

Article

Structural Properties of the Wyner–Ziv Rate Distortion Function: Applications for Multivariate Gaussian Sources [†]

Michail Gkagkos ¹  and Charalambos D. Charalambous ^{2,*} 

¹ Department of Electrical and Computer Engineering, Texas A & M University, College Station, TX 77843, USA; gkagkos@tamu.edu

² Department of Electrical and Computer Engineering, University of Cyprus, P.O. Box 20537, CY-1678 Nicosia, Cyprus

* Correspondence: chadcha@ucy.ac.cy

[†] Preliminary results of this paper are published in 2021 IEEE International Symposium on Information Theory (ISIT), Melbourne, Victoria, Australia.

Abstract: The main focus of this paper is the derivation of the structural properties of the test channels of Wyner’s operational information rate distortion function (RDF), $\bar{R}(\Delta_X)$, for arbitrary abstract sources and, subsequently, the derivation of additional properties for a tuple of multivariate correlated, jointly independent, and identically distributed Gaussian random variables, $\{X_t, Y_t\}_{t=1}^\infty$, $X_t : \Omega \rightarrow \mathbb{R}^{n_x}$, $Y_t : \Omega \rightarrow \mathbb{R}^{n_y}$, with average mean-square error at the decoder and the side information, $\{Y_t\}_{t=1}^\infty$, available only at the decoder. For the tuple of multivariate correlated Gaussian sources, we construct optimal test channel realizations which achieve the informational RDF, $\bar{R}(\Delta_X) \triangleq \inf_{\mathcal{M}(\Delta_X)} I(X; Z|Y)$, where $\mathcal{M}(\Delta_X)$ is the set of auxiliary RVs Z such that $\mathbf{P}_{Z|X,Y} = \mathbf{P}_{Z|X}$, $\hat{X} = f(Y, Z)$, and $\mathbb{E}\{\|X - \hat{X}\|^2\} \leq \Delta_X$. We show the following fundamental structural properties: (1) Optimal test channel realizations that achieve the RDF and satisfy conditional independence, $\mathbf{P}_{X|\hat{X},Y,Z} = \mathbf{P}_{X|\hat{X},Y} = \mathbf{P}_{X|\hat{X}}$, $\mathbb{E}\{X|\hat{X}, Y, Z\} = \mathbb{E}\{X|\hat{X}\} = \hat{X}$. (2) Similarly, for the conditional RDF, $R_{X|Y}(\Delta_X)$, when the side information is available to both the encoder and the decoder, we show the equality $\bar{R}(\Delta_X) = R_{X|Y}(\Delta_X)$. (3) We derive the water-filling solution for $R_{X|Y}(\Delta_X)$.

Keywords: Wyner’s side information; multivariate Gaussian sources; test channel distributions



Citation: Gkagkos, M.; Charalambous, C.D. Structural Properties of the Wyner–Ziv Rate Distortion Function: Applications for Multivariate Gaussian Sources. *Entropy* **2024**, *26*, 306. <https://doi.org/10.3390/e26040306>

Academic Editor: Eduard Jorswieck

Received: 4 January 2024

Revised: 7 March 2024

Accepted: 20 March 2024

Published: 29 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction, Problem Statement, and Main Results

1.1. The Wyner and Ziv Lossy Compression Problem and Generalizations

Wyner and Ziv [1] derived an operational information definition for the lossy compression problem in Figure 1 with respect to a single-letter fidelity of reconstruction. The joint sequence of random variables (RVs) $\{(X_t, Y_t) : t = 1, 2, \dots\}$ takes values in sets of finite cardinality, $\{\mathcal{X}, \mathcal{Y}\}$, and it is generated independently according to the joint probability distribution function $\mathbf{P}_{X,Y}$. Wyner [2] generalized [1] to RVs $\{(X_t, Y_t) : t = 1, 2, \dots\}$ that take values in abstract alphabet spaces $\{\mathcal{X}, \mathcal{Y}\}$ and hence include continuous-valued RVs.

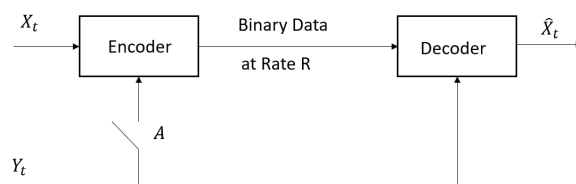


Figure 1. The Wyner and Ziv [1] block diagram of lossy compression. If switch A is closed, then the side information is available at both the encoder and the decoder; if switch A is open, the side information is only available at the decoder.

(A) *Switch “A” Closed:* When the side information $\{Y_t : t = 1, 2, \dots\}$ is available non-causally at both the encoder and the decoder, Wyner [2] (see also Berger [3]) characterized the infimum of all achievable operational rates (denoted by $\bar{R}_1(\Delta_X)$ in [2]), subject to a single-letter fidelity with average distortion less than or equal to $\Delta_X \in [0, \infty)$. The rate is given by the single-letter operational information theoretic conditional RDF:

$$R_{X|Y}(\Delta_X) \triangleq \inf_{\mathcal{M}_0(\Delta_X)} I(X; \hat{X}|Y) \in [0, \infty], \quad \Delta_X \in [0, \infty) \quad (1)$$

$$= \inf_{\mathbf{P}_{\hat{X}|X,Y} : \mathbf{E}\{d_X(X, \hat{X})\} \leq \Delta_X} I(X; \hat{X}|Y) \quad (2)$$

where $\mathcal{M}_0(\Delta_X)$ is the set specified by

$$\mathcal{M}_0(\Delta_X) \triangleq \left\{ \hat{X} : \Omega \rightarrow \hat{\mathcal{X}} : \mathbf{P}_{X,Y,\hat{X}} \text{ is the joint measure on } \mathcal{X} \times \mathcal{Y} \times \hat{\mathcal{X}}, \right. \\ \left. \mathbf{E}\{d_X(X, \hat{X})\} \leq \Delta_X \right\}, \quad (3)$$

and $\hat{X}$ is the reproduction of X . $I(X; \hat{X}|Y)$ is the conditional mutual information between X and $\hat{X}$ conditioned on Y , and $d_X(\cdot, \cdot)$ is the fidelity criterion between x and $\hat{x}$. The infimum in (1) is over all elements of $\mathcal{M}_0(\Delta_X)$ with induced joint distributions $\mathbf{P}_{X,Y,\hat{X}}$ of the RVs $(X, Y, \hat{X})$ such that the marginal distribution $\mathbf{P}_{X,Y}$ is the fixed joint distribution of the source (X, Y) . This problem is equivalent to (2) [4].

(B) *Switch “A” Open:* When the side information is available non-causally only at the decoder, Wyner [2] characterized the infimum of all achievable operational rates (denoted by $R^*(\Delta_X)$ in [2]), subject to a single-letter fidelity with average distortion less than or equal to Δ_X . The rate is given by the single-letter operational information theoretic RDF as a function of an auxiliary RV $Z : \Omega \rightarrow \mathcal{Z}$:

$$\bar{R}(\Delta_X) \triangleq \inf_{\mathcal{M}(\Delta_X)} \left\{ I(X; Z) - I(Y; Z) \right\} \in [0, \infty], \quad \Delta_X \in [0, \infty) \quad (4)$$

$$= \inf_{\mathcal{M}(\Delta_X)} I(X; Z|Y) \quad (5)$$

where $\mathcal{M}(\Delta_X)$ is specified by the set of auxiliary RVs Z and defined as:

$$\mathcal{M}(\Delta_X) \triangleq \left\{ Z : \Omega \rightarrow \mathcal{Z} : \mathbf{P}_{X,Y,Z,\hat{X}} \text{ is the joint measure on } \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \hat{\mathcal{X}}, \right.$$

$$\left. \mathbf{P}_{Z|X,Y} = \mathbf{P}_{Z|X}, \exists \text{ meas. fun. } f : \mathcal{Y} \times \mathcal{Z} \rightarrow \hat{\mathcal{X}}, \hat{X} = f(Y, Z), \mathbf{E}\{d_X(X, \hat{X})\} \leq \Delta_X \right\}. \quad (6)$$

Wyner’s realization of the joint measure $\mathbf{P}_{X,Y,Z,\hat{X}}$ induced by the RVs $(X, Y, Z, \hat{X})$ is illustrated in Figure 2, where Z is the output of the “test channel”, $\mathbf{P}_{Z|X}$. Clearly, $\bar{R}(\Delta_X)$ involves two strategies, i.e., $f(\cdot, \cdot)$ and $\mathbf{P}_{Z|X,Y} = \mathbf{P}_{Z|X}$. This makes it a much more complex problem compared to $R_{X|Y}(\Delta_X)$ (which involves only $\mathbf{P}_{\hat{X}|X,Y}$).

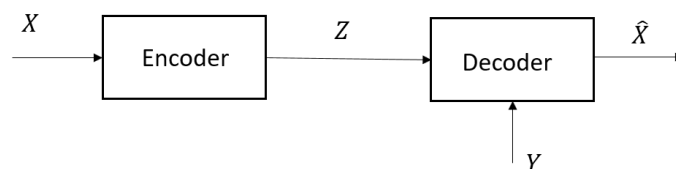


Figure 2. Test channel when side information is only available to the decoder.

Throughout [2], the following assumption is imposed.

Assumption 1. $I(X; Y) < \infty$ (see [2]).

Wyner [2] considered scalar-valued jointly Gaussian RVs (X, Y) with square-error distortion and constructed the optimal realizations $\hat{X}$ and $(Z, \hat{X})$ and the function $f(X, Z)$ from the sets $\mathcal{M}_0(\Delta_X)$ and $\mathcal{M}(\Delta_X)$, respectively. Also, it is shown that these realizations achieve the characterizations of the RDFs $R_{X|Y}(\Delta_X)$ and $\bar{R}(\Delta_X)$, respectively, and that the two rates are equal, i.e., $\bar{R}(\Delta_X) = R_{X|Y}(\Delta_X)$.

(C) *Marginal RDF*: If there is no side information, $\{Y_t : t = 1, 2, \dots\}$, or the side information is independent of the source, $\{X_t : t = 1, 2, \dots\}$, the RDFs $R_{X|Y}(\Delta_X)$ and $\bar{R}(\Delta_X)$ degenerate to the marginal RDF $R_X(\Delta_X)$, defined by

$$R_X(\Delta_X) \triangleq \inf_{\mathbf{P}_{\hat{X}|X}: \mathbf{E}\{d_X(X, \hat{X})\} \leq \Delta_X} I(X; \hat{X}) \in [0, \infty], \quad \Delta_X \in [0, \infty). \quad (7)$$

(D) *Gray's Lower Bounds*: A lower bound on $R_{X|Y}(\Delta_X)$ is given by Gray in [4] [Theorem 3.1]. This bound connects $R_{X|Y}(\Delta_X)$ with the marginal RDF and the mutual information between X and Y as follows:

$$R_{X|Y}(\Delta_X) \geq R_X(\Delta_X) - I(X; Y). \quad (8)$$

Clearly, the lower bound is trivial for values of $\Delta_X \in [0, \infty)$ such that $R_X(\Delta_X) - I(X; Y) < 0$.

1.2. Main Contributions of the Paper

We first consider Wyner's [2] RDFs $R_{X|Y}(\Delta_X)$ and $\bar{R}(\Delta_X)$ for arbitrary RVs (X, Y) defined on abstract alphabet spaces, and we derive structural properties of the realizations that achieve the two optimal test channels. Subsequently, we generalize Wyner's [2] results to multivariate-valued jointly Gaussian RVs (X, Y) . In other words, we construct the optimal multivariate-valued realizations $\hat{X}$ and $(\hat{X}, Z)$ and the function $f(X, Z)$ which achieve the RDFs $R_{X|Y}(\Delta_X)$ and $\bar{R}(\Delta_X)$, respectively. In the literature, it is often called achievability of the converse coding theorem. In addition, we use the realizations to prove the equality $\bar{R}(\Delta_X) = R_{X|Y}(\Delta_X)$ and to derive the water-filling solution. Along the way, we verify that our results reproduce, for scalar-valued RVs (X, Y) , Wyner [2] RDFs and the optimal realizations. However, to our surprise, the existing results from the literature [[5], Theorem 4 and Abstract and [6], Theorem 3A], which deal with the more general multivariate-valued remote sensor problem (the RDF of the remote sensor problem is a generalization of Wyner's RDF $\bar{R}(\Delta_X)$, with the encoder observing a noisy version of the RVs generated by the source), do not degenerate to Wyner's [2] RDFs, when specialized to scalar-valued RVs (we verify this in Remark 5 by also checking the correction suggested in <https://tiangroup.engr.tamu.edu/publications/>) (accessed on 3 January 2024). In Section 1.3, we give a detailed account of the main results of this paper. We should emphasize that preliminary results of this paper appeared in [7], mostly without the details of the proofs. This paper is extended [7] and contains complete proofs of the preliminary results of [7], which in some cases are lengthy (see, for example, Section 4, proofs of Theorems 3–5, Corollaries 1 and 2, etc.).

1.3. Problem Statement and Main Results

(a) We consider a tuple of jointly independent and identically distributed (i.i.d.) arbitrary RVs $(X^n, Y^n) = \{(X_t, Y_t) : t = 1, 2, \dots, n\}$ defined on abstract alphabet spaces, and we derive the following results.

(a.1) **Lemma 1**: Achievable lower bound on the conditional mutual information $I(X; \hat{X}|Y)$, which strengthens Gray's lower bound (8) [[4], Theorem 3.1].

(a.2) **Theorem 2**: Structural properties of the optimal reconstruction $\hat{X}$, which achieves a lower bound on $R_{X|Y}(\Delta_X)$ for mean-square error distortion. Theorem 2 strengthens the conditions for the equality to hold, $R_{X|Y}(\Delta_X) = \bar{R}(\Delta_X)$, given by Wyner [2] [Remarks, p. 65]

(see Remark 1). However, for finite-alphabet-valued sources with Hamming distance distortion, it might be the case that $R_{X|Y}(\Delta_X) < \bar{R}(\Delta_X)$, as pointed out by Wyner and Ziv [1] [Section 3] for the doubly symmetric binary source.

(b) We consider a tuple of jointly i.i.d. multivariate Gaussian RVs $(X^n, Y^n) = \{(X_t, Y_t) : t = 1, 2, \dots, n\}$, with respect to the square-error fidelity, as defined below.

$$X_t : \Omega \rightarrow \mathbb{R}^{n_x} = \mathcal{X}, Y_t : \Omega \rightarrow \mathbb{R}^{n_y} = \mathcal{Y}, t = 1, 2, \dots, n, \quad (9)$$

$$X_t \in N(0, Q_X), \quad Y_t \in N(0, Q_Y), \quad (10)$$

$$Q_{(X_t, Y_t)} = \mathbb{E} \left\{ \begin{bmatrix} X_t \\ Y_t \end{bmatrix} \begin{bmatrix} X_t \\ Y_t \end{bmatrix}^T \right\} = \begin{bmatrix} Q_X & Q_{X,Y} \\ Q_{X,Y}^T & Q_Y \end{bmatrix}, \quad (11)$$

$$\mathbf{P}_{X_t, Y_t} = \mathbf{P}_{X,Y} \text{ multivariate Gaussian distribution,} \quad (12)$$

$$\hat{X}_t : \Omega \rightarrow \mathbb{R}^{n_x} = \mathcal{X}, \quad t = 1, 2, \dots, n, \quad (13)$$

$$D_X(x^n, \hat{x}^n) = \frac{1}{n} \sum_{t=1}^n \|x_t - \hat{x}_t\|_{\mathbb{R}^{n_x}}^2, \quad (14)$$

where n_x, n_y are arbitrary positive integers, $X \in N(0, Q_X)$ means X is a Gaussian RV, with zero mean and covariance matrix Q_X , and $\|\cdot\|_{\mathbb{R}^{n_x}}^2$ is the Euclidean distance on $\mathbb{R}^{n_x}$. To give additional insight we often consider the following realization of side information (the condition $DD^T \succ 0$ ensures $I(X; Y) < \infty$, and hence, Assumption 1 is respected).

$$Y_t = CX_t + DV_t, \quad (15)$$

$$V_t \in N(0, Q_V), \quad (16)$$

$$C \in \mathbb{R}^{n_y \times n_x}, \quad D \in \mathbb{R}^{n_y \times n_y}, \quad DD^T \succ 0, \quad Q_V = I_{n_y}, \quad (17)$$

$$V^n \text{ independent of } X^n, \quad (18)$$

where I_{n_y} denotes the $n_y \times n_y$ identity matrix. For the above specification of the source and distortion criterion, we derive the following results.

(b.1) Theorems 3 and 4: Structural properties of optimal realization of $\hat{X}$, which achieves $R_{X|Y}(\Delta_X)$, its closed form expression.

(b.2) Theorem 5: Structural properties of optimal realization of $\hat{X}$ and $\hat{X} = f(Y, Z)$, which achieve $\bar{R}(\Delta_X)$ and the closed form expression of $\bar{R}(\Delta_X)$.

(b.3) A proof that $\bar{R}(\Delta_X)$ and $R_{X|Y}(\Delta_X)$ coincide: Calculation of the distortion region such that Gray's lower bound (8) holds with equality.

In Remark 4, we consider the tuple of scalar-valued, jointly Gaussian RVs (X, Y) with square error distortion function and verify that our optimal realizations of $\hat{X}$ and the closed form expressions for $R_{X|Y}(\Delta_X)$ and $\bar{R}(\Delta_X)$ are identical to Wyner's [2] realizations and RDFs.

We should emphasize that our methodology is different from past studies in the sense that we focus on the structural properties of the realizations of the test channels, that achieve the characterizations of the two RDFs (i.e., verification of the converse coding theorem). Our derivations are generic and bring new insight into the construction of realizations that induce the optimal test channels of other distributed source coding problems (i.e., establishing the achievability of the converse coding theorem).

1.4. Additional Generalizations of the Wyner-Ziv [1] and Wyner [2] RDFs

Below, we discuss additional generalizations of Wyner and Ziv [1] and Wyner's [2] RDFs.

(A) *Draper and Wornell [8] Distributed Remote Source Coding Problem:* Draper and Wornell [8] generalized the RDF $\bar{R}(\Delta_X)$, when the source to be estimated at the decoder is $S : \Omega \rightarrow \mathcal{S}$, and it is not directly observed at the encoder. Rather, the encoder observes a RV $X : \Omega \rightarrow \mathcal{X}$ (which is correlated with S), while the decoder observes another RV, as side information, $Y : \Omega \rightarrow \mathcal{Y}$, which provides information on (S, X) . The aim is to reconstruct S at the decoder by $\hat{S} : \Omega \rightarrow \hat{\mathcal{S}}$, subject to an average distortion $\mathbb{E}\{d_S(S, \hat{S})\} \leq \Delta_S$, by

a function $\hat{S} = f(Y, Z)$. The RDF for this problem, called the distributed remote source coding problem, is defined by [8]

$$\bar{R}^{PO}(\Delta_S) = \inf_{\mathcal{M}^{PO}(\Delta_S)} I(X; Z|Y) \in [0, \infty], \quad (19)$$

where $\mathcal{M}^{PO}(\Delta_S)$ is specified by the set of auxiliary RVs Z , and defined as:

$$\begin{aligned} \mathcal{M}^{PO}(\Delta_S) \triangleq \{ & Z : \Omega \rightarrow \mathcal{Z} : \mathbf{P}_{S,X,Y,Z,\hat{X}} \text{ is the joint measure on } \mathcal{S} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \hat{\mathcal{X}}, \\ & \mathbf{P}_{Z|S,X,Y} = \mathbf{P}_{Z|X}, \exists \text{ measurable function } f^{PO} : \mathcal{Y} \times \mathcal{Z} \rightarrow \hat{\mathcal{S}}, \\ & \hat{S} = f^{PO}(Y, Z), \mathbf{E}\{d_S(S, \hat{S})\} \leq \Delta_S \}. \end{aligned} \quad (20)$$

Clearly, if $S = X$ -a.s (almost surely), then $\bar{R}^{PO}(\Delta_S)$ degenerates (this implies the optimal test channel that achieves the characterization of the RDF $\bar{R}^{PO}(\Delta_S)$ should degenerate to the optimal test channel that achieves the characterization of the RDF $\bar{R}(\Delta_X)$) to $\bar{R}(\Delta_X)$. For scalar-valued jointly Gaussian RVs $(S, X, Y, Z, \hat{X})$ with square-error distortion, Draper and Wornell [8] [Equation (3) and Appendix A.1] derived the characterization of the RDF $\bar{R}^{PO}(\Delta_S)$ and constructed the optimal realization $\hat{S} = f^{PO}(Y, Z)$, which achieves this characterization.

In [5,6], the authors investigated the RDF $\bar{R}^{PO}(\Delta_S)$ of [8] for the multivariate jointly Gaussian RVs $(S, X, Y, Z, \hat{X})$, with square-error distortion, and derived a characterization for the RDF $\bar{R}^{PO}(\Delta_S)$ in [[5], Theorem 4] and [[6], Theorem 3A] (see [[6], Equation (26)]). However, it will become apparent in Remark 5 that, when $S = X$ -almost surely (a.s.), and hence $\bar{R}^{PO}(\Delta_S) = \bar{R}(\Delta_X)$, the RDFs given in [[5], Theorem 4] and [[6], Theorem 3A], do not produce Wyner's [2] value. We also show in Remark 5 that the same technical issues occur for the correction suggested in <https://tiangroup.engr.tamu.edu/publications/> (accessed on 3 January 2024). Similarly, when $S = X$ -a.s. and $Y = X$ -a.s. [[5], Theorem 4] and [[6], Theorem 3A], do not produce the classical RDF $R_X(\Delta_X)$ of the Gaussian source X .

(B) *Additional Literature Review:* The formulation of Figure 1 is generalized to other multiterminal or distributed lossy compression problems, such as relay networks, sensor networks, etc., under various code formulations and assumptions. Oohama [9] analyzed lossy compression problems for a tuple of scalar correlated Gaussian memoryless sources with square error distortion criterion. Also, he determined the rate-distortion region, in the special case when one source provides partial side information to the other source. Furthermore, Oohama in [10] analyzed separate lossy compression problems for $L + 1$ scalar correlated Gaussian memoryless sources, when L of the sources provide partial side information at the decoder for the reconstruction of the remaining source and gave a partial answer to the rate distortion region. Additionally, ref. [10] proved that the problem of [10] includes, as a special case, the additive white Gaussian CEO problem analyzed by Viswanathan and Berger [11]. Extensions of [10] are derived by Ekrem and Ulukus [12] and Wang and Chen [13], where an outer bound on the rate region is derived for the vector Gaussian multiterminal source. Additional works are [14–16] and the references therein.

The vast literature on multiterminal or distributed lossy compression of jointly Gaussian sources with square-error distortion (including the references mentioned above), is often confined to scalar-valued correlated RVs. Moreover, as easily verified, not much emphasis is given in the literature on the structural properties of the realizations of RVs that induce the optimal test channels that achieve the characterizations of the RDFs.

The rest of the paper is organized as follows. In Section 2, we review Wyner's [2] operational definition of lossy compression. We also state a fundamental theorem on mean-square estimation that we use throughout the paper regarding the analysis of (b). The main Theorems are presented in Section 3; some of the proofs, including the structural properties, are given in Section 4. Connections between our results and the past literature

are provided in Section 5. A simulation to show the gap between the two rates is given in the same section.

2. Preliminaries

In this section, we review the Wyner [2] source coding problems with fidelity in Figure 1. We begin with the notation, which follows closely [2].

2.1. Notation

Let $\mathbb{Z} \triangleq \{\dots, -1, 0, 1, \dots\}$ the set of all integers, $\mathbb{N} \triangleq \{0, 1, 2, \dots\}$ the set of natural integers, $\mathbb{Z}_+ \triangleq \{1, 2, \dots\}$. For $n \in \mathbb{Z}_+$, denote the following finite subset of the above defined set, $\mathbb{Z}_n \triangleq \{1, 2, \dots, n\}$. Denote the real numbers by $\mathbb{R}$ and the set of positive and of strictly positive real numbers, by $\mathbb{R}_+ = [0, \infty)$ and $\mathbb{R}_{++} = (0, \infty)$, respectively.

For any matrix $A \in \mathbb{R}^{p \times m}$, $(p, m) \in \mathbb{Z}_+ \times \mathbb{Z}_+$, we denote its kernel by $\ker(A)$ its transpose by A^T , and for $m = p$, we denote its trace by $\text{trace}(A)$, and by $\text{diag}\{A\}$, the matrix with diagonal entries A_{ii} , $i \in \mathbb{Z}_p$, and zero elsewhere. The determinant of a square matrix A is denoted by $\det(A)$. The identity matrix with dimensions $p \times p$ is designated as I_p . Denote an arbitrary set or space by $\mathcal{U}$ and the product space formed by n copies of it by $\mathcal{U}^n \triangleq \times_{i=1}^n \mathcal{U}$. $u^n \in \mathcal{U}^n$ denotes the set of n -tuples $u^n \triangleq (u_1, u_2, \dots, u_n)$, where $u_k \in \mathcal{U}$, $k \in \mathbb{Z}_n$ are its coordinates. Denote a probability space by $(\Omega, \mathcal{F}, \mathbb{P})$. For a sub-sigma-field $\mathcal{G} \subseteq \mathcal{F}$, and $A \in \mathcal{F}$, denote by $\mathbb{P}(A|\mathcal{G})$ the conditional probability of A given $\mathcal{G}$; i.e., $\mathbb{P}(A|\mathcal{G}) = \mathbb{P}(A|\mathcal{G})(\omega)$, $\omega \in \Omega$ is a measurable function on Ω .

On the above probability space, consider two-real valued random variables (RV) $X : \Omega \rightarrow \mathcal{X}$, $Y : \Omega \rightarrow \mathcal{Y}$, where $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ are arbitrary measurable spaces. The measure (or joint distribution if $\mathcal{X}, \mathcal{Y}$ are Euclidean spaces) induced by (X, Y) on $\mathcal{X} \times \mathcal{Y}$ is denoted by $\mathbf{P}_{X,Y}$ or $\mathbf{P}(dx, dy)$ and their marginals on $\mathcal{X}$ and $\mathcal{Y}$ by $\mathbf{P}_X$ and $\mathbf{P}_Y$, respectively. The conditional measure of RV X conditioned on Y is denoted by $\mathbf{P}_{X|Y}$ or $\mathbf{P}(dx|y)$, when $Y = y$ is fixed. On the above probability space, consider three-real values RVs $X : \Omega \rightarrow \mathcal{X}$, $Y : \Omega \rightarrow \mathcal{Y}$, $Z : \Omega \rightarrow \mathcal{Z}$. We say that RVs (Y, Z) are conditional independent given RV X if $\mathbf{P}_{Y,Z|X} = \mathbf{P}_{Y|X}\mathbf{P}_{Z|X}$ -a.s (almost surely) or equivalently $\mathbf{P}_{Z|X,Y} = \mathbf{P}_{Z|X}$ -a.s; the specification a.s is often omitted. We often denote the above conditional independence by the Markov chain (MC) $Y \leftrightarrow X \leftrightarrow Z$.

Finally, for RVs X, Y , etc., $H(X)$ denotes differential entropy of X , $H(X|Y)$ conditional differential entropy of X given Y , and $I(X; Y)$ the mutual information between X and Y , as defined in standard books on information theory [17,18]. We use $\log(\cdot)$ to denote the natural logarithm. The notation $X \in N(0, Q_X)$ means X is a Gaussian distributed RV with zero mean and covariance $Q_X \succeq 0$, where $Q_X \succeq 0$ (resp. $Q_X \succ 0$) means Q_X is positive semidefinite (respectively, positive definite). We denote the covariance of X and Y by

$$Q_{X,Y} \triangleq \text{cov}(X, Y). \quad (21)$$

We denote the covariance of X conditioned on Y by

$$\begin{aligned} Q_{X|Y} &\triangleq \text{cov}(X, X|Y) \\ &= \mathbf{E}\left\{\left(X - \mathbf{E}(X|Y)\right)\left(X - \mathbf{E}(X|Y)\right)^T\right\} \text{ if } (X, Y) \text{ is jointly Gaussian,} \end{aligned} \quad (22)$$

where the second equality is due to a property of jointly Gaussian RVs.

2.2. Mean-Square Estimation of Conditionally Gaussian RVs

Below, we state a well-known property of conditionally Gaussian RVs from [19], which we use in our derivations.

Proposition 1. *Conditionally Gaussian RVs [19]. Consider a pair of multivariate RVs $X = (X_1, \dots, X_{n_x})^T : \Omega \rightarrow \mathbb{R}^{n_x}$ and $Y = (Y_1, \dots, Y_{n_y})^T : \Omega \rightarrow \mathbb{R}^{n_y}$, $(n_x, n_y) \in \mathbb{Z}_+ \times \mathbb{Z}_+$, defined on some probability distribution $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{G} \subseteq \mathcal{F}$ be a sub- σ -algebra. Assume the conditional distribution of (X, Y) conditioned on $\mathcal{G}$, i.e., $\mathbf{P}(dx, dy|\mathcal{G})$ is $\mathbb{P}$ -a.s. (almost surely) Gaussian, with conditional means*

$$\mu_{X|\mathcal{G}} \triangleq \mathbf{E}\{X|\mathcal{G}\}, \quad \mu_{Y|\mathcal{G}} \triangleq \mathbf{E}\{Y|\mathcal{G}\}, \quad (23)$$

and conditional covariances

$$Q_{X|\mathcal{G}} \triangleq \text{cov}(X, X|\mathcal{G}), \quad Q_{Y|\mathcal{G}} \triangleq \text{cov}(Y, Y|\mathcal{G}), \quad (24)$$

$$Q_{X,Y|\mathcal{G}} \triangleq \text{cov}(X, Y|\mathcal{G}). \quad (25)$$

Then, the vectors of conditional expectations $\mu_{X|Y,\mathcal{G}} \triangleq \mathbf{E}\{X|Y, \mathcal{G}\}$ and matrices of conditional covariances $Q_{X|Y,\mathcal{G}} \triangleq \text{cov}(X, X|Y, \mathcal{G})$ are given, $\mathbb{P}$ -a.s., by the following expressions (If $Q_{Y|\mathcal{G}} \succ 0$ then the inverse exists and the pseudoinverse is $Q_{Y|\mathcal{G}}^+ = Q_{Y|\mathcal{G}}^{-1}$):

$$\mu_{X|Y,\mathcal{G}} = \mu_{X|\mathcal{G}} + Q_{X,Y|\mathcal{G}} Q_{Y|\mathcal{G}}^+ (Y - \mu_{Y|\mathcal{G}}), \quad (26)$$

$$Q_{X|Y,\mathcal{G}} \triangleq Q_{X|\mathcal{G}} - Q_{X,Y|\mathcal{G}} Q_{Y|\mathcal{G}}^+ Q_{X,Y|\mathcal{G}}^T. \quad (27)$$

If $\mathcal{G}$ is the trivial information, i.e., $\mathcal{G} = \{\Omega, \emptyset\}$, then $\mathcal{G}$ is removed from the above expressions.

Note that, if $\mathcal{G} = \{\Omega, \emptyset\}$, then (26) and (27) reduce to the well-known conditional mean and conditional covariance of X conditioned on Y .

For Gaussian RVs, we make use of the following properties.

Proposition 2. *Let $X : \Omega \rightarrow \mathbb{R}^n$, $n \in \mathbb{Z}_+$, $X \in N(0, Q_X)$, $Q_X \succeq 0$, $S \in \mathbb{R}^{n_1 \times n}$, $n_1 \in \mathbb{Z}_+$, and denote by $\mathcal{F}^X$ and $\mathcal{F}^{SX}$ the σ -algebra generated by the RVs X and SX , respectively. The following hold.*

(a) $\mathcal{F}^{SX} \subseteq \mathcal{F}^X$.

(b) $\mathcal{F}^{SX} = \mathcal{F}^X$ if and only if $\ker(Q_X) = \ker(SQ_X)$.

Proof. This is well-known in measure theory, see [20]. $\square$

Proposition 3. *Let $X : \Omega \rightarrow \mathbb{R}^n$, $n \in \mathbb{Z}_+$, $X \in N(0, Q_X)$, $Q_X \succeq 0$, $\text{rank}(Q_X) = n_1$, $n_1 \in \mathbb{Z}_+$, $n_1 < n$. Then, there exists a linear transformation $S \in \mathbb{R}^{n_1 \times n}$ such that, if $X_1 : \Omega \rightarrow \mathbb{R}^{n_1}$, $X_1 = SX$, then $X_1 \in N(0, Q_{X_1})$, $Q_{X_1} \succ 0$, $\mathcal{F}^X = \mathcal{F}^{X_1}$.*

Proof. This is well-known in probability theory, see [20]. $\square$

2.3. Wyner's Coding Theorems with Side Information at the Decoder

For the sake of completeness, we introduce certain results from Wyner's work in [2], which we use in this paper. On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, consider a tuple of jointly i.i.d. RVs $(X^n, Y^n) = \{(Y_t, X_t) : t \in \mathbb{Z}_n\}$,

$$X_t : \Omega \rightarrow \mathcal{X}, \quad Y_t : \Omega \rightarrow \mathcal{Y}, \quad t \in \mathbb{Z}_n, \quad (28)$$

with induced distribution $\mathbf{P}_{X_t, Y_t} = \mathbf{P}_{X, Y}, \forall t$. Consider also the measurable function $d_X : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow [0, \infty)$, for a measurable space $\hat{\mathcal{X}}$. Let

$$\mathcal{I}_M \triangleq \{0, 1, \dots, M-1\}, \quad M \in \mathbb{Z}_M, \quad (29)$$

be a finite set.

A code (n, M, D_X) , when switch “A” is open (see Figure 1), is defined by two measurable functions, the encoder F_E and the decoder F_D , with average distortion, as follows.

$$F_E : \mathcal{X}^n \longrightarrow \mathcal{I}_M, \quad F_D : \mathcal{I}_M \times \mathcal{Y}^n \longrightarrow \hat{\mathcal{X}}^n, \quad (30)$$

$$\frac{1}{n} \mathbf{E} \left\{ \sum_{t=1}^n d_X(X_t, \hat{X}_t) \right\} = D_X, \quad (31)$$

where $\hat{X}^n$ is again a sequence of RVs, $\hat{X}^n = F_D(F_E(X^n), Y^n) \in \hat{\mathcal{X}}^n$. A non-negative rate distortion pair (R, Δ_X) is said to be *achievable* if for every $\epsilon > 0$, and n sufficiently large, there exists a code (n, M, D_X) such that

$$M \leq 2^{n(R+\epsilon)}, \quad D_X \leq \Delta_X + \epsilon. \quad (32)$$

Let $\mathcal{R}$ denote the set of all achievable pairs (R, Δ_X) , and define, for $\Delta_X \geq 0$, the infimum of all achievable rates by

$$R^*(\Delta_X) = \inf_{(R, \Delta_X) \in \mathcal{R}} R. \quad (33)$$

If for some Δ_X there is no $R < \infty$ such that $(R, \Delta_X) \in \mathcal{R}$, then set $R^*(\Delta_X) = \infty$. For arbitrary abstract spaces Wyner [2] characterized the infimum of all achievable rates $R^*(\Delta_X)$ by the single-letter RDF, $\bar{R}(\Delta_X)$ given by (5) and (6), in terms of an auxiliary RV $Z : \Omega \rightarrow \mathcal{Z}$. Wyner’s realization of the joint measure $\mathbf{P}_{X, Y, Z, \hat{X}}$ induced by the RVs $(X, Y, Z, \hat{X})$ is illustrated in Figure 2, where Z is the output of the “test channel”, $\mathbf{P}_{Z|X}$. Wyner proved the following coding theorems.

Theorem 1. Wyner [[2], Theorems, pp. 64–65]. Suppose Assumption 1 holds.

(a) Converse Theorem. For any $\Delta_X \geq 0$, $R^*(\Delta_X) \geq \bar{R}(\Delta_X)$.

(b) Direct Theorem. If the conditions stated in [2, pages 64–65, (i), (ii)] hold, then $R^*(\Delta_X) \leq \bar{R}(\Delta_X)$, $0 \leq \Delta_X < \infty$.

In Figure 1, when switch A is closed and the tuple of jointly independent and identically distributed RVs (X^n, Y^n) is defined as in Section 2.3, Wyner [2] generalized Berger’s [3] characterization of all achievable pairs (R, Δ_X) , from finite alphabet spaces to abstract alphabet spaces.

A code (n, M, D_X) , when switch “A” is closed, (see Figure 1), is defined as in Section 2.3, with the encoder F_E , replaced by

$$F_E : \mathcal{X}^n \times \mathcal{Y}^n \longrightarrow \mathcal{I}_M. \quad (34)$$

Let $\mathcal{R}_1$ denote the set of all achievable pairs (R, Δ_X) , again as defined in Section 2.3. For $\Delta_X \geq 0$, define the infimum of all achievable rates by

$$\bar{R}_1(\Delta_X) = \inf_{(R, \Delta_X) \in \mathcal{R}_1} R. \quad (35)$$

Wyner [2] characterized the infimum of all achievable rates $\bar{R}_1(\Delta_X)$ by the single-letter RDF $R_{X|Y}(\Delta_X)$ given by (1) and (3). The coding Theorems are given by Theorem 1 with $R^*(\Delta_X)$ and $\bar{R}(\Delta_X)$ replaced by $\bar{R}_1(\Delta_X)$ and $R_{X|Y}(\Delta_X)$, respectively. That is, $\bar{R}_1(\Delta_X) = R_{X|Y}(\Delta_X)$ (using Wyner’s notation [[2], Appendix A.1]) These coding theorems generalized earlier work of Berger [3] for finite alphabet spaces. Wyner also derived a fundamental lower bound on $R^*(\Delta_X)$ in terms of $\bar{R}_1(\Delta_X)$, as stated in the next remark.

Remark 1. Wyner [[2], Remarks, p. 65]

(A) For $Z \in \mathcal{M}(\Delta_X)$, $\hat{X} = f(Y, Z)$, and thus $\mathbf{P}_{Z|X,Y} = \mathbf{P}_{Z|X}$. Then, by a property of conditional mutual information and the data processing inequality:

$$I(X; Z|Y) = I(X; Z, f(Y, Z)|Y) \geq I(X; \hat{X}|Y) \geq R_{X|Y}(\Delta_X), \quad (36)$$

where the last equality is defined since $\hat{X} \in \mathcal{M}_0(\Delta_X)$ (see [[2], Remarks, p. 65]). Moreover, minimizing (36) over $Z \in \mathcal{M}(\Delta_X)$ gives

$$R^*(\Delta_X) \geq R_{X|Y}(\Delta_X). \quad (37)$$

(B) Inequality (37) holds with equality, i.e., $R^*(\Delta_X) = R_{X|Y}(\Delta_X)$ if $\hat{X} \in \mathcal{M}_0(\Delta_X)$, which achieves $I(X; \hat{X}|Y) = R_{X|Y}(\Delta_X)$ can be generated as in Figure 2 with $I(X; Z|Y) = I(X; \hat{X}|Y)$. This occurs if and only if $I(X; Z|\hat{X}, Y) = 0$, and follows from the identity and lower bound

$$I(X; Z|Y) = I(X; Z, \hat{X}|Y) = I(X; Z|Y, \hat{X}) + I(X; \hat{X}|Y) \quad (38)$$

$$\geq I(X; \hat{X}|Y), \quad (39)$$

where the inequality holds with equality if and only if $I(X; Z|\hat{X}, Y) = 0$.

3. Main Theorems and Discussion

In this section, we state the main results of this paper. These are the achievable lower bounds of Lemma 1 and Theorem 2, which hold for RVs defined on general abstract alphabet spaces, and Theorems 4 and 5, which hold for multivariate Gaussian RVs.

3.1. Side Information at Encoder and Decoder for an Arbitrary Source

We start with the following achievable lower bound on the conditional mutual information $I(X; \hat{X}|Y)$, which appears in the definition of $R_{X|Y}(\Delta_X)$ of (1); this strengthens Gray's lower bound (8) [[4], Theorem 3.1].

Lemma 1. *Achievable lower bound on conditional mutual information. Let $(X, Y, \hat{X})$ be a triple of arbitrary RVs taking values in the abstract spaces $\mathcal{X} \times \mathcal{Y} \times \hat{\mathcal{X}}$, with distribution $\mathbf{P}_{X,Y,\hat{X}}$ and joint marginal the fixed distribution $\mathbf{P}_{X,Y}$ of (X, Y) . Then, the following hold.*

(a) *The inequality holds:*

$$I(X; \hat{X}|Y) \geq I(X; \hat{X}) - I(X; Y). \quad (40)$$

Moreover, the equality holds

$$I(X; \hat{X}|Y) = I(X; \hat{X}) - I(X; Y) \in [0, \infty), \quad (41)$$

if and only if

$$\mathbf{P}_{X|\hat{X},Y} = \mathbf{P}_{X|\hat{X}} - \text{a.s.} \quad \text{or equivalently } Y \leftrightarrow \hat{X} \leftrightarrow X \text{ is a MC.} \quad (42)$$

(b) *If $Y \leftrightarrow \hat{X} \leftrightarrow X$ is a Markov chain then the equality holds*

$$R_{X|Y}(\Delta_X) = R_X(\Delta_X) - I(X; Y), \quad \Delta_X \in \mathcal{D}_C(X|Y), \quad (43)$$

i.e., for all Δ_X that belong to strictly positive set $\mathcal{D}_C(X|Y) \subseteq [0, \infty)$.

Proof. See Appendix A.1. $\square$

The next theorem which holds for arbitrary RVs is further used to derive the characterization of $R_{X|Y}(\Delta_X)$ for multivariate Gaussian RVs.

Theorem 2. Achievable lower bound on conditional mutual information and mean-square error estimation

(a) Let $(X, Y, \hat{X})$ be a triple of arbitrary RVs on the abstract spaces $\mathcal{X} \times \mathcal{Y} \times \hat{\mathcal{X}}$, with distribution $\mathbf{P}_{X,Y,\hat{X}}$ and joint marginal the fixed distribution $\mathbf{P}_{X,Y}$ of (X, Y) .

Define the conditional mean of X conditioned on $(\hat{X}, Y)$ by

$$\bar{X}^{cm} \triangleq \mathbf{E}(X|Y, \hat{X}) = e(Y, \hat{X}), \quad (44)$$

for some measurable function $f : \mathcal{Y} \times \hat{\mathcal{X}} \rightarrow \mathcal{X}$.

(1) The inequality holds:

$$I(X; \hat{X}|Y) \geq I(X; \bar{X}^{cm}|Y). \quad (45)$$

(2) The equality holds, $I(X; \hat{X}|Y) = I(X; \bar{X}^{cm}|Y)$ if anyone of the conditions (i) or (ii) holds.

$$(i) \bar{X}^{cm} = \hat{X} - a.s \quad (46)$$

$$(ii) \text{ For a fixed } y \in \mathcal{Y} \text{ the function } e(y, \cdot) : \hat{\mathcal{X}} \rightarrow \mathcal{X}, e(y, \hat{x}) = \bar{x}^{cm} \text{ uniquely defines } \hat{x} \text{ i.e., } e(y, \cdot) \text{ is an injective function on the support of } \hat{x}. \quad (47)$$

(b) In part (a) let $(X, Y, \hat{X})$ be a triple of arbitrary RVs on $\mathcal{X} \times \mathcal{Y} \times \hat{\mathcal{X}} = \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \times \mathbb{R}^{n_x}$, $(n_x, n_y) \in \mathbb{Z}_+ \times \mathbb{Z}_+$.

For all measurable functions $(y, \hat{x}) \mapsto g(y, \hat{x}) \in \mathbb{R}^{n_x}$, the mean-square error satisfies

$$\mathbf{E}\left\{\|X - g(Y, \hat{X})\|_{\mathbb{R}^{n_x}}^2\right\} \geq \mathbf{E}\left\{\|X - \mathbf{E}(X|Y, \hat{X})\|_{\mathbb{R}^{n_x}}^2\right\}, \quad \forall g(\cdot). \quad (48)$$

Proof. See Appendix A.2. $\square$

3.2. Side Information at Encoder and Decoder for Multivariate Gaussian Source

The characterizations of the RDFs $R_{X|Y}(\Delta_X)$ and $\bar{R}(\Delta_X)$ for a multivariate Gaussian source are encapsulated in Theorems 3–5; these are proved in Section 4. These theorems include the structural properties of optimal test channels or realizations of $(\hat{X}, Z)$, which induce joint distributions. Furthermore, they achieve the RDFs; the closed form expressions of the RDFs are based on a water-filling. The realization of the optimal test channel of $R_{X|Y}(\Delta_X)$ is shown in Figure 3.

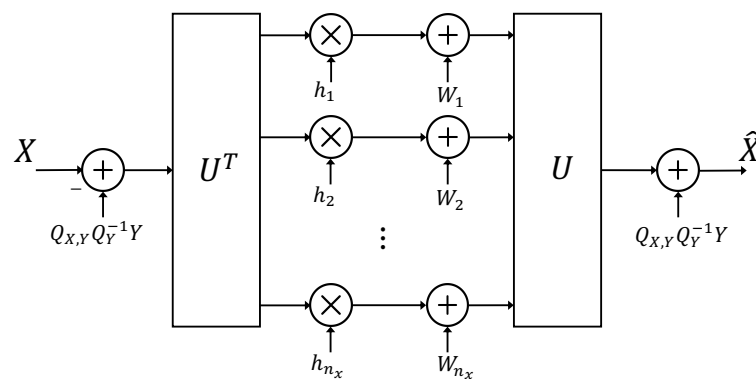


Figure 3. $R_{X|Y}(\Delta_X)$: A realization of optimal reproduction $\hat{X}$ over parallel additive Gaussian noise channels of Theorem 4, where $h_i \triangleq 1 - \frac{\delta_i}{\lambda_i} \geq 0, i = 1, \dots, n_x$ are the diagonal element of the spectral decomposition of the matrix $H = U \text{diag}\{h_1, \dots, h_{n_x}\} U^T$, and $W_i \in N(0, h_i \delta_i), i = 1, \dots, n_x$, the additive noise introduced due to compression.

The following theorem gives a parametric realization of optimal test channel that achieves the characterization of the RDF $R_{X|Y}(\Delta_X)$.

Theorem 3. Characterization of $R_{X|Y}(\Delta_X)$ by test channel realization. Consider the RDF $R_{X|Y}(\Delta_X)$ defined by (1), for the multivariate Gaussian source with mean-square error distortion defined by (9)–(18). The following hold.

(a) The optimal realization $\hat{X}$ that achieves $R_{X|Y}(\Delta_X)$ is parametrized by the matrices (H, Q_W) and represented by

$$\hat{X} = H(X - Q_{X,Y}Q_Y^{-1}Y) + Q_{X,Y}Q_Y^{-1}Y + W \quad (49)$$

$$= H(X - Q_{X,Y}Q_Y^{-1}Y) + Q_{X,Y}Q_Y^{-1}Y + H\Psi, \quad \text{if } H^{-1} \text{ exists,} \quad (50)$$

where

$$HQ_{X|Y} = Q_{X|Y}H^T \triangleq Q_{X|Y} - \Sigma_\Delta \succeq 0, \quad (51)$$

$$W \text{ independent of } (X, Y), \quad Q_W \in N(0, Q_W), \quad (52)$$

$$Q_W \triangleq HQ_{X|Y} - HQ_{X|Y}H^T = H\Sigma_\Delta = \Sigma_\Delta - \Sigma_\Delta Q_{X|Y}^{-1}\Sigma_\Delta = \Sigma_\Delta H \succeq 0, \quad (53)$$

$$W = H\Psi, \quad \Psi \in N(0, Q_\Psi), \quad Q_\Psi \triangleq \Sigma_\Delta H^{-1} = H^{-1}\Sigma_\Delta, \quad \text{if } H^{-1} \text{ exists,} \quad (54)$$

$$\Sigma_\Delta \triangleq \mathbf{E}\{(X - \hat{X})(X - \hat{X})^T\}, \quad (55)$$

$$Q_{\hat{X}|Y} = Q_{X|Y} - \Sigma_\Delta \succeq 0, \quad (56)$$

$$Q_{X|Y} = Q_X - Q_{X,Y}Q_Y^{-1}Q_{X,Y}^T \succ 0, \quad Q_{X,Y} = Q_X C^T, \quad Q_Y = CQ_X C^T + DD^T. \quad (57)$$

Moreover, the optimal parametric realization of $\hat{X}$ satisfies the following structural properties.

$$(i) \quad \mathbf{P}_{X|\hat{X},Y} = \mathbf{P}_{X|\hat{X}}, \quad \text{if } Q_X \succ \Sigma_\Delta, \quad (58)$$

$$(ii) \quad \mathbf{E}\{X|Y\} = \mathbf{E}\{\hat{X}|Y\}, \quad \text{if } Q_X \succeq \Sigma_\Delta, \quad (59)$$

$$(iii) \quad \text{cov}(X, \hat{X}|Y) = \text{cov}(\hat{X}, \hat{X}|Y), \quad \text{if } Q_{X|Y} \succ \Sigma_\Delta, \quad (60)$$

$$(iv) \quad \mathbf{E}\{X|\hat{X}, Y\} = \mathbf{E}\{X|\hat{X}\} = \hat{X}, \quad \text{if } Q_{X|Y} \succ \Sigma_\Delta. \quad (61)$$

(b) The RDF $R_{X|Y}(\Delta_X)$ is given by

$$R_{X|Y}(\Delta_X) = \inf_{\Sigma_\Delta \succeq 0, Q_{X|Y} - \Sigma_\Delta \succeq 0, \text{trace}(\Sigma_\Delta) \leq \Delta_X} \frac{1}{2} \log \max \{1, \det(Q_{X|Y}\Sigma_\Delta^{-1})\}. \quad (62)$$

Proof. The proof is given in Section 4. $\square$

The next theorem gives additional structural properties of the optimal test channel realization of Theorem 3 and uses these properties to characterize RDF $R_{X|Y}(\Delta_X)$ via a water-filling solution.

Theorem 4. Characterization of $R_{X|Y}(\Delta_X)$ via water-filling solution. Consider the RDF $R_{X|Y}(\Delta_X)$ defined by (1), for the multivariate Gaussian source with mean-square error distortion defined by (9)–(18), and its characterization in Theorem 3. The following hold.

(a) The matrices of the parametric realization of $\hat{X}$,

$$\{\Sigma_\Delta, Q_{X|Y}, H, Q_W\} \text{ have spectral decompositions with respect to the same unitary matrix } UU^T = I_{n_x}, U^T U = I_{n_x}, \quad (63)$$

where the realization coefficients are

$$Q_W = H\Sigma_\Delta = U\text{diag}(\sigma_{W_1}^2, \dots, \sigma_{W_{n_x}}^2)U^T, \quad \Sigma_\Delta = U\text{diag}(\delta_1, \dots, \delta_{n_x})U^T, \quad (64)$$

$$H = I_{n_x} - Q_{X|Y}^{-1}\Sigma_\Delta = U\text{diag}(h_1, \dots, h_{n_x})U^T, \quad Q_{X|Y} = U\text{diag}(\lambda_1, \dots, \lambda_{n_x})U^T, \quad (65)$$

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{n_x} > 0, \quad \delta_1 \geq \delta_2 \geq \dots \geq \delta_{n_x} > 0, \quad (66)$$

$$\sigma_{W_1}^2 \geq \sigma_{W_2}^2 \geq \dots \geq \sigma_{W_{n_x}}^2 \geq 0, \quad h_1 \geq h_2 \geq \dots \geq h_{n_x} \geq 0, \quad \sigma_{W_i}^2 = h_i \delta_i, \quad h_i \triangleq 1 - \frac{\delta_i}{\lambda_i}, \quad (67)$$

and the eigenvalues $\sigma_{W_i}^2$ and h_i are given by

$$\sigma_{W_i}^2 = \frac{\min(\lambda_i, \delta_i)(\lambda_i - \min(\lambda_i, \delta_i))}{\lambda_i}, \quad h_i = \frac{\lambda_i - \min(\lambda_i, \delta_i)}{\lambda_i}, \quad \sum_{i=1}^{n_x} \min(\lambda_i, \delta_i) = \Delta_X. \quad (68)$$

Moreover, if $\sigma_{W_i}^2 = 0$, then $h_i = 0$, and vice versa.

(b) The RDF $R_{X|Y}(\Delta_X)$ is given by the water-filling solution:

$$R_{X|Y}(\Delta_X) = \frac{1}{2} \log \max \{1, \det(Q_{X|Y}\Sigma_\Delta^{-1})\} = \frac{1}{2} \sum_{i=1}^{n_x} \log \frac{\lambda_i}{\delta_i}, \quad (69)$$

where

$$\mathbf{E}\{\|X - \hat{X}\|_{\mathbb{R}^{n_x}}^2\} = \text{trace}(\Sigma_\Delta) = \sum_{i=1}^{n_x} \delta_i = \Delta_X, \quad \delta_i = \begin{cases} \mu, & \text{if } \mu < \lambda_i \\ \lambda_i, & \text{if } \mu \geq \lambda_i \end{cases} \quad (70)$$

and $\mu \in (0, \infty)$ is a Lagrange multiplier (obtained from the Kuch–Tucker conditions).

(c) Figure 3 depicts the parallel channel scheme that realizes the optimal $\hat{X}$ of parts (a), (b), which achieves $R_{X|Y}(\Delta_X)$.

(d) If X and Y are independent or Y is replaced by a RV that generates the trivial information, i.e., the σ -algebra of Y is $\sigma\{Y\} = \{\Omega, \emptyset\}$ (or $C = 0$ in (15)), then (a)–(c) hold with $Q_{X|Y} = Q_X$, $Q_{X,Y} = 0$, and $R_{X|Y}(\Delta_X) = R_X(\Delta_X)$, i.e., reduces to the marginal RDF of X .

Proof. The proof is given in Section 4. $\square$

The proof of Theorem 4 (see Section 4) is based on the identification of structural properties of the test channel distribution. Some of the implications are briefly described below.

Conclusion 1: The construction and the structural properties of the optimal test channel $\mathbf{P}_{X|\hat{X},Y}$ that achieves the water-filling characterization of the RDF $R_{X|Y}(\Delta_X)$ of Theorems 3 and 4 are not documented elsewhere in the literature.

(i) Structural properties (58) and (61) strengthen Gray's inequality [[4], Theorem 3.1], (see proof of (8)) to the equality. That is, structural property (58) implies that Gray's [[4], Theorem 3.1] lower bound (8) holds with equality for a strictly positive surface (See Gray [4] for definition) $\Delta_X \in \mathcal{D}_C(X|Y) \subseteq [0, \infty)$, i.e.,

$$R_{X|Y}(\Delta_X) = R_X(\Delta_X) - I(X;Y), \quad \Delta_X \in \mathcal{D}_C(X|Y) \triangleq \{\Delta_X \in [0, \infty) : \Delta_X \leq n_x \lambda_{n_x}\}. \quad (71)$$

The set $\mathcal{D}_C(X|Y)$ excludes values of $\Delta_X \in [0, \infty)$ for which water-filling is active in (69) and (70).

By the realization of the optimal reproduction $\hat{X}$, it follows that the subtraction of equal quantities $\mathbf{E}\{X|Y\}$ at the encoder and decoder does not affect the information measure, noting that $\mathbf{E}\{X|Y\} = \mathbf{E}\{\hat{X}|Y\}$.

Theorem 4 points (a) and (b) are obtained with the aid of Theorem 3 and Hadamard's inequality, which shows $Q_{X|Y}$ and Σ_Δ have the same eigenvectors.

(ii) Structural properties of realizations of Theorems 3 and 4: The matrices $\{\Sigma_\Delta, Q_{X|Y}, H, Q_W\}$ are nonnegative symmetric and have a spectral decomposition with respect to the same unitary matrix $UU^T = I_{n_x}$ [21]. This implies that the test channel is equivalently represented by parallel additive Gaussian noise channels (subject to pre-processing and post-processing at the encoder and decoder).

(iii) In Remark 4, we show that the realization of optimal $\hat{X}$ in Figure 3, which achieves the RDF of Theorem 4, degenerates to Wyner's [2] optimal realization, which attains the RDF $R_{X|Y}(\Delta_X)$, for the tuple of scalar-valued, jointly Gaussian RVs (X, Y) with square error distortion function.

3.3. Side Information Only at Decoder for Multivariate Gaussian Source

Theorem 5 gives the optimal test channel that achieves the characterization of the RDF $\bar{R}(\Delta_X)$ and further states that there is no loss of compression rate if side information is only available at the decoder. That is, although in general, $\bar{R}(\Delta_X) \geq R_{X|Y}(\Delta_X)$, an optimal reproduction $\hat{X} = f(Y, Z)$ of X , where $f(\cdot, \cdot)$ is linear, is constructed such that the inequality holds with equality.

Theorem 5. *Characterization and water-filling solution of $\bar{R}(\Delta_X)$. Consider the RDF $\bar{R}(\Delta_X)$ defined by (5) for the multivariate Gaussian source with mean-square error distortion, defined by (9)–(18). Then, the following hold.*

(a) *The characterization of the RDF, $\bar{R}(\Delta_X)$ satisfies*

$$\bar{R}(\Delta_X) \geq R_{X|Y}(\Delta_X), \quad (72)$$

where $R_{X|Y}(\Delta_X)$ is given in Theorem 4b.

(b) *The optimal realization $\hat{X} = f(Y, Z)$, which achieves the lower bound in (72), i.e., $\bar{R}(\Delta_X) = R_{X|Y}(\Delta_X)$, is represented by*

$$\hat{X} = f(Y, Z) \quad (73)$$

$$= (I - H)Q_{X,Y}Q_Y^{-1}Y + Z, \quad (74)$$

$$Z = HX + W, \quad (75)$$

$$(H, Q_W) \text{ given by (51)–(57), and (63) holds.} \quad (76)$$

Moreover, the following structural properties hold:

(1) *The optimal test channel satisfies*

$$(i) \quad \mathbf{P}_{X|\hat{X},Y,Z} = \mathbf{P}_{X|\hat{X},Y} \stackrel{(\alpha)}{=} \mathbf{P}_{X|\hat{X}}, \text{ where } (\alpha) \text{ holds if } Q_X \succ \Sigma_\Delta, \quad (77)$$

$$(ii) \quad \mathbf{E}\{X|\hat{X}, Y, Z\} = \mathbf{E}\{X|\hat{X}, Y\} \stackrel{(\beta)}{=} \mathbf{E}\{X|\hat{X}\} \stackrel{(\gamma)}{=} \hat{X}, \text{ where } (\beta), (\gamma) \text{ hold if } Q_{X|Y} \succ \Sigma_\Delta, \quad (78)$$

$$(iii) \quad \mathbf{P}_{Z|X,Y} = \mathbf{P}_{Z|X}. \quad (79)$$

(2) *Structural property (2) of Theorem 4a holds.*

Proof. It is given in Section 4. $\square$

The proof of Theorem 5 is based on the derivation of the structural properties and Theorem 4. Some implications are discussed below.

Conclusion 2: The optimal reproduction $\hat{X} = f(X, Z)$ or test channel distribution $\mathbf{P}_{X|\hat{X},Y,Z}$, which achieves $\bar{R}(\Delta_X)$ of Theorem 5, are not reported in the literature.

(i) From the structural property (1) of Theorem 5, i.e., (77), it follows that the lower bound $\bar{R}(\Delta_X) \geq R_{X|Y}(\Delta_X)$ is achieved by the realization $\hat{X} = f(Y, Z)$ of Theorem 5b; i.e., for a given $Y = y$, then $\hat{X}$ uniquely defines Z .

(ii) If X is independent of Y or Y generates trivial information, then the RDFs $\bar{R}(\Delta_X) = \bar{R}_{X|Y}(\Delta_X)$ degenerate to the classical RDF of the source X , i.e., $R_X(\Delta_X)$, as expected. This is easily verified from (73) and (76), i.e., $Q_{X,Y} = 0$, which implies $\hat{X} = Z$. For scalar-valued RVs, $X : \Omega \rightarrow \mathbb{R}, Y : \Omega \rightarrow \mathbb{R}, X \in N(0, \sigma_X^2)$, and X independent of Y , then the optimal realization reduces to

$$\hat{X} = Z = \left(1 - \frac{\Delta_X}{\sigma_X^2}\right)X + \sqrt{\left(1 - \frac{\Delta_X}{\sigma_X^2}\right)\Delta_X}\bar{W}, \quad \bar{W} \in N(0, 1), \quad \sigma_X^2 \leq \Delta_X, \quad (80)$$

$$Q_{\hat{X}} = Q_Z = \sigma_X^2 = \sigma_X^2 - \Delta_X \geq 0, \quad (81)$$

as expected.

(iii) In Remark 4, we show that the realization of optimal $\hat{X} = f(Y, Z)$, which achieves the RDF $\bar{R}(\Delta_X)$ of Theorem 5, degenerates to Wyner's [2] realization that attains the RDF $\bar{R}(\Delta_X)$, of the tuple of scalar-valued, jointly Gaussian RVs (X, Y) , with the square error distortion function.

4. Proofs of Theorems 3–5

In this section, we derive the statements of Theorems 3–5 by making use of Theorem 2 (which holds for general abstract alphabet spaces) by restricting attention to multivariate jointly Gaussian (X, Y) .

4.1. Side Information at Encoder and Decoder

For jointly Gaussian RVs $(X, Y, \hat{X})$, in the next theorem we identify simple sufficient conditions for the lower bound of Theorem 2 to be achievable.

Theorem 6. *Sufficient conditions for the lower bounds of Theorem 2 to be achievable. Consider the statement of Theorem 2 for a triple of jointly Gaussian RVs $(X, Y, \hat{X})$ on $\mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \times \mathbb{R}^{n_x}$, $(n_x, n_y) \in \mathbb{Z}_+ \times \mathbb{Z}_+$, i.e., $\mathbf{P}_{X,Y,\hat{X}} = \mathbf{P}_{X,Y,\hat{X}}^G$ and joint marginal the fixed Gaussian distribution $\mathbf{P}_{X,Y} = \mathbf{P}_{X,Y}^G$ of (X, Y) . Then,*

$$\bar{X}^{cm} \triangleq \mathbf{E}(X|Y, \hat{X}) = e^G(Y, \hat{X}), \quad (82)$$

$$e^G(Y, \hat{X}) = \mathbf{E}(X|Y) + \text{cov}(X, \hat{X}|Y) \left\{ \text{cov}(\hat{X}, \hat{X}|Y) \right\}^{-1} \left(\hat{X} - \mathbf{E}(\hat{X}|Y) \right). \quad (83)$$

Moreover, the following hold.

Case (i). $\text{cov}(\hat{X}, \hat{X}|Y) \succ 0$, that is, $\text{rank}(Q_{\hat{X}|Y}) = n_x$. Condition (84) is sufficient for $I(X; \hat{X}|Y) = I(X; \bar{X}^{cm}|Y)$.

$$\bar{X}^{cm} \triangleq \mathbf{E}(X|Y, \hat{X}) = e^G(Y, \hat{X}) = \hat{X} - a.s. \quad (84)$$

In addition, Conditions 1 and 2 below are sufficient for (84) to hold.

$$\text{Condition 1. } \mathbf{E}(X|Y) = \mathbf{E}(\hat{X}|Y) \quad (85)$$

$$\text{Condition 2. } \text{cov}(X, \hat{X}|Y) \text{cov}(\hat{X}, \hat{X}|Y)^{-1} = I_{n_x} \quad (86)$$

Case (ii). $\text{cov}(\hat{X}, \hat{X}|Y) \succeq 0$ but not $\text{cov}(\hat{X}, \hat{X}|Y) \succ 0$; that is, $\text{rank}(Q_{\hat{X}|Y}) = n_1 < n_x$. Condition (87) is sufficient for $I(X; \hat{X}|Y) = I(X; \bar{X}^{cm}|Y)$.

$$e^G(\cdot, \cdot) \text{ defined by (83) satisfies (47)}. \quad (87)$$

In addition, a sufficient condition for (87) to hold is, for a fixed $Y = y \in \mathcal{Y}$, the σ -algebras satisfy $\mathcal{F}^{\hat{X}} = \mathcal{F}^{e^G(y, \hat{X})}$.

Proof. Note that identity (83) follows from Proposition 1, (26), by letting $Y = \hat{X}$ and $\mathcal{G}$ be the information generated by Y . Consider Case (i); If (84) holds then $I(X; \hat{X}|\bar{X}^{cm}, Y) = 0$. By (83), Conditions 1 and 2 are sufficient for (84) to hold. Consider Case (ii). Sufficient condition (87) follows from Theorem 2, and implies $I(X; \hat{X}|\bar{X}^{cm}, Y) = 0$. The statement below (87) follows from Proposition 2. $\square$

Now, we turn our attention to the optimization problem $R_{X|Y}(\Delta_X)$ defined by (1) for the multivariate Gaussian source with mean-square error distortion defined by (9)–(18). In the next lemma, we derive a *preliminary parametrization* of the optimal reproduction distribution $\mathbf{P}_{\hat{X}|X,Y}$ of the RDF $R_{X|Y}(\Delta_X)$.

Lemma 2. *Preliminary parametrization of optimal reproduction distribution of $R_{X|Y}(\Delta_X)$. Consider the RDF $R_{X|Y}(\Delta_X)$ defined by (1) for the multivariate Gaussian source, i.e., $\mathbf{P}_{X,Y} = \mathbf{P}_{X,Y}^G$, with mean-square error distortion defined by (9)–(18).*

- (a) *For every joint distribution $\mathbf{P}_{X,Y,\hat{X}}$ there exists a jointly Gaussian distribution denoted by $\mathbf{P}_{X,Y,\hat{X}}^G$ with marginal the fixed distribution $\mathbf{P}_{X,Y}^G$, which minimizes $I(X; \hat{X}|Y)$ and satisfies the average distortion constraint, i.e., with $d_X(x, \hat{x}) = \|x - \hat{x}\|_{\mathbb{R}^{n_x}}^2$.*
- (b) *The conditional reproduction distribution of the RDF $R_{X|Y}(\Delta_X)$ is $\mathbf{P}_{\hat{X}|X,Y} = \mathbf{P}_{\hat{X}|X,Y}^G$ and induced by the parametric realization of $\hat{X}$ (in terms of H, G, Q_W),*

$$\hat{X} = HX + GY + W, \quad (88)$$

$$H \in \mathbb{R}^{n_x \times n_x}, \quad G \in \mathbb{R}^{n_x \times n_y}, \quad (89)$$

$$W \in N(0, Q_W), \quad Q_W \succeq 0, \quad (90)$$

$$W \text{ independent of } (X, Y), \quad (91)$$

and $\hat{X}$ is a Gaussian RV.

- (c) $R_{X|Y}(\Delta_X)$ is characterized by the optimization problem.

$$R_{X|Y}(\Delta_X) \triangleq \inf_{\mathcal{M}_0^G(\Delta_X)} I(X; \hat{X}|Y), \quad \Delta_X \in [0, \infty), \quad (92)$$

where $\mathcal{M}_0^G(\Delta_X)$ is specified by the set

$$\mathcal{M}_0^G(\Delta_X) \triangleq \left\{ \hat{X} : \Omega \rightarrow \hat{\mathcal{X}} : (88) - (91) \text{ hold, and } \mathbf{E}\{\|X - \hat{X}\|_{\mathbb{R}^{n_x}}^2\} \leq \Delta_X \right\}. \quad (93)$$

- (d) *If there exists (H, G, Q_W) such that (84) or (87) hold, then a further lower bound on $R_{X|Y}(\Delta_X)$ is achieved in the subset $\mathcal{M}_0^{G,o}(\Delta_X) \subseteq \mathcal{M}_0^G(\Delta_X)$ defined by*

$$\mathcal{M}_0^{G,o}(\Delta_X) \triangleq \left\{ \hat{X} : \Omega \rightarrow \hat{\mathcal{X}} : (88) - (91) \text{ hold, (84) or (87) hold, } \mathbf{E}\{\|X - \hat{X}\|_{\mathbb{R}^{n_x}}^2\} \leq \Delta_X \right\}, \quad (94)$$

and the corresponding characterization of the RDF is

$$R_{X|Y}(\Delta_X) \triangleq \inf_{\mathcal{M}_0^{G,o}(\Delta_X)} I(X; \hat{X}|Y), \quad \Delta_X \in [0, \infty). \quad (95)$$

Proof. (a) This is omitted since it is similar to the classical unconditional RDF $R_X(\Delta_X)$ of a Gaussian message $X \in N(0, Q_X)$. (b) By (a), the conditional distribution $\mathbf{P}_{\hat{X}|X,Y}^G$ is such that, its conditional mean is linear in (X, Y) , its conditional covariance is nonrandom, i.e., constant, and for fixed $(X, Y) = (x, y)$, $\mathbf{P}_{\hat{X}|X,Y}^G$ is Gaussian. Such a distribution is induced by the parametric realization (88)–(91). (c) Follows from parts (a) and (b). (d) Follows from Theorem 6 and (48) due to the achievability of the lower bounds. $\square$

In the next theorem, we identify the optimal triple (H, G, Q_W) such that (84) or (87) hold (i.e., establish its existence), characterize the RDF by $R_{X|Y}(\Delta_X) \triangleq \inf_{\mathcal{M}_0^{G,o}(\Delta_X)} I(X; \hat{X}|Y)$, and construct a realization $\hat{X}$ that achieves it.

Theorem 7. *Characterization of RDF $R_{X|Y}(\Delta_X)$. Consider the RDF $R_{X|Y}(\Delta_X)$, defined by (1), for the multivariate Gaussian source with mean-square error distortion, defined by (9)–(18). The characterization of the RDF $R_{X|Y}(\Delta_X)$ is*

$$R_{X|Y}(\Delta_X) \triangleq \inf_{\mathcal{Q}(\Delta_X)} I(X; \hat{X}|Y) \quad (96)$$

$$= \inf_{\mathcal{M}_0^{G,o}(\Delta_X)} I(X; \hat{X}|Y) \quad (97)$$

$$= \inf_{\mathcal{Q}(\Delta_X)} \frac{1}{2} \log \left\{ \det(Q_{X|Y} \Sigma_\Delta^{-1}) \right\}, \quad (98)$$

where

$$\mathcal{Q}(\Delta_X) \triangleq \left\{ \Sigma_\Delta \succeq 0 : Q_{X|Y} - \Sigma_\Delta \succeq 0, \text{ trace}(\Sigma_\Delta) \leq \Delta_X \right\}, \quad (99)$$

$$\Sigma_\Delta \triangleq \mathbf{E} \left\{ (X - \hat{X})(X - \hat{X})^T \right\}, \quad (100)$$

$$Q_{X|Y} = Q_X - Q_{X,Y} Q_Y^{-1} Q_{X,Y}^T, \quad (101)$$

$$Q_{X,Y} = Q_X C^T, \quad Q_Y = C Q_X C^T + D D^T. \quad (102)$$

The realization of the optimal reproduction $\hat{X} \in \mathcal{M}_0^{G,o}(\Delta_X)$, which achieves $R_{X|Y}(\Delta_X)$, is given in Theorem 3a, also satisfies the properties stated under Theorem 3a. (i)–(iv).

Proof. See Appendix A.3. $\square$

Remark 2. *Structural properties of the optimal realization of Theorem 4a. For the characterization of the RDF $R_{X|Y}(\Delta_X)$ of Theorem 7, which is achieved by $\hat{X}$ defined in Theorem 3a in terms of the matrices $\{\Sigma_\Delta, Q_{X|Y}, H, Q_W\}$, we show in Corollary 2, the statements of Theorem 4a, i.e.,*

$$(i) \quad H = H^T \succeq 0, \quad (103)$$

$$(ii) \quad \{\Sigma_\Delta, \Sigma_{X|Y}, H, Q_W\} \text{ have spectral repres. with respect to the same unitary matrix } U U^T = I_{n_X}. \quad (104)$$

To prove the structural property of Remark 2, we use the next corollary, which is a degenerate case of [[22], Lemma 2] (i.e., the structural properties of test channel of Gorbunov and Pinsker [23] nonanticipatory RDF of Markov sources).

Corollary 1. *Structural properties of realization of optimal $\hat{X}$ of Theorem 4a. Consider the characterization of the RDF $R_{X|Y}(\Delta_X)$ of Theorem 7. Suppose $Q_{X|Y} \succ 0$ and $\Sigma_\Delta \succeq 0$ commute, that is,*

$$Q_{X|Y}\Sigma_\Delta = \Sigma_\Delta Q_{X|Y}. \quad (105)$$

Then,

$$(1) \ H = I_{n_x} - \Sigma_\Delta Q_{X|Y}^{-1} = H^T, \quad Q_W = \Sigma_\Delta H^T = \Sigma_\Delta H = H\Sigma_\Delta = Q_W^T \succeq 0 \quad (106)$$

$$(2) \ \{\Sigma_\Delta, Q_{X|Y}, H, Q_W\} \text{ have spectral} \\ \text{decompositions with respect to the same unitary matrix } UU^T = I_{n_x}, \ U^T U = I_{n_x}. \quad (107)$$

that is, the following hold.

$$Q_{X|Y} = U \text{diag}\{\lambda_1, \dots, \lambda_{n_x}\} U^T, \quad \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{n_x} > 0, \quad (108)$$

$$\Sigma_\Delta = U \text{diag}\{\delta_1, \dots, \delta_{n_x}\} U^T, \quad \delta_1 \geq \delta_2 \geq \dots \geq \delta_{n_x} \geq 0, \quad (109)$$

$$H = U \text{diag}\{1 - \frac{\delta_1}{\lambda_1}, \dots, 1 - \frac{\delta_{n_x}}{\lambda_{n_x}}\} U^T, \quad (110)$$

$$Q_W = U \text{diag}\{(1 - \frac{\delta_1}{\lambda_1})\delta_1, \dots, (1 - \frac{\delta_{n_x}}{\lambda_{n_x}})\delta_{n_x}\} U^T, \text{ and } (1 - \frac{\delta_k}{\lambda_k})\delta_k \geq 0. \quad (111)$$

Proof. See Appendix A.4. $\square$

In the next corollary, we re-express the realization of $\hat{X}$ of Theorem 4a, which characterizes the RDF of Theorem 7 using a translation of X and $\hat{X}$ by subtracting their conditional means with respect to Y , making use of property $\mathbf{E}\{X|Y\} = \mathbf{E}\{\hat{X}|Y\}$ of (78). This is the realization shown in Figure 3.

Corollary 2. *Equivalent characterization of $R_{X|Y}(\Delta_X)$. Consider the characterization of the RDF $R_{X|Y}(\Delta_X)$ of Theorem 7 and the realization of $\hat{X}$ of Theorem 3a and Theorem 4a. Define the translated RVs*

$$\mathbf{X} \triangleq X - \mathbf{E}\{X|Y\} = X - Q_{X,Y}Q_Y^{-1}Y, \quad \hat{\mathbf{X}} \triangleq \hat{X} - \mathbf{E}\{\hat{X}|Y\} = \hat{X} - Q_{X,Y}Q_Y^{-1}Y. \quad (112)$$

Let

$$Q_{X|Y} = U \text{diag}\{\lambda_1, \dots, \lambda_{n_x}\} U^T, \quad UU^T = I_{n_x}, \ U^T U = I_{n_x}, \quad \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{n_x}, \quad (113)$$

$$\bar{\mathbf{X}} \triangleq U^T \mathbf{X}, \quad \hat{\bar{\mathbf{X}}} \triangleq U^T \hat{\mathbf{X}}. \quad (114)$$

Then,

$$\hat{\mathbf{X}} = H\mathbf{X} + W, \quad (115)$$

$$I(X; \hat{X}|Y) = I(\mathbf{X}; \hat{\mathbf{X}}) = I(U^T \mathbf{X}; U^T \hat{\mathbf{X}}), \quad (116)$$

$$\mathbf{E}\|X - \hat{X}\|_{\mathbb{R}^{n_x}}^2 = \mathbf{E}\|\mathbf{X} - \hat{\mathbf{X}}\|_{\mathbb{R}^{n_x}}^2 = \mathbf{E}\|U^T \mathbf{X} - U^T \hat{\mathbf{X}}\|_{\mathbb{R}^{n_x}}^2 = \text{trace}(\Sigma_\Delta), \quad (117)$$

where (H, Q_W) are given in Theorem 3a.

Further, the characterization of the RDF $R_{X|Y}(\Delta_X)$ (98) satisfies the following equalities and inequality:

$$R_{X|Y}(\Delta_X) \triangleq \inf_{Q(\Delta_X)} I(X; \hat{X}|Y) = \inf_{Q(\Delta_X)} \frac{1}{2} \log \max \left\{ 1, \det(Q_{X|Y} \Sigma_\Delta^{-1}) \right\} \quad (118)$$

$$= \inf_{\mathbf{E} \|\mathbf{X} - \hat{\mathbf{X}}\|_{\mathbb{R}^{n_X}}^2 \leq \Delta_X} I(\mathbf{X}; \hat{\mathbf{X}}) \quad (119)$$

$$= \inf_{\mathbf{E} \|U^T \mathbf{X} - U^T \hat{\mathbf{X}}\|_{\mathbb{R}^{n_X}}^2 \leq \Delta_X} I(U^T \mathbf{X}; U^T \hat{\mathbf{X}}) \quad (120)$$

$$\geq \inf_{\mathbf{E} \|U^T \mathbf{X} - U^T \hat{\mathbf{X}}\|_{\mathbb{R}^{n_X}}^2 \leq \Delta_X} \sum_{t=1}^{n_X} I(\bar{\mathbf{X}}_t; \hat{\bar{\mathbf{X}}}_t) \quad (121)$$

Moreover, the inequality (121) is achieved if $Q_{X|Y} \succeq 0$ and $\Sigma_\Delta \succeq 0$ commute; that is, if (105) holds, then

$$R_{X|Y}(\Delta_X) = \inf_{\sum_{i=1}^{n_X} \delta_i \leq \Delta_X} \frac{1}{2} \sum_{i=1}^{n_X} \log \max \left\{ 1, \frac{\lambda_i}{\delta_i} \right\} \quad (122)$$

where

$$\text{diag}\{\mathbf{E}(U^T \mathbf{X} - U^T \hat{\mathbf{X}})(U^T \mathbf{X} - U^T \hat{\mathbf{X}})^T\} = \text{diag}\{\delta_1, \delta_2, \dots, \delta_{n_X}\}. \quad (123)$$

Proof. By Theorem 3a,

$$\hat{\mathbf{X}} = H\mathbf{X} + (I - H)Q_{X,Y}Q_Y^{-1}\mathbf{Y} + W \quad (124)$$

$$= H(X - Q_{X,Y}Q_Y^{-1}\mathbf{Y}) + Q_{X,Y}Q_Y^{-1}\mathbf{Y} + W \quad (125)$$

$$\implies \hat{\mathbf{X}} - Q_{X,Y}Q_Y^{-1}\mathbf{Y} = H(X - Q_{X,Y}Q_Y^{-1}\mathbf{Y}) + W \quad (126)$$

$$\implies \hat{\mathbf{X}} = H\mathbf{X} + W. \quad (127)$$

The last equation establishes (115). By properties of conditional mutual information and the properties of optimal realization $\hat{\mathbf{X}}$, the following equalities hold.

$$I(X; \hat{\mathbf{X}}|Y) = I(X - Q_{X,Y}Q_Y^{-1}\mathbf{Y}; \hat{\mathbf{X}} - Q_{X,Y}Q_Y^{-1}\mathbf{Y}|Y) \quad (128)$$

$$= I(\mathbf{X}; \hat{\mathbf{X}}|Y), \quad \text{by (112)} \quad (129)$$

$$= H(\hat{\mathbf{X}}|Y) - H(\hat{\mathbf{X}}|Y, \mathbf{X}) \quad (130)$$

$$= H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|Y, \mathbf{X}), \quad \text{by indep. of } (\mathbf{X}, W) \text{ and } Y \quad (131)$$

$$= H(\hat{\mathbf{X}}) - H(\hat{\mathbf{X}}|\mathbf{X}), \quad \text{by indep. of } W \text{ and } Y \text{ for fixed } X \quad (132)$$

$$= I(\mathbf{X}; \hat{\mathbf{X}}) \quad (133)$$

$$= I(U^T \mathbf{X}; U^T \hat{\mathbf{X}}) \quad (134)$$

$$= I(\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2, \dots, \bar{\mathbf{X}}_{n_X}; \hat{\bar{\mathbf{X}}}_1, \hat{\bar{\mathbf{X}}}_2, \dots, \hat{\bar{\mathbf{X}}}_{n_X}) \quad (135)$$

$$\geq \sum_{t=1}^{n_X} I(\bar{\mathbf{X}}_t; \hat{\bar{\mathbf{X}}}_t), \quad \text{by mutual independence of } \bar{\mathbf{X}}_t, t = 1, 2, \dots, n_X. \quad (136)$$

Moreover, inequality (136) holds with equality if $(\bar{\mathbf{X}}_t; \hat{\bar{\mathbf{X}}}_t), t = 1, 2, \dots, n_X$ are jointly independent. The average distortion function is then given by

$$\mathbf{E} \|\mathbf{X} - \hat{\mathbf{X}}\|_{\mathbb{R}^{n_X}}^2 = \mathbf{E} \|\mathbf{X} - \hat{\mathbf{X}} - Q_{X,Y}Q_Y^{-1}\mathbf{Y} + Q_{X,Y}Q_Y^{-1}\mathbf{Y}\|_{\mathbb{R}^{n_X}}^2 \quad (137)$$

$$= \mathbf{E} \|\mathbf{X} - \hat{\mathbf{X}}\|_{\mathbb{R}^{n_X}}^2, \quad \text{by (112)} \quad (138)$$

$$= \mathbf{E} \|U^T \mathbf{X} - U^T \hat{\mathbf{X}}\|_{\mathbb{R}^{n_X}}^2 = \text{trace}(\Sigma_\Delta), \quad \text{by } UU^T = I_{n_X}. \quad (139)$$

By Corollary 1, if (105) holds, that is, $Q_{X|Y} \succ 0$ and $\Sigma_\Delta \succeq 0$ satisfy $Q_{X|Y}\Sigma_\Delta = \Sigma_\Delta Q_{X|Y}$ (i.e., commute), then (106)–(111) hold, and by (122) we obtain

$$\widehat{\mathbf{X}} \triangleq U^T \widehat{\mathbf{X}} = U^T H \mathbf{X} + U^T W = U^T \widehat{\mathbf{X}} = U^T H U U^T \mathbf{X} + U^T W \quad (140)$$

$$= U^T H U \widehat{\mathbf{X}} + U^T W, \quad U^T H U \text{ is diagonal and } U^T W \text{ has indep. components.} \quad (141)$$

Hence, if (105) holds, then the lower bound in (136) holds with equality because $(\widehat{\mathbf{X}}_t; \widehat{\mathbf{X}}_t), t \in \mathbb{Z}_{n_x}$ are jointly independent. Moreover, if (105) holds, then from, say, (118), the expressions (122) and (123) are obtained. The above equations establish all claims. $\square$

Proposition 4. Theorem 4 is correct.

Proof. By invoking Corollary 2, Theorem 7 and the convexity of $R_{X|Y}(\Delta_X)$ given by (122), then we arrive at the statements of Theorem 4, which completely characterize the RDF $R_{X|Y}(\Delta_X)$ and construct a realization of the optimal $\widehat{X}$ that achieves it. $\square$

Next, we discuss the degenerate case, when the statements of Theorems 3, 4 and 7 reduce to the RDF $R_X(\Delta_X)$ of a Gaussian RV X with square-error distortion function. We illustrate that the identified structural property of the realization matrices $\{\Sigma_\Delta, Q_{X|Y}, H, Q_W\}$ leads to the well-known water-filling solution.

Remark 3. Degenerate case of Theorem 7 and realization $\widehat{X}$ of Theorem 4a. Consider the characterization of the RDF $R_{X|Y}(\Delta_X)$ of Theorem 7, the realization of $\widehat{X}$ Theorem 3a, Theorem 3, and assume X and Y are independent or Y generates the trivial information; i.e., the σ -algebra of Y is $\sigma\{Y\} = \{\Omega, \emptyset\}$ or $C = 0$ in (15)–(18).

(a) By the definitions of $Q_{X,Y}, Q_{X|Y}$ then

$$Q_{X,Y} = 0, \quad Q_{X|Y} = Q_X. \quad (142)$$

Substituting (142) into the expressions of Theorem 7, the RDF $R_{X|Y}(\Delta_X)$ reduces to

$$R_{X|Y}(\Delta_X) = R_X(\Delta_X) \triangleq \inf_{Q(\Delta_X)} I(X; \widehat{X}) \quad (143)$$

$$= \inf_{Q^m(\Delta_X)} \frac{1}{2} \log \left\{ \det(Q_X \Sigma_\Delta^{-1}) \right\}, \quad (144)$$

where

$$Q^m(\Delta_X) \triangleq \left\{ \Sigma_\Delta \succeq 0 : Q_X \succeq \Sigma_\Delta, \text{ trace}(\Sigma_\Delta) \leq \Delta_X \right\}, \quad (145)$$

and the optimal reproduction $\widehat{X}$ reduces to

$$\widehat{X} = (I_{n_x} - \Sigma_\Delta Q_X^{-1})X + W, \quad Q_X \succeq \Sigma_\Delta, \quad (146)$$

$$Q_W = (I_{n_x} - \Sigma_\Delta Q_X^{-1})\Sigma_\Delta \succeq 0. \quad (147)$$

Thus, $R_X(\Delta_X)$ is the well-known RDF of a multivariate memoryless Gaussian RV X with square-error distortion.

(b) For the RDF $R_X(\Delta_X)$ of part (a), it is known [24] that Σ_Δ and Q_X have a spectral decomposition with respect to the same unitary matrix, that is,

$$Q_X = U \Lambda_X U^T, \quad \Sigma_\Delta = U \Delta U^T, \quad U U^T = I, \quad (148)$$

$$\Lambda_X = \text{diag}\{\lambda_{X,1}, \dots, \lambda_{X,n_x}\}, \quad \Delta = \text{diag}\{\delta_1, \dots, \delta_{n_x}\}, \quad (149)$$

where the entries of (Λ_X, Δ) are in decreasing order.
Define

$$\mathbf{X}^p \triangleq \mathbf{U}^T \mathbf{X}, \quad \widehat{\mathbf{X}}^p \triangleq \mathbf{U}^T \widehat{\mathbf{X}}, \quad \mathbf{W}^p \triangleq \mathbf{U}^T \mathbf{W}. \quad (150)$$

Then, a parallel channel realization of the optimal reproduction $\widehat{\mathbf{X}}^p$ is obtained as follows:

$$\widehat{\mathbf{X}}^p = \mathbf{H} \mathbf{X}^p + \mathbf{W}^p, \quad (151)$$

$$\mathbf{H} = \mathbf{I}_{n_x} - \Delta \Lambda_X^{-1} = \text{diag}\{1 - \frac{\delta_1}{\lambda_{X,1}}, \dots, 1 - \frac{\delta_{n_x}}{\lambda_{X,n_x}}\}, \quad (152)$$

$$\mathbf{Q}_{\mathbf{W}^p} = \mathbf{H} \Delta = \text{diag}\{(1 - \frac{\delta_1}{\lambda_{X,1}}) \delta_1, \dots, (1 - \frac{\delta_{n_x}}{\lambda_{X,n_x}}) \delta_{n_x}\}. \quad (153)$$

The RDF $R_X(\Delta_X)$ is then computed from the reverse water-filling equations as follows.

$$R_X(\Delta_X) = \frac{1}{2} \sum_{i=1}^{n_x} \log \frac{\lambda_{X,i}}{\delta_i}, \quad (154)$$

where

$$\sum_{i=1}^{n_x} \delta_i = \Delta_X, \quad \delta_i = \begin{cases} \mu, & \text{if } \mu < \lambda_{X,i} \\ \lambda_{X,i}, & \text{if } \mu \geq \lambda_{X,i} \end{cases} \quad (155)$$

and $\mu \in [0, \infty)$ is a Lagrange multiplier (obtained from the Kuch–Tucker conditions).

4.2. Side Information Only at Decoder

In general, when the side information is available only at the decoder, the achievable operational rate $R^*(\Delta_X)$ is greater than the achievable operational rate $\bar{R}_1(\Delta_X)$ when the side information is available to the encoder and the decoder [2]. By Remark 1, $\bar{R}(\Delta_X) \geq R_{X|Y}(\Delta_X)$, and equality holds if $I(X; Z|\widehat{X}, Y) = 0$.

In view of the characterization of $R_{X|Y}(\Delta_X)$ and the realization of the optimal reproduction $\widehat{X}$ of Theorem 3, which is presented in Figure 3, we observe that we can re-write (49) as follows.

$$\widehat{X} = \mathbf{H} \mathbf{X} + (\mathbf{I}_{n_x} - \mathbf{H}) \mathbf{Q}_{X,Y} \mathbf{Q}_Y^{-1} \mathbf{Y} + \mathbf{W}, \quad (156)$$

$$= (\mathbf{I}_{n_x} - \mathbf{H}) \mathbf{Q}_{X,Y} \mathbf{Q}_Y^{-1} \mathbf{Y} + \mathbf{Z} \quad (157)$$

$$= f(Y, Z) \quad (158)$$

$$\mathbf{Z} = \mathbf{H} \mathbf{X} + \mathbf{W}, \quad (159)$$

$$\mathbf{H} = \mathbf{I}_{n_x} - \Sigma_{\Delta} \mathbf{Q}_{X|Y}^{-1}, \quad \mathbf{Q}_W = \mathbf{H} \Sigma_{\Delta}, \quad \text{defined by (51)–(63)}, \quad (160)$$

$$\mathbf{P}_{Z|X,Y} = \mathbf{P}_{Z|X}, \quad (\widehat{X}, Y) \text{ uniquely define } Z, \text{ which implies } I(X; Z|\widehat{X}, Y) = 0. \quad (161)$$

Proposition 5. Theorem 5 is correct.

Proof. From the above realization of $\widehat{X} = f(Y, Z)$, we have the following. (a) By Wyner, see Remark 1, then the inequalities (36) and (37) hold, and equalities hold if $I(X; Z|\widehat{X}, Y) = 0$. That is, for any $\widehat{X} = f(Y, Z)$, and by the properties of conditional mutual information, then

$$I(X; Z|Y) \stackrel{(\alpha)}{=} I(X; Z, \widehat{X}|Y) \quad (162)$$

$$\stackrel{(\beta)}{=} I(X; Z|\widehat{X}, Y) + I(X; \widehat{X}|Y) \quad (163)$$

$$\stackrel{(\gamma)}{\geq} I(X; \widehat{X}|Y), \quad (164)$$

where (α) is due to $\hat{X} = f(Y, Z)$, (β) is due to the chain rule of mutual information, and (γ) is due to $I(X; Z|\hat{X}, Y) \geq 0$. Hence, (72) is obtained (as in Wyner [2] for a tuple of scalar jointly Gaussian RVs). (b) Equality holds in (164) if there exists an $\hat{X} = f(Y, Z)$ such that $I(X; Z|\hat{X}, Y) = 0$, and the average distortion is satisfied. Taking $\hat{X} = f(Y, Z) = (I_{n_x} - H)Q_{X,Y}Q_Y^{-1}Y + Z$, where $Z = g(X, W)$ is specified by (156)–(160), then $I(X; Z|\hat{X}, Y) = 0$ and the average distortion is satisfied. Since the realization (156)–(160) is identical to the realization (73)–(76), then part (b) is also shown. (c) This follows directly from the optimal realization. $\square$

5. Connection with Other Works and Simulations

In this section, we illustrate that for the special case of scalar-valued jointly Gaussian RVs (X, Y) , our results reproduce Wyner's [2] results. In addition, we show that the characterizations of the RDFs of the more general problems considered in [5,6] (i.e., where a noisy version of source is available at the encoder) do not reproduce Wyner's [2] results. Finally, we present simulations.

5.1. Connection with Other Works

Remark 4. The degenerate case to Wyner's [2] optimal test channel realizations. Now, we verify that for the tuple of scalar-valued, jointly Gaussian RVs (X, Y) , with square error distortion function specified below, our optimal realizations of $\hat{X}$ and closed form expressions for $R_{X|Y}(\Delta_X)$ and $\bar{R}(\Delta_X)$ are identical to Wyner's [2] realizations and RDFs (see Figure 4). Let us define:

$$X : \Omega \rightarrow \mathcal{X} \triangleq \mathbb{R}, \quad Y : \Omega \rightarrow \mathcal{Y} \triangleq \mathbb{R}, \quad \hat{X} : \Omega \rightarrow \hat{\mathcal{X}} \triangleq \mathbb{R}, \quad (165)$$

$$d_X(x, \hat{x}) = (x - \hat{x})^2, \quad (166)$$

$$X \in N(0, \sigma_X^2), \quad \sigma_X^2 > 0, \quad Y = \alpha(X + U), \quad (167)$$

$$U \in N(0, \sigma_U^2), \quad \sigma_U^2 > 0, \quad \alpha > 0. \quad (168)$$

(a) RDF $R_{X|Y}(\Delta_X)$: By Theorem 4a applied to (165)–(168), we obtain

$$Q_X = \sigma_X^2, \quad Q_{X,Y} = \alpha\sigma_X^2, \quad Q_Y = \sigma_Y^2 = \alpha^2\sigma_X^2 + \alpha^2\sigma_U^2, \quad Q_{X|Y} = c\sigma_U^2, \quad c \triangleq \frac{\sigma_X^2}{\sigma_X^2 + \sigma_U^2}, \quad (169)$$

$$H = 1 - \Delta_X Q_{X|Y}^{-1} = \frac{c\sigma_U^2 - d}{c\sigma_U^2} \equiv a, \quad Q_{X,Y}Q_Y^{-1} = \frac{c}{\alpha}, \quad H Q_{X,Y}Q_Y^{-1} = \frac{ac}{\alpha}, \quad (170)$$

$$W = H\Psi = a\Psi, \quad Q_\Psi = H^{-1}\Delta_X = \frac{\Delta_X}{a} = \frac{c\sigma_U^2\Delta_X}{c\sigma_U^2 - \Delta_X}, \quad c\sigma_U^2 - \Delta_X > 0. \quad (171)$$

Moreover, by Theorem 4b the optimal reproduction $\hat{X} \in \mathcal{M}_0(d)$ and $R_{X|Y}(d)$ are,

$$\hat{X} = a(X - \frac{c}{\alpha}Y) + \frac{c}{\alpha}Y + a\Psi, \quad c\sigma_U^2 - \Delta_X > 0 \quad (172)$$

$$R_{X|Y}(\Delta_X) = \begin{cases} \frac{1}{2} \log \frac{c\sigma_U^2}{\Delta_X}, & 0 < \Delta_X < c\sigma_U^2 \\ 0, & \Delta_X \geq c\sigma_U^2. \end{cases} \quad (173)$$

This shows our realization of Figure 3 degenerates to Wyner's [2] realization of Figure 4a.

(b) RDF $\bar{R}(\Delta_X)$: By Theorem 5b applied to (165)–(168), and using the calculations (169)–(172), we obtain

$$\hat{X} = f(Y, Z) = \frac{c}{\alpha}(1 - a)Y + Z \quad \text{by (172), (175),} \quad (174)$$

$$Z = a(X + \Psi), \quad (a, \Psi) \quad \text{defined in (170), (171)} \quad (175)$$

$$\bar{R}(\Delta_X) = R_{X|Y}(\Delta_X) = (173) \quad \text{by evaluating } I(X; Z) - I(Y; Z), \text{ using (4) and (175).} \quad (176)$$

This shows our value of $\bar{R}(\Delta_X)$ and optimal realization $\hat{X} = f(Y, Z)$ reproduce Wyner's optimal realization and the value of $\bar{R}(\Delta_X)$ given in [2] (i.e., Figure 4b).

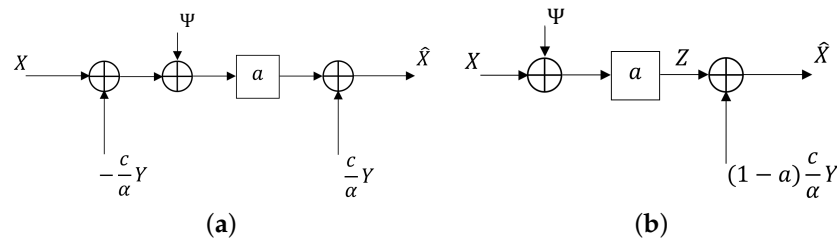


Figure 4. Wyner's realizations of optimal reproductions for RDFs $R_{X|Y}(\Delta_X)$ and $\bar{R}(\Delta_X)$. (a) RDF $R_{X|Y}(\Delta_X)$: Wyner's [2] optimal realization of $\hat{X}$ for RDF $R_{X|Y}(\Delta_X)$ of (165)–(168). (b) RDF $\bar{R}(\Delta_X)$: Wyner's [2] optimal realization $\hat{X} = f(X, Z)$ for RDF $\bar{R}(\Delta_X)$ of (165)–(168).

In the following Remark, we show that, when $S = X$ -a.s., the realization of the auxiliary RV Z , which is used in the proofs in [5,6] to show the converse coding theorem does not coincide with Wyner's realization [2]. Also, their realizations do not reproduce Wyner's RDF (this observation is verified for modified realization given in the correction note without proof in <https://tiangroup.engr.tamu.edu/publications/> (accessed on 3 January 2024)). The deficiency of the realizations in [5,6] to show the converse was first pointed out in [7], using an alternative proof.

Remark 5. Optimal test channel realization of [5,6]

(a) The derivation of [[5], Theorem 4], uses the following representation of RVs (see [[5], Equation (4)] adopted to our notation using (19)):

$$X = (K_{xs}K_{sy} + K_{xy})Y + K_{xs}N_1 + N_2, \quad S = K_{sy}Y + N_1,$$

where N_1 and N_2 are independent Gaussian RVs with zero mean, N_1 is independent Y and N_2 is independent of (S, Y) .

To reduce [5,6] to the Wyner and Ziv RDF, we set $X = S$ -a.s., which then implies, $K_{xs} = I$, $N_2 = 0$ -a.s and $K_{xy} = 0$. According to the derivation of the converse [[5], Theorem 4] (see [[5], 3 lines above Equation (32)] using our notation), the optimal realization of the auxiliary RV Z_T used to achieve the RDF is

$$Z_T = U^T X + N_3, \quad (177)$$

where $Q_{X|Y} = U \text{diag}(\lambda_1, \dots, \lambda_n) U^T$ and U is a unitary matrix, $N_3 \in N(0, Q_{N_3})$ such that Q_{N_3} is a diagonal covariance matrix, with elements given by (for the value of $\sigma_{3,i}^2$, we considered the one given in the correction note in <https://tiangroup.engr.tamu.edu/publications/> (accessed on 3 January 2024) (although no derivation is given), where it is stated that $\sigma_{3,i}^2$ that appeared in the derivation [[5], proof of theorem 4] should be multiplied by λ_i),

$$\sigma_{3,i}^2 = \frac{\min(\lambda_i, \delta_i)}{\lambda_i - \min(\lambda_i, \delta_i)} \lambda_i, \quad \sum_{i=1}^n \min(\lambda_i, \delta_i) = \Delta_X. \quad (178)$$

(b) It is easy to verify that the above realization of Z_T that uses the correction of footnote 6 is precisely the realization given in [[6], Theorem 3A].

(c) Special Case: For scalar-valued RVs the auxiliary RV Z_T reduces to

$$Z_T = X + N_3, \quad N_3 \in N(0, \frac{\Delta_X Q_{X|Y}}{Q_{X|Y} - \Delta_X}), \quad Q_{X|Y} > \Delta_X \quad (179)$$

Now, we examine whether the realization (179) corresponds to Wyner's realization and induces Wyner's RDF. Recall that the Wyner's [2] RDF, denoted by $R_{X;Z|Y}(\Delta_X)$ and corresponding to auxiliary RV Z , is

$$Z = HX + W, \quad H = \frac{Q_{X|Y} - \Delta_X}{Q_{X|Y}}, \quad W \in N(0, H\Delta_X), \quad (180)$$

$$R_{X;Z|Y}(\Delta_X) = I(X;Z|Y) = \frac{1}{2} \log \left(\frac{Q_{X|Y}}{\Delta_X} \right), \quad \Delta_X \leq Q_{X|Y}. \quad (181)$$

Clearly, the two realizations (179) and (180) are different. Let $\hat{R}_{X;Z_T|Y}(\Delta_X)$ denote the RDF corresponding to the realization Z_T . Then $\hat{R}_{X;Z_T|Y}(\Delta_X)$ can be computed using $I(X;Z_T|Y) = I(X;Z_T) - I(Y;Z_T) = -H(Z_T|X) + H(Z_T|Y)$ where $H(\cdot|\cdot)$ denotes the conditional differential entropy. Then, by using

$$Q_{Z_T|X} = Q_{N_3} = \frac{\Delta_X Q_{X|Y}}{Q_{X|Y} - \Delta_X}, \quad (182)$$

$$Q_{Z_T|Y} = Q_{N_3} + Q_{X|Y}. \quad (183)$$

it is straightforward to show that

$$\hat{R}_{X;Z_T|Y}(\Delta_X) = -H(Z_T|X) + H(Z_T|Y) \quad (184)$$

$$= -\frac{1}{2} \log(2\pi e \frac{\Delta_X Q_{X|Y}}{Q_{X|Y} - \Delta_X}) + \frac{1}{2} \log(2\pi e \frac{Q_{X|Y}^2}{Q_{X|Y} - \Delta_X}), \quad \Delta_X < Q_{X|Y} \quad (185)$$

However, we note that (i) unlike Wyner's RDF given in (181), which gives $R_{X;Z|Y}(\Delta_X) = 0$ at $\Delta_X = Q_{X|Y}$, the corresponding $\hat{R}_{X;Z_T|Y}(\Delta_X) = -\infty + \infty$ at $\Delta_X = Q_{X|Y}$, and (ii) Wyner's test channel realization is $Z = HX + W$, $H = \frac{Q_{X|Y} - \Delta_X}{Q_{X|Y}}$ and $W \in N(0, H\Delta_X)$, which is different from the test channel realization in (179). In particular, if $Q_{X|Y} = \Delta_X$, then $H = 0 \Rightarrow W \in N(0, 0)$ and $Z = 0$ -a.s. On the other hand, for the test channel in (179), if $Q_{X|Y} = \Delta_X$, then $N_3 \in N(0, +\infty)$, and thus the variance of Z_T in (179) is not zero.

Further, in Proposition 6, we prove that for the multi-dimensional source, the test channel realization in (179) does not achieve the RDF when water-filling is active, i.e., when at least one component of the source is not reproduced.

(d) *Special Case Classical RDF*: The classical RDF is obtained as a special case if we assume X and Y are independent or Y generates the trivial information $\{\Omega, \emptyset\}$; i.e., Y is nonrandom. Clearly, in this case, the RDF $\hat{R}_{S;Z_T|Y}(\Delta_X)$ should degenerate to the classical RDF of the source X , i.e., $R_X(\Delta_X)$, and it should be that $\hat{X} = Z_T$. However, for this case, (179) gives $Q_{Z_T} = Q_X + \frac{\Delta_X Q_X}{Q_X - \Delta_X} = \frac{Q_X^2}{Q_X - \Delta_X}$, which is fundamentally different from Wyner's degenerate, and correct values $Q_{\hat{X}} = Q_Z = \max\{0, Q_X - \Delta_X\}$.

Proposition 6. When $S = X$ -a.s., Wyner's [2] auxiliary RV Z and the auxiliary RV Z_T given in (177) i.e., the degenerate case of [5,6] (with the correction of footnote 6), are not related by an invertible function. As a result, the computed RDF based on the two realizations are different.

Proof. Recall that, if the two auxiliary RVs Z_T and Z are not related by an invertible function, i.e., $Z = f(Z_T)$, where $f(\cdot)$ is invertible and both f and its inverse are measurable, then $I(X;Z_T) - I(Y;Z_T) \neq I(X;Z) - I(Y;Z)$. It was shown earlier in this paper (and also in [7]) that for the multivariate Wyner's RDF, the auxiliary RV takes the form

$$Z = HX + W, \quad W \in N(0, Q_W), \quad (186)$$

where $Q_W = H\Sigma_\Delta = U\text{diag}(\sigma_{w,1}^2, \dots, \sigma_{w,n}^2)U^T$, $\Sigma_\Delta = U\text{diag}(\delta_1, \dots, \delta_n)U^T$, $H = I - Q_{X|Y}^{-1}\Sigma_\Delta = U\text{diag}(h_1, \dots, h_n)U^T$ and $Q_{X|Y} = U\text{diag}(\lambda_1, \dots, \lambda_n)U^T$, where U is a unitary matrix. The eigenvalues $\sigma_{w,i}^2$ and h_i are given by

$$\sigma_{w,i}^2 = \frac{\min(\lambda_i, \delta_i)(\lambda_i - \min(\lambda_i, \delta_i))}{\lambda_i}, \quad (187)$$

$$h_i = \frac{\lambda_i - \min(\lambda_i, \delta_i)}{\lambda_i}, \quad (188)$$

where $\sum_{i=1}^n \min(\lambda_i, \delta_i) = \Delta_X$. Hence, Equations (186), (187), and (188), imply that if $\sigma_{w,i}^2 = 0$ then $h_i = 0$, and vice versa. Such zero values correspond to compression with water-filling.

On the other hand, from (177) and (178), if water-filling is active, then $\sigma_{3,i}^2 = \frac{\lambda_i^2}{\lambda_i - \lambda_i}$. Moreover, by comparing Equations (187) with (178) and (188) with (177), it is straightforward to show that $f(\cdot) = HU$. If HU is not an invertible matrix for all values of the distortion Δ_X , then $I(X; Z_T) - I(Y; Z_T) \neq I(X; Z) - I(Y; Z)$.

By (188) it is easy to show that if $\min(\lambda_i, \delta_i) = \lambda_i$, HU is not invertible. This implies $I(X; Z_T) - I(Y; Z_T) \neq I(X; Z) - I(Y; Z)$. $\square$

5.2. Simulations

In this section, we provide an example to show the gap between the classical rate distortion $R_X(\Delta_X)$ defined in (154), the conditional distortion function $R_{X|Y}(\Delta_X)$ (69), and to verify the validity of Gray's lower bound (8). Note that in Theorem 5 it is shown that $R_{X|Y}(\Delta_X) = \bar{R}(\Delta_X)$, and hence the plot for $\bar{R}(\Delta_X)$ is omitted. For the evaluation, we pick a joint covariance matrix (11) given by

$$Q_{(X,Y)} = \begin{bmatrix} 2.5000 & 1.1250 & 0.4750 & 0.6125 \\ 1.1250 & 0.8125 & 0.2750 & 0.3063 \\ 0.4750 & 0.2750 & 0.1525 & 0.1625 \\ 0.6125 & 0.3063 & 0.1625 & 0.2031 \end{bmatrix}, \quad X: \Omega \rightarrow \mathbb{R}^2, \quad Y: \Omega \rightarrow \mathbb{R}^2.$$

In order to compute the rates, we first have to find Q_X, Q_Y, Q_{XY} and $Q_{X|Y}$. From the definition of $Q_{(X,Y)}$ given in (11), it is easy to see that the covariance of X, Y , and the joint covariance of X and Y are equal to

$$Q_X = \begin{bmatrix} 2.5000 & 1.1250 \\ 1.1250 & 0.8125 \end{bmatrix}, \quad Q_Y = \begin{bmatrix} 0.1525 & 0.1625 \\ 0.1625 & 0.2031 \end{bmatrix}, \quad Q_{XY} = \begin{bmatrix} 0.4750 & 0.6125 \\ 0.2750 & 0.3063 \end{bmatrix}.$$

Then, the conditional covariance $Q_{X|Y}$, which appears in $R_{X|Y}(\Delta_X)$, can be computed from (27). Using Singular Value Decomposition (SVD), we can calculate the eigenvalues of $Q_{X|Y}$. For this case, the eigenvalues of the conditional covariance are $\{0.7538, 0.2\}$. Similarly, the eigenvalues of Q_X can be determined. Finally, the eigenvalues of Q_X and $Q_{X|Y}$ are passed to the water-filling to compute the $R_X(\Delta_X)$ and $R_{X|Y}(\Delta_X)$, respectively.

The classical rate distortion, the conditional RDF, and the Gray's lower bound for the joint covariance above are illustrated in Figure 5. It is clear that $R_{X|Y}(\Delta_X)$ is smaller, and as the distortion Δ_X increases, the gap between the classical and conditional RDF becomes larger. Gray's lower bound is achievable for some positive distortion values, as provided in (71), i.e., for $\Delta_X \in \{\Delta_X \in [0, \infty) : \Delta_X \leq n_x \lambda_{n_x}\}$. Recall that the set of eigenvalues of $Q_{X|Y}$ is $\{0.7538, 0.2\}$, and the lower bound is achievable for $\Delta_X \leq 2 \cdot 0.2 = 0.4$; i.e., for these values $R_{X|Y}(\Delta_X) = R_X(\Delta_X) - I(X; Y)$.

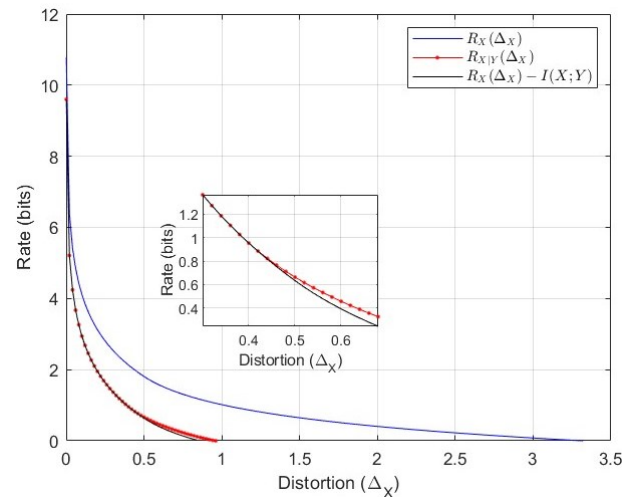


Figure 5. Comparison of classical RDF, $R_X(\Delta_X)$, conditional RDF $R_{X|Y}(\Delta_X) = \bar{R}(\Delta_X)$, and Gray's lower bound $R_X(\Delta_X) - I(X;Y)$ (solid green line).

6. Conclusions

We derived nontrivial structural properties of the optimal test channel realizations that achieve the optimal test channel distributions of the characterizations of RDFs for a tuple of multivariate jointly independent and identically distributed Gaussian random variables with mean-square error fidelity for two cases. Initially, the side information was available at the encoder and decoder, and then it was only available at the decoder. Using the realizations of the optimal test channels, we showed that when the side information is known to the encoder and the decoder, it does not achieve a better compression compared to when side information is only known to the decoder.

Author Contributions: M.G. and C.D.C. contributed to the conceptualization, methodology, and writing of the manuscript. All authors have read and agreed to the published version of the manuscript.

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

Data Availability Statement: Data are contained within the article.

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

Appendix A

Appendix A.1. Proof of Lemma 1

(a) By the chain rule of mutual information,

$$I(X; \hat{X}, Y) = I(X; Y | \hat{X}) + I(X; \hat{X}) \quad (A1)$$

$$= I(X; \hat{X} | Y) + I(X; Y) \quad (A2)$$

Since $I(X; Y | \hat{X}) \geq 0$, then from the above it follows

$$I(X; \hat{X}) \leq I(X; \hat{X} | Y) + I(X; Y) \quad (A3)$$

$$I(X; \hat{X} | Y) \geq I(X; \hat{X}) - I(X; Y) \quad (A4)$$

The above shows (40). However, the inequality holds with equality if and only if $I(X; Y|\hat{X}) = 0$, and this quantity is zero if and only if $\mathbf{P}_{X|\hat{X},Y} = \mathbf{P}_{X|\hat{X}}$. Alternatively, we note the following:

$$\begin{aligned} I(X; \hat{X}|Y) &= \mathbf{E} \left[\log \frac{\mathbf{P}_{X|\hat{X},Y}}{\mathbf{P}_{X|Y}} \right] \\ &= \mathbf{E} \left[\log \frac{\mathbf{P}_{X|\hat{X},Y}}{\mathbf{P}_{X|Y}} \frac{\mathbf{P}_X}{\mathbf{P}_X} \right] \\ &= \mathbf{E} \left[\log \frac{\mathbf{P}_{X|\hat{X},Y}}{\mathbf{P}_X} - \log \frac{\mathbf{P}_{X|Y}}{\mathbf{P}_X} \right] \\ &= \mathbf{E} \left[\log \frac{\mathbf{P}_{X|\hat{X}}}{\mathbf{P}_X} - \log \frac{\mathbf{P}_{X|Y}}{\mathbf{P}_X} \right], \text{ if and only if } \mathbf{P}_{X|\hat{X},Y} = \mathbf{P}_{X|\hat{X}}. \end{aligned}$$

This completes the statement of equality of (40); i.e., it establishes equality (41). (b) Consider a test channel $\mathbf{P}_{X|\hat{X},Y}$ such that $\mathbf{E}\{\|X - \hat{X}\|_{\mathbb{R}^{n_x}}^2 \leq \Delta_X\}$, i.e., $\hat{X} \in \mathcal{M}_0(\Delta_X)$, and such that $\mathbf{P}_{X|\hat{X},Y} = \mathbf{P}_{X|\hat{X}}$, for $\Delta_X \in \mathcal{D}_C(X|Y) \subseteq [0, \infty)$. By (41) taking the infimum of both sides over $\hat{X} \in \mathcal{M}_0(\Delta_X)$ such that $\mathbf{P}_{X|\hat{X},Y} = \mathbf{P}_{X|\hat{X}}$, then (43) is obtained on a nontrivial surface, i.e., $\Delta_X \in \mathcal{D}_C(X|Y)$, which exists due to continuity and convexity of $R_X(\Delta_X)$ for $\Delta_X \in (0, \infty)$. This completes the proof.

Appendix A.2. Proof of Theorem 2

(a) (1) By properties of conditional mutual information [18],

$$I(X; \hat{X}|Y) \stackrel{(\alpha)}{=} I(X; \hat{X}, \bar{X}^{cm}|Y) \quad (\text{A5})$$

$$\stackrel{(\beta)}{=} I(X; \hat{X}|\bar{X}^{cm}, Y) + I(X; \bar{X}^{cm}|Y) \quad (\text{A6})$$

$$\stackrel{(\gamma)}{\geq} I(X; \bar{X}^{cm}|Y) \quad (\text{A7})$$

where (α) is due to $\bar{X}^{cm}$ being a function of $(Y, \hat{X})$, and a well-known property of the mutual information [18] (β) is due to the chain rule of mutual information [18], and (γ) is due to $I(X; \hat{X}|\bar{X}^{cm}, Y) \geq 0$. Hence, inequality (45) is shown. (2) If (i) holds, i.e., $\hat{X} = \bar{X}^{cm} - a.s.$, then $I(X; \hat{X}|\bar{X}^{cm}, Y) = 0$, and hence the inequality (45) becomes an equality. If (ii) holds, since for fixed $y \in \mathcal{Y}$ the function $e(y, \cdot) : \mathcal{X} \rightarrow \mathcal{X}$, $e(y, \hat{x}) = \bar{x}^{cm}$ uniquely defines $\hat{x}$, then $I(X; \hat{X}|\bar{X}^{cm}, Y) = 0$, and the inequality (45) becomes an equality.

(b) The inequality (48) is well known due to the orthogonal projection theorem.

Appendix A.3. Proof of Theorem 7

Consider the realization (88). We identify the triple (H, G, Q_W) such that (84) or (87) hold; i.e., we characterize the set $\mathcal{M}_0^{G,o}(\Delta_X)$.

Case (i). $\text{cov}(\hat{X}, \hat{X}|Y) \succ 0$, that is, $\text{rank}(Q_{\hat{X}|Y}) = n_x$. By Theorem 6, Case (i), we seek the triple (H, G, Q_W) such that (84) holds, i.e., $\bar{X}^{cm} = \hat{X} - a.s.$ Recall that Conditions 1 and 2 of Theorem 6 are sufficient for $\hat{X} = \bar{X}^{cm}$.

Condition 1, i.e., (85). The left-hand side part of (85) is given by (this follows from mean-square estimation theory, or an application of (26) with $\mathcal{G} = \{\Omega, \emptyset\}$)

$$\mathbf{E}(X|Y) = \mathbf{E}(X) + \text{cov}(X, Y)\text{cov}(Y, Y)^{-1}(Y - \mathbf{E}(Y)) \quad (\text{A8})$$

$$\begin{aligned} &= \text{cov}(X, Y)\text{cov}(Y, Y)^{-1}Y \\ &= Q_{X,Y}Q_Y^{-1}Y \end{aligned} \quad (\text{A9})$$

$$= Q_X C^T Q_Y^{-1}Y \quad \text{by model (15)–(18)}. \quad (\text{A10})$$

Similarly, the right hand side of (85) is given by

$$\mathbf{E}(\hat{X}|Y) = \mathbf{E}(\hat{X}) + \text{cov}(\hat{X}, Y) \text{cov}(Y, Y)^{-1} (Y - \mathbf{E}(Y)) \quad (\text{A11})$$

$$\begin{aligned} &= \text{cov}(\hat{X}, Y) \text{cov}(Y, Y)^{-1} Y \\ &= (HQ_{X,Y} + GQ_Y) Q_Y^{-1} Y \end{aligned} \quad (\text{A12})$$

$$= (HQ_X C^T + GQ_Y) Q_Y^{-1} Y \quad \text{by (15)–(18)} \quad (\text{A13})$$

Equating (A9) and (A12), then

$$\mathbf{E}(X|Y) = \mathbf{E}(\hat{X}|Y) \quad (\text{A14})$$

$$\Rightarrow Q_{X,Y} Q_Y^{-1} Y = (HQ_{X,Y} + GQ_Y) Q_Y^{-1} Y \quad \text{by (A12)} \quad (\text{A15})$$

$$\Rightarrow G = (I - H) Q_{X,Y} Q_Y^{-1} \quad (\text{A16})$$

$$\Rightarrow G = (I - H) Q_X C^T Q_Y^{-1} \quad \text{by (15)–(18)}. \quad (\text{A17})$$

Hence, G is obtained, and the reproduction is represented by

$$\hat{X} = HX + (I - H) Q_{X,Y} Q_Y^{-1} Y + W, \quad (\text{A18})$$

$$\mathbf{E}(\hat{X}|Y) = Q_{X,Y} Q_Y^{-1} Y = \mathbf{E}(X|Y), \quad (\text{A19})$$

$$\hat{X} - \mathbf{E}(\hat{X}|Y) = HX - HQ_{X,Y} Q_Y^{-1} Y + W. \quad (\text{A20})$$

Condition 2, i.e., (86). To apply (86), the following calculations are needed.

$$Q_{X|Y} \triangleq \text{cov}(X, X|Y) \quad (\text{A21})$$

$$\begin{aligned} &= \mathbf{E}\left\{ \left(X - \mathbf{E}(X|Y) \right) \left(X - \mathbf{E}(X|Y) \right)^T \right\} \\ &= Q_X - Q_{X,Y} Q_Y^{-1} Q_{X,Y}^T \end{aligned} \quad (\text{A22})$$

$$= Q_X - Q_X C^T Q_Y^{-1} C Q_X \quad \text{by (15)–(18)} \quad (\text{A23})$$

$$\begin{aligned} \text{cov}(X, \hat{X}|Y) &\triangleq \mathbf{E}\left\{ \left(X - \mathbf{E}(X|Y) \right) \left(\hat{X} - \mathbf{E}(\hat{X}|Y) \right)^T \right\} \\ &= \mathbf{E}\left\{ \left(X - \mathbf{E}(X|Y) \right) \left(\hat{X} - \mathbf{E}(X|Y) \right)^T \right\} \quad \text{by (A19)} \end{aligned} \quad (\text{A24})$$

$$= \mathbf{E}\left\{ \left(X - \mathbf{E}(X|Y) \right) \left(\hat{X} \right)^T \right\} \quad \text{by orthogonality} \quad (\text{A25})$$

$$= Q_X H^T - Q_{X,Y} Q_Y^{-1} Q_{Y,X} H^T \quad \text{by (A18), (A19)} \quad (\text{A26})$$

$$= Q_X H^T - Q_X C^T Q_Y^{-1} C Q_X H^T \quad \text{by (15)–(18)} \quad (\text{A27})$$

$$\begin{aligned} &= (Q_X - Q_X C^T Q_Y^{-1} C Q_X) H^T \\ &= Q_{X|Y} H^T. \end{aligned} \quad (\text{A28})$$

$$\text{cov}(\hat{X}, \hat{X}|Y) \triangleq \mathbf{E}\left\{\left(\hat{X} - \mathbf{E}(\hat{X}|Y)\right)\left(\hat{X} - \mathbf{E}(\hat{X}|Y)\right)^T\right\} \quad (\text{A29})$$

$$= HQ_X H^T + Q_W - HQ_{X,Y} Q_Y^{-1} Q_{Y,X} H^T \quad \text{by (A20)}$$

$$= HQ_X H^T + Q_W - HQ_X C^T Q_Y^{-1} C Q_X H^T \quad \text{by (15)–(18)} \quad (\text{A30})$$

$$= H(Q_X - Q_X C^T Q_Y^{-1} C Q_X) H^T + Q_W$$

$$= HQ_{X|Y} H^T + Q_W. \quad (\text{A31})$$

By Condition 2 and (A28) and (A31),

$$\text{cov}(X, \hat{X}|Y) \text{cov}(\hat{X}, \hat{X}|Y)^{-1} = I_{n_x} \quad (\text{A32})$$

$$\implies Q_{X|Y} H^T (HQ_{X|Y} H^T + Q_W)^{-1} = I_{n_x}$$

$$\implies Q_W = Q_{X|Y} H^T - HQ_{X|Y} H^T \quad (\text{A33})$$

$$\implies Q_W = (I_{n_x} - H) Q_{X|Y} H^T. \quad (\text{A34})$$

It remains to show $Q_W = Q_W^T$. This will follow shortly by identifying the equation for H as follows. Conditions 1 and 2 imply

$$\Sigma_\Delta = \text{cov}(X, X|Y, \hat{X}) \quad (\text{A35})$$

$$= \text{cov}(X, X|Y) - \text{cov}(X, \hat{X}|Y) \text{cov}(\hat{X}, \hat{X}|Y)^{-1} \text{cov}(X, \hat{X}|Y)^T, \text{ mboxby Proposition 1, (26)} \quad (\text{A36})$$

$$= \text{cov}(X, X|Y) - \text{cov}(X, \hat{X}|Y)^T, \quad \text{by (86)} \quad (\text{A37})$$

$$= Q_{X|Y} - HQ_{X|Y}, \quad \text{by (A28)}. \quad (\text{A38})$$

$$\implies HQ_{X|Y} = Q_{X|Y} - \Sigma_\Delta \quad (\text{A39})$$

$$\implies H = I - \Sigma_\Delta Q_{X|Y}^{-1} \quad (\text{A40})$$

By (A39), it then follows from (A33) that $Q_W = Q_W^T$. From the specification of G the equation of Q_W given by (A33) and $HQ_{X|Y}$, H given by (A39), and (A40) then follows the realization of Theorem 4.(a) for the case $Q_{X|Y} - \Sigma_\Delta \succ 0$. Properties (58)–(61) are easily verified.

Case (ii). $\text{cov}(\hat{X}, \hat{X}|Y) \succeq 0$ but not $\text{cov}(\hat{X}, \hat{X}|Y) \succ 0$, that is, $\text{rank}(Q_{\hat{X}|Y}) = n_1 < n_x$. We can verify that the stated realization in Theorem 4.(a) is such that Condition (87) holds. By (83) and the above calculations, we have

$$\bar{X}^{cm} = e^G(Y, \hat{X}) = \mathbf{E}(X|Y) + \text{cov}(X, \hat{X}|Y) \left\{ \text{cov}(\hat{X}, \hat{X}|Y) \right\}^+ (\hat{X} - \mathbf{E}(\hat{X}|Y)) \quad (\text{A41})$$

$$= \mathbf{E}(X|Y) + (Q_{X|Y} - \Sigma_\Delta) (Q_{X|Y} - \Sigma_\Delta)^+ (\hat{X} - \mathbf{E}(\hat{X}|Y)). \quad (\text{A42})$$

Since $Q_{\hat{X}|Y} = Q_{X|Y} - \Sigma_\Delta$, $\mathbf{E}\left\{\left(\hat{X} - \mathbf{E}(\hat{X}|Y)\right)\left(\hat{X} - \mathbf{E}(\hat{X}|Y)\right)^T\right\} = Q_{X|Y} - \Sigma_\Delta$, and $\text{rank}(L) = n_1$, where $L = (Q_{X|Y} - \Sigma_\Delta) (Q_{X|Y} - \Sigma_\Delta)^+$, then an application of Proposition 3, implies that Condition (87) holds. Thus, we have established Theorem 3.(a) and the properties stated under Theorem 4.(a). (i)–(iv). Finally, (96)–(101) are obtained from the realization, and hence Theorem 3.(b) is achievable.

Appendix A.4. Proof of Corollary 1

(a) This part is a special case of a related statement in [22]. However, we include it for completeness. By linear algebra [21], given two matrices $A \in \mathcal{S}_+^{k \times k}$, $B \in \mathcal{S}_+^{k \times k}$, then the following statements are equivalent: (1) AB is normal and (2) $AB \succeq 0$, where AB normal means $(AB)(AB)^T = (AB)^T(AB)$. Note that AB is normal if and only if $AB = BA$; i.e., they

commute. Let $A = U_A D_A U_A^T, B = U_B D_B U_B^T, U_A U_A^T = I_k, U_B U_B^T = I_k$; i.e., there exists a spectral representation of A, B in terms of unitary matrices U_A, U_B and diagonal matrices D_A, D_B . Then, $AB \succeq 0$ if and only if the matrices A and B commute; i.e., $AB = BA$, and A and B commute if and only if $U_A = U_B$.

Suppose (105) holds. Letting $A = Q_{X|Y}, B = \Sigma_\Delta$, then $A = U_A D_A U_A^T, B = U_B D_B U_B^T, U_A U_A^T = I_{n_x}, U_B U_B^T = I_{n_x}, U_A = U_B$. Since $Q_{X|Y}^{-1} = A^{-1} = U_A D_A^{-1} U_A^T$, then $\Sigma_\Delta Q_{X|Y}^{-1} = Q_{X|Y}^{-1} \Sigma_\Delta$; i.e., they commute. Hence,

$$\begin{aligned} H_t^T &= (I_{n_x} - (\Sigma_\Delta Q_{X|Y}^{-1})^T)^T = I_{n_x} - (Q_{X|Y}^{-1})^T \Sigma_\Delta^T = I_{n_x} - Q_{X|Y}^{-1} \Sigma_\Delta \\ &= I_{n_x} - \Sigma_\Delta Q_{X|Y}^{-1} = H \quad \text{since } Q_{X|Y} \text{ and } \Sigma_\Delta \text{ commute.} \end{aligned} \quad (\text{A43})$$

By the definition of Q_W given in Theorem 4.(a), we have

$$Q_W = \Sigma_\Delta H^T = Q_W^T = H \Sigma_\Delta. \quad (\text{A44})$$

Substituting (A43) into (A44), then

$$Q_W = \Sigma_\Delta H. \quad (\text{A45})$$

Hence, $\{\Sigma_\Delta, \Sigma_{X|Y}, H, Q_W\}$ are all elements of $\mathcal{S}_+^{p \times p}$ having a spectral decomposition with respect to the same unitary matrix $UU^T = I_{n_x}$.

References

- Wyner, A.; Ziv, J. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Inf. Theory* **1976**, *22*, 1–10. [\[CrossRef\]](#) [\[CrossRef\]](#)
- Wyner, A. The rate-distortion function for source coding with side information at the decoder-II: General sources. *Inf. Control* **1978**, *38*, 60–80. [\[CrossRef\]](#) [\[CrossRef\]](#)
- Berger, T. *Rate Distortion Theory: A Mathematical Basis for Data Compression*; Englewood Cliffs: Prentice-Hall, NJ, USA, 1971.
- Gray, R.M. A new class of lower bounds to information rates of stationary sources via conditional rate-distortion functions. *IEEE Trans. Inf. Theory* **1973**, *19*, 480–489. [\[CrossRef\]](#) [\[CrossRef\]](#)
- Tian, C.; Chen, J. Remote Vector Gaussian Source Coding With Decoder Side Information Under Mutual Information and Distortion Constraints. *IEEE Trans. Inf. Theory* **2009**, *55*, 4676–4680. [\[CrossRef\]](#) [\[CrossRef\]](#)
- Zahedi, A.; Ostergaard, J.; Jensen, S.H.; Naylor, P.; Bech, S. Distributed remote vector gaussian source coding with covariance distortion constraints. In Proceedings of the 2014 IEEE International Symposium on Information Theory, Honolulu, HI, USA, 29 June–4 July 2014; pp. 586–590.
- Gkagkos, M.; Charalambous, C.D. Structural Properties of Test Channels of the RDF for Gaussian Multivariate Distributed Sources. In Proceedings of the 2021 IEEE International Symposium on Information Theory (ISIT), Melbourne, Australia, 12–20 July 2021; pp. 2631–2636. [\[CrossRef\]](#)
- Draper, S.C.; Wornell, G.W. Side information aware coding strategies for sensor networks. *IEEE J. Sel. Areas Commun.* **2004**, *22*, 966–976. [\[CrossRef\]](#) [\[CrossRef\]](#)
- Oohama, Y. Gaussian multiterminal source coding. *IEEE Trans. Inf. Theory* **1997**, *43*, 1912–1923. [\[CrossRef\]](#)
- Oohama, Y. Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder. *IEEE Trans. Inf. Theory* **2005**, *51*, 2577–2593. [\[CrossRef\]](#) [\[CrossRef\]](#)
- Viswanathan, H.; Berger, T. The quadratic Gaussian CEO problem. *IEEE Trans. Inf. Theory* **1997**, *43*, 1549–1559. [\[CrossRef\]](#) [\[CrossRef\]](#)
- Ekrem, E.; Ulukus, S. An outer bound for the vector Gaussian CEO problem. In Proceedings of the 2012 IEEE International Symposium on Information Theory Proceedings, Cambridge, MA, USA, 1–6 July 2012; pp. 576–580.
- Wang, J.; Chen, J. Vector Gaussian Multiterminal Source Coding. *IEEE Trans. Inf. Theory* **2014**, *60*, 5533–5552. [\[CrossRef\]](#) [\[CrossRef\]](#)
- Xu, Y.; Guang, X.; Lu, J.; Chen, J. Vector Gaussian Successive Refinement With Degraded Side Information. *IEEE Trans. Inf. Theory* **2021**, *67*, 6963–6982. [\[CrossRef\]](#) [\[CrossRef\]](#)
- Renna, F.; Wang, L.; Yuan, X.; Yang, J.; Reeves, G.; Calderbank, R.; Carin, L.; Rodrigues, M.R.D. Classification and Reconstruction of High-Dimensional Signals From Low-Dimensional Features in the Presence of Side Information. *IEEE Trans. Inf. Theory* **2016**, *62*, 6459–6492. [\[CrossRef\]](#) [\[CrossRef\]](#)
- Salehkalaibar, S.; Phan, B.; Khisti, A.; Yu, W. Rate-Distortion-Perception Tradeoff Based on the Conditional Perception Measure. In Proceedings of the 2023 Biennial Symposium on Communications (BSC), Montreal, QC, Canada, 4–7 July 2023; pp. 31–37.
- Gallager, R.G. *Information Theory and Reliable Communication*; John Wiley & Sons, Inc.: New York, NY, USA, 1968.

18. Pinsker, M.S. *The Information Stability of Gaussian Random Variables and Processes*; Holden-Day, Inc.: San Francisco, CA, USA, 1964; Volume 133, pp. 28–30.
19. Aries, A.; Liptser, R.; Shirayev, A. *Statistics of Random Processes II: Applications; Stochastic Modelling and Applied Probability*; Springer: New York, NY, USA, 2013.
20. van Schuppen, J. *Control and System Theory of Discrete-Time Stochastic Systems*; Number 923 in Communications and Control Engineering; Springer: Berlin/Heidelberg, Germany, 2021. [[CrossRef](#)]
21. Horn, R.A.; Johnson, C.R. (Eds.) *Matrix Analysis*, 2nd ed.; Cambridge University Press: New York, NY, USA, 2013.
22. Charalambous, C.; Charalambous, T.; Kourtellaris, C.; van Schuppen, J. Structural Properties of Nonanticipatory Epsilon Entropy of Multivariate Gaussian Sources. In Proceedings of the 2020 IEEE International Symposium on Information Theory, Los Angeles, CA, USA, 21–26 June 2020; pp. 586–590.
23. Gorbunov, A.K.; Pinsker, M.S. Prognostic Epsilon Entropy of a Gaussian Message and a Gaussian Source. *Probl. Inf. Transm.* **1974**, *10*, 93–109.
24. Ihara, S. *Information Theory for Continuous Systems*; World Scientific: Singapore, 1993.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.