilde{N}\) is Gaussian, independent of \((X, Z)\) and \(K_X|Z, \tilde{Z} = K_X|Z, \tilde{Z}\). Since \((X, Z, \tilde{Z}, E[X|Z, \tilde{Z}])\) are jointly Gaussian, we can write

\[
\tilde{Z} = B \begin{pmatrix} X \\ Z \\ E[X|Z, \tilde{Z}] \end{pmatrix} + \tilde{N}_Z,
\]

for some matrix \(B\) where \(\tilde{N}_Z\) is independent of \((X, Z, E[X|Z, \tilde{Z}])\) and Gaussian with some covariance matrix \(K_{\tilde{N}_Z}\).

Observe that the orthogonality principle and the equation \(K_{X|Z, \tilde{Z}} = K_{X|Z, \tilde{Z}}\) together imply that

\[
K_{E[X|Z, \tilde{Z}]} = K_{E[X|Z, \tilde{Z}]}.
\]

Orthogonality also implies that \(K_X|E[X|Z, \tilde{Z}] = K_E[X|Z, \tilde{Z}]\) and \(K_X|E[X|Z, \tilde{Z}] = K_E[X|Z, \tilde{Z}]\).

Hence,

\[
K_{X|E[X|Z, \tilde{Z}]} = K_{E[X|Z, \tilde{Z}]}.
\]

Likewise, orthogonality implies that \(K_{E[X|Z, \tilde{Z}]} = K_{XZ}\) and \(K_{E[X|Z, \tilde{Z}]} = K_{XZ}\).

Thus,

\[
K_{E[X|Z, \tilde{Z}]} = K_{E[X|Z, \tilde{Z}]}.
\]

Then \(66, 67, \) and \(68\) imply that \((X, Z, E[X|Z, \tilde{Z}])\) and \((X, Z, E[X|Z, \tilde{Z}])\) are equal in distribution. Now given \((X, \tilde{Z}, Z)\), create \(\tilde{Z}\) via

\[
\hat{Z} = B \begin{pmatrix} X \\ Z \\ E[X|Z, \tilde{Z}] \end{pmatrix} + \tilde{N}_Z,
\]

where \(\tilde{N}_Z\) is Gaussian with covariance matrix \(K_{\tilde{N}_Z}\) and is independent of \((X, E[X|Z, \tilde{Z}], Z)\). Then,

\[(X, Z, \hat{Z}, E[X|Z, \tilde{Z}]) = (X, Z, \hat{Z}, E[X|Z, \tilde{Z}]),\]

in distribution and so \(\hat{Z} \leftrightarrow X \leftrightarrow Z\) and \(E[X|Z, \hat{Z}] = E[X|Z, \tilde{Z}]\) almost surely.

Proof of Lemma 8 Let \((W, W_G, X, \tilde{Z}, Z)\) be as in the statement. Then by Lemma 8 we can form a random vector \(\tilde{Z} = A_\tilde{Z}X + N_\tilde{Z}\), where \(N_\tilde{Z}\) is independent of \((X, \tilde{Z})\), such that \(\tilde{Z} \leftrightarrow X \leftrightarrow Z\), \(K_X|X, \tilde{Z} = K_X|X, \tilde{Z}\) and \(E[X|Z, \tilde{Z}] = E[X|Z, \tilde{Z}]\) almost surely. Since for any \(W\) such that \(W \leftrightarrow X \leftrightarrow (\tilde{Z}, \tilde{Z}, Z)\) we have \(K_X|W, \tilde{Z} = K_X|W, \tilde{Z}\) and \(K_X|W, E[X|Z, \tilde{Z}] = K_X|W, E[X|Z, \tilde{Z}]\), it suffices to prove the result for the special case in which \(Z = (Z, Z)\) so we shall assume that \(\tilde{Z}\) has this form. Also, let \(\bar{X} = E[X|W, \tilde{Z}]\). We will write the covariance matrix of the best linear estimate of \(X\) using \(\bar{X}\) and \(\tilde{Z}\) in terms of \(K_X|W, Z\) and \(K_X|\bar{Z}\) by applying the procedure of [13]. Then we can write

\[
K(X, \bar{X}, \tilde{Z}) = \begin{pmatrix} K_X & K_{\bar{X}} & K_X A_{\bar{X}}^T \\ K_{\bar{X}} & K_{\bar{X}} & K_{\bar{X}} A_{\bar{X}}^T \\ A_{\bar{X}} K_X & A_{\bar{X}} K_X & A_{\bar{X}} K_X A_{\bar{X}}^T + K_{N_{\bar{X}}\bar{X}} \end{pmatrix}
\]
where \( K_X = (K_X - K_X|W,z) \). Note that \( K_X \) may not be invertible, meaning that some of the elements of \( \tilde{X} \) can be determined as linear combinations of others. Thus it is enough to consider only the components of \( \tilde{X} \) or linear combinations of them, denoted by \( \tilde{X} = Q\tilde{X} \), such that the resulting covariance matrix, denoted as \( K_X^* = QK_XQ^T \), is invertible. Then we can write,

\[
K(X, \tilde{X}, \tilde{Z}) = \begin{pmatrix}
K_X & K_{\tilde{X}} Q^T & K_{\tilde{X}} A_{\tilde{Z}}^T \\
QK_{\tilde{X}} & K_{\tilde{X}} & QK_{\tilde{X}} A_{\tilde{Z}}^T \\
A_{\tilde{Z}}K_X & A_{\tilde{Z}}K_{\tilde{X}} Q^T & A_{\tilde{Z}}K_X A_{\tilde{Z}}^T + K_{N_\alpha}
\end{pmatrix}.
\]

The covariance matrix of a linear estimate of \( X \) using \( \tilde{X} \) and \( \tilde{Z} \) is

\[
K_{(X|\tilde{X},\tilde{Z})l} = K_{(X|\tilde{X},\tilde{Z})l} = K_X - (K_{\tilde{X}} Q^T K_X A_{\tilde{Z}}^T) C^{-1} \begin{pmatrix} QK_{\tilde{X}} \\ A_{\tilde{Z}}K_X \end{pmatrix}
\]

where

\[
C = \begin{pmatrix} K_{\tilde{X}} & QK_{\tilde{X}} A_{\tilde{Z}}^T \\
A_{\tilde{Z}}K_{\tilde{X}} Q^T & A_{\tilde{Z}}K_X A_{\tilde{Z}}^T + K_{N_\alpha}\end{pmatrix}.
\]

By the matrix inversion lemma we have

\[
K_{(X|\tilde{X},\tilde{Z})l}^{-1} = K_X^{-1} + K_X^{-1} (K_{\tilde{X}} Q^T K_X A_{\tilde{Z}}^T) E^{-1} \begin{pmatrix} QK_{\tilde{X}} \\ A_{\tilde{Z}}K_X \end{pmatrix} K_X^{-1}
\]

where

\[
E = C - \begin{pmatrix} QK_{\tilde{X}} \\ A_{\tilde{Z}}K_X \end{pmatrix} K_X^{-1} (K_{\tilde{X}} Q^T K_X A_{\tilde{Z}}^T)
\]

\[
= C - \begin{pmatrix} Q(I - K_X|W,z K_X^{-1}) \\ A_{\tilde{Z}} \end{pmatrix} (K_{\tilde{X}} Q^T K_X A_{\tilde{Z}}^T)
\]

\[
= C - \begin{pmatrix} Q(K_{\tilde{X}} - K_X|W,z K_X^{-1} K_X) Q^T QK_{\tilde{X}} A_{\tilde{Z}}^T \\ A_{\tilde{Z}} K_{\tilde{X}} Q^T \end{pmatrix}
\]

\[
= \begin{pmatrix} K_{\tilde{X}} & QK_{\tilde{X}} A_{\tilde{Z}}^T \\
A_{\tilde{Z}} K_{\tilde{X}} Q^T & A_{\tilde{Z}} K_X A_{\tilde{Z}}^T + K_{N_\alpha}\end{pmatrix} - \begin{pmatrix} Q(K_{\tilde{X}} - K_X|W,z K_X^{-1} K_X) Q^T QK_{\tilde{X}} A_{\tilde{Z}}^T \\ A_{\tilde{Z}} K_{\tilde{X}} Q^T \end{pmatrix}
\]

\[
= \begin{pmatrix} Q(K_X|W,z - K_X|W,z K_X^{-1} K_X|W,z) Q^T 0 \\ 0 K_{N_\alpha}\end{pmatrix}.
\]

Then

\[
K_{(X|\tilde{X},\tilde{Z})l}^{-1} = K_X^{-1} + K_X^{-1} (K_{\tilde{X}} Q^T K_X A_{\tilde{Z}}^T) \begin{pmatrix} K_{(X|\tilde{X})l}^{-1} & 0 \\ 0 K_{Z|\tilde{X}}^{-1} \end{pmatrix} \begin{pmatrix} QK_{\tilde{X}} \\ A_{\tilde{Z}}K_X \end{pmatrix} K_X^{-1}
\]

\[
= K_X^{-1} + K_X^{-1} (K_{\tilde{X}} Q^T K_{(X|\tilde{X})l}^{-1} K_X A_{\tilde{Z}}^T K_{Z|\tilde{X}}^{-1} K_X) \begin{pmatrix} QK_{\tilde{X}} \\ A_{\tilde{Z}}K_X \end{pmatrix} K_X^{-1}
\]

\[
= K_X^{-1} + K_X^{-1} (K_{\tilde{X}} Q^T K_{(X|\tilde{X})l}^{-1} QK_{\tilde{X}} K_X A_{\tilde{Z}}^T K_{Z|\tilde{X}}^{-1} A_{\tilde{Z}} K_X) K_X^{-1}
\]

\[
= K_{(X|\tilde{X})l}^{-1} + K_{X|\tilde{Z}} - K_X^{-1}, \text{ by matrix inversion lemma}
\]

\[
K_{X|\tilde{X}}^{-1} + K_{X|\tilde{Z}} - K_X^{-1}
\]

\[
K_{X|\tilde{X}}^{-1} + K_{X|\tilde{Z}} - K_X^{-1}
\]

\[
= K_{X|\tilde{X}}^{-1} + K_{X|\tilde{Z}} - K_X^{-1}.
\]

Hence we have

\[
K_{(X|\tilde{X},\tilde{Z})l}^{-1} = K_{X|W,Z}^{-1} + K_{X|\tilde{Z}}^{-1} - K_X^{-1} + K_{X|W,G,z}^{-1} - K_{X|W,G,z}^{-1}.
\]

(69)

Note that \( K_X|W,z \leq K_{(X|\tilde{X},\tilde{Z})l} \) so \( K_{X|W,z,\tilde{Z}} \geq K_{(X|\tilde{X},\tilde{Z})l}^{-1} \). Then, from (69) we have

\[
K_{X|W,z,\tilde{Z}}^{-1} \geq K_{(X|\tilde{X},\tilde{Z})l}^{-1} + K_{X|W,z} - K_{X|W,G,z}^{-1}.
\]

(70)

Thus, by (70) if \( K_{X|W,G,z} = K_{X|W,z} \) then \( K_{X|W,z,\tilde{Z}} \leq K_{X|W,G,z} \) and if \( K_{X|W,z} = K_{X|W,G,z} \) then \( K_{X|W,z,\tilde{Z}} \leq K_{X|W,z} \).
Corollary 1. Let \((W, X, Z, \tilde{Z})\) be random vectors such that \(X, Z\), and \(\tilde{Z}\) are jointly Gaussian, \(W \leftrightarrow X \leftrightarrow \tilde{Z} \leftrightarrow Z\) and \(K_{X|W,Z} > 0\). Also, let \(\tilde{D} = (D^{-1} + K_{X|Z}^{-1} - K_{X|Z}^{-1})^{-1}\). If \(K_{X|W,Z} = \tilde{D}\) then \(K_{X|W,Z} \leq \tilde{D}\).

Proof: We can find \(W_G\) jointly Gaussian with \((X, \tilde{Z}, Z)\) such that \((W, W_G) \leftrightarrow W \leftrightarrow \tilde{Z} \leftrightarrow Z\) and \(K_{X|W_G,Z} = K_{X|W,Z} = \tilde{D}\). Then \(K^{-1}_{X|W_G,Z} = K^{-1}_{X|W,G,Z} + K^{-1}_{X|Z} - K^{-1}_{X|Z} = K^{-1}_{X|W,Z} + K^{-1}_{X|W,Z} - K^{-1}_{X|Z}\). Lemma \([7]\) then implies the result.

APPENDIX B

Proof of Lemma \([7]\). We show that without loss of optimality, the auxiliary random vectors can be chosen to be jointly Gaussian with \((X, Y_1, \tilde{Y}_2, Y)\) in ELB and Enhanced-ELB. Let \(Y \in S_G\), \((W, U, V) \in \tilde{C}_1\) of \((W, U, V) \in \tilde{C}_1\) for Enhanced-ELB) be given and \(R_{t_{01}} = I(X; W, U|Y) + I(X; V|W, U, Y)\) as defined before. Note that without loss of generality we can write \(Y = A_YX + B_YY_1 + N_Y\), where \(N_Y\) is a Gaussian vector that is independent of the pair \((X, Y_1)\). Then

\[
R_{t_{01}} = h(X|Y_1) - h(X|W, U, Y_1) + h(X|W, U, Y) - h(X|W, U, V, Y)
\]

\[
= h(X|Y_1) - h(X|W, U, Y_1) + h(X|W, U, Y_1, Y) - h(X|W, U, V, Y_1, Y), \text{ since } X - Y - Y_1
\]

\[
= h(X|Y_1) - I(X; Y|W, U, Y_1) - h(X|W, U, Y, Y_1)
\]

\[
= h(X|Y_1) + h(Y|X, Y_1) - h(Y|W, U, Y_1) - h(X|W, U, V, Y_1)
\]

\[
= h(X|Y_1) + h(Y|X, Y_1) - h(A_YX + N_Y|W, U, Y_1)
\]

\[
= h(X|W, U, Y, Y_1), \text{ since } Y = A_YX + B_YY_1 + N_Y
\]

\[
\geq \frac{1}{2} \log \left[ \frac{|K_{X|Y_1}| |K_{Y|X,Y_1}|}{|K_{A_YX+N_Y|W,U,Y_1}| |K_{X|W,U,V,Y_1,Y}|} \right]
\]

where \(K_{A_YX+N_Y|W,U,Y_1} = A_YK_{X|W,U,Y_1} + K_{Y|X,Y_1} + K_{X|W,U,Y_1}\) and equality holds if \((W, U, V)\) is Gaussian. We can find \((W_G, U_G)\) that are jointly Gaussian with \((X, Y_1, Y_2, Y)\) such that \((W_G, U_G) \leftrightarrow X \leftrightarrow Y \leftrightarrow (Y_1, Y_2)\) and \(K_{X|W,G,U,G,Y} = K_{X|W,U,Y_1}\). Then by Lemma \([7]\), \(K_{X|W,G,U,G,Y} \geq K_{X|Y,Y_1,Y_2} \geq K_{X|W,U,V,Y}\). Thus we can find a \(V_G\) that is jointly Gaussian with \((W_G, U_G, X, Y_1, Y_2, Y)\) such that \((W_G, U_G, V_G) \leftrightarrow X \leftrightarrow Y \leftrightarrow (Y_1, Y_2)\) and \(K_{X|W,G,U,G,Y} = K_{X|W,U,V,Y}\), giving \((W_G, U_G, V_G) \in \tilde{C}_1\), \((W_G, U_G, V_G) \in \tilde{C}_1\) for Enhanced-ELB). Therefore, one can choose the auxiliary random vectors to be jointly Gaussian with \((X, Y_1, Y_2, Y)\) without loss of optimality in \(R_{t_{01}}\). The same argument applies to \(R_{t_{02}}\) as well.

APPENDIX C

Lemma 9. \(R'_{t_0}(D, Y)\) is a convex function with respect to \(D\).

Proof of Lemma \([7]\). To prove the lemma, we use a similar argument to \([14]\). Let \(\epsilon > 0\) be given. We can find \((\tilde{W}, \tilde{U}, \tilde{V})\) and \((\bar{W}, \bar{U}, \bar{V})\) in \(C_{\tilde{D}}(D)\) and \(C_{\bar{D}}(D)\) respectively such that

\[
\max\{I(X; \tilde{W}, \tilde{U}|Y_1) + I(X; \tilde{V}|\tilde{W}, \tilde{U}, Y) + I(X; \tilde{W}|\tilde{V}|\tilde{W}, \tilde{U}, Y)\} \leq R'_{t_0}(\tilde{D}, Y) + \epsilon
\]

\[
\max\{I(X; \bar{W}, \bar{U}|Y_1) + I(X; \bar{V}|\bar{W}, \bar{U}, Y) + I(X; \bar{W}|\bar{V}|\bar{W}, \bar{U}, Y)\} \leq R'_{t_0}(\bar{D}, Y) + \epsilon.
\]

Now we construct \((\bar{W}, U, V)\) and show that it is in \(C_{\bar{D}}(D) + (1 - \lambda)\bar{D}\). Let \(T\) be a binary random variable with \(P(T = 1) = \lambda\) and independent of \((\tilde{W}, U, V, \bar{W}, \bar{U}, \bar{V}, X, Y_1, Y_2, Y)\). Then we define

\[
W = (\tilde{W}, T) \text{ if } T = 1,
\]

\[
W = (\tilde{W}, T) \text{ if } T = 0,
\]

\[
\tilde{U} = (\tilde{U}, T) \text{ if } T = 1,
\]

\[
\tilde{U} = (\tilde{U}, T) \text{ if } T = 0,
\]

\[
V = (\tilde{V}, T) \text{ if } T = 1,
\]

\[
V = (\tilde{V}, T) \text{ if } T = 0,
\]

and

\[
g_1(W, U, Y_1) = E[X|\tilde{W}, \tilde{U}, Y_1] \text{ if } T = 1,
\]

\[
g_1(W, U, Y_1) = E[X|\tilde{W}, \tilde{U}, Y_1] \text{ if } T = 0,
\]

\[
g_2(W, V, Y_2) = E[X|\tilde{W}, \tilde{V}, Y_2] \text{ if } T = 1,
\]

\[
g_2(W, V, Y_2) = E[X|\tilde{W}, \tilde{V}, Y_2] \text{ if } T = 0.
\]

Note that \(K_{X|g_1(W, U, Y_1)} = \lambda K_{X|\tilde{W}, \tilde{U}, Y_1} + (1 - \lambda) K_{X|\tilde{W}, \tilde{U}, Y_1}\) and since \(\Gamma_1\) is a linear operator, \(\Gamma_1(K_{X|g_1(W, U, Y_1)}) \leq \lambda \tilde{D}_1 + (1 - \lambda) \tilde{D}_1\). Similarly, that \(K_{X|g_2(W, V, Y_2)} = \lambda K_{X|\tilde{W}, \tilde{V}, Y_2} + (1 - \lambda) K_{X|\tilde{W}, \tilde{V}, Y_2}\) gives \(\Gamma_2(K_{X|g_2(W, V, Y_2)}) \leq \lambda \tilde{D}_2 + (1 - \lambda) \tilde{D}_2\).
Hence, $(W, U, V) \in C_{i}(\lambda D + (1 - \lambda)\bar{D})$. We can write
\[
R'_i(\lambda D + (1 - \lambda)\bar{D}, Y)
\]
\[
\leq \max\{I(X; W, U|Y_1) + I(X; V|W, U, Y), I(X; W, V|Y_2) + I(X; U|W, Y, V, Y)\}
\]
\[
= \max\{I(X; W, U, T|Y_1) + I(X; V|W, U, T, Y), I(X; W, V, T|Y_2) + I(X; U|W, V, T, Y)\}
\]
\[
= \max\{\lambda I(X; \tilde{W}, \tilde{U}|Y_1) + (1 - \lambda)I(X; \tilde{W}, \tilde{U}|Y_1) + \lambda I(X; \tilde{V}|\tilde{W}, \tilde{U}, Y) + (1 - \lambda)I(X; \tilde{V}|\tilde{W}, \tilde{U}, Y),
\]
\[
\lambda I(X; \tilde{W}, \tilde{V}|Y_2) + (1 - \lambda)I(X; \tilde{W}, \tilde{V}|Y_2) + \lambda I(X; \tilde{U}|\tilde{W}, \tilde{V}, Y) + (1 - \lambda)I(X; \tilde{U}|\tilde{W}, \tilde{V}, Y)\}
\]
\[
\leq \lambda \max\{I(X; \tilde{W}, \tilde{U}|Y_1) + I(X; \tilde{V}|\tilde{W}, \tilde{U}, Y), I(X; \tilde{W}, \tilde{V}|Y_2) + I(X; \tilde{U}|\tilde{W}, \tilde{V}, Y)\}
\]
\[
+ (1 - \lambda) \max\{I(X; \tilde{W}, \tilde{U}|Y_1) + I(X; \tilde{V}|\tilde{W}, \tilde{U}, Y), I(X; \tilde{W}, \tilde{V}|Y_2) + I(X; \tilde{U}|\tilde{W}, \tilde{V}, Y)\}
\]
\[
\leq \lambda R_i(\lambda \bar{D}, Y) + (1 - \lambda)R'_i(D, Y) + \epsilon.
\]
By letting $\epsilon \to 0$, we conclude that $R'_i(D, Y)$ is a convex function of $D$.

**APPENDIX D**

**Proof of Lemma**

First we consider $A > 0$. Using the matrix inversion lemma, we can write
\[
([M^{-1} + A]^{-1})_{\text{diag}} = (A^{-1} - A^{-1}[M + A^{-1}]^{-1}A^{-1})_{\text{diag}}
\]
\[
= A^{-1} - A^{-1}([M + A^{-1}]_{\text{diag}}A^{-1}
\]
\[
\leq A^{-1} - A^{-1}[M_{\text{diag}} + A^{-1}]^{-1}A^{-1},
\]
since $(M_{\text{diag}})^{-1} \preceq (M^{-1})_{\text{diag}}$, [15 Theorem 7.7.8]. By the matrix inversion lemma, the right hand side of the last inequality is $([M_{\text{diag}}]^{-1} + A)^{-1}$.

Now, we consider $A \succeq 0$. Without loss of generality we can assume that all positive diagonal entries are on the upper left corner of $A$. Hence we can write
\[
A = \begin{pmatrix}
A_1 & 0 \\
0 & 0
\end{pmatrix},
\]
where $A_1 > 0$ is an $m_1 \times m_1$ matrix, where $m_1 \leq m$. Also we can represent $M$ in terms of block matrices,
\[
M = \begin{pmatrix}
M_1 & M_2 \\
M_2^T & M_3
\end{pmatrix},
\]
where $M_1 > 0$, is an $m_1 \times m_1$ matrix, and $M_3 > 0$ is an $(m - m_1) \times (m - m_1)$ matrix. Then we can write inverse of $M$ as
\[
M^{-1} = \begin{pmatrix}
\bar{M}_1 & \bar{M}_2 \\
\bar{M}_2 & \bar{M}_3
\end{pmatrix},
\]
where
\[
\bar{M}_1 = (M_1 - M_2M_3^{-1}M_2^T)^{-1}
\]
\[
\bar{M}_3 = (M_3 - M_2^T(M_3^{-1}M_2)^{-1})^{-1}
\]
\[
\bar{M}_2 = -M_1^{-1}M_2(M_1 - M_2M_3^{-1}M_2^T)^{-1}.
\]

Also,
\[
[M^{-1} + A] = \begin{pmatrix}
\bar{M}_1 + A_1 & \bar{M}_2 \\
\bar{M}_2 & \bar{M}_3
\end{pmatrix}.
\]
When we take the inverse of $[M^{-1} + A]$ we have,
\[
[M^{-1} + A]^{-1} = \begin{pmatrix}
\bar{M}_1 & \bar{M}_2 \\
\bar{M}_2 & \bar{M}_3
\end{pmatrix}.
\]
where $\bar{M}_2$ is a matrix in terms of $\bar{M}_1$, $\bar{M}_2$, $\bar{M}_3$, $A$ and
\[
\bar{M}_1 = (\bar{M}_1 + A_1 - \bar{M}_2\bar{M}_3^{-1}\bar{M}_2^T)^{-1}
\]
\[
\bar{M}_3 = (\bar{M}_3 - \bar{M}_2^T(\bar{M}_1 + A_1)^{-1}\bar{M}_2)^{-1}.
\]
Since \( M_1 = (\bar{M}_1 - \bar{M}_2 \bar{M}_3^{-1} \bar{M}_2^T)^{-1} \) and \( M_3 = (\bar{M}_3 - \bar{M}_2^T \bar{M}_3^{-1} \bar{M}_2)^{-1} \), we can write
\[
\tilde{M}_1 = [M_1^{-1} + A_1]^{-1} \\
\tilde{M}_3 \preceq M_3.
\]
Then utilizing the inequalities above we can write
\[
([M^{-1} + A]^{-1})_{\text{diag}} = \begin{pmatrix} \tilde{M}_1 & \tilde{M}_2 \\ \tilde{M}_2 & M_3 \end{pmatrix}_{\text{diag}} \preceq \begin{pmatrix} ([M_1^{-1} + A_1]^{-1})_{\text{diag}} & 0 \\ 0 & (M_3)_{\text{diag}} \end{pmatrix} \preceq ([M_{\text{diag}}]^{-1} + A)^{-1}.
\]

\[\text{References}\]

[1] C. Heegard and T. Berger, “Rate distortion when side information may be absent,” *Information Theory, IEEE Transactions on*, vol. 31, no. 6, pp. 727–734, 1985.

[2] Y. Birk and T. Kol, “Informed-source coding-on-demand (iscod) over broadcast channels,” in *INFOCOM '98. Seventeenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3, 1998, pp. 1257–1264 vol.3.

[3] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol, “Index coding with side information,” in *Foundations of Computer Science, 2006. FOCS '06. 47th Annual IEEE Symposium on*, 2006, pp. 197–206.

[4] S. Unal and A. B. Wagner, “A rate-distortion approach to index coding,” *CoRR*, vol. abs/1402.0258, 2014. [Online]. Available: http://arxiv.org/abs/1402.0258

[5] R. Timo, T. Chan, and A. Grant, “Rate distortion with side-information at many decoders,” *Information Theory, IEEE Transactions on*, vol. 57, no. 8, pp. 5240–5257, 2011.

[6] S. Vishwanath, G. Kramer, S. Shamai, S. Jafar, and A. Goldsmith, “Capacity bounds for Gaussian vector broadcast channels,” *DIMACS SERIES IN DISCRETE MATHEMATICS AND THEORETICAL COMPUTER SCIENCE*, vol. 62, pp. 107–122, 2004.

[7] S. Watanabe, “The rate-distortion function for product of two sources with side-information at decoders,” in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, July 2011, pp. 2761–2765.

[8] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Orlando, FL, USA: Academic Press, Inc., 1982.

[9] R. Timo, T. Oechtering, and M. Wigger, “Source coding problems with conditionally less noisy side information,” *Information Theory, IEEE Transactions on*, vol. 60, no. 9, pp. 5516–5532, Sept 2014.

[10] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 2.1,” [http://cvxr.com/cvx](http://cvxr.com/cvx), Mar. 2014.

[11] ———, “Graph implementations for nonsmooth convex programs,” pp. 95–110, 2008, [http://stanford.edu/~boyd/graphdcp.html](http://stanford.edu/~boyd/graphdcp.html)

[12] J. W. Eaton, D. Bateman, and S. Hauberg, *GNU Octave version 3.0.1 manual: a high-level interactive language for numerical computations*. CreateSpace Independent Publishing Platform, 2009, ISBN 1441413006. [Online]. Available: [http://www.gnu.org/software/octave/doc/interpreter](http://www.gnu.org/software/octave/doc/interpreter)

[13] J. Wang, J. Chen, and X. Wu, “On the sum rate of Gaussian multiterminal source coding: New proofs and results,” *Information Theory, IEEE Transactions on*, vol. 56, no. 8, pp. 3946–3960, Aug 2010.

[14] A. D. Wyner, “The rate-distortion function for source coding with side information at the decoder-ii. general sources,” *Information and Control*, vol. 38, no. 1, pp. 60–80, 1978. [Online]. Available: [http://dx.doi.org/10.1016/S0019-9958(78)90034-7](http://dx.doi.org/10.1016/S0019-9958(78)90034-7)

[15] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York, NY, USA: Cambridge University Press, 1986.