Distributed source coding in dense sensor networks

Akshay Kashyap[1], Luis Alfonso Lastras-Montaño[2], Cathy Xia[3], and Zhen Liu[3]

[1]Dept. of ECE, UIUC, Urbana, IL. Tel: 217-766-2537, Email: kashyap@uiuc.edu.

[2]IBM T.J. Watson Research Center, Yorktown Heights, NY. Email: lastrasl@us.ibm.com.

[3]IBM T.J. Watson Research Center, Hawthorne, NY. Emails: cathyx@us.ibm.com, zhenl@us.ibm.com.

Abstract

We study the problem of the reconstruction of a Gaussian field defined in $[0, 1]$ using N sensors deployed at regular intervals. The goal is to quantify the total data rate required for the reconstruction of the field with a given mean square distortion. We consider a class of two-stage mechanisms which a) send information to allow the reconstruction of the sensor's samples within sufficient accuracy, and then b) use these reconstructions to estimate the entire field. To implement the first stage, the heavy correlation between the sensor samples suggests the use of distributed coding schemes to reduce the total rate. Our main contribution is to demonstrate the existence of a distributed block coding scheme that achieves, for a given fidelity criterion for the sensor's measurements, a total information rate that is within a constant, independent of N, of the minimum information rate required by an encoder that has access to all the sensor measurements simultaneously. The constant in general depends on the autocorrelation function of the field and the desired distortion criterion for the sensor samples.

1 Introduction

Recent advances in wireless communications and micro-electro mechanical systems have fueled the development of technologies on wide area sensor networks. Low-powered inexpensive sensors can be dispersed on a large geographical scale and organized into networks so as to monitor physical phenomena over a large field. However, the design of such distributed sensing and networking schemes involves many challenges pertaining to the scarcity of power, bandwidth, and computing resources. Some natural questions include: How many sensors should be deployed? At what rate should data be sampled and how should the data be compressed? Is the communication network formed by the sensing nodes capable of transferring the generated data rate? Or in general, is such sensor network feasible? Effective design of distributed sensor networks therefore requires a fundamental understanding of the tradeoffs between the communication resources (such as the effective data rate between a typical sensor and the data collector), and the amount of data required for effective sensing.

In this paper we give an upper bound to the total data rate required for the reconstruction of the random field. The model under consideration is one in which a large number of sensors are deployed for data collection, while the processing of data is performed at a central location. This model has recently received much attention in the context of data gathering sensor networks [1], [2]. A question that has been the central subject in [1], [2] is whether the wireless network formed by the links between the sensor nodes is capable of supporting the communication of the data gathered by the sensor nodes to the fusion center.

It is not difficult to see that in such a scenario, there is a communication bottleneck at the central processing unit. In fact, under any model of communication in which there is an upper bound on the number of bits that can be conveyed to the center per unit time (for example, the protocol model in [3]), the average data rate that can be guaranteed between a sensor node and the central processing unit scales at best as $\Theta(\frac{1}{N})$ with the number N of sensors. This observation has led to pessimistic results regarding the feasibility of a sensor network composed of nodes charged with both sampling and communication [1], [2].

In this paper, we take the viewpoint that given this communication constraint, the compression of sensed data is of central importance. The main contribution of this paper is demonstrating the utility of distributed source coding as a means of reducing the sum rate. As the number of sensors increases, they are packed more densely, and the data of sensors located close together becomes increasingly correlated. Reducing this redundancy in data using the knowledge of the statistical correlation between sensor observations is therefore attractive. We prove that for a given distortion requirement, the rate required by distributed coding stays no more than a constant away from the rate required by joint coding of all the samples as the number of sensors becomes large. The increasing correlation between the data can be utilized in such a way that the rate-penalty of distributed coding does not grow unboundedly as the number of samples being coded grows. Our proof has the pleasing feature of showing directly that as the number (and therefore the density) of samples increases, we can make do with increasingly coarsely quantized estimates of those samples at the fusion center.

Finally, a note about the organization of this paper. We define the problem precisely in Section 2, and also discuss a simple reconstruction scheme. In Section 3, we consider different coding schemes, point to point, distributed, and centralized, and find expressions for the coding rates of the point to point and centralized schemes. The rate of distributed coding is analyzed in detail in Section 4. We present some numerical examples in Section 5. We make some conclusing remarks in Section 6.

2 Preliminaries

2.1 Problem formulation

We take a discrete time model, and assume (as in [1]) that at each time i, the field of interest is modeled by a stationary (in space) Gaussian random process $X^{(i)}(s)$ defined on the segment $[0,1]$. We assume that the process $X^{(i)}$ at time i is independent of the process $X^{(j)}$ at time $j \neq i$, and has identical statistics. In the rest of the paper, we omit the time superscript when doing so does not lead to any confusion.

Consider the scenario where n sensors are placed at regular intervals in the segment $[0,1]$, with sensor k being placed at $s_k = \frac{2k+1}{2N}$ for $k = [0, 1, \ldots, N-1]$. Each can observe the value $X^{(i)}(s_k)$ of the field at the point where it is located. Sensor k encodes a block of m observations, $[X^{(1)}(s_k), X^{(2)}(s_k), \ldots, X^{(m)}(s_k)]$ into an index I_k chosen from the set $\{1, 2, \ldots, 2^{mR_k}\}$, where R_k is the rate of sensor k, which we state in the units of nats per discrete time unit (which we refer to as a "snapshot" of the field). We assume that the blocklength is the same at all sensors and that all sensors are synchronized. The messages of the sensors are assumed to be communicated to the fusion center over a shared, rate constrained, noiseless channel. The fusion center then uses the received data to produce a reconstruction $\tilde{X}^{(i)}(s)$ of the field.

We will refer to $\mathcal{E}(X^{(i)}(s) - \tilde{X}^{(i)}(s))^2$ as the mean square error (MSE) of the reconstruction of the field at point s and time i. We are interested in the average (over a blocklength) integrated MSE of the reconstruction defined as

$$J_{MSE}(m) = \frac{1}{m}\sum_{i=1}^{m}\int_0^1 \mathcal{E}\left(X^{(i)}(s) - \tilde{X}^{(i)}(s)\right)^2 ds. \tag{1}$$

The goal is to use the quantized measurements available at the fusion center to reconstruct the entire field in such a way that for some specified positive constraint D_{net},

$$\lim_{m\to\infty} J_{MSE}(m) \le D_{net}.$$

Due to communication constraints, we want to achieve this goal with the least total rate $\sum_{k=0}^{N-1} R_k$ of communication from all the sensors to the fusion center.

In addition to the stationary requirement, we make some very mild assumptions on the random field $X(\cdot)$. For simplicity, we take $\mathcal{E}\left[X(s)\right] = 0$ and $\text{Var}(X(s)) = 1$ for all $s \in [0,1]$. Thus the covariance matrix of $X(\cdot)$ is the same as the autocorrelation matrix. Denote the autocorrelation function of $X(\cdot)$ by $\rho(\tau) = \mathcal{E}\left[X(s)X(s+\tau)\right]$. Clearly, $\rho(0) = 1$. We assume that the process $X(\cdot)$ is mean square continuous, and that $\rho(\tau)$ is positive and non increasing in a neighborhood of $\tau = 0$. Note that the mean square continuity of $X(\cdot)$ is equivalent [4] to the continuity of $\rho(\tau)$ at $\tau = 0$.

While we focus exclusively on one dimensional fields in this paper, it will be clear from our proofs that the results only rely on the fact that data located at points that are close together is increasingly correlated, and therefore the results proved here extend to higher dimensions.

2.2 Data Reconstruction Model

We consider a two step reconstruction model in which the fusion center decodes the sensor information to obtain quantized estimates of the sensor measurements, and then uses these values to estimate the entire field.

For a precise description it suffices to consider the reconstruction of the field at time i; we omit the time superscript in the following. Denote the N samples seen by the sensors as a vector $\mathbf{X} = [X(s_0), \ldots, X(s_{N-1})]^T$. If the exact values of the sensor samples $\mathbf{X}$ are given, then under the assumption of Gaussianity, the minimum mean square error (MMSE) estimator for any point s in the field is the linear minimum mean squared error estimator: $\mathcal{E}[X(s)|\mathbf{X}] = \Sigma_{X(s)\mathbf{X}}\Sigma_{\mathbf{X}}^{-1}\mathbf{X}$, where $\Sigma_{X(s)\mathbf{X}} = \mathcal{E}[X(s)\mathbf{X}^T]$ and $\Sigma_{\mathbf{X}} = \mathcal{E}[\mathbf{X}\mathbf{X}^T]$. This estimator minimizes the MSE for every s, and therefore minimizes J_{MSE}.

For simplicity of analysis, however, consider a simpler (in general sub-optimal) estimator, which estimates the field $X(s)$ at any point s using only the measurement taken at the sampling location $n(s)$ closest to s. Formally, for any s, define $n(s) = \frac{2k+1}{2N}$ if $s \in [\frac{k}{N}, \frac{k+1}{N})$. Using this notation, the estimator we consider is $\mathcal{E}[X(s)|X(n(s))] = \rho(s-n(s))X(n(s))$.

Denote the noisy estimates of the sensor samples by $\{\tilde{X}\left(\frac{2k+1}{2N}\right)\}_k$. If $s \neq n(s)$, our reconstruction function uses the quantized estimates in place of the true measurements in the estimator derived above: $\tilde{X}(s) = \rho(s-n(s))\tilde{X}(n(s)),\ s \neq n(s)$.

2.3 Error analysis

The error attained by this procedure can be written as

$$\begin{aligned} X(s) - \tilde{X}(s) &= [X(s) - \rho(s-n(s))X(n(s))] + \left[\rho(s-n(s))(X(n(s)) - \tilde{X}(n(s))\right] \\ &= E_S(s) + E_Q(s), \end{aligned}$$

which can be interpreted as saying that the total error can be written as the sum of the error in interpolation from the sample value and the error in the estimation of the sample value itself. Averaging over time, $J_{MSE}(m) = \frac{1}{m}\sum_{i=1}^{m}\int_0^1 \mathcal{E}\left(E_S^{(i)}(s) + E_Q^{(i)}(s)\right)^2 ds.$

In the coding schemes that we consider in this paper, two cases arise, as we shall see in Section 3. In the first case $E_S(s)$ and $E_Q(s)$ are independent, and so J_{MSE} can be written as

$$\begin{aligned} J_{MSE}(m) &= \frac{1}{m}\sum_{i=1}^{m}\int_0^1 \mathcal{E}\left(E_S^{(i)}(s)\right)^2 + \mathcal{E}\left(E_Q^{(i)}(s)\right)^2 \\ &= \frac{1}{m}\sum_{i=1}^{m}\int_0^1 \left[1-\rho^2(n(s))\right] + \mathcal{E}\left[X^{(i)}(n(s))) - \tilde{X}^{(i)}(n(s))\right]^2 \rho^2(s-n(s)). \end{aligned}$$

Further, since in each segment $\left(\frac{k}{N}, \frac{k+1}{N}\right)$, $k = 0, \ldots, N-1$, $1-\rho^2\left(s-n(s)\right) \leq 1-\rho^2\left(\frac{1}{2N}\right)$ and $\rho^2\left(s-n(s)\right) \leq 1$ we get

$$J_{MSE}(m) \leq \left\{1-\rho^2\left(\frac{1}{2N}\right)\right\} + J'_{MSE}(m), \tag{2}$$

where

$$J'_{MSE}(m) = \frac{1}{N}\sum_{k=0}^{N-1}\frac{1}{m}\sum_{i=1}^{m}\mathcal{E}\left[X^{(i)}(s_k) - \tilde{X}^{(i)}(s_k)\right]^2. \tag{3}$$

In the second case, independence between $E_S(s)$ and $E_Q(s)$ does not hold. In this case using the Cauchy-Schwarz inequality [4], $E|XY| \leq \sqrt{EX^2EY^2}$, we obtain

$$\mathcal{E}\left(E_S^{(i)}(s) + E_Q^{(i)}(s)\right)^2 \leq \mathcal{E}\left(E_S^{(i)}(s)\right)^2 + \mathcal{E}\left(E_Q^{(i)}(s)\right)^2 + 2\sqrt{\mathcal{E}\left(E_S^{(i)}(s)\right)^2\mathcal{E}\left(E_Q^{(i)}(s)\right)^2}.$$

Averaging over the time index, using the Jensen's inequality and the concavity of $y(x) = \sqrt{x}$, and assuming that N is large enough so that $\rho^2\left(\frac{1}{2N}\right) \geq \frac{1}{2}$ (so that for $s \in \left(\frac{k}{N}, \frac{k+1}{N}\right)$ $\rho^2(s-n(s))(1-\rho^2(s-n(s)) \leq \rho^2(\frac{1}{2N})(1-\rho^2(\frac{1}{2N}))$), we get

$$J_{MSE}(m) \leq \left\{1-\rho^2\left(\frac{1}{2N}\right)\right\} + J'_{MSE}(m) + \sqrt{\rho^2(\frac{1}{2N})(1-\rho^2(\frac{1}{2N}))J'_{MSE}(m)}, \tag{4}$$

with $J'_{MSE}(m)$ as in (3).

3 Data compression: rates of three coding schemes

We consider three schemes for the coding of the sensor samples. At one extreme is point to point coding, in which the coding scheme at the sensors does not make any use of the correlation between the samples. At the other extreme is the centralized coding rate, which an encoder that had access to all the sample values would have coded at. A distributed coding scheme, on the other hand, makes use of the statistical correlation of the data so that the sensors can, while encoding their samples without any collaboration, can achieve smaller rates than the rate of point to point coding.

In this section, we use the bounds on J_{MSE} derived in Section 2.3 to find the error constraint that is imposed on coding in these schemes.

3.1 Point to point coding

The scheme considered here uses N encoders $\{f_k\}_{k=0}^{N-1}$ and N decoders $\{g_k\}_{k=0}^{N-1}$ with the property that $\tilde{X}^{(1,\cdots,m)}(s_k) = g_k(f_k(X^{(1,\cdots,m)}(s_k)))$. The encoding/decoding pairs will be chosen to be a good standard lossy block code for Gaussian m-vectors with parameters to be discussed shortly.

Due to the orthogonality principle [4] and the stochastic independence of the field realizations at different time instances, for every $s \in [0,1]$ and all times $i, j \in \{1, \ldots, m\}$, the estimation error $E_S^{(i)}(s)$ is independent of the data $X^{(j)}(n(s))$, and hence also of the estimates $\tilde{X}^{(j)}(n(s))$ and the quantization error $E_Q^{(i)}(s)$. Therefore, we can use the error bound in (2). Now, if N is large enough then $1-\rho^2\left(\frac{1}{2N}\right) < D_{net}$. Therefore, from (2), the rate R_k at sensor k is that required for $\frac{1}{m}\sum_{i=1}^{m} \mathcal{E}\left[X^{(i)}(s_k) - \tilde{X}^{(i)}(s_k)\right]^2 \leq D'(N)$ where $D'(N) = D_{net} - \left(1-\rho^2\left(\frac{1}{2N}\right)\right)$. This can be found from the rate distortion function of a Gaussian source with a quadratic error distortion criterion, which for a source variance σ^2 is given by $R(D) = \frac{1}{2}\log\frac{\sigma^2}{D}$, for $0 < D \leq \sigma^2$, (and $R(D) = 0$ for $D > \sigma^2$). Therefore, if each sensor uses b nats per observation, then we require that $b = -\frac{1}{2}\log\left(D_{net} - \left(1-\rho^2\left(\frac{1}{2N}\right)\right)\right)$, in which case the sum rate is

$$Nb = -\frac{N}{2}\log\left[D_{net} - \left(1-\rho^2\left(\frac{1}{2N}\right)\right)\right] \text{ nats.} \tag{5}$$

3.2 Distributed coding of the sensor measurements

As before, we have N encoder and decoder pairs. Each encoder produces its codewords by observing sensor samples at a its own location. Nevertheless, in this case the compression mechanism takes into account the correlation of the samples with the goal of reducing the rate requirements and thus in general the decoders may use the information sent by all the encoders to produce their reconstruction:

$$\tilde{X}^{(1,\cdots,m)}(s_k) = g_k(f_0(X^{(1,\cdots,m)}(s_0)), \cdots, f_{N-1}(X^{(1,\cdots,m)}(s_{N-1})))$$

Since all observations are used in coming up with the estimate $\tilde{X}(n(s))$, independence between $E_S(s)$ and $E_Q(s)$ does not hold in general, and therefore, we use the bound in (4). So, $\lim_m J_{MSE}(m) < D_{net}$ is met if $\lim_m J'_{MSE}(m) \leq D'(N)$, where now, using (4), $D'(N) = \left(\sqrt{D_{net} - \left(1-\rho(\frac{1}{2N})^2\right)^2} - \sqrt{\rho^2(\frac{1}{2N})(1-\rho^2(\frac{1}{2N}))}\right)^2$, given that N is large enough so that $1-\rho^2\left(\frac{1}{2N}\right) < D_{net}$.

3.3 Joint encoding of the sensor measurements

In this case we have one encoder f, which has access to all the sensor samples, and N decoders $\{g_k\}_{k=0}^{N-1}$ (at the fusion center) with the property that

$$\tilde{X}^{(1,\cdots,m)}(s_k) = g_k(f(X^{(1,\cdots,m)}(s_0, \cdots, s_{N-1}))).$$

As in the case of distributed coding, independence between $E_S(s)$ and $E_Q(s)$ does not hold, and thus we will rely on (4) and the distortion constraint imposed on coding would be the value of $D'(N)$ found in Section 3.2.

The study of this case provides us with a lower bound on the rates achievable through distributed coding. The minimal rate for a given distortion criterion for this centralized coding scenario is given by the classical rate distortion function:

$$R_N^*(D) = \inf_{p(\tilde{\mathbf{X}}|\mathbf{X})} I(\tilde{\mathbf{X}};\mathbf{X}), \text{ subject to } \frac{1}{N}\mathcal{E}\left[\|\mathbf{X}-\tilde{\mathbf{X}}\|_2^2\right] \le D, \tag{6}$$

where $\|.\|_2$ indicates the Euclidean distance. This is the rate distortion of a vector Gaussian source with a common distortion criterion, and can be computed numerically [5].

4 Redundancy of Distributed source coding

Slepian and Wolf [6] proved that the optimal sum rate of distributed source coding is the same as the optimal rate of joint coding: for noiseless coding of discrete sources there is no inherent loss in rate in distributed coding. The lossy distributed source coding (L-DSC) problem is, however, still unsolved. In general, it is possible that the minimum total rate required by the best lossy distributed coding is greater than the minimum total rate required by a joint encoding of the sources. Moreover, this rate loss might increase with the number of samples being coded. For example, the redundancy of the quantization scheme in [7] increases linearly with the number of samples. We state the L-DSC problem and review relevant results below.

4.1 An achievable region for multi-terminal source coding

The rate region $\mathcal{R}(D)$ is defined as the set of all $N-$tuples of rates $(R_0, R_1, \ldots, R_{N-1})$ for which $\lim_m \frac{1}{m}\sum_{i=1}^m \frac{1}{N}\sum_{k=0}^{N-1}\mathcal{E}\left[\left(X^{(i)}(s_k)-\tilde{X}^{(i)}(s_k)\right)^2\right] \le D$. If a rate vector belongs to the rate region, we say that the corresponding set of rates is achievable.

While the rate region for L-DSC is not known even for discrete sources, inner and outer bounds on the rate region are known. An achievable region for two discrete sources appeared in [8], and was extended to continuous sources in [7]. The extension to a general number of Gaussian sources appears in [9]. Though the result is stated in [9] for individual distortion constraints on the sources, the extension to a more general distortion constraint is straightforward. We state the achievable region for distributed source coding in the form most useful to us in Theorem 1 below. In the statement of the theorem, we use $A \leftrightarrow B \leftrightarrow C$ to denote a Markov-chain relationship between random variables A, B and C, that is, conditioned on B, A is independent of C. Also for any $S \subset \{0, 1, \ldots, N-1\}$, $\mathbf{X}_S$ denotes the vector of those sources the indexes of which lie in the set S and S^c denotes the complement of the set S.

Theorem 1 $\mathcal{R}(D) \supset \mathcal{R}_{in}(D)$, *where* $\mathcal{R}_{in}(D)$ *is the set of* $N-$*tuples of rates for which there exists a* $N-$ *vector* $\mathbf{U}$ *of random variables that satisfies the following conditions.*

1. $\forall\, S \subseteq \{0, 1, \ldots, N-1\}, \quad \mathbf{U}_S \leftrightarrow \mathbf{X}_S \leftrightarrow \mathbf{X}_{S^c} \leftrightarrow \mathbf{U}_{S^c}$.
2. $\forall\, S \subseteq \{0, 1, \ldots, N-1\}, \quad \sum_{i\in S} R_i \ge I(\mathbf{X}_S; \mathbf{U}_S|\mathbf{U}_{S^c})$.
3. $\exists\ \tilde{\mathbf{X}}(\mathbf{U})$ *such that*

$$\frac{1}{N}\sum_{i=0}^{N-1}\mathcal{E}\left[\left(X(s_i)-\tilde{X}(s_i)(\mathbf{U})\right)^2\right] \le D. \tag{7}$$

Note that all the rate-constraints in Theorem 1 are tight [9] (in particular, the constraint on the sum rate is not implied by any other set of constraints).

Constructing a vector $\mathbf{U}$ satisfying the conditions of Theorem 1 corresponds to the usual construction of a forward channel for proving achievability in a rate-distortion problem. For each i, U_i can be thought of as the encoding of $X(s_i)$.

We now study a distributed coding scheme in which the sensors utilize the knowledge of the correlation between the data samples to encode them, and the fusion center jointly decodes the messages it receives from the sensors.

4.2 Bounding the rate loss

We now use techniques similar to those in [10] to bound the redundancy of distributed coding over the rate of joint coding. Consider a random vector $\mathbf{Z}$ that is distributed $N(0, pI)$, and is independent of $\mathbf{X}$, so that the vector $\mathbf{X}+\mathbf{Z}$ satisfies the Markov chain constraints of Theorem 1.

Say a vector $\mathbf{V}$ achieves the joint rate distortion in (6):

$$\mathbf{V} = \arg\min_{p(\mathbf{V}|\mathbf{X})} I(\mathbf{X};\mathbf{V}), \text{subject to } \frac{1}{N}\mathcal{E}\left[\|\mathbf{X}-\mathbf{V}\|_2^2\right] \leq D. \tag{8}$$

Now, expanding $I(\mathbf{X};\mathbf{X}+\mathbf{Z},\mathbf{V})$ in two ways, we get $I(\mathbf{X};\mathbf{X}+\mathbf{Z})+I(\mathbf{X};\mathbf{V}|\mathbf{X}+\mathbf{Z}) = I(\mathbf{X};\mathbf{V})+I(\mathbf{X};\mathbf{X}+\mathbf{Z}|\mathbf{V})$, so that

$$\begin{aligned} I(\mathbf{X};\mathbf{X}+\mathbf{Z})-I(\mathbf{X};\mathbf{V}) &\leq I(\mathbf{X};\mathbf{X}+\mathbf{Z}|\mathbf{V}) \\ &= I((\mathbf{X}-\mathbf{V});(\mathbf{X}-\mathbf{V})+\mathbf{Z}|\mathbf{V}). \end{aligned} \tag{9}$$

Since $\mathbf{V} \leftrightarrow (\mathbf{X}-\mathbf{V}) \leftrightarrow (\mathbf{X}-\mathbf{V})+\mathbf{Z}$, we have $I((\mathbf{X}-\mathbf{V});(\mathbf{X}-\mathbf{V})+\mathbf{Z}|\mathbf{V}) \leq I((\mathbf{X}-\mathbf{V});(\mathbf{X}-\mathbf{V})+\mathbf{Z})$. Subject to the constraint in (8), $I((\mathbf{X}-\mathbf{V});(\mathbf{X}-\mathbf{V})+\mathbf{Z})$ is upper bounded by the capacity of a parallel Gaussian channel, with noise $\mathbf{Z}$ and input $\mathbf{W}=\mathbf{X}-\mathbf{V}$, the power constraint on which is given by $\frac{1}{N}\mathcal{E}[\|\mathbf{W}\|^2] \leq D$. The capacity of this channel is [5] $C = \frac{N}{2}\log\left(1+\frac{D}{p}\right)$. Therefore, from (9) and the definition (8) of $\mathbf{V}$ as the rate-distortion achieving random vector, we get

$$I(\mathbf{X};\mathbf{X}+\mathbf{Z}) - R_N^*(D) \leq \frac{N}{2}\log\left(1+\frac{D}{p}\right). \tag{10}$$

So, if we can construct an estimator $\tilde{\mathbf{X}}(\mathbf{X}+\mathbf{Z})$ that satisfies condition (7), then equation (10) gives us an upper bound on the redundancy of distributed coding over joint coding. Further to minimize the sum rate needed by this choice of $\mathbf{U}$, we want to make the forward channel as noisy as possible, that is, make p in (10) as large as possible, while still making sure that the estimator $\tilde{\mathbf{X}}(\mathbf{X}+\mathbf{Z})$ satisfies (7).

For constructing the estimator $\tilde{\mathbf{X}}$ of Theorem 1, we use the optimum MMSE estimator $\tilde{\mathbf{X}}(\mathbf{X}+\mathbf{Z}) = \Sigma_{\mathbf{X}(\mathbf{X}+\mathbf{Z})}\Sigma_{\mathbf{X}+\mathbf{Z}}^{-1}(\mathbf{X}+\mathbf{Z})$, which achieves an error of

$$D_a = \frac{1}{N}\mathrm{tr}\left(\Sigma_{\mathbf{X}} - \Sigma_{\mathbf{X}(\mathbf{X}+\mathbf{Z})}\Sigma_{\mathbf{X}+\mathbf{Z}}^{-1}\Sigma_{(\mathbf{X}+\mathbf{Z})\mathbf{X}}\right) \tag{11}$$

We now wish to find the largest p for which $D_a \leq D_{net}$ (we are considering the large N case here, so that $D'(N) \simeq D_{net}$) Let us call this largest value of p, which would be a function of the number of samples N, as $p_{\max}(N)$.

Lemma 1 *Let $\rho(\tau)$ be a symmetric autocorrelation function ($\rho(0) = 1$) such that a threshold $\theta > 0$ exists with the properties that a) $\rho(\tau)$ is positive and non-increasing with increasing values of $|\tau|$ for $\tau \in (0, \theta)$ b) the inequality $1 - \rho^2(\theta)/4 \leq D$ holds. Then $p_{\max}(N) \geq \theta N$ if $N \geq 4/\theta$.*

Proof: We call a value of p allowable if the expected reconstruction error D_a in (11) is less than D. We find the largest p for the error criterion: $\mathcal{E}[(\tilde{X}(s_i) - X(s_i))^2] \leq D$ for each $i \in \{0, \ldots, N-1\}$, which is more stringent than the average error requirement.

Let us consider the estimation of $X(s_0)$. Since $\tilde{X}(s_i)$ is the best linear estimate of $X(s_i)$ from the data $\mathbf{X} + \mathbf{Z}$, any other linear estimator cannot result in a smaller expected MSE. We take advantage of this observation and choose a linear estimator that although suboptimal, is simple to analyze and yet suffices to establish the lemma.

Our estimator for $X(s_0)$ shall be the scaled average $\alpha \sum_{1 \leq i \leq N\theta} X(s_i) + Z_i$, where α will be a parameter optimized shortly. To estimate $X(s_i)$ for $i \neq 0$, simply substitute the samples used with those whose indexes lie in the set $\{i+1, \cdots, i+N\theta\}$ (or, for samples at the right edge of the interval $[0, 1]$, $\{i - N\theta, \cdots, i-1\}$; this does not lead to any change in what follows because of the stationarity of the field).

It is not difficult to see that

$$\begin{aligned}
&\mathcal{E}\left(X(s_0) - \alpha \sum_{1 \leq i \leq N\theta} X(s_i) + Z_i\right)^2 \\
&= \mathcal{E}\left[X(s_0)^2\right] - 2\alpha \sum_{1 \leq i \leq N\theta} \rho(i/N) + \alpha^2 \mathcal{E}\left(\sum_{1 \leq i \leq N\theta} X(s_i)\right)^2 + \alpha^2 \mathcal{E}\left(\sum_{1 \leq i \leq N\theta} Z_i\right)^2 \\
&\leq 1 - 2\alpha(N\theta - 1)\rho(\theta) + \alpha^2 N^2 \theta^2 + \alpha^2 N\theta p \\
&= \left[1 - 2\alpha N\theta\rho(\theta) + \alpha^2 N^2\theta^2 + \alpha^2 N\theta p\right] + 2\alpha\rho(\theta) \qquad (12)
\end{aligned}$$

by using the inequality $1 \geq \rho(\tau) \geq \rho(\theta)$ for $\tau \in (0, \theta)$ and the fact that the greatest integer not greater than $N\theta$ is at least $N\theta - 1$. The value of α that makes the bracketed expression smallest is equal to $\alpha^* = \frac{\rho(\theta)}{N\theta + p}$ (we do not optimize the entire expression for simplicity). Substitution of this value yields $1 - \frac{\rho^2(\theta)}{1+p/(N\theta)}\left(1 - \frac{2}{N\theta}\right) \leq 1 - \frac{1}{2}\frac{\rho^2(\theta)}{1+p/(N\theta)}$ as an upper bound in (12), where the latter is a consequence of the assumption that $N\theta \geq 4$. Now suppose that $p/N = \theta$, then $1 - \frac{1}{2}\frac{\rho^2(\theta)}{1+p/(N\theta)} = 1 - \frac{\rho^2(\theta)}{4} \leq D$ by the assumption on θ stated at the beginning. Note that this discussion suggests means of obtaining better (higher) allowed values for the p/N ratio by incorporating precise knowledge of the autocorrelation function and doing an optimization. $\diamond$

The purpose of this Lemma is only to establish that $p_{\max}(N)$ grows at least linearly with N. The constants presented were chosen for simplicity of presentation.

Intuitively, as we look at more and more samples, due to the increasing correlation between them, we can allow the looks to become more and more noisy. We now use this lemma to bound the redundancy of distributed coding.

Proposition 1 *For any given D_{net}, under the assumptions of Lemma 1, the sum rate of distributed coding is no more than a constant (that is independent of N) away from the rate of joint coding.*

Proof: From (10) we get $I(\mathbf{X}; \mathbf{X} + \mathbf{Z}) - R_N^*(D_{net}) \leq \frac{N}{2}\log\left(1 + \frac{D_{net}}{\theta N}\right) \simeq \frac{D_{net}}{2\theta}$. Substituting (using Theorem 1) $\sum_{i=1}^N R_i$, the sum rate for $I(\mathbf{X}; \mathbf{X} + \mathbf{Z})$, we see that $\sum_{i=1}^N R_i \leq R_N^*(D_{net}) + \frac{D_{net}}{2\theta}$, which completes the proof. $\diamond$

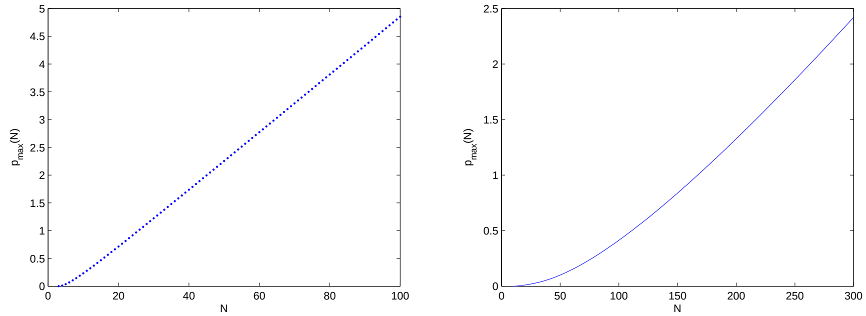

Figure 1: Linear increase of p_{max} for large N: $\rho(\tau) = \mathrm{sinc}(\tau)$ (left) and $\rho(\tau) = \exp\{-|\tau|\}$ (right). $D_{net} = 0.1$.

Note also that for $\mathbf{V}$ as in (8), we have $R_N^*(D_{net}) = I(\mathbf{X};\mathbf{V}) \leq I(\mathbf{X};\tilde{\mathbf{X}}(\mathbf{X}+\mathbf{Z}))$ from (8). Using the data processing inequality [5], $I(\mathbf{X};\mathbf{V}) \leq I(\mathbf{X};\tilde{\mathbf{X}}(\mathbf{X}+\mathbf{Z})) \leq I(\mathbf{X};\mathbf{X}+\mathbf{Z}) \leq \frac{N}{2}\log(1+1/(\theta N))$, which converges to $1/(2\theta)$, thus proving that $R_N^*(D_{net})$ is bounded as a function of N.

5 Some numerical examples

The rate found in (5) for N sensors can be minimized as a function of N to find the best possible sum rate achievable using point-to-point coding. For example, for $D_{net} = 0.1$, the least sum rate thus achieved is 6.00 nats per snapshot for $\rho(\tau) = \mathrm{sinc}(\tau) = \sin(\pi\tau)/(\pi\tau)$, attained with 4 sensors, and 23.46 nats for $\rho(\tau) = \exp(-|\tau|)$, attained with 12 sensors. Let us now compare this with the rate of joint and distributed coding.

In Figure 2 we illustrate (dashed curve) the value of $R_N^*(D'(N))$, the rate for joint coding of the samples, for $D_{net} = 0.1$, with $D'(N)$ as in Section 3.2. As N becomes large, $D'(N)$ increases to D_{net}, and therefore $R_N^*(D'(N))$ decreases. It can be seen that the limit of $R_N^*(D'(N))$ for large N is significantly smaller than the rate for point to point coding found above. The extra rate used by distributed coding (over $R_N^*(D'(N))$) has been bounded in Theorem 1. We use the MMSE estimator (11) and find the value of $p_{\max}$ numerically. We show in Figure 1 the variation of $p_{\max}$ with N, and it can be seen that the allowable growth in $p_{\max}$ is indeed linear in N. Using this value of p in (10), we get the bound on the rate of distributed coding, which is also illustrated (dotted curve) in Figure 2.

6 Conclusion

We have quantified the data required for the reconstruction of a random field using samples taken by sensors. Also, we have shown that the sum rate can be reduced greatly by using distributed source coding techniques. By showing that the sum rate required for achieving $J_{MSE} \leq D_{net}$ is a constant that depends only on the correlation structure and D_{net}, we have also answered the feasibility question posed in [1], [2] positively: the scaling of the number of bits per sensor is $\Theta(\frac{1}{N})$ which is the same as that of the bandwidth per sensor.

In proving tight bounds on the sum rate of distributed coding, we made full use to the heavy correlation between data observed at positions that are close together. When the number of sensors is large, the redundancy in their data can be utilized by coding more and more coarsely: this corresponds to more noisy samples, and is manifested in the growth of the noise $p_{\max}$ in the forward channel in Section 4.2. We believe that this technique of bounding the sum rate is of independent interest.

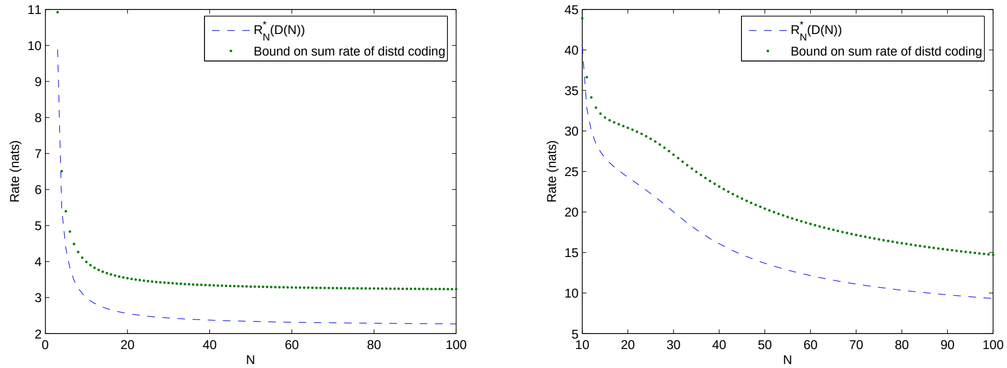

Figure 2: Rates of joint and distributed coding (in nats per snapshot) vs. number of sensors N: $\rho(\tau) = \mathrm{sinc}(\tau)$ (left) and $\rho(\tau) = \exp\{-|\tau|\}$ (right). $D_{net} = 0.1$.

While we have presented a bound on the sum rate, we have not considered any coding schemes for achieving these rates. For two sensors, dithered quantization has been studied as a technique for distributed coding in [7]. Study of such schemes is an attractive area of future research.

References

[1] D. Marco, E. J. Duarte-Melo, M. Liu, and D. L. Neuhoff, "On the many-to-one transport capacity of a dense wireless sensor network and the compressibility of its data," in *Lecture notes in Computer Science*, L. J. Guibas and F. Zhao, Eds. Springer, 2003, pp. 1–16.

[2] A. Scaglione and S. D. Servetto, "On the interdependence of routing and data compression in multi-hop sensor networks," in *Proceedings of MOBICOM*, 2002.

[3] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46(2), pp. 388–404, March 2000.

[4] B. Hajek, *An Exploration of Random Processes for Engineers*. Available online at `http://tesla.csl.uiuc.edu/~hajek/Papers/randomprocMay04.pdf`, 2004.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[6] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *Transactions on Information Theory*, vol. IT-19, pp. 471–480, July 1973.

[7] R. Zamir and T. Berger, "Multiterminal source coding with high resolution," *IEEE Transactions on Information Theory*, vol. 45, pp. 106–117, January 1999.

[8] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communication*, G. Longo, Ed. Springer, 1977.

[9] P. Viswanath, "Sum rate of multiterminal Gaussian source coding," in *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 2002.

[10] R. Zamir, "Rate loss in the Wyner-Ziv problem," *IEEE Transactions on Information Theory*, vol. 42, no. 6, pp. 2073–2084, November 1996.