Distortion-Rate Analysis for Distributed Estimation with Wireless Sensor Networks†

Ioannis D. Schizas, Georgios B. Giannakis and Nihar Jindal
Dept. of ECE, University of Minnesota
200 Union Str. SE, Minneapolis, MN, 55455, USA
{schizas,georgios,nihar}@ece.umn.edu

Abstract

We deal with centralized and distributed rate-constrained estimation of random signal vectors using a network of wireless sensors communicating with a fusion center. Specifically, we determine lower and upper bounds for the Distortion-Rate (D-R) function. The lower bound is obtained by considering centralized estimation with a single-sensor setup, for which we determine the D-R function in closed-form, and specify an estimator that achieves it. Furthermore, we derive a novel alternating scheme that numerically determines an achievable upper bound of the D-R function for general distributed estimation using multiple sensors. Numerical examples confirm that the algorithm performs well and yields D-R upper bounds which are tight.

Keywords: Distortion-Rate Function, Distributed Estimation, Wireless Sensor Networks

1 Introduction

Stringent bandwidth and energy constraints that wireless sensor networks (WSNs) must adhere to motivate efficient compression and encoding schemes when estimating random signals or parameter vectors of interest. In such networks, it is of paramount importance to determine bounds on the minimum achievable distortion between the signal of interest and its estimate formed at the fusion center (FC) using the encoded information transmitted by the sensors subject to rate constraints.

In the *reconstruction* scenario, the FC wishes to accurately estimate the sensor observations. In the *estimation* scenario, the FC is interested in accurately estimating an underlying random vector which is correlated with, but not equal to, the sensor observations. In the single sensor setting, single-letter characterizations of the D-R function for both scenarios are known: the reconstruction scenario is the standard distortion-rate problem [4, p. 336], and the estimation problem, which is also referred to as rate-distortion with a remote source, has also been determined [1, p. 78]. In the distributed scenario, where there are multiple sensors with correlated observations, neither problem is well understood. The best analytical inner and outer bounds for the D-R function for reconstruction can be found in [2]. An iterative scheme has been developed in [5], which numerically determines an achievable upper bound for distributed reconstruction but not for signal estimation.

† Prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

For the general problem of estimating a parameter vector which has continuous-valued entries and is correlated with sensor observations, most of the existing literature examines Gaussian data and Gaussian parameters. Specifically, when each sensor observes a common *scalar* random parameter contaminated with Gaussian noise, the D-R function for estimating this parameter has been determined in [3,7,11] to solve the so called CEO problem; see also [6]. Additionally, D-R bounds for a linear-Gaussian data model have been derived in [8] when the number of parameters equals the number of all sensor observations. In this paper, we pursue D-R analysis for distributed estimation with WSNs without constraining the number of observations and/or the number of random parameters we want to estimate.

We first determine the D-R function for estimating a *vector* parameter when applying rate-constrained encoding to the observation data, in closed form for the single-sensor case (Section 3). Without assuming that the number of parameters equals the number of observations, we prove that the optimal scheme achieving the D-R function amounts to first computing the minimum mean square error (MMSE) estimate at the sensor, and then optimally compressing the estimate via reverse water-filling (rwf). The D-R function for the single-sensor setup serves as a non-achievable lower D-R bound for rate constrained estimation in the multi-sensor setup. Next, we develop an alternating scheme that numerically determines an achievable D-R upper bound for the multi-sensor scenario (Section 4). Different from [5], which deals with WSN-based distributed reconstruction, our approach aims for general estimation problems. Combining the lower bound of Section 3 with the numerically determined upper bound of Section 4, we specify a region where the D-R function for distributed estimation lies in.

2 Problem Statement

With reference to Fig. 1 (Left), consider a WSN comprising L sensors that communicate with an FC. Each sensor, say the ith, observes an $N_i \times 1$ vector $\mathbf{x}_i(t)$ which is correlated with a $p \times 1$ random signal (parameter vector) of interest $\mathbf{s}(t)$, where t denotes discrete time. Similar to [7, 8, 11], we assume that:

(a1) *No information is exchanged among sensors and the links with the FC are noise-free.*

(a2) *The random vector $\mathbf{s}(t)$ is generated by a stationary Gaussian vector memoryless source with $\mathbf{s}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{ss})$; the sensor data $\{\mathbf{x}_i(t)\}_{i=1}^{L}$ adhere to the linear-Gaussian model $\mathbf{x}_i(t) = \mathbf{H}_i\mathbf{s}(t) + \mathbf{n}_i(t)$, where $\mathbf{n}_i(t)$ denotes additive white Gaussian noise (AWGN); i.e., $\mathbf{n}_i(t) \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$; noise $\mathbf{n}_i(t)$ is uncorrelated across sensors, time and with $\mathbf{s}$; and $\mathbf{H}_i$ as well as (cross-) covariance matrices $\mathbf{\Sigma}_{ss}$, $\mathbf{\Sigma}_{sx_i}$ and $\mathbf{\Sigma}_{x_ix_j}$ are known $\forall i, j \in \{1, \ldots, L\}$.*

Notice that (a1) assumes that sufficiently strong channel codes are used; while whiteness of $\mathbf{n}_i(t)$ and the zero-mean assumptions in (a2) are made without loss of generality. The linear model in (a2) is commonly encountered in estimation and in a number of cases it even accurately approximates non-linear mappings; e.g., via a first-order Taylor expansion in target tracking applications. Although confining ourselves to Gaussian vectors $\mathbf{x}_i(t)$ is of interest on its own, following arguments similar to those in [1, p. 134] we can show that the D-R functions obtained in this paper bound from above their counterparts for non-Gaussian sensor data $\mathbf{x}_i(t)$.

Blocks $\mathbf{x}_i^{(n)} := \{\mathbf{x}_i(t)\}_{t=1}^{n}$, comprising n consecutive time instantiations of the vector $\mathbf{x}_i(t)$, are encoded per sensor to yield each encoder's output $\mathbf{u}_i^{(n)} = \mathbf{f}_i^{(n)}(\mathbf{x}_i^{(n)})$, $i = 1, \ldots, L$. These outputs are communicated through ideal orthogonal channels to the FC. There, $\mathbf{u}_i^{(n)}$'s are decoded to obtain an estimate of $\mathbf{s}^{(n)} := \{\mathbf{s}(t)\}_{t=1}^{n}$ denoted as $\hat{\mathbf{s}}_R^{(n)}(\mathbf{u}_1^{(n)}, \ldots, \mathbf{u}_L^{(n)}) = \mathbf{g}_R^{(n)}(\mathbf{x}_1^{(n)}, \ldots, \mathbf{x}_L^{(n)})$, since $\mathbf{u}_i^{(n)}$ is a function of $\mathbf{x}_i^{(n)}$. The rate constraint is imposed through a bound on the cardinality of the range of the sensor encoding functions, i.e., the cardinality of

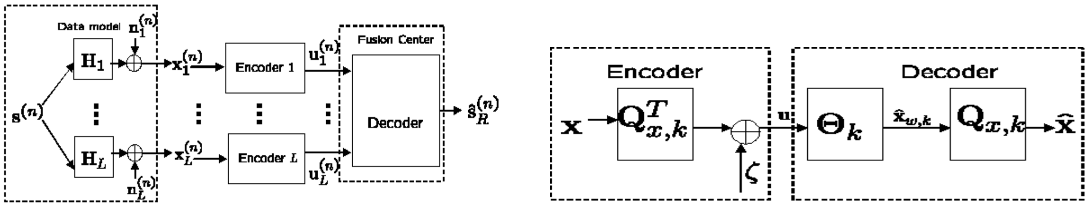


Figure 1: *Left*: Distributed setup.; *Right*: Test channel for $\mathbf{x}$ Gaussian in a point-to-point link.

the range of $\mathbf{f}_i^{(n)}$ must be no larger than 2^{nR_i}, where R_i is the available rate at the encoder of the ith sensor. The sum rate satisfies the constraint $\sum_{i=1}^{L} R_i \leq R$, where R is the total available rate shared by the L sensors. Under this rate constraint, we want to determine the minimum possible MSE distortion $(1/n)\sum_{t=1}^{n} E[\|\mathbf{s}(t) - \hat{\mathbf{s}}_R(t)\|^2]$ for estimating $\mathbf{s}$ in the limit of infinite block-length n. When $L = 1$, a single-letter information theoretic characterization is known for the latter, but no simplification is known for the distributed multi-sensor scenario.

3 Distortion-Rate for Centralized Estimation

We will first determine the D-R function for estimating s(t) in a *single-sensor* setup. The single-letter characterization of the D-R function in this setup allow us to drop the time index. Here, all $\{\mathbf{x}_i\}_{i=1}^{L} := \mathbf{x}$ are available to a single sensor, and $\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}$. We let $\rho := \text{rank}(\mathbf{H})$ denote the rank of matrix $\mathbf{H}$. The D-R function in such a scenario provides a lower (non-achievable) bound on the MMSE that can be achieved in a multi-sensor distributed setup, where each $\mathbf{x}_i$ is observed by a different sensor. Existing works treat the case $N = p$ [9, 12], but here we look for the D-R function regardless of N, p, in the linear-Gaussian model framework.

3.1 Background on D-R analysis for Reconstruction

The D-R function for encoding $\mathbf{x}$, which has probability density function (pdf) $p(\mathbf{x})$, with rate R at an individual sensor, and reconstructing it (in the MMSE sense) as $\hat{\mathbf{x}}$ at the FC, is given by [4, p. 342]

$$D_x(R) = \min_{\substack{p(\hat{\mathbf{x}}|\mathbf{x}) \\ I(\mathbf{x};\hat{\mathbf{x}}) \leq R}} E_{p(\hat{\mathbf{x}},\mathbf{x})}[\|\mathbf{x} - \hat{\mathbf{x}}\|^2], \;\; \mathbf{x} \in \mathbb{R}^N, \;\; \hat{\mathbf{x}} \in \mathbb{R}^N, \tag{1}$$

where the minimization is w.r.t. the conditional pdf $p(\hat{\mathbf{x}}|\mathbf{x})$. Let $\mathbf{\Sigma}_{xx} = \mathbf{Q}_x \mathbf{\Lambda}_x \mathbf{Q}_x^T$ denote the eigenvalue decomposition of $\mathbf{\Sigma}_{xx}$, where $\mathbf{\Lambda}_x = \text{diag}(\lambda_{x,1} \cdots \lambda_{x,N})$ and $\lambda_{x,1} \geq \cdots \geq \lambda_{x,N} > 0$.

For $\mathbf{x}$ Gaussian, $D_x(R)$ can be determined by applying rwf to the pre-whitened vector $\mathbf{x}_w := \mathbf{Q}_x^T \mathbf{x}$ [4, p. 348]. For a prescribed rate R, it turns out that $\exists\, k$ such that the first k entries $\{\mathbf{x}_w(i)\}_{i=1}^{k}$ of $\mathbf{x}_w$, are encoded and reconstructed independently from each other using rate $\{R_i = 0.5 \log_2(\lambda_{x,i}/d(k,R))\}_{i=1}^{k}$, where $d(k,R) = \left(\prod_{i=1}^{k} \lambda_{x,i}\right)^{1/k} 2^{-2R/k}$ with $R = \sum_{i=1}^{k} R_i$; and the last $N-k$ entries of $\mathbf{x}_w$ are assigned no rate; i.e., $\{R_i = 0\}_{i=k+1}^{N}$. The corresponding MMSE for encoding $\mathbf{x}_w(i)$, the ith entry of $\mathbf{x}_w$, under a rate constraint R_i, is $D_i = E[\|\mathbf{x}_w(i) - \hat{\mathbf{x}}_w(i)\|^2] = d(k,R)$ when $i = 1, \ldots, k$ and $D_i = \lambda_{x,i}$ when $i = k+1 \ldots, N$. The resultant MMSE (D-R function) is

$$D_x(R) = E[\|\mathbf{x} - \hat{\mathbf{x}}\|^2] = E[\|\mathbf{x}_w - \hat{\mathbf{x}}_w\|^2] = kd(k,R) + \sum_{i=k+1}^{N} \lambda_{x,i}. \tag{2}$$

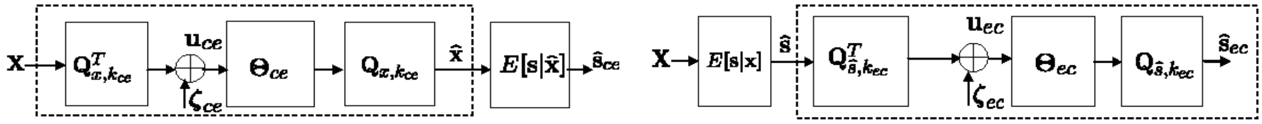


Figure 2: *Left*: Test channel for the CE scheme.; *Right*: Test channel for the EC scheme.

Especially for $d(k,R)$, it holds that $\max(\{\lambda_{x,i}\}_{i=k+1}^N) \leq d(k,R) < \min(\{\lambda_{x,i}\}_{i=1}^k)$. Intuitively, $d(k,R)$ is a threshold distortion determining which entries of $\mathbf{x}_w$ are assigned with nonzero rate. The first k entries of $\mathbf{x}_w$ with variance $\lambda_{x,i} > d(k,R)$ are encoded with non-zero rate, but the last $N-k$ ones are discarded in the encoding procedure (are set to zero).

Associated with the rwf principle is the so called test channel; see e.g., [4, p. 345]. The encoder's MSE optimal output is $\mathbf{u} = \mathbf{Q}_{x,k}^T\mathbf{x} + \boldsymbol{\zeta}$, where $\mathbf{Q}_{x,k}$ is formed by the first k columns of $\mathbf{Q}_x$, and $\boldsymbol{\zeta}$ models the distortion noise that results due to the rate-constrained encoding of $\mathbf{x}$. The zero-mean AWGN $\boldsymbol{\zeta}$ is uncorrelated with $\mathbf{x}$ and its diagonal covariance matrix $\boldsymbol{\Sigma}_{\zeta\zeta}$ has entries $[\boldsymbol{\Sigma}_{\zeta\zeta}]_{ii} = \lambda_{x,i}D_i/(\lambda_{x,i} - D_i)$. The part of the test channel that takes as input $\mathbf{u}$ and outputs $\hat{\mathbf{x}}$, models the decoder. The reconstruction $\hat{\mathbf{x}}$ of $\mathbf{x}$ at the decoder output is

$$\hat{\mathbf{x}} = \mathbf{Q}_{x,k}\boldsymbol{\Theta}_k\mathbf{u} = \mathbf{Q}_{x,k}\boldsymbol{\Theta}_k\mathbf{Q}_{x,k}^T\mathbf{x} + \mathbf{Q}_{x,k}\boldsymbol{\Theta}_k\boldsymbol{\zeta}, \tag{3}$$

where $\boldsymbol{\Theta}_k$ is a diagonal matrix with non-zero entries $[\boldsymbol{\Theta}_k]_{ii} = (\lambda_{x,i} - D_i)/\lambda_{x,i}$, $i = 1, \ldots, k$.

3.2 D-R analysis for Estimation

The D-R function for estimating source $\mathbf{s}$ given observation $\mathbf{x}$ (where the source and observation are probabilistically drawn from the joint pdf $p(\mathbf{x}, \mathbf{s})$) with rate R at an individual sensor, and reconstructing it (in the MMSE sense) as $\hat{\mathbf{x}}$ at the FC is given by [1, p. 79]

$$D_s(R) = \min_{\substack{p(\hat{\mathbf{s}}_R|\mathbf{x}) \\ I(\mathbf{x};\hat{\mathbf{s}}_R) \leq R}} E_{p(\hat{\mathbf{s}}_R,\mathbf{s})}[\|\mathbf{s} - \hat{\mathbf{s}}_R\|^2], \;\; \mathbf{s} \in \mathbb{R}^N, \;\; \hat{\mathbf{s}}_R \in \mathbb{R}^N, \tag{4}$$

where the minimization is w.r.t. the conditional pdf $p(\hat{\mathbf{s}}_R|\mathbf{x})$. In order to achieve the D-R function, one might be tempted to first compress $\mathbf{x}$ by applying rwf at the sensor, without taking into account the data model relating $\mathbf{s}$ with $\mathbf{x}$, and subsequently use the reconstructed $\hat{\mathbf{x}}$ to form the MMSE estimate $\hat{\mathbf{s}}_{ce} = E[\mathbf{s}|\hat{\mathbf{x}}]$ at the FC. An alternative option would be to first form the MMSE estimate $\hat{\mathbf{s}} = E[\mathbf{s}|\mathbf{x}]$, encode the latter using rwf at the sensor, and after decoding at the FC, obtain the reconstructed estimate $\hat{\mathbf{s}}_{ec}$. Referring to the former option as *Compress-Estimate* (CE), and to the latter as *Estimate-Compress* (EC), we are interested in determining which one yields the smallest MSE under a rate constraint R. Another interesting question is whether any of the CE and EC schemes enjoys MMSE optimality (i.e., achieves (4)). With subscripts ce and ec corresponding to these two options, let us also define the errors $\tilde{\mathbf{s}}_{ce} := \mathbf{s} - \hat{\mathbf{s}}_{ce}$ and $\tilde{\mathbf{s}}_{ec} := \mathbf{s} - \hat{\mathbf{s}}_{ec}$.

For CE, we depict in Fig. 2 (Left) the test channel for encoding $\mathbf{x}$ via rwf, followed by MMSE estimation of $\mathbf{s}$ based on $\hat{\mathbf{x}}$. Suppose that when applying rwf to $\mathbf{x}$ with prescribed rate R, the first k_{ce} components of $\mathbf{x}_w$ are assigned with non-zero rate and the rest are discarded. The MMSE optimal encoder's output for encoding $\mathbf{x}$ is given, as in Section 3.1, by $\mathbf{u}_{ce} = \mathbf{Q}_{x,k_{ce}}^T\mathbf{x} + \boldsymbol{\zeta}_{ce}$. The covariance matrix of $\boldsymbol{\zeta}_{ce}$ has diagonal entries $[\boldsymbol{\Sigma}_{\zeta_{ce}\zeta_{ce}}]_{ii} = \lambda_{x,i}D_i^{ce}/(\lambda_{x,i} - D_i^{ce})$ for $i = 1, \ldots, k_{ce}$, where $D_i^{ce} := E[(\mathbf{x}_w(i) - \hat{\mathbf{x}}_w(i))^2]$. Recalling that $D_i^{ce} = \left(\prod_{i=1}^{k_{ce}} \lambda_{x,i}\right)^{1/k_{ce}} 2^{-2R/k_{ce}}$ when $i = 1, \ldots, k_{ce}$ and $D_i^{ce} = \lambda_{x,i}$, when $i = k_{ce}+1, \ldots, N$, the reconstructed $\hat{\mathbf{x}}$ in CE is [c.f. (3)]

$$\hat{\mathbf{x}} = \mathbf{Q}_{x,k_{ce}}\boldsymbol{\Theta}_{ce}\mathbf{Q}_{x,k_{ce}}^T\mathbf{x} + \mathbf{Q}_{x,k_{ce}}\boldsymbol{\Theta}_{ce}\boldsymbol{\zeta}_{ce}, \tag{5}$$

where $[\boldsymbol{\Theta}_{ce}]_{ii} = (\lambda_{x,i} - D_i^{ce})/\lambda_{x,i}$, for $i = 1, \ldots, k_{ce}$. Letting $\check{\mathbf{x}} := \mathbf{Q}_x^T \hat{\mathbf{x}} = [\check{\mathbf{x}}_1^T \;\; \mathbf{0}_{1\times(N-k_{ce})}]^T$, with $\check{\mathbf{x}}_1 := \boldsymbol{\Theta}_{ce}\mathbf{Q}_{x,k_{ce}}^T \mathbf{x} + \boldsymbol{\Theta}_{ce}\boldsymbol{\zeta}_{ce}$, we have for the MMSE estimate

$$\hat{\mathbf{s}}_{ce} = E[\mathbf{s}|\hat{\mathbf{x}}] = E[\mathbf{s}|\mathbf{Q}_x^T\hat{\mathbf{x}}] = E[\mathbf{s}|\check{\mathbf{x}}_1] = \boldsymbol{\Sigma}_{s\check{x}_1}\boldsymbol{\Sigma}_{\check{x}_1\check{x}_1}^{-1}\check{\mathbf{x}}_1, \tag{6}$$

since $\mathbf{Q}_x^T$ is unitary and the last $N - k_{ce}$ entries of $\check{\mathbf{x}}$ are useless for estimating $\mathbf{s}$. We have shown in [10] that the covariance matrix $\boldsymbol{\Sigma}_{\tilde{s}_{ce}\tilde{s}_{ce}} := E[(\mathbf{s} - \hat{\mathbf{s}}_{ce})(\mathbf{s} - \hat{\mathbf{s}}_{ce})^T]$ of $\tilde{\mathbf{s}}_{ce}$ is

$$\boldsymbol{\Sigma}_{\tilde{s}_{ce}\tilde{s}_{ce}} = \boldsymbol{\Sigma}_{ss} - \boldsymbol{\Sigma}_{s\check{x}_1}\boldsymbol{\Sigma}_{\check{x}_1\check{x}_1}^{-1}\boldsymbol{\Sigma}_{\check{x}_1 s} = \boldsymbol{\Sigma}_{ss} - \boldsymbol{\Sigma}_{sx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xs} + \boldsymbol{\Sigma}_{sx}\mathbf{Q}_x\boldsymbol{\Delta}_{ce}\mathbf{Q}_x^T\boldsymbol{\Sigma}_{xs}, \tag{7}$$

where $\boldsymbol{\Delta}_{ce} := \text{diag}\left(D_1^{ce}\lambda_{x,1}^{-2} \cdots D_N^{ce}\lambda_{x,N}^{-2}\right)$.

In Fig. 2 (Right) we depict the test channel for the EC scheme. The MMSE estimate $\hat{\mathbf{s}} = E[\mathbf{s}|\mathbf{x}]$ is followed by the test channel that results when applying rwf to a pre-whitened version of $\hat{\mathbf{s}}$, with rate R. Let $\boldsymbol{\Sigma}_{\hat{s}\hat{s}} = \mathbf{Q}_{\hat{s}}\boldsymbol{\Lambda}_{\hat{s}}\mathbf{Q}_{\hat{s}}^T$ be the eigenvalue decomposition for the covariance matrix of $\hat{\mathbf{s}}$, where $\boldsymbol{\Lambda}_{\hat{s}} = \text{diag}(\lambda_{\hat{s},1} \cdots \lambda_{\hat{s},p})$ and $\lambda_{\hat{s},1} \geq \cdots \geq \lambda_{\hat{s},p}$. Suppose now that the first k_{ec} entries of $\hat{\mathbf{s}}_w = \mathbf{Q}_{\hat{s}}^T\hat{\mathbf{s}}$ are assigned with non-zero rate and the rest are discarded. The MSE optimal encoder's output is given by $\mathbf{u}_{ec} = \mathbf{Q}_{\hat{s},k_{ec}}^T\hat{\mathbf{s}} + \boldsymbol{\zeta}_{ec}$, and the estimate $\hat{\mathbf{s}}_{ec}$ is

$$\hat{\mathbf{s}}_{ec} = \mathbf{Q}_{\hat{s},k_{ec}}\boldsymbol{\Theta}_{ec}\mathbf{Q}_{\hat{s},k_{ec}}^T\hat{\mathbf{s}} + \mathbf{Q}_{\hat{s},k_{ec}}\boldsymbol{\Theta}_{ec}\boldsymbol{\zeta}_{ec}, \tag{8}$$

where $\mathbf{Q}_{\hat{s},k_{ec}}$ is formed by the first k_{ec} columns of $\mathbf{Q}_{\hat{s}}$. For the $k_{ec} \times k_{ec}$ diagonal matrices $\boldsymbol{\Theta}_{ec}$ and $\boldsymbol{\Sigma}_{\zeta_{ec}\zeta_{ec}}$ we have $[\boldsymbol{\Theta}_{ec}]_{ii} = (\lambda_{\hat{s},i} - D_i^{ec})/\lambda_{\hat{s},i}$ and $[\boldsymbol{\Sigma}_{\zeta_{ec}\zeta_{ec}}]_{ii} = \lambda_{\hat{s},i}D_i^{ec}/(\lambda_{\hat{s},i} - D_i^{ec})$, where $D_i^{ec} := E[(\hat{\mathbf{s}}_w(i) - \hat{\mathbf{s}}_{ec,w}(i))^2]$, and $\hat{\mathbf{s}}_{ec,w} := \mathbf{Q}_{\hat{s}}^T\hat{\mathbf{s}}_{ec}$. Recall also that $D_i^{ec} = \left(\prod_{i=1}^{k_{ec}}\lambda_{\hat{s},i}\right)^{1/k_{ec}} 2^{-2R/k_{ec}}$ when $i = 1, \ldots, k_{ec}$ and $D_i^{ec} = \lambda_{\hat{s},i}$, for $i = k_{ec}+1, \ldots, p$. Upon defining $\boldsymbol{\Delta}_{ec} := \text{diag}\left(D_1^{ec} \cdots D_p^{ec}\right)$, the covariance matrix of $\tilde{\mathbf{s}}_{ec}$ is given by [10]

$$\boldsymbol{\Sigma}_{\tilde{s}_{ec}\tilde{s}_{ec}} = \boldsymbol{\Sigma}_{ss} - \boldsymbol{\Sigma}_{sx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xs} + \mathbf{Q}_{\hat{s}}\boldsymbol{\Delta}_{ec}\mathbf{Q}_{\hat{s}}^T. \tag{9}$$

The MMSE associated with CE and EC is given, respectively, by [c.f. (7) and (9)]

$$D_{ce}(R) := \text{trace}(\boldsymbol{\Sigma}_{\tilde{s}_{ce}\tilde{s}_{ce}}) = J_o + \epsilon_{ce}(R), \text{ and } D_{ec}(R) := \text{trace}(\boldsymbol{\Sigma}_{\tilde{s}_{ec}\tilde{s}_{ec}}) = J_o + \epsilon_{ec}(R), \tag{10}$$

where $\epsilon_{ce}(R) := \text{trace}(\boldsymbol{\Sigma}_{sx}\mathbf{Q}_x\boldsymbol{\Delta}_{ce}\mathbf{Q}_x^T\boldsymbol{\Sigma}_{xs})$, $\epsilon_{ec}(R) := \text{trace}(\mathbf{Q}_{\hat{s}}\boldsymbol{\Delta}_{ec}\mathbf{Q}_{\hat{s}}^T)$, and $J_o := \text{trace}(\boldsymbol{\Sigma}_{ss} - \boldsymbol{\Sigma}_{sx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xs})$ is the MMSE achieved when estimating $\mathbf{s}$ based on $\mathbf{x}$, without source encoding ($R \to \infty$). Since J_o is common to both EC and CE it is important to compare $\epsilon_{ce}(R)$ with $\epsilon_{ec}(R)$ in order to determine which estimation scheme achieves the smallest MSE. The following theorem provides such an asymptotic comparison:

Theorem 1: *If* $R > R_{th} := 0.5\max\left\{\log_2\left(\left(\prod_{i=1}^{\rho}\lambda_{x,i}\right)/\sigma^{2\rho}\right), \log_2\left(\left(\prod_{i=1}^{\rho}\lambda_{\hat{s},i}\right)/(\lambda_{\hat{s},\rho})^{\rho}\right)\right\}$, *then* $\epsilon_{ce}(R) = \gamma_1 2^{-2R/N}$ *and* $\epsilon_{ec}(R) = \gamma_2 2^{-2R/\rho}$, *where* γ_1 *and* γ_2 *are constants.*

An immediate consequence of Theorem 1 is that the MSE for EC converges as $R \to \infty$ to J_o with rate $O(2^{-2R/\rho})$. The MSE of CE converges likewise, but with rate $O(2^{-2R/N})$. For the typical case $N > \rho$, EC approaches the lower bound J_o faster than CE, implying correspondingly a more efficient usage of the available rate R. This is intuitively reasonable since CE compresses $\mathbf{x}$, which contains the noise $\mathbf{n}$. Since the last $N - \rho$ eigenvalues of $\boldsymbol{\Sigma}_{xx}$ equal the noise variance σ^2, part of the available rate is consumed to compress the noise. On the contrary, the MMSE estimator $\hat{\mathbf{s}}$ in EC suppresses significant part of the noise.
Let us examine now some special cases to gain more insight about Theorem 1.

Scalar model ($p = 1$, $N = 1$): Let $x = hs + n$, where h is fixed, while s, n are uncorrelated with $s \sim \mathcal{N}(0, \sigma_s^2)$, $n \sim \mathcal{N}(0, \sigma_n^2)$, and $\sigma_x^2 = h^2\sigma_s^2 + \sigma_n^2$. With $\sigma^2_{\tilde{s}_{ce}}$ and $\sigma^2_{\tilde{s}_{ec}}$ denoting the variances of $\tilde{s}_{ce}$ and $\tilde{s}_{ec}$, respectively, we have shown in [10] that:

Proposition 1: *For $N = p = 1$, it holds that $\sigma^2_{\tilde{s}_{ce}} = \sigma^2_{\tilde{s}_{ec}}$ and hence the D-R functions for EC and CE are identical; i.e., $D_{ec}(R) = D_{ce}(R)$.*

Vector model ($p = 1$, $N > 1$): With $\mathbf{x} = \mathbf{h}s + \mathbf{n}$ and $R_{th} := 0.5\log_2(1 + \sigma_s^2\|\mathbf{h}\|^2/\sigma^2)$, we have established that [10]:

Proposition 2: *For $R \leq R_{th}$ it holds that $\epsilon_{ce}(R) = \epsilon_{ec}(R)$ and thus $D_{ec}(R) = D_{ce}(R)$. For $R > R_{th}$, we have $\epsilon_{ce}(R) > \epsilon_{ec}(R)$ and thus EC uses more efficiently the available rate.*

Matrix-vector model ($N > 1$, $p > 1$ and $\mathbf{\Sigma}_{ss} = \sigma_s^2\mathbf{I}_p$): For this setup, we have $\mathbf{\Sigma}_{sx} = \sigma_s^2\mathbf{H}^T$ and $\mathbf{\Sigma}_{xx} = \sigma_s^2\mathbf{H}\mathbf{H}^T + \sigma^2\mathbf{I}$. Letting $\mathbf{H} = \mathbf{U}_h\mathbf{\Sigma}_h\mathbf{V}_h^T$ be the SVD of $\mathbf{H}$, where $\mathbf{\Sigma}_h$ is an $N \times p$ diagonal matrix $\mathbf{\Sigma}_h = \text{diag}(\sigma_{h,1}\cdots\sigma_{h,p})$, we have proved that [10]

Proposition 3: *If either $N > \rho$, or, $N = \rho$ and $\exists\ i, j \in [1, \rho]$ with $i \neq j$ such that $\sigma_{h,i} \neq \sigma_{h,j}$ and $R > R_{th}$, with*

$$R_{th} := \frac{1}{2}\max\left\{\log_2\left(\prod_{i=1}^{\rho}\left(1 + \frac{\sigma_s^2\sigma_{h,i}^2}{\sigma^2}\right)\right), \log_2\left(\frac{\prod_{i=1}^{\rho}\sigma_{h,i}^2/(\sigma_{h,i}^2\sigma_s^2 + \sigma^2)}{(\sigma_{h,\rho}^{2\rho}/(\sigma_{h,\rho}^2\sigma_s^2 + \sigma^2)^\rho)}\right)\right\}, \tag{11}$$

then it holds that $\epsilon_{ce}(R) > \epsilon_{ec}(R)$, implying that the EC is more rate efficient than CE. If $N = \rho$ and $\sigma_{h,1} = \ldots = \sigma_{h,\rho}$, then $\epsilon_{ce}(R) = \epsilon_{ec}(R)$ and consequently $D_{ec}(R) = D_{ce}(R)$.

We define the signal-to-noise ratio (SNR) as SNR $= \text{trace}(\mathbf{H}\mathbf{\Sigma}_{ss}\mathbf{H}^T)/(N\sigma^2)$, and compare in Fig. 3 the MMSE when estimating s using the CE and EC schemes. With $\mathbf{\Sigma}_{ss} = \sigma_s^2\mathbf{I}_p$, $p = 4$ and $N = 40$, we observe that beyond a threshold rate, the distortion of EC converges to J_o faster than that of CE, which corroborates Theorem 1. Notice also that the gap between the EC and CE curves for SNR $= 2$ is larger than the gap for SNR $= 4$. This is true because as the noise power increases, the portion of the rate allocated to noise terms in CE increases accordingly. However, thanks to the MMSE estimator, EC cancels part of the noise and utilizes more efficiently the available rate.

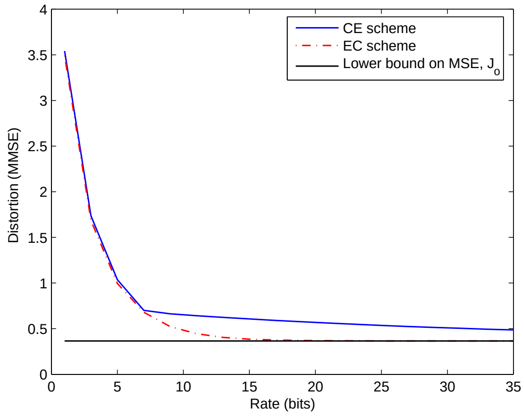
Our analysis so far raises the question whether EC is MSE optimal. We have shown that this is the case when estimating s with a given rate R and without forcing any assumption about N and p. A related claim has been reported in [9, 12] for $N = p$, but the extension to $N \neq p$ is not obvious. We have established that [10]:

Theorem 2: *The D-R function when estimating* $\mathbf{s}$ *based on* $\mathbf{x}$ *can be expressed as*

$$D_s(R) = \min_{\substack{p(\hat{\mathbf{s}}_R|\mathbf{x})\\ I(\mathbf{x};\hat{\mathbf{s}}_R)\leq R}} E[\|\mathbf{s} - \hat{\mathbf{s}}_R\|^2] = E[\|\tilde{\mathbf{s}}\|^2] + \min_{\substack{p(\hat{\mathbf{s}}_R|\hat{\mathbf{s}})\\ I(\hat{\mathbf{s}};\hat{\mathbf{s}}_R)\leq R}} E[\|\hat{\mathbf{s}} - \hat{\mathbf{s}}_R\|^2], \tag{12}$$

where $\hat{\mathbf{s}} = \mathbf{\Sigma}_{sx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{x}$ is the MMSE estimator, and $\tilde{\mathbf{s}}$ is the corresponding MMSE.

Theorem 2 reveals that the optimal means of estimating s is to first form the optimal MMSE estimate $\hat{\mathbf{s}}$ and then apply optimal rate-distortion encoding to this estimate. The lower bound on this distortion when $R \to \infty$, is $J_o = E[\|\tilde{\mathbf{s}}\|^2]$, which is intuitively appealing. The D-R function in (12) is achievable, because the rightmost term in (12) corresponds to the D-R function for reconstructing the MMSE estimate $\hat{\mathbf{s}}$ which is known to be achievable using random coding; see e.g., [1, p. 66].



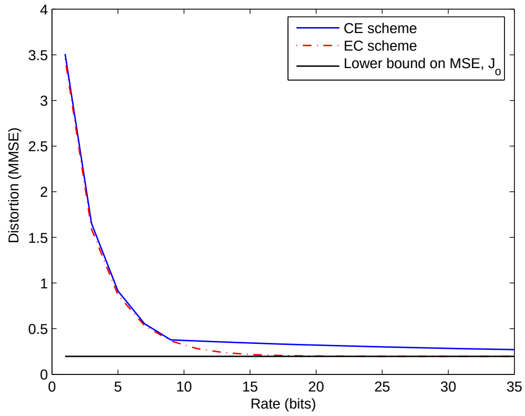


Figure 3: D-R region for EC and CE at SNR $= 2$ (Left) and SNR $= 4$ (Right).

4 Distortion-Rate for Distributed Estimation

Let us now consider the D-R function for estimating $\mathbf{s}$ in a multi-sensor setup, under a total available rate R which has to be shared among all sensors. Because analytical specification of the D-R function in this case remains intractable, we will develop an alternating algorithm that numerically determines an achievable upper bound for it. Combining this upper bound with the non-achievable lower bound corresponding to an equivalent single-sensor setup, and applying the MMSE optimal EC scheme, will provide a (hopefully tight) region where the D-R function lies in. For simplicity in exposition, we confine ourselves to a two-sensor setup, but our results apply to any finite $L > 2$.

To this end, we consider the following single-letter characterization of the upper bound on the D-R function:

$$\bar{D}(R) = \min_{\substack{p(\mathbf{u}_1|\mathbf{x}_1), p(\mathbf{u}_2|\mathbf{x}_2), \hat{\mathbf{s}}_R \\ I(\mathbf{x};\mathbf{u}_1,\mathbf{u}_2) \leq R}} E_{p(\mathbf{s},\mathbf{u}_1,\mathbf{u}_2)}[\|\mathbf{s} - \hat{\mathbf{s}}_R(\mathbf{u}_1,\mathbf{u}_2)\|^2], \tag{13}$$

where the minimization is w.r.t. $\{p(\mathbf{u}_i|\mathbf{x}_i)\}_{i=1}^2$ and $\hat{\mathbf{s}}_R := \hat{\mathbf{s}}_R(\mathbf{u}_1,\mathbf{u}_2)$. Achievability of $\bar{D}(R)$ can be established by readily extending to the vector case the scalar results in [3]. To carry out the minimization in (13), we develop an alternating scheme whereby $\mathbf{u}_2$ is treated as side information that is available at the decoder when optimizing (13) w.r.t. $p(\mathbf{u}_1|\mathbf{x}_1)$ and $\hat{\mathbf{s}}_R(\mathbf{u}_1,\mathbf{u}_2)$. The side information $\mathbf{u}_2$ is considered as the output of an optimal rate-distortion encoder applied to $\mathbf{x}_2$ for estimating $\mathbf{s}$, without taking into account $\mathbf{x}_1$. Since $\mathbf{x}_2$ is Gaussian, the side information will have the form (c.f. Section 3.2) $\mathbf{u}_2 = \mathbf{Q}_2\mathbf{x}_2 + \boldsymbol{\zeta}_2$, where $\mathbf{Q}_2 \in \mathbb{R}^{k_2 \times N_2}$ and $k_2 \leq N_2$, due to the rate constrained encoding of $\mathbf{x}_2$. Recall that the $k_2 \times 1$ vector $\boldsymbol{\zeta}_2$ is uncorrelated with $\mathbf{x}_2$ and Gaussian; i.e., $\boldsymbol{\zeta}_2 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\zeta_2\zeta_2})$.

Based on $\boldsymbol{\psi} := [\mathbf{x}_1^T \;\; \mathbf{u}_2^T]^T$, the optimal estimator for $\mathbf{s}$ is the MMSE one: $\hat{\mathbf{s}} = E[\mathbf{s}|\mathbf{x}_1,\mathbf{u}_2] = \boldsymbol{\Sigma}_{s\psi}\boldsymbol{\Sigma}_{\psi\psi}^{-1}\boldsymbol{\psi} = \mathbf{L}_1\mathbf{x}_1 + \mathbf{L}_2\mathbf{u}_2$, where $\mathbf{L}_1, \mathbf{L}_2$ are $p \times N_1$ and $p \times k_2$ matrices such that $\boldsymbol{\Sigma}_{s\psi}\boldsymbol{\Sigma}_{\psi\psi}^{-1} = [\mathbf{L}_1 \;\; \mathbf{L}_2]$. If $\tilde{\mathbf{s}}$ is the corresponding MSE, then $\mathbf{s} = \hat{\mathbf{s}} + \tilde{\mathbf{s}}$, where $\tilde{\mathbf{s}}$ is uncorrelated with $\boldsymbol{\psi}$ due to the orthogonality principle. Noticing also that $\hat{\mathbf{s}}_R(\mathbf{u}_1,\mathbf{u}_2)$ is uncorrelated with $\tilde{\mathbf{s}}$ because it is a function of $\mathbf{x}_1$ and $\mathbf{u}_2$, we have $E[\|\mathbf{s} - \hat{\mathbf{s}}_R(\mathbf{u}_1,\mathbf{u}_2)\|^2] = E[\|\hat{\mathbf{s}} - \hat{\mathbf{s}}_R(\mathbf{u}_1,\mathbf{u}_2)\|^2] + E[\|\tilde{\mathbf{s}}\|^2]$, or,

$$E[\|\mathbf{s} - \hat{\mathbf{s}}_R(\mathbf{u}_1,\mathbf{u}_2)\|^2] = E[\|\mathbf{L}_1\mathbf{x}_1 - (\hat{\mathbf{s}}_R(\mathbf{u}_1,\mathbf{u}_2) - \mathbf{L}_2\mathbf{u}_2)\|^2] + E[\|\tilde{\mathbf{s}}\|^2]. \tag{14}$$

Clearly, it holds that $I(\mathbf{x};\mathbf{u}_1,\mathbf{u}_2) = R_2 + I(\mathbf{x}_1;\mathbf{u}_1) - I(\mathbf{u}_2;\mathbf{u}_1)$, where $R_2 := I(\mathbf{x};\mathbf{u}_2)$ is the rate consumed to form the side information $\mathbf{u}_2$ and the rate constraint in (13) becomes

$I(\mathbf{x};\mathbf{u}_1,\mathbf{u}_2) \leq R \Leftrightarrow I(\mathbf{x}_1;\mathbf{u}_1) - I(\mathbf{u}_2;\mathbf{u}_1) \leq R - R_2 := R_1$. The new signal of interest in (14) is $\mathbf{L}_1\mathbf{x}_1$; thus, $\mathbf{u}_1$ has to be a function of $\mathbf{L}_1\mathbf{x}_1$. Then, $\mathbf{x}_1 \to \mathbf{L}_1\mathbf{x}_1 \to \mathbf{u}_1$, constitutes a Markov chain, which implies that $I(\mathbf{x}_1;\mathbf{u}_1) \leq I(\mathbf{L}_1\mathbf{x}_1;\mathbf{u}_1)$. Using the latter, we obtain

$$I(\mathbf{x}_1;\mathbf{u}_1) - I(\mathbf{u}_2;\mathbf{u}_1) \leq I(\mathbf{L}_1\mathbf{x}_1;\mathbf{u}_1) - I(\mathbf{u}_2;\mathbf{u}_1). \tag{15}$$

From the RHS of (15), we deduce the stricter constraint $I(\mathbf{L}_1\mathbf{x}_1;\mathbf{u}_1) - I(\mathbf{u}_2;\mathbf{u}_1) \leq R_1$. Combining the latter with (14) and (13), we arrive at the D-R upper bound

$$\bar{\bar{D}}(R_1) = \min_{\substack{p(\mathbf{u}_1|\mathbf{L}_1\mathbf{x}_1),\hat{\mathbf{s}}_R \\ I(\mathbf{L}_1\mathbf{x}_1;\mathbf{u}_1)-I(\mathbf{u}_1;\mathbf{u}_2)\leq R_1}} E[\|\mathbf{L}_1\mathbf{x}_1 - (\hat{\mathbf{s}}_R(\mathbf{u}_1,\mathbf{u}_2) - \mathbf{L}_2\mathbf{u}_2)\|^2] + E[\|\tilde{\mathbf{s}}\|^2], \tag{16}$$

through which we can determine an achievable D-R region, having available rate R_1 at the encoder and side information $\mathbf{u}_2$ at the decoder. Since $\mathbf{x}_1$ and $\mathbf{u}_2$ are jointly Gaussian, we can apply the Wyner-Ziv result [13], which allows us to consider that $\mathbf{u}_2$ is available both at the decoder and the encoder. This, in turn, permits re-writing the first expectation in (16) as:

$$\min_{\substack{p(\hat{\mathbf{s}}_R|\mathbf{L}_1\mathbf{x}_1,\mathbf{u}_2) \\ I(\mathbf{L}_1\mathbf{x}_1;\hat{\mathbf{s}}_R|\mathbf{u}_2)\leq R_1}} E[\|\mathbf{L}_1\mathbf{x}_1 - [\hat{\mathbf{s}}_R(\mathbf{u}_1,\mathbf{u}_2) - \mathbf{L}_2\mathbf{u}_2]\|^2]. \tag{17}$$

If $\hat{\mathbf{s}}_1 := E[\mathbf{L}_1\mathbf{x}_1|\mathbf{u}_2] = \mathbf{L}_1\boldsymbol{\Sigma}_{x_1u_2}\boldsymbol{\Sigma}_{u_2u_2}^{-1}\mathbf{u}_2$ and $\tilde{\mathbf{s}}_1$ is the corresponding MSE, then we can write $\mathbf{L}_1\mathbf{x}_1 = \hat{\mathbf{s}}_1 + \tilde{\mathbf{s}}_1$. For the rate constraint in (17), we have

$$I(\mathbf{L}_1\mathbf{x}_1;\hat{\mathbf{s}}_R|\mathbf{u}_2) = I(\mathbf{L}_1\mathbf{x}_1 - \hat{\mathbf{s}}_1;\hat{\mathbf{s}}_R - \mathbf{L}_2\mathbf{u}_2 - \hat{\mathbf{s}}_1|\mathbf{u}_2) = I(\tilde{\mathbf{s}}_1;\hat{\mathbf{s}}_R - \mathbf{L}_2\mathbf{u}_2 - \hat{\mathbf{s}}_1), \tag{18}$$

where the first equality is true because $\mathbf{u}_2$ is known; while the second one holds since $\mathbf{u}_2$ is uncorrelated with $\tilde{\mathbf{s}}_1$, due to the orthogonality principle, and likewise $\mathbf{u}_2$ is uncorrelated with $\hat{\mathbf{s}}_{R,12}(\mathbf{u}_1,\mathbf{u}_2) := \hat{\mathbf{s}}_R(\mathbf{u}_1,\mathbf{u}_2) - \mathbf{L}_2\mathbf{u}_2 - \hat{\mathbf{s}}_1$. Utilizing (17) and (18), we arrive at:

$$\bar{\bar{D}}(R_1) = \min_{\substack{p(\hat{\mathbf{s}}_{R,12}|\tilde{\mathbf{s}}_1) \\ I(\tilde{\mathbf{s}}_1;\hat{\mathbf{s}}_{R,12})\leq R_1}} E[\|\tilde{\mathbf{s}}_1 - \hat{\mathbf{s}}_{R,12}(\mathbf{u}_1,\mathbf{u}_2)\|^2] + E[\|\tilde{\mathbf{s}}\|^2]. \tag{19}$$

Notice that (19) is the D-R function for reconstructing the MSE $\tilde{\mathbf{s}}_1$ with rate R_1. Since $\tilde{\mathbf{s}}_1$ is Gaussian, we can readily apply rwf to the pre-whitened $\mathbf{Q}_{\tilde{s}_1}^T\tilde{\mathbf{s}}_1$ for determining $\bar{\bar{D}}(R_1)$ and the corresponding test channel that achieves $\bar{\bar{D}}(R_1)$. Through the latter, and considering the eigenvalue decomposition $\boldsymbol{\Sigma}_{\tilde{s}_1\tilde{s}_1} = \mathbf{Q}_{\tilde{s}_1}\text{diag}(\lambda_{\tilde{s}_1,1}\cdots\lambda_{\tilde{s}_1,p})\mathbf{Q}_{\tilde{s}_1}^T$, we find that the first encoder's output that minimizes (13) has the form

$$\mathbf{u}_1 = \mathbf{Q}_{\tilde{s}_1,k_1}^T\mathbf{L}_1\mathbf{x}_1 + \boldsymbol{\zeta}_1 = \mathbf{Q}_1\mathbf{x}_1 + \boldsymbol{\zeta}_1, \tag{20}$$

where $\mathbf{Q}_{\tilde{s}_1,k_1}$ denotes the first k_1 columns of $\mathbf{Q}_{\tilde{s}_1}$, k_1 is the number of $\mathbf{Q}_{\tilde{s}_1}^T\tilde{\mathbf{s}}_1$ entries that are assigned with non-zero rate, and $\mathbf{Q}_1 := \mathbf{Q}_{\tilde{s}_1,k_1}^T\mathbf{L}_1$. The $k_1 \times 1$ AWGN $\boldsymbol{\zeta}_1 \sim \mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}_{\zeta_1\zeta_1})$ is uncorrelated with $\mathbf{x}_1$. Additionally, we have $[\boldsymbol{\Sigma}_{\zeta_1\zeta_1}]_{ii} = \lambda_{\tilde{s}_1,i}D_i^1/(\lambda_{\tilde{s}_1,i} - D_i^1)$, where $D_i^1 = \left(\prod_{i=1}^{k_1}\lambda_{\tilde{s}_1,i}\right)^{1/k_1} 2^{-2R_1/k_1}$, for $i = 1,\ldots,k_1$, and $D_i^1 = \lambda_{\tilde{s}_1,i}$ when $i = k_1+1,\ldots,p$. This way, we are able to determine also $p(\mathbf{u}_1|\mathbf{x}_1)$. The reconstruction function has the form

$$\hat{\mathbf{s}}_R(\mathbf{u}_1,\mathbf{u}_2) = \mathbf{Q}_{\tilde{s}_1,k_1}\boldsymbol{\Theta}_1\mathbf{u}_1 - \mathbf{Q}_{\tilde{s}_1,k_1}\boldsymbol{\Theta}_1\mathbf{Q}_{\tilde{s}_1,k_1}^T\mathbf{L}_1\boldsymbol{\Sigma}_{x_1u_2}\boldsymbol{\Sigma}_{u_2u_2}^{-1}\mathbf{u}_2 + \mathbf{L}_1\boldsymbol{\Sigma}_{x_1u_2}\boldsymbol{\Sigma}_{u_2u_2}^{-1}\mathbf{u}_2 + \mathbf{L}_2\mathbf{u}_2, \tag{21}$$

where $[\boldsymbol{\Theta}_1]_{ii} = \lambda_{\tilde{s}_1,i}D_i^1/(\lambda_{\tilde{s}_1,i} - D_i^1)$, and the MMSE is $\bar{\bar{D}}(R_1) = \sum_{j=1}^{p} D_j^1 + E[\|\tilde{\mathbf{s}}\|^2]$.

The approach in this subsection can be applied in an alternating fashion from sensor to sensor in order to determine appropriate $p(\mathbf{u}_i|\mathbf{x}_i)$, for $i = 1, 2$, and $\hat{\mathbf{s}}_R(\mathbf{u}_1,\mathbf{u}_2)$ that at best globally

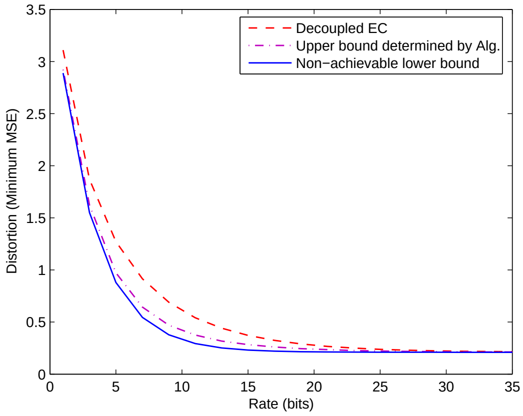


Figure 4: Distortion-rate bounds for estimating $\mathbf{s}$ in a two-sensor setup.

minimize (16). The conditional pdfs can be determined by finding the appropriate covariances $\mathbf{\Sigma}_{\zeta_i\zeta_i}$. Furthermore, by specifying the optimal $\mathbf{Q}_1$ and $\mathbf{Q}_2$, we have a complete characterization of the encoders' structure. The resultant algorithm is summarized next:

Initialize $\mathbf{Q}_1^{(0)}, \mathbf{Q}_2^{(0)}, \mathbf{\Sigma}_{\zeta_1\zeta_1}^{(0)}, \mathbf{\Sigma}_{\zeta_2\zeta_2}^{(0)}$ *by applying optimal D-R encoding to each sensor's test channel independently. For a total rate* R*, generate* M *random increments* $\{r(m)\}_{m=0}^{M}$*, such that* $0 \leq r(m) \leq R$ *and* $\sum_{m=0}^{M} r(m) = R$*. Set* $R_1(0) = R_2(0) = 0$*, and for* $j = 1, \ldots, M$*, set* $R(j) = \sum_{l=0}^{j} r(l)$

for $i = 1, 2$

$\bar{i} = \mod(i, 2) + 1$ *%The complementary index*

$R_0(j) = I(\mathbf{x}; \mathbf{u}_{\bar{i}}^{(j)})$

We use $\mathbf{Q}_{\bar{i}}^{(j-1)}, \mathbf{\Sigma}_{\zeta_{\bar{i}}\zeta_{\bar{i}}}^{(j-1)}, R(j), R_0(j)$ *to determine* $\mathbf{Q}_i^{(j)}$, $\mathbf{\Sigma}_{\zeta_i\zeta_i}^{(j)}$ *and distortion* $\bar{\bar{D}}(R_i(j))$

end

Update matrices $\mathbf{Q}_l^{(j)}, \mathbf{\Sigma}_{\zeta_l\zeta_l}^{(j)}$ *that result the smallest distortion* $\bar{\bar{D}}(R_l(j))$*, with* $l \in [1, 2]$

Set $R_l(j) = R(j) - I(\mathbf{x}; \mathbf{u}_{\bar{l}}^{(j)})$ *and* $R_{\bar{l}}(j) = I(\mathbf{x}; \mathbf{u}_{\bar{l}}^{(j)})$.

In Fig. 4, we plot the non-achievable lower bound which corresponds to one sensor having available the entire $\mathbf{x}$ and using the optimal EC scheme. Moreover, we plot an achievable D-R upper bound determined by letting the i-th sensor form its local estimate $\hat{\mathbf{s}}_i = E[\mathbf{s}|\mathbf{x}_i]$, and then apply optimal rate-distortion encoding to $\hat{\mathbf{s}}_i$. If $\hat{\mathbf{s}}_{R,1}$ and $\hat{\mathbf{s}}_{R,2}$ are the reconstructed versions of $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$, respectively, then the decoder at the FC forms the final estimate $\hat{\mathbf{s}}_R = E[\mathbf{s}|\hat{\mathbf{s}}_{R,1}, \hat{\mathbf{s}}_{R,2}]$. We also plot the achievable D-R region determined numerically by the alternating algorithm. For each rate, we keep the smallest distortion returned after 500 executions of the algorithm simulated with $\mathbf{\Sigma}_{ss} = \mathbf{I}_p$, $p = 4$, and $N_1 = N_2 = 20$, at SNR $= 2$. We observe that the algorithm provides a tight upper bound for the achievable D-R region. Using also the non-achievable lower bound (solid line), we have effectively reduced the 'uncertainty region' where the D-R function lies.

5 Conclusions

We determined the D-R function for estimating a random vector in a single-sensor setup and established the optimality of the estimate-first compress-afterwards (EC) approach along with the suboptimality of a compress-first estimate-afterwards (CE) alternative. When it comes to estimation using multiple sensors, the corresponding D-R function can be bounded from below using the single-sensor D-R function achieved using the EC scheme. An alternating algorithm was also derived for determining numerically an achievable D-R upper bound in the distributed multi-sensor setup. Simulations demonstrated that the numerically determined upper bound is more tight than analytically found alternatives. Using this upper bound in combination with the non-achievable lower bound we have obtained a tight region, where the D-R function for distributed estimation lies in.

Issues of interest not accounted by our analysis include the incorporation of fading channels with additive noise at the FC, and general (possibly non-linear) dynamical data models where the distribution of the observation data is no longer stationary or Gaussian.[1]

References

[1] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Prentice Hall, 1971.

[2] T. Berger, "Multiterminal Source Coding," *Lectures Presented at CISM Summer School on the Info. Theory Approach to Comm.*, July 1977.

[3] J. Chen, X. Zhang, T. Berger and S. B. Wicker, "An Upper Bound on the Sum-Rate Distortion Function and Its Corresponding Rate Allocation Schemes for the CEO Problem," *IEEE Journal on Sel. Areas in Comm.*, pp. 406–411, August 2004.

[4] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd Ed., John Wiley and Sons, 1991.

[5] M. Gastpar, P. L. Draggoti and M. Vetterli, "The Distributed Karhunen-Loève Transform," submitted to *IEEE Trans. on Info. Theory*, November 2004; http://www.eecs.berkeley.edu/∼gastpar/

[6] P. Ishwar, R. Puri, K. Ramchadran and S.S. Pradhan, "On Rate-Constrained Distributed Estimation in Unreliable Sensor Networks," *IEEE Journal on Sel. Areas in Comm.*, pp. 765–775, April 2005.

[7] Y. Oohama, "The Rate-Distortion Function for the Quadratic Gaussian CEO Problem," *IEEE Trans. On Info. Theory*, pp. 1057–1070, May 1998.

[8] A. Pandya, A. Kansal, G. Pottie and M. Srivastava, "Fidelity and Resource Sensitive Data Gathering,"*Proc. of the 42nd Allerton Conference,* Allerton, IL, September 2004.

[9] D. J. Sakrison, "Source encoding in the presence of random disturbance," *IEEE Trans. on Info. Theory*, pp. 165–167, January 1968.

[10] I. D. Schizas and G. B. Giannakis, "Distortion-Rate Analysis for Distributed Estimation with Wireless Sensor Networks," submitted to *IEEE Trans. on Info. Theory*, 2005; http://spincom.ece.umn.edu/

[11] H. Viswanathan and T. Berger, "The Quadratic Gaussian CEO Problem," *IEEE Trans. on Info. Theory*, pp. 1549–1559, September 1997.

[12] J. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Trans. on Info. Theory*, pp. 406–411, July 1970.

[13] A. Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder," *IEEE Trans. on Info. Theory*, No. 1, pp. 1–10, January 1976.

[1]The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies of the Army Research Laboratory or the U. S. Government.