Methods for Generalized Rate Distortion Problems in Sensor Networks

Francesco Lorenzi, February 2022

***Abstract*—Starting from the results obtained in Gkagkos et al. [6], where a generalization of Wyner-Ziv source coding is presented, we study past work to find the most relevant methods in rate distortion theory to tackle the sensor network source coding problem. In particular, a single-sensor problem with side information, in the case of multivariate Gaussian source and quadratic distortion is defined. It is analyzed using methods that generalize classical techniques. The concept of structural properties is described, and they are derived using properties of jointly Gaussian vectors.**

I. Introduction

The recent availability of inexpensive sensors equipped with communication interfaces opens the possibility of realizing large-scale sensor networks that may outperform localized measurement techniques as well as open new measurement opportunities.

From an Information Theory point of view, it is interesting to model a sensor network in order to find bounds on the information that is possible to collect from the group of sensors, and eventually develop efficient codes that approach those bounds. The similarity of this scenario to other problems in Information Theory, as multiple access channel, source coding of correlated sources, etc. facilitates the usage of known techniques.

Naturally, many different aspects of a network are to be included in the design of such system, as network-level protocols, physical level implementation and channel estimation. The interplay between those aspects may be untangled by some reasonable simplifying assumptions, as the non-cyclicity of the network.

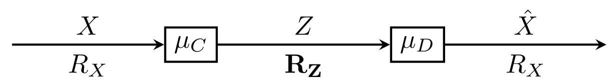
As a starting point, this essay focuses on the results of Gkagkos et al. [6], which in turn is an application of theoretical development by Gkagkos and Charlambous [5]. In this works light is shed onto the rate distortion problem in presence of side information, for the problem of Gaussian multivariate source coding. Several properties of Gaussian random vectors are used to establish an achievable lower bound on the rate, and to characterize the corresponding optimal test channel. The interesting point of view of *structural properties* is adopted.

The essay aims to shed light on the common methodological aspects behind rate distortion problems, and in general the problem of modelling and analyzing an information systems. As it is often the case, methods developed for specific problems are applicable by analogy in various scenarios, so a unifying point of view is useful for the solution of a rich class of problems.

A. Related work

Seminal work in rate distortion theory is done in Shannon [7, pp. 47-50], where for the first time the concept of *transmission with a fidelity criterion* was adopted. In this work, starting from a transmission problem, the problem of finding the rate of a source, subject to an average distortion bound was proposed. The original formulation of rate distortion theory needs to be interpreted in a multi-stage transmission scheme: in Dobrushin and Tsybakov [2] transmission in presence of distortion of the source was discussed, and the concept of *indirect* problem was introduced.

A result similar to the source channel separation theorem was derived in presence of noise. A method for *reducing* the distortion function over the stages of the system was here adopted, effectively reducing the problem to the one schematized in Fig.1, with $\mathbf{R_Z}$ information rate of Z and $d(\cdot,\cdot)$ a distance between realizations of $X, \hat{X}$.



Minimize $\mathbf{R_Z}$ under the constraint $\mathbf{E}[d(X,\hat{X})] \leq D$

Fig. 1: Reduced rate-distortion problem

In the work of Witsenhausen [9] this very concept of reduction is elaborated further. Using probabilistic results from stochastic control [10], a general principle, called the *disconnection principle* for conditional expectation is provided, without proof. Using this principle, the average distortion across all the system is determined by conditioning on the channel variables. This concept is very useful, especially in the presence of *side information*.

A thoroughly studied rate distortion problem is the encoder-decoder scheme in Fig.2, in which side information can be present at the encoder and the decoder, or only at the decoder. The idea beyond the use of side information is that, if we have a variable that is dependent from the source information, it can improve the source recovery at the decoder, i.e. lower the rate at the same distortion level, even if we are not interested in decoding the variable itself.

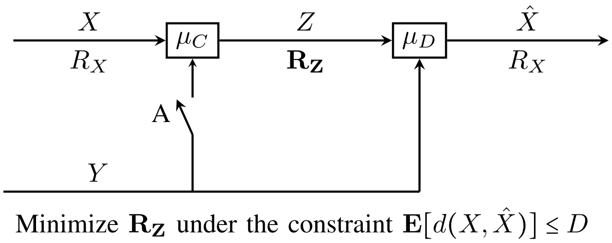


Fig. 2: Side information rate distortion problem (reduced)

The study of the scenario with side information only at the decoder presents some difficulties, and was carried out by Wyner and Ziv [11], where a lower bound is found and connected to other scenarios of side information. In Wyner [13] the result is extended to continuous variables, and it is established that, for the Gaussian-quadratic case, the encoder access to side information *does not allow for rate reduction*. The rate distortion coding has many aspects in common with the channel coding. In Cover and Chiang [1] the duality is studied, even in the case of side information, and a unified RDF expression for all the cases of side information is found. Finally, the duality transformation is formalized. An interesting concept regarding this is the one of *test channel* for a rate distortion scheme: the relationship between the estimated variable $\hat{X}$ and the source variable X is described by a transition probability $f(X|\hat{X})$ between symbols, and by similarity to channel one it is called *test channel*. The problem of finding the RDF is closely related to the characterization of the test channel, which represents the $\arg\min$ in the constrained optimization problem. As argued [5], the problem of finding the RDF is only a part of the complete determination of the variables in the coding scheme, as structural properties, which are properties of the interplay between variables in the coding scheme, are useful as well.

Side information is particularly interesting, as a model, in the case of distributed sensing, as shown in Draper and Wornell [3]. A sensor in a network has to communicate the information coming from other sensors to a reference node, encoding it with measurements the sensor itself performs. If the goal is to decode separately the measurements of *each* sensor, this framework is equivalent to the multiple access scenario, and the corresponding rate distortion problem can be addressed starting from the analysis of Slepian-Wolf achievable rate region [8]. It constitutes a multiple-user problem. However, if the goal is to decode the information coming from a single source, and measured through multiple sensors, the problem simplifies. We focus on the last case, in particular to the single-step communication between two network nodes. The result in [6] is the characterization of RDF and test channel in this scenario, in the case of Gaussian vector sensing, with Gaussian side information, with respect to MSE distortion.

B. Essay organization

The essay is organized as follows: in section II we formalize the problem, and recall some important results from previous works, as the formalism of conditional mutual information. The reduction procedure for average distortion and the disconnection principle are commented. In section III the assumption of Gaussianity proposed in [6] is commented, along with known properties of jointly Gaussian vectors. A lower bound on mutual conditional information is obtained. In section IV the original solution and characterization proposed in [6] and [5] is computed from the assumptions and conditions of III. In V we conclude and also we propose some future work.

Finally, section VI is an appendix, where calculations regarding the quadratic-Gaussian case are given to clarify some of the calculations in the essay. Moreover, some comments are made on the formalism of conditional expectation.

C. Notation

In this essay we indicate with $p_{X,Y}$ the joint probability distribution function of the random variables X, Y. R_A is the symbol rate of a specific link which transmit i.i.d. symbols distributed as the A random variable. A random variable is denoted by uppercase letter X, its alphabet as the same letter calligraphic uppercase $\mathfrak{X}$, and one of its element as a lowercase letter x. The number of dimensions of a random vector X is indicated with n_X. For Gaussian vectors, Q indicates a covariance matrix ($Q = Q^T$), $Q \succ 0$ indicates positive definiteness, and $Q \succeq 0$ indicates positive semi-definiteness. If the Gaussian vector is composed of sub-vectors X and Y, with Gaussian marginals, the total covariance matrix is decomposed in blocks with this notation

$$Q = \left[\begin{array}{c|c} Q_X & Q_{XY} \\ \hline Q_{XY}^T & Q_Y \end{array}\right] \tag{1}$$

where each matrix is defined as

$$Q_{UV} = \mathrm{cov}(U, V) = \mathbf{E}[(U - \mathbf{E}[U])(V - \mathbf{E}[V])^T]. \tag{2}$$

Conditional covariance is defined naturally through conditional expectation. A special notation is used for conditional independence. X is independent of Y given Z is written as

$$(X \perp\!\!\!\perp Y)|Z \tag{3}$$

and it is equivalent to the equation for distributions

$$p_{X|ZY} = p_{X|Z}. \tag{4}$$

If $X \leftrightarrow Y \leftrightarrow \hat{X}$ is a Markov chain, then

$$(X \perp\!\!\!\perp \hat{X})|Y. \tag{5}$$

II. Formalization and important results

Let us proceed by successive generalizations: the class of problems addressed in [6] is broader than the standard problem of rate distortion theory, so, starting from the standard problem, we introduce the elements which generalize the tractation.

A. Classical rate distortion problems

The classical rate distortion problem is formulated in the following way: let there be a fixed source of i.i.d. symbols

$\{X_k\}$, and let μ_C and μ_D an encoder-decoder pair such that the rate of the code is R. Let d be a distortion measure between the input and reproduced symbols, often defined as Hamming distance or euclidean norm (MSE). The question to be answered is which pairs (R, D) of rate and average distortion are achievable, which means that there exists a sequence of encoder-decoder pairs for which, in the limit, we have rate R and distortion D.

The rate distortion theorem states that

Theorem 1. *For a given source X, a pair (R, D), is achievable if and only if $R \geq \alpha(D)$, where α is the rate distortion function associated with the source, and defined as*

$$\alpha(D) = \min_{p_{X|\hat{X}} \in \mathcal{M}(D)} I(\hat{X}; X) \tag{6}$$

where

$$\mathcal{M}(D) = \left\{ p_{X|\hat{X}} \,\middle|\, \mathbf{E}[d(\hat{X}, X)] \leq D \right\} \tag{7}$$

notice that the rate distortion function is calculated with respect to a single symbol. Let us recall the classical rate distortion function for Gaussian variables. Let $X \sim \mathcal{N}(0, \sigma^2)$

$$\alpha(D) = \begin{cases} \frac{1}{2}\log\left(\frac{\sigma^2}{D}\right) & \text{when } 0 \leq D \leq \sigma^2, \\ 0 & \text{when } D > \sigma^2. \end{cases} \tag{8}$$

in section IV a similar expression will be shown for our generalized problem.

The first generalization is the presence of side information.

B. Side information at the decoder

The usage of side information give rise to an alteration of the scheme of rate distortion theory. The main results are given in [11], and then generalized in [13]. Since we are interested in sensor network application, the scheme contains a souce p_X, which is the magnitude to be sensed by the network, and the transducer itself, that generates the measure, as in Fig.3

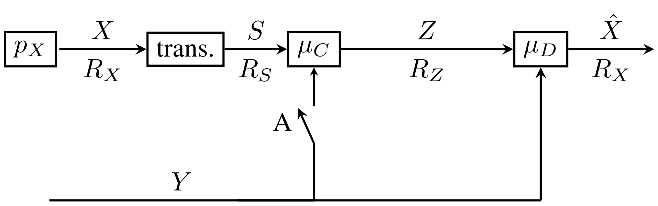


Fig. 3: Sensing and estimation in a single node of a sensor network

A fundamental Markov relation is that Z and Y are conditionally independent given X, namely

$$(Z \perp\!\!\!\perp Y)|X. \tag{9}$$

Consider the problem to find the minimum rate at which we can transmit X and recover it with an average distortion less than a fixed quantity. This solved finding the *rate distortion function* for the coding scheme. We call such functions α_0 and α_1, respectively for the open A switch and for the closed one. The problem of expressing α_0, which indicates the presence of side information *only at the decoder* has been solved[11]. Referring to the scheme in Fig.3, with A open, it holds

$$\alpha_0(D) = \min_{\mathcal{M}_0(D)} I(S; Z) - I(Y; Z) \tag{10}$$

where the constrain set is the set of all the distribution, or equivalently the set of all the variables, for which

$$\mathcal{M}_0(D) := \{Z : (Z \perp\!\!\!\perp X)|S, Y; \tag{11}$$
$$\exists f : \hat{X} = f(Y, Z), \mathbf{E}[d(X, \hat{X})] < D\} \tag{12}$$

it is important to notice that the expression (10) can be written as a conditional mutual information, with condition on the side information

$$I(X; Z) - I(Y; Z) \tag{13}$$
$$= H(Z) - H(Z|X) - H(Z) + H(Z|Y) = \tag{14}$$
$$\overset{(*)}{=} H(Z|Y) - H(Z|XY) = \tag{15}$$
$$= I(Z; X|Y) \tag{16}$$

where $(*)$ holds because of conditional independence in (9).

As for α_1, the solution is less difficult, and summarized in [11]

$$\alpha_1(D) = \min_{\mathcal{M}_1(D)} I(S; \hat{X}|Y) \tag{17}$$

where the constrain set is

$$\mathcal{M}_1(D) := \{\hat{X} : (\hat{X} \perp\!\!\!\perp X)|S, Y; \tag{18}$$
$$\exists f : \hat{X} = f(Y, Z), \mathbf{E}[d(X, \hat{X})] < D\} \tag{19}$$

Equations (10) and (17) are the starting point for our discussion on sensor source coding problem.

C. Indirect problem and disconnection principle

When considering the full transmission problem between sensors, we consider the entropy of the source and the capacity of the transmission channel. For a single user scheme, the *source-channel separation theorem*[1] in the presence of noise holds: transmission is possible if and only if the source entropy rate (written as the rate distortion function, if in presence of allowed distortion), is less than the channel capacity [7]. This theorem, as originally formulated, used a simple scheme of channel coding for a given source. In a full transmission situation we could have the *sensing* part C_{in} at the transmitting end which produces a measurement error, as well as an *estimation* part C_{out} at the receiving end, which produces an estimation error. This formulation is called the *indirect problem*. The influence of the input and output disturbances, can be taken into account by a reduction method [2] [10]. In [9] the indirect problem of rate distortion was described as in Fig.4.

[1]there is no such general result for multi-user problems.

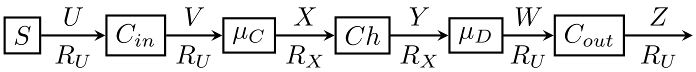


Fig. 4: Setup of a composite rate distortion problem

In order to compute the rate distortion function, used for the assessment of feasibility of transmission, we decouple the variables. Using the Markov chain relation

$$U \leftrightarrow V \leftrightarrow X \leftrightarrow Y \leftrightarrow W \leftrightarrow Z \tag{20}$$

and conditional independence conditions,

$$\mathbf{E}[d(U,Z)] = \tag{21}$$
$$= \sum_{\mathfrak{U},\mathfrak{Z}} p_{U,Z}(u,z)d(u,z) \tag{22}$$
$$= \sum_{\mathfrak{U},\mathfrak{Z},\mathfrak{V},\mathfrak{W}} p_{U,Z|V,W}(u,z|v,w)\underbrace{p_{V,W}(v,w)}_{Ch}d(u,z) \tag{23}$$
$$= \sum_{\mathfrak{U},\mathfrak{Z},\mathfrak{V},\mathfrak{W}} \underbrace{p_{V|U}(v|u)}_{C_{in}}\underbrace{p_{Z|W}(z|w)}_{C_{out}}\frac{p_U(u)}{p_V(v)}\underbrace{p_{V,W}(v,w)}_{Ch}d(u,z) \tag{24}$$

So it is possible to define an *amended* distortion function between symbols, independent from the channel

$$\hat{d}(v,w) \coloneqq \frac{1}{p_V(v)}\sum_{\mathfrak{U},\mathfrak{Z}} p_{V|U}(v|u)p_{Z|W}(z|w)p_U(u)d(u,z) \tag{25}$$

and we recognize in (25) the expression as a conditional expectation, so

$$\hat{d}(V,W) \coloneqq \mathbf{E}[d(U,Z)|V,W] \tag{26}$$

this is related to the original by the tower property

$$\mathbf{E}[d(U,Z)] \coloneqq \mathbf{E}[\mathbf{E}\left[\hat{d}(U,Z)|V,W\right]] \tag{27}$$

this shows that the indirect problem constraint on distortion can be reduced to an equivalent direct problem which uses an amended distortion function.

Using an argument on probability space, it is possible to generalize the above result, and to state the *disconnection principle*. This concepts highlights the fact that the computation of the conditional expectation is not functionally dependent from the realization of noise in the channel side [10, sec. IV]. The principle is general, and can be applied also to side information schemes [10, Fig.6].

III. Lower bound

Let us show some important results for jointly Gaussian vectors and quadratic distortion

Theorem 2. *For a couple of jointly distributed random vectors (X,Y), it holds*

$$\arg\min_{g(\cdot)} \mathbf{E}[||X - g(Y)||^2] = \mathbf{E}[X|Y] \tag{28}$$

where minimization is over all possible deterministic functions. In the case of jointly Gaussian distributed vectors

$$\mathbf{E}[X|Y] = \mathbf{E}[X] + cov(X,Y)cov(Y,Y)^\dagger(Y - \mathbf{E}[Y]) \tag{29}$$

where the operation denoted by † is pseudoinversion.

The proof is shown for the invertible case in VI.

Let us consider the minimization problem described in (17). In order to find a lower bound on the conditional mutual information as in (10), a fundamental variable, the *conditional mean* is defined as

$$\hat{X}^{cm} = \mathbf{E}[X|\hat{X}Y] \tag{30}$$

it follows from theorem (2) that the conditional mean is the function of $(\hat{X},Y)$ that minimizes quadratic distortion,

$$\arg\min_{g(\cdot)} \mathbf{E}[||X - g(Y,\hat{X})||^2] = \hat{X}^{cm} \tag{31}$$

and using (29)

$$\hat{X}^{cm} = \mathbf{E}[X|Y] + \text{cov}(X,\hat{X}|Y)\text{cov}(\hat{X},\hat{X}|Y)^\dagger(\hat{X} - \mathbf{E}[\hat{X}|Y]). \tag{32}$$

Furthermore, being that $\hat{X}^{cm}$ is a deterministic function of $(\hat{X},Y)$, we have the lower bound

$$I(X;\hat{X}|Y) = I(X;\hat{X},f(Y,\hat{X})|Y) = \tag{33}$$
$$= I(X;\hat{X},\hat{X}^{cm}|Y) = \tag{34}$$
$$= I(X;\hat{X}|\hat{X}^{cm}Y) + I(X;\hat{X}^{cm}|Y) \tag{35}$$
$$\geq I(X;\hat{X}^{cm}|Y) \tag{36}$$

this can be viewed as an application of data processing inequality in conditional mutual information [12]. In order for the equality to hold, i.e. to achieve the lower bound, we need to have

$$\hat{X} = \hat{X}^{cm} \quad a.s. \tag{37}$$

In the case of an invertible covariance $\text{cov}(\hat{X},\hat{X}|Y)$, the above achievement is ensured by the following conditions

$$\mathbf{E}[X|Y] = \mathbf{E}[\hat{X}|Y] \tag{38}$$

and

$$\text{cov}(X,\hat{X}|Y)\text{cov}(\hat{X},\hat{X}|Y)^{-1} = I \tag{39}$$

as follows by inspection on equation (32).

IV. Structural properties

In general, as argued in [5], when finding a RDF, it is often the case that the minimization problem is tacked using mathematical optimization techniques. The concept of *structural properties*, which are properties of a test channel, may be useful and interesting as the RDF itself, as they describe the nature of the variables involved in the test channel. In this case the formulation of structural properties is even more interesting, as the Gaussian multivariate source X is partially observable, and side information Y is available.

A. Switch A closed

Starting from the lower bound formulated in the previous section, we want to characterize the expression of the test channel realization which achieves the RDF α_1, in the case of side information at the encoder and decoder. By finding a suitable parametric expression of the estimation $\hat{X}$, we are able to use the conditions (38) and (39) that allows us to achieve the bound given in (36). Moreover, the *constrained* optimization is achieved, as the optimization space include the parametrized expression of the estimator, as discussed below.

All the following use results on Gaussian vectors that are recalled in VI. Let us consider, using the same argument as classical rate distortion theory, that the variables that achieve the bound,

$$(X, S, Y, \hat{X}) \text{ are jointly Gaussian.} \tag{40}$$

The marginal distribution of (X, Y, S) is fixed as this is assumed to be the generalized source distribution. In order to characterize the test channel, we need to describe $\hat{X}$. In fact, $\hat{X}$ depends only on S, Y, as X is not directly observable. By property (40), the conditional probability $p_{\hat{X}|SY}$, has expected value which is linear in S, Y, and a covariance which do not depend from these variables. So it is possible to write a parametrized version of $\hat{X}$ as follows

$$\hat{X} = HS + GY + W \tag{41}$$

where

$$H \in \mathbb{R}^{n_x, n_s} \tag{42}$$
$$G \in \mathbb{R}^{n_x, n_y} \tag{43}$$

the variable W stands for independent white noise, and has a covariance matrix

$$Q_W \geq 0 \tag{44}$$

in order to find the RDF, we must know that such parametrized solution respect the distortion constraint. However, by applying conditions (38) (39), we will have that $\hat{X}^{cm} = \hat{X}$ a.s., so by the property (31), the distribution induced by such variable gives an average distortion less than every other choice of estimator. So our proposed solution satisfy the constraint.

Let us find the parameters (H, G, Q_W). Starting from (38), and writing the left side

$$\mathbf{E}[X|Y] = \tag{45}$$
$$= \text{cov}(X, X)\text{cov}(Y, Y)^{-1}Y = \tag{46}$$
$$= Q_{XY}Q_Y^{-1}Y \tag{47}$$
$$\tag{48}$$

and the right side

$$\mathbf{E}[\hat{X}|Y] = \tag{49}$$
$$= \text{cov}(\hat{X}, Y)\text{cov}(Y, Y)^{-1}Y = \tag{50}$$
$$= \text{cov}(HS + GY + W, Y)Q_Y^{-1}Y = \tag{51}$$
$$= (HQ_{SY} + GQ_Y)Q_Y^{-1}Y = \tag{52}$$
$$= HQ_{SY}Q_Y^{-1}Y + GY \tag{53}$$

using independence of noise. By equating the two sides, G is obtained

$$G = (Q_{XY} - HQ_{SY})Q_Y^{-1} \tag{54}$$

Using condition (39), starting from the first term of the left side and remembering (39)

$$\text{cov}(X, \hat{X}|Y) = \tag{55}$$
$$= \mathbf{E}[(X - \mathbf{E}[X|Y])(\hat{X} - \mathbf{E}[X|Y])^T] = \tag{56}$$
$$= \mathbf{E}[(X - \mathbf{E}[X|Y])(\hat{X})^T] = \tag{57}$$
$$= \mathbf{E}[(X - \mathbf{E}[X|Y])(HS + GY + W)^T] = \tag{58}$$
$$= Q_{SX}H^T + Q_{XY}G^T - \tag{59}$$
$$- Q_{XY}Q_Y^{-1}Q_{SY}^T H^T - Q_{XY}G^T = \tag{60}$$
$$= Q_{XS|Y}H^T \tag{61}$$
$$\tag{62}$$

the last equation is justified as

$$Q_{XS|Y} = cov(X, S|Y) = \tag{63}$$
$$= \mathbf{E}[(X - \mathbf{E}[X|Y])(S - \mathbf{E}[S|Y])^T] = \tag{64}$$
$$= \mathbf{E}[(X - Q_{XY}Q_Y^{-1}Y)(S - Q_{SY}Q_Y^{-1}Y)^T] = \tag{65}$$
$$= Q_{SX} - Q_{XY}Q_Y^{-1}Q_{SY}^T \tag{66}$$
$$\tag{67}$$

the second term of the left side is

$$\text{cov}(\hat{X}, \hat{X}|Y)^{-1} = \tag{68}$$
$$= \mathbf{E}[(\hat{X} - \mathbf{E}[X|Y])(\hat{X} - \mathbf{E}[X|Y])^T]^{-1} = \tag{69}$$
$$= (HQ_S H^T + HQ_{SY}G^T + \tag{70}$$
$$GQ_{SY}H^T + GQ_Y G^T + Q_W)^{-1} = \tag{71}$$
$$= (HQ_S H^T - HQ_{SY}Q_Y^{-1}Q_{SY}^T H^T + Q_W)^{-1} = \tag{72}$$
$$= (HQ_{S|Y}H^T + Q_W)^{-1} \tag{73}$$

Imposing the product to be the identity matrix we characterize the noise covariance

$$HQ_{S|Y}H^T + Q_W = Q_{XS|Y}H^T \tag{74}$$
$$\implies Q_W = Q_{XS|Y}H^T - HQ_{S|Y}H^T. \tag{75}$$

When the above conditions hold, we have $\hat{X} = \hat{X}^{cm}$ a.s., so it holds

$$\hat{X} = \mathbf{E}[X|\hat{X}Y] \tag{76}$$
$$\implies \mathbf{E}[X|\hat{X}Y] = \mathbf{E}[X|\hat{X}] \tag{77}$$
$$\implies p_{X|\hat{X}Y} = p_{X|\hat{X}} \tag{78}$$

the last implication is valid as for jointly Gaussian vectors uncorrelation implies independence.

By the Markov chain relations we have $(X \perp\!\!\!\perp \hat{X})|Y$, and $(X \perp\!\!\!\perp \hat{X})|SY$ so it is possible to complete the characterization writing

$$p_{\hat{X}|SYX} = p_{\hat{X}|SY} \tag{79}$$

The equations from (76) to (79), and their implications, consitute the *structural properties*. We can now characterize the RDF: let us introduce for simplicity the matrix Σ, defined as

$$\Sigma = cov(X, X|\hat{X}, Y) = \quad (80)$$

$$= \mathbf{E}[(X - \mathbf{E}[X|\hat{X}, Y)])(X - \mathbf{E}[X|\hat{X}, Y)])^T] = \quad (81)$$

$$= \mathbf{E}[(X - \hat{X})(X - \hat{X})^T] \quad (82)$$

$$= Q_{X|Y} - HQ_{XS|Y} \quad (83)$$

it is possible to write

$$HQ^T_{XS|Y} = Q_{X|Y} - \Sigma. \quad (84)$$

and also the average distortion for a particular choice of $\hat{X}$ can be computed as

$$D = tr(\Sigma). \quad (85)$$

The corresponding RDF can be written computing the mutual entropy inside the framework of the structural properties

$$\alpha_1(D) = \inf_{tr(\Sigma)\le D} \left[\frac{1}{2} \log \left(\frac{\det(Q_{\hat{X}|Y})}{\det(Q_W)} \right) \right] \quad (86)$$

with $Q_{\hat{X}|Y}$ that can be expressed as

$$Q_{\hat{X}|Y} = HQ_{S|Y}H^T + Q_W = Q_{X|Y} - \Sigma \ge 0 \quad (87)$$

As shown in [6, remarks III.1, 2] this result is justified as it is possible to derive less general well known results, as the one of Wyner [13], with fully observable source.

B. Switch A open

Finally, we need to extend the structural properties to the case of side information only at the decoder. For jointly Gaussian variables, we know from [13] that $\alpha_0(D) = \alpha_1(D)$. Nonetheless, the properties are formulated in a different manner, as it is different the encoding scheme. Let us assume the following structure, using G, H as before

$$\hat{X} = GY + Z \quad (88)$$

$$Z = HS + W \quad (89)$$

interesting considerations are made in [13, remarks at page 65], and [12, equation 3.11], in which is is stated that, if $S \leftrightarrow Y, Z \leftrightarrow \hat{X}$ is a Markov chain, as in our case, then

$$I(S; Z|Y) = I(S; \hat{X}|Y) \iff I(S; Z|\hat{X}Y) \quad (90)$$

being that $\hat{X} = f(Y, Z)$, the condition on the left is satisfied, and we can conclude $\alpha_0(D) = \alpha_1(D)$. In Fig.5 the scheme is drawn.

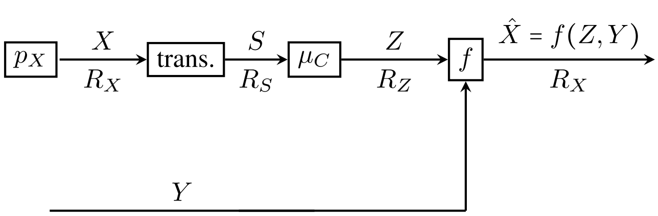


Fig. 5: Structure for characterization of $\alpha_0(D)$

The structural properties for the test channel follow from the properties derived in the case of switch A closed, using the additional properties

$$Q_{\hat{X}|Y} = Q_{Z|Y} \quad (91)$$

$$Q_{S|ZY} = Q_{S|\hat{X}Y} \quad (92)$$

V. Conclusions

In this essay, we studied different problems related to rate distortion theory and its applications in sensor networks. The multi-user framework was not used, as the focus was on a single communication problem in case of side information. Starting from well-known results for side information schemes, we computed the characterization of a particular test channel in the jointly Gaussian vector source. Being that the source is partially observable, we have shown methods for reducing generic indirect rate distortion problems to direct ones. Using several properties of Gaussian variables, a lower bound on conditional mutual information, valid for both cases of information at the encoder and the decoder, and only at the decoder, was calculated. Finally, utilizing achievability conditions and parametric expression of the estimation variable, the structural properties of such test channel were derived. In all the derivation we employed algebraic manipulations on Gaussian expectation and covariance matrices.

Further work may involve the proof of the disconnection principle [10], and the exploration of its implications in the scenario of sensor networks, and general problems. The full-fledged network information theory problem for sensor network remains untouched and may be a topic for further investigations. Structural properties can be used to obtain metrics on different realistic scenarios. Lastly, from the formal point of view, it may be interesting to refine the formalism using conditions over sub σ-algebras: that may give additional insight on the passages of characterization of the test channel.

VI. Appendix

The results shown in theorem (2) are well known results for random vectors. These results are important and well-known, but are also non trivial, so in this section some intuition about the proofs of those properties are given in A and B. In part C we comment on the formalism of conditioning over a sub σ-algebra, as done in Witsenhausen [9]. In the following, suppose X and Y are respectively a n, and m dimensional absolutely continuous random vectors.

A. Optimality of conditional expectation

Let us start from the first part of theorem (2). The tower property for conditional expectation asserts that

$$\mathbf{E}[||X - g(Y)||^2] = \mathbf{E}[\mathbf{E}[||X - g(Y)||^2|Y]] \quad (93)$$

we aim to minimize the left side of (93), by minimizing the conditional expectation for every $Y = y$. Using the definition of conditional expectation, with $f_{X|Y}$ conditional probability

$$\mathbf{E}[\|X - g(Y)\|^2 | Y = y] = \tag{94}$$

$$= \int_{\mathbb{R}^n} dx f_{X|Y}(x|y) \|x - g(y)\|^2 \tag{95}$$

$$= \int_{\mathbb{R}^n} dx f_{X|Y}(x|y) \big(x^T x - x^T g(y) - \tag{96}$$

$$- g(y)^T x + g(y)^T g(y) \big). \tag{97}$$

The minimization problem is in functional sense, but the expression to be minimized allows us to do a pointwise optimization, i.e. find the optimal values for $g = g(y)$, as the expression which involves g contains only algebraic operations[2]. By pointwise derivation we obtain the optimality condition on the gradient.

$$\nabla_g \left(\mathbf{E}[\|X - g(Y)\|^2 | Y = y] \right) = \tag{98}$$

$$= \int_{\mathbb{R}^n} dx f_{X|Y}(x|y) \left(-2x + 2g \right) = \tag{99}$$

$$= -2\mathbf{E}[X|Y = y] + 2g = 0. \tag{100}$$

Since the functional to be minimized is convex, we conclude that the minimum is

$$g(y) = \mathbf{E}[X|Y = y] \tag{101}$$

B. Conditional expectation in Gaussian case

Let us assume that the variables X, Y, as defined before, are also *jointly Gaussian*. Assume the Gaussian vector Z to be

$$Z = \left(\frac{X}{Y} \right) \sim \mathcal{N}(0, Q) \tag{102}$$

and the covariance matrix is

$$Q = \left[\begin{array}{c|c} Q_X & Q_{XY} \\ \hline Q_{XY}^T & Q_Y \end{array} \right] \tag{103}$$

To compute the conditional expectation we compute first the conditional probability $p_{X|Y}$. It turns out that such distribution is Gaussian, and its expectation is the conditional expectation. We assume all the covariance matrices to be nonsingular for simplicity. Assuming ξ to be a normalization constant

$$p_{X|Y}(x|y) = \frac{p_{XY}(x, y)}{p_Y(y)} = \tag{104}$$

$$= \xi \exp \left[-\frac{1}{2} \left(Z^T Q^{-1} Z - Y^T Q_Y^{-1} Y \right) \right]. \tag{105}$$

Recall now a simple algebraic result for the inverse of a block matrix

$$Q^{-1} = \left[\begin{array}{c|c} Q_X & Q_{XY} \\ \hline Q_{XY}^T & Q_Y \end{array} \right]^{-1} =: \left[\begin{array}{c|c} R_X & R_{XY} \\ \hline R_{XY}^T & R_Y \end{array} \right] \tag{106}$$

[2]The problem can be addressed also using Gateaux derivative, or, equivalently, using Euler-Lagrange equations: all the methods yield similar computations.

with

$$R_X = (Q_X - Q_{XY} Q_Y^{-1} Q_{XY}^T)^{-1} \tag{107}$$

$$R_{XY} = -(Q_X - Q_{XY} Q_Y^{-1} Q_{XY}^T)^{-1} Q_{XY} Q_Y^{-1} \tag{108}$$

$$R_Y = (Q_Y - Q_{XY}^T Q_X^{-1} Q_{XY})^{-1} \tag{109}$$

from this result it is simple to deduce

$$R_X^{-1} R_{XY} = -Q_{XY} Q_Y^{-1} \tag{110}$$

Starting from (105), it is possible to compute a new covariance matrix using a *square completion* argument

$$p_{X|Y}(x|y) = \tag{111}$$

$$= \xi' \exp \left[-\frac{1}{2} \left((x - Q_{XY} Q_Y^{-1} y)^T R_X (x - Q_{XY} Q_Y^{-1} y) \right) \right] \tag{112}$$

In conclusion, the conditioned variable has the following properties

$$\mathbf{E}[X|Y] = Q_{XY} Q_Y^{-1} Y \tag{113}$$

$$\mathbf{E}[X^2|Y] = Q_{X|Y} = Q_X - Q_{XY} Q_Y^{-1} Q_{XY}^T \tag{114}$$

From this result, with some algebraic passages we can also obtain the property, for a triple of jointly Gaussian vectors (X, Y, Z)

$$\text{cov}(X, Y|Z) = Q_{XY|Z} = Q_{XY} - Q_{XZ} Q_Z^{-1} Q_{YZ}^T \tag{115}$$

Finally, noticing that the conditional covariance is constant and do not depend from the conditioning variable, i.e. it is a degenerate variable, it is possible to write, in the Gaussian case

$$\text{cov}(X, Y|Z) = \tag{116}$$

$$= \mathbf{E}[(X - \mathbf{E}[X|Z])(Y - \mathbf{E}[Y|Z])^T | Z] = \tag{117}$$

$$= \mathbf{E}[(X - \mathbf{E}[X|Z])(Y - \mathbf{E}[Y|Z])^T]. \tag{118}$$

C. Conditional expectation with respect to sub σ-algebras

Some of the cited works in the essay use the formalism of conditioning with respect to a sub σ-algebra [10], [12]. The two representations are equivalent, given that any conditioning random variable Y is a measurable function on a probability space $(\Omega, \mathscr{F}, P)$, and the corresponding sub σ-algebra, which will be a subset of $\mathscr{F}$, is determined by all of the sets which corresponds to any particular realization of the variable Y. Inside any of these sets, any other variable $X(\omega)$ will have a particular distribution, and so a particular entropy for example. The general knowledge of Y induces a knowledge on the probability space, that is used to compute the value of other variables.

In Wyner [12] the conditioning is always written with respect to a sub σ-algebra: for example $P(A|\mathscr{B})$ is the probability of event A given the sub σ-algebra $\mathscr{B}$, and $P(A|\mathscr{B})$ is a $\mathscr{B}$-measurable function, i.e. a random variable over the probability space $(\Omega, \mathscr{B}, P)$. Conditional entropy is defined in the discrete case passing through the random variable $H(X|\mathscr{Y})$

$$H(X|\mathscr{Y}) = -\sum_{x \in \mathfrak{X}} P(X = x|\mathscr{Y}) \log(P(X = x|\mathscr{Y})) \tag{119}$$

which is a $\mathscr{Y}$-measurable function. Classical conditional entropy is obtained as an expectation, as it is well-known

$$H(X|Y) = \mathbf{E}[H(X|\mathscr{Y})]. \tag{120}$$

In the same way we can define also conditional mutual information [12], which is the main instrument for side information source coding.

From an operative point of view there is a compromise in the usage of this formalism: the augmented mathematical abstraction may be not justified by the results that it allows for. Mathematical details can be found in Kallenberg [4], and it may be interesting to explore further the notation possibilities of sub σ-algebras in application to Information Theory.

References

[1] T.M. Cover and Mung Chiang. "Duality between channel capacity and rate distortion with two-sided state information". In: *IEEE Transactions on Information Theory* 48.6 (June 2002), pp. 1629–1638.

[2] R. Dobrushin and B. Tsybakov. "Information transmission with additional noise". In: *IEEE Transactions on Information Theory* 8.5 (Sept. 1962), pp. 293–304.

[3] S.C. Draper and G.W. Wornell. "Side Information Aware Coding Strategies for Sensor Networks". In: *IEEE Journal on Selected Areas in Communications* 22.6 (Aug. 2004), pp. 966–976.

[4] O. Kallenberg. *Foundations of modern probability theory*. Springer, 2021.

[5] C. D. Charalambous M. Gkagkos. "Structural Properties of Optimal Test Channels for Distributed Source Coding with Decoder Side Information for Multivariate Gaussian Sources with Square-Error Fidelity".

[6] C. D. Charalambous M. Gkagkos E. Stylianou. "Structural Properties of Optimal Test Channels for Gaussian Multivariate Partially Observable Distributed Sources".

[7] C. E. Shannon. "A Mathematical Theory of Communication". In: *Bell System Technical Journal* 27.3 (July 1948), pp. 379–423.

[8] D. Slepian and J. Wolf. "Noiseless coding of correlated information sources". In: *IEEE Transactions on Information Theory* 19.4 (July 1973), pp. 471–480.

[9] H. Witsenhausen. "Indirect rate distortion problems". In: *IEEE Transactions on Information Theory* 26.5 (Sept. 1980), pp. 518–521.

[10] H.S. Witsenhausen. "On policy independence of conditional expectations". In: *Information and Control* 28.1 (May 1975), pp. 65–75.

[11] A. Wyner and J. Ziv. "The rate-distortion function for source coding with side information at the decoder". In: *IEEE Transactions on Information Theory* 22.1 (Jan. 1976), pp. 1–10.

[12] A.D. Wyner. "A definition of conditional mutual information for arbitrary ensembles". In: *Information and Control* 38.1 (July 1978), pp. 51–59.

[13] A.D. Wyner. "The rate-distortion function for source coding with side information at the decoder\3-II: General sources". In: *Information and Control* 38.1 (July 1978), pp. 60–80.