A Realization Approach to Lossy Network Compression of a Tuple of Correlated Multivariate Gaussian RVs†

Charalambos D. Charalambous 1,* and Jan H. van Schuppen 2,‡

1 Department of Electrical and Computer Engineering, University of Cyprus, P.O. Box 20537, CY-1678 Nicosia, Cyprus
2 Van Schuppen Control Research, Gouden Leeuw 143, 1103 KB Amsterdam, The Netherlands
* Correspondence: chadcha@ucy.ac.cy
† Preliminary results are presented in part at the 2020 IEEE International Symposium on Information Theory (ISIT), Los Angeles, CA, USA, 21–26 June 2020.

Abstract: Examined in this paper is the Gray and Wyner source coding for a simple network of correlated multivariate Gaussian random variables, \( Y_1 : \Omega \rightarrow \mathbb{R}^p_1 \) and \( Y_2 : \Omega \rightarrow \mathbb{R}^p_2 \). The network consists of an encoder that produces two private rates \( R_1 \) and \( R_2 \), and a common rate \( R_0 \), and two decoders, where decoder 1 receives rates \((R_1,R_0)\) and reproduces \( Y_1 \) by \( \hat{Y}_1 \), and decoder 2 receives rates \((R_2,R_0)\) and reproduces \( Y_2 \) by \( \hat{Y}_2 \), with mean-square error distortions \( \mathbb{E}||Y_i - \hat{Y}_i||^2_{\Omega} \leq \Delta_i \in [0,\infty], i = 1,2 \). Use is made of the weak stochastic realization and the geometric approach of such random variables to derive test channel distributions, which characterize the rates that lie on the Gray and Wyner rate region. Specific new results include: (1) A proof that, among all continuous or finite-valued random variables, \( W : \Omega \rightarrow \mathbb{R}^n \), Wyner’s common information, \( C(Y_1,Y_2) = \inf_{P_{Y_1Y_2W}W} P_{Y_1W}P_{Y_2W}:I(Y_1,Y_2;W) \), is achieved by a Gaussian random variable, \( W : \Omega \rightarrow \mathbb{R}^n \) of minimum dimension \( n \), which makes the two components of the tuple \( (Y_1,Y_2) \) conditionally independent according to the weak stochastic realization of \((Y_1,Y_2)\), and a the formula \( C(Y_1,Y_2) = \frac{1}{2} \sum_{i=1}^{n} \ln \left( \frac{1+d_i}{1-d_i} \right) \), where \( d_i \in (0,1), i = 1,\ldots, n \) are the canonical correlation coefficients of the correlated parts of \( Y_1 \) and \( Y_2 \), and a realization of \((Y_1,Y_2,W)\) which achieves this. (2) The parameterization of rates that lie on the Gray and Wyner rate region, and several of its subsets. The discussion is largely self-contained and proceeds from first principles, while connections to prior literature is discussed.

Keywords: Gray–Wyner network; Wyner’s lossy common information; weak realizations of conditional independence; canonical variable form of multivariate Gaussian random variables; multi-user communication

1. Introduction

In their seminal paper, Source Coding for a Simple Network [1], Gray and Wyner characterized the lossless rate region for a tuple of finite-valued random variables, and the lossy rate region for a tuple of arbitrary distributed random variables. Many extensions and generalizations followed Gray and Wyner’s fundamental work. Wyner [2] introduced an operational definition of the common information between a tuple of sources that generate symbols with values in finite spaces. Wyner’s operational definition of common information is defined as the minimum achievable common message rate on the Gray and Wyner lossless rate region. Witsenhausen [3] investigated bounds for Wyner’s common information, and sequences of pairs of random variables in this regard [4]. Gács and Körner [5] introduced another definition of common randomness between a tuple of jointly independent and identically distributed random variables. Benammar and Zaidi [6,7] characterized the Gray–Wyner rate region, when there is side information at the decoders, under various scenarios that include both receivers and reproduce the source symbols without distortion. Insightful application examples for binary sources are considered in [7] (Section 4.2). In...
their previous work, Benammar and Zaidi [8,9] characterized the rate distortion function of the Heegard and Berger [10] problem, with two sources and side information at the two decoders (under a degraded set-up). Connections between the Gray and Wyner lossy source coding network and the notions of empirical and strong coordination capacity for arbitrary networks were developed by Cuff, Permuter and Cover [11] and the references therein, where the authors elaborated on the usefulness of the common information between the different network nodes.

Viswanatha, Akyol and Rose [12], and Xu, Liu and Chen [13], explored the connection of Wyner’s common information and the Gray and Wyner lossy rate region, to generalize Wyner’s common information to its lossy counterpart, for random variables taking values in arbitrary spaces. They characterized Wyner’s lossy common information as the minimum common message rate on the Gray and Wyner lossy rate region, when the sum rate is arbitrarily close to the rate distortion function with joint decoding for the Gray and Wyner lossy network. Applications to encryption and secret key generation are discussed by Viswanatha, Akyol and Rose in [12] (and references therein).

The current paper is focused on the calculations of rates that lie in the Gray and Wyner rate region [1], for two sources that generate symbols, according to the model of jointly independent and identically distributed multivariate correlated Gaussian random variables $Y_1: \Omega \to \mathbb{R}^{p_1}$, $Y_2: \Omega \to \mathbb{R}^{p_2}$, and square-error fidelity at the two decoders. The current literature on methods and algorithms to compute such rates are subject to a number of limitations which often prevent their practical usefulness:

1. Rates that lie in the Gray and Wyner rate region are only known for the special case of a tuple of scalar-valued Gaussian random variables with square error distortion, i.e., $p_1 = p_2 = 1$ [1,12,13].

2. Wyner’s lossy common information is only computed in closed form, for the special cases of a tuple of scalar-valued Gaussian random variables, [12,13].

3. Important generalizations to a tuple of sources that generate multivariate Gaussian symbols, require new derivations often of considerable difficulty.

4. Realizations of the optimal test channel distributions and their structural properties of the various rate distortion functions (RDFs), which are involved in the Gray and Wyner characterization of the rate region, are not developed.

5. A proof that the Gray and Wyner for jointly Gaussian sources, is characterized by a Gaussian auxiliary random variable $W$ is still missing from past literature.

It is known from [1] that the Gray and Wyner rate region can be parameterized by an auxiliary random variable $W: \Omega \to \mathbb{W}$, via several rate distortion functions. Moreover, subsets of the Gray and Wyner rate region are parameterized by $W$ which satisfies conditional independence (1).

$$P_{Y_1,Y_2|W} = P_{Y_1|W} P_{Y_2|W}.$$  

(1)

The current paper makes use of the canonical variable form and the weak stochastic realization of the tuple of random variables $(Y_1, Y_2)$, introduced in Section 2 to characterize subsets of the Gray and Wyner rate region, which are parameterized by jointly Gaussian random variables $(Y_1, Y_2, W)$ with $W: \Omega \to \mathbb{R}^n$, where $n$ is a finite number, while in some cases, the minimum dimension of $W$ is clarified. The weak stochastic realization is developed to deal with the fundamental issue that, for the Gray and Wyner network, one is given the joint distribution $P_{Y_1,Y_2}$, while the characterization of the RDFs involves the specification of the test channel distributions, that achieve these RDFs, and the actual construction of realizations of all random variables involved, that induce the test channel distributions. Furthermore, Wyner’s common information between $Y_1$ and $Y_2$ involves the construction of a joint distribution $P_{Y_1,Y_2,W}$ where $W$ is the auxiliary random variable that makes $Y_1$ and $Y_2$ conditionally independent, i.e., (1) holds.

The rest of the section serves mainly to review the Gray and Wyner characterization of the lossy rate region and the characterization of Wyner’s lossy common information.
1.1. Literature Review

(a) The Gray and Wyner source coding for a simple network [1].

Consider the Gray and Wyner source coding for a simple network, as shown in Figure 1, for a tuple of jointly independent and identically distributed multivariate Gaussian random variables \((Y^N_1, Y^N_2)\) \(\{ (Y^N_{1,i}, Y^N_{2,i}) : i = 1, 2, \ldots, N \}\),

\[
Y_{1,i} : \Omega \rightarrow \mathbb{R}^{p_1} = \mathbb{Y}_1, \quad Y_{2,i} : \Omega \rightarrow \mathbb{R}^{p_2} = \mathbb{Y}_2, \quad i = 1, \ldots, N
\]

with square error distortion functions at the two decoders,

\[
D_{Y_1}(y_1^N, \hat{y}_1^N) = \frac{1}{N} \sum_{i=1}^{N} ||y_{1,i} - \hat{y}_{1,i}||^2_{\mathbb{R}^{p_1}}, \quad D_{Y_2}(y_2^N, \hat{y}_2^N) = \frac{1}{N} \sum_{i=1}^{N} ||y_{2,i} - \hat{y}_{2,i}||^2_{\mathbb{R}^{p_2}}
\]

where \(|| \cdot ||_{\mathbb{R}^{p_i}}\) are Euclidean distances on \(\mathbb{R}^{p_i}, i = 1, 2\).

\[\text{Figure 1. The Gray and Wyner source coding for a simple network [1] \(Y_{1,j}, Y_{2,j} \sim P_{Y_{1,j}Y_{2,j}}, i = 1, \ldots, N\).}\]

The encoder takes as its input the data sequences \((Y^N_1, Y^N_2)\) and produces at its output three messages, \((S_0, S_1, S_2)\), with binary bit representations \((NR_0, NR_1, NR_2)\), respectively. There are three channels, channel 0, channel 1, channel 2, with capacities \((C_0, C_1, C_2)\) (in bits per second), respectively, to transmit the messages to two decoders. Channel 0 is a common channel and channel 1 and channel 2 are the private channels which connect the encoder to each of the two decoders. Message \(S_0\) is a common or public message that is transmitted through the common channel 0 with capacity \(C_0\) to decoder 1 and decoder 2; \(S_1\) is a private message which is transmitted through the private channel 1 with capacity \(C_1\) to decoder 1; and \(S_2\) is a private message, which is transmitted through the private channel 2 with capacity \(C_2\) to decoder 2.

Decoder 1 aims to reproduce \(Y_1^N\) by \(\hat{Y}_1^N\) subject to an average distortion and decoder 2 aims to reproduce \(Y_2^N\) by \(\hat{Y}_2^N\), subject to an average distortion, where \((\hat{Y}_{1,i}, \hat{Y}_{2,i}) = (\hat{y}_{1,i}, \hat{y}_{2,i}) \in \hat{\mathbb{Y}}_1 \times \hat{\mathbb{Y}}_2 \subseteq \mathbb{Y}_1 \times \mathbb{Y}_2, i = 1, \ldots, N\), that is,

\[
\mathbb{E}\{D_{Y_1}(Y_1^N, \hat{Y}_1^N)\} \leq \Delta_1, \quad \mathbb{E}\{D_{Y_2}(Y_2^N, \hat{Y}_2^N)\} \leq \Delta_2, \quad (\Delta_1, \Delta_2) \in [0, \infty] \times [0, \infty].
\]

Gray and Wyner characterized the rate region, denoted by \(R_{GW}(\Delta_1, \Delta_2)\), via a coding scheme that uses the auxiliary random variable \(W\), and the family of probability distributions

\[
P \triangleq \{ P_{Y_1, Y_2, W}(y_1, y_2, w) : y_1 \in \mathbb{Y}_1, y_2 \in \mathbb{Y}_2, w \in \mathbb{W}, P_{Y_1, Y_2, W}(y_1, y_2, \infty) = P_{Y_1, Y_2}(y_1, y_2) \}
\]

such that the joint probability distribution \(P_{Y_1, Y_2, W}(y_1, y_2, w)\) on \(\mathbb{Y}_1 \times \mathbb{Y}_2 \times \mathbb{W}\), has a \((Y_1, Y_2)\) – marginal probability distribution \(P_{Y_1, Y_2}(y_1, y_2)\) on \(\mathbb{Y}_1 \times \mathbb{Y}_2\) that coincides with the probability distribution of \((Y_1, Y_2)\).
For the source and distortion function specified in (2) and (3), we apply the weak stochastic realization to construct the family of distributions $\mathcal{P}$, which is parameterized by the auxiliary random variable $W$. Use is made of the characterization of $\mathcal{R}_{GW}(\Delta_1, \Delta_2)$ described in terms of an auxiliary random variable, as follows.

**Theorem 1** (Theorem 8 in [1]). Let $\mathcal{R}_{GW}(\Delta_1, \Delta_2)$ denote the Gray and Wyner rate region of the simple network shown in Figure 1. Suppose there exists $\bar{q}_i \in \mathbb{F}$ such that $E\{d_{Y_i}(Y_i, \hat{Y}_i)\} < \infty$, for $i = 1, 2$. For each $P_{Y_1,Y_2,W} \in \mathcal{P}$ and $\Delta_1 \geq 0, \Delta_2 \geq 0$, define the subset of Euclidean 3--dimensional space

$$\mathcal{R}_{GW}^{Y_1,Y_2,W}(\Delta_1, \Delta_2) = \left\{ (R_0, R_1, R_2) \left| R_0 = I(Y_1,Y_2,W), \ R_1 = R_{Y_1|W}(\Delta_1), \ R_2 = R_{Y_2|W}(\Delta_2) \right\} \right.$$  

where $R_{Y_i|W}(\Delta_i)$ is the conditional rate distortion function of $Y_i^N$, conditioned on $W^N$, at decoder $i$, for $i = 1, 2$, and $R_{Y_1,Y_2}(\Delta_1, \Delta_2)$ is the joint rate distortion function of the joint decoding of $(Y_1^N, Y_2^N)$ (all single letters). Let

$$\mathcal{R}_{GW}^*(\Delta_1, \Delta_2) = \left( \bigcup_{P_{Y_1,Y_2,W} \in \mathcal{P}} \mathcal{R}_{GW}^{Y_1,Y_2,W}(\Delta_1, \Delta_2) \right)^c$$

where $(\cdot)^c$ denotes the closure of the indicated set. The achievable Gray–Wyner lossy rate region is given by

$$\mathcal{R}_{GW}(\Delta_1, \Delta_2) = \mathcal{R}_{GW}^*(\Delta_1, \Delta_2).$$

Gray and Wyner [1] (Theorem 6) also showed that, if $(R_0, R_1, R_2) \in \mathcal{R}_{GW}(\Delta_1, \Delta_2)$, then

$$R_0 + R_1 + R_2 \geq R_{Y_1,Y_2}(\Delta_1, \Delta_2),$$

$$R_0 + R_1 \geq R_{Y_1}(\Delta_1),$$

$$R_0 + R_2 \geq R_{Y_2}(\Delta_2)$$

where $R_{Y_i}(\Delta_i)$ is the rate distortion function of $Y_i^N$ at decoder $i$, for $i = 1, 2$, and $R_{Y_1,Y_2}(\Delta_1, \Delta_2)$ is the joint rate distortion function of $(Y_1^N, Y_2^N)$ at the two decoders. The inequality in (9) is called the *Pangloss Bound* of the Gray–Wyner lossy rate region $\mathcal{R}_{GW}(\Delta_1, \Delta_2)$. The set of triples $(R_0, R_1, R_2) \in \mathcal{R}_{GW}(\Delta_1, \Delta_2)$ that satisfy the equality $R_0 + R_1 + R_2 = R_{Y_1,Y_2}(\Delta_1, \Delta_2)$ is called the *Pangloss Plane* of the Gray–Wyner lossy rate region $\mathcal{R}_{GW}(\Delta_1, \Delta_2)$.

Gray and Wyner proved that $\mathcal{R}_{GW}(\Delta_1, \Delta_2)$, is also determined from [1] (4) of page 1703, Equation (42),

$$T(\alpha_1, \alpha_2) = \inf_{P_{Y_1,Y_2,W} \in \mathcal{P}} \left\{ I(Y_1,Y_2;W) + \alpha_1 R_{Y_1|W}(\Delta_1) + \alpha_2 R_{Y_2|W}(\Delta_2) \right\}$$

where $0 \leq \alpha_i \leq 1$, $i = 1, 2$, $\alpha_1 + \alpha_2 \geq 1$, and where for each $P_{Y_1,Y_2,W} \in \mathcal{P}$, the conditional distribution $P_{Y_i|W}$ is defined, from which follows the $Y_i$--marginals $P_{Y_i|W}$, $i = 1, 2$.

(b) **Wyner’s common Information of finite-valued random variables.**

Wyner [2] introduced an operational definition of the common information between a tuple of random variables $(Y_1^N, Y_2^N)$ that takes values in finite spaces.

The first approach of Wyner’s operational definition of common information between sequences $Y_1^N$ and $Y_2^N$ is defined as the minimum achievable common message rate $R_0$ on the Gray–Wyner network of Figure 1.

Wyner’s single letter information theoretic characterization of the infimum of all achievable message rates $R_0$, called Wyner’s common information, is defined by

$$C(Y_1, Y_2) = \inf_{P_{Y_1,Y_2,W} : P_{Y_1,Y_2,W} = P_{Y_1|W}P_{Y_2|W}} I(Y_1,Y_2;W).$$
Here, $P_{Y_1, Y_2, W}$ is any joint probability distribution on $\mathbb{Y}_1 \times \mathbb{Y}_2 \times \mathbb{W}$ with $(Y_1, Y_2)$—marginal $P_{Y_1, Y_2}$, such that $W$ makes $Y_1$ and $Y_2$ conditionally independent, that is $P_{Y_1, Y_2, W} \in \mathcal{P}$.

(c) Minimum common message rate and Wyner’s lossy common information for arbitrary random variables.

Viswanatha, Akyol and Rose [12], and Xu, Liu and Chen [13] explored the connection of Wyner’s common information and the Gray–Wyner lossy rate region, to provide a new interpretation of Wyner’s common information to its lossy counterpart. The following characterization was derived by Xu, Liu and Chen [13] (an equivalent characterization was also derived by Viswanatha, Akyol and Rose [12]).

**Theorem 2** (Theorem 4 in [13]). Suppose there exists $\hat{y}_i \in \mathcal{Y}_i$ such that $E\{d_Y(Y_i, \hat{y}_i)\} < \infty$, for $i = 1, 2$.

Let $C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2)$ denote the minimum common message rate $R_0$ on the Gray–Wyner lossy rate region $\mathcal{R}_{GW}(\Delta_1, \Delta_2)$, with a sum rate not exceeding the joint rate distortion function, $\sum_{i=0}^{2} R_i \geq R_{Y_1, Y_2}(\Delta_1, \Delta_2)$, while satisfying the average distortions. Then, $C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2)$ is characterized by the optimization problem

$$C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2) = \inf \ I(Y_1, Y_2; W)$$

such that the following identity holds

$$R_{Y_1|W}(\Delta_1) + R_{Y_2|W}(\Delta_2) + I(Y_1, Y_2; W) = R_{Y_1, Y_2}(\Delta_1, \Delta_2)$$

where the infimum is over all random variables $W$ taking values in $\mathcal{W}$, which parameterize the source distribution via $P_{Y_1, Y_2, W}$, having a $\mathbb{Y}_1 \times \mathbb{Y}_2$—marginal source distribution $P_{Y_1, Y_2}$, and induce joint distributions $P_{W, Y_1, Y_2, \hat{Y}_1, \hat{Y}_2}$ which satisfy the constraints.

It is shown in [12,13] that there exists a distortion region such that $C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2) = C_W(Y_1, Y_2)$, i.e., it is independent of the distortions $(\Delta_1, \Delta_2)$, and $C_W(Y_1, Y_2) = C(Y_1, Y_2)$, i.e., it is equal to Wyner’s information theoretic characterization of common information between $Y_1$ and $Y_2$, defined by (13). However, their proofs that $W$ is finite-dimensional Gaussian relies on the assumption that $W$ is continuous-valued.

The next theorem is derived by Xu, Liu and Chen [13].

**Theorem 3** (Theorem 5 in [13]). Let $(Y_1, Y_2)$ be a pair of random variables with distribution $P_{Y_1, Y_2}$ on the alphabet space $\mathbb{Y}_1 \times \mathbb{Y}_2$, where $\mathbb{Y}_1$ and $\mathbb{Y}_2$ are arbitrary measurable spaces that can be discrete or continuous.

Let $W$ be any random variable achieving $C(Y_1, Y_2)$ defined by (13).

Let the reproduction alphabet $\mathcal{Y}_1 = \mathbb{Y}_1$, $\mathcal{Y}_2 = \mathbb{Y}_2$ and two per-letter distortion measures $d_{Y_1}(y_1, \hat{y}_1), d_{Y_2}(y_2, \hat{y}_2)$ satisfy

$$d_{Y_1}(y_1, \hat{y}_1) > d_{Y_1}(y_1, y_1) = 0, \ y_i \neq \hat{y}_i, \ i = 1, 2$$

If the following conditions are satisfied:

1. For any $y_1 \in \mathbb{Y}_1$, $y_2 \in \mathbb{Y}_2$ and $w \in \mathcal{W}$, $P_{W|Y_1, Y_2} > 0$;
2. There exists an $\hat{y}_i \in \mathcal{Y}_i$, such that $E\{d_{Y_i}(Y_i, \hat{y}_i)\} < \infty$, $i = 1, 2$

then there exists a strictly positive vector $\gamma = (\gamma_1, \gamma_2) \in (0, \infty) \times (0, \infty)$, such that, for $0 \leq (\Delta_1, \Delta_2) \leq \gamma$,

$$C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2) = C_W(Y_1, Y_2) = C(Y_1, Y_2).$$
Moreover, $C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2)$ is constant on $D_W = \{ (\Delta_1, \Delta_2) \in [0, \infty] \times [0, \infty] : 0 \leq (\Delta_1, \Delta_2) \leq \gamma \}$. 

The analog of the above theorem is also derived by Viswanatha, Akyol and Rose in [12] (Lemma 1). A subset of the Pangloss plane is derived by Gray and Wyner [1] (Theorem 9). For a bivariate Gaussian random variables, i.e., $p_1 = p_2 = 1$, with square-error distortions, Viswanatha, Akyol and Rose in [12], and Xu, Liu and Chen [13] computed $C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2)$, by using Xiao's and Luo's [14] (Theorem 6) the closed-form expression of joint rate distortion function $R_{Y_1,Y_2}(\Delta_1, \Delta_2)$. In addition, for the bivariate Gaussian random variables, with symmetric square-error distortions, i.e., $\Delta_1 = \Delta_2 = \Delta$, Gray and Wyner [1] (Section 2.5, (B)), computed a rate-triple $(R_0, R_1, R_2) \in R_{GW}(\Delta_1, \Delta_2)$ that lies on the Pangloss plane.

1.2. Main Theorems and Discussion

What follows is a brief summary of the main theorems derived in this paper, and relations to the literature.

Theorem 9 shows that, among all joint distributions $P_{Y_1, Y_2, W}$ induced by a tuple of multivariate correlated Gaussian random variables $(Y_1, Y_2)$, and an arbitrary random variable $W : \Omega \rightarrow \mathbb{W}$, continuous or discrete-valued, Wyner’s common information $C(Y_1, Y_2)$, defined by (13), is minimized by a tuple $(Y_1, Y_2, W)$ which induces a jointly Gaussian distribution $P_{Y_1, Y_2, W}$, and $W : \Omega \rightarrow \mathbb{W} = \mathbb{R}^n$ is a finite-dimensional Gaussian random variable. In particular, Theorem 9 gives the weak stochastic realization of $(Y_1, Y_2)$, and the construction of the random variable $W$, which induce a joint distribution $P_{Y_1, Y_2, W}$ that achieves the minimum of $I(Y_1, Y_2; W)$ such that $W$ makes $Y_1$ and $Y_2$ conditionally independent.

Then, use is made of Theorem 9, Section 2.2, such as Definition 1 of the canonical variable form and the weak stochastic realization to derive Wyner’s common information $C(Y_1, Y_2)$ defined by (13), and the optimal realization of the triple $(Y_1, Y_2, W^*) = (Y_1, Y_2, W^*)$ that achieves $C(Y_1, Y_2)$, as stated in the next theorem.

Theorem 4. Consider a tuple of Gaussian random variables $Y_i : \Omega \rightarrow \mathbb{R}^p_i$, with $Q_{Y_i} > 0$, for $i = 1, 2$, $(Y_1, Y_2) \in G(0, Q_{(Y_1, Y_2)})$, $Q_{(Y_1, Y_2)} \geq 0$, and apply Algorithm A1 (and the notation therein) to decompose and transform the random variables into a canonical variable form (with abuse of notation, the transform random variables are denoted by $(Y_1, Y_2) \in G(0, Q_{CVF})$, $(Y_1, Y_2) \in G(0, Q_{CVF})$, using the material and notation of Section 2.2, i.e., Definition 1.

(a) Then,

$$C(Y_1, Y_2) = C(Y_{11}, Y_{21}) + C(Y_{12}, Y_{22}) + C(Y_{13}, Y_{23})$$

$$= \left\{ \begin{array}{ll}
0, & \text{if } p_{13} > 0, p_{23} > 0, p_{11} = p_{12} = p_{21} = p_{22} = 0, \\
\frac{1}{2} \sum_{i=1}^n \ln \left( \frac{1 + d_i}{1 - d_i} \right), & \text{if } p_{12} = p_{22} > 0, p_{11} = p_{21} = 0, p_{13} \geq 0, p_{23} \geq 0, \\
+\infty, & \text{if } p_{11} = p_{21} > 0
\end{array} \right.$$ 

where $(p_{11}, p_{12}, p_{13})$ and $(p_{21}, p_{22}, p_{23})$ are the dimensions of the canonical variable decomposition of the tuple $(Y_1, Y_2)$, and

$$C(Y_{11}, Y_{21}) = +\infty, \text{ if } p_{11} = p_{21} > 0;$$

$$C(Y_{13}, Y_{23}) = 0, \text{ if } p_{13} > 0 \text{ and } p_{23} > 0;$$

$$C(Y_{12}, Y_{22}) = \frac{1}{2} \sum_{i=1}^n \ln \left( \frac{1 + d_i}{1 - d_i} \right), \text{ if } n = p_{12} = p_{22} > 0.$$
Thus, $C(Y_{12}, Y_{22})$ is the most interesting value if defined.

(b) The random variable $W^*$ defined below is such that $C(Y_1, Y_2)$ of part (a) is attained.

$$W^* : \Omega \to \mathbb{R}^n, \quad n \in \mathbb{Z}_+,$$

$$n_1 = p_{11} = p_{21}, \quad n_2 = p_{12} = p_{22}, \quad n_1 + n_2 = n,$$

$$W^* = \begin{pmatrix} W_1^* \\ W_2^* \end{pmatrix}, \quad W_1^* : \Omega \to \mathbb{R}^{n_1}, \quad W_2^* : \Omega \to \mathbb{R}^{n_2},$$

$$W_1^* = Y_{11} = Y_{21},$$

$$W_2^* = L_1 Y_{12} + L_2 Y_{22} + L_3 V, \quad \text{see Theorem 11. (b) for the formulas of } L_1, L_2, L_3$$

where the following properties hold:

then $(Y_1, Y_2, W^*) \in G(0, Q_0(I))$, see (81) for $Q_0(I)$,

$$(F_{Y_{11}, Y_{12}, Y_{13}}, F_{Y_{21}, Y_{22}, Y_{23}}) \subseteq \mathcal{F}_{W_1^*, W_2^*} \in \mathcal{C}I,$$

$$(F_{W_1^*} \subseteq (F_{Y_{11}} \lor F_{Y_{21}}), F_{W_2^*} \subseteq (F_{Y_{12}} \lor F_{Y_{22}}),$$

$$C(Y_{11}, Y_{22}) = I(Y_1, Y_2; W^*).$$

(c) The following operations are defined, using (a),

$$W^* = \begin{pmatrix} W_1^* \\ W_2^* \end{pmatrix},$$

$$W_1^* = Y_{11} = Y_{21},$$

$$W_2^* = L_1 Y_{12} + L_2 Y_{22} + L_3 V, \quad \text{see (103),(104) for the formulas of } L_1, L_2, L_3;$$

$$Z_{12} = Y_{12} - E[Y_{12}|F_{W_1^*}] = Y_{12} - Q_{Y_{12}, W_2^*} Q_{W_2^*}^{-1} W_2^*,$$

$$Z_{22} = Y_{22} - E[Y_{22}|F_{W_1^*}] = Y_{22} - Q_{Y_{22}, W_2^*} Q_{W_2^*}^{-1} W_2^*,$$

$$Z_{13} = Y_{13}, \quad Z_{23} = Y_{23}, \quad \text{(the components } Z_{11} \text{ and } Z_{21} \text{ do not exist)},$$

$$Z_{1} = \begin{pmatrix} Z_{12} \\ Z_{13} \end{pmatrix}, \quad Z_{2} = \begin{pmatrix} Z_{22} \\ Z_{23} \end{pmatrix}$$

and these imply,

$$Y_{11} = W_1^* = Y_{21},$$

$$Y_{12} = Q_{Y_{12}, W_2^*} Q_{W_2^*}^{-1} W_2^* + Z_{12}, \quad Y_{22} = Q_{Y_{22}, W_2^*} Q_{W_2^*}^{-1} W_2^* + Z_{22},$$

$$Y_{13} = Z_{13}, \quad Y_{23} = Z_{23};$$

equivalently

$$\begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \end{pmatrix} = \begin{pmatrix} I_{n_1} & 0 & 0 \\ 0 & Q_{Y_{12}, W_2^*} Q_{W_2^*}^{-1} & 0 \\ 0 & 0 & I_{n_1 - n_1 - n_2} \end{pmatrix} \begin{pmatrix} W_1^* \\ W_2^* \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & I_{n_2} & 0 \\ 0 & 0 & I_{n_1 - n_1 - n_2} \end{pmatrix} \begin{pmatrix} Z_{12} \\ Z_{13} \end{pmatrix},$$

$$\begin{pmatrix} Y_{21} \\ Y_{22} \\ Y_{23} \end{pmatrix} = \begin{pmatrix} I_{n_1} & 0 & 0 \\ 0 & Q_{Y_{22}, W_2^*} Q_{W_2^*}^{-1} & 0 \\ 0 & 0 & I_{n_1 - n_1 - n_2} \end{pmatrix} \begin{pmatrix} W_1^* \\ W_2^* \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & I_{n_2} & 0 \\ 0 & 0 & I_{n_1 - n_1 - n_2} \end{pmatrix} \begin{pmatrix} Z_{22} \\ Z_{23} \end{pmatrix}.$$

The derivation of Theorem 4 is presented in Section 3.2, after several of the tools are presented, such as, weak stochastic realizations and minimal realizations.

Remark 1. Relation of Theorem 4 to the literature.

(a) Corollary 1 in [15] gives an expression analogous to the case (22), which is expressed in terms of the correlation coefficients, $\rho_i \in (-1, 1)$ and not the canonical correlation coefficients $d_i \in (0, 1)$. 

Similarly, [16], under Lemma 1, reproduces Corollary 1 in [15], with the correlation coefficients $\rho_i$ replaced by their absolute values $|\rho_i|$.

(b) The derivation in [15,16] is based on the use of rate distortion functions of Gaussian random variables with square-error distortion functions, which presupposes that the auxiliary RV $W \to \mathbb{W}$ takes continuous values.

(c) Refs. [15,16] do not provide a realization of the triple $(Y_1, Y_2, W^*)$, as given in Theorem 4 (which is based on applying the parametrization of Theorem 8).

On the other hand, the derivation of Theorem 4 is based on Theorem 9, which shows that, among all joint distributions $P_{Y_1, Y_2, W}$ induced by a tuple of multivariate correlated Gaussian random variables $(Y_1, Y_2)$, and an arbitrary random variable $W : \Omega \to \mathbb{W}$, continuous or discrete-valued, Wyner’s common information $C(Y_1, Y_2)$, defined by (13), is minimized by a triple $(Y_1, Y_2, W)$ which induces a jointly Gaussian distribution $P_{Y_1, Y_2, W}$, and $W : \Omega \to \mathbb{W} = \mathbb{R}^n$ is finite-dimensional Gaussian random variable.

(d) The derivation of Theorem 4 contains many intermediate results which are applicable to the problems considered in [15,16], such as Relaxed Wyner’s Common Information in [17]. These are discussed in Section 4.3.

Theorem 5 gives a parametric characterization of the Gray and Wyner rate region $\mathcal{R}_{GW}(\Delta_1, \Delta_2)$, with respect to the variance matrix of the tuple of jointly Gaussian random variables $(Y_1, Y_2, W)$.

**Theorem 5.** Consider a tuple of Gaussian random variables $Y_i : \Omega \to \mathbb{R}^{p_i}$, with $Q_{Y_i} > 0$, for $i = 1, 2$, $(Y_1, Y_2) \in G(0, Q_{(Y_1,Y_2)})$ (not necessarily in canonical variable form), with induced Gaussian measure $P_0 = G(0, Q_{(Y_1,Y_2)})$ on the space $(\mathbb{R}^{p_1} \times \mathbb{R}^{p_2}, B(\mathbb{R}^{p_1}) \otimes B(\mathbb{R}^{p_2}))$, and square error distortion functions $D_{Y_1}(y_1, \hat{y}_1) = ||y_1 - \hat{y}_1||^2_{B^{p_1}}, D_{Y_2}(y_2, \hat{y}_2) = ||y_2 - \hat{y}_2||^2_{B^{p_2}}$.

The following hold.

(a) There exists a Gaussian measure $P_1 = G(0, Q_{(Y_1,Y_2),W})$ defined on the space $(\mathbb{R}^{p_1} \times \mathbb{R}^{p_2} \times \mathbb{R}^n, B(\mathbb{R}^{p_1}) \otimes B(\mathbb{R}^{p_2}) \otimes B(\mathbb{R}^n))$, $n \in \mathbb{Z}_+$, associated with the Gaussian random variables $(Y_1, Y_2)$, $W : \Omega \to \mathbb{R}^n$, $W \in G(0, Q_{W})$ such that $P_1|_{\mathbb{R}^{p_1} \times \mathbb{R}^{p_2}} = G(0, Q_{(Y_1,Y_2)})$. Moreover, a realization of the random variables $(Y_1, Y_2, W)$ with induced measure $P_1 = G(0, Q_{(Y_1,Y_2),W})$ is

\begin{align*}
\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} &= Q_{(Y_1,Y_2),W} Q_W^t W + \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \quad (39) \\
(Z_1, Z_2) &\in G(0, Q_{(Z_1,Z_2)}), \quad (Z_1, Z_2) \text{ independent of } W, \quad (40) \\
Q_{(Y_1,Y_2),W} &= \mathbb{E}[\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} W^T] = \begin{pmatrix} Q_{Y_1,W} \\ Q_{Y_2,W} \end{pmatrix}. \quad (41)
\end{align*}
where $Q^\dagger_W$ is the pseudoinverse of $Q_W$.

(b) For Gaussian auxiliary random variables given in part (a), the Gray–Wyner rate region $R_{GW}(\Delta_1, \Delta_2)$ is determined from

$$T^G(\alpha_1, \alpha_2) = \inf_{(Y_1, Y_2, W) \in G(0, Q_{(Y_1,Y_2,W)})} \left\{ I(Y_1, Y_2; W) \right\}$$

$$= a_1 R_{Y_1|W}(\Delta_1) + a_2 R_{Y_2|W}(\Delta_2)$$

$$= \inf_{(Y_1, Y_2, W) \in G(0, Q_{(Y_1,Y_2,W)})} \left\{ I(Y_1, Y_2; W) \right\}$$

$$= a_1 R_{Y_1|W}(\Delta_1) + a_2 R_{Y_2|W}(\Delta_2)$$

subject to $Q_{(Y_1,Y_2)} = Q_{(Y_1,Y_2,W)}$, $Q_{Y_1,Y_2,W} = Q_{Y_1|W} - Q_{Y_1,Y_2|W}Q^\dagger_{Y_2|W}Q_{Y_2|W}$

where $0 \leq \alpha_i \leq 1$, $i = 1, 2$, $a_1 + a_2 \geq 1$, $I(Y_1, Y_2; W) = H(Y_1, Y_2) - H(Y_1|Y_2, W) - H(Y_2|W)$, and $R_{Y_i|W}(\Delta_i), i = 1, 2$ are given in Theorem 13.(b), and $(\cdot)^+ = \max\{1, \cdot\}$.

The derivation of Theorem 5 is presented in Section 4.4, after the structural properties of RDFs, $R_{Y_1,Y_2}(\Delta_1, \Delta_2), R_{Y_i|W}(\Delta_i), R_{Y_i}(\Delta_i), i = 1, 2$, of Theorem 12, Theorem 13, Theorem 14 are presented. From Theorem 5, follow simplified characterizations of subsets of the rate region $R_{GW}(\Delta_1, \Delta_2)$, such as rates that lie on Pangloss Plane, and rates that correspond to $W$ that make $Y_1$ and $Y_2$ conditional independent, i.e., $W$ is such that $P_{Y_1,Y_2|W} = P_{Y_1|W}P_{Y_2|W}$.

Utilizing the structural properties of RDFs, $R_{Y_i,Y_2}(\Delta_1, \Delta_2), R_{Y_i|W}(\Delta_i), R_{Y_i}(\Delta_i), i = 1, 2$, of Theorem 12, Theorem 13, Theorem 14, and Theorem 4, the next theorem is obtained, which gives the formula of Wyner’s lossy common information $C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2) = C_{W}(Y_1, Y_2)$.

**Theorem 6.** Consider a tuple $(Y_1, Y_2)$ of Gaussian random variables in the canonical variable form of Definition 1. Restrict attention to the correlated parts of these random variables, as defined in Theorem 8, by (77)–(79), and subset of the distortion region is defined by

$$D_W = \left\{ (\Delta_1, \Delta_2) \in [0, \infty] \times [0, \infty] \mid 0 \leq \Delta_1 \leq n(1 - d_1), \ 0 \leq \Delta_2 \leq n(1 - d_1) \right\},$$

$$\forall j \in Z_n, \ d_j \in (0, 1).$$

Then, Wyner’s lossy common information (calculation of expression in Theorem 3) is given by

$$C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2) = C_{W}(Y_1, Y_2)$$

$$= C(Y_1, Y_2) = \frac{1}{2} \sum_{j=1}^{n} \ln \left( \frac{1 + d_j}{1 - d_j} \right), \ (\Delta_1, \Delta_2) \in D_W$$

The derivation of Theorem 6 is presented in Section 4.2 and makes use of a degenerate version of the realization of the triple $(Y_1, Y_2, W^*)$ given in Theorem 4, and the RDFs $R_{Y_i,Y_2}(\Delta_1, \Delta_2), R_{Y_i|W}(\Delta_i), i = 1, 2$.

**Remark 2.** By Theorem 5, a subset of the Gray–Wyner rate region is obtained by replacing $(Y_1, Y_2, W) \in G(0, Q_{(Y_1,Y_2,W)})$ of (39) and (40) by $W$ that makes $Y_1$ and $Y_2$ conditionally independent, i.e., $(Z_1, Z_2) \in G(0, Q_{(Z_1,Z_2)})$ and $(Z_1, Z_2, W)$ mutually independent (e.g., $Q_{(Z_1,Z_2)}$ is block-diagonal).
1.3. Structure of the Paper

Section 2 introduces the mathematical tools of the geometric approach to Gaussian random variables, the weak stochastic realization of conditional independence (Section 2.4).

Section 3 contains the problem statement, the solution procedure and the weak realization of a tuple of multivariable random variables \((Y_1, Y_2)\) such that another multivariate Gaussian random variable \(W\) makes \(Y_1\) and \(Y_2\) conditionally independent (Section 2.5). \(C_W(Y_1, Y_2) = C(Y_1, Y_2)\).

Section 4 is concerned with the characterization of the Gray–Wyner rate region \(R_{GW}(\Delta_1, \Delta_2)\), the characterization of rates that lie on the Pangloss Plane, and Wyner’s lossy common information. This section includes calculations of the rate distortion functions \(R_{Y_1,Y_2}(\Delta_1, \Delta_2), R_{Y_i|W}(\Delta_i), R_{Y_i}(\Delta_i), i = 1, 2\), the weak stochastic realizations of the random variables \((Y_1, Y_2, \hat{Y}_1, \hat{Y}_2, W)\) which achieve these rate distortion functions, for jointly multivariate Gaussian random variables with square-error distortion functions.

Section 5 includes remarks on possible extensions.

Appendix A.3 makes use of a matrix equality and a determinant inequality first obtained by Hua LooKeng in 1952, which are used to carry out the optimization problem of Wyner’s lossy common information \(C_W(Y_1, Y_2) = C(Y_1, Y_2)\).

2. Probabilistic Properties of Tuples of Random Variables

The reader finds in this section the basic properties associated with:

(1) the transformation of a tuple of Gaussian multivariate random variables \((Y_1, Y_2)\) in their canonical variable form, and

(2) The parameterization of all jointly Gaussian distributions \(P_{Y_1,Y_2,W}(y_1,y_2,w)\) by a zero mean Gaussian random variables \(W \colon \Omega \to \mathbb{R}^k \equiv \mathbb{W}\) such that (a) \(W\) makes the multivariate random variables \((Y_1, Y_2)\) conditional independent, and (b) The marginal distribution \(P_{Y_1,Y_2,W}(y_1,y_2,\infty) = P_{Y_1,Y_2}(y_1,y_2)\) coincides with the joint distribution of the multivariate random variables \((Y_1, Y_2)\).

2.1. Notation of Elements of Probability Theory

The notation used in this paper is briefly specified. Denote by \(Z_+ = \{1, 2, \ldots, \}\) the set of the integers and by \(\mathbb{N} = \{0, 1, 2, \ldots, \}\) the set of the natural integers. For \(n \in Z_+\) denote the following finite subsets of the above defined sets by \(Z_n = \{1, 2, \ldots, n\}\) and \(\mathbb{N}_n = \{0, 1, 2, \ldots, n\}\).

Denote the real numbers by \(\mathbb{R}\) and the set of positive and of strictly positive real numbers, respectively, by \(\mathbb{R}_+ = [0, \infty)\) and \(\mathbb{R}_{++} = (0, \infty) \subset \mathbb{R}\). The vector space of \(n\)-tuples of real numbers is denoted by \(\mathbb{R}^n\). Denote the Borel \(\sigma\)-algebra on this vector space by \(B(\mathbb{R}^n)\), hence, \((\mathbb{R}^n, B(\mathbb{R}^n))\) is a measurable space.

The expression \(\mathbb{R}^{n \times m}\) denotes the set of \(n\) by \(m\) matrices with elements in the real numbers, for \(n, m \in \mathbb{Z}_+\). For the symmetric matrix \(Q \in \mathbb{R}^{n \times n}\), the inequality \(Q \geq 0\) denotes that for all vectors \(u \in \mathbb{R}^n\) the inequality \(u^TQu \geq 0\) holds. Similar, \(Q > 0\) denotes that for all \(u \in \mathbb{R}^n\setminus\{0\}\), \(u^TQu > 0\). The notation \(Q_1 \leq Q_2\) denotes that \(Q_2 - Q_1 \geq 0\).

Consider a probability space denoted by \((\Omega, F, \mathbb{P})\) consisting of a set \(\Omega\), a \(\sigma\)-algebra \(F\) of subsets of \(\Omega\), and a probability measure \(\mathbb{P} : F \to [0,1]\).

A real-valued random variable is a function \(X : \Omega \to \mathbb{R}\) such that the following set belongs to the indicated \(\sigma\)-algebra, \(\{\omega \in \Omega | X(\omega) \in (-\infty,u]\} \in F\) for all \(u \in \mathbb{R}\). A random variable taking values in an arbitrary measurable space \((\mathbb{X}, B(\mathbb{X}))\) is defined correspondingly by \(X : \Omega \to \mathbb{X}\) and \(X^{-1}(A) = \{\omega \in \Omega | X(\omega) \in A\} \in B(\mathbb{X})\), for all \(A \in B(\mathbb{X})\). The measure (or distribution if \(\mathbb{X}\) is an Euclidean space) induced by the random variable on \((\mathbb{X}, B(\mathbb{X}))\) is denoted by \(P_X\) or \(P(dx)\). The \(\sigma\)-algebra generated by a random variable \(X : \Omega \to \mathbb{X}\) is defined as the smallest \(\sigma\)-algebra containing the subsets \(X^{-1}(A) \in F\) for all \(A \in B(\mathbb{X})\). It is denoted by \(F^X\). The real-valued random variable \(X\) is called \(\mathbb{G}\)-measurable for a \(\sigma\)-algebra \(\mathbb{G} \subseteq F\) if the subset \(\{\omega \in \Omega | X(\omega) \in (-\infty,u]\} \in \mathbb{G}\) for all \(u \in \mathbb{R}\).
Denote the set of positive random variables which are measurable on a sub-$\sigma$-algebra $G \subseteq F$ by,

$$L_+(G) = \{ X : \Omega \to \mathbb{R}_+ = [0, \infty) | X \text{ is } G\text{-measurable} \}.$$  

The tuple of sub-$\sigma$-algebras $F_1, F_2 \subseteq F$ is called independent if $\mathbb{E}[X_1 X_2] = \mathbb{E}[X_1] \mathbb{E}[X_2]$ for all $X_1 \in L_+(F_1)$ and all $X_2 \in L_+(F_2)$. The definition can be extended to any finite set of independent sub-$\sigma$-algebras.

2.2. Geometric Approach of Gaussian Random Variables and Canonical Variable Form

The purpose of this section is to introduce the geometric approach of a tuple of finite-dimensional Gaussian random variables using the canonical variable form of the tuple introduced by H. Hotelling, [18]. The use of the geometric approach of two Gaussian random variables with respect to the computation of mutual information is elaborated by Gelfand and Yaglom in [19], making reference to an insight due to Kolmogorov. However, the canonical variable form is not given in [19].

An $\mathbb{R}^n$-valued Gaussian random variable with as parameters the mean value $m_X \in \mathbb{R}^n$ and the variance $Q_X \in \mathbb{R}^{n \times n}$, $Q_X = Q_X^T \geq 0$, is a function $X : \Omega \to \mathbb{R}^n$ which is a random variable with respect to the computation of mutual information. The measure of this random variable equals a Gaussian measure described by its characteristic function,

$$\mathbb{E}[\exp(iu^T X)] = \exp(iu^T m_X - \frac{1}{2} u^T Q_X u), \ \forall \ u \in \mathbb{R}^n.$$  

Note that this definition includes the case in which the random variable is almost surely equal to a constant in which case $Q_X = 0$. A Gaussian random variable with these parameters is denoted by $X \in G(m_X, Q_X)$.

The effective dimension of the random variable is denoted by $\text{dim}(X) = \text{rank}(Q_X)$.

Any tuple of random variables $X_1, \ldots, X_k$ is called jointly Gaussian if the vector $(X_1^T, X_2^T, \ldots, X_k^T)^T$ is a Gaussian random variable. A tuple of Gaussian random variables $(Y_1, Y_2)$ will be denoted this way to save space, rather than by

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}.$$  

Then, the variance matrix of this tuple is denoted by

$$(Y_1, Y_2) \in G(0, Q_{(Y_1, Y_2)}), \ Q_{(Y_1, Y_2)} = \begin{pmatrix} Q_{Y_1} & Q_{Y_1, Y_2} \\ Q_{Y_1, Y_2}^T & Q_{Y_2} \end{pmatrix} \in \mathbb{R}^{(p_1+p_2) \times (p_1+p_2)}.$$  

The reader should distinguish the variance matrices $Q_{(Y_1, Y_2)}$ and $Q_{(Y_1, Y_2)} \in \mathbb{R}^{p_1 \times p_2}$. Any such tuple of Gaussian random variables is independent if and only if $Q_{Y_1, Y_2} = 0$.

**Definition 1.** The canonical variable form.
Consider a tuple of Gaussian random variables $Y_i : \Omega \to \mathbb{R}^{p_i}$, with $Q_{Y_i} > 0$, for $i = 1, 2$, $(Y_1, Y_2) \in G(0, Q_{(Y_1, Y_2)}), \ Q_{(Y_1, Y_2)} \geq 0$. Define the canonical variable form of these random
variables if a basis has been chosen and a transformation of the random variables to this basis has been carried out such that with respect to the new basis, one has the representation

\[(Y_1, Y_2) \in G(0, Q_{\text{cvf}}), \quad \text{where},\]

\[
Q_{\text{cvf}} = \begin{pmatrix}
I_{p_1} & 0 & 0 & I_{p_1} & 0 & 0 \\
0 & I_{p_2} & 0 & 0 & D & 0 \\
0 & 0 & I_{p_3} & 0 & 0 & 0 \\
I_{p_4} & 0 & 0 & I_{p_4} & 0 & 0 \\
0 & D & 0 & 0 & I_{p_2} & 0 \\
0 & 0 & 0 & 0 & 0 & I_{p_3}
\end{pmatrix} \in \mathbb{R}^{p \times p},
\]

\(p, p_1, p_2, p_{11}, p_{12}, p_{13}, p_{21}, p_{22}, p_{23} \in \mathbb{N},\)

\(p = p_1 + p_2, \quad p_1 = p_{11} + p_{12} + p_{13}, \quad p_2 = p_{21} + p_{22} + p_{23}, \quad p_{11} = p_{21}, \quad p_{12} = p_{22},\)

\(D = \text{Diag}(d_1, \ldots, d_{p_{12}}), \quad 1 > d_1 \geq d_2 \geq \ldots \geq d_{p_{12}} > 0,\)

One then says that \((Y_{11}, \ldots, Y_{k_1}), (Y_{21}, \ldots, Y_{2k_2})\) are the canonical variables and \((d_1, \ldots, d_{k_{12}})\) the canonical correlation coefficients.

If \(Q_{(Y_1, Y_2)} > 0\) then necessarily \(p_{11} = p_{21} = 0.\)

Appendix A.1 gives Algorithm A1 to transform the variance matrix \(Q_{(Y_1, Y_2)} \geq 0\) by two nonsingular transformations \(S_i \in \mathbb{R}^{p_i \times p_i}, \quad i = 1, 2,\) to its canonical variable form \(Q_{\text{cvf}}\) of Definition 1, such that

\[
Y = \begin{pmatrix}
Y_1 \\
Y_2
\end{pmatrix} = \begin{pmatrix}
Y_{11} & Y_{12} & Y_{13} \\
Y_{21} & Y_{22} & Y_{23}
\end{pmatrix}, \quad Y_{ij} : \Omega \rightarrow \mathbb{R}^{p_i}, \quad i = 1, 2, j = 1, 2, 3.
\]  

**Proposition 1. Properties of components of the canonical variable form.**

Consider a tuple \((Y_1, Y_2) \in G(0, Q_{\text{cvf}})\) of Gaussian random variables in the canonical variable form.

(a) The three components \(Y_{11}, Y_{12}, Y_{13}\) of \(Y_1\) are independent random variables. Similarly, the three components \(Y_{21}, Y_{22}, Y_{23}\) of \(Y_2\) are independent random variables.

(b) The equality \(Y_{11} = Y_{21}\) of these random variables holds almost surely.

(c) The tuple of random variables \((Y_{12}, Y_{22})\) is correlated as shown by the formula

\[
\mathbb{E}[Y_{12}Y_{22}^T] = D.
\]

\(Y_{11} = Y_{21} - a.s.,\)  

\(E[Y_{12}Y_{22}^T] = D.
\]

Note that the different components of \(Y_{12}\) and of \(Y_{22}\) are independent random variables; thus, \(Y_{12,j}\) and \(Y_{12,j}\) are independent, and \(Y_{22,j}\) and \(Y_{22,j}\) are independent, and \(Y_{12,j}\) and \(Y_{22,j}\) are independent, for all \(i \neq j;\) and that \(Y_{12,j}\) and \(Y_{22,j}\) for \(j = 1, \ldots, p_{12} = p_{22}\) are correlated.

(d) The random variable \(Y_{13}\) is independent of \(Y_2.\) Similarly, the random variable \(Y_{23}\) is independent of \(Y_1.\)

**Proof.** The results are immediately obvious from the fact that the random variables are all jointly Gaussian and from the variance formula (50) of the canonical variable form.

Next, the interpretation of the various components of the canonical variable form is defined, as in [20].
Definition 2. Interpretation of components of the canonical variable form.
Consider a tuple of jointly Gaussian random variables \((Y_1, Y_2) \in G(0, Q_{cvf})\) in the canonical variable form of Definition 1. Call the various components as defined in the next table.

| \(Y_{11}\) | \(Y_{21}\) | identical information of \(Y_1\) and \(Y_2\) |
| \(Y_{12}\) | correlated information of \(Y_1\) with respect to \(Y_2\) |
| \(Y_{13}\) | private information of \(Y_1\) with respect to \(Y_2\) |
| \(Y_{21}\) | \(Y_{11}\) | identical information of \(Y_1\) and \(Y_2\) |
| \(Y_{22}\) | correlated information of \(Y_2\) with respect to \(Y_1\) |
| \(Y_{23}\) | private information of \(Y_2\) with respect to \(Y_1\) |

For \(Y_{11} = Y_{21} - a.s.\) the term identical information is used.

Theorem 7 is a formula of mutual information \(I(Y_1; Y_2)\) for a general tuple of finite-dimensional Gaussian random variables \((Y_1, Y_2) \in G(0, Q_{Y_1,Y_2})\). This formula is the subject of much discussion in Gelfand and Yaglom [19] (see Equation (2.8’) and Chapter II).

Theorem 7. Consider a tuple of finite-dimensional Gaussian random variables \((Y_1, Y_2) \in G(0, Q_{Y_1,Y_2})\) with \(Q_{Y_1} > 0, i = 1, 2\).
Compute the canonical variable form of the tuple of Gaussian random variables according to Algorithm A1. This yields the indices \(p_{11} = p_{21}, p_{12} = p_{22}, p_{13} = p_{23}\), and \(n = p_{11} + p_{12} = p_{21} + p_{22}\) and the diagonal matrix \(D\) with canonical correlation coefficients or singular values \(d_i \in (0, 1)\) for \(i = 1, \ldots, n\).

Then, mutual information \(I(Y_1; Y_2)\) is computed according to the formula

\[
I(Y_1; Y_2) = \begin{cases} 
0, & \text{if } 0 = p_{11} = p_{12} = p_{21} = p_{22}, \ p_{13} > 0, \ p_{23} > 0, \\
-\frac{1}{2} \sum_{i=1}^{n} \log(1 - d_i^2), & \text{if } 0 = p_{11} = p_{21}, \ p_{12} = p_{22} > 0, \ p_{13} > 0, \ p_{23} > 0, \\
-\infty, & \text{if } p_{11} = p_{21} > 0, \ p_{12} = p_{22} > 0, \ p_{13} > 0, \ p_{23} > 0.
\end{cases}
\]  

(56)

where \(d_i\) are the canonical correlation coefficients, i.e.,

\[
d_i = d_i(Y_{12,i}Y_{22,i}) = \frac{\mathbf{E}[Y_{12,i}Y_{22,i}]^2}{\sqrt{\mathbf{E}[Y_{12,i}]^2 \mathbf{E}[Y_{22,i}]^2}} = \mathbf{E}[Y_{12,i}Y_{22,i}], \ i = 1, \ldots, n.
\]  

(57)

Proof. The derivation given in Appendix A.3 of [21] (since it is not given in [19]).

By the last entry of (56), it is appropriate to consider to only \((Y_1, Y_2) \in G(0, Q_{Y_1,Y_2})\) such that \(p_{11} = p_{21} = 0\), i.e., by removing the identical components prior to the analysis of mutual information problems.

Remark 3. The material discussed in Section 1.2 makes use of the concepts of this section. The main point to be made is that, in lossy source coding problems, the source distribution is fixed, while the optimal reproduction distribution needs to be found and realized. Then, a pre-encoder can be used by invoking Algorithm A1.

2.3. Conditional Independence of a Triple of Gaussian Random Variables

The concept of conditional independence is basic to the entire paper. The definition is provided below. The characterization of a Gaussian measure on a triple of Gaussian random variables having the conditional independence property is stated.

Definition 3. Conditional independence.
Consider a probability space \((\Omega, F, \mathbb{P})\) and three sub-\(\sigma\)-algebras \(F_1, F_2, G \subseteq F\). Call the sub-\(\sigma\)-algebras \(F_1\) and \(F_2\) conditionally independent given, or conditioned on, the sub-\(\sigma\)-algebra \(G\) if the following factorization property holds:

\[
\mathbf{E}[Y_1Y_2|G] = \mathbf{E}[Y_1|G]\mathbf{E}[Y_2|G], \ \forall \ Y_1 \in L^+(F_1), \ Y_2 \in L^+(F_2).
\]  

(58)
Denote this property by \((F_1, F_2 | G) \in \text{CI}\).

For Gaussian random variables, the definition of minimality of a Gaussian random variable \(X\) that makes two Gaussian random variables \((Y_1, Y_2)\) conditionally independent is needed. The definition is introduced below.

**Definition 4. Minimality of conditional independence of Gaussian random variables.**

Consider three random variables, \(Y_i : \Omega \to \mathbb{R}^{n_i}\) for \(i = 1, 2\) and \(X : \Omega \to \mathbb{R}^n\).

Call the random variables \(Y_1\) and \(Y_2\) Gaussian conditionally independent conditioned on or given \(F_X\) if:

1. \((F_{Y_1}, F_{Y_2} | F_X) \in \text{CI}\);
2. \((Y_1, Y_2, X)\) are jointly Gaussian random variables.

The notation \((Y_1, Y_2 | X) \in \text{CIG}\) is used to denote this property.

Call the random variables \((Y_1, Y_2 | X)\) minimally Gaussian conditionally independent if:

1. They are Gaussian conditionally independent;
2. There does not exist another tuple \((Y_1, Y_2 | X_1)\) with \(X_1 : \Omega \to \mathbb{R}^{n_1}\) such that \((Y_1, Y_2 | X_1) \in \text{CIG}\) and \(n_1 < n\).

This property is denoted by \((Y_1, Y_2 | X_1) \in \text{CIG}_{\text{min}}\).

There exists a simple equivalent condition for the conditional independence of tuple of Gaussian random variables by a third Gaussian random variable. This condition is expressed in terms of parameterizing the variance matrix of the tuple as presented in the next proposition.

**Proposition 2.** [22] (Proposition 3.4) Equivalent condition for the conditional independence of the tuple of Gaussian random variables.

Consider a triple of jointly Gaussian random variables denoted as \((Y_1, Y_2, X) \in \mathcal{G}(0, Q)\) with \(Q_X > 0\). This triple is Gaussian conditionally independent if and only if

\[
Q_{Y_1,Y_2} = Q_{Y_1,X}Q_X^{-1}Q_{X,Y_2}.
\]  

(59)

It is minimally Gaussian conditionally independent if and only if, in addition, \(n = \text{dim}(X) = \text{rank}(Q_{Y_1,Y_2})\).

It will become apparent in Section 4.4 that the Gray and Wyner lossy rate region \(\mathcal{R}_{GW}(\Delta_1, \Delta_2)\) is parameterized by a triple of jointly Gaussian random variables \((Y_1, Y_2, W)\), but not necessarily such that \(W\) makes \(Y_1\) and \(Y_2\) conditionally independent. However, subsets of \(\mathcal{R}_{GW}(\Delta_1, \Delta_2)\), are characterized by a triple \((Y_1, Y_2, W)\), such that \(W\) makes \(Y_1\) and \(Y_2\) conditionally independent.

### 2.4. Weak Realization of a Gaussian Probability Measure on a Tuple of Random Variables

This section is motivated by Theorem 9, which states that, among all joint distributions \(P_{Y_1, Y_2, W}\) induced by a tuple of multivariate correlated Gaussian random variables \((Y_1, Y_2)\), and an arbitrary random variable \(W : \Omega \to \mathbb{W}\), continuous or discrete-valued, Wyner’s common information \(C(Y_1, Y_2)\), defined by (13), is minimized by a triple \((Y_1, Y_2, W)\) which induces a jointly Gaussian distribution \(P_{Y_1, Y_2, W}\), and \(W : \Omega \to \mathbb{W} = \mathbb{R}^n\) is finite-dimensional Gaussian random variable.

To develop the above results, use is made of the solution of the problem of the weak Gaussian stochastic realization of a tuple of Gaussian random variables. Specifically, to determine a Gaussian probability measure on a triple of Gaussian random variables such that:

1. The measure restricted to the first two Gaussian random variables is equal to the considered probability measure;
2. The third Gaussian random variable makes the other two random variables conditionally independent. This problem does not have a unique solution, there is a
set of Gaussian probability measures which meets those conditions. Needed is the parameterization of this set of solutions.

Below, the problem is stated in more detail. Its solution is provided in the next section.

**Problem 1.** Weak stochastic realization of a tuple of conditionally independent Gaussian random variables.

Weak stochastic realization problem of a Gaussian random variable. Consider a Gaussian measure $P_0 = G(0, Q_0)$ on the space $(\mathbb{R}^{p_1+p_2}, B(\mathbb{R}^{p_1+p_2}))$. Determine the integer $n \in \mathbb{N}$ and construct all Gaussian measures on the space $(\mathbb{R}^{p_1+p_2+n}, B(\mathbb{R}^{p_1+p_2+n}))$ such that, if $P_1 = G(0, Q_1)$ is such a measure with $(Y_1, Y_2, X) \in G(0, Q_1)$, then:

1. $G(0, Q_1)_{|\mathbb{R}^{p_1+p_2}} = G(0, Q_0)$;
2. $(Y_1, Y_2, X) \in \text{CIG}_{\text{min}}$.

Here, the indicated random variables $(Y_1, Y_2, X)$ are constructed having the measure $G(0, Q_1)$ with the dimensions $p_1, p_2, n \in \mathbb{Z}_+$, respectively.

The next definition and proposition are about the weak Gaussian stochastic realization of a tuple of jointly Gaussian multivariate random variables and its weak stochastic realization.

**Definition 5.** Minimality of weak stochastic realization of a tuple of conditionally independent Gaussian random variables.

Consider a Gaussian measure $P_0 = G_0(0, Q_{(Y_1, Y_2)})$ with zero mean values for a tuple $(Y_1, Y_2)$ of random variables on the product space $(\mathbb{R}^{p_1} \times \mathbb{R}^{p_2}, B(\mathbb{R}^{p_1}) \otimes B(\mathbb{R}^{p_2}))$ for $p_1, p_2 \in \mathbb{Z}_+$ with

$$Q_{(Y_1, Y_2)} = \left(\begin{array}{cc}
Q_{Y_1} & Q_{Y_1Y_2} \\
Q_{Y_1Y_2}^T & Q_{Y_2}
\end{array}\right), \quad Q_{Y_1} > 0, Q_{Y_2} > 0.$$

(a) A weak Gaussian stochastic realization of the Gaussian measure $G_0(0, Q_{(Y_1, Y_2)})$ is defined to be a Gaussian measure $P_1 = G_1$ if there exists an integer $n \in \mathbb{Z}_+$ such that the Gaussian measure $G_1$ is defined on the space $(\mathbb{R}^{p_1} \times \mathbb{R}^{p_2}, B(\mathbb{R}^{p_1}) \otimes B(\mathbb{R}^{p_2}) \otimes B(\mathbb{R}^n))$ associated with random variables in the three spaces denoted, respectively, by $Y_1$, $Y_2$, and $X$, and such that:

1. $G_1|_{\mathbb{R}^{p_1} \times \mathbb{R}^{p_2}} = G_0(0, Q_{(Y_1, Y_2)})$;
2. $Q_X > 0$;
3. Conditional independence holds: $P_{Y_1Y_2|x} = P_{Y_1|x}P_{Y_2|x}$, where these are Gaussian measures, with means which are linear functions of the random variable $X$ and deterministic variance matrices.

(b) The weak Gaussian stochastic realization is called minimal if the dimension $n$ of the random variable $X$ is the smallest possible over all weak Gaussian stochastic realizations as defined in (a).

(c) A Gaussian random variable representation of a weak Gaussian stochastic realization $G_1$ is defined as a triple of random variables satisfying the following relations

$$\begin{align*}
(Y_1, Y_2, X, V_1, V_2), & \quad \nu_{V_1}, \nu_{V_2} \in \mathbb{Z}_+, \nu_{V_1} \geq p_1, \nu_{V_2} \geq p_2, \\
Y_1 : \Omega & \to \mathbb{R}^{p_1}, Y_2 : \Omega \to \mathbb{R}^{p_2}, V_1 : \Omega \to \mathbb{R}^{p_1}, V_2 : \Omega \to \mathbb{R}^{p_2}, X : \Omega \to \mathbb{R}^n,
\end{align*}$$

(60)

$$\begin{align*}
(V_1, V_2, X) & \in G, \text{ and these are zero mean independent random variables} \\
Q_{V_1} & > 0, Q_{V_2} > 0, Q_X > 0; \\
C_1 & \in \mathbb{R}^{p_1 \times n}, C_2 \in \mathbb{R}^{p_2 \times n}, N_1 \in \mathbb{R}^{p_1 \times p_1}, N_2 \in \mathbb{R}^{p_2 \times p_2},
\end{align*}$$

(61)

$$\begin{align*}
Y_1 & = C_1 X + N_1 V_1, \\
Y_2 & = C_2 X + N_2 V_2,
\end{align*}$$

(62)

$$\begin{align*}
Q_{Y_1} & = C_1 Q_X C_1^T + N_1 Q_{V_1} N_1^T, \\
Q_{Y_2} & = C_2 Q_X C_2^T + N_2 Q_{V_2} N_2^T,
\end{align*}$$

(63)

$$\begin{align*}
Q_{Y_1Y_2} & = C_1 Q_X C_2^T, \\
G_0(0, Q_{(Y_1, Y_2)}) & = G_1|_{\mathbb{R}^{p_1} \times \mathbb{R}^{p_2}}.
\end{align*}$$

(64)
From the assumptions, it then follows that \((Y_1, Y_2)\) are Gaussian random variables, hence the last equality makes sense.

(d) A minimal Gaussian random variable representation of a weak Gaussian stochastic realization is defined as a triple of random variables as in (c) except that, in addition, it is required that,

\[
\text{rank}(C_1) = n = \text{rank}(C_2).
\]

The case \(Q_X \geq 0\) in (a),(2) is similar.

The next proposition shows the equivalence of weak Gaussian stochastic realizations of Definition 5. (a), (b) to Definition 5. (c), (d), respectively.

**Proposition 3.** Consider the setting of Definition 5 with \((Y_1, Y_2) \in G(0, Q_{(Y_1, Y_2)})\) with the representation of (62) and (63).

(a) A weak Gaussian stochastic realization in terms of a measure \(P_1 = G_1\) as defined in Definition 5. (a) is equivalent to a Gaussian random variable representation of Definition 5. (c).

(b) The minimal weak Gaussian stochastic realization of Definition 5. (b) is equivalent to a minimal weak Gaussian random variable representation of Definition 5. (d).

**Proof.** The derivation given in Appendix A.5 of [21].

Consider Figure 2. The two signals \(Y_1, Y_2\) are to be reproduced at the two decoders by \(\hat{Y}_1, \hat{Y}_2\) subject to the square-error distortion functions. According to Gray and Wyner, the characterization of the lossy rate region is described by a single coding scheme that uses the auxiliary random variable \(W\), which is common to both \(Y_1, Y_2\). A subset of the rate triples on the Gray and Wyner rate region, which is achieved by a triple that satisfies \((F_{Y_1}, F_{Y_2}|FW) \in \text{CIG}\). Below, this conditional independence is further detailed in terms of the mathematical framework of weak stochastic realization such that \((F_{Y_1}, F_{Y_2}|FW) \in \text{CIG}\).

**Definition 6.** The model for a triple of Gaussian random variables.

Consider a tuple of Gaussian random variables specified by \(Y = (Y_1, Y_2) \in G(0, Q_{(Y_1, Y_2)})\) with \(Y_i : \Omega \rightarrow \mathbb{R}^p\) for \(i = 1, 2\). Take a jointly Gaussian measure \(G(0, Q_{(Y_1, Y_2)}, W)\) for the triple \((Y_1, Y_2, W), W : \Omega \rightarrow \mathbb{R}^q, W \in G(0, Q_W)\), such that the marginal measure on \((Y_1, Y_2)\) is equal to the considered measure, with

\[
Q_{(Y_1, Y_2, W)} = \begin{pmatrix}
Q_{Y_1} & Q_{Y_1, Y_2} & Q_{Y_1, W} \\
Q_{Y_1, Y_2}^T & Q_{Y_2} & Q_{Y_2, W} \\
Q_{Y_1, W}^T & Q_{Y_2, W} & Q_W
\end{pmatrix}.
\]
Denote the parameterized joint measure with respect to $W$, by $(Y_1, Y_2, W) \in G(0, Q_{(Y_1,Y_2,W)})$. This parameterized joint measure $(Y_1, Y_2, W) \in G(0, Q_{(Y_1,Y_2,W)})$ also includes the subset such that the conditional independence holds, $(F^{Y_1}, F^{Y_2}|F^W) \in \text{CIG}.$

In the following subsections, it will be shown how such a random variable $W$ can be constructed in a number of cases.

Algorithm 1. (a) gives the general case, while (b) gives the special case when the joint measure by $(Y_1, Y_2, W) \in G(0, Q_{(Y_1,Y_2,W)})$ such that $(F^{Y_1}, F^{Y_2}|F^W) \in \text{CIG}$ via weak stochastic realization.

**Algorithm 1.** Consider the model of a tuple of Gaussian random variables of Definition 6.

(a) General case.

1. At the encoder, first compute the variables,

$$
\left( \begin{array}{c}
Z_1 \\
Z_2
\end{array} \right) = \left( \begin{array}{c}
Y_1 \\
Y_2
\end{array} \right) - \mathbb{E}[\left( \begin{array}{c}
Y_1 \\
Y_2
\end{array} \right) | W] = \left( \begin{array}{c}
Y_1 \\
Y_2
\end{array} \right) - Q_{(Y_1,Y_2),W} Q_W^{-1} W
$$

$$
Q_{(Y_1,Y_2),W} = \mathbb{E}[\left( \begin{array}{c}
Y_1 \\
Y_2
\end{array} \right) W^T] = \left( \begin{array}{c}
Q_{Y_1,W} \\
Q_{Y_2,W}
\end{array} \right);
$$

then, the triple $(Z_1, Z_2, W)$ of jointly Gaussian random variables are such that $(Z_1, Z_2) \in G(0, Q_{(Z_1,Z_2)})$ and $(Z_1, Z_2)$ independent of $W$.

2. The tuple of random variables $(Y_1, Y_2)$ are represented according to

$$
\left( \begin{array}{c}
Y_1 \\
Y_2
\end{array} \right) = Q_{(Y_1,Y_2),W} Q_W^{-1} W + \left( \begin{array}{c}
Z_1 \\
Z_2
\end{array} \right)
$$

(b) Special case. Consider $(F^{Y_1}, F^{Y_2}|F^W) \in \text{CIG}$, and assume $Q_W > 0$.

1. At the encoder, compute first the variables,

$$
Z_1 = Y_1 - \mathbb{E}[Y_1 | F^W] = Y_1 - Q_{Y_1,W} Q_W^{-1} W,
$$

$$
Z_2 = Y_2 - \mathbb{E}[Y_2 | F^W] = Y_2 - Q_{Y_2,W} Q_W^{-1} W;
$$

then the triple $(Z_1, Z_2, W)$ of jointly Gaussian random variables are independent.

2. The tuple of random variables $(Y_1, Y_2)$ are represented according to,

$$
Y_1 = Q_{Y_1,W} Q_W^{-1} W + Z_1, \quad Y_2 = Q_{Y_2,W} Q_W^{-1} W + Z_2.
$$

We emphasize that $Y_1$ and $Y_2$ are conditionally independent condition on $W$ if and only if $Z_1$ and $Z_2$ are independent.

The validity of the statements of the algorithm follow from the next proposition.

**Proposition 4.** Consider the model of a tuple of Gaussian random variables of Definition 6, for cases (a), (b).

(a) At the encoder, the conditional expectations are correct and the definitions of $Z_1$ and of $Z_2$ are well defined.

(b) The three random variables $(Z_1, Z_2, W)$ are independent. Consequently, the three sequences $(W^N, Z_1^N, Z_2^N)$, and messages generated by the Gray–Wyner encoder,

$$
f^E(Y_1^N, Y_2^N) = f^E(W^N, Z_1^N, Z_2^N) = (S_0, S_1, S_2)$$

are independent.

**Proof.** Case (a). This follows from realization theory (since no constraints are imposed).

Case (b). This is a specific application of Proposition 3. □
For the definition of $C(Y_1, Y_2)$, use is made of the construction of the actual family of measures such that $(Y_1, Y_2|W) \in \mathcal{CIG}$ holds, and the weak stochastic realization. These are presented in Theorem 8 and Corollary 1.

2.5. Characterization of Minimal Conditional Independence of a Triple of Gaussian Random Variables

Introduce the notation of the parameterization of the family of Gaussian probability distributions

$$\mathcal{P}^{\mathcal{CIG}} = \{ \mathbf{P}_{Y_1, Y_2, W}(y_1, y_2, w) \mid \mathbf{P}_{Y_1, Y_2|W}(y_1, y_2|w) = \mathbf{P}_{Y_1|W}(y_1|w)\mathbf{P}_{Y_2|W}(y_2|w) \},$$

$$\mathbf{P}_{Y_1, Y_2, W}(y_1, y_2, w) = \mathbf{P}_{Y_1, Y_2}(y_1, y_2)\delta(w), \quad (Y_1, Y_2, W) \text{ is jointly Gaussian} \}.$$  (75)

A subset of the set $\mathcal{P}^{\mathcal{CIG}}$ is the set of distributions $\mathcal{P}^{\mathcal{CIG}}_{\text{min}}$, with the additional constraint that the dimension of the random variable $W$ is minimal while all other conditions hold, defined by

$$\mathcal{P}^{\mathcal{CIG}}_{\text{min}} = \{ \mathbf{P}_{Y_1, Y_2, W}(y_1, y_2, w) \in \mathcal{P}^{\mathcal{CIG}} \mid (Y_1, Y_2) \in \mathcal{CIG}_{\text{min}} \} \subseteq \mathcal{P}^{\mathcal{CIG}}.$$  (76)

The parameterization of the family of Gaussian probability distributions $\mathcal{P}^{\mathcal{CIG}}$ and $\mathcal{P}^{\mathcal{CIG}}_{\text{min}}$ require the solution of the weak stochastic realization problem of Gaussian random variables defined by Problem 1. This problem is solved in [22] (Theorem 4.2). For the readers’ convenience, it is stated below.

**Theorem 8.** Ref. [22] (Theorem 4.2) Consider a tuple $(Y_1, Y_2)$ of Gaussian random variables in the canonical variable form of Definition 1. Restrict attention to the correlated parts of these random variables. Thus, the random variables $Y_1, Y_2$ have the same dimension $n = p_1 = p_2$, and their covariance matrix $D \in \mathbb{R}^{n \times n}$ is a nonsingular diagonal matrix with the diagonal ordered real-numbers in the interval $(0, 1)$. Hence,

$$(Y_1, Y_2) \in \mathcal{G}(0, \mathcal{Q}(Y_1, Y_2)) = \mathbf{P}_0, \quad Y_1, Y_2 : \Omega \to \mathbb{R}^n, \quad n \in \mathbb{Z}_+,$$  (77)

$$\mathcal{Q}(Y_1, Y_2) = \left( \begin{array}{cc} I & D \\ D & I \end{array} \right),$$  (78)

$$D = \text{Diag}(d_1, d_2, \ldots, d_n) \in \mathbb{R}^{n \times n}, \quad 1 > d_1 \geq d_2 \geq \ldots \geq d_n > 0.$$  (79)

That is, $p_{11} = p_{21} = 0, p_{13} = p_{23} = 0$.

(a) There exists a probability measure $\mathbf{P}_1$, and a triple of Gaussian random variables $Y_1, Y_2, W : \Omega \to \mathbb{R}^n$ defined on it, such that (i) $\mathbf{P}_1 \| (Y_1, Y_2) = \mathbf{P}_0$ and (ii) $(Y_1, Y_2, W) \in \mathcal{CIG}_{\text{min}}$.

(b) There exist a family of Gaussian measures denoted by $\mathbf{P}_{ci} \subseteq \mathcal{P}^{\mathcal{CIG}}_{\text{min}}$ that satisfy (i) and (ii) of (a), and moreover, this family is parameterized by the matrices and sets, as follows.

$$\mathcal{G}(0, \mathcal{Q}_s(\mathcal{Q}_W)), \quad \mathcal{Q}_W \in \mathcal{Q}_W.$$  (80)

$$\mathcal{Q}_s = \mathcal{Q}_s(\mathcal{Q}_W) = \left( \begin{array}{ccc} I & D \\ D & I \end{array} \right),$$  (81)

$$\mathcal{Q}_W = \left\{ \mathcal{Q}_W \in \mathbb{R}^{n \times n} \mid \mathcal{Q}_W = \mathcal{Q}_W^T, 0 < D \leq \mathcal{Q}_W \leq D^{-1} \right\},$$  (82)

$$\mathbf{P}_{ci} = \left\{ \mathcal{G}(0, \mathcal{Q}_s(\mathcal{Q}_W)) \text{ on } (\mathbb{R}^{3n}, B(\mathbb{R}^{3n})) \mid \mathcal{Q}_W \in \mathcal{Q}_W \right\} \subseteq \mathcal{P}^{\mathcal{CIG}}_{\text{min}}.$$  (83)

Furthermore, for any measure $\mathbf{P}_1 \in \mathcal{P}^{\mathcal{CIG}}_{\text{min}}$, there exists a triple of state transformation of the form $(Y_1, Y_2, W) \to (S_1 Y_1, S_2 Y_2, S_W W)$ for nonsingular square matrices $S_1$, $S_2$, $S_W$ such that the corresponding measure of the three transformed variables belongs to $\mathbf{P}_{ci}$.
The application of Theorem 8 is discussed in the next remark, in the context of parameterizing any rate-triple on the Gray–Wyner lossy rate region \((R_0, R_1, R_2) \in \mathcal{R}_{GW}(\Delta_1, \Delta_2)\) that lies on the Pangloss plane.

**Remark 4. Applications of Theorem 8.**

(a) Theorem 8 is a parameterization of the family of Gaussian measures \(P_c \subseteq \mathcal{P}_{Gi} \) by the entries of the covariance matrix \(Q_W\). Hence, it is at most an \((n+1)/2\)-dimensional parameterization;

(b) It is shown in Section 4.4 that only a subset of the achievable rate region \(\mathcal{R}_{GW}(\Delta_1, \Delta_2) = \mathcal{R}^*_GW(\Delta_1, \Delta_2)\) is generated from distributions \(P_c \subseteq \mathcal{P}_{Gi} \subseteq \mathcal{P}\).

The next corollary is useful to the calculation of \(C(Y_1, Y_2)\), since by Theorem 9, an achievable lower bound on \(I(Y_1, Y_2; W)\) is incurred by a Gaussian random variable \(W\), such that the distribution \(P_{Y_1,Y_2,W} \in P_c \subseteq \mathcal{P}_{Gi}\), corresponding to \(W \in G(0, Q_W)\). By Theorem 9, and since \(C(Y_1, Y_2)\) is invariant with respect to nonsingular transformations applied to \((Y_1, Y_2, W)\), the next corollary gives the realization of \((Y_1, Y_2)\) as defined in Theorem 8, by (77)–(79), expressed in terms of an arbitrary Gaussian random variable \(W \in G(0, Q_W)\).

**Corollary 1.** Consider a tuple \((Y_1, Y_2)\) of Gaussian random variables in the canonical variable form of Definition 1. Restrict the attention to the correlated parts of these random variables, as defined in Theorem 8, by (77)–(79).

Then, a realization of the random variables \((Y_1, Y_2)\) which induce the family of measures \(P_c \subseteq \mathcal{P}_{Gi}\), defined by (80)–(83), is

\[
Y_1 = Q_{Y_1,W}Q_W^{-1}W + Z_1 \tag{84}
\]
\[
Q_{Y_1,W} = D^{1/2}, \quad Z_1 \in G(0, (I - D^{1/2}Q_W^{-1}D^{1/2})), \tag{85}
\]
\[
Y_2 = Q_{Y_2,W}Q_W^{-1}W + Z_2 \tag{86}
\]
\[
Q_{Y_2,W} = D^{1/2}Q_W, \quad Z_2 \in G(0, (I - D^{1/2}Q_WD^{1/2})), \tag{87}
\]

\((Z_1, Z_2, W)\), are independent. \tag{88}

Furthermore, the mutual information \(I(Y_1, Y_2; W)\) is given by

\[
I(Y_1, Y_2; W) = H(Y_1, Y_2) - H(Y_1|W) - H(Y_2|W) \tag{89}
\]
\[
= \frac{1}{2} \sum_{i=1}^{n} \ln(1 - d_i^2) - \frac{1}{2} \ln(\det((I - D^{1/2}Q_W^{-1}D^{1/2})(I - D^{1/2}Q_WD^{1/2})))) \tag{90}
\]

and it is parameterized by \(Q_W \in Q_W\), where \(Q_W\) is defined by the set of Equation (82).

**Proof.** The correctness of the realization is due to Proposition 2 and Theorem 8. The calculation of mutual information follows from the realization. \(\square\)

3. Wyner’s Common Information

This section is devoted to the calculation of Wyner’s common information \(C(Y_1, Y_2)\), defined by (13), for \(P_{Y_1,Y_2} = G(0, Q(Y_1,Y_2))\), and the construction of the weak stochastic realization of \((Y_1, Y_2, W)\) that achieves this.

3.1. Reduction of the Calculation of Wyner’s Common Information

First, we show Theorem 9, which states: given a tuple of multivariate correlated Gaussian random variables \((Y_1, Y_2)\), and an arbitrary random variable \(W\) (i.e., taking continuous or discrete values), Wyner’s common information \(C(Y_1, Y_2)\), defined by (13), is minimized by a triple \((Y_1, Y_2, W)\) which induces a jointly Gaussian distribution \(P_{Y_1,Y_2,W}\), and \(W : \Omega \rightarrow \mathbb{W} = \mathbb{R}^n\) is finite-dimensional Gaussian random variable.
Theorem 9. Consider a tuple of multivariate-correlated Gaussian random variables $Y_1 : \Omega \to \mathbb{R}^{p_1}, Y_2 : \Omega \to \mathbb{R}^{p_2}, p_i \in \mathbb{Z}_+, i = 1, 2$ with the variance matrix of this tuple denoted by

$$ (Y_1, Y_2) \in G(0, Q_{(Y_1,Y_2)}), \quad Q_{(Y_1,Y_2)} = \begin{pmatrix} Q_{Y_1} & Q_{Y_1,Y_2} \\ Q_{Y_1,Y_2} & Q_{Y_2} \end{pmatrix} \in \mathbb{R}^{(p_1+p_2) \times (p_1+p_2)} \quad (91) $$

and, without loss of generality, assume that $Q_{(Y_1,Y_2)}$ is a positive definite matrix. Let $W : \Omega \to \mathbb{W}$ be any auxiliary random variable, with $\mathbb{W}$ being an arbitrary measurable space, and $P_{Y_1,Y_2,W}$ any joint probability distribution of the triple $(Y_1, Y_2, W)$ on the product space $(\mathbb{R}^{p_1} \times \mathbb{R}^{p_2} \times \mathbb{W}, B(\mathbb{R}^{p_1}) \otimes B(\mathbb{R}^{p_2}) \otimes B(\mathbb{W}))$ with $(Y_1, Y_2)$-marginal $P_{Y_1,Y_2}$ the Gaussian distribution $P_{Y_1,Y_2} = G(0, Q_{(Y_1,Y_2)})$. The following hold.

(a) Define the random variables $Z_1, Z_2$ by

$$ Z_i = Y_i - \mathbb{E}[Y_i|W], \quad Z_i : \Omega \to \mathbb{R}^{p_i}, \; i = 1, 2. \tag{92} $$

The inequalities hold:

$$ I(Y_1, Y_2; W) = H(Y_1, Y_2) - H(Y_1, Y_2|W) \tag{93} $$

$$ = H(Y_1, Y_2) - H(Y_1|Y_2, W) - H(Y_2|W) \tag{94} $$

$$ \geq H(Y_1, Y_2) - H(Y_1|W) - H(Y_2|W) \tag{95} $$

$$ = H(Y_1, Y_2) - H(Y_1 - \mathbb{E}[Y_1|W]|W) - H(Y_2 - \mathbb{E}[Y_2|W]|W) \tag{96} $$

$$ \geq H(Y_1, Y_2) - H(Y_1 - \mathbb{E}[Y_1|W]) - H(Y_2 - \mathbb{E}[Y_2|W]) \tag{97} $$

$$ = H(Y_1, Y_2) - H(Z_1) - H(Z_2) \tag{98} $$

$$ \geq H(Y_1, Y_2) - H(Z_1) - H(Z_2) \text{ if } Z_1, Z_2 \text{ have finite variances, and Gaussian.} \tag{99} $$

(b) If:

(i) $W : \Omega \to \mathbb{W} = \mathbb{R}^n$ is an $n$-dimensional, $n \in \mathbb{Z}_+$, Gaussian random variable;

(ii) $(Z_1, Z_2, W)$ are mutually independent jointly Gaussian random variables, then all inequalities in (93)–(99) hold with equality, and $(Y_1, Y_2, W)$ induces a family of jointly probability distributions $P_{Y_1,Y_2,W}$ with $(Y_1, Y_2)$-marginal $P_{Y_1,Y_2}$, such that $W$ makes $Y_1$ and $Y_2$ conditionally independent, that is $P_{Y_1,Y_2|W} = P_{Y_1|W}P_{Y_2|W}$;

(c) Among all joint distributions, $P_{Y_1,Y_2,W}$ induced by the jointly Gaussian random variables $(Y_1, Y_2) \in G(0, Q_{(Y_1,Y_2)})$ of (91), and an arbitrary random variable $W : \Omega \to \mathbb{W}$, such that $(Y_1, Y_2)$-marginal $P_{Y_1,Y_2}$ is the Gaussian distribution $P_{Y_1,Y_2} = G(0, Q_{(Y_1,Y_2)})$, and $P_{Y_1,Y_2|W} = P_{Y_1|W}P_{Y_2|W}$, a jointly Gaussian distribution achieves the lower bounds of $I(Y_1, Y_2; W)$ in part (a), i.e., achieves Wyner’s common information $C(Y_1, Y_2)$, defined by (13), and $P_{Y_1,Y_2,W}$ is induced by an $n$-dimensional, $n \in \mathbb{Z}_+$, Gaussian random variable $W : \Omega \to \mathbb{W} = \mathbb{R}^n$, and $(Y_1, Y_2) \in G(0, Q_{(Y_1,Y_2)})$, such a distribution is induced by the triple $(Y_1, Y_2, W)$ represented by

$$ Y_1 = \mathbb{E}[Y_1|W] + Z_1, \quad Y_2 = \mathbb{E}[Y_2|W] + Z_2, \tag{100} $$

$$ (W, Z_1, Z_2) \text{ are mutually independent, Gaussian random variables.} \tag{101} $$

Proof. (a) (93) is due to an identity of mutual information, (94) is due to the chain rule of entropy, (95) due to conditioning reduces entropy, (96) due to a property of conditional entropy, (97) due to conditioning reduces entropy, (98) is due to definition (92) and (99), is due to maximum entropy principle. (b) Since, $Y_i = \mathbb{E}[Y_i|W] + Z_i, i = 1, 2$ and $(Y_1, Y_2) \in G(0, Q_{(Y_1,Y_2)})$, if (i) and (ii) hold, then all inequalities hold with equality, and the statements are easily verified. (c) Follows from part (b). \hfill \Box

Remark 5. Theorem 9 shows that, among all random variables $W$ which induce a joint distribution $P_{Y_1,Y_2,W}$ with $(Y_1, Y_2)$-marginal $P_{Y_1,Y_2}$ the Gaussian distribution $P_{Y_1,Y_2} = G(0, Q_{(Y_1,Y_2)})$, then for the Wyner’s common information $C(Y_1, Y_2)$ problem, it suffices to consider a jointly Gaussian triple $(Y_1, Y_2, W)$ such that $W$ makes $Y_1$ and $Y_2$ conditionally independent.
3.2. Wyner’s Common Information of Correlated Random Variables

Assume that the tuple of multivariate correlated Gaussian random variables $(Y_1, Y_2) \in G(0, Q_{Y_1, Y_2})$ of Theorem 9 is already transformed to the canonical variable representation, see Definition 1, using Algorithm A1, i.e., by the nonsingular transformation, $S = \text{Block-diag}(S_1, S_2)$. Mutual information is invariant with respect to nonsingular transformations, and $I(Y_1, Y_2; W) = I(S_1Y_1, S_2Y_2; W)$.

By Theorem 9, (c), the joint probability distributions $P_{Y_1, Y_2|W}(y_1, y_2, w)$ are jointly Gaussian, and parameterized by the random variable $W$. This family of distributions is parameterized by the multidimensional random variable $W$, such that $(Y_1, Y_2)$ are conditionally independent, conditioned on $W$, the marginal distribution $P_{Y_1, Y_2|W}(y_1, y_2, \infty) = P_{Y_1, Y_2}(y_1, y_2)$ coincides with the distribution of $(Y_1, Y_2)$, and to represent $(Y_1, Y_2)$.

Using the above construction, one obtains the next theorem.

**Theorem 10.** Consider a tuple of Gaussian random variables $(Y_1, Y_2) \in G(0, Q_{\text{cvf}})$ as described and decomposed according to Algorithm A1. Restrict attention to the correlated parts of these random variables, as described in Theorem 8, (77)–(79) (i.e., only components $(Y_{12}, Y_{22})$ are present).

(a) Theorem 8 holds, and in particular, the family of jointly Gaussian distributions $P_{Y_1, Y_2, W}$ induced by $(Y_1, Y_2) \in G(0, Q_{\text{cvf}})$ and a Gaussian random variable $W : \Omega \to \mathbb{R}^n$, with minimum dimension $n$, such that the $(Y_1, Y_2)$–marginal is $P_{Y_1, Y_2} = G(0, G(0, Q_{\text{cvf}}))$, and $P_{Y_1, Y_2|W} = P_{Y_1|W} P_{Y_2|W}$ is parameterized by the family of Theorem 8. (b), i.e., (80)–(83); (b) Corollary 1, (84)–(88) characterizes the family of realizations of $(Y_1, Y_2, W)$, parameterized by $W$, which induce jointly Gaussian distributions, such that $W : \Omega \to \mathbb{R}^n$ is a Gaussian random variable with minimum dimension $n$, $P_{Y_1, Y_2|W} = P_{Y_1|W} P_{Y_2|W}$, and $(Y_1, Y_2) \in G(0, Q_{\text{cvf}}))$. Moreover, Wyner’s common information $C(Y_1, Y_2)$ is computed from the expression $I(Y_1, Y_2; W)$ of Corollary 1, (90), optimized over $Q_W \in Q_W$, where $Q_W$ is defined by the set of Equation (82).

**Remark 6.** It is apparent that the proof of the formula $C(Y_1, Y_2; W)$ in [15,16] is based on rate distortion function, i.e., they do not directly address Wyner’s optimization problem (13), as in Theorem 9, which first shows, among all continuous or discrete random variables, $W$ is Gaussian, and there is no parameterization of the set of distributions $P_{Y_1, Y_2, W}$ achieving conditional independence $P_{Y_1, Y_2|W} = P_{Y_1|W} P_{Y_2|W}$, i.e., the optimization over the parameterized family of Gaussian measures of Theorem 8 is not given.

In the next theorem, the family of measures $P_{\text{ci}} \subseteq P^\text{CIG}_{\text{min}}$, defined by (80)–(83), which leads to realization of $(Y_1, Y_2)$, given in Corollary 1, is ordered for the determination of a single joint distribution $P_{Y_1, Y_2, W^*} \in P_{\text{ci}} \subseteq P^\text{CIG}_{\text{min}}$, which achieves $C(Y_1, Y_2)$. This leads to the realization of $(Y_1, Y_2)$ expressed in terms of $W^*$ and vectors of independent Gaussian random variables $(Z_1, Z_2)$, one for each realization, each having independent components.

**Theorem 11.** Consider a tuple $(Y_1, Y_2)$ of Gaussian random variables in the canonical variable form of Definition 1. Restrict attention to the correlated parts of these random variables, as defined in Theorem 8, defined by (77)–(79).

The following hold.

(a) The information quantity $C(Y_1, Y_2)$ is given by

$$C(Y_1, Y_2) = \frac{1}{2} \sum_{i=1}^{n} \ln \left( \frac{1 + d_i}{1 - d_i} \right) = \frac{1}{2} \sum_{i=1}^{n} \ln \left( 1 + \frac{2d_i}{1 - d_i} \right) \in (0, \infty). \quad (102)$$
(b) The realizations of the random variables \((Y_1, Y_2, W^*)\) that achieve \(C(Y_1, Y_2)\) are represented by
\[
V : \Omega \to \mathbb{R}^n, \quad V \in G(0, I), \quad \text{the vector } V \text{ has independent components},
\]
\[
F^V, F^{Y_1} \vee F^{Y_2}, \quad \text{are independent } \sigma\text{-algebras},
\]
\[
L_1 = L_2 = D^{1/2}(I + D)^{-1} \in \mathbb{R}^{n \times n}, \quad (103)
\]
\[
L_3 = (I - D)^{1/2}(I + D)^{-1/2} \in \mathbb{R}^{n \times n}, \quad L_1, L_2, L_3, \quad \text{are diagonal matrices}, \quad (104)
\]
\[
W^* = L_1Y_1 + L_2Y_2 + L_3V, \quad W^* : \Omega \to \mathbb{R}^n, \quad (105)
\]
\[
Z_1 = Y_1 - D^{1/2}W^*, \quad Z_1 : \Omega \to \mathbb{R}^n, \quad (106)
\]
\[
Z_2 = Y_2 - D^{1/2}W^*, \quad Z_2 : \Omega \to \mathbb{R}^n. \quad (107)
\]

Then:
\[
Z_1 \in G(0, (I - D)), \quad Z_2 \in G(0, (I - D)), \quad W^* \in G(0, I); \quad (108)
\]
\[
(Z_1, Z_2, W^*), \quad \text{are independent and (109)}
\]
\[
Y_1 = D^{1/2}W^* + Z_1, \quad Y_2 = D^{1/2}W^* + Z_2 \quad (110)
\]

hence, the variables \((Y_1, Y_2, W^*)\) induce a distribution \(P_{Y_1, Y_2, W^*} \in \mathcal{P}_{\text{CIG}} \subseteq \mathcal{P}_{\text{CIG min}}\). Note that, in addition, each of the random variables \(Z_1, Z_2,\) and \(W^*\) have independent components.

(c) The variables \((Y_1, Y_2, W^*)\) defined in (b) induce a distribution \(P_{Y_1, Y_2, W^*} \in \mathcal{P}_{\text{CIG min}} \subseteq \mathcal{P}_{\text{CIG}}\) which achieves \(C(Y_1, Y_2)\),
\[
C(Y_1, Y_2) = I(Y_1, Y_2; W^*). \quad (111)
\]

**Proof.** By Theorem 9, the random variables \((Y_1, Y_2, W)\) are restricted to jointly Gaussian random variables. Since mutual information \(I(Y_1, Y_2; W)\) is invariant with respect to nonsingular transformations \(S_1, S_2\), i.e., \(I(Y_1, Y_2; W) = I(S_1Y_1, SY_2; W)\), and \((F^{Y_1}, F^{Y_2}|F^W) \in \text{CIG}\) is equivalent to \((F^{Y_1}, F^{Y_2}|F^W) \in \text{CIG}\), then it suffices to consider the canonical variable form of Definition 1, and to construct a measure that carries a triple of jointly Gaussian random variables \(Y_1, Y_2, W : \Omega \to \mathbb{R}^n\) such that \((F^{Y_1}, F^{Y_2}|F^W) \in \text{CIG}\).

(a) (1) Take a probability measure \(P_1\) such that there exists a triple of Gaussian random variables \(Y_1, Y_2, W : \Omega \to \mathbb{R}^n\) with \(P_1|(y_1, y_2)\) = \(P_0\) and \((F^{Y_1}, F^{Y_2}|F^W) \in \text{CIG}\). It will first be proven that attention can be restricted to those state random variables \(W\) of which the dimension equals \(n = p_{12} = p_{22}\).

Suppose that there exists a state random variable \(W : \Omega \to \mathbb{R}^{n_1}\) such that \((F^{Y_1}, F^{Y_2}|F^W) \in \text{CIG}\) and \(n_1 > n\). Hence, \(W\) does not make \((Y_1, Y_2)\) minimally conditionally independent. Construct a minimal vector which makes the tuple minimally conditionally independent according to the procedure of [22] (Proposition 3.5). Thus,
\[
W_1 = E[Y_1|F^W] = L_{11}W, \quad L_{11} \in \mathbb{R}^{n \times n_1},
\]
\[
W_2 = E[Y_2|F^W] = L_{12}W_1, \quad L_{12} \in \mathbb{R}^{n \times n_1}.
\]

Then, \((F^{Y_1}, F^{Y_2}|F^{W_2}) \in \text{CIG}_{\text{min}}\) and the dimension of \(W_2\) is \(n = p_{12} = p_{22}\). Determine a linear transformation of \(W_2\) by a matrix \(L_{15} \in \mathbb{R}^{n \times n}\) such that
\[
W_3 = L_{15}W_2 = L_{15}L_{12}L_{11}W = L_{13}W, \quad L_{13} = L_{15}L_{12}L_{11}, \quad W_3 \in G(0, Q_3), \quad Q_3 = I_n = L_{13}Q_WL_{13}^T.
\]

It is then possible to construct a matrix \(L_{14} \in \mathbb{R}^{(n_1 \times n_1) \times n_1}\) such that
\[
W_4 = L_{14}W, \quad W_4 \in G(0, Q_4), \quad Q_4 = I, \quad L_{14}Q_WL_{14}^T = 0; \quad \left(\begin{array}{c}
W_3 \\
W_4
\end{array}\right) \in G(0, I_{n_1}), \quad \text{rank}\left(\begin{array}{c}
L_{13} \\
L_{14}
\end{array}\right) = n_1,
\]
and, due to $L_{14}Q_W L_{13}^T = 0$, $W_3$, $W_4$ are independent random variables. See [23] [Theorem 4.9] for a theorem with which the existence of $L_4$ can be proven. Note further that $F^W = F_{W_3}^W$. Hence, the random variables $W_3$, $W_4$ are independent, $(F_{Y_1}, F_{Y_2} | F_W) \in \mathcal{C}_{\min}$, and $I(Y_1, Y_2; W) = I(Y_1, Y_2; W_3, W_4)$.

By properties of mutual information, it now follows that

$$I(Y_1, Y_2; W_3, W_4) - I(Y_1, Y_2; W_3)$$

$$= H(Y_1, Y_2) + H(W_3, W_4) - H(Y_1, Y_2, W_3, W_4) - H(Y_1, Y_2) - H(W_3) + H(Y_1, Y_2, W_3)$$

$$= H(Y_1, Y_2, W_3) + H(W_4) - H(Y_1, Y_2, W_3, W_4),$$

by independence of $W_3$ and $W_4$;

$$= I(Y_1, Y_2, W_3; W_4) \geq 0.$$  

Thus, for the computation of $C(Y_1, Y_2)$, attention can be restricted to those state variables $W$ which are of minimal dimension.

(2) Take a probability measure $P_1$ such that there exists a triple of Gaussian random variables $Y_1, Y_2, W : \Omega \rightarrow \mathbb{R}^n$ with $P_1|_{(Y_1, Y_2)} = P_0$ and $(F_{Y_1}, F_{Y_2} | F_W) \in \mathcal{C}_{\min}$.

According to [22] (Theorem 4.2), there exist in general many such measures which are parameterized by the matrices and the sets, as stated in Theorem 8, (b), and defined by (80)–(83).

(3) Then, the mutual information of the triple of Gaussian random variables is calculated, using Theorem 8, (b) for any choice of $Q_W \in \mathcal{Q}_W$, where $\mathcal{Q}_W$ is given by (82). Then

$$I(Y_1, Y_2; W) = H(Y_1, Y_2) - H(Y_1 | W) - H(Y_2 | W).$$

The following calculations are then obvious:

$$\det(Q_{(Y_1, Y_2)}) = \det \left( \begin{array}{cc} I & D \\ D & I \end{array} \right) = \det(I - D^2) = \prod_{i=1}^{n}(1 - d_i^2);$$

$$H(Y_1, Y_2) = \frac{1}{2} \ln(\det(Q_{(Y_1, Y_2)})) + \frac{1}{2} (2n) \ln(2\pi e) = \frac{1}{2} \sum_{i=1}^{n} \ln(1 - d_i^2) + n \ln(2\pi e);$$

$$P_{Y_1 | W}(y_1 | W) \in G(E[Y_1 | F_W], Q_{Y_1 | W}),$$

$$E[Y_1 | F_W] = Q_{Y_1 | W} Q_W^{-1} W = D^{1/2} Q_{W}^{-1} W;$$

$$Q_{Y_1 | W} = I - Q_{Y_1 | W} Q_{W}^{-1} Q_W Q_{W}^{-1} T_{Y_1 | W} = I - D^{1/2} Q_{W}^{-1} D^{1/2} W;$$

$$H(Y_1 | W) = \frac{1}{2} \ln(\det(I - D^{1/2} Q_{W}^{-1} D^{1/2})) + \frac{1}{2} n \ln(2\pi e);$$

$$E[Y_2 | F_W] = Q_{Y_2 | W} Q_W^{-1} W = D^{1/2} Q_{W}^{-1} Q_W Q_{W}^{-1} T_{Y_1 | W} = I - D^{1/2} Q_{W}^{-1} D^{1/2};$$

$$Q_{Y_2 | W} = I - Q_{Y_2 | W} Q_{W}^{-1} Q_W Q_{W}^{-1} T_{Y_1 | W} = I - D^{1/2} Q_{W}^{-1} D^{1/2} W;$$

$$H(Y_2 | W) = \frac{1}{2} \ln(\det(I - D^{1/2} Q_{W}^{-1} D^{1/2})) + \frac{1}{2} n \ln(2\pi e).$$

From the above calculations, it then follows

$$I(Y_1, Y_2; W) = H(Y_1, Y_2) - H(Y_1 | W) - H(Y_2 | W)$$

$$= \frac{1}{2} \sum_{i=1}^{n} \ln(1 - d_i^2) + n \ln(2\pi e)$$

$$- \frac{1}{2} \ln(\det(I - D^{1/2} Q_{W}^{-1} D^{1/2})) - \frac{1}{2} n \ln(2\pi e)$$

$$- \frac{1}{2} \ln(\det(I - D^{1/2} Q_{W}^{-1} D^{1/2})) - \frac{1}{2} n \ln(2\pi e)$$

$$= \frac{1}{2} \sum_{i=1}^{n} \ln(1 - d_i^2) - \frac{1}{2} \ln(\det([I - D^{1/2} Q_{W}^{-1} D^{1/2}][I - D^{1/2} Q_{W}^{-1} D^{1/2}])).$$  

(112)
The above calculations verify the statements of Corollary 1. 

(4) The computation of \( C(Y_1, Y_2) \) requires the solution of an optimization problem.

\[
C(Y_1, Y_2) = \inf_{P_1 \in \mathcal{P}_1} I(Y_1, Y_2; W) \\
= \inf_{Q_W \in \mathcal{Q}_W} \left\{ \frac{1}{2} \sum_{i=1}^n \ln(1 - d_i^2) - \frac{1}{2} \ln(\det([I - D^{1/2}Q_W^{-1}D^{1/2}][I - D^{1/2}Q_WD^{1/2}])] \right\}. \tag{115}
\]

Since the first term in (115), \( \frac{1}{2} \sum_{i=1}^n \ln(1 - d_i^2) \), does not depend on \( Q_W \) and the natural logarithm is a strictly increasing function, then

\[
C(Y_1, Y_2) \text{ is equivalent to: } \sup_{Q_W \in \mathcal{Q}_W} \det\left([I - D^{1/2}Q_W^{-1}D^{1/2}][I - D^{1/2}Q_WD^{1/2}]) \right]. \tag{116}
\]

Define:

\[
L_1(Q_W) = (I - D^{1/2}Q_W^{-1}D^{1/2})(I - D^{1/2}Q_WD^{1/2}), \\
f_1(Q_W) = \det(L_1(Q_W)). \tag{117}
\]

Note that the expression \( L_1(Q_W) \in \mathbb{R}^{n \times n} \) is a non-symmetric square matrix in general.

It will be proven that

\[
f_1(Q_W) = \det(L_1(Q_W)) \leq \det([I - D]^2), \quad \forall Q_W \in \mathcal{Q}_W, \tag{119}
\]

\[
\det(L_1(Q_W)) = \det([I - D]^2) \quad \text{if and only if} \quad Q_W = I. \tag{120}
\]

From these two relations, it follows that \( Q_W^* = I \in \mathbb{R}^{n \times n} \) is the unique solution of the supremization problem.

The inequality in (119) follows from Proposition A4. The equality of (120) is proven in two steps. If \( Q_W = I \), then the equality of (120) holds as follows from direct substitution in (117). The converse is proven by contradiction. Suppose that \( Q_W \neq I \). Then, it again follows from Proposition A4 that strict inequality holds in (119). Hence, the equality is proven.

(5) Finally, the value of \( C(Y_1, Y_2) \) is computed for \( Q_W^* = I \).

\[
C(Y_1, Y_2) = \frac{1}{2} \sum_{i=1}^n \ln(1 - d_i^2) - \frac{1}{2} \ln(\det(I - D^{1/2}Q_W^{-1}D^{1/2})) - \frac{1}{2} \ln(\det(I - D^{1/2}Q_WD^{1/2})) \\
= \frac{1}{2} \sum_{i=1}^n \ln(1 - d_i^2) - \frac{1}{2} \ln(\det(I - D)) \\
- \frac{1}{2} \sum_{i=1}^n \ln(1 - d_i^2) - \frac{1}{2} \sum_{i=1}^n \ln((1 - d_i)^2) - \frac{1}{2} \sum_{i=1}^n \ln\left(\frac{1 - d_i^2}{(1 - d_i)^2}\right) \\
= \frac{1}{2} \sum_{i=1}^n \ln\left(\frac{1 + d_i}{1 - d_i}\right) = \frac{1}{2} \sum_{i=1}^n \ln\left(1 + \frac{2d_i}{1 - d_i}\right). 
\]

(b) It follows from part (a) of the theorem that \( C(Y_1, Y_2) \) is attained as the mutual information \( I(Y_1, Y_2; W) \) for a random variable \( W \) with \( Q_W = Q = I \). Consider now a triple of random variables \( (Y_1, Y_2, W) \in G(0, Q_s(I)) \) as defined in (80)–(83), hence, \( Q_W = I \). Denote the random variable \( W \) from now on by \( W^* \) to indicate that it achieves the infimum of the definition of \( C(Y_1, Y_2) \). Thus, \( Q_{W^*} = I \) and

\[
(Y_{12}, Y_{22}, W^*) \in G(0, Q_s(I)), \\
Q_s(I) = \begin{pmatrix} I & D & D^{1/2} \\ D & I & D^{1/2} \\ D^{1/2} & D^{1/2} & I \end{pmatrix} > 0. \tag{121}
\]
Let $V : \Omega \to \mathbb{R}^{n_2}$ be a Gaussian random variable with $V \in G(0, I)$ which is independent of $(Y_1, Y_2, W)$.

Define the new state variable $\tilde{W} = L_1 Y_1 + L_2 Y_2 + L_3 V$. Then, $(Y_1, Y_2, V, W^*)$ are jointly Gaussian and it has to be shown that then $Q_{\tilde{W}} = I$, $Q_{Y_1, \tilde{W}} = D^{1/2}$, and $Q_{Y_2, \tilde{W}} = D^{1/2}$. These equalities follow from simple calculations using the expressions of $L_1, L_2$, and $L_3$ which calculations are omitted. It then follows from those calculations and the definition of the Gaussian measure $G(0, Q_y(I))$ that, almost surely, $\tilde{W} = W^*$.

The signals are then represented by
\begin{align*}
Z_1 &= Y_1 - E[Y_1 | F^{W^*}] = Y_1 - Q_{Y_1, W^*}^{-1}(Q_{W^*})^{-1}W^* = Y_1 - D^{1/2}W^*, \\
Z_2 &= Y_2 - E[Y_2 | F^{W^*}] = Y_2 - Q_{Y_2, W^*}^{-1}(Q_{W^*})^{-1}W^* = Y_2 - D^{1/2}W^*.
\end{align*}

It is proven that the triple of random variables $(Z_1, Z_2, W^*)$ are independent.

\begin{align*}
E[Z_1(W^*)^T] &= E[Y_1(W^*)^T] - D^{1/2}E[W^*(W^*)^T] = D^{1/2} - D^{1/2} = 0, \\
E[Z_2(W^*)^T] &= E[(Y_1 - D^{1/2}W^*)(Y_2 - D^{1/2}W^*)^T] = 0.
\end{align*}

Hence, the original signals are represented as shown by the formulas,
\begin{align*}
Y_1 &= Z_1 + D^{1/2}W^*, \text{ by (121), } Q_{y_1, W^*}Q_{W^*}^{-1} = D^{1/2}, \text{ and by definition of } Z_1, \\
Y_2 &= Z_2 + D^{1/2}W^*, \text{ similarly.}
\end{align*}

3.3. Wyner’s Common Information of Arbitrary Gaussian Random Variables

First, the two special cases of (1) a tuple of independent Gaussian random variables and (2) a tuple of identical Gaussian random variables are analyzed. From those results and that of the previous subsection, one can then prove Wyner’s common information for arbitrary Gaussian random variables.

The special case of the canonical variable form with only private parts is presented below.

**Proposition 5.** Consider the case of a tuple of Gaussian vectors with only private parts. Hence, the Gaussian measure is

\begin{equation}
(Y_{13}, Y_{23}) \in G(0, Q_{Y_{13}, Y_{23}}), \quad Q_{Y_{13}, Y_{23}} = \begin{pmatrix}
I & 0 \\
0 & I
\end{pmatrix}, \quad Y_{13} : \Omega \to \mathbb{R}^{p_3}, \quad Y_{23} : \Omega \to \mathbb{R}^{p_3}.
\end{equation}

(a) The minimal $\sigma$-algebra $F^W$ which makes $Y_{13}, Y_{23}$ conditionally independent is the trivial $\sigma$-algebra denoted by $F_0 = \{\emptyset, \Omega\}$. Thus, $(F^{13}, F^{123} | F_0) \in CI$. The random variable $W$, in this case, is the constant $W_3 = 0 \in \mathbb{R}$, hence $F^{W_3} = F_0$.

(b) Then, $W^* = W_3$ and

\begin{equation}
C(Y_1, Y_2) = C(Y_{13}, Y_{23}) = I(Y_{13}, Y_{23}; W_3) = 0.
\end{equation}

(c) The weak stochastic realization that achieves $C(Y_{13}, Y_{23}) = 0$ is

\begin{equation}
Z_1 = Y_{13}, \quad Z_2 = Y_{23}, \quad W_3 = 0.
\end{equation}

The special case of canonical variable form with only identical parts is presented below.
Proposition 6. Consider the case of a tuple of Gaussian vectors with only the identical part. Hence the Gaussian measure is,

\[ Y_{11} : \Omega \to \mathbb{R}^{p_{11}}, \ Y_{21} : \Omega \to \mathbb{R}^{p_{21}}, \ p_{11} = p_{21}, \]

\[ (Y_{11}, Y_{21}) \in G(0, Q(\eta_{11}, \eta_{21})), \ Q(\eta_{11}, \eta_{21}) = \begin{pmatrix} I & I \\ I & I \end{pmatrix}, \ Y_{11} = Y_{21} \ a.s. \] (127)

(a) The only minimal \( \sigma \)-algebra which makes \( Y_{11} \) and \( Y_{21} \) Gaussian conditional-independent is \( F_{Y_{11}} = F_{Y_{21}} \). The state variable is thus, \( W_{1} = Y_{11} = Y_{21} \) and \( F_{W} = F_{Y_{11}} = F_{Y_{21}} \).

(b) Then \( C(Y_{1}, Y_{2}) = C(Y_{11}, Y_{21}) = +\infty \).

(c) The weak stochastic realization is again simple, the variable \( W \) equals the identical component and there is no need to use the signals \( Z_1 \) and \( Z_2 \). Thus, the representations are,

\[ Z_1 = 0 \in \mathbb{R}, \ Z_2 = 0 \in \mathbb{R}, \ W = Y_{11} = Y_{21}. \] (128)

Theorem 4 is now proven. Thus, the setting is that of a tuple of arbitrary Gaussian random variables, not necessarily restricted to the correlated parts of these random variables of Theorem 8, by (77)–(79). It is shown that \( C(Y_{1}, Y_{2}) \) is computed by a decomposition and by the use of the formulas previously obtained in Section 3.2.

Proof of Theorem 4. (a)

\[ C(Y_{1}, Y_{2}) = \inf_{(Y_{1}, Y_{2}, W) \in \text{CIG}} I(Y_{1}, Y_{2}; W) \]

\[ = \inf \left\{ I(Y_{11}, Y_{21}; W_{1}) + I(Y_{12}, Y_{22}; W_{2}) + I(Y_{13}, Y_{23}; 0) \right\}, \] by Proposition A1,

\[ \geq \inf_{(Y_{1}, Y_{2}, W) \in \text{CIG}} I(Y_{11}, Y_{21}; W_{1}) + \inf_{(Y_{1}, Y_{2}, W) \in \text{CIG}} I(Y_{12}, Y_{22}; W_{2}) + \inf_{(Y_{1}, Y_{2}, W) \in \text{CIG}} I(Y_{13}, Y_{23}; 0) \]

\[ = C(Y_{11}, Y_{21}; W_{1}) + C(Y_{12}, Y_{22}; W_{2}) + C(Y_{13}, Y_{23}; 0) \]

\[ = \begin{cases} 0, & \text{if} \ p_{11} = 0, p_{23} = 0, p_{12} = p_{21} = p_{22} = 0, \\ \frac{1}{2} \sum_{n=1}^{p_{12}} \ln \left( \frac{1 + \rho_{12}}{1 - \rho_{12}} \right), & \text{if} \ p_{12} = p_{22} > 0, p_{11} = p_{21} = 0, p_{13} > 0, p_{23} > 0, \\ +\infty, & \text{else}. \end{cases} \] (129)

The latter equality follows from, respectively, Proposition 6, Theorem 11 and Proposition 5 (a and b). It will be shown that \( C(Y_{1}, Y_{2}) \) is less than or equal to the right-hand side of Equation (129). From the latter inequality and the above inequality then follows the expression according to Equation (129).

To be specific, it will be proven that \( C(Y_{1}, Y_{2}) \) is less than the expression \( I(Y_{1}, Y_{2}; W^{*}) \) where \( W^{*} \) is defined in statement (b) of the proposition. It then follows from the proof of Theorem 11 that \( (F_{Y_{12}}, F_{Y_{22}}|F_{W_{2}}) \in \text{CIG}_{\text{min}} \).

Then:

\[ C(Y_{1}, Y_{2}) = \inf_{(Y_{1}, Y_{2}) \in \text{CIG}} I(Y_{1}, Y_{2}; W) \leq I(Y_{1}, Y_{2}; W^{*}) \]

\[ = I(Y_{11}, Y_{21}; W_{1}) + I(Y_{12}, Y_{22}; W_{2}) + I(Y_{13}, Y_{23}; 0) \]

\[ = \begin{cases} 0, & \text{if} \ p_{11} = 0, p_{23} = 0, p_{11} = p_{12} = p_{21} = p_{22} = 0, \\ \frac{1}{2} \sum_{n=1}^{p_{12}} \ln \left( \frac{1 + \rho_{12}}{1 - \rho_{12}} \right), & \text{if} \ p_{12} = p_{22} > 0, p_{11} = p_{21} = 0, p_{13} > 0, p_{23} > 0, \\ +\infty, & \text{else}. \end{cases} \]

The latter equality is proven as follows. In the first case, when \( p_{11} > 0, p_{23} > 0, \) and \( p_{11} = p_{12} = p_{21} = p_{22} = 0, \) then \( Y_{1} = Y_{13} \) and \( Y_{2} = Y_{23} \) are independent random variables. It then follows from Proposition 5 that \( I(Y_{1}, Y_{2}; 0) = I(Y_{13}, Y_{23}; 0) = 0. \) In the second case,
when \( p_{12} = p_{22} > 0, p_{13} \geq 0, p_{23} \geq 0, \) and \( p_{11} = p_{21} = 0, \) it follows from Proposition A1 and from Theorem 11 that

\[
I(Y_1, Y_2; W^*) = I(Y_{12}, Y_{22}; W_2^2) + I(Y_{13}, Y_{23}; 0) = \sum_{i=1}^{n} \ln \left( 1 + \frac{d_i}{1 - d_i} \right).
\]

In the third case, when \( p_{11} = p_{21} > 0 \) and the other \( p_{ij} \) indices are arbitrary, then \( I(Y_1, Y_2; W^*) = +\infty. \) Hence, the inequality \( C(Y_1, Y_2) \leq \) right-hand side is proven and hence equality holds.

(c) This directly follows from Proposition 4. See also Section 3.6 of [21].

A procedure for the numerical calculations of Wyner’s information common information is given in Section 3.7 of [21].

4. Parametrization of Gray and Wyner Rate Region and Wyner’s Lossy Common Information

This section is devoted to the characterizations of rates that lie in the Gray–Wyner rate region for a tuple of Gaussian random variables with square error distortion functions.

By Gray–Wyner [1] (Theorem 8), reproduced in Theorem 1 to characterize rate triples \((R_0, R_1, R_2) \in R_{GW}(\Delta_1, \Delta_2),\) it is necessary to:

(i) Characterize the rate distortion functions \( R_{Y_i}(\Delta_i), R_{Y_i|W}(\Delta_i) i = 1,2 \) and \( R_{Y_1,Y_2}(\Delta_1, \Delta_2) \);

(ii) Construct the realizations that induce the test channels of \( R_{Y_i}(\Delta_i), R_{Y_i|W}(\Delta_i), i = 1,2 \) and \( R_{Y_1,Y_2}(\Delta_1, \Delta_2),\) and understand the structural properties of the realizations.

4.1. Characterizations of Joint, Conditional and Marginal RDFs

Theorem 12 is the characterization of the joint RDF \( R_{Y_1,Y_2}(\Delta_1, \Delta_2) \) from [24].

**Theorem 12.** Ref. [24] Consider a tuple of Gaussian random variables \( Y_i : \Omega \to \mathbb{R}^{p_i}, \) with \( (Y_1, Y_2) \in G(0, Q_{(Y_1,Y_2)}) , Q_{(Y_1,Y_2)} \geq 0 \) (which implies \( Q_{Y_i} > 0, \) for \( i = 1,2 \)). Consider the joint RDF \( R_{Y_1,Y_2}(\Delta_1, \Delta_2) \) with square error distortion functions \( D_{Y_1}(y_1, \hat{y}_1) = ||y_1 - \hat{y}_1||_{R_1}^2, \)

\( D_{Y_2}(y_2, \hat{y}_2) = ||y_2 - \hat{y}_2||_{R_2}^2. \) Then, the following hold:

(a) The mutual information \( I(Y_1, Y_2; \hat{Y}_1, \hat{Y}_2) \) satisfies

\[
I(Y_1, Y_2; \hat{Y}_1, \hat{Y}_2) \geq I(Y_1, Y_2; E\{Y_1|\hat{Y}_1, \hat{Y}_2\}, E\{Y_2|\hat{Y}_1, \hat{Y}_2\}) \tag{130}
\]

and the mean square error satisfies

\[
E\left\{ ||Y_i - \hat{Y}_i||_{R_i}^2 \right\} \geq E\left\{ ||Y_i - E\{Y_i|\hat{Y}_1, \hat{Y}_2\}||_{R_i}^2 \right\}, \quad i = 1,2. \tag{131}
\]

Moreover, inequalities in (130) and (131) hold with equality if there exists a jointly Gaussian realization of \((\hat{Y}_1, \hat{Y}_2)\) or a Gaussian test channel distribution \( P_{\hat{Y}_1, \hat{Y}_2|Y_1,Y_2} \) such that the joint distribution \( P_{\hat{Y}_1, \hat{Y}_2|Y_1,Y_2} \) is jointly Gaussian, and such that the following identities both hold:

\[
E\{Y_1|\hat{Y}_1, \hat{Y}_2\} = E\{Y_1|\hat{Y}_1\} = \hat{Y}_1, \quad E\{Y_2|\hat{Y}_1, \hat{Y}_2\} = E\{Y_2|\hat{Y}_2\} = \hat{Y}_2. \tag{132}
\]
(b) A realization that achieves the lower bounds of part (a), i.e., satisfies (132), is the Gaussian realization of \((Y_1, Y_2, \hat{Y}_1, \hat{Y}_2)\) given by

\[
\begin{align*}
\left( \hat{Y}_1, \hat{Y}_2 \right) &= H \left( Y_1 \mid Y_2 \right) + \left( V_1 \mid V_2 \right) \\
(V_1, V_2) &\in G(0, Q_{(V_1, V_2)}), \quad (V_1, V_2) \text{ independent of } (Y_1, Y_2),
\end{align*}
\] (133)

\[H_Q(Y_1, Y_2) = Q_{(Y_1, Y_2)} - Q_{(E_1, E_2)} \geq 0,\] (134)

\[Q_{(V_1, V_2)} = Q_{(E_1, E_2)} H^T = H_Q(Y_1, Y_2) - H_Q(Y_1, Y_2) H^T \geq 0,
\] (135)

\[Q_{(V_1,V_2)}^1 = Q_{(E_1, E_2)} Q_{(Y_1,Y_2)} Q_{(E_1,E_2)} \geq 0,
\] (136)

if \(Q_{(Y_1,Y_2)} > 0\) then
\[H = I_{p_1+p_2} - Q_{(E_1,E_2)} Q_{(Y_1,Y_2)}^1,\] (137)

where \((E_1, E_2)\) is the error tuple, that satisfies the structural property,
\[E_i = Y_i - E\{Y_i\mid F_{\hat{Y}_i,Y_2}\} = Y_i - \hat{Y}_i, \quad i = 1, 2
\] (138)

and the variance matrix of this tuple is,
\[(E_1, E_2) \in G(0, Q_{(E_1,E_2)}), \quad Q_{(E_1,E_2)} = \begin{pmatrix} Q_{E_1} & Q_{E_1,E_2} \\ Q_{E_1,E_2} & Q_{E_2} \end{pmatrix} \in \mathbb{R}^{(p_1+p_2) \times (p_1+p_2)}.
\]

(c) The joint RDF \(R_{Y_1,Y_2}(\Delta_1, \Delta_2)\) is characterized by

\[R_{Y_1,Y_2}(\Delta_1, \Delta_2) = \inf_{E\{||E_i||_{\mathbb{R}^{p_i}}\} = \text{trace}(Q_{E_i}) \leq \Delta_i, i = 1, 2} \frac{1}{2} \ln \left( \frac{\det(Q_{(Y_1,Y_2)})}{\det(Q_{(E_1,E_2)})} \right) + \in [0, \infty],
\] (139)

such that \(Q_{(Y_1,Y_2)} = Q_{(Y_1,Y_2)} - Q_{(E_1,E_2)} \geq 0
\] (140)

where the test channel distribution \(P_{\hat{Y}_1,Y_2\mid Y_1,Y_2}\) or the joint distribution \(P_{\hat{Y}_1,Y_2,Y_1,Y_2}\) is induced by the realization of part (b).

The conditional rate distortion function, derived in [25,26], is also required.

**Theorem 13.** Ref. [25] (Theorem 1, Thorem 4), [26] Consider a triple of random variables \(Y_i : \Omega \rightarrow \mathbb{R}^{p_i}, i = 1, 2, W : \Omega \rightarrow \mathbb{W}\), where \(W\) is continuous or finite-valued, with joint distribution \(P_{Y_1,Y_2,W}\), and marginal distributions \(P_{Y_1,Y_2}\) and \(P_{Y_i}\), \(i = 1, 2\), the jointly Gaussian distribution \((Y_1, Y_2) \in G(0, Q_{(Y_1,Y_2)}), Q_{(Y_1,Y_2)} \geq 0\) and \(Y_i \in G(0, Q_{Y_i}), Q_{Y_i} > 0, i = 1, 2,\) respectively. Consider the conditional RDFs \(R_{Y_i\mid W}(\Delta_i), i = 1, 2\) with square error distortion functions \(D_{Y_i}(y_i, \hat{y}_i) = ||y_i - \hat{y}_i||_{\mathbb{R}^{p_i}}^2, i = 1, 2\). Then, the following hold.

(a) For an arbitrary random variable, \(W : \Omega \rightarrow \mathbb{W}\), the mutual information \(I(Y_i; \hat{Y}_i\mid W)\) satisfies

\[I(Y_i; \hat{Y}_i\mid W) \geq I(Y_i; \mathbb{E}\{Y_i\mid F_{\hat{Y}_i,W}\}\mid W), \quad i = 1, 2
\] (141)

and the mean square error satisfies

\[\mathbb{E}\{||Y_i - \hat{Y}_i||_{\mathbb{R}^{p_i}}^2\} \geq \mathbb{E}\{||Y_i - \mathbb{E}\{Y_i\mid F_{\hat{Y}_i,W}\}||_{\mathbb{R}^{p_i}}^2\}, \quad i = 1, 2.
\] (142)

Moreover, inequalities in (141) and (142) hold with equality, if there exists a realization of \(\hat{Y}_i\) of the test channel distribution \(P_{\hat{Y}_i\mid Y_i,W}\), such that the joint distribution \(P_{\hat{Y}_i,Y_i,W}\) satisfies the identity:

\[\hat{X}_i^{cm} = \mathbb{E}\{Y_i\mid F_{\hat{Y}_i,W}\} = Y_i, \quad i = 1, 2.
\] (143)
(b) Suppose \( P_{Y_1,Y_2,W} \) is a jointly Gaussian distribution and \( W : \Omega \to \mathbb{R}^n \) is Gaussian. A realization that achieves the lower bounds of part (a), i.e., satisfies (143), is the Gaussian realization of \((Y_i, W, \hat{Y}_i)\) given by

\[
\begin{align*}
\hat{Y}_i &= H_i Y_i + (I_{W_i} - H_i) Q_{Y_i,W} Q_{W} W + V_i, \quad i = 1, 2, \\
V_i &\in G(0, Q_{Y_i}), \quad V_i \text{ independent of } Y_i, \\
H_i Q_{Y_i|W} &= Q_{Y_i|W} - Q_{E_i} \geq 0, \\
Q_{V_i} &= H_i Q_{E_i} + H_i Q_{Y_i|W} H_i^T \geq 0,
\end{align*}
\]

where \(^t\) denotes the pseudoinverse of a matrix, \( E_i \) is the error, that satisfies the structural property,

\[
E_i = Y_i - E \{ Y_i | F_{\hat{Y}_i,W} \} = Y_i - \hat{Y}_i, \quad E_i \in G(0, Q_{E_i}), \quad i = 1, 2.
\]

The RDF \( R_{Y_i|W}(\Delta_i) \) for jointly Gaussian \((Y_1, Y_2, W)\) is characterized by

\[
R_{Y_i|W}(\Delta_i) = \inf_{||E_i||^2_{\mathbb{R}^n}} \left\{ \frac{1}{2} \ln \left( \frac{\det(Q_{Y_i|W})}{\det(Q_{E_i})} \right) \right\} = \inf_{\Omega} \left\{ \frac{1}{2} \ln \left( \frac{\det(Q_{Y_i|W})}{\det(Q_{E_i})} \right) \right\} \quad i = 1, 2
\]

such that \( Q_{Y_i} = Q_{Y_i|W} - Q_{E_i} \geq 0 \),

where the test channel distribution \( P_{\hat{Y}_i|Y_i,W} \) or the joint distribution \( P_{\hat{Y}_i,Y_i,W} \) is induced by the above realization.

The following is stated as a conjectured, because it is not shown in this paper; it can be shown using Theorem 13.

**Conjecture 1.** Conditional RDF of Gaussian sources with arbitrary conditioning RV

Consider a triple of random variables \( Y_i : \Omega \to \mathbb{R}^n, i = 1, 2, W : \Omega \to \mathbb{W}, \) where \( W \) is continuous or finite-valued, with joint distribution \( P_{Y_1,Y_2,W} \) and marginal distributions \( P_{Y_1,Y_2} \) and \( P_{Y_i}, i = 1, 2 \), the jointly Gaussian distribution \((Y_1, Y_2) \in G(0, Q_{(Y_1,Y_2)}), Q_{(Y_1,Y_2)} \geq 0 \) and \( Y_i \in G(0, Q_{Y_i}), Q_{Y_i} > 0, i = 1, 2 \), respectively. Consider the conditional RDFs \( R_{Y_i|W}(\Delta_i), i = 1, 2 \) with square error distortion functions \( D_{Y_i}(y_i, \hat{y}_i) = ||y_i - \hat{y}_i||^2_{\mathbb{R}^n}, i = 1, 2 \).

Then, the following hold.

(a) For an arbitrary random variable, \( W : \Omega \to \mathbb{W}, \) and \( X_i^{cm} \) satisfying (143), the following lower bounds hold.

\[
I(X_i; \hat{X}_i|W) \geq I(X_i; \hat{X}_i^{cm}|W), \quad i = 1, 2
\]

\[
\geq \int_W I(X_i; \hat{X}_i^{cm}|W = w) P_W(dw)
\]

\[
\geq \inf_{w \in W} I(X_i; \hat{X}_i^{cm}|W = w)
\]

\[
E[D_{X_i}(X_i, \hat{X}_i)] = \int_W \left( \int_{\mathbb{R}^n \times \mathbb{R}^n} D_{X_i}(x_i, \hat{x}_i) P_{X_i,\hat{X}_i|W}(x_i, \hat{x}_i|w) \right) P_W(dw)
\]

\[
= \int_W \Delta_i(w) P_W(dw),
\]

\[
\geq \int_W \Delta_i^{cm}(w) P_W(dw),
\]

\[
\geq \inf_{w \in W} \Delta_i^{cm}(w), \quad i = 1, 2,
\]

where

\[
\Delta_i(w) \doteq E[D_{X_i}(X_i, \hat{X}_i)|W = w], \quad \Delta_i^{cm}(w) \doteq E[D_{X_i}(X_i, \hat{X}_i^{cm})|W = w], \quad i = 1, 2.
\]
Moreover, the inequalities in (153), (157), are achieved if, (i) (141) holds, and (ii) the mutual information I(\(X_i; \hat{X}_i|W = w\)) and \(\Delta_i(w)\) for \(i = 1, 2\), are independent of \(w \in \mathbb{W}\).

(b) The rate distortion function \(R_{X_i|W}(\Delta_i)\) for \(W : \Omega \rightarrow \mathbb{W}\) a continuous or finite-valued, achieves a minimum value if, (i) \(W : \Omega \rightarrow \mathbb{R}^n, n \in \mathbb{Z}_+\) is Gaussian, and \(P_{Y_i|W}\) is jointly Gaussian, (ii) \((\hat{X}_i, X_i, W)\) is given by the realization of Theorem 13(b) for \(i = 1, 2\).

The characterization of the marginal RDFs \(R_{Y_i}(\Delta_i), i = 1, 2\)—which are well-known, and can be found in many books—is also needed; the weak realization of the test channel, which follows from Theorem 13 (see also [25]), as a degenerate case, and summarized in the next theorem, is important in this paper.

**Theorem 14.** Ref. [25] (Theorem 1, Theorem 4) Consider a tuple of Gaussian random variables \(Y_i : \Omega \rightarrow \mathbb{R}^p, i = 1, 2\), with \((Y_1, Y_2) \in \mathbb{G}(0, Q_{(Y_1, Y_2)}^1), Q_{(Y_1, Y_2)} > 0, Q_{Y_i} > 0\), for \(i = 1, 2\).

For the marginal RDFs \(R_{Y_i}(\Delta_i), i = 1, 2\) with square error distortion functions \(D_{Y_i}(y_i, \hat{y}_i) = ||y_i - \hat{y}_i||^2_{\mathbb{R}^p}\), the statements of Theorem 13 hold with \(W\), generating the trivial information, i.e., \(P_{W} = \{\Omega, \emptyset\}\). That is, the marginal RDFs \(R_{Y_i}(\Delta_i)\) are characterized by

\[
R_{Y_i}(\Delta_i) = \inf_{E \in \{||E_i||^{p_i}\} = \text{tr}(Q_{E_i}) \leq \Delta_i} \frac{1}{2} \ln \left( \frac{\det(Q_{Y_i})}{\det(Q_{E_i})} \right)^{+} \in [0, \infty], \quad i = 1, 2,
\]

such that \(Q_{\hat{Y}_i} = Q_{Y_i} - Q_{E_i} \geq 0\) (159)

where the test channel distribution \(P_{Y_i|Y_i}\) or the joint distribution \(P_{\hat{Y}_i Y_i}\) is induced by the realization

\[
\hat{Y}_i = H_i Y_i + V_i, \quad i = 1, 2,
\]

\(V_i \in \mathbb{G}(0, Q_{V_i}), \quad V_i \text{ independent of } Y_i\), (160)

\(H_i Q_{Y_i} = Q_{Y_i} - Q_{E_i} \geq 0\), (161)

\(Q_{V_i} = H_i Q_{E_i} = H_i Q_{Y_i} - H_i Q_{Y_i} H_i^T \geq 0\). (162)

and where \(E_i\) is the error that satisfies the structural property

\[
E_i = Y_i - E\{Y_i|\hat{Y}_i\} = Y_i - \hat{Y}_i, \quad E_i \in \mathbb{G}(0, Q_{E_i}), \quad i = 1, 2.
\]

(163)

Then, we express the characterization of the joint RDF of Theorem 12, using the canonical variable form, and the canonical correlation coefficients. The special case when \(Q_{(E_1, E_2)}\) is block-diagonal is given in [27].

**Theorem 15.** Consider the statement of Theorem 12. Compute the canonical variable form of the tuple of Gaussian random variables \((Y_1, Y_2) \in \mathbb{G}(0, Q_{(Y_1, Y_2)}^1), Q_{Y_i} > 0\), according to Algorithm A1. This yields the indices \(p_{11} = p_{21}, p_{12} = p_{22}, p_{13} = p_{31}, p_{23} = p_{32}, \) and \(n = p_{11} + p_{12} = p_{21} + p_{22}\), the diagonal matrix \(D_4\) with canonical correlation coefficients \(d_{4i} \in (0, 1)\) for \(i = 1, \ldots, p_{12}\), and decompositions (see Algorithm A1, 1–4)

\[
Q_{Y_1} = U_1 D_1 U_1^T, \quad Q_{Y_2} = U_2 D_2 U_2^T,
\]

with \(U_i \in \mathbb{R}^{p_i \times p_i}\) orthogonal \((U_i U_i^T = I_{p_i} = U_i^T U_i), i = 1, 2\), singular-value decomposition of \(D_4^{-\frac{1}{2}} U_1^T Q_{Y_1} U_2 D_2^{-\frac{1}{2}} = U_3 D_3 U_3^T\), (165)
with $U_3 \in \mathbb{R}^{p_1 \times p_1}, U_4 \in \mathbb{R}^{p_2 \times p_2}$ orthogonal,

$$D_3 = \begin{pmatrix} I_{p_1} & 0 & 0 \\ 0 & D_4 & 0 \\ 0 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{p_1 \times p_2},$$

$$D_4 = \text{Diag}(d_{4,1}, ..., d_{4,p_2}) \in \mathbb{R}^{p_1 \times p_2}, \quad 1 > d_{4,1} \geq d_{4,2} \geq \ldots \geq d_{4,p_2} > 0.$$  

(168)

(169)

Define the new variance matrix of $Q_{(Y_1,Y_2)}$ according to

$$Q_{\text{cvf}} = \begin{pmatrix} I_{p_1} & D_3 \\ D_3^T & I_{p_2} \end{pmatrix}.$$

(170)

Compute the canonical variable form of the tuple of Gaussian error random variables $(E_1, E_2) \in \mathbb{G}(0,Q_{(E_1,E_2)})$ of Theorem 12.(b), according to Algorithm A1. This yields the indices $\pi_{11} = \pi_{21}, \pi_{12} = \pi_{22}, \pi_{13}, \pi_{23}$, and $\pi = \pi_{11} + \pi_{12} = \pi_{21} + \pi_{22}$ and the diagonal matrix $D_4$ with canonical correlation coefficients $\tilde{d}_{4,i} \in (0,1)$ for $i = 1, \ldots, p_2$, and decompositions (see Algorithm A1, 1–4),

$$Q_{E_1} = U_1 D_1 U_1^T, \quad Q_{E_2} = U_2 D_2 U_2^T,$$

$$D_1 = \text{Diag}(\tilde{d}_{1,1}, \ldots, \tilde{d}_{1,p_1}), \quad \tilde{d}_{1,1} \geq \tilde{d}_{1,2} \geq \ldots \geq \tilde{d}_{1,p_1} > 0, \quad i = 1, 2,$$

(171)

(172)

with $U_i \in \mathbb{R}^{p_i \times p_i}$ orthogonal ($U_i U_i^T = I_{p_i} = U_i^T U_i$), $i = 1, 2$, singular-value decomposition of

$$D_1^{-\frac{1}{2}} U_1^T Q_{E_1} U_2 D_2^{-\frac{1}{2}} = U_3 D_3 U_4^T,$$

(173)

with $U_3 \in \mathbb{R}^{p_1 \times p_1}, U_4 \in \mathbb{R}^{p_2 \times p_2}$ orthogonal,

$$D_3 = \begin{pmatrix} I_{p_1} & 0 & 0 \\ 0 & D_4 & 0 \\ 0 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{p_1 \times p_2},$$

$$D_4 = \text{Diag}(\tilde{d}_{4,1}, ..., \tilde{d}_{4,p_2}) \in \mathbb{R}^{p_1 \times p_2}, \quad 1 > \tilde{d}_{4,1} \geq \tilde{d}_{4,2} \geq \ldots \geq \tilde{d}_{4,p_2} > 0.$$  

(174)

(175)

Define the new variance matrix of $Q_{(E_1,E_2)}$ according to,

$$Q_{\text{cvf}} = \begin{pmatrix} I_{p_1} & D_3 \\ D_3^T & I_{p_2} \end{pmatrix}.$$

(176)

The joint RDF $R_{Y_1,Y_2}(\Delta_1, \Delta_2)$ of Theorem 12. (c) is equivalently characterized by

$$R_{Y_1,Y_2}(\Delta_1, \Delta_2) = \inf_{Q_{(E_1,E_2)}} \left\{ \begin{array}{l} Q_{(E_1,E_2)} \geq 0, \quad \sum_{j=1}^{p_1} \sum_{j'=1}^{p_2} \tilde{d}_{1,j} \leq \Delta_1, \sum_{j=1}^{p_1} \sum_{j'=1}^{p_2} \tilde{d}_{2,j'} \leq \Delta_2 \end{array} \right\} \frac{1}{2} \ln \left( \frac{\det(D_1)}{\det(D_1)} \frac{\det(D_2)}{\det(D_2)} \frac{\det(Q_{\text{cvf}})}{\det(Q_{\text{cvf}})} \right) \in [0,\infty].$$

(177)

such that $Q_{(E_1,E_2)} - Q_{(E_1,E_2)} \geq 0,$

(178)

where

$$\det(Q_{\text{cvf}}) = \det(I_{p_1} - D_3 D_3^T)$$

$$= \begin{cases} 1, & \text{if } p_{13} > 0, p_{23} > 0, p_{11} = p_{12} = p_{21} = p_{22} = 0, \\
\prod_{i=1}^{n} \left(1 - \tilde{d}_{4,i}^2\right), & \text{if } p_{11} = p_{21} = 0, p_{12} = p_{22} = n, p_{13} \geq 0, p_{23} \geq 0, \\
0, & \text{if } p_{11} = p_{21} > 0, p_{12} = p_{22} \geq 0, p_{13} \geq 0, p_{23} \geq 0, \end{cases}$$

(179)

(180)
\[ \det(Q_{\text{cvf}}) = \det(I_{p_1} - D_3 D_3^T) \]  
\[ = \left\{ \begin{array}{ll}
1, & \text{if } \mathcal{P}_{13} > 0, \mathcal{P}_{23} > 0, \mathcal{P}_{11} = \mathcal{P}_{12} = \mathcal{P}_{21} = \mathcal{P}_{22} = 0, \\
\prod_{i=1}^n (1 - \mathcal{P}_{ii}^2), & \text{if } \mathcal{P}_{11} = \mathcal{P}_{21} = 0, \mathcal{P}_{12} = \mathcal{P}_{22} = 1, \mathcal{P}_{13} \geq 0, \mathcal{P}_{23} \geq 0, \\
0, & \text{if } \mathcal{P}_{11} = \mathcal{P}_{21} > 0, \mathcal{P}_{12} = \mathcal{P}_{22} \geq 0, \mathcal{P}_{13} \geq 0, \mathcal{P}_{23} \geq 0.
\end{array} \right. \]  
(181)  
(182)

Moreover, a necessary condition for \( R_{Y_1,Y_2}(\Delta_1,\Delta_2) < +\infty \) is \( \mathcal{P}_{11} = \mathcal{P}_{21} = 0 \).

**Proof.** First, apply Algorithm A1 to the tuple of Gaussian random variables \((Y_1, Y_2) \in G(0, Q_{(Y_1,Y_2)})\), and then to the Gaussian random variables \((E_1, E_2) \in G(0, Q_{(E_1,E_2)})\) of Theorem 12. (b). This gives (166)–(176). Then, (177) follows from (139) using (166)–(176), and the standard properties of a determinant of a matrix. The remaining equations are obtained from (166)–(176). The last statement follows from the values of \( \det(Q_{\text{cvf}}), \det(Q_{\text{cvf}}) \).

**Remark 7.** By Theorem 15, since \( R_{Y_1,Y_2}(\Delta_1,\Delta_2) \in [0,\infty) \), by (180), it suffices to consider \( Q_{(Y_1,Y_2)} > 0 \), which implies \( p_{11} = p_{21} = 0, Q_{Y_i} > 0, i = 1,2 \). Furthermore, to ensure \( R_{Y_1,Y_2}(\Delta_1,\Delta_2) \in [0,\infty) \), it suffices to also consider \( Q_{(E_1,E_2)} > 0 \), which implies that \( \mathcal{P}_{11} = \mathcal{P}_{21} = 0, Q_{E_i} > 0, i = 1,2 \).

From Theorem 15, the next corollary directly follows which identifies the subset of the distortion region such that Gray’s lower bound [28], \( R_{Y_1,Y_2}(\Delta_1,\Delta_2) \geq R_{Y_2|Y_1}(\Delta_2) + R_{Y_1}(\Delta_1) \) holds with equality.

**Corollary 2.** Consider the statement of Theorem 15, and without loss of generality, assume \((Y_1, Y_2) \in G(0, Q_{(Y_1,Y_2)})\), with \( Q_{(Y_1,Y_2)} > 0 \) (and hence \( Q_{Y_i} > 0, i = 1,2 \)). The joint RDF \( R_{Y_1,Y_2}(\Delta_1,\Delta_2) \) with \( D_1, D_2, D_1, D_2, Q_{\text{cvf}}, Q_{\text{cvf}} \), defined in Theorem 15, and corresponding to \( p_{11} = p_{21} = \mathcal{P}_{11} = \mathcal{P}_{21} = 0 \) satisfies the lower bound \( (183) \) is obtained from Theorem 13 by letting \( W = Y_1 \).

\[ R_{Y_1,Y_2}(\Delta_1,\Delta_2) \geq R_{Y_2|Y_1}(\Delta_2) + R_{Y_1}(\Delta_1) \]  
\[ = \mathbb{E}\left\{ \inf_{\|E_2\|_{Q_{(E_2)}}^2} \text{trace}(Q_{(E_2)}) \leq \Delta_2 \right\} \sum_{i=1}^\infty \frac{1}{2} \ln \left( \frac{\det(Q_{Y_i})}{\det(Q_{E_i})} \right) + \inf_{\|E_1\|_{Q_{(E_1)}}^2} \text{trace}(Q_{(E_1)}) \leq \Delta_1 \right\} \sum_{i=1}^\infty \frac{1}{2} \ln \left( \frac{\det(D_1)}{\det(D_1)} \right) \]  
(183)  
(184)  
(185)

that is, \( \mathcal{P}_{12} = \mathcal{P}_{22} = 0 \).

Moreover, the inequalities (184) and (185) hold with the equalities, on the strictly positive surface \( D_C(Y_1,Y_2) \), defined by

\[ D_C(Y_1,Y_2) = \left\{ (\Delta_1,\Delta_2) \in [0,\infty) \times [0,\infty) \mid Q_{(Y_1,Y_2)} - Q_{(E_1,E_2)} > 0 \right\}. \]  
(186)

**Proof.** The lower bound (183) is due to Gray [28]. The equality in (184) follows by using the values of the rate distortion functions in the right hand side of (183). Equality (185) follows from the singular value decomposition of the matrices given in Theorem 15, using \( Q_{Y_2|Y_1} = Q_{Y_2} - Q_{Y_2|Y_1} Q_{Y_1}^{-1} Q_{Y_2|Y_1}^T \). To establish the equalities, note that (177) with \( \det(Q_{\text{cvf}}) = 1 \) equivalently \( \mathcal{P}_{12} = \mathcal{P}_{22} = 0 \), is precisely (185). Moreover, it can be easily verified that \( \mathcal{P}_{12} = \mathcal{P}_{22} = 0 \) for the distortion region \( D_C(Y_1,Y_2) \). \( \square \)
4.2. Wyner’s Lossy Common Information of Correlated Gaussian Vectors

Derived in this section are the characterizations of $C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2)$ via Theorem 2, for jointly Gaussian random variables with square-error distortion, as well as $C_W(Y_1, Y_2)$ via Theorem 3.

**Definition 7.** Wyner’s lossy common information of a tuple of Gaussian multivariate random variables. Consider a tuple of jointly Gaussian random variables $Y_1 : \Omega \rightarrow \mathbb{R}^{p_1} \equiv Y_1, Y_2 : \Omega \rightarrow \mathbb{R}^{p_2} \equiv Y_2$, in terms of the notation $(Y_1, Y_2) \in G(0, Q(Y_1, Y_2)), Q_Y > 0, i = 1, 2$, and square error distortion functions between $(y_1, y_2)$, and its reproduction $(\hat{y}_1, \hat{y}_2)$, given by

$$D_{Y_1}(y_1, \hat{y}_1) = ||y_1 - \hat{y}_1||^2_{\mathbb{R}^{p_1}}, \quad D_{Y_2}(y_2, \hat{y}_2) = ||y_2 - \hat{y}_2||^2_{\mathbb{R}^{p_2}}$$

(187)

where $|| \cdot ||^2_{\mathbb{R}_p}$ denotes Euclidean distances on $\mathbb{R}^p, i = 1, 2$. 

(a) Wyner’s common information (information definition) of the tuple of Gaussian random variables $(Y_1, Y_2)$ is defined by the expression

$$C(Y_1, Y_2) = \inf_{W: \Omega \rightarrow \mathbb{R}^n, (F^{(1)}, F^{(2)}: F^W) \in \text{CIG}} I(Y_1, Y_2; W) \in [0, \infty].$$

(188)

Call any random variable $W$ as defined above such that $(Y_1, Y_2, W) \in G(0, Q(Y_1, Y_2, W))$ and $(F^{(1)}, F^{(2)}: F^W) \in \text{CIG}$ a state of the tuple $(Y_1, Y_2)$.

If there exists a random variable $W^* : \Omega \rightarrow \mathbb{R}^n$ with $n^* \in \mathbb{Z}_+ = \{1, 2, \ldots, n\}$ which attains the infimum; thus, if $C(Y_1, Y_2) = I(Y_1, Y_2; W^*)$, then call that random variable a minimal information state of the tuple $(Y_1, Y_2)$.

(b) Wyner’s common information (operational definition) is defined for a tuple of strictly positive real numbers $\gamma = (\gamma_1, \gamma_2) \in \mathbb{R}_+ \times \mathbb{R}_+ = (0, \infty) \times (0, \infty)$ such that, for all $0 \leq (\Delta_1, \Delta_2) \leq \gamma$,

$$C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2) = C_W(Y_1, Y_2) = C(Y_1, Y_2),$$

for $(\Delta_1, \Delta_2) \in D_W = \{ (\Delta_1, \Delta_2) \in [0, \infty] \times [0, \infty] : 0 \leq (\Delta_1, \Delta_2) \leq \gamma \}$

(189)

provided identity (15) holds, i.e., $R_{Y_1|W}(\Delta_1) + R_{Y_2|W}(\Delta_2) + I(Y_1, Y_2; W) = R_{Y_1, Y_2}(\Delta_1, \Delta_2)$.

By the above definition, the problem of calculating Wyner’s lossy common information via (18) is decomposed into the characterization of $C(Y_1, Y_2)$ such that identity (15) is satisfied. This follows from the fact that the only difference between $C_W(Y_1, Y_2)$ and $C(Y_1, Y_2)$ is the specification of the region $D_W$ such that $C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2) = C_W(Y_1, Y_2) = C(Y_1, Y_2)$ is constant for $(\Delta_1, \Delta_2) \in D_W$.

In the next theorem, we make use of the characterizations of the various rate distortion functions, and the test channel realizations to identify subsets of the rate region that lie on the Pangloss plane, and are consistent with the characterization of Viswanatha, Akyol and Rose [12] (Theorem 1, Equations (19) and (20)).

**Theorem 16.** Consider a tuple $(Y_1, Y_2)$ of Gaussian random variables in the canonical variable form of Definition 1. Restrict the attention to the correlated parts of these random variables, as defined in Theorem 8, by (77)–(79). Furthermore, consider a realization of the random variables $(Y_1, Y_2)$ which induces the family of measures $P_{ei} \subseteq P_{\text{CIG}}^{\text{G}}$, as defined in Corollary 1, by (84)–(88). Then, the following hold.

(a) The joint rate distortion function $R_{Y_1, Y_2}(\Delta_1, \Delta_2)$ of $(Y_1, Y_2)$ with square error distortion satisfies

$$R_{Y_1, Y_2}(\Delta_1, \Delta_2) = \inf_{\Sigma_i \equiv 1, \Delta_{i} \leq \Delta_1, \Delta_{j} \leq \Delta_2} \frac{1}{2} n \sum_{j=1}^{n} \ln \left( \frac{1 - d_j^2}{\Delta_{1,j} \Delta_{2,j}} \right), (\Delta_1, \Delta_2) \in D_C(Y_1, Y_2),$$

(190)

$$\text{trace}(Q_{E_1}) = E||Y_1 - \hat{Y}_1||^2_{\mathbb{R}^n} = \sum_{j=1}^{n} \Delta_{1,j}, \quad \text{trace}(Q_{E_2}) = E||Y_2 - \hat{Y}_2||^2_{\mathbb{R}^n} = \sum_{j=1}^{n} \Delta_{2,j}$$

(191)
where $D_C(Y_1, Y_2)$ is a strictly positive surface, defined by

$$D_C(Y_1, Y_2) = \left\{ (Y_1, Y_2) \in [0, \infty] \times [0, \infty] \mid Q_{(Y_1, Y_2)} - Q_{(E_1, E_2)} > 0 \right\} \quad (192)$$

and where $Q_{(E_1, E_2)}$ is the variance of the errors $E_i = Y_i - \bar{Y}_i, i = 1, 2$, with parameters $P_{11} = P_{21} = P_{12} = P_{22} = 0$, and $P_{13} = P_{23} = n$.

The conditional rate distortion functions $R_{Y_i|W}(\Delta_i)$ of $Y_i$ conditioned on $W$ with square error distortion, and mutual information $I(Y_1, Y_2; W)$ satisfy

$$R_{Y_1|W}(\Delta_1) = \inf_{\text{trace}(Q_{E_1}) \leq \Delta_1} \frac{1}{2} \ln \left( \frac{\det(I - D^{1/2}Q_W^{-1}D^{1/2})}{\det(Q_{E_1})} \right), \quad \Delta_1 \in [0, \infty) \quad (193)$$

$$R_{Y_2|W}(\Delta_2) = \inf_{\text{trace}(Q_{E_2}) \leq \Delta_2} \frac{1}{2} \ln \left( \frac{\det(I - D^{1/2}Q_W^{-1}D^{1/2})}{\det(Q_{E_2})} \right), \quad \Delta_2 \in [0, \infty) \quad (194)$$

$$I(Y_1, Y_2; W) = \frac{1}{2} \ln \left( \frac{\det(I - D^{1/2}D_W^{-1}D^{1/2})}{\det(I - D^{1/2}D_WD^{1/2})} \right). \quad (195)$$

where $\text{trace}(Q_{E_i}), i = 1, 2$ are defined as in (191).

(b) The representations of reproductions (the reader may verify that the realization satisfies the conditions given in Viswanatha, Akyol and Rose [12], Theorem 1, Equations (19) and (20), $(\hat{Y}_1, \hat{Y}_2)$ of $(Y_1, Y_2)$ at the output of decoder 1 and decoder 2, which achieve the joint rate distortion functions $R_{Y_i|W}(\Delta_1, \Delta_2), R_{Y_i|W}(\Delta_i), i = 1, 2$ of part (a), are

$$Y_1 = D^{1/2}Q_W^{-1}W + Z_1, \quad (196)$$

$$Y_2 = D^{1/2}W + Z_2, \quad (197)$$

$$\hat{Y}_1 = Y_1 - Q_{E_1}(I - D^{1/2}Q_W^{-1}D^{1/2})^{-1}Z_1 + V_1, \quad (198)$$

$$= D^{1/2}Q_W^{-1}W + A_1Z_1 + V_1, \quad (199)$$

$$\hat{Y}_2 = Y_2 - Q_{E_2}(I - D^{1/2}Q_WD^{1/2})^{-1}Z_2 + V_2, \quad (200)$$

$$= D^{1/2}W + A_2Z_2 + V_2, \quad (201)$$

$$Z_1 \in G(0, (I - D^{1/2}Q_W^{-1}D^{1/2})), \quad Z_2 \in G(0, (I - D^{1/2}Q_WD^{1/2})), \quad W \in G(0, Q_W), \quad (202)$$

$$Q_{E_1} = E\{Y_1 - \hat{Y}_1 \} (Y_1 - \hat{Y}_1)^T, \quad Q_{E_2} = E\{Y_2 - \hat{Y}_2 \} (Y_2 - \hat{Y}_2)^T, \quad (203)$$

$$V_1 \in G(0, Q_{E_1}A_1^T), \quad V_2 \in G(0, Q_{E_2}A_2^T), \quad (204)$$

$$A_1 = I - Q_{E_1}(I - D^{1/2}Q_W^{-1}D^{1/2})^{-1}, \quad A_2 = I - Q_{E_2}(I - D^{1/2}Q_WD^{1/2})^{-1}, \quad (205)$$

$$Q_{E_i} = U_i\Lambda_iU_i^T, \quad \Lambda_i = \text{Diag}(\Lambda_{i1}, \ldots, \Lambda_{in}) \in \mathbb{R}^{n \times n}, \quad U_iU_i^T = U_i^TU_i = I, \quad i = 1, 2, \quad (206)$$

$$Q_{Y_i|W} = I - D^{1/2}Q_W^{-1}D^{1/2} = U_i\Lambda_{Y_i|W}U_i^T, \quad \Lambda_{Y_i|W} = \text{Diag}(\Lambda_{Y_1|W, t}, \ldots, \Lambda_{Y_1|W, n}), \quad (207)$$

$$Q_{Y_2|W} = I - D^{1/2}Q_WD^{1/2} = U_2\Lambda_{Y_2|W}U_2^T, \quad \Lambda_{Y_2|W} = \text{Diag}(\Lambda_{Y_2|W, t}, \ldots, \Lambda_{Y_2|W, n}), \quad (208)$$

$$(V_1, V_2, Z_1, Z_2, W), \quad \text{are independent} \quad (209)$$

and are parameterized by $Q_W \in Q_W$, where $Q_W$ is defined by the set of Equation (82).

Moreover, the joint distribution $P_{Y_1, Y_2, \hat{Y}_1, \hat{Y}_2, W}$ satisfies (the reader may verify that conditions (210) are identical to Viswanatha, Akyol and Rose [12], Theorem 1, Equations (19) and (20)) for rates that lie on the Pangloss plane

$$P_{\hat{Y}_1, \hat{Y}_2|W} = P_{Y_1|W}P_{Y_2|W}P_{Y_1, Y_2|W} = P_{Y_1, Y_2|Y_1, Y_2} \quad (210)$$

$$P_{Y_1, Y_2, \hat{Y}_1, \hat{Y}_2, W} = P_{Y_1|W}P_{Y_2|Y_1, W}P_{\hat{Y}_1, \hat{Y}_2|Y_1, \hat{Y}_2}P_{W|Y_1, Y_2} \quad (211)$$

$$P_{Y_1, Y_2, \hat{Y}_1, \hat{Y}_2, W} = P_{Y_1|W}P_{Y_2|W}P_{\hat{Y}_1, \hat{Y}_2|W}P_{W|Y_1, Y_2} \quad (212)$$
(c) Consider part (a) and the realization of part (b). Then, \( R_{Y_1|W}^{(\Delta_1)} + R_{Y_2|W}^{(\Delta_2)} + I(Y_1, Y_2; W) = R_{Y_1, Y_2|W}^{(\Delta_1, \Delta_2)} \) on the subset \( \mathcal{D}_C(Y_1, Y_2) \), defined by (192) such that (210) holds.

(d) Suppose \( Q_W = Q_W^* \in \mathcal{Q}_W \) is diagonal, i.e., \( Q_W = \text{Diag}(Q_{W_1}, \ldots, Q_{W_n}) \), \( d_i \leq Q_{W_i} \leq d_i^{-1}, \forall i \). Then, the conditional RDFs \( R_{Y_1|W}^{(\Delta)} \) are given by

\[
R_{Y_1|W}^{(\Delta)}(\Delta_1) = \inf_{\sum_{i=1}^n \Delta_{1,i} = \Delta_1} \frac{1}{2} \sum_{j=1}^n \ln \left( \frac{1 - d_j/Q_{W_j}^*}{\Delta_{1,j}} \right),
\]

\[
R_{Y_2|W}^{(\Delta)}(\Delta_2) = \inf_{\sum_{i=1}^n \Delta_{2,i} = \Delta_2} \frac{1}{2} \sum_{j=1}^n \ln \left( \frac{1 - d_j/Q_{W_j}^*}{\Delta_{2,j}} \right),
\]

and the optimal \( \Delta_{1,j}, \Delta_{2,j} \) are obtained from the water-filling equations,

\[
\Delta_{1,j} = \begin{cases} 
\lambda_j, & \lambda < 1 - d_j/Q_{W_j}^*, \\
1 - d_j, & \lambda \geq 1 - d_j/Q_{W_j}^*, 
\end{cases} \quad \Delta_1 \in [0, \infty),
\]

\[
\Delta_{2,j} = \begin{cases} 
\lambda_j, & \lambda < 1 - d_j/Q_{W_j}^*, \\
1 - d_j, & \lambda \geq 1 - d_j/Q_{W_j}^*, 
\end{cases} \quad \Delta_2 \in [0, \infty).
\]

and the representations of part (b) hold, with \( Q_{E,i}, Q_{Y,i|W} \) diagonal matrices.

**Proof.** (a) Since the attention is restricted to the correlated parts of these random variables, as defined in Theorem 8, by (77)–(79), then the statements of joint RDF \( R_{Y_1, Y_2}^{(\Delta_1, \Delta_2)} \) of part (a) are a special case of Theorem 12, (c), and obtained from Corollary 2. Similarly, expressions (193)–(195) follow from (13). However, as demonstrated shortly, these also follow, from the derivation of part (b). (b) Recall that the joint rate distortion function is achieved by a jointly Gaussian distribution \( P_{Y_1, Y_2, \hat{Y}_1, \hat{Y}_2} \) such that the average square-error distortions are satisfied. Consider the realization of the random variables \( (Y_1, Y_2) \) which induce the family of measures \( P_{\hat{Y}_1, \hat{Y}_2|Y_1, Y_2} \) as defined in Corollary 1, by (84)–(88). By properties of mutual information, then

\[
I(Y_1, Y_2; \hat{Y}_1, \hat{Y}_2) = H(Y_1, Y_2) - I(Y_1, Y_2; \hat{Y}_1, \hat{Y}_2)
\]

\[
= H(Y_1, Y_2) - H(Y_1|\hat{Y}_1, \hat{Y}_2, Y_2) - H(Y_2|\hat{Y}_1, \hat{Y}_2)
\]

\[
= H(Y_1, Y_2) - H(Y_2|Y_1, \hat{Y}_2, Y_1) - H(Y_1|Y_1, \hat{Y}_2)
\]

\[
\geq H(Y_1, Y_2) - H(Y_1|\hat{Y}_1) - H(Y_2|\hat{Y}_2), \quad \text{con. reduces entropy},
\]

\[
= \frac{1}{2} \sum_{i=1}^n \ln(1 - d_i^2) + n \ln(2\pi e) - H(Y_1|\hat{Y}_1) - H(Y_2|\hat{Y}_2)
\]

\[
\geq \frac{1}{2} \sum_{i=1}^n \ln(1 - d_i^2) + n \ln(2\pi e) - \frac{1}{2} \sum_{i=1}^n \ln(\Delta_{1,i}) - \frac{1}{2} \sum_{i=1}^n \ln(\Delta_{2,i})
\]

\[
= \frac{1}{2} \sum_{i=1}^n \left( \frac{1}{\Delta_{1,i} \Delta_{2,i}} \right) \ln(1 - d_i^2)
\]

where \( \sum_{i=1}^n \Delta_{1,i} = E[ ||Y_1 - \hat{Y}_1||_2^2] \leq \Delta_1 \) and \( \sum_{i=1}^n \Delta_{2,i} = E[ ||Y_2 - \hat{Y}_2||_2^2] \leq \Delta_2 \). The average distortion satisfies

\[
\Delta_1 \geq E[ ||Y_1 - \hat{Y}_1||_2^2] \geq E[ ||Y_1 - E[Y_1|F\hat{Y}_1]||_2^2],
\]

\[
\Delta_2 \geq E[ ||Y_2 - \hat{Y}_2||_2^2] \geq E[ ||Y_1 - E[Y_1|F\hat{Y}_1]||_2^2].
\]
Furthermore,

\[
\begin{align*}
\text{if } P_{Y_1,Y_2|\hat{Y}_1,\hat{Y}_2} &= P_{Y_1|\hat{Y}_1} P_{Y_2|\hat{Y}_2} \text{ then inequality (220) holds with equality, } & (226) \\
\text{if } E[Y_1|F^{\hat{Y}_1}] = \hat{Y}_1 \text{ then inequality (224) holds with equality, } & (227) \\
\text{if } E[Y_2|F^{\hat{Y}_2}] = \hat{Y}_2 \text{ then inequality (225) holds with equality. } & (228)
\end{align*}
\]

It can be verified that the representations (196)–(209) satisfy \( P_{Y_1,Y_2|\hat{Y}_1,\hat{Y}_2} = P_{Y_1|\hat{Y}_1} P_{Y_2|\hat{Y}_2} \), \( E[Y_1|F^{\hat{Y}_1}] = \hat{Y}_1 \), and \( E[Y_2|F^{\hat{Y}_2}] = \hat{Y}_2 \), and that all inequalities become equalities. The decomposition of the joint distribution according to (210) follows from the representations of \((\hat{Y}_1, \hat{Y}_2)\), and similarly for (211) and (212). The conditional RDFs \( R_{Y_i|W}(\Delta_i) \), \( i = 1, 2 \) are shown as above. (c) This is easily verified, because (210) holds and hence rates lie on the Pangloss plane, for the strictly positive surface, \( D_C(Y_1, Y_2) \). (d) This follows directly from parts (a)–(c). □

**Remark 8.** We should emphasize that Theorem 16 does not fully characterize the Pangloss plane, i.e., the subset of the rate regions such that \( R_{Y_i|W}(\Delta_1) + R_{Y_i|W}(\Delta_2) + I(Y_1, Y_2; W) = R_{Y_1,Y_2}(\Delta_1, \Delta_2) \) is larger than \( D_C(Y_1, Y_2) \). To determine the entire set that characterizes the Pangloss plane, we need to consider the rate distortion function (177) with (178), and the general realization (39). We do not pursue this further, because it requires the closed-form solution of \( R_{Y_i,Y_2}(\Delta_1, \Delta_2) \), which is currently an open and challenging problem, and beyond the scope of this paper. We should mention that the analysis of the scalar-valued Gaussian example in [12,13], i.e., when \( p_1 = p_2 = 1 \), made use of closed-form expression of \( R_{Y_1,Y_2}(\Delta_1, \Delta_2) \) due to [13].

**Proof of Theorem 6.** One way to prove the statement is to compute the characterizations of the rate distortion functions \( R_X(\Delta_1), R_{Y_i|W}(\Delta_i), i = 1, 2 \) and \( R_{Y_1,Y_2}(\Delta_1, \Delta_2) \), using the realization of the random variables \((Y_1, Y_2)\) which induce the family of measures \( P_{\eta} \subseteq P_{\text{min}}^C \), as defined in Corollary 1, by (84)–(88). In view of Definition 7. (b), it suffices to verify that identity (15) holds, i.e., \( R_{Y_i|W}(\Delta_1) + R_{Y_i|W}(\Delta_2) + I(Y_1, Y_2; W) = R_{Y_1,Y_2}(\Delta_1, \Delta_2) \) for \((\Delta_1, \Delta_2) \in D_W \), for the choice \( W = W^* \in G(0, I) \) which achieves the minimum in (188) (i.e., due to Theorem 11. (b)).

Similar to Theorem 16, it can be shown that the conditional RDFs \( R_{Y_i|W^*}(\Delta_i), i = 1, 2 \) are given by

\[
\begin{align*}
R_{Y_1|W^*}(\Delta_1) &= \inf_{\Sigma_1 \leq \Delta_1} \frac{1}{2} \sum_{j=1}^n \ln \left( \frac{1-d_{1j}}{\Delta_{1j}} \right), \; W^* \in G(0, I) \quad (229) \\
R_{Y_2|W^*}(\Delta_2) &= \inf_{\Sigma_2 \leq \Delta_2} \frac{1}{2} \sum_{j=1}^n \ln \left( \frac{1-d_{2j}}{\Delta_{2j}} \right), \; W^* \in G(0, I), \quad (230) \\
E[|Y_1 - \hat{Y}_1|_{\mathbb{R}^n}] &= \sum_{j=1}^n \Delta_{1j}, \quad E[|Y_2 - \hat{Y}_2|_{\mathbb{R}^n}] = \sum_{j=1}^n \Delta_{2j}. \quad (231)
\end{align*}
\]

The pay-off of the joint RDF \( R_{Y_1,Y_2}(\Delta_1, \Delta_2) \) in (190) is related to the pay-offs of the conditional RDFs \( R_{Y_1|W}(\Delta_1), R_{Y_2|W}(\Delta_2) \), and \( C(Y_1, Y_2) = I(Y_1, Y_2; W^*) \) in (102), via the identity

\[
\frac{1}{2} \sum_{j=1}^n \ln \left( \frac{1-d_{1j}^2}{\Delta_{1j} \Delta_{2j}} \right) = \frac{1}{2} \sum_{j=1}^n \ln \left( \frac{1-d_{1j}}{\Delta_{1j}} \right) + \frac{1}{2} \sum_{j=1}^n \ln \left( \frac{1-d_{2j}}{\Delta_{2j}} \right) + \frac{1}{2} \sum_{j=1}^n \ln \left( \frac{1+d_{1j}}{1-d_{1j}} \right) \quad (232).
\]

For \((\Delta_1, \Delta_2) \in D_W \) defined by (47), it then follows from (232), the identity

\[
R_{Y_1,Y_2}(\Delta_1, \Delta_2) = \inf_{\Sigma_1 \leq \Delta_1, \Sigma_2 \leq \Delta_2} \left( \frac{1}{2} \sum_{j=1}^n \ln \left( \frac{1-d_{1j}^2}{\Delta_{1j} \Delta_{2j}} \right) \right) \quad (233)
\]

\[
= R_{Y_1|W^*}(\Delta_1) + R_{Y_2|W^*}(\Delta_2) + I(Y_1, Y_2; W^*), \quad (\Delta_1, \Delta_2) \in D_W \text{ defined by (47).} \quad (234)
\]
This completes the proof. □

4.3. Applications to Problems of the Literature [15–17]

The next two corollaries illustrate the application of the results developed in this paper to the optimization problems analyzed in [15–17].

**Corollary 3.** Applications to problems in [15]

Consider the Gaussian secure source coding and Wyner’s common information [15], defined by the optimization problem [15] (see Equation (18), Section IV.B),

\[
\arg\min_{P_{Y_1,Y_2,W} : P_{Y_1,Y_2,W} = P_{Y_1|W}P_{Y_2|W}} \left\{ \lambda I(Y_1; W) + I(Y_1, Y_2; W) \right\}, \quad \lambda \in [0, \infty) \tag{235}
\]

where the tuple \((Y_1, Y_2)\) are the zero mean jointly Gaussian, and \(W : \Omega \rightarrow \mathbb{W}\) is continuous or discrete-valued random variable (The derivation of the formula for (235) in [15] makes use of rate distortion functions, [15] (Equation (47))).

Then, the following hold.

For any jointly distributed random variables \((Y_1, Y_2, W)\) that minimize the expression in (235), there exists a jointly Gaussian triple \((Y_1, Y_2, W)\) such that \(W \mapsto \mathbb{R}^n\) is a Gaussian random variable, which achieves the same minimum value.

Moreover, the following characterization of (235) holds.

\[
\arg\min_{P_{Y_1,Y_2,W} : P_{Y_1,Y_2,W} = P_{Y_1|W}P_{Y_2|W}} \left\{ \lambda I(Y_1; W) + I(Y_1, Y_2; W) \right\} = \arg\min_{Q_W \in \mathcal{Q}_W} \left\{ \frac{\lambda}{2} \ln \det(I - D^{1/2}Q_W D^{1/2}) + \frac{1}{2} \sum_{i=1}^n \ln(1 - d_i^2) \right\} 
\]

(236)

\[
\arg\min_{Q_W \in \mathcal{Q}_W} \left\{ \frac{\lambda}{2} \sum_{i=1}^n \ln(1 - \frac{d_i}{q_i})^{-1} + \frac{1}{2} \sum_{i=1}^n \ln(1 - d_i^2) \right\} 
\]

(237)

where

\[
Q_W = D_W = \text{Diag}(q_1, q_2, \ldots, q_n) \in \mathcal{Q}_W, \quad q_1 \geq q_2 \geq \ldots \geq q_n > 0, \quad \mathcal{Q}_W = (82). \tag{238}
\]

**Proof.** By the use Theorem 9. (c), it suffices to restrict attention to jointly Gaussian random variables \((Y_1, Y_2, W)\). Transform the tuple \((Y_1, Y_2)\) in the canonical variable form of Definition 1. Restrict attention to the correlated parts of these random variables, as defined in Theorem 8, by (77)–(79), and consider the realization of the transformed random variables of Corollary 1. Then, the value of \(I(Y_1; W) + I(Y_2, Y_2; W)\) is identical to the value of the same expression, evaluated using the realization of Corollary 1. By simple evaluation, using the realization of Corollary 1,

\[
\lambda I(Y_1; W) + I(Y_2, Y_2; W) = \frac{\lambda}{2} \ln \det(I - D^{1/2}Q_W^{-1} D^{1/2}) 
\]

(239)

\[
+ \frac{1}{2} \sum_{i=1}^n \ln(1 - d_i^2) \tag{239}
\]

and it is parameterized by \(Q_W \in \mathcal{Q}_W\), where \(\mathcal{Q}_W\) is defined by the set of Equation (82). By Hadamard’s determinant inequality, an achievable lower bound on the first right-hand side term of (239), holds if \(Q_W \in \mathcal{Q}_W\) and \((I - D^{1/2}Q_W^{-1} D^{1/2})\) is a diagonal matrix, and this lower bound is achieved by a diagonal \(Q_W \in \mathcal{Q}_W\). Furthermore, by recalling the derivation of Theorem 11, an achievable lower bound on the second right-hand side term of (239)
holds, i.e., of $I(Y_1, Y_2; W)$, when $Q_W \in Q_W$ is diagonal. Hence, both lower bounds are achieved simultaneously, by $Q_W \in Q_W$ and $Q_W$ a diagonal matrix. Then, an achievable lower bound on (239) is obtained, if $Q_W$ is specified by (238). ☐

The remaining optimization problem in (237) is easily carried out, and hence omitted. Corollary 4 illustrates the application of the results developed in this paper to the Gaussian relaxed Wyner’s common information [16,17] (Definition 2 and Section III).

**Corollary 4.** Applications to problems in [16,17]
Consider the Gaussian relaxed Wyner’s common information considered in [16,17] (see Definition 2 and Section III of [17])

$$C_Y(Y_1, Y_2) = \min_{P_{W|Y_1, Y_2}: I(Y_1, Y_2; W) \leq \gamma} I(Y_1, Y_2; W)$$

(240)

where the tuple $(Y_1, Y_2)$ are zero mean jointly Gaussian, and $W: \Omega \to W$ is continuous or discrete-valued random variable (the value of (240) computed in [16,17], Theorem 4, is different from (241); moreover, the derivation in [16,17], Section III.A, is different from the derivation presented below). Then

$$C_Y(Y_1, Y_2) = C(Y_1, Y_2) = (19), \quad \forall \gamma \in (0, \infty).$$

(241)

**Proof.** By the use Theorem 9, (c), it suffices to restrict the attention to jointly Gaussian random variables $(Y_1, Y_2, W)$. By Proposition 3 or Corollary 1, there exists a family of realizations of $(Y_1, Y_2)$ parameterized by a Gaussian random variable $W$, which induces conditional independence $P_{Y_1, Y_2|W} = P_{Y_1|W}P_{Y_2|W}$, and hence the lower bound $I(Y_1, Y_2; W) \geq H(Y_1, Y_2) - H(Y_1|W) - H(Y_2|W)$ is achieved, i.e., the constraint in (240) is always satisfied, because the minimizer is such that $I(Y_1; Y_2|W) = 0$, i.e., the constraint is not active. Hence, the general solution of (240) is the one given in Theorem 4. ☐

**Remark 9.** Corollary 4 implies that the definition of the relaxed Gaussian Wyner’s common information considered in [16,17] (see Definition 2 and Section III of [17]) should be replaced by $\min_{P_{W|Y_1, Y_2}: I(Y_1, Y_2; W) = 1} I(Y_1, Y_2; W)$, i.e., the inequality is replaced by an equality, so that the constraint is active for all $\gamma \in (0, \infty)$.

4.4. Characterization and Parameterization of the Gray and Wyner Rate Region by Jointly Gaussian RV’s

Derived in this section, for jointly Gaussian random variables with square-error distortion, using [1] (4) of page 1703, Equation (42)), i.e., (12), and the RDFs, $R_{Y_iY_j}(\Delta_1, \Delta_2)$, $R_{Y_i|W}(\Delta_1)$, $R_{Y_i}(\Delta_1)$, $i = 1, 2$, of Theorems 12, 13, 14, and Theorem 9, are:

(1) Theorem 5—the characterizations of the rate region $R_{GW}(\Delta_1, \Delta_2)$,
(2) The characterization of rates that lie on Pangloss Plane.

**Proof of Theorem 5.** (a) This follows from the realization of random variables that induce Gaussian measures, by repeating Theorem 9, without requiring $W$ makes $Y_1$ and $Y_2$ conditionally independent, i.e., the stated realization with random variables $Z_1$ and $Z_2$ correlated, achieves a lower bound on $I(Y_1, Y_2; W)$, among all random variables $W$. (b) The stated characterization (42) follows from the discussion prior to the Theorem, i.e., by an application of [1] (4) of page 1703, Equation (42)), i.e., $T(\alpha_1, \alpha_2)$, and Theorem 9, Theorem 13, which imply that the infimum in $T(\alpha_1, \alpha_2)$ is over the parameterized set of jointly Gaussian random variables $(Y_1, Y_2, W) \in G(0, Q_{(Y_1, Y_2, W)})$ with joint distribution (68). From part (a), then (43) follows. (44) is due to identity $I(Y_1, Y_2; W) = H(Y_1, Y_2) - H(Y_1|Y_2, W) - H(Y_2|W)$ and that the values $R_{Y_i|W}(\Delta_1)$ and $R_{Y_i}(\Delta_2)$ depend only on $Q_{Y_i|W}$ and $Q_{Y_i}$, and the errors (see Theorem 13). ☐
Theorem 17 gives the parameterization of a subset of the Pangloss Plane, as a degenerate case of Theorem 5.

**Theorem 17.** Consider the statement of Theorem 5.

(a) Rate triples \((R_0, R_1, R_2)\) that lie on the Pangloss Plane, is determined by the subset of the rate region \(R_{GW}(\Delta_1, \Delta_2)\) of Theorem 5. (b), such that the joint distribution \(P_{W,Y_1,Y_2}\) satisfies the conditions,

\[
P_{\hat{Y}_1,\hat{Y}_2|W} = P_{\hat{Y}_1|WP_{\hat{Y}_2|W}, P_{Y_1,Y_2|\hat{Y}_1,\hat{Y}_2} = P_{Y_1,Y_2|\hat{Y}_1,\hat{Y}_2}.
\]

Specifically, the Pangloss Plane is characterized by

\[
T^G(a_1, a_2) = \inf_{(Y_1,Y_2,W) \in G(0, Q_{Y_1,Y_2,W})} \left\{ I(Y_1,Y_2;W) + a_1 R_{Y_1|W}(\Delta_1) + a_2 R_{Y_2|W}(\Delta_2) \right\}
\]

\[
= \inf_{Q_{Y_1|Y_2,W}, Q_{Y_2|W}} \left\{ \frac{1}{2} \ln \frac{\det(Q_{Y_1,Y_2})}{\det(Q_{Y_1|W}) \det(Q_{Y_2|W})} + a_1 R_{Y_1|W}(\Delta_1) + a_2 R_{Y_2|W}(\Delta_2) \right\}
\]

such that

\[
P_{W,Y_1,Y_2,\hat{Y}_1,\hat{Y}_2} \text{ satisfies (242), and its marginals induce the test channels}
\]

of the RDFs, \(R_{Y_1,Y_2}(\Delta_1, \Delta_2), R_{Y_i|W}(\Delta_i), R_{Y_i}(\Delta_i), i = 1, 2\), of Theorems 12, 13, 14. (245)

(b) A subset of the rate triple \((R_0, R_1, R_2)\) that lie on the Pangloss Plane is determined by the restriction of part (a) to \((Y_1,Y_2,W) \in CIG\), i.e.,

\[
T^{GCI}(a_1, a_2) = \inf_{(Y_1,Y_2,W) \in G(0, Q_{Y_1,Y_2,W}), (Y_1,Y_2,W) \in CIG} \left\{ I(Y_1,Y_2;W) + a_1 R_{Y_1|W}(\Delta_1) + a_2 R_{Y_2|W}(\Delta_2) \right\}
\]

\[
= \inf_{Q_{Y_1|Y_2,W}, Q_{Y_2|W}} \left\{ \frac{1}{2} \ln \frac{\det(Q_{Y_1,Y_2})}{\det(Q_{Y_1|W}) \det(Q_{Y_2|W})} + a_1 R_{Y_1|W}(\Delta_1) + a_2 R_{Y_2|W}(\Delta_2) \right\}
\]

such that (245) holds, where \(I(Y_1,Y_2;W) = H(Y_1,Y_2) - H(Y_1|W) - H(Y_2|W)\), and \(R_{Y_i|W}(\Delta_i), i = 1, 2\) are given in Theorem 13. (c).

**Proof.** (a) Condition (242) characterizes the rates \((R_0, R_1, R_2) \in R_{GW}(\Delta_1, \Delta_2)\) that lie on the Pangloss plane, and these are derived in [12] (Theorem 1, Equations (19) and (20)). Hence, the statement follows from Theorem 5.

(b) That (246) defines a subset of \(R_{GW}(\Delta_1, \Delta_2)\), follows from the fact that the set of joint distributions \((Y_1,Y_2,W) \in G(0, Q_{Y_1,Y_2,W})\) and \((Y_1,Y_2,W) \in CIG\) is a subset of the set of joint distributions \((Y_1,Y_2,W) \in G(0, Q_{Y_1,Y_2,W})\). Moreover, by part (a) and \((Y_1,Y_2,W) \in CIG\) implies \(I(Y_1,Y_2;W) = H(Y_1,Y_2) - H(Y_1|W) - H(Y_2|W) = H(Y_1,Y_2) - H(Y_1|W) - H(Y_2|W)\) and that the values \(R_{Y_1|W}(\Delta_1)\) and \(R_{Y_2|W}(\Delta_2)\) only depend on \(Q_{Y_1|W}\) and \(Q_{Y_2|W}\), and the errors, as shown in Theorem 13. (c). Hence, the statement holds. \(\square\)

From Theorem 16 and Theorem 17. (b) follows a simpler parameterization of rates that lie on the Pangloss Plane of the Gray–Wyner rate region \(R_{GW}(\Delta_1, \Delta_2)\), when \((Y_1,Y_2)\) are in canonical variable form.
Corollary 5. Consider the statement of Theorem 16 with square error distortion functions, i.e., a tuple \((Y_1, Y_2)\) is the canonical variable form. A subset of the rate triple \((R_0, R_1, R_2)\) that lie on the Pangloss Plane corresponding to the restriction \((Y_1, Y_2|W) \in \text{CIG}\), is determined from

\[
T_{\text{cvf}}^{\text{MG}}(a_1, a_2) = \inf_{Q_W} \left\{ I(Y_1, Y_2; W) + a_1 R_{Y_1|W}(\Delta_1) + a_2 R_{Y_2|W}(\Delta_2) \right\}
\]

\[
= \inf_{Q_W} \left\{ \frac{1}{2} \sum_{i=1}^{n} \ln(1 - d_i^2) - \frac{1}{2} \ln(\det(I - D^{1/2}Q_W^{-1/2}D^{1/2})) \right. \\
\left. + a_1 R_{Y_1|W}(\Delta_1) + a_2 R_{Y_2|W}(\Delta_2) \right\}
\]

such that \((245)\) holds, where \(0 \leq a_i \leq 1, i = 1, 2, a_1 + a_2 \geq 1\), and where \(R_{Y_i|W}(\Delta_i), i = 1, 2\) are given in Theorem 16, and the infimum is taken over \(Q_W \in \mathcal{Q}_W\), defined by the set of Equation (82).

**Proof.** The stated characterization \((248)\) is the application of \([1]\) ((4) of page 1703, Equation (42)) and the results of Theorem 16 and Theorem 17. (b). \(\square\)

In view of Theorem 1 (i.e., Theorem 8 in \([1]\)), additional parameterizations of the Gray–Wyner rate region \(R_{GW}(\Delta_1, \Delta_2)\) directly follow from the expressions already derived, i.e., the joint RDF \(R_{Y_1, Y_2}(\Delta_1, \Delta_2)\) of Theorem 12, the conditional RDFs \(R_{Y_1|Y_2}(\Delta_1), R_{Y_2|Y_1}(\Delta_2)\) of Theorem 13, the marginal RDFs \(R_{Y_1}(\Delta_1), R_{Y_2}(\Delta_2)\) of Theorem 14, and the values of \(I(Y_1, Y_2; W)\), based on Gaussian realization of Theorem 5. (a).

5. Conclusions

This paper formulates the classical Gray and Wyner source coding for a simple network with a tuple of multivariate, correlated Gaussian random variables, with square-error fidelity at the two decoders, from the geometric approach of a Gaussian random variables and the weak stochastic realization of correlated Gaussian random variables. This approach leads to a parameterization of the Gray–Wyner rate region, with respect to variance matrix of the jointly Gaussian triple \((Y_1, Y_2, W)\), where \(W\) a Gaussian auxiliary random variable. However, much remains to be achieved, from the computation point of view, for this problem, and to exploit the new approach to other multi-user problems of information theory.

**Author Contributions:** C.D.C. and J.H.v.S. contributed to the conceptualization, methodology, and writing of the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** The work of C.D. Charalambous was co-funded by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation (Project: EXCELLENCE/1216/0296).

**Data Availability Statement:** Numerical evaluations of Wyner’s common information, based on the implementation of the canonical variable form, and the calculation of the canonical variable coefficients are found in \([21]\), Section 3.7.

**Acknowledgments:** The second author is grateful to H.S. Witsenhausen (formerly affiliated with Bell Laboratories) for contacts about the problem of common information in the early 1980s. This paper is an answer to their questions about the problem of Wyner’s common information. The authors are very grateful to the University of Cyprus for the partial financial support which made their cooperation possible. The authors are also grateful to Guo Lei (Chinese Academy of Sciences, Institute for Mathematics) and to Xi Kaihua (Shandong University, Jinan, Shandong Province, China; formerly of Delft University of Technology) for help with obtaining copies of the papers of Hua LooKeng.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

**Appendix A**

**Appendix A.1. Algorithm to Generate the Canonical Variable Form**

The algorithm that generates \(Q_{\text{CVF}}\) is presented below.
Algorithm A1. Transformation of a variance matrix into its canonical variable form.
Data: \( p_1, p_2 \in \mathbb{Z}_+ \), \( Q \in \mathbb{R}^{(p_1 + p_2) \times (p_1 + p_2)} \), satisfying \( Q = Q^T \geq 0 \) (it is noted that \( Q > 0 \) implies \( Q_{11} > 0, Q_{22} > 0 \), with decomposition

\[
Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{12}^T & Q_{22} \end{pmatrix}, \quad Q_{11} \in \mathbb{R}^{p_1 \times p_1}, \quad Q_{22} \in \mathbb{R}^{p_2 \times p_2}, \quad Q_{12} \in \mathbb{R}^{p_1 \times p_2}, \quad Q_{11} > 0, Q_{22} > 0.
\]

1. Perform singular-value decompositions:

\[
Q_{11} = U_1 D_1 U_1^T, \quad Q_{22} = U_2 D_2 U_2^T,
\]

with \( U_1 \in \mathbb{R}^{p_1 \times p_1} \) orthogonal \((U_1 U_1^T = I = U_1^T U_1)\) and

\[
D_1 = \text{Diag}(d_{1,1}, \ldots, d_{1,p_1}) \in \mathbb{R}^{p_1 \times p_1}, \quad d_{1,1} \geq d_{1,2} \geq \ldots \geq d_{1,p_1} > 0,
\]

and \( U_2, D_2 \) satisfying corresponding conditions.

2. Perform a singular-value decomposition of

\[
D_1^{-\frac{1}{2}} U_1^T Q_{12} U_2 D_2^{-\frac{1}{2}} = U_3 D_3 U_4^T,
\]

with \( U_3 \in \mathbb{R}^{p_1 \times p_1}, U_4 \in \mathbb{R}^{p_2 \times p_2} \) orthogonal and

\[
D_3 = \begin{pmatrix}
I_{p_1} & 0 & 0 \\
0 & D_4 & 0 \\
0 & 0 & 0
\end{pmatrix} \in \mathbb{R}^{p_1 \times p_2},
\]

\[
D_4 = \text{Diag}(d_{4,1}, \ldots, d_{4,p_2}) \in \mathbb{R}^{p_2 \times p_2}, \quad 1 > d_{4,1} \geq d_{4,2} \geq \ldots \geq d_{4,p_2} > 0.
\]

3. Compute the new variance matrix according to

\[
Q_{\text{cvf}} = \begin{pmatrix} I_{p_1} & D_3 \\ D_3^T & I_{p_2} \end{pmatrix},
\]

4. The transformation to the canonical variable representation

\((Y_1 \mapsto S_1 Y_1, \quad Y_2 \mapsto S_2 Y_2)\) is then

\[
S_1 = U_3^T D_1^{-\frac{1}{2}} U_1^T, \quad S_2 = U_4^T D_2^{-\frac{1}{2}} U_2^T.
\]

If \( Q > 0 \) then \( Q_{11} > 0 \) and \( Q_{22} > 0 \), and \( D_3 \) does not contain the first row and first column, i.e., these are removed.

Appendix A.2. Information Theory

In this appendix, the reader finds two formulas of information theory which are used in the body of the paper. These are obtained from [29–31].

The first equality is proven in [29] (p. 19, Th. 2.4.1) and [30] (p. 31, (2.4.26), (2.4.28)).

Proposition A1. Consider random variables \( Y_{1,1}, Y_{1,2}, Y_{2,1}, Y_{2,2}, X_1, X_2 \) such that the following two triples are independent random variables, \((Y_{1,1}, Y_{2,1}, X_1)\) and \((Y_{1,2}, Y_{2,2}, X_2)\). Then, the mutual information expression additively decomposes,

\[
I(Y_{1,1}, Y_{1,2}, Y_{2,1}, Y_{2,2}; X_1, X_2) = I(Y_{1,1}, Y_{2,1}; X_1) + I(Y_{1,2}, Y_{2,2}; X_2).
\]

Proof. For completeness, the proof is given in [21] (Proposition A.2). \(\square\)
Consider a tuple of jointly Gaussian random variables $(X, Y) \in \mathcal{N}(0, Q_{(X,Y)})$ with $X : \Omega \to \mathbb{R}^n$, $Y : \Omega \to \mathbb{R}^p$, $Q_{(X,Y)} > 0$, and $Q_{(X,Y)} = \begin{pmatrix} Q_X & Q_{XY} \\ Q_{XY}^T & Q_Y \end{pmatrix}$. Then, 

\[ H(X) = \frac{1}{2} \ln \left( \det(Q_X) \right) + \frac{1}{2} n \ln(2\pi e), \tag{A2} \]

\[ H(Y|X) = H(X,Y) - H(X) = \frac{1}{2} \ln \left( \det(Q_Y - Q_{XY}Q_X^{-1}Q_{XY}) \right) + \frac{1}{2} p \ln(2\pi e), \tag{A3} \]

\[ I(Y; X) = -\frac{1}{2} \ln \left( \frac{\det(Q_{(X,Y)})}{\det(Q_Y)\det(Q_X)} \right). \tag{A4} \]

Appendix A.3. An Inequality for Determinants

An inequality for matrices is derived in this appendix which is needed in the body of the paper.

**Lemma A1.** Consider the real-valued matrices $A, B \in \mathbb{R}^{n \times n}$. Assume that 

\[ 0 \leq I - A^T A, \quad 0 < I - B^T B, \quad \text{and rank}(B) = n. \tag{A5} \]

Then, 

\[ \det([I - A^T A][I - B^T B]) \leq (\det(I - A^T B))^2. \tag{A6} \]

A related result is mentioned at [32] (Theorem 9.E.6). In that book, the proof of corresponding result refers to paper [33]. That reference was received by the authors but they could not read it because the paper is in Mandarin. However, they could read the formulas of the paper. Hua LooKeng developed these results to calculate an orthonormal basis for a function of one complex variable. A more recent reference for this inequality is [34] (Theorem 7.19).

The proof of Lemma A1 below is analogous to that of Hua LooKeng in [33]. The main differences are in the assumptions.

**Lemma A2.** Ref. [33] (pp. 464, 470). Consider the matrices $A, B \in \mathbb{R}^{n \times n}$. Assume that $I - B^T B$ is a nonsingular matrix and that rank$(B) = n$. Then 

\[ (I - A^T A) - (I - A^T B)[I - B^T B]^{-1}(I - A^T B)^T \]

\[ = -(A - B)[I - B^T B]^{-1}(A - B)^T. \tag{A7} \]

**Proof.** For completeness, the proof is given in [21] (Lemma A.4).

**Proposition A2.** Ref. [33] (Equation (2)). Consider the symmetric positive-definite matrices $Q_1, Q_2, Q \in \mathbb{R}^{n \times n}$ such that $Q_1 + Q_2 = Q$. Then 

\[ \det(Q_1) + \det(Q_2) \leq \det(Q). \tag{A8} \]
Proof. Proof of Lemma A1. By the assumptions, 0 \leq (I - AT)A, 0 < (I - BT)B, and rank(B) = n, from Lemma A2 follows that

\[
(A - B)\begin{bmatrix} I - BT B \end{bmatrix}^{-1}(A - B)^T + [I - A^T A] = (I - AT)B[I - BT B]^{-1}(I - AT B)^T;
\]

\[
0 \leq \text{det}(I - AT A), \text{by the assumption on } A,
\]

\[
\leq \text{det}((A - B)\begin{bmatrix} I - BT B \end{bmatrix}^{-1}(A - B)^T) + \text{det}([I - AT A]), \text{by an assumption on } B,
\]

\[
\leq \text{det}((I - AT B)[I - BT B]^{-1}(I - AT B)^T)
\]

by Lemma A2, Proposition A2, and by the assumptions,

\[
= \left(\text{det}(I - AT B)\right)^2 |\text{det}([I - BT B])|^{-1},
\]

\[
\Rightarrow \text{det}([I - AT A][I - BT B]) = \text{det}([I - AT A]) \text{det}([I - BT B]) \leq \left(\text{det}([I - BT B])\right)^2.
\]

\[\square\]

Another preliminary result is needed.

Proposition A3. Consider the matrix \(Q_X = Q_X^T \in \mathbb{R}^{n \times n}\) of Proposition 2 and the matrix \(D \in \mathbb{R}^{n \times n}\) of Definition 1. Thus, both \(Q_X > 0\) and \(D > 0\). Then,

\[
D \leq Q_X^{-1} \leq D^{-1} \iff D \leq Q_X \leq D^{-1}, \tag{A9}
\]

\[
D < Q_X^{-1} < D^{-1} \iff D < Q_X < D^{-1}. \tag{A10}
\]

Proof. For completeness, the proof is given in [21] (Proposition A.6).  \[\square\]

Proposition A4. Consider the matrices defined in Definition 1. Thus, \(D \in \mathbb{R}^{n \times n}\), is a diagonal matrix satisfying \(0 < D\), and the matrix \(Q_X \in \mathbb{R}^{n \times n}\) satisfies \(Q_X = Q_X^T\) and \(0 < D \leq Q_X \leq D^{-1}\). Then,

\[
\text{det}\left([I - D^{1/2}Q_X^{-1}D^{1/2}][I - D^{1/2}Q_XD^{1/2}]\right) \leq \text{det}([I - D]^2), \forall Q_X, \tag{A11}
\]

\[
\text{det}\left([I - D^{1/2}Q_X^{-1}D^{1/2}][I - D^{1/2}Q_XD^{1/2}]\right) < \text{det}([I - D]^2), \text{if } Q_X \neq I. \tag{A12}
\]

Proof. (1) If \(Q_X < D^{-1}\) then \(\text{det}(D^{-1} - Q_X) > 0\). Consider the case of in which \(Q_X\) satisfies \(Q_X \leq D^{-1}\) but not \(Q_X < D^{-1}\). Then, \(\text{det}(D^{-1} - Q_X) = 0\). Hence,

\[
\text{det}(I - D^{1/2}Q_XD^{1/2}) = \text{det}(D^{1/2}(I - Q_X)D^{1/2}) = \text{det}(D^{1/2})\text{det}(D^{-1} - Q_X)\text{det}(Q^{1/2}) = 0.
\]

Then, \(0 < D < I\) implies that \(\text{det}([I - D]^2) \geq 0\), hence that the inequality (A11) holds.

(2) If \(D < Q_X\) then \(\text{det}(D - Q_X) < 0\). If \(D \leq Q_X\) but not \(D < Q_X\) then by Proposition A3 \(Q_X^{-1} \leq D^{-1}\) but not \(Q_X^{-1} < D^{-1}\). Then, \(\text{det}(D^{-1} - Q_X^{-1}) = 0\). Hence,

\[
\text{det}(I - D^{1/2}Q_X^{-1}D^{1/2}) = \text{det}(D^{1/2}(I - Q_X^{-1})D^{1/2}) = \text{det}(D^{1/2})\text{det}(D^{-1} - Q_X^{-1})\text{det}(D^{1/2}) = 0.
\]

In this case, the inequality (A11) also holds.

(3) Then consider the case in which \(D < Q_X < D^{-1}\). Lemma A1 will be used to prove the result. Define, therefore,

\[
A = (Q_X^{-1/2})D^{1/2}, \quad B = (Q_X^{1/2})D^{1/2}.
\]
First, it is proven that the assumptions of the lemma are satisfied. Note that $0 < Q_X$ implies that $\text{rank}(Q_X) = n$. This and the fact that $\text{rank}(D) = n$ imply that $\text{rank}(Q_X^{-1} D^{1/2}) = n$. Further note that

$$I - A^T A = I - D^{1/2} Q_X^{-1} D^{1/2},$$

$$0 < D \leq Q_X \leq D^{-1} \text{ by assumption, } \Rightarrow$$

$$0 < D \leq Q_X^{-1} \leq D^{-1}, \text{ by Proposition A3},$$

$$0 < D^2 \leq D^{1/2} Q_X^{-1} D^{1/2} \leq I \Rightarrow$$

$$0 < I - D^{1/2} Q_X^{-1} D^{1/2} = I - A^T A;$$

$$0 < D < Q_X < D^{-1} \text{ by the case considered},$$

$$0 < D^2 < D^{1/2} Q_X D^{1/2} < I, \Rightarrow$$

$$0 < I - D^{1/2} Q_X D^{1/2} = I - B^T B;$$

$$I - A^T B = I - D^{1/2} Q_X^{-1} D^{1/2} = I - D.$$

From Lemma A1 it then follows that

$$\det \left( \left( I - D^{1/2} Q_X^{-1} D^{1/2} \right) \left[ I - D^{1/2} Q_X D^{1/2} \right] \right) \leq \det ([I - D]^2).$$

(4) Then suppose that, in addition, $Q_X \neq I$. Then

$$A - B = Q_X^{-1/2} D^{1/2} - Q_X^{1/2} D^{1/2} = Q_X^{-1/2} [I - Q_X] D^{1/2} \neq 0, \text{ using that, } Q_X \neq I,$$

$$0 < (A - B) [I - B^T B]^{-1} (A - B)^T, \text{ because } 0 < I - B^T B, \text{ and } A - B \neq 0,$$

$$\text{det}(I - A^T A)$$

$$< \text{det}((A - B) [I - B^T B]^{-1} (A - B)) + \text{det}(I - A^T A),$$

$$\leq \text{det}((I - A^T B) [I - B^T B]^{-1} (I - A^T B)^T), \text{ by Lemma A2},$$

$$\Rightarrow \text{det}((I - A^T A) (I - B^T B)) = \text{det}(I - A^T A) \text{det}(I - B^T B)$$

$$< \text{det}(I - A^T B)^2, \text{ by Lemma A2 and its proof},$$

$$\Rightarrow \text{det}((I - D^{1/2} Q_X^{-1} D^{1/2}) [I - D^{1/2} Q_X D^{1/2}]) < \text{det} ([I - D]^2), \text{ by sub. of } A, B.$$

References

1. Gray, R.M.; Wyner, A.D. Source coding for a simple network. *Bell Syst. Tech. J.* 1974, 53, 1681–1721. [CrossRef]
2. Wyner, A.D. The common information of two dependent random variables. *IEEE Trans. Inf. Theory* 1975, 21, 163–179. [CrossRef]
3. Witsenhausen, H.S. Values and bounds for common information of two discrete variables. *SIAM J. Appl. Math.* 1976, 31, 313–333. [CrossRef]
4. Witsenhausen, H.S. On sequences of pairs of dependent random variables. *SIAM J. Appl. Math.* 1975, 28, 100–113. [CrossRef]
5. Gacs, P.; Korner, J. Common information is much less than mutual information. *Probl. Control Inf. Theory* 1973, 2, 149–162.
6. Benammar, M.; Zaidi, A. Rate-distortion region of a Gray-Wyner model with side information. *Entropy* 2018, 20, 2. [CrossRef] [PubMed]
7. Benammar, M.; Zaidi, A. Rate-distortion region of a Gray-Wyner problem with side information. In Proceedings of the IEEE International Symposium on Information Theory (ISIT.2017), Aachen, Germany, 25–30 June 2017; pp. 106–110.
8. Benammar, M.; Zaidi, A. Rate-distortion function for a Heegard-Berger problem with two sources and degraded reconstruction sets. *IEEE Trans. Inf. Theory* 2016, 62, 5080–5092. [CrossRef]
9. Benammar, M.; Zaidi, A. Rate-distortion function for a Heegard-Berger problem with common reconstruction constraint. In Proceedings of the IEEE International Theory Workshop (ITW.2015), Jeju Island, Korea, 11–15 October 2015.
10. Heegard, C.; Berger, T. Rate distortion when side information may be absent. *IEEE Trans. Inf. Theory* 1985, 31, 727–734. [CrossRef]
11. Cuff, P.W.; Permuter, H.H.; Cover, T.M. Coordination Capacity. *IEEE Trans. Inf. Theory* 2010, 56, 4181–4206. [CrossRef]
12. Kumar B.; Viswanatha, E.A.; Rose, K. The lossy common information of correlated sources. *IEEE Trans. Inf. Theory* 2014, 60, 3238–3253.
13. Xu, G.; Liu, W.; Chen, B. A lossy source coding interpretation of Wyner’s common information. *IEEE Trans. Inf. Theory* 2016, 62, 754–768. [CrossRef]
14. Xiao, J.-J.; Luo, Z.-Q. Compression of Correlated Gaussian Sources under Individual Distortion Criteria. In Proceedings of the 43rd Annual Allerton Conference on Communication, Control and Computing, Monticello, IL, USA, 28–30 September 2005; pp. 438–447.

15. Satpathy, S.; Cuff, P. Gaussian secure source coding and Wyner’s common information. In Proceedings of the IEEE International Symposium on Information Theory (ISIT.2015), Hong Kong, China, 14–19 July 2015; pp. 116–120.

16. Veld, G.J.O.; Gastpar, M.C. Total correlation of Gaussian vector sources on the Gray-Wyner network. In Proceedings of the 54th Annual Allerton Conference on Communication, Control and Computing (Allerton), Monticello, IL, USA, 27–30 September 2016; pp. 385–392.

17. Sula, E.; Gastpar, M. Relaxed Wyner’s common information. arXiv 2019, arXiv:1912.07083.

18. Hotelling, H. Relation between two sets of variates. Biometrika 1936, 28, 321–377. [CrossRef]

19. Gelfand, M.I.; Yaglom, M. Calculation of the amount of information about a random function contained in another such function. Am. Math. Soc. Transl. 1959, 2, 199–246.

20. van Schuppen, J.H. Common, correlated, and private information in control of decentralized systems. In Coordination Control of Distributed Systems; van Schuppen, J.H., Villa, T., Eds.; Number 456 in Lecture Notes in Control and Information Sciences; Springer International Publishing: Cham, Switzerland, 2015; pp. 215–222.

21. Charalambous, C.D.; van Schuppen, J.H. A new approach to lossy network compression of a tuple of correlated multivariate Gaussian RVs. arXiv 2019, arXiv:1905.12695.

22. van Putten, C.; van Schuppen, J.H. The weak and strong Gaussian probabilistic realization problem. J. Multivar. Anal. 1983, 13, 118–137. [CrossRef]

23. Noble, B. Applied Linear Algebra; Prentice-Hall: Englewood Cliffs, NJ, USA, 1969.

24. Stylianou, E.; Charalambous, C.D.; Charalambous, T. Joint rate distortion function of a tuple of correlated multivariate Gaussian sources with individual fidelity criteria. In Proceedings of the 2021 IEEE International Symposium on Information Theory (ISIT.2021), Melbourne, Australia, 12–20 July 2021; pp. 2167–2172.

25. Gkagkos, M.; Charalambous, C.D. Structural properties of optimal test channels for distributed source coding with decoder side information for multivariate Gaussian sources with square-error fidelity. arXiv 2020, arXiv:2011.10941.

26. Gkangos, M.; Charalambous, C.D. Structural properties of test channels of the RDF for Gaussian multivariate distributed sources. In Proceedings of the 2021 IEEE International Symposium on Information Theory (ISIT.2021), Melbourne, Australia, 12–20 July 2021; pp. 2631–2636.

27. Charalambous, C.D.; van Schuppen, J.H. Characterization of conditional independence and weak realizations of multivariate gaussian random variables: Applications to networks. In Proceedings of the IEEE International Symposium on Information Theory (ISIT.2020), Los Angeles, CA, USA, 21–26 June 2020.

28. Gray, R.M. A new class of lower bounds to information rates of stationary via conditional rate-distortion functions. IEEE Trans. Inf. Theory 1973, 19, 480–489. [CrossRef]

29. Cover, T.M.; Thomas, J.A. Elements of Information Theory; John Wiley & Sons: New York, NY, USA, 1991.

30. Gallager, R.G. Information Theory and Reliable Communication; John Wiley & Sons: New York, NY, USA, 1968.

31. Yaglom, A.M.; Yaglom, I.M. Probability and Information; D. Reidel Publishing Company: Dordrecht, The Netherlands, 1983.

32. Marshall, A.W.; Olkin, I. Inequalities: Theory of Majorization and Its Applications; Academic Press: New York, NY, USA, 1979.

33. Hua, L.K. Inequalities involving determinants. Acta Math. Sin. 1955, 5, 463–470. (In Chinese; English summary)

34. Zhang, F. Positive semidefinite matrices. In Matrix Theory: Basic Results and Techniques; Springer Science+Business Media: New York, NY, USA, 2011; pp. 199–252.