AFRL-RY-WP-TR-2011-1086

INTEGRATED MULTI-MODAL RF SENSING

Sensor Integration and Joint PDF Construction for Distributed Detection and Classification

Muralidhar Rangaswamy

RF Space Exploitation Technology Branch
Layered Sensing Exploitation Division

Steven Kay and Quan Ding

University of Rhode Island

JANUARY 2011
Interim Report

AIR FORCE RESEARCH LABORATORY
SENSORS DIRECTORATE
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7320
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

NOTICE AND SIGNATURE PAGE

AFRL-RY-WP-TR-2011-1086 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

*//Signature//

BRADLEY MORTENSON
Section Chief

//Signature//

JEFFREY SANDERS
Branch Chief

*Disseminated copies will show “//Signature//” stamped or typed above the signature blocks.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
January 2011	Interim	26 January 2009 – 26 January 2011

4. TITLE AND SUBTITLE
Integrated Multi-Modal RF Sensing: Sensor Integration and Joint PDF Construction for Distributed Detection and Classification

5a. CONTRACT NUMBER
In-house

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Muralidhar Rangaswamy (AFRL/RYAP)
Steven Kay and Quan Ding (University of Rhode Island)

5d. PROJECT NUMBER
2305

5e. TASK NUMBER
HE

5f. WORK UNIT NUMBER
2305HE00

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

RF Space Exploitation Technology Branch
Layered Sensing Exploitation Division
Air Force Research Laboratory, Sensors Directorate
Wright-Patterson Air Force Base, OH 45433-7320
Air Force Materiel Command, United States Air Force

University of Rhode Island
4 East Alumni Ave.
Kingston, RI 02881

8. PERFORMING ORGANIZATION REPORT NUMBER
AFRL-RY-WP-TR-2011-1086

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory
Sensors Directorate
Wright-Patterson Air Force Base, OH 45433-7320
Air Force Materiel Command
United States Air Force

10. SPONSORING/MONITORING AGENCY ACRONYM(S)
AFRL/RYAP

11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S)
AFRL-RY-WP-TR-2011-1086

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES
PAO case number 88ABW-11-0386, cleared 28 Jan 2011. Report contains color.

14. ABSTRACT
With multiple sensors in distributed systems, one is expected to make better decisions than with a single sensor. We investigate the problem of sensor integration to combine all the available information. In this paper, we propose a novel method of constructing the joint probability density function (PDF) based on the exponential family. This method does not require the knowledge of the marginal PDFs and hence is useful in many practical cases. We prove that our method is asymptotically optimal in Kullback-Leibler (KL) divergence. It is shown that the performance of our method is the same as existing methods, while it requires less information.

15. SUBJECT TERMS
Distributed detection and classification, exponential family, joint probability density function, Kullback-Leibler divergence, sensor integration

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT:	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON (Monitor)
Unclassified	Unclassified	Unclassified	SAR	28	Muralidhar Rangaswamy
					19b. TELEPHONE NUMBER *(Include Area Code)* N/A

Sensor Integration and Joint PDF Construction for Distributed Detection and Classification

Steven Kay, *Fellow, IEEE,* Quan Ding, *Student Member, IEEE,* and Muralidhar Rangaswamy, *Fellow, IEEE*

Abstract

With multiple sensors in distributed systems, one is expected to make better decisions than with a single sensor. We investigate the problem of sensor integration to combine all the available information. In this paper, we propose a novel method of constructing the joint probability density function (PDF) based on the exponential family. This method does not require the knowledge of the marginal PDFs and hence is useful in many practical cases. We prove that our method is asymptotically optimal in Kullback-Leibler (KL) divergence. It is shown that the performance of our method is the same as existing methods, while it requires less information.

Index Terms

Distributed detection and classification, exponential family, joint probability density function, Kullback-Leibler divergence, sensor integration

I. Introduction

Distributed systems and information fusion have been widely studied and used in engineering, finance, and scientific research. Such applications are radar, sonar, biomedical analysis, stock prediction, weather forecast, and chemical, biological, radiological, and nuclear (CBRN) detection, to name a few. If the joint probability density functions (PDFs) under each candidate hypothesis are known, we would easily

This work was supported by the Sensors Directorate of the Air Force Research Laboratory (AFRL/RYRT) under contract number FA8650-08-D-1303 to Dynetics, Inc.

Steven Kay and Quan Ding are with the Department of Electrical, Computer, and Biomedical Engineering, University of Rhode Island, Kingston, RI 02881, USA. E-mail: kay@ele.uri.edu, dingqqq@ele.uri.edu.

Muralidhar Rangaswamy is with the Sensors Directorate of the Air Force Research Laboratory (AFRL/RYRT), WPAFB, OH 45433, USA. E-mail: Muralidhar.Rangaswamy@wpafb.af.mil.

obtain the optimal performance by the Neyman-Pearson rule for detection (binary hypothesis testing) and by the maximum a posteriori probability (MAP) rule for classification (multiple hypothesis testing) [1]. However in practice, this information may not be available. This usually happens when the dimensionality of the sample space is high and we do not have enough training samples to have an accurate estimate of the joint PDF. This is also recognized as "curse of dimensionality" in pattern recognition and machine learning. Hence it is important to construct an appropriate joint PDF when it is not available. One common approach is to assume that the measurements from different sensors are independent [2], [3]. This approach has been widely used due to its simplicity, since the joint PDF is then the product of the marginal PDFs. This is also known as the "product rule" in combining classifiers [4]. In spite of its popularity, the independence assumption may not be a good one if the measurements are actually correlated. Furthermore, as stated in [4], the product rule is severe because "it is sufficient for a single recognition engine to inhibit a particular interpretation by outputting a close to zero probability for it". Hence people have studied other methods that consider the correlation among the measurements. A copula based frame work is proposed in [5], [6] to construct the joint PDF. The exponentially embedded families (EEFs) are used in [7] to estimate the joint PDF that is asymptotically closest to the true one in Kullback-Leibler (KL) divergence.

Note that the above methods all require the knowledge of marginal PDFs. In this paper, we consider the case when the marginal PDFs are not available or accurate, which can happen due to high dimensional sample space and insufficient training data. We present a new way of constructing the joint PDF without the knowledge of marginal PDFs but only a reference PDF. The constructed joint PDF takes the form of the exponential family and it incorporates all the available information. The maximum likelihood estimate (MLE) [8] of the unknown parameters can be easily solved based on the properties of the exponential family. It is shown that the constructed PDF is asymptotically the optimal one in the sense that it is asymptotically closest to the true PDF in KL divergence. Since there is no Gaussian distribution assumption on the reference PDF, this method can be very useful when the underlying distributions are non-Gaussian. We start with the detection problem, and then extend our method to the classification problem. For detection, it is shown that under some conditions, our detection statistics are the same as the the clairvoyant generalized likelihood ratio test (GLRT). For classification, our classifier also has the same performance as the estimated MAP classifier. Both the clairvoyant GLRT and the estimated MAP classifier assume that the true PDFs under each candidate hypothesis are known except for the usual unknown parameters.

The paper is organized as follows. In Section II, we introduce a distributed detection/classification problem. In Section III, we construct the joint PDF by an exponential family and apply it to the problem in Section II. The KL divergence between the true PDF and the constructed PDF is examined in Section IV, and the result shows that the constructed PDF is asymptotically optimal. Examples for distributed detection are given in Section V, and examples for distributed classification are given in VI. Simulation results to compare the performance of our method with existing methods are shown in Section VII. In Section VIII, we draw the conclusions.

II. Problem Statement

Consider the distributed detection/classification problem when we observe the outputs of two sensors, $\mathbf{T}_1(\mathbf{x})$ and $\mathbf{T}_2(\mathbf{x})$ which are transformations of the underlying samples $\mathbf{x}$ that are unobservable. We choose two sensors for simplicity. All the results in this paper are valid for multiple sensors. For detection, we want to distinguish between two hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ based on the outputs of the two sensors, and for classification, we have M candidate hypotheses $\mathcal{H}_i$ for $i = 1, 2, \ldots, M$.

Assume that we have enough training data $\mathbf{T}_{1_i}(\mathbf{x})$'s and $\mathbf{T}_{2_i}(\mathbf{x})$'s under $\mathcal{H}_0$ when there is no signal present. Hence we have a good estimate of the joint PDF of $\mathbf{T}_1$ and $\mathbf{T}_2$ under $\mathcal{H}_0$ [9], and thus we assume $p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1, \mathbf{t}_2; \mathcal{H}_0)$ is completely known. Under $\mathcal{H}_1$ or $\mathcal{H}_i$ for $i = 1, 2, \ldots, M$ when a signal is present, we may not even have enough training data to estimate the marginal PDFs. This is especially the case in the radar scenario, where the target is present for only a small portion of the time. So our goal is to use as much information as we have to construct an appropriate $p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1, \mathbf{t}_2; \mathcal{H}_1)$ under $\mathcal{H}_1$ for detection or $p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1, \mathbf{t}_2; \mathcal{H}_i)$ under each $\mathcal{H}_i$ for classification. A simple illustration is shown in Figure 1.

III. Joint PDF Construction by Exponential Family and Its Application in Distributed Systems

To start with, we consider the detection problem, where we wish to construct $p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1, \mathbf{t}_2; \mathcal{H}_1)$. The result will then be extended to the classification problem.

Since $p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1, \mathbf{t}_2; \mathcal{H}_1)$ cannot be uniquely specified based on $p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1, \mathbf{t}_2; \mathcal{H}_0)$, we need the following reasonable assumptions to construct the joint PDF.

1) Under $\mathcal{H}_1$ the signal is small and thus $p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1, \mathbf{t}_2; \mathcal{H}_1)$ is close to $p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1, \mathbf{t}_2; \mathcal{H}_0)$.

2) $p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1, \mathbf{t}_2; \mathcal{H}_1)$ can be parameterized by some signal parameters $\boldsymbol{\theta}$ such that

$$p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1, \mathbf{t}_2; \mathcal{H}_1) = p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1, \mathbf{t}_2; \boldsymbol{\theta})$$

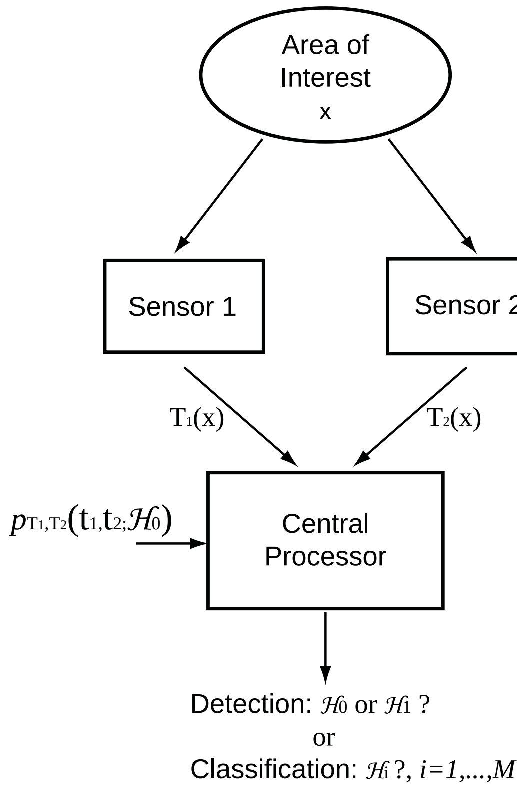

Fig. 1. Distributed detection/classification system with two sensors

$$p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1,\mathbf{t}_2;\mathcal{H}_0) = p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1,\mathbf{t}_2;\mathbf{0})$$

Note that since $\boldsymbol{\theta}$ represents signal amplitudes, $\boldsymbol{\theta} \neq \mathbf{0}$ under $\mathcal{H}_1$. Therefore, the detection problem is to select between

$$\begin{aligned} \mathcal{H}_0 : &\quad \boldsymbol{\theta} = \mathbf{0} \\ \mathcal{H}_1 : &\quad \boldsymbol{\theta} \neq \mathbf{0} \end{aligned}$$

To simplify the notation, let

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_1 \\ \mathbf{T}_2 \end{bmatrix}$$

so that we can write the joint PDF $p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1,\mathbf{t}_2;\boldsymbol{\theta})$ as $p_{\mathbf{T}}(\mathbf{t};\boldsymbol{\theta})$. With the small signal assumptions, it has been shown in [10] that by using a first order Taylor expansion on the log-likelihood function $\ln p_{\mathbf{T}}(\mathbf{t};\boldsymbol{\theta})$, we can construct the PDF of $\mathbf{T}$ as

$$p_{\mathbf{T}}(\mathbf{t};\boldsymbol{\theta}) = \exp\left[\boldsymbol{\theta}^T\mathbf{t} - K(\boldsymbol{\theta}) + \ln p_{\mathbf{T}}(\mathbf{t};\mathbf{0})\right] \tag{1}$$

where

$$K(\boldsymbol{\theta}) = \ln E_0 \left[\exp\left(\boldsymbol{\theta}^T \mathbf{T}\right)\right] \tag{2}$$

is the cumulant generating function of $p_{\mathbf{T}}(\mathbf{t};\mathbf{0})$, and it normalizes the PDF to integrate to 1. Since $\mathbf{T}$ is a sufficient statistic for the constructed exponential PDF in (1), this PDF incorporates all the information from sensors. Note that only $p_{\mathbf{T}}(\mathbf{t};\mathbf{0})$ is required in (1) to construct $p_{\mathbf{T}}(\mathbf{t};\boldsymbol{\theta})$, and is assumed that $p_{\mathbf{T}}(\mathbf{t};\mathbf{0})$ is available or it can be estimated with reasonable accuracy. Also note that if $\mathbf{T}_1$, $\mathbf{T}_2$ are statistically dependent under $\mathcal{H}_0$, they will also be dependent under $\mathcal{H}_1$.

The next step is to estimate the unknown parameters $\boldsymbol{\theta}$. We resort to the MLE [8] by maximizing (1) over $\boldsymbol{\theta}$. Note that $K(\boldsymbol{\theta})$ is convex by Holder's inequality [11]. Since maximizing (1) is equivalent to maximizing $\boldsymbol{\theta}^T\mathbf{t} - K(\boldsymbol{\theta})$, this becomes a convex optimization problem and many existing methods can be readily utilized [12], [13]. Also, the MLE of $\boldsymbol{\theta}$ will satisfy

$$\mathbf{t} = \frac{\partial K(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{3}$$

When the MLE $\hat{\boldsymbol{\theta}}$ is found, we will use $p_{\mathbf{T}}(\mathbf{t};\hat{\boldsymbol{\theta}})$ as our estimated PDF under $\mathcal{H}_1$. Hence similar to the GLRT [1], we will decide $\mathcal{H}_1$ if

$$\ln \frac{p_{\mathbf{T}}(\mathbf{t};\hat{\boldsymbol{\theta}})}{p_{\mathbf{T}}(\mathbf{t};\mathbf{0})} = \hat{\boldsymbol{\theta}}^T\mathbf{t} - K(\hat{\boldsymbol{\theta}}) > \tau \tag{4}$$

where τ is a threshold. We will show in the next section that $p_{\mathbf{T}}(\mathbf{t};\hat{\boldsymbol{\theta}})$ is asymptotically the optimal in the sense of KL divergence.

To extend our method to classification, the above two assumptions can be simply modified as

1) The signal is small under each $\mathcal{H}_i$ and hence $p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1,\mathbf{t}_2;\mathcal{H}_i)$ is close to $p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1,\mathbf{t}_2;\mathcal{H}_0)$.

2) Under each $\mathcal{H}_i$, the joint PDF can be parameterized by some signal parameters $\boldsymbol{\theta}_i$ so that

$$p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1,\mathbf{t}_2;\mathcal{H}_i) = p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1,\mathbf{t}_2;\boldsymbol{\theta}_i)$$

$$p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1,\mathbf{t}_2;\mathcal{H}_0) = p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1,\mathbf{t}_2;\mathbf{0})$$

Similar to (1), as shown in [14], we can construct the PDF of $\mathbf{T}$ under $\mathcal{H}_i$ as

$$p_{\mathbf{T}}(\mathbf{t};\boldsymbol{\theta}_i) = \exp\left[\boldsymbol{\theta}_i^T\mathbf{t} - K(\boldsymbol{\theta}_i) + \ln p_{\mathbf{T}}(\mathbf{t};\mathbf{0})\right] \tag{5}$$

where

$$K(\boldsymbol{\theta}_i) = \ln E_0 \left[\exp\left(\boldsymbol{\theta}_i^T \mathbf{T}\right)\right] \tag{6}$$

is the cumulant generating function of $p_{\mathbf{T}}(\mathbf{t};\mathbf{0})$ that normailzes the constructed PDF. When the MLE of $\boldsymbol{\theta}_i$ is found by maximizing $p_{\mathbf{T}}(\mathbf{t};\boldsymbol{\theta}_i)$ over $\boldsymbol{\theta}_i$, we consider $p_{\mathbf{T}}(\mathbf{t};\hat{\boldsymbol{\theta}}_i)$ as our estimate of $p_{\mathbf{T}}(\mathbf{t};\mathcal{H}_i)$ where

$\hat{\boldsymbol{\theta}}_i$ is the MLE of $\boldsymbol{\theta}_i$. Hence similar to the MAP rule [1], we will decide $\mathcal{H}_i$ for which the following is maximum over i:

$$p(\mathcal{H}_i|\mathbf{t}) = \frac{p_{\mathbf{T}}(\mathbf{t};\mathcal{H}_i)p(\mathcal{H}_i)}{p_{\mathbf{T}}(\mathbf{t})} = \frac{p_{\mathbf{T}}(\mathbf{t};\hat{\boldsymbol{\theta}}_i)p(\mathcal{H}_i)}{p_{\mathbf{T}}(\mathbf{t})} \tag{7}$$

When we assume that the prior probabilities of each candidate hypothesis, i.e., $p(\mathcal{H}_1) = \ldots, = p(\mathcal{H}_M) = 1/M$, $p(\mathcal{H}_i)$ cancels and we can equivalently decide $\mathcal{H}_i$ for which the following is maximum over i:

$$\ln\frac{p_{\mathbf{T}}(\mathbf{t};\hat{\boldsymbol{\theta}}_i)}{p_{\mathbf{T}}(\mathbf{t};\mathbf{0})} = \hat{\boldsymbol{\theta}}_i^T\mathbf{t} - K(\hat{\boldsymbol{\theta}}_i) \tag{8}$$

IV. KL Divergence Between The True PDF and The Constructed PDF

The KL divergence is a non-symmetric measure of difference between two PDFs. For two PDFs p_1 and p_0, it is defined as

$$D\left(p_1 \,\|p_0\,\right) = \int p_1(\mathbf{x})\ln\frac{p_1(\mathbf{x})}{p_0(\mathbf{x})}d\mathbf{x}$$

It is well known that the $D\left(p_1 \,\|p_0\,\right) \geq 0$ with equality if and only if $p_1 = p_0$ almost everywhere [15]. By Stein's lemma [16], the KL divergence measures the asymptotic performance for detection. An extended result to classification is recently presented in [17]. Next we will show that $p_{\mathbf{T}}(\mathbf{t};\hat{\boldsymbol{\theta}})$ is the optimal under both hypotheses. That is, if it is under $\mathcal{H}_0$, $p_{\mathbf{T}}(\mathbf{t};\hat{\boldsymbol{\theta}}) = p_{\mathbf{T}}(\mathbf{t};\mathbf{0})$ asymptotically, and if it is under $\mathcal{H}_1$, $p_{\mathbf{T}}(\mathbf{t};\hat{\boldsymbol{\theta}})$ is asymptotically the closest to the true PDF in KL divergence. Similar results and arguments have been shown in [7], [18].

Assume that we observe independent and identically distributed (IID) $\mathbf{T}_i$'s with

$$\mathbf{T}_i = \begin{bmatrix} \mathbf{T}_{1_i} \\ \mathbf{T}_{2_i} \end{bmatrix}$$

for $i = 1, 2, \ldots, M$. Without abuse of notation, we will write $p_{\mathbf{T}_1,\mathbf{T}_2,\ldots,\mathbf{T}_M}(\mathbf{t}_1,\mathbf{t}_2,\ldots,\mathbf{t}_M;\boldsymbol{\theta})$ as $p(\mathbf{t}_1,\mathbf{t}_2,\ldots,\mathbf{t}_M;\boldsymbol{\theta})$. The constructed PDF can be easily extended as

$$\begin{aligned} &p(\mathbf{t}_1,\mathbf{t}_2,\ldots,\mathbf{t}_M;\boldsymbol{\theta}) \\ &= \exp\left[\boldsymbol{\theta}^T\textstyle\sum_{i=1}^M\mathbf{t}_i - MK(\boldsymbol{\theta}) + \ln p(\mathbf{t}_1,\mathbf{t}_2,\ldots,\mathbf{t}_M;\mathbf{0})\right] \end{aligned} \tag{9}$$

So we want to maximize

$$\frac{1}{M}\ln\frac{p(\mathbf{t}_1,\mathbf{t}_2,\ldots,\mathbf{t}_M;\boldsymbol{\theta})}{p(\mathbf{t}_1,\mathbf{t}_2,\ldots,\mathbf{t}_M;\mathbf{0})} = \frac{1}{M}\boldsymbol{\theta}^T\sum_{i=1}^M\mathbf{t}_i - K(\boldsymbol{\theta}) \tag{10}$$

and $\hat{\boldsymbol{\theta}}$ is found by solving

$$\frac{1}{M}\sum_{i=1}^M\mathbf{t}_i = \frac{\partial K(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \tag{11}$$

Now we consider two cases. First, if the true PDF is under $\mathcal{H}_0$, then by the law of large numbers,

$$\frac{1}{M}\sum_{i=1}^{M}\mathbf{t}_i \to E_0(\mathbf{t})$$

as $M \to \infty$. Note that

$$\frac{\partial K(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\mathbf{0}} = E_0(\mathbf{t})$$

Since the solution of (11) is unique, asymptotically we have

$$\hat{\boldsymbol{\theta}} = \mathbf{0}$$

and hence $p(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_M; \hat{\boldsymbol{\theta}}) = p(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_M; \mathbf{0})$.

Secondly, if the true PDF is under $\mathcal{H}_1$, then by the law of large numbers,

$$\frac{1}{M}\sum_{i=1}^{M}\mathbf{t}_i \to E_1(\mathbf{t})$$

as $M \to \infty$. From (10), we are asymptotically maximizing

$$\boldsymbol{\theta}^T E_1(\mathbf{t}) - K(\boldsymbol{\theta}) \tag{12}$$

To avoid confusion, we will denote the underlying true PDF under $\mathcal{H}_1$ as $p(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_M; \mathcal{H}_1)$ and our constructed PDF as $p(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_M; \boldsymbol{\theta})$. Since

$$\begin{aligned}&\ln\frac{p(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_M; \mathcal{H}_1)}{p(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_M; \boldsymbol{\theta})}\\ &= -\left(\boldsymbol{\theta}^T\sum_{i=1}^{M}\mathbf{t}_i - MK(\boldsymbol{\theta})\right) + \ln\frac{p(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_M; \mathcal{H}_1)}{p(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_M; \mathbf{0})}\end{aligned}$$

the KL divergence between the true PDF and the constructed one is

$$\begin{aligned}&D\left(p(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_M; \mathcal{H}_1)||p(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_M; \boldsymbol{\theta})\right)\\ &= E_1\left[-\left(\boldsymbol{\theta}^T\sum_{i=1}^{M}\mathbf{t}_i - MK(\boldsymbol{\theta})\right) + \ln\frac{p(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_M; \mathcal{H}_1)}{p(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_M; \mathbf{0})}\right]\\ &= -M\left[\boldsymbol{\theta}^T E_1(\mathbf{t}) - K(\boldsymbol{\theta})\right]\\ &\quad + D\left(p(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_M; \mathcal{H}_1)||p(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_M; \mathbf{0})\right)\end{aligned} \tag{13}$$

Since $D\left(p(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_M; \mathcal{H}_1)||p(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_M; \mathbf{0})\right)$ is fixed, $D\left(p(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_M; \mathcal{H}_1)||p(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_M; \boldsymbol{\theta})\right)$ is minimized by maximizing (12). This shows that $p(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_M; \hat{\boldsymbol{\theta}})$ is asymptotically the closest to $p(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_M; \mathcal{H}_1)$ in KL divergence.

V. EXAMPLES-DISTRIBUTED DETECTION

In this section, we compare our method with the clairvoyant GLRT for some detection problems. The clairvoyant GLRT assumes that we know the true PDF of $\mathbf{T}$ under $\mathcal{H}_1$ except for the underlying unknown parameters $\boldsymbol{\alpha}$, and it decides $\mathcal{H}_1$ if

$$\ln \frac{p_{\mathbf{T}}(\mathbf{t};\hat{\boldsymbol{\alpha}})}{p_{\mathbf{T}}(\mathbf{t};\mathbf{0})} > \tau \tag{14}$$

A. Partially Observed Linear Model with Gaussian Noise

Suppose we have the linear model with

$$\mathbf{x} = \mathbf{H}\boldsymbol{\alpha} + \mathbf{w} \tag{15}$$

with

$$\begin{aligned} \mathcal{H}_0 : &\quad \boldsymbol{\alpha} = \mathbf{0} \\ \mathcal{H}_1 : &\quad \boldsymbol{\alpha} \neq \mathbf{0} \end{aligned}$$

where $\mathbf{x}$ is an $N \times 1$ vector of the underlying unobservable samples, $\mathbf{H}$ is an $N \times p$ observation matrix with full column rank, $\boldsymbol{\alpha}$ is an $p \times 1$ vector of the unknown signal amplitudes, and $\mathbf{w}$ is an $N \times 1$ vector of white Gaussian noise with known variance σ^2. We observe two sensor outputs

$$\begin{aligned} \mathbf{T}_1(\mathbf{x}) &= \mathbf{H}_1^T \mathbf{x} \\ \mathbf{T}_2(\mathbf{x}) &= \mathbf{H}_2^T \mathbf{x} \end{aligned} \tag{16}$$

where $\mathbf{H}_1$ is $N \times p_1$ and $\mathbf{H}_2$ is $N \times p_2$. Note that $[\mathbf{H}_1, \mathbf{H}_2]$ does not have to be $\mathbf{H}$. This model is called a partially observed linear model.

Let $\mathbf{G} = [\mathbf{H}_1, \mathbf{H}_2]$. We assume that $\mathbf{G}$ has full column rank so that there is no redundant measurements of the sensors. Then we have

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_1(\mathbf{x}) \\ \mathbf{T}_2(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1^T \mathbf{x} \\ \mathbf{H}_2^T \mathbf{x} \end{bmatrix} = \mathbf{G}^T \mathbf{x} \tag{17}$$

So $\mathbf{T}$ is also Gaussian and

$$\mathbf{T} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{G}^T \mathbf{G}\right) \quad \text{under } \mathcal{H}_0$$

Let $p = p_1 + p_2$, and we can see that $\mathbf{T}$ is $p \times 1$. As a result, we construct the PDF as in (1) with

$$K(\boldsymbol{\theta}) = \ln E_0 \left[\exp\left(\boldsymbol{\theta}^T \mathbf{t}\right)\right] = \frac{1}{2}\sigma^2 \boldsymbol{\theta}^T \mathbf{G}^T \mathbf{G} \boldsymbol{\theta} \tag{18}$$

Hence the constructed PDF is

$$
\begin{aligned}
&p_{\mathbf{T}}(\mathbf{t};\boldsymbol{\theta}) \\
&= \exp\left[\boldsymbol{\theta}^T\mathbf{t} - K(\boldsymbol{\theta}) + \ln p_{\mathbf{T}}(\mathbf{t};\mathbf{0})\right] \\
&= \frac{1}{(2\pi\sigma^2)^{\frac{p}{2}}\det^{\frac{1}{2}}\left(\mathbf{G}^T\mathbf{G}\right)}\exp\left(-\frac{\mathbf{t}^T\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{t}}{2\sigma^2}\right) \\
&\quad\cdot\exp\left[\boldsymbol{\theta}^T\mathbf{t} - \frac{1}{2}\sigma^2\boldsymbol{\theta}^T\mathbf{G}^T\mathbf{G}\boldsymbol{\theta}\right]
\end{aligned} \tag{19}
$$

which can be simplified as

$$
\mathbf{T} \sim \mathcal{N}\left(\sigma^2\mathbf{G}^T\mathbf{G}\boldsymbol{\theta}, \sigma^2\mathbf{G}^T\mathbf{G}\right) \quad \text{under } \mathcal{H}_1 \tag{20}
$$

Note that $\boldsymbol{\theta}$ is the vector of the unknown parameters in the constructed PDF, and it is different from the truly unknown parameters $\boldsymbol{\alpha}$. From (3) and (18), the MLE of $\boldsymbol{\theta}$ satisfies

$$
\mathbf{t} = \frac{\partial K(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \sigma^2\mathbf{G}^T\mathbf{G}\boldsymbol{\theta}
$$

So

$$
\hat{\boldsymbol{\theta}} = \frac{1}{\sigma^2}\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{t}
$$

and the test statistic becomes

$$
\hat{\boldsymbol{\theta}}^T\mathbf{t} - K(\hat{\boldsymbol{\theta}}) = \frac{1}{2\sigma^2}\mathbf{t}^T\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{t} \tag{21}
$$

Next we consider the clairvoyant GLRT. That is the GLRT when we know the true PDF of $\mathbf{T}$ under $\mathcal{H}_1$ except for the truly underlying unknown parameters $\boldsymbol{\alpha}$. It is considered as the suboptimal test by plugging the MLE of $\boldsymbol{\alpha}$ into the true PDF parameterized by $\boldsymbol{\alpha}$. Since the constructed PDF may not be the true PDF, the clairvoyant GLRT requires more information than our method. From (17) we know that

$$
\mathbf{T} \sim \mathcal{N}\left(\mathbf{G}^T\mathbf{H}\boldsymbol{\alpha}, \sigma^2\mathbf{G}^T\mathbf{G}\right) \quad \text{under } \mathcal{H}_1 \tag{22}
$$

Note that (20) is the constructed PDF while (22) is the true PDF. We write the true PDF under $\mathcal{H}_1$ as $p_{\mathbf{T}}(\mathbf{t};\boldsymbol{\alpha})$. The MLE of $\boldsymbol{\alpha}$ is found by maximizing

$$
\begin{aligned}
&\ln\frac{p_{\mathbf{T}}(\mathbf{t};\boldsymbol{\alpha})}{p_{\mathbf{T}}(\mathbf{t};\mathbf{0})} \\
&= -\frac{1}{2\sigma^2}\left(\mathbf{t}-\mathbf{G}^T\mathbf{H}\boldsymbol{\alpha}\right)^T\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\left(\mathbf{t}-\mathbf{G}^T\mathbf{H}\boldsymbol{\alpha}\right) \\
&\quad + \frac{1}{2\sigma^2}\mathbf{t}^T\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{t}
\end{aligned}
$$

If $q \leq p$, i.e., the length of $\mathbf{t}$ is less than the length of $\boldsymbol{\alpha}$. Then the MLE $\hat{\boldsymbol{\alpha}}$ may not be unique but since $\left(\mathbf{t} - \mathbf{G}^T\mathbf{H}\boldsymbol{\alpha}\right)^T \left(\mathbf{G}^T\mathbf{G}\right)^{-1} \left(\mathbf{t} - \mathbf{G}^T\mathbf{H}\boldsymbol{\alpha}\right) \geq 0$, we could always find $\hat{\boldsymbol{\alpha}}$ such that $\mathbf{t} = \mathbf{G}^T\mathbf{H}\hat{\boldsymbol{\alpha}}$ and hence $\left(\mathbf{t} - \mathbf{G}^T\mathbf{H}\hat{\boldsymbol{\alpha}}\right)^T \left(\mathbf{G}^T\mathbf{G}\right)^{-1} \left(\mathbf{t} - \mathbf{G}^T\mathbf{H}\hat{\boldsymbol{\alpha}}\right) = 0$. Hence the clairvoyant GLRT statistic becomes

$$\ln \frac{p_{\mathbf{T}}(\mathbf{t}; \hat{\boldsymbol{\alpha}})}{p_{\mathbf{T}}(\mathbf{t}; \mathbf{0})} = \frac{1}{2\sigma^2} \mathbf{t}^T \left(\mathbf{G}^T\mathbf{G}\right)^{-1} \mathbf{t}$$

which is the same as our test statistic when $q \leq p$.

B. Partially Observed Linear Model with Gaussian Mixture Noise

The partially observed linear model remains the same as in the previous subsection except for instead of assuming that $\mathbf{w}$ is white Gaussian, we will assume that $\mathbf{w}$ has a Gaussian mixture distribution with two components, i.e.,

$$\mathbf{w} \sim \pi\mathcal{N}(\mathbf{0}, \sigma_1^2\mathbf{I}) + (1-\pi)\mathcal{N}(\mathbf{0}, \sigma_2^2\mathbf{I}) \tag{23}$$

where π, σ_1^2 and σ_2^2 are known ($0 < \pi < 1$). The following derivation can be easily extended when $\mathbf{w} \sim \sum_{i=1}^{L} \pi_i \mathcal{N}(\mathbf{0}, \sigma_i^2\mathbf{I})$.

Since $\mathbf{w}$ has a Gaussian mixture distribution, $\mathbf{T} = \mathbf{G}^T\mathbf{x}$ is also Gaussian mixture distributed and

$$\mathbf{T} \sim \pi\mathcal{N}(\mathbf{0}, \sigma_1^2\mathbf{G}^T\mathbf{G}) + (1-\pi)\mathcal{N}(\mathbf{0}, \sigma_2^2\mathbf{G}^T\mathbf{G}) \quad \text{under } \mathcal{H}_0$$

So we have

$$\begin{aligned} K(\boldsymbol{\theta}) &= \ln E_0 \left[\exp\left(\boldsymbol{\theta}^T\mathbf{t}\right)\right] \\ &= \ln\left(\pi e^{\frac{1}{2}\sigma_1^2\boldsymbol{\theta}^T\mathbf{G}^T\mathbf{G}\boldsymbol{\theta}} + (1-\pi)e^{\frac{1}{2}\sigma_2^2\boldsymbol{\theta}^T\mathbf{G}^T\mathbf{G}\boldsymbol{\theta}}\right) \end{aligned} \tag{24}$$

Hence the constructed PDF is

$$\begin{aligned} & p_{\mathbf{T}}(\mathbf{t}; \boldsymbol{\theta}) \\ &= \exp\left[\boldsymbol{\theta}^T\mathbf{t} - K(\boldsymbol{\theta}) + \ln p_{\mathbf{T}}(\mathbf{t}; \mathbf{0})\right] \\ &= \left[\frac{\pi}{\left(2\pi\sigma_1^2\right)^{\frac{p}{2}} \det^{\frac{1}{2}}\left(\mathbf{G}^T\mathbf{G}\right)} \exp\left(-\frac{\mathbf{t}^T\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{t}}{2\sigma_1^2}\right) + \frac{1-\pi}{\left(2\pi\sigma_2^2\right)^{\frac{p}{2}} \det^{\frac{1}{2}}\left(\mathbf{G}^T\mathbf{G}\right)} \exp\left(-\frac{\mathbf{t}^T\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{t}}{2\sigma_2^2}\right)\right] \\ & \quad \cdot \exp\left(\boldsymbol{\theta}^T\mathbf{t}\right) / \left(\pi e^{\frac{1}{2}\sigma_1^2\boldsymbol{\theta}^T\mathbf{G}^T\mathbf{G}\boldsymbol{\theta}} + (1-\pi)e^{\frac{1}{2}\sigma_2^2\boldsymbol{\theta}^T\mathbf{G}^T\mathbf{G}\boldsymbol{\theta}}\right) \end{aligned} \tag{25}$$

Although this constructed PDF cannot be further simplified, we can still find the MLE by solving

$$\begin{aligned}\mathbf{t} &= \frac{\partial K(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ &= \Big(\pi e^{\frac{1}{2}\sigma_1^2 \boldsymbol{\theta}^T \mathbf{G}^T \mathbf{G} \boldsymbol{\theta}} \cdot \sigma_1^2 \mathbf{G}^T \mathbf{G} \boldsymbol{\theta} \\ &\quad + (1-\pi) e^{\frac{1}{2}\sigma_2^2 \boldsymbol{\theta}^T \mathbf{G}^T \mathbf{G} \boldsymbol{\theta}} \cdot \sigma_2^2 \mathbf{G}^T \mathbf{G} \boldsymbol{\theta}\Big) / \\ &\quad \Big(\pi e^{\frac{1}{2}\sigma_1^2 \boldsymbol{\theta}^T \mathbf{G}^T \mathbf{G} \boldsymbol{\theta}} + (1-\pi) e^{\frac{1}{2}\sigma_2^2 \boldsymbol{\theta}^T \mathbf{G}^T \mathbf{G} \boldsymbol{\theta}}\Big) \end{aligned} \tag{26}$$

Our test statistic is just

$$\begin{aligned} &\hat{\boldsymbol{\theta}}^T \mathbf{t} - K(\hat{\boldsymbol{\theta}}) \\ &= \hat{\boldsymbol{\theta}}^T \mathbf{t} - \ln\left(\pi e^{\frac{1}{2}\sigma_1^2 \hat{\boldsymbol{\theta}}^T \mathbf{G}^T \mathbf{G} \hat{\boldsymbol{\theta}}} + (1-\pi) e^{\frac{1}{2}\sigma_2^2 \hat{\boldsymbol{\theta}}^T \mathbf{G}^T \mathbf{G} \hat{\boldsymbol{\theta}}}\right) \end{aligned} \tag{27}$$

where $\hat{\boldsymbol{\theta}}$ satisfies (26). Although no analytical solution of the MLE of $\boldsymbol{\theta}$ exists, it can be found using convex optimization techniques [12], [13]. Moreover, an analytical solution exists when $||\boldsymbol{\theta}|| \to 0$. To see this, we will show that

$$\lim_{||\boldsymbol{\theta}|| \to 0} \frac{\partial K(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} ./ \left(\pi \sigma_1^2 \mathbf{G}^T \mathbf{G} \boldsymbol{\theta} + (1-\pi) \sigma_2^2 \mathbf{G}^T \mathbf{G} \boldsymbol{\theta}\right) = \mathbf{1} \tag{28}$$

where ./ means element-by-element division.

To prove (28), we have

$$\lim_{||\boldsymbol{\theta}|| \to 0} \left(\pi e^{\frac{1}{2}\sigma_1^2 \boldsymbol{\theta}^T \mathbf{G}^T \mathbf{G} \boldsymbol{\theta}} + (1-\pi) e^{\frac{1}{2}\sigma_2^2 \boldsymbol{\theta}^T \mathbf{G}^T \mathbf{G} \boldsymbol{\theta}}\right) = 1 \tag{29}$$

and

$$\begin{aligned} &\lim_{||\boldsymbol{\theta}|| \to 0} \Big(\pi e^{\frac{1}{2}\sigma_1^2 \boldsymbol{\theta}^T \mathbf{G}^T \mathbf{G} \boldsymbol{\theta}} \cdot \sigma_1^2 \mathbf{G}^T \mathbf{G} \boldsymbol{\theta} \\ &\qquad + (1-\pi) e^{\frac{1}{2}\sigma_2^2 \boldsymbol{\theta}^T \mathbf{G}^T \mathbf{G} \boldsymbol{\theta}} \cdot \sigma_2^2 \mathbf{G}^T \mathbf{G} \boldsymbol{\theta}\Big) ./ \\ &\qquad \left(\pi \sigma_1^2 \mathbf{G}^T \mathbf{G} \boldsymbol{\theta} + (1-\pi) \sigma_2^2 \mathbf{G}^T \mathbf{G} \boldsymbol{\theta}\right) \\ &= \mathbf{1} \end{aligned} \tag{30}$$

by L'Hopital's rule. Dividing (30) by (29) and from (26), (28) is proved. As a result of (26) and (28), the MLE of $\boldsymbol{\theta}$ satisfies

$$\mathbf{t} = \pi \sigma_1^2 \mathbf{G}^T \mathbf{G} \boldsymbol{\theta} + (1-\pi) \sigma_2^2 \mathbf{G}^T \mathbf{G} \boldsymbol{\theta}$$

as $||\boldsymbol{\theta}|| \to 0$ and $\hat{\boldsymbol{\theta}}$ can be easily found as

$$\hat{\boldsymbol{\theta}} = \frac{1}{\pi \sigma_1^2 + (1-\pi) \sigma_2^2} \left(\mathbf{G}^T \mathbf{G}\right)^{-1} \mathbf{t} \tag{31}$$

Since

$$\lim_{||\boldsymbol{\theta}||\to 0} K(\boldsymbol{\theta})/\left(\frac{1}{2}\pi\sigma_1^2\boldsymbol{\theta}^T\mathbf{G}^T\mathbf{G}\boldsymbol{\theta}+\frac{1}{2}(1-\pi)\sigma_2^2\boldsymbol{\theta}^T\mathbf{G}^T\mathbf{G}\boldsymbol{\theta}\right)$$
$$=1$$

by using L'Hopital's rule twice, as $||\boldsymbol{\theta}||\to 0$, our test statistic becomes

$$\hat{\boldsymbol{\theta}}^T\mathbf{t}-\left(\frac{1}{2}\pi\sigma_1^2\hat{\boldsymbol{\theta}}^T\mathbf{G}^T\mathbf{G}\hat{\boldsymbol{\theta}}+\frac{1}{2}(1-\pi)\sigma_2^2\hat{\boldsymbol{\theta}}^T\mathbf{G}^T\mathbf{G}\hat{\boldsymbol{\theta}}\right)$$
$$=\frac{1}{2\left(\pi\sigma_1^2+(1-\pi)\sigma_2^2\right)}\mathbf{t}^T\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{t}$$

To find the clairvoyant GLRT statistic, we know that under $\mathcal{H}_1$

$$\mathbf{T}\sim\mathbf{G}^T\mathbf{H}\boldsymbol{\alpha}+\pi\mathcal{N}(\mathbf{0},\sigma_1^2\mathbf{G}^T\mathbf{G})+(1-\pi)\mathcal{N}(\mathbf{0},\sigma_2^2\mathbf{G}^T\mathbf{G}) \tag{32}$$

Note the difference between (25) and (32) since (25) is the constructed PDF and (32) is the true PDF. The MLE of $\boldsymbol{\alpha}$ is found by maximizing

$$\begin{aligned}
&p_{\mathbf{T}}(\mathbf{t};\boldsymbol{\alpha})\\
&=\frac{\pi}{(2\pi)^{q/2}\det^{1/2}\left(\sigma_1^2\mathbf{G}^T\mathbf{G}\right)}\\
&\quad\cdot\exp\left[-\frac{1}{2}(\mathbf{t}-\mathbf{G}^T\mathbf{H}\boldsymbol{\alpha})^T\frac{\left(\mathbf{G}^T\mathbf{G}\right)^{-1}}{\sigma_1^2}(\mathbf{t}-\mathbf{G}^T\mathbf{H}\boldsymbol{\alpha})\right]\\
&\quad+\frac{1-\pi}{(2\pi)^{q/2}\det^{1/2}\left(\sigma_2^2\mathbf{G}^T\mathbf{G}\right)}\\
&\quad\cdot\exp\left[-\frac{1}{2}(\mathbf{t}-\mathbf{G}^T\mathbf{H}\boldsymbol{\alpha})^T\frac{\left(\mathbf{G}^T\mathbf{G}\right)^{-1}}{\sigma_2^2}(\mathbf{t}-\mathbf{G}^T\mathbf{H}\boldsymbol{\alpha})\right]
\end{aligned}$$

When $q\le p$, the MLE of $\boldsymbol{\alpha}$ may not be unique but satisfies $\mathbf{t}=\mathbf{G}^T\mathbf{H}\hat{\boldsymbol{\alpha}}$. As a result, $p_{\mathbf{T}}(\mathbf{t};\hat{\boldsymbol{\alpha}})$ is a constant and the clairvoyant GLRT statistic becomes

$$-\ln p_{\mathbf{T}}(\mathbf{t};\mathbf{0})$$

Note that $p_{\mathbf{T}}(\mathbf{t};\mathbf{0})$ is decreasing as $\mathbf{t}^T\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{t}$ increases, the clairvoyant GLRT statistic

$$\mathbf{t}^T\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{t} \tag{33}$$

which is the same as our test statistic as $||\boldsymbol{\theta}||\to 0$.

Note that the noise in (23) is uncorrelated but not independent. We consider a general case when the noise can be correlated with PDF

$$\mathbf{w}\sim\pi\mathcal{N}(\mathbf{0},\mathbf{C}_1)+(1-\pi)\mathcal{N}(\mathbf{0},\mathbf{C}_2) \tag{34}$$

TABLE I

COMPARISON OF OUR TEST STATISTIC AND THE CLAIRVOYANT GLRT

	Our Method	Clairvoyant GLRT
Gaussian Noise	$\mathbf{t}^T\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{t}$	$\mathbf{t}^T\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{t}$
Uncorrelated Non-Gaussian Noise	$\max_{\boldsymbol{\theta}}\left[\boldsymbol{\theta}^T\mathbf{t}-\ln\left(\pi e^{\frac{1}{2}\sigma_1^2\boldsymbol{\theta}^T\mathbf{G}^T\mathbf{G}\boldsymbol{\theta}}+(1-\pi)e^{\frac{1}{2}\sigma_2^2\boldsymbol{\theta}^T\mathbf{G}^T\mathbf{G}\boldsymbol{\theta}}\right)\right]$	$\mathbf{t}^T\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{t}$
Correlated Non-Gaussian Noise	$\max_{\boldsymbol{\theta}}\left[\boldsymbol{\theta}^T\mathbf{t}-\ln\left(\pi e^{\frac{1}{2}\boldsymbol{\theta}^T\mathbf{G}^T\mathbf{C}_1\mathbf{G}\boldsymbol{\theta}}+(1-\pi)e^{\frac{1}{2}\boldsymbol{\theta}^T\mathbf{G}^T\mathbf{C}_2\mathbf{G}\boldsymbol{\theta}}\right)\right]$	$-\ln\left(\frac{\pi}{\det^{1/2}(\mathbf{C}_1)}\exp\left[-\frac{1}{2}\mathbf{t}^T\left(\mathbf{G}^T\mathbf{C}_1\mathbf{G}\right)^{-1}\mathbf{t}\right]+\frac{1-\pi}{\det^{1/2}(\mathbf{C}_2)}\exp\left[-\frac{1}{2}\mathbf{t}^T\left(\mathbf{G}^T\mathbf{C}_2\mathbf{G}\right)^{-1}\mathbf{t}\right]\right)$

It can be shown that similar to (27), our test statistic is

$$\hat{\boldsymbol{\theta}}^T\mathbf{t}-\ln\left(\pi e^{\frac{1}{2}\hat{\boldsymbol{\theta}}^T\mathbf{G}^T\mathbf{C}_1\mathbf{G}\hat{\boldsymbol{\theta}}}+(1-\pi)e^{\frac{1}{2}\hat{\boldsymbol{\theta}}^T\mathbf{G}^T\mathbf{C}_2\mathbf{G}\hat{\boldsymbol{\theta}}}\right) \tag{35}$$

and the clairvoyant GLRT statistic is

$$-\ln\left(\frac{\pi}{\det^{1/2}(\mathbf{C}_1)}\exp\left[-\frac{1}{2}\mathbf{t}^T\left(\mathbf{G}^T\mathbf{C}_1\mathbf{G}\right)^{-1}\mathbf{t}\right]\right.$$
$$\left.+\frac{1-\pi}{\det^{1/2}(\mathbf{C}_2)}\exp\left[-\frac{1}{2}\mathbf{t}^T\left(\mathbf{G}^T\mathbf{C}_2\mathbf{G}\right)^{-1}\mathbf{t}\right]\right) \tag{36}$$

when $q \leq p$.

C. Summary

We have considered a partially observed linear model with both Gaussian and non-Gaussian noise. Table I compares our test statistic with the clairvoyant GLRT.

In Gaussian noise, $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}))$. The test statistics are exactly the same.

In uncorrelated non-Gaussian noise, $\mathbf{w} \sim \pi\mathcal{N}(\mathbf{0}, \sigma_1^2\mathbf{I}) + (1-\pi)\mathcal{N}(\mathbf{0}, \sigma_2^2\mathbf{I})$. The test statistics are the same as $\boldsymbol{\theta} \to \mathbf{0}$.

In correlated non-Gaussian noise, $\mathbf{w} \sim \pi\mathcal{N}(\mathbf{0}, \mathbf{C}_1) + (1-\pi)\mathcal{N}(\mathbf{0}, \mathbf{C}_2)$. Although we cannot show the equivalence between these two test statistics, we will see in Section VII that their performances appear to be the same.

VI. EXAMPLES-DISTRIBUTED CLASSIFICATION

In this section, we compare our method with the estimated MAP classifier for some classification problems. The estimated MAP classifier assumes that the PDF of $\mathbf{T}$ under $\mathcal{H}_i$ is known except for some unknown underlying parameters $\boldsymbol{\alpha}_i$. We assume equal prior probability of the candidate hypothesis, i.e., $p(\mathcal{H}_1) = \ldots, = p(\mathcal{H}_M) = 1/M$. So the estimated MAP classifier reduces to the estimated maximum likelihood classifier [1], which finds the MLE of $\boldsymbol{\alpha}_i$ and chooses $\mathcal{H}_i$ for which the following is maximum over i:

$$p_{\mathbf{T}}(\mathbf{t}; \hat{\boldsymbol{\alpha}}_i) \tag{37}$$

where $\hat{\boldsymbol{\alpha}}_i$ is the MLE of $\boldsymbol{\alpha}_i$.

A. Linear Model with Known Variance

Consider the following classification model:

$$\mathcal{H}_i : \quad \mathbf{x} = A_i \mathbf{s}_i + \mathbf{w} \tag{38}$$

where $\mathbf{s}_i$ is an $N \times 1$ known signal vector with the same length as $\mathbf{x}$, A_i is the unknown signal amplitude, and $\mathbf{w}$ is white Gaussian noise with known variance σ^2. Assume that instead of observing $\mathbf{x}$, we can only observe the measurements of two sensors

$$\begin{aligned} \mathbf{T}_1 &= \mathbf{H}_1^T \mathbf{x} \\ \mathbf{T}_2 &= \mathbf{H}_2^T \mathbf{x} \end{aligned} \tag{39}$$

where $\mathbf{H}_1$ is $N \times p_1$ and $\mathbf{H}_2$ is $N \times p_2$. Here p_1 and p_2 are the length for vectors $\mathbf{T}_1$ and $\mathbf{T}_2$ respectively. We can write (39) as

$$\mathbf{T} = \mathbf{G}^T \mathbf{x} \tag{40}$$

by letting

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_1 \\ \mathbf{T}_2 \end{bmatrix}$$

and

$$\mathbf{G} = [\mathbf{H}_1 \ \mathbf{H}_2]$$

where $\mathbf{G}$ is $N \times (p_1 + p_2)$ with $p_1 + p_2 \leq N$. We assume that $\mathbf{G}$ has full column rank so that there are no redundant measurements of the sensors. Note that $\mathbf{G}$ can be any matrix with full column rank.

Let $\mathcal{H}_0$ be the reference hypothesis when there is noise only, i.e.,

$$\mathcal{H}_0 : \quad \mathbf{x} = \mathbf{w} \tag{41}$$

Since $\mathbf{x}$ is Gaussian under $\mathcal{H}_0$, according to (40), $\mathbf{T}$ is also Gaussian and

$$\mathbf{T} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{G}^T \mathbf{G}\right)$$

under $\mathcal{H}_0$. We construct the PDF under $\mathcal{H}_i$ as in (1) with

$$K(\boldsymbol{\theta}_i) = \ln E_0\left[\exp\left(\boldsymbol{\theta}_i^T \mathbf{T}\right)\right] = \frac{1}{2}\sigma^2 \boldsymbol{\theta}_i^T \mathbf{G}^T \mathbf{G} \boldsymbol{\theta}_i \tag{42}$$

Hence the constructed PDF is

$$\begin{aligned} & p_{\mathbf{T}}(\mathbf{t}; \boldsymbol{\theta}_i) \\ &= \exp\left[\boldsymbol{\theta}_i^T \mathbf{t} - K(\boldsymbol{\theta}_i) + \ln p_{\mathbf{T}}(\mathbf{t}; \mathbf{0})\right] \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{p_1+p_2}{2}} \det^{\frac{1}{2}}\left(\mathbf{G}^T\mathbf{G}\right)} \exp\left(-\frac{\mathbf{t}^T \left(\mathbf{G}^T \mathbf{G}\right)^{-1} \mathbf{t}}{2\sigma^2}\right) \\ & \quad \cdot \exp\left[\boldsymbol{\theta}_i^T \mathbf{t} - \frac{1}{2}\sigma^2 \boldsymbol{\theta}_i^T \mathbf{G}^T \mathbf{G} \boldsymbol{\theta}_i\right] \end{aligned} \tag{43}$$

which can be simplified as

$$\mathbf{T} \sim \mathcal{N}\left(\sigma^2 \mathbf{G}^T \mathbf{G} \boldsymbol{\theta}_i, \sigma^2 \mathbf{G}^T \mathbf{G}\right) \quad \text{under } \mathcal{H}_i \tag{44}$$

The next step is to find the MLE of $\boldsymbol{\theta}_i$. Note that the MLE of $\boldsymbol{\theta}_i$ is found by maximizing ${\boldsymbol{\theta}_i}^T \mathbf{t} - K(\boldsymbol{\theta}_i)$ over $\boldsymbol{\theta}_i$. If this optimization procedure is carried without any constraint, then $\hat{\boldsymbol{\theta}}_i$ would be the same for all i. Hence we need some implicit constraints in finding the MLE. Since $\boldsymbol{\theta}_i$ represents the signal under $\mathcal{H}_i$, we should have

$$\boldsymbol{\theta}_i = A_i \mathbf{G}^T \mathbf{s}_i = E_{\mathcal{H}_i}(\mathbf{T}) \tag{45}$$

which is the mean of $\mathbf{T}$ under $\mathcal{H}_i$. As a result, (44) can be written as

$$\mathbf{T} \sim \mathcal{N}\left(\sigma^2 A_i \mathbf{G}^T \mathbf{G} \mathbf{G}^T \mathbf{s}_i, \sigma^2 \mathbf{G}^T \mathbf{G}\right) \quad \text{under } \mathcal{H}_i \tag{46}$$

Thus, instead of finding the MLE of $\boldsymbol{\theta}_i$ by maximizing

$$\boldsymbol{\theta}_i^T \mathbf{t} - K(\boldsymbol{\theta}_i) = \boldsymbol{\theta}_i^T \mathbf{t} - \frac{1}{2}\sigma^2 \boldsymbol{\theta}_i^T \mathbf{G}^T \mathbf{G} \boldsymbol{\theta}_i \tag{47}$$

with the constraint in (45), we can find the MLE of A_i in (46) and then plug it into (45). It can be found that

$$\hat{A}_i = \frac{\mathbf{s}_i^T \mathbf{G} \mathbf{t}}{\sigma^2 \mathbf{s}_i^T \mathbf{G} \mathbf{G}^T \mathbf{G} \mathbf{G}^T \mathbf{s}_i} \tag{48}$$

and

$$\hat{\boldsymbol{\theta}}_i = \frac{\mathbf{G}^T \mathbf{s}_i \mathbf{s}_i^T \mathbf{G} \mathbf{t}}{\sigma^2 \mathbf{s}_i^T \mathbf{G} \mathbf{G}^T \mathbf{G} \mathbf{G}^T \mathbf{s}_i} \tag{49}$$

Hence by removing the constant factors, the test statistic of our classifier for $\mathcal{H}_i$ is

$$\frac{(\mathbf{s}_i^T \mathbf{G} \mathbf{t})^2}{(\mathbf{G}^T \mathbf{s}_i)^T \mathbf{G}^T \mathbf{G} (\mathbf{G}^T \mathbf{s}_i)} \tag{50}$$

according to (8).

Next we consider the estimate MAP classifier. In this case, we assume that we know the true PDF

$$\mathbf{T} \sim \mathcal{N}\left(A_i \mathbf{G}^T \mathbf{s}_i, \sigma^2 \mathbf{G}^T \mathbf{G}\right) \quad \text{under } \mathcal{H}_i \tag{51}$$

Note that (51) is the true PDF of $\mathbf{T}$ under $\mathcal{H}_i$ and (46) is the constructed PDF. It can be found that the MLE of A_i in the true PDF under $\mathcal{H}_i$ is

$$\hat{A}_i = \frac{\mathbf{s}_i^T \mathbf{G}\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{t}}{\mathbf{s}_i^T \mathbf{G}\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{s}_i} \tag{52}$$

By removing the constant terms, the test statistic of the estimated MAP classifier for $\mathcal{H}_i$ is

$$\frac{(\mathbf{s}_i^T \mathbf{G}\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{t})^2}{\left(\mathbf{G}^T\mathbf{s}_i\right)\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\left(\mathbf{G}^T\mathbf{s}_i\right)} \tag{53}$$

according to (37). Note that (48) and (52) are different because (48) is the MLE of A_i under the constructed PDF and (52) is the MLE of A_i under the true PDF. Also note that if $\mathbf{G}^T\mathbf{G}$ is a scaled identity matrix, test statistics in (50) and (53) are equivalent, and hence our method coincides with the estimated MAP classifier.

B. Linear Model with Unknown Variance

To extend the above example, we consider the above linear model with unknown noise variance σ^2. As we have shown in (46), the constructed PDF is still

$$\mathbf{T} \sim \mathcal{N}\left(\sigma^2 A_i \mathbf{G}^T\mathbf{G}\mathbf{G}^T \mathbf{s}_i, \sigma^2 \mathbf{G}^T \mathbf{G}\right) \quad \text{under } \mathcal{H}_i \tag{54}$$

except for that σ^2 is unknown. Let $B_i = \sigma^2 A_i$, we have

$$\mathbf{T} \sim \mathcal{N}\left(B_i \mathbf{G}^T\mathbf{G}\mathbf{G}^T \mathbf{s}_i, \sigma^2 \mathbf{G}^T \mathbf{G}\right) \quad \text{under } \mathcal{H}_i \tag{55}$$

Instead of finding the MLEs of A_i and σ^2, we can find the MLEs of B_i and σ^2. Let $\mathbf{h}_i = \mathbf{G}^T\mathbf{G}\mathbf{G}^T\mathbf{s}_i$ and $\mathbf{C} = \mathbf{G}^T\mathbf{G}$. It can be shown that

$$\hat{B}_i = (\mathbf{h}_i^T\mathbf{C}^{-1}\mathbf{h}_i)^{-1}\mathbf{h}_i^T\mathbf{C}^{-1}\mathbf{t} \tag{56}$$

and

$$\hat{\sigma}^2 = \frac{1}{p_1 + p_2}(\mathbf{t} - \mathbf{h}_i\hat{B}_i)^T\mathbf{C}^{-1}(\mathbf{t} - \mathbf{h}_i\hat{B}_i) \tag{57}$$

By removing the constant factors, it can also be shown that the test statistic is equivalent to

$$\frac{\mathbf{t}^T\mathbf{C}^{-1}\mathbf{h}_i(\mathbf{h}^T\mathbf{C}^{-1}\mathbf{h}_i)^{-1}\mathbf{h}_i^T\mathbf{C}^{-1}\mathbf{t}}{\mathbf{t}^T\left[\mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{h}_i(\mathbf{h}_i^T\mathbf{C}^{-1}\mathbf{h})_i^{-1}\mathbf{h}_i^T\mathbf{C}^{-1}\right]\mathbf{t}} \tag{58}$$

TABLE II

COMPARISON OF OUR TEST STATISTIC AND THE ESTIMATED MAP CLASSIFIER

	Our Method	Estimated MAP
Known σ^2	$\frac{(\mathbf{s}_i^T\mathbf{G}\mathbf{t})^2}{(\mathbf{G}^T\mathbf{s}_i)^T\mathbf{G}^T\mathbf{G}(\mathbf{G}^T\mathbf{s}_i)}$	$\frac{(\mathbf{s}_i^T\mathbf{G}\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{t})^2}{(\mathbf{G}^T\mathbf{s}_i)\left(\mathbf{G}^T\mathbf{G}\right)^{-1}(\mathbf{G}^T\mathbf{s}_i)}$
Unknown σ^2	$\frac{\mathbf{t}^T\mathbf{C}^{-1}\mathbf{h}_i(\mathbf{h}^T\mathbf{C}^{-1}\mathbf{h}_i)^{-1}\mathbf{h}_i^T\mathbf{C}^{-1}\mathbf{t}}{\mathbf{t}^T\left[\mathbf{C}^{-1}-\mathbf{C}^{-1}\mathbf{h}_i(\mathbf{h}_i^T\mathbf{C}^{-1}\mathbf{h})_i^{-1}\mathbf{h}_i^T\mathbf{C}^{-1}\right]\mathbf{t}}$	$\frac{\mathbf{t}^T\mathbf{C}^{-1}\mathbf{g}_i(\mathbf{g}_i^T\mathbf{C}^{-1}\mathbf{g}_i)^{-1}\mathbf{h}^T\mathbf{C}^{-1}\mathbf{t}}{\mathbf{t}^T\left[\mathbf{C}^{-1}-\mathbf{C}^{-1}\mathbf{g}_i(\mathbf{g}_i^T\mathbf{C}^{-1}\mathbf{g}_i)^{-1}\mathbf{g}_i^T\mathbf{C}^{-1}\right]\mathbf{t}}+$

where $\mathbf{h}_i = \mathbf{G}^T\mathbf{G}\mathbf{G}^T\mathbf{s}_i$, $\mathbf{g}_i = \mathbf{G}^T\mathbf{s}_i$ and $\mathbf{C} = \mathbf{G}^T\mathbf{G}$.

Next we consider the estimated MAP classifier. So the true PDF is still

$$\mathbf{T} \sim \mathcal{N}\left(A_i\mathbf{G}^T\mathbf{s}_i, \sigma^2\mathbf{G}^T\mathbf{G}\right) \quad \text{under } \mathcal{H}_i \tag{59}$$

with unknown A_i and σ^2. Let $\mathbf{g}_i = \mathbf{G}^T\mathbf{s}_i$ and $\mathbf{C} = \mathbf{G}^T\mathbf{G}$. Similar to (56), (57) and (58), it can be shown that

$$\hat{A}_i = (\mathbf{g}_i^T\mathbf{C}^{-1}\mathbf{g}_i)^{-1}\mathbf{g}_i^T\mathbf{C}^{-1}\mathbf{t} \tag{60}$$

$$\hat{\sigma}^2 = \frac{1}{p_1+p_2}(\mathbf{t}-\mathbf{g}_i\hat{A}_i)^T\mathbf{C}^{-1}(\mathbf{t}-\mathbf{g}_i\hat{A}_i) \tag{61}$$

and the test statistic of the estimated MAP classifier is

$$\frac{\mathbf{t}^T\mathbf{C}^{-1}\mathbf{g}_i(\mathbf{g}_i^T\mathbf{C}^{-1}\mathbf{g}_i)^{-1}\mathbf{h}^T\mathbf{C}^{-1}\mathbf{t}}{\mathbf{t}^T\left[\mathbf{C}^{-1}-\mathbf{C}^{-1}\mathbf{g}_i(\mathbf{g}_i^T\mathbf{C}^{-1}\mathbf{g}_i)^{-1}\mathbf{g}_i^T\mathbf{C}^{-1}\right]\mathbf{t}} \tag{62}$$

Note that if $\mathbf{G}^T\mathbf{G}$ is a scaled identity matrix, since $\mathbf{h}_i = \mathbf{G}^T\mathbf{G}\mathbf{g}_i$, the test statistics in (58) and (62) are equivalent. Hence our method is exactly the same as the estimated MAP classifier if $\mathbf{G}^T\mathbf{G}$ is a scaled identity matrix.

C. Summary

We have considered a linear model both known and unknown noise variance. Table II compares our test statistic with the estimated MAP classifier. If $\mathbf{G}^T\mathbf{G}$ is a scaled identity matrix, our method and the estimated MAP classifier are exactly the same.

VII. SIMULATIONS

A. Distributed Detection

Since our test statistic coincides with the clairvoyant GLRT under Gaussian noise as shown in subsection V-A, we will only compare the performances under non-Gaussian noise (both uncorrelated noise as in (23) and correlated noise as in (34)).

Consider the model where

$$x[n] = A_1 + A_2 r^n + A_3 \cos(2\pi f n + \phi) + w[n] \tag{63}$$

for $n = 0, 1, \ldots, N-1$ with known r and frequency f but unknown amplitudes A_1, A_2, A_3 and phase ϕ. This is a linear model as in (15) where

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & r & \cos(2\pi f) & \sin(2\pi f) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & r^{N-1} & \cos(2\pi f(N-1)) & \sin(2\pi f(N-1)) \end{bmatrix}$$

and $\boldsymbol{\alpha} = [A_1, A_2, A_3 \cos\phi, -A_3 \sin\phi]^T$.

Let $\mathbf{w}$ have an uncorrelated Gaussian mixture distribution as in (23). For the partially observed linear model, we observe two sensor outputs as in (16). We compare the GLRT in (27) with the clairvoyant GLRT in (33). Note that the MLE of $\boldsymbol{\theta}$ in (27) is found numerically, not by the asymptotic approximation in (31). In the simulation, we use $N = 20$, $A_1 = 2$, $A_2 = 3$, $A_3 = 4$, $\phi = \pi/4$, $r = 0.95$, $f = 0.34$, $\pi = 0.9$, $\sigma_1^2 = 50$, $\sigma_2^2 = 500$, and $\mathbf{H}_1$ and $\mathbf{H}_2$ are the first and third columns in $\mathbf{H}$ respectively, i.e., $\mathbf{H}_1 = [1, 1, \ldots, 1]^T$, $\mathbf{H}_2 = [1, \cos(2\pi f), \ldots, \cos(2\pi f(N-1))]^T$. As shown in Figure 2, the performances are almost the same which justifies their equivalence under small signals assumption shown in Section V.

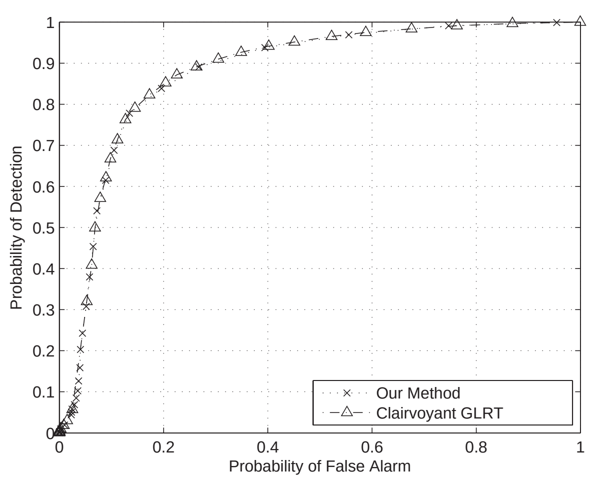

Fig. 2. ROC curves for the GLRT using the constructed PDF and the clairvoyant GLRT with uncorrelated Gaussian mixture noise.

Next for the same model in (63), let $\mathbf{w}$ have a correlated Gaussian mixture distribution as in (23). We compare performances of the GLRT using the constructed PDF as in (35) and the clairvoyant GLRT as in (36). We use $N = 20$, $A_1 = 3$, $A_2 = 4$, $A_3 = 3$, $\phi = \pi/7$, $r = 0.9$, $f = 0.46$, $\pi = 0.7$, $\mathbf{H}_1 = [1, 1, \ldots, 1]^T$, $\mathbf{H}_2 = [1, \cos(2\pi f), \ldots, \cos(2\pi f(N-1))]^T$. The covariance matrices $\mathbf{C}_1$, $\mathbf{C}_2$ are generated using $\mathbf{C}_1 = \mathbf{R}_1^T \times \mathbf{R}_1$, $\mathbf{C}_2 = \mathbf{R}_2^T \times \mathbf{R}_2$, where $\mathbf{R}_1$, $\mathbf{R}_2$ are full rank $N \times N$ matrices. As shown in Figure 3, the performances are still very similar.

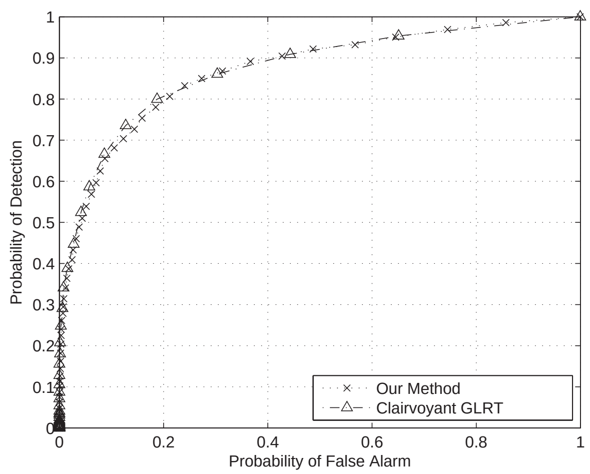

Fig. 3. ROC curves for the GLRT using the constructed PDF and the clairvoyant GLRT with correlated Gaussian mixture noise.

B. Distributed Classification

For the model in (38)

$$\mathcal{H}_i : \quad \mathbf{x} = A_i \mathbf{s}_i + \mathbf{w}$$

we first consider a case when $\mathbf{G}^T\mathbf{G}$ is approximately a scaled identity matrix. Let $A_1 = 0.4$, $A_2 = 1.2$, $A_3 = 0.9$ and

$$s_1(n) = \cos(2\pi f_1 n)$$

$$s_2(n) = \cos(2\pi f_2 n)$$

$$s_3(n) = \cos(2\pi f_3 n)$$

where $n = 0, 1, \ldots, N-1$ with $N = 25$, and $f_1 = 0.14$, $f_2 = 0.34$, $f_3 = 0.41$. Let $p(\mathcal{H}_1) = p(\mathcal{H}_2) = p(\mathcal{H}_3) = 1/3$. Assume that there are two sensors, each with an observation matrix as follows respectively:

$$\mathbf{H}_1 = \begin{bmatrix} 1 & \cos(2\pi f_1) & \cdots & \cos\left(2\pi f_1(N-1)\right) \\ 1 & \cos(2\pi f_2) & \cdots & \cos\left(2\pi f_2(N-1)\right) \end{bmatrix}^T$$
$$\mathbf{H}_2 = \begin{bmatrix} 1 & \cos(2\pi f_3) & \cdots & \cos\left(2\pi f_3(N-1)\right) \end{bmatrix}^T$$

We use (50) and (53) as our test statistics for the two methods respectively when σ^2 is known. Test statistics in (58) and (62) are used when σ^2 is unknown. The probabilities of correct classification are plotted versus $\ln(1/\sigma^2)$ in Figure 4. We see that our method has the same performance with the estimated MAP classifier with known or unknown σ^2, and probabilities of correct classification goes to 1 as $\sigma^2 \to 0$.

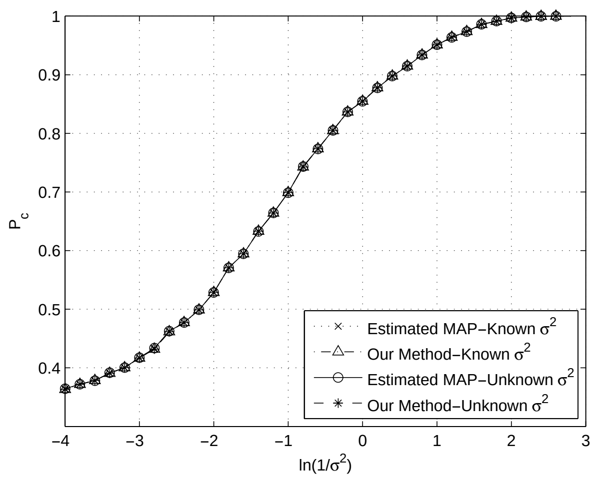

Fig. 4. Probability of correct classification for both methods.

Next we consider a case when $\mathbf{G}^T\mathbf{G}$ is not a scaled identity matrix. Let $A_1 = 0.5$, $A_2 = 1$, $A_3 = 1$ and

$$s_1(n) = \cos(2\pi f_1 n) + 1$$
$$s_2(n) = \cos(2\pi f_2 n) + 0.5$$
$$s_3(n) = \cos(2\pi f_3 n)$$

where $n = 0, 1, \ldots, N-1$ with $N = 20$, and $f_1 = 0.17$, $f_2 = 0.28$, $f_3 = 0.45$. Let $p(\mathcal{H}_1) = p(\mathcal{H}_2) = p(\mathcal{H}_3) = 1/3$. Assume that there are three sensors (this is an extension of the two sensor assumption), each with an observation matrix as follows respectively:

$$\mathbf{H}_1 = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T$$

$$\mathbf{H}_2 = \begin{bmatrix} 1 & \cos(2\pi f_1) & \cdots & \cos\left(2\pi f_1(N-1)\right) \\ 1 & \cos(2\pi f_2) & \cdots & \cos\left(2\pi f_2(N-1)\right) \end{bmatrix}^T$$

$$\mathbf{H}_3 = \left[1 \quad \cos\left(2\pi(f_3+0.02)\right) \quad \cdots \right.$$
$$\left. \cos\left(2\pi(f_3+0.02)(N-1)\right)\right]^T$$

Note that in $\mathbf{H}_3$, we set the frequency to $f_3+0.02$. This is the case when the knowledge of the frequency is not accurate. We also see in Figure 4 that the performances of both methods are the same with known or unknown σ^2, and probabilities of correct classification goes to 1 as $\sigma^2 \to 0$.

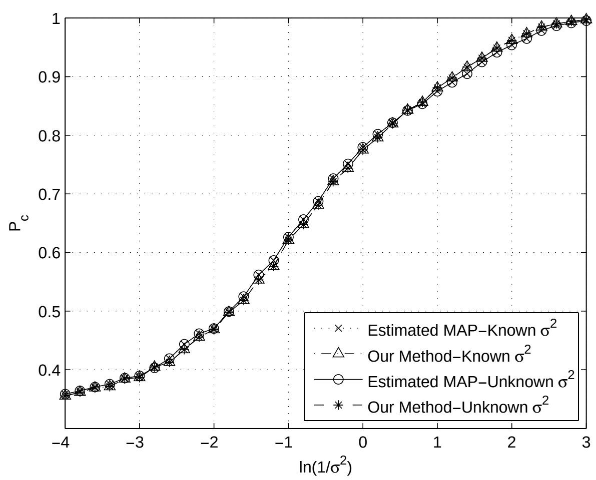

Fig. 5. Probability of correct classification for both methods.

VIII. Conclusions

A novel method of constructing the joint PDF of the measurements from multiple sensors distributed systems has been proposed. Only a reference PDF is needed in the construction. The constructed PDF is asymptotically optimal in KL divergence. The performance of our method has shown to be as good as existing methods for both detection and classification, while less information is needed for our method.

References

[1] S. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1998.

[2] S. Thomopoulos, R. Viswanathan, and D. Bougoulias, “Optimal distributed decision fusion,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 25, pp. 761–765, Sep. 1989.

[3] Z. Chair and P. Varshney, “Optimal data fusion in multiple sensor detection systems,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, pp. 98–101, Jan. 1986.

[4] J. Kittler, M. Hatef, R. Duin, and J. Matas, “On combining classifiers,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, pp. 226–239, Mar. 1998.

[5] A. Sundaresan, P. Varshney, and N. Rao, “Distributed detection of a nuclear radioactive source using fusion of correlated decisions,” in *Information Fusion, 2007 10th International Conference on*, 2007, pp. 1–7.

[6] S. Iyengar, P. Varshney, and T. Damarla, “A parametric copula based framework for multimodal signal processing,” in *ICASSP*, 2009, pp. 1893–1896.

[7] S. Kay and Q. Ding, “Exponentially embedded families for multimodal sensor processing,” in *ICASSP*, Mar. 2010, pp. 3770–3773.

[8] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[9] S. Kay, A. Nuttall, and P. Baggenstoss, “Multidimensional probability density function approximations for detection, classification, and model order selection,” *IEEE Trans. Signal Process.*, vol. 49, pp. 2240–2252, Oct. 2001.

[10] S. Kay, Q. Ding, and D. Emge, “Joint pdf construction for sensor fusion and distributed detection,” in *International Conference on Information Fusion*, Jun. 2010.

[11] L. Brown, *Fundamentals of Statistical Exponential Families*. Institute of Mathematical Statistics, 1986.

[12] S. Boyd and L.Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[13] D. Luenberger, *Linear and Nonlinear Programming*, 2nd ed. Springer, 2003.

[14] S. Kay, Q. Ding, and M. Rangaswamy, “Sensor integration for classification,” in *Asilomar Conference on Signals, Systems, and Computers*, Nov. 2010.

[15] S. Kullback, *Information Theory and Statistics*, 2nd ed. Courier Dover Publications, 1997.

[16] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley and Sons, 2006.

[17] M. Westover, “Asymptotic geometry of multiple hypothesis testing,” *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3327–3329, Jul. 2008.

[18] S. Kay, “Exponentially embedded families - new approaches to model order estimation,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, pp. 333–345, Jan. 2005.