

On Side-Informed Coding of Noisy Sensor Observations

Chao Yu and Gaurav Sharma

ECE Dept, University of Rochester,
Rochester NY 14627

ABSTRACT

In this paper, we consider the problem of side-informed coding in applications where the data available at the encoder consists of indirect noisy observations of the signal desired at the decoder. In these scenarios, under Gaussian statistics we show that for a mean-squared distortion metric, the side-informed encoding problem can be decomposed into a “side-informed” minimum mean-squared error (MMSE) estimation followed by side-informed coding of the MMSE estimate, without incurring any rate-distortion penalty. By recursively exploiting this decomposition, we develop a sequential framework for side-informed coding in multi-sensor networks, where each sensor observes linear noise-corrupted measurements. We construct a practical realization of this encoder using a Karhunen-Loève transform with 1-D scalar coset codes. Simulations demonstrate that simple code constructions based on the estimate-then-code partitioned structure provide improvements over their counterparts that perform the encoding directly without a pre-processing estimation step.

1. INTRODUCTION

In this paper, we consider the problem of side-informed coding where the data available at transmitters consists of indirect noise-corrupted observations of the desired signal at the decoder. Under Gaussian statistics and a mean-squared-error distortion criterion, we demonstrate that *without incurring a rate-distortion penalty* the encoder may be decomposed into two stages comprising of a “side-informed” minimum mean-squared error (MMSE) estimation followed by side-informed coding of the MMSE estimates. This generalizes the analogous result of Sakrison [1] for the non-side-informed scenario. The decomposition paves the way for the development of side-informed coding methods for our scenario, where the encoder is structured in the aforementioned two stage fashion, consisting of a first side-informed estimation stage that is followed by Wyner-Ziv encoder constructions developed for scenarios where the observations are noise-free. By recursively exploiting the decomposition, we extend our methodology to the multi-sensor scenario leading to a sequential es-

timization and coding scheme. A realization of the sequential encoding system is developed by using the side-informed estimation (based on the decomposition) in conjunction with a simple side-informed coder consisting of a Karhunen-Loève transformation (KLT) and 1-D scalar quantization with memoryless cosets [2].

2. A DECOMPOSITION FOR SIDE-INFORMED CODING OF NOISY OBSERVATIONS

We begin with the model shown in Fig.1. Two observations \mathbf{y}_0 and \mathbf{y}_1 pertaining to a vector \mathbf{x} are recorded, the first locally at a central processor (CP) and the second at a remote node (RN). The RN is connected to the CP by a channel over which R bits can be communicated noiselessly. The RN encodes its measurement \mathbf{y}_1 as a message m that is communicated to the CP over the channel, where without loss of generality we assume that $m \in \{1, 2, \dots, 2^R\}$ so as to satisfy the channel rate constraint. The receiver utilizes the local sensor measurement \mathbf{y}_0 and the message m from the RN in a “Decoder” to obtain an estimate $\tilde{\mathbf{X}} = \tilde{\mathbf{x}}_m(\mathbf{y}_0)$ of the vector \mathbf{x} . The resulting mean-squared distortion for the decoder is $D = E[\|\mathbf{X} - \tilde{\mathbf{X}}\|^2]$, where the upper-case symbols represent the random variables for which the corresponding lower-case symbols are sample realizations, both being bold when these correspond to vectors¹. For this scenario, the rate distortion function $R(D)$ represents the infimum of the channel rates for which a mean-squared distortion less than or equal to D can be achieved in the system of Fig.1, where it is assumed that the encoding may be performed over arbitrarily long blocks of *iid* observations and the rate is assumed to be amortized over the number of blocks in this process. The rate-distortion function thus establishes an upper-bound on the performance of side-informed coding methods.

In order to accomplish our desired decomposition, we consider an encoding/decoding scheme in the framework of Fig.1, defined by an encoder mapping $\mathbf{y}_1 \rightarrow m \stackrel{\text{def}}{=} m(\mathbf{y}_1)$ and a decoder mapping function $\tilde{\mathbf{x}}_m(\mathbf{y}_0)$. The mean-squared distortion for encoder-decoder chain is given by

$$D = E[\|\mathbf{X} - \tilde{\mathbf{X}}\|^2] = E[\|\mathbf{X} - \tilde{\mathbf{x}}_M(\mathbf{Y}_0)\|^2] \quad (1)$$

This work is supported in part by the National Science Foundation under grant number ECS-0428157.

¹We adopt this convention throughout this paper.

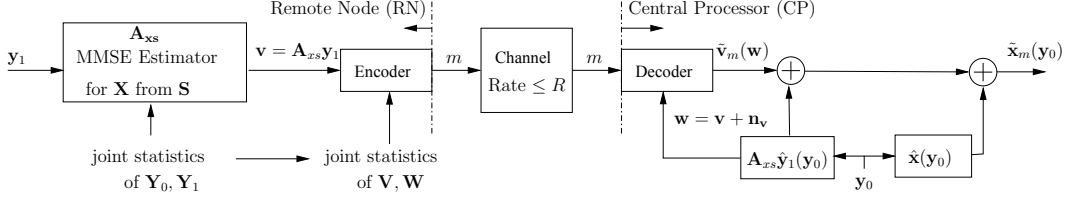


Fig. 2. Sensor observation coding in the presence of decoder side-information: Encoder and Decoder are decomposed into side-informed Estimation and side-informed Coding Stages.

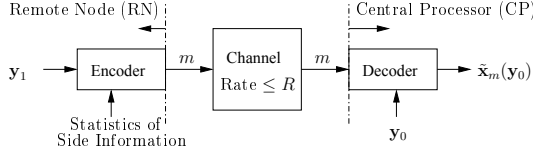


Fig. 1. Sensor observation coding in the presence of decoder side-information: Single stage Encoder and Decoder.

This mean-squared distortion may be expressed in an alternate form as

$$D = \int E \left[\|\mathbf{X} - \tilde{\mathbf{x}}_{m(\mathbf{y}_1)}(\mathbf{Y}_0)\|^2 \mid \mathbf{y}_1 \right] f_{\mathbf{Y}_1}(\mathbf{y}_1) d\mathbf{y}_1 \quad (2)$$

where $f_{\mathbf{Y}_1}(\cdot)$ denotes the probability density function of the random vector \mathbf{Y}_1 . It follows that the optimum encoding for the observation vector \mathbf{y}_1 is

$$m^*(\mathbf{y}_1) = \underset{k \in \{1, 2, \dots, 2^R\}}{\operatorname{argmin}} E \left[\|\mathbf{X} - \tilde{\mathbf{x}}_k(\mathbf{Y}_0)\|^2 \mid \mathbf{y}_1 \right] \quad (3)$$

We assume (without loss of generality) that this encoder is utilized at the RN in the system of Fig.1 and show that we can construct a corresponding partitioned encoder in the structure of Fig.2, where the latter has an expected distortion value that is no larger than D . For this purpose, we denote by $\hat{\mathbf{x}}(\mathbf{y}_0, \mathbf{y}_1)$ the minimum mean-squared error (MMSE) estimate for \mathbf{X} obtained from the observation $\mathbf{y}_0, \mathbf{y}_1$ (in the absence of any other information). We note that $\hat{\mathbf{x}}(\mathbf{y}_0, \mathbf{y}_1) = E[\mathbf{X} \mid \mathbf{y}_0, \mathbf{y}_1]$. Now in terms of this notation, for any $k \in \{1, 2, \dots, 2^R\}$, we can write

$$\begin{aligned} & E \left[\|\mathbf{X} - \tilde{\mathbf{x}}_k(\mathbf{Y}_0)\|^2 \mid \mathbf{y}_1 \right] \\ &= E \left[\|\mathbf{X} - \hat{\mathbf{x}}(\mathbf{Y}_0, \mathbf{y}_1)\|^2 \mid \mathbf{y}_1 \right] \\ &+ E \left[\|\hat{\mathbf{x}}(\mathbf{Y}_0, \mathbf{y}_1) - \tilde{\mathbf{x}}_k(\mathbf{Y}_0)\|^2 \mid \mathbf{y}_1 \right] \end{aligned} \quad (4)$$

where (as in [1]) we have used the orthogonality property of the MMSE estimators that ensures the prediction error $(\mathbf{X} - \hat{\mathbf{x}}(\mathbf{Y}_0, \mathbf{y}_1))$ is orthogonal to any function of the predictor inputs. Using (4) in (3), and observing that the first term in (4) is a constant independent of k and the optimal

selection rule (3) can be instead be formulated as:

$$m^*(\mathbf{y}_1) = \underset{k \in \{1, 2, \dots, 2^R\}}{\operatorname{argmin}} E \left[\|\hat{\mathbf{x}}(\mathbf{Y}_0, \mathbf{y}_1) - \tilde{\mathbf{x}}_k(\mathbf{Y}_0)\|^2 \mid \mathbf{y}_1 \right] \quad (5)$$

Under joint Gaussian assumption, we have

$$\hat{\mathbf{x}}(\mathbf{Y}_0, \mathbf{y}_1) = \hat{\mathbf{x}}(\mathbf{Y}_0) + \hat{\mathbf{x}}(\mathbf{y}_1 - \hat{\mathbf{y}}_1(\mathbf{Y}_0)) \quad (6)$$

Furthermore, under the assumption of joint Gaussian statistics, the MMSE estimator is a linear operator. We denote by \mathbf{A}_{xs} the estimator from $\mathbf{S} \stackrel{\text{def}}{=} (\mathbf{y}_1 - \hat{\mathbf{y}}_1(\mathbf{Y}_0))$ to \mathbf{x} , then $\hat{\mathbf{x}}(\mathbf{y}_1 - \hat{\mathbf{y}}_1(\mathbf{Y}_0)) = \mathbf{A}_{xs} \mathbf{y}_1 - \mathbf{A}_{xs} \hat{\mathbf{y}}_1(\mathbf{Y}_0)$. Note that \mathbf{S} can be viewed as the *innovation* in \mathbf{Y}_1 given \mathbf{Y}_0 and \mathbf{A}_{xs} can be viewed as a “side-informed” estimator of \mathbf{x} from \mathbf{y}_1 given \mathbf{y}_0 . Now (5) can be rewritten as:

$$\begin{aligned} & E \left[\|\hat{\mathbf{x}}(\mathbf{Y}_0, \mathbf{y}_1) - \tilde{\mathbf{x}}_k(\mathbf{Y}_0)\|^2 \mid \mathbf{y}_1 \right] \\ &= E \left[\|\mathbf{V} - \tilde{\mathbf{v}}_k(\mathbf{Y}_0)\|^2 \mid \mathbf{v} \right] \end{aligned} \quad (7)$$

where $\mathbf{v} \stackrel{\text{def}}{=} \mathbf{A}_{xs} \mathbf{y}_1$ and

$$\tilde{\mathbf{v}}_k(\mathbf{Y}_0) \stackrel{\text{def}}{=} \tilde{\mathbf{x}}_k(\mathbf{Y}_0) - \hat{\mathbf{x}}(\mathbf{Y}_0) + \hat{\mathbf{v}}(\mathbf{y}_0) \quad (8)$$

where $\hat{\mathbf{v}}(\mathbf{y}_0) = \mathbf{A}_{xs} \hat{\mathbf{y}}_1(\mathbf{Y}_0)$ is the MMSE estimate of \mathbf{v} from \mathbf{y}_0 . The decoding rule $\tilde{\mathbf{v}}_k(\cdot)$ can be viewed as a side-informed decoder for \mathbf{v} , using \mathbf{y}_0 as side information, induced by the corresponding side-informed decoding rule $\tilde{\mathbf{x}}_k(\cdot)$. Comparing (8) and (3) we see that the optimum encoding rule for \mathbf{v} (with \mathbf{y}_0 as side information) also defines the optimal encoding rule for \mathbf{y}_1 (with \mathbf{y}_0 as side information). Next considering the decoder for the former problem, and observing that $\mathbf{w} \stackrel{\text{def}}{=} \hat{\mathbf{v}}(\mathbf{y}_0)$ is trivially a sufficient statistic of \mathbf{y}_0 for estimating \mathbf{v} , we see that the optimal encoding and decoding rules for encoding of \mathbf{v} with \mathbf{w} as side information also define optimal encoding and decoding rules for \mathbf{v} with \mathbf{y}_0 as side information. For the optimal decoder thus

$$E \left[\|\mathbf{V} - \tilde{\mathbf{v}}_k(\mathbf{Y}_0)\|^2 \mid \mathbf{v} \right] = E \left[\|\mathbf{V} - \tilde{\mathbf{v}}_k(\mathbf{W})\|^2 \mid \mathbf{v} \right] \quad (9)$$

From (7) and (9), these induce the corresponding optimal encoder and decoder for our original problem of Fig.1. This establishes that the decomposition shown in Fig.2 can

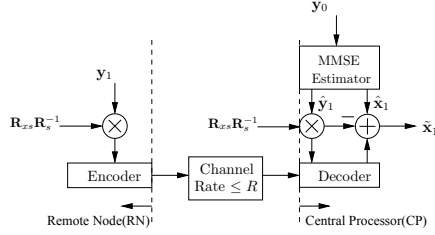


Fig. 3. Detailed encoding and decoding structure for the scenario described in Fig.1

be accomplished without any rate distortion penalty. We observe that the side-informed encoding problem in Fig.2 corresponding to the scenario where at the decoder, \mathbf{v} is to be encoded where $\mathbf{w} = \mathbf{v} + \mathbf{n}_v$ is available at the decoder, where $\mathbf{n}_v \stackrel{\text{def}}{=} \mathbf{w} - \mathbf{v} = \hat{\mathbf{v}}(\mathbf{y}_0) - \mathbf{v}$. This is the form of the well researched side-informed coding problem [3, 2, 4].

The decomposition of Fig.2 relies specifically on the Gaussian statistics and mean-squared distortion metric and therefore does come at the loss of some generality with respect to the Wyner-Ziv setting which applies to a broader class of distortion measures. Also note that since the decomposition of Fig.2 is in fact a specific instance of the general side-informed coder in Fig.1, the former cannot offer an improvement in rate-distortion performance over the latter. The decomposition of Fig.2 is useful nonetheless because it provides a structure in which practical codecs may be constructed for scenarios involving indirect and noise-corrupted sensor observations using side-informed encoder designed for the noiseless case.

For jointly Gaussian signals we have $\hat{\mathbf{x}}(\mathbf{y}_0) = \mathbf{R}_{XY_0} \mathbf{R}_{Y_0}^{-1} \mathbf{y}_0$, $\hat{\mathbf{y}}_1(\mathbf{y}_0) = \mathbf{R}_{Y_1 Y_0} \mathbf{R}_{Y_0}^{-1} \mathbf{y}_0$ and $\mathbf{A}_{xs} = \mathbf{R}_{xs} \mathbf{R}_s^{-1}$, where $\mathbf{R}_{ab} = E[\mathbf{a}\mathbf{b}^T]$ denotes cross-covariance of random vectors \mathbf{a} and \mathbf{b} and $\mathbf{R}_a = E[\mathbf{a}\mathbf{a}^T]$ denotes auto-covariance of \mathbf{a} . The partitioned decoder of Fig.2 takes the specific form shown in Fig.3.

The decomposition above had an interesting interpretation by virtue of its analogy with the Kalman filter [5], which we will demonstrate by considering the situation as if *the RN also shared the side information \mathbf{y}_0 that is available only at the CP* and regarding the observations \mathbf{y}_0 and \mathbf{y}_1 as sequential observations.

3. MULTI-SENSOR SCENARIO: A SEQUENTIAL APPROACH

The Kalman filter analogy of the preceding section suggests a sequential extension of the estimate-then-encode framework developed in the preceding section for multi-sensor scenarios. We consider a multi-sensor network as illustrated in Fig.4. A set of N RNs are deployed to estimate a desired signal \mathbf{x} , the sensor at the i^{th} RN records the measurement \mathbf{y}_i , which it

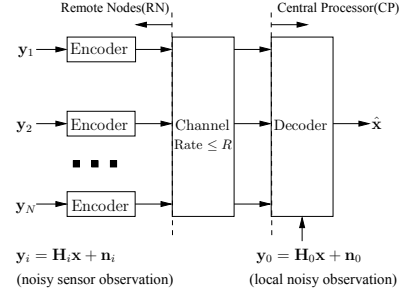


Fig. 4. Generalization of the single-sensor scenario described in Fig.1, observations from N sensors are to be encoded

pre-processes and transmits over the channel to the decoder. The decoder utilizes the messages it receives from the RNs along with its own local measurement \mathbf{y}_0 to estimate \mathbf{x} . If we assume that the encoded sensor observations arrive in sequence (for instance, if the sensors are ordered in increasing hop distance from the decoder), the received information from the lower ordered sensors may be considered as part of the side information for subsequent (higher-ordered) sensors. The estimation and coding at these sensors can therefore be performed sequentially through a recursive exploit of the decomposition of Section 2.²

For our specific development we assume all observations are modeled as linear measurements corrupted by additive noise $\mathbf{y}_i = \mathbf{H}_i \mathbf{x} + \mathbf{n}_i$, where \mathbf{H}_i is a linear observation matrix and \mathbf{n}_i is additive noise which is assumed to be independent of \mathbf{x} . \mathbf{x}, \mathbf{n}_i are assumed to be jointly Gaussian. The linear observation model *in absence of noise* has been previously considered in [3, 6]. We assume \mathbf{y}_0 is available at the decoder initially and an MMSE estimate of \mathbf{x} from \mathbf{y}_0 is obtained as $\hat{\mathbf{x}}_0$ with prediction error $\mathbf{t}_0 \stackrel{\text{def}}{=} \mathbf{x} - \hat{\mathbf{x}}_0$. When \mathbf{y}_1 is observed, the pre-processed observation $\mathbf{v}_1 = \mathbf{A}_1 \mathbf{y}_1$ is decoded as $\tilde{\mathbf{v}}_1$ provided the side information $\hat{\mathbf{v}}_1 = \mathbf{A}_1 \hat{\mathbf{y}}_1$, where $\mathbf{A}_1 = \mathbf{R}_{xs_1} \mathbf{R}_{s_1}^{-1}$, $\mathbf{s}_1 = \mathbf{y}_1 - \hat{\mathbf{y}}_1$ represents the innovation, $\hat{\mathbf{y}}_1 = \mathbf{H}_1 \hat{\mathbf{x}}_0$ denotes the MMSE estimate of \mathbf{y}_1 from \mathbf{y}_0 . In the last stage the estimate of \mathbf{x} is generated by $\tilde{\mathbf{x}}_1 = \hat{\mathbf{x}}_0 + \tilde{\mathbf{v}}_1 - \hat{\mathbf{v}}_1$ and the estimation error is $\mathbf{t}_1 = \mathbf{x} - \tilde{\mathbf{x}}_1$. The encoding and decoding for \mathbf{y}_0 reflects the process illustrated in Fig.3.

To generalize the analysis to subsequent sensors, we denote the estimate by the first k observations as $\tilde{\mathbf{x}}_k$, when \mathbf{y}_{k+1} is observed, the innovation can be represented as: $\mathbf{s}_{k+1} = \mathbf{y}_{k+1} - \mathbf{H}_{k+1} \tilde{\mathbf{x}}_k$, and the pre-processed observation $\mathbf{v}_{k+1} = \mathbf{A}_{k+1} \mathbf{y}_{k+1}$, where $\mathbf{A}_{k+1} = \mathbf{R}_{xs_{k+1}} \mathbf{R}_{s_{k+1}}^{-1}$. \mathbf{v}_{k+1} is received as $\tilde{\mathbf{v}}_{k+1}$ provided the side information $\hat{\mathbf{v}}_{k+1} = \mathbf{A}_{k+1} \hat{\mathbf{y}}_{k+1}$. The estimate of \mathbf{x} is updated by $\tilde{\mathbf{x}}_{k+1} = \tilde{\mathbf{x}}_k + \tilde{\mathbf{v}}_{k+1} - \hat{\mathbf{v}}_{k+1}$, resulting an estimation error $\mathbf{t}_{k+1} = \mathbf{x} - \tilde{\mathbf{x}}_{k+1}$. We use $\{\mathbf{K}_i\}_{i=0}^N$ to denote the auto-

²In actuality the CP receives only a quantized version of the information recorded at previous sensors. In our development, however, we ignore the impact of quantization error.

covariance of $\{\mathbf{t}_i\}_{i=0}^N$, and it can be shown that

$$\begin{aligned}\mathbf{A}_{k+1} &= \mathbf{K}_k \mathbf{H}_{k+1}^T (\mathbf{H}_{k+1} \mathbf{K}_k \mathbf{H}_{k+1}^T + \mathbf{R}_{\mathbf{n}_{k+1}})^{-1} \\ \mathbf{K}_{k+1} &= \mathbf{K}_k - \mathbf{R}_{\mathbf{x}_{s_{k+1}}} \mathbf{R}_{s_{k+1}}^{-1} \mathbf{R}_{s_{k+1}} \mathbf{x} \\ &= \mathbf{K}_k - \mathbf{K}_k \mathbf{H}_{k+1}^T (\mathbf{H}_{k+1} \mathbf{K}_k \mathbf{H}_{k+1}^T + \mathbf{R}_{\mathbf{n}_{k+1}})^{-1} \mathbf{H}_{k+1} \mathbf{K}_k\end{aligned}$$

$\mathbf{A}_{k+1}, \mathbf{K}_{k+1}$ will be used for the processing of next available sensor.

The pre-processed signal \mathbf{v}_k is transmitted by a side-informed Wyner-Ziv codec using $\hat{\mathbf{v}}_k = \mathbf{A}_k \hat{\mathbf{y}}_k$, equivalently represented by $\mathbf{w}_k = \mathbf{v}_k + \mathbf{n}_k$, as the side information. Note that the noise contained in the side information $\mathbf{n}_k = \mathbf{A}_k \mathbf{s}_k$, whose auto-covariance matrix can be represented as $\mathbf{R}_{\mathbf{n}_k} = \mathbf{R}_{\mathbf{x}_{s_k}} \mathbf{R}_{s_k}^{-1} \mathbf{R}_{s_k} \mathbf{x}$. For the purpose of noise-whitening, we use a KLT based transform code [7]. The transform matrix, \mathbf{U}_k^T , is generated by the KLT of $\mathbf{R}_{\mathbf{n}_k}$:

$$\mathbf{U}_k \Lambda_k \mathbf{U}_k^T = \mathbf{R}_{\mathbf{n}_k}$$

\mathbf{U}_k is correspondingly applied to \mathbf{v}_k and \mathbf{w}_k . The noise-whitened observations are encoded by 1-D scalar quantization with memoryless cosets [2] in our simulations. The readers are referred to [3, 2, 4] for a detailed description of various implementations of Wyner-Ziv codec.

4. IMPLEMENTATION AND SIMULATION RESULTS

In our simulations, we set the maximum number of RNs as $N = 3$. We consider a 4×1 signal vector \mathbf{x} with Toeplitz auto-covariance matrix \mathbf{R}_x with $[1, \rho, \rho^2, \rho^3]$ as the first row, where $\rho = 0.8$. The measurement matrices \mathbf{H}_i are generated with i.i.d random coefficients each distributed as $\mathcal{N}(0, 1)$. Observation noises \mathbf{n}_i are i.i.d $\mathcal{N}(0, \sigma^2)$ random variables. Different values of σ are considered in order to evaluate the proposed coding scheme under high noise and low noise regimes. We use 1-D scalar quantization and memoryless coset code [2] in our implementation of the side-informed encoder and decoder. We assume a uniform allocation of the total available rate between the RNs. The available rate at an individual RN is distributed over the channels in its KLT based coder using a simple greedy heuristic briefly summarized in Table. 1, which assumes a 6 dB reduction in distortion for the channel, per allocated bit.

<i>Inputs:</i>	
Total available bit rate R	
Transform channel variances: $\{\sigma_k^2, k = 1, 2, \dots, M\}$	
<i>Initialization:</i>	
Bit rates $\{\mathbf{r}_k = 0, k = 1, 2, \dots, M\}$	
idx = 0	
<i>Iterations:</i>	
while (idx < R)	
$i = \arg \max \sigma_k, k = 1, 2, \dots, M$	
$\sigma_i = \sigma_i / 2$	
idx = idx + 1	
endwhile	

Table 1. Simple rate-allocation heuristic based on a 6 dB reduction in distortion per allocated bit

Figure 5 presents simulation results comparing the rate distortion performance of different encoder constructions. These comparisons illustrate the practical utility of the proposed estimate-then-encode framework. In each of the sub-figures in Fig.5, the abscissa is the average bit rate per RN and the ordinate is the decoder MSE distortion, expressed in dB as $10 \log_{10} \frac{E\|\mathbf{x} - \hat{\mathbf{x}}\|^2}{E\|\mathbf{x}\|^2}$. All results presented represent averages over 200 Monte Carlo simulations. Since the full rate-distortion behavior of the problem (which is commonly referred to as the CEO problem [8]) is an open problem thus far [9], we use a looser, though more readily computable, lower bound on the distortion as our benchmark, i.e. the MSE of $E(\mathbf{x} | \{\mathbf{y}_i\}_{i=0}^N)$ which is the MMSE estimate when the observations are all available locally at the central processor. These lower bounds are represented in Fig.5 by the three horizontal dashed lines (respectively in the order going from upper to lower). We define the measurement SNR in dB for the i^{th} sensor as $10 \log_{10} \frac{E\|\mathbf{H}_i \mathbf{x}\|^2}{E\|\mathbf{n}_i\|^2} = \frac{1}{\sigma_i^2}$ and utilize a value of $\sigma_i^2 = 0.05 \equiv 13\text{dB}$ for the high SNR regime and $\sigma^2 = 1 \equiv 0\text{dB}$ for the low SNR regime.

Figure 5(a) compares the performance of the proposed estimate-then-encoder (E-E) as described in Section 3 with and without a side-informed coder for the high SNR regime. We show the benefit of side-informed coding by comparing it with conventional coding where signals are quantized by a probability density function (pdf) optimized Lloyd-Max [10] quantizer. We observe side-informed coding always offers better performance. In particular, when $R = 3$, the improvement is about 2dB when 3 sensor observations are available. In high rate regimes ($R \geq 5$), both coding scheme approach the lower bound defined earlier.

To illustrate the benefit that the proposed “side-informed” estimate can offer as a pre-processing technique for practical side-informed coding methods, we use our simple side-informed coder (1-D scalar quantization with memoryless cosets) and compare the scenarios where the encoder either: a) uses the E-E scheme, or b) directly encode (D-ENC) the sensor observations. In the latter scheme, only the pre-processing is omitted, however, the side-informed coding, noise-whitening, and rate allocation all remain the same. Fig.5(b) shows the simulation results for the high SNR scenario. We observe the proposed E-E scheme outperforms all the time. With a bit rate $R = 3$, the improvement is about 0.5dB invariant of number of sensors. As expected, both schemes converge to the lower bounds in the high rate regime.

Figure 5(c) represents the low SNR regime counterpart for Fig.5(a) wherein the encoding of pre-processed observation according to Section 3 is compared for encoders with and without side information. In this case the results indicate that our simple side-informed encoder constructions are far from optimal and the non-side informed, pdf optimized encoder performs better. The high level of noise renders the side information irrelevant and the pdf optimization therefore

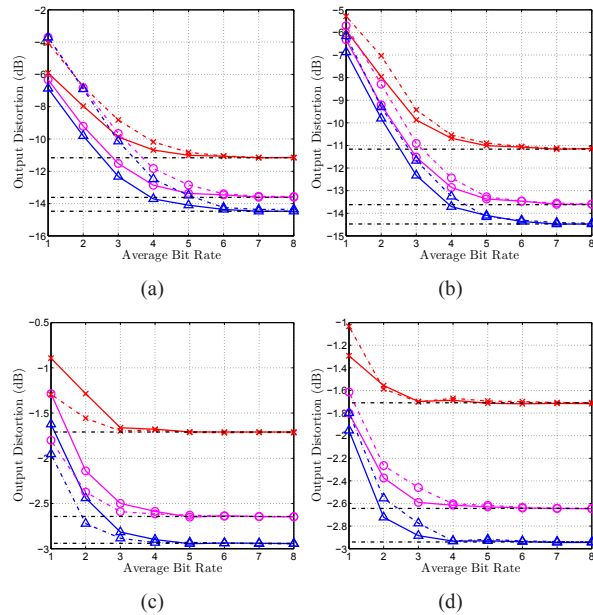


Fig. 5. Rate-distortion performance for different encoder constructions. Results are shown for various number of sensors: “×”: 1 sensor; “o”: 2 sensors; “Δ”: 3 sensors. a). HSNR, E-E with (Solid) vs. without (Dashed) SI. b). HSNR, E-E (Solid) vs. D-ENC (Dashed). Both utilize SI. c). LSNR, E-E with (Solid) vs. without (Dashed) SI. d). LSNR, E-E (Solid) vs. D-ENC (Dashed). Both do not utilize SI.

offers better performance than our elementary side-informed coder. A better design of the Wyner-Ziv codec in this case would performance comparable to (possibly slightly better) than the non-side informed approach.

Figure 5(d) examines performance of the E-E scheme in the low SNR regime. In this case, the encoder for both methods is the non side-informed encoder since this offered better performance in 5(a). The plots clearly show that the E-E scheme, despite the irrelevance of the SI, offers better performance.

5. CONCLUSION

We consider the problem of side-informed coding of sources whose observations are indirect noise-corrupted observations of the desired signal at the decoder. We show that under Gaussian statistics, for a mean-squared distortion metric, the side-informed encoding problem can be decomposed into a “side-informed” minimum mean-squared error (MMSE) estimation followed by side-informed coding of the MMSE estimate, without a degradation in rate-distortion performance. The decomposition motivates the development of practical side-informed coding methods for these scenarios. We develop a sequential scheme for communication of noise-corrupted linear measurements from multiple sensors to a central processor by recursively using the decomposition. Simulations per-

formed with a practical encoder implementation combining this framework with a simple side-informed encoder demonstrate promising performance: simple constructions based on the partitioned structure provide improvements over their counterparts that perform the encoding directly without a pre-processing estimation step.

6. REFERENCES

- [1] D. Sakrison, “Source encoding in the presence of random disturbance,” *IEEE Trans. Info. Theory*, vol. 14, pp. 165–167, 1968.
- [2] S. Pradhan and K. Ramchandran, “Distributed source coding using syndromes (DISCUS): design and construction,” *IEEE Trans. Info. Theory*, vol. 49, no. 3, pp. 626–643, 2003.
- [3] R. Zamir, S. Shamai, and U. Erez, “Nested linear/lattice codes for structured multiterminal binning,” *IEEE Trans. Info. Theory*, vol. 48, no. 6, pp. 1250–1276, 2002.
- [4] Z. Xiong, A. Liveris, and S. Cheng, “Distributed source coding for sensor networks,” *IEEE Sig. Proc. Mag.*, vol. 21, no. 5, pp. 80–94, 2004.
- [5] S. Haykin, *Adaptive Filter Theory*, 4th ed. NJ: Prentice Hall, 2002.
- [6] A. Vosoughi and A. Scaglione, “Precoding and decoding paradigms for distributed vector data compression,” *IEEE Trans. Signal Proc.*, pp. 1445–1460, Apr 2007.
- [7] V. Goyal, “Theoretical foundations of transform coding,” *Signal Processing Magazine, IEEE*, vol. 18, no. 5, pp. 9–21, 2001.
- [8] H. Viswanathan and T. Berger, “The quadratic Gaussian CEO problem,” *IEEE Trans. Info. Theory*, vol. 43, no. 5, pp. 1549–1559, 1997.
- [9] Y. Oohama, “The rate-distortion function for the quadratic Gaussian CEO problem,” *IEEE Trans. Info. Theory*, vol. 44, no. 3, pp. 1057–1070, 1998.
- [10] S. P. Lloyd, “Least squares quantization in PCM,” *IEEE Trans. Info. Theory*, vol. 28, no. 3, pp. 129–137, Mar. 1982.