

Modulation Classification for MIMO-OFDM Signals via Gibbs Sampling

Yu Liu*, Osvaldo Simeone*, Alexander M. Haimovich*, and Wei Su†

*CWCSR, New Jersey Institute of Technology, Newark, 07102, New Jersey, USA

†U.S. Army Communication-Electronics Research Development and Engineering Center, I2WD,
Aberdeen Proving Ground, MD 21005, USA

Email: {yl227, osvaldo.simeone, haimovic}@njit.edu, wei.su@ieee.org

Abstract—The problem of modulation classification for a multiple-antenna (MIMO) system employing orthogonal frequency division multiplexing (OFDM) is investigated under the assumptions of unknown frequency-selective fading channels and signal-to-noise ratio (SNR). The classification problem is formulated as a Bayesian inference task and a solution is proposed based on a selection of the prior distributions that adopts a latent Dirichlet model for the modulation type and on the Bayesian network formalism. The proposed Gibbs sampling method converges to the optimal Bayesian solution and the speed of convergence is shown to improve via annealing and random restarts. While most of the existing modulation classification techniques works under the assumptions that the channels are flat fading and that a large amount of observed data symbols is available, the proposed approach performs well under more general conditions. Finally, the proposed Bayesian method is demonstrated to improve over existing non-Bayesian approaches based on independent component analysis.

Index Terms—Modulation classification; MIMO-OFDM; Gibbs sampling.

I. INTRODUCTION

A major task in cognitive radios [1] is the classification of the modulation format of unknown received signals. Modulation classification methods are generally classified as inference-based or pattern recognition-based [1]. The inference-based approaches fall into two categories, namely Bayesian and non-Bayesian methods [2]. Bayesian approaches model unknown parameters as random variables, which are assigned some prior distributions, and aim at evaluating the posterior probability of the modulation type. Non-Bayesian approaches, instead, model unknown parameters as nuisance variables that need to be estimated before performing modulation classification. With pattern recognition-based methods, specific features are extracted from the received signal and then used to discriminate among the candidate modulations. Compared to the pattern recognition-based approaches, inference-based methods generally achieve better classification performance at the cost of a higher computational complexity [1].

Many classification algorithms have been developed for single-antenna (SISO) systems [1]-[4], while only few publications address multiple-antenna (MIMO) systems [5]-[7]. In [5], a non-Bayesian inference-based approach, referred to as ICA-PC, is proposed, whereby the channel matrix required for the calculation of the hypotheses test is estimated blindly

by independent component analysis. Several related pattern recognition-based algorithms are introduced in [6]-[7]. As for MIMO systems employing orthogonal frequency division multiplexing (OFDM), a non-Bayesian approach is proposed in [8] based on ICA-PC that assumes the invariance of the frequency-domain channels across the coherence bandwidth.

In this work, we develop a Bayesian modulation classification technique for MIMO-OFDM systems operating over frequency-selective fading channels, assuming unknown channels and signal-to-noise ratio (SNR). The proposed method adopt the latent Dirichlet BN introduced in [4] for the selection of the prior distribution in SISO systems. The adopted BN model enables the application of Gibbs sampling techniques, while avoiding the convergence issues associated with the presence of zeros in the joint distribution (see [3], [4]). Based on this model, a Bayesian solution is developed based on Gibbs sampling [9]. Specifically, the proposed Gibbs sampling method converges to the optimal Bayesian solution and its speed of convergence is generally improved by multiple random restarts and annealing [10], [11]. While the reviewed existing modulation classification algorithms for MIMO-OFDM systems work under the assumptions that the channels are flat fading [5]-[7], and/or that the number of samples is large (as for pattern recognition-based methods) [6], [7], the proposed method achieves satisfactory performance under more general conditions.

Notation: The superscripts T and H denote matrix or vector transpose and Hermitian, respectively. Lower case bold letters and upper case bold letters are used to denote column vectors and matrices, respectively. The notation $\mathbf{b} \setminus b_i$, where $\mathbf{b} = [b_1, \dots, b_n]^T$ and $i \in \{1, \dots, n\}$, denotes the vector composed of all the elements of \mathbf{b} except b_i . We use an angle bracket $\langle \cdot \rangle$, to represent the expectation with respect to the random variables indicated in the subscript. The notations $\mathbf{1}(\cdot)$ stand for the indicator function. The cardinality of a set \mathcal{B} is denoted $|\mathcal{B}|$. The notations $\mathcal{CN}(\boldsymbol{\mu}, \mathbf{C})$ and $\mathcal{IG}(a, b)$ represent the the circularly symmetric complex Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix \mathbf{C} and the inverse gamma distribution with shape parameter a and scale parameter b , respectively.

II. SYSTEM MODEL

Consider a MIMO-OFDM system operating over a frequency-selective fading channel with N subcarriers, M_t transmit antennas, M_r receive antennas and a coherence period of K OFDM symbols. All frequency-domain transmitted symbols during the coherence period are taken from a finite constellation $A \in \mathcal{A}$, such as M -PSK or M -QAM, where \mathcal{A} is the (finite) set containing all possible constellations. We focus on the problem of detecting the constellation A in the absence of information about the signal-to-noise ratio (SNR), the transmitted symbols and the fading channel coefficients.

After matched filtering and sampling, assuming that time synchronization has been successfully performed at least within the error margin afforded by the cyclic prefix, the frequency-domain received samples $\mathbf{y}[n, k] = [y_1[n, k], \dots, y_{M_r}[n, k]]^T$, across the M_r receive antennas at the n -th subcarrier of the k -th OFDM frame, can be expressed as

$$\mathbf{y}[n, k] = \mathbf{H}[n]\mathbf{s}[n, k] + \mathbf{z}[n, k], \quad (1)$$

where $\mathbf{H}[n]$ is the $M_r \times M_t$ frequency-domain channel matrix associated with the n -th subcarrier; $\mathbf{s}[n, k]$ is the $M_t \times 1$ vector composed of the symbols transmitted by the M_t antennas, i.e., $\mathbf{s}[n, k] = [s_1[n, k], \dots, s_{M_t}[n, k]]^T$ with $s_{m_t}[n, k] \in A$ being the symbol transmitted by the m_t -th transmit antenna over the n -th subcarrier of the k -th OFDM symbol; and $\mathbf{z}[n, k] = [z_1[n, k], \dots, z_{M_r}[n, k]]^T \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$ is complex white Gaussian noise, which is independent over indices n and k . The frequency-domain channel matrix $\mathbf{H}[n]$ can be written as

$$\mathbf{H}[n] = \begin{bmatrix} \tilde{h}_{1,1}[n] & \cdots & \tilde{h}_{M_t,1}[n] \\ \vdots & \ddots & \vdots \\ \tilde{h}_{1,M_r}[n] & \cdots & \tilde{h}_{M_t,M_r}[n] \end{bmatrix}, \quad (2)$$

where $\tilde{\mathbf{h}}_{m_t, m_r} = [\tilde{h}_{m_t, m_r}[1], \dots, \tilde{h}_{m_t, m_r}[N]]^T$ denotes the $N \times 1$ frequency-domain channel vector between the m_t -th transmit antenna and the m_r -th receive antenna. Assuming that the channels for every pair (m_t, m_r) have at most L symbol spaced taps, we write $\tilde{\mathbf{h}}_{m_t, m_r} = \mathbf{W}\mathbf{h}_{m_t, m_r}$, with \mathbf{h}_{m_t, m_r} being the $L \times 1$ time-domain channel vector and \mathbf{W} being the $N \times L$ matrix composed of the first L columns of the DFT matrix of size N . Note that the channel is a constant within the coherence frame of K OFDM symbols.

According to (1) and (2), the $NK \times 1$ received frequency-domain signals $\mathbf{y}_{m_r} = [\mathbf{y}_{m_r}[1]^T, \dots, \mathbf{y}_{m_r}[K]^T]^T$ at the m_r -th receive antenna is given by

$$\mathbf{y}_{m_r} = \sum_{m_t=1}^{M_t} \mathbf{D}_{m_t} \tilde{\mathbf{h}}_{m_t, m_r} + \mathbf{z}_{m_r}, \quad m_r = 1, \dots, M_r, \quad (3)$$

where $\mathbf{y}_{m_r}[k] = [y_{m_r}[1, k], \dots, y_{m_r}[N, k]]^T$; $\mathbf{D}_{m_t} = [\mathbf{D}_{m_t, 1}, \dots, \mathbf{D}_{m_t, K}]^T$ is an $NK \times N$ matrix representing the transmitted symbols with $\mathbf{D}_{m_t, k}$ being an $N \times N$ diagonal matrix whose (n, n) element is $s_{m_t}[n, k]$; and $\mathbf{z}_{m_r} = [\mathbf{z}_{m_r}[1]^T, \dots, \mathbf{z}_{m_r}[K]^T]^T$ with $\mathbf{z}_{m_r}[k] = [z_{m_r}[1, k], \dots, z_{m_r}[N, k]]^T$.

Let us further define the $NKM_t \times 1$ vector $\mathbf{s} = [\mathbf{s}_1, \dots, \mathbf{s}_K]^T$ containing all the transmitted symbols with $\mathbf{s}_k = [s[1, k]^T, \dots, s[N, k]^T]^T$; the $LM_tM_r \times 1$ vector $\mathbf{h} = [\mathbf{h}_1^T, \dots, \mathbf{h}_{M_r}^T]^T$ for the time domain channels associated with all the transmit-receive antenna pairs, where $\mathbf{h}_{m_r} = [\mathbf{h}_{1, m_r}^T, \dots, \mathbf{h}_{M_t, m_r}^T]^T$; and the $NKM_r \times 1$ receive signal vector $\mathbf{y} = [\mathbf{y}_1^T, \dots, \mathbf{y}_{M_r}^T]^T$. The task of modulation classification is for the receiver to correctly detect the modulation format A given only the received samples \mathbf{y} , while being uninformed about the symbols \mathbf{s} , the channel \mathbf{h} and the noise power σ^2 . Using (1) and (3), the likelihood function $p(\mathbf{y}|A, \mathbf{s}, \mathbf{h}, \sigma^2)$ of the observation is given by

$$\begin{aligned} & p(\mathbf{y}|A, \mathbf{s}, \mathbf{h}, \sigma^2) \\ &= \prod_{n, k} p(\mathbf{y}[n, k] | \mathbf{s}[n, k], \mathbf{H}[n], \sigma^2) \\ &= \prod_{m_r} p(\mathbf{y}_{m_r} | \mathbf{s}, \mathbf{h}_{m_r}, \sigma^2), \end{aligned} \quad (4)$$

with $\mathbf{y}_{m_r} | (\mathbf{s}, \mathbf{h}_{m_r}, \sigma^2) \sim \mathcal{CN}(\sum_{m_t=1}^{M_t} \mathbf{D}_{m_t} \mathbf{W} \mathbf{h}_{m_t, m_r}, \sigma^2 \mathbf{I})$ and $\mathbf{y}[n, k] | (\mathbf{s}[n, k], \mathbf{H}[n], \sigma^2) \sim \mathcal{CN}(\mathbf{H}[n] \mathbf{s}[n, k], \sigma^2 \mathbf{I})$.

III. PRELIMINARIES

In this section, we review some necessary preliminary concepts. Specifically, we start by introducing the general task of Bayesian inference in Sec. III-A; we review the definition of BN, which is a useful graphical tool to represent knowledge about the structure of a joint distribution, in Sec. III-B; and, finally, we review a approximate solution to the Bayesian inference task, namely, Gibbs sampling in Sec. III-C.

A. Bayesian Inference

Bayesian inference aims at computing the posterior probability of the variables of interest given the evidence, where the evidence is a subset of random variables in the model. Specifically, given the values of some evidence variables $\Theta_e = \theta_e$, one wishes to estimate the posterior distribution of a subset of the unknown variables $\Theta_u = [\Theta_1, \dots, \Theta_G]^T$. We assume here for simplicity of exposition that all variables are discrete with finite cardinality. However, the extension to continuous variables with pdfs is immediate as it will be argued. The conditional pmf of Θ_u given the evidence $\Theta_e = \theta_e$ is proportional to the product of a prior distribution $p(\Theta_u)$ on the unknown variables Θ_u and of the likelihood of the evidence $p(\Theta_e | \Theta_u)$:

$$p(\Theta_u | \Theta_e = \theta_e) \propto p(\Theta_u) p(\Theta_e = \theta_e | \Theta_u). \quad (5)$$

If one is interested in computing the posterior distribution of the unknown variable Θ_j , then a direct approach would be to write

$$p(\Theta_j = \theta_j | \Theta_e = \theta_e) = \sum_{\Theta_u \setminus \Theta_j} p(\Theta_u = \theta_u | \Theta_e = \theta_e). \quad (6)$$

The inference task (6) is made difficult in practice by the multidimensional summation over all the values of the variables $\Theta_u \setminus \Theta_j$. Note also that, if the variables are continuous, the

operation of summation is replaced by integration and a similar discussion applies. Next, we discuss the BN model.

B. Bayesian Network

A BN is an acyclic graph that can be used to represent useful aspects of the structure of a joint distribution. Each node in the graph represents a random variable, while the directed edges between the nodes encode the probabilistic influence of one variable on another. Node Θ_i is defined to be a parent of Θ_j , if an edge from node Θ_i to node Θ_j exists in the graph. According to the BN's chain rule [9], the influence encoded in a BN for a set of variables $\Theta = [\Theta_1, \dots, \Theta_J]^T$ can be interpreted as the factorization of the joint distribution in the form

$$p(\Theta) = \prod_{j=1}^J p(\Theta_j | \text{Pa}_{\Theta_j}), \quad (7)$$

where we use Pa_{Θ_j} to denote the set of parent variables of variable Θ_j . Note that (7) encodes the fact that each variable Θ_j is independent of its ancestors in the BN, when conditioning on its parent variables Pa_{Θ_j} . In the following, we will find it useful to rewrite (7) in a more abstract way as [9]

$$p(\Theta) = \prod_{\phi} \phi(\mathcal{B}_{\phi}), \quad (8)$$

where the product is taken over all J factor $\phi(\mathcal{B}_{\phi}) = p(\Theta_j | \text{Pa}_{\Theta_j})$ with $\mathcal{B}_{\phi} = \{\Theta_j, \text{Pa}_{\Theta_j}\}$.

C. Gibbs Sampling

Markov chain Monte Carlo (MCMC) techniques provide effective iterative approximate solutions to the Bayesian inference task (6) that are based on randomization and can obtain increasingly accurate posterior distribution as the number of iterations increases. The goal of these techniques is to generate M random samples $\theta_u^{(1)}, \dots, \theta_u^{(M)}$ from the desired posterior distribution $p(\Theta_u | \Theta_e = \theta_e)$. This is done by running a Markov chain whose equilibrium distribution is $p(\Theta_u | \Theta_e = \theta_e)$. As a result, the multidimensional summation (or integration) (6) can be approximated by an ensemble average by the law of large numbers. In particular, the marginal distribution of any particular variable Θ_j in Θ_u can be estimated as

$$p(\Theta_j = \theta_j | \Theta_e = \theta_e) \approx \frac{1}{M} \sum_{m=M_0+1}^M \mathbf{1}(\theta_j^{(m)} = \theta_j), \quad (9)$$

where $\theta_j^{(m)}$ is the m -th sample for Θ_j generated by the Markov chain, and M_0 denotes the number of samples used as burn-in period to reduce the correlations with the initial values [13].

Gibbs sampling is a classical MCMC algorithm that defines the aforementioned Markov chain by sampling all the variables in Θ_u one-by-one. Specifically, the algorithm begins with a set of arbitrary feasible values for Θ_u . Then, at step m , a sample for a given variable Θ_j is drawn from the conditional distribution $p(\Theta_j | \Theta_u \setminus \Theta_j, \Theta_e)$. Whenever a sample is generated for a variable, the value of that variable is updated within the vector Θ_u . It can be shown that the required conditional distributions

$p(\Theta_j | \Theta_u \setminus \Theta_j, \Theta_e)$ can be calculated by multiplying all the factors in the factorization (8) that contain the variable of interest and then normalizing the resulting distribution, i.e., we have

$$p(\Theta_j | \Theta_u \setminus \Theta_j, \Theta_e) \propto \prod_{\phi: \Theta_j \in \mathcal{B}_{\phi}} \phi(\mathcal{B}_{\phi}), \quad (10)$$

where the right-hand side of (10) is the product of the factors in (8) that involve the variable Θ_j .

Remark 1: A sufficient condition for asymptotic correctness of Gibbs sampling is that the conditional distributions $p(\Theta_j | \Theta_u \setminus \Theta_j, \Theta_e)$ are strictly positive in their domains for all j [9, Ch. 12].

Remark 2: When applying Gibbs sampling to practical problems, in particular those with high-dimensional and multimodal posterior distribution $p(\Theta_j | \Theta_u \setminus \Theta_j, \Theta_e)$, slow convergence may be encountered due to the local nature of the updates. One approach to address this issue is to run Gibbs sampling with *multiple random restarts* that are initialized with different feasible solutions [13]. Moreover, within each run, *simulated annealing* may be used to avoid low-probability “traps.” Accordingly, the prior probability, or the likelihood, may be parametrized by a temperature parameter T , such that a large temperature implies a lower reliance on the evidence aimed at exploring more thoroughly the range of the variables. Samples are generated, starting with a high temperature and ending with a low temperature [10], [11].

IV. BAYESIAN INFERENCE FOR MODULATION CLASSIFICATION

In this section, we tackle the problem of detecting the modulation $A \in \mathcal{A}$ by adopting a Bayesian inference formulation. First, in Sec. IV-A, we discuss the problem of selecting a proper prior distribution, and argue that a latent Dirichlet model inspired by [12] and first used for modulation classification in [4], provides an effective choice. Then, based on this prior model, we develop a solution based on Gibbs sampling in Sec. IV-B.

A. Latent Dirichlet Bayesian Network

According to (5), the joint distribution of the unknown variables $(A, \mathbf{s}, \mathbf{h}, \sigma^2)$ may be expressed

$$p(A, \mathbf{s}, \mathbf{h}, \sigma^2 | \mathbf{y}) \propto p(\mathbf{y} | A, \mathbf{s}, \mathbf{h}, \sigma^2) p(A, \mathbf{s}, \mathbf{h}, \sigma^2), \quad (11)$$

where the likelihood function $p(\mathbf{y} | A, \mathbf{s}, \mathbf{h}, \sigma^2)$ is given in (4), and the term $p(A, \mathbf{s}, \mathbf{h}, \sigma^2)$ stands for the prior information on the unknown quantities. The prior is assumed to factorize as

$$p(A, \mathbf{s}, \mathbf{h}, \sigma^2) = p(A) \left\{ \prod_{n,k,m_t} p(s_{m_t}[n,k] | A) \right\} \times \prod_{m_t, m_r} p(\mathbf{h}_{m_t, m_r}) p(\sigma^2), \quad (12)$$

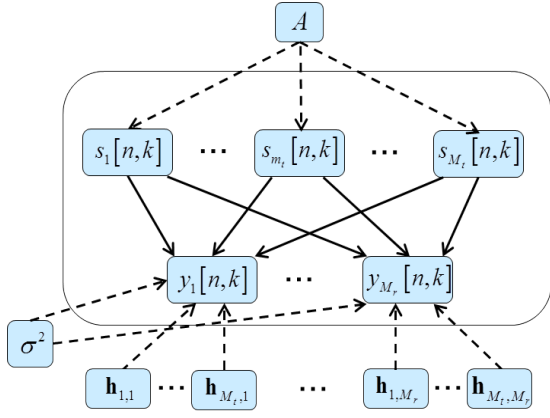


Figure 1. BN \mathcal{G}_1 for the modulation classification scheme based on the factorization (11). The nodes inside the rectangle are repeated NK times.

1) *Conventional Prior*: A natural choice for the prior distribution of the unknown variables $(A, \mathbf{s}, \mathbf{h}, \sigma^2)$ is given by $A \sim \text{uniform}(\mathcal{A})$, $s_{m_t}[n, k] | A \sim \text{uniform}(A)$, $\mathbf{h}_{m_t, m_r} \sim \mathcal{CN}(\mathbf{0}, \alpha \mathbf{I})$, and $\sigma^2 \sim \mathcal{IG}(\alpha_0, \beta_0)$ with fixed parameters $(\alpha, \alpha_0, \beta_0)$ [4]. Recall that the inverse Gamma distribution is the conjugate prior for the Gaussian likelihood at hand, and that uninformative priors can be obtained by selecting sufficiently large α and β_0 and sufficiently small α_0 [13]. The BN \mathcal{G}_1 that encodes the factorization given by (11), along with (4) and (12), is shown in Fig. 1.

The Bayesian inference task for modulation classification of MIMO-OFDM is to compute the posterior probability of the modulation A conditioned on the received signal \mathbf{y} , namely

$$p(A|\mathbf{y}) = \sum_{\mathbf{s}} \int p(A|\mathbf{s}, \mathbf{h}, \sigma^2|\mathbf{y}) d\mathbf{h}d\sigma^2. \quad (13)$$

Following the discussion in Sec. III, the calculation in (13) is intractable because of the multidimensional summation and integration. Gibbs sampling (Sec. III-C) offer a feasible solution. However, the prior distribution (12) does not satisfy the sufficient condition mentioned in *Remark 1*, since some of the conditional distributions required for Gibbs sampling are not strictly positive in their domains. In particular, the conditional distribution term $p(s_{m_t}[n, k]|A = a)$ is zero for all values of $s_{m_t}[n, k]$ not belonging to the constellation a , i.e., $p(s_{m_t}[n, k]|a) = 0$ for $s_{m_t}[n, k] \notin a$. Therefore, the Gibbs sampler generally fails to converge to the posterior distribution.

2) *Latent Dirichlet BN*: In order to alleviate the problem outlined above, we propose to adopt a prior distribution encoded on a latent Dirichlet BN \mathcal{G}_2 shown in Fig. 2. Accordingly, each transmitted symbol $s_{m_t}[n, k]$ is distributed as a random mixture of uniform distributions on the different constellations in the set \mathcal{A} . Specifically, a random vector \mathbf{p}_A of length $|\mathcal{A}|$ is introduced to represent the mixture weights, with $\mathbf{p}_A(a)$ being the probability that each symbol $s_{m_t}[n, k]$ belongs to the constellation $a \in \mathcal{A}$. Given the mixture weights \mathbf{p}_A , the transmitted symbols $s_{m_t}[n, k]$ are mutually independent and distributed according to a mixture of uniform dis-

tributions, i.e., $p(s_{m_t}[n, k]|\mathbf{p}_A) = \sum_{a: s_{m_t}[n, k] \in a} \mathbf{p}_A(a) / |a|$. The Dirichlet distribution is selected as the prior distribution of \mathbf{p}_A in order to simplify the development of the proposed solutions, as shown in the following subsections. In particular, given a set of nonnegative parameters $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_{|\mathcal{A}|}]^T$, we have $\mathbf{p}_A \sim \text{Dirichlet}(\boldsymbol{\gamma})$ [9].

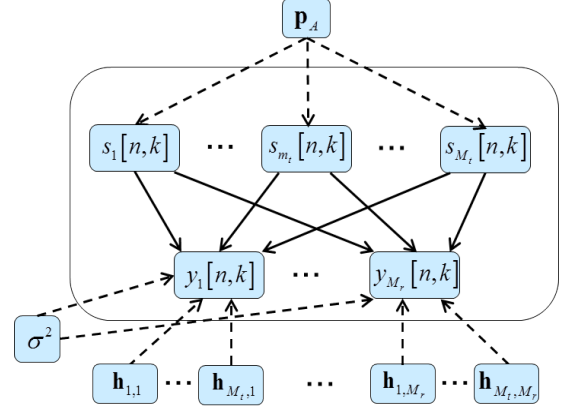


Figure 2. BN \mathcal{G}_2 for the modulation classification scheme based on the Dirichlet latent variable \mathbf{p}_A . The nodes inside the rectangle are repeated NK times.

The BN \mathcal{G}_2 encodes a factorization of the conditional distribution $p(\mathbf{p}_A, \mathbf{s}, \mathbf{h}, \sigma^2|\mathbf{y})$

$$p(\mathbf{p}_A, \mathbf{s}, \mathbf{h}, \sigma^2|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{p}_A, \mathbf{s}, \mathbf{h}, \sigma^2) p(\mathbf{p}_A) \prod_{m_t, m_r} p(\mathbf{h}_{m_t, m_r}) \left\{ \prod_{n, k, m_t} p(s_{m_t}[n, k]|\mathbf{p}_A) \right\} p(\sigma^2), \quad (14)$$

where we have $\mathbf{p}_A \sim \text{Dirichlet}(\boldsymbol{\gamma})$ with a set of nonnegative parameters $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_{|\mathcal{A}|}]^T$ [9], $p(s_{m_t}[n, k]|\mathbf{p}_A) = \sum_{a: s_{m_t}[n, k] \in a} \mathbf{p}_A(a) / |a|$, and the other distributions are as in (4) and (12). The Bayesian inference task for modulation classification is to compute the posterior probability of the mixture weight vector \mathbf{p}_A conditional on the received signal \mathbf{y} , namely

$$p(\mathbf{p}_A|\mathbf{y}) = \sum_{\mathbf{s}} \int p(\mathbf{p}_A|\mathbf{s}, \mathbf{h}, \sigma^2|\mathbf{y}) d\mathbf{h}d\sigma^2, \quad (15)$$

and then to estimate A as the value that maximize the a posteriori mean of \mathbf{p}_A :

$$\hat{A} = \arg \max_{a \in \mathcal{A}} \langle \mathbf{p}_A(a) | \mathbf{y} \rangle_{p(\mathbf{p}_A|\mathbf{y})}. \quad (16)$$

The proposed approach guarantees that all the conditional distributions needed for Gibbs sampling based on the BN \mathcal{G}_2 are non-zero, and therefore the aforementioned convergence problem for the inference based on BN \mathcal{G}_1 is avoided.

B. Modulation Classification via Gibbs Sampling

In this subsection, we elaborate on Gibbs sampling for modulation classification. As explained in Sec. III-C, in order to sample from the joint posterior distribution (14), the

distribution of each variable conditioned on all other variables is needed. According to (10), we have:

$$p(\mathbf{p}_A | \mathbf{s}, \mathbf{h}, \sigma^2, \mathbf{y}) \sim \text{Dirichlet}(\boldsymbol{\gamma} + \mathbf{c}), \quad (17)$$

where $\mathbf{c} = [c_1, \dots, c_{|\mathcal{A}|}]^T$, and c_a is the number of samples of transmitted symbols in constellation $a \in \mathcal{A}$;

$$p(s_{m_t}[n, k] | \mathbf{p}_A, \mathbf{s} \setminus s_{m_t}[n, k], \mathbf{h}, \sigma^2, \mathbf{y}) \\ \propto p(s_{m_t}[n, k] | \mathbf{p}_A) p(\mathbf{y}[n, k] | \mathbf{s}[n, k], \mathbf{H}[n], \sigma^2), \quad (18)$$

$$\mathbf{h}_{m_t, m_r} | (\mathbf{P}_A, \mathbf{s}, \mathbf{h} \setminus \mathbf{h}_{m_t, m_r}, \sigma^2, \mathbf{y}) \\ \sim \mathcal{CN}(\hat{\mathbf{h}}_{m_t, m_r}, \hat{\boldsymbol{\Sigma}}_{m_t, m_r}), \quad (19)$$

$$\text{and } \sigma^2 | \mathbf{p}_A, \mathbf{s}, \mathbf{h}, \mathbf{y} \sim \mathcal{IG}(\alpha, \beta), \quad (20)$$

where we have

$$(\hat{\boldsymbol{\Sigma}}_{m_t, m_r})^{-1} = \frac{1}{\sigma^2} \mathbf{W}^H \mathbf{D}_{m_t}^H (\mathbf{D}_{m_t} \mathbf{W}), \quad (21)$$

$$\hat{\mathbf{h}}_{m_t, m_r} \\ = \hat{\boldsymbol{\Sigma}}_{m_t, m_r} \left[\frac{1}{\sigma^2} \mathbf{W}^H \mathbf{D}_{m_t}^H \left(\mathbf{y}_{m_r} - \sum_{m'_t \neq m_t} \mathbf{D}_{m'_t} \tilde{\mathbf{h}}_{m'_t, m_r} \right) \right]; \quad (22)$$

$\alpha = \alpha_0 + NK M_r$ and $\beta = \beta_0 + \sum_{m_r} \left\| \mathbf{y}_{m_r} - \sum_{m_t} \mathbf{D}_{m_t} \tilde{\mathbf{h}}_{m_t, m_r} \right\|^2$. Note that (17) is a consequence of the fact that Dirichlet distribution is the conjugate prior of the categorical likelihood [9]; (19) can be derived by following from standard MMSE channel estimation results [4]; and (20) follows the fact that the inverse Gamma distribution is the conjugate prior for the Gaussian distribution [13].

Remark 3: When the SNR is high, the convergence speed is severely limited by the close-to-zero probabilities in the conditional distribution (18). This is due to the fact that, in this regime, the samples of σ^2 tend to be small making the relationship between $\mathbf{y}[n, k]$ and $s_{m_t}[n, k]$ almost deterministic. As discussed in *Remark 2*, the strategy of Gibbs sampling with multiple random restarts and annealing may be adopted to address this issue. For simulated annealing, we substitute the conditional distribution (20) for σ^2 with an iteration dependent prior given as [10]

$$\sigma^2 | \mathbf{p}_A, \mathbf{s}, \mathbf{h}, \mathbf{y} \sim \mathcal{IG}(\alpha', \beta), \quad (23)$$

where we have $\alpha'(m) = (1 - (1 - p_0) \exp(-m/m_0))\alpha$, with m denoting the current iteration index, $p_0 = 0.1$ and $m_0 = 0.3M$, where M is the total number of iterations. For multiple restarts, we propose to use the entropy of the pmf $\langle \mathbf{p}_A \rangle_{p(\mathbf{p}_A | \mathbf{y})}$, estimated in a run as the metric, to choose among the N_{run} runs of Gibbs sampling which one should be used in (16). Specifically, the run with the minimum entropy estimate

$\langle \mathbf{p}_A \rangle_{p(\mathbf{p}_A | \mathbf{y})}$ is selected. The rationale of this choice is that an estimate $\langle \mathbf{p}_A \rangle_{p(\mathbf{p}_A | \mathbf{y})}$ with a smaller entropy identifies a specific modulation type with low uncertainty than an estimate $\langle \mathbf{p}_A \rangle_{p(\mathbf{p}_A | \mathbf{y})}$ with higher entropy (i.e., closer to a uniform distribution).

V. NUMERICAL RESULTS AND DISCUSSIONS

In this section, we evaluate the performance of the proposed modulation classification schemes for the detection of three possible modulation formats: QPSK, 8-PSK and 16-QAM within a MIMO-OFDM system. The performance criterion is the probability of correct classification assuming that the three modulations are equally likely. Normalized Rayleigh fading channels are assumed such that $E[\|\mathbf{h}_{m_t, m_r}\|^2] = 1$. We define the average SNR as $10 \log(M_t/\sigma^2)$. Unless stated otherwise, the following conditions are assumed: *i*) $M_t = M_r = 2$ antennas; *ii*) $K = 2$ OFDM symbols; and *iii*) $L = 5$ taps with relative powers given by [0 dB, -4.2 dB, -11.5 dB, -17.6 dB, -21.5 dB].

A. Performance of Gibbs Sampling

We first investigate the performance of the proposed Gibbs sampling algorithms with or without multiple random restarts and simulated annealing within each run (see *Remark 3*). The number of runs in each process of Gibbs sampling with multiple random restarts is selected to be $N_{run} = 5$, and the number of iterations in each run is $M = 2000$, where $M_0 = 0.85M$ initial samples are used as burn-in period.¹ All elements of the vector parameter $\boldsymbol{\gamma}$ of the prior distribution $\mathbf{p}_A \sim \text{Dirichlet}(\boldsymbol{\gamma})$ are selected to be equal to a parameter γ . As also reported in [4], it may be shown, via numerical results, that the modulation classification performance is not sensitive to the choice of parameter γ as long as the value of the virtual observation γ (see Sec. IV-A2) is not very small ($\gamma < 1$). For the numerical experiments in this paper, we select the values of γ to be equal to 8% of the total number of symbols, e.g., in this example $\gamma = [0.08NK M_t] = 40$.

In Fig. 3, the performance with or without multiple random restarts and simulated annealing within each run is plotted as a function of SNR. It can be seen that both strategies of multiple random restarts and annealing improve the success rate, and that the best performance is achieved by Gibbs sampling with both random restarts and annealing. As discussed in *Remark 3*, annealing is seen to be especially effective in the high-SNR regime.

B. Comparison of Gibbs Sampling and ICA-PC[8]

Here, we compare the classification results achieved by the proposed Gibbs sampling scheme with the ICA-PC approach of [8], which extends to MIMO-OFDM the techniques studied in [5]. The approach in [8] exploits the invariance of the frequency-domain channels across the coherence bandwidth to perform classification. Specifically, the subcarriers are grouped in sets of D adjacent subcarriers whose frequency-domain

¹The samples in the burn-in period are not used to evaluate the average in (16).

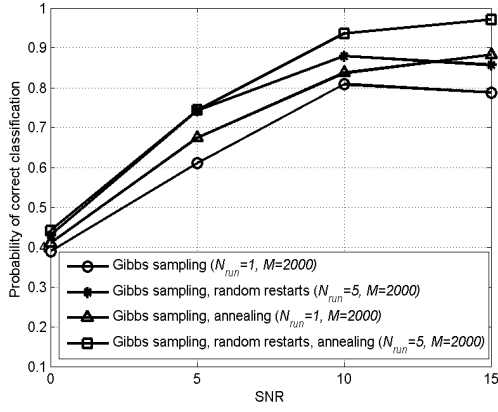


Figure 3. Probability of correct classification using Gibbs sampling versus SNR ($N = 128$, $M_t = M_r = 2$, $K = 2$ and $L = 5$).

channel matrices are assumed to be identical. Let us denote the frequency-domain channel matrix and the received samples for the i -th group by \mathbf{H}_i and \mathbf{y}_i respectively, $i = 1, \dots, N/D$. To compute the likelihood function $p(\mathbf{y}_i|A = a, \mathbf{H}_i)$ of the received samples \mathbf{y}_i over the subcarriers within group i , an estimate $\hat{\mathbf{H}}_i$ of the channel matrix \mathbf{H}_i is first obtained using ICA-PC, and then the likelihood $p(\mathbf{y}_i|A = a, \hat{\mathbf{H}}_i)$ is approximated as $p(\mathbf{y}_i|A = a, \hat{\mathbf{H}}_i)$. Accordingly, the likelihood function $p(\mathbf{y}|A = a, \mathbf{H})$ of all the received samples \mathbf{y} is approximated as $p(\mathbf{y}|A = a, \hat{\mathbf{H}}) = \prod_i p(\mathbf{y}_i|A = a, \hat{\mathbf{H}}_i)$, where $\hat{\mathbf{H}} = \{\hat{\mathbf{H}}_i\}_{i=1}^{N/D}$. The detected modulation is selected as $\hat{A} = \arg \max_{a \in \mathcal{A}} p(\mathbf{y}|A = a, \hat{\mathbf{H}})$.

In Fig. 4, we plot the performance of the approach based on ICA-PC with different values of D and Gibbs sampling with random restarts and annealing. The number of runs are $N_{run} = 5$, and the annealing schedule is (23). It can be seen from Fig. 4 that Gibbs sampling significantly outperforms ICA-PC. In this regard, note that, with $D = 4$, the accuracy in ICA-PC is poor due to the insufficient number of observed data samples; while with $D = 16$, the model mismatch problem becomes more severe due to the assumption of equal channel matrices in each subcarrier group.

VI. CONCLUSIONS

In this paper, we have proposed a Bayesian modulation classification scheme for MIMO-OFDM systems based on a selection of the prior distributions that adopts a latent Dirichlet model and on the Bayesian network formalism. The proposed Gibbs sampling method converges to the optimal Bayesian solution and its speed of convergence is shown to improve by multiple random restarts and annealing. The technique is seen to overcome the performance limitation of state-of-the-art non-Bayesian schemes based on ICA. In fact, while the mentioned existing modulation classification algorithms rely on the assumptions that the channels are flat fading, and/or that a large amount of samples are available (as for pattern recognition-based methods), the proposed scheme achieves

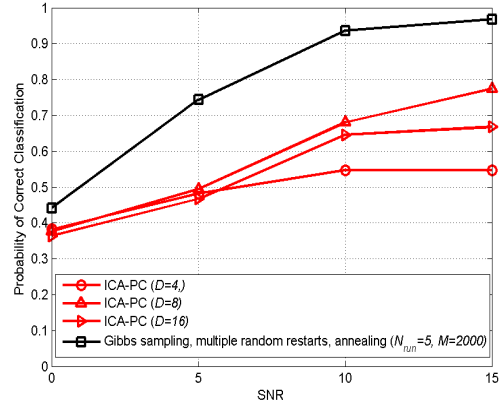


Figure 4. Probability of correct classification using Gibbs sampling with multiple random restarts and annealing and approach of [8] based on ICA-PC versus SNR ($N = 128$, $M_t = M_r = 2$, $K = 2$ and $L = 5$).

satisfactory performance under more general conditions. For example, with $M_t = 2$ transmit antennas and under frequency selective fading channels with $L = 5$ taps, a correct classification rate of above 97% may be attained with $M_r = 2$ receive antennas and with 256 received samples at each antenna.

REFERENCES

- [1] O. A. Dobre, A. Abdi, Y. Bar-Ness and W. Su, "A survey of automatic modulation classification techniques: classical approaches and new developments," *IET Communications*, vol.1, no. 2, pp. 137-156, Apr. 2007.
- [2] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari and D. B. Rubin, *Bayesian Data Analysis*, CRC Press, 2013.
- [3] T. A. Drumright and Z. Ding, "QAM constellation classification based on statistical sampling for linear distortive channels," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1575-1586, May 2006.
- [4] Y. Liu, O. Simeone, A. Haimovich and W. Su, "Modulation classification via Gibbs sampling based on a latent Dirichlet Bayesian network," *IEEE Signal Proc. Letters*, vol. 21, no. 9, pp. 1135-1139, Sep. 2014.
- [5] V. Choqueuse, S. Azou, K. Yao, L. Collin and G. Burel, "Blind Modulation recognition for MIMO Systems," *MTA Review*, vol. XIX, no. 2, pp. 183-195, June 2009.
- [6] M. S. Muhlhaus, M. Oner, O. A. Dobre, H. U. Jakel and F. K. Jondral, "Automatic modulation classification for MIMO systems using fourth-order cumulants," in *Proc. IEEE Vehic. Technology Conf.* Fall, pp. 1-5, Quebec City, QC, CAN., Sep. 2012.
- [7] M. S. Muhlhaus, M. Oner, O. A. Dobre and F. K. Jondral, "A low complexity modulation classification algorithm for MIMO systems," *IEEE Communications Letters*, no. 10, pp. 1881-1884, Oct. 2013.
- [8] A. Agirman-Tosun, Y. Liu, A. M. Haimovich, O. Simeone, W. Su, j. Dabin and E. Kanterakis, "Modulation classification of MIMO-OFDM signals by independent component analysis and support vector machines," *IEEE Asilomar Conference*, CA, Nov. 2011.
- [9] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*, MIT Press, 2009.
- [10] C. Fevotte and S. J. Godsill, "A Bayesian approach for blind separation of sparse sources," *IEEE Trans. Audio, Speech and Language Process.*, vol. 14, no. 6, pp. 2174-2188, Nov. 2006.
- [11] Y. Nourani and B. Andresen, "A comparison of simulated annealing cooling strategies," *J. Phys. A*, vol. 31, no. 41, pp. 8373-8385, July 1998.
- [12] D. Blei, A. Ng, and M. Jordan, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993-1022, Jan. 2003.
- [13] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer, 1999.