Distortion-Rate for Non-Distributed and Distributed Estimation Problems

Ioannis Schizas, 3332862
ECE Department, University of Minnesota
May 7, 2005

Abstract

Wireless sensor networks are characterized by strict energy and bandwidth constraints, which motivate efficient compression schemes. The distortion incurred due to compression is an important metric for determining performance of reconstruction or estimation algorithms implemented at the fusion center based on observations collected at individual sensors. For the single-sensor link with the fusion center (non-distributed case), the distortion-rate (D-R) function when reconstructing Gaussian observations has been fully characterized. But this is not true for signal estimation problems with either single-sensor, or, multi-sensor observations (distributed case). In this work, we derive novel D-R results pertaining to signal estimation in a centralized point-to-point link, and we offer an interesting extension of an iterative procedure for numerically determining strict (achievable) upper bounds on the D-R region in a distributed estimation setup.

I. Introduction

Bandwidth and energy constraints in sensor networks call for efficient compression and encoding schemes.With a prescribed rate, it is of paramount importance to determine bounds on the minimum achievable distortion. Through these bounds, we can assess the loss in quality of the received data at the fusion center under pre-specified bandwidth constraints. When it comes to compressing and reconstructing sensor observations the best analytical inner and outer bounds for the D-R region can be found in [8]. An iterative scheme is also developed in [1] for determining an achievable D-R region, or, at best the D-R function for the distributed reconstruction setup.

A related problem emerges when the sensor observation $\mathbf{x}$ entail a random signal (or parameter vector) $\mathbf{s}$ that we wish to estimate in the presence of additive noise $\mathbf{n}$. Most frequently in practice, the data adhere to the linear model $\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}$, where $\mathbf{H}$ is a known deterministic matrix, while the signal $\mathbf{s}$ and the noise $\mathbf{n}$ are uncorrelated and Gaussian. The D-R function when $\mathbf{H}$ corresponds to an all-one vector and $\mathbf{s}$ is scalar has been treated in [5], [4], [9] and [10], and is known as the CEO problem; see also [11]. An interesting question is whether it is better from a distortion perspective to first compress sensor observations $\mathbf{x}$ and then use them to estimate $\mathbf{s}$, or, to first form an estimate $\hat{\mathbf{s}}_\infty$ based $\mathbf{x}$ and then compress this estimate. For the single-sensor case, we have obtained strict results addressing this question. Concerning the distributed case, we follow an approach similar to [1] in order to numerically determine an achievable D-R region, or, at best the D-R function for estimating $\mathbf{s}$.

This work is organized as follows. In Section II, we provide preliminaries concerning the D-R function in the reconstruction setup, which is pertinent to the method in [1]. In Section III-A, we consider the D-R function for non-distributed estimation in point-to-point links (single-sensor case), where we conclude that first estimating and then compressing is optimal. This offers an interesting extension of [6] and [7] for the estimation setup. Continuing with the distributed case in Section III-B, we find that treating all but one encoder's outputs as side information establishes optimality of the estimate-first and compress-afterwards approach. We further derive an iterative algorithm for determining an achievable D-R region for the estimation setup under rate constraints which offers a novel extension of [1] to the estimation setup. Finally, we summarize our results in Section IV.

II. Distortion-Rate for Reconstruction

A. Single-sensor setup

Considering the mean-square error (MSE) as a distortion metric, we provide here basic definitions for the D-R function in three cases:

i) The D-R function when encoding an $N \times 1$ real vector $\mathbf{x}$ at an individual sensor and reconstruct it at the fusion center as $\hat{\mathbf{x}}$, under a rate constraint R, is given by [3]

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

Fig. 1. *Left*: Test channel for $\mathbf{x}$ Gaussian; *Right*: Distributed setup.

where the minimization is carried with respect to (w.r.t.) the conditional pdf $p(\hat{\mathbf{x}}|\mathbf{x})$.
ii) If side information $\mathbf{u}_2 \in \mathbb{R}^{k_2}$ is available at the decoder, the D-R function is provided by [3], [2]

$$D^*(R) = \min_{\substack{p(\mathbf{u}_1|\mathbf{x}),\ \mathbf{f} \\ s.t.\ I(\mathbf{x};\mathbf{u}_1) - I(\mathbf{u}_2;\mathbf{u}_1) \leq R}} E_{p(\mathbf{x},\mathbf{u}_1,\mathbf{u}_2)} \|\mathbf{x} - \mathbf{f}(\mathbf{u}_1,\mathbf{u}_2)\|^2, \tag{2}$$

where $\mathbf{u}_1 \in \mathbb{R}^{k_1}$ denotes the encoder output and the minimization is w.r.t. both $p(\mathbf{u}_1|\mathbf{x})$ and the reconstruction function $\hat{\mathbf{x}} = \mathbf{f}(\mathbf{u}_1, \mathbf{u}_2)$, where $\mathbf{f}(\mathbf{u}_1, \mathbf{u}_2) : \mathbb{R}^{k_1} \times \mathbb{R}^{k_2} \to \mathbb{R}^N$.
iii) The D-R function when $\mathbf{u}_2$ is available *both* at the encoder and the decoder is [2]

$$D_{\mathbf{x}|\mathbf{u}_2}(R) = \min_{\substack{p(\hat{\mathbf{x}}|\mathbf{x},\mathbf{u}_2) \\ s.t.\ I(\mathbf{x};\hat{\mathbf{x}}|\mathbf{u}_2) \leq R}} E_{p(\hat{\mathbf{x}},\mathbf{x},\mathbf{u}_2)} \|\mathbf{x} - \hat{\mathbf{x}}\|^2. \tag{3}$$

Consider now that $\mathbf{x}$ is Gaussian, i.e., $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{xx})$, and let $\mathbf{\Sigma}_{xx} = \mathbf{Q}_x \mathbf{\Lambda}_x \mathbf{Q}_x^T$ be the eigenvalue decomposition of its covariance matrix $\mathbf{\Sigma}_{xx}$, where $\mathbf{\Lambda}_x := \text{diag}(\lambda_{x,1}, \cdots, \lambda_{x,N})$, is a diagonal matrix with the nonzero entries ordered as: $\lambda_{x,1} \geq \cdots \geq \lambda_{x,N} > 0$. The D-R function in (1) can now be determined by applying the reverse-water-filling (rwf) principle to the pre-whitened vector $\mathbf{x}_w = \mathbf{Q}_x^T \mathbf{x}$. Notice that since $\mathbf{Q}_x$ is orthogonal, the D-R function under the MSE metric for $\mathbf{x}_w$ coincides with that of $\mathbf{x}$. If the rate is constrained, i.e., $I(\mathbf{x}; \hat{\mathbf{x}}) \leq R$ then $\exists\ k$ such that the distortion for the ith entry of $\mathbf{x}_w$ is expressed as

$$D_i = \begin{cases} \mu(k, R) & \text{for } i = 1, \cdots, k \\ \lambda_{x,i} & \text{for } i = k+1, \cdots, N \end{cases}, \tag{4}$$

while the total distortion is $D(R) = k\mu(k, R) + \sum_{i=k+1}^{N} \lambda_{x,i}$. Since the rate can be expressed, according to the rwf principle, as $R = (1/2) \sum_1^k \log_2 (\lambda_{x,i}/\mu)$, it follows that $\mu(k, R) = \left(\prod_{i=1}^{k} \lambda_{x,i}\right)^{1/k} 2^{-2R/k}$, where k is the largest integer in the interval $\{1, \ldots, N\}$ such that $\mu(k, R) \leq \lambda_{x,i}$ for $i = 1, \ldots, k$.

With reference to Fig. 1 (left), we can use the matrices

$$\begin{aligned} \mathbf{A} &= \text{diag}\left((\lambda_{x,1} - D_1)/\lambda_{x,1}, \cdots, (\lambda_{x,N} - D_N)/\lambda_{x,N}\right) = \text{diag}\left(\mathbf{A}_k, \mathbf{0}_{N-k}\right) \\ \mathbf{\Sigma}_{zz} &= \text{diag}\left(\lambda_{x,1} D_1/(\lambda_{x,1} - D_1), \cdots, \lambda_{x,N} D_N/(\lambda_{x,N} - D_N)\right) \end{aligned}. \tag{5}$$

to construct a test channel, as in [3], for which the D-R function (1) is achieved. Eq. (5) implies that among the N parallel Gaussian channels, only the first k are active. The remaining $N-k$ channels are inactive in the sense that rwf does not assign any rate to them. This can be seen also from the fact that for $i > k$, we have $\mathbf{A}(i, i) = 0$ and $\mathbf{\Sigma}_{zz}(i, i) = \infty$; thus, the last $N-k$ channels transmit no information. The latter can be confirmed by the fact that the last $N-k$ elements of $\mathbf{y} = \mathbf{A}\mathbf{Q}_x^T \mathbf{x} + \mathbf{A}\mathbf{z}$, are always zero. To check the validity of this test channel, let us express the reconstructed vector as $\hat{\mathbf{x}} = \mathbf{Q}_x \mathbf{y} = \mathbf{Q}_x \mathbf{A} \mathbf{Q}_x^T \mathbf{x} + \mathbf{Q}_x \mathbf{A} \mathbf{z} = \mathbf{Q}_{x,k} \mathbf{A}_k \mathbf{Q}_{x,k}^T \mathbf{x} + \mathbf{Q}_{x,k} \mathbf{A}_k \mathbf{z}_k$, where $\mathbf{Q}_{x,k}$ ($\mathbf{Q}_{x,N-k}$) denotes the first k (last $N-k$) columns of $\mathbf{Q}_x$, and $\mathbf{z}_k$ are the first k entries of $\mathbf{z}$. It can be readily verified that $E\|\mathbf{x} - \hat{\mathbf{x}}\|^2 = D(R) = k\mu(k, R) + \sum_{i=k+1}^{N} \lambda_{x,i}$ and that $I(\mathbf{x}; \hat{\mathbf{x}}) = R$, which shows that indeed the D-R function (5) is achievable for $\mathbf{x}$ Gaussian.

B. Multisensor Setup

For brevity, we will consider only two sensors/encoders communicating to a fusion center, but the results can be easily extended to more than two sensors. Consider the setup where sensor i observes the $N_i \times 1$ vector $\mathbf{x}_i$, $i = 1, 2$. Each sensor encodes its observations under a total rate constraint R and sends the encoded information to the decoder through an ideal channel. At the decoder the $N \times 1$ vector $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T]^T$ is to be reconstructed as $\hat{\mathbf{x}}$; see also Fig. 1 (right). We clearly have $N = N_1 + N_2$, and assume that $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{xx})$, where in block matrix form $\mathbf{\Sigma}_{xx} = \left[\mathbf{\Sigma}_{x_i x_j}\right]$ for $i, j = 1, 2$.

Determining the D-R region for such a setup is analytically intractable. A non-achievable lower bound can be found by pretending that the entire vector $\mathbf{x}$ is available per sensor, based on which we can apply the rwf solution. An upper bound can also be obtained by having each sensor independently and optimally compressing its observations without considering the cross-correlation matrix $\mathbf{\Sigma}_{x_1x_2}$. It will turn out that a tighter upper bound is possible numerically by the method that we will present here; see also [1].

Theorem 1 ([1]): *If for a fixed total rate R the auxiliary random vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ satisfy*[1] $p(\mathbf{u}_1, \mathbf{u}_2, \mathbf{x}) = p(\mathbf{u}_1|\mathbf{x}_1)p(\mathbf{u}_2|\mathbf{x}_2)p(\mathbf{x})$, *then the D-R function is*

$$D(R) = \min_{\substack{p(\mathbf{u}_i|\mathbf{x}_i) \ ,i=1,2 \ , \ \hat{\mathbf{x}} \\ s.t. \ I(\mathbf{x};\mathbf{u}_1,\mathbf{u}_2)\le R}} E_{p(\mathbf{x},\mathbf{u}_1,\mathbf{u}_2)} \|\mathbf{x} - \hat{\mathbf{x}}(\mathbf{u}_1, \mathbf{u}_2)\|^2. \tag{6}$$

A closed-form solution for the minimization problem in (6) appears impossible. The approach we follow is to consider that $\mathbf{u}_2$ is given, and then try to minimize (6) wrt $p(\mathbf{u}_1|\mathbf{x}_1)$. Let as denote the given $\mathbf{u}_2$ as $\mathbf{u}_{2g}$. This can be considered as the output of an optimal rate-distortion encoder applied to $\mathbf{x}_2$ without taking into account $\mathbf{x}_1$. Once again, the encoder is defined with the aid of a test channel. Specifically, a pre-whitened version of $\mathbf{y}$, denoted as $\tilde{\mathbf{y}} = \text{diag}(\mathbf{A}_k^{-1}, \mathbf{0}_{N-k})\mathbf{y}$, can be considered as the output of the encoder; see also Fig.1 (left). The scaling in the first k components of $\mathbf{y}$ is not important and can be ignored; but the fact that the last $N-k$ entries are zero, is important. Since $\mathbf{x}_2$ is Gaussian, an optimal rate-distortion encoding scheme would produce a vector $\mathbf{u}_{2g} = \mathbf{C}_2\mathbf{x}_2 + \mathbf{z}_2$, where $\mathbf{C}_2 \in \mathbb{R}^{k_2\times N_2}$. The $k_2 \times 1$ vector $\mathbf{z}_2$ is Gaussian distributed, $\mathbf{z}_2 \sim \mathcal{N}\ (\mathbf{0}, \mathbf{\Sigma}_{z_2z_2})$, and is independent of $\mathbf{x}_2$. The fact that $k_2 \le N_2$, indicates the reduced dimensionality of the encoder output due to rate constraints. Given $\mathbf{u}_2$, we have that $I(\mathbf{x};\mathbf{u}_1,\mathbf{u}_{2g}) = I(\mathbf{x};\mathbf{u}_{2g}) + I(\mathbf{x}_1;\mathbf{u}_1|\mathbf{u}_{2g}) + I(\mathbf{x}_2;\mathbf{u}_1|\mathbf{x}_1;\mathbf{u}_{2g}) = R_0 + I(\mathbf{x}_1;\mathbf{u}_1|\mathbf{u}_{2g})$, where $R_0 := I(\mathbf{x};\mathbf{u}_{2g})$. In addition, we have that $I(\mathbf{x}_1;\mathbf{u}_1|\mathbf{u}_{2g}) = I(\mathbf{x}_1;\mathbf{u}_1) - I(\mathbf{u}_{2g};\mathbf{u}_1)$, since $\mathbf{u}_1$ is a function of $\mathbf{x}_1$. We now need to find

$$D(R_1) = \min_{\substack{p(\mathbf{u}_1|\mathbf{x}_1) \ and \ \hat{\mathbf{x}} \\ s.t. \ I(\mathbf{x}_1;\mathbf{u}_1)-I(\mathbf{u}_{2g};\mathbf{u}_1)\le(R-R_0)=R_1}} E_{p(\mathbf{x},\mathbf{u}_1,\mathbf{u}_{2g})} \|\mathbf{x} - \hat{\mathbf{x}}(\mathbf{u}_1, \mathbf{u}_{2g})\|^2. \tag{7}$$

Since the available information is $\mathbf{x}_1$ and $\mathbf{u}_{2g}$, the best estimate we can have for $\mathbf{x}$ is the MMSE estimator

$$\hat{\mathbf{x}}_\infty = E[\mathbf{x}|\mathbf{x}_1, \mathbf{u}_{2g}] = \left[\mathbf{\Sigma}_{xx_1}\mathbf{\Sigma}_{xu_{2g}}\right] \cdot \underbrace{\begin{bmatrix} \mathbf{\Sigma}_{x_1x_1} & \mathbf{\Sigma}_{x_1u_{2g}} \\ \mathbf{\Sigma}_{u_{2g}x_1} & \mathbf{\Sigma}_{u_2u_{2g}} \end{bmatrix}^{-1}}_{\check{\mathbf{\Sigma}}} = \underbrace{\begin{pmatrix} \mathbf{I}_{N_1} \\ \mathbf{\Gamma}_1 \end{pmatrix} \mathbf{x}_1}_{\hat{\mathbf{x}}_{1,\infty}} + \underbrace{\begin{pmatrix} \mathbf{0}_{N_1\times k_2} \\ \mathbf{\Gamma}_2 \end{pmatrix} \mathbf{u}_{2g}}_{\hat{\mathbf{x}}_{2,\infty}},$$

where the ∞ subscript signifies no compression (infinite rate), and $[\mathbf{\Gamma}_1 \ \ \mathbf{\Gamma}_2] = [\mathbf{\Sigma}_{x_2x_1} \ \ \mathbf{\Sigma}_{x_2u_{2g}}]\check{\mathbf{\Sigma}}$ with $\mathbf{\Gamma}_1$ a $N_2 \times N_1$ matrix and $\mathbf{\Gamma}_2$ a $N_2 \times k_2$ matrix. We can write $\mathbf{x} = \hat{\mathbf{x}}_\infty + \tilde{\mathbf{x}}_\infty$, where $\tilde{\mathbf{x}}_\infty$ is the MMSE, which is known to be independent of $\mathbf{x}_1$ and $\mathbf{u}_{2g}$. Taking advantage of the latter, we have $E\|\mathbf{x} - \hat{\mathbf{x}}(\mathbf{u}_1, \mathbf{u}_{2g})\|^2 = E\|\hat{\mathbf{x}}_\infty + \tilde{\mathbf{x}}_\infty - \hat{\mathbf{x}}(\mathbf{u}_1, \mathbf{u}_{2g})\|^2 = E\|\hat{\mathbf{x}}_\infty - \hat{\mathbf{x}}(\mathbf{u}_1, \mathbf{u}_{2g})\|^2 + E\|\tilde{\mathbf{x}}_\infty\|^2 = E\|\hat{\mathbf{x}}_{1,\infty} - \hat{\mathbf{x}}'(\mathbf{u}_1, \mathbf{u}_{2g})\|^2 + E\|\tilde{\mathbf{x}}_\infty\|^2$, where $\hat{\mathbf{x}}'(\mathbf{u}_1, \mathbf{u}_{2g}) = \hat{\mathbf{x}}(\mathbf{u}_1, \mathbf{u}_{2g}) - \hat{\mathbf{x}}_{2,\infty}$. Based on this expression for the distortion, we can rewrite (7) as

$$D(R_1) = \min_{\substack{p(\mathbf{u}_1|\hat{\mathbf{x}}_{1,\infty}) \ and \ \hat{\mathbf{x}} \\ s.t. \ I(\hat{\mathbf{x}}_{1,\infty};\mathbf{u}_1)-I(\mathbf{u}_{2g};\mathbf{u}_1)\le R_1}} E_{p(\hat{\mathbf{x}}_{1,\infty},\mathbf{u}_1,\mathbf{u}_{2g})} \|\hat{\mathbf{x}}_{1,\infty} - \hat{\mathbf{x}}'(\mathbf{u}_1, \mathbf{u}_{2g})\|^2 + E\|\tilde{\mathbf{x}}_\infty\|^2. \tag{8}$$

We can see that the D-R function is lower bounded by the MMSE. Using the fact that $\mathbf{x}_1$ and $\mathbf{u}_{2g}$ are jointly Gaussian, we can apply the Wyner-Ziv result [2], according to which the same D-R function results whether side information is available only at the encoder, or, both at the encoder and decoder; thus, (8) becomes [c.f. (3)]

$$D(R_1) = \min_{\substack{p(\hat{\mathbf{x}}'|\hat{\mathbf{x}}_{1,\infty},\mathbf{u}_{2g}) \\ s.t. \ I(\hat{\mathbf{x}}_{1,\infty};\hat{\mathbf{x}}'|\mathbf{u}_{2g})\le R_1}} E_{p(\hat{\mathbf{x}}_{1,\infty},\hat{\mathbf{x}}',\mathbf{u}_{2g})} \|\hat{\mathbf{x}}_{1,\infty} - \hat{\mathbf{x}}'\|^2 + E\|\tilde{\mathbf{x}}_\infty\|^2. \tag{9}$$

The next step is to remove the side information from the constraints so that we can readily apply rwf. We have $\hat{\mathbf{x}}_{1,\infty} = \hat{\mathbf{x}}'_{1,\infty} + \tilde{\mathbf{x}}_{1,\infty}$, where $\hat{\mathbf{x}}'_{1,\infty} = E[\hat{\mathbf{x}}_{1,\infty}|\mathbf{u}_{2g}] = \mathbf{\Sigma}_{\hat{x}_{1,\infty}u_{2g}}\mathbf{\Sigma}_{u_{2g}u_{2g}}^{-1}\mathbf{u}_{2g} = \mathbf{B}\mathbf{u}_{2g}$ and $\mathbf{B} = \mathbf{\Sigma}_{\hat{x}_{1,\infty}u_{2g}}\mathbf{\Sigma}_{u_{2g}u_{2g}}^{-1}$. Recall

[1]Intuitively, with $\mathbf{u}_i$ being a function of $\mathbf{x}_i$, we obatin a Markov chain $\mathbf{x}_j \to \mathbf{x}_i \to \mathbf{u}_i$ for $i \neq j$.

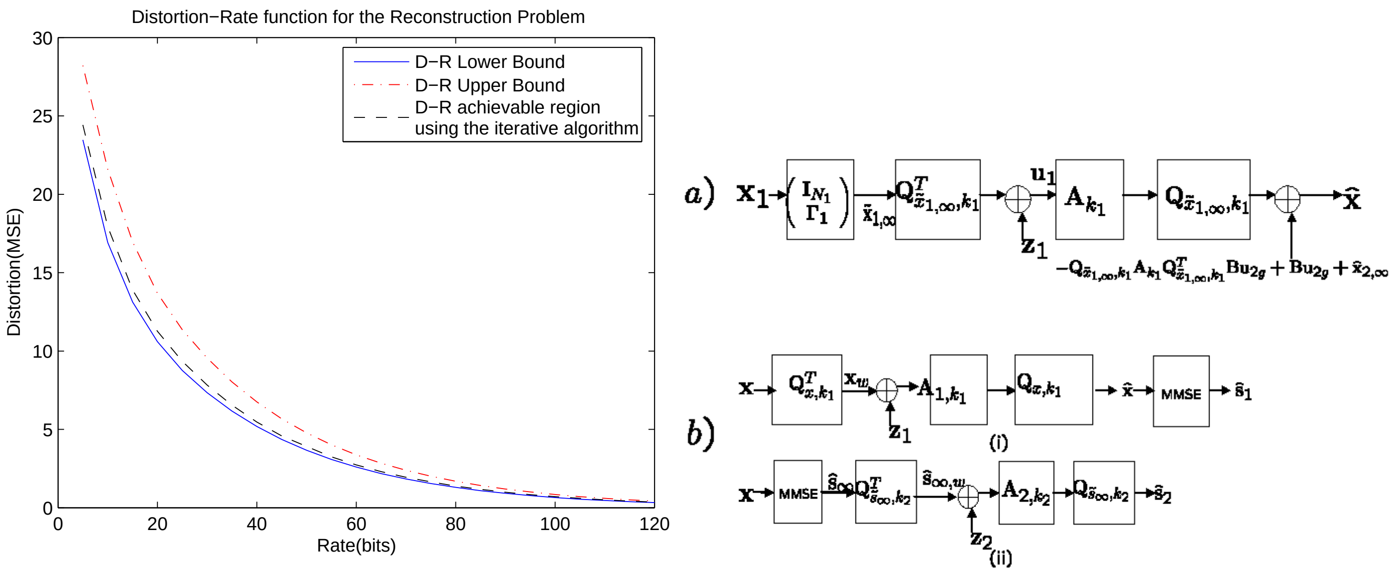

Fig. 2. *Left*: D-R region bounds for the reconstruction problem; *Right*: (a) Test channel with side information at the decoder; (b) Test channels for the two estimation schemes.

that $\tilde{\mathbf{x}}_{\infty,1}$ is the MMSE which is independent of $\hat{\mathbf{x}}'_{1,\infty}$. We can write $E\|\mathbf{x}_{1,\infty}-\hat{\mathbf{x}}'\|^2 = E\|\tilde{\mathbf{x}}_{1,\infty}-(\hat{\mathbf{x}}'-\mathbf{B}\mathbf{u}_{2g})\|^2 = E\|\tilde{\mathbf{x}}_{1,\infty}-\hat{\mathbf{x}}''\|^2$, where $\hat{\mathbf{x}}'' = \hat{\mathbf{x}}' - \mathbf{B}\mathbf{u}_{2g}$. Then, (8) becomes

$$
\begin{aligned}
D(R_1) &= \min_{\substack{p(\hat{\mathbf{x}}''|\tilde{\mathbf{x}}_{1,\infty},\mathbf{u}_{2g}) \\ s.t.\;\; I(\tilde{\mathbf{x}}_{1,\infty};\hat{\mathbf{x}}''|\mathbf{u}_{2g})\le R_1}} E_{p(\hat{\mathbf{x}}_{1,\infty},\hat{\mathbf{x}}'',\mathbf{u}_{2g})}\|\tilde{\mathbf{x}}_{1,\infty}-\hat{\mathbf{x}}''\|^2 + E\|\tilde{\mathbf{x}}_{\infty}\|^2 \\
&= \min_{\substack{p(\hat{\mathbf{x}}''|\tilde{\mathbf{x}}_{1,\infty}) \\ s.t.\;\; I(\tilde{\mathbf{x}}_{1,\infty};\hat{\mathbf{x}}'')\le R_1}} E_{p(\hat{\mathbf{x}}_{1,\infty},\hat{\mathbf{x}}'',\mathbf{u}_{2g})}\|\tilde{\mathbf{x}}_{1,\infty}-\hat{\mathbf{x}}''\|^2 + E\|\tilde{\mathbf{x}}_{\infty}\|^2,
\end{aligned}
\tag{10}
$$

where the last equality comes form the fact that the optimum $\hat{\mathbf{x}}''$ can be independent of $\mathbf{u}_{2g}$, since $\tilde{\mathbf{x}}_{1,\infty}$ is independent from $\hat{\mathbf{x}}''$. Let the eigenvalue decomposition for the MMSE covariance matrix be $\mathbf{\Sigma}_{\tilde{x}_{1,\infty}} = \mathbf{Q}_{\tilde{x}_{1,\infty}}\mathbf{\Lambda}_{\tilde{x}_{1,\infty}}\mathbf{Q}^T_{\tilde{x}_{1,\infty}}$, with $\mathbf{\Lambda}_{\tilde{x}_{1,\infty}} = \text{diag}(\lambda_{\tilde{x}_{1,\infty},1},\cdots,\lambda_{\tilde{x}_{1,\infty},N_1},0,...,0)$, and $\lambda_{\tilde{x}_{1,\infty},1}\ge\cdots\ge\lambda_{\tilde{x}_{1,\infty},N_1}>0$. According to subsection II-A, we have that $D(R_1) = k_1\mu(k_1,R_1)+\sum_{i=k_1+1}^{N}\lambda_{\tilde{x}_{1,\infty},i}+E\|\tilde{\mathbf{x}}_{\infty}\|^2$, where k_1 is the largest integer in $\{1,\cdots,N_1\}$ such that $\mu(k_1,R_1) = \left(\prod_{i=1}^{k_1}\lambda_{\tilde{x}_{1,\infty},i}\right)^{1/k_1} 2^{-2R/k_1} \le \lambda_{\tilde{x}_{1,\infty},i}$, for $i=1,\dots,k_1$.

Fig. 2 (Right (a)) depicts the test channel where we can see how $\hat{\mathbf{x}}$ is created by using both $\mathbf{x}_1$ and the available side information $\mathbf{u}_{2g}$ at the decoder. The output of the encoder $\mathbf{u}_1$ is $\mathbf{u}_1 = \mathbf{Q}^T_{\tilde{x}_{1,\infty},k_1}[\mathbf{I}_{N_1}\mathbf{\Gamma}_1^T]^T\mathbf{x}_1+\mathbf{z}_1 = \mathbf{C}_1\mathbf{x}_1+\mathbf{z}_1$, where $\mathbf{C}_1 = \mathbf{Q}^T_{\tilde{x}_{1,\infty},k_1}[\mathbf{I}_{N_1}\mathbf{\Gamma}_1^T]^T$. Let $\mathbf{Q}_{\tilde{x}_{1,\infty},k_1}$ denote the first k_1 columns of $\mathbf{Q}_{\tilde{x}_{1,\infty}}$. Notice that the $k_1\times 1$ vector $\mathbf{z}_1$ is independent of $\mathbf{x}_1$, and $\mathbf{z}_1\sim\mathcal{N}\left(\mathbf{0},\mathbf{\Sigma}_{z_1z_1}\right)$, offers the distribution $p(\mathbf{u}_1|\mathbf{x}_1)$ that minimizes (7). Specifically, we have that [c.f. (5)]

$$
\mathbf{A}_{k_1} = \text{diag}\left((\lambda_{\tilde{x}_{1,\infty},1}-D_1)/\lambda_{\tilde{x}_{1,\infty},1},\cdots,(\lambda_{\tilde{x}_{1,\infty},k_1}-D_{k_1})/\lambda_{\tilde{x}_{1,\infty},k_1}\right), \tag{11}
$$

$$
\mathbf{\Sigma}_{z_1z_1} = \text{diag}\left((\lambda_{\tilde{x}_{1,\infty},1}D_1)/(\lambda_{\tilde{x}_{1,\infty},1}-D_1),\cdots,(\lambda_{\tilde{x}_{1,\infty},k_1}D_{k_1})/(\lambda_{\tilde{x}_{1,\infty},k_1}-D_{k_1})\right), \tag{12}
$$

where D_i's are defined as in (4) using this time $\lambda_{\tilde{x}_{1,\infty},i}$ and $k=k_1$.

The approach in this subsection can be applied iteratively on a per sensor basis in order to determine appropriate $p(\mathbf{u}_i|\mathbf{x}_i)$ for $i=1,2$ and a function, which is specified by the test channel, $\hat{\mathbf{x}}(\mathbf{u}_1,\mathbf{u}_2)$ that at best globally minimizes (7). Generally, the following iterative procedure is guaranteed to converge only to a local minimum, thus determining an achievable D-R region and at best the D-R function.

Algorithm 1: *Initialization:* Determine[2] $\mathbf{C}_1^{(0)}, \mathbf{C}_2^{(0)}, \mathbf{\Sigma}_{z_1 z_1}^{(0)}, \mathbf{\Sigma}_{z_2 z_2}^{(0)}$ using the test channels that are produced by applying optimal distortion-rate encoding to each sensor independently. For a total rate constraint R, generate M random increments $r(i)\ \ i = 0, \cdots, M$, such that $0 \leq r(i) \leq R$ and $\sum_{i=0}^{M} r(i) = R$. Assign to each sensor rate $R_1(0) = R_2(0) = 0$.

$for\ \ j = 1, \cdots, M$

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 $\mathbf{C}_{\bar{i}}^{(j-1)}, \mathbf{\Sigma}_{z_{\bar{i}} z_{\bar{i}}}^{(j-1)}, R(j), R_0(j)$ to determine $\mathbf{C}_i^{(j)}$ and $\mathbf{\Sigma}_{z_i z_i}^{(j)}$ using the previously described procedure.

Determine the distortion $D_i^{(j)} = k_i^{(j)} D_i^{(j)} + \sum_{l=k_i^{(j)}+1}^{N_1} \lambda_{\tilde{x}_{1,\infty}, l}^{(j)} + E\|\tilde{\mathbf{x}}_{\infty}^{(j)}\|^2$

end

Find the sensor that results minimum distortion $D_l^{(j)}$ with $l = 1, 2$ and update corresponding matrices $\mathbf{C}_l^{(j)}, \mathbf{\Sigma}_{z_l z_l}^{(j)}$

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

end

In Fig. 2 (left), we plot the lower bound which is determined by assuming that joint encoding of $\mathbf{x}$ is possible. We plot also the upper bound which is obtained by having each sensor apply optimal rate-distortion encoding to its observation vector $\mathbf{x}_i$ independently from the other, and the achievable D-R region which is numerically determined by the algorithm. For a certain rate, we keep the smallest distortion returned after 500 executions of the algorithm. We have used $\mathbf{\Sigma}_{xx} =$ Toeplitz$([1, ..., \rho^N])$, with $\rho = 0.7$ and $N_1 = N_2 = 20$. We observe that the algorithm provides a strict upper bound for the achievable D-R region which is very close to the lower bound of the joint encoding scheme.

III. Distortion-Rate For Estimation

In the previous section, we determined the D-R function for reconstructing a vector $\mathbf{x}$ either jointly or in a distributed manner. In this section, we suppose $\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}$, where for the $p \times 1$ vector $\mathbf{s}$ we have $\mathbf{s} \sim N(\mathbf{0}, \mathbf{\Sigma}_{ss})$. The $N \times p$ matrix $\mathbf{H}$ is assumed full column rank and fixed. The $N \times 1$ vector $\mathbf{n}$ denotes additive white Gaussian noise (AWGN) independent of $\mathbf{s}$, and $\mathbf{n} \sim \mathcal{N}\ (\mathbf{0}, \sigma^2 \mathbf{I}_N)$. Clearly, it holds that $\mathbf{\Sigma}_{xx} = \mathbf{H}\mathbf{\Sigma}_{ss}\mathbf{H}^T + \sigma^2 \mathbf{I}_N$. We are interested in determining the distortion incurred when estimating $\mathbf{s}$ based on $\mathbf{x}$, under a rate constraint R. We will first compare the distortion of two different scenarios in the single-sensor case, where we: i) either encode-compress $\mathbf{x}$ using rwf and use the reconstructed vector $\hat{\mathbf{x}}$ to form the MMSE estimate $\hat{\mathbf{s}}_1 = E[\mathbf{s}|\hat{\mathbf{x}}]$; or, ii) form the MMSE estimate $\hat{\mathbf{s}}_\infty = E[\mathbf{s}|\mathbf{x}]$, encode-compress $\hat{\mathbf{s}}$ using rwf, and after decoding, obtain the reconstructed estimate $\hat{\mathbf{s}}_2$. We will henceforth refer to the first scheme as $C(ompress) - E(stimate)$, scheme and to the second scheme as $E(stimate) - C(ompress)$.

A. Single-sensor case

First we analyze the case where the entire vector $\mathbf{x}$ is available to one sensor. We will denote the estimation error for the first and second scheme as $\tilde{\mathbf{s}}_1 = \mathbf{s} - \hat{\mathbf{s}}_1$ and $\tilde{\mathbf{s}}_2 = \mathbf{s} - \hat{\mathbf{s}}_2$, respectively. For the C-E scheme we depict in Fig. 2(right (b)) the test channel for encoding $\mathbf{x}$, followed by MMSE estimation. We have $\hat{\mathbf{x}} = \mathbf{Q}_{x,k_1}\mathbf{A}_{1,k_1}\mathbf{Q}_{x,k_1}^T\mathbf{x} + \mathbf{Q}_{x,k_1}\mathbf{A}_{1,k_1}\mathbf{z}_1$. We consider the vector $\hat{\hat{\mathbf{x}}} = \mathbf{Q}_x^T\hat{\mathbf{x}} = \left[\hat{\hat{\mathbf{x}}}_1^T, \mathbf{0}_{1\times(N-k)}\right]^T$, where $\hat{\hat{\mathbf{x}}}_1 = \mathbf{A}_{1,k_1}\mathbf{Q}_{x,k_1}^T\mathbf{x} + \mathbf{A}_{1,k_1}\mathbf{z}$, and $\hat{\mathbf{s}}_1 = E[\mathbf{s}|\hat{\mathbf{x}}] = E[\mathbf{s}|\mathbf{Q}_x^T\hat{\mathbf{x}}] = E[\mathbf{s}|\hat{\hat{\mathbf{x}}}_1]$. The covariance matrix of $\tilde{\mathbf{s}}_1$ is given by

$$\begin{aligned}\mathbf{\Sigma}_{\tilde{s}_1\tilde{s}_1} = E[(\mathbf{s} - \hat{\mathbf{s}}_1)(\mathbf{s} - \hat{\mathbf{s}}_1)^T] = \mathbf{\Sigma}_{ss} - \mathbf{\Sigma}_{s\hat{\hat{x}}_1}\mathbf{\Sigma}_{\hat{\hat{x}}_1\hat{\hat{x}}_1}^{-1}\mathbf{\Sigma}_{\hat{\hat{x}}_1 s} = \\ \mathbf{\Sigma}_{ss} - \mathbf{\Sigma}_{sx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{xs} + \mathbf{\Sigma}_{sx}\mathbf{Q}_x\mathbf{\Delta}_1\mathbf{Q}_x^T\mathbf{\Sigma}_{xs},\end{aligned} \tag{13}$$

where $\mathbf{\Delta}_1 = \text{diag}\left(D_1^1\lambda_{x,1}^{-2} \cdots D_N^1\lambda_{x,N}^{-2}\right)$, and D_i^1 for $i = 1, \ldots, N$ is the distortion for the i-th element of $\mathbf{x}_{w,N} = \mathbf{Q}_x^T\mathbf{x}$, [c.f. (4)]. The integer $k_1 \in \{1, \cdots, N\}$ determines the number of elements of $\mathbf{x}_{w,N}$ that are going to be assigned with rate (c.f. Section II-A). The test channel for the E-C scheme is depicted in Fig. 2(Right (b)). Let us consider the eigenvalue decomposition $\mathbf{\Sigma}_{\hat{s}_\infty\hat{s}_\infty} = \mathbf{Q}_{\hat{s}_\infty}\mathbf{\Lambda}_{\hat{s}_\infty}\mathbf{Q}_{\hat{s}_\infty}^T$,

[2]Those matrices determine also $\mathbf{u}_1^{(0)}$ and $\mathbf{u}_2^{(0)}$.

where $\mathbf{\Lambda}_{\hat{s}_\infty} = \text{diag}(\lambda_{\hat{s}_\infty,1} \cdots \lambda_{\hat{s}_\infty,N})$ and $\lambda_{\hat{s}_\infty,1} \geq \cdots \geq \lambda_{\hat{s}_\infty,N} > 0$. The estimate of $\mathbf{s}$ is given by $\hat{\mathbf{s}}_2 = \mathbf{Q}_{\hat{s}_\infty,k_2}\mathbf{A}_{2,k_2}\mathbf{Q}^T_{\hat{s}_\infty,k_2}\hat{\mathbf{s}}_\infty + \mathbf{Q}_{\hat{s}_\infty,k_2}\mathbf{A}_{2,k_2}\mathbf{z}_2$, and the covariance matrix of $\tilde{\mathbf{s}}_2$ is

$$\mathbf{\Sigma}_{\tilde{s}_2\tilde{s}_2} = \mathbf{\Sigma}_{ss} - \mathbf{\Sigma}_{sx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{xs} + \mathbf{Q}_{\hat{s}_\infty}\mathbf{\Delta}_2\mathbf{Q}_{\hat{s}_\infty}, \tag{14}$$

where $\mathbf{\Delta}_2 = \text{diag}\left(D_1^2, \cdots, D_p^2\right)$ and D_i^2 for $i = 1, \cdots, p$ is the distortion for the i-th element for the vector $\mathbf{s}^p_{\infty,w} = \mathbf{Q}^T_{\hat{s}_\infty}\hat{\mathbf{s}}_\infty$ [c.f. (4)]. The integer $k_2 \in \{1, \cdots, p\}$ determines the number of elements of $\mathbf{s}^p_{\infty,w}$ that are going to be assigned with rate (c.f. Section II-A). The first k_2 components of $\mathbf{s}^p_{\infty,w}$ are denoted as $\mathbf{s}_{\infty,w}$. Consider the matrices $\mathbf{E}_1 = \mathbf{\Sigma}_{sx}\mathbf{Q}_x\mathbf{\Delta}_1\mathbf{Q}_x^T\mathbf{\Sigma}_{xs}$, $\mathbf{E}_2 = \mathbf{Q}_{\hat{s}_\infty}\mathbf{\Delta}_2\mathbf{Q}^T_{\hat{s}_\infty}$, and let $D_0 = \text{trace}(\mathbf{\Sigma}_{ss} - \mathbf{\Sigma}_{sx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{xs})$. Furthermore, for $i = 1, 2$ we have $d_i(R) = \text{trace}(\mathbf{\Sigma}_{\tilde{s}_i\tilde{s}_i}) = D_0 + \epsilon_i(R)$ with $\epsilon_i(R) = \text{trace}(\mathbf{E}_i)$. Notice that the distortion of both schemes contains a common term which is due to the MMSE incurred when estimating $\mathbf{s}$ based on $\mathbf{x}$, and it is achieved when $R \to \infty$. Our first result[3] compares the distortion terms $\epsilon_i(R)$.

Theorem 2: *If* $R > \max\left\{(1/2)\log_2\left(\left(\prod_{i=1}^p \lambda_{x,i}\right)/(\sigma^2)^p\right), (1/2)\log_2\left(\left(\prod_{i=1}^p \lambda_{\hat{s}_\infty,i}\right)\right)/(\lambda_{\hat{s}_\infty,p})^p)\right\}$, *then* $\epsilon_1(R) = \gamma_1 2^{-2R/N}$ *and* $\epsilon_2(R) = \gamma_2 2^{-2R/p}$, *where* γ_1 *and* γ_2 *are constants independent of R.*

An immediate consequence of Theorem 2 is that the distortion for the E-C scheme converges asymptotically to D_0 as $R \to \infty$ with rate $O(2^{-2R/p})$. The C-E scheme converges likewise, but with rate $O(2^{-2R/N})$. If $N \gg p$, then the E-C scheme is approaching the lower bound D_0 much faster than C-E, implying correspondingly a more efficient usage of the available rate R. Let us examine now some special cases for which we have obtained stronger results.

Scalar case ($p = 1$, $N = 1$): Here we have the model $x = hs + n$, where h is fixed, while s, n are independent 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 $\sigma^2_{\tilde{s}_1}$ and $\sigma^2_{\tilde{s}_2}$ denoting the variances of $\tilde{s}_1 = s - \hat{s}_1$ and $\tilde{s}_2 = s - \hat{s}_2$, respectively, we have shown that:

Proposition 1: *For the scalar case, we have that* $\sigma^2_{\tilde{s}_1} = \sigma^2_{\tilde{s}_2}$ *and hence the two schemes have identical D-R functions.*

Vector case ($p = 1$, $N > 1$): In this case, we have the model $\mathbf{x} = \mathbf{h}s + \mathbf{n}$. With $R_t = (1/2)\log_2\left(1 + (\sigma_s^2\|\mathbf{h}\|^2)/\sigma^2\right)$, we have established that

Proposition 2: *For* $R \leq R_t$ *it holds that* $\epsilon_1(R) = \epsilon_2(R)$ *and the two schemes have identical D-R. For* $R > R_t$, *we have that* $\epsilon_1(R) > \epsilon_2(R)$ *and thus the* $E - C$ *scheme uses more efficiently the available rate.*

Matrix case I ($N \geq p > 1$ 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 $\mathbf{\Sigma}_{xx} = \sigma_s^2\mathbf{H}\mathbf{H}^T + \sigma^2\mathbf{I}$. Let $\mathbf{H} = \mathbf{U}_h\mathbf{\Sigma}_h\mathbf{V}_h^T$ be the SVD of $\mathbf{H}$, where $\mathbf{U}_h : N \times N$, $\mathbf{V}_h : p \times p$, and $\mathbf{\Sigma}_h$ is an $N \times p$ diagonal matrix $\mathbf{\Sigma}_h = \text{diag}(\sigma_{h,1} \cdots \sigma_{h,p})$ with $\sigma_{h,1} \geq, \cdots, \geq \sigma_{h,p} > 0$. Then we have

Proposition 3: *If*

$$R > \max\left\{(1/2)\log_2\left(\prod_{i=1}^p(1 + (\sigma_s^2\sigma_{h,i}^2)/\sigma^2)\right), (1/2)\log_2\left(\left(\prod_{i=1}^p(\sigma_{h,i}^2/(\sigma_{h,i}^2\sigma_s^2 + \sigma^2))\right) / \left((\sigma_{h,p}^2)^p/(\sigma_{h,p}^2\sigma_s^2 + \sigma^2)^p\right)\right)\right\}, \tag{15}$$

and either $N > p$ *with* $\sigma_{h,1} \geq, \ldots, \geq \sigma_{h,p} > 0$, *or,* $N = p$ *with* $\sigma_{h,1} >, \cdots, > \sigma_{h,p} > 0$, *it holds that* $\epsilon_1(R) > \epsilon_2(R)$, *implying that the E-C is more rate efficient than C-E. If* $N = p$ *and* $\sigma_{h,1} =, \ldots, = \sigma_{h,p} > 0$, *then* $\epsilon_1(R) = \epsilon_2(R)$ *and the two schemes have identical distortions.*

Matrix case II ($N \geq p > 1$, $\mathbf{\Sigma}_{ss} = \text{diag}\left(\sigma_{s,1}^2, \cdots, \sigma_{s,p}^2\right)$ and $\mathbf{H}^T\mathbf{H} = \text{diag}(\|\mathbf{h}_1\|^2, \cdots, \|\mathbf{h}_p\|^2)$) : With the notation $\mathbf{H} = \bar{\mathbf{H}}\mathbf{D}_h$, where $\bar{\mathbf{H}}^T\bar{\mathbf{H}} = \mathbf{I}_p$ and $\mathbf{D}_h = \text{diag}(\|\mathbf{h}_1\|, \cdots, \|\mathbf{h}_p\|)$, we have proved that:

Proposition 4: *If*

$$R > \max\left\{(1/2)\log_2 \prod_{i=1}^p(1 + (\sigma_{s,i}^2\|\mathbf{h}_i\|^2)/\sigma^2)\right), (1/2)\log_2 \prod_{i=1}^p(\sigma_{s,i}^4\|\mathbf{h}_i\|^2/(\|\mathbf{h}_i\|^2\sigma_{s,i}^2 + \sigma^2))\right) / \ (\sigma_{s,p}^4\|\mathbf{h}_p\|^2)^p/(\|\mathbf{h}_p\|^2\sigma_{s,p}^2 + \sigma^2)^p \ \left.\right)\right\} \tag{16}$$

[3]Proof of this and subsequent results are provided in the Appendix.

and either $N > p$, or, $N = p$ and $\exists\ k$ such that $\|\mathbf{h}_k\| \neq \|\mathbf{h}_i\|$ for $i = 1, \ldots, p$ and $i \neq k$, it holds that $\epsilon_1(R) > \epsilon_2(R)$, implying that the E-C is more rate efficient than C-E. If $N = p$ and $\sigma_{h,1} =, \ldots, = \sigma_{h,p} > 0$, then $\epsilon_1(R) > \epsilon_2(R)$, and hence the E-C scheme is more rate-efficient than the C-E one. If $N = p$ and $\|\mathbf{h}_i\| = \|\mathbf{h}\|$ for $i = 1, \ldots, p$ then $\epsilon_1(R) = \epsilon_2(R)$, and the two schemes have the same distortion.

In Fig. 3, we compare the distortion when estimating $\mathbf{s}$ using the C-E and E-C schemes. With $\mathbf{\Sigma}_{ss} = \sigma_s^2 \mathbf{I}_p$, $p = 4$ and $N = 40$, we define the signal-to-noise ratio (SNR) as SNR $= \text{trace}(\mathbf{H}\mathbf{\Sigma}_{ss}\mathbf{H}^T)/(N\sigma^2)$. Observe that after a certain rate, the E-C scheme incurs a strictly smaller distortion than the C-E scheme. As $R \to \infty$, the distortion of E-C approaches D_{e0} with a rate faster than C-E and this corroborates Theorem 2. The rate after which the distortion of the E-C scheme falls down in a much faster rate than the one for C-E increases, as the SNR increases. This is intuitively justified since as the SNR increases, σ^2 becomes smaller and thus a larger rate is required for having $k_1 = N$. This can be verified also from (15), where the threshold increases as σ^2 decreases.

Our analysis so far raises the question whether the E-C scheme is optimal w.r.t. the distortion achieved for a given rate. We will show that this is the case for estimating $\mathbf{s}$ under a rate constraint R. A related claim has been reported in [6] and [7] for $N = p$, without restricting consideration to a Gaussian linear model. When $N > p$, extension of [6] and [7] is not obvious and we are going to examine this under the linearity and Gaussianity assumptions. Considering the model $\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}$ and assuming $N \geq p$, we have:

Theorem 3: *The D-R function for estimating $\mathbf{s}$ under the constraint $I(\mathbf{x}; \hat{\mathbf{s}}) \leq R$ can be expressed as*

$$D_e(R) = \min_{\substack{p(\hat{\mathbf{s}}|\mathbf{x}) \\ s.t.\ \ I(\mathbf{x};\hat{\mathbf{s}}) \leq R}} E\|\mathbf{s} - \hat{\mathbf{s}}\|^2 = \min_{\substack{p(\hat{\mathbf{s}}|\hat{\mathbf{s}}_\infty) \\ s.t.\ \ I(\hat{\mathbf{s}}_\infty;\hat{\mathbf{s}}) \leq R}} E\|\hat{\mathbf{s}}_\infty - \hat{\mathbf{s}}\|^2 + E\|\tilde{\mathbf{s}}_\infty\|^2, \tag{17}$$

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

Theorem 3 shows that the optimal setup for estimating the parameter vector $\mathbf{s}$ is first to form the optimal MMSE estimate $\hat{\mathbf{s}}_\infty$ and then apply optimal rate-distortion encoding to this estimate. The lower bound on this distortion when $R \to \infty$, is $D_{e0} = E\|\tilde{\mathbf{s}}_\infty\|^2$, which is intuitively appealing.

B. Distributed Case

In this subsection we will consider the problem of determining the D-R function when estimating $\mathbf{s}$ at the fusion center based on distributed observations from two sensors. Consider that the i-th sensor observes the $N_i \times 1$ vector $\mathbf{x}_i = \mathbf{H}_i\mathbf{s} + \mathbf{n}_i$, $i = 1, 2$, where $N_i \geq p$ and $N_1 + N_2 = N$. Furthermore, assume that $[\mathbf{n}_1^T, \mathbf{n}_2^T]^T \sim \mathcal{N}\ (\mathbf{0}, \sigma^2\mathbf{I}_N)$, and $\mathbf{s} \sim \mathcal{N}\ (\mathbf{0}, \mathbf{\Sigma}_{ss})$ with $\mathbf{s}$ independent of $[\mathbf{n_1}^T, \mathbf{n_2}^T]^T$. We wish to determine the D-R function

$$D(R) = \min_{\substack{p(\mathbf{u}_i|\mathbf{x}_i)\ \ for\ \ i=1,2\ \ and\ \ \hat{\mathbf{s}} \\ s.t.\ \ I(\mathbf{x};\mathbf{u}_1,\mathbf{u}_2) \leq R}} E_{p(\mathbf{s},\mathbf{u}_1,\mathbf{u}_2)}\|\mathbf{s} - \hat{\mathbf{s}}(\mathbf{u}_1, \mathbf{u}_2)\|^2. \tag{18}$$

Following the same procedure as in Section II-B, we will treat $\mathbf{u}_2$ as side information given at the decoder. Let us assume $\mathbf{u}_2 = \mathbf{u}_{2g} = \mathbf{C}_2\mathbf{x}_2 + \mathbf{z}_2$, as in Section II-B. Likewise, we find that $I(\mathbf{x}; \mathbf{u}_1, \mathbf{u}_{2g}) = R_0 + I(\mathbf{x}_1; \mathbf{u}_1|\mathbf{u}_{2g})$ and $I(\mathbf{x}_1; \mathbf{u}_1|\mathbf{u}_{2g}) = I(\mathbf{x}_1; \mathbf{u}_1) - I(\mathbf{u}_{2g}; \mathbf{u}_1)$, since $\mathbf{u}_1$ is a function of $\mathbf{x}_1$. Letting $\psi = [\mathbf{x}_1^T, \mathbf{u}_{2g}^T]^T$, we have $\hat{\mathbf{s}}_\infty = E[\mathbf{s}|\mathbf{x}_1, \mathbf{u}_{2g}] = \mathbf{\Sigma}_{s\psi}\mathbf{\Sigma}_{\psi\psi}^{-1}\psi = \mathbf{L}_1\mathbf{x}_1 + \mathbf{L}_2\mathbf{u}_{2g}$, where $\mathbf{\Sigma}_{s\psi}\mathbf{\Sigma}_{\psi\psi}^{-1} = [\mathbf{L}_1, \mathbf{L}_2]$ and $\mathbf{L}_1 : p \times N_1$ and $\mathbf{L}_2 : p \times k_2$. We can write $\mathbf{s} = \hat{\mathbf{s}}_\infty + \tilde{\mathbf{s}}_\infty$ and after that $E\|\mathbf{s} - \hat{\mathbf{s}}(\mathbf{u}_1, \mathbf{u}_{2g})\|^2 = E\|\hat{\mathbf{s}}_\infty - \hat{\mathbf{s}}(\mathbf{u}_1, \mathbf{u}_{2g})\|^2 + E\|\tilde{\mathbf{s}}_\infty\|^2$, since $\tilde{\mathbf{s}}_\infty$ is independent of $\hat{\mathbf{s}}_\infty$ and $\hat{\mathbf{s}}$ can be independent of $\tilde{\mathbf{s}}_\infty$ without affecting the distortion.

Lemma 1: It holds that $I(\mathbf{L}\mathbf{x}_1; \mathbf{u}_1) = I(\mathbf{x}_1; \mathbf{u}_1)$.

Using our results so far, we can perform the minimization required in

$$D(R_1) = \min_{\substack{p(\mathbf{u}_1|\mathbf{L}_1\mathbf{x}_1)\ \ and\ \ \hat{\mathbf{s}} \\ s.t.\ \ I(\mathbf{L}_1\mathbf{x}_1;\mathbf{u}_1) - I(\mathbf{u}_1;\mathbf{u}_{2g}) \leq R_1}} E_{p(\mathbf{L}_1\mathbf{x}_1,\mathbf{u}_1,\mathbf{u}_{2g})}\|\mathbf{L}_1\mathbf{x}_1 - (\hat{\mathbf{s}}(\mathbf{u}_1, \mathbf{u}_{2g}) - \mathbf{L}_2\mathbf{u}_{2g})\|^2 + E\|\tilde{\mathbf{s}}_\infty\|^2, \tag{19}$$

with $R_1 = R - R_0$. As in Section II-B, we can apply the Wyner-Ziv result to re-write (19) as

$$D(R_1) = \min_{\substack{p(\hat{\mathbf{s}}|\mathbf{L}_1\mathbf{x}_1,\mathbf{u}_{2g}) \\ s.t.\;\; I(\mathbf{L}_1\mathbf{x}_1;\hat{\mathbf{s}}|\mathbf{u}_{2g})\le R_1}} E_{p(\mathbf{L}_1\mathbf{x}_1,\hat{\mathbf{s}},\mathbf{u}_{2g})}\|\mathbf{L}_1\mathbf{x}_1 - (\hat{\mathbf{s}} - \mathbf{L}_2\mathbf{u}_{2g})\|^2 + E\|\tilde{\mathbf{s}}_\infty\|^2. \tag{20}$$

Again, we want to somehow remove the side information from the rate constraints in order to arrive at a minimization problem where the rwf solution is applicable. To this end, notice that $\mathbf{L}_1\mathbf{x}_1 = \hat{\mathbf{s}}_{\infty,1} + \tilde{\mathbf{s}}_{\infty,1}$ where $\hat{\mathbf{s}}_{\infty,1} = E[\mathbf{L}_1\mathbf{x}_1|\mathbf{u}_{2g}] = \mathbf{L}_1\mathbf{\Sigma}_{x_1u_{2g}}\mathbf{\Sigma}^{-1}_{u_{2g}u_{2g}}\mathbf{u}_{2g}$, and $\tilde{\mathbf{s}}_{\infty,1}$ is independent of $\mathbf{u}_{2g}$. Clearly, we then have that $E\|\mathbf{L}_1\mathbf{x}_1 - (\hat{\mathbf{s}} - \mathbf{L}_2\mathbf{u}_{2g})\|^2 = E\|\tilde{\mathbf{s}}_{\infty,1} - \hat{\mathbf{s}}'\|^2$ and $\hat{\mathbf{s}}' = \hat{\mathbf{s}} - \mathbf{L}_2\mathbf{u}_{2g} - \hat{\mathbf{s}}_{\infty,1}$. It also follows that $I(\mathbf{L}_1\mathbf{x}_1;\hat{\mathbf{s}}|\mathbf{u}_{2g}) = I(\tilde{\mathbf{s}}_{\infty,1};\hat{\mathbf{s}}'|\mathbf{u}_{2g}) = I(\tilde{\mathbf{s}}_{\infty,1};\hat{\mathbf{s}}')$, where the last equality holds because $\hat{\mathbf{s}}'$ can be independent from $\mathbf{u}_{2g}$ since we want to minimize $E\|\tilde{\mathbf{s}}_{\infty,1} - \hat{\mathbf{s}}'\|^2$. As a result, (20) can be written as

$$D(R_1) = \min_{\substack{p(\hat{\mathbf{s}}'|\tilde{\mathbf{s}}_{\infty,1}) \\ s.t.\;\; I(\tilde{\mathbf{s}}_{\infty,1};\hat{\mathbf{s}}')\le R_1}} E_{p(\tilde{\mathbf{s}}_{\infty,1},\hat{\mathbf{s}}')}\|\tilde{\mathbf{s}}_{\infty,1} - \hat{\mathbf{s}}'\|^2 + E\|\tilde{\mathbf{s}}_\infty\|^2, \tag{21}$$

and since $\tilde{\mathbf{s}}_{\infty,1}$ is Gaussian, we can apply rwf. Consider the covariance matrix $\mathbf{\Sigma}_{\tilde{s}_{\infty,1}\tilde{s}_{\infty,1}} = \mathbf{Q}_{\tilde{s}_{\infty,1}}\mathbf{\Lambda}_{\tilde{s}_{\infty,1}}\mathbf{Q}^T_{\tilde{s}_{\infty,1}}$ with $\mathbf{\Lambda}_{\tilde{s}_{\infty,1}} = \text{diag}\left(\tilde{\lambda}_{s,1},\cdots,\tilde{\lambda}_{s,p}\right)$. Using again the notion of a test channel, the optimal choice for $\mathbf{u}_1$ and $\hat{\mathbf{s}}$ in (19) is $\mathbf{u}_1 = \mathbf{Q}^T_{\tilde{s}_{\infty,1,k}}\mathbf{L}_1\mathbf{x}_1 + \mathbf{z}_1 = \mathbf{C}_1\mathbf{x}_1 + \mathbf{z}_1$, where $\mathbf{Q}_{\tilde{s}_{\infty,1,k}}$ is formed using the first k columns of $\mathbf{Q}_{\tilde{s}_{\infty,1}}$, and $\mathbf{C}_1 = \mathbf{Q}^T_{\tilde{s}_{\infty,1,k}}\mathbf{L}_1$. In this way, we are able to determine also $p(\mathbf{u}_1|\mathbf{x}_1)$. Additionally, we have that $\hat{\mathbf{s}} = \mathbf{Q}_{\tilde{s}_{\infty,1,k}}\mathbf{A}_{1,k}\mathbf{u}_1 - \mathbf{Q}_{\tilde{s}_{\infty,1,k}}\mathbf{A}_{1,k}\mathbf{Q}^T_{\tilde{s}_{\infty,1,k}}\mathbf{L}_1\mathbf{\Sigma}_{x_1u_{2g}}\mathbf{\Sigma}^{-1}_{u_{2g}u_{2g}}\mathbf{u}_{2g} + \mathbf{L}_1\mathbf{\Sigma}_{x_1u_{2g}}\mathbf{\Sigma}^{-1}_{u_{2g}u_{2g}}\mathbf{u}_{2g} + \mathbf{L}_2\mathbf{u}_{2g}$, where the matrices $\mathbf{A}_{1,k}$ and $\mathbf{\Sigma}_{z_1z_1}$ are specified in (5) using $\tilde{\lambda}_{s,i}$ for $i = 1,\cdots,k$. It follows from Section II-A that $D(R_1) = k\mu(k,R) + \sum_{l=k+1}^{p}\tilde{\lambda}_{s,l} + E\|\tilde{\mathbf{s}}_\infty\|^2$ and $\mu(k,R)$ is the threshold of the rwf solution. The form of $\mathbf{u}_1$ reveals that an optimal scheme for estimating $\mathbf{s}$ under a rate constraint R and available side information $\mathbf{u}_{2g}$ is to use the $\mathbf{L}_1\mathbf{x}_1$ part of $E[\mathbf{s}|\mathbf{x}_1,\mathbf{u}_2]$; subsequently, we can apply the prewhitening matrix $\mathbf{Q}^T_{\tilde{s}_{\infty,1,k}}$, and finally compress the resulting vector $\mathbf{Q}^T_{\tilde{s}_{\infty,1,k}}\mathbf{L}_1\mathbf{x}_1$.

The approach we outlined here can also be applied alternately in each sensor in order to determine appropriate $p(\mathbf{u}_i|\mathbf{x}_i)$ for $i = 1, 2$ and a function $\hat{\mathbf{s}}(\mathbf{u}_1,\mathbf{u}_2)$ which at best minimizes (18). The resultant iterative procedure is guaranteed again to converge only to a local minimum, thus determining an achievable D-R region and at best the D-R function.We can use for our purposes Algorithm 1 in order to obtain a tighter upper bound for the D-R region or at best determine the D-R function in the our estimation problem. Concerning rate allocation, we certainly follow an identical approach, but the matrices $\mathbf{C}_1^{(j)}$, $\mathbf{C}_2^{(j)}$, $\mathbf{\Sigma}^{(j)}_{z_1z_1}$, $\mathbf{\Sigma}^{(j)}_{z_2z_2}$ are now determined as we described in this section.

In Fig. 4, we plot the lower bound when assuming that one sensor has available the entire vector $\mathbf{x}$ and we apply the optimal E-C scheme. The upper bound is determined by letting the i-th sensor form its local estimate $\hat{\mathbf{s}}_{\infty,i} = E[\mathbf{s}|\mathbf{x}_i]$, and then apply rate-distortion encoding regardless of the other sensor in $\hat{\mathbf{s}}_{\infty,i}$. If $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$ are the compressed versions of $\hat{\mathbf{s}}_{\infty,1}$ and $\hat{\mathbf{s}}_{\infty,2}$, respectively, then the decoder forms the final estimate $\hat{\mathbf{s}} = E[\mathbf{s}|\hat{\mathbf{s}}_1,\hat{\mathbf{s}}_2]$. We also plot the achievable D-R region determined by the 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 in estimating $\mathbf{s}$.

IV. Conclusions

In this project, we first analyzed the D-R function for estimating a vector of parameters in a non-distributed setup. We also used an iterative scheme to determine an achievable D-R region or at best the D-R function of a similar problem in a distributed setup. For the single-sensor case, we carried out a comparison between two different schemes for estimating $\mathbf{s}$ under a rate constraint. In the first scheme, we encode-compress the observation vector and at the decoder we apply MMSE. In the second scheme, we form an MMSE estimate before the encoder and subsequently compress this estimate. The decoder output yields the final estimate of $\mathbf{s}$. The last scheme, referred to as estimate-compress, has been shown to be optimal in the sense that the D-R function is determined by a constant which is the MMSE for estimating $\mathbf{s}$ given $\mathbf{x}$ and also by the distortion incurred when encoding this MMSE estimate.

For the distributed scheme, we investigated a two sensor scenario. We considered that the output of the second encoder (and thus the encoder itself) is given and we treated this as side information in order to determine the

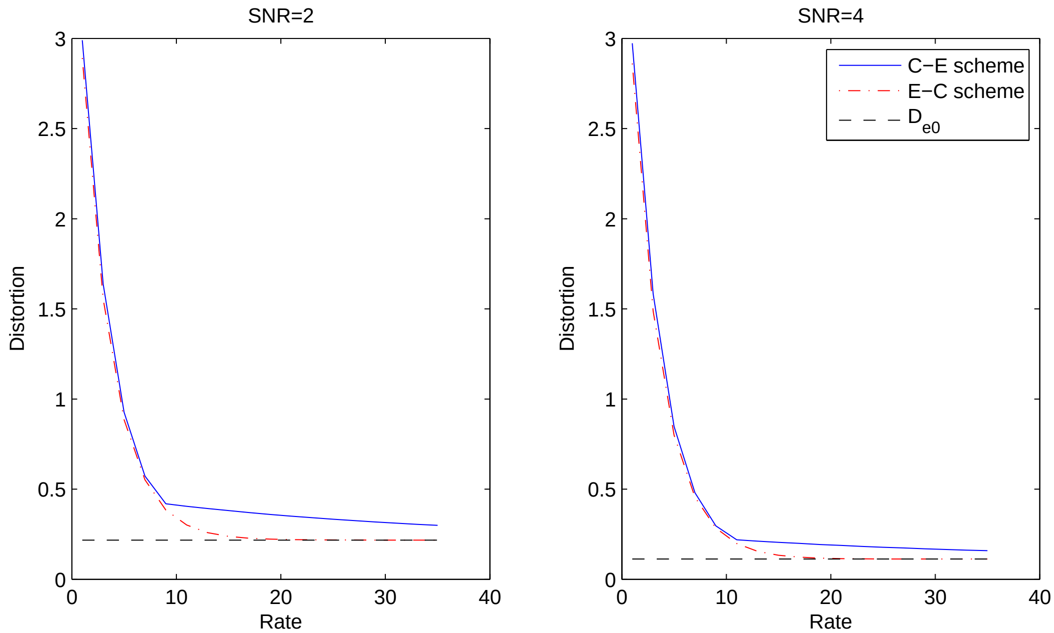

Fig. 3. Distortion-rate region for the C-E and E-C with $\mathbf{\Sigma}_{ss} = \mathbf{I}_4$, and $N_1 = N_2 = 20$.

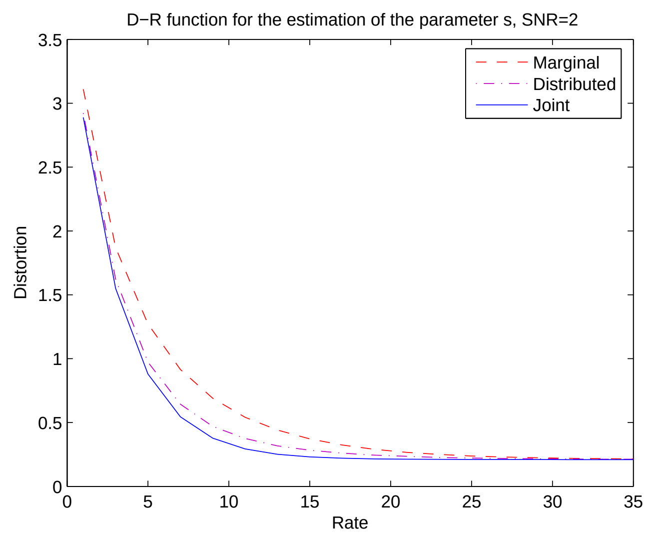

Fig. 4. Distortion-rate bounds for estimating $\mathbf{s}$ with $\mathbf{\Sigma}_{ss} = \mathbf{I}_4$, and $N_1 = N_2 = 20$.

optimal structure for the first encoder. A similar approach was utilized for the reconstruction problem in [1] which was also examined in this project. In this setup, we also proved that again it is better to first estimate and then compress. Finally, we applied an iterative scheme similar to [1] in order to derive upper bounds for the D-R region, or, at best determine the D-R function. The numerically determined upper bounds were compared with clairvoyant analytical upper bounds and found to be lower and closer to the lower bound of the single-sensor case where joint encoding is possible.

Acknowledgements: I would like to thank Prof. Georgios B. Giannakis and Prof. Nihar Jindal for the interesting discussions that we had on this project.

References

[1] M. Gastpar, P.L. Draggoti and M. Vetterli, "The Distributed Karhunen-Loeve Transform," submitted to *IEEE Transactions on Information Theory*, November 2004; downloadable from http://www.eecs.berkeley.edu/∼gastpar/

[2] A. Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder," *IEEE Transactions on Information Theory*, Vol. IT-22, No. 1, pp. 1–10, January 1976.

[3] T. Cover and J. Thomas, *Elements of Information Theory*, Second Edition, John Wiley and Sons, 1991.

[4] H. Viswanathan and T. Berger, "The Quadratic Gaussian CEO Problem," *IEEE Transactions on Information Theory*, Vol. 43, No. 5, pp. 1549–1559, September 1997.

[5] Y. Oohama, "The Rate-Distortion Function for the Quadratic Gaussian CEO Problem," *IEEE Transactions On Information Theory*, Vol. 44, No. 3, pp. 1057–1070, May, 1998.

[6] D. J. Sakrison, "Source encoding in the presence of random disturbance," *IEEE Transactions on Information Theory*, Vol. IT-14, No. 1, pp. 165–167, January 1968.

[7] J. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Transactions on Information Theory*, Vol. 16, No. 4, pp. 406–411, July 1970.

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

[9] J. Chen, X. Zhang, T. Berger, 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 Selected Areas in Communications*, Vol. 22, No. 6, pp. 406–411, August 2004.

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

[11] P. Ishwar, R. Puri, K. Ramchadran, S.S. Pradhan, "On Rate-Constrained Distributed Estimation in Unreliable Sensor Networks," *IEEE Journal on Selected Areas in Communications*, Vol. 23, No. 4, pp. 765–775, April 2005.

V. Appendix

Proof of Theorem 2: For the C-E scheme with $k_1 = N$, the rwf threshold is given by $\mu_1(N,R) = \left(\prod_{i=1}^{p} \lambda_{x,i}\right)^{1/N} (\sigma^2)^{(N-p)/N} 2^{-2R/N}$. Since $k_1 = N$, it must hold that $\mu_1(N,R) < \sigma^2 \Leftrightarrow R > (1/2)\log_2\left(\left(\prod_{i=1}^{p} \lambda_{x,i}\right)/(\sigma^2)^p\right)$. For the E-C scheme with $k_2 = p$ the rwf threshold is $\mu_2(p,R) = \left(\prod_{i=1}^{p} \lambda_{\hat{s}_\infty,i}\right)^{1/p} 2^{-2R/p}$ and it must hold that

$$\mu_2(p,R) < \lambda_{\hat{s}_\infty,p} \Leftrightarrow R > (1/2)\log_2\left(\left(\prod_{i=1}^{p} \lambda_{\hat{s}_\infty,i}\right)\right)/(\lambda_{\hat{s}_\infty,p})^p.$$

Thus, when $R > \max\left\{(1/2)\log_2\left(\left(\prod_{i=1}^{p} \lambda_{x,i}\right)/(\sigma^2)^p\right), (1/2)\log_2\left(\left(\prod_{i=1}^{p} \lambda_{\hat{s}_\infty,i}\right)\right)/(\lambda_{\hat{s}_\infty,p})^p\right\}$, we have that $k_1 = N$ and $k_2 = p$. We have $\mathbf{\Delta}_1 = \mu_1(N,R)\text{diag}\left(\lambda_{x,1}^{-2}, \cdots, \lambda_{x,N}^{-2}\right) = 2^{-2R/N}\alpha_1\text{diag}(\lambda_{x,1}^{-2}, \cdots, \lambda_{x,N}^{-2})$ and $\mathbf{\Delta}_2 = \mu_2(p,R)\mathbf{I}_p = \left(\prod_{i=1}^{p} \lambda_{\hat{s}_\infty,i}\right)^{1/p} 2^{-2R/p}\mathbf{I}_p$, and $\alpha_1 = \left(\prod_{i=1}^{p} \lambda_{x,i}\right)^{1/N} (\sigma^2)^{(N-p)/N}$.

Because $\mathbf{Q}_x, \mathbf{\Lambda}_x, \mathbf{Q}_{\hat{s}_\infty}$ do not depend on the rate R it follows that

$$\epsilon_1(R) = 2^{-2R/N}\text{trace}(\mathbf{\Sigma}_{sx}\mathbf{Q}_x\alpha_1\text{diag}(\lambda_{x,1}^{-2}, \cdots, \lambda_{x,N}^{-2})\mathbf{Q}_x^T\mathbf{\Sigma}_{xs}) = 2^{-2R/N}\gamma_1$$

and $\epsilon_2(R) = 2^{-2R/p}$ trace $\left(\left(\prod_{i=1}^{p} \lambda_{\hat{s}_\infty,i}\right)^{1/p}\mathbf{I}_p\right) = \gamma_2 2^{-2R/p}$ where γ_1, γ_2 does not depend on the rate R. Q.E.D.

Proof of Proposition 1: Using (13), we have for the C-E scheme that $\Delta_1 = \sigma_x^2 2^{-2R} = (h^2\sigma_s^2 + \sigma_n^2)2^{-2R}$ and the variance of the error $\tilde{s}_1 = s - \hat{s}_1$ is $\sigma_{\tilde{s}_1}^2 = \sigma_s^2 - h^2\sigma_s^4(\sigma_x^2)^{-1} + h^2\sigma_s^4(\Delta_1/(\sigma_x^2)^2) = \sigma_s^2 - h^2\sigma_s^4(\sigma_x^2)^{-1} + h^2\sigma_s^4(2^{-2R}/(h^2\sigma_s^2 + \sigma_n^2))$.
Likewise, using (14) we have for the E-C scheme that $\Delta_2 = \sigma_{\hat{s}_\infty}^2 2^{-2R} = h^2\sigma_s^4(h^2\sigma_s^2+\sigma_n^2)^{-1}\ 2^{-2R}$ and the variance of the error $\tilde{s}_2 = s - \hat{s}_2$ is $\sigma_{\tilde{s}_2}^2 = \sigma_s^2 - h^2\sigma_s^4(\sigma_x^2)^{-1} + \Delta_2 = \sigma_s^2 - h^2\sigma_s^4(\sigma_x^2)^{-1} + h^2\sigma_s^4(2^{-2R}/(h^2\sigma_s^2 + \sigma_n^2))$. It then follows immediately that $\sigma_{\tilde{s}_1}^2 = \sigma_{\tilde{s}_2}^2$. Q.E.D.

Proof of Proposition 2: We can readily verify that $\mathbf{\Lambda}_x = \text{diag}\left(\sigma^2 + \sigma_s^2\|\mathbf{h}\|^2, \sigma^2, \cdots, \sigma^2\right)$ and $\mathbf{Q}_x = [\mathbf{q}_{x,1}, \cdots, \mathbf{q}_{x,N}]$, where $\mathbf{q}_{x,1} = (\mathbf{h}/\|\mathbf{h}\|) \perp \mathbf{q}_{x,j}$, for $j \neq 1$. For the C-E scheme, k_1 can be either 1 or N and this happens because $\lambda_{x,i} = \sigma^2$ for $i = 2, \ldots N$; thus, when $\mu_1(k_1, R) < \sigma^2$ all the elements of $\mathbf{x}_{w,N}$ have to be assigned nonzero rate. For $k_1 = 1$, the threshold is $\mu_1(1,R) = (\sigma_s^2\|\mathbf{h}\|^2 + \sigma^2)2^{-2R}$. Continuing, it must hold that $\mu_1(1,R) \geq \sigma^2 \Leftrightarrow R \leq (1/2)\log_2\left(1 + (\sigma_s^2\|\mathbf{h}\|^2)/\sigma^2\right) = R_t$.
If $R > R_t$, we have that $k_1 = N$ and the threshold is $\mu_1(N,R) = (\sigma_s^2\|\mathbf{h}\|^2 + \sigma^2)^{(1/N)}\left(\sigma^2\right)^{(1-1/N)} 2^{-2R/N}$, while the distortion term $\epsilon_1(R) = \text{trace}\left(\mathbf{\Sigma}_{sx}\mathbf{Q}_x\mathbf{\Delta}_1\mathbf{Q}_x^T\mathbf{\Sigma}_{xs}\right)$ is given by

$$\epsilon_1(R) = \text{trace}(\sigma_s^4\mathbf{h}^T\mathbf{Q}_x\mathbf{\Delta}_1\mathbf{Q}_x^T\mathbf{h}) = \begin{cases} \beta 2^{-2R} & ,R \leq R_t \\ \beta\left(\sigma^2/(\sigma_s^2\|\mathbf{h}\|^2 + \sigma^2)\right)^{(1-1/N)} 2^{-2R/N} & ,R > R_t \end{cases} \tag{22}$$

where $\beta = \sigma_s^4\|\mathbf{h}\|^2/(\sigma_s^2\|\mathbf{h}\|^2 + \sigma^2)$. We have for the E-C scheme that $\sigma_{\hat{s}_\infty}^2 = \sigma_s^2\mathbf{h}^T\mathbf{Q}_x\mathbf{\Lambda}_x^{-1}\mathbf{Q}_x^T\mathbf{h} = \beta$. Since we compress the Gaussian scalar random variable $\hat{\mathbf{s}}_\infty$, we have that $\epsilon_2(R) = \beta 2^{-2R}$, $\forall\ R$. The result follows immediately with a direct comparison between $\epsilon_1(R)$ and $\epsilon_2(R)$ when $R \leq R_t$, and when $R < R_t$, respectively. Q.E.D.

Proof of Proposition 3: It follows immediately that $\mathbf{Q}_x = \mathbf{U}_h$ and $\mathbf{\Lambda}_x = \text{diag}(\sigma_s^2\sigma_{h,1}^2 + \sigma^2, \cdots, \sigma_s^2\sigma_{h,p}^2 + \sigma^2, \sigma^2, \cdots, \sigma^2)$. With $\mathbf{\Sigma}_{\hat{s}_\infty\hat{s}_\infty} = \sigma_s^2\mathbf{V}_h\mathbf{\Sigma}_h^T(\sigma_s^2\mathbf{\Sigma}_h\mathbf{\Sigma}_h^T + \sigma^2\mathbf{I})^{-1}\mathbf{\Sigma}_h\mathbf{V}_h^T$, it can be easily verified that $\mathbf{Q}_{\hat{s}_\infty} = \mathbf{V}_h$ and $\mathbf{\Lambda}_{\hat{s}_\infty} = \text{diag}\left((\sigma_s^4\sigma_{h,1}^2)/(\sigma_s^2\sigma_{h,1}^2 + \sigma^2), \cdots, (\sigma_s^4\sigma_{h,p}^2)/(\sigma_s^2\sigma_{h,p}^2 + \sigma^2)\right)$.

For the C-E scheme, we have $\mathbf{E}_1 = \sigma_s^4\mathbf{V}_h\mathbf{\Sigma}_h^T\mathbf{\Delta}_1\mathbf{\Sigma}_h\mathbf{V}_h^T$ and $\epsilon_1(R) = \text{trace}(\mathbf{E}_1) = \sigma_s^4\sum_{i=1}^{p}(\sigma_{h,i}^2 D_i^1)/(\sigma_s^2\sigma_{h,i}^2+\sigma^2)^2$. If $k_1 = N$, we have $D_i^1 = \mu_1(N,R) = \left(\prod_{i=1}^{p}(\sigma_s^2\sigma_{h,i}^2 + \sigma^2)\right)^{1/N} (\sigma^2)^{(N-p)/N} 2^{-2R/N}$ for $i = 1, \ldots, N$, and $\mu_1(N,R) < \sigma^2 \Leftrightarrow R > (1/2)\log_2\left(\prod_{i=1}^{p}(1 + (\sigma_s^2\sigma_{h,i}^2)/\sigma^2)\right)$. While for the E-C scheme, $\mathbf{E}_2 = \mathbf{V}_h\mathbf{\Delta}_2\mathbf{V}_h^T$ and $\epsilon_2(R) = \text{trace}(\mathbf{E}_2) = \text{trace}(\mathbf{\Delta}_2) = \sum_{i=1}^{p} D_i^2$. For $k_2 = p$, it follows for $j = 1, \cdots, p$ that

$$D_j^2 = \mu_2(p,R) = \left(\prod_{i=1}^{p}(\sigma_s^4\sigma_{h,i}^2)/(\sigma_s^2\sigma_{h,i}^2 + \sigma^2)\right)^{1/p} 2^{-2R/p},$$

and for this to hold, we must have

$$\mu_2(p,R) < (\sigma_s^4\sigma_{h,p}^2)/(\sigma_s^2\sigma_{h,p}^2+\sigma^2) \Leftrightarrow R > (1/2)\log_2\left(\prod_{i=1}^{p}(\sigma_{h,i}^2/(\sigma_{h,i}^2\sigma_s^2+\sigma^2))/((\sigma_{h,p}^2)^p/(\sigma_{h,p}^2\sigma_s^2+\sigma^2)^p)\right).$$

Accordingly when (15) is satisfied we have $k_1 = N, k_2 = p$,

$$\epsilon_1(R) = \sigma_s^4\sigma^2 2^{-2R/N}\left(\prod_{j=1}^{p}((\sigma_s^2\sigma_{h,j}^2)/\sigma^2+1)\right)^{1/N}\sum_{i=1}^{p}\sigma_{h,i}^2/(\sigma_s^2\sigma_{h,i}^2+\sigma^2)^2,$$

and $\epsilon_2(R) = \sigma_s^4 2^{-2R/p}\left(\prod_{j=1}^{p}\sigma_{h,j}^2/(\sigma_s^2\sigma_{h,j}^2+\sigma^2)\right)^{1/p} p$. When $N > p$, in order for $\epsilon_1(R) > \epsilon_2(R)$ to hold we must have

$$R > (1/2)\log_2\left(\prod_{i=1}^{p}(1+(\sigma_s^2\sigma_{h,i}^2)/\sigma^2)\right) + (Np/2(N-p))\log_2\gamma, \tag{23}$$

where

$$\gamma = \left(\prod_{i=1}^{p}\sigma_{h,i}^2/(\sigma_{h,i}^2\sigma_s^2+\sigma^2)^2\right)^{1/p} \Big/ \left((1/p)\sum_{j=1}^{p}(\sigma_{h,j}^2/(\sigma_{h,j}^2\sigma_s^2+\sigma^2)^2)\right) \leq 1,$$

from the AM-GM[4] inequality. But (23) is satisfied, since (15) holds. For the case where $N = p$ it follows that

$$\epsilon_1(R) \geq \epsilon_2(R) \Leftrightarrow \gamma \leq 1.$$

We have already mentioned that $\gamma \leq 1$ and strict inequality holds when the singular values of $\mathbf{H}$ denoted as $\sigma_{h,i}$ are not equal. If the opposite holds then $\epsilon_1(R) = \epsilon_2(R)$. Q.E.D.

Proof of Proposition 4: It follows that $\mathbf{Q}_x = [\bar{\mathbf{H}}, \mathbf{q}_{x,p+1}, \cdots, \mathbf{q}_{x,N}]$, where $\mathbf{q}_{x,p+1}, \cdots, \mathbf{q}_{x,N} \perp$ range($\bar{\mathbf{H}}$), and $\mathbf{\Lambda}_x = \text{diag}\left(\sigma_{s,1}^2\|\mathbf{h}_1\|^2+\sigma^2, \cdots, \sigma_{s,p}^2\|\mathbf{h}_p\|^2+\sigma^2, \sigma^2, \cdots, \sigma^2, \cdots, \sigma^2\right)$. Continuing we get

$$\mathbf{\Sigma}_{\hat{s}_\infty\hat{s}_\infty} = \mathbf{\Sigma}_{ss}\mathbf{H}^T\mathbf{Q}_x\mathbf{\Lambda}_x^{-1}\mathbf{Q}_x^T\mathbf{H}\mathbf{\Sigma}_{ss} = \mathbf{\Sigma}_{ss}\mathbf{D}_h\bar{\mathbf{H}}^T\mathbf{Q}_x(\mathbf{D}_h^2\mathbf{\Sigma}_{\mathbf{ss}}+\sigma^2\mathbf{I}_p)^{-1}\mathbf{Q}_x^T\bar{\mathbf{H}}\mathbf{D}_h\mathbf{\Sigma}_{ss} =$$

$\text{diag}\left((\sigma_{s,1}^4\|\mathbf{h}_1\|^2)/(\sigma_{s,1}^2\|\mathbf{h}_1\|^2+\sigma^2), \cdots, (\sigma_{s,p}^4\|\mathbf{h}_p\|^2)/(\sigma_{s,p}^2\|\mathbf{h}_p\|^2+\sigma^2)\right)$. As with the proof of Prop. 3, we have $\epsilon_1(R) = \sum_{i=1}^{p}(\sigma_{s,i}^4\|\mathbf{h}_i\|^2 D_i^1)/(\sigma_{s,i}^2\|\mathbf{h}_i\|^2+\sigma^2)^2$.

When $k_1 = N$, we have $D_i^1 = \mu_1(N,R) = \left(\prod_{i=1}^{p}(\sigma_{s,i}^2\|\mathbf{h}_i\|^2+\sigma^2)\right)^{1/N} (\sigma^2)^{(N-p)/N}2^{-2R/N}$ for $i = 1, \ldots, N$, and it must hold that

$$\mu_1(N,R) < \sigma^2 \Leftrightarrow R > (1/2)\log_2\left(\prod_{i=1}^{p}(1+(\sigma_{s,i}^2\|\mathbf{h}_i\|^2)/\sigma^2)\right).$$

For the E-C scheme, we have $\epsilon_2(R) = \sum_{i=1}^{p} D_i^2$. If $k_2 = p$, it follows for $j = 1, \cdots, p$ that, $D_j^2 = \mu_2(p,R) = \left(\prod_{i=1}^{p}(\sigma_{s,i}^4\|\mathbf{h}_i\|^2)/(\sigma_{s,i}^2\|\mathbf{h}_i\|^2+\sigma^2)\right)^{1/p} 2^{-2R/p}$, and in order this to hold we must have

$$\mu_2(p,R) < (\sigma_{s,p}^4\|\mathbf{h}_p\|^2)/(\sigma_{s,p}^2\|\mathbf{h}_p\|^2+\sigma^2)$$

or equivalently,

$$R > (1/2)\log_2\left(\prod_{i=1}^{p}(\sigma_{s,i}^4\|\mathbf{h}_i\|^2/(\|\mathbf{h}_i\|^2\sigma_{s,i}^2+\sigma^2))/((\sigma_{s,p}^4\|\mathbf{h}_p\|^2)^p/(\|\mathbf{h}_p\|^2\sigma_{s,p}^2+\sigma^2)^p)\right). \tag{24}$$

When (16) is satisfied, we have $k_1 = N, k_2 = p$,

$$\epsilon_1(R) = \sigma^2 2^{-2R/N}\left(\prod_{j=1}^{p}((\sigma_{s,j}^2\|\mathbf{h}_j\|^2)/\sigma^2+1)\right)^{1/N}\sum_{i=1}^{p}\sigma_{s,i}^4\|\mathbf{h}_i\|^2/(\sigma_{s,i}^2\|\mathbf{h}_i\|^2+\sigma^2)^2$$

[4] Arithmetic Mean-Geometric Mean inequality.

and $\epsilon_2(R) = 2^{-2R/p}\left(\prod_{j=1}^{p}(\sigma_{s,j}^4\|\mathbf{h}_j\|^2)/(\sigma_{s,j}^2\|\mathbf{h}_j\|^2+\sigma^2)\right)^{1/p} p$.

In order for $\epsilon_1(R) > \epsilon_2(R)$ to hold, when $N > p$, we must have

$$R > (1/2)\log_2\left(\prod_{i=1}^{p}(1+(\sigma_{s,i}^2\|\mathbf{h}_i\|^2)/\sigma^2)\right) + (Np/2(N-p))\log_2\gamma', \tag{25}$$

where

$$\gamma' = \left((\prod_{i=1}^{p}\sigma_{s,i}^4\|\mathbf{h}_i\|^2/(\sigma_{s,i}^2\|\mathbf{h}_i\|^2+\sigma^2)^2)^{1/p}\right) / \left((1/p)\sum_{j=1}^{p}(\sigma_{s,j}^2\|\mathbf{h}_j\|^2/(\|\mathbf{h}_j\|^2\sigma_{s,j}^2+\sigma^2)^2)\right) \leq 1,$$

from the AM-GM inequality. But (25) is satisfied, since (16) holds. For the case where $N = p$ it follows that

$$\epsilon_1(R) \geq \epsilon_2(R) \Leftrightarrow \gamma' \leq 1.$$

We have already mentioned that $\gamma' \leq 1$ and strict inequality holds when the euclidean norms $\|\mathbf{h}_i\|$ are not equal. If the opposite holds then $\epsilon_1(R) = \epsilon_2(R)$. Q.E.D.

Proof of Theorem 3: Let $\mathbf{s} = \hat{\mathbf{s}}_\infty + \tilde{\mathbf{s}}_\infty$, where $\tilde{\mathbf{s}}_\infty$ is independent of $\mathbf{x}$. We have then, $E\|\mathbf{s}-\hat{\mathbf{s}}\|^2 = E\|\hat{\mathbf{s}}_\infty - \hat{\mathbf{s}}\|^2 + E\|\tilde{\mathbf{s}}_\infty\|^2$, because $\hat{\mathbf{s}}_\infty$ and $\hat{\mathbf{s}}$ are independent of $\tilde{\mathbf{s}}_\infty$ since they are functions of $\mathbf{x}$. Consider now an invertible matrix $\mathbf{T}$, and recall that $I(\mathbf{x};\hat{\mathbf{s}}) = I(\mathbf{Tx};\hat{\mathbf{s}})$. The distortion in estimating $\mathbf{s}$ using either the data $\mathbf{Tx}$ or $\mathbf{x}$ is the same, since the matrix $\mathbf{T}$ is invertible. Consider the SVD decomposition of $\mathbf{H} = \mathbf{U}_h\mathbf{\Sigma}_h\mathbf{V}_h^T$, where $\mathbf{\Sigma}_h = [\mathbf{\Sigma}_{h,p}^T, \mathbf{0}_{p\times(N-p)}]^T$ and $\mathbf{\Sigma}_{h,p} = \text{diag}(\sigma_{h,1},\cdots,\sigma_{h,p})$ is a $p\times p$ diagonal matrix with non-zero entries, since $\mathbf{H}$ is full column rank. We clearly have $\mathbf{\Sigma}_{xx} = \mathbf{H}\mathbf{\Sigma}_{ss}\mathbf{H}^T + \sigma^2\mathbf{I}_p = \mathbf{U}_h\left(\mathbf{\Sigma}_h\mathbf{V}_h^T\mathbf{\Sigma}_{ss}\mathbf{V}_h\mathbf{\Sigma}_h^T + \sigma^2\mathbf{I}_N\right)\mathbf{U}_h^T$, and $\mathbf{\Sigma}_h\mathbf{V}_h^T\mathbf{\Sigma}_{ss}\mathbf{V}_h\mathbf{\Sigma}_h^T + \sigma^2\mathbf{I}_N = \text{diag}\left(\mathbf{\Sigma}_{h,p}\mathbf{V}_h^T\mathbf{\Sigma}_{ss}\mathbf{V}_h\mathbf{\Sigma}_{h,p}^T + \sigma^2\mathbf{I}_p, \sigma^2\mathbf{I}_{(N-p)}\right) = \text{diag}\left(\mathbf{B}_p, \sigma^2\mathbf{I}_{(N-p)}\right)$, where $\mathbf{B}_p = \mathbf{\Sigma}_{h,p}\mathbf{V}_h^T\mathbf{\Sigma}_{ss}\mathbf{V}_h\mathbf{\Sigma}_{h,p}^T + \sigma^2\mathbf{I}_p$. Because $\mathbf{H}$ is a full column rank matrix, $\mathbf{B}_p$ is invertible. Let us now consider the matrix $\mathbf{\Sigma}_{sx}\mathbf{\Sigma}_{xx}^{-1} = \mathbf{\Sigma}_{ss}\mathbf{H}^T\mathbf{U}_h\text{diag}\left(\mathbf{B}_p^{-1}, \sigma^{-2}\mathbf{I}_{(N-p)}\right)\mathbf{U}_h^T = \mathbf{\Sigma}_{ss}\mathbf{V}_h\mathbf{\Sigma}_{h,p}^T\mathbf{B}_p^{-1}\mathbf{U}_{h,p}^T$, where $\mathbf{U}_{h,p}$ denotes the first p columns of $\mathbf{U}_h$. Since rank$(\mathbf{\Sigma}_{ss}\mathbf{V}_h\mathbf{\Sigma}_{h,p}^T\mathbf{B}_p^{-1}) = p$, we have that[5] range$(\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{xs})$ = range$(\mathbf{U}_{h,p})$ = span$(\mathbf{U}_{h,p})$.

Now let the matrix $\mathbf{T} = [(\mathbf{\Sigma}_{sx}\mathbf{\Sigma}_{xx}^{-1})^T, \mathbf{R}^T]^T$. If we select $\mathbf{R} = \mathbf{U}_{h,(N-p)}$, which are the last $N-p$ columns of $\mathbf{U}_h$, then matrix $\mathbf{T}$ is guaranteed to be invertible. It follows that $\mathbf{Rx} = \mathbf{RHs} + \mathbf{Rn} = \mathbf{Rn}$, because range$(\mathbf{R}^T) \perp$ range$(\mathbf{H})$ = range$(\mathbf{U}_{h,p})$ = span$(\mathbf{U}_{h,p})$. Then $\mathbf{Tx} = [(\mathbf{\Sigma}_{sx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{x})^T, (\mathbf{Rn})^T]^T = [\hat{\mathbf{s}}_\infty^T, \mathbf{Rn}^T]^T$. Another important fact is that $\hat{\mathbf{s}}_\infty$ and $\mathbf{Rn}$ are independent. This happens because both $\hat{\mathbf{s}}_\infty$ and $\mathbf{Rn}$ are Gaussian and $E[\hat{\mathbf{s}}_\infty(\mathbf{Rn})^T] = E[\hat{\mathbf{s}}_\infty\mathbf{n}^T\mathbf{R}^T] = E[\mathbf{\Sigma}_{sx}\mathbf{\Sigma}_{xx}^{-1}(\mathbf{Hs}+\mathbf{n})\mathbf{n}^T\mathbf{R}^T] = \mathbf{\Sigma}_{sx}\mathbf{\Sigma}_{xx}^{-1}E[\mathbf{nn}^T]\mathbf{R}^T = \sigma^2\mathbf{\Sigma}_{ss}\mathbf{V}_h\mathbf{\Sigma}_{h,p}^T\mathbf{B}_p^{-1}\mathbf{U}_{h,p}^T\mathbf{U}_{h,(N-p)} = \mathbf{0}_{p\times(N-p)}$. It then follows $I(\mathbf{x};\hat{\mathbf{s}}) = I(\mathbf{Tx};\hat{\mathbf{s}}) = I(\hat{\mathbf{s}}_\infty, \mathbf{Rn};\hat{\mathbf{s}}) = I(\mathbf{Rn};\hat{\mathbf{s}}) + I(\hat{\mathbf{s}}_\infty;\hat{\mathbf{s}}|\mathbf{Rn})$. Now the optimal $\hat{\mathbf{s}}$ is independent of $\mathbf{Rn}$ without affecting the distortion, since the noise does not contain any information for the estimation of $\mathbf{s}$. Using the previous argument and the independence of $\hat{\mathbf{s}}_\infty$ and $\mathbf{Rn}$ we obtain $I(\mathbf{Rn};\hat{\mathbf{s}}) = 0$ and $I(\hat{\mathbf{s}}_\infty;\hat{\mathbf{s}}|\mathbf{Rn}) = I(\hat{\mathbf{s}}_\infty;\hat{\mathbf{s}})$. Using the previous arguments arrive at the desired result

$$D_e(R) = \min_{\substack{p(\hat{\mathbf{s}}|\mathbf{x}) \\ s.t.\ I(\mathbf{x};\hat{\mathbf{s}})\leq R}} E\|\mathbf{s}-\hat{\mathbf{s}}\|^2 = \min_{\substack{p(\hat{\mathbf{s}}|\hat{\mathbf{s}}_\infty) \\ s.t.\ I(\hat{\mathbf{s}}_\infty;\hat{\mathbf{s}})\leq R}} E\|\hat{\mathbf{s}}_\infty-\hat{\mathbf{s}}\|^2 + E\|\tilde{\mathbf{s}}_\infty\|^2. \text{ Q.E.D.}$$

Proof of Lemma 1: With $\mathbf{L}_1 = \left(\mathbf{\Sigma}_{sx_1} - \mathbf{\Sigma}_{su_{2g}}\mathbf{\Sigma}_{u_{2g}u_{2g}}^{-1}\mathbf{\Sigma}_{u_{2g}x_1}\right)\left(\mathbf{\Sigma}_{x_1x_1} - \mathbf{\Sigma}_{x_1u_{2g}}\mathbf{\Sigma}_{u_{2g}u_{2g}}^{-1}\mathbf{\Sigma}_{u_{2g}x_1}\right)^{-1}$, consider the SVD of $\mathbf{H}_1 = \mathbf{U}_{h_1}\mathbf{\Sigma}_{h_1}\mathbf{V}_{h_1}^T$, where $\mathbf{\Sigma}_{h_1} = [\mathbf{\Sigma}_{h_{1,p}}, \mathbf{0}_{p\times(N_1-p)}]$, $\mathbf{U}_{h_1} = [\mathbf{U}_{h_{1,p}}, \mathbf{U}_{h_{1,N_1-p}}]$ and $\mathbf{\Sigma}_{h_{1,p}} = \text{diag}(\sigma_{h,1},\cdots,\sigma_{h,p})$. Moreover, we have that

$$\mathbf{\Sigma}_{sx_1} - \mathbf{\Sigma}_{su_{2g}}\mathbf{\Sigma}_{u_{2g}u_{2g}}^{-1}\mathbf{\Sigma}_{u_{2g}x_1} = \left(\mathbf{\Sigma}_{ss}\mathbf{V}_{h_1} - \mathbf{\Sigma}_{su_{2g}}\mathbf{\Sigma}_{u_{2g}u_{2g}}^{-1}\mathbf{\Sigma}_{u_{2g}s}\right)\mathbf{\Sigma}_{h_1}^T\mathbf{U}_{h_1}^T, \tag{26}$$

[5]The range$(\mathbf{M})$ denotes the vector subspace, where the columns of matrix $\mathbf{M}$ lie. The span$(\mathbf{M})$ denotes the vector subspace which is created by any linear combination of the columns of $\mathbf{M}$.

where $\mathbf{\Omega}_1 = \mathbf{V}_{h_1}^T \mathbf{\Sigma}_{ss} \mathbf{V}_{h_1}$, and $\mathbf{\Omega}_1 = \mathbf{V}_{h_1}^T \mathbf{\Sigma}_{su_{2g}} \mathbf{\Sigma}_{u_{2g}u_{2g}}^{-1} \mathbf{\Sigma}_{u_{2g}s} \mathbf{V}_{h_1}$. Eq. (26) can be re-written as

$$\mathbf{\Sigma}_{sx_1} - \mathbf{\Sigma}_{su_{2g}} \mathbf{\Sigma}_{u_{2g}u_{2g}}^{-1} \mathbf{\Sigma}_{u_{2g}x_1} = \left(\mathbf{\Sigma}_{ss} \mathbf{V}_{h_1} - \mathbf{\Sigma}_{su_{2g}} \mathbf{\Sigma}_{u_{2g}u_{2g}}^{-1} \mathbf{\Sigma}_{u_{2g}s} \right) \mathbf{\Sigma}_{h_1}^T \mathbf{U}_{h_1}^T$$

and since $\mathbf{L}_1 = \left(\mathbf{\Sigma}_{ss} \mathbf{V}_{h_1} - \mathbf{\Sigma}_{su_{2g}} \mathbf{\Sigma}_{u_{2g}u_{2g}}^{-1} \mathbf{\Sigma}_{u_{2g}s} \right) [\mathbf{\Sigma}_{h_{1,p}}, \mathbf{0}_{p \times (N_1 - p)}] \mathbf{\Omega}^{-1} \mathbf{U}_{h_1}^T$ with

$$\mathbf{\Omega} = \text{diag} \left(\mathbf{\Sigma}_{h_{1,p}} (\mathbf{\Omega_1} + \mathbf{\Omega_2}) \mathbf{\Sigma}_{h_{1,p}} + \sigma^2 \mathbf{I}_p, \;\; \sigma^2 \mathbf{I}_{(N_1 - p)} \right),$$

it follows that range$(\mathbf{L}_1^T)$ = span$(\mathbf{U}_{h_{1,p}})$. With $\mathbf{G} = \mathbf{U}_{h_{1,(N_1-p)}}^T$, let us construct the invertible matrix $\mathcal{T} = [\mathbf{L}_1^T, \mathbf{G}^T]^T$. Note also that $E[\mathbf{L}_1 \mathbf{x}_1 (\mathbf{G}\mathbf{n}_1)^T] = \mathbf{0}_{p \times (N_1 - p)}$. Using the data $\mathcal{T}\mathbf{x}_1$, and the given side information $\mathbf{u}_{2g}$ will not affect the distortion, since $\mathcal{T}$ is invertible. As a result, we find that $I(\mathbf{x}_1; \mathbf{u}_1) = I(\mathcal{T}\mathbf{x}_1; \mathbf{u}_1) = I(\mathbf{L}_1 \mathbf{x}_1, \mathbf{G}\mathbf{n}_1; \mathbf{u}_1) = I(\mathbf{G}\mathbf{n}_1; \mathbf{u}_1) + I(\mathbf{L}_1 \mathbf{x}_1; \mathbf{u}_1 | \mathbf{G}\mathbf{n}_1) = I(\mathbf{L}_1 \mathbf{x}_1; \mathbf{u}_1)$ since $\mathbf{u}_1$ is independent of $\mathbf{G}\mathbf{n}_1$, and $\mathbf{L}_1 \mathbf{x}_1$ is independent of $\mathbf{G}\mathbf{n}_1$. Q.E.D.