NETWORK PARAMETER ESTIMATION WITH DETECTION FAILURES

Emre Ertin, Randolph L. Moses and Lee C. Potter

The Ohio State University
Department of Electrical Engineering
2015 Neil Avenue, Columbus, OH 43210
{ertine, randy, potter }@ee.eng.ohio-state.edu

ABSTRACT

Distributed sensing systems fuse information from various local measurements to form joint estimates of physical phenomena. We study Bayesian parameter estimation with measurement failures, where the probability of failing to make an observation depends on the unknown parameters. Both the absence and the existence of the measurements provide information about the parameters. We present lower bounds for minimum mean square error estimation. The bounds are applied to the analysis of sensor arrays for source localization and the distributed estimation problem with binary detectors.

1. INTRODUCTION

There has been renewed interest in the notion of deploying large numbers of networked sensors for applications ranging from environmental monitoring to surveillance [1]. In a typical scenario a number of sensors are distributed in a region of interest. Each sensor is equipped with sensing, processing and communication capabilities. The information gathered from the sensors can be used to detect, track and classify objects of interest. On-board transducers detect energy (acoustic, seismic, magnetic, pressure, etc.) and convert to an electrical signal. Because transmission of raw data would consume to much energy, each sensor further process data to obtain summarized decision statistics and communicates them to fusion center(s) for joint decisions. For example sensors equipped with an acoustic array can detect acoustic energy and locally process it to obtain time of arrival (TOA) and direction of arrival (DOA) measurements. These local measurements can be processed jointly to locate the source of the acoustic energy.

In practice, some of the sensors may fail to detect a source signal and the probability of failing to make an observation depends on the unknown parameters (*e.g.*, the sensor might fail to detect the source signal to compute a reliable TOA if the distance to the source is large). A suboptimal approach would be to use only the available measurements in estimation. The fact that some observations failed (and some were successful) provide information about the unknown parameters in addition to the measurements themselves. This additional information could be utilized for improved estimates.

In this paper we characterize the parameter estimation performance of distributed sensing systems with detection failures using lower bounds on minimum mean square error estimation. We then discuss the application of the derived bounds to the analysis of binary detector arrays for source parameter estimation. Two examples are presented. The first example uses a simple one dimensional sensor network to illustrate the application of the bounds to stationary target localization. The second example discusses localization with a dense set of binary detectors.

2. MODEL

Let $x_1, x_2, \ldots, x_n$ be the set of potential measurements. The random variables $\{x_i\}$ are conditionally independent and distributed with density $f_i(y|\theta)$, where $\theta \in \Theta$ represent an unknown random parameter with prior density $g(\theta)$. In addition we assume each measurement is an independent Bernoulli trial with success and failure probabilities given by $p_i(\theta)$ and $q_i(\theta) = 1 - p_i(\theta)$ respectively. In this model the absence (or the existence) of a measurement provides information about the unknown parameters θ. In particular, let $\mathcal{I}$, a subset of all possible observations $\mathcal{N} = \{1, 2, \ldots, n\}$ denote the set of successful measurements with the associated measurements $x^{\mathcal{I}} = \{x_i\}_{i \in \mathcal{I}}$. The posterior density of θ is given by Bayes' Rule as

$$g\left(\theta | \{x_i\}_{i \in \mathcal{I}}\right) = \frac{\mathsf{P}(x^{\mathcal{I}}|\theta)}{\mathsf{P}(x^{\mathcal{I}})} g(\theta) \ ,$$

where

$$\begin{aligned} \mathsf{P}(x^{\mathcal{I}}|\theta) &= \mathsf{P}[\mathcal{I}|\theta] \prod_{i \in \mathcal{I}} f_i(x_i|\theta) \\ &= \prod_{i \in \mathcal{I}} p_i(\theta) f_i(x_i|\theta) \prod_{j \in \mathcal{N}/\mathcal{I}} q_j(\theta) \end{aligned}$$

$$\mathsf{P}(x^{\mathcal{I}}) \;=\; \int_{\Theta} \mathsf{P}(x^{\mathcal{I}}|\theta)g(\theta)d\theta\,.$$

We observe that the posterior density is affected by the measured values $x^{\mathcal{I}}$ as well as the index set itself through the probability mass function $\mathsf{P}[\mathcal{I}|\theta]$. In the next section we quantify the effect of $\mathsf{P}[\mathcal{I}|\theta]$ on the estimation performance using lower bounds on mean square error.

3. PARAMETER ESTIMATION WITH MEASUREMENT FAILURES

We consider parameter estimation problems where $\Theta \subset \mathbb{R}^M$. We define the error vector as $\theta_\epsilon = \hat{\theta} - \theta$ and derive constraints on the correlation matrix of the error vector $\mathcal{R}_\epsilon = E[\theta_\epsilon\theta_\epsilon^T]$, where the expectation is taken over all $[\theta, x]$. The following theorem combines the Fisher information from the observations, existence/absence of measurements and the prior to give a lower bound on $\mathcal{R}_\epsilon$.

Theorem 1

$$\mathcal{R}_\epsilon \geq \sum_{\mathcal{I}\subset\mathcal{N}} \mathsf{P}[\mathcal{I}]\Big(\mathbf{J}_g^{\mathcal{I}} + \sum_{j\in\mathcal{N}\setminus\mathcal{I}} \mathbf{J}_{q_j}^{\mathcal{I}} + \sum_{i\in\mathcal{I}}(\mathbf{J}_{p_i}^{\mathcal{I}} + \mathbf{J}_{x_i}^{\mathcal{I}})\Big)^{-1} \quad (1)$$

*where $\geq$ refers to the positive semidefinite (psd) partial order on Hermitian matrices. (*i.e. *$A \geq B$ if $A - B$ is psd), $\mathsf{P}[\mathcal{I}]$ is the probability mass function over possible index sets*

$$\mathsf{P}[\mathcal{I}] = \int_{\Theta}\Big(\prod_{i\in\mathcal{I}} p_i(\theta)\Big)\Big(\prod_{j\in\mathcal{N}\setminus\mathcal{I}} q_j(\theta)\Big)\, g(\theta)\; d\theta$$

and the Fisher information matrices of prior, existence, absence and measurement information are computed as

$$\begin{aligned}
\left[\mathbf{J}_g^{\mathcal{I}}\right]_{kl} &= -E\left[\frac{\partial^2 \ln g(\theta)}{\partial\theta_k\,\partial\theta_l}\,\Big|\, \mathcal{I}\right]\\
\left[\mathbf{J}_{p_i}^{\mathcal{I}}\right]_{kl} &= -E\left[\frac{\partial^2 \ln p_i(\theta)}{\partial\theta_k\,\partial\theta_l}\,\Big|\, \mathcal{I}\right]\\
\left[\mathbf{J}_{q_j}^{\mathcal{I}}\right]_{kl} &= -E\left[\frac{\partial^2 \ln q_j(\theta)}{\partial\theta_k\,\partial\theta_l}\,\Big|\, \mathcal{I}\right]\\
\left[\mathbf{J}_{x_i}^{\mathcal{I}}\right]_{kl} &= -E\left[\frac{\partial^2 \ln f_i(x_i|\theta)}{\partial\theta_k\,\partial\theta_l}\,\Big|\, \mathcal{I}\right]
\end{aligned}$$

Proof: The error correlation matrix for an estimator $\hat{\theta}$ can be computed as:

$$\mathcal{R}_\epsilon = \sum_{\mathcal{I}\subset\mathcal{N}} \mathsf{P}[\mathcal{I}]E\left[\theta_\epsilon\theta_\epsilon^T|\mathcal{I}\right]$$

Under regularity conditions on $f(x^{\mathcal{I}}, \theta|\mathcal{I})$ [2], Cramer-Rao lower bound applies to each term as:

$$E\left[\theta_\epsilon\theta_\epsilon^T\,\big|\,\mathcal{I}\right] \geq E\left[\{\nabla \ln f(x^{\mathcal{I}},\theta)\}\{\nabla \ln f(x^{\mathcal{I}},\theta)\}^T\,\big|\,\mathcal{I}\right]^{-1}$$

This implies:

$$\mathcal{R}_\epsilon \geq \sum_{\mathcal{I}\subset\mathcal{N}} \mathsf{P}[\mathcal{I}]E\left[\{\nabla \ln f(x^{\mathcal{I}},\theta)\}\{\nabla \ln f(x^{\mathcal{I}},\theta)\}^T|\mathcal{I}\right]^{-1}$$

Then (1) follows from:

$$f(x^{\mathcal{I}},\theta|\mathcal{I}) = \frac{1}{\mathsf{P}[\mathcal{I}]}\Big[\prod_{i\in\mathcal{I}} p_i(\theta)f_i(x^i|\theta)\Big]\Big[\prod_{j\in\mathcal{N}\setminus\mathcal{I}} q_j(\theta)\Big]\, g(\theta)$$

The computation of the bound in (1) can be computationally prohibitive if the number of potential measurements is large. In particular, for a problem with n potential measurements, (1) requires computation of 2^n expectation terms. A computationally attractive bound on $\mathcal{R}_\epsilon$ which is linear in n is given next.

Theorem 2

$$\mathcal{R}_\epsilon \geq \Big(\mathbf{J}_g + \sum_{i=1}^{n} \mathsf{p_i}(\mathbf{J}_{p_i} + \mathbf{J}_{x_i}) + \mathsf{q_i}\mathbf{J}_{q_i}\Big)^{-1} \qquad (2)$$

where the prior probability of success and failure for measurement i is given as:

$$\mathsf{p_i} = \int p_i(\theta)g(\theta)d\theta \qquad \mathsf{q_i} = \int q_i(\theta)g(\theta)d\theta$$

and the relevant Fisher information matrices are computed as:

$$\begin{aligned}
\left[\mathbf{J}_g\right]_{kl} &= -E\left[\frac{\partial^2 \ln g(\theta)}{\partial\theta_k\,\partial\theta_l}\right]\\
\left[\mathbf{J}_{p_i}\right]_{kl} &= -E\left[\frac{\partial^2 \ln p_i(\theta)}{\partial\theta_k\,\partial\theta_l}\,\Big|\, i\in\mathcal{I}\right]\\
\left[\mathbf{J}_{q_i}\right]_{kl} &= -E\left[\frac{\partial^2 \ln q_j(\theta)}{\partial\theta_k\,\partial\theta_l}\,\Big|\, i\notin\mathcal{I}\right]\\
\left[\mathbf{J}_{x_i}\right]_{kl} &= -E\left[\frac{\partial^2 \ln f_i(x_i|\theta)}{\partial\theta_k\,\partial\theta_l}\,\Big|\, i\in\mathcal{I}\right]
\end{aligned}$$

Proof: By applying the matrix-version of the harmonic arithmetic mean inequality [3] to right hand side of (1) we get:

$$\begin{aligned}
\mathcal{R}_\epsilon \geq &\sum_{\mathcal{I}\subset\mathcal{N}} \mathsf{P}[\mathcal{I}]\Big(\mathbf{J}_g^{\mathcal{I}} + \sum_{j\in\mathcal{N}\setminus\mathcal{I}} \mathbf{J}_{q_j}^{\mathcal{I}} + \sum_{i\in\mathcal{I}}(\mathbf{J}_{p_i}^{\mathcal{I}} + \mathbf{J}_{x_i}^{\mathcal{I}})\Big)^{-1}\\
\geq &\Big(\sum_{\mathcal{I}\subset\mathcal{N}} \mathsf{P}[\mathcal{I}]\Big(\mathbf{J}_g^{\mathcal{I}} + \sum_{j\in\mathcal{N}\setminus\mathcal{I}} \mathbf{J}_{q_j}^{\mathcal{I}} + \sum_{i\in\mathcal{I}}(\mathbf{J}_{p_i}^{\mathcal{I}} + \mathbf{J}_{x_i}^{\mathcal{I}})\Big)\Big)^{-1}
\end{aligned}$$

Then, (2) follows by rewriting the expectation integrals as:

$$\begin{aligned}
\sum_{\mathcal{I}\subset\mathcal{N}} \mathsf{P}[\mathcal{I}]\mathbf{J}_g^{\mathcal{I}} &= \mathbf{J}_g\\
\sum_{\mathcal{I}\subset\mathcal{N}} \mathsf{P}[\mathcal{I}] \sum_{j\in\mathcal{N}\setminus\mathcal{I}} \mathbf{J}_{q_j}^{\mathcal{I}} &= \sum_{i=1}^{n} \mathsf{q_i}\mathbf{J}_{q_i}
\end{aligned}$$

$$\sum_{\mathcal{I}\subset\mathcal{N}} \mathsf{P}[\mathcal{I}] \sum_{i\in\mathcal{I}} \mathbf{J}_{p_i}^{\mathcal{I}} = \sum_{i=1}^{n} \mathsf{p}_i \mathbf{J}_{p_i}$$

$$\sum_{\mathcal{I}\subset\mathcal{N}} \mathsf{P}[\mathcal{I}] \sum_{i\in\mathcal{I}} \mathbf{J}_{x_i}^{\mathcal{I}} = \sum_{i=1}^{n} \mathsf{p}_i \mathbf{J}_{x_i}$$

4. DISTRIBUTED PARAMETER ESTIMATION WITH BINARY DETECTORS

Large scale sensor networks are envisioned for surveillance applications such as detection, localization and tracking of personnel. In some applications, complexity and cost constraints may dictate usage of simple binary detectors for detection and estimation tasks. For example binary tripwire detectors can be used for detection and localization of acoustic or seismic energy. In this section we apply the results of Section 3, to analyze binary detector arrays for estimation of parameters of a single source. The fusion of multiple binary decisions for joint detection decisions has been studied extensively in the literature [4, 5, 6]. Here we study the related problem of fusion of binary decisions for joint estimation of source parameters. We limit our discussion here to optimal fusion of binary decisions and don't consider joint optimization of the fusion and local detection rules.

We consider parameter estimation using n binary detectors and assume that each detector makes a local measurement of a source signal characterized by parameter θ. The parameter vector θ models the characteristics of the source such as its location, amplitude or target class. Using its local measurement, each sensor makes an independent binary decision characterized by the operating point $p_i^D(\theta)$ (*i.e.* the probability of sensor i declaring a detection for a source with parameter θ). We denote the probability of miss for sensor i as $p_i^M(\theta) = 1 - p_i^D(\theta)$. We further assume that the source parameter has a prior distribution given by the density function $g(\theta)$. With this prior, the fusion task can be posed as a Bayesian paramater estimation problem with binary measurements. The MAP estimator for θ is given by:

$$\hat{\theta}_{MAP}(\mathcal{I}) = \arg\max_{\theta} \Big[g(\theta) \prod_{i\in\mathcal{I}} p_i^D(\theta) \prod_{i\in\mathcal{N}/\mathcal{I}} p_i^M(\theta) \Big] \quad (3)$$

The results derived in the previous section can be used to bound the correlation matrix of the errors of the MAP estimator. Typically binary detector systems are deployed in large numbers to compensate for the small amount information provided by each detector. Therefore, the bound given in Theorem 1 is not practical for most applications. However, the bound in Theorem 2 is linear in the number of detectors and can be useful in the analysis of such large scale systems.

Corollary 3 *For a binary detector with n sensor nodes employing local decision rules $\{p_i^D(\theta), p_i^M(\theta)\}_{i\in\mathcal{N}}$ the error*

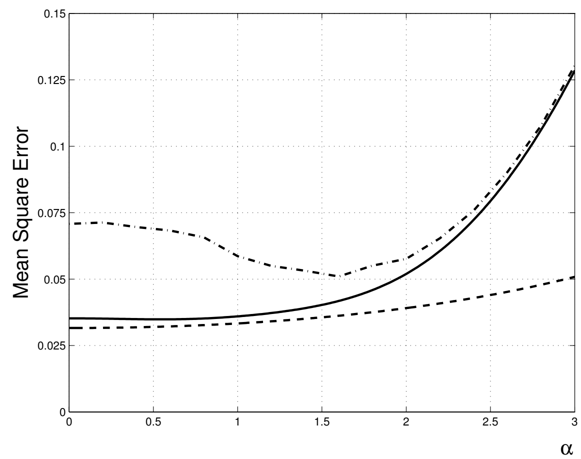

Fig. 1. Mean square error (MSE) for MAP estimator (dash-dot) and the MSE-Bounds (Eqn.(1)- solid, Eqn.(2)- dashed)

correlation matrix $\mathcal{R}_\epsilon$ is bounded as:

$$\mathcal{R}_\epsilon \geq \left(\mathbf{J}_g + \sum_{i=1}^{n} \mathsf{p}_i^D \mathbf{J}_i^D + \mathsf{p}_i^M \mathbf{J}_i^M \right)^{-1} \quad (4)$$

where the prior probability of detection and miss for detector i is given as:

$$\mathsf{p}_\mathsf{i}^\mathsf{D} = \int p_i^D(\theta) g(\theta)\, d\theta \qquad \mathsf{p}_\mathsf{i}^\mathsf{M} = \int p_i^M(\theta) g(\theta)\, d\theta$$

and the Fisher information matrices associated with prior, detection and miss information is computed as:

$$\big[\mathbf{J}_g\big]_{kl} = -E\left[\frac{\partial^2 \ln g(\theta)}{\partial\theta_k\, \partial\theta_l}\right]$$

$$\big[\mathbf{J}_i^D\big]_{kl} = -E\left[\frac{\partial^2 \ln p_i^D(\theta)}{\partial\theta_k\, \partial\theta_l} \;\Big|\; i \in \mathcal{I}\right]$$

$$\big[\mathbf{J}_i^M\big]_{kl} = -E\left[\frac{\partial^2 \ln p_i^M(\theta)}{\partial\theta_k\, \partial\theta_l} \;\Big|\; i \notin \mathcal{I}\right]$$

5. EXPERIMENTS

Experiment A
We consider a one dimensional source localization problem with time and direction of arrival measurements. A single source is placed at a point θ on a one dimensional line, with known prior density

$$g(\theta) = \frac{1}{\sqrt{2\pi\sigma_g^2}} \exp[-\frac{\theta^2}{2{\sigma_g}^2}]$$

Two sensors are located at $\{-\alpha, \alpha\}$ and make time and direction of arrival measurements for an acoustic signal emitted by the source. The measurement for the sensor i can be modeled as:

$$f_i(x_i|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big[-\frac{(x_i - (\theta - (-1)^i \alpha))^2}{(2\sigma^2)}\Big] ,$$

under the assumption that the signal emission time is known. Each sensor can fail to detect the source with probability

$$p_i(x) = \exp\Big[-\frac{(\theta - (-1)^i \alpha)^2}{r_0^2}\Big] .$$

Mean Square Error performance of the maximum aposteriori probability (MAP) estimator is obtained using 10,000 Monte Carlo simulations for $\sigma_g = 1, \sigma = 0.25$ and $r_0 = 4$. Bounds given by (1) and (2) and the MSE performance of the MAP estimator are plotted in Figure 1 for various values of α.

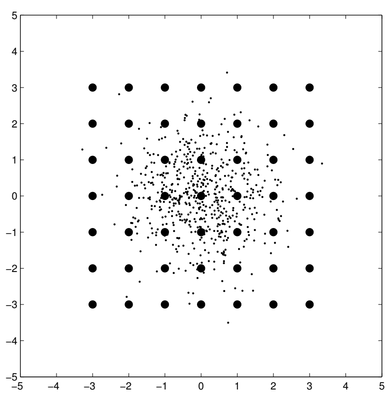

Fig. 2. Location of the binary detectors and scatter plot of the source distribution for Example 2

Experiment B
We consider a source localization problem with $n = 49$ binary detectors placed on a uniform grid covering a 7×7 square area centered at the origin. The parameter vector $\theta = [x, y]$ denotes the source location and has a prior density

$$g(x, y) = \frac{1}{2\pi} \exp\big[-\frac{x^2 + y^2}{2}\big] .$$

The location of the binary detectors and scatter plot of the source distribution is given in Figure 2. The probability of detection for a sensor located at $[s_x, s_y]$ is specified by

$$p_i^D(\theta) = \exp\Big[-\frac{(x - s_x)^2 + (y - s_y)^2}{r_0^2}\Big] ,$$

where r_0 is a user selected parameter specifying the sensitivity of the detector. We computed the performance of two estimators using 10,000 MonteCarlo simulations as measured by the average radius of the uncertainty ellipse. The first estimator is the MAP estimator given in (3). The second estimator estimates the source location as the centroid of the detectors which declare a detection. The performance of the estimators is plotted for various values of r_0 in Figure 3. The radius computed from (4) is also shown. We note that for both estimators there is an optimal value for the sensitivity parameter r_0, illustrating the tradeoff between the two sources of information – sensors which report detection and sensors which report miss. We also observe that for a wide range of r_0, (4) provides a good approximation for the MSE error for the MAP estimator.

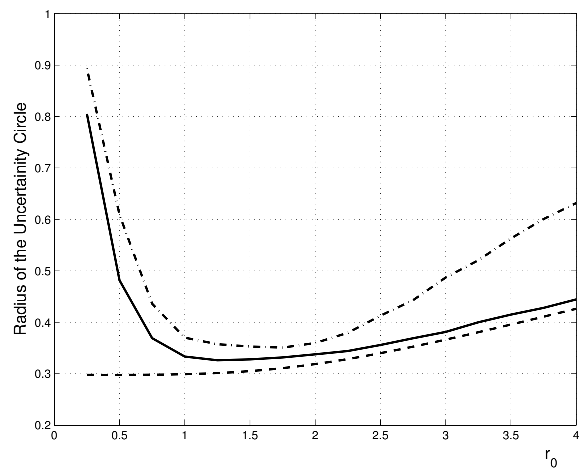

Fig. 3. Average uncertainity radius for the MAP estimator (solid) and the Centroid estimator (dash-dot) and the bound in Eqn.(4) - (dashed)

6. REFERENCES

[1] S. Kumar, D. Shepherd, and F. Zhao, "Collaborative signal and information processing in micro-sensor networks," *IEEE Signal Processing Magazine*, vol. 19, no. 1, 2002.

[2] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, vol. 1, Wiley, New York, 1968.

[3] K. V. Bhagwat and R. Subramanian, "Inequalities between means of positive operators," *Math. Proc. Cambridge Philos. Soc*, vol. 83, pp. 393–401, 1978.

[4] R. R. Tenney and N. R. Sandell, "Detection with distributed sensors," *IEEE Trans. AES*, vol. 17, 98-101 1981.

[5] I. Y. Hoballah and P. K. Varshney, "Distributed bayesian signal detection," *IEEE Trans. Inform. Theory*, vol. 35, pp. 995–1000, 1989.

[6] V. V. Veeravalli, T. Basar, and H. V. Poor, "Minimax robust decentralized detection," *IEEE Trans. Inform. Theory*, vol. 40, pp. 35–40, 1994.