Source Coding in Networks with Covariance Distortion Constraints

Adel Zahedi, Jan Østergaard, *Senior Member, IEEE,* Søren Holdt Jensen, *Senior Member, IEEE,* Patrick A. Naylor, *Senior Member, IEEE,* and Søren Bech

***Abstract*—We consider a source coding problem with a network scenario in mind, and formulate it as a remote vector Gaussian Wyner-Ziv problem under covariance matrix distortions. We define a notion of minimum for two positive-definite matrices based on which we derive an explicit formula for the rate-distortion function (RDF). We then study the special cases and applications of this result. We show that two well-studied source coding problems, i.e. remote vector Gaussian Wyner-Ziv problems with mean-squared error and mutual information constraints are in fact special cases of our results. Finally, we apply our results to a joint source coding and denoising problem. We consider a network with a centralized topology and a given weighted sum-rate constraint, where the received signals at the center are to be fused to maximize the output SNR while enforcing no linear distortion. We show that one can design the distortion matrices at the nodes in order to maximize the output SNR at the fusion center. We thereby bridge between denoising and source coding within this setup.**

***Index Terms*—Source coding, Sensor networks, Covariance matrix distortions, Rate-distortion functions, Noise reduction**

I. Introduction

It is an inherent property of sensor networks that there are several observations of the desired source due to the availability of several sensors. If the observations at different nodes are correlated, it makes sense to make use of the correlation among the measurements to reduce the communication costs. This is particularly important for wireless networks where the resources, i.e. power and bandwidth are limited. In networks with a centralized topology, as shown in Fig. 1 (a), the measurements from all the nodes are to be sent directly to a fusion center. In this case, one could think of the data that is already available at the fusion center (sent by other nodes) as side information. For an ad hoc topology with communication among the nodes, as shown in Fig.1 (b), the observation at the receiving node could be treated as side information. In either case, due to the correlation between the side information at the receiver and the data to be sent, the required transmission rate could be reduced using distributed source coding [2], [3]. The block diagram of the resulting source coding problem with side information is shown in Fig. 2. The vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are the desired source, the observation at the encoder, and the side information available to the decoder, respectively. As shown in Fig. 2, an optimal estimation $\hat{\mathbf{x}}$ of $\mathbf{x}$ is obtained from the received data and the side information at the decoder. This setup is commonly referred to as the Wyner-Ziv problem after the celebrated work by Wyner and Ziv [4]. Note that this setup is less general than those shown in Fig. 1, where there may be several sources. However, we focus on this setup for simplicity.

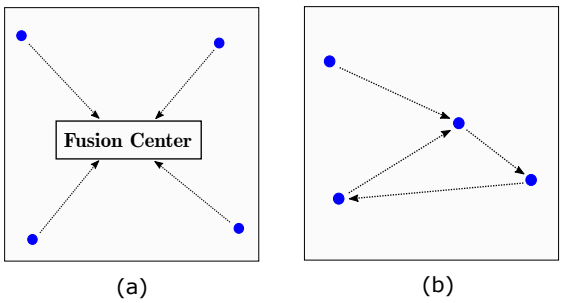


Fig. 1. Examples of wireless sensor networks (a) with a centralized topology, (b) with an ad hoc topology

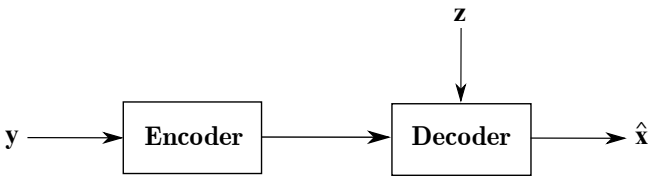


Fig. 2. Block diagram of the remote source coding problem

In this work, we consider i.i.d. vector sources. Note that in practice, the sources are correlated in time. By considering i.i.d. vector sources, we allow for correlation within each vector. This partially models the memory in the sources. To make full use of the memory in the signals, one has to either avoid the i.i.d. assumption, or let the dimension of the vector sources tend to infinity. For our derivations, we use the i.i.d. assumption and finite dimensions, and therefore accept some level of suboptimality.

The Wyner-Ziv problem first appeared as an attempt to generalize the Slepian and Wolf's results in [5] for lossless coding of two digital sources to the case of lossy source

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° ITN-GA-2012-316969. The work of J. Østergaard is financially supported by VILLUM FONDEN Young Investigator Programme, Project No. 10095. This work was presented in part at the 2014 IEEE International Symposium on Information Theory [11] and the IEEE Data Compression Conference [1].

A. Zahedi, J. Østergaard, and S. H. Jensen (email: {adz, jo, shj}@es.aau.dk) are with the Department of Electronic Systems, Aalborg University, Denmark. P. Naylor (email: p.naylor@imperial.ac.uk) is with the Electrical and Electronic Engineering Department, Imperial College London, UK, and S. Bech (email: sbe@es.aau.dk) is with Bang & Olufsen, Denmark and the Department of Electronic Systems, Aalborg University, Denmark.

coding. The main result in [5] was that, it is possible to encode two digital sources losslessly and separately with a sum-rate that asymptotically achieves the joint entropy of the two sources given that the decoding is performed jointly. In [4], Wyner and Ziv considered a similar problem for lossy source coding, where one source is to be encoded and sent to the decoder and the other one is available only to the decoder and serves as the side information. The results were later extended to continuous sources in [6]. It was shown that in general, this setup incurs a loss compared to the case where the side information is available at the encoder, but if the sources are Gaussian, no loss occurs.

Since in our setup the observation $\mathbf{y}$ is in general not the same as the desired source $\mathbf{x}$, we refer to it as the remote Wyner-Ziv problem. In this paper, we consider the remote Wyner-Ziv problem with a covariance matrix distortion constraint. We derive a closed-form formula for the rate-distortion function (RDF), and then study the special cases and applications of this result to source coding problems in networks. In order to formally define the problem, we introduce the distortion constraints which are of interest for this work in Section I-B, but before that, we need to define our notation.

A. Notation

We denote matrices and vectors by boldface uppercase and lowercase letters, respectively. We consider zero-mean stationary Gaussian sources, which generate independent vectors $\mathbf{x} \in \mathbb{R}^{n_x}$, $\mathbf{y} \in \mathbb{R}^{n_y}$, and $\mathbf{z} \in \mathbb{R}^{n_z}$, as shown in Fig. 2. The ordered set of vectors $(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N)$ is denoted by $\{\mathbf{x}_i\}_{i=1}^N$ for compactness. We denote the expectation operation by $E[\cdot]$. The covariance and cross-covariance matrices are denoted by the symbol $\mathbf{\Sigma}$ followed by an appropriate subscript. As an example, for random vectors $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$, the conditional cross-covariance of $\mathbf{x}$ and $\mathbf{y}$ given $\mathbf{z}$ and the conditional covariance of $\mathbf{x}$ given $\mathbf{y}$ and $\mathbf{z}$ are denoted by $\mathbf{\Sigma_{xy|z}}$ and $\mathbf{\Sigma_{x|yz}}$, respectively, and are defined as:

$$\mathbf{\Sigma_{xy|z}} \triangleq E\left[(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}|\mathbf{z}})(\mathbf{y}-\boldsymbol{\mu}_{\mathbf{y}|\mathbf{z}})^T|\mathbf{z}\right],$$
$$\mathbf{\Sigma_{x|yz}} \triangleq E\left[(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y},\mathbf{z}})(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y},\mathbf{z}})^T|\mathbf{y},\mathbf{z}\right],$$

where the superscript ^{T} denotes the transpose of a matrix, and $\boldsymbol{\mu}$ denotes the mean. For instance, $\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y},\mathbf{z}} = E[\mathbf{x}|\mathbf{y},\mathbf{z}]$ is the conditional mean of $\mathbf{x}$ given $\mathbf{y}$ and $\mathbf{z}$.

Markov chains are denoted by two-headed arrows as in $\mathbf{x} \leftrightarrow \mathbf{y} \leftrightarrow \mathbf{u}$, which means that if $\mathbf{y}$ is given, then $\mathbf{x}$ and $\mathbf{u}$ are independent. The information theoretic operations of differential entropy and mutual information are denoted by $h(\cdot)$ and $I(\cdot;\cdot)$, respectively. The trace operation is denoted by $\mathrm{tr}(\cdot)$. The set of symmetric positive-definite matrices is denoted by $\mathbb{S}^+$. The statement $\mathbf{A} \succeq \mathbf{B}$ ($\mathbf{A} \succ \mathbf{B}$) means that $\mathbf{A}-\mathbf{B}$ is positive semidefinite (definite). The $n \times n$ identity matrix is denoted by $\mathbf{I}_n$. We use $\mathrm{diag}\{\lambda_i, i = 1, \ldots, n\}$ to show an $n \times n$ diagonal matrix having the elements $\lambda_1, \ldots, \lambda_n$ on its main diagonal. We also make use of the following notations for briefness:

$$(a)^+ \triangleq \max(a, 1), \quad (a)^- \triangleq \min(a, 1).$$

B. Distortion Constraints

Consider the remote Wyner-Ziv problem where the decoder makes an estimation of the desired source $\mathbf{x} \in \mathbb{R}^{n_x}$ using the received message $\mathbf{w} \in \mathbb{R}^{n_w}$ and the side information $\mathbf{z} \in \mathbb{R}^{n_z}$. For the special case where the sources are scalar, the well-known mean-squared error distortion constraint is in the form of $\sigma^2_{x|wz} \leq D$, where D is a given target distortion, and $\sigma^2_{x|wz}$ is the variance of the reconstruction error. With this distortion constraint, the scalar Gaussian remote Wyner-Ziv problem was treated in [7]–[9], and the rate-distortion function was derived in a rather simple form.

For the vector Gaussian case, there are several options to define a quadratic distortion constraint. The most obvious one is the mean-squared error distortion constraint, which is the generalization of the above-mentioned distortion constraint for the case of vector sources. It is defined as follows:

$$\mathrm{tr}\left(\mathbf{\Sigma_{x|wz}}\right) \leq n_x D,$$

where D is a given target distortion. Another constraint, which is related to a relay network problem as discussed in Section III-B, is a mutual information constraint defined as:

$$I(\mathbf{x};\mathbf{w}|\mathbf{z}) \geq R_I$$

where R_I is a given constant. For both the mean-squared error distortion and mutual information constraints, the vector Gaussian problem was solved in [10], and turned out to have a parametric form, resembling the familiar water-filling solutions for some other Gaussian rate-distortion problems.

The last constraint we discuss is the covariance matrix distortion constraint, which is defined as:

$$\mathbf{\Sigma_{x|wz}} \preceq \mathbf{D}, \tag{1}$$

where $\mathbf{D}$ is a given positive-definite target distortion matrix. For the vector Gaussian Wyner-Ziv problem with the covariance matrix distortion constraint, the rate-distortion function can be derived using standard techniques for a limited set of distortion matrices (e.g. $\mathbf{D} \preceq \mathbf{\Sigma_{x|z}}$) [11]. However, the general case does not appear to be manageable to solve using standard techniques, and no closed-form statement is available for the general RDF in the literature. In this paper, we treat this problem and study some of its applications. We elaborate on some aspects of this problem and our motivation in the next subsection.

C. Motivation

During the recent years, there has been a shift from the traditional mean-squared error distortion constraint to covariance matrix distortions in the area of multiterminal source coding [12]–[17].

In [12], a multiple description scenario is considered where L encoders send their descriptions of a vector Gaussian source

to L individual and one joint decoders. Each decoder makes an estimation of the source within a given covariance matrix distortion constraint. The sum-rate of the multiple descriptions problem is specified in [12] under certain assumptions on the distortion matrices.

The Gaussian one-helper problem with the covariance matrix distortion constraint was studied in [13], [14]. Here in addition to a main encoder which observes the direct source and sends a message to the decoder with rate R_1, there is a helper with noisy observations, which sends a message with rate R_2 to the decoder. The decoder estimates the source using the received messages. The rate regions for this problem for the scalar and general helpers were derived in [13] and [14], respectively.

In addition to the above works, there are also several results which bound the rate-distortion regions of source coding problems under the covariance matrix distortion constraint [15]–[17]. In the so-called CEO problem [18], a number of *agents* encode and send their observations of a source with certain rates to a *CEO*. The CEO makes an estimation of the source using the received messages from the agents. The rate region for the vector Gaussian CEO problem under the covariance matrix distortion constraint was outer-bounded in [15], [16]. Finally, in [17], a set of L noisy observations of K correlated scalar Gaussian sources are encoded separately and sent with the rates $R_1, \cdots, R_L$ to a joint decoder. The decoder makes an estimation of the sources using the received messages. The rate-distortion regions for this problem are inner- and outer-bounded for three different distortion constraints including the covariance matrix distortions.

Although the vector Gaussian remote Wyner-Ziv problem has a rather simple setup, we could not find any closed-form solution for the general problem under the covariance matrix distortion constraint in the literature. Neither does such a closed-form solution follow as a special case of the above-mentioned results.

In general, with the covariance matrix distortion constraint, the matrix form of the target distortion gives rise to new issues compared to the scalar target distortions, which make the problem harder to solve. For the Wyner-Ziv problem, the covariance matrix of the unknown part of the source at the decoder is $\mathbf{\Sigma_{x|z}}$. With a scalar target distortion d, the following two cases might arise:

- $\text{tr}(\mathbf{\Sigma_{x|z}}) \leq n_x d$, for which the required rate is zero.
- $\text{tr}(\mathbf{\Sigma_{x|z}}) > n_x d$, which means that some nonzero rate has to be spent. In this case, with an optimal coding scheme, the variance of the reconstruction error at the decoder would be the same as the target distortion.

For a matrix target distortion $\mathbf{D}$ on the other hand, it might happen that none of the two cases $\mathbf{\Sigma_{x|z}} \preceq \mathbf{D}$ and $\mathbf{\Sigma_{x|z}} \succeq \mathbf{D}$ hold. In general, there are two sets of directions in $\mathbb{R}^{n_x}$. In one set of directions, the distortion constraint is already satisfied at the decoder using the side information $\mathbf{z}$. In the other set of directions, the distortion constraint cannot be satisfied without some help from the encoder, making the minimum required rate a nonzero value. In such cases, the covariance matrix of the reconstruction error at the decoder is not guaranteed to be equal to $\mathbf{D}$.

The argument for considering the covariance matrix distortion constraints despite the above-mentioned issues is the generality of the results. For the remote Wyner-Ziv problem, the resulting rate-distortion function $R(\mathbf{D})$ would be a mapping from $\mathcal{D} = \left\{\mathbf{D} \in \mathbb{S}^+ | \, \mathbf{D} \succ \mathbf{\Sigma_{x|yz}}\right\}$ to $\mathbb{R}$. For a given problem with a constraint on $\mathbf{D}$, there is in general a subset of $\mathcal{D}$ for which the constraint is satisfied. We can then choose one member of this subset which minimizes the rate. As an example (among other examples that will be studied in this paper), we will show later on, that the ubiquitous case of the RDF with the mean-squared error distortion is simply equivalent to $R(\mathbf{D})$ at a particular distortion matrix $\mathbf{D}^*$ in $\mathcal{D}$. This particular $\mathbf{D}^*$ is the one that minimizes the rate $R(\mathbf{D})$ subject to a mean-squared error constraint on $\mathbf{D}$.

In addition to generality, the covariance matrix distortion constraint has other advantages. Defining the distortion with respect to the covariance matrix of the reconstruction error allows for formulating new problems. In many sensor network setups, there are constraints on the sum-rate or weighted sum-rate of the network. As we will show in Section III, this can be formulated as a constraint on the determinant of the distortion matrices. One can then optimize a function (such as SNR) over the network subject to this constraint in order to optimally allocate the rates and distortions to the sensors.

D. Overview of the Paper

Section II is dedicated to the vector Gaussian remote Wyner-Ziv problem under covariance matrix distortion constraints. We start with defining a minimum for a pair of symmetric positive-definite matrices based on the joint diagonalization of the matrices. We then derive some properties for this notion of minimum of two matrices, based on which we find an explicit formula for $R(\mathbf{D})$. This minimum of two matrices seems to be natural to our source coding problem, and in addition to the RDF itself, it appears also in the coding schemes and the reconstruction error at the decoder. To derive the RDF, we first lower-bound $R(\mathbf{D})$ using the properties of the minimum of two matrices and information-theoretic arguments. Next we upper-bound it by suggesting a linear coding scheme. The lower and upper bounds coincide, thus yielding the RDF.

In Section III, we present examples of applications and special cases of the results obtained in Section II. We consider three applications. First, we consider a similar source coding problem under the mean-squared error distortion constraint, and will show that the resulting RDF is a special case of $R(\mathbf{D})$ for a certain choice of the distortion matrix $\mathbf{D}$.

Next, we consider a relay network scenario where in addition to the data transmitted to the center from a main transmitter, there is a relay which would like to help the main transmitter by sending its own observation to the center with a certain rate. The problem could be formulated as a source coding problem giving a rate-information function [10]. We will show that similar to the previous case, the resulting rate-information function will be given by $R(\mathbf{D})$ for a certain choice of the distortion matrix $\mathbf{D}$. We then show how one could implement this scenario using the results from Section II.

Finally, as the third and most elaborated example of the applications, we consider a sensor network with a centralized topology. The observation from each sensor is encoded and transmitted to the fusion center with a certain rate. We assume that the the weighted sum-rate of the network is limited to a given amount. At the fusion center one would like to fuse the received data in a manner that the output experiences no linear distortion. We will show that under the given weighted sum-rate constraint, one could design and allocate the distortion matrices to the sensor nodes in order to maximize the output signal to noise ratio at the fusion center.

We conclude the paper in Section IV by discussing the main results and possibilities for future work.

II. Vector Gaussian Remote Wyner-Ziv Problem

In this section, we solve the source coding problem in networks which was simplified to the block diagram in Fig. 2. Later in Section III, we will show how we can use the results for optimizing certain functions in networks. We start with a formal definition of the source coding problem that was introduced in Section I. Assume that $\{\mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i\}_{i=1}^N$ is a sequence of i.i.d. zero-mean random vectors such that $\mathbf{x}_i \in \mathbb{R}^{n_x}$, $\mathbf{y}_i \in \mathbb{R}^{n_y}$, and $\mathbf{z}_i \in \mathbb{R}^{n_z}$ are jointly Gaussian for $i = 1, \cdots, N$. The encoder observes $\{\mathbf{y}_i\}_{i=1}^N$, and using an encoding function:

$$\phi^{(N)} : \mathbb{R}^{n_y N} \to \{1, \cdots, M^{(N)}\},$$

sends a message to the decoder. The decoder observes $\{\mathbf{z}_i\}_{i=1}^N$ and receives the message from the encoder, based on which makes an estimation of $\{\mathbf{x}_i\}_{i=1}^N$ using a decoding function:

$$\psi^{(N)} : \{1, \cdots, M^{(N)}\} \times \mathbb{R}^{n_z N} \to \mathbb{R}^{n_x N}.$$

Definition 1. *The rate-distortion pair $(R, \mathbf{D})$ is achievable if there exists a block length N and the encoding and decoding functions $\phi^{(N)}$ and $\psi^{(N)}$, such that for $\{\hat{\mathbf{x}}_i\}_{i=1}^N = \psi^{(N)}\left(\phi^{(N)}(\{\mathbf{y}_i\}_{i=1}^N), \{\mathbf{z}_i\}_{i=1}^N\right)$:*

$$R \geq \frac{1}{N} \log M^{(N)},$$
$$\mathbf{D} \succeq \frac{1}{N} \sum_{i=1}^{N} E\left[(\mathbf{x}_i - \hat{\mathbf{x}}_i)(\mathbf{x}_i - \hat{\mathbf{x}}_i)^T\right].$$

Definition 2. *The rate-distortion region for the above problem is the closure of the set of achievable rate-distortion pairs $(R, \mathbf{D})$.*

Definition 3. *Let $\mathcal{R}$ be the rate-distortion region for the above problem. The RDF $R(\mathbf{D})$ is then defined as:*

$$R(\mathbf{D}) = \inf_{(R,\mathbf{D}) \in \mathcal{R}} R.$$

Notice that $\mathbf{\Sigma_{x|yz}}$ is the covariance matrix of the reconstruction error at the decoder when in addition to the side information $\mathbf{z}$, the uncoded observation $\mathbf{y}$ of the encoder is also available at the decoder. Since this is not achievable with finite rates, we must have $\mathbf{\Sigma_{x|yz}} \prec \mathbf{D}$, otherwise the rate would become infinite. We thus assume that it holds.

Suppose that the auxiliary random variable $\mathbf{u}$ satisfies the Markov chain $\mathbf{u} \leftrightarrow \mathbf{y} \leftrightarrow (\mathbf{x}, \mathbf{z})$. From the results in [6], it follows that the operational RDF defined above is equal to the following information RDF:

$$R(\mathbf{D}) = \min_{\mathbf{u}} I(\mathbf{y}; \mathbf{u}|\mathbf{z}) \quad \text{s.t.} \quad \begin{cases} \mathbf{\Sigma_{x|uz}} \preceq \mathbf{D}, \\ \mathbf{u} \leftrightarrow \mathbf{y} \leftrightarrow (\mathbf{x}, \mathbf{z}). \end{cases} \tag{2}$$

We write $\mathbf{x}$ in terms of its linear estimation from $\mathbf{y}$ and $\mathbf{z}$ and the estimation error $\mathbf{n}$ as follows:

$$\mathbf{x} = \mathbf{A}\mathbf{y} + \mathbf{B}\mathbf{z} + \mathbf{n}, \tag{3}$$

where $\mathbf{n}$ is independent of $\mathbf{y}$ and $\mathbf{z}$, and $\mathbf{\Sigma_n} = \mathbf{\Sigma_{x|yz}}$. See Appendix A for a quick derivation of the matrices $\mathbf{A}$ and $\mathbf{B}$.

Defining $\mathbf{y}'$ as:

$$\mathbf{y}' = \mathbf{A}\mathbf{y}, \tag{4}$$

and using (3) and the Markov chain in (2), one can write:

$$\mathbf{\Sigma_{x|z}} = \mathbf{\Sigma_{y'|z}} + \mathbf{\Sigma_{x|yz}}, \tag{5}$$
$$\mathbf{\Sigma_{x|uz}} = \mathbf{\Sigma_{y'|uz}} + \mathbf{\Sigma_{x|yz}}. \tag{6}$$

We solve Problem (2) in the rest of this section. To do so, we first introduce a notion of minimum for two positive-definite matrices in the next subsection.

A. Minimum of Two Matrices

Based on the joint diagonalization of two matrices, we define a minimum for two positive-definite matrices and derive some important properties of this definition. These properties will be crucial in the derivation of the RDF in the next two subsections. Although any joint diagonalization would work for this matter, we focus on a particular case which is defined in the sequel.

Consider the ordered pair $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2)$ of two $n \times n$ symmetric positive-definite matrices. Denote the eigenvalue decomposition of $\mathbf{\Sigma}_1$ by:

$$\mathbf{\Sigma}_1 = \mathbf{U}^T \mathbf{\Lambda} \mathbf{U},$$

where $\mathbf{\Lambda} = \text{diag}\{\lambda_i, i = 1, \ldots, n\}$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Consider the principal square-root $\mathbf{\Sigma}_1^{\frac{1}{2}} = \mathbf{U}^T \mathbf{\Lambda}^{\frac{1}{2}} \mathbf{U}$ of $\mathbf{\Sigma}_1$ and denote the eigenvalue decomposition of $\mathbf{\Sigma}_1^{-\frac{1}{2}} \mathbf{\Sigma}_2 \mathbf{\Sigma}_1^{-\frac{1}{2}}$ by:

$$\mathbf{\Sigma}_1^{-\frac{1}{2}} \mathbf{\Sigma}_2 \mathbf{\Sigma}_1^{-\frac{1}{2}} = \mathbf{W}^T \mathbf{\Gamma} \mathbf{W},$$

such that $\mathbf{\Gamma} = \text{diag}\{\gamma_i, i = 1, \ldots, n\}$ and $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_n$. We then define the joint diagonalizer of $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2)$ as follows.

Definition 4. *The joint diagonalizer $\mathbf{V}$ of the ordered pair $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2)$ is defined as:*

$$\mathbf{V} = \mathbf{\Lambda}^{\frac{1}{2}} \mathbf{W} \mathbf{\Sigma}_1^{-\frac{1}{2}}.$$

One can verify that:

$$\mathbf{V} \mathbf{\Sigma}_1 \mathbf{V}^T = \mathbf{\Lambda},$$

$$\mathbf{V}\boldsymbol{\Sigma}_2\mathbf{V}^T = \boldsymbol{\Lambda}',$$

where the diagonal matrix $\boldsymbol{\Lambda}' = \text{diag}\{\lambda_i', i = 1, \ldots, n\}$ is defined as $\boldsymbol{\Lambda}' = \boldsymbol{\Lambda}\boldsymbol{\Gamma}$.

This variant of joint diagonalization is based on another form of diagonalization found in [19, Theorem 8.3.1]. Based on the above joint diagonalization, we define the minimum of $(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$ as follows:

Definition 5. *The minimum of the pair $(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$ of symmetric positive-definite matrices with the joint diagonalizer $\mathbf{V}$ is defined as:*

$$\min(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = \mathbf{V}^{-1}\text{diag}\{\min(\lambda_i, \lambda_i'), i = 1, \ldots, n\}\mathbf{V}^{-T}.$$

Some properties of the above definitions are summarized in the following lemmas.

Lemma 1. *For the pair $(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$ of symmetric positive-definite matrices with the joint diagonalizer $\mathbf{V}$ the following properties hold:*

1) $|\mathbf{V}| = 1$.
2) $\lambda_1' \geq \lambda_2' \geq \cdots \geq \lambda_n'$.
3) $\min(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) \preceq \boldsymbol{\Sigma}_1$, *and* $\min(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) \preceq \boldsymbol{\Sigma}_2$.
4) $\min(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = \boldsymbol{\Sigma}_1$ *if and only if* $\boldsymbol{\Sigma}_1 \preceq \boldsymbol{\Sigma}_2$, *and* $\min(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = \boldsymbol{\Sigma}_2$ *if and only if* $\boldsymbol{\Sigma}_2 \preceq \boldsymbol{\Sigma}_1$.

Proof: The proof follows immediately from the definitions and is thus left out. ∎

The proof of the following lemma follows from the results in [20].

Lemma 2. *Consider the following optimization problem:*

$$a^* = \max_{\boldsymbol{\Sigma}\in\mathbb{S}^+} |\boldsymbol{\Sigma}| \quad \textit{s.t.} \quad \begin{cases} \boldsymbol{\Sigma} \preceq \boldsymbol{\Sigma}_1, \\ \boldsymbol{\Sigma} \preceq \boldsymbol{\Sigma}_2, \end{cases}$$

where $\boldsymbol{\Sigma}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2 \in \mathbb{S}^+$. We then have $a^* = |\min(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)|$.

B. Rate Distortion Function

The following theorem is the main result of this section.

Theorem 1. *The rate-distortion function for problem (2) for $\mathbf{D} \succ \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}}$ is given by:*

$$R(\mathbf{D}) = \frac{1}{2}\log\frac{\left|\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}}\right|}{\left|\min(\mathbf{D} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}}, \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}})\right|}. \quad (7)$$

Proof: We prove the theorem by deriving coinciding upper and lower bounds on the RDF in Sections II-C and II-D, respectively. ∎

C. Lower Bound on $R(\mathbf{D})$

Let $\mathbf{V}$ be the joint diagonalizer of $(\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}}, \mathbf{D} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}})$ as given in Definition 4, such that: [1]

$$\mathbf{V}(\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}})\mathbf{V}^T = \boldsymbol{\Lambda}, \quad (8)$$

[1]We assume that $\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}}$ is of full rank. The rank-deficient case can be converted to an equivalent problem with full-rank matrices using an appropriate transformation. See e.g. Appendix A in [13].

$$\mathbf{V}(\mathbf{D} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}})\mathbf{V}^T = \boldsymbol{\Lambda}'. \quad (9)$$

Define the function $\tilde{R}(\mathbf{D})$ as:

$$\tilde{R}(\mathbf{D}) = \frac{1}{2}\log\frac{\left|\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}}\right|}{\left|\min(\mathbf{D} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}}, \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}})\right|}. \quad (10)$$

The following lemma is then the main result of this subsection.

Lemma 3. *The RDF $R(\mathbf{D})$ defined in (2) is lower-bounded as $R(\mathbf{D}) \geq \tilde{R}(\mathbf{D})$.*

Proof: See Appendix B. ∎

D. Upper Bound on $R(\mathbf{D})$

In this subsection, we upper bound $R(\mathbf{D})$ with the same function $\tilde{R}(\mathbf{D})$ which appeared in the lower bound in the previous subsection. This thus in combination with the lower-bound, gives the RDF in a closed form. It is important to note though that the knowledge of the RDF does not necessarily specify how one could encode the observations to achieve $R(\mathbf{D})$. The upper bound derived in this subsection is constructive, in the sense that it is based on an achievable coding scheme. Such a scheme should suggest a way to transform $\mathbf{y}$ into another variable $\mathbf{u}$, such that for a given distortion $\mathbf{D}$:

- The required rate for delivering $\mathbf{u}$ to the decoder is no higher than $R(\mathbf{D})$.
- The decoder could make an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$ using $\mathbf{u}$ and $\mathbf{z}$, such that the distortion constraint (1) is satisfied.

The role of $\mathbf{u}$ here is to model the quantization effect on $\mathbf{y}$ (or a transformation of $\mathbf{y}$). It is thus typically composed of a linear transformation of $\mathbf{y}$ and an additive noise term that represents the quantization noise. We suggest such a scheme in the sequel. Suppose that $\mathbf{V}$, $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \cdots, \lambda_n\}$, and $\boldsymbol{\Lambda}' = \text{diag}\{\lambda_1', \cdots, \lambda_n'\}$ are defined as in (8) and (9). Suppose also that $\mathbf{U}$ is the orthogonal matrix of the eigenvectors of $\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}}$ giving the eigenmatrix $\boldsymbol{\Lambda}$. The result of this subsection is then given by the following lemma.

Lemma 4. *The following Gaussian test channel achieves $\tilde{R}(\mathbf{D})$ given by (10):*

$$\mathbf{u}^* = \mathbf{U}\mathbf{A}\mathbf{y} + \boldsymbol{\nu}, \quad (11)$$

where $\mathbf{A}$ is defined in (3), and the the coding noise $\boldsymbol{\nu}$ is independent of $\mathbf{y}$ with the covariance matrix given by:

$$\boldsymbol{\Sigma}_\nu = \mathbf{U}\mathbf{V}^{-1}\text{diag}\left\{\frac{\lambda_i \min(\lambda_i, \lambda_i')}{\lambda_i - \min(\lambda_i, \lambda_i')}, i = 1, \ldots, n_x\right\}\mathbf{V}^{-T}\mathbf{U}^T. \quad (12)$$

Moreover, the covariance matrix of the reconstruction error at the optimal decoder is given by:

$$\boldsymbol{\Sigma}_\mathbf{e} = \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{u}^*\mathbf{z}} = \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}} + \min\left(\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}}, \mathbf{D} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}}\right). \quad (13)$$

Proof: See Appendix C. ∎

Note that the coding scheme given by (11) involves two steps:

- Linear transformation of the observation $\mathbf{y}$ into $\mathbf{UAy}$,
- Quantization of the resulting sequence such that the quantization noise $\boldsymbol{\nu}$ becomes Gaussian and independent of $\mathbf{y}$ with the covariance matrix given by (12).

The second step can be performed using a high-dimensional dithered vector quantizer. The dithered quantizer guarantees that $\boldsymbol{\nu}$ will be independent of $\mathbf{y}$, and the high dimension allows for the Gaussianity of $\boldsymbol{\nu}$ [21].

Finally, notice that an immediate result of the above lemma is that $R(\mathbf{D}) \leq \tilde{R}(\mathbf{D})$, which establishes the desired upper bound on the RDF.

Remark: The first and second terms at the right-hand side of (13) are related to the error due to the remoteness of the source and the coding artifacts, respectively. This separability resembles the Wolf and Ziv's result in [22], where it is shown that for a point-to-point remote joint source-channel coding problem with mean-squared error distortions, the end-to-end distortion can be decomposed as the sum of two terms. One term is due to the remoteness of the source and the other one is due to coding. Note however, that it was shown in [23], that this result is not optimal for the multiterminal case.

III. Applications and Special Cases

A. Mean-Squared Error Constraint

For the vector Gaussian remote Wyner-Ziv problem with a scalar mean-squared error constraint, the rate-distortion function is given by [10]:

$$R(D)=\min_{\mathbf{u}} I(\mathbf{y};\mathbf{u}|\mathbf{z}) \quad \text{s.t.} \quad \text{tr}\left(\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{uz}}\right) \leq n_x D, \mathbf{u}\leftrightarrow\mathbf{y}\leftrightarrow(\mathbf{x},\mathbf{z}).$$

It was shown in [10] that $R(D)$ is given by:

$$R(D) = \frac{1}{2}\sum_{i=1}^{n_x} \log\left(\frac{\lambda_i}{\lambda}\right)^+,$$

where $\lambda_i; i=1,2,\ldots,n_x$ are the eigenvalues of $\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}}-\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}}$, and λ is a constant (water level) satisfying the following constraint:

$$\sum_{i=1}^{n_x} \min\left(\lambda,\lambda_i\right) = n_x D - \text{tr}\left(\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}}\right). \tag{14}$$

We will show that this is equivalent to a special case of $R(\mathbf{D})$ for a particular choice of the distortion matrix $\mathbf{D}$. Denote the eigenvalue decomposition of $\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}}$ by $\mathbf{U}^T\text{diag}\{\lambda_i, i=1,\ldots,n_x\}\mathbf{U}$, and define:

$$\mathbf{D}^* = \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}} + \mathbf{U}^T\text{diag}\{\min(\lambda,\lambda_i), i=1,\ldots,n_x\}\mathbf{U},$$

where λ is defined by (14). It is clear that $\mathbf{D}^* \succ \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}}$ and is thus a valid distortion matrix. Substituting $\mathbf{D}^*$ in (7) yields:

$$\begin{aligned} R(\mathbf{D}^*) &= \frac{1}{2}\sum_{i=1}^{n_x}\log\left(\frac{\lambda_i}{\min\left(\lambda,\lambda_i\right)}\right) \\ &= \frac{1}{2}\sum_{i=1}^{n_x}\log\left(\frac{\lambda_i}{\lambda}\right)^+ = R(D). \end{aligned}$$

B. Relay Networks

Consider a relay network where a main transmitter uses n_x antennas to transmit the Gaussian signal $\mathbf{x}$ to the end receiver which has its own noisy observation $\mathbf{z}$ of $\mathbf{x}$ using n_z antennas. There is also a relay that makes the noisy observation $\mathbf{y}$ using its n_y antennas and transmits to the end receiver with a given maximum rate. It is desired to find the minimum rate at which the system can reliably transmit. Assuming that the statistics and channel state information for $\mathbf{x}$ are known to the relay and receiver, it was shown in [10] that this problem is equivalent to the following optimization problem:

$$R(R_I) = \min_{\mathbf{u}} I\left(\mathbf{y};\mathbf{u}|\mathbf{z}\right) \quad \text{s.t.} \quad I(\mathbf{x};\mathbf{u}|\mathbf{z}) \geq R_I, \mathbf{u}\leftrightarrow\mathbf{y}\leftrightarrow(\mathbf{x},\mathbf{z}),$$

for a given R_I, and the resulting rate-mutual information function $R(R_I)$ for $0 \leq R_I \leq \frac{1}{2}\log\left(\frac{|\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}}|}{|\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}}|}\right)$ is given by:

$$R\left(R_I\right) = \frac{1}{2}\sum_{i=1}^{n_x}\log\left(\mu_i\left[\left(\frac{1-\mu_i}{1-\gamma}\right)^- - (1-\mu_i)\right]^{-1}\right), \tag{15}$$

where μ_i, $i=1,\ldots,n_x$ are the eigenvalues of $\mathbf{I}_{n_x} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}}^{-\frac{1}{2}}\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}}\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}}^{-\frac{1}{2}}$, [2] and $\gamma \in [0,1)$ satisfies the following:

$$-\frac{1}{2}\sum_{i=1}^{n_x}\log\left(\frac{1-\mu_i}{1-\gamma}\right)^- = R_I. \tag{16}$$

See [10] and the references therein for more details on the problem setup and formulation. We will first show that $R\left(R_I\right)$ is a special case of $R(\mathbf{D})$ for a particular choice of the distortion matrix $\mathbf{D}$, and then using the results of Section II, we will show how to design a coding scheme that achieves $R(R_I)$.

Suppose that the eigenvalue decomposition of $\mathbf{I}_{n_x} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}}^{-\frac{1}{2}}\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}}\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}}^{-\frac{1}{2}}$ is given by:

$$\mathbf{I}_{n_x} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}}^{-\frac{1}{2}}\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}}\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}}^{-\frac{1}{2}} = \mathbf{W}^T\text{diag}\{\mu_i, i=1,\ldots,n_x\}\mathbf{W}. \tag{17}$$

Define:

$$\mathbf{D}^* = \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}}^{\frac{1}{2}}\mathbf{W}^T\text{diag}\left\{\min\left(1,\frac{1-\mu_i}{1-\gamma}\right), i=1,\ldots,n_x\right\}\mathbf{W}\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}}^{\frac{1}{2}}, \tag{18}$$

where γ is given by (16). From the facts that $\gamma < 1$ and $\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}}^{-\frac{1}{2}}\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}}\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}}^{-\frac{1}{2}} \preceq \mathbf{I}_{n_x}$, and using (17) and (18), it follows that $\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}}^{-\frac{1}{2}}\mathbf{D}^*\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}}^{-\frac{1}{2}} \succ \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}}^{-\frac{1}{2}}\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}}\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}}^{-\frac{1}{2}}$, or equivalently $\mathbf{D}^* \succ \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}}$. Therefore, $\mathbf{D}^*$ is a valid distortion matrix. Substituting $\mathbf{D}^*$ in (7), noting that $\mathbf{D}^* \preceq \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}}$, and using property 4 in Lemma 1, we have:

[2] In [10], μ_i are defined as the nonzero eigenvalues of $\boldsymbol{\Sigma}_{\mathbf{y}|\mathbf{z}}^{1/2}\mathbf{A}^T\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}}^{-1}\mathbf{A}\boldsymbol{\Sigma}_{\mathbf{y}|\mathbf{z}}^{1/2}$. Note however that for any matrix $\mathbf{Q}$ the zero-excluded multispectrum is the same for $\mathbf{QQ}^T$ and $\mathbf{Q}^T\mathbf{Q}$. One can then see that μ_i are also the eigenvalues of $\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}}^{-1/2}\mathbf{A}\boldsymbol{\Sigma}_{\mathbf{y}|\mathbf{z}}\mathbf{A}^T\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}}^{-1/2}$. Using (4) and (5), this simplifies to $\mathbf{I}_{n_x} - \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}}^{-\frac{1}{2}}\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{yz}}\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{z}}^{-\frac{1}{2}}$.

$$
\begin{aligned}
R(\mathbf{D}^*) &= \frac{1}{2}\log\frac{\left|\mathbf{\Sigma_{x|z}}-\mathbf{\Sigma_{x|yz}}\right|}{\left|\min(\mathbf{D}^*-\mathbf{\Sigma_{x|yz}},\mathbf{\Sigma_{x|z}}-\mathbf{\Sigma_{x|yz}})\right|}\\
&= \frac{1}{2}\log\frac{\left|\mathbf{\Sigma_{x|z}}-\mathbf{\Sigma_{x|yz}}\right|}{\left|\mathbf{D}^*-\mathbf{\Sigma_{x|yz}}\right|}\\
&= \frac{1}{2}\log\frac{\left|\mathbf{I}_{n_x}-\mathbf{\Sigma}_{\mathbf{x|z}}^{-\frac{1}{2}}\mathbf{\Sigma_{x|yz}}\mathbf{\Sigma}_{\mathbf{x|z}}^{-\frac{1}{2}}\right|}{\left|\mathbf{\Sigma}_{\mathbf{x|z}}^{-\frac{1}{2}}\mathbf{D}^*\mathbf{\Sigma}_{\mathbf{x|z}}^{-\frac{1}{2}}-\mathbf{\Sigma}_{\mathbf{x|z}}^{-\frac{1}{2}}\mathbf{\Sigma_{x|yz}}\mathbf{\Sigma}_{\mathbf{x|z}}^{-\frac{1}{2}}\right|}\\
&= \frac{1}{2}\sum_{i=1}^{n_x}\log\left(\frac{\mu_i}{1-\mu_i}\frac{1-\gamma}{\gamma}\right)^+ \qquad (19)\\
&= R(R_I). \qquad (20)
\end{aligned}
$$

where (20) follows from some straightforward manipulation of (15) and comparing the result with (19). Recall that for any valid distortion matrix $\mathbf{D}$, the coding scheme (11) achieves $R(\mathbf{D})$. Since $\mathbf{D}^*$ given by (18) is a valid distortion matrix, to implement a coding scheme that yields $R(R_I) = R(\mathbf{D}^*)$, one could set $\mathbf{D} = \mathbf{D}^*$ and use (11).

C. Noise Reduction in Sensor Networks

A simple diagram of a sensor array as well as a centralized network of wireless sensors is shown in Fig. 3. In both cases, there is a central processing unit which makes an estimation of the desired source $\mathbf{x}_d$ using the received signals. For the sensor array, the signals which are fed to the processing unit are impaired versions of the desired source affected by the environment. One can then design the processing unit so as to minimize the noise at the output. For the sensor network, the sensor signals which are already impaired by the environment effects, have to be transmitted to the center with limited rates via wireless links. This incurs extra noise terms which are due to digital transmission (coding noise). Although this extra noise term degrades the performance, its impact can be alleviated by the fact that the coding noise can be controlled by the designer. In other words, in addition to the processing unit, one can make use of the distortion matrices as design parameters in order to minimize the output noise at the processing center. Let us elaborate on this idea in the sequel.

Consider an array of N sensors, where sensor i, $i = 1, \cdots, N$ observes a noisy version $y_i(k)$ of a Gaussian source $x_i(k)$ given by:

$$y_i(k) = x_i(k) + n_i(k), \qquad (21)$$

where k is the time index, and $n_i(k)$ is the additive noise which is assumed to be Gaussian and independent of the source $x_i(k)$ and the noise terms at the other nodes. Define $\mathbf{y}_i \in \mathbb{R}^n$ as $\mathbf{y}_i = [y_i(k-n+1) \ \cdots \ y_i(k)]^T$ (and similarly for $\mathbf{x}_i$ and $\mathbf{n}_i$). One could then rewrite (21) in the vector form as:

$$\mathbf{y}_i = \mathbf{x}_i + \mathbf{n}_i, \qquad (22)$$

where we drop the dependency on k for simplicity of notation. The sources $\mathbf{x}_i; i = 1, \cdots, N$ are related to a desired source $\mathbf{x}_d$ as:

$$\mathbf{x}_i = \mathbf{W}_i^T\,\mathbf{x}_d,\ i = 1, ..., N, \qquad (23)$$

where $\mathbf{W}_i,\ i = 1, \cdots, N$ are $n \times n$ matrices. Defining $\mathbf{W}$ as $\mathbf{W} = [\mathbf{W}_1\ ...\ \mathbf{W}_N]$, we rewrite (23) as:

$$\mathbf{x} = \mathbf{W}^T\mathbf{x}_d.$$

where $\mathbf{x} = [\mathbf{x}_1^T\ ...\ \mathbf{x}_N^T]^T$ (and we also define $\mathbf{y}$ and $\mathbf{n}$ similarly). At the processing unit, as shown in Fig. 3 (a), the output $\mathbf{z}$ is produced using the following filtering operation:

$$\mathbf{z} = \sum_{i=1}^{N}\mathbf{H}_i\mathbf{y}_i = \mathbf{H}\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{H}\mathbf{n},$$

where $\mathbf{H} = [\mathbf{H}_1\ ...\ \mathbf{H}_N]$. The objective is to maximize the output SNR of the filter $\mathbf{H}$ while enforcing only additive distortion to the reconstruction (no linear distortion). A typical example of this scenario with no linear distortion is in microphone arrays, where the linear distortion would cause undesired artefacts in the resulting audio or speech signal [24]. Another example of applications is in Networked Control Systems, where the feedback data is transmitted via a communication network. In certain cases, because of the implementation constraints the source coding scheme used to compress the feedback information should incur no linear distortion [25].

To formulate the problem, notice that the output error of the filter $\mathbf{H}$ is given by:

$$\mathbf{e} = \mathbf{z} - \mathbf{x}_d = \left(\mathbf{H}\mathbf{W}^T - \mathbf{I}\right)\mathbf{x}_d + \mathbf{H}\mathbf{n} = \mathbf{e_x} + \mathbf{e_n},$$

where $\mathbf{e_x}$ is the linear distortion of the signal and $\mathbf{e_n}$ is the additive distortion. To minimize the additive distortion while suppressing the linear distortion, one could choose $\mathbf{H} = \mathbf{H}_{nld}$, where

$$\mathbf{H}_{nld} = \arg\min_{\mathbf{H}} \mathrm{tr}(E[\mathbf{e_n}\mathbf{e_n}^T]) \quad s.t. \quad \mathbf{H}\mathbf{W}^T = \mathbf{I}. \qquad (24)$$

There is a closed form solution for (24) given by [24]:

$$\mathbf{H}_{nld} = \left(\mathbf{W}\mathbf{\Sigma}_{\mathbf{n}}^{-1}\mathbf{W}^T\right)^{-1}\mathbf{W}\mathbf{\Sigma}_{\mathbf{n}}^{-1}, \qquad (25)$$

which gives the following SNR at the output:

$$\mathrm{SNR}(\mathbf{H}_{nld}) = \frac{\mathrm{tr}\left(\mathbf{\Sigma}_{\mathbf{x}_d}\right)}{\mathrm{tr}\left\{\left(\mathbf{W}\mathbf{\Sigma}_{\mathbf{n}}^{-1}\mathbf{W}^T\right)^{-1}\right\}}. \qquad (26)$$

Now suppose that instead of N array sensors, we have a network of N wireless sensors as shown in Fig. 3 (b). Node i encodes its observation $\mathbf{y}_i$ and transmits it with the rate $R(\mathbf{D}_i)$ for a target distortion $\mathbf{D}_i$. This RDF is equivalent to (7) for the special case where $n_x = n_y = n_z = n$, $\mathbf{x} = \mathbf{y}$ and $\mathbf{z} = \mathbf{0}$. The result is given by:

$$R(\mathbf{D}_i) = \frac{1}{2}\log\frac{|\mathbf{\Sigma}_{\mathbf{y}_i}|}{|\min(\mathbf{D}_i, \mathbf{\Sigma}_{\mathbf{y}_i})|}.$$

We further assume that $\mathbf{D}_i \preceq \mathbf{\Sigma}_{\mathbf{y}_i}$. From property 4 in Lemma 1 it then follows that:

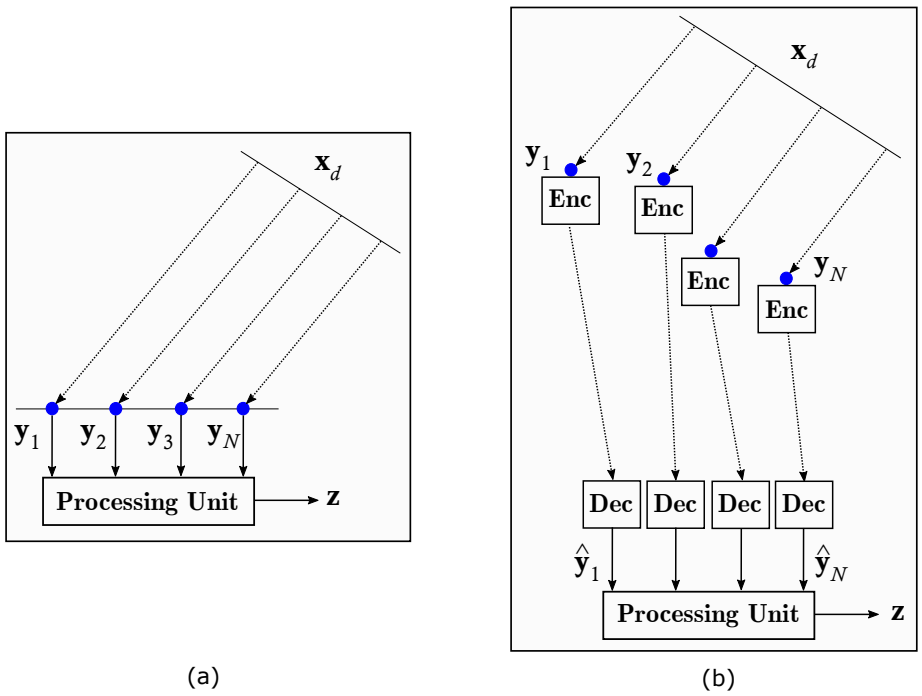


Fig. 3. (a) A sensor array, (b) a centralized wireless sensor network.

$$R(\mathbf{D}_i) = \frac{1}{2}\log\frac{|\mathbf{\Sigma}_{\mathbf{y}_i}|}{|\mathbf{D}_i|}. \tag{27}$$

Note that we have assumed independent decoding of the messages received from the sensors. This is in general sub-optimal. An optimal strategy requires a joint decoding similar to the CEO problem [18], which would be difficult to solve analytically and also to implement.

From the definition of the distortion constraint (1) and the Gaussianity of the achievable coding scheme (see also the proof of Lemma 4), it follows that the reconstruction $\hat{\mathbf{y}}_i'$ at the decoder satisfies $\mathbf{y}_i - \hat{\mathbf{y}}_i' = \boldsymbol{\nu}_i$, with $\boldsymbol{\nu}_i$ being independent of $\hat{\mathbf{y}}_i'$ and $\mathbf{\Sigma}_{\boldsymbol{\nu}_i} = \mathbf{D}_i$. Equivalently, one could write:

$$\hat{\mathbf{y}}_i' = \mathbf{A}_i\mathbf{y}_i + \boldsymbol{\nu}_i',$$

where $\boldsymbol{\nu}_i'$ and $\mathbf{y}_i$ are independent, and:

$$\begin{aligned}\mathbf{A}_i &= (\mathbf{\Sigma}_{\mathbf{y}_i} - \mathbf{D}_i)\mathbf{\Sigma}_{\mathbf{y}_i}^{-1}\\ \mathbf{\Sigma}_{\boldsymbol{\nu}_i'} &= \mathbf{A}_i\mathbf{D}_i.\end{aligned}$$

We apply the invertible map $\mathbf{A}_i^{-1}$ to $\hat{\mathbf{y}}_i'$ to obtain the following equivalent (but preferred) version of the reconstruction:

$$\hat{\mathbf{y}}_i = \mathbf{A}_i^{-1}\hat{\mathbf{y}}_i' = \mathbf{y}_i + \boldsymbol{\nu}_i, \tag{28}$$

with:

$$\mathbf{\Sigma}_{\boldsymbol{\nu}_i} = \left(\mathbf{D}_i^{-1} - \mathbf{\Sigma}_{\mathbf{y}_i}^{-1}\right)^{-1}. \tag{29}$$

Combining (22) and (28) we have:

$$\hat{\mathbf{y}}_i = \mathbf{x}_i + \mathbf{n}_i + \boldsymbol{\nu}_i = \mathbf{x}_i + \mathbf{v}_i, \tag{30}$$

where $\mathbf{v}_i = \mathbf{n}_i + \boldsymbol{\nu}_i$. Recall that the objective is to maximize the SNR at the output of the processing unit using the received noisy signals $\hat{\mathbf{y}}_i$, while incurring no linear distortion to the output signal. Comparing (30) to (22), it is clear that this problem is similar to the case of the sensor array. One can thus apply the same filter as (25) to $\hat{\mathbf{y}}_i; i = 1, \cdots, N$ to maximize the output SNR. To do so, we rewrite (30) as:

$$\hat{\mathbf{y}} = \mathbf{x} + \mathbf{v}, \tag{31}$$

where $\hat{\mathbf{y}} = [\hat{\mathbf{y}}_1^T \ ... \ \hat{\mathbf{y}}_N^T]^T$, and $\mathbf{x}$ and $\mathbf{v}$ are defined similarly. The output SNR similar to (26) is then given by:

$$\mathrm{SNR}(\mathbf{H}_{nld}, \mathbf{D}_1, ..., \mathbf{D}_N) = \frac{\mathrm{tr}\left(\mathbf{\Sigma}_{\mathbf{x}_d}\right)}{\mathrm{tr}\left\{\left[\mathbf{W}\left(\mathbf{\Sigma}_{\mathbf{v}}(\mathbf{D}_1, ..., \mathbf{D}_N)\right)^{-1}\mathbf{W}^T\right]^{-1}\right\}}. \tag{32}$$

Notice that due to the wireless link between the sensors and the center and the limited transmission rate, in addition to the additive noise $\mathbf{n}_i$ in the measurements, there is also a reconstruction error $\boldsymbol{\nu}_i$ due to quantization and coding. Therefore, the covariance matrix of the noise term $\mathbf{v}$ in (31) depends on the set of distortions as emphasized by the notation in (32). While the quantization error is inevitable, one could shape the covariance matrix $\mathbf{\Sigma}_{\mathbf{v}}(\mathbf{D}_1, ..., \mathbf{D}_N)$ in (32) by manipulating the distortions $\mathbf{D}_1, ..., \mathbf{D}_N$ in order to maximize the output SNR.

To be more specific, suppose that it is required that the weighted sum-rate of the network $\sum_{i=1}^N \alpha_i R(\mathbf{D}_i)$ does not exceed a given amount R for a given set of weights $\alpha_1, ..., \alpha_N$. The problem can then be formulated as follows:

$$\max_{\mathbf{D}_1,...,\mathbf{D}_N} \text{SNR}(\mathbf{H}_{nld}, \mathbf{D}_1, ..., \mathbf{D}_N) \quad s.t. \quad \sum_{i=1}^{N} \alpha_i R(\mathbf{D}_i) = R. \tag{33}$$

The weights $\alpha_1, ..., \alpha_N$ could be chosen for example to equalize the power consumption at the sensors, or to give the network sum-rate by being set equal to each other, or to satisfy any other criterion specified by the user. Without loss of generality, we assume that:

$$\sum_{i=1}^{N} \alpha_i = 1. \tag{34}$$

It is worth mentioning that this problem includes the allocation of the rates to the nodes, but is not limited to it, since rate allocation is equivalent to the specification of the determinant of the distortion matrices $\mathbf{D}_i$ according to (27), while in (33) the whole matrices $\mathbf{D}_1, ..., \mathbf{D}_N$ must be chosen optimally to maximize the output SNR. This possibility is a result of having the target distortions in the matrix form.

Problem (33) is equivalent to the following problem:

$$\min_{\mathbf{D}_1,...,\mathbf{D}_N} \text{tr}\left\{ \left[\mathbf{W} \left(\mathbf{\Sigma}_{\mathbf{v}}(\mathbf{D}_1, ..., \mathbf{D}_N) \right)^{-1} \mathbf{W}^T \right]^{-1} \right\}$$
$$s.t. \quad \sum_{i=1}^{N} \alpha_i R(\mathbf{D}_i) = R. \tag{35}$$

Substituting (27) in the constraint in (35), noting that $\mathbf{\Sigma}_{\mathbf{v}_i} = \mathbf{\Sigma}_{\mathbf{n}_i} + \mathbf{\Sigma}_{\boldsymbol{\nu}_i}$, and using (29), we rewrite (35) as:

$$\min_{\mathbf{D}_1,...,\mathbf{D}_N} \text{tr}\left\{ \left(\sum_{i=1}^{N} \mathbf{W}_i \left[\mathbf{\Sigma}_{\mathbf{n}_i} + \left(\mathbf{D}_i^{-1} - \mathbf{\Sigma}_{\mathbf{y}_i}^{-1} \right)^{-1} \right]^{-1} \mathbf{W}_i^T \right)^{-1} \right\}$$
$$s.t. \quad \prod_{i=1}^{N} |\mathbf{D}_i|^{\alpha_i} = \beta, \tag{36}$$

where β is defined as:

$$\beta = e^{-2R} \prod_{i=1}^{N} |\mathbf{\Sigma}_{\mathbf{y}_i}|^{\alpha_i}. \tag{37}$$

For simplicity, we assume that $\mathbf{W}_i$ in (23) are invertible. Notice that even if the relationship between $\mathbf{x}_i$ and $\mathbf{x}_d$ is not invertible, one could find an invertible matrix $\mathbf{W}_i$ which minimizes for example the mean-squared error between $\mathbf{x}_i$ and $\mathbf{W}_i^T \mathbf{x}_d$. In this case, (23) would be an approximation (See [24] for more details). However, the approximation error could be added to the additive noise $\mathbf{n}_i$ in (22). One could thus assume that $\mathbf{y}_i = \mathbf{W}_i^T \mathbf{x}_d + \mathbf{n}_i$ always holds with an invertible $\mathbf{W}_i$.

We define $\mathbf{Z}_i$ and $\mathbf{C}_i$ as:

$$\mathbf{Z}_i = \mathbf{W}_i \mathbf{\Sigma}_{\mathbf{n}_i}^{-1} \left(\mathbf{\Sigma}_{\mathbf{n}_i}^{-1} + \mathbf{D}_i^{-1} - \mathbf{\Sigma}_{\mathbf{y}_i}^{-1} \right)^{-1} \mathbf{\Sigma}_{\mathbf{n}_i}^{-1} \mathbf{W}_i^T, \tag{38}$$
$$\mathbf{C}_i = \mathbf{W}_i \mathbf{\Sigma}_{\mathbf{n}_i}^{-1} \left(\mathbf{\Sigma}_{\mathbf{n}_i}^{-1} - \mathbf{\Sigma}_{\mathbf{y}_i}^{-1} \right)^{-1} \mathbf{\Sigma}_{\mathbf{n}_i}^{-1} \mathbf{W}_i^T. \tag{39}$$

The following set of equations then follow from the KKT conditions for Problem (36) (See Appendix D for the derivation):

$$\alpha_i \lambda \mathbf{A}^2 = \mathbf{Z}_i - \mathbf{Z}_i \mathbf{C}_i^{-1} \mathbf{Z}_i, \tag{40}$$
$$\mathbf{A} = \sum_{i=1}^{N} \mathbf{W}_i \mathbf{\Sigma}_{\mathbf{n}_i}^{-1} \mathbf{W}_i^T - \sum_{i=1}^{N} \mathbf{Z}_i \tag{41}$$
$$\prod_{i=1}^{N} \left| \mathbf{\Sigma}_{\mathbf{n}_i}^{-1} \mathbf{W}_i^T \mathbf{Z}_i^{-1} \mathbf{W}_i \mathbf{\Sigma}_{\mathbf{n}_i}^{-1} - \mathbf{\Sigma}_{\mathbf{n}_i}^{-1} + \mathbf{\Sigma}_{\mathbf{y}_i}^{-1} \right|^{-\alpha_i} = \beta. \tag{42}$$

In general, it is not easy to solve (40)–(42) analytically to find the unknowns $\mathbf{A}$, λ, and $\mathbf{Z}_i; i = 1, \cdots, N$. Instead, one could proceed by using numerical methods. In the rest of this subsection, we consider special cases of this problem where further analysis is possible and leads to interesting results.

1) High-Rate Regime: We assume that (40) could be approximated as:

$$\alpha_i \lambda \mathbf{A}^2 \approx \mathbf{Z}_i. \tag{43}$$

From (38) and (39), this means that the distortion matrices $\mathbf{D}_i$ in (38) have small eigenvalues, or equivalently, $\mathbf{D}_i^{-1}$ has large eigenvalues. This holds, if the rates $R(\mathbf{D}_i)$ are high enough. We will show later, that in fact in order for the results to hold, the rates need not be very high, and depending on the setup, even for relatively low rates the results will be valid. The high-rate assumption can also be interpreted as:

$$\mathbf{D}_i^{-1} + \mathbf{\Sigma}_{\mathbf{n}_i}^{-1} - \mathbf{\Sigma}_{\mathbf{y}_i}^{-1} \approx \mathbf{D}_i^{-1}. \tag{44}$$

We will use this form later to find the optimal value of the parameter λ. Combining (41) and (43), using (34), and defining:

$$\mathbf{S} = \sum_{i=1}^{N} \mathbf{W}_i \mathbf{\Sigma}_{\mathbf{n}_i}^{-1} \mathbf{W}_i^T, \tag{45}$$

yields:

$$\lambda \mathbf{A}^2 + \mathbf{A} - \mathbf{S} \approx \mathbf{0}. \tag{46}$$

From (46), it is easy to see that $\mathbf{A}$ and $\mathbf{S}$ share the same eigenvectors. Denote the eigenvalue decomposition of $\mathbf{S}$ by:

$$\mathbf{S} = \mathbf{U}_s \text{diag}\{s_i, i = 1, \cdots, n\} \mathbf{U}_s^T. \tag{47}$$

One could then write $\mathbf{A}$ as:

$$\mathbf{A} \approx \mathbf{U}_s \text{diag}\{a_i, i = 1, \cdots, n\} \mathbf{U}_s^T. \tag{48}$$

To find $a_i, i = 1, \cdots, n$, we substitute (48) and (47) in (46) to obtain:

$$\lambda a_i^2 + a_i - s_i \approx 0; \;\; i = 1, \cdots, n,$$

which gives the following solution:

$$a_i \approx \frac{\sqrt{1 + 4\lambda s_i} - 1}{2\lambda}; \;\; i = 1, \cdots, n. \tag{49}$$

Note that the only unknown parameter in (49) is λ. To find λ, we use (44) and (38) to rewrite (42) as:

$$\prod_{i=1}^{N} \left|\mathbf{\Sigma}_{\mathbf{n}_i}^{-1}\mathbf{W}_i^T\mathbf{Z}_i^{-1}\mathbf{W}_i\mathbf{\Sigma}_{\mathbf{n}_i}^{-1}\right|^{-\alpha_i} \approx \beta. \tag{50}$$

Defining γ as:

$$\gamma = \beta \prod_{i=1}^{N} \left|\alpha_i \mathbf{\Sigma}_{\mathbf{n}_i}\mathbf{W}_i^{-T}\mathbf{W}_i^{-1}\mathbf{\Sigma}_{\mathbf{n}_i}\right|^{-\alpha_i}, \tag{51}$$

substituting (43) in (50), and simplifying the result yields:

$$\lambda^n \left|\mathbf{A}^2\right| \approx \gamma. \tag{52}$$

Finally, substituting (49) in (52) leads to the following equation for λ:

$$\prod_{i=1}^{N} \left(\frac{\sqrt{1+4\lambda s_i}-1}{\sqrt{4\lambda}}\right)^2 \approx \gamma. \tag{53}$$

To summarize the results, given the parameters $\mathbf{W}_i$, α_i, $\mathbf{\Sigma}_{\mathbf{n}_i}$ for $i = 1, \cdots, n$, one should take the following steps to find the optimal allocation of the distortions $\mathbf{D}_i$ to the nodes:

- Calculate $\mathbf{S}$ using (45), and find $\mathbf{U}_s$ and $s_i; i = 1, \cdots, n$ from (47).
- Calculate γ from (51) and then find λ by solving (53).
- Use (49) and (48) to find $\mathbf{A}$.
- Calculate $\mathbf{Z}_i$ using (43).
- Calculate $\mathbf{D}_i$ using (38).

Note that the whole process depends on finding a solution for (53). Define R_{min} as:

$$R_{min} = \frac{1}{2}\log \frac{\prod_{i=1}^{N} \left|\alpha_i \mathbf{\Sigma}_{\mathbf{n}_i}\mathbf{W}_i^{-T}\mathbf{W}_i^{-1}\mathbf{\Sigma}_{\mathbf{n}_i}\right|^{-\alpha_i} |\mathbf{\Sigma}_{\mathbf{y}_i}|^{\alpha_i}}{|\mathbf{S}|}.$$

We then prove the following proposition:

Proposition 1. *If $R \geq R_{min}$, there is a unique solution to (53), otherwise (53) has no solution.*

Proof: See Appendix E. ∎

We perform a few simulations to further study the behaviour of the output SNR as a function of the distortion matrices $\mathbf{D}_i$. We assume that there are two nodes $N = 2$, and $\mathbf{W}_1 = \mathbf{W}_2 = \mathbf{I}_n$. We consider the following structure for the covariance matrix of both signal and noise:

$$\mathbf{\Sigma}_{(i,j)} = \nu\rho^{|i-j|},$$

where $\mathbf{\Sigma}_{(i,j)}$ is the (i,j)th element of $\mathbf{\Sigma}$. For the signal $\mathbf{x}_d$, we choose $\nu = 1$ and $\rho = 0.9$. For the noise terms, we use ν and ρ as control parameters. The dimension and the weighted sum-rate are kept fixed on $n = 32$ and $R = 2.5n$, respectively, unless otherwise stated.

We check the optimality of the results and the gain that can be achieved. We first specify the parameters α_1, α_2, and the covariance matrix of the noise terms. We find the distortion matrices $\mathbf{D}_i^*; i = 1, 2$ suggested by the procedure described above and their corresponding SNR. We then compare the resulting SNR to the SNR given by other choices of the distortion matrices. Assuming that the eigenvalue decomposition of $\mathbf{D}_i^*$ is given by $\mathbf{D}_i^* = \mathbf{U}_i\mathbf{\Lambda}_i\mathbf{U}_i^T$, we use the following formula to generate other distortion matrices:

$$\mathbf{D}_i = \mathbf{U}_i \left(\beta_i\mathbf{\Lambda}_i + \eta_i\mathbf{\Theta}_i\right)\mathbf{U}_i^T, \tag{54}$$

where β_i and η_i are weighting parameters, and $\mathbf{\Theta}_i$ is a diagonal matrix whose diagonal elements are drawn from a uniform distribution in the interval $[0, 5\iota_i]$, with ι_i being the largest element of $\mathbf{\Lambda}_i$. By choosing β_i close to 1 and η_i close to 0, the resulting distortion matrices $\mathbf{D}_i$ will be slightly perturbed versions of $\mathbf{D}_i^*$. On the other hand, by choosing β_i close to 0 and η_i close to 1, the resulting $\mathbf{D}_i$ will be random. Using (54), we generate L valid distortion pairs $(\mathbf{D}_1, \mathbf{D}_2)$ using the following algorithm:

Initialization: Set $j = 1$.
Iterations: While $j \leq L$ perform the following steps:

1) Generate $\mathbf{D}_1$ using (54)
2) If the result is not a valid distortion matrix, go back to 1, otherwise Generate $\mathbf{D}_2$ using (54).
3) Normalize $\mathbf{D}_2$ such that the constraint in (33) holds. If the result is not a valid distortion matrix, go back to 2.
4) Calculate the SNR given by $(\mathbf{D}_1, \mathbf{D}_2)$.
5) Increase j by 1.

We set $L = 10^3$ for the rest of this subsection. To study the behavior of the output SNR around $(\mathbf{D}_1^*, \mathbf{D}_2^*)$, we set $\beta_1 = \beta_2 = 0.999$ and $\eta_1 = \eta_2 = 0.001$. Figure 4 illustrates the results for different choices of the parameters. As seen from Fig. 4, the theoretical results from the high-rate approximation indeed give the local maximum for the output SNR.

To verify that $(\mathbf{D}_1^*, \mathbf{D}_2^*)$ is the global maximizer of the output SNR and to see the gap between the maximum SNR and the one resulting from a random allocation of the distortion matrices, we set $\beta_1 = \beta_2 = 0$ and $\eta_1 = \eta_2 = 1$. The results are illustrated in Fig. 5. As seen from the plots, the gap between the optimal SNR and the one resulting from a random allocation could in some cases be as large as 5 dB.

To show the scalability in sense of the number of sensors, we repeat the simulation in Fig. 5 (a) with 4 sensors. Two sensors have similar parameters with sensor 1 in Fig. 5 (a), and the other two sensors have similar parameters with sensor 2 in the same figure. The weights are set as $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0.25$. The result is shown in Fig. 6, which shows that for 4 sensors, as expected, the output SNR is higher compared to 2 sensors (Fig. 5 (a)) due to the higher number of the available noisy signals. Moreover, comparison of the figures shows that the gap between the optimal performance and the one resulting from the random allocation of the distortion matrices is similar for the 4 sensors and 2 sensors scenarios.

Next we study the validity of the high-rate assumption. Notice that even with low rates, the procedure described to find the optimal distortion allocation might give a solution. The problem, however is that if the rates are not high enough, the approximation in (43) will not hold, and the resulting distortion matrices might violate the sum-rate constraint in (33). The approximation error is negligible if $\alpha_1 R(\mathbf{D}_1^*) + \alpha_2 R(\mathbf{D}_2^*) \approx R$. The difference between $\alpha_1 R(\mathbf{D}_1^*) + \alpha_2 R(\mathbf{D}_2^*)$

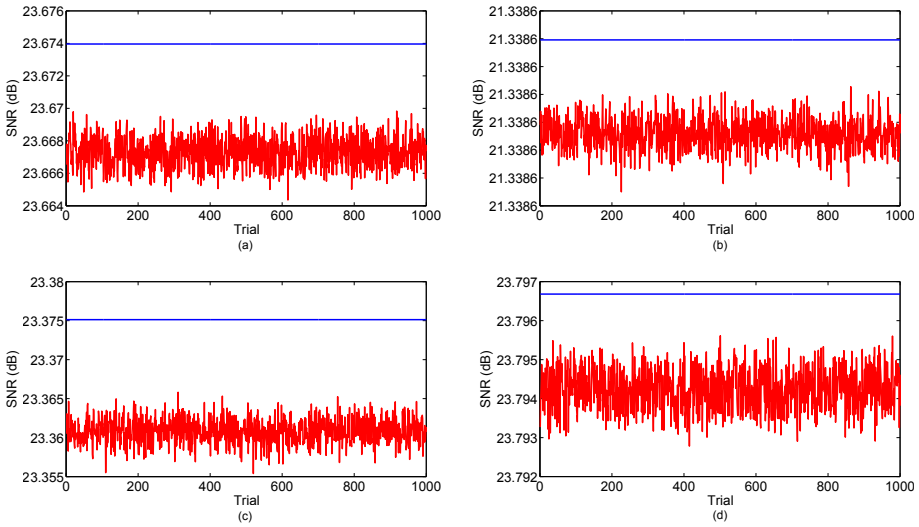


Fig. 4. Output SNR for 1000 trials where the distortion matrices $(\mathbf{D}_1, \mathbf{D}_2)$ are slightly perturbed versions of $(\mathbf{D}_1^*, \mathbf{D}_2^*)$. The blue line indicates the output SNR at $(\mathbf{D}_1^*, \mathbf{D}_2^*)$. (a) with $\alpha_1 = \alpha_2 = 0.5$, and $\rho_1 = 0.9, \rho_2 = 0.3, \nu_1 = 0.01, \nu_2 = 0.02$ for the covariance matrix of the noise terms, (b) $\alpha_1 = \alpha_2 = 0.5$, and $\rho_1 = \rho_2 = 0$, $\nu_1 = 0.01, \nu_2 = 0.02$, (c) $\alpha_1 = 0.7, \alpha_2 = 0.3$, and $\rho_1 = 0.9, \rho_2 = 0.3$, $\nu_1 = 0.01, \nu_2 = 0.02$, (d) $\alpha_1 = 0.3, \alpha_2 = 0.7$, and $\rho_1 = 0.9, \rho_2 = 0.3$, $\nu_1 = 0.01, \nu_2 = 0.02$.

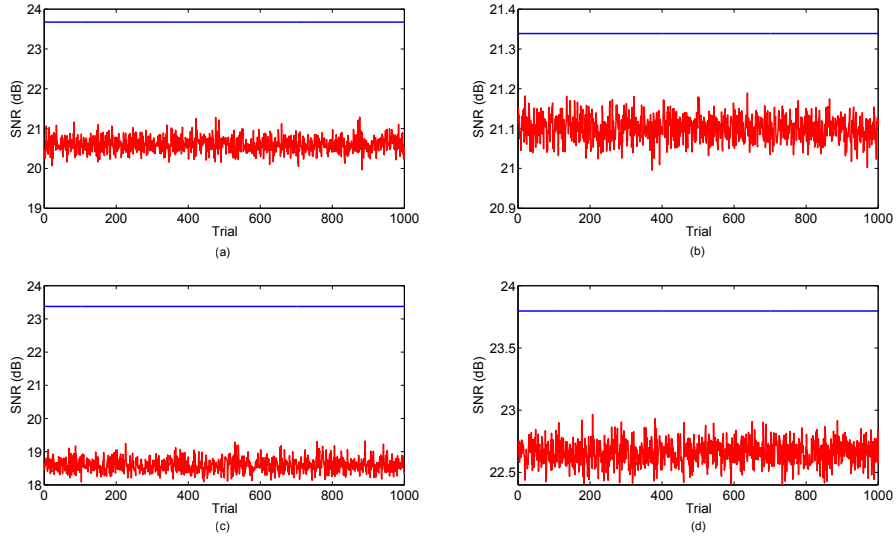


Fig. 5. Output SNR for 1000 trials where the distortion matrices $(\mathbf{D}_1, \mathbf{D}_2)$ are chosen randomly. The blue line indicates the output SNR at $(\mathbf{D}_1^*, \mathbf{D}_2^*)$. (a) with $\alpha_1 = \alpha_2 = 0.5$, and $\rho_1 = 0.9, \rho_2 = 0.3, \nu_1 = 0.01, \nu_2 = 0.02$ for the covariance matrix of the noise terms, (b) $\alpha_1 = \alpha_2 = 0.5$, and $\rho_1 = \rho_2 = 0$, $\nu_1 = 0.01, \nu_2 = 0.02$, (c) $\alpha_1 = 0.7, \alpha_2 = 0.3$, and $\rho_1 = 0.9, \rho_2 = 0.3$, $\nu_1 = 0.01, \nu_2 = 0.02$, (d) $\alpha_1 = 0.3, \alpha_2 = 0.7$, and $\rho_1 = 0.9, \rho_2 = 0.3$, $\nu_1 = 0.01, \nu_2 = 0.02$.

and R as a function of R is plotted in Fig. 7 for $\alpha_1 = \alpha_2 = 0.5$, and $\rho_1 = \rho_2 = 0.8$ and $\nu_1 = 0.1, \nu_2 = 0.2$ for the covariance matrices of the noise terms. Note that for a weighted sum-rate R that is as low as 25 nats per block (equivalent to 0.78 nats per sample or 1.13 bits per sample) the high-rate assumption still holds with a negligible error.

2) Scalar Sources: We assume that $n = 1$, which means that all the sources are scalar. We further assume that there are only two nodes. While the high-rate case studied in the previous subsection is of interest for practical applications, the scalar case with two nodes may be of less practical relevance. However, it leads to interesting analytical results that makes it worth studying.

We denote the scalar version of $\mathbf{D}_i$ and $\mathbf{W}_i$ by D_i and w_i, respectively. The variance of random variables is denoted by Σ followed by a subscript. We assume that we are interested in the sum-rate (so $\alpha_1 = \alpha_2 = 0.5$) and $w_1 = w_2$ to further simplify the problem for analytical derivations. Applying these assumptions to (38)–(42), and defining Σ_1 and Σ_2 as:

$$\Sigma_i = \Sigma_{n_i} - \frac{{\Sigma_{n_i}}^2}{\Sigma_{y_i}}, \quad i = 1, 2,$$

one can show that the target distortions D_1^* and D_2^* are given by:

$$D_1^* = \beta \frac{\Sigma_{n_1}}{\Sigma_{n_2}} \left(\frac{\Sigma_{n_1}\Sigma_{n_2} - \Sigma_2 \beta}{\Sigma_{n_1}\Sigma_{n_2} - \Sigma_1 \beta} \right), \tag{55}$$

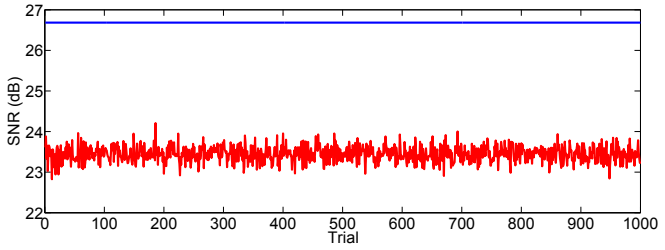


Fig. 6. Output SNR at 1000 trials for a simulation where there are 4 sensors and the distortion matrices $(\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3, \mathbf{D}_4)$ are chosen randomly. The blue line indicates the output SNR at $(\mathbf{D}_1^*, \mathbf{D}_2^*, \mathbf{D}_3^*, \mathbf{D}_4^*)$, which gives the optimal solution. The choice of parameters is similar to the simulation in Fig. 5 (a).

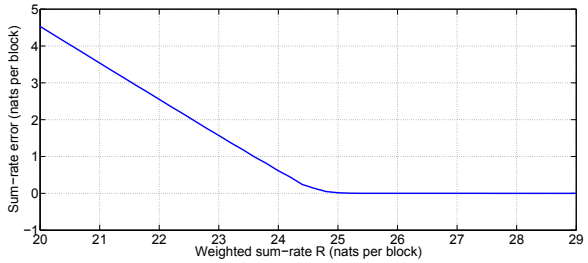


Fig. 7. The difference between the desired weighted sum-rate R and the resulting weighted sum-rate due to the high-rate approximation as a function of R

$$D_2^* = \beta \frac{\Sigma_{n_2}}{\Sigma_{n_1}} \left(\frac{\Sigma_{n_1}\Sigma_{n_2} - \Sigma_1 \beta}{\Sigma_{n_1}\Sigma_{n_2} - \Sigma_2 \beta} \right). \tag{56}$$

Based on this, and defining the following parameters:

$$R_{max} = \frac{1}{4} \log \frac{\left[\max(\Sigma_1, \Sigma_2)\right]^2}{(\Sigma_{n_1} - \Sigma_1)(\Sigma_{n_2} - \Sigma_2)}, \tag{57}$$

$$R_{min} = \frac{1}{4} \log \frac{\left[\min(\Sigma_1, \Sigma_2)\right]^2}{(\Sigma_{n_1} - \Sigma_1)(\Sigma_{n_2} - \Sigma_2)}, \tag{58}$$

we prove the following proposition.

Proposition 2. *The stationary point (D_1^*, D_2^*) given by (55)-(56) is the unique maximizer of the output SNR, if $R > R_{max}$, and is the unique minimizer, if $R < R_{min}$.*

Proof: See Appendix F ■

To observe the behaviour of the output SNR as a function of the distortions, consider the following example. Let $w_1 = w_2 = 1$, $\Sigma_{x_d} = 1$, $\Sigma_{y_1} = 1.2$, $\Sigma_{n_1} = 0.2$, $\Sigma_{y_2} = 1.1$, and $\Sigma_{n_2} = 0.1$. From (57) and (58) we have $R_{max} = 1.13$ and $R_{min} = 0.83$. We make use of the weighted sum-rate constraint $\frac{1}{2}R(D_1) + \frac{1}{2}R(D_2) = R$ to write $\text{SNR}(D_1, D_2)$ as a function of D_1 or D_2 only (denoted by $\text{SNR}(D_1)$ and $\text{SNR}(D_2)$, respectively). We plot $\text{SNR}(D_1)$ and $\text{SNR}(D_2)$ for three different values $0.5, 1, 2$ of the weighted sum-rate R. The result is shown in Fig. 8. As seen in Fig. 8 (a), for $R = 0.5 < R_{min}$ the output SNR is minimized at $(D_1 = D_1^*, D_2 = D_2^*)$, which is not desired. The maximum SNR should then be on the boundary, which means that the whole rate should be given to one of the nodes. It can be seen in the figure that the highest SNR is achieved for the smallest D_2, suggesting that the whole rate should be allocated to node 2. This is in agreement with intuition, since node 2 is less noisy than node 1. For $R_{min} < R < R_{max}$, one could see in Fig. 8 (b) that $(D_1 = D_1^*, D_2 = D_2^*)$ is not a feasible point. Similar to the previous case, the optimal solution should again be on the boundary. Finally, for $R = 2 > R_{max}$, $(D_1 = D_1^*, D_2 = D_2^*)$ maximizes the output SNR.

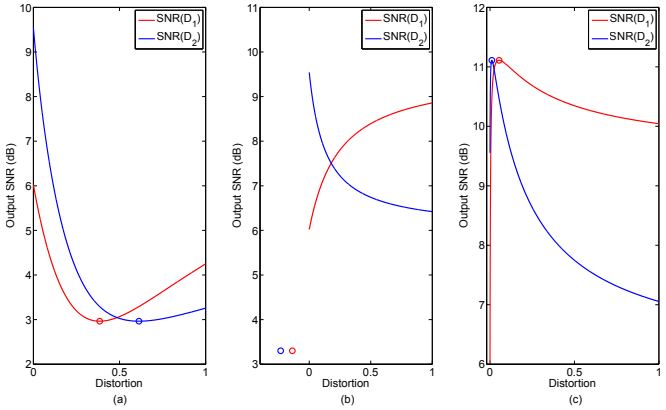


Fig. 8. An example of the output SNR as a function of D_1 and D_2 for the scalar case with two nodes for three different values of the network sum-rate, (a) $R < R_{min}$, (b) $R_{min} < R < R_{max}$, and (c) $R > R_{max}$. The circles show the stationary points suggested by the KKT conditions.

Note that the above example suggests that the distortion allocation has a water-filling form. The critical rate R_{max} acts as the water level. If R is above this level, the rate is split between the nodes based on their noise levels. If R is below the water level, the noisier node is omitted, and the whole rate is given to the other node.

IV. Conclusions and Discussion

We considered a source coding problem in a networked setup under covariance matrix distortion constraints. We modelled the problem as a vector Gaussian remote Wyner-Ziv problem and solved the problem by deriving an explicit formula for the rate-distortion function and designing coding schemes that asymptotically achieve the rate-distortion function. We then studied some applications of the results. In particular, we showed that the rate-distortion function for the equivalent Wyner-Ziv problem with mean-squared error distortions and the rate-information function modelling a relay network source coding problem are special cases of our results. Finally, we considered a centralized sensor network with a weighted sum-rate constraint where each node transmits its observation with a certain rate and distortion, and the received data is fused at the center to maximize the output SNR without enforcing linear distortions. For this problem we bridged between noise reduction and source coding by showing that the distortion matrices and the rates at the individual nodes could be designed to maximize the output SNR at the center. We considered special cases such as the high-rate case or the case of scalar sources in order to obtain analytical results. Further work could be other possible special cases of the noise reduction problem where analytical results can be obtained.

APPENDIX A

From the fact that $\mathbf{n}$ is uncorrelated with $\mathbf{y}$ and $\mathbf{z}$ and:

$$\mathbf{x} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \end{pmatrix} \begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix} + \mathbf{n}, \tag{59}$$

it follows that:

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \end{pmatrix} = \begin{pmatrix} \mathbf{\Sigma_{xy}} & \mathbf{\Sigma_{xz}} \end{pmatrix} \begin{pmatrix} \mathbf{\Sigma_y} & \mathbf{\Sigma_{yz}} \\ \mathbf{\Sigma_{zy}} & \mathbf{\Sigma_z} \end{pmatrix}^{-1}$$
$$= \begin{pmatrix} \mathbf{\Sigma_{xy}}\mathbf{\Delta_1^{-1}} - \mathbf{\Sigma_{xz}}\mathbf{\Delta_2^{-1}}\mathbf{\Sigma_{zy}}\mathbf{\Sigma_y^{-1}} & \mathbf{\Sigma_{xz}}\mathbf{\Delta_2^{-1}} - \mathbf{\Sigma_{xy}}\mathbf{\Delta_1^{-1}}\mathbf{\Sigma_{yz}}\mathbf{\Sigma_z^{-1}} \end{pmatrix},$$

where $\mathbf{\Delta}_1$ and $\mathbf{\Delta}_2$ are defined as:

$$\mathbf{\Delta}_1 = \mathbf{\Sigma_y} - \mathbf{\Sigma_{yz}}\mathbf{\Sigma_z^{-1}}\mathbf{\Sigma_{zy}}, \quad \mathbf{\Delta}_2 = \mathbf{\Sigma_z} - \mathbf{\Sigma_{zy}}\mathbf{\Sigma_y^{-1}}\mathbf{\Sigma_{yz}}.$$

APPENDIX B
PROOF OF LEMMA 3

We start from the following chain of inequalities:

$$\begin{aligned}
R(\mathbf{D}) &= \min_{\mathbf{u}} I(\mathbf{y};\mathbf{u}|\mathbf{z}) \ \text{ s.t. } \mathbf{\Sigma_{x|uz}} \preceq \mathbf{D}, \mathbf{u} \leftrightarrow \mathbf{y} \leftrightarrow (\mathbf{x},\mathbf{z}) \\
&\geq \min_{\mathbf{u}} I(\mathbf{y};\mathbf{u}|\mathbf{z}) \ \text{ s.t. } \mathbf{\Sigma_{x|uz}} \preceq \mathbf{D}, \mathbf{u} \leftrightarrow (\mathbf{y},\mathbf{z}) \leftrightarrow \mathbf{x} && (60) \\
&= \min_{\mathbf{u}} I(\mathbf{y};\mathbf{u}|\mathbf{z}) \ \text{ s.t. } \mathbf{\Sigma_{x|uz}} \preceq \mathbf{D}, \mathbf{u} \leftrightarrow (\mathbf{y}',\mathbf{z}) \leftrightarrow \mathbf{x} && (61) \\
&\geq \min_{\mathbf{u}} I(\mathbf{y}';\mathbf{u}|\mathbf{z}) \ \text{ s.t. } \mathbf{\Sigma_{x|uz}} \preceq \mathbf{D}, \mathbf{u} \leftrightarrow (\mathbf{y}',\mathbf{z}) \leftrightarrow \mathbf{x} && (62) \\
&= \min_{\mathbf{u}} h(\mathbf{y}'|\mathbf{z}) - h(\mathbf{y}'|\mathbf{u},\mathbf{z}) \ \text{ s.t. } \begin{cases} \mathbf{\Sigma_{x|uz}} \preceq \mathbf{D}, \\ \mathbf{u} \leftrightarrow (\mathbf{y}',\mathbf{z}) \leftrightarrow \mathbf{x} \end{cases} \\
&\geq \min_{\mathbf{\Sigma_{y'|uz}}} \frac{1}{2} \log \frac{\left|\mathbf{\Sigma_{y'|z}}\right|}{\left|\mathbf{\Sigma_{y'|uz}}\right|} \text{ s.t. } \begin{cases} \mathbf{\Sigma_{y'|uz}} \preceq \mathbf{\Sigma_{y'|z}} \\ \mathbf{\Sigma_{x|uz}} \preceq \mathbf{D} \\ \mathbf{u} \leftrightarrow (\mathbf{y}',\mathbf{z}) \leftrightarrow \mathbf{x} \end{cases} && (63) \\
&\geq \min_{\mathbf{\Sigma_{y'|uz}}} \frac{1}{2} \log \frac{\left|\mathbf{\Sigma_{y'|z}}\right|}{\left|\mathbf{\Sigma_{y'|uz}}\right|} \text{ s.t. } \mathbf{\Sigma_{y'|uz}} \preceq \mathbf{\Sigma_{y'|z}}, \mathbf{\Sigma_{x|uz}} \preceq \mathbf{D} \\
&= \min_{\mathbf{\Sigma_{y'|uz}}} \frac{1}{2} \log \frac{\left|\mathbf{\Sigma_{y'|z}}\right|}{\left|\mathbf{\Sigma_{y'|uz}}\right|} \text{ s.t. } \begin{cases} \mathbf{\Sigma_{y'|uz}} \preceq \mathbf{\Sigma_{x|z}} - \mathbf{\Sigma_{x|yz}} \\ \mathbf{\Sigma_{y'|uz}} \preceq \mathbf{D} - \mathbf{\Sigma_{x|yz}} \end{cases} && (64) \\
&= \frac{1}{2} \log \frac{\left|\mathbf{\Sigma_{y'|z}}\right|}{a^*}, && (65)
\end{aligned}$$

where (60) is because the knowledge of the side information $\mathbf{z}$ at the encoder cannot increase the rate, (61) is because from (3), when $\mathbf{z}$ is given, $\mathbf{y}'$ is a sufficient statistic of $\mathbf{y}$ for the estimation of $\mathbf{x}$, (62) follows from (4) and the data processing inequality, (63) is because choosing $\mathbf{y}'|\mathbf{uz}$ to be Gaussian maximizes the differential entropy, and we have added the constraint $\mathbf{\Sigma_{y'|uz}} \preceq \mathbf{\Sigma_{y'|z}}$, since any valid $\mathbf{\Sigma_{y'|uz}}$ must satisfy this condition, (64) follows from (5)–(6), and a^* is defined as:

$$a^* = \max_{\mathbf{\Sigma_{y'|uz}}} \left|\mathbf{\Sigma_{y'|uz}}\right| \text{ s.t. } \begin{cases} \mathbf{\Sigma_{y'|uz}} \preceq \mathbf{D} - \mathbf{\Sigma_{x|yz}}, \\ \mathbf{\Sigma_{y'|uz}} \preceq \mathbf{\Sigma_{x|z}} - \mathbf{\Sigma_{x|yz}}. \end{cases}$$

From Lemma 2 it follows that $a^* = \left|\min(\mathbf{D} - \mathbf{\Sigma_{x|yz}}, \mathbf{\Sigma_{x|z}} - \mathbf{\Sigma_{x|yz}})\right|$. Substituting this and (5) in (65) yields:

$$R(\mathbf{D}) \geq \frac{1}{2} \log \frac{\left|\mathbf{\Sigma_{x|z}} - \mathbf{\Sigma_{x|yz}}\right|}{\left|\min(\mathbf{D} - \mathbf{\Sigma_{x|yz}}, \mathbf{\Sigma_{x|z}} - \mathbf{\Sigma_{x|yz}})\right|}.$$

APPENDIX C
PROOF OF LEMMA 4

Based on the results in [5], [4], it is enough to show that for $\mathbf{u}^*$ defined in (11), we have $I(\mathbf{y};\mathbf{u}^*|\mathbf{z}) = \tilde{R}(\mathbf{D})$, and the covariance matrix $\mathbf{\Sigma_{x|u^*z}}$ of the reconstruction error satisfies $\mathbf{\Sigma_{x|u^*z}} \preceq \mathbf{D}$. We start with the following chain of equalities:

$$\begin{aligned}
I(\mathbf{y};\mathbf{u}^*|\mathbf{z}) &= h(\mathbf{u}^*|\mathbf{z}) - h(\mathbf{u}^*|\mathbf{y},\mathbf{z}) \\
&= \frac{1}{2} \log \left(\frac{\left|\mathbf{\Sigma_{u^*|z}}\right|}{\left|\mathbf{\Sigma_{u^*|yz}}\right|} \right) \\
&= \frac{1}{2} \log \left(\frac{\left|\mathbf{U}\mathbf{\Sigma_{y'|z}}\mathbf{U}^T + \mathbf{\Sigma}_\nu\right|}{\left|\mathbf{\Sigma}_\nu\right|} \right) && (66) \\
&= \frac{1}{2} \log \left(\frac{\left|\mathbf{\Lambda} + \mathbf{\Sigma}_\nu\right|}{\left|\mathbf{\Sigma}_\nu\right|} \right). && (67)
\end{aligned}$$

where (66) follows from (11) and (4), and (67) follows from (5). Rewriting $\mathbf{\Lambda}$ as:

$$\mathbf{\Lambda} = \mathbf{U}(\mathbf{\Sigma_{x|z}} - \mathbf{\Sigma_{x|yz}})\mathbf{U}^T = \mathbf{U}\mathbf{V}^{-1}\mathbf{\Lambda}\mathbf{V}^{-T}\mathbf{U}^T, \tag{68}$$

and substituting (68) and (12) in (67) and simplifying the result, we get:

$$\begin{aligned}
&I(\mathbf{y};\mathbf{u}^*|\mathbf{z}) \\
&= \frac{1}{2} \log \left(\frac{\left|\mathbf{V}^{-1}\mathbf{\Lambda}\mathbf{V}^{-T}\right|}{\left|\mathbf{V}^{-1}\mathrm{diag}\left\{\min(\lambda_i, \lambda_i'), i = 1, \ldots, n_x\right\}\mathbf{V}^{-T}\right|} \right) \\
&= \frac{1}{2} \log \frac{\left|\mathbf{\Sigma_{x|z}} - \mathbf{\Sigma_{x|yz}}\right|}{\left|\min(\mathbf{D} - \mathbf{\Sigma_{x|yz}}, \mathbf{\Sigma_{x|z}} - \mathbf{\Sigma_{x|yz}})\right|} = \tilde{R}(\mathbf{D}).
\end{aligned}$$

Next, we derive the covariance matrix of the reconstruction error. Similar to (3), we can write $\mathbf{x}$ as:

$$\mathbf{x} = \mathbf{C}\mathbf{u}^* + \mathbf{G}\mathbf{z} + \mathbf{n}_1, \tag{69}$$

where $\mathbf{n}_1$ is independent of $\mathbf{u}^*$ and $\mathbf{z}$. Form (69) it follows that $\mathbf{\Sigma_{xu^*|z}} = \mathbf{C}\mathbf{\Sigma_{u^*|z}}$, or:

$$\mathbf{C} = \mathbf{\Sigma_{xu^*|z}}\mathbf{\Sigma_{u^*|z}^{-1}}. \tag{70}$$

From (11), (4), and (5) we have:

$$\mathbf{\Sigma_{u^*|z}} = \mathbf{\Lambda} + \mathbf{\Sigma}_\nu. \tag{71}$$

Also note that:

$$\begin{aligned}
\mathbf{\Sigma_{xu^*|z}} &= \mathbf{\Sigma_{xy|z}}\mathbf{A}^T\mathbf{U}^T && (72) \\
&= \mathbf{A}\mathbf{\Sigma_{y|z}}\mathbf{A}^T\mathbf{U}^T && (73) \\
&= \mathbf{\Sigma_{y'|z}}\mathbf{U}^T && (74) \\
&= \left(\mathbf{\Sigma_{x|z}} - \mathbf{\Sigma_{x|yz}}\right)\mathbf{U}^T, && (75)
\end{aligned}$$

where (72), (73), (74) and (75) follow from (11), (3), (4) and (5), respectively. The covariance matrix of the reconstruction error can then be written as:

$$\mathbf{\Sigma_{x|u^*z}} = \mathbf{\Sigma_{n_1}} \tag{76}$$

$$= \mathbf{\Sigma_{x|z}} - \mathbf{C}\mathbf{\Sigma_{u^*|z}}\mathbf{C}^T \tag{77}$$

$$= \mathbf{\Sigma_{x|z}} - \mathbf{\Sigma_{xu^*|z}}\mathbf{\Sigma}^{-1}_{\mathbf{u^*|z}}\mathbf{\Sigma}^T_{\mathbf{xu^*|z}} \tag{78}$$

$$= \mathbf{\Sigma_{x|z}} - \left(\mathbf{\Sigma_{x|z}} - \mathbf{\Sigma_{x|yz}}\right)\mathbf{U}^T(\mathbf{\Lambda}+\mathbf{\Sigma}_\nu)^{-1}\mathbf{U}\left(\mathbf{\Sigma_{x|z}} - \mathbf{\Sigma_{x|yz}}\right) \tag{79}$$

$$\begin{aligned}= \mathbf{\Sigma_{x|yz}} &+ \left(\mathbf{\Sigma_{x|z}} - \mathbf{\Sigma_{x|yz}}\right) \\ &- \left(\mathbf{\Sigma_{x|z}} - \mathbf{\Sigma_{x|yz}}\right)\mathbf{U}^T(\mathbf{\Lambda} + \mathbf{\Sigma}_\nu)^{-1}\mathbf{U}\left(\mathbf{\Sigma_{x|z}} - \mathbf{\Sigma_{x|yz}}\right)\end{aligned}$$

$$\begin{aligned}= \mathbf{\Sigma_{x|yz}} &+ \mathbf{V}^{-1}\mathbf{\Lambda}\mathbf{V}^{-T} \\ &- \mathbf{V}^{-1}\mathbf{\Lambda}\mathbf{V}^{-T}\mathbf{U}^T(\mathbf{\Lambda} + \mathbf{\Sigma}_\nu)^{-1}\mathbf{U}\mathbf{V}^{-1}\mathbf{\Lambda}\mathbf{V}^{-T}\end{aligned} \tag{80}$$

$$\begin{aligned}= \mathbf{\Sigma_{x|yz}} &+ \mathbf{V}^{-1}\mathbf{\Lambda}\mathbf{V}^{-T} \\ &-\mathbf{V}^{-1}\mathbf{\Lambda}\left(\mathbf{\Lambda}+\mathrm{diag}\left\{\frac{\lambda_i \min\left(\lambda_i,\lambda_i'\right)}{\lambda_i-\min\left(\lambda_i,\lambda_i'\right)}, i=1,\ldots,n_x\right\}\right)^{-1}\mathbf{\Lambda}\mathbf{V}^{-T}\end{aligned} \tag{81}$$

$$= \mathbf{\Sigma_{x|yz}} + \mathbf{V}^{-1}\mathrm{diag}\left\{\min\left(\lambda_i,\lambda_i'\right), i=1,\ldots,n_x\right\}\mathbf{V}^{-T}$$

$$= \mathbf{\Sigma_{x|yz}} + \min\left(\mathbf{\Sigma_{x|z}} - \mathbf{\Sigma_{x|yz}}, \mathbf{D} - \mathbf{\Sigma_{x|yz}}\right) \tag{82}$$

where (76) and (77) follow from (69), (78) is the result of substituting (70) in (77), (79) follows from (71) and (75), (80) follows from (8), and (81) follows from (68) and (12).

Finally, rewriting (82) as:

$$\mathbf{\Sigma_{x|u^*z}} - \mathbf{\Sigma_{x|yz}} = \min\left(\mathbf{\Sigma_{x|z}} - \mathbf{\Sigma_{x|yz}}, \mathbf{D} - \mathbf{\Sigma_{x|yz}}\right), \tag{83}$$

and using Property 3 in Lemma 1, it is clear that $\mathbf{\Sigma_{x|u^*z}} \preceq \mathbf{D}$, as desired. This completes the proof.

Appendix D
Derivation of the KKT Conditions

We start with defining $\mathbf{A}$ as in (41). Writing $\mathbf{D}_i$ in terms of $\mathbf{Z}_i$ using (38), and substituting in the constraint in (36) yields (42). We derive (40) to complete the proof. Applying the matrix inversion lemma to $\left[\mathbf{\Sigma_{n}}_i + \left(\mathbf{D}_i^{-1} - \mathbf{\Sigma}_{\mathbf{y}_i}^{-1}\right)^{-1}\right]^{-1}$ in (36), and using the definitions (38) and (41), we rewrite the optimization problem as follows:

$$\begin{aligned}&\min_{\mathbf{Z}_1,\cdots,\mathbf{Z}_N} \mathrm{tr}\left\{\mathbf{A}^{-1}\right\} \\ &s.t. \quad \prod_{i=1}^{N}\left|\mathbf{\Sigma}_{\mathbf{y}_i}^{-1} - \mathbf{\Sigma}_{\mathbf{n}_i}^{-1} + \mathbf{\Sigma}_{\mathbf{n}_i}^{-1}\mathbf{W}_i^T\mathbf{Z}_i^{-1}\mathbf{W}_i\mathbf{\Sigma}_{\mathbf{n}_i}^{-1}\right|^{-\alpha_i} = \beta.\end{aligned} \tag{84}$$

For a given $i \in \{1,\cdots,N\}$, we differentiate the Lagrangian of (84) with respect to $\mathbf{Z}_i$. To do so, we first apply a logarithm function to the constraint in (84), and then write the Lagrangian form of the problem as:

$$\mathcal{L}(\lambda, \mathbf{Z}_1, \cdots, \mathbf{Z}_N) = \mathrm{tr}\left\{\left(\mathbf{A}_i - \mathbf{Z}_i\right)^{-1}\right\} + \lambda\left(\sum_{j=1}^{N} f_j(\mathbf{Z}_j) - \log\beta\right), \tag{85}$$

where $\mathbf{A}_i$ and $f_j(\mathbf{Z}_j)$ are defined as:

$$\mathbf{A}_i = \mathbf{A} + \mathbf{Z}_i,$$

$$f_j(\mathbf{Z}_j) = -\alpha_j \log\left|\mathbf{\Sigma}_{\mathbf{y}_j}^{-1} - \mathbf{\Sigma}_{\mathbf{n}_j}^{-1} + \mathbf{\Sigma}_{\mathbf{n}_j}^{-1}\mathbf{W}_j^T\mathbf{Z}_j^{-1}\mathbf{W}_j\mathbf{\Sigma}_{\mathbf{n}_j}^{-1}\right|.$$

Note that from (41) it follows that $\mathbf{A}_i$ does not depend on $\mathbf{Z}_i$. Using (39), we rewrite $f_j(\mathbf{Z}_j)$ as:

$$\begin{aligned}f_j(\mathbf{Z}_j) &= -\alpha_j \log\left|\mathbf{\Sigma}_{\mathbf{n}_j}^{-1}\mathbf{W}_j^T\left(\mathbf{Z}_j^{-1} - \mathbf{C}_j^{-1}\right)\mathbf{W}_j\mathbf{\Sigma}_{\mathbf{n}_j}^{-1}\right| \\ &= -\alpha_j \log\left|\mathbf{\Sigma}_{\mathbf{n}_j}^{-1}\mathbf{W}_j^T\mathbf{W}_j\mathbf{\Sigma}_{\mathbf{n}_j}^{-1}\right| - \alpha_j\log\left|\mathbf{C}_j^{-1}\right| \\ &\quad - \alpha_j\log\left|\mathbf{Z}_j^{-1}\right| - \alpha_j\log|\mathbf{C}_j - \mathbf{Z}_j|.\end{aligned} \tag{86}$$

Substituting (86) in (85), differentiating with respect to $\mathbf{Z}_i$ while taking into account that $\mathbf{Z}_i$ is symmetric [26], and setting the derivative equal to 0 yields:

$$\begin{aligned}&2(\mathbf{A}_i - \mathbf{Z}_i)^{-2} - (\mathbf{A}_i - \mathbf{Z}_i)^{-2} \circ \mathbf{I}_n + 2\lambda\alpha_i\left[\mathbf{Z}_i^{-1} + (\mathbf{C}_i - \mathbf{Z}_i)^{-1}\right] \\ &- \lambda\alpha_i\left[\mathbf{Z}_i^{-1} + (\mathbf{C}_i - \mathbf{Z}_i)^{-1}\right] \circ \mathbf{I}_n = \mathbf{0},\end{aligned} \tag{87}$$

where $\circ$ denotes the Hadamard product. From (87) it follows that:

$$(\mathbf{A}_i - \mathbf{Z}_i)^{-2} + \lambda\alpha_i\left[\mathbf{Z}_i^{-1} + (\mathbf{C}_i - \mathbf{Z}_i)^{-1}\right] = \mathbf{0}. \tag{88}$$

Replacing $\mathbf{A}_i - \mathbf{Z}_i$ by $\mathbf{A}$ in (88) and simplifying the result yields (40).

Appendix E
Proof of Proposition 1

Let us define:

$$f_i(\lambda) = \left(\frac{\sqrt{1 + 4\lambda s_i} - 1}{\sqrt{4\lambda}}\right)^2; i = 1, \cdots, N.$$

Using basic calculus one can show that $f_i(\lambda)$ is a monotonically increasing and strictly concave function of λ, which varies from 0 to s_i when λ goes from 0 to ∞. It then follows that the function $f(\lambda) = \prod_{i=1}^{N} f_i(\lambda) - \gamma$ is monotonically increasing with the range $[-\gamma, |\mathbf{S}| - \gamma[$. One could then see that $f(\lambda)$ has one root, if and only if:

$$|\mathbf{S}| - \gamma > 0, \tag{89}$$

otherwise it has no root. Substituting (51) in (89) and simplifying the result yields $R > R_{min}$.

Appendix F
Proof of Proposition 2

Applying the assumptions $\alpha_1 = \alpha_2 = 0.5$, $w_1 = w_2$, and $n = 1$, the optimization problem is simplified to:

$$\begin{aligned}\max_{D_1,D_2} &\frac{1}{\Sigma_{n_1}} - \frac{1}{{\Sigma_{n_1}}^2}\left(\frac{1}{\frac{1}{\Sigma_{n_1}} + \frac{1}{D_1} - \frac{1}{\Sigma_{y_1}}}\right) \\ &+ \frac{1}{\Sigma_{n_2}} - \frac{1}{{\Sigma_{n_2}}^2}\left(\frac{1}{\frac{1}{\Sigma_{n_2}} + \frac{1}{D_2} - \frac{1}{\Sigma_{y_2}}}\right) \quad s.t.\ D_1D_2 = \beta^2.\end{aligned} \tag{90}$$

Applying the following change of variables:

$$D_i' = \Sigma_i + \frac{{\Sigma_{n_i}}^2}{D_i}, \quad i = 1, 2, \tag{91}$$

and defining β' as:

$$\beta' = (\Sigma_{n_1} - \Sigma_1)(\Sigma_{n_2} - \Sigma_2)e^{4R}, \tag{92}$$

one could rewrite (90) as:

$$\min_{D_1', D_2'} \left(\frac{1}{D_1'} + \frac{1}{D_2'} \right) \quad s.t. \quad (D_1' - \Sigma_1)(D_2' - \Sigma_2) = \beta'. \tag{93}$$

Using the constraint in (93), we then write D_2' in terms of D_1' and substitute it in the cost function to obtain an unconstrained function of D_1'. We call this function $f_1(D_1')$. Differentiating $f_1(D_1')$ with respect to D_1' yields:

$$\frac{df_1(D_1')}{dD_1'} = \frac{\left[\sqrt{\beta'}D_1' + \beta' + \Sigma_2(D_1' - \Sigma_1)\right]\left[(\sqrt{\beta'} - \Sigma_2)D_1' - (\beta' - \Sigma_1\Sigma_2)\right]}{{(D_1')}^2\left(\Sigma_2 D_1' - \Sigma_1\Sigma_2 + \beta'\right)^2}, \tag{94}$$

which is zero at $D_1'^* = \Sigma_1 + \frac{{\Sigma_{n_1}}^2}{D_1^*}$ (equivalent to $D_1 = D_1^*$ using (91)). The sign of the derivative around this point determines whether there is an extremum or not. The denominator and the first term in the numerator in (94) are positive. We thus study only the second term in the numerator. We replace D_1' by $\Sigma_1 + \frac{{\Sigma_{n_1}}^2}{D_1^*} + \epsilon$. The result is $(\sqrt{\beta'} - \Sigma_2)\epsilon$, which implies that if $\sqrt{\beta'} > \Sigma_2$, the derivative is positive for $\epsilon > 0$ and negative for $\epsilon < 0$. This means that if $\sqrt{\beta'} > \Sigma_2$, the minimum is at $D_1' = \Sigma_1 + \frac{{\Sigma_{n_1}}^2}{D_1^*}$, or equivalently $D_1 = D_1^*$ is the global maximizer of the SNR. Similarly, one could write D_1' in terms of D_2' and substitute in the cost function to obtain $f_2(D_2')$. Considering the sign of the derivative around $D_2 = D_2^*$ leads to the conclusion that if $\sqrt{\beta'} > \Sigma_1$, the SNR is maximized at $D_2 = D_2^*$. Combining the two conditions, the following should hold in order to have the SNR maximized at (D_1^*, D_2^*):

$$\sqrt{\beta'} > \max(\Sigma_1, \Sigma_2). \tag{95}$$

Substituting (92) in (95) yields:

$$R > \frac{1}{4} \log \frac{\left[\max(\Sigma_1, \Sigma_2)\right]^2}{(\Sigma_{n_1} - \Sigma_1)(\Sigma_{n_2} - \Sigma_2)} = R_{max}. \tag{96}$$

Similarly one can show that for $R < R_{min}$, (D_1^*, D_2^*) minimizes the output SNR. This completes the proof.

REFERENCES

[1] A. Zahedi, J. Østergaard, S. H. Jensen, P. Naylor, and S. Bech, *Coding and enhancement in wireless acoustic sensor networks*, IEEE Data Compression Conference, Snowbird, UT, Apr. 2015.

[2] P. L. Dragotti, M. Gastpar, *Distributed source coding, theory, algorithms and applications*, Elsevier, 1st edition, 2009.

[3] A. Zahedi, J. Østergaard, S. H. Jensen, P. Naylor, and S. Bech, *Audio coding in wireless acoustic sensor networks*, Signal Processing, vol. 107, pp. 141–152, Feb. 2015.

[4] A. Wyner and J. Ziv, *The rate-distortion function for source coding with side information at the decoder*, IEEE Transactions on Information Theory, vol. 22, no. 1, pp.1-10, Jan. 1976.

[5] D. Slepian and J.Wolf, *Noiseless coding of correlated information sources*, IEEE Transactions on Information Theory, vol. 19, no. 4, pp.4711-480, Jul. 1973.

[6] A. Wyner, *The rate-distortion function for source coding with side information at the decoder–II:General sources*, Information and Control, vol. 38, pp. 60-80, 1978.

[7] S. C. Draper and G. W. Wornell, *Side information aware coding strategies for estimation under communication constraints*, IEEE Journal on Selected Areas in Communications, vol. 22, no. 6, pp. 1-11, Aug. 2004.

[8] H. Yamamoto, *Wyner-Ziv theory for a general function of the correlated sources*, IEEE Transactions on Information Theory, vol. IT-28, no. 5, pp. 803-807, Sep. 1982.

[9] H. Yamamoto and K. Itoh, *Source coding theory for multiterminal communication systems with a remote source*, IEICE Transactions, vol. E63, no. 10, pp.700-706, Oct. 1980.

[10] C. Tian and J. Chen, *Remote vector Gaussian source coding with decoder side information under mutual information and distortion constraints*, IEEE Transactions on Information Theory, vol. 55, no. 10, pp.4676-4680, Oct. 2009.

[11] A. Zahedi, J. Østergaard, S. H. Jensen, P. Naylor, and S. Bech, *Distributed remote vector Gaussian source coding with covariance distortion constraints*, IEEE International Symposium on Information Theory (ISIT), Honolulu, HI, USA, Jul. 2014.

[12] H. Wang and P. Viswanath, *Vector Gaussian multiple description with individual and central receivers*, IEEE Transactions on Information Theory, vol. 53, no. 6, pp.2133-2153, Jun. 2007.

[13] M. S. Rahman and A. Wagner, *Rate region of the Gaussian scalar-help-vector source-coding problem*, IEEE Transactions on Information Theory, vol. 58, no. 1, pp.172-188, Jan. 2012.

[14] M. S. Rahman and A. B. Wagner, *Rate region of the vector Gaussian one-helper source-coding problem*, IEEE Transactions on Information Theory, vol. 61, no. 5, pp.2708-2728, May 2015.

[15] J. Wang and J. Chen, Vector Gaussian multiterminal source coding, IEEE Transactions on Information Theory, vol. 60, no. 9, pp.5533-5552, Sep. 2014.

[16] E. Ekrem and S. Ulukus, *An outer bound for the vector Gaussian CEO problem*, IEEE Transactions on Information Theory, vol. 60, no. 11, pp.6870-6887, Nov. 2014.

[17] Y. Oohama, *Indirect and direct Gaussian distributed source coding problems*, IEEE Transactions on Information Theory, vol. 60, no. 12, pp.7506-7539, Dec. 2014.

[18] T. Berger, Z. Zhang, and H. Viswanathan, *The CEO problem*, IEEE Transactions on Information Theory, vol. 42, no. 3, pp. 887-902, May 1996.

[19] D. Bernstein, *Matrix mathematics, theory, facts and formulas*, Princeton University Press, 2nd edition, 2009.

[20] C. Li and L. Rodman, *Some extremal problems for positive definite matrices and operators*, Linear Algebra and its Applications, vol. 140, pp.139-154, Oct. 1990.

[21] R. Zamir and M. Feder, *On lattice quantization noise*, IEEE Transactions on Information Theory vol. 42, no. 4, pp.1152-1159, July 1996.

[22] J. K. Wolf and J. Ziv, *Transmission of noisy information to a noisy receiver with minimum distortion*, IEEE Transactions on Information Theory, vol. IT-16, pp. 406-411, Jul. 1970.

[23] A. Lapidoth and I.-H. Wang, *Communicating remote Gaussian sources over Gaussian multiple access channels*, IEEE International Symposium on Information Theory (ISIT), Saint Petersburg, Russia, Aug. 2011.

[24] J. Benesty, J. Chen, and Y. Huang, *Microphone array signal processing*, Springer-Verlag, 2008.

[25] E. I. Silva, M. S. Derpich, and J. Østergaard, *An Achievable Data-Rate Region Subject to a Stationary Performance Constraint for LTI Plants*, IEEE Transactions on Automatic Control, vol. 56, no. 8, pp. 1968 - 1973, Aug. 2011.

[26] K. B. Petersen and M. S. Pedersen, *The matrix cookbook*, 2012, http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=3274.