2015 Data Compression Conference

Coding and Enhancement in Wireless Acoustic Sensor Networks

Adel Zahedi*, Jan Østergaard*, Søren Holdt Jensen*, Patrick Naylor†, and Søren Bech*‡
*Department of Electronic Systems
Aalborg University, 9220 Aalborg, Denmark
Email: {adz, jo, shj, sbe}@es.aau.dk
†Electrical and Electronic Engineering Department
London Imperial College, London SW7 2AZ, UK
Email: p.naylor@imperial.ac.uk
‡Bang & Olufsen
7600 Struer, Denmark

Abstract

We formulate a new problem which bridges between source coding and enhancement in wireless acoustic sensor networks. We consider a network of wireless microphones, each of which encoding its own measurement under a covariance matrix distortion constraint and sending it to a fusion center. To process the data at the center, we use a recent spatio-temporal prediction filter. We assume that a weighted sum-rate for the network is specified. The problem is to allocate optimal distortion matrices to the nodes in order to achieve a maximum output SNR at the fusion center after processing the received data, while the weighted sum-rate for the network is no more than the specified value. We formulate this problem as an optimization problem for which we derive a set of equalities imposed on the solution by studying the KKT conditions. In particular, for the special case of scalar sources with two microphones and a sum-rate constraint, we derive the distortion allocation in closed form and will show that if the given sum-rate is higher than a critical value, the stationary points from the KKT conditions lead to distortion allocations which maximize the output SNR of the filter.

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.

1068-0314/15 $31.00 © 2015 IEEE
DOI 10.1109/DCC.2015.20

IEEE computer society

I. INTRODUCTION

A wireless acoustic sensor network (WASN) is a set of wireless microphones equipped with communications and signal processing units, and distributed over an environment to record sound and transmit it to a fusion center for further processing. Thanks to the microphones being wireless and the distributed nature, WASNs are capable of being used as a much more promising alternative to the classical microphone arrays for a variety of applications, including ambient intelligence, vehicle tracking, hearing aids, etc [1].

To minimize the demands on the bandwidth and transmission power over the network, the recorded data at the microphones should be quantized and encoded before being transmitted to the center. The problem of source coding for WASNs has been addressed in [2]– [5]. Here, we assume that the set of rate-distortion functions for the nodes and the optimal achieving coding schemes are known. What we are interested in is to allocate the distortion constraints in such a way that for a given weighted sum-rate for the network, the output SNR after processing the data at the fusion center is maximized.

The rest of the paper is organized as follows. In Section II, the spatio-temporal prediction approach for enhancement in microphone arrays, and the rate-distortion functions used in this paper are briefly introduced. In Section III, the problem is formulated and the optimality conditions on the solution are studied. In Section IV, the closed-form solution for the nodes distortions for the special but insightful case of two microphones with scalar sources is presented. Section V concludes the paper.

II. PRELIMINARIES

This section is dedicated to a brief introduction to the enhancement filter and the rate-distortion functions we use in this paper. We assume that all random vectors are of zero mean. The covariance matrix of a random vector is denoted by the symbol $\mathbf{\Sigma}$ followed by an appropriate subscript. All covariance matrices are assumed to be of full rank. The trace and expectation operations are denoted by $\mathrm{tr}(\cdot)$ and $E[\cdot]$, respectively. To indicate that $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$ is positive-semidefinite, we write $\mathbf{\Sigma}_1 \succeq \mathbf{\Sigma}_2$.

A. Spatio-Temporal Prediction Filter

Consider an array of N microphones, where microphone $i,\ i = 1, ..., N$ records a noisy version $y_i(k)$ of a sound source $x_i(k)$ given by:

$$y_i(k) = x_i(k) + n_i(k), \tag{1}$$

where k is a time index, and the noise term $n_i(k)$ is assumed to be uncorrelated with the source $x_i(k)$. In the vector form, (1) can be written as:

$$\mathbf{y}_i = \mathbf{x}_i + \mathbf{n}_i, \tag{2}$$

where $\mathbf{y}_i$ is a vector in $\mathbb{R}^n$ defined as $\mathbf{y}_i = [y_i(k-n+1)\ ...\ y_i(k)]^T$ (and similarly for $\mathbf{x}_i$ and $\mathbf{n}_i$), and the dependency on k is dropped for simplicity of notation. At the processing unit, as shown in Fig. 1, the output $\mathbf{z}$ is produced using the following filtering operation:

$$\mathbf{z} = \sum_{i=1}^{N} \mathbf{H}_i \mathbf{y}_i = \mathbf{H}\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{H}\mathbf{n}, \tag{3}$$

where $\mathbf{H} = [\mathbf{H}_1 \ ... \ \mathbf{H}_N]$ and $\mathbf{y} = [\mathbf{y}_1^T \ ... \ \mathbf{y}_N^T]^T$ (and similarly for $\mathbf{x}$ and $\mathbf{n}$). Consider a reference signal $\mathbf{x}_r$ related to $\mathbf{x}_i,\ i = 1, ..., N$. From the information-theoretic point of view, the optimal filter could be thought of as one that minimizes the mean-squared error between the output $\mathbf{z}$ and the reference $\mathbf{x}_r$. This is equivalent to the well-known Wiener filter. This filter, however, distorts the audio signal. Perceptually, it is more desirable to avoid this linear distortion in price of more additive distortion (lower SNR). The goal is to find the optimal filter $\mathbf{H}$ that maximizes the SNR at the output without distorting the signal. A very comprehensive technique is the spatio-temporal prediction approach [6], which in addition to the time-domain predictability of the signals, also utilizes the spatial correlation of the signals at different microphones. This filter is briefly introduced in the rest of this subsection.

Assuming that there is a set of $n \times n$ matrices $\mathbf{W}_i,\ i = 1, ..., N$ so that we have: [1]

$$\mathbf{x}_i = \mathbf{W}_i \, \mathbf{x}_r,\ i = 1, ..., N, \tag{4}$$

or equivalently:

$$\mathbf{x} = \mathbf{W}^T \mathbf{x}_r, \tag{5}$$

where $\mathbf{W}$ is defined as:

$$\mathbf{W} = [\mathbf{W}_1 \ ... \ \mathbf{W}_N], \tag{6}$$

the error at the output of the filter $\mathbf{H}$ is given by:

$$\mathbf{e} = \mathbf{z} - \mathbf{x}_r = \left(\mathbf{H}\mathbf{W}^T - \mathbf{I}\right) \mathbf{x}_r + \mathbf{H}\mathbf{n} = \mathbf{e}_\mathbf{x} + \mathbf{e}_\mathbf{n}, \tag{7}$$

where $\mathbf{e}_\mathbf{x}$ is the distortion of the signal and $\mathbf{e}_\mathbf{n}$ is the output noise. To minimize the noise without distorting the signal, one can use the filter $\mathbf{H}_{ST}$ given by the following problem:

$$\mathbf{H}_{ST} = \arg\min_{\mathbf{H}} \mathrm{tr}(E[\mathbf{e}_\mathbf{n} \mathbf{e}_\mathbf{n}^T]) \quad s.t. \quad \mathbf{H}\mathbf{W}^T = \mathbf{I}, \tag{8}$$

There is a closed form solution for (8) given by [6]:

$$\mathbf{H}_{ST} = \left(\mathbf{W}\mathbf{\Sigma}_\mathbf{n}^{-1}\mathbf{W}^T\right)^{-1} \mathbf{W}\mathbf{\Sigma}_\mathbf{n}^{-1}, \tag{9}$$

which leads to the following SNR at the output:

[1] This means that there is a deterministic relationship between the remote sources $\mathbf{x}_i$ and the reference signal $\mathbf{x}_r$. When the relationship is not deterministic, one can choose $\mathbf{W}_i$ to minimize the mean-squared error between $\mathbf{x}_i$ and $\mathbf{x}_r$. In this case, (4) would be an approximation. See [6] for more details.

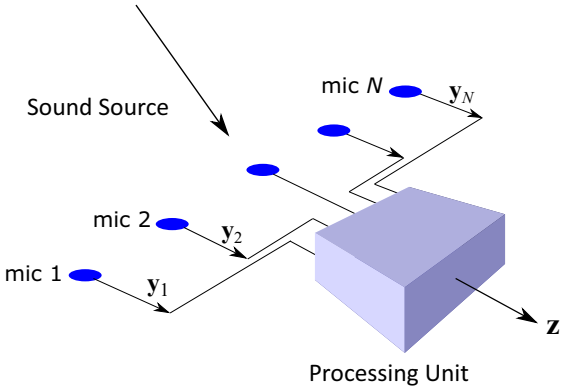

Figure 1: A microphone array

$$\text{SNR}(\mathbf{H}_{ST}) = \frac{\text{tr}\left(\mathbf{\Sigma}_{\mathbf{x}_r}\right)}{\text{tr}\left\{\left(\mathbf{W}\mathbf{\Sigma}_{\mathbf{n}}^{-1}\mathbf{W}^T\right)^{-1}\right\}}. \tag{10}$$

B. Rate-Distortion Functions

Suppose that a Gaussian source with samples $\mathbf{y} \in \mathbb{R}^n$ is to be encoded and transmitted with a finite rate, such that the reconstruction $\hat{\mathbf{y}}'$ at the decoder satisfies the following distortion constraint:

$$E\left[\left(\mathbf{y}-\hat{\mathbf{y}}'\right)\left(\mathbf{y}-\hat{\mathbf{y}}'\right)^T\right] \preceq \mathbf{D} \tag{11}$$

where the target distortion $\mathbf{D}$ is symmetric and positive definite. As a special case of the results in [7], the minimum rate yielding a given distortion $\mathbf{D}$ is given by:

$$R\left(\mathbf{D}\right) = \frac{1}{2}\log\left(\frac{|\mathbf{\Sigma}_{\mathbf{y}}|}{|\mathbf{D}|}\right), \tag{12}$$

for $\mathbf{0} \prec \mathbf{D} \preceq \mathbf{\Sigma}_{\mathbf{y}}$, and is zero for $\mathbf{D} \succeq \mathbf{\Sigma}_{\mathbf{y}}$. For other choices of $\mathbf{D}$ there is no closed-form formula for the rate-distortion function. The reconstruction $\hat{\mathbf{y}}'$ satisfies the following [7]:

$$\hat{\mathbf{y}}' = \mathbf{A}\mathbf{y} + \boldsymbol{\nu}', \tag{13}$$

where $\boldsymbol{\nu}'$ and $\mathbf{y}$ are uncorrelated and Gaussian, and we have:

$$\mathbf{A} = (\mathbf{\Sigma}_{\mathbf{y}} - \mathbf{D})\mathbf{\Sigma}_{\mathbf{y}}^{-1} \tag{14}$$

$$\mathbf{\Sigma}_{\boldsymbol{\nu}'} = \mathbf{A}\mathbf{D}. \tag{15}$$

We apply the invertible map $\mathbf{A}^{-1}$ to $\hat{\mathbf{y}}'$ to obtain the following equivalent version of the reconstruction which is more convenient to work with:

$$\hat{\mathbf{y}} = \mathbf{A}^{-1}\hat{\mathbf{y}}' = \mathbf{y} + \boldsymbol{\nu}, \tag{16}$$

with [2]:

$$\boldsymbol{\Sigma}_{\boldsymbol{\nu}} = \left(\mathbf{D}^{-1} - \boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\right)^{-1}. \tag{17}$$

III. Optimal Distortion Allocation

Consider a WASN with N sensor nodes, where node i records its own measurement $\mathbf{y}_i$ given by (2), and encodes it for a target distortion $\mathbf{D}_i$ leading to the rate $R(\mathbf{D}_i)$ given by (12). At the decoder, the reconstruction $\hat{\mathbf{y}}_i$ is obtained which satisfies (16)–(17). A filter is then applied to the reconstructions, yielding the output $\mathbf{z}$. This scenario is illustrated in Fig. 2.

From (2) and (16) the reconstructions can be written as:

$$\hat{\mathbf{y}}_i = \mathbf{x}_i + \mathbf{n}_i + \boldsymbol{\nu}_i = \mathbf{x}_i + \mathbf{v}_i, \tag{18}$$

or:

$$\hat{\mathbf{y}} = \mathbf{x} + \mathbf{v}, \tag{19}$$

where $\hat{\mathbf{y}}$ (and similarly $\mathbf{x}$ and $\mathbf{v}$) is defined as $\hat{\mathbf{y}} = [\hat{\mathbf{y}}_1^T \ ... \ \hat{\mathbf{y}}_N^T]^T$ and $\mathbf{v} = \mathbf{n}+\boldsymbol{\nu}$. Comparing (18) to (2), it is clear that the problem is similar to the case of the microphone array. One can thus apply the spatio-temporal prediction filter introduced in Section II-A to the set of reconstructions to produce the desired enhanced output. The output SNR similar to (10) is given by:

$$\text{SNR}(\mathbf{H}_{ST}, \mathbf{D}_1, ..., \mathbf{D}_N) = \frac{\text{tr}\left(\boldsymbol{\Sigma}_{\mathbf{x}_r}\right)}{\text{tr}\left\{\left[\mathbf{W}\left(\boldsymbol{\Sigma}_{\mathbf{v}}(\mathbf{D}_1, ..., \mathbf{D}_N)\right)^{-1}\mathbf{W}^T\right]^{-1}\right\}}. \tag{20}$$

where the dependency of the covariance matrix of the noise term $\mathbf{v}$ on the set of distortions is emphasized by the notation in $\boldsymbol{\Sigma}_{\mathbf{v}}(\mathbf{D}_1, ..., \mathbf{D}_N)$. The difference between the WASN and the microphone array in this context is now clear: Due to the limited transmission rate, in addition to the additive noise in the measurements (which is in common between the two cases), in the WASN case there is also a reconstruction error due to quantization and coding. While this error is inevitable, we can take the role of the distortions $\mathbf{D}_1, ..., \mathbf{D}_N$ in (20) to the advantage of shaping the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{v}}(\mathbf{D}_1, ..., \mathbf{D}_N)$ in order to maximize the output SNR. To do so, suppose that for a given set of weights $\alpha_1, ..., \alpha_N$, the weighted sum-rate of the network $\sum_{i=1}^{N} \alpha_i R(\mathbf{D}_i)$ is required to be no more than R. The problem can then be formulated as follows:

[2]Note that in this case the linear distortion term $\mathbf{A}$ is removed from the reconstruction, resulting in a variation in the additive distortion. One can see from (17) that if the inequality $\mathbf{D} \prec \boldsymbol{\Sigma}_{\mathbf{y}}$ is not strict, the covariance matrix of the noise term blows up. This is in agreement with the results in [8]

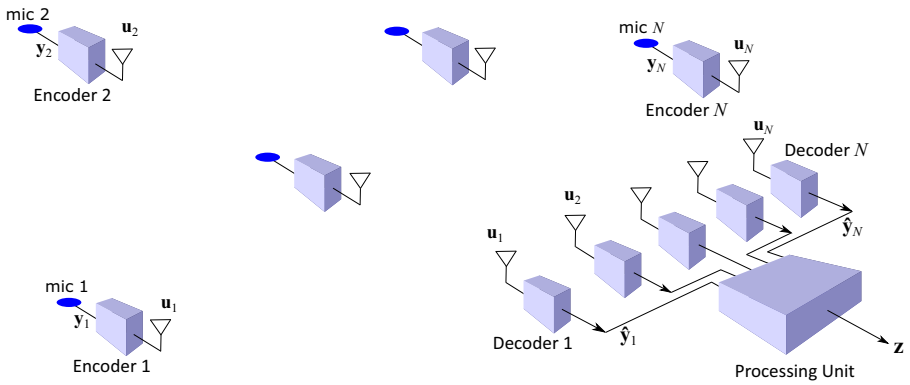

Figure 2: A wireless acoustic sensor network

$$\max_{\mathbf{D}_1,...,\mathbf{D}_N} \text{SNR}(\mathbf{H}_{ST}, \mathbf{D}_1, ..., \mathbf{D}_N) \quad s.t. \quad \sum_{i=1}^{N} \alpha_i R(\mathbf{D}_i) = R, \tag{21}$$

The set of weights $\alpha_1, ..., \alpha_N$ could be chosen for example to equalize the power consumption at the nodes, or to give the network sum-rate by simply being set equal to 1, or to satisfy whatever other criteria specified by the user. Notice that this problem is beyond the mere allocation of the rates to the nodes (although including it), since rate allocation is equivalent to the specification of the determinant of the distortion matrices $\mathbf{D}_1, ..., \mathbf{D}_N$ according to (12), while in (21) the whole matrices $\mathbf{D}_1, ..., \mathbf{D}_N$ must be chosen optimally to shape the covariance matrix $\mathbf{\Sigma}_{\mathbf{v}}(\mathbf{D}_1, ..., \mathbf{D}_N)$ of the noise in the desired way to maximize the output SNR. This possibility is a result of using vector source coding with covariance matrix distortion constraints.

Problem (21) is hard to solve analytically. In the sequel, we make some simplifying assumptions enabling us to study the KKT conditions for the solution of the optimization problem. In particular, we assume that the additive noise terms $\mathbf{n}_i,\ i = 1, ..., N$ in the measurements are mutually uncorrelated.[3] We further assume that the distortions $\mathbf{D}_i,\ i = 1, ..., N$ are in the interval $\mathbf{0} \prec \mathbf{D}_i \preceq \mathbf{\Sigma}_{\mathbf{y}_i}$, so the rate-distortion functions are given by (12). The problem (21) can then be rewritten as:

$$\min_{\mathbf{D}_1,...,\mathbf{D}_N} \text{tr}\left\{ \left[\mathbf{W} \left(\mathbf{\Sigma}_{\mathbf{v}}(\mathbf{D}_1, ..., \mathbf{D}_N) \right)^{-1} \mathbf{W}^T \right]^{-1} \right\} \quad s.t. \quad \prod_{i=1}^{N} |\mathbf{D}_i|^{\alpha_i} = \beta, \tag{22}$$

where $\beta = e^{-2R} \prod_{i=1}^{N} |\mathbf{\Sigma}_{\mathbf{y}_i}|^{\alpha_i}$. From (18) and (19), we have $\mathbf{\Sigma}_{\mathbf{v}} = \mathbf{\Sigma}_{\mathbf{n}} + \mathbf{\Sigma}_{\boldsymbol{\nu}}$, and thus from the assumption that the additive noise terms at different nodes are mutually uncorrelated, we conclude that $\mathbf{\Sigma}_{\mathbf{v}}$ is block-diagonal, and from (17), the ith matrix on the diagonal

[3] Even if the noise terms at different nodes come from the same source, in presence of room effects the recorded versions of the noise at different microphones would have experienced different impulse responses, leading to reduced correlations.

is given by:

$$\mathbf{\Sigma}_{\mathbf{v}_i} = \mathbf{\Sigma}_{\mathbf{n}_i} + \left(\mathbf{D}_i^{-1} - \mathbf{\Sigma}_{\mathbf{y}_i}^{-1}\right)^{-1}, \tag{23}$$

from which (22) can be written as:

$$\min_{\mathbf{D}_1,...,\mathbf{D}_N} \text{tr}\left\{\left(\sum_{i=1}^{N} \mathbf{W}_i\left[\mathbf{\Sigma}_{\mathbf{n}_i} + \left(\mathbf{D}_i^{-1} - \mathbf{\Sigma}_{\mathbf{y}_i}^{-1}\right)^{-1}\right]^{-1}\mathbf{W}_i^T\right)^{-1}\right\} \quad s.t. \quad \prod_{i=1}^{N}|\mathbf{D}_i|^{\alpha_i} = \beta, \tag{24}$$

Assuming that $\mathbf{W}_i,\ i = 1, ..., N$ are invertible, the KKT conditions for (24) leads to the following set of equations:

$$\alpha_i \mathbf{W}_i^{-T}\left[\mathbf{X}_i^{-1} + (\mathbf{C}_i - \mathbf{X}_i)^{-1}\right]\mathbf{W}_i^{-1} = \mathbf{K}, \quad i = 1, ..., N, \tag{25}$$

where $\mathbf{K}$ is a fixed matrix, and $\mathbf{X}_i$ and $\mathbf{C}_i$ are defined as:

$$\mathbf{X}_i = \mathbf{\Sigma}_{\mathbf{n}_i}^{-1}\left(\mathbf{\Sigma}_{\mathbf{n}_i}^{-1} + \mathbf{D}_i^{-1} - \mathbf{\Sigma}_{\mathbf{y}_i}^{-1}\right)^{-1}\mathbf{\Sigma}_{\mathbf{n}_i}^{-1}, \tag{26}$$

$$\mathbf{C}_i = \mathbf{\Sigma}_{\mathbf{n}_i}^{-1}\left(\mathbf{\Sigma}_{\mathbf{n}_i}^{-1} - \mathbf{\Sigma}_{\mathbf{y}_i}^{-1}\right)^{-1}\mathbf{\Sigma}_{\mathbf{n}_i}^{-1}. \tag{27}$$

Since $\mathbf{D}_i$ can be calculated from $\mathbf{X}_i$ (and vice versa) using (26), one can use (25)-(27) together with the constraint in (24), and $\mathbf{0} \prec \mathbf{D}_i \preceq \mathbf{\Sigma}_{\mathbf{y}_i},\ i = 1, ..., N$ to allocate the optimal distortion matrices. We leave out the derivation of (25) due to space limitations. See [9] as a useful guide for differentiating functions of matrices.

IV. Scalar Case

In this section, we consider a special case of the problem described in Section III, where $n = 1$. We denote by D_i and w_i the scalar version of the matrices $\mathbf{D}_i$ and $\mathbf{W}_i$, respectively. We use the symbol Σ followed by an appropriate subscript to denote the variance of random variables. To further simplify the problem for analytical derivations, we assume that there are only 2 nodes, and $\alpha_1 = \alpha_2$ (which means that we are interested in the sum-rate), and $w_1 = w_2$. Although for most applications this case may not be of practical interest, it leads to intuitive results which might help to understand the more complicated cases.

Applying the above assumptions to (25) and using the constraint in (24), one can show that the target distortions D_1^* and D_2^* are given by:

$$D_1^* = \sqrt{\beta}\frac{\Sigma_{n_1}}{\Sigma_{n_2}}\left(\frac{\Sigma_{n_1}\Sigma_{n_2} - \Sigma_2\sqrt{\beta}}{\Sigma_{n_1}\Sigma_{n_2} - \Sigma_1\sqrt{\beta}}\right), \tag{28}$$

$$D_2^* = \sqrt{\beta}\frac{\Sigma_{n_2}}{\Sigma_{n_1}}\left(\frac{\Sigma_{n_1}\Sigma_{n_2} - \Sigma_1\sqrt{\beta}}{\Sigma_{n_1}\Sigma_{n_2} - \Sigma_2\sqrt{\beta}}\right), \tag{29}$$

where Σ_1 and Σ_2 are given by:

$$\Sigma_i = \Sigma_{n_i} - \frac{{\Sigma_{n_i}}^2}{\Sigma_{y_i}}, \quad i = 1, 2. \tag{30}$$

Defining the following critical sum-rates:

$$R_1^c = \frac{1}{2} \log \frac{\left[\max(\Sigma_1, \Sigma_2)\right]^2}{(\Sigma_{n_1} - \Sigma_1)(\Sigma_{n_2} - \Sigma_2)}, \tag{31}$$

$$R_2^c = \frac{1}{2} \log \frac{\left[\min(\Sigma_1, \Sigma_2)\right]^2}{(\Sigma_{n_1} - \Sigma_1)(\Sigma_{n_2} - \Sigma_2)}, \tag{32}$$

we prove the following proposition.

Proposition 4.1: The solution (28)-(29) is the unique maximizer of the output SNR, if $R > R_1^c$, and is the unique minimizer, if $R < R_2^c$.

Proof: Reducing the cost function in (24) to the scalar case for two nodes and using the assumptions $\alpha_1 = \alpha_2$ and $w_1 = w_2$, the resulting problem will be equivalent to the following:

$$\min_{D_1, D_2} \frac{1}{\Sigma_{n_1}} - \frac{1}{{\Sigma_{n_1}}^2} \left(\frac{1}{\frac{1}{\Sigma_{n_1}} + \frac{1}{D_1} - \frac{1}{\Sigma_{y_1}}} \right) + \frac{1}{\Sigma_{n_2}} - \frac{1}{{\Sigma_{n_2}}^2} \left(\frac{1}{\frac{1}{\Sigma_{n_2}} + \frac{1}{D_2} - \frac{1}{\Sigma_{y_2}}} \right) \quad s.t. \;\; D_1 D_2 = \beta. \tag{33}$$

Removing the fixed terms $\frac{1}{\Sigma_{n_1}}$ and $\frac{1}{\Sigma_{n_2}}$ and using the following change of variables:

$$D_i' = \Sigma_i + \frac{{\Sigma_{n_i}}^2}{D_i}, \quad i = 1, 2, \tag{34}$$

problem (33) can be rewritten as:

$$\min_{D_1', D_2'} \left(\frac{1}{D_1'} + \frac{1}{D_2'} \right) \quad s.t. \quad (D_1' - \Sigma_1)(D_2' - \Sigma_2) = \beta', \tag{35}$$

where β' is defined as:

$$\beta' = (\Sigma_{n_1} - \Sigma_1)(\Sigma_{n_2} - \Sigma_2) e^{2R}. \tag{36}$$

Writing D_2' in terms of D_1' using the constraint in (35) and substituting in the cost function, we obtain an unconstrained function of D_1' referred to as $f_1(D_1')$. Differentiating this function with respect to D_1' yields:

$$\frac{df_1(D_1')}{dD_1'} = \frac{\left[\sqrt{\beta'} D_1' + \beta' + \Sigma_2 (D_1' - \Sigma_1)\right] \left[(\sqrt{\beta'} - \Sigma_2) D_1' - (\beta' - \Sigma_1 \Sigma_2)\right]}{{(D_1')}^2 \left(\Sigma_2 D_1' - \Sigma_1 \Sigma_2 + \beta'\right)^2}, \tag{37}$$

which is zero at $D_1' = \Sigma_1 + \frac{{\Sigma_{n_1}}^2}{D_1^*}$ (equivalent to $D_1 = D_1^*$ using (34)). We study the sign of the derivative around this point. Since the whole denominator and the first term in the numerator in (37) are positive, we consider only the second term in the numerator and replace D_1' by $\Sigma_1 + \frac{{\Sigma_{n_1}}^2}{D_1^*} + \epsilon$. The result is $(\sqrt{\beta'} - \Sigma_2)\epsilon$, which means that if $\sqrt{\beta'} > \Sigma_2$, the derivative is negative for $\epsilon < 0$ and positive for $\epsilon > 0$. Therefore if $\sqrt{\beta'} > \Sigma_2$, the global minimum is at $D_1' = \Sigma_1 + \frac{{\Sigma_{n_1}}^2}{D_1^*}$, or equivalently the global maximizer of the SNR is at $D_1 = D_1^*$. Similarly, one can write D_1' in terms of D_2' and substitute in the cost function to obtain the function $f_2(D_2')$, and then study the sign of the derivative around $D_2 = D_2^*$. This leads to the condition $\sqrt{\beta'} > \Sigma_1$ for having the SNR maximized. Combining the two conditions, the following should hold to have a maximized SNR at $D_1 = D_1^*$ and $D_2 = D_2^*$:

$$\sqrt{\beta'} > \max(\Sigma_1, \Sigma_2). \tag{38}$$

Using (36), (38) can be written as:

$$R > \frac{1}{2} \log \frac{\left[\max(\Sigma_1, \Sigma_2)\right]^2}{(\Sigma_{n_1} - \Sigma_1)(\Sigma_{n_2} - \Sigma_2)} = R_1^c. \tag{39}$$

Similarly one can show that for $R < R_2^c$ the output SNR is minimized. This completes the proof.

■

As an example, let $\Sigma_{y_1} = 1.15$, $\Sigma_{n_1} = 0.15$, $\Sigma_{y_2} = 1.1$, $\Sigma_{n_2} = 0.1$, $\Sigma_{x_r} = 1$, and $w_1 = w_2 = 1$. Then we have $R_1^c = 2.28$ and $R_2^c = 1.92$. We plot $f_1(D_1)$ and $f_2(D_2)$ for three different values of R, namely $1, 2, 3$. The result is shown in Fig. 3. From Fig. 3(a) it is seen that for $R = 1 < R_2^c$ the output SNR at $D_1 = D_1^*$ and $D_2 = D_2^*$ is minimized, which is not desired. The optimal choice of the distortions should then be on the boundary, which means that the whole rate should be allocated to one of the nodes. From the figure, it can be seen that the highest SNR is achieved when D_2 is as small as possible. Thus the whole rate should be allocated to node 2, which is in agreement with intuition, since $\Sigma_{n_2} < \Sigma_{n_1}$. For $R_2^c < R < R_1^c$ as shown in Fig. 3 (b), $(D_1 = D_1^*, D_2 = D_2^*)$ is not a feasible solution. The optimal solution is again on the boundary similar to the previous case. Finally, for $R = 3 > R_1^c$, $(D_1 = D_1^*, D_2 = D_2^*)$ indeed maximizes the output SNR. In this case, we have $(D_1^* = 0.1082, D_2^* = 0.029)$ corresponding to $R(D_1^*) = 1.182, R(D_2^*) = 1.818$, which means that higher rate should be spent at node 2.

V. Conclusions

We considered a wireless acoustic sensor network with centralized topology and formulated a novel problem which relates source coding to noise reduction. To encode the measurements at the nodes, we considered vector Gaussian source coding with covariance matrix distortion constraints. For noise reduction we applied a spatio-temporal prediction filter to the set of the messages received at the center. The objective was then to determine the optimal target distortion matrices for each of the nodes such that the resulting SNR at the center is maximized while for a given set of weights, the weighted sum-rate of the

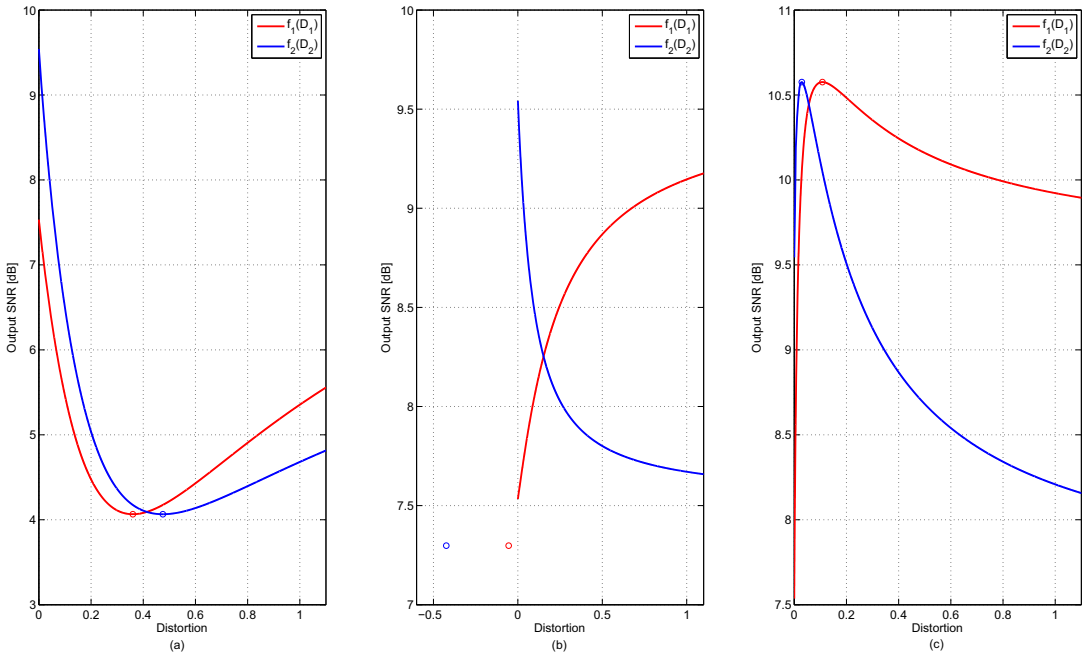

Figure 3: An example of the output SNR as a function of D_1 and D_2 for the scalar case with two nodes for three different values of the network sum-rate, (a) $R < R_2^c$, (b) $R_2^c < R < R_1^c$, and (c) $R > R_1^c$. The circles show the stationary points suggested by the KKT conditions.

network does not go beyond a prespecified value. For the resulting optimization problem we studied the KKT conditions and derived the optimality conditions for the distortion matrices. We further derived the closed-form solution for the special case with scalar sources and two nodes, leading to the intuitive result that the derived distortion values uniquely maximize the output SNR, if the prespecified sum-rate is higher that a critical value.

REFERENCES

[1] A. Bertrand, *Applications and trends in wireless acoustic sensor networks: a signal processing perspective*, in Proc. IEEE Symposium on Communications and Vehicular Technology (SCVT), Nov. 2011.

[2] J. Østergaard and M. S. Derpich, *Sequential remote source coding in wireless acoustic sensor networks*, European Signal Processing Conference (Eusipco), pp. 1269-1273, Aug. 2012.

[3] A. Zahedi, J. Østergaard, S. H. Jensen, P. Naylor, and S. Bech, *Distributed remote vector Gaussian source coding for wireless acoustic sensor networks*, IEEE Data Compression Conference, Snowbird, UT, Mar. 2014.

[4] 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.

[5] A. Majumdar, K. Ramchandran, and I. Kozintsev, *Distributed coding for wireless audio sensors,* IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA, Oct. 2003.

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

[7] 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, June-July 2014.

[8] M. S. Derpich, J. Østergaard, and G. C. Goodwin, *The Quadratic Gaussian Rate-Distortion Function for Source Uncorrelated Distortions*, IEEE Data Compression Conference (DCC), Snowbird, UT, pp. 73 - 82, Mar. 2008.

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