COMMUNICATION-ESTIMATION TRADEOFFS IN WIRELESS SENSOR NETWORKS

S.-H. Son, S. R. Kulkarni, S. C. Schwartz[*]

Princeton University
Department of Electrical Engineering
Princeton, NJ 08544

M. Roan[†]

The Pennsylvania State University
Applied Research Laboratory
State College, PA 16804

ABSTRACT

The distributed nature of wireless sensor networks illustrates well classical engineering tradeoffs: how to minimize communication (and possibly computation) cost, and thus energy dissipation, while maintaining acceptable performance levels in estimation and inference applications. We study a simple sensor network under *dependent* Gaussian noise and develop strategies for parameter estimation in a variety of communication scenarios. From an energy point of view, sending all data to a fusion center is the most costly, but leads to optimum performance results. Processing data at each sensor and sending parameter estimates and associated quality measures is a reasonable communication saving procedure and yet, in some cases, may lead to performance equivalent to sending all data to the fusion center. A sequential procedure is most parsimonious in terms of communication cost and especially effective in large wireless sensor networks. We explore those conditions for which little, or no loss in performance is encountered with this sequential procedure. Specifically, we provide analytical expressions for the maximum likelihood estimator under "geometric" dependent noise. We show, by means of analysis and simulations, that the performance is only marginally degraded when the noise is assumed to be independent.

1. INTRODUCTION

The emergence of wireless sensor networks (WSNs) has created a resurgence in the fields of distributed detection and estimation (see e.g. [1], [2] and references therein). In the canonical WSN application, densely deployed sensors with limited range, resolution and power are expected to either detect a common occurrence or estimate parameters of interest. The sensors are able to collaborate amongst themselves or with the fusion center. The main objective is to accurately detect or estimate the state of nature while minimizing the energy resource expenditure.

While numerous works on distributed detection in WSNs exist, the literature on distributed estimation in WSNs is not as prevalent. In the present setting of the distributed estimation problem, each sensor collects observations based on a parameter of interest. Either the observations or sufficient statistics, if available, are shared amongst the sensors or passed to the fusion center. To further reduce the energy cost for communication, the observations may be quantized before transmission. Ultimately, the estimate of the parameter is obtained by optimizing a non-linear function based on the received observations (e.g., maximum likelihood or minimum mean-square error).

Deriving the maximum likelihood estimator (MLE) in a WSN setting has been studied under various contexts. The simplest approach is to send the full set of unprocessed observations to the fusion center where the clairvoyant MLE can be computed. This approach is not feasible for WSNs due to the high communication cost, but provides a benchmark for accuracy performance. Another approach is to develop procedures that take into account the power and bandwidth constraints. In [3, 4], the focus is on finding a class of MLEs that attain a variance that is close to the clairvoyant estimator when the observations are quantized to one bit. In [5], an iterative information sharing procedure based on the Fisher scoring method is explored. Both approaches lead to significant reduction in terms of energy expenditure and incur little or no loss of information.

While most of the results on distributed estimation assume that the sensor observations are conditionally independent, less is known about the broader, more difficult problem in which the sensor observations are *conditionally dependent*. In many practical applications, a large number of sensors are deployed over a finite region. Hence, some spatial correlation most likely exists among the sensor observations due to the dependent noise. The issue of distributed estimation in dependent noise was studied in [6], where the authors implemented suboptimal estimates to show that their scheme outperforms procedures which neglect dependency.

In this paper, we consider a deterministic mean location parameter estimation problem and we study the effect of dependent noise in the estimation-accuracy tradeoff. In the presence of independent Gaussian noise, deriving the MLE is straightforward and highly energy efficient. The full observation set from all the sensors does not need to be present at the fusion center when the parameter estimate is calculated. Instead, only specific quality measures based on local sensor observations are necessary. Hence, if all the sensors only send quality measures to the fusion center, $O(N)$ bit-meters of transport energy cost is required, where N is the number of sensors. Furthermore, the specific quality measures from each sensor can be passed sequentially from sensor to sensor. Given the individual sensor quality measures, a cumulative sum can be computed where each sensors adds its own local contribution to the previous cumulative sum. This sequential procedure requires $O(\sqrt{N})$ bit-meters of transport energy cost. Refer to [7] for details.

On the other hand, deriving the MLE for the dependent observation case is not analytically feasible. In addition, the MLE under dependent noise requires all the data observations to be present at the fusion center, hence costing $O(NM)$ bit-meters in terms of transport cost, where M is the number of local sensor observations. To address these issues, we explore a particular model

[*]This research was supported in part by the ONR under grant N00014-03-1-0290, the Army Research Office under grant DAAD19-00-1-0466, Draper Laboratory under IR&D 6002 grant DL-H-546263, and the National Science Foundation under grant CCR-0312413.

[†]This author performed the work while at Princeton University.

where the covariance matrix has special structure. This allows the clairvoyant MLE and the corresponding Cramer-Rao lower bound (CRLB) to have explicit analytic forms. Then a comparison is made between the variance of MLE under dependent noise with the variance of the MLE assuming independent noise to see the improvement in accuracy by incorporating dependency in the estimate. We show that the MLE assuming independent noise is asymptotically equivalent to the MLE assuming dependent noise. Hence, assuming independence drastically reduces the communication cost while only marginally reducing the accuracy performance.

The organization of this paper is as follows. Section 2 has the problem formulation with analytical results for several different sensor network architectures. Section 3 provides the simulation results and conclusion.

2. PROBLEM FORMULATION

Parameter	Description
M	# of measurements per sensor
N	# of sensors
i	measurement index: $i = 1, \dots, M$
j	sensor index : $j = 1, \dots, N$
θ	scalar parameter of interest
$x_{i,j}$	i^{th} observation from j^{th} sensor
$\boldsymbol{x}_i$	i^{th} observation vector from all the sensors

Consider a wireless sensor network (WSN) comprised of N sensors where each sensor collects M measurements. While we assume that the sensor observations are independent from measurement to measurement, they are not necessarily independent from sensor to sensor. Hence, the observations at time index i of N sensors are modeled as

$$\boldsymbol{x}_i = \theta \mathbf{1} + \boldsymbol{w}_i \tag{1}$$

where

$$f(\boldsymbol{w}_i|\Sigma) = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \boldsymbol{w}_i^T \Sigma^{-1} \boldsymbol{w}_i\right) . \tag{2}$$

We will let $f(\boldsymbol{w}_i|\Sigma)$ denote the noise probability density function (Gaussian) with covariance matrix Σ. We assume θ is fixed but unknown.

If all the observations from the WSN, $\{\boldsymbol{x}_i\}_{i=1}^{M}$, are available, the MLE of θ is given by

$$\widehat{\theta}_{MLE} = \frac{1}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} \frac{1}{M} \sum_{i=1}^{M} \mathbf{1}^T \Sigma^{-1} \boldsymbol{x}_i . \tag{3}$$

The variance of the estimator in Eq. (3) is

$$\text{Var}\,(\widehat{\theta}_{MLE}) = \frac{1}{M\,(\mathbf{1}^T \Sigma^{-1} \mathbf{1})} , \tag{4}$$

and the Fisher information is

$$\text{I}\,(\widehat{\theta}_{MLE}) = M \left(\mathbf{1}^T \Sigma^{-1} \mathbf{1}\right) . \tag{5}$$

Hence, the Cramer-Rao lower bound (CRLB) is achieved by the estimator.

The estimator in Eq. (3) represents the most general form of the MLE of θ. Depending on the structure of the covariance matrix, Σ, simplifications can be made on the form of the MLE. The simplest case is when the spatial independence is imposed on the sensor observations. If the noise is independent from sensor to sensor and $\sigma_j^2 = \sigma^2$ for all j, then $\Sigma = \sigma^2 \mathbf{I}$. The MLE of θ is given by

$$\widehat{\theta}_{SAE} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} x_{i,j} . \tag{6}$$

We will refer to the estimator found in Eq. (6) as the sample average estimator (SAE).

By observing the structure of the SAE in Eq. (6), it is evident that a sequential procedure can be implemented where each sensor only passes certain statistics of their own data from sensor to sensor. As the statistics traverse throughout the WSN, each sensor updates the current statistic values based on their own observation data. For the independent Gaussian noise case with the covariance matrix $\Sigma = \sigma^2 \mathbf{I}$, the sequential procedure only requires the sample average, $\widehat{\mu}_j$, to be computed and passed from sensor to sensor, where $\widehat{\mu}_j = \frac{1}{M} \sum_{i=1}^{M} x_{i,j}$.

This example is the motivation behind our sequential procedure. Without any loss of accuracy in the MLE estimate of θ, we save energy by implementing a sequential procedure over a centralized procedure. Now, we adopt a similar approach for a particular class of dependent noise.

2.1. Dependent Noise, Fixed Spacing

Assume that the sensor nodes are equally spaced apart and each individual sensor has the same variance. As more sensors are added, the space the sensors cover grows accordingly. The elements of the covariance matrix have a geometric form: $\Sigma_{i,j} = \sigma^2 \rho^{|i-j|}$. Then, for the 1-dimensional sensor array, the covariance matrix will be:

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{N-1} \\ \rho & 1 & \rho & \dots & \rho^{N-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \dots & 1 \end{bmatrix} . \tag{7}$$

The matrix in Eq. (7) is referred to as the Kac-Murdock-Szegö matrix [8,9] which has a simple tridiagonal inverse

$$\Sigma^{-1} = \frac{1}{\sigma^2(1-\rho^2)} \begin{bmatrix} 1 & -\rho & 0 & \dots & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 \\ & \ddots & \ddots & \ddots & \\ 0 & \dots & -\rho & 1+\rho^2 & -\rho \\ 0 & \dots & 0 & -\rho & 1 \end{bmatrix} ,$$

when $N \geq 3$. The MLE of θ is given by

$$\begin{aligned} \widehat{\theta}_{MLE} = & \frac{1}{M(N(1-\rho)+2\rho)} \\ & \times \left[\sum_{i=1}^{M} \sum_{j=1}^{N} x_{i,j} - \rho \sum_{i=1}^{M} \sum_{j=2}^{N-1} x_{i,j} \right] . \end{aligned} \tag{8}$$

The corresponding variance of the estimator is

$$\text{Var}(\widehat{\theta}_{MLE}) = \frac{\sigma^2(1+\rho)}{M\big(N(1-\rho)+2\rho\big)} . \tag{9}$$

To implement the MLE, all the observations must be sent to the fusion center. A distributed technique is not apparent since correlation exists amongst the observations. Thus, the energy expenditure in terms of transport cost is $O(MN)$ bit-meters. However, if we only calculate the sample average, a sequential procedure can be used that only costs $0(\sqrt{N})$ bit-meters. So the following questions arise. Is it worth all this extra energy cost to achieve the best MLE? What is the accuracy performance gain in terms of the variance of both estimators under dependent noise? To answer these questions, the variance of the SAE needs to be calculated under the noise conditions where Σ is Eq. (7). The difference between the variance of the MLE and the variance of the SAE provides a measure for the accuracy improvement.

Under the same noise conditions, the variance of the estimator in (6) is found to be

$$\begin{aligned}\mathrm{Var}\,(\widehat{\theta}_{SAE}) = &\frac{\sigma^2}{NM} + \frac{2\sigma^2\rho}{N^2M(1-\rho)} \\ &\times \left[(N-1) - \frac{\rho}{1-\rho}\left(1-\rho^{N-1}\right)\right] .\end{aligned} \tag{10}$$

A straightforward computation shows the following proposition.

Proposition 1 *For the covariance matrix given by Eq. (7), if $N \geq 3$ and $\rho \in (0,1)$, we have*

$$\frac{\mathrm{Var}\,(\widehat{\theta}_{\mathrm{SAE}}) - \mathrm{Var}\,(\widehat{\theta}_{\mathrm{MLE}})}{\mathrm{Var}\,(\widehat{\theta}_{\mathrm{MLE}})} = O(N^{-1})\,. \tag{11}$$

Thus, $\widehat{\theta}_{SAE}$ is asymptotically optimal. For example, when $\rho = 0.1$, $M = 10$, $\sigma^2 = 10$, $N = 10$, $\frac{\mathrm{Var}\,(\widehat{\theta}_{SAE}) - \mathrm{Var}\,(\widehat{\theta}_{MLE})}{\mathrm{Var}\,(\widehat{\theta}_{MLE})} = 0.16\%$. However, when $N = 20$, $\frac{\mathrm{Var}\,(\widehat{\theta}_{SAE}) - \mathrm{Var}\,(\widehat{\theta}_{MLE})}{\mathrm{Var}\,(\widehat{\theta}_{MLE})} = 0.090\%$. Hence, even for small values of N, the performance of $\widehat{\theta}_{SAE}$ is not much worse than that of $\widehat{\theta}_{MLE}$.

2.2. Dependent Noise, Proximity Spacing

A different model is considered when the space the sensors cover is fixed. As more sensors are added, the sensors get closer and thus, more correlated. If we add sensors, equally spaced, on a unit straight line, the maximum distance between adjacent sensors is $d = \frac{1}{N-1}$. Then the elements of the covariance matrix $A_{i,j} = \sigma^2\rho^{|i-j|d}$.

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho^{\frac{1}{N-1}} & \rho^{\frac{2}{N-1}} & \dots & \rho \\ \rho^{\frac{1}{N-1}} & 1 & \rho^{\frac{1}{N-1}} & \dots & \rho^{\frac{N-2}{N-1}} \\ \rho^{\frac{2}{N-1}} & \rho^{\frac{1}{N-1}} & 1 & \dots & \rho^{\frac{N-3}{N-1}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho^{\frac{N-2}{N-1}} & \rho^{\frac{N-3}{N-1}} & \dots & 1 \end{bmatrix} . \tag{12}$$

The MLE for the covariance matrix in (12) is

$$\widehat{\theta}_{MLE} = \frac{1}{N(1-\rho^{\frac{1}{N-1}}) + 2\rho^{\frac{1}{N-1}}} \left[\sum_{j=1}^{N} \widehat{\mu}_j - \rho^{\frac{1}{N-1}} \sum_{j=2}^{N-1} \widehat{\mu}_j\right] \tag{13}$$

with

$$\mathrm{Var}\,(\widehat{\theta}_{MLE}) = \frac{\sigma^2(1+\rho^{\frac{1}{N-1}})}{M\left(N(1-\rho^{\frac{1}{N-1}}) + 2\rho^{\frac{1}{N-1}}\right)} . \tag{14}$$

Note, as $N \to \infty$, the variance of the MLE approaches $\frac{2\sigma^2}{M\left(2-\log\rho\right)}$.

However, the variance of the SAE under the same covariance matrix is

$$\begin{aligned}\mathrm{Var}\,(\widehat{\theta}_{SAE}) = &\frac{\sigma^2}{NM} + \frac{2\sigma^2\rho^{\frac{1}{N-1}}}{N^2M(1-\rho^{\frac{1}{N-1}})} \\ &\times \left[(N-1) - \frac{\rho^{\frac{1}{N-1}}}{1-\rho^{\frac{1}{N-1}}}(1-\rho)\right] .\end{aligned} \tag{15}$$

Given the analytical expressions for the variance of both estimators, we have the following propositions.

Proposition 2 *For the covariance matrix given by Eq. (12), if $N \geq 3$ and $\rho \in (0,1)$, we have*

$$\begin{aligned}&\sup_N \left[\mathrm{Var}\,(\widehat{\theta}_{\mathrm{SAE}}) - \mathrm{Var}\,(\widehat{\theta}_{\mathrm{MLE}})\right] = \\ &-\frac{2\sigma^2}{M}\left[\frac{1}{\log\rho} + \frac{1-\rho}{\left(\log\rho\right)^2} + \frac{1}{2-\log\rho}\right] ,\end{aligned} \tag{16}$$

and

$$\sup_{N,\rho} \left[\mathrm{Var}\,(\widehat{\theta}_{\mathrm{SAE}}) - \mathrm{Var}\,(\widehat{\theta}_{\mathrm{MLE}})\right] \leq 0.072\frac{\sigma^2}{M} . \tag{17}$$

Proposition 3 *For the covariance matrix given by Eq. (12), if $N \geq 3$ and $\rho \in (0,1)$, we have*

$$\begin{aligned}&\sup_N \left[\frac{\mathrm{Var}\,(\widehat{\theta}_{\mathrm{SAE}}) - \mathrm{Var}\,(\widehat{\theta}_{\mathrm{MLE}})}{\mathrm{Var}\,(\widehat{\theta}_{\mathrm{MLE}})}\right] = \\ &\left(1 - \frac{2}{\log\rho}\right)\left(1 + \frac{1-\rho}{\log\rho}\right) - 1 ,\end{aligned} \tag{18}$$

and

$$\sup_{N,\rho} \left[\frac{\mathrm{Var}\,(\widehat{\theta}_{\mathrm{SAE}}) - \mathrm{Var}\,(\widehat{\theta}_{\mathrm{MLE}})}{\mathrm{Var}\,(\widehat{\theta}_{\mathrm{MLE}})}\right] \leq 0.14 . \tag{19}$$

Propositions 2 and 3 claim that for any N, the absolute and relative performance losses compared to the performance of the MLE are bounded by a function of ρ, as seen by the right hand side of Eqs. (16) and (18), respectively. Also, for all N and ρ, the bound on absolute performance loss is found to be $7.2\% \times \frac{\sigma^2}{M}$, while the bound on the relative performance loss is found to be 14%.

3. SIMULATION RESULTS AND CONCLUSION

We simulated a WSN under geometric dependent noise where each sensor takes 10 measurements and has the same variance, $\sigma_j^2 = 10$. All the figures represent the relationship between the variance of both estimators and ρ for various N. In Figs. (1) and (2), the distance between the sensors are fixed. However, in Fig. (1), the sensors are placed on a line, while in Fig. (2), a more realistic scenario is encountered where the sensors are placed on a square grid. As stated in Prop. 1 and evident from the graph in Fig. (1), as the number of sensors, N, increases, the gap between the variance of both estimators decreases. Also, for any fixed N, the gap between variance of both estimators is relatively small. For the case where the sensors are placed on a square grid, an analytical closed form expression has yet to be found. Unlike Fig. (1), in Fig. (2), as N increases, the gap between the variance of the two estimators does not seem to shrink. However, the difference in

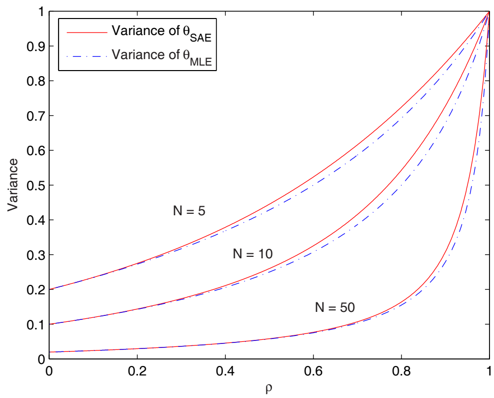

Fig. 1. Plot of variance versus ρ for the MLE and the SAE for the cases where $N = 5, 10, 50$. The entries of the covariance matrix are given by $\Sigma_{i,j} = \sigma^2 \rho^{|i-j|}$, $(M = 10, \sigma^2 = 10)$.

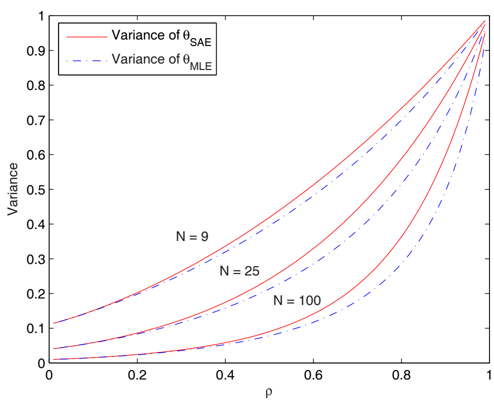

Fig. 2. Plot of the variance versus ρ for the MLE and the SAE for the cases where $N = 9, 25, 100$. The entries of the covariance matrix are given by $\Sigma_{i,j} = \sigma^2 \rho^{|i-j|}$, $(M = 10, \sigma^2 = 10)$.

the performance is not too large. As for Fig. (3), the sensors are placed on a line, but the distance between the sensors gets closer as more sensors are added. When N increases, the gap between the variances of both estimators also increases. This relationship holds since as N increases, the sensor observations get more correlated. Hence, the independent assumption is less valid. For the MLE, when $N > 5$, the variance curves are indistinguishable. Also, evident from the graph, both Props. 2 and 3 hold.

For the covariance matrices studied in this paper, there does not appear to be much degradation if you assume that the noise is independent and use the SAE. More precisely, we showed that if the dependent noise structure has the form in Eq. (7), the SAE is asymptotically equivalent to the MLE. Also, if the noise covariance has the form in Eq. (12), we found numerical and analytical bounds on the performance loss. While we incur a small loss in performance by using the SAE instead of the optimal MLE, there are considerable energy savings due to the sequential nature of calculating the SAE.

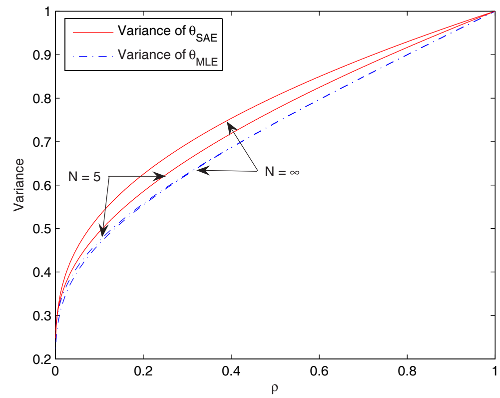

Fig. 3. Plot of the variance versus ρ for the MLE and the SAE for the cases where $N = 5, \infty$. The entries of the covariance matrix are given by $\Sigma_{i,j} = \sigma^2 \rho^{\frac{|i-j|}{N-1}}$, $(M = 10, \sigma^2 = 10)$.

4. REFERENCES

[1] S. D. Servetto, R. Knopp, A. Ephremides, S. Verdú, and S. B. Wicker editors, "Fundamental performance limits of wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, August 2004.

[2] S. Kumar, F. Zhao, and D. Shepherd editors, "Special issue on collaborative information processing," *IEEE Signal Processing Magazine*, vol. 19, no. 2, March 2002.

[3] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks, part 1: Gaussian case," *IEEE Transactions on Signal Processing*, July 2004 (submitted).

[4] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks, part 2: Unknown pdf," *IEEE Transactions on Signal Processing*, July 2004 (submitted).

[5] D. Blatt and A. Hero, "Distributed maximum likelihood estimation for sensor networks," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, May 2004, pp. 929–932.

[6] E. Geraniotis Y. A. Chau, "Distributed multisensor parameter estimation in dependent noise," *IEEE Transactions on Communications*, vol. 40, no. 2, pp. 373–384, February 1992.

[7] S.-H. Son, M. Chiang, S. R. Kulkarni, and S. C. Schwartz, "The value of clustering in distributed estimation for sensor networks," in *Information Processing in Sensor Networks*, 2004 (submitted).

[8] M. Dow, "Explicit inverses of toeplitz and associated matrices," *ANZIAM Journal*, vol. 44 (E), pp. E185–E215, January 2003.

[9] M. Kac, W. L. Murdock, and G. Szegö, "On the eigenvalues of certain hermitian forms," *Journal of Rational and Mechanical Analysis*, vol. 2, pp. 767–800, 1953.