Technical Research Report

SYSTEMS RESEARCH CENTER

Supported by the National Science Foundation Engineering Research Center Program (NSFD CD 8803012), Industry and the University

Distributed Estimation of A Location Parameter in Dependent Noise

by Y.A. Chau and E. Geraniotis

TR 89-78

DISTRIBUTED ESTIMATION OF A LOCATION PARAMETER IN DEPENDENT NOISE

Yawgeng A. Chau and Evaggelos Geraniotis

Department of Electrical Engineering
and Systems Research Center
University of Maryland
College Park, MD 20742

ABSTRACT

We address the problem of distributed estimation from dependent observations involving two sensors ($k=1,2$) that collect observations $X_i^{(k)}$ ($i=1,2,\ldots,N$) of the same nonrandom location parameter θ in additive noise. We consider two cases of interest, the case of independent observations across sensors and the case of correlated observations across sensors. The estimation schemes of the sensors are chosen so as to minimize a common cost function consisting of the weighted sum of the mean square errors of the estimates $\theta_N^{(k)}$ ($k=1,2$) of the two sensors and the mean square of their difference.

The observations of the two sensors are modeled as two m-dependent or ϕ-mixing sequences. The correlation between the two observation sequences is also characterized by an m-dependent or ϕ-mixing sequence. Because high-order statistics of dependent observations are generally difficult to characterize, maximum-likelihood estimates may be impossible to derive or implement; instead, suboptimal estimates, which solve $\min_{\theta_N^{(k)}} \phi_k(X_i^{(k)} - \theta_N^{(k)})$ or, equivalently, $\sum_{i=1}^{N} g_k(X_i^{(k)} - \theta_N^{(k)}) = 0$ on the basis of the memoryless nonlinearities $g_k(\cdot)$, are employed by the sensors. With sensor test statistics, minimizing the above cost function with respect to $\theta_N^{(k)}$ is equivalent to minimizing it with respect to the nonlinearities $g_k(\cdot)$, thus resulting in integral equations. By solving these integral equations, we obtain optimal nonlinearities within this suboptimal framework. Examples for m-dependent Cauchy noise are provided in support of our analysis.

This research was supported in part by the Office of Naval Research under contract N00014-89-J-1375 and in part by the Systems Research Center at the University of Maryland, College Park, through the National Science Foundation's Engineering Research Centers Program: NSF CDR 8803012.

I. INTRODUCTION

Distributed estimation has attracted considerable attention in recent years (see [1]-[4]). However, the dependencies in the observations of each sensor or across sensors have not been taken into consideration. In many practical situations, where the sampling rate increases or the distributed estimators (sensors) are relatively close to each other geographically, it is necessary to consider the dependency of observations.

In this paper, we address the following problem for asymptotic (including of a large sample size N) distributed estimation of a location parameter in dependent noises across sensors: two sensors without communication make observations $X_i^{(k)}$ (with i and k denoting the ith observation and the kth sensor, respectively) of the same nonrandom location parameter θ in additive noise and collectively make their decisions. The forms of the estimation structures are obtained by minimizing the following cost function

$$\begin{aligned} J &= c_1E[(\theta_N^{(1)}-\theta)^2]+c_2E[(\theta_N^{(2)}-\theta)^2]+c_3E[(\theta_N^{(1)}-\theta_N^{(2)})^2] \\ &= (c_1+c_3)E[(\theta_N^{(1)}-\theta)^2]+(c_2+c_3)E[(\theta_N^{(2)}-\theta)^2]+2c_3E[(\theta_N^{(1)}-\theta)(\theta_N^{(2)}-\theta)] \qquad (1) \end{aligned}$$

where c_1, c_2 and c_3 are constants and $\theta_N^{(k)}$ $(k=1,2)$ is the estimate of the location parameter θ at the sensor k with sample size N. This choice of cost function pertains to the consensus of the two sensors. As is well known in centralized (single sensor) estimation problems, the estimate is obtained by optimizing a particular nonlinear function (likelihood function or mean square error) of the N observations. Similarly, we characterize the estimation structures of the sensors by nonlinear functions $\varphi_k(\cdot)$ $(k=1,2)$ with derivatives $g_k(\cdot)$, whose form is to be determined by minimizing the cost function (1). Therefore, for large N and with these particular estimation structures, minimizing (1) with respect to $\theta_N^{(k)}$ is equivalent to performing a minimization of (1) with respect to the nonlinearities $g_k(\cdot)$ $(k=1,2)$.

Two cases are of interest. First we consider the case, in which the two sensors collect two sequences of **dependent** observations but without any correlation across sensors; then we consider the case, in which the two sequences of observations are **correlated** across sensors and time is considered. In both cases, the dependency of noise is characterized by either m-dependent or ϕ-mixing process (see [5]). Since the ϕ-mixing model contains the m-dependent model as a special case, we use, throughout this paper, the ϕ-mixing model for the description of dependence in noise. Let us now state the definition of a $\phi-$mixing sequence. A stationary sequence $\{Y_k\}_{k=1}^{\infty}$ is said to be $\phi-$mixing, if, for $i \geq 1$, $j \geq 1$ and $B_1 \in \mathcal{F}_{i+j}^{\infty}$, there is a real sequence $\{\phi_k\}_{k=0}^{\infty}$ such that

$$\sup_{B_2 \in \mathcal{F}_1^i} |P(B_1 \cap B_2) - P(B_1)P(B_2)| \leq \phi_j P(B_1)$$

and

$$\lim_{k \to \infty} \phi_k = 0$$

where $\mathcal{F}_{i+j}^{\infty}$ is the σ-field generated by $\{Y_k\}_{k=i+j}^{\infty}$ and $\mathcal{F}_1^i$ is the σ-field generated by $\{Y_k\}_{k=1}^{i}$.

Under this $\phi-$mixing model, the optimal estimation structure should be based on statistics involving nth ($n > 2$) order moments (i.e. nonlinearities with memory) which are difficult to characterize. Therefore, suboptimal decision structures based on memoryless nonlinearities are used. This type of structure for processing dependent data has been useful in detection problems (see [6] and [7]). Although this memoryless structure of estimation is not optimal, optimizing the cost function (1) with respect to the nonlinearities renders a scheme with a better performance than the one that does not take into account the dependence across sensors and/or time.

The remainder of this paper is organized as follows: In Section II, we first cite the Central Limit Theorem for ϕ-mixing sequences and then use it to characterize the asymptotic means and variances of the quantities $\theta_N^{(k)} - \theta$ $(k = 1, 2)$; then we express the cost function (1) as a function of (g_1, g_2) for large N. In Section III, we consider the case in which each sensor collects dependent observations but without dependence across sensors. In Section IV, we consider the case with dependent observations across sensors and time; the special case in which the two sequences of observations have the same first-order and second-order statistics is of particular interest. In Section V, we draw the conclusions.

II. PRELIMINARIES

Let the observations of the two sensors be described by

$$X_i^{(k)} = \theta S_i + W_i^{(k)} \tag{2}$$

where $W_i^{(k)}$ ($k = 1, 2$ and $i = 1, \cdots, N$) are two stationary ϕ-mixing processes with univariate densities $f_k(\cdot)$. We may assume, without loss of generality, that $S_1 = S_2 = \cdots = S_N = 1$. As mentioned in Section I, when the sampling rate increases, $\{W_i^{(k)}\}_{i=1}^{\infty}$ and $\{W_{j+i}^{(k)}\}_{i=1}^{\infty}$ become correlated, for $k = 1, 2$ and $j = 1, 2 \cdots, m$. Actually, as $m \to \infty$, the dependence is characterized by a stationary $\phi-$mixing sequence. Furthermore, the estimates of θ are based on the structures with nonlinear functions $\varphi_k(\cdot)$ ($k = 1, 2$) as follows

$$\theta_N^{(k)} = arg \min_{\theta} \sum_{i=1}^{N} \varphi_k(X_i^{(k)} - \theta) \tag{3}$$

where φ_k are nonconstant functions to be determined (for example, $\varphi_k(\cdot) = -ln f_k(\cdot)$ in the well known case, in which the two sensors i.i.d. sequences and the noise sequences $W_i^{(1)}$ and $W_i^{(2)}$ are also independent for all i). In this particular case, the estimator for each sensor is termed by Huber the M-estimator, in his framework of robust estimations (see [8]). Let $g_k = \varphi_k'$ be the derivatives of φ_k (for k=1,2), then $\theta_N^{(k)}$ equivalently satisfies

$$\sum_{i=1}^{N} g_k(X_i^{(k)} - \theta_N^{(k)}) = 0, \ k = 1, 2. \tag{4}$$

Thus, we can derive the forms of g_k ($k = 1, 2$).

For a $\phi-$mixing sequence and the associated nonlinearities g_k we cite the following Central Limit Theorem (see [5]).

Theorem 1: Suppose that, for $k = 1, 2$, $\{Y_j^{(k)}\}_{j=1}^{\infty}$ are stationary $\phi-$mixing sequences with

$$\sum_{j=1}^{\infty} (\phi_j^{(k)})^{\frac{1}{2}} < \infty$$

and that g_k are measurable functions satisfying

$$E[g_k(Y_1^{(k)})] = \mu_k, \ \ var[g_k(Y_1^{(k)})] < \infty$$

Then the series

$$\sigma_k^2(g) = var[g_k(Y_1^{(k)})] + 2\sum_{j=1}^{\infty} cov[g_k(Y_1^{(k)})g_k(Y_{j+1}^{(k)})] \tag{5}$$

converges absolutely. Furthermore, if $\sigma_k^2 > 0$,

$$G_n^{(k)} = \frac{1}{\sqrt{n}}\left[\sum_{j=1}^{n} g_k(Y_j^{(k)}) - n\mu_k\right]$$

converges in distribution to a normal distribution with mean zero and variance σ_k^2; the pair $(G_n^{(1)}, G_n^{(2)})$ are also asymptotically jointly Gaussian distributed.

Throughout this paper we make the following regularity assumptions:

(i) the nonlinearities g_k is continuous and has a uniformly continuous derivative g_k'

(ii) φ_k are continuous convex functions, hence $g_k' > 0$;

(iii) as N increases, $\theta_N^{(k)} \to \theta$

(iv) $\sigma_k^2(g) > 0$.

Moreover, we may assume, without loss of generality, that $\mu_k = E[g_k] = 0$ $(k = 1, 2)$ in the following formulation. For situations in which $\mu_k \neq 0$, we replace $g_k(\cdot)$ by $\bar{g}_k(\cdot) = g_k(\cdot) - \mu_k$ and all results in this paper still hold.

Next we show that the statistic $\sqrt{N}(\theta_N^{(k)} - \theta)$ (for $k = 1, 2$) can be written in an explicit function of g and g'. First, let $T_N^{(k)} = \theta_N^{(k)} - \theta$. Then, from (2) and (5), we have

$$\sum_{i=1}^{N} g(W_i^{(k)} - T_N^{(k)}) = 0.$$

Under the above assumptions and with the help of the mean value theorem, we find a λ with $0 \leq \lambda \leq 1$, such that

$$\sum_{i=1}^{N} g_k(W_i^{(k)}) - T_N^{(k)} \sum_{i=1}^{N} g_k'(W_i^{(k)} - \lambda T_N^{(k)}) = 0$$

for $k = 1, 2$. Therefore,

$$\sqrt{N} T_N^{(k)} = \frac{\sum_{i=1}^{N} g_k(W_i^{(k)})/\sqrt{N}}{\sum_{i=1}^{N} g_k'(W_i^{(k)} - \lambda T_N^{(k)})/N}. \tag{6}$$

The denominator in (6) asymptotically approachs to $E[g_k']$, as N increases (see Lemma 5 of [8]). Thus,

$$\begin{aligned} & E[\sqrt{N}(\theta_N^{(1)} - \theta) \cdot \sqrt{N}(\theta_N^{(2)} - \theta)] = E[\sqrt{N} T_N^{(1)} \cdot \sqrt{N} T_N^{(2)}] \\ \rightarrow \quad & \frac{E\left[\sum_{i=1}^{N} g_1(W_i^{(1)})/\sqrt{N} \cdot \sum_{i=1}^{N} g_2(W_i^{(2)})/\sqrt{N}\right]}{E[g_1']E[g_2']} \end{aligned} \tag{7}$$

As the following result is used in this paper extensively, we write it as another theorem (see [8]).

Theorem 2: Suppose that $\theta_N^{(k)} \rightarrow \theta$ (for $k = 1, 2$), as N increases. Then $\sqrt{N}(\theta_N^{(k)} - \theta)$ (for $k = 1, 2$) is asymptotically normal with asymptotic mean zero and asymptotic variance

$$V_k(g_k) = \frac{\sigma_k^2(g_k)}{\left(\int g_k(w) f_k'(w) dw\right)^2} \tag{8}$$

where $\sigma_k^2(g_k)$ is defined in Theorem 1.

Proof: Because $T_N^{(k)} \rightarrow 0$ and $T_N^{(k)}$ is given by (6), we immediately have the above result from Lemma 5 of [8] and Theorem 1 by noticing that $(E[g_k'])^2 = (-\int g_k(w) f'(w) dw)^2$.

Since $\sqrt{N}\theta$ is the asymptotic mean of $\sqrt{N}\theta_N^{(k)}$ under assumption (iii) as N increases, we have $(\sqrt{N}\theta_N^{(k)} - \sqrt{N}\theta)^2 = V_k$ asymptotically as N increases. Thus, using the form of

$$\sqrt{N}T_N^{(k)} = \sqrt{N}(\theta_N^{(k)} - \theta)$$

given by (6) we can write J given by (1) as

$$\begin{aligned}
J(g_1, g_2) &= \frac{1}{N}\{(c_1 + c_3)E[(\sqrt{N}\theta_N^{(1)} - \sqrt{N}\theta)^2] + (c_2 + c_3)E[(\sqrt{N}\theta_N^{(2)} - \sqrt{N}\theta)^2] \\
&\quad +2c_3 E[(\sqrt{N}\theta_N^{(1)} - \sqrt{N}\theta)(\sqrt{N}\theta_N^{(2)} - \sqrt{N}\theta)]\} \\
&= \{\frac{(c_1 + c_3)\sigma_1^2(g_1)}{N\left(\int g_1(x) f_1'(x)dx\right)^2} + \frac{(c_2 + c_3)\sigma_2^2(g_2)}{N\left(\int g_2(x) f_2'(x)dx\right)^2} \\
&\quad +2c_3 \frac{E\left[\sum_{i=1}^N g_1(W_i^{(1)})/\sqrt{N} \cdot \sum_{i=1}^N g_2(W_i^{(2)})/\sqrt{N}\right]}{N(\int g_1(w) f_1'(w)dw)(\int g_2(w) f_2'(w)dw)}\} \\
&= \frac{1}{N}\left\{(c_1 + c_3)V_1(g_1) + (c_2 + c_3)V_2(g_2) + 2c_3\rho_{12}(g_1, g_2)\sqrt{V_1(g_1)}\sqrt{V_2(g_2)}\right\} \qquad (9)
\end{aligned}$$

where ρ_{12} is the correlation coefficient of $\sum_{i=1}^N \frac{g_1(W_i^{(1)})}{\sqrt{N}}$ and $\sum_{i=1}^N \frac{g_2(W_i^{(2)})}{\sqrt{N}}$, i.e.,

$$\rho_{12}(g_1, g_2) = \frac{E\left[\sum_{i=1}^N g_1(W_i^{(1)})/\sqrt{N} \cdot \sum_{i=1}^N g_2(W_i^{(2)})/\sqrt{N}\right]}{\sqrt{\sigma_1^2(g_1)\sigma_2^2(g_2)}}. \qquad (10)$$

The numerator of the last term in the last expression of (9) asymptotically has, for large N, the following form (see [9])

$$\begin{aligned}
E\left[\sum_{i=1}^N g_1(W_i^{(1)})/\sqrt{N} \cdot \sum_{i=1}^N g_2(W_i^{(2)})/\sqrt{N}\right] &= \frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N E[g_1(W_i^{(1)})g_2(W_j^{(2)})] \\
= \quad E[g_1(W_1^{(1)})g_2(W_1^{(2)})] + 2\sum_{j=1}^m E[g_1(W_1^{(1)})g_2(W_{j+1}^{(2)})] & \qquad (11)
\end{aligned}$$

where $m \to \infty$ for the ϕ-mixing noise case. We notice that (11) is an extension of (5).

III. THE CASE OF DEPENDENCE ACROSS TIME

In this section, we consider the case in which the observations of the two sensors are two dependent sequences with dependency across them only. Then,

$$cov[g_1(W_1^{(1)}), g_2(W_{j+1}^{(2)})] = 0; \quad j = 1, 2, \cdots$$

since the two sequences of observations are uncorrelated across time. In addition, let $c_1 = c_2 = c_3 = 1$. Then the cost function for present situation has the form

$$J(g_1, g_2) = \left\{ \frac{2\sigma_1^2(g_1)}{N\left(\int g_1(x) f_1'(x) dx\right)^2} + \frac{2\sigma_2^2(g_2)}{N\left(\int g_2(x) f_2'(x) dx\right)^2} \right\} \tag{12}$$

where

$$\sigma_k^2(g_k) = \int g_k^2(x) f_k(x) dx + 2 \sum_{j=1}^{m} \int \int g_k(x) g_k(y) f_{W_1^{(k)}, W_{j+1}^{(k)}}(x, y) dx dy \tag{13}$$

with $f_{W_1^{(k)}, W_{j+1}^{(k)}}(x, y)$ being the joint probability density of $W_1^{(k)}$ and $W_{j+1}^{(k)}$ (for $k = 1, 2$) and $m \to \infty$, for ϕ−mixing dependence,

Because the value of the above cost function is invariant under the scaling of $g_k(\cdot)$ (meaning that $J(g_1, g_2) = J(s_1 g_1, s_2 g_2)$ with s_k scaling factors), minimizing it with respect to $g_k(\cdot)$ is equivalent to maximizing the following forms (see [5])

$$H_k(g_k) = -\int g_k(x) f_k'(x) dx + \lambda_L \sigma_k^2(g_k) \tag{14}$$

for all N and $k = 1, 2$, where λ_L is the Lagrange multiplier and $-\int g_k(x) f_k'(x) dx = \int g_k'(x) f_k(x) dx > 0$ for convex function $\varphi(\cdot)$. The minimization of (14) with respect to $g_k(\cdot)$, for each k, has been shown in [5] and the equation to be solved for optimum $g_k(\cdot)$ is (with $\lambda_L = -1/2$)

$$-f_k'(x) - \int_{-\infty}^{\infty} g_k(y) \sum_{j=1}^{m} f_j^{(k)}(x, y) dy = g_k(x) f_k(x) \tag{15}$$

where $m \to \infty$ for the ϕ−mixing noise W_i $(i = 1, 2, \cdots, N)$ and

$$f_j^{(k)}(x, y) = f_{W_1^{(k)}, W_{j+1}^{(k)}}(x, y) + f_{W_1^{(k)}, W_{j+1}^{(k)}}(y, x). \tag{16}$$

Since the search for optimum g_k $(k = 1, 2)$ for this case is decoupled for the two sensors, the cost function is the sum of two subcosts, whose inverses are the efficacies of two asymptotic relative efficiencies (AREs) in [5].

IV. THE CASE OF DEPENDENCE ACROSS TIME AND SENSORS

In situations in which the two sensors are close geographically, it is necessary to consider the dependence of observations across sensors. Thus, the two sequences of observations are no longer uncorrelated across sensors.

In this section, we characterize the dependence of the observations across sensors as $\phi-$mixing dependence. In particular, we consider the situation in which $f_1(x) = f_2(x) = f(x)$ and the second order statistics of the two observation sequences are identical, i.e.,

$$f_{W_1^{(1)},W_{j+1}^{(1)}}(x,y) = f_{W_1^{(2)},W_{j+1}^{(2)}}(x,y) = f_{W_1,W_{j+1}}(x,y). \tag{17}$$

Under this symmetric condition, the optimum nonlinearities satisfy $g_1(\cdot) = g_2(\cdot) = g(\cdot)$.

A. Under the Condition $c_1 = c_2 = c_3 = 1$

The cost function (1) for this case has the form

$$\tilde{J}(g) = \frac{4\sigma^2(g) + 4E[g(W_1^{(1)})g(W_1^{(2)})] + 2\sum_{j=1}^{m} E[g(W_1^{(1)})g(W_{j+1}^{(2)})]}{N(\int g(x)f'(x)dx)^2} \tag{18}$$

with $m \to \infty$, where the form of $\sigma^2(g)$ is given from equation (5) of Section III and

$$E[g(W_i^{(1)})g(W_j^{(2)})] = \int\int g(x)g(y)\tilde{f}_{W_i^{(1)},W_j^{(2)}}(x,y)dxdy \tag{19}$$

with $\tilde{f}_{W_i^{(1)},W_j^{(2)}}$ the joint probability distribution of $W_i^{(1)}$ and $W_j^{(2)}$ for all $i,j = 1,2,\cdots$, due to the stationarity of $\tilde{f}_{W_i^{(1)},W_j^{(2)}}$. Since the above cost function is also invariant under the scaling of $g(\cdot)$, minimizing it with respect to $g(\cdot)$ is equivalent to maximizing

$$\tilde{H}(g) = -\int g(x)f'(x)dx + \lambda_L \left\{ 2\sigma^2(g) + E[g(W_1^{(1)})g(W_1^{(2)})] + 2\sum_{j=1}^{m} E[g(W_1^{(1)})g(W_{j+1}^{(2)})] \right\} \tag{20}$$

with respect to $g(\cdot)$. Minimization similar to the above one has been shown in [9] with $\sigma^2(g)$ replacing $2\sigma^2(g)$. The necessary condition for the maximization of (20) is $\partial\tilde{H}(g+\epsilon\delta g)/\partial|_{\epsilon=0} = 0$, so that

$$\int\{-f'(x)+2\lambda_L\left[2g(x)f(x)+\sum_{j=1}^{m}\int_{-\infty}^{\infty}g(y)(2f_j(x,y)+\tilde{f}_0(x,y)/2+\tilde{f}_j(x,y))dy\right]\}\delta g(x)dx=0 \tag{21}$$

where

$$\tilde{f}_j(x,y)=\tilde{f}_{W_1^{(1)},W_{j+1}^{(2)}}(x,y)+\tilde{f}_{W_1^{(1)},W_{j+1}^{(2)}}(y,x) \tag{22}$$

with $\tilde{f}_{W_1^{(1)},W_{j+1}^{(2)}}$ defined after (16). The sufficient condition for maximization of (20) is $\partial^2\tilde{H}(g+\epsilon\delta g)/\partial^2|_{\epsilon=0}<0$, namely (see [9]),

$$2\lambda_L[2\sigma^2(\delta g)+\rho_{12}(\delta g)\sigma^2(\delta g)]<0 \tag{23}$$

where $\rho_{12}(\delta g)=\rho_{12}(\delta g,\delta g)$. Therefore, (21) with negative λ_L is the necessary and sufficient condition for optimum $g(\cdot)$ to minimize (20). We notice that λ_L in (20) is a scaling factor of $g(\cdot)$. Furthermore, $g(x)=-f'(x)/f(x)$ (using a likelihood function $\varphi(x)=-\ln f(x)$) for i.i.d. observations across time and sensors. Hence we may set $\lambda_L=-1/4$ to be consistent with the conventional condition for $g(x)=-f'(x)/f(x)$ in the i.i.d. case. Thus, the final form of the integral equation satisfied by the optimum $g(\cdot)$ is

$$-f'(x)-\int_{-\infty}^{\infty}\tilde{K}(x,y)g(y)=g(x)f(x) \tag{24}$$

where

$$\tilde{K}(x,y)=\sum_{j=1}^{m}[2f_{W_1,W_{j+1}}(x,y)+\tilde{f}_{W_1^{(1)},W_{j+1}^{(2)}}(x,y)]+\tilde{f}_{W_1^{(1)},W_1^{(2)}}(x,y)/2. \tag{25}$$

B. Under the Conditions $c_1=c_2=0$ and $c_3=1$

With $c_1=c_2=0$ and $c_3=1$, the cost function has the form

$$\tilde{J}(g)=E[(\theta_N^{(1)}-\theta_N^{(2)})^2]=\frac{2\sigma^2(g)+2E[g(W_1^{(1)})g(W_1^{(2)})]+4\sum_{j=1}^{m}E[g(W_1^{(1)})g(W_{j+1}^{(2)})]}{N\left(\int g(x)f'(x)dx\right)^2}. \tag{26}$$

The minimization of above cost function with respect to g can be obtained in a way similar to the one in the previous section. We will not repeat the formulation procedure here. The final form of the integral equation satisfied by the optimum $g(\cdot)$ is

$$-f'(x) - \int_{-\infty}^{\infty} \bar{K}(x,y)g(y) = g(x)f(x) \tag{27}$$

where

$$\bar{K}(x,y) = 2\sum_{j=1}^{m}[f_{W_1,W_{j+1}}(x,y) + \tilde{f}_{W_1^{(1)},W_{j+1}^{(2)}}(x,y)] + \tilde{f}_{W_1^{(1)},W_1^{(2)}}(x,y). \tag{28}$$

C. Examples

To illustrate the performance of the above two-sensor schemes, we consider a noise process with a Cauchy probability distribution. This Cauchy noise has a first-order probability density given by

$$f(x) = \frac{1}{\pi[1+(x-\gamma)^2]}$$

where $-\infty < \gamma < \infty$ is the median of the Cauchy noise.

The first-order and second-order probability densities of the Cauchy noise can be obtained from a nonlinear transformation of a Gaussian process, whose first-order density is given by (see [5])

$$f_G(x) = \frac{\exp[-(x-\gamma)^2/2]}{\sqrt{2\pi}}$$

with autocorrelation coefficients $E[X_1^{(k)}X_{j+1}^{(k)}] = \rho_j$ (for $j = 0, 1, \cdots$ and $k = 1, 2$) across time and $E[X_1^{(1)}X_{j+1}^{(2)}] = \rho_c\rho_j$ across sensors, where $m \to \infty$ for ϕ-mixing noise and m is finite for m-dependent noise. Note that, for the dependent Gaussian noise given above, the optimal nonlinearities for dependent noise (across time and/or sensors) are equal to the ones obtained for independent noise (across time and/or sensors) except for a scaling factor (see [5]). Thus, the corresponding optimal cost functions given by (1) are equal. Since the procedure of obtaining the nonlinear transformation is described in [5], we shall not repeat it here. In the following examples, we set

$$\rho_j = \begin{cases} 1 - |j|/(m+1) & \text{if } |j| < m \\ 0 & \text{if } |j| \geq m+1 \end{cases}$$

and $\rho_c = 0.7$; we evaluate the cost functions for different values of m (i.e. for the m-dependent model). Note that $E[X_1^{(k)} X_{j+1}^{(k)}] = 0, |j| \geq m+1$ for a Gaussian process implies that $X_1^{(k)}$ and $X_{j+1}^{(k)}$ are independent of each other, as a result of which the induced Cauchy random variables are also independent.

Example 1: This example illustrates the case with $c_1 = c_2 = c_3 = 1$ and $\mu(g) = 0$. The left part (Columns 2 and 3) of Table 1 gives the comparison of the cost functions with different nonlinearities. The second column of this table represents the ratio between the cost J_{opt} with the optimal $g(\cdot)$ and the cost J_{ias} with the optimal nonlinearities obtained after ignoring the dependence across sensors for different m. The third column of this table represents the ratio of the cost J_{opt} and the cost J_{iid} with the optimal nonlinearities (i.e., $g_{lo}(x) = -f'(x)/f(x)$) that are obtained by ignoring the dependences across time and sensors.

m	$c_1 = c_2 = c_3 = 1$		$c_1 = c_2 = 0, c_3 = 1$	
	J_{ias}/J_{opt}	J_{iid}/J_{opt}	J_{ias}/J_{opt}	J_{iid}/J_{opt}
1	1.030	1.157	1.074	1.252
2	1.040	1.305	1.090	1.437
3	1.045	1.415	1.099	1.570
4	1.048	1.494	1.104	1.667
5	1.050	1.552	1.108	1.737
10	1.054	1.690	1.114	1.902
15	1.055	1.736	1.115	1.957
50	1.056	1.783	1.117	2.011
100	1.056	1.788	1.117	2.016

Table 1: two cases for different values of c_1, c_2 and c_3

Example 2: This example illustrates the case with $c_1 = c_2 = 0$, $c_3 = 1$ and $\mu(g) = 0$. The right part (Columns 4 and 5) of Table 1 gives the same type of comparison as the one in the left part of Table 1 for this case.

V. CONCLUSIONS AND EXTENSIONS

Using the cost function defined by (1) we derived an asymptotic scheme with large sample size N for the distributed (two-sensor) estimation of a parameter θ in dependent noise described by ϕ-mixing or m-dependent sequences. The estimation structures of the two sensors are characterized by nonlinear functions of the observations $\varphi_k(\cdot)$ $(k = 1,2)$. When the sensors employ the nonlinearities $g_k(\cdot) = \varphi'_k(\cdot)$, we have shown that, as $N \to \infty$, minimizing the cost with respect to the estimates $\theta_N^{(k)}(k = 1,2)$ of the parameter θ is equivalent to minimizing it with respect to the nonlinearities $g_k(\cdot)$. The optimum $g_k(\cdot)$ are obtained by solving linear integral equations.

This scheme uses estimation structures similar to the M-estimators of [8] which are proposed in the context of robustness. The mean of the nonlinearities need not to be zero; thus the univariate probability density of the noise need not to be symmetric. Examples with dependent Cauchy noise are provided to illustrate the analysis. Table 1 shows that the scheme proposed in this paper has better performance than the one obtained by ignoring the dependence across time and/or sensors.

Although the analytical results in the above sections are derived for a constant parameter (i.e., $S_1 = S_2 = \cdots = S_N$), the scheme proposed here can be used for a nonrandom parameter with time-varying value under the assumption

$$C = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} S_i^2 < \infty.$$

In this case, we only need replace $g_k(X_i - \theta_N^{(k)})$ with $g_k(X_i - \theta_N^{(k)} S_i)$ in the analysis (for $i = 1, 2, \cdots, N$ and $k = 1,2$). Then the asymptotic variances in Theorem 2 take the form (see [10])

$$V_k(g_k) = \frac{\sigma_k^2(g_k)}{C\left(\int g_k(w) f'_k(w) dw\right)^2}.$$

Therefore, the optimum nonlinearities will satisfy integral equations similar to those derived in

Sections III and IV. Finally, to evaluate the optimum nonlinearities we need complete knowledge of the first- and the second-order statistics of the sensor observations, which may be difficult to acquire in reality. Therefore, the robustification of our schemes to uncertainties in the first- and second-order statistics is of interest and it will be the subject of further work in this area.

References

[1] C. Y. Chong. "Hierarchical Estimation," *Proceedings of the 2nd MIT/ONR Workshop on Distributed Communication and Decision Problems*, July 1979.

[2] A. S. Willsky, M. Bello, D. A. Castanon, B. C. Levy and G. Verghese. "Combining and Updating of Local estimates and Regional Maps along Sets of One-Dimensional Tracks." *IEEE Transactions on Automatic Control*, Vol. AC-27, pp. 799-813, Aug. 1982.

[3] D. A. Castanon and D. Teneketzis. "Distributed Estimation Algorithms for Nonlinear Systems." *IEEE Transactions Automatic Control*, Vol. AC-30, pp. 418-425, May 1985.

[4] D. A. Castanon and D. Teneketzis. "Further Results on the Asymptotic Agreement Problem." *IEEE Transactions on Automatic Control*, Vol. AC-33, pp. 515-523, June 1988.

[5] H. V. Poor and J. B. Thomas. "Memoryless Discrete-Time Detection of a Constant Signal in m-Dependent Noise. *IEEE Transactions Information Theory,* Vol. IT-25, pp. 54-61, Jan. 1979.

[6] D. R. Halverson and G. L. Wise. "Discrete-Time Detection in $\phi-$mixing Noise." *IEEE Transactions on Information Theory*, Vol. IT-26, pp. 189-198, 1980.

[7] P. Billingsley. *Convergence of Probability Measures.* New York: Wiley, 1968.

[8] P. J. Huber. "Robust Estimation of a Location Parameter." *Annals of Mathematical Statist.* Vol. 35 pp. 73-101, 1964.

[9] E. Geraniotis and Y. A. Chau. "Distributed Detection of Weak Signals from Multiple Sensors with Correlated Observations." *Proceedings of the 27th IEEE Conference on Decision and Control*, Austin, TX, Dec. 7-9, 1988.

[10] A. H. El-Sawy and V. D. VandeLinde. "Robust Detection of Known Signals." *IEEE Transactions on Information Theory,* Vol. IT-23, pp.722-727, Nov. 1977.