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

Charalambos D. Charalambous
Department of Electrical and Computer Engineering, University of Cyprus
P.O.Box 20537, CY-1678 Nicosia, Cyprus
chadcha@ucy.ac.cy

Jan H. van Schuppen
Van Schuppen Control Research
Gouden Leeuw 143, 1103 KB Amsterdam, The Netherlands
jan.h.van.schuppen@xs4all.nl

May 31, 2019

The classical Gray and Wyner source coding for a simple network for sources that generate a tuple of multivariate, correlated Gaussian random variables \( Y_1 : \Omega \rightarrow \mathbb{R}^{p_1} \) and \( Y_2 : \Omega \rightarrow \mathbb{R}^{p_2} \), is re-examined using the geometric approach of Gaussian random variables, and the weak stochastic realization of correlated Gaussian random variables. New results are:

(1) The formulation, methods and algorithms to parametrize all random variables \( W : \Omega \rightarrow \mathbb{R}^n \) which make the two components of the tuple \( (Y_1, Y_2) \) conditionally independent, according to the weak stochastic realization of \( (Y_1, Y_2) \).

(2) The formulation of random variables \( (Y_1, Y_2) \) via non-singular transformations \( (S_1, S_2) \), into their canonical variable form as \( S_1Y_1 = (Y_1', Y_2') = ((Y_{11}, Y_{12}), Y_{13}), S_2Y_2 = (V_1, V_2') = ((V_{21}, V_{22}), V_{23}) \), where \( Y_{11} = Y_{21} \)-a.s., \( Y_{13} \) and \( Y_{23} \) are independent and each of these has independent components, \( Y_{12} \) and \( Y_{22} \) are correlated and each of these has independent components, \( E[Y_{12}Y_{23}'] = D \), for some diagonal matrix \( D \) with diagonal entries \( 1 > d_1 \geq d_2 \geq \ldots \geq d_n > 0 \) in \((0, 1)\) called the canonical correlation coefficients. A formula of mutual information \( I(Y_1; Y_2) \) between \( Y_1 \) and \( Y_2 \), expressed in terms of the canonical correlation coefficients.

(3) A formula for Wyner’s lossy common information for joint decoding with mean-square error distortions \( E\|Y_1 - \hat{Y}_1\|_{\mathbb{R}^{p_1}} \leq \Delta_1 \in [0, \infty) \) and \( E\|Y_2 - \hat{Y}_2\|_{\mathbb{R}^{p_2}} \leq \Delta_2 \in [0, \infty) \), where \( (\hat{Y}_1, \hat{Y}_2) \) are the reproductions of \( (Y_1, Y_2) \), given by \( C_W(Y_1, Y_2) = \frac{1}{2} \sum_{j=1}^{n} \ln \left( \frac{1+d_j}{1-d_j} \right) \), where the distortion region is defined by \( 0 \leq \Delta_1 \leq \sum_{j=1}^{n} (1-d_j), 0 \leq \Delta_2 \leq \sum_{j=1}^{n} (1-d_j) \).

(4) The methods are shown to be of fundamental importance to the parametrization of the lossy rate region of the Gray and Wyner source coding problem, and the calculation of the smallest common message rate \( R_0 \) on the Gray and Wyner source problem, when the sum rate \( R_1 + R_1 + R_2 \) is arbitrary close to the joint rate distortion function \( R_{Y_1,Y_2}(\Delta_1, \Delta_2) \) of joint decoding.

The methods and algorithms may be applicable to other problems of multi-user communication, such as, the multiple access channel, etc.

The discussion is largely self-contained and proceeds from first principles; basic concepts of weak stochastic realization theory of multivariate correlated Gaussian random variables are reviewed, while certain results are developed to meet the requirement of results (1)-(4).

**Keywords and Phrases:** Gray-Wyner network, Wyner’s lossy common information, weak realizations of conditional independence, canonical variable form of multivariate Gaussian random variables, multi-user communication.
An important class of theoretical and practical problems in communication is of a multi-user nature. Such problems are (i) lossless and lossy network source coding for compression of data generated by multiple sources, and (ii) network channel coding for transmission of data generated by multiple sources over noisy channels. A sub-class of network source coding problems deals with two sources that generate at each time instant, symbols that are stationary memoryless, multivariate, correlated, and jointly Gaussian distributed. A sub-class of network channel coding problems deals with stationary memoryless Gaussian multiple access channels (MAC) with two or more multivariate correlated sources and a multivariate output. This paper is focused on a new approach, which is applied to such network source coding problems, while a brief discussion on the importance of the method to network channel coding problems is included.

Gray and Wyner in their seminal paper, *source coding for a simple network* [10], 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 [29], 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 [28] has investigated bounds for Wyner’s common information, and sequences of pairs of random variables in this regard [27]. Gacs and Körner [7] introduced another definition of common randomness between a tuple of jointly independent and identically distributed random variables. Viswanatha, Akyol and Rose [26], and Xu, Liu, and Chen [31], 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 arbitrary 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 [26] (and references therein).

Similarly, extensions and generalizations of the fundamental work of Slepian and Wolf on MACs are found in Kramer [13] (and references therein). Specific examples are given in the paper [14], where the authors refer to the concept of Wyner’s common information and provide a bound on the transmission rate over Gaussian MACs.

In all these works, the search for methods and algorithms to compute rates that lie on the Gray and Wyner rate region, the Slepian and Wolf rate region, Wyner’s common information, has been an active area since
the early 1970s. Present methods and algorithms to compute such rates, for specific application examples 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, and for the special case of the doubly symmetric binary source with Hamming distance distortion [10] (Section 2.5).

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, with square error distortion, and for the special case of the doubly symmetric binary source [26] (Section III.C), and [31] (Section IV.C).

3. Rates that lie on Slepian and Wolf rate region of memoryless Gaussian MACs (with common message) are only computed in closed form, for the special case of a tuple of scalar-valued inputs and a scalar-valued output [15] (Chapter 8).

4. Important generalizations, of the above application examples to a tuple of sources that generate multivariate correlated Gaussian symbols, and to Gaussian MAC with common message, with multivariate channel inputs and multivariate channel output, require new derivations often of considerable difficulty. The mathematics of the derivations of closed form expressions for such generalizations are not transparent, and often require new mathematical tools.

This paper introduces a new approach at the assemblage of problems of multi-user information theory, in which achievable rate regions or rates, are parametrized by Gaussian auxiliary random variables $W : \Omega \rightarrow \mathbb{R}^n$, that make two multivariate correlated Gaussian random variables $Y_1 : \Omega \rightarrow \mathbb{R}^{p_1}, Y_2 : \Omega \rightarrow \mathbb{R}^{p_2}$ conditionally independent

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

The focus of the current paper is on multi-user problems of information theory, for which the operational definitions of achievable rate regions or rates are characterized by their information theoretic definitions, via mutual information, conditional mutual information, etc., that depend on the joint distribution $P_{Y_1, Y_2, W}$ of the random variables $(W, Y_1, Y_2)$.

One application of the methods and algorithms that emerge from the new approach, is the Gray and Wyner source coding for a simple network [10]. Another application is the stationary memoryless multiple access channel (MAC), with or without a common message [21, 6, 13]. For two sources that generate symbols, according to the model of jointly independent and identically distributed multivariate correlated Gaussian random variables $(Y_1, Y_2)$, the Gray and Wyner rate region can be parametrized by an auxiliary random variable $W$ that satisfies conditional independence (1). The rate region of stationary memoryless Gaussian MAC channels with a common message, with two multivariate channel inputs $X_1 : \Omega \rightarrow \mathbb{R}^{p_1}, X_2 : \Omega \rightarrow \mathbb{R}^{p_2}$, and a multivariate channel output $Y : \Omega \rightarrow \mathbb{R}^{p_2}$, is parametrized by an auxiliary random variable $W$ that satisfies conditional independence (1), with $(Y_1, Y_2)$ replaced by $(X_1, X_2)$.

The following are the highlights of the methods and algorithms of the new approach of the paper:

**Method 1.(a).**

In the context of the Gray and Wyner source coding problem for a simple network, this method utilizes [24] to

(i) parametrize the family of Gaussian probability distributions $P_{Y_1, Y_2, W}(y_1, y_2, w)$ by the multidimensional random variable $W$ such that $(Y_1, Y_2)$ are conditionally independent, conditioned on $W$, that is, (1) holds, and 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

(ii) represent the random variables $(Y_1, Y_2)$ using the weak stochastic realization, expressed in terms of $W$ and two independent random variables $V_1 : \Omega \rightarrow \mathbb{R}^{p_1}, V_2 : \Omega \rightarrow \mathbb{R}^{p_2}$, which are independent of $W$.

**Method 1.(b).**

In the context of the stationary memoryless Gaussian MACs with or without a common message, this method (i) parametrizes the family of Gaussian channel input probability distributions $P_{X_1, X_2, W}(x_1, x_2, w)$ by the multidimensional random variable $W$ such that $(X_1, X_2)$ are conditionally independent, conditioned on $W$, that
is, $P_{X_1,X_2|W}(x_1,x_2|w) = P_{X_1|W}(x_1|w)P_{X_2|W}(x_2|w)$, and the average power constraints $\text{trace}(E[X_iX_i^T]) \leq \kappa_i \in [0,\infty], i = 1,2$ are satisfied, and

(ii) represents the random variables $(X_1,X_2)$ using the weak stochastic realization, expressed in terms of $W$ and two independent random variables $V_1 : \Omega \rightarrow \mathbb{R}^{p_1}, V_2 : \Omega \rightarrow \mathbb{R}^{p_2}$, which are independent of $W$.

**Method 2.**

This method applies to both, the Gray and Wyner source coding for a simple network, and to the stationary memoryless Gaussian MACs. The method utilizes the geometric approach to Gaussian random variables [11, 1, 9], where the underlying geometric object of a Gaussian random variable $Y : \Omega \rightarrow \mathbb{R}^p$ is the $\sigma$-algebra $F_Y$ generated by $Y$. A basis transformation of such a random variable is then the transformation defined by a non-singular matrix $S \in \mathbb{R}^{p \times p}$, and directly follows that $F_Y = F_{SY}$. For a tuple of jointly Gaussian multivariate random variables $(Y_1,Y_2)$, a basis transformation of this tuple consists of a matrix

$$S = \text{Block-diag}(S_1, S_2),$$

$$S_1, S_2 \text{ square and non-singular matrices, and hence the spaces satisfy}$$

$$F_{Y_1} = F^{S_1}Y_1, \quad F_{Y_2} = F^{S_2}Y_2. \quad \text{(4)}$$

This transformation introduces an equivalence relation on the representation of the tuple of random variables $(Y_1,Y_2)$. Thus one can speak about a canonical form for these spaces with the property (the full specification is given in Definition 2.2)

$$S_1Y_1 = (V_1,Y_1') = ((Y_{11},Y_{12}), Y_{13}), \quad S_2Y_2 = (V_2,Y_2') = ((Y_{21},Y_{22}), Y_{23}), \quad \text{(5)}$$

$$Y_{11} = Y_{21} - a.s., \quad \text{(6)}$$

$$Y_{13} \text{ and } Y_{23} \text{ are independent and each of these has independent components,} \quad \text{(7)}$$

$$Y_{12} \text{ and } Y_{22} \text{ are correlated and each of these has independent components,} \quad \text{(8)}$$

$$E[Y_{12}Y_{22}^T] = D \quad \text{(9)}$$

for some diagonal matrix $D$ with diagonal entries in $(0, 1)$ called the canonical correlation coefficients. Component $Y_{13}' = Y_{13}$ is the private component of $S_1Y_1$, and component $V_1 = (Y_{11}, Y_{12})$ is the correlated component of $S_1Y_1$, with respect to $S_2Y_2$, and similarly for $V_2 = (Y_{21}, Y_{22})$. This method is equivalent to pre-processing of the tuple of correlated random variables $(Y_{1j}, Y_{2j}) \sim P_{Y_1,Y_2}, i = 1, \ldots, N$, with the aid of a linear pre-encoder transformation $S = \text{Block-diag}(S_1, S_2)$. The power of methods 1 and 2 is more apparent in the actual calculations of rates and the development of algorithms to compute such rates. Examples of such type will be discussed later, and include an expression of mutual information $I(Y_1; Y_2)$ between $Y_1$ and $Y_2$, expressed only in terms of the canonical correlation coefficients, i.e., the elements of $D$.

The rest of the section serves mainly to review the Gray and Wyner characterization of lossy rate region and the characterization of Wyner’s lossy common information, for the purpose of linking these to Methods 1 and 2. Then the power of methods 1 and 2 with respect to the computations of rates that lie on the Gray and Wyner characterization of lossy rate region is further discussed. The characterization of the rate regions of stationary memoryless Gaussian MAC with or without common message, is included and briefly discussed, to illustrate the importance of Methods 1 and 2 to other multi-user problems.

1.1 Literature Review

(a) The Gray and Wyner source coding for a simple network [10].

Consider the Gray and Wyner source coding for a simple network shown Fig. 1, for a tuple of jointly independent and identically distributed multivariate Gaussian random variables $(Y_1^N, Y_2^N) = \{(Y_{1,i}, Y_{2,i}) : i = 1,2, \ldots, N\},$

$$Y_{1,i} : \Omega \rightarrow \mathbb{Y}_1, \quad Y_{2,i} : \Omega \rightarrow \mathbb{Y}_2, \quad i = 1, \ldots, N \quad \text{(10)}$$
with square error distortion functions at the two decoders,

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

(11)

where \( || \cdot ||_2 \) are Euclidean distances on \( \mathbb{R}^N, i = 1, 2 \).

The encoder takes as its input the data sequences \((Y_1^N, Y_2^N)\) and produces at its output three messages \((S_0, S_1, S_2)\), with binary bit representations \((NR_0, NR_1, NR_2)\). There are three channels, Channel 0, Channel 1, and 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 public channel and channel 1 and channel 2 are the private channels, which connect the encoder to each of the two decoders. Of the three messages, message \(S_0\) is a common or public message that is transmitted through the public 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 has as objective to reproduce \(Y_1^N\) by \(\hat{Y}_1^N\) subject to an average distortion and decoder 2 has as objective to reproduce \(Y_2^N\) by \(\hat{Y}_2^N\), subject to an average distortion, where \((\hat{Y}_{1,i}, \hat{Y}_{2,i}) = (\hat{y}_{1,i}, \hat{y}_{2,i}) \in \hat{Y}_1 \times \hat{Y}_2 \subseteq \mathbb{Y}_1 \times \mathbb{Y}_2, i = 1, \ldots, N\), that is,

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

(12)

The fundamental question posed by Gray and Wyner is: determine which channel capacity triples \((C_0, C_1, C_2)\) are necessary and sufficient for each sequence \((Y_1^N, Y_2^N)\) to be reliably reproduced at the intended decoders, while satisfying the average distortions.

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

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

(13)

for some auxiliary random variable \(W\), i.e., 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 \((Y_1, Y_2)\).

The characterization of \(R_{GW}(\Delta_1, \Delta_2)\) is described in terms of an auxiliary random variable, as follows.

**Theorem 1.1** (Theorem 8 in [10])

Let \(R_{GW}(\Delta_1, \Delta_2)\) denote the Gray and Wyner rate region of the simple network shown in Fig. 1.

Suppose there exists \(\tilde{y}_i \in \hat{Y}_i\) such that \(\mathbb{E}\{d_{Y_i}(y_i, \tilde{y}_i)\} < \infty, \) 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

\[ R_{GW}^{P_{Y_1,Y_2,W}}(\Delta_1, \Delta_2) = \left\{ (R_0, R_1, R_2): R_0 \geq I(Y_1, Y_2; W), R_1 \geq R_{Y_1|W}(\Delta_1), R_2 \geq R_{Y_2|W}(\Delta_2) \right\} \]  

(14)

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 joint decoding of \((Y_1^N, Y_2^N)\) (all single letters).

Let

\[ R_{GW}^* = \left( \bigcup_{P_{Y_1,Y_2,W} \in \mathcal{P}} R_{GW}^{P_{Y_1,Y_2,W}}(\Delta_1, \Delta_2) \right)^c \]  

(15)

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

\[ R_{GW}(\Delta_1, \Delta_2) = R_{GW}^*(\Delta_1, \Delta_2). \]  

(16)
Gray and Wyner [10] (Theorem 6) also showed that, if \((R_0, R_1, R_2) \in \mathcal{R}_{GW}(\Delta_1, \Delta_2)\), then
\[
\begin{align*}
R_0 + R_1 + R_2 & \geq R_{Y_1, Y_2}(\Delta_1, \Delta_2), \\
R_0 + R_1 & \geq R_{Y_1}(\Delta_1), \\
R_0 + R_2 & \geq R_{Y_2}(\Delta_2)
\end{align*}
\] (17) (18) (19)
where \(R_i(\Delta_i)\) is the rate distortion function of \(Y_i^N\) at decoder \(i\), for \(i = 1, 2\). The inequality in (17) 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 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 \(\mathcal{R}_{GW}(\Delta_1, \Delta_2)\).

For the special case of a tuple of scalar-valued (bivariate) Gaussian random variables, i.e., \(p_1 = p_2 = 1\), with square error distortion and \(\Delta_1 = \Delta_2 = \Delta\), Gray and Wyner [10] (Section 2.5, (B)), showed that the choice of scalar-valued Gaussian random variable \(W\) that satisfies conditional independence (1), ensures a corresponding rate triple \((R_0, R_1, R_2) \in \mathcal{R}_{GW}(\Delta_1, \Delta_2)\) that lies on Pangloss Plane, and they derived explicitly the formulae of these rates. Similarly for the the doubly symmetric binary source with Hamming distance distortion function.

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

Wyner [29], 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 Fig. 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,
\[
C(Y_1, Y_2) = \inf_{P_{Y_1Y_2W}: P_{Y_1|W} = P_{Y_1}, P_{Y_2|W} = P_{Y_2}} I(Y_1, Y_2; W).
\] (20)

Here \(P_{Y_1Y_2W}\) is any joint probability distribution on \(\mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{W}\) with \((Y_1, Y_2)\)—marginal \(P_{Y_1}, Y_2\), such that \(W\) makes \(Y_1\) and \(Y_2\) conditionally independent, that is \(P_{Y_1Y_2W} \in \mathcal{P}\).

Although, for continuous-valued random variables, such as, jointly Gaussian, the Wyner’s common information \(C(Y_1, Y_2)\) is meaningful mathematically, its operational interpretation for arbitrary random variables, is shown recently by Viswanatha, Akyol and Rose [26], and Xu, Liu, and Chen [31], to correspond to a specific point on the Gray-Wyner lossy rate region \(\mathcal{R}_{GW}(\Delta_1, \Delta_2)\).

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

Viswanatha, Akyol and Rose [26], and Xu, Liu, and Chen [31], 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. They first defined and characterized the minimum common message rate \(R_0\) on the Gray-Wyner lossy rate region, when the sum rate is arbitrary close to the rate distortion function with joint decoding for the Gray-Wyner lossy network.

The following characterization is derived by Xu, Liu, and Chen [31] (an equivalent characterization is also derived by Viswanatha, Akyol and Rose [26]).

**Theorem 1.2 (Theorem 4 in [31])**

Suppose there exists \(\hat{y}_i \in \hat{\mathcal{Y}}_i\) such that \(E\{d_i(Y_i, \hat{y}_i)\} < \infty\), 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 sum rate not exceeding the joint rate distortion function \(R_{Y_1, Y_2}(\Delta_1, \Delta_2)\), while satisfying the average distortions.

Then \(C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2)\) is characterized by the optimization problem
\[
C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2) = \inf I(Y_1, Y_2; W)
\] (21)
such that the following identity holds

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

(22)

where the infimum is over all random variables \( W \) taking values in \( \mathcal{W} \), which parametrize the source distribution via \( P_{Y_1, Y_2, W} \), having a \( Y_1 \times Y_2 \)-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 constraint.

A necessary condition for the equality constraint (22) to hold is (see Appendix B in [31]) is \( R_{Y_1, Y_2|W}(\Delta_1, \Delta_2) = R_{Y_1|W}(\Delta_1) + R_{Y_2|W}(\Delta_2) \), and sufficient condition for this equality to hold is the conditional independence condition [31]: \( P_{Y_1, Y_2|W} = P_{Y_1|W}P_{Y_2|W} \). Hence, a sufficient condition for any rate \( (R_0, R_1, R_2) \in \mathcal{R}_{GW}(\Delta_1, \Delta_2) \) to lie on the Pangloss plane is the conditional independence.

Further, it is shown in [26, 31], 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 equals to the Wyner’s information theoretic characterization of common information between \( Y_1 \) and \( Y_2 \), defined by (20).

The next theorem is derived by Xu, Liu, and Chen [31].

**Theorem 1.3** (Theorem 5 in [31])

Let \( (Y_1, Y_2) \) be a pair of random variables with distribution \( P_{Y_1, Y_2} \) on the alphabet space \( \mathcal{Y}_1 \times \mathcal{Y}_2 \), where \( \mathcal{Y}_1 \) and \( \mathcal{Y}_2 \) are arbitrary measurable spaces that can be discrete or continuous.

Let \( W \) be any random variable achieving \( C(Y_1, Y_2) \) defined by (20).

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

\[
d_{Y_i}(y_i, \hat{y}_i) > d_{Y_i}(y_i, y_i), \quad y_i \neq \hat{y}_i, \quad i = 1, 2
\]

If the following conditions are satisfied:

1) for any \( y_1 \in \mathcal{Y}_1, y_2 \in \mathcal{Y}_2 \) and \( w \in \mathcal{W}, \ P_{W|Y_1,Y_2} > 0; \)

2) there exists an \( \hat{y}_i \in \hat{\mathcal{Y}}_i, \) such that

\[
E\{d_{Y_i}(Y_i, \hat{Y}_i)\} < \infty, \quad i = 1, 2
\]

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 \( \mathcal{D}_W = \{ (\Delta_1, \Delta_2) \in [0, \infty] \times [0, \infty]: 0 \leq (\Delta_1, \Delta_2) \leq \gamma \}. \)

It should be mentioned that the analog of the above theorem is also derived by Viswanath, Akyol and Rose in [26] in Lemma 1.

Hence, by the work of Viswanatha, Akyol and Rose [26], and Xu, Liu, and Chen [31], then Wyner’s lossy common information is precisely the smallest message rate \( C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2) \) on the Gray-Wyner lossy rate region, for certain distortion levels, when the total rate, \( R_0 + R_1 + R_2 \), is arbitrary close to the rate distortion function with joint decoding, i.e., \( R_{Y_1, Y_2}(\Delta_1, \Delta_2) \).

For a bivariate Gaussian random variables with square-error distortion, the rate-triple \( (R_0, R_1, R_2) \in \mathcal{R}_{GW}(\Delta_1, \Delta_2) \) that lies on the Pangloss plane is computed by Gray and Wyner [10] (see also Section 2.5, (B)). Further, Viswanatha, Akyol and Rose in [26] (Section III.C), and Xu, Liu, and Chen [31] (Theorem 7 and Proposition 3) computed \( C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2) \) and \( C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2) = C_W(Y_1, Y_2) = C(Y_1, Y_2) \). The formulae for \( C(Y_1, Y_2) \) is given below for the purpose of comparing it to Wyner’s lossy common information formulae of a tuple of multivariate Gaussian random variables derived in this paper.
Theorem 1.4 [10, 26, 31]
Consider two jointly Gaussian scalar-valued random variables \((Y_1, Y_2)\) with zero mean \(E[Y_1] = E[Y_2] = 0\), unit variance \(E[Y_1^2] = E[Y_2^2] = 1\), and \(E[Y_1 Y_2] = \rho \in [-1, 1]\), that is, the variance matrix \(^1\) is

\[
Q(Y_1, Y_2) = \begin{bmatrix}
\sigma_{Y_1}^2 & \rho \sigma_{Y_1} \sigma_{Y_2} \\
\rho \sigma_{Y_1} \sigma_{Y_2} & \sigma_{Y_2}^2
\end{bmatrix} = \begin{bmatrix}
1 & \rho \\
\rho & 1
\end{bmatrix}
\]

with eigenvalues \(\lambda_1 = 1 - |\rho|, \lambda_2 = 1 + |\rho|\), and square error distortion \(d_i(Y_i, \hat{Y}_i) = (Y_i - \hat{Y}_i)^2, i = 1, 2\).
For \(\rho \in [0, 1)\) then

\[
C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2) = C_W(Y_1, Y_2) = C(Y_1, Y_2) = \frac{1}{2} \log \frac{1 + \rho}{1 - \rho}, \quad 0 \leq \Delta_i \leq 1 - \rho, \quad i = 1, 2.
\]

It should be mentioned that for the Gaussian example of Theorem 1.4, Gray and Wyner [10] in Section 2.5, (B), computed the rate-triple \((R_0, R_1, R_2) \in R_{GW}(Y_1, Y_2; \Delta, \Delta)\), i.e., for the symmetric case, that lies on the Pan-gloss plane, and that \(C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2) = C_W(Y_1, Y_2) = R_{Y_1 Y_2} (1 - p, 1 - p)\) for \(0 \leq \Delta \leq 1 - \rho\).
Viswanatha, Akyol and Rose in [26], and Xu, Liu, and Chen [31] computed \(C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2)\) for all distortion regions, by using the expression of joint rate distortion function \(R_{Y_1 Y_2}(\Delta_1, \Delta_2)\) derived in [30].

The search for methods and algorithms to compute rate-triples \((R_0, R_1, R_2)\) that lie on the Gray-Wyner lossy rate region \(R_{GW}(\Delta_1, \Delta_2)\), \(C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2)\), Wyner’s lossy common information \(C_W(Y_1, Y_2)\), and Wyner’s common information \(C(Y_1, Y_2)\) (for finite valued random variables), beyond the above two application examples is limited.

However, applications of Wyner’s common information in other problems of information theory are many. Several applications of Wyner’s common information are found in the characterizations of rate regions of multi-user information theory by Kramer in [13] (and references therein). Specific examples are given in the paper [14], where the authors refer to the concept of Wyner’s common information and provide a bound on the transmission over multi-access channels.

But the formula to compute any rate triple \((R_0, R_1, R_2)\) that lies on the Gray-Wyner rate region \(R_{GW}(\Delta_1, \Delta_2)\), \(C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2)\), and Wyner’s lossy common information \(C_W(Y_1, Y_2)\) for the general case of multivariate Gaussian random variables was never published to the best of the authors’ knowledge.

(d) Stationary and memoryless Gaussian MACs with a common message.
Consider the stationary and memoryless Gaussian MAC, with a common message, with two multivariate inputs \((X_{1t}^N, X_{2t}^N)\)

\[
X_{1t} : \Omega \to \mathbb{R}^P, \quad X_{2t} : \Omega \to \mathbb{R}^P, \quad t = 1, \ldots, N
\]

three sources generating independent messages \((S_0, S_1, S_2)\), that take values in message sets \(\mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2\) with binary bit representations, \((NR_0, NR_1, NR_2)\), and a multivariate output, \(Y^N, Y_t : \Omega \to \mathbb{R}^P, t = 1, \ldots, N\), defined by the conditional distributions

\[
P_{Y_t|X_{1t}, X_{2t}, S_0, S_1, S_2} = P(Y_t|X_{1t}, X_{2t}). \quad t = 1, \ldots, N.
\]

The channel inputs are subject to average power constraints,

\[
\frac{1}{N} E\left\{\sum_{t=1}^N \|X_{1t}(S_0, S_i)\|^2_{L^2}\right\} \leq \kappa_i \in [0, \infty], \quad i = 1, 2.
\]

Messages \(S_0\), called the common messages, are available to both encoders, while messages \(S_1\) are available to encoder 1 and messages \(S_2\) are available to encoder 2. The decoder’s input is \(Y^N\) and its output are the three estimates \((\hat{S}_0, \hat{S}_1, \hat{S}_2)\) of \((S_0, S_1, S_2)\).

\(^1\)Gray and Wyner [10] discussed the case \(\rho \in [0, 1)\).
The problem posed by Slepian and Wolf is: determine the rate-triple \((R_0, R_1, R_2)\) such that the error probability \(P^{(N)} = \text{Prob}\{\hat{S}_1 \neq S_1, \hat{S}_2 \neq S_2\}\) is arbitrarily small for large enough \(N\).

Let \(\mathcal{R}_{\text{MAC}}(\kappa_1, \kappa_2)\) denote the rate region of the MAC, when the available power is \((\kappa_1, \kappa_2)\). Define the parametrized family of probability distributions,

\[
\mathcal{P}^{\text{CIG}} = \left\{ P_{X_1, X_2, W}(x_1, x_2, w) : \begin{align*}
P_{X_1, X_2|W}(x_1, x_2|w) &= P_{X_1|W}(x_1|w)P_{X_2|W}(x_2|w), \\
P_{X_1, X_2,W}(x_1, x_2, \infty) &= P_{X_1,X_2}(x_1, x_2), \quad (X_1, X_2, W) \text{ jointly Gaussian, } W : \Omega \to \mathbb{R}^n \end{align*} \right\}. \tag{31}
\]

For each \(P_{Y_1, Y_2} \in \mathcal{P}^{\text{CIG}}\) and \(\kappa_1 \geq 0, \kappa_2 \geq 0\), define the subset of Euclidean 3-dimensional space

\[
\mathcal{R}_{\text{MAC}-C}^{P_{Y_1, Y_2}}(\kappa_1, \kappa_2) = \left\{ (R_0, R_1, R_2) : 0 \leq R_1 \leq I(X_1; Y|X_2, W), \quad 0 \leq R_2 \leq I(X_2; Y|X_1, W), \quad \begin{align*}
R_1 + R_2 &\leq I(X_1, X_2; Y|W), \\
R_0 + R_1 + R_2 &\leq I(X_1, X_2; Y), \\
P_{W|X_1, X_2, Y} &= P_W P_{X_1|W} P_{X_2|W} P_Y |X_1, X_2, \quad \text{tr}(E[X_i X_i^T]) \leq \kappa_i \in [0, \infty], \quad i = 1, 2 \end{align*} \right\} \tag{32}
\]

where \(I(X_1; Y|X_2, W), I(X_2; Y|X_1, W), I(X_1, X_2; Y)\) are single letter conditional and mutual information measures. In (32), use is made of the condition that \(Y\) and \(W\) are conditionally independent given \((X_1, X_2)\).

Slepian’s and Wolf characterization of the capacity region, is defined by single letter information theoretic quantities, by the set of rate triples (see also Chapter 8 in [13], and specifically equations (8.9)-(8.12), and Section 8.3 in [13]):

\[
\mathcal{R}_{\text{MAC}}^{\text{MAC}-C}(\kappa_1, \kappa_2) = \left\{ \bigcup_{P_{Y_1, Y_2} \in \mathcal{P}^{\text{CIG}}} \mathcal{R}_{\text{MAC}-C}^{P_{Y_1, Y_2}}(\kappa_1, \kappa_2) \right\}^c = \mathcal{R}_{\text{MAC}}(\kappa_1, \kappa_2) \tag{33}
\]

(e) The stationary memoryless Gaussian multiple access channels (MACs) without common message.

This is a special case of (d), with two multivariate inputs, \((X_1^N, X_2^N)\), as in (28), two sources generating independent messages \((S_1, S_2)\), that take values in message sets \(\mathcal{M}_1 \times \mathcal{M}_2\), with binary bit representations \((NR_1, NR_2)\), and a multivariate output, \(Y^N, Y : \Omega \to \mathbb{R}^p, t = 1, \ldots, N\), defined by the conditional distributions

\[
P_{Y|X_1, X_2, S_1, S_2} = P_Y(x_1, x_2), \quad t = 1, \ldots, N \tag{34}
\]

subject to average power constraints,

\[
\frac{1}{N} E\left\{ \sum_{t=1}^N \|X_i|S_i\|^2 \right\} \leq \kappa_i \in [0, \infty], \quad i = 1, 2 \tag{35}
\]

Messages \(S_1\) are available to encoder 1 and messages \(S_2\) are available to encoder 2. The decoder’s input is \(Y^N\) and its output are the two estimates \((\hat{S}_1, \hat{S}_2)\) of \((S_1, S_2)\).

The problem is to determine the rate tuple \((R_1, R_2)\) such that the error probability \(P^{(N)} = \text{Prob}\{\hat{S}_1 \neq S_1, \hat{S}_2 \neq S_2\}\) is arbitrarily small for large enough \(N\).

Let \(\mathcal{R}_{\text{MAC}}(\kappa_1, \kappa_2)\) denote the rate region of the MAC. For each \(P_{Y_1, Y_2} \in \mathcal{P}^{\text{CIG}}\) and \(\kappa_1 \geq 0, \kappa_2 \geq 0\), define the subset of Euclidean 3-dimensional space

\[
\mathcal{R}_{\text{MAC}}^{P_{Y_1, Y_2}}(P_1, P_2) = \left\{ (R_1, R_2) : \begin{align*}
0 \leq R_1 &\leq I(X_1; Y|X_2, W), \quad 0 \leq R_2 \leq I(X_2; Y|X_1, W), \\
R_1 + R_2 &\leq I(X_1, X_2; Y|W), \quad \text{tr}(E[X_i X_i^T]) \leq \kappa_i \in [0, \infty], \quad i = 1, 2 \end{align*} \right\}. \tag{36}
\]

The characterization of the capacity region, is defined by single letter information theoretic quantities, by the set of rate tuples (see Section 14.3 in [6], and specifically Theorem 14.3.3):

\[
\mathcal{R}_{\text{MAC}}(\kappa_1, \kappa_2) = \left\{ \bigcup_{P_{Y_1, Y_2} \in \mathcal{P}^{\text{CIG}}} \mathcal{R}_{\text{MAC}}^{P_{Y_1, Y_2}}(\kappa_1, \kappa_2) \right\}^c = \mathcal{R}_{\text{MAC}}(\kappa_1, \kappa_2). \tag{37}
\]
For the scalar-valued Gaussian MAC with common message, defined by $Y_i = X_{1,t} + X_{2,t} + Z_t, t = 1, \ldots, N$, i.e., $p_1 = p_2 = p_3$, with $Z_t, t = 1, \ldots, N$, independent identically distributed, zero mean unit variance Gaussian noise, it is known that any rate triple $(R_0, R_1, R_2)$ satisfies the inequalities [13]:

\begin{align}
R_1 &\leq \frac{1}{2} \log \left( 1 + \kappa_1 (1 - \rho_{11}^2) \right), \\
R_2 &\leq \frac{1}{2} \log \left( 1 + \kappa_2 (1 - \rho_{22}^2) \right), \\
R_1 + R_2 &\leq \frac{1}{2} \log \left( 1 + \kappa_1 (1 - \rho_{11}^2) + \kappa_2 (1 - \rho_{22}^2) \right), \\
R_0 + R_1 + R_2 &\leq \frac{1}{2} \log \left( 1 + \kappa_1 + \kappa_2 + 2 \sqrt{\kappa_1 \kappa_2 \rho_1 \rho_2} \right)
\end{align}

where $\rho_i \in [0, 1], i = 1, 2$ are parameters to be optimized. For the special case without a common messages, i.e., messages $S_1$ and $S_2$ are independent, and hence $R_0$ is absent, then similar inequalities for any rate tuple $(R_1, R_2)$ are known [5].

The search for methods and algorithms to compute rates that lie on the rate regions of Gaussian MAC, $\mathcal{R}_{\text{MAC}}(\kappa_1, \kappa_2)$, $\mathcal{R}_{\text{MAC-C}}(\kappa_1, \kappa_2)$, to device coding schemes for Gaussian and digital messages, and to develop joint source-channel matching techniques, is currently very active. Early progress is reported in [17, 18, 19] for scalar-valued Gaussian MACs with feedback, without common message, and more recently in [14, 3], based on joint-source matching techniques. But the formula to compute rates that lie on $\mathcal{R}_{\text{MAC}}(\kappa_1, \kappa_2)$, $\mathcal{R}_{\text{MAC-C}}(\kappa_1, \kappa_2)$, for multiple-input multiple-output (MIMO) Gaussian channels with common message, is still missing.

The authors provide in this paper and in its companion paper [4], methods and algorithms to compute various rates that lie in the characterization of Gray and Wyner lossy rate region, for a tuple of multivariate correlated Gaussian random variables, using concepts of conditional independence, weak stochastic realization theory, and the canonical variable form of a tuple of Gaussian random variables. However, it is recognized that the concepts, methods and algorithms used in this paper are also of interest to other Gaussian multi-user problems of information theory, such as, the stationary memoryless Gaussian MAC channels with two channel inputs $(X_1, X_2)$. Hence, their presentation is pedagogical, while their applications to other fundamental problems of multi-user information theory, although discussed, it requires additional investigations.

1.2 Discussion of Method 1 and Method 2

Applications of Method 1.(a) and Method 2 to calculate rates that lie on the Gray-Wyner lossy rate region. Method 2, when applied to the Gray-Wyner lossy source coding for a simple network, is equivalent to a pre-processing at the encoder. It should be mentioned that if the joint distribution of $P_{Y_1, Y_2}$ is replaced by the joint probability mass function of random variables $(Y_1, Y_2)$ taking finite values, then this pre-processing shown in is equivalent to the invariance property of Wyner’s common information of $(Y_1, Y_2)$, that is discuss by Wyner throughout the seminal paper [29], and more specifically in Remark F.

Method 2 can then be used prior to applying Method 1.(a) to derive a parametrization of the family of jointly Gaussian probability distributions $P_{Y_1, Y_2}$, having the properties of Method 2.(a), (i) and (ii).

Method 1.(a) and Method 2 can be used to compute the marginal rate distortion functions $R_{Y_i}(\Delta_i), i = 1, 2$, of multivariate jointly Gaussian random variables, subject to square-error distortions, the conditional rate distortion functions $R_{Y_i|W}(\Delta_1)$ and $R_{Y_j|W}(\Delta_2), i = 1, 2$, and joint rate distortion function $R_{Y_1, Y_2}(\Delta_1, \Delta_2)$, all expressed in terms of the elements of the covariance matrix $Q_{cvf}$, and water-filling solutions of independent random variables.

Method 1.(a) and Method 2 can be used to construct the weak stochastic realizations of the reproductions $(\hat{Y}_1, \hat{Y}_2)$ of $(Y_1, Y_2)$, which achieve the rate distortion functions, $R_{Y_i}(\Delta_i), R_{Y_j|W}(\Delta_1), i = 1, 2$ and $R_{Y_i, Y_2}(\Delta_1, \Delta_2)$, in which the realizations are described by parallel additive Gaussian noise channels.

Method 1.(a) and method 2, can be used to compute a rate triple that lie on the Gray-Wyner rate region, $(R_0, R_1, R_2) \in \mathcal{R}_{\text{GW}}(\Delta_1, \Delta_2)$, to compute the minimum common message rate on the Gray-Wyner lossy rate
region $R_0 \in \mathcal{R}_{GW}(\Delta_1, \Delta_2)$, for the rate triple $(R_0, R_1, R_2) \in \mathcal{R}_{GW}(\Delta_1, \Delta_2)$, with sum rate $R_0 + R_1 + R_2$ that is equal to the joint rate distortion function of joint decoding, that is, the sum rate lies on the Pangloss Plane [10] of $\mathcal{R}_{GW}(\Delta_1, \Delta_2)$.

Method 1.(a) and method 2, can be used to identify the range of values of $(\Delta_1, \Delta_2)$ such that $C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2) = C_W(Y_1, Y_2) = I(Y_1, Y_2; W^\ast)$ is independent of the distortion levels $(\Delta_1, \Delta_2)$ and corresponds to Wyner’s lossy common information. Further, these methods can be used to determine the weak stochastic realizations of the random variables $(W^\ast, Y_1, Y_2, \hat{Y}_1, \hat{Y}_2)$, which achieve $C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2)$ and $C_W(Y_1, Y_2)$. These are given in the current paper (with the exception of $C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2)$).

Applications to the MACs with and without common message.
Method 1.(b) and method 2 can be used to compute the information measures that define the rate regions of the stationary Gaussian MACS, $\mathcal{R}_{MAC}(P_1, P_2)$ and $\mathcal{R}_{MAC-C}(P_1, P_2)$, and to determine weak stochastic realizations of the random variables, $(W, X_1, X_2, Y)$ which maximize information measures, in a similar spirit to the one discussed above for the Gray-Wyner source coding for a simple network.

Method 1 and Method 2 may be applicable to the development of coding schemes for the above MACs for Gaussian messages and digital messages, and for the development of joint source-channel matching techniques, such as the ones discussed in [17, 18, 19, 14, 3].

For other applications, such as database applications, secret key generation, and the lossy extension of the Gacs and Korner common information, the reader is advised to read Viswanatha, Akyol and Rose [26], specifically Section IV.

In general, the method 1 and method 2 and accompanied algorithms are useful in engineering problems of signal processing, communication and control, of processing a tuple of multivariate signals or with the analysis of a tuple of signals, which are Gaussian distributed. Based on experience, it is clear that any tuple of Gaussian distributed signals has relations that are often described as common signals, correlated signals, and private signals, not necessarily related to the information theoretic notions of common and private information. The common signals are often defined by those components of the signals that are identical. The correlated signals are often defined by those components of the signals that are correlated. The private signals are often defined by those signals of the signals that satisfy probabilistic independence of the signals. However, the decomposition of a tuple of multivariate Gaussian distributed signals, into their common signal, correlated signals, and private signals was not discussed in the literature so far, especially with respect to the computation of rates that lie the rate region, in problems of multi-user information theory, such as the ones discussed above.

This paper, however is devoted to the development of the methods, and their applications to the solution of the optimization problem $C(Y_1, Y_2) = C_W(Y_1, Y_2)$ for the region $\mathcal{D}_W$, for a tuple of multivariate random variables, $Y_i: \Omega \rightarrow \mathbb{R}^{P_i}, i = 1, 2$.

1.3 Structure of the Paper
The terminology of the paper is simplified. When a tuple of Gaussian random variables is mentioned it refers to a tuple of multivariate jointly-Gaussian random variables. The long expressions have been abbreviated to save on space of the paper. In the area of stochastic system theory and of control theory the expression common information is now also used and it refers to the identical components of a tuple of processes rather than to the concept of common information used in information theory. The authors prefer to use the expression for common information of stochastic system theory and therefore refer to Wyner’s common information with this expression.

Section 2 introduces the mathematical tools of the geometric approach to Gaussian random variables (Section 3.1), the weak stochastic realization of conditional independence (Section 2.3). Section 3 contains the problem statement, the solution procedure (Section 3.1), 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 3.2). Section 3.3 derives the formulae for Wyner’s lossy common information $C_W(Y_1, Y_2) = C(Y_1, Y_2)$. This section also provides the calculations of rate distortion functions $R_{Y_i}(\Delta_i), i = 1, 2,$
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 4 includes remarks on possible extensions.

Appendix B 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 solution procedure for two fundamental concepts of probability theory.

1) The transformation of tuple of Gaussian multivariate random variables \((Y_1, Y_2)\), via nonsingular transformations \((S_1, S_2)\) such that the transform random variables \(Y_1 \mapsto S_1Y_1, Y_2 \mapsto S_2Y_2\) are represented in their canonical variable form.

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

From the mathematical concepts (1) and (2) one can formulate Method 1.(a), Method 1.(b), and Method 2, discussed in Section 1.2.

These two mathematical concepts are introduced and discussed in Section 2.3 and Section 3.1.

### 2.1 Notation of Elements of Probability Theory

The notation used in the paper is briefly specified. Denote by \(\mathbb{Z}_+ = \{1, 2, \ldots\}\) the set of the integers and by \(\mathbb{N} = \{0, 1, 2, \ldots\}\) the set of the natural integers. For \(n \in \mathbb{Z}_+\) denote the following finite subsets of the above defined sets by \(\mathbb{Z}_n = \{1, 2, \ldots, n\}\) 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 \(m \times n\) matrices 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 \(u^T Qu \geq 0\) holds. Similar, \(Q > 0\) denotes that for all \(u \in \mathbb{R}^n \setminus \{0\}\), \(u^T Qu > 0\). The notation \(Q_1 \leq Q_2\) denotes that \(Q_2 - Q_1 \geq 0\).

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

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

\[
L_+(G) = \{X : \Omega \rightarrow [0, \infty) \mid X \text{ is } G\text{-measurable}\}.
\]

The tuple of sub-\(\sigma\)-algebras \(F_1, F_2 \subseteq \mathbb{F}\) is called independent if \(E[X_1X_2] = E[X_1]E[X_2]\) 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, and to
random variables taking values in arbitrary measurable spaces \((X, \mathcal{B}(X))\).

2.2 Geometric Approach of Gaussian Random Variables and Canonical Variable Form

The purpose of this section is to introduce concepts and results on the canonical variable form for a tuple of finite-dimensional Gaussian random variables, with emphasis on the geometric approach of Gaussian random variables, and to briefly discuss them in the context of multi-user communication. Thus the spaces generated by these random variables are the main objects of study while the actual random variables are only representations with respect to a basis. The concept of a canonical variable decomposition was introduced by H. Hotelling, [11].

The use of the geometric approach of two Gaussian random variables with respect to the computation of mutual information is further elaborated in subsequent parts of the paper.

The geometric objects are the \(\sigma\)-algebras of the Gaussian random variables. With respect to a basis, the probability distribution of a Gaussian random variable is characterized by its mean and its covariance matrix. Because the central theme of this paper is Gaussian random variables, the exposition is a combination of the geometric approach and the basis representations.

A \(\mathbb{R}^n\)-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 \(X : \Omega \rightarrow \mathbb{R}^n\) which is a random variable and such that the measure of this random variable equals a Gaussian measure described by its characteristic function,

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

Note that this definition includes the case when 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 \(X \in G(m_X, Q_X)\).

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

Any tuple of random variables \(X_1, \ldots, X_k\) is called jointly Gaussian if the vector \((X_1, X_2, \ldots, X_k)^T\) 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 \(Q_{Y_1,Y_2} = 0\).

The conditional expectation of a Gaussian random variable \(X : \Omega \rightarrow \mathbb{R}^n\) conditioned on the \(\sigma\)-algebra generated by another Gaussian random variable \(Y : \Omega \rightarrow \mathbb{R}^p\) with \((X, Y) \in G(m, Q_{X,Y})\) and with \(Q_Y > 0\) is, as is well known, again a Gaussian random variable with characteristic function,

\[
E[\exp(\imath u^T X)|F^Y] = \exp(\imath u^T E[X|F^Y] - \frac{1}{2} u^T Q_{X|Y} u), \quad \forall u \in \mathbb{R}^n, \quad \text{where},
\]

\[
Q_{X|Y} = \begin{pmatrix}
Q_X & Q_{X,Y} \\
Q_{X,Y}^T & Q_Y
\end{pmatrix},
\]

\[
E[X|F^Y] = m_X + Q_{X,Y} Q_Y^{-1}(Y - m_Y),
\]

\[
Q_{X|Y} = Q_X - Q_{X,Y} Q_Y^{-1} Q_{Y,X}.
\]
Next the geometric approach is defined. A random variable as defined above is always defined with respect to a particular basis of the linear space. The underlying geometric object of a Gaussian random variable \( Y : \Omega \to \mathbb{R}^p \) is the \( \sigma \)-algebra \( F^Y \). In this paper the authors prefer the \( \sigma \)-algebra as the geometric object rather than the linear space generated by the random variable. A basis transformation of such a random variable is then the transformation defined by a nonsingular matrix \( S \in \mathbb{R}^{p \times p} \), which implies \( F^Y = F^{SY} \) and the random variable \( SY \) is Gaussian.

Next consider a tuple of jointly Gaussian random variables \((Y_1, Y_2)\). A basis transformation of this tuple consists of a matrix \( S = \text{Block-diag}(S_1, S_2) \) with \( S_1, S_2 \) square and nonsingular matrices, which then implies that spaces satisfy \( F^{Y_1} = F^{S_1Y_1} \) and \( F^{Y_2} = F^{S_2Y_2} \), and \( S_1Y_1 \) and \( S_2Y_2 \) are Gaussian. This transformation introduces an equivalence relation on the representation of the tuple of random variables \((Y_1, Y_2)\). Further, it is known that mutual information, is invariant with respect to such transformations [20]. Thus one can speak about a canonical form for these spaces which is introduced next.

The concept of a canonical form has been defined in algebra, see [2]. It is defined for a set with an equivalence relation defined on it. A canonical form is then a subset of the considered set such that every element of the set is equivalent to a unique element of the subset, called then the canonical form of the element. The canonical variable form of a tuple of Gaussian random variables defined below is not a canonical form in the above defined sense. It is a form but there is a remaining invariance left in the form which is properly described. The expression canonical variable form will be used with the understanding that it is not a proper canonical form but with a remaining invariant.

Below the following problem is considered and the solution provided.

**Problem 2.1** Consider the tuple of jointly Gaussian random variables \( Y_1 : \Omega \to \mathbb{R}^{p_1} \) and \( Y_2 : \Omega \to \mathbb{R}^{p_2} \), with \((Y_1, Y_2) \in G(0, Q_{Y_1, Y_2})\). Determine a canonical form for the spaces \( F^{Y_1}, F^{Y_2} \) up to linear basis transformations.

The above problem has been posed and solved by H. Hotelling [11] as described below. Other references are the books [1, 9]. The reader may identify the elements of equations (5)-(9) with the next definition. Below the assumption is used that the variance of the random variable \( Y_1 \) is such that \( Q_{Y_1} > 0 \). It could be the case the random variable \( Y_1 \) has a variance which satisfies \( Q_{Y_1} \geq 0 \) but does not satisfy \( Q_{Y_1} > 0 \). In this case one can transform the basis to the following form,

\[
\left( \begin{array}{cc} Q_{Y_1} & 0 \\ 0 & 0 \end{array} \right) \in \mathbb{R}^{p_1 \times p_1},
\]

\[p_{11} = \text{rank}(Q_{Y_1}) = \text{rank}(Q_{11}), \ Q_{11} \in \mathbb{R}^{p_{11} \times p_{11}},
\]

With respect to the new basis, the second component of the transformed \( Y_1 \) is a constant which can be disregarded because it has no variance. Then \( \text{rank}(Q_{11}) = p_{11} \) implies that \( Q_{11} > 0 \) and one continues the procedure with the first component of the transformed variable \( Y_1 \). Of course, a corresponding procedure can be carried out for the random variable \( Y_2 \) if necessary.

**Definition 2.2** 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 \( i = 1, 2 \), \((Y_1, Y_2) \in G(0, Q_{Y_1, Y_2})\).

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
the bivariate Gaussian random variables, where,

\[
Q_{cvf} = \begin{pmatrix}
I_{p11} & 0 & 0 & I_{p21} & 0 & 0 \\
0 & I_{p12} & 0 & 0 & D & 0 \\
0 & 0 & I_{p13} & 0 & 0 & 0 \\
0 & D & 0 & 0 & I_{p22} & 0 \\
0 & 0 & 0 & 0 & 0 & I_{p23}
\end{pmatrix} \in \mathbb{R}^{p \times p},
\]  

(45)

\[p, p_1, p_2, p_{11}, p_{12}, p_{13}, p_{21}, p_{22}, p_{23} \in \mathbb{N},\]
\[p = p_1 + p_2, \quad p_1 = p_{11} + p_{12} + p_{13}, \quad p_2 = p_{21} + p_{22} + p_{23}, \quad p_{11} + p_{12} = p_{21}, \quad p_{12} = p_{22},\]
\[D = \text{Diag}(d_1, \ldots, d_{p_{12}}), \quad 1 \geq d_1 \geq d_2 \geq \ldots \geq d_{p_{12}} > 0,\]

(46)

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

(47)

One then says that \((Y_{11}, \ldots, Y_{1k_1}), (Y_{21}, \ldots, Y_{2k_2})\) are the canonical variables and \((d_1, \ldots, d_{k_1})\) the canonical correlation coefficients.

The next theorem states that there always exists a nonsingular basis transformation such that with respect to the new basis \((S_1 Y_1, S_2 Y_2) \in G(0, Q_{cvf})\) has the canonical variable form presented in Definition 2.2, and gives some of the properties.

**Theorem 2.3** Existence of the canonical variable form and properties.

Let \(Y_1 : \Omega \to \mathbb{R}^{p_1}\) and \(Y_2 : \Omega \to \mathbb{R}^{p_2}\) be jointly Gaussian random variables with \((Y_1, Y_2) \in G(0, Q_{cvf})\) and \(Q_{Y_i} > 0, \text{ for } i = 1, 2.\)

(a) Then there exists a nonsingular basis transformation

\[S = \text{Block-diag}(S_1, S_2) \in \mathbb{R}^{(p_1 + p_2) \times (p_1 + p_2)}\]

(50)

such that with respect to the new basis \((S_1 Y_1, S_2 Y_2) \in G(0, Q_{cvf})\) has the canonical variable form presented in Definition 2.2. It could be that one or more components of the canonical variable form are not present for a particular probability distribution.

(b) Assume that the pair \((Y_1, Y_2) \in G(0, Q_{cvf})\) is in canonical variable form. Then the basis transformation \(S = \text{Block-diag}(S_1, S_2)\) leaves the canonical variable form invariant if and only if when,

\[D = \text{Block-diag}(D_1, \ldots, D_m), \text{ with,}\]
\[D_i = \text{Diag}(d_{i1}, \ldots, d_{im}) = d_{ij}, \quad i \neq j \Rightarrow d_i \neq d_j, \text{ then,}\]

\[S_1 = \text{Block-diag}(S_{11}, \ldots, S_{1m}, S_{1m+1}),\]
\[S_2 = \text{Block-diag}(S_{21}, \ldots, S_{2m}, S_{2m+1}),\]

15
are both compatible with the block decomposition of $D$, then for all $i \in \mathbb{Z}_m$

$$S_{1,i}^T S_{1,i} = I, \quad S_{2,i}^T S_{2,i} = I, \quad D_i S_{2,i} = S_{1,i} D_i,$$

and $S_{1,m+1}^T S_{1,m+1} = I, \quad S_{2,m+1}^T S_{2,m+1} = I$. In case the canonical variables are all distinct, $(\Leftrightarrow (i \neq j \Rightarrow d_i \neq d_j))$, then

$$S_1 = \text{Block-diag}(S_{1,1}, S_{1,m+1}), \quad S_2 = \text{Block-diag}(S_{2,1}, S_{2,m+1}),$$

with $S_{1,1}, S_{2,1}$ sign matrices. A sign matrix is a diagonal matrix with on the diagonal only elements of the set $\{-1, +1\} \subset \mathbb{R}$.

**Proof** The result is due to H. Hotelling, [11]. A formulation may be found in the book [1]. \qed

Various properties of the components of of the canonical variable form are identified in the next proposition.

**Proposition 2.4 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

$$E[Y_{12} Y_{22}^T] = D = \text{Diag}(d_1, \ldots, d_{p_1}).$$

Note that the different components of $Y_{12}$ and of $Y_{22}$ are independent random variables; thus $Y_{12,i}$ and $Y_{12,j}$ are independent, and $Y_{22,i}$ and $Y_{22,j}$ are independent, and $Y_{12,i}$ and $Y_{22,j}$ are independent for all $i \neq j$; and that $Y_{12,i}$ and $Y_{22,j}$ for $j = 1, \ldots, p_1 = p_2$ 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 (45) of the canonical variable form. \qed

**Example 2.5** Consider Definition 2.2, for the bivariate Gaussian random variables (26) of Theorem 1.4, i.e., let $p_1 = p_2 = 1$. Then the following results hold for three cases.

(i) $E[Y_{12} Y_{22}] = d_1 \in (0,1)$. Then $(Y_1, Y_2) \in G(0, Q_{\text{cvf}})$ are described by $Y_1 = Y_{12}, Y_2 = Y_{22}$ and hence the components $Y_{11}, Y_{13}$ and $Y_{21}, Y_{23}$ of $Y_1$ and $Y_2$, respectively, are absent, and $E[Y_{12} Y_{22}] = d_1$. Hence,

$$Q_{\text{cvf}} = Q_{(Y_{12}, Y_{22})} = \begin{bmatrix} 1 & d_1 \\ d_1 & 1 \end{bmatrix}.$$  \hspace{1cm} (52)

(ii) $d_1 = 0$. Then $(Y_1, Y_2) \in G(0, Q_{\text{cvf}})$ are described by $Y_1 = Y_{13}, Y_2 = Y_{23}$ and hence the components $Y_{11}, Y_{12}$, $Y_{21}, Y_{22}$ are absent, and hence $Y_{13}, Y_{23}$ are independent Gaussian $G(0, 1)$.

(iii) $d = 1$. Then $(Y_1, Y_2) \in G(0, Q_{\text{cvf}})$ are described by $Y_1 = Y_{11}, Y_2 = Y_{21}, Y_{11} = Y_{21} - \text{a.s.}, Y_1, Y_2 \in G(0, 1)$, and hence the components $Y_{12}, Y_{13}, Y_{22}, Y_{23}$ are absent.

The previous example illustrates that, even for the bivariate Gaussian random variables (26) of Theorem 1.4, the main elements are the canonical correlation coefficients, and not the correlation coefficients.

Next the interpretation of the various components of the canonical variable form is defined, as in [25].
**Definition 2.6** 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 2.2. Call the various components as defined in the next table.

| Component | Interpretation |
|-----------|----------------|
| \(Y_{11} = Y_{21} - \text{a.s.}\) | 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 \(Y_2\) |
| \(Y_{21} = Y_{11} - \text{a.s.}\) | 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 \(Y_{11} = Y_{21} - \text{a.s.}\) the term identical information is used.

The next remark recalls the mutual information \(I(Y_1; Y_2)\) between two scalar-valued Gaussian random variables, expressed in terms of their correlation coefficient. Then \(I(Y_1; Y_2)\) is generalized to a tuple of multivariate random variables, expressed in terms of the canonical correlation coefficients.

**Remark 2.7** Consider two jointly Gaussian scalar-valued random variables \((Y_1, Y_2)\) with variance matrix defined by (26). Then mutual information, for any \(\rho \in [-1, 1]\), is given by

\[
I(Y_1; Y_2) = -\frac{1}{2} \ln \left(1 - \rho^2\right) = \begin{cases} 
+\infty, & \text{if } \rho \in \{-1, 1\}, \\
-\frac{1}{2} \ln \left(1 - \rho^2\right) \in (0, \infty), & \text{if } \rho \in (-1, 1), \rho \neq 0, \\
0, & \text{if } \rho = 0.
\end{cases}
\]

In particular, (53) shows that for Gaussian random variables \((Y_1, Y_2)\) then \(I(Y_1; Y_2) = +\infty\) if and only if the correlation coefficient takes the values \(\rho = +1, \rho = -1\). In the case, \(Y_1 = Y_2\) with probability one then \(I(Y_1; Y_2) = +\infty\).

Remark 2.7 is now extended to a general tuple of finite-dimensional Gaussian random variables \((Y_1, Y_2) \in G(0, Q_{Y_1,Y_2})\).

**Theorem 2.8** Consider a tuple of finite-dimensional Gaussian random variables \((Y_1, Y_2) \in G(0, Q_{Y_1,Y_2})\). Compute the canonical variable form of the tuple of Gaussian random variables according to Algorithm 2.10. 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,

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

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

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

**Proof of Theorem 2.8** By Theorem 2.3 and Proposition 2.4 there always exists a nonsingular basis transformation such that with respect to the new basis \((S_1Y_1, S_2Y_2) \in G(0, Q_{cvf})\) has the canonical variable form presented in Definition 2.2. Further, since the basis transformation is nonsingular then

\[
I(Y_1; Y_2) = I(S_1Y_1; S_2Y_2) = I(Y_{11}, Y_{12}, Y_{21}, Y_{22}, Y_{23}).
\]
If \( p_{11} = p_{21} > 0 \) then the components \( Y_{11}, Y_{21} \) are present and since \( Y_{11} = Y_{21} \) a.s., then the last right hand side entry in (54) is obtained. If \( 0 = p_{11} = p_{12} = p_{21} = p_{22} = p_{13} > 0, p_{23} > 0 \), then by (56), \( I(Y_1; Y_2) = I(Y_{13}; Y_{23}) = 0 \), by independence of \( Y_{13} \) and \( Y_{23} \). Hence, the first right hand side entry in (54) is obtained. Suppose \( p_{11} = p_{21} = 0 \), so that \( Y_{11}, Y_{21} \) are not present. By the chain rule of mutual information, then

\[
I(Y_1; Y_2) = I(Y_1; Y_{13}; Y_{23}) = I(Y_1; Y_{12}; Y_{22}, Y_{23}) + I(Y_{13}; Y_{22}, Y_{23})
\]

By Proposition 2.4, the variance matrix is

\[
Q_{Y_{12}, Y_{22}} = \begin{pmatrix} I & D \\ D & I \end{pmatrix},
\]

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

Hence, the second right hand side entry in (54) is obtained, and (55) follows as well.

The next remark describes the connection of the canonical variable form to the computation of the information definition of Gacs and Korner [7] common information, and Wyner’s information definition of common information \( C(Y_1, Y_2) \).

**Remark 2.9** A related notion to Wyner’s common information is the Gacs and Korner [7] definition of common randomness between a tuple of jointly independent and identically distributed random variables \( \{(Y_{1,i}, Y_{2,i}) : i = 1, \ldots, \} \). Let \( \{f_N^{(E)}(\cdot), g_N^{(E)}(\cdot) : N = 1, \ldots, \} \) be the encoder mappings which generate messages \( (S_1, S_2) = (s_1, s_2) \), each with values in a message set \( \mathcal{M} \), defined by

\[
f_N^{(E)}(Y_{1,N}) = S_1, \quad g_N^{(E)}(Y_{2,N}) = S_2.
\]

Define \( \epsilon_N = \text{Prob}\{S_1 \neq S_2\} \) and \( \rho_N = \frac{1}{N} H(S_1) \). Let \( \{(f_N^{(E)},g_N^{(E)}) : N = 1, \ldots, \} \) be the sequence of encoder mappings such that \( \lim_{N \to \infty} \epsilon_N = 0 \) and \( \rho = \sup_{N \to \infty} \frac{1}{N} H(S_1) \). Then it is possible to independently extract approximately \( \rho \) bits per symbol by observing either one of the two sequences \( Y_{1,N}, Y_{2,N} \), as \( N \to \infty \), Gacs and Korner [7] common randomness is defined by \( C_{\text{GK}}(Y_1, Y_2) = \sup \rho \), where the supremum is taken over all encoder sequences \( \{(f_N^{(E)},g_N^{(E)}) : N = 1, \ldots, \} \) such that \( \lim_{N \to \infty} \epsilon_N = 0 \).

It is known from [29], Remark E, that

\[
C_{\text{GK}}(Y_1, Y_2) \leq I(Y_1; Y_2) \leq C(Y_1, Y_2)
\]

Moreover,

\[
C_{\text{GK}}(Y_1, Y_2) = C(Y_1, Y_2) = I(Y_1; Y_2) \quad \text{if and only if it is possible to represent}
\]

\[
Y_1 = (V, Y_{1}', Y_{1}'), \quad Y_2 = (V, Y_{2}', Y_{2}''), \quad \text{where } Y_{1}' \text{ and } Y_{2}'' \text{ are conditionally independent given } V.
\]

The above condition is equivalent to the condition that the joint probability mass function of random variables \( X, Y \) can be expressed, with a relabeling of the rows and columns, in terms of a new joint probability mass function, with specific block elements on its diagonal and zero on its off diagonal elements. Such a
transformation is analogous to a pre-processing, by an invariance transformation that leaves Wyner’s common information of $(Y_1, Y_2)$, unchanged.

In the special case discussed by Wyner [29] throughout the paper,

$$Y_1 = (V, Y'_1), \quad Y_2 = (V, Y'_2), \text{ where } Y'_1, Y'_2, V \text{ are independent, then}$$

$$C_{GK}(Y_1, Y_2) = C(Y_1, Y_2) = H(V)$$

(68)  \hspace{1cm} (69)

where $H(V)$ is the entropy of $V$.

Next, two special cases of the canonical variable form of Definition 2.2, are discussed.

Case (1). If $Y_{11}$ is absent then $Y_{12}$ is absent, and vice-versa, that is, $p_{11} = p_{21} = 0$, and then

$$S_1 Y_1 = (V_1, Y'_1) = (Y_{12}, Y_{13}), \quad S_2 Y_2 = (V_2, Y'_2) = (Y_{22}, Y_{23}),$$

$$Y_{12} \text{ and } Y_{22} \text{ are correlated and each of these has independent components},$$

$$Y_{12}, Y_{23}, Y_{13} \text{ are independent and }, Y_{12}, Y_{13}, Y_{23} \text{ are independent},$$

$$E[Y_{12} Y_{22}] = D.$$  \hspace{1cm} (70)  \hspace{1cm} (71)  \hspace{1cm} (72)  \hspace{1cm} (73)  \hspace{1cm} (74)

Case (2). If $Y_{12}, Y_{22}$ are absent, that is, $p_{12} = p_{22} = 0$, then

$$S_1 Y_1 = (V_1, Y'_1) = (Y_{11}, Y_{13}), \quad S_2 Y_2 = (V_2, Y'_2) = (Y_{21}, Y_{23}),$$

$$Y_{11} = Y_{21} - a.s. \iff V_1 = V_2 - a.s.$$  \hspace{1cm} (75)  \hspace{1cm} (76)

$$Y_{11}, Y_{13}, Y_{23} \text{ are independent and each of these has independent components},$$

$$Y_{21}, Y_{23}, Y_{13} \text{ are independent and each of these has independent components}.$$  \hspace{1cm} (77)  \hspace{1cm} (78)

Hence, Case (2) is the analog of (68), i.e., special case discussed by Wyner [29], hence it follows that $C_{GK}(Y_1, Y_2) = C(Y_1, Y_2) = H(Y_{11}) = H(Y_{21})$, and $Y_{11} = Y_{21} - a.s. \in G(0, I_{p_{11}})$.

In Section 2.3, it will be shown how to construct a probability measure that carries an auxiliary Gaussian random variable $W$, such that conditional independence holds: $P_{S_1 Y_1, S_2 Y_2|W} = P_{S_1 Y_1|W} P_{S_2 Y_2|W}$, from which $C(Y_1, Y_2) = C(S_1 Y_1, S_2 Y_2)$ can be computed.

It is remarked that the canonical variable form is directly applicable to the lossy extension of Gacs and Korner common information derived by Viswanatha, Akyol and Rose [26] in Section IV. Specifically, the information theoretic characterization of Theorem 3 in [26].

The algorithm that generates $Q_{CVF}$ is presented below.

**Algorithm 2.10 Transformation of a variance matrix to its canonical variable form.**

*Data: $p_1, p_2 \in Z^+$, $Q \in \mathbb{R}^{(p_1+p_2) \times (p_1+p_2)}$, satisfying $Q = Q^T > 0$, with decomposition $Q = \left( \begin{array}{cc} Q_{11} & \bar{Q}_{12} \\ \bar{Q}_{12}^T & Q_{22} \end{array} \right), \quad Q_{11} \in \mathbb{R}^{p_1 \times p_1}, \quad Q_{22} \in \mathbb{R}^{p_2 \times p_2}, \quad Q_{12} \in \mathbb{R}^{p_1 \times p_2}.$*

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

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

and $U_2, D_2$ satisfying corresponding conditions.

---

It is noted that $Q > 0$ implies $Q_{11} > 0, Q_{22} > 0$. 

19
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
\[
D_3 = \begin{pmatrix}
I_{p_{11}} & 0 & 0 \\
0 & D_4 & 0 \\
0 & 0 & 0
\end{pmatrix} \in \mathbb{R}^{p_1 \times p_2},
\]
\[ D_4 = \text{Diag}(d_{4,1}, ..., d_{4,p_{12}}) \in \mathbb{R}^{p_{12} \times p_{12}}, \quad 1 > d_{4,1} \geq d_{4,2} \geq \ldots \geq d_{4,p_{12}} > 0. \]

3. Compute the new variance matrix according to,
\[
Q_{\text{cvf}} = \begin{pmatrix}
I_{p_1} & D_3 \\
D_3 & I_{p_2}
\end{pmatrix}.
\]

4. The transformation to the canonical variable representation
\[
(Y_1 \mapsto S_1 Y_1, Y_2 \mapsto S_2 Y_2) \text{ is then}
\]
\[ S_1 = U_1^T D_1^{-\frac{1}{2}} U_1^T, \quad S_2 = U_4^T D_2^{-\frac{1}{2}} U_2^T. \]

**Remark 2.11** The material discussed in Section 1.2, related to Method 2 are applications 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 its realization. Conversely, in channel capacity problems the channel distribution is fixed, while the optimal channel input distribution needs to be found.
In source coding problems one application of Method 2 is the pre-encoder, which is constructed by invoking Algorithm 2.10.

### 2.3 Weak Realization of Conditional Independent Gaussian Random Variables

The purpose of this section is to introduce concepts and results as stated in Section 2, under 2), on conditional independence of a tuple of finite-dimensional Gaussian random variables \((Y_1, Y_2)\), conditioned on another Gaussian random variable \(W\), and the weak realization of these random variables. Then to briefly discuss them in the context of multi-user communication.

The concept of conditional independence and its connection to the weak realization problem is stated below. It is shown in the companion paper [4] that the rate region \(R_{GW}(\Delta_1, \Delta_2)\) is parametrized by a random variable \(W\) that makes \(Y_1\) and \(Y_2\) conditionally independent. The fact that, for the calculation of \(C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2)\) and \(C_W(Y_1, Y_2)\), it is sufficient to consider only random variables \(W\) that make \(Y_1\) and \(Y_2\) conditionally independent is discussed in Section 3.

However, the current paper is only concerned with the use of the material of this section (and last section) to compute the information quantity \(C(Y_1, Y_2)\), which can be shown to be Wyner’s lossy common information, for certain distortion regions.

**Definition 2.12** Conditional independence.
Consider a probability space \((\Omega, F, P)\) 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,
\[
E[Y_1 Y_2 | G] = E[Y_1 | G] E[Y_2 | G], \quad \forall \ Y_1 \in L_+(F_1), \ Y_2 \in L_+(F_2).
\]
Denote this property by \((F_1, F_2 | G) \in CI\).
Examples of triples of sub-σ-algebras which are conditional independent are (1) \((F_1, F_2|F_1) \in CI\); 
(2) \((F_1, F_2,F_2) \in CI\); (3) \((F_1,F_2|\{\emptyset, \Omega\}) \in CI\) if \(F_1\) and \(F_2\) are independent.

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 2.13**  
**Minimality of conditional independence of Gaussian random variables.**

Consider three random variables, \(Y_i : \Omega \to \mathbb{R}^{p_i}\) 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) \((F^{Y_1}, F^{Y_2}|F^X) \in CI\) and 
(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 and 
(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 conditional independence of tuple of Gaussian random variables by a third Gaussian random variable. This condition is expressed in terms of parametrizing the variance matrix of the tuple as presented in the next proposition.

**Proposition 2.14**  
[23, Prop. 3.4]  
**Equivalent condition for conditional independence of tuple of Gaussian random variables.**

Consider a triple of jointly Gaussian random variables denoted as \((Y_1,Y_2,X) \in G(0,Q)\) with \(Q > 0\). 
This triple is Gaussian conditionally independent if and only if
\[Q_{Y_1,Y_2} = Q_{Y_1,X}Q_X^{-1}Q_{X,Y_2} .\] (80)

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

As mentioned earlier, the calculation of \(C_{GW}(Y_1,Y_2;\Delta_1,\Delta_2)\) and \(C_W(Y_1,Y_2), C(Y_1,Y_2)\) is directly related to the solution of the weak stochastic realization problem of random variables \((W,Y_1,Y_2,\hat{Y}_1,\hat{Y}_2)\), that achieve the rate distortion functions \(R_y(\Delta_1), R_{y|w}(\Delta_1), i = 1, 2,\) and \(R_{y_1,y_2}(\Delta_1,\Delta_2)\), that is, the construction of the joint distribution \(P_{W,Y_1,Y_2,\hat{Y}_1,\hat{Y}_2}\) with marginal the source distribution \(P_{Y_1,Y_2}\), which achieves these rate distortion functions. These rate distortion functions are much easier to calculate, if a tuple of Gaussian random variables \(Y_i : \Omega \to \mathbb{R}^{p_i}\) with \(Q_{Y_i} > 0\), for \(i = 1, 2, (Y_1,Y_2) \in G(0,Q_{Y_1,Y_2})\), is transformed to the canonical variable form of Definition 2.2, i.e. \((Y_1,Y_2) \in G(0,Q_{\text{CVF}})\), and then Gray and Wyner’s lossy rate region \(r_{GW}(\Delta_1,\Delta_2)\) is parametrized by the random variable \(W\) that makes \(Y_1\) and \(Y_2\) conditionally independent.

**Problem 2.15**  
**Weak and strong stochastic realizations.**

(a) **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) \(G(0,Q_1)|_{\mathbb{R}^{p_1+p_2}} = G(0,Q_0)\) and 
(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.

(b) **Strong stochastic realization problem of a Gaussian random variable.** Consider a probability space \((\Omega,F,\mathbb{P})\) with defined on it a tuple of random variables denoted \((Y_1,Y_2)\) with the dimensions of \(p_1,p_2\) respectively. 
Construct all random variables \(X : \Omega \to \mathbb{R}^n\) such that
(1) \((Y_1, Y_2|X) \in \text{CIG}_{\text{min}}\) and
(2) \(F^X \subseteq (F^{Y_1} \lor F^{Y_2})\).

Note that for the weak stochastic realization Problem 2.15, (a) one is asked to construct a measure while for the strong stochastic realization Problem 2.15, (b) one is asked to construct a random variable on a prespecified probability space already containing the random variables \(Y_1\) and \(Y_2\).

**Remark 2.16** For the calculation of rates that lie on the Gray-Wyner lossy rate region and of rates that lie on the rate region of the stationary Gaussian MACs, the concept of weak stochastic realization is required. It should be mentioned that Proposition 2.14 is directly applicable to stationary and memoryless Gaussian MACs, described in Section 1, because the rate regions \(R_{\text{MAC}}(\kappa_1, \kappa_2)\) and \(R_{\text{MAC}-C}(\kappa_1, \kappa_2)\) are defined with respect to three jointly Gaussian random variables \((X_1, X_2, W)\) such that \(W\) makes \(X_1\) and \(X_2\) conditionally independent. That is, \(P_{X_1, X_2, W} = P_{X_1|W}P_{X_2|W}P_W\). In such applications, Proposition 2.14 is a characterization of conditional independence, while one is asked to solve the weak stochastic realization Problem 2.15, subject to the average power constraints.

Further to Remark 2.16, 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 2.17** Minimality of weak stochastic realization of Gaussian random variables. Consider a Gaussian measure \(G_0 = G_0(0, Q_{(Y_1, Y_2)})\) 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)} = \begin{pmatrix}
Q_{Y_1}
& Q_{Y_1, Y_2}

Q_{Y_1, Y_2}^T
& Q_{Y_2}
\end{pmatrix},
\]

where \(Q_{Y_1} > 0\), \(Q_{Y_2} > 0\).

(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 \(G_1 = G_1\) if there exists an integer \(n \in \mathbb{Z}_+\) such that the Gaussian measure \(G_1\) is defined on the space \((\mathbb{R}^{p_1} \times \mathbb{R}^{p_2} \times \mathbb{R}^n, B(\mathbb{R}^{p_1}) \otimes B(\mathbb{R}^{p_2}) \otimes B(\mathbb{R}^n))\), 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) \(Q_X > 0\); and

(3) conditional independence holds: \(P_{Y_1, Y_2|X} = P_{Y_1|X}P_{Y_2|X}\), where these are Gaussian measures, with means which are linear functions of the random variable \(X\) and deterministic variance matrices, i.e., similar to (42)-(44) with appropriate changes.

(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,

\[(Y_1, Y_2, X, V_1, V_2), \quad p_{V_1}, p_{V_2} \in \mathbb{Z}_+, \quad p_{V_1} \geq p_1, \quad p_{V_2} \geq p_2;\]

\[Y_1 : \Omega \to \mathbb{R}^{p_1}, \quad Y_2 : \Omega \to \mathbb{R}^{p_2}, \quad V_1 : \Omega \to \mathbb{R}^{p_1}, \quad V_2 : \Omega \to \mathbb{R}^{p_2}, \quad X : \Omega \to \mathbb{R}^n,\]

\[(V_1, V_2, X) \in G, \text{ and these are zero mean independent random variables,}\]

\[Q_{V_1} > 0, \quad Q_{V_2} > 0, \quad Q_X > 0;\]

\[C_1 \in \mathbb{R}^{p_1 \times n}, \quad C_2 \in \mathbb{R}^{p_2 \times n}, \quad N_1 \in \mathbb{R}^{p_1 \times p_{V_1}}, \quad N_2 \in \mathbb{R}^{p_2 \times p_{V_2}},\]

\[Y_1 = C_1 X + N_1 V_1, \]

\[Y_2 = C_2 X + N_2 V_2,\]

\[Q_{Y_1} = C_1 Q_X C_1^T + N_1 Q_{V_1} N_1^T,\]

\[Q_{Y_2} = C_2 Q_X C_2^T + N_2 Q_{V_2} N_2^T,\]

\[Q_{Y_1, Y_2} = C_1 Q_X C_1^T,\]

\[G_0(0, Q_{Y_1, Y_2}) = G_1 |_{\mathbb{R}^{p_1} \times \mathbb{R}^{p_2}}.\]

From the assumptions 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 next proposition shows equivalence of weak Gaussian stochastic realizations of Definition 2.17.(a), (b) to Definition 2.17.(c), (d), respectively.

**Proposition 2.18** Consider the setting of Definition 2.17 with \((Y_1, Y_2) \in G(0, Q_{(Y_1, Y_2)})\) with the representation of \((83), (84)\).

\[(a)\] A weak Gaussian stochastic realization in terms of a measure \(P_1 = G_1\) as defined in Definition 2.17.(a) is equivalent with a Gaussian random variable representation of Definition 2.17.(c).

\[(b)\] The minimal weak Gaussian stochastic realization of Definition 2.17.(b) is equivalent to a minimal weak Gaussian random variable representation of Definition 2.17.(d).

**Proof** \((a)\) (\(\iff\)) From the assumption that each of the following three random variables \(V_1, \ V_2, \ X\) is Gaussian and that they are independent, follows that \((Y_1, Y_2, X)\) are jointly Gaussian. From the assumptions follows that \(G_0 = G_1 |_{\mathbb{R}^{p_1} \times \mathbb{R}^{p_2}},\) i.e., the specified restriction of \(G_1\) is \(G_0\).

That conditional independence holds for \((Y_1, Y_2 | X) \in \text{CIG}\) follows from,

\[
E[Y_1 | X] = E[C_1 X + N_1 V_1 | X] = C_1 X + N_1 E[V_1] = C_1 X = Q_{Y_1, X} Q_X^{-1} X,
\]

\[C_1 = Q_{Y_1, X} Q_X^{-1}, \quad \text{because} \quad Q_X > 0,\]

where the well known formula is used for conditional expectation of two jointly Gaussian random variables for \(E[Y_1 | X] = Q_{Y_1, X} Q_X^{-1} X\). Similarly, \(C_2 = Q_{Y_2, X} Q_X^{-1}\). From this follows that,

\[
Q_{Y_1, X} Q_X^{-1} Q_{Y_2, X}^T = C_1 Q_X C_1^T = Q_{Y_1, Y_2},
\]

where the last equality follows from equation \((87)\). From the obtained equality follows that the conditional independence of \((Y_1, Y_2 | X) \in \text{CIG}\) holds.
(a) \(\implies\) Consider a triple of jointly Gaussian random variables \((Y_1, Y_2, X)\) such that conditional independence holds, \([Y_1, Y_2|X] \in CIG\). From the conditional independence and Proposition 2.14 follows that, \(Q_{Y_1,Y_2} = Q_{X|Y_1}Q^{-1}_{Y_1|X}Q_{X,Y_1}^T\).

Because these random variables are jointly Gaussian the following conditional expectations have the form

\[
\begin{align*}
\mathbb{E}[Y_1|X] &= Q_{Y_1|X}Q^{-1}_{X}X = C_1X, \quad C_1 = Q_{Y_1|X}Q^{-1}_{X}, \\
N_1V_1 &= Y_1 - C_1X, \\
\mathbb{E}[Y_2|X] &= Q_{Y_2|X}Q^{-1}_{X}X = C_2X, \quad C_2 = Q_{Y_2|X}Q^{-1}_{X}, \\
N_2V_2 &= Y_2 - C_2X,
\end{align*}
\]

(91) (92) (93) (94)

by the formulas for \(C_1, C_2\), and the one for conditional independence. Note that the above also establishes that \((X, V_1, V_2)\) are independent random variables. Thus one obtains the representation of Definition 2.17.\(\implies\) By Proposition 2.14 (shown in [23]) a weak Gaussian realization is minimal if and only if \(\text{rank}(Q_{Y_1,Y_2}) = n\). From the proof of (a) \(\iff\) follows that,

\[
Q_{Y_1,Y_2} = Q_{X|Y_1}Q^{-1}_{Y_1|X}Q^T_{X,Y_1} = C_1Q_XC^T_2.
\]

From the above quoted characterization of minimality follows that,

\[
n = \text{rank}(Q_{Y_1,Y_2}) = \text{rank}(C_1Q_XC^T_2),
\]

(99)

hence that \(\text{rank}(C_1) = n = \text{rank}(C_2)\).

(b) \(\iff\) From the rank assumption and the obtained relation follows that,

\[
n = \text{rank}(C_1Q_XC^T_2) = \text{rank}(Q_{Y_1,Y_2}),
\]

(100)

thus minimality holds. \(\square\)

### 3 Calculation of Wyner’s Lossy Common Information

For the calculation of \(C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2)\) via Theorem 1.2, and \(C_W(Y_1, Y_2)\) via Theorem 1.3, it is sufficient to impose the conditional independence \((F_{Y_1}, F_{Y_2}|F^W) \in CI\), that is, \(P_{Y_1,Y_2|W} = P_{Y_1|W}P_{Y_2|W}\). This is due to the following.

1. The well-known inequality

\[
I(Y_1, Y_2; W) = H(Y_1, Y_2) - H(Y_1, Y_2|W) \geq H(Y_1, Y_2) - H(Y_1|W) - H(Y_2|W)
\]

(101)

which is achieved if \(P_{Y_1,Y_2|W} = P_{Y_1|W}P_{Y_2|W}\), and

2. a necessary condition for the equality constraint (22) to hold is (see Appendix B in [31]) is

\[
R_{Y_1,Y_2|W}(\Delta_1, \Delta_2) = R_{Y_1|W}(\Delta_1) + R_{Y_2|W}(\Delta_2).
\]

(102)

Further, a sufficient condition for (102) to hold is the conditional independence condition [31]: \(P_{Y_1,Y_2|W} = P_{Y_1|W}P_{Y_2|W}\).

Hence, a sufficient condition for any rate \((R_0, R_1, R_2) \in R_{GW}(\Delta_1, \Delta_2)\) to lie on the Pangloss plane is the conditional independence.

Further,

3. for jointly Gaussian random variables \((Y_1, Y_2)\) with square-error distortion, then by the maximum entropy principle the optimal joint distribution \(P_{Y_1,Y_2;\hat{Y}_1,\hat{Y}_2,W}\) of the optimization problem \(C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2)\) is confined to a jointly Gaussian distribution.

Thus one arrives at the definition of Wyner’s lossy common information given below.
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 \((Y_1, Y_2) \in G(0, \mathcal{Q}_{(Y_1, Y_2)}),\) and square error distortion functions between \((Y_1, Y_2),\) and its reproduction \((\hat{Y}_1, \hat{Y}_2),\) given by

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

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

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

\[
C(Y_1, Y_2) = \inf_{W: \Omega \rightarrow \mathbb{R}^{n^*}, (F_{Y_1}, F_{Y_2}|F_W) \in \text{CIG}} I(Y_1, Y_2; W) \in [0, \infty].
\]

Call any random variable \(W\) as defined above such that \((Y_1, Y_2, W) \in G\) and \((F_{Y_1}, F_{Y_2}|F_W) \in \text{CIG}\) a state of the tuple \((Y_1, Y_2).\)

If there exists a random variable \(W^* : \Omega \rightarrow \mathbb{R}^{n^*}\) with \(n^* \in \mathbb{Z}_+ = \{1, 2, \ldots, \}\) 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 lossy common information (operational definition) is defined for a strictly positive 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 \((\Delta_1, \Delta_2) \in D_W = \{ (\Delta_1, \Delta_2) \in [0, \infty] \times [0, \infty] : 0 \leq (\Delta_1, \Delta_2) \leq \gamma \}\)

provided identity (22) 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 (25) is decomposed into the characterization of \(C(Y_1, Y_2)\) such that identity (22) 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 \(C_{GW}(Y_1, Y_2; \Delta_1, \Delta_2) = C_W(Y_1, Y_2) = C(Y_1, Y_2),\) i.e., it is constant for \((\Delta_1, \Delta_2) \in D_W.\) An alternative derivation of this decomposition is provided in [4].

The current paper is mainly devoted to the optimization problem \(C(Y_1, Y_2)\) defined by (104), in terms of the joint distribution \(P_{Y_1,Y_2,W}\) which achieves the infimum in (104). It is then shown that the weak realization of the tuple \((Y_1, Y_2),\) expressed in terms of the random variable \(W^*\), ensures the validity of identity (22) of Theorem 1.2, with \(W\) replaced by \(W^*\). Hence, by Theorem 1.3 there exists a \(D_W,\) such that Wyner's lossy common information (operational definition) is computed from (105).

Alternatively, the reader can verify that the weak realization of the tuple \((Y_1, Y_2),\) expressed in terms of the random variable \(W^*\), which is obtained from the characterization of Wyner's common information \(C(Y_1, Y_2)\) defined by (104), ensures that all conditions of Theorem 1 in [26] are satisfied.

It should be mentioned that \(C(Y_1, Y_2)\) is allowed to take the value \(+\infty,\) because mutual information \(I(Y_1, Y_2 : W)\) is defined via relative entropy which is allowed to take the value \(+\infty.\) Further, if one evaluates the definition for two identical random variables then one arrives at the conclusion that the answer for \(I(Y_1, Y_2 : W)\) is best the value \(+\infty.\) Because mutual information takes only values in the positive real numbers, then \(C(Y_1, Y_2)\) has to be positive.

In general there are many random variables \(W\) which make the tuple of random variables \((Y_1, Y_2)\) conditionally independent. Therefore one wants to infimize the mutual information overall random variables \(W\) which make \(Y_1\) and \(Y_2\) conditionally independent.

In the next sections the Wyner lossy common information of finite-dimensional Gaussian random variables will be computed.
Figure 2: Weak stochastic realization of \( (Y_1,i,Y_2,i) \sim P_{Y_1,Y_2,i}, i = 1,\ldots,N \) and \( (\hat{Y}_1,i,\hat{Y}_2,i), i = 1,\ldots,N \) at the encoder and decoder with respect to the common and private random variables \( (W_N, Z_1^N, Z_2^N), (\hat{W}_N, \hat{Z}_1^N, \hat{Z}_2^N) \).

### 3.1 Solution Procedure

From the Definition 3.1 follows directly the procedure to compute Wyner’s common information (information definition) \( C(Y_1,Y_2) \).

**Procedure 3.2** Computation of Wyner’s common information \( C(Y_1,Y_2) \).

1. Determine a parametrization of all random variables \( W \) which make the two components of the tuple \( (Y_1,Y_2) \) conditionally independent, thus so that \( (F_{Y_1,Y_2}|F_W) \in \text{CIG} \), according to the weak stochastic realization of Problem 2.15(a).

2. Solve Wyner’s common information problem (information definition)

\[
C(Y_1,Y_2) = \inf_{(Y_1,Y_2,W) \in G, (Y_1,Y_2|W) \in \text{CIG}} I(Y_1,Y_2;W)
\]  

(106)

over the set of all measures determined in Step 1.

In Procedure 3.2, use is made of the concepts of Sections 2.2, 2.3, and of a result that is found in a paper co-authored by the second-named author [23].

### 3.2 Preliminary Characterization of Conditional Independence

Consider the pre-encoder of Fig. 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 lossy rate region is described by a single coding scheme that uses the auxiliary random variable \( W \), which is common to both \( Y_1, Y_2 \). Below this engineering interpretation is further detailed in terms of the mathematical framework of weak stochastic realization such that \( (F_{\hat{Y}_1,\hat{Y}_2}|F_W) \in \text{CIG} \).

**Definition 3.3** 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 \( Y_i : \Omega \to \mathbb{R}^{p_i} \) for \( i = 1,2 \). Take a jointly Gaussian measure \( G(0,Q_{(Y_1,Y_2,W)}) \) for the triple \( (Y_1,Y_2,W) \), \( W : \Omega \to \mathbb{R}^n, W \in G(0,Q_W) \), such that the marginal measure on \( (Y_1,Y_2) \) equals the considered measure, and the conditional
independence holds, \((F_{Y_1}, F_{Y_2} | F^W) \in \text{CIG}\). Denote the joint measure by \((Y_1, Y_2, W) \in G(0,Q_{(Y_1,Y_2,W)})\) with
\[
Q_{(Y_1,Y_2,W)} = \begin{pmatrix}
Q_{Y_1} & Q_{Y_1,Y_2} & Q_{Y_1,W} \\
Q_{Y_1,Y_2}^T & Q_{Y_2} & Q_{Y_2,W} \\
Q_{Y_1,W}^T & Q_{Y_2,W}^T & Q_W
\end{pmatrix}.
\]

In the two following sections it will be shown how to construct such a random variable \(W\) in a number of cases.

The algorithm that generates the joint measure by \((Y_1, Y_2, W) \in G(0,Q_{(Y_1,Y_2,W)})\) via weak stochastic realization is given below.

**Algorithm 3.4** Consider the model of a tuple of Gaussian random variables of Def. 3.3. Assume that \(Q_W > 0\).

1. At the encoder, compute first the variables,
\[
Z_1 = Y_1 - E[Y_1 | F^W] = Y_1 - Q_{Y_1,W}Q_W^{-1}W, \tag{107}
\]
\[
Z_2 = Y_2 - E[Y_2 | F^W] = Y_2 - Q_{Y_2,W}Q_W^{-1}W, \tag{108}
\]

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,
\[
Y_1 = Q_{Y_1,W}Q_W^{-1}W + Z_1, \quad Y_2 = Q_{Y_2,W}Q_W^{-1}W + Z_2. \tag{109}
\]

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

**Proposition 3.5** Consider the model of a tuple of Gaussian random variables of Def. 3.3.

(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** This is a specific application of Proposition 2.18.

For the definition of \(C(Y_1,Y_2)\), use is made of the construction of the actual family of measures such that \((Y_1,Y_2|W) \in \text{CIG}\) holds, and the weak stochastic realization. These are presented in Theorem 3.8 and Corollary 3.10.

### 3.3 The Expression of Wyner’s Lossy Common Information of Multivariate Gaussian Random Variables

The reader is at this point is assumed to be familiar with the concepts of a Gaussian random variable, of the canonical variable form of a tuple of jointly-Gaussian random variables (of identical, correlated, and private parts), and of conditionally-independent Gaussian random variables, as presented in Section 2.2 and Section 2.3.

First, the calculation of the formulae for \(C(Y_1,Y_2)\) is stated in the form of an algorithm. The algorithm makes use of the canonical variable form of Section 2.2.
Algorithm 3.6 Consider a tuple of finite-dimensional Gaussian random variables \((Y_1, Y_2) \in G(0, Q_{Y_1Y_2})\) as defined in Section 3.1.

1. Compute the canonical variable form of the tuple of Gaussian random variables according to Algorithm 2.10. 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 singular values \(d_i \in (0, 1)\) for \(i = 1, \ldots, n\).

2. Compute the information quantity \(C(Y_1, Y_2)\) according to the formula,
\[
C(Y_1, Y_2) = \begin{cases} 
0, & \text{if } 0 = p_{11} = p_{12} = p_{21} = p_{22}, p_{13} > 0, p_{23} > 0, \\
\frac{1}{2} \sum_{i=1}^{n} \ln \left(\frac{1 + d_i}{1 - d_i}\right), & \text{if } 0 = p_{11} = p_{12}, p_{12} = p_{22} > 0, p_{13} \geq 0, p_{23} \geq 0, \\
& \text{if } p_{11} = p_{21} > 0, p_{12} = p_{22} \geq 0, p_{23} \geq 0, p_{23} \geq 0.
\end{cases}
\]

(110)

Theorem 3.7 Consider a tuple of Gaussian random variables. Algorithm 3.6 is correct and produces the quantity \(C(Y_1, Y_2)\) defined by (104) of the tuple \((Y_1, Y_2)\).

Proof The proof of the theorem is provided in Section 3.8 after several special cases of the result have been derived, i.e., when \(p_{11} = p_{21} = 0\), Section 3.4. \(\square\)

The computation of the information quantity \(C(Y_1, Y_2)\) is structured by the concepts of identical, correlated, and private components of the two vectors considered; see Section 3.1 for the definitions of these concepts. The quantity \(C(Y_1, Y_2)\)

(i) in the first case of equation (110) covers the case in which the random variables \((Y_1, Y_2)\) are independent random variables and there are neither identical nor correlated components,

(ii) in the second case of equation (110) covers the case in which there is no identical component, but there are nontrivial correlated components, and possibly, correlated and independent components.

(iii) in the last case of equation (110) covers the case when there is a nontrivial identical component and, successively the reader will be shown the computation of information quantity \(C(Y_1; Y_2)\) for:

1. the correlated components;
2. the private components; and
3. the identical information components.

The general case is then a combination of the above three special cases.

3.4 Wyner’s Information Common Information of Correlated Gaussian Vectors

This section is devoted to the application of Method 1 (a) and Method 2 to calculate \(C(Y_1, Y_2)\), and to present the weak stochastic realization of \((Y_1, Y_2, W^*)\) that achieve this. The identification of the random variable \(W^*\), such that \((Y_1, Y_2, W^*)\) achieves \(C(Y_1, Y_2)\), is given in the next theorem. The theorem utilizes the parametrization of the family of Gaussian probability distributions

\[
P_{Y_1Y_2W} = \left\{ P_{Y_1Y_2W} : \begin{array}{l}
P_{Y_1W}(y_1|w) = P_{y_1|w}(y_1)P_{y_2|w}(y_2), \\
P_{Y_1Y_2}(y_1, y_2) = P_{Y_1}(y_1)P_{Y_2}(y_2), \\
Y_1, Y_2, W \text{ jointly Gaussian} \end{array} \right\}.
\]

(111)

A subset of the set \(P_{Y_1Y_2W}^\text{CIG}\) is the set of distributions \(P_{Y_1Y_2W}^\text{CIG}\), with the additional constraint that the dimension of the random variable \(W\) is minimal while all other conditions hold, defined by

\[
P_{Y_1Y_2W}^\text{CIG} = \left\{ P_{Y_1Y_2W} : \begin{array}{l}
P_{Y_1Y_2W} \in P_{Y_1Y_2W}^\text{CIG}, \\
Y_1, Y_2W \in CIG_{\text{min}} \end{array} \right\}.
\]

(112)
Use is made of the canonical variable form as defined in Def. 2.2, since $C(Y_1,Y_2)$ is invariant with respect to the transformation of initial random variables $(Y_1,Y_2)$ to their canonical variable form. Further, use is made of minimal Gaussian conditional independence as defined in Def. 2.13, and of related results of Section 2.3.

Def. 2.13 is needed, because, for the computation of $C(Y_1,Y_2)$, the attention can be restricted to those state variables $W$ which are of minimal dimension.

The parametrization of the family of Gaussian probability distributions $\mathcal{P}^{\text{CIG}}_{\text{min}}$ require the solution of the weak stochastic realization problem of Gaussian random variables defined by Problem 2.15. This problem is solved in [23, Th. 4.2]. For the readers convenience it is stated below.

**Theorem 3.8** [23, Theorem 4.2] Consider a tuple $(Y_1,Y_2)$ of Gaussian random variables in the canonical variable form of Def. 2.2. 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 on the diagonal ordered real-numbers in the interval $(0,1)$. Hence,

$$(Y_1,Y_2) \in G(0,Q_{(Y_1,Y_2)}), \quad (113)$$

$${Q}_{(Y_1,Y_2)} = \begin{pmatrix} I & D \\ D & I \end{pmatrix}, \quad (114)$$

$$D = \text{Diag}(d_1, d_2, \ldots, d_n) \in \mathbb{R}^{n \times n}, \quad 1 > d_1 \geq d_2 \geq \ldots \geq d_n > 0. \quad (115)$$

That is, $p_{11} = p_{21} = 0$, $p_{13} = p_{33} = 0$.

(a) There exists a probability measure $P_1$, and a triple of Gaussian random variables $Y_1,Y_2$, $W : \Omega \to \mathbb{R}^n$ defined on it, such that (i) $P_1|_{(Y_1,Y_2)} = P_0$ and (ii) $(F^{Y_1}, F^{Y_2}| F^W) \in \text{CIG}_{\text{min}}$.

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

$$G(0, Q_\iota(Q_W)), \quad Q_W \in \mathcal{Q}_W, \quad (116)$$

$${Q}_\iota = Q_\iota(Q_W) = \begin{pmatrix} I & D \\ D & I \end{pmatrix}, \quad D^{1/2} \begin{pmatrix} Q_W \\ Q_W^T \end{pmatrix} D^{1/2} \quad (117)$$

$$Q_W = \left\{ Q_W \in \mathbb{R}^{n \times n} | Q_W = Q_W^T, \quad 0 < D \leq Q_W \leq D^{-1} \right\}, \quad (118)$$

$$P_{\Delta} = \left\{ G(0, Q_\iota(Q_W)) \mid (\mathbb{R}^{3n}, B(\mathbb{R}^{3n}))| Q_W \in \mathcal{Q}_W \right\} \subseteq \mathcal{P}^{\text{CIG}}_{\text{min}} \quad (119)$$

The application of Theorem 3.8 is discussed in the next remark, in the context of parametrizing 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, and in the context of parametrizing the rate region of the Gaussian MACs of Section 1.1.

**Remark 3.9** Theorem 3.8 is a parametrization of the family of Gaussian measures $P_{\Delta} \subseteq \mathcal{P}^{\text{CIG}}_{\text{min}}$ by the entries of the covariance matrix $Q_W$. Hence, it is at most, an $(n + 1)/2$–dimensional parametrization.

(a) Theorem 3.8, when applied to the Gaussian MACs of Section 1.1, by also incorporating the average power constraint, it parametrizes the rate regions $C_{\text{MAC}}(\kappa_1,\kappa_2)$ and $C_{\text{MAC--C}}(\kappa_1,\kappa_2)$ by at most an $(n + 1)/2$–dimensional parametrization. The weak stochastic realization of random variables $(Y_1,Y_2)$, in terms of the random variable $W$ is given in Corollary 3.10.

(b) By analogy to (a), it is expected that the achievable rate region $\mathcal{R}_{GW}(\Delta_1,\Delta_2) = \mathcal{R}_{GW}^*(\Delta_1,\Delta_2)$ is generated from distributions $P_{\Delta} \subseteq \mathcal{P}^{\text{CIG}}_{\text{min}} \subseteq \mathcal{P}$. This issue is addressed in the companion paper [4].

The next corollary is preliminary to a subsequent theorem (i.e., Theorem 3.11), that shows $C(Y_1, Y_2)$ is achieved by the distribution $P_{W|Y_1,Y_2} \subseteq P_{\Delta}$, corresponding to $W^* \in G(0,I)$, i.e., with covariance the identity matrix. The next corollary gives the realization of $(Y_1, Y_2)$, expressed in terms of an arbitrary Gaussian random variable $W \in G(0,Q_W)$.

29
Corollary 3.10 Consider a tuple \((Y_1, Y_2)\) of Gaussian random variables in the canonical variable form of Def. 2.2. Restrict attention to the correlated parts of these random variables, as defined in Theorem 3.8, by (113)-(115).

Then a realization of the random variables \((Y_1, Y_2)\) which induce the family of measures \(P_{ci} \subseteq \mathcal{P}^{\text{CIG}}\), defined by (116)-(119), is

\[
Y_1 = Q_{Y_1,W} Q_W^{-1} W + Z_1
\]

\[
Q_{Y_1,W} = D^{1/2}, \quad Z_1 \in G(0,(I - D^{1/2}Q_W^{-1}D^{1/2})),
\]

\[
Y_2 = Q_{Y_2,W} Q_W^{-1} W + Z_2
\]

\[
Q_{Y_2,W} = D^{1/2} Q_W, \quad Z_2 \in G(0,(I - D^{1/2}Q_W D^{1/2})),
\]

\((Z_1, Z_2, W)\), are independent.

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

\[
I(Y_1, Y_2; W) = H(Y_1, Y_2) - H(Y_1|W) - H(Y_2|W)
\]

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

and it is parametrized by \(Q_W \in Q_W\), where \(Q_W\) is defined by the set of equation (118).

Proof of Corollary 3.10 The correctness of the realization is due to Proposition 2.14 and Theorem 3.8. The calculation of mutual information follows from the realization.

\[\square\]

In the next theorem the family of measures \(P_{ci} \subseteq \mathcal{P}^{\text{CIG}}\), defined by (116)-(119), which leads to realization of \((Y_1, Y_2)\), given in Corollary 3.10, is ordered for the determination of a single joint distribution \(P_{Y_1, Y_2, W^*} \in P_{ci} \subseteq \mathcal{P}^{\text{CIG}}\), 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 3.11 Consider a tuple \((Y_1, Y_2)\) of Gaussian random variables in the canonical variable form of Def. 2.2. Restrict attention to the correlated parts of these random variables, as defined in Theorem 3.8, defined by (113)-(115).

The following hold.

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

\[
C(Y_1, Y_2) = \frac{1}{2} \sum_{i=1}^{n} \ln \left( \frac{1+d_i}{1-d_i} \right) = \frac{1}{2} \sum_{i=1}^{n} \ln \left( 1 + \frac{2d_i}{1-d_i} \right) \in (0, \infty).
\]

(b) The realizations of the random variables \((Y_1, Y_2, W^*)\) that achieve \(C(Y_1, Y_2)\) are represented by

\[V : \Omega \rightarrow \mathbb{R}^n, \quad V \in G(0,I),\]

\[F^V, F^{Y_1} \vee F^{Y_2}, \text{ are independent } \sigma\text{-algebras},\]

\[L_1 = L_2 = D^{1/2}(I+D)^{-1} \in \mathbb{R}^{n \times n}, \quad L_3 = (I-D)^{1/2}(I+D)^{-1/2} \in \mathbb{R}^{n \times n}, \quad L_1, L_2, L_3, \text{ are diagonal matrices},\]

\[W^* = L_1 Y_1 + L_2 Y_2 + L_3 V, \quad W^* : \Omega \rightarrow \mathbb{R}^n,\]

\[Z_1 = Y_1 - D^{1/2}W^*, \quad Z_1 : \Omega \rightarrow \mathbb{R}^n,\]

\[Z_2 = Y_2 - D^{1/2}W^*, \quad Z_2 : \Omega \rightarrow \mathbb{R}^n.\]
Then
\[ Z_1 \in G(0,(I-D)), \quad Z_2 \in G(0,(I-D)), \quad W^* \in G(0,I); \]
\[ (Z_1,Z_2,W^*), \text{ are independent and} \]
\[ Y_1 = D^{1/2}W^* + Z_1, \quad Y_2 = D^{1/2}W^* + Z_2 \]
\[ (133) \]
\[ (134) \]
\[ (135) \]
hence the variables \((Y_1,Y_2,W^*)\) induce a distribution \(P_{Y_1,Y_2,W^*} \subseteq P_{\text{min}} \). Note that, in addition, each of the random variables \(Z_1, Z_2,\) and \(W^*\) has independent components.

(c) The variables \((Y_1,Y_2,W^*)\) defined in (b) induce a distribution \(P_{Y_1,Y_2,W^*} \subseteq P_{\text{min}} \) which achieves \(C(Y_1,Y_2)\)
\[ C(Y_1,Y_2) = I(Y_1,Y_2;W^*). \]
\[ (136) \]

**Proof of Theorem 3.11** Since mutual information \(I(Y_1,Y_2;W)\) is invariant with respect to nonsingular transformations, then by Theorem 2.3, (b), it suffices to consider the canonical variable form of Def. 2.2, 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 \(P_{Y_1,Y_2,W} = P_0\) 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 [23, Proposition 3.5]. Thus,
\[ W_1 = \mathbb{E}[Y_1|F_W] = L_{11} W, \quad L_{11} \in \mathbb{R}^{n \times n_1}, \]
\[ W_2 = \mathbb{E}[Y_2|F_W] = L_{12} W_1, \quad L_{12} \in \mathbb{R}^{n \times n_1}. \]

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 \(L_{15} \in \mathbb{R}^{n \times n_2}\) such that,
\[ W_3 = L_{15} W_2 = L_{15} L_{12} L_{11} W = L_{13} W, \quad L_{13} = L_{15} L_{12} L_{11}, \quad W_3 \in G(0,Q_3), \quad Q_3 = I_n = L_{13} Q_W L_{13}^T. \]

It is then possible to construct a matrix \(L_{14} \in \mathbb{R}^{(n_1-n_2) \times n_1}\) such that,
\[ W_4 = L_{14} W, \quad W_4 \in G(0,Q_4), \quad Q_4 = I, \quad L_{14} Q_W L_{13}^T = 0; \]
\[ \begin{pmatrix} W_3 \\ W_4 \end{pmatrix} \in G(0, I_{n_1}), \quad \text{rank} \left( \begin{pmatrix} L_{13} \\ L_{14} \end{pmatrix} \right) = n_1, \]
and, due to \(L_{14} Q_W L_{13}^T = 0\), \(W_3, W_4\) are independent random variables. See [16, Th. 4.9] for a theorem with which the existence of \(L_4\) can be proven. Note further that \(F_W = F_{W_4}\).

Hence the random variables \(W_3, W_4\) are independent, \((F_{Y_1},F_{Y_2}|F_{W_4}) \in \text{CIG}_{\text{min}}\), and
\[ I(Y_1,Y_2;W) = I(Y_1,Y_2;W_3,W_4). \]

From equation (191) now follows that,
\[ I(Y_1,Y_2;W_3,W_4) - I(Y_1,Y_2;W_3) \]
\[ = H(Y_1,Y_2) + H(W_3,W_4) - H(Y_1,Y_2,W_3,W_4) - H(Y_1,Y_2) - H(W_3) + H(Y_1,Y_2,W_3) \]
\[ = H(Y_1,Y_2,W_3) + H(W_4) - H(Y_1,Y_2,W_3,W_4), \text{ by independence of } W_3 \text{ and } W_4; \]
\[ = I(Y_1,Y_2,W_3,W_4) \geq 0. \]
The above inequality is also related to Pinsker’s chain rule, [20]. 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 $Y_1, Y_2, W : \Omega \to \mathbb{R}^n$ with $P_1(Y_1, Y_2) = P_0$ and $(F^{W_1}, F^{W_2}) \in \mathcal{C}_{\text{Gmin}}$.

According to [23, Th. 4.2] there exist in general many such measures which are parametrized by the matrices and the sets, as stated in Theorem 3.8, (b), and defined by (116)-(119).

(3) Next the mutual information of the triple of Gaussian random variables is calculated for any choice of $Q_W \in Q_W$, where $Q_W$ is given by (118). From (194) follows that

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

The following calculations are then obvious,

$$\det(Q(Y_1, Y_2)) = \det \begin{pmatrix} I & D \\ D & I \end{pmatrix} = \det(I - D^2) = \prod_{i=1}^{n}(1 - d_i^2);$$

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

$$P(Y_1 | W) \in G(E[Y_1 | F^W], Q_Y | W),$$

$$E[Y_1 | F^W] = Q_{Y_1, w} Q_w^{-1} W = D^{1/2} Q_w^{-1} W; \quad \text{by (117)}$$

$$Q_{Y_1, w} = I - Q_{Y_1, w} Q_w^{-1} W = D^{1/2} Q_w^{-1} W; \quad \text{by (117)}$$

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

$$E[Y_2 | F^W] = Q_{Y_2, w} Q_w^{-1} W = D^{1/2} Q_w Q_w^{-1} W = D^{1/2} W;$$

$$Q_{Y_2, w} = I - Q_{Y_2, w} Q_w^{-1} W = D^{1/2} Q_w Q_w^{-1} W; \quad \text{by (117)}$$

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

From the above calculations it then follows,

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

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

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

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

$$= \frac{1}{2} \sum_{i=1}^{n} \ln(1 - d_i^2) - \frac{1}{2} \ln(\det([I - D^{1/2} Q_w^{-1} W D^{1/2}][I - D^{1/2} Q_w W D^{1/2}])), \quad (139)$$

The above calculations verify the statements of Corollary 3.10.

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

$$C(Y_1, Y_2) = \inf_{P_1 \in P_d} I(Y_1, Y_2; W)$$

$$= \inf_{Q_w \in Q_W} \left\{ \frac{1}{2} \sum_{i=1}^{n} \ln(1 - d_i^2) - \frac{1}{2} \ln(\det([I - D^{1/2} Q_w^{-1} W D^{1/2}][I - D^{1/2} Q_w W D^{1/2}])) \right\}. \quad (140)$$

Since the first term in (140), $\frac{1}{2} \sum_{i=1}^{n} \ln(1 - d_i^2)$, does not depend on $Q_w$ and the natural logarithm is a strictly increasing function, then

$$C(Y_1, Y_2)$$

is equivalent to the problem: $\sup_{Q_w \in Q_W} \det \left[ (I - D^{1/2} Q_w^{-1} W D^{1/2})(I - D^{1/2} Q_w W D^{1/2}) \right]. \quad (141)$
Let $Q_W \in \mathbb{R}^{n \times n}$ be a non-symmetric square matrix in general.

It will be proven that,

$$f_1(Q_W) = \det(L_1(Q_W)) \leq \det([I - D])^2, \forall Q_W \in \mathbb{Q}_W,$$

$$\det(L_1(Q_W)) = \det([I - D])^2 \text{ if and only if } Q_W = I. \quad (145)$$

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

The inequality in (144) follows from Proposition B.5. The equality of (145) is proven in two steps. If $Q_W = I$ then equality of (145) holds as follows from direct substitution in (142). The converse is proven by contradiction. Suppose that $Q_W \neq I$. Then it follows again from Proposition B.5 that strict inequality holds in (144). Hence the equality is proven.

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

$$C(Y_1, Y_2) = \frac{1}{2} \sum_{i=1}^{n} \ln(1 - d_i^2) - \frac{1}{2} \ln(\det(I - D^{1/2}(Q_W)^{-1}D^{1/2})) - \frac{1}{2} \ln(\det(I - D^{1/2}(Q_W^*)D^{1/2}))$$

$$= \frac{1}{2} \sum_{i=1}^{n} \ln(1 - d_i^2) - \frac{1}{2} \ln(\det(I - D))$$

$$= \frac{1}{2} \sum_{i=1}^{n} \ln(1 - d_i^2) - \frac{1}{2} \sum_{i=1}^{n} \ln((1 - d_i)^2) = \frac{1}{2} \sum_{i=1}^{n} \ln \left( \frac{1 - d_i^2}{(1 - d_i)^2} \right)$$

$$= \frac{1}{2} \sum_{i=1}^{n} \ln \left( 1 + \frac{2d_i}{1 - d_i} \right).$$

(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 $Q_W = I(L)$. Consider now a triple of random variables $(Y_1, Y_2, W) \in G(0, Q_s(I))$ as defined in (116)-(119), 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,

$$(Y_{12}, Y_{22}, W^*) \in G(0, Q_s(I)),$$

$$Q_s(I) = \begin{pmatrix} I & D & D^{1/2} \\ D & I & D^{1/2} \\ D^{1/2} & D^{1/2} & I \end{pmatrix} > 0. \quad (146)$$

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 $\hat{W} = L_1Y_1 + L_2Y_2 + L_3V$. Then $(Y_1, Y_2, V, W^*)$ are jointly Gaussian and it has to be shown that then $Q_{\hat{W}} = I, Q_{Y_1, \hat{W}} = D^{1/2},$ and $Q_{Y_2, \hat{W}} = D^{1/2}$. 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 $G(0, Q_s(I))$ that $\hat{W} = W^*$ almost surely.

The signals are then represented by,

$$Z_1 = Y_1 - \mathbb{E}[Y_1|W^*] = Y_1 - Q_{Y_1, W^*}(Q_{W^*})^{-1}W^* = Y_1 - D^{1/2}W^*, \quad (147)$$

$$Z_2 = Y_2 - \mathbb{E}[Y_2|W^*] = Y_2 - Q_{Y_2, W^*}(Q_{W^*})^{-1}W^* = Y_2 - D^{1/2}W^*. \quad (148)$$

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

$$\mathbb{E}[Z_1(W^*)^T] = \mathbb{E}[Y_1(W^*)^T] - D^{1/2}\mathbb{E}[W^*(W^*)^T] = D^{1/2} - D^{1/2} = 0, \quad \mathbb{E}[Z_2(W^*)^T] = 0,$$

$$\mathbb{E}[Z_1Z_2^T] = \mathbb{E}[(Y_1 - D^{1/2}W^*)(Y_2 - D^{1/2}W^*)^T] = 0.$$
Hence, the original signals are represented as shown by the formulas,
\[ Y_1 = Z_1 + D_1^{1/2}W^*, \quad \text{by (146), } Q_{Y_1, W^*} Q_{W^*}^{-1} = D_1^{1/2}, \] and by def. of \( Z_1 \),
\[ Y_2 = Z_2 + D_2^{1/2}W^*, \quad \text{similarly.} \]

\[ \square \]

Example 3.12 (below) is introduced to illustrate some subtle issues related to Theorem 1.4 that computes
the lossy common information for the bivariate Gaussian random variables (26), i.e., \( p_1 = p_2 = 1 \), given in
Theorem 1.4, found in many references, such as, [10, 26, 31].

**Example 3.12** Consider an application of Theorem 3.11 to the bivariate Gaussian random variables (26) of
Theorem 1.4, i.e., \( p_1 = p_2 = 1 \). Assume that the random variables have been transformed to the canonical
variable form with in this case a single canonical correlation coefficient \( d_1 \in [0, 1] \), see Example 2.5.(i). Note
that if the correlation coefficient of \( Y_1 \) and \( Y_2 \) is negative, \( \rho_{Y_1 Y_2} < 0 \) then that representation is not in the
canonical variable form so the theorem above does not apply. The theorem above requires the canonical
variable form with a positive correlation coefficient.

Distinguish the cases:
(i) \( d_1 \in (0, 1) \). Hence, the Gaussian measure in canonical variable form is

\[
(Y_{12}, Y_{22}) \in G(0, Q_{Y_{12}, Y_{22}}), \quad Q_{(Y_{12}, Y_{22})} = \begin{pmatrix}
1 & d_1 \\
0 & 1
\end{pmatrix},
\]
\[ Y_{12} : \Omega \to \mathbb{R}^{p_{12}}, \quad Y_{22} : \Omega \to \mathbb{R}^{p_{22}}, \quad p_{12} = p_{22} = 1. \]

(a) The minimal \( \sigma \)-algebra which makes \( Y_{12} \) and \( Y_{22} \) Gaussian conditional-independent is \( F^W \), where \( W : \Omega \to \mathbb{R}, W \in G(0, Q_W), \) i.e., \( W \) is the state variable.

(b) Then the random variable which achieves \( C(Y_1, Y_2) \) is \( W = W^* \in G(0, 1) \), and

\[
C(Y_1, Y_2) = C(Y_{12}, Y_{22}) = I(Y_{12}, Y_{22}; W^*) = \frac{1}{2} \ln \left( \frac{1 + d_1}{1 - d_1} \right).
\]

(c) The weak stochastic realization that achieves (150) is

\[
Y_{12} = \sqrt{d_1} W^* + \sqrt{1 - d_1} Z_1, \quad \text{(151)}
\]
\[
Y_{22} = \sqrt{d_1} W^* + \sqrt{1 - d_1} Z_2, \quad \text{(152)}
\]
\[
Z_1 \in G(0, 1), \quad Z_2 \in G(0, 1), \quad W^* \in G(0, 1), \quad \text{(153)}
\]
\[
W^*, Z_1, Z_2 \quad \text{independent.} \quad \text{(154)}
\]

(ii) \( d_1 = 0 \). This follows from the special cases discussed in Proposition 3.13.

(iii) \( d_1 = 1 \). This follows from the special cases discussed in Proposition 3.14.

### 3.5 Wyner’s Information Common Information in Case of Private Gaussian Vectors

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

**Proposition 3.13** Consider the case of a tuple of Gaussian vectors with only private parts. Hence the Gaussian
measure is

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

34
Formally, to compute the Gaussian measure is, \( + \) random variables takes the value \( + \). Consider the case of a tuple of Gaussian vectors with only the identical part. Hence the special case of canonical variable form with only identical parts is presented below.

3.6 Wyner’s Information Common Information in Case of Identical Gaussian Vectors

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

**Proposition 3.14** Consider the case of a tuple of Gaussian vectors with only the identical part. Hence the Gaussian measure is,

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

\[
(Y_{11}, Y_{21}) \in G(0, Q_{(Y_{11}, Y_{21})}), \quad Q_{(Y_{11}, Y_{21})} = \begin{pmatrix} I & I \\ I & I \end{pmatrix}, \quad Y_{11} = Y_{21} \text{ a.s.}
\]

(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, \( W = Y_{11} = Y_{21} \) and \( F^W = F_{Y_{11}} = F_{Y_{21}} \).

(b) Then \( C(Y_{11}, Y_{21}) = C(Y_{11}, Y_{21}) = +\infty \). See the comment below.

(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,

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

The amount of Wyner’s common information of \( C(Y_{11}, Y_{21}) \) for this case is \(+\infty\) and this requires comments. Formally, to compute \( I(Y_{11}, Y_{21}; W) \), one needs to evaluate the expression (197). But in this case the determinant \( \det(Q(Y_{11}, Y_{21}, W)) \) is zero. Thus,

\[
C(Y_{11}, Y_{21}) = I(Y_{11}, Y_{21}; W) = -\ln\left(\frac{\det(Q(Y_{11}, Y_{21}, W))}{\det(Q_{Y_{11}, Y_{21}}) \det(Q_W)}\right) = -\ln(0).
\]

One may extend the definition of the natural logarithm from the domain \((0, +\infty)\) to the value at zero by the limit \( \lim_{x \to 0} \ln(x) = -\infty \). The authors therefore propose to define that \( C(Y_{11}, Y_{21}) \) of identical Gaussian random variables takes the value \(+\infty\), hence \( C(Y_{11}, Y_{21}) = +\infty \). This definition also makes sense considered from an information theoretic interpretation, in which mutual information is defined via relative entropy which admits the value of \(+\infty\).
3.7 Wyner’s Lossy Common Information of Correlated Gaussian Vectors

Consider a tuple \( (Y_1, Y_2) \) of Gaussian random variables in the canonical variable form of Def. 2.2. Restrict attention to the correlated parts of these random variables, as defined in Theorem 3.8, defined by (113)-(115).

Consider the tuple of jointly Gaussian random variables \( Y_1 : \Omega \rightarrow \mathbb{R}^{p_1}, Y_2 : \Omega \rightarrow \mathbb{R}^{p_2} \) with square error distortion functions of Definition 3.1. Then the following statements hold.

(a) The minimization of \( I(Y_1, Y_2; W) \) over joint distributions \( P_{Y_1,Y_2,W}(y_1,y_2,w) \) having a marginal distribution \( P_{Y_1,Y_2,W}(y_1,y_2,w) = P_{Y_1,Y_2}(y_1,y_2) \) can be confined to Gaussian random variables \( W \) such that \( (Y_1,Y_2,W) \) are jointly Gaussian distributed, and \( (Y_1,Y_2,W) \in \text{CIG} \).

(b) The calculation of the rate distortion functions \( R_Y(\Delta_i), R_{Y|W}(\Delta_i), \) \( i = 1, 2 \) and \( R_{Y_1,Y_2}(\Delta_1, \Delta_2) \) can be confined to jointly Gaussian random variables \( (Y_1,Y_2,Y_1',Y_2',W) \).

(c) The characterization of the Gray-Wyner lossy rate region \( \mathcal{R}_{GW}(\Delta_1, \Delta_2) \) of Theorem 1.1, in terms of the joint distributions \( P_{Y_1,Y_2,W} \in \mathcal{P} \), can be confined to Gaussian random variables \( W \) such that \( (Y_1,Y_2,W) \) are jointly Gaussian distributed.

(d) The characterization of Wyner’s lossy common information \( C_W(Y_1,Y_2) \) is invariant with respect to nonsingular basis transformation

\[
S = \text{Block-diag}(S_1, S_2) \in \mathbb{R}^{(p_1+p_2) \times (p_1+p_2)}
\]

such that with respect to the new basis \( (S_1Y_1, S_2Y_2) \) is jointly Gaussian.

(e) Any rate triple \( (R_0, R_1, R_2) \) that belongs to the characterization of the Gray-Wyner lossy rate region \( \mathcal{R}_{GW}(\Delta_1, \Delta_2) \) of Theorem 1.1, is equivalently computed by transforming the tuple \( (Y_1,Y_2) \) of Gaussian random variables into their canonical variable form of Def. 2.2.

In view of (a)-(e) above, the closed form expression of Wyner’s lossy common information is now obtained.

**Theorem 3.15** Consider a tuple \( (Y_1, Y_2) \) of Gaussian random variables in the canonical variable form of Def. 2.2. Restrict attention to the correlated parts of these random variables, as defined in Theorem 3.8, by (113)-(115).

Then Wyner’s lossy common information \( C_W(Y_1,Y_2) \) is given by

\[
C_{GW}(Y_1,Y_2;\Delta_1, \Delta_2) = C_W(Y_1,Y_2) = C(Y_1,Y_2) = \frac{1}{2} \sum_{j=1}^{n} \ln \left( \frac{1+d_j}{1-d_j} \right), \quad (\Delta_1, \Delta_2) \in \mathcal{D}_W \tag{162}
\]

where the distortion region is defined by

\[
\mathcal{D}_W = \left\{ (\Delta_1, \Delta_2) \in [0,\infty] \times [0,\infty] \mid 0 \leq \Delta_1 \leq \sum_{j=1}^{n} (1-d_j), \ 0 \leq \Delta_2 \leq \sum_{j=1}^{n} (1-d_j) \right\}, \tag{163}
\]

\[\forall j \in \mathbb{Z}, \ d_j \in (0,1).\]

**Proof of Theorem 3.15** A direct way to prove the statement is to compute the characterizations of the rate distortion functions \( R_Y(\Delta_i), R_{Y|W}(\Delta_i), i = 1, 2 \) and \( R_{Y_1,Y_2}(\Delta_1, \Delta_2) \), using the weak stochastic realization of Theorem 3.11(b), and then verify that identity (22) holds, i.e., \( R_{Y|W}(\Delta_1) + R_{Y|W}(\Delta_2) + I(Y_1,Y_2;W) = R_{Y_1,Y_2}(\Delta_1, \Delta_2) \) for \( (\Delta_1, \Delta_2) \in \mathcal{D}_W \), for the choice \( W = W^* \) (given in Theorem 3.11(b)). An alternative method is also given in [4]. \[\square\]

It is noted that Theorem 1.4, that corresponds to bivariate Gaussian random variables, i.e., \( p_1 = p_2 = 1 \), first derived by Gray and Wyner [10], and subsequently in [26, 31], is a strict special case of the above theorem.

36
3.8 Wyner’s Information Common Information of Arbitrary Gaussian Vectors

Consider finally a tuple of arbitrary Gaussian random variables, not necessarily red restricted to the correlated parts of these random variables, as defined in Theorem 3.8, by (113)-(115). Then $C(Y_1, Y_2)$ is computed by a decomposition and by the use of the formulas obtained earlier in Section 3.4.

Assume that the tuple is already transformed to the canonical variable representation, see Def. 2.2. Note that then

the three tuples of random variables $(Y_{11}, Y_{21})$, $(Y_{12}, Y_{22})$, $(Y_{13}, Y_{23})$ are independent.  \(\text{(64)}\)

**Theorem 3.16** Consider a tuple of Gaussian random variables $(Y_1, Y_2) \in G(0, Q_{\text{cvf}})$ as described and decomposed according to Algorithm 3.6.

(a) Then,

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

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

In particular cases one computes the canonical variable decomposition of the tuple $(Y_1, Y_2)$ and obtains the indices, $(p_{11}, p_{12}, p_{13})$ and $(p_{21}, p_{22}, p_{23})$. Then,

$$C(Y_{11}, Y_{21}) = +\infty, \text{ if } p_{11} = p_{12} > 0;$$

$$C(Y_{31}, Y_{32}) = 0, \text{ if } p_{31} > 0 \text{ and } p_{23} > 0;$$

$$C(Y_{12}, Y_{22}) = \frac{1}{2} \sum_{i=1}^{n} \ln \left( \frac{1 + d_i}{1 - d_i} \right), \text{ if } p_{12} = p_{22} > 0. \quad \text{(66)}$$

Thus $C(Y_{12}, Y_{22})$ is the most interesting value if defined.

(b) The random variable $W^*$ defined below, is such that $C(Y_1, Y_2)$ is attained by the mutual information for this random variable.

$$W^* : \Omega \to \mathbb{R}^n, \ n \in \mathbb{Z}_+, \quad \text{n}_1 = p_{11} = p_{21}, \ n_2 = p_{12} = p_{22}, \ n_1 + n_2 = n,$$

$$W^* = \begin{pmatrix} W_{11}^* \\ W_{22}^* \end{pmatrix}, \quad W_{11}^* : \Omega \to \mathbb{R}^{n_1}, \quad W_{22}^* : \Omega \to \mathbb{R}^{n_2},$$

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

$$W_{22}^* = L_1 Y_{12} + L_2 Y_{22} + L_3 V, \ \text{see Theorem 3.11.(b) for the formulas of } L_1, L_2, L_3; \quad \text{(69)}$$

then

$$(Y_1, Y_2, W^*) \in G(0, Q_4(I)), \quad (F^{Y_1, Y_2, Y_{13}, Y_{12}, Y_{23} | W^*_1, W^*_2}) \in CI, \ \text{see (117) for } Q_4(I).$$

$$F^{W^*_1} \subseteq (F^{Y_1} \lor F^{Y_2}), F^{W^*_2} \subseteq (F^{Y_{12}} \lor F^{Y_{22}}); \quad \text{(70)}$$

then also,

$$C(Y_1, Y_2) = I(Y_1, Y_2; W^*). \quad \text{(71)}$$
(c) The following operations are made, using (a),

\[
W^* = \left( \begin{array}{c}
W^*_1 \\
W^*_2
\end{array} \right),
\]

(174)

\[
W^*_1 = Y_{11} = Y_{21},
\]

(175)

\[
W^*_2 = L_1 Y_{12} + L_2 Y_{22} + L_3 V,
\]

(176)

see (128), (129) for the formulas of \(L_1, L_2, L_3\);

\[
Z_{12} = Y_{12} - E[Y_{12}|F^{W^*_2}],
\]

(177)

\[
Z_{22} = Y_{22} - E[Y_{22}|F^{W^*_2}],
\]

(178)

\[
Z_{13} = Y_{13}, \ Z_{23} = Y_{23}, \ (the \ components \ Z_{11} \ and \ Z_{21} \ do \ not \ exist),
\]

(179)

\[
Z_1 = \left( \begin{array}{c}
Z_{12} \\
Z_{13}
\end{array} \right), \ Z_2 = \left( \begin{array}{c}
Z_{22} \\
Z_{23}
\end{array} \right).
\]

(180)

Hence,

\[
Y_{11} = W^*_1 = Y_{21},
\]

(181)

\[
Y_{12} = Q_{Y_{12},W^*_2} Q_{W^*_2}^{-1} W^*_2 + Z_{12}, \ Y_{22} = Q_{Y_{22},W^*_2} Q_{W^*_2}^{-1} W^*_2 + Z_{22},
\]

(182)

\[
Y_{13} = Z_{13}, \ Y_{23} = Z_{23}.
\]

(183)

**Proof of Theorem 3.16 (a)**

\[
C(Y_1, Y_2) = \inf_{(Y_1, Y_2) \in \text{CIG}} I(Y_1, Y_2; W)
\]

\[
= \inf_{(Y_1, Y_2) \in \text{CIG}} \left\{ I(Y_{11}, Y_{21}; W_1) + I(Y_{12}, Y_{22}; W_2) + I(Y_{13}, Y_{23}; 0) \right\}, \ \text{by Proposition A.1},
\]

\[
\geq \inf_{(Y_1, Y_2, W) \in \text{CIG}} I(Y_{11}, Y_{21}; W_1) + \inf_{(Y_1, Y_2, W) \in \text{CIG}} I(Y_{12}, Y_{22}; W_2) + \inf_{(Y_1, Y_2, W) \in \text{CIG}} I(Y_{13}, Y_{23}; 0)
\]

\[
= \inf_{(Y_{11}, Y_{21}, W_1) \in \text{CIG}} I(Y_{11}, Y_{21}; W_1) + \inf_{(Y_{12}, Y_{22}, W_2) \in \text{CIG}} I(Y_{12}, Y_{22}; W_2) + I(Y_{13}, Y_{23}; 0)
\]

\[
= C(Y_{11}, Y_{21}; W_1) + C(Y_{12}, Y_{22}; W_2) + C(Y_{13}, Y_{23}; 0)
\]

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

(184)

The latter equality follows from, respectively, Proposition 3.14, Theorem 3.11, and Proposition 3.13.

(a & b) It will be shown that \(C(Y_1, Y_2)\) is less or equal to the right-hand side of equation (184). From the latter inequality and the above inequality then follows the expression according to equation (184).

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 3.11 that \((F^{Y_{12}}, F^{Y_{22}}|F^{W^*_2}) \in \text{CIG}_{\text{min}}\).

Then,

\[
C(Y_1, Y_2) = \inf_{(Y_1, Y_2, W) \in \text{CIG}} I(Y_1, Y_2; W) \leq I(Y_1, Y_2; W^*)
\]

\[
= I(Y_{11}, Y_{21}; W^*_1) + I(Y_{12}, Y_{22}; W^*_2) + I(Y_{13}, Y_{23}; 0)
\]

\[
= \left\{ \begin{array}{ll}
0, & \text{if } p_{13} > 0, \ p_{23} > 0, \ p_{11} = p_{12} = p_{21} = p_{22} = 0,
\frac{1}{2} \sum_{i=1}^{p} \ln \left( \frac{1 + d - d^2}{1 - d} \right), & \text{if } p_{12} = p_{22} > 0, \ p_{11} = p_{21} = 0, \ p_{13} \geq 0, \ p_{23} \geq 0,
+\infty, & \text{else}.
\end{array} \right.
\]
The latter equality is proven as follows. In the first case, when \( p_{13} > 0, p_{23} > 0, \) 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 3.13 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 A.1 and from Theorem 3.11 that 
\[
I(Y_1, Y_2; W^+) = I(Y_{12}, Y_{22}; W^2) + I(Y_{13}, Y_{23}; 0) = \frac{1}{2} \sum_{j=1}^{n} \ln \left( \frac{1 + d_i}{1 - d_i} \right).
\]
In the third case, when \( p_{11} = p_{21} > 0 \) and the other \( p_i \) indices are arbitrary, then \( I(Y_1, Y_2; W^+) = +\infty. \) Hence the inequality \( C(Y_1, Y_2) \leq \text{right-hand side} \) is proven and hence equality holds.

(c) This follows directly from Proposition 3.5.

The reader is advised to carry out an approximation of the covariance matrix based on the computations of Algorithm 2.10 as explained next. Consider the case in which the dimensions of the two vectors are ordered as \( p_1 \geq p_2. \) The case of \( p_1 < p_2 \) can be obtained by interchanging the indices. The outcome of the computation of the algorithm is then of the form,

\[
Q_{\text{cvf}} = \begin{pmatrix} I_{p_1} & Q_{12} \\ Q_{12}^T & I_{p_2} \end{pmatrix}, \quad Q_{12} = \begin{pmatrix} D \\ 0 \end{pmatrix} \in \mathbb{R}^{p_1 \times p_2},
\]

\[
D = \text{Diag}(d_1, \ldots, d_{p_2}) \in \mathbb{R}^{p_2 \times p_2}, 1 \geq d_1 \geq d_2 \geq \ldots \geq d_{p_2} \geq 0.
\]

Next the values on the diagonal of the matrix \( D \) are partitioned in three groups depending on two threshold values \( h_1, h_2 \in (0, 1) \) with \( h_1 > h_2, \) according to,

\[
\bar{d}_i = \begin{cases} 
1.0000, & \text{if } d_i \in (h_1, 1], \\
0.0000, & \text{if } d_i \in [0, h_2), \\
d_i, & \text{if } d_i \in [h_2, h_1],
\end{cases}
\]

for all \( i \in \mathbb{Z}_{p_2}. \) For example, if \( h_1 = 0.9990 \) and \( d_1 = 0.9993 \) then set \( \bar{d}_1 = 1.0000; \) and if \( h_2 = 0.0001 \) and \( d_{p_2} = 0.00006 \) then set \( \bar{d}_{p_2} = 0.0000. \) Denote then the number of elements of the vector \( \bar{d} \) which are equal to 1.0000 by \( p_{11} \) and the number of elements of that vector which are equal to 0.0000 by \( p_{23}. \) Define further, \( p_{21} = p_{11}, p_{22} = p_2 - p_{21} - p_{23}, p_{12} = p_{22}, \) and \( p_{13} = p_1 - p_{11} - p_{12}. \)

\[
\bar{Q} = \text{Diag}(\bar{d}_{p_{11}+1}, \ldots, \bar{d}_{p_{11}+p_{21}}) \in \mathbb{R}^{p_{12} \times p_{12}},
\]

\[
\bar{Q}_{12} = \begin{pmatrix} I_{p_{11}} & 0 \\ 0 & \bar{D} \end{pmatrix} \in \mathbb{R}^{p_{23} \times p_{23}}, \quad \bar{Q}_{12} = \begin{pmatrix} I_{p_{11}} & \bar{Q}_{12} \\ \bar{Q}_{12}^T & I_{p_{22}} \end{pmatrix} \in \mathbb{R}^{p_{12} \times p_{12}}.
\]

The indices \((p_{11}, p_{12}, p_{13})\) and \((p_{21}, p_{22}, p_{23})\) now satisfy the relations, \( p_1 = p_{11} + p_{12} + p_{13}, p_2 = p_{21} + p_{22} + p_{23}, p_{11} = p_{21}, \) and \( p_{12} = p_{22}. \) The reader can use either \( Q_{\text{cvf}} \) or its approximation \( \bar{Q}_{\text{cvf}}. \)

**Example 3.17** Consider the tuple of Gaussian random variables,

\[
(Y_1, Y_2) \in \mathcal{G}(0, Q_{(Y_1, Y_2)}), \quad p_1 = 3, \quad p_2 = 3,
\]

\[
Q_{(Y_1, Y_2)} = \begin{pmatrix} I_{p_1} & Q_{Y_1, Y_2} \\ Q_{Y_1, Y_2}^T & I_{p_2} \end{pmatrix},
\]

\[
Q_{Y_1, Y_2} = \begin{pmatrix} 0.8 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.1 \end{pmatrix} \in \mathbb{R}^{p_1 \times p_2}.
\]
A computation then yields,

\[ (p_{11}, p_{12}, p_{13}) = (0, 3, 0), \quad (p_{21}, p_{22}, p_{23}) = (0, 3, 0), \]

\[ D = \begin{pmatrix} 0.8 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.1 \end{pmatrix} \in \mathbb{R}^{3 \times 3}, \]

\[ C(Y_1, Y_2) = 5.0444 \text{ bits}. \]

**Example 3.18** Consider the tuple of Gaussian random variables,

\[ (Y_1, Y_2) \in \mathcal{G}(0, Q_{(Y_1, Y_2)}), \quad p_1 = 6, \quad p_2 = 5, \]

\[ Q_{(Y_1, Y_2)} = \begin{pmatrix} I_{p_1} & Q_{Y_1, Y_2} \\ Q_{Y_1, Y_2}^T & I_{p_2} \end{pmatrix}, \]

\[ Q_{Y_1, Y_2} = \begin{pmatrix} 0.999998 & 0 & 0 & 0 \\ 0 & 0.999992 & 0 & 0 \\ 0 & 0 & 0.8 & 0 \\ 0 & 0 & 0 & 0.3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{6 \times 6}. \]

A computation then yields,

\[ (p_{11}, p_{12}, p_{13}) = (2, 2, 2), \quad (p_{21}, p_{22}, p_{23}) = (2, 2, 1), \]

\[ D = \begin{pmatrix} 0.8 & 0 \\ 0 & 0.3 \end{pmatrix} \in \mathbb{R}^{2 \times 2}, \]

\[ C(Y_1, Y_2) = +\infty \text{ bits}, \]

\[ C(Y_{12}, Y_{22}) = 4.0630 \text{ bits}. \]

**Example 3.19** Consider a tuple of Gaussian random variables for which the variance matrix is generated by a random number generator. Generate the matrix \( L \in \mathbb{R}^{p \times p} \) such that every element has a normal distribution with parameters \( \mathcal{G}(0, 1) \) and such that all elements of the matrix are independent. Then define \( Q = LL^T \) to guarantee that the matrix \( Q \) is semi-positive-definite. Then,

\[ (Y_1, Y_2) \in \mathcal{G}(0, Q_{(Y_1, Y_2)}), \quad p_1 = 5, \quad p_2 = 4, \quad p = p_1 + p_2 = 9; \]

\[ Q_{(Y_1, Y_2)}, \text{ randomly generated as described above, values not displayed.} \]

The outcome of a computation is then that,

\[ (p_{11}, p_{12}, p_{13}) = (0, 4, 1); (p_{21}, p_{22}, p_{23}) = (0, 4, 0); \]

\[ C(Y_{12}, Y_{22}) = 13.1597 \text{ bits}. \]

**Example 3.20** Consider a tuple of Gaussian random variables of which the variance is generated as in the previous example.

\[ (Y_1, Y_2) \in \mathcal{G}(0, Q_{(Y_1, Y_2)}), \quad p_1 = 5, \quad p_2 = 4, \quad p = p_1 + p_2 = 9; \]

\[ Q_{(Y_1, Y_2)}, \text{ randomly generated as described above.} \]

Then

\[ (p_{11}, p_{12}, p_{13}) = (0, 4, 1); \quad (p_{21}, p_{22}, p_{23}) = (0, 4, 0); \]

\[ C(Y_{12}, Y_{22}) = 13.9962 \text{ bits}. \]
Remark 3.21 As a final comment it is mentioned that the methods of this paper can be used to characterize the rate region $R_{MAC-C}(\kappa_1, \kappa_2)$ of stationary memoryless multiple-input multiple-output (MIMO) additive Gaussian MAC with common message, (28)-(33), using the weak stochastic realization of Gaussian random variables, of Definition 2.17, and Proposition 2.18, and then specialized to the rate region $R_{MAC}(\kappa_1, \kappa_2)$ without common message, defined by (34)-(37).

4 Concluding Remarks

This paper formulates the classical Gray and Wyner source coding for a simple network with a tuple of multivariate, correlated Gaussian random variables, from the geometric approach of a Gaussian random variables, and the weak stochastic realization of correlated Gaussian random variables. This approach leads to a very general treatment of such problems, and provides insight which is missing when attacked by other methods. A closed-form expression is derived for Wyner’s lossy common information between two multivariate correlated Gaussian random variables, when the decoders apply mean-square error decoding; it corresponds to the minimum common message rate $R_0$ on the Gray and Wyner rate region with sum rate $R_0 + R_1 + R_2$ equal to the joint rate distortion function, which is constant over a specific rate region. However, much remains to be done to exploit the new approach to other multi-user problems of information theory.

Acknowledgements

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 1980’s. This paper is an answer to his 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 Prof. Guo Lei (Chinese Academy of Sciences, Institute for Mathematics) and to Dr. Xi Kaihua (Delft University of Technology) for help with obtaining copies of the papers of Hua LooKeng.

The second named author thanks the Department of Applied Mathematics of Delft University of Technology for the hospitality arrangement at the department during which the paper was written.
References

[1] T.W. Anderson. An introduction to multivariate statistical analysis. Wiley, New York, 1958.

[2] G. Birkhoff and A. MacLane. A survey of modern algebra, fourth edition. MacMillan Publ. Co. Inc., New York, 1977.

[3] S. N. Diggavi S. Shamai C. Tian, J. Chen. Optimality and approximate optimality of source-channel separation in networks. IEEE Trans. Information Theory, 60:904–918, 2014.

[4] Charalampos D. Charalambous and Jan H. van Schuppen. Rates for the Gray and Wyner lossy source coding for a simple network of a tuple of correlated multivariate Gaussian RVs. IEEE Transactions on Information Theory, preprint, 2019.

[5] T.M. Cover. An algorithm for maximizing expected log investment return. IEEE Trans. Inform. Theory, 30:369–373, 1984.

[6] T.M. Cover and J.A. Thomas. Elements of information theory. John Wiley & Sons, New York, 1991.

[7] P. Gacs and J. Körner. Common information is much less than mutual information. in Problems of Control and Information Theory, 2:149–162, 1973.

[8] R.G. Gallager. Information theory and reliable communication. John Wiley & Sons, New York, 1968.

[9] R.D. Gitens. Canonical analysis - A review with applications in ecology. Springer-Verlag, Berlin, 1985.

[10] R. M. Gray and A.D. Wyner. Source coding for a simple network. Bell System Techn. J., 53:1681–1721, 1974.

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

[12] L.K. Hua. Inequalities involving determinants (in chinese; english summary). Acta Math. Sinica, 5:463–470, 1955.

[13] G. Kramer. Topics in multi-user information theory. Foundations and Trends in Communication and Information Theory, 4:265–444, 2008.

[14] Amos Lapidoth and Michèle Wigger. A necessary condition for the transmissibility of correlated sources over a MAC. In Proc. 2016 IEEE International Symposium on Information Theory (ISIT 2016), pages 2024–2028, New York, 2016. IEEE Press.

[15] A.W. Marshall and I. Olkin. Inequalities: Theory of majorization and its applications. Academic Press, New York, 1979.

[16] B. Noble. Applied linear algebra. Prentice-Hall, Englewood Cliffs, NJ, 1969.

[17] L. Ozarow. On a source-coding problem with two channels and three receivers. Bell System Techn. J., 59:1909–1921, 1980.

[18] L Ozarow. The capacity of the white gaussian multiple access channel with feedback. IEEE Trans. Information Theory, 30, July 1984.

[19] L Ozarow and S. Leung-Yan-Cheong. An achievable region and outer bound for the gaussian broadcast channel with feedback. IEEE Transactions on Information Theory, 30(4):667–671, July 1984.

[20] M.S. Pinsker. Information and information stability of random variables. Holden-Day, San Francisco, 1964.

[21] D. Slepian and J. K. Wolf. A coding theorem for multiple access channels with correlated sources. Bell Syst. Techn. J., 52:1037–1076, 1973.

[22] A.A. Stoorvogel and J.H. van Schuppen. System identification with information theoretic criteria. In S. Bittanti and G. Picci, editors, Identification, adaptation, learning, pages 289–338. Springer, Berlin, 1996.

[23] C. van Putten and J.H. van Schuppen. The weak and strong gaussian probabilistic realization problem. J. Multivariate Anal., 13:118–137, 1983.

[24] C. van Putten and J.H. van Schuppen. Invariance properties of the conditional independence relation. Ann. Probab., 13:934–945, 1985.

[25] Jan H. van Schuppen. Common, correlated, and private information in control of decentralized systems. In Jan H. van Schuppen and Tiziano Villa, editors, Coordination control of distributed systems, number 456 in Lecture Notes in Control and Information Sciences, pages 215–222. Springer International Publishing, Cham, Switzerland, 2015.

[26] Kumar B. Viswanatha, Emrah Akyol, and Kenneth Rose. The lossy common information of correlated sources. IEEE Trans. Information Theory, 60:3238–3253, 2014.

[27] A.D. Wyner. The common information of two dependent random variables. SIAM J. Appl. Math., 28:100–113, 1975.

[28] A.D. Wyner. Values and bounds for common information of two discrete variables. SIAM J. Applied Mathematics, 31:313–333, 1976.

[29] J.-J. Xiao and Z.-Q. Luo. Compression of correlated Gaussian sources under individual distortion criteria, pages 438–447. University of Illinois, 2005.

[30] Ge Xu, Wei Liu, and Biao Chen. A lossy source coding interpretation of wyner’s common information. IEEE Trans. Information Theory, 62:754–768, 2016.

[31] A.M. Yaglom and I.M. Yaglom. Probability and Information. D. Reidel Publ. Co., Dordrecht, 1983.
A Information Theory

The reader finds in this appendix a few formulas of information theory which are used in the body of the paper. Most of these are known in the literature, and are found in standard books on information theory such as, [6, 8, 32].

Entropy of a continuous-valued random variable \( X : \Omega \rightarrow \mathbb{R} \) with a probability density with respect to Lebesgue measure and with support on the real numbers, is defined by the formula, [8, p. 31, (2.4.24)].

\[
H(X) = \int_{-\infty}^{\infty} p(v) \log \left( \frac{1}{p(v)} \right) dv.
\]

(185)

The term differential entropy is used for the entropy of a continuous-valued random variable in the literature but this term will not be used further in this paper.

If the random variables \( X \) and \( Y \) are independent then by [6, Th. 2.6.6],

\[
H(X, Y) = H(X) + H(Y).
\]

(186)

If for two random variables \((X, Y)\) the conditional distribution of the random variable \( Y \) conditioned on the random variable \( X \) exists then the conditional entropy is defined by the formulas,

\[
H(Y|X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{Y|X}(v, w) \log \left( \frac{1}{p_{Y|X}(v|w)} \right) dv dw.
\]

(187)

For these formulas see, [8, p. 31, (2.4.25)] and [6, p. 16, (2.12)]. It then follows from [8, p. 31] that

\[
H(Y|X) = H(X, Y) - H(X).
\]

(188)

Mutual information of a tuple of random variables \((X, Y)\) is defined by the formulas, see [8, p. 31, (2.4.28)],

\[
I(X; Y) = \int \int p_{X,Y}(v, w) \log \left( \frac{p_{X,Y}(v, w)}{p_X(v)p_Y(w)} \right) dv dw \in [0, \infty].
\]

(189)

It follows from [8, Th. 2.3.2] and [6, p. 27] that \(I(X; Y) = 0\) if and only if the random variables \(X\) and \(Y\) are independent.

Mutual information is known to be invariant with respect to scaling with nonsingular transformations. Thus for continuous-valued random variables \( X : \Omega \rightarrow \mathbb{R}^n \) and \( Y : \Omega \rightarrow \mathbb{R}^p \) and nonsingular matrices \( S_x \in \mathbb{R}^{n \times n}, S_y \in \mathbb{R}^{p \times p}, \) it is true that \((S_x X; S_y Y) = I(X; Y). \) See [20].

Several relations between entropy and mutual information are,

\[
I(Y; X) = H(Y) - H(Y|X) = H(Y) + H(X) - H(X, Y),
\]

(190)

\[
I(Y_1, Y_2; X) = H(Y_1, Y_2) - H(Y_1, Y_2|X) = H(Y_1, Y_2) + H(X) - H(Y_1, Y_2, X).
\]

(191)

The first equality is proven in [8, p. 31, (2.4.26), (2.4.28)] and [6, p. 19, Th. 2.4.1].

Proposition A.1 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 decomposes additively,

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

(192)

Proof The independence of the random variables and the following calculations establish the equality,

\[
I(Y_{1,1}, Y_{1,2}, Y_{2,1}, Y_{2,2}; X_1, X_2)
= H(Y_{1,1}, Y_{1,2}, Y_{2,1}, Y_{2,2}) + H(X_1, X_2) - H(Y_{1,1}, Y_{1,2}, Y_{2,1}, Y_{2,2}, X_1, X_2) \text{ by equation (190),}
= H(Y_{1,1}, Y_{2,1}) + H(Y_{2,1}) + H(X_1, X_2) - H(Y_{1,1}, Y_{2,1}, X_1, X_2)
= \left\{ H(Y_{1,1}, Y_{1,2}) + H(X_1) - H(Y_{1,1}, Y_{1,2}, X_1) \right\} + \left\{ H(Y_{1,2}, Y_{2,2}) + H(X_2) - H(Y_{1,2}, Y_{2,2}, X_2) \right\}
= I(Y_{1,1}, Y_{2,1}; X_1) + I(Y_{1,2}, Y_{2,2}; X_2).
\]

43
Proposition A.2 Consider the random variables $Y_1, Y_2, X$. Assume that conditional dependence holds, $(F_{Y_1}, F_{Y_2}|F_X) \in CI$. Then,

\[
H(Y_1, Y_2|X) = H(Y_1|X) + H(Y_2|X),
\]

\[
I(Y_1, Y_2; X) = H(Y_1, Y_2) - H(Y_1|X) - H(Y_2|X).
\]

Proof From conditional independence and for sets $A_1 \in F_{Y_1}$ and $A_2 \in F_{Y_2}$ follows that

\[
P_{Y_1,Y_2}(A_1 \times A_2|F_X) = P_{Y_1}(A_1|F_X)P_{Y_2}(A_2|F_X),
\]

\[
\log(P_{Y_1,Y_2}(A_1 \times A_2|F_X)) = \log(P_{Y_1}(A_1|F_X)) + \log(P_{Y_2}(A_2|F_X)),
\]

from which the result follows.

\[
I(Y_1, Y_2; X) = H(Y_1, Y_2) - H(Y_1|X) - H(Y_2|X), \quad \text{by \ (191)},
\]

\[
= H(Y_1, Y_2) - H(Y_1|X) - H(Y_2|X), \quad \text{by \ (193)}.
\]

There follow several formulas for entropy and for mutual information of Gaussian random variables. The main reference for this is the paper [20]. See also, [22, Appendix B]. Consider a tuple of jointly Gaussian random variables $(X, Y) \in G(0, Q_{(X,Y)})$ 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,

\[
H(X) = \frac{1}{2} \ln \left( \det(Q_X) \right) + \frac{1}{2} n \ln(2\pi e), \quad \text{by \ (195)}
\]

\[
H(Y|X) = H(X, Y) - H(X) = \frac{1}{2} \ln \left( \det(Q_Y - Q_{XY} Q_{XY}^{-1} Q_{XY}) \right) + \frac{1}{2} p \ln(2\pi e), \quad \text{by \ (196)}
\]

\[
I(Y; X) = - \frac{1}{2} \ln \left( \frac{\det(Q_{(X,Y)})}{\det(Q_Y) \det(Q_X)} \right), \quad \text{by \ (197)}
\]

□
B An Inequality for Determinants

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

Lemma B.1 Consider the real-valued matrices $A, B \in \mathbb{R}^{n \times n}$. Assume that,

$$0 \leq I - A^T A, \quad 0 < I - B^T B, \quad \text{and \ rank}(B) = n. \quad (198)$$

Then,

$$\det([I - A^T A][I - B^T B]) \leq (\det(I - A^T B))^2. \quad (199)$$

A related result is mentioned at [15, Th. 9.E.6]. In that book the proof of corresponding result is referred to the paper [12]. That reference has been received by the authors but they cannot read it because the paper is in Chinese. However, the formulas of the paper they can read. Hua LooKeng developed these results to calculate an orthonormal basis for a function of one complex variable. The proof of Lemma B.1 below is analogous to that of Hua LooKeng in [12]. The main differences are in the assumptions.
Lemma B.2 [12, p. 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 $\text{rank}(B) = n$. Then

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

Proof Obviously,

\[
B^T (I - B B^T) = (I - B^T B)B^T, \quad (I - B^T B)B = B(I - B^T B).
\]

From this equation, the assumptions that $I - B^T B$ is nonsingular and that $\text{rank}(B) = n$ follows that $I - B B^T$ is nonsingular.

Note the following calculations.

\[
B^T (I - B B^T) = (I - B^T B)B^T,
\]

\[
\Leftrightarrow B^T (I - B B^T)^{-1} = (I - B^T B)^{-1}B^T,
\]

\[
\Leftrightarrow (I - B B^T)^{-1}B = B(I - B^T B)^{-1};
\]

\[
I = I - B^T B + B^T B
\]

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

\[
= (I - B^T B) + B^T (I - B B^T)^{-1}B(I - B^T B), \text{ by (201)},
\]

\[
\Leftrightarrow (I - B^T B)^{-1} = I + B^T (I - B B^T)^{-1}B;
\]

\[
I = I - B B^T + B B^T
\]

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

\[
= (I - B B^T) + B(I - B^T B)^{-1}B^T (I - B B^T), \text{ by (203)},
\]

\[
\Leftrightarrow (I - B B^T)^{-1} = I + B(I - B^T B)^{-1}B^T;
\]

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

\[
= I - [I - B B^T]^{-1} + AB^T[I - B B^T]^{-1} + [I - B B^T]^{-1}BA^T + A[I + B^T (I - BB^T)^{-1}B]A^T,
\]

\[
= -B[I - B^T B]^{-1}B^T + A[I - B^T B]^{-1}B + B[I - B^T B]^{-1}A^T + A[I - B^T B]^{-1}A^T
\]

using respectively (205), (202), (203), and (204),

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

□
**Proposition B.3** [12, eq. (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
\[
\det(Q_1) + \det(Q_2) \leq \det(Q) \quad \text{(206)}
\]

**Proof** By [16, Th. 12.7] and because \( Q_1 \) and \( Q_2 \) are positive-definite, there exists a matrix \( S \in \mathbb{R}^{n \times n} \) and a diagonal matrix \( D \in \mathbb{R}^{n \times n} \) with \( D = \text{Diag}(d_1, \ldots, d_n) \geq 0 \) such that \( Q_1 = SS^T \) and \( Q_2 = SDS^T \). Then,
\[
Q = Q_1 + Q_2 = S(I + D)S^T,
\]
\[
\det(Q) = \det(S) \det(I + D) \det(S^T) = \det(S) \prod_{i=1}^n (1 + d_i) \det(S^T)
\]
\[
\geq \det(S) [1 + \prod_{i=1}^n d_i] \det(S^T) = \det(SS^T) + \det(SDS^T) = \det(Q_1) + \det(Q_2).
\]

\[\square\]

**Proof of Lemma B.1.** By the assumptions, \( 0 \leq (I - A^T A), \ 0 < (I - B^T B) \), and \( \text{rank}(B) = n \), from Lemma B.2 follows that
\[
(A - B)[I - B^T B]^{-1}(A - B)^T + [I - A^T A]
\]
\[
= (I - A^T B)[I - B^T B]^{-1}(I - A^T B)^T;
\]
\[
0 \leq \det(I - A^T A), \ \text{by the assumption on } A,
\]
\[
\leq \det((A - B)[I - B^T B]^{-1}(A - B)^T) + \det([I - A^T A]), \ \text{by an assumption on } B,
\]
\[
\leq \det((I - A^T B)[I - B^T B]^{-1}(I - A^T B)^T)
\]
by Lemma B.2, Proposition B.3, and by the assumptions,
\[
= (\det(I - A^T B))^2 [\det([I - B^T B])]^{-1},
\]
\[
\Rightarrow \det([I - A^T A][I - B^T B]) = \det([I - A^T A]) \det([I - B^T B]) \leq (\det([I - A^T B]))^2.
\]

\[\square\]

Another preliminary result is needed.

**Proposition B.4** Consider the matrix \( Q_X \in \mathbb{R}^{n \times n}_{\text{spd}} \) of Prop. 2.14 and the matrix \( D \in \mathbb{R}^{n \times n} \) of Def. 2.2. Thus both \( Q_X > 0 \) and \( D > 0 \). Then,
\[
D \leq Q_X^{-1} \leq D^{-1} \iff D \leq Q_X \leq D^{-1} \quad \text{(207)}
\]
\[
D < Q_X^{-1} < D^{-1} \iff D < Q_X < D^{-1}. \quad \text{(208)}
\]

**Proof** Consider a symmetric matrix \( Q \in \mathbb{R}^{n \times n} \) satisfying \( 0 < Q < I \). Note that in general \( Q \neq Q_X \). Its square root exists and satisfies, \( 0 < Q^{1/2} = (Q^{1/2})^T \in \mathbb{R}^{n \times n} \). Then
\[
Q \leq I \iff I = Q^{-1/2}Q^{1/2} = Q^{-1/2}Q^{1/2} \leq Q^{-1} \iff I \leq Q^{-1}. \quad \text{(209)}
\]

Hence,
\[
Q_X \leq D^{-1} \iff D^{1/2}Q_XD^{1/2} \leq I
\]
\[
\iff I \leq D^{-1/2}Q_X^{-1}D^{-1/2}, \ \text{by equation (209)},
\]
\[
D \geq Q_X^{-1},
\]
\[
D \leq Q_X \iff I \leq D^{-1/2}Q_XD^{-1/2}
\]
\[
\iff D^{1/2}Q_X^{-1}D^{1/2} \leq I, \ \text{by equation (209)},
\]
\[
\iff Q_X^{-1} \leq D^{-1}.
\]

The proof for the case with strict inequalities is similar. \[\square\]
Proposition B.5  Consider the matrices defined in Def. 2.2. 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,
\[
\det \left((I - D^{1/2}Q_X^{-1}D^{1/2})[I - D^{1/2}Q_XD^{1/2}]\right) \leq \det([I - D])^2, \forall Q_X \text{ considered, while}, \quad (210)
\]
\[
\det \left((I - D^{1/2}Q_X^{-1}D^{1/2})[I - D^{1/2}Q_XD^{1/2}]\right) < \det([I - D])^2, \text{ if } Q_X \neq I. \quad (211)
\]

Proof First consider the case of $D \leq Q_X$ but not $D < Q_X$. Then
\[
\det(I - D^{1/2}Q_XD^{1/2}) = \det(D^{1/2}[D - Q_X]D^{1/2}) = 0.
\]
Because $0 < D < I$ the inequality (210) holds. If $Q_X \leq D$ but not $Q_X < D$ then by Proposition B.4 $D \leq Q_X^{-1}$ but not $D < Q_X^{-1}$. Further,
\[
\det(I - D^{1/2}Q_X^{-1}D^{1/2}) = \det(D^{1/2}[D - Q_X^{-1}]D^{1/2}) = 0,
\]
In this case the inequality (210) also holds.

Next consider the case in which $D < Q_X < D^{-1}$. Lemma B.1 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 rank($Q_X$) = $n$.
This and the fact that rank($D$) = $n$ imply that rank($B$) = rank($Q_X^{1/2}D^{1/2}$) = $n$. Note further that,
\[
I - A^T A = I - D^{1/2}Q_X^{-1}D^{1/2},
\]
\[
0 < D \leq Q_X \leq D^{-1} \text{ by assumption}, \Rightarrow \quad 0 < D \leq Q_X^{-1} \leq D^{-1}, \text{ by Proposition B.4},
\]
\[
0 < D^2 \leq D^{1/2}Q_X^{-1}D^{1/2} \leq I \Rightarrow \quad 0 \leq I - D^{1/2}Q_X^{-1}D^{1/2} = I - A^T A;
\]
\[
0 < D < Q_X < D^{-1} \text{ by the case considered}, \quad 0 < D^2 < D^{1/2}Q_XD^{1/2} < I, \Rightarrow \quad 0 < I - D^{1/2}Q_XD^{1/2} = I - B^T B;
\]
\[
I - A^T B = I - D^{1/2}Q_X^{-1/2}Q_X^{1/2}D^{1/2} = I - D.
\]
From Lemma B.1 then follows that,
\[
\det \left((I - D^{1/2}Q_X^{-1}D^{1/2})[I - D^{1/2}Q_XD^{1/2}]\right) \leq \det([I - D])^2.
\]
Next suppose that in addition $Q_X \neq I$. Then
\[
A - B = Q_X^{-1/2}D^{1/2} - Q_X^{1/2}D^{1/2} = Q_X^{-1/2}[I - Q_X]D^{1/2} \neq 0, \text{ using that, } Q_X \neq I,
\]
\[
0 < (A - B)[I - B^T B]^{-1}(A - B)^T, \text{ because } 0 < I - B^T B, \text{ and } A - B \neq 0,
\]
\[
\det(I - A^T A) < \det((A - B)[I - B^T B]^{-1}(A - B)) + \det(I - A^T A),
\]
\[
\leq \det((I - A^T B)[I - B^T B]^{-1}(I - A^T B)^T), \text{ by Lemma B.2},
\]
\[
\Rightarrow \det(I - A^T A)(I - B^T B) = \det(I - A^T A)\det(I - B^T B)
\]
\[
< (\det(I - A^T B))^2, \text{ by Lemma B.2 and its proof},
\]
\[
\Rightarrow \det((I - D^{1/2}Q_X^{-1}D^{1/2})[I - D^{1/2}Q_XD^{1/2}] < \det([I - D])^2, \text{ by substitution of } A \text{ and } B.
\]
Contents

1 Introduction
  1.1 Literature Review ............................................ 4
  1.2 Discussion of Method 1 and Method 2 ......................... 10
  1.3 Structure of the Paper ....................................... 11

2 Probabilistic Properties of Tuples of Random Variables .... 12
  2.1 Notation of Elements of Probability Theory .................. 12
  2.2 Geometric Approach of Gaussian Random Variables and Canonical Variable Form .......... 13
  2.3 Weak Realization of Conditional Independent Gaussian Random Variables ............. 20

3 Calculation of Wyner’s Lossy Common Information .......... 24
  3.1 Solution Procedure .......................................... 26
  3.2 Preliminary Characterization of Conditional Independence ......................... 26
  3.3 The Expression of Wyner’s Lossy Common Information of Multivariate Gaussian Random Variables ................................................. 27
  3.4 Wyner’s Information Common Information of Correlated Gaussian Vectors ............ 28
  3.5 Wyner’s Information Common Information in Case of Private Gaussian Vectors .......... 34
  3.6 Wyner’s Information Common Information in Case of Identical Gaussian Vectors ....... 35
  3.7 Wyner’s Lossy Common Information of Correlated Gaussian Vectors ................ 36
  3.8 Wyner’s Information Common Information of Arbitrary Gaussian Vectors ............ 37

4 Concluding Remarks .............................................. 41

A Information Theory .............................................. 43

B An Inequality for Determinants ................................ 45