Dimension Reduction and Expansion: Distributed Source Coding in a Noisy Environment

Anna N. Kim* & Fredrik Hekland†

*Centre for Quantifiable Quality of Service in Communication Systems (Q2S)
Norwegian University of Science and Technology, NO-7491 Trondheim, Norway
†ABB Corporate Research Centre (CRC), NO-1375 Billingstad, Norway
E-mail: {anna.kim@q2s.ntnu.no; fredrik.hekland@no.abb.com}

Abstract

We studied the problem of distributed coding and transmission of inter-correlated sources with memory. Different from the conventional distributed source coding structure which relies on design of effective channel codes to model the inter-correlation and quantizer, the proposed system utilizes distributed compressed sensing [1] for signal dimension reduction through linear matrix operations and dimension expansion for protection against channel noise through a hybrid scalar quantizer linear coder [2]. The proposed system is optimized for minimum end-to-end distortion under a transmission energy constraint. Its performance is verified through simulation and can serve as a good starting point for designing similar analogue based dimension reduction-expansion schemes for applications in sensor networks.

1 Introduction

In a wireless sensor network, the sensors collect information and communicate it to the sink over wireless channels. When a group of sensors are deployed to observe the same physical phenomenon, for example, sensing a temperature field over a period of time or monitoring an event using image or acoustic signals, observations from the different sensors are often correlated. Distributed source coding (DSC) can be applied in this case to exploit the correlation between the sources, where the correlated observations can be encoded separately and decoded jointly for more energy efficient communication without sacrificing much performance. Figure 1(a) illustrates DSC of two correlated sources \(x_1, x_2\), where there is no communication between the two encoders. Slepian and Wolf (S-W) [3] proved that when \(x_1, x_2\) are correlated memoryless discrete sources, the joint entropy \(H(x_1, x_2)\) is sufficient to encode them losslessly, which is as efficient as joint encoding. Wyner and Ziv later derived the corresponding bounds for lossy coding with a distortion criterion [4].

Much of the existing work in DSC adopt the structure of combining source codes (quantization) [5] with S-W coding using effective channel codes [6],[7],[8]. The inter-correlation between the sources is generally modeled as a binary channel and capacity achieving codes can then be used to reach the Slepian-Wolf bound by transmitting only the syndrome bits. When transmitting the encoded observation over a noisy channel, the channel codes then must also take error protection into account. The resulting system often have to pay price either in terms of large block length (delay) for better performance [7] or typically perform closer to the theoretic limit over poor quality channels [9].

1This work was supported by the Research Council of Norway, under Project TAIL I/O
Another challenge in DSC is to design general, practical coding schemes for correlated sources with memory. A conventional approach would be to first de-correlated the source by using for example transform coding, followed by quantization and S-W coding as in the case of DSC of memoryless sources. It is also a very challenging task to design the optimal channel code which take both inter- and intra-correlations, plus the channel noise effect into account [10]. And the associated increase in complexity, delay and power consumption can restrict their application in sensor networks.

Recently, from the statistics community, the approach of data acquisition through \textit{compressed sensing} (CS) [11] generated a series of promising results and possible applications [12],[13],[14]. In brief, given a discrete time, real valued observation vector $x \in \mathbb{R}^N$, which is known to be compressible in some transform domain of basis functions $\Psi$. Compressed sensing states that instead of applying the standard transform coding, i.e. first calculating all $N$ coefficients and then discarding many to achieve compression, successful signal recovery can be achieved by having a small number of \textit{direct} measurements $m \in \mathbb{R}^n$ of $x$, using a random projection matrix $\Phi$ of dimension $n \times N$ which satisfies a set of requirements and $m = \Phi x$. Signal dimension reduction or compression is achieved when $n \ll N$. Recovery of the source from $m$ can be done by solving the convex optimization problem:

$$
\min_\theta ||\theta||_1, \quad \text{subject to} \quad \Phi \Psi \theta = m. \tag{1}
$$

In other words, the decoded signal $\hat{x}$ can be constructed from a set of transform coefficients $\theta$ with the smallest $\ell_1$ norm and some basis functions $\Psi$ (e.g. wavelet or Gabor) known at the decoder, under the constraint the resulting measurement is same as that of $m$. As CS emerges as an efficient and versatile tool for acquiring data with memory, a natural extension is make it operate in a distributed manner for application in sensor networks. In [1], a distributed compressed sensing (DCS) framework was proposed. The authors outlined the independent encoding using CS, the joint decoding schemes and derived analogous achievable \textit{measurement rate}\footnote{It is not the source coding rate in bits per symbol, but the dimension reduction ratio $n/N$.} re-
gions. DCS in a noisy environment is however a problem that is yet to be addressed. Within CS, on the other hand, come a number of reconstruction algorithms designed for noisy measurements, at the expense of having more measurements or increased complexity [15],[16].

In this paper we propose a distributed source coding structure for transmission over noisy channels. It combines signal dimension reduction through CS and dimension expansion using a simple hybrid coding scheme proposed in [2] for better protection against channel noise. The system does not consist of any channel coding blocks. The performance gain of a dimension expansion system can be explained as follows. In a generic, point-to-point wireless communication system, the optimal performance theoretically attainable (OPTA) is prescribed by equating the rate-distortion function of the source with the channel capacity [17]. Different OPTA curves can be obtained for different source-channel bandwidth ratios, which are in practice obtained by combining (or distributing) $S$ source samples into (or over) $C$ channel samples. We illustrate in Figure 1(b) for the case of independently identically distributed (i.i.d.) Gaussian source over additive white Gaussian noise (AWGN) channel. Clearly, for the same level of source symbol-to-noise ratio (SNR), dimension reduction requires much better quality channels (high channel SNR) and conversely, for the same channel SNR, dimension expansion can offer much better source SNR at the expense of greater demand of channel bandwidth. This large gain in performance motivates the design of effective dimension expansion systems that can perform close to OPTA. However the task is not a simple one, especially for the case of large dimension expansion ratio (greater than 1:2). Nonetheless, simple and robust analogue dimension changing schemes can turn out to be better alternatives to their digital counterparts in joint source-channel coding [18].

Since the proposed system does not produce bits as in a digital system, instead of optimization in the distortion-rate sense, i.e. minimizing the average end-to-end distortion at given rate constraint, we optimize the system performance with a transmission energy constraint, which can also be interpreted as analogous to that of rate budget. The rest of paper is organized as follows. We first describe the problem at hand and introduce the proposed system structure in detail in section 2. We then analyze the end-to-end distortion in section 3 and outline the optimization routine. Simulation results and discussion are presented in section 4 and we conclude in section 5 with possible future directions.

2 Problem Formulation and System Structure

Consider the following communication scenario. A group of sensors are observing a common source from a distance. Observation $x_j \in \mathbb{R}^N$ is a discrete time, continuous amplitude signal from sensor $s_j$. The observations share a common component $x_C$ plus an individual innovation component due to for example temporal or spatial correlated sensing: $x_j = x_C + z_j, j \in \{1, 2, ..., J\}$. Both $z_j$ and $x_C$ are said to be compressible in the same transform basis. This situation is referred to as joint sparsity model type-1 in [1]. Each sensor encodes its observation independently and transmit it to the sink over channels corrupted with additive white noise $\eta \sim \mathcal{N}(0, \sigma^2)$. We wish to recover each sensor’s observation with a squared error fidelity criterion, averaged over all observation samples: $D_j = \mathbb{E}[\|x_j - \tilde{x}_j\|^2], j \in \{1, 2\}$, where $x_j$ and $\tilde{x}_j$ are the original and decoded source vectors respectively.

The proposed transmission system has the structure depicted in Figure 2. We
consider the simple case of a sensor pair $s_1$ and $s_2$. The structure can be repeated over multiple pairs.

On the encoder side, sensor $s_j$ measures its observation using a random projection matrix $\Phi_j$ of size $n \times N$ and $n \ll N$. The project matrix $\Phi_j$ is made of i.i.d. Gaussian entries with variance $1/n$ and is constructed as stated in Theorem 7 in [1]. We refer the reader to [1] for a detailed description of the algorithm and only utilize the results here. The top portion of the project matrix is for reconstruction of the signal difference $x_1 - x_2$, which is common for the two encoders and the bottom for recovering signal average $\frac{1}{2}x_1 + \frac{1}{2}x_2$, which is unique for each encoder. The resulting measurement vector is then:

$$m_j = \Phi_j x_j; \quad \Phi_j^T = [\Phi_D^T \Phi_{A_j}^T] \quad \text{with} \quad m_j \in \mathbb{R}^n; \quad x_j \in \mathbb{R}^N \quad \text{and} \quad j \in \{1, 2\} \quad (2)$$

We encode the measurement vector $m_j$ using the hybrid scalar quantizer linear coder (HSQLC) [2]. It is first quantized by a scalar quantizer $Q$. Both the quantization indices and the quantization noise are transmitted over the AWGN channel. For each source symbol $m_{ij}$ of measurement vector $m_j$, there are corresponding two channel symbols: $y_{ij}^1$ (quantization indices) and $y_{ij}^2$ (quantization error). Hence, the resulting system has a source-to-channel space dimension conversion ratio of 1:2. The quantization indices are transmitted as multilevel PAM (pulse amplitude modulation) symbols and quantization error as direct PAM. Proper power allocation between $y_{ij}^1$ and $y_{ij}^2$ is applied while satisfying the channel power constraint.

At the decoder side, the received channel symbols of quantization indices are re-quantized with $\hat{Q}$, which suppresses the additive channel noise. $\hat{y}_{ij}^1$ is then added to the quantization error $\hat{y}_{ij}^2$, which is decoded via an optimal linear decoder (Wiener filter), to obtain the decoded measurement vector $\hat{m}_j$.

To recover the correlated observations, we utilize the optimization principle Basis Pursuit (BP) [20], which is proven to be nearly optimal for solving the $\ell_1$ optimization problem in Eqn.1 and can be implemented efficiently using linear programming. Joint decoding of the two correlated observations are performed as proposed in [1]. We apply BP for reconstructing the difference signal $\tilde{x}_1 - \tilde{x}_2$ and the average $\frac{1}{2}\tilde{x}_1 + \frac{1}{2}\tilde{x}_2$, from which $\tilde{x}_1$ and $\tilde{x}_2$ immediately follow.
3 Distortion Minimization with Power Constraint

We express the total average end-to-end distortion $D_{tot}^{j}$ as the sum of the distortions from dimension reduction (DCS) and dimension expansion (HSQLC), assuming the two components are uncorrelated. Given the total transmission energy budget $E_{tot}^{j}$ for transmitting the measurement vector $m_j$ from sensor $s_j$, we want to solve:

$$\min_{n,Q,\tilde{Q},K_1,K_2} D_{tot}^{j} = D_{red}^{j}(n) + D_{exp}^{j}(Q,\tilde{Q},K_1,K_2,\hat{P}_j)$$ (3)

subject to $\sum_n P_j \leq E_{total}^{j}$, $j \in \{1,2\}$ (4)

where $n$ is the dimension of the measurement vector, $P_j$ is the average transmission power in joule per channel use, $Q$ and $\tilde{Q}$ are the quantizers for HSQLC at encoder and decoder side respectively and $K_1, K_2$ are determined by the power allocation between the quantization error and quantization indices.

One important result from Donoho’s paper on compressed sensing is that for a compressible signal $x \in \mathbb{R}^N$ with the reconstruction $\hat{x}$ obtained by solving Eqn. 1, the squared reconstruction error is bounded by:

$$||x - \hat{x}||_2^2 \leq C(n/\log(N))^{1-2/p} \quad 0 < p < 2,$$ (5)

where again $n$ is dimension of the measurement vector, $C$ is a constant greater than zero, and $p$ governs the sparsity of transform coefficients $\theta$ of $x$ in some orthonormal transform basis $\Psi$. The case $p = 2$ is when $\theta$ is no longer sparse. We can see that the accuracy of signal approximation improves with increasing number of measurements $n$. When $n = \mathcal{O}(K \log(N))$, with a reasonable constant $C$, CS’s performance is comparable to that of approximation using $K$ largest transform coefficients.

At the same time, due to the transmission energy budget, the available average power for channel symbols becomes lower as $n$ increases. This results in an increase of $D_{exp}^{j}$. The optimization essentially seeks for the trade-off between two distortion components that results in a minimum overall distortion, while satisfying the power constraint.

3.1 Distortion from Dimension Reduction

At the receiver side, the joint decoding first recover the difference signal $\hat{x}_1 - \hat{x}_2$ from solving

$$\min_{\theta_{diff}} ||\theta_{diff}||_1 \quad \text{subject to} \quad \Phi_D(\hat{x}_1 - \hat{x}_2) = \hat{m}_1 - \hat{m}_2$$ (6)

using BP, where $\Phi_D$ is the portion of the projection matrix $\Phi$ that is constructed for the difference of the correlated observations and $\hat{m}_1$ and $\hat{m}_2$ are the decoded measurements from dimension expansion. We denote the resulting distortion $D_{diff}$. After knowing $\hat{x}_1 - \hat{x}_2$, $\hat{m}_1$, $\hat{m}_2$, and the relation:

$$x_1 - \frac{1}{2}(x_1 - x_2) = \frac{1}{2}x_1 + \frac{1}{2}x_2 = x_2 + \frac{1}{2}(x_1 - x_2)$$ (7)

we use BP to solve:

$$\min_{\theta_{ave}} ||\theta_{ave}||_1 \quad \text{subject to} \quad \Phi_1\left(\frac{1}{2}\hat{x}_1 + \frac{1}{2}\hat{x}_2\right) = \hat{m}_1 - \frac{1}{2}\Phi_1(\hat{x}_1 - \hat{x}_2).$$ (8)
and we denote the distortion as $D_{ave}$ for constructing $\frac{1}{2}\tilde{x}_1 + \frac{1}{2}\tilde{x}_2$. Then $D^c_j$ can be expressed as:

$$D^c_j = \frac{1}{2}D_{diff} + D_{ave}, \ j \in \{1, 2\}$$

(9)

after obtaining $\tilde{x}_1$ and $\tilde{x}_2$ from the average and difference signal reconstructions.

### 3.2 Distortion from Dimension Expansion

As from [2], we utilize the simplified receiver structure, and have the $D_{exp}^j$ as:

$$D_{exp}^j = \sum_{k=0}^{L-1} \sum_{q=0}^{L-1} p_{k,q} \left[ (1 - K_1)^2 \int_{d_k}^{d_{k+1}} m_j^2 f_{M_j}(m_j) dm_j + 2(1 - K_1)(K_1 a_k - a_q) \int_{d_k}^{d_{k+1}} m_j f_{M_j}(m_j) dm_j \right]$$

$$+ \int_{d_k}^{d_{k+1}} m_j f_{M_j}(m_j) dm_j + \left( (K_1 a_k - a_q)^2 + \frac{K_1^2 \sigma_j^2}{K_2} \right) \int_{d_k}^{d_{k+1}} f_{M_j}(m_j) dm_j$$

(10)

where $a_k$ is the quantization indices of $m_j$ after being quantized by $Q$ with decision levels $d_k$, $f_{M_j}(m_j)$ is the probability distribution function of $m_j$, $\sigma_j^2$ is the channel noise variance, $p_{k,q}$ is the probability of receiving $a_q$ when $a_k$ was transmitted.

The above distortion expression can be optimized numerically over different channel noise variance. The resulting $Q$ and $\tilde{Q}$ can be approximated as a scalar quantizer of fixed step sizes [21].

### 3.3 Optimization Routine

There is however, to our knowledge, not yet an explicit, closed form operational relation between $n$ and the resulting mean-squared-distortion under BP with a given random projection matrix. We outline an iterative optimization algorithm for solving Eqn. 3~4 numerically.

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>S0</strong></td>
<td>Choose a value for $n$. Initialize the $\Phi_j$ and obtain the measurement vector $m_j, j \in {1, 2}$ using Eqn 2.</td>
</tr>
<tr>
<td><strong>S1</strong></td>
<td>Scale the average channel power constraint according to the size of $m_j$ and $E_j^{tot}$. Minimize $D_{exp}^j$ for the given channel noise variance $\sigma_j^2$.</td>
</tr>
<tr>
<td><strong>S2</strong></td>
<td>Jointly decode $\tilde{x}_j$. Calculate $D_j^{tot}$.</td>
</tr>
<tr>
<td><strong>S3</strong></td>
<td>Increase $n$. Repeat <strong>S0~S2</strong> until the current $D_j^{tot}$ becomes greater than the last, stop.</td>
</tr>
</tbody>
</table>

Note that the above described process can be performed off-line, which can be desirable in sensor network applications.

### 4 Simulation and Discussion

We examine the performance of the proposed system using simulations. We look into the following simple case. The correlated observations are sparse signals, which are synthetically generated with only a few number of nonzero transform coefficients in certain transform basis. In other words, we can write $x$ in matrix form: $x = \Psi \theta$, where $\|\theta\|_0 = \mathcal{K}$ and the $\ell_0$-th norm denotes the number of nonzero elements. The
correlated observations $x_1$ and $x_2$ share a common sparse component with normalized sparsity $S_c = 0.26$ and each have individual innovations sparsity component of sparsity $S_1 = S_2 = 0.077$. This corresponds to the case of symmetric rate, i.e. the same number of measurements are taken from two observations. We present the results where the nonzero elements of $\theta_j$ are i.i.d. Gaussian distributed random entries. Similar experiments can be conducted with other distributions. All simulations are done in MATLAB with BP implementation from the SparseLab (http://sparselab.standford.edu). The observations $x_1, x_2$ are of length 80,000, which are partitioned into segments of length 800.

With the fixed transmission energy budget, at a given value of channel noise variance, we ran the optimization routine over a range of measurements to obtain the one that produces minimum end-to-end distortion (or maximum observation SNR).

Figure 3(a) shows the simulation results in terms of observation signal-to-noise versus the corresponding channel SNR. From this plot, with a known transmission energy budget and the channel noise variance, one can determine the average transmission power (which also translates into the measurement vector size) needed for the obtaining the desired observation SNR level. The $D_{exp}^j$ minimization also outputs the corresponding quantization $Q$ and $	ilde{Q}$ and power allocation parameter values used for transmission.

As we stated earlier in the introduction, there are also reconstruction algorithms in CS that are capable of effective source recovery directly from noisy measurements. We therefore setup the following a simple communication chain for comparison. It consists of the same dimension reduction from DCS, but the measurement vectors are transmitted directly over the AWGN channel using PAM, which is shown to reach OPTA [22], where each transmitted channel symbol correspond to each source symbol. We compare the two systems under the same transmission energy budget. The DCS-PAM scheme also has measurements optimized for maximum observation SNR. We use BPDN (basis pursuit denoising) [20] for reconstruction from measurements with noise. It solves the following optimization problem:

$$\min_\theta \frac{1}{2} ||m - \Phi \Psi \theta||^2_2 + \lambda \cdot ||\theta||_1$$

(11)

with appropriately chosen $\lambda$. We compare DCS-PAM with BP and DCS-PAM with BPDN and the proposed DCS-HSQLC with BP. Results are presented in Figure 3(b) for recovery of observation $x_1$. Same as in Figure 3(a), we plot the observation SNR versus channel SNR.

Here we can see that the overall performance of DCS-PAM is well below that of DCS-HSQLC with the same transmission energy budget. This is expected since as Figure 1(b) indicates dimension expansion of 1:2 has substantial gain over the 1:1 system. When the channel SNR is low, the gain from BPDN is over BP without denoising is relatively small and the performance is several dB below that with dimension expansion using BP without denoising. The gain of BPDN improves with an increasing channel SNR, however its gap to DCS-HSQLC also increases. It appears that BPDN does provide considerable benefits, however when combined with signal PAM transmission it does not match up to that of DCS-HSQLC using only BP. Bear in mind that the proposed system can also gain from using BPDN in the reconstruction process.
5 Concluding Remarks

Distributed source coding in an noisy environment is a challenging problem. In this paper, we proposed a simple and effective communication structure that ties two existing analogue type compression and transmission schemes. While distributed compressed sensing can effectively perform independent encoding and joint decoding of correlated sources with memory, the connecting bandwidth expansion scheme ensures distortion from channel noise is minimized and the two parts are jointly optimized under a total transmission energy constraint. Simulation results show the clear benefit over system operating with equal source-to-channel symbol space conversion.

An immediate step for our proposed structure is to examine real world signals and combine them with possibly improved HSQCLC suggested in [23]. Comparisons to competing digital systems are also needed.

To improve the overall system performance, both the signal dimension reduction and expansion parts can be further explored. In the case of distributed compressed sensing, much is needed to connect the proposed framework with the information theoretic bounds to effectively analyze, quantify and subsequently narrow the performance gap and making it easier to compare with other DSC schemes. [24] is a step towards that direction. At the same time, design of dimension expansion of higher ratio can be incorporated in the process, for example, using nonlinear mappings [19]. From a practical perspective, applying DSC in wireless sensor networks means that energy consumption, complexity and delay are important constraints and should be well addressed. Naturally, the use of analogue type dimension expansion in DSC is not restricted to DCS type compression schemes. It can be combined with for example Distributed KL transform [25].

References


