Rate Allocation in Distributed Sensor Network*

Jun Chen Xin Zhang Toby Berger Stephen B. Wicker

School of Electrical and Computer Engineering
Cornell University
Ithaca, NY 14853
{jc353, xz56}@cornell.edu
{berger, wicker}@ece.cornell.edu

Abstract

We consider a distributed sensing system in which several observations are communicated to the fusion center using limited transmission rate. The observations must be separately encoded so that the target can be estimated with minimum average distortion. We get an achievable region of rates and average distortion. The quadratic Gaussian case is discussed in detail and the results are applied to the quadratic Gaussian CEO problem to derive an upper bound on the sum-rate distortion function.

I. Introduction

In this paper, we consider the following distributed sensing system (see Fig.1).

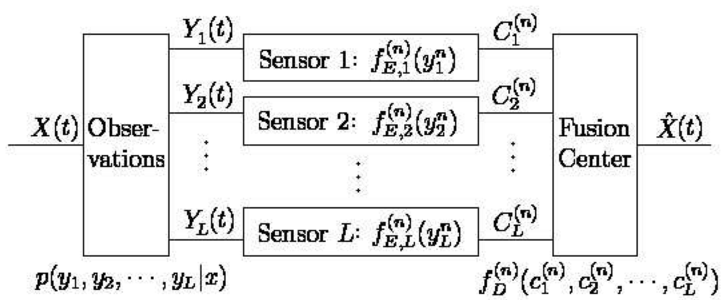

Figure 1: Model of distributed sensing system

$\{X(t)\}_{t=1}^{\infty}$ is the target data sequence that the fusion center is interested in. This data sequence cannot be observed directly. L sensors are deployed, which observe independently corrupted versions of $\{X(t)\}_{t=1}^{\infty}$. The data rate at which sensor i $(i = 1, 2, \cdots, L)$ may communicate information about its observations to the fusion center is limited to R_i bits per second [1]. Due to wide geographical separation of the sensors or other reasons, the

*This research is supported in part by NSF Grant No. CCR-9980616, ARO P-40116-PH-MUR, DAAD 19-99-1-0215 and the Multidisciplinary University Research Initiative (MURI) under the Office of Naval Research Contract N00014-00-1-0564.

[1]The rate constraints may come from the restrictions on the resources, say bandwidth, power, etc., that are available to sensors

sensors are not permitted to communicate with each other, i.e. sensor i has to send data based solely on its own noisy observations $\{Y_i(t)\}_{t=1}^{\infty}$. Finally, the decision $\{\hat{X}(t)\}_{t=1}^{\infty}$ is computed from the combined data at fusion center. This model has been studied in [1] in the case of two sensors. So part of our work can be viewed as a direct generalization of [1]. A closely related problem, called CEO problem, was discussed in [2] for discrete case, and [3, 4] for quadratic gaussian case. Here we mainly follow the notations adopted in [3].

Let $\{X(t), Y_1(t), \cdots, Y_L(t)\}_{t=1}^{\infty}$ be a temporally memoryless source with instantaneous joint probability distribution $P(x, y_1, y_2, \cdots, y_L)$ on $\mathcal{X} \times \mathcal{Y}^L$, where L is the number of sensors, $\mathcal{X}$ is the common alphabet of the random variables $X(t)$ for $t = 1, 2, \cdots$, $\mathcal{Y}$ is the common alphabet of the random variables $Y_i(t)$ for $i = 1, \cdots, L;\ t = 1, 2, \cdots$. In this paper, we let U^n denote random vector $[U(1), U(2), \cdots, U(n)]$, and $u^n = [u(1), u(2), \cdots, u(n)]$ be its realization.

For $i = 1, \cdots, L$, sensor i encodes a block $y_i^n = [y_i(1), \cdots, y_i(n)]$ of length n from its observed data using a source code $c_i^{(n)} = f_{E,\,i}^{(n)}(y_i^n)$ of rate $R_i^{(n)} = \frac{1}{n}\log|\mathcal{C}_i^{(n)}|$. The codewords from the L sensors, $c_1^{(n)}, \cdots, c_L^{(n)}$, are sent to the fusion center. The task of fusion center is to recover the target data sequence $x^n = [x(1), \cdots, x(n)]$ with small expected distortion defined as $d^{(n)} = \frac{1}{n}Ed(X^n, \hat{X}^n)$, where $d(x, \hat{x})$ is a given distortion measure and $\hat{X}^n$ is the estimate of random target sequence X^n. The fusion center implements a mapping $f_D^{(n)}:\ \mathcal{C}_1^{(n)} \times \cdots \times \mathcal{C}_L^{(n)} \to \mathcal{X}^n$, i.e. the estimate at fusion center is of the form $\hat{x}^n = f_D^{(n)}(c_1^{(n)}, \cdots, c_L^{(n)})$.

The rest part of this paper is divided into four sections. In Section II we use multiterminal source coding techniques to find an achievable region of rates and average distortion $(R_1, R_2, \cdots, R_L, D)$. The proof of achievability is outlined, with the detailed proof left to the Appendix. In Section III, we evaluate this region for correlated memoryless Gaussian observations and squared distortion measure. The results are shown to coincide with known results in special cases. In Section IV, the results got in Section III are applied to study quadratic gaussian CEO problem. Areas for extension of the results are suggested in Section V which serves as conclusion.

II. Achievable Region

Definition: A $(L{+}1)$-tuple of rates and distortion $(R_1, R_2, \cdots, R_L, D)$ is said achievable if for any $\varepsilon > 0$, there exists an n_0 such that for $n > n_0$ there exist encoders:

$$\begin{array}{ll} f_{E,1}^{(n)} : \mathcal{Y}_1^n \to \mathcal{C}_1^{(n)} & \log|\mathcal{C}_1^{(n)}| \le n(R_1 + \varepsilon) \\ f_{E,2}^{(n)} : \mathcal{Y}_2^n \to \mathcal{C}_2^{(n)} & \log|\mathcal{C}_2^{(n)}| \le n(R_2 + \varepsilon) \\ \vdots & \vdots \\ f_{E,L}^{(n)} : \mathcal{Y}_L^n \to \mathcal{C}_L^{(n)} & \log|\mathcal{C}_L^{(n)}| \le n(R_L + \varepsilon) \end{array}$$

and a decoder:

$$f_D^{(n)} : \mathcal{C}_1^{(n)} \times \mathcal{C}_2^{(n)} \times \cdots \times \mathcal{C}_L^{(n)} \to \mathcal{X}^n$$

such that $\frac{1}{n}E[\sum_{t=1}^{n} d(X(t), \hat{X}(t))] \le D + \varepsilon$.

Theorem: Given the joint distribution of the discrete random variables $(X, Y_1, Y_2, \cdots, Y_L)$: $P(x, y_1, y_2, \cdots, y_L)$ and the bounded distortion measure $d : \mathcal{X} \times \mathcal{X} \to [0, d_{max}]$, a rates-distortion $(L+1)$-tuple $(R_1, R_2, \cdots, R_L, D)$ is said achievable if there exist random variables $W_1, W_2, \cdots, W_L$ on spaces $\mathcal{W}_1, \mathcal{W}_2, \cdots, \mathcal{W}_L$ respectively with $W_i \to Y_i \to (X, Y_{\mathcal{I}_L \setminus \{i\}}, W_{\mathcal{I}_L \setminus \{i\}})$ for all $i \in \mathcal{I}_L$; and a function $g : \mathcal{W}_1 \times \mathcal{W}_2 \times \cdots \times \mathcal{W}_L \to \mathcal{X}$ such that:

$$\sum_{i \in \mathcal{A}} R_i \geq I(Y_{\mathcal{A}}; W_{\mathcal{A}} | W_{\mathcal{I}_L \setminus \mathcal{A}}) \tag{1}$$

$$Ed(X, \hat{X}) \leq D \tag{2}$$

for any $\mathcal{A} \subseteq \mathcal{I}_L$, where $\hat{X} = g(W_1, W_2, \cdots, W_L)$. Here:

(i) The notion $A \to B \to C$ means that A, B, C form a Markov chain;

(ii) $\mathcal{I}_L \triangleq \{1, 2, \cdots, L\}$;

(iii) $\mathcal{I}_L \setminus \mathcal{A} \triangleq \mathcal{I}_L \bigcap \mathcal{A}^c$;

(iv) If $\mathcal{B} = \{i_1, i_2, \cdots, i_k\}$, then $W_{\mathcal{B}} \triangleq (W_{i_1}, W_{i_2}, \cdots, W_{i_k})$. Specially, $W_\emptyset \triangleq \emptyset$ and $I(U; V|\emptyset) \triangleq I(U; V)$.

Moreover, if T is the region of $(L+1)$-tuple satisfying the above conditions, then any $(L+1)$-tuple in the convex hull of T is achievable[2].

Outline of Proof:
Our proof closely follows the proof in [7] which is based on strong typicality and Cover's binning technique [8].
Let $(W_1, W_2, \cdots, W_L)$ and function g satisfy the conditions given in the theorem. Construct the random codebooks $\{\vec{\mathcal{C}}^{(n)} = (\mathcal{C}_1^{(n)}, \mathcal{C}_2^{(n)}, \cdots, \mathcal{C}_L^{(n)})\}$[3] (where $\mathcal{C}_i^{(n)}$ denotes the codebook of encoder i) as follows:
At encoder i, generate M_i i.i.d. codewords W_i^n according to $\prod_{l=1}^{n} p(w_i(l))$ and index them $W_i^n(j), j = 1, 2, \cdots, M_i$. Let $\mathcal{C}_i^{(n)} = \{W_i^n(j)\}_{j=1}^{M_i}$. Randomly assign the indices of the codewords to one of 2^{nR_i} bins[4] using a uniform distribution over the indices of the bins such that every bin contains $N_i = M_i 2^{-nR_i}$ codewords . For the simplicity of future proof, we suppose every generated W_i^n satisfy strong typicality. By weak law of large numbers, this assumption holds with probability close to one when n is large enough. Or we can simply draw W_i^n from strongly typical set to ensure this assumption.

The coding and decoding procedures are described as follows:
Encoding Scheme:
The coding procedure is divided into two operations, called precoding and coding.
At Encoder i, given observations y_i^n, if it's typical, map it onto the $w_i^n(j) \in \mathcal{C}_i^{(n)}$ with the smallest index j such that $(y_i^n, w_i^n(j))$ is jointly typical. Let $w_i^n(y_i^n)$ denotes the w_i^n onto which y_i^n is mapped.
For coding, the index of the bin which contains $w_i^n(y_i^n)$ is sent. Let $b_i(y_i^n)$ denote this

[2]This follows through a time-sharing argument.

[3]Here $\mathcal{C}_i^{(n)}$ actually is not the $\mathcal{C}_i^{(n)}$ stated in the definition at the end of last page. As we will see, we will not send the codewords in $\mathcal{C}_i^{(n)}$ directly. Instead, we will send the index of bin. That's why we have 2^{nR_i} bins at encoder i, while generally $|\mathcal{C}_i^{(n)}| > 2^{nR_i}$.

[4]We suppose 2^{nR_i} is an integer. When n is large enough, this assumption causes no essential loss.

bin index. If y_i^n is not typical or there does not exist $w_i^n(j) \in \mathcal{C}_i^{(n)}$ such that $(y_i^n, w_i^n(j))$ is jointly typical, then a special error symbol is sent. This special error symbol does not increase the rate R_i in the limit of large n, so we may safely ignore it.

Decoding Scheme:

Given $(b_1, b_2, \cdots, b_L)$, if there exists a unique $(w_1^n, w_2^n, \cdots, w_L^n)$ such that $w_i^n \in \mathcal{B}_i(b_i)$ (where $\mathcal{B}_i(b_i)$ denotes the bin with index b_i at the encoder i) and $(w_1^n, w_2^n, \cdots, w_L^n)$ is jointly typical, then call it $(\hat{w}_1^n, \hat{w}_2^n, \cdots, \hat{w}_L^n)$; otherwise declare an error and incur the maximum distortion d_{max}. If the received vector contains special error symbol, also declare an error and incur the maximum distortion d_{max}. Assuming no error, produce the estimate $\hat{x}(k) = g(\hat{w}_1(k), \hat{w}_2(k), \cdots, \hat{w}_L(k))$ for $k = 1, 2, \cdots, n$.

See appendix for detailed analysis of the probability of decoding error.

III. Gaussian Case[5]

For simplicity, in this section, we consider two-sensor case. Let $\{X(t), Y_1(t), Y_2(t)\}_{t=1}^{\infty}$ be i.i.d Gaussian vectors with $p(y_1, y_2|x) = p_1(y_1|x) \cdot p_2(y_2|x)$, i.e. Y_1 and Y_2 are independent conditioning on X. We let two auxiliary random variables W_1, W_2 be joint Gaussian with X, Y_1, Y_2, which makes it possible to get explicit formulae in quadratic Gaussian case. As usual, squared distortion measure is used here.

A. Parametric Representation

Since $W_1 \to (Y_1, Y_2) \to W_2$ and $Y_1 \to X \to Y_2$, we can get the following two equations[6]:

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} X + \begin{pmatrix} N_1 \\ N_2 \end{pmatrix} \tag{3}$$

$$\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} = \begin{pmatrix} l_1 & 0 \\ 0 & l_2 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} + \begin{pmatrix} T_1 \\ T_2 \end{pmatrix} \tag{4}$$

where N_1 and N_2 are independent Gaussian noises at the two sensors with variance $\sigma_{N_1}^2$ and $\sigma_{N_2}^2$ respectively; l_1 and l_2 are two scalars; T_1 and T_2 are two mutually independent Gaussian random variables with variance $\sigma_{T_1}^2$ and $\sigma_{T_2}^2$. (N_1, N_2) are independent of X and (T_1, T_2) are independent of (Y_1, Y_2).

1. Distortion:

   We rewrite (3) in the form:

   $$X = (a_1 \quad a_2) \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} + V$$

   where $(a_1\ a_2) = R_{YX}^T R_{YY}^{-1}$ and V is a Gaussian r.v. with variance $\sigma_V^2 = \sigma_X^2 - R_{YX}^T R_{YY}^{-1} R_{YX}$ and independent of (Y_1, Y_2). Due to the fact that $X \to (Y_1, Y_2) \to$

[5] Although we only prove the theorem in discrete case and bounded distortion measure. Our result can be extended straightforward to Gaussian case and squared distortion measure by some standard techniques [5], [6]. Specifically, the Markov lemma which is fundamental in this proof has been generalized by [6] to Gaussian case.

[6] We can also let $\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} k_1 \\ k_2 \end{pmatrix} X + \begin{pmatrix} N_1 \\ N_2 \end{pmatrix}$. But since all are zero-mean, (3) can always be got by scaling. In the case when $k_1 = 0$ (or $k_2 = 0$), we can let $\sigma_{N_1} = \infty$ (or $\sigma_{N_2} = \infty$) in (3).

(W_1, W_2), (T_1, T_2) and V are independent. In the Gaussian case, the optimal estimate of X from (W_1, W_2), i.e. $E(X|W_1, W_2)$, is linear MMSE estimate. So we have:

$$\hat{X}(W_1, W_2) = R_{WX}^T R_{WW}^{-1} \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}, \quad E(X - \hat{X}(W_1, W_2))^2 = \sigma_X^2 - R_{WX}^T R_{WW}^{-1} R_{WX}$$

Substitute the covariances in it, we get:

$$\frac{1}{D} \le \frac{1}{\sigma_X^2} + \frac{\mu_1^2}{\mu_1^2 \sigma_{N_1}^2 + 1} + \frac{\mu_2^2}{\mu_2^2 \sigma_{N_2}^2 + 1} \tag{5}$$

where $\mu_i = \frac{l_i}{\sigma_{T_i}^2}$, for $i = 1, 2$.

Clearly, a nontrivial D should be in the range of $D_0 \le D \le \sigma_X^2$, where $D_0 = \frac{1}{(\frac{1}{\sigma_X^2} + \frac{1}{\sigma_{N_1}^2} + \frac{1}{\sigma_{N_2}^2})}$, which is the MMSE of X given (Y_1, Y_2).

2. Rates:

   The theorem gives:

$$\begin{aligned} R_1 \ge I(Y_1; W_1|W_2) \qquad & R_2 \ge I(Y_2; W_2|W_1) \\ R_1 + R_2 \ & \ge \ I(Y_1, Y_2; W_1, W_2) \\ \text{subject to: } E(X - \hat{X}(W_1, W_2))^2 \ & \le \ D. \end{aligned}$$

   For joint Gaussian random vectors $\vec{X}$, $\vec{Y}$, $\vec{Z}$, we have:

$$I(\vec{X}; \vec{Y}) = \frac{1}{2} \log^+ \frac{\det R_{\vec{X}} \det R_{\vec{Y}}}{\det R_{\vec{X}\vec{Y}}}; \quad I(\vec{X}; \vec{Y}|\vec{Z}) = \frac{1}{2} \log^+ \frac{\det R_{\vec{X}\vec{Z}} \det R_{\vec{Y}\vec{Z}}}{\det R_{\vec{X}\vec{Y}\vec{Z}} \det R_{\vec{Z}}}$$

   So, we get:

$$R_1 \ \ge \ \frac{1}{2} \log^+ \frac{(\mu_1^2 \sigma_{Y_1}^2 + 1)(\mu_2^2 \sigma_{Y_2}^2 + 1) - \mu_1^2 \mu_2^2 \rho^2 \sigma_{Y_1}^2 \sigma_{Y_2}^2}{\mu_2^2 \sigma_{Y_2}^2 + 1} \tag{6}$$

$$R_2 \ \ge \ \frac{1}{2} \log^+ \frac{(\mu_1^2 \sigma_{Y_1}^2 + 1)(\mu_2^2 \sigma_{Y_2}^2 + 1) - \mu_1^2 \mu_2^2 \rho^2 \sigma_{Y_1}^2 \sigma_{Y_2}^2}{\mu_1^2 \sigma_{Y_1}^2 + 1} \tag{7}$$

$$R_1 + R_2 \ \ge \ \frac{1}{2} \log^+ [(\mu_1^2 \sigma_{Y_1}^2 + 1)(\mu_2^2 \sigma_{Y_2}^2 + 1) - \mu_1^2 \mu_2^2 \rho^2 \sigma_{Y_1}^2 \sigma_{Y_2}^2] \tag{8}$$

   subject to (5), where $R_{Y_1 Y_2} = \begin{pmatrix} \sigma_{Y_1}^2 & \rho \sigma_{Y_1} \sigma_{Y_2} \\ \rho \sigma_{Y_1} \sigma_{Y_2} & \sigma_{Y_2}^2 \end{pmatrix}$ and $\rho = \frac{E(XY)}{\sigma_{Y_1} \sigma_{Y_2}}$. Note that we always have:

$$\begin{aligned} & \frac{1}{2} \log^+ \frac{(\mu_1^2 \sigma_{Y_1}^2 + 1)(\mu_2^2 \sigma_{Y_2}^2 + 1) - \mu_1^2 \mu_2^2 \rho^2 \sigma_{Y_1}^2 \sigma_{Y_2}^2}{\mu_2^2 \sigma_{Y_2}^2 + 1} \\ + \ & \frac{1}{2} \log^+ \frac{(\mu_1^2 \sigma_{Y_1}^2 + 1)(\mu_2^2 \sigma_{Y_2}^2 + 1) - \mu_1^2 \mu_2^2 \rho^2 \sigma_{Y_1}^2 \sigma_{Y_2}^2}{\mu_1^2 \sigma_{Y_1}^2 + 1} \\ \le \ & \frac{1}{2} \log^+ [(\mu_1^2 \sigma_{Y_1}^2 + 1)(\mu_2^2 \sigma_{Y_2}^2 + 1) - \mu_1^2 \mu_2^2 \rho^2 \sigma_{Y_1}^2 \sigma_{Y_2}^2], \end{aligned}$$

   so whenever (R_1, R_2) is optimal, (8) will be an equality, while the equalities in (6) and (7) generally can not be satisfied simultaneously (except when $\mu_1 \mu_2 \rho = 0$).Similarly, due to the continuity of the distortion measure, (5) will turn out to be an equality whenever (R_1, R_2) is optimal. Since we are only interested in the boundary of the achievable region, we will view (5), (8) as equations later on.

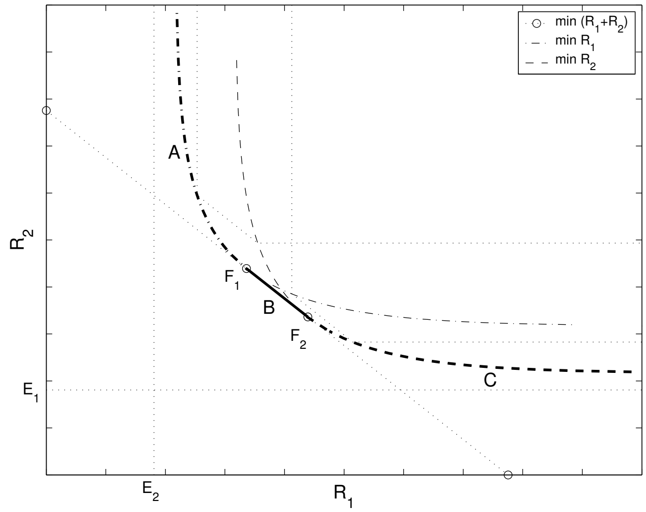

Figure 2: Achievable Region

The thick dash-dot line shows the boundary given by (5), (6), (8) and the thick dash line (5), (7), (8). The solid line is where $\min(R_1 + R_2)$ lies. The lines A, B, C together define the boundary of the achievable rates (R_1, R_2) for a fixed distortion D. The region above the thin dotted lines is achievable (R_1, R_2) for some pair of (μ_1^2, μ_2^2) that satisfies (5). The lower convex envelop for all of them will be exactly the curve formed by A, B and C.

B. Special Case Study

1. Gaussian Multiterminal Source Coding with one distortion constraint:

   If we let $Y_1 = X$, the problem is reduced to that discussed in [6]. Apply $\sigma_{Y_1}^2 = \sigma_X^2$, $\sigma_{Y_2}^2 = \sigma_X^2 + \sigma_{N_2}^2$, $\rho^2 = \frac{\sigma_X^2}{\sigma_X^2+\sigma_{N_2}^2}$ and $R_2 = \frac{1}{2}\log^+(\mu_2^2\sigma_{Y_2}^2 + 1)$ to (5), (6) and (8), we get:

$$R_1(D) \geq \frac{1}{2}\log^+ \frac{\sigma_X^2}{D}(1 - \rho^2 + \rho^2 2^{-2R_2}).$$

2. Noisy Wyner-Ziv problem:

   The successively structured CEO problem [9] is formulated as the data fusion is done in a serial fashion. Sensors communicate one to the next over rate-constraint channels. Each sensor does the optimal encoding to the combination of the received data from the previous sensor and its own observed data, which is exactly the case when setting $W_2 = Y_2$, i.e. $\mu_2^2 = \infty$ in our problem. The main theorem in [9] is termed "noisy" Wyner-Ziv problem. It's given in [9] that:

$$R_{DW}(D) = \frac{1}{2}\log^+ \frac{\sigma_{X|Y_2}^2 - \sigma_{X|Y_1Y_2}^2}{D - \sigma_{X|Y_1Y_2}^2}$$

   Represent $\sigma_{X|Y_2}^2$ and $\sigma_{X|Y_1Y_2}^2$ by σ_{N_1}, σ_{N_2} and σ_x, we get:

$$R_{DW}(D) = \frac{1}{2}\log^+ \frac{\sigma_X^4\sigma_{N_2}^4}{(\sigma_X^2 + \sigma_{N_2}^2)[D(\sigma_X^2\sigma_{N_1}^2 + \sigma_X^2\sigma_{N_2}^2 + \sigma_{N_1}^2\sigma_{N_2}^2) - \sigma_X^2\sigma_{N_1}^2\sigma_{N_2}^2]}$$

   While our solution gives:

$$R_1 = \frac{1}{2}\log^+[\mu_1^2(\sigma_{Y_1}^2 - \frac{\rho^2\sigma_{Y_1}^2\sigma_{Y_2}^2}{\sigma_{Y_2}^2}) + 1], \text{ subject to } \frac{1}{D} = \frac{1}{\sigma_X^2} + \frac{\mu_1^2}{\mu_1^2\sigma_{N_1}^2 + 1} + \frac{1}{\sigma_{N_2}^2}$$

It's easy to verify that $R_1(D) = R_{DW}(D)$. It is worth noting that $R_1(D)$ corresponds to point E_2 in Fig. 2.

By further setting $Y_1 = X$, we get $R_1(D) = R_{DW}(D) = \frac{1}{2}\log^+ \frac{\sigma_X^2\sigma_{N_2}^2}{(\sigma_X^2+\sigma_{N_2}^2)D}$, which is a classical result first discovered by Wyner and Ziv in [10].

3. Extreme cases:

We consider two cases: $\rho = 1$ and $\rho = 0$ to get some intuitive view of the problem.

(a) $\rho = 1$

This is the case: $Y_1 = Y_2 = X$. (5)-(8) could be simplified to:

$$R_1 \geq \frac{1}{2}\log^+ \frac{1+\sigma_X^2(\mu_1^2+\mu_2^2)}{1+\mu_2^2\sigma_X^2} \quad R_2 \geq \frac{1}{2}\log^+ \frac{1+\sigma_X^2(\mu_1^2+\mu_2^2)}{1+\mu_1^2\sigma_X^2}$$
$$R_1 + R_2 \geq \frac{1}{2}\log^+ [1+\sigma_X^2(\mu_1^2+\mu_2^2)], \text{ subject to } \frac{\sigma_X^2}{D} \leq 1+\sigma_X^2(\mu_1^2+\mu_2^2)$$

We directly get: $R_1 + R_2 \geq \frac{1}{2}\log^+ \frac{\sigma_X^2}{D}$. This is an interesting result, i.e. no matter how we distribute rates between two sensors, it looks as if all rates are assigned to one sensor. This phenomenon can be explained by the following argument. Since two sensors have same observations, they can use a common codebook which contains $2^{n(R_1+R_2)}$ codewords (i.e. indices). Represent every index in the form of $A \times B$ where $A \in \mathcal{A}, B \in \mathcal{B}, |\mathcal{A}| = 2^{nR_1}, |\mathcal{B}| = 2^{nR_2}$. Sensor 1 transmits part A and sensor 2 transmits part B (there is no confusion since both sensors have the same observations and codebook). Fusion center then combines A and B to get the original index. Apparently, this argument holds for more general case when $Y_1 = Y_2 = \cdots = Y_L = X + N$.

(b) $\rho = 0$

This is the case: $Y_1 = X + N_1$ and $Y_2 = N_2$ (or $Y_1 = N_1, Y_2 = N_2$), i.e. one of the sensor observes pure noise. We have:

$$R_1 \geq \frac{1}{2}\log^+ (\mu_1^2\sigma_{Y_1}^2+1), \;\; R_2 \geq \frac{1}{2}\log^+ (\mu_2^2\sigma_{N_2}^2+1)$$
$$R_1 + R_2 \geq \frac{1}{2}\log^+ [(\mu_1^2\sigma_{Y_1}^2+1)(\mu_2^2\sigma_{N_2}^2+1)], \text{ subject to } \frac{\sigma_X^2}{D} \leq \frac{\mu_1^2\sigma_{Y_1}^2+1}{\mu_1^2\sigma_{N_1}^2+1}$$

Obviously, distortion constraint does not rely on $\sigma_{N_2}^2$ or μ_2. To minimize the rates, we set $\mu_1^2 = \frac{\sigma_X^2 - D}{D(\sigma_X^2+\sigma_{N_1}^2)-\sigma_X^2\sigma_{N_1}^2}$, $\mu_2^2 = 0$ and get $R_1 = \frac{1}{2}\log^+ (\frac{\sigma_X^4}{D(\sigma_X^2+\sigma_{N_1}^2)-\sigma_X^2\sigma_{N_1}^2})$, $R_2 = 0$. So all rate should be assigned to sensor 1.

IV. Application to CEO problem

We apply the results in Section III to quadratic Gaussian CEO problem, i.e. trying to find the minimal total rate R to achieve the distortion less than or equal to D. Define:

$$\begin{aligned} G(\mu_1^2, \mu_2^2) &= (\mu_1^2\sigma_{Y_1}^2+1)(\mu_2^2\sigma_{Y_2}^2+1) - \mu_1^2\mu_1^2\rho^2\sigma_{Y_1}^2\sigma_{Y_2}^2 \\ &\quad +\lambda(1+\frac{\mu_1^2\sigma_X^2}{\mu_1^2\sigma_{N_1}^2+1}+\frac{\mu_2^2\sigma_X^2}{\mu_2^2\sigma_{N_2}^2+1}-\frac{\sigma_X^2}{D}) \end{aligned}$$

Use Lagrange Multiplier, we get:

$$\min(R_1+R_2)=\begin{cases}\frac{1}{2}\log^+\frac{4\sigma_X^2}{D\sigma_{N_1}^2\sigma_{N_2}^2(\frac{1}{D_0}-\frac{1}{D})^2}, & \text{if } D_0<D<\min(\frac{1}{\frac{1}{\sigma_X^2}+\frac{1}{\sigma_{N_1}^2}-\frac{1}{\sigma_{N_2}^2}},\frac{1}{\frac{1}{\sigma_X^2}+\frac{1}{\sigma_{N_2}^2}-\frac{1}{\sigma_{N_1}^2}});\\ \frac{1}{2}\log^+\frac{\sigma_X^4}{D(\sigma_X^2+\sigma_{N_1}^2)-\sigma_X^2\sigma_{N_1}^2}, & \text{if } \frac{1}{\frac{1}{\sigma_X^2}+\frac{1}{\sigma_{N_1}^2}-\frac{1}{\sigma_{N_2}^2}}<D<\sigma_X^2 \text{ and } \sigma_{N_1}^2<\sigma_{N_2}^2;\\ \frac{1}{2}\log^+\frac{\sigma_X^4}{D(\sigma_X^2+\sigma_{N_2}^2)-\sigma_X^2\sigma_{N_2}^2}, & \text{if } \frac{1}{\frac{1}{\sigma_X^2}+\frac{1}{\sigma_{N_2}^2}-\frac{1}{\sigma_{N_1}^2}}<D<\sigma_X^2 \text{ and } \sigma_{N_1}^2>\sigma_{N_2}^2.\end{cases} \quad (9)$$

(9) shows that if the distortion constraint is loose or the total amount of available rate is very limited, then the optima[7] scheme is to use the more reliable sensor only. It's easy to see that the optimal rate allocation scheme corresponds to the points on line "B" (i.e. segment from F_1 to F_2) in Fig.2, where

$$F_1=\left(\frac{1}{2}\log\frac{4\sigma_X^2\sigma_{N_2}^2}{D\sigma_{N_1}^2(\frac{1}{D_0}-\frac{1}{D})(2\sigma_{N_2}^2+2\sigma_X^2-\sigma_X^2\sigma_{N_2}^2(\frac{1}{D_0}-\frac{1}{D}))},\ \frac{1}{2}\log\frac{2\sigma_{N_2}^2+2\sigma_X^2-\sigma_X^2\sigma_{N_2}^2(\frac{1}{D_0}-\frac{1}{D})}{\sigma_{N_2}^4(\frac{1}{D_0}-\frac{1}{D})}\right),$$

$$F_2=\left(\frac{1}{2}\log\frac{2\sigma_{N_1}^2+2\sigma_X^2-\sigma_X^2\sigma_{N_1}^2(\frac{1}{D_0}-\frac{1}{D})}{\sigma_{N_1}^4(\frac{1}{D_0}-\frac{1}{D})},\ \frac{1}{2}\log\frac{4\sigma_X^2\sigma_{N_1}^2}{D\sigma_{N_2}^2(\frac{1}{D_0}-\frac{1}{D})(2\sigma_{N_1}^2+2\sigma_X^2-\sigma_X^2\sigma_{N_1}^2(\frac{1}{D_0}-\frac{1}{D}))}\right).$$

It is clear that the optimal rate allocation scheme generally is not unique except when one of the latter two cases in (9) happens. In these cases, "B" reduces to a point on R_1 axis $(\frac{1}{2}\log^+\frac{\sigma_X^4}{D(\sigma_X^2+\sigma_{N_1}^2)-\sigma_X^2\sigma_{N_1}^2},0)$ or R_2 axis $(0,\frac{1}{2}\log^+\frac{\sigma_X^4}{D(\sigma_X^2+\sigma_{N_2}^2)-\sigma_X^2\sigma_{N_2}^2})$.

V. Conclusion

In this paper, we derive the achievable region of rates and distortion in a distributed sensing system. Future work will be devoted to evaluating this result in quadratic gaussian case when $L>2$. Whether or not this achievable region contains all the legitimate points of rates and distortion is still an open problem. The key to the complete solution is to establish a converse coding theorem which is currently under investigation. This converse coding theorem, if found, will be a generalization of the results in [6].

APPENDIX: Proof of the Theorem

Consider the following exhaustive error events.

E_1 : $(X^n,Y_1^n,Y_2^n,\cdots,Y_L^n)$ are not jointly typical.

E_2 : $\bigcup_{i=1}^{L}E_{2,i}$, where $E_{2,i}=E_1^c\bigcap F_i$, for $i=1,2,\cdots,L$; F_i : (Y_i^n,W_i^n) not typical for all $W_i^n\in\mathcal{C}_i^{(n)}$.

E_3 : $E_1^c\bigcap E_2^c\bigcap F_{L+1}$ where F_{L+1}: there does not exist $(W_1^n,W_2^n,\cdots,W_L^n)$ such that $W_i^n\in\mathcal{B}_i(b_i)$ and $(W_1^n,W_2^n,\cdots,W_L^n)$ is jointly typical.

E_4 : $E_1^c\bigcap E_2^c\bigcap E_3^c\bigcap F_{L+2}$ where F_{L+2} : $(\hat{W}_1^n,\hat{W}_2^n,\cdots,\hat{W}_L^n)$ not unique.

E_5 : $E_1^c\bigcap E_2^c\bigcap E_3\bigcap E_4^c\bigcap F_{L+3}$ where F_{L+3} : $\frac{1}{n}d(X^n,\hat{X}^n)>D+\varepsilon$.

Let $\bar{P}_e$ denote the probability of decoding error, where the bar over P_e indicates that the probability is averaged over the ensemble codebooks. It is clear that

[7] We only address the optimal schem in the achievable region.

$$\bar{P}_e = P(\bigcup_{i=1}^{5} E_i) \leq \sum_{i=1}^{5} P(E_i)$$

(i) $P(E_1) \to 0$ as $n \to \infty$ by weak law of large numbers.

(ii) $P(E_2) \leq \sum_{i=1}^{L} P(E_{2,i})$

$P(E_{2,i}) = P(E_1^c \cap F_i) = P(F_i|E_1^c)P(E_1^c) \leq P(F_i|E_1^c)$

$= P\{(Y_i^n, W_i^n)$ not typical for all $W_i^n \in \mathcal{C}_i|(Y_1^n, Y_2^n, \cdots, Y_L^n)$ typical$\}$

$= [1 - P\{(Y_i^n, W_i^n)$ typical for randomly chosen $W_i^n|Y_i^n$ typical $\}]^{M_i}$

$\leq [1 - 2^{-n[I(Y_i;W_i)+\varepsilon_i]}]^{M_i} \leq \exp(-M_i 2^{-n[I(Y_i;W_i)+\varepsilon_i]}) \to 0$

as $n \to \infty$ if $M_i \geq 2^{n[I(Y_i;W_i)+2\varepsilon_i]}$, where ε_i could be made arbitrarily small as $n \to \infty$.

(iii) $P(E_3) \to 0$ as $n \to \infty$ by the Markov lemma on typicality.

(iv) $P(E_4) \leq P(F_{L+2}|\bigcup_{j=1}^{3} E_j^c) = P\{(\hat{W}_1^n, \hat{W}_2^n, \cdots, \hat{W}_L^n)$is not unique$|\bigcup_{j=1}^{3} E_j^c\}$

$\leq \sum_{\mathcal{A}\subseteq\mathcal{I}_L} [(\prod_{i\in\mathcal{A}} N_i)2^{-n(\Sigma_{\mathcal{A}}-\varepsilon_{\mathcal{A}})}] \to 0$ if $\prod_{i\in\mathcal{A}} N_i \leq 2^{n(\Sigma_{\mathcal{A}}-2\varepsilon_{\mathcal{A}})}$ for all $\mathcal{A} \subseteq \mathcal{I}_L$

Here WLOG suppose $\mathcal{A} = \{i_1, i_2, \cdots, i_k\}$, then

$\Sigma_{\mathcal{A}} \triangleq \sum_{l=1}^{k} I(W_{\mathcal{I}_L\backslash\mathcal{A}_{k+1-l}}; W_{i_l})$; $\mathcal{A}_l \triangleq \{i_{k-l+1}, i_{k-l+2}, \cdots, i_k\}$, i.e. the last l elements of $\mathcal{A}$; $\varepsilon_{\mathcal{A}}$ can be made arbitrarily small as $n \to \infty$.

(v) $P(E_5) \to 0$ as $n \to \infty$ by the jointly strong typicality of $(X^n, \hat{W}_1^n, \hat{W}_2^n, \cdots, \hat{W}_L^n)$, the definition of g and the boundedness of d (i.e. $d_{max} < \infty$).

In summary, if $M_i \geq 2^{n[I[Y_i;W_i]+2\varepsilon_i]}$ and $\prod_{i\in\mathcal{A}} N_i \leq 2^{n(\Sigma_{\mathcal{A}}-2\varepsilon_{\mathcal{A}})}$ for all $i \in \mathcal{I}_L$ and $\mathcal{A} \subseteq \mathcal{I}_L$, then $\bar{P}_e \to 0$ as $n \to \infty$. So we have $2^{n\sum_{i\in\mathcal{A}} R_i} = \frac{\prod_{i\in\mathcal{A}} M_i}{\prod_{i\in\mathcal{A}} N_i} \geq 2^{n[\sum_{i\in\mathcal{A}} I(Y_i;W_i)-\Sigma_{\mathcal{A}}-\Delta_{\mathcal{A}}]}$, where $\Delta_{\mathcal{A}} \triangleq 2\sum_{i\in\mathcal{A}} \varepsilon_i + 2\varepsilon_{\mathcal{A}}$. Since $\Delta_{\mathcal{A}}$ can be arbitrarily small, it follows that

$$\sum_{i\in\mathcal{A}} R_i \geq \sum_{i\in\mathcal{A}} I(Y_i;W_i) \ - \ \Sigma_{\mathcal{A}} \tag{10}$$

Since $W_i \to Y_i \to (X, Y_{\mathcal{I}_L\backslash\{i\}}, W_{\mathcal{I}_L\backslash\{i\}})$, we have

$$\begin{aligned}
I(Y_{\mathcal{A}}; W_{\mathcal{A}}|W_{\mathcal{I}_L\backslash\mathcal{A}}) &= H(W_{\mathcal{A}}|W_{\mathcal{I}_L\backslash\mathcal{A}}) - H(W_{\mathcal{A}}|W_{\mathcal{I}_L\backslash\mathcal{A}}, Y_{\mathcal{A}}) \\
&= \sum_{l=1}^{k} H(W_{i_l}|W_{\mathcal{I}_L\backslash\mathcal{A}_{k+1-l}}) - \sum_{l=1}^{k} H(W_{i_l}|Y_{i_l}) \\
&= \sum_{l=1}^{k} [H(W_{i_l}|W_{\mathcal{I}_L\backslash\mathcal{A}_{k+1-l}}) - H(W_{i_l})] - \sum_{l=1}^{k} [H(W_{i_l}|Y_{i_l}) - H(W_{i_l})] \\
&= -\sum_{l=1}^{k} I(W_{i_l}; W_{\mathcal{I}_L\backslash\mathcal{A}_{k+1-l}}) + \sum_{l=1}^{k} I(W_{i_l}; Y_{i_l}) \\
&= \sum_{j\in\mathcal{A}} I(Y_j; W_j) \ - \ \Sigma_{\mathcal{A}}
\end{aligned}$$

Here as before we suppose $\mathcal{A} = \{i_1, i_2, \cdots, i_k\}$.
So we can rewrite (10) as $\sum_{i\in\mathcal{A}} R_i \geq \sum_{i\in\mathcal{A}} I(Y_{\mathcal{A}}; W_{\mathcal{A}}|W_{\mathcal{I}_L\setminus\mathcal{A}})$ for all $\mathcal{A} \subseteq \mathcal{I}_L$. QED.

References

[1] T. J. Flynn and R. M. Gray, "Encoding of correlated observations," *IEEE Trans. Inform. Theory*, vol. 33, pp. 773-787, Nov. 1987.

[2] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Trans. Inform. Theory*, vol. 42, pp. 887-902, May 1996.

[3] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1549-1559, Sept. 1997.

[4] Yasutada Oohama, "The Rate-Distortion Function for the Quadratic Gaussian CEO Problem," *IEEE Trans. on Inform. Theory*, vol. 44, no. 3, pp. 1057-1070, May 1998.

[5] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder-II: General sources," *Inform. Contr.*, vol. 38, pp. 60-80, Jul. 1978.

[6] Yasutada Oohama, "Gaussian Multiterminal Source Coding," *IEEE Trans. on Inform. Theory*, vol. 43, no. 6, pp. 1912-1923, Nov. 1997.

[7] S. Y. Tung, "Multiterminal source coding," Ph.D. dissertation, School of Electrical Engineering, Cornell Univ., Ithaca, NY, May 1978.

[8] T. M. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Inform. Theory*, vol. 21, pp 226-228, Mar. 1975.

[9] Stark C. Draper, Gregory W. Wornell, "Successively Structured CEO Problems," *ISIT 2002*, pp. 65, Lausanne, Switzerland, June 30-July 5, 2002

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