18th International Conference on Information Fusion
Washington, DC - July 6-9, 2015

Distributed Classification under Statistical Dependence with Application to Automatic Modulation Classification

Hao He*, Sora Choi*, Pramod K. Varshney* and Wei Su†

*Department of EECS, Syracuse University, Syracuse, NY 13244, USA
†U.S. Army Communications-Electronics RD&E Center, Aberdeen Proving Ground, MD 21005, USA

***Abstract*—In this paper, we consider the distributed classification of discrete random signals in wireless sensor networks (WSNs). Observing the same random signal makes sensors' observations conditionally dependent which complicates the design of distributed classification systems. In the literature, this dependence has been ignored for simplicity although this may significantly affect the performance of the classification system. We derive the necessary conditions for the optimal decision rules at the sensors and the fusion center (FC) by introducing a "hidden" random variable. Furthermore, we introduce an iterative algorithm to search for the optimal decision rules. The proposed scheme is applied to a distributed Automatic Modulation Classification (AMC) problem. It is shown to attain superior performance in comparison with other approaches which disregard the inter-sensor dependence.**

Keywords: distributed classification, dependent observations, wireless sensor networks, automatic modulation classification

I. INTRODUCTION

Classification using multiple sensors is generally more reliable, and has been widely studied in several engineering applications like, target recognition, and identification. Distributed processing approaches for classification are desired in wireless sensor networks (WSNs) because gathering all sensors' observations at the FC requires large communication bandwidth. In this approach, the sensors process and analyze their observed raw data and transmit only the compressed information to the FC, which then generates the final decision [1]. Distributed hypothesis testing schemes for classification have received significant attention, but most research has focused on cases where the observations at different sensors are independent [2]–[4]. However, when sensors observe the same random signals, their observations may not be independent.

The Automatic Modulation Classification (AMC) problem serves as a good example of the above scenario, where under each modulation scheme (hypothesis), the communication signal can be viewed as a discrete random variable taking values from the corresponding set of constellation symbols. AMC is a signal processing method that is used to classify the modulation scheme corresponding to the received noisy communication signals and plays a key role in various civilian and military applications. Extensive research has been done on AMC methods with a single receiver [5]–[9], whose performance depends heavily on the channel quality. Thus, network centric methods for AMC using multiple sensors have been motivated and investigated [10]–[15]. In many cases, due to the scarcity of transmission resources including channel bandwidth and local sensors' energy, distributed processing of locally sensed signals is desirable. Accordingly, local decision rules and the fusion rule in such environments are investigated in [12]–[14]. In [13], each sensor conducts a test based on the likelihood ratio of its observations, which, according to [16], is optimal only with conditionally independent sensor observations. To the best of our knowledge, no work has tackled the problem of distributed modulation classification considering conditionally dependent observations, which is the topic studied in this paper.

Some recent efforts on distributed detection with conditionally dependent observations are discussed in [17]–[20], where [20] focuses on only the fusion aspect of the problem and [19] emphasizes a very general theoretical framework. The main contribution of this paper is that we derive the optimal rules at the sensors and the FC in the Bayesian framework with dependent observations for the distributed classification problem shown in Figure 1. Our approach is based on the introduction of a new "hidden" random variable as proposed in [19], through which a hierarchical conditional independence model is built. We derive the necessary condition for optimal decision rules at the sensors and the FC, based on which, an iterative optimization algorithm is proposed. We address the implementation issue of the iterative algorithm by discretizing observation space of the local sensors. Through simulation, we are able to show that in the distributed AMC problem, our proposed method outperforms the other approaches that assume conditionally independent observations.

The paper is organized as follows. In Section II, the problem of distributed classification of discrete random signals is formulated. In Section III, the necessary conditions for the optimal sensor rules and the optimal fusion rule with dependent observations are derived. An iterative optimization algorithm is proposed in Section IV and the results of applying our proposed algorithm to a distributed AMC problem are provided in Section V. Section VI concludes this work and

discusses potential extensions.

II. PROBLEM FORMULATION

We consider a distributed hypothesis testing problem in a wireless sensor network. Suppose there are K candidate hypotheses represented by $H_0, \ldots, H_{K-1}$, with prior probabilities $\pi_0, \ldots, \pi_{K-1}$. The signal is represented by a discrete random variable s. Under H_i, the signal set is given as $\mathcal{S}_i$, i.e., $s \in \mathcal{S}_i = \{I_i^1, \ldots, I_i^{M_i}\}$, where $I_i^{m_i}$ is the m_i-th symbol. It is assumed that there is no prior information available about the probability of each symbol for a given hypothesis H_i. Thus, the probability of each symbol is considered equal, i.e.,

$$H_i : P(s = I_i^{m_i}) = \frac{1}{M_i}, \quad m_i = 1, \ldots, M_i \tag{1}$$

for $i = 0, \ldots, K-1$, where $P(\cdot)$ denotes the probability mass function (PMF). It is noted that the symbol sets under different hypotheses may overlap, i.e., $\mathcal{S}_i \cap \mathcal{S}_j \neq \emptyset, \forall i \neq j$.

We consider a general signal reception scenario with multiple sensors where the received observation of sensor l at time n can be written as:

$$r_{nl} = h_l s_n + w_{nl} \tag{2}$$

where h_l is the channel gain and $\{s_n\}_{n=1}^N$ is the discrete random signal sequence. We assume w_{nl} to be independent and identically distributed (i.i.d) noise whose distribution is known. Thus, the probability density function (PDF) of r_{nl} conditioned on the hypothesis H_i and the symbol s_n can be obtained and written as $p_i(r_{nl}|s_n)$, where $p_i(\cdot)$ denotes the PDF under hypothesis H_i.

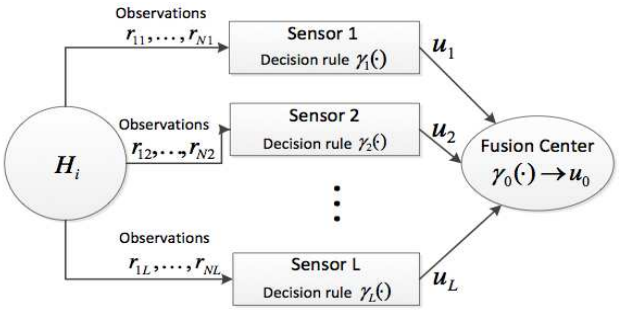

Fig. 1. Distributed modulation classification using a wireless sensor network

As shown in Figure 1, each sensor l locally processes the observation sequence $\mathbf{r}_l := [r_{1l}, \ldots, r_{Nl}]^T$ and makes a decision $u_l = \gamma_l(\mathbf{r}_l)$, where $u_l \in \mathcal{K} := \{0, \ldots, K-1\}$, where $\gamma_l(\cdot)$ is the decision rule at sensor l. When the FC receives the local decisions $\mathbf{u} := [u_1, \ldots, u_L]$ from all sensors, it makes the final decision $u_0 = \gamma_0(\mathbf{u})$, where $u_0 \in \mathcal{K}$ and $\gamma_0(\cdot)$ is the fusion rule. We assume that the true underlying hypothesis remains unchanged during the collection of N observations.

Our goal is to design the set of local decision rules and the fusion rule, i.e., $\boldsymbol{\gamma} := \{\gamma_1 \ldots, \gamma_L, \gamma_0\}$ to maximize the classification performance, in terms of probability of correct classification P_c, considering the fact that the sensors observing the same discrete random signals have dependent observations, i.e.,

$$p_i(\mathbf{r}_1, \ldots, \mathbf{r}_L) \neq \prod_{l=1}^{L} p_i(\mathbf{r}_l) \tag{3}$$

III. DESIGN OF OPTIMAL SENSOR RULES AND FUSION RULE

In the distributed classification system, the random variables involved form the following Markov chain:

$$H \rightarrow \mathbf{R} \rightarrow \mathbf{U} \rightarrow U_0 \tag{4}$$

where H represents the hypotheses, $\mathbf{R}$ is the observation matrix, $\mathbf{U}$ is the vector of sensor decisions, and U_0 is the final decision. Independence between $\mathbf{R}$ and H makes the derivation of decision rules easy by allowing factorization of $p(\mathbf{R}|H)$, but it does not hold here as discussed in the previous section. However, according to the observation model in (2), for a given symbol s_n which takes a value from the set $\mathcal{S} := \cup_{i=0}^{K-1} \mathcal{S}_i$, the variables $[R_{n1}, \ldots, R_{nL}]$ are independently distributed, for all $n = 1, \ldots, N$. Thus, we define a random vector $\mathbf{Y} := [s_1, \ldots, s_N]$ to represent the underlying symbols. And by introducing $\mathbf{Y}$ between H and $\mathbf{R}$, into the Markov chain in (4), the following results hold:

1) Given $\mathbf{Y}$, $\mathbf{R}$ is independent of H, namely, the following Markov chain is valid:

$$H \rightarrow \mathbf{Y} \rightarrow \mathbf{R} \rightarrow \mathbf{U} \rightarrow \mathbf{U}_0 \tag{5}$$

2) The introduction of random variable $\mathbf{Y}$ makes different sensors' observations conditionally independent of each other, and thus facilitates the design of sensor decision rules. In other words, $\mathbf{R}_1, \ldots, \mathbf{R}_L$ are independent conditioned on $\mathbf{Y}$, i.e.,

$$p(\mathbf{r}_1, \ldots, \mathbf{r}_L|\mathbf{Y}) = \prod_{l=1}^{L} p(\mathbf{r}_l|\mathbf{Y}). \tag{6}$$

We aim to derive the sensor decision rules and the fusion rule that maximize the probability of correct classification P_c

$$P_c = \sum_{u_0=0}^{K-1} \sum_{i=0}^{K-1} c_{u_0,i} p_i(u_0) \pi_i \tag{7}$$

where, $c_{u_0,i}$ indicates the cost of deciding u_0 when the true hypothesis is H_i. We consider the particular cost assignment where $c_{u_0,i} = 1$ if and only if $u_0 = i$, otherwise, $c_{u_0,i} = 0$, which means that a correct classification occurs when the decision at the FC matches with the true hypothesis. Thus, P_c can be further simplified as

$$P_c = \sum_{u_0=0}^{K-1} p(u_0|H_{u_0}) \pi_{u_0} \tag{8}$$

We can express P_c with respect to sensor l as follows:

$$\begin{aligned} P_c &= \int_{\mathbf{R}} \sum_{\mathbf{u}} \sum_{u_0=0}^{K-1} \pi_{u_0} p(u_0|\mathbf{u}) p(\mathbf{u}|\mathbf{r}) p_{u_0}(\mathbf{r}) d\mathbf{r} \\ &= \int_{\mathbf{R}_l} \sum_{u_l} p(u_l|\mathbf{r}_l) f_l(u_l, \mathbf{r}_l) d\mathbf{r}_l \end{aligned} \quad (9)$$

where $\mathbf{r} := [\mathbf{r}_1, \ldots, \mathbf{r}_L]$ and

$$f_l(u_l, \mathbf{r}_l) = \sum_{\mathbf{u}^l} \sum_{u_0=0}^{K-1} \pi_{u_0} p(u_0|\mathbf{u}^l, u_l) \int_{\mathbf{R}^l} p(\mathbf{u}^l|\mathbf{r}^l) p_{u_0}(\mathbf{r}) d\mathbf{r}^l \quad (10)$$

where a vector (matrix) with a superscript denotes it without the lth element (column), for example, $\mathbf{R}^l$ represents the vector $\mathbf{R} \setminus R_l = [R_1, \ldots, R_{l-1}, R_{l+1}, \ldots, R_L]$.

To maximize P_c, the optimal decision rule at sensor l given fixed decision rules at all the other sensors and the FC is to make a decision u_l such that $f_l(u_l, \mathbf{r}_l)$ is maximized, namely

$$\gamma_l(\mathbf{r}_l) = \arg\max_{u_l} f_l(u_l, \mathbf{r}_l) \quad (11)$$

for all $\mathbf{r}_l$. This is because for given $\mathbf{r}_l$, $p(u_l|\mathbf{r}_l) = 1$ only for $u_l = \gamma_l(\mathbf{r}_l)$, otherwise $p(u_l|\mathbf{r}_l) = 0$, thus the decision rule in (11) maximizes the probability of making a correct decision when $\mathbf{r}_l$ is observed.

With the introduction of the hidden random vector $\mathbf{Y}$, the joint distribution of all sensors' observations $\mathbf{r}$ conditioned on hypothesis H_i, i.e., $p_i(\mathbf{r})$ ($p_{u_0}(\mathbf{r})$ in (10)), can be written as follows:

$$\begin{aligned} p_i(\mathbf{r}) &= \sum_{\mathbf{Y}} P(\mathbf{Y} = \mathbf{y}|H_i) \prod_{l=1}^{L} p(\mathbf{r}_l|\mathbf{Y} = \mathbf{y}) \\ &= \sum_{\mathbf{Y}} \left(\frac{1}{M_i}\right)^N \prod_{l=1}^{L} p(\mathbf{r}_l|\mathbf{Y} = \mathbf{y}) \end{aligned} \quad (12)$$

where $\sum_{\mathbf{Y}} = \sum_{\mathbf{y} \in \mathcal{S}^N}$. By combining (12) and (10), $f_l(u_l, \mathbf{r}_l)$ can be simplified as follows:

$$f_l(u_l, \mathbf{r}_l) = \sum_{\mathbf{Y}} \beta_l(u_l, \mathbf{y}) p(\mathbf{r}_l|\mathbf{y}) \quad (13)$$

where

$$\begin{aligned} \beta_l(u_l, \mathbf{y}) &= \sum_{\mathbf{u}^l} \sum_{u_0=0}^{K-1} \pi_{u_0} p_{u_0}(\mathbf{y}) p(u_0|\mathbf{u}^l, u_l) \\ &\quad \times \int_{\mathbf{R}^l} p(\mathbf{u}^l|\mathbf{r}^l) p(\mathbf{r}^l|\mathbf{y}) d\mathbf{r}^l \\ &= \sum_{\mathbf{u}^l} \sum_{u_0=0}^{K-1} \pi_{u_0} p_{u_0}(\mathbf{y}) p(u_0|\mathbf{u}^l, u_l) \\ &\quad \times \prod_{h=1, h\neq l}^{L} \int_{\mathbf{R}_h} p(u_h|\mathbf{r}_h) p(\mathbf{r}_h|\mathbf{y}) d\mathbf{r}_h \end{aligned} \quad (14)$$

is a scalar function of u_l and $\mathbf{y}$. Thus, the optimal sensor rule γ_l is

$$\gamma_l(\mathbf{r}_l) = \arg\max_{u_l} \sum_{\mathbf{Y}} \beta_l(u_l, \mathbf{y}) p(\mathbf{r}_l|\mathbf{y}). \quad (15)$$

For a binary hypothesis testing problem, i.e, $K = 2$, the optimal decision rule in (15) can be written as follows:

$$\gamma_l(\mathbf{r}_l) = \mathbb{1}\left\{ \sum_{\mathbf{Y}} [\beta_l(1, \mathbf{y}) - \beta_l(0, \mathbf{y})] p(\mathbf{r}_l|\mathbf{y}) \right\} \quad (16)$$

where $\mathbb{1}\{\cdot\}$ is defined as follows:

$$\mathbb{1}\{x\} = \begin{cases} 1, & x \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (17)$$

Next, the optimal fusion rule at the FC that maximizes the probability of correct classification P_c is investigated. We have

$$\begin{aligned} P_c &= \sum_{u_0=0}^{K-1} p(u_0|H_{u_0}) \pi_{u_0} \\ &= \sum_{\mathbf{u}} \sum_{u_0=0}^{K-1} p(u_0|\mathbf{u}) p(\mathbf{u}|H_{u_0}) \pi_{u_0} \end{aligned} \quad (18)$$

To maximize P_c, the optimal fusion rule is to make a decision u_0 such that such $\pi_{u_0} p(\mathbf{u}|H_{u_0})$ is maximized, namely

$$\begin{aligned} \gamma_0(\mathbf{u}) &= \arg\max_{u_0} \pi_{u_0} p(\mathbf{u}|H_{u_0}) \\ &= \arg\max_{u_0} \int_{\mathbf{R}} \prod_{l=1}^{L} p(u_l|\mathbf{r}_l) p_{u_0}(\mathbf{r}) \pi_{u_0} d\mathbf{r} \end{aligned} \quad (19)$$

for any local decision vector $\mathbf{u}$. Because for a given $\mathbf{u}$, $p(u_0|\mathbf{u})$ takes the value either 0 or 1, the fusion rule in (19) maximizes P_c given in (18). For a binary hypothesis testing problem, the following fusion rule can be obtained:

$$\gamma_0(\mathbf{u}) = \mathbb{1}\left\{ \pi_1 p_1(\mathbf{u}) - \pi_0 p_0(\mathbf{u}) \right\} \quad (20)$$

Proposition 1: Let $\{\gamma_1(\mathbf{r}_1), \ldots, \gamma_L(\mathbf{r}_L)\}$ and $\gamma_0(\mathbf{u})$ be a set of optimal sensor decision rules and an optimal fusion rule in a distributed classification system that maximizes the probability of correct classification P_c. Then they must satisfy the following conditions:

1) For all local decision rules $\gamma_l, l = 1, \ldots, L$:

$$\gamma_l(\mathbf{r}_l) = \arg\max_{u_l} \sum_{\mathbf{Y}} \beta_l(u_l, \mathbf{y}|\boldsymbol{\gamma}^l, \gamma_0) p(\mathbf{r}_l|\mathbf{y}) \quad (21)$$

where $\beta_l(\cdot)$ is given as follows:

$$\begin{aligned} &\beta_l(u_l, \mathbf{y}|\boldsymbol{\gamma}^l, \gamma_0) \\ &= \sum_{\mathbf{u}^l} \sum_{u_0=0}^{K-1} \pi_{u_0} p_{u_0}(\mathbf{y}) \delta\left(\gamma_0(\mathbf{u}^l, u_l) - u_0\right) \\ &\quad \times \prod_{h=1, h\neq l}^{L} \int_{\mathbf{R}_h} \delta\left(\gamma_h(\mathbf{r}_h) - u_h\right) p(\mathbf{r}_h|\mathbf{y}) d\mathbf{r}_h \end{aligned} \quad (22)$$

where the notation of $\beta(\cdot|\boldsymbol{\gamma}^l, \gamma_0)$ is to emphasize that the value of β is conditioned on the given decision rules $\boldsymbol{\gamma}^l, \gamma_0$, and $\delta(\cdot)$ is defined as follows:

$$\delta(x) = \begin{cases} 1, & x = 0 \\ 0, & \text{otherwise.} \end{cases} \quad (23)$$

2) For the fusion rule:

$$\gamma_0(\mathbf{u}) = \arg\max_{u_0} \int_{\mathbf{R}} \prod_{l=1}^{L} \delta\left(\gamma_l(\mathbf{r}_l) - u_l\right) p_{u_0}(\mathbf{r}) \pi_{u_0} d\mathbf{r}. \quad (24)$$

The necessary conditions to determine the optimal rules that maximize the probability of correct classification naturally is obtained. To search for the optimal rules, we adapt the idea of Gauss-Seidel iterative algorithm. We present a computationally efficient iterative algorithm for obtaining discrete approximations of the optimal rules in the next section.

IV. Computational Algorithm

We propose an iterative algorithm based on Proposition 1, by considering the following Gauss-Seidel iterative process [21]. Let the sensor decision rules and the fusion rule at the kth stage of iteration be denoted by $(\gamma_1^{(k)}, \ldots, \gamma_L^{(k)}, \gamma_0^{(k)})$ with the initial set $(\gamma_1^{(0)}, \ldots, \gamma_L^{(0)}, \gamma_0^{(0)})$. At the $k+1$th iteration, after the decision rule of sensor $l-1$, i.e. $\gamma_{l-1}^{(k+1)}$, is updated, the decision rule of sensor l is updated according to the following equation

$$\begin{aligned}\gamma_l^{(k+1)}(\mathbf{r}_l) = \\ \arg\max_{u_l} \sum_{\mathbf{Y}} \beta_l(u_l, \mathbf{y} | \gamma_1^{(k+1)}, \ldots, \gamma_{l-1}^{(k+1)}, \\ \gamma_{l+1}^{(k)}, \ldots, \gamma_L^{(k)}, \gamma_0^{(k)}) p(\mathbf{r}_l | \mathbf{y}). \end{aligned} \quad (25)$$

Once every sensor's decision rule is updated, the fusion rule is obtained by

$$\begin{aligned}\gamma_0^{(k+1)}(\mathbf{u}) = \\ \arg\max_{u_0} \int_{\mathbf{R}} \prod_{l=1}^{L} \delta\left(\gamma_l^{(k+1)}(\mathbf{r}_l) - u_l\right) p_{u_0}(\mathbf{r}) \pi_{u_0} d\mathbf{r}. \end{aligned} \quad (26)$$

This algorithm involves obtaining sensor rules $(\gamma_1^{(k)}, \ldots, \gamma_L^{(k)})$ that are continuous functions. Thus, discretizing the input space, and thus the continuous functions, is necessary for obtaining a solution in practice. For illustration purposes, we present the algorithm with each sensor making its decision based on the observation at a single time instant, i.e., $N = 1$. And for notational simplicity, x_{nl} will be written as x_l by omitting the time index n. The corresponding hidden random variable Y is a scalar variable in this case. We define these functions on equally discretized grids of $\{r_{1,1}, \ldots, r_{1,t_1}, \ldots, r_{1,T_1}\}$, $\{r_{2,1}, \ldots, r_{2,t_2}, \ldots, r_{2,T_2}\}$, $\ldots$, $\{r_{L,1}, \ldots, r_{L,t_L}, \ldots, r_{L,T_L}\}$ with Δ_l being the discretization step size of r_l. The following discretized Gauss-Seidel iterative algorithm is obtained:

Step 1: Initialize L sensor rules and the fusion rule respectively, for $l = 1, \ldots, L$.

$$\gamma_l^{(0)}(r_{l,t_l}) = i \in \mathcal{K} \quad \forall t_l = 1, \ldots, T_l \quad (27)$$

$$\gamma_0^{(0)}(\mathbf{u}) = i \in \mathcal{K} \quad \forall \mathbf{u} \in \mathcal{K}^L \quad (28)$$

Step 2: Iteratively, update L sensor rules and the fusion rule for better system performance. The $(k+1)$th stage of the iteration is as follows:

For $t_1 = 1, \ldots, T_1$,

$$\begin{aligned}\gamma_1^{(k+1)}(r_{1,t_1}) = \\ \arg\max_{u_1} \sum_{Y} \beta_1(u_1, y | \gamma_2^{(k)}, \ldots, \gamma_L^{(k)}, \gamma_0^{(k)}) \\ \times p(r_{1,t_1} | y) \end{aligned} \quad (29)$$

with

$$\begin{aligned}\beta_1(u_1, y | \gamma_2^{(k)}, \ldots, \gamma_L^{(k)}, \gamma_0^{(k)}) = \\ \sum_{\mathbf{u}^1} \sum_{u_0=0}^{K-1} \pi_{u_0} p_{u_0}(y) \delta\left(\gamma_0^{(k)}(\mathbf{u}^1, u_1) - u_0\right) \\ \times \prod_{l=2}^{L} \left[\sum_{t_l=1}^{T_l} \delta\left(\gamma_l^{(k)}(r_{l,t_l}) - u_l\right) p(r_{l,t_l} | y) \Delta r_l\right] \end{aligned} \quad (30)$$

For $t_2 = 1, \ldots, T_2$,

$$\begin{aligned}\gamma_2^{(k+1)}(r_{2,t_2}) = \\ \arg\max_{u_2} \sum_{Y} \beta_2(u_2, y | \gamma_1^{(k+1)}, \gamma_3^{(k)}, \ldots, \gamma_L^{(k)}, \gamma_0^{(k)}) \\ \times p(r_{2,t_2} | y) \end{aligned} \quad (31)$$

with

$$\begin{aligned}\beta_2(u_2, y | \gamma_1^{(k+1)}, \gamma_3^{(k)}, \ldots, \gamma_L^{(k)}, \gamma_0^{(k)}) = \\ \sum_{\mathbf{u}^2} \sum_{u_0=0}^{K-1} \pi_{u_0} p_{u_0}(y) \delta\left(\gamma_0^{(k)}(\mathbf{u}^2, u_2) - u_0\right) \\ \times \left[\sum_{t_1=1}^{T_1} \delta\left(\gamma_1^{(k+1)}(r_{1,t_1}) - u_1\right) p(r_{1,t_1} | y) \Delta r_1\right] \\ \times \prod_{l=3}^{L} \left[\sum_{t_l=1}^{T_l} \delta\left(\gamma_l^{(k)}(r_{l,t_l}) - u_l\right) p(r_{l,t_l} | y) \Delta r_l\right] \end{aligned} \quad (32)$$

$\ldots\ldots$

For $t_L = 1, \ldots, T_L$,

$$\begin{aligned}\gamma_L^{(k+1)}(r_{L,t_L}) = \\ \arg\max_{u_L} \sum_{Y} \beta_L(u_L, y | \gamma_1^{(k+1)}, \ldots, \gamma_{L-1}^{(k+1)}, \gamma_0^{(k)}) \\ \times p(r_{L,t_L} | y) \end{aligned} \quad (33)$$

with

$$\begin{aligned}\beta_L(u_L, y | \gamma_1^{(k+1)}, \ldots, \gamma_{L-1}^{(k+1)}, \gamma_0^{(k)}) = \\ \sum_{\mathbf{u}^L} \sum_{u_0=0}^{K-1} \pi_{u_0} p_{u_0}(y) \delta\left(\gamma_0^{(k)}(\mathbf{u}^L, u_L) - u_0\right) \\ \times \prod_{l=1}^{L-1} \left[\sum_{t_l=1}^{T_l} \delta\left(\gamma_l^{(k+1)}(r_{l,t_l}) - u_l\right) p(r_{l,t_l} | y) \Delta r_l\right] \end{aligned} \quad (34)$$

For $\mathbf{u} \in \mathcal{K}^L$,

$$\gamma_0^{(k+1)}(\mathbf{u}) = \arg\max_{u_0} \sum_{t_1}^{T_1} \cdots \sum_{t_L}^{T_L} \prod_{l=1}^{L} \delta\left(\gamma_l^{(k+1)}(r_{l,t_l}) - u_l\right) \times p_{u_0}(r_{1,t_1}, \ldots, r_{L,t_L})\pi_{u_0} \quad (35)$$

Step 3: The local decision rules and the fusion rule are updated iteratively until the following termination criterion is satisfied:

$$\begin{aligned} &\gamma_1^{(k+1)}(r_{1,t_1}) = \gamma_1^{(k)}(r_{1,t_1}), \\ &\gamma_2^{(k+1)}(r_{2,t_2}) = \gamma_2^{(k)}(r_{2,t_2}), \ldots, \\ &\gamma_L^{(k+1)}(r_{L,t_L}) = \gamma_L^{(k)}(r_{L,t_L}); \end{aligned} \quad (36)$$

$\forall t_1, \ldots, t_L$. And for all $\mathbf{u} \in \mathcal{K}^L$

$$\gamma_0^{(k+1)}(\mathbf{u}) = \gamma_0^{(k)}(\mathbf{u}) \quad (37)$$

The above algorithm terminates, since the search is done on a discretized space. After these decision rules $\gamma_0, \gamma_1, \ldots, \gamma_L$ are obtained, they are employed to make local decisions and final decision are made.

V. Example

In this section, we apply our proposed approach to a distributed AMC problem. For the candidate hypothesis being a M-PSK signal, the constellation symbol set is given as $\mathcal{S}_i = \{e^{j2\pi m_i/M_i} | m_i = 0, \ldots, M_i - 1\}$ while for the candidate hypothesis being a M-QAM signal, the constellation symbol set is $\mathcal{S}_i = \{b_{m_i} e^{j\theta_{m_i}} | m_i = 1, \ldots, M_i\}$, where b_{m_i} is the amplitude of the m_i-th symbol. Binary hypothesis testing is considered in our example, i.e., $K = 2$.

We assume that the wireless channel between the unknown transmitter and each sensor undergoes flat block fading, i.e., the channel impulse response is

$$h(t) = \alpha e^{j\theta} \delta(t)$$

where α and θ are the channel (or the signal) gain and the channel (or the signal) phase, respectively, which are assumed known in this work. We assume the observation noise w_{nl} to be independent and identically distributed (i.i.d) circularly symmetric complex Gaussian with real and imaginary parts of variance $N_0/2$, i.e., $w_{nl} \sim \mathcal{CN}(0, N_0)$. The PDF of r_{nl} conditioned on the modulation format i and the symbol s_n can be written as

$$p_i(r_{nl}|s_n) = \frac{1}{\pi N_0} \exp\left(-\frac{1}{N_0} |r_{nl} - \alpha e^{j\theta} s_n|^2\right) \quad (38)$$

To evaluate the performance of the decision rules obtained from the iterative algorithm for distributed AMC, denoted by IADA, we consider several binary modulation classification problems in this section. We first consider a WSN consisting of two receiving sensors ($L = 2$), for each of which, the length of the local decision window is $N = 1$.

In the initialization step, local sensor decisions $\gamma_l(r_{l,t_l})$ are chosen randomly and the fusion rule is the majority voting rule. Through multiple experiments, we are able to observe that different initializations of the local rules eventually lead to the same set of decision rules, while different initializations of the fusion rule results in different outcomes. Thus, multiple initializations of the fusion rule are needed for comparison purposes and majority voting rule leads to the best performance in our case. Then, sensor decision rules and the fusion rule are updated iteratively until the termination conditions in (36), (37) are satisfied. After the rules are obtained, $100,000$ Monte Carlo trials are conducted to test the performance of these rules, in terms of the probability of correct classification. Following the above procedure, the performance of the algorithm in Section IV, is obtained for different signal to noise ratio (SNR) values.

In the first simulation experiment, the two candidate hypotheses are both PSK and equal priors are assumed, i.e., $\pi_0 = \pi_1 = 0.5$. Many sets of simulation have been conducted for different combinations of candidate hypotheses. Due to the similarity of the results, the performance of IADA is given for only two sets of simulations (BPSK vs. QPSK and QPSK vs. 8PSK) in Figure 2, where it is compared with the likelihood ratio based method (LRBM) derived under the independence assumption [13]. Our proposed algorithm achieves a much better performance in distinguishing two different PSKs under any SNR.

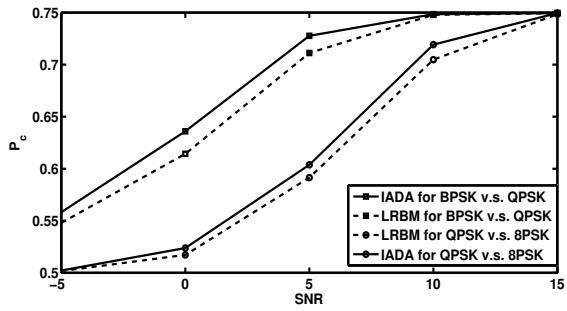

Fig. 2. P_c vs. SNR for testing between two PSK modulation schemes

In the second simulation experiment, we test a PSK signal against a QAM signal, and the priors are set to $\pi_0 = 0.3, \pi_1 = 0.7$. The performance of testing 16QAM against BPSK and testing 16QAM against QPSK is shown in Figure 3, demonstrating the superiority of the decision rules obtained by our algorithm compared to the independence-assumption based method.

The previous simulations about binary modulation classification were conducted for a two-sensor network. We further test our proposed sensor rules and fusion rule in networks with multiple sensors ($L \geq 2$). As is shown in Figure 4, the classification performance improves with the increase in the size of sensor network.

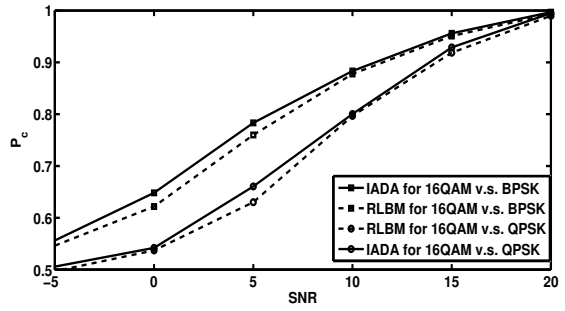

Fig. 3. P_c vs. SNR for testing a PSK against a QAM

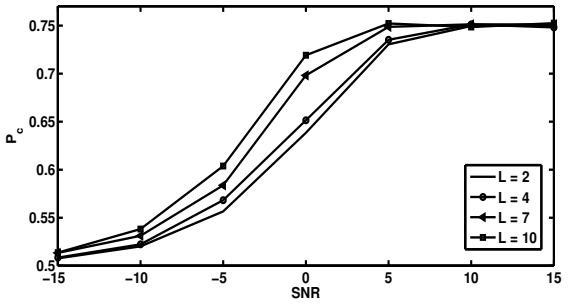

Fig. 4. P_c vs. SNR for testing BPSK against QPSK in sensor networks with different sizes

VI. CONCLUSION

In this work, we studied the problem of distributed classification of random signals in sensor networks. With the introduction of a "hidden" random variable, we successfully derived the necessary conditions for the optimal sensor compression rules and fusion rule under conditionally dependent observations. An iterative algorithm, which is easy to implement, was proposed to generate the sensor rules and the fusion rule. We applied the method that we proposed in this work to a modulation classification problem and it exhibits better performance than the other approaches, as shown in the numerical results. The convergence of our proposed algorithm will be studied in the future work.

REFERENCES

[1] P. K. Varshney, *Distributed detection and data fusion*. Springer-Verlag New York, Inc., 1996.

[2] T.-Y. Wang, Y. Han, B. Chen, and P. Varshney, "A combined decision fusion and channel coding scheme for distributed fault-tolerant classification in wireless sensor networks," *Wireless Communications, IEEE Trans. on*, vol. 5, no. 7, pp. 1695–1705, July 2006.

[3] A. D'Costa, V. Ramachandran, and A. Sayeed, "Distributed classification of Gaussian space-time sources in wireless sensor networks," *Selected Areas in Communications, IEEE Journal on*, vol. 22, no. 6, pp. 1026–1036, August 2004.

[4] S. Gutta and Q. Cheng, "Data-based distributed classification and its performance analysis," in *Information Fusion (FUSION), 2012 15th International Conference on*, July 2012, pp. 1519–1526.

[5] O. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: classical approaches and new trends," *Communications, IET*, vol. 1, no. 2, pp. 137–156, April 2007.

[6] J. Sills, "Maximum-likelihood modulation classification for PSK/QAM," in *Military Communications Conference Proceedings, MILCOM IEEE*, vol. 1, 1999, pp. 217–220.

[7] P. Panagiotou, A. Anastasopoulos, and A. Polydoros, "Likelihood ratio tests for modulation classification," in *MILCOM 2000. 21st Century Military Communications Conference Proceedings*, vol. 2, 2000, pp. 670–674.

[8] O. Dobre and F. Hameed, "Likelihood-based algorithms for linear digital modulation classification in fading channels," in *Electrical and Computer Engineering, 2006. CCECE '06. Canadian Conference on*, May 2006, pp. 1347–1350.

[9] F. Hameed, O. Dobre, and D. Popescu, "On the likelihood-based approach to modulation classification," *Wireless Communications, IEEE Trans. on*, vol. 8, no. 12, pp. 5884–5892, December 2009.

[10] O. Ozdemir, R. Li, and P. Varshney, "Hybrid maximum likelihood modulation classification using multiple radios," *Communications Letters, IEEE*, vol. 17, no. 10, pp. 1889–1892, October 2013.

[11] O. Ozdemir, P. K. Varshney, W. Su, and A. L. Drozd, "Asymptotic properties of likelihood based linear modulation classification systems," *arXiv preprint arXiv:1211.6631*, 2012.

[12] W. Su and J. Kosinski, "Framework of network centric signal sensing for automatic modulation classification," in *Networking, Sensing and Control (ICNSC), 2010 International Conference on*, April 2010, pp. 534–539.

[13] J. L. Xu, W. Su, and M. Zhou, "Distributed automatic modulation classification with multiple sensors," *Sensors Journal*, vol. 10, no. 11, pp. 1779–1785, 2010.

[14] Y. Zhang, N. Ansari, and W. Su, "Optimal decision fusion based automatic modulation classification by using wireless sensor networks in multipath fading channel," in *Global Telecommunications Conference (GLOBECOM)*. IEEE, 2011, pp. 1–5.

[15] P. A. Forero, A. Cano, and G. B. Giannakis, "Distributed feature-based modulation classification using wireless sensor networks," in *Military Communications Conference, MILCOM*. IEEE, 2008, pp. 1–7.

[16] I. Hoballah and P. Varshney, "Distributed Bayesian signal detection," *Information Theory, IEEE Trans. on*, vol. 35, no. 5, pp. 995–1000, September 1989.

[17] Y. Zhu, R. Blum, Z.-Q. Luo, and K. M. Wong, "Unexpected properties and optimum-distributed sensor detectors for dependent observation cases," *Automatic Control, IEEE Trans. on*, vol. 45, no. 1, pp. 62–72, January 2000.

[18] X. Shen, Y. Zhu, L. He, and Z. You, "A near-optimal iterative algorithm via alternately optimizing sensor and fusion rules in distributed decision systems," *Aerospace and Electronic Systems, IEEE Trans. on*, vol. 47, no. 4, pp. 2514–2529, October 2011.

[19] H. Chen, B. Chen, and P. Varshney, "A new framework for distributed detection with conditionally dependent observations," *Signal Processing, IEEE Trans. on*, vol. 60, no. 3, pp. 1409–1419, 2012.

[20] S. G. Iyengar, R. Niu., and P. K. Varshney, "Fusing dependent decisions for hypothesis testing with heterogeneous sensors," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4888 – 4897, September 2012.

[21] Z.-B. Tang, K. Pattipati, and D. Kleinman, "An algorithm for determining the decision thresholds in a distributed detection problem," *Systems, Man and Cybernetics, IEEE Trans. on*, vol. 21, no. 1, pp. 231–237, January 1991.