Joint PDF Construction for Sensor Fusion and Distributed Detection

Steven Kay and **Quan Ding**
Department of Electrical, Computer, and Biomedical Engineering
University of Rhode Island
Kingston, RI, U.S.A.
{kay,dingqqq}@ele.uri.edu

Darren Emge
Edgewood Chemical Biological Center
Aberdeen Proving Ground, MD, U.S.A.
darren.emge@us.army.mil *

Abstract – *A novel method of constructing a joint PDF under $\mathcal{H}_1$, when the joint PDF under $\mathcal{H}_0$ is known, is developed. It has direct application in distributed detection systems. The construction is based on the exponential family and it is shown that asymptotically the constructed PDF is optimal. The generalized likelihood ratio test (GLRT) is derived based on this method for the partially observed linear model. Interestingly, the test statistic is equivalent to the clairvoyant GLRT, which uses the true PDF under $\mathcal{H}_1$, even if the noise is non-Gaussian.*

Keywords: Distributed detection, data fusion, joint PDF, exponential family, Gaussian mixture.

1 Introduction

Data fusion or sensor fusion in distributed detection systems has been widely studied over the years. By combining the data from different sensors, better performance can be expected than using a single sensor alone. The optimal detection performance can be obtained if the joint probability density function (PDF) of the measurements from different sensors under each hypothesis is completely known. However in practice, this joint PDF is usually not available. So a key issue in this area is how to construct the joint PDF of the measurements from different sensors. One common approach is to assume that the measurements are independent [1], [2]. This approach has been widely used due to its simplicity, since the joint PDF is then the product of the marginal PDFs. This leads to the product rule in combining classifiers, and it is effectively a severe rule as stated in [3] that "it is sufficient for a single recognition engine to inhibit a particular interpretation by outputting a close to zero probability for it". Moreover, the independence is a strong assumption and the measurements can be correlated in many cases. The dependence between measurements has been considered in [4, 5, 6]. A copula based framework is used in [4, 5] to estimate the joint PDF from the marginal PDFs. The exponentially embedded families (EEFs) are proposed in [6] to asymptotically minimize the Kullback-Leibler (KL) divergence between the true PDF and the estimated one.

Note that all the above methods are based on the assumption that we know the marginal PDFs of the measurements. But in many cases, the marginal PDFs may not be available or accurate. This could happen when we do not have enough training data. In this paper, we will present a new way of constructing a 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 the maximum likelihood estimate (MLE) of the unknown parameters can be easily solved based on the exponential family. Since there is no Gaussian distribution assumption on the reference PDF, this method can be very useful when the underlying distributions are non-Gaussian. In the examples when we apply this method to the detection problem, under some conditions, the detection statistics can be shown to be the same as the the clairvoyant generalized likelihood ratio test (GLRT), which is the test when the true PDF under $\mathcal{H}_1$ is known except for the usual unknown parameters.

The paper is organized as follows. Section 2 formulates the detection problem. The construction of the joint PDF is presented and is applied to the detection problem in Section 3. The KL divergence between the true PDF to the constructed PDF is examined in Section 4. We give two examples in Section 5. In Section 6, some simulation results are shown. Conclusions are given in Section 7.

*This work was supported by the Air Force Office of Scientific Research through the Air Force Research Laboratory Sensors Directorate under Contract FA8650-08-D-1303 and by the Edgewood Chemical Biological Center under Contract W91ZLK-08-P-1214.

2 Problem Statement

Consider the detection 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 (see Figure 1). All the results are valid for any number of sensors. We just choose two for simplicity. 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 $\mathcal{H}_0$ (see [7]), and thus we assume $p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1,\mathbf{t}_2;\mathcal{H}_0)$ is completely known. Under $\mathcal{H}_1$ when a signal is present, we may not have enough training data to estimate the joint PDF under $\mathcal{H}_1$. So our goal is to construct an appropriate $p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1,\mathbf{t}_2;\mathcal{H}_1)$ and use it for detection. 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) $p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1,\mathbf{t}_2;\mathcal{H}_1)$ depends on signal parameters $\boldsymbol{\theta}$ so that

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

and

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

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

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

3 Construction of Joint PDF for Detection

To simplify the notation, let

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

so that the joint PDF $p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1,\mathbf{t}_2;\boldsymbol{\theta})$ can be written as $p_{\mathbf{T}}(\mathbf{t};\boldsymbol{\theta})$. Since we assume that $||\boldsymbol{\theta}||$ is small, we expand the log-likelihood function using a first order Taylor expansion.

$$\ln p_{\mathbf{T}}(\mathbf{t};\boldsymbol{\theta}) = \ln p_{\mathbf{T}}(\mathbf{t};\mathbf{0}) + \boldsymbol{\theta}^T \frac{\partial \ln p_{\mathbf{T}}(\mathbf{t};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\mathbf{0}} + o(||\boldsymbol{\theta}||) \tag{1}$$

We omit the $o(||\boldsymbol{\theta}||)$ term but in order for $p_{\mathbf{T}}(\mathbf{t};\boldsymbol{\theta})$ to be a valid PDF, we normalize the PDF to integrate to one as

$$\begin{aligned} & p_{\mathbf{T}}(\mathbf{t};\boldsymbol{\theta}) \\ & = \exp\left[\boldsymbol{\theta}^T \frac{\partial \ln p_{\mathbf{T}}(\mathbf{t};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\mathbf{0}} - K(\boldsymbol{\theta}) + \ln p_{\mathbf{T}}(\mathbf{t};\mathbf{0})\right] \end{aligned} \tag{2}$$

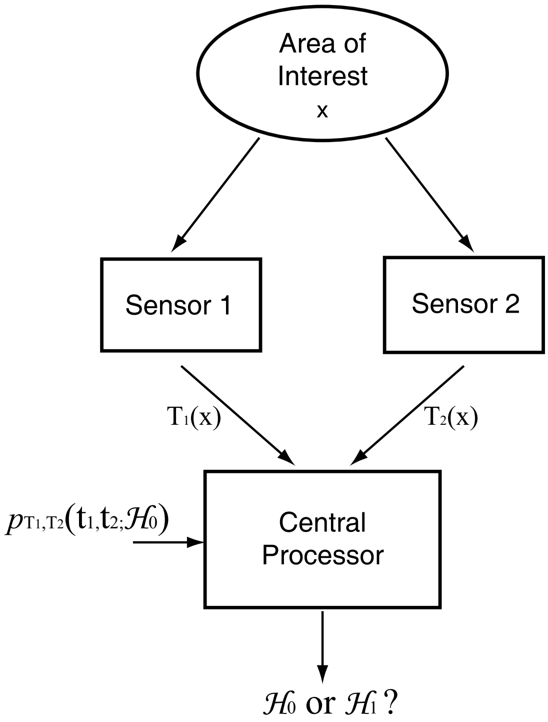

Figure 1: Distributed detection system with two sensors

where

$$K(\boldsymbol{\theta}) = \ln E_0\left[\exp\left(\boldsymbol{\theta}^T \frac{\partial \ln p_{\mathbf{T}}(\mathbf{t};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\mathbf{0}}\right)\right] \tag{3}$$

Here E_0 denotes the expected value under $\mathcal{H}_0$.

Next we assume that the sensor outputs are the score functions, i.e.,

$$\mathbf{t} = \frac{\partial \ln p_{\mathbf{T}}(\mathbf{t};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\mathbf{0}} \tag{4}$$

and are sufficient statistics for the constructed PDF under $\mathcal{H}_1$. This will be true if $p_{\mathbf{T}}(\mathbf{t};\boldsymbol{\theta})$ is in the exponential family with

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

where

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

and $E_0(\mathbf{T}) = \mathbf{0}$. This can be easily verified since by (5), we have

$$\frac{\partial \ln p_{\mathbf{T}}(\mathbf{t};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\mathbf{0}} = \mathbf{t} - \frac{\partial K(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\mathbf{0}}$$

and

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

as well known properties of the exponential family. Note that even if $E_0(\mathbf{T}) \neq \mathbf{0}$, we still have

$$\mathbf{t} - E_0(\mathbf{T}) = \frac{\partial \ln p_{\mathbf{T}}(\mathbf{t};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} |_{\boldsymbol{\theta}=\mathbf{0}}$$

We can use $\mathbf{t} - E_0(\mathbf{T})$ instead of $\mathbf{t}$ as the sensor outputs and hence still satisfy (4) and (5). As a result, we will use (5) as our constructed PDF. This implies that $\mathbf{t}$ is a sufficient statistic for the constructed exponential PDF, and hence this PDF incorporates all the sensor information. Note that if $\mathbf{T}_1$, $\mathbf{T}_2$ are statistically dependent under $\mathcal{H}_0$, they will also be dependent under $\mathcal{H}_1$. Also note that only $p_{\mathbf{T}}(\mathbf{t};\mathbf{0})$ is required in (5). It is assumed in practice that this can be estimated or found analytically [7] with reasonable accuracy.

Since $\boldsymbol{\theta}$ is unknown, the GLRT is used for detection [8]. We want to maximize $p_{\mathbf{T}}(\mathbf{t};\boldsymbol{\theta})$ or $\ln \frac{p_{\mathbf{T}}(\mathbf{t};\boldsymbol{\theta})}{p_{\mathbf{T}}(\mathbf{t};\mathbf{0})} = \boldsymbol{\theta}^T\mathbf{t} - K(\boldsymbol{\theta})$ over $\boldsymbol{\theta}$. This is a convex optimization problem since $K(\boldsymbol{\theta})$ is convex by Holder's inequality [9]. Hence many convex optimization techniques can be utilized [10, 11]. By taking the derivative with respect to $\boldsymbol{\theta}$, the MLE of $\boldsymbol{\theta}$ is found by solving

$$\mathbf{t} = \frac{\partial K(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{7}$$

Also because $K(\boldsymbol{\theta})$ is convex, the MLE $\hat{\boldsymbol{\theta}}$ is unique. Then we decide $\mathcal{H}_1$ if

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

where τ is a threshold.

4 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 $D\left(p_1 \| p_0\right) \geq 0$ with equality if and only if $p_1 = p_0$ [12]. By Stein's lemma [13], the KL divergence measures the asymptotic performance for detection.

It can be shown 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 [6, 14].

5 Examples

In this section, we will apply the the constructed PDF of (5) to some detection problems. We will start with the simple case with Gaussian noise, and then we will extend the result to the more general case with Gaussian mixture noise.

5.1 Partially Observed Linear Model with Gaussian Noise

Suppose we have the linear model with

$$\mathbf{x} = \mathbf{H}\boldsymbol{\alpha} + \mathbf{w} \tag{9}$$

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

$$\begin{aligned} \mathbf{T}_1(\mathbf{x}) &= \mathbf{H}_1^T\mathbf{x} \\ \mathbf{T}_2(\mathbf{x}) &= \mathbf{H}_2^T\mathbf{x} \end{aligned} \tag{10}$$

where $\mathbf{T}_1$ and $\mathbf{T}_2$ could be any subset of columns of $\mathbf{H}$. Note that $[\mathbf{H}_1, \mathbf{H}_2]$ does not have to be $\mathbf{H}$. This model is called a partially observed linear model. Note that a sufficient statistic is $\mathbf{H}^T\mathbf{x}$, so there is some information loss over the case when $\mathbf{x}$ is observed, unless $\mathbf{H} = [\mathbf{H_1}, \mathbf{H_2}]$.

Let $\mathbf{G} = [\mathbf{H}_1, \mathbf{H}_2]$, then we have

$$\mathbf{T} = \left[\begin{array}{c} \mathbf{T}_1(\mathbf{x}) \\ \mathbf{T}_2(\mathbf{x}) \end{array}\right] = \left[\begin{array}{c} \mathbf{H}_1^T\mathbf{x} \\ \mathbf{H}_2^T\mathbf{x} \end{array}\right] = \mathbf{G}^T\mathbf{x} \tag{11}$$

Therefore, $\mathbf{T}$ is also Gaussian with PDF

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

and $\mathbf{T}_1$, $\mathbf{T}_2$ are seen to be correlated for $\mathbf{H}_1^T\mathbf{H}_2 \neq \mathbf{0}$. As a result, we construct the PDF as in (5) with

$$K(\boldsymbol{\theta}) = \ln E_0\left[\exp\left(\boldsymbol{\theta}^T\mathbf{T}\right)\right] = \frac{1}{2}\sigma^2\boldsymbol{\theta}^T\mathbf{G}^T\mathbf{G}\boldsymbol{\theta} \tag{12}$$

Note that $\boldsymbol{\theta}$ is the vector of the unknown parameters in the constructed PDF, and it is different from the unknown parameters $\boldsymbol{\alpha}$ in the linear model.

By (7) and (12), 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

$$\hat{\boldsymbol{\theta}}^T \mathbf{t} - K(\hat{\boldsymbol{\theta}}) = \frac{1}{2\sigma^2}\mathbf{t}^T\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{t} \tag{13}$$

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 underlying unknown parameters $\boldsymbol{\alpha}$. From (11) we know that

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

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

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

Let $\mathbf{t}$ be $q\times 1$. If $q \le p$, 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. Since $\left(\mathbf{t}-\mathbf{G}^T\mathbf{H}\boldsymbol{\alpha}\right)^T\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\left(\mathbf{t}-\mathbf{G}^T\mathbf{H}\boldsymbol{\alpha}\right) \ge 0$, 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 the GLRT on our constructed PDF (see (13)) when $q \le p$.

5.2 Partially Observed Linear Model with Non-Gaussian Noise

The partially observed linear model remains the same as in the previous subsection except 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.,

$$\mathbf{w} \sim \pi\mathcal{N}(\mathbf{0},\sigma_1^2\mathbf{I}) + (1-\pi)\mathcal{N}(\mathbf{0},\sigma_2^2\mathbf{I}) \tag{14}$$

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

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

It can be shown that the GLRT statistic is

$$\max_{\boldsymbol{\theta}}\left[\boldsymbol{\theta}^T\mathbf{t} - \ln\left(\pi e^{\frac{1}{2}\sigma_1^2\boldsymbol{\theta}^T\mathbf{G}^T\mathbf{G}\boldsymbol{\theta}} + (1-\pi)e^{\frac{1}{2}\sigma_2^2\boldsymbol{\theta}^T\mathbf{G}^T\mathbf{G}\boldsymbol{\theta}}\right)\right] \tag{15}$$

Although no analytical solution of the MLE of $\boldsymbol{\theta}$ exists, it can be found using convex optimization techniques [10, 11]. Moreover, an analytical solution exists as $||\boldsymbol{\theta}|| \to 0$. It can be shown that

$$\hat{\boldsymbol{\theta}} = \frac{1}{\pi\sigma_1^2 + (1-\pi)\sigma_2^2}\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{t} \tag{16}$$

and the GLRT statistic becomes

$$\frac{1}{2\left(\pi\sigma_1^2 + (1-\pi)\sigma_2^2\right)}\mathbf{t}^T\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{t} \tag{17}$$

as $||\boldsymbol{\theta}|| \to 0$.

The clairvoyant GLRT statistic can be shown to be equivalent to

$$\mathbf{t}^T\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{t} \tag{18}$$

when $q \le p$. Hence the clairvoyant GLRT coincides with the GLRT using the constructed PDF as $||\boldsymbol{\theta}|| \to 0$.

Note that the noise in (14) is uncorrelated but not independent. We consider a general case when the noise can be correlated with PDF

$$\mathbf{w} \sim \pi\mathcal{N}(\mathbf{0},\mathbf{C}_1) + (1-\pi)\mathcal{N}(\mathbf{0},\mathbf{C}_2) \tag{19}$$

It can be shown that for the GLRT using the constructed PDF, the test statistic is

$$\max_{\boldsymbol{\theta}}\left[\boldsymbol{\theta}^T\mathbf{t} - \ln\left(\pi e^{\frac{1}{2}\boldsymbol{\theta}^T\mathbf{G}^T\mathbf{C}_1\mathbf{G}\boldsymbol{\theta}} + (1-\pi)e^{\frac{1}{2}\boldsymbol{\theta}^T\mathbf{G}^T\mathbf{C}_2\mathbf{G}\boldsymbol{\theta}}\right)\right] \tag{20}$$

and the clairvoyant GLRT statistic is

$$\begin{aligned}
-\ln\Bigg(&\frac{\pi}{\det^{1/2}(\mathbf{C}_1)}\exp\left[-\frac{1}{2}\mathbf{t}^T\left(\mathbf{G}^T\mathbf{C}_1\mathbf{G}\right)^{-1}\mathbf{t}\right]\\
&+\frac{1-\pi}{\det^{1/2}(\mathbf{C}_2)}\exp\left[-\frac{1}{2}\mathbf{t}^T\left(\mathbf{G}^T\mathbf{C}_2\mathbf{G}\right)^{-1}\mathbf{t}\right]\Bigg)
\end{aligned} \tag{21}$$

when $q \le p$.

6 Simulations

Since the GLRT using the constructed PDF coincides with the clairvoyant GLRT under Gaussian noise as shown in subsection 5.1, we will only compare the performances under non-Gaussian noise (both uncorrelated noise as in (14) and correlated noise as in (19)).

Consider the model where

$$x[n] = A_1 + A_2 r^n + A_3\cos(2\pi f n + \phi) + w[n] \tag{22}$$

for $n = 0, 1, \dots, N-1$ 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 (9) 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 (14). For the partially observed linear model, we observe two sensor outputs as in (10). We compare the GLRT in (15) with the clairvoyant GLRT in (18). Note that the MLE of $\boldsymbol{\theta}$ in (15) is found numerically, not by the asymptotic approximation in (16). 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 5.

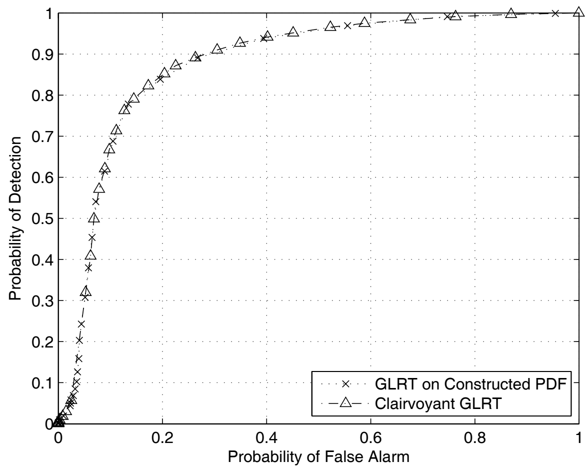

Figure 2: ROC curves for the GLRT using the constructed PDF and the clairvoyant GLRT with uncorrelated Gaussian mixture noise.

Next for the same model in (22), let $\mathbf{w}$ have a correlated Gaussian mixture distribution as in (14). We compare performances of the GLRT using the constructed PDF as in (20) and the clairvoyant GLRT as in (21). 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.

7 Conclusions

A novel method of combining sensor outputs for detection based on the exponential family has been proposed. It does not require the joint PDF under $\mathcal{H}_1$. The constructed PDF has been shown to be optimal in KL divergence. The GLRT statistic based on this method can be shown to be equivalent to the clairvoyant GLRT statistic for the partially observed linear model with both Gaussian or non-Gaussian noise. The equivalence is also shown in simulations.

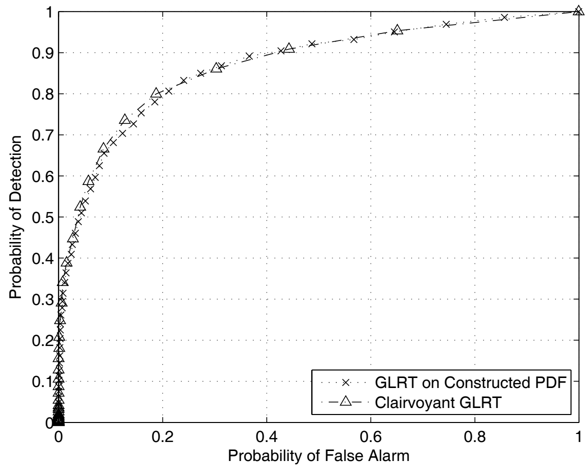

Figure 3: ROC curves for the GLRT using the constructed PDF and the clairvoyant GLRT with correlated Gaussian mixture noise.

References

[1] S.C.A. Thomopoulos, R. Viswanathan, and D.K. Bougoulias, "Optimal distributed decision fusion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 25, pp. 761–765, Sep. 1989.

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

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

[4] A. Sundaresan, P.K. Varshney, and N.S.V. 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.

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

[6] S. Kay and Q. Ding, "Exponentially embedded families for multimodal sensor processing," in *ICASSP*, 2010.

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

[8] S. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, Prentice-Hall, 1998.

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

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

[11] D.G. Luenberger, *Linear and Nonlinear Programming*, Springer, 2 edition, 2003.

[12] S. Kullback, *Information Theory and Statistics*, Courier Dover Publications, second edition, 1997.

[13] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley and Sons, second edition, 2006.

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