INFINITE FACTORIAL UNBOUNDED HIDDEN MARKOV MODEL FOR BLIND MULTIUSER CHANNEL ESTIMATION

Isabel Valera, Francisco J. R. Ruiz, Fernando Perez-Cruz

Department of Signal Theory and Communications, University Carlos III in Madrid
Fernando Perez-Cruz is also a Machine Learning Scientist at Amazon

ABSTRACT

Bayesian nonparametric models allow solving estimation and detection problems with an unbounded number of degrees of freedom. In multiuser multiple-input multiple-output (MIMO) communication systems we might not know the number of active users and the channel they face, and assuming maximal scenarios (maximum number of transmitters and maximum channel length) might degrade the receiver performance. In this paper, we propose a Bayesian nonparametric prior and its associated inference algorithm, which is able to detect an unbounded number of users with an unbounded channel length. This generative model provides the dispersive channel model for each user and a probabilistic estimate for each transmitted symbol in a fully blind manner, i.e., without the need of pilot (training) symbols.

Index Terms— Hidden Markov models, Bayesian non parametrics, Markov chain Monte Carlo, multiple-input multiple-output (MIMO), channel estimation, user detection.

1. INTRODUCTION

In the last two decades, many researchers have focused on multiple-input multiple-output (MIMO) communication systems, due to their high channel capacity at comparatively low bandwidth consumption [1, 2]. When digital symbols are transmitted over frequency-selective channels, inter-symbol interference (ISI) occurs, degrading the performance of the receiver in terms of symbol detection error probability. To improve the performance, channel estimation is applied to mitigate the effects of ISI. Before detecting the transmitted symbols, the channel state information (CSI) needs to be estimated at the receiver by sending pilot symbols as training data. Blind channel estimation involves symbol detection without the use of training data, which allows a more efficient communication as the total bandwidth becomes available for the user's data. This can be accomplished either without explicit estimation of the channel parameters or by joint symbol detection and channel parameter estimation.

We address the problem of blind joint channel parameter and data estimation in a MIMO communication channel. Specifically, we tackle the case where neither the number of transmitters nor the length of the channel impulse response (channel length) is known. Further details on the problem statement are given in Section 2. Up to our knowledge, all previous works consider that at least one of these quantities is fixed and known, and no efforts have been made to address simultaneously the channel length estimation and the user activity detection problems. This paper aims to solve this limitation making use of Bayesian nonparametric (BNP) tools, which constitutes a novel contribution. The symbols sent by each transmitter can be viewed as a time sequence that the receiver tries to reconstruct, leading us to hidden Markov models (HMM) [3]. Our approach, based on BNP tools, consists on modeling all the transmitters as an unbounded number of independent chains in a factorial HMM (FHMM). In the literature, many nonparametric extensions of standard time series models can be found. The hierarchical Dirichlet process (HDP) has been proposed to define an HMM with an infinite number of latent states called HDP-HMM [4]. The nonparametric extension of the FHMM in [5] is the infinite factorial (binary) HMM (IFHMM) [6], which defines a probability distribution over an unbounded number of binary Markov chains. We extend this model to allow for any number of states and develop a new algorithm to infer both the number of Markov chains and the number of states. Due to its nonparametric nature, our model becomes flexible enough to account for any number of transmitters and channel length in any communication scenario, without the need of additional previous knowledge or bounds. Our model is suitable for binary phase-shift keying (BPSK) multiuser noncoherent scenarios in which neither the transmitters nor the receiver know the CSI and this information is blindly estimated. The channel model is selective in time and frequency, underspread and locally widesense stationary [7]. Although we focus on additive white Gaussian noise (AWGN) channels, it can be readily adapted to reckon with other probability distributions over the noise.

2. MIMO CHANNEL

Assume a MIMO system with N_t transmitters and N_r receiving antennas, in which each receiver observes a linear combination of all the transmitted data sequences, due to the ISI, under additive white Gaussian noise. Specifically, the row observation vector compound of the observations at all the receiving antennas at time instant t is given by $\mathbf{y}_t = \sum_{k=0}^{K-1} \mathbf{x}_{t-k}\mathbf{H}^{(k)} + \mathbf{n}_t$, where $\mathbf{x}_{t-k}$ is a vector with the symbols transmitted by all the transmitters at instant $t-k$, $\mathbf{H}^{(k)}$ is a $N_t \times N_r$ matrix which contains the channel coefficients corresponding to tap k ($k \in \{0, \ldots, K-1\}$, being K the channel length for all the transmitter-receiver pairs), and $\mathbf{n}_t$ is the N_r-dimensional noise vector. Let us denote each element of matrix $\mathbf{H}^{(k)}$ as $h_{ji}^{(k)}$.

We assume the use of a BPSK constellation (i.e., each symbol $x_{tn} = \pm 1$ with equal probability) and therefore, $\mathbf{n}_t$, $\{\mathbf{H}^{(k)}\}_{k=0}^{K-1}$ and $\mathbf{y}_t$ contain all real values. We consider that the noise $\mathbf{n}_t$ is Gaussian distributed with zero mean and covariance matrix $\sigma_n^2 \mathbf{I}_{N_r}$, being $\mathbf{I}_{N_r}$ the identity matrix of size N_r, and the channel coefficients $h_{ji}^{(k)}$ are Gaussian distributed with zero mean and variance 1.

Our goal is to infer both the number of transmitters and the transmitted symbols (as well as the channel length and the channel coefficients) using the observations collected during T time steps, i.e., the observation vectors $\mathbf{y}_t$ for $t = 1, \ldots, T$ that we gather in a $T \times N_r$ matrix $\mathbf{Y}$. This problem can be found, for instance, in a code-division multiple access (CDMA) context where a set of terminals wish to communicate with a common access point (AP). Each terminal accesses the channel randomly, and the AP receives the superposition of signals from the active terminals only. The AP is interested in determining both the active terminals and the transmitted symbols. Another example arises in the context of wireless sensor networks, where the communication nodes can often switch on and off asynchronously during operation, and a fusion center collects the signals from a subset of them. Again, the fusion center faces the problem of determining both the number of active sensors and the symbols that each sensor transmits [8]. Finally, another example can be found in cooperation schemes, such as interference alignment [9], in which the reuse of frequencies in nearby cells creates an interference channel between the users and the base stations.

Here, we propose an infinite factorial unbounded HMM, in which each parallel chain represents a transmitter and the state at each time instant in the Markov chain corresponds to the state of the channel between that transmitter and all the receivers, where the state of the channel is determined by the set of the last K symbols sent by the transmitter. Hence, the set of unknowns is composed of the number of transmitters N_t, the symbols sent by each transmitter, the channel length K and the channel coefficients $\{\mathbf{H}^{(k)}\}_{k=0}^{K-1}$.

3. INFINITE FACTORIAL UNBOUNDED HMM

Assume a factorial nonbinary HMM shown in Figure 1. In this figure, $s_{tm} \in \{0, 1, \ldots, Q-1\}$ represents the hidden state at time instant t in the m-th chain and all the states s_{tm} are grouped together in a $T \times M$ matrix denoted by $\mathbf{S}$. For simplicity, we assume that $s_{0m} = 0$ for all the Markov chains.

Under this model, each chain m represents an HMM with transition probabilities contained in the $Q \times Q$ matrix $\mathbf{A}^m$, whose rows are denoted by $\mathbf{a}_q^m$ ($q = 0, \ldots, Q-1$). Hence, $\mathbf{a}_q^m$ corresponds to the transition probability vector from state q in chain m. Note that the transition probability matrices $\mathbf{A}^m$ are independently distributed for each Markov chain $m = 1, \ldots, M$, and, since the variables s_{tm} follow an HMM, we can write that

$$s_{tm} | s_{(t-1)m}, \mathbf{A}^m \sim \mathbf{a}_{s_{(t-1)m}}^m. \tag{1}$$

As shown in Figure 1 and in order to ensure the reproducibility of the model when the number of chains M tend to infinity, the transition probability vectors $\mathbf{a}_q^m$ are differently distributed for $q = 0$ (inactive state) than for the rest of the states, i.e.,

$$\mathbf{a}_q^m | Q, \beta_0, \beta \sim \text{Dirichlet}(\beta_0, \beta, \ldots, \beta), \tag{2}$$

for $q = 1, \ldots, Q-1$, and

$$\mathbf{a}_0^m = \begin{bmatrix} a^m & (1-a^m)p_1^m & \ldots & (1-a^m)p_{Q-1}^m \end{bmatrix}, \tag{3}$$

where β_0 and β model the *a priori* information about the transition probabilities from states other than 0, and a^m and $\mathbf{p}^m = [p_1^m, \ldots, p_{Q-1}^m]$ are in turn random variables distributed as

$$a^m | \alpha \sim \text{Beta}\left(1, \frac{\alpha}{M}\right), \text{ and} \tag{4}$$

$$\mathbf{p}^m | Q, \gamma \sim \text{Dirichlet}(\gamma). \tag{5}$$

The elements of vector $\mathbf{a}_0^m$ are denoted by a_{0i}^m, for $i = 0, \ldots, Q-1$.

The number of hidden states Q is Poisson distributed with parameter λ, namely,

$$p(Q|\lambda) = \frac{\lambda^{Q-2} e^{-\lambda}}{(Q-2)!}, \qquad Q = 2, \ldots, \infty. \tag{6}$$

The choice of Q sampled from a Poisson distribution is not a limitation in the application of MIMO channel estimation, where the memory of the channel, as well as the channel coefficients, are assumed to be invariant during the observation period.

We can obtain the probability distribution over the matrix $\mathbf{S}$ by integrating out the transition probabilities. However, as the number of independent Markov chains M tends to infinity, the probability of a single matrix $\mathbf{S}$ vanishes. This is not a limitation, since we are interested in the probability of the whole equivalence class of $\mathbf{S}$. The equivalence classes are defined with respect to a function on integer-valued matrices, called *lof*$(\cdot)$ (left-ordered form), which is obtained by sorting the columns of the matrix $\mathbf{S}$ from left to right by the history of that column, defined as the magnitude of the base-Q number expressed by that column, taking the first row as the most significant value. Since the elements of matrix $\mathbf{S}$ can be arbitrarily relabeled, we say that two matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ with elements in $\{0, \ldots, Q-1\}$ are in the same equivalence class if there exists a bijective permutation function $f(\cdot)$ on the set $\{0, \ldots, Q-1\}$, subject to $f(0) = 0$, such that, when applied to all the elements of $\mathbf{S}_2$ to obtain $\mathbf{S}_2'$, $lof(\mathbf{S}_1) = lof(\mathbf{S}_2')$. The element 0 cannot be relabeled, since it represents the inactive state and therefore requires special treatment.

Finally, remark that this model is exchangeable in the columns and in the labels of the states, and Markov exchangeable in the rows.

In order to fit the likelihood model to MIMO systems we propose the observation model shown in Figure 1, in which the observation matrix $\mathbf{Y}$ is distributed as a Gaussian random matrix, i.e.,

$$p(\mathbf{Y}|\mathbf{S}, \mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_{Q-1}, \sigma_y^2) = \frac{1}{(2\pi\sigma_y^2)^{\frac{TD}{2}}} \exp\Bigg\{ -\frac{1}{2\sigma_y^2} \times \text{trace}\Bigg[\left(\mathbf{Y} - \sum_{q=1}^{Q-1} \mathbf{Z}_q \mathbf{\Phi}_q \right)^\top \left(\mathbf{Y} - \sum_{q=1}^{Q-1} \mathbf{Z}_q \mathbf{\Phi}_q \right) \Bigg] \Bigg\}, \tag{7}$$

where $\mathbf{Z}_q$ is defined as a binary $T \times M_+$ matrix with elements $(\mathbf{Z}_q)_{tm} = 1$ if $s_{tm} = q$ (assuming $\mathbf{S}$ is expressed in its left-ordered form) and zero otherwise, and $\mathbf{\Phi}_q$ are $M_+ \times D$ matrices, with M_+ being the number of nonzero columns in $\mathbf{S}$. Thus, the mean value of $\mathbf{y}_t$ depends on the additive contribution of all chains (transmitters) at time instant t.

Under this model, the matrices $\mathbf{\Phi}_q$ are closely related to the channel coefficients. Specifically, each row of the matrix $\mathbf{\Phi}_q$ corresponds to the linear combination of the K channel coefficients corresponding to a particular state q of the channel between one transmitter and all the receivers. Note that the number of coefficients that are linearly combined to obtain the matrices $\mathbf{\Phi}_q$ coincides with the channel length K, which is represented in this model by the number of states Q, such that $Q = 2^K + 1$.

We place a normal-inverse-gamma prior over both the observation noise variance and the set of matrices $\{\mathbf{\Phi}_q\}_{q=1}^{Q-1}$, i.e.,

$$\begin{aligned} p(\{\mathbf{\Phi}_q\}_{q=1}^{Q-1}, \sigma_y^2 | \mathbf{S}, \xi, \nu, \tau) &= \mathrm{N} - \Gamma^{-1}(0, \xi, \nu, \tau) \\ &= \frac{\nu^\tau}{\Gamma(\tau)} \left(\frac{1}{\sigma_y^2} \right)^{\tau+1} \exp\left\{ -\frac{\nu}{\sigma_y^2} \right\} \\ &\quad \times \prod_{q=1}^{Q-1} \frac{1}{\left(2\pi\sigma_y^2/\xi\right)^{\frac{DM_+}{2}}} \exp\left\{ -\frac{\xi}{2\sigma_y^2} \text{trace}\left[\mathbf{\Phi}_q^\top \mathbf{\Phi}_q \right] \right\}. \end{aligned} \tag{8}$$

Note that the parameters $\mathbf{\Phi}_q$ for any particular state q through all the chains are assumed as independent and we have also placed a prior over the observation noise variance. The independence among the variables $\mathbf{\Phi}_q$ is assumed to properly fit the MIMO channel, where the channel coefficients are all independent, while the prior over the noise variance makes the inference less sensitive to this parameter.

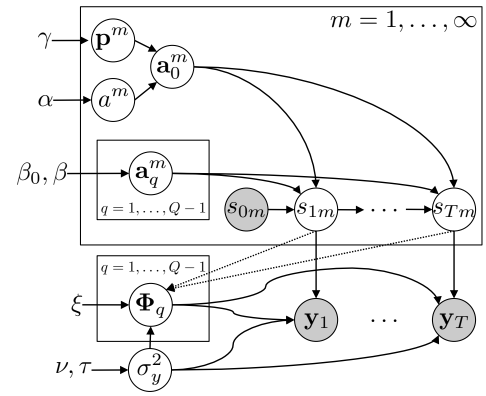

Fig. 1. Graphical observation model for the nonbinary infinite factorial HMM.

4. INFERENCE

We propose an MCMC sampler that obtains samples from the target distribution $p([\mathbf{S}], Q|\mathbf{Y})$ by iteratively proceeding as follows:

- **Step 1:** Update the allocation matrix $\mathbf{S}$ for a given value of Q via Gibbs sampling.
- **Step 2:** Consider splitting a component into two or merging two into one.
- **Step 3:** Consider the birth of a new state or the death of an empty state.

Remark that the number of active parallel Markov chains is updated in first step, while the two latter steps allow increasing or decreasing the number of states Q by one. For clarity, throughout this subsection we drop the dependence on the hyperparameters in the notation.

Step 1: Gibbs sampler

The first step involves just a sweep of the Gibbs sampler. The algorithm iteratively samples the value of each element s_{tm} given the remaining variables, i.e., it samples from

$$p(s_{tm} = k|\mathbf{Y}, \mathbf{S}_{\neg tm}, Q) \propto p(s_{tm} = k|\mathbf{S}_{\neg tm}, Q) p(\mathbf{Y}|\mathbf{S}), \tag{9}$$

where $\mathbf{S}_{\neg tm}$ represents the matrix $\mathbf{S}$ without the element s_{tm}.

An analytical expression for the term $p(s_{tm} = k|\mathbf{S}_{\neg tm}, Q)$ in Eq. 9 can be derived. Additionally, the second term in Eq. 9 can be obtained by integrating out the observation noise variance σ_y^2 and all matrices $\mathbf{\Phi}_q$ in (7).

Hence, for $t = 1, \ldots, T$, the Gibbs sampler proceeds as follows:

1. For $m = 1, \ldots, M_+$, sample element s_{tm} from (9).

2. Draw M_{new} columns of $\mathbf{S}$ with states s_{tm} ($m = M_+ + 1, \ldots, M_+ + M_{new}$) from a distribution where the prior is Poisson$(M_{new}|\frac{\alpha}{T}) \times \frac{1}{(Q-1)^{M_{new}}}$, and update M_+.

Step 2: Split and Merge moves

In the second step, we choose to split with probability b_Q or to merge with probability $d_Q = 1 - b_Q$. Naturally, $d_2 = 0$, and we use $b_Q = d_Q = 1/2$ for $Q = 3, \ldots, \infty$. In the merge move, we start from a matrix $\tilde{\mathbf{S}}$ and $Q+1$ states and we randomly select two of the nonzero states, q_1 and q_2, and combine them into a single state q_*, thus creating a matrix $\mathbf{S}$ with Q states. In the split move, in which we start from a matrix $\mathbf{S}$ and Q states, a nonzero state q_* is randomly chosen and split into two new ones, q_1 and q_2, ending with a new matrix $\tilde{\mathbf{S}}$ and $Q+1$ states. The acceptance probabilities for the split and merge moves are given by $\min(1, R)$ and $\min(1, R^{-1})$, respectively, where

$$R = \frac{p(\mathbf{Y}|[\tilde{\mathbf{S}}])}{p(\mathbf{Y}|[\mathbf{S}])} \frac{p([\tilde{\mathbf{S}}]|Q+1)}{p([\mathbf{S}]|Q)} \frac{p(Q+1)}{p(Q)} \frac{d_{Q+1} 2/Q}{b_Q P_{alloc}}, \quad (10)$$

being P_{alloc} the probability of making the particular allocation of the elements in matrix $\tilde{\mathbf{S}}$. Therefore, P_{alloc} depends on how the elements in $\mathbf{S}$ taking value q_* are split into q_1 and q_2. Although the simplest allocation method could consist on splitting completely at random, we decide to apply several sweeps of a restricted Gibbs sampling scheme for those states in $\mathbf{S}$ taking value q_*, so as to increase the acceptance probability.

Step 3: Birth and Death moves

In the third step, we first randomly choose between the birth or the death of a state with probabilities b_Q and d_Q, respectively. The removal of a state is accomplished by randomly selecting an empty component and deleting it, thereby jumping from $Q+1$ states to Q. Matrix $\tilde{\mathbf{S}}$ is relabeled so that its elements belong to the set $\{0, \ldots, Q-1\}$, resulting in matrix $\mathbf{S}$. In the birth move, we start from a model with Q states and we create a new empty component. Matrix $\mathbf{S}$ is unaltered in this process, i.e., $\tilde{\mathbf{S}} = \mathbf{S}$. The acceptance probabilities for the birth and death moves are $\min(1, R)$ and $\min(1, R^{-1})$, respectively, where in this case R is given by

$$R = \frac{p([\tilde{\mathbf{S}}]|Q+1)}{p([\mathbf{S}]|Q)} \frac{p(Q+1)}{p(Q)} \frac{d_{Q+1}}{b_Q(Q_0+1)}, \quad (11)$$

with Q_0 being the number of empty components before the birth of a new empty state.

5. EXPERIMENTS

We now generate a series of examples to assess the performance of the proposed model. To this end, we simulate a MIMO system for different scenarios. Simulated data is standard for analyzing digital communication systems and widely accepted in the research community (see, e.g., [10, 11, 12]), because the models accurately represent the wireless channels. We try different values for the number of transmitters N_t, the number of receivers N_r, the channel length K, and the signal-to-noise ratio (SNR), which is defined as $\text{SNR(dB)} = -10\log(\sigma_n^2)$. In particular, we consider three multiuser MIMO scenarios:

- *Scenario A*: Flat channel ($K = 1$) with two transmitters ($N_t = 2$).
- *Scenario B*: Channel length $K = 2$ and $N_t = 2$.
- *Scenario C*: Channel length $K = 2$ and $N_t = 3$.

For these three cases, we vary the SNR value as well as the number of receivers N_r. In order to generate the observations, we assume a number of transmitters N_t that send a burst of BPSK symbols during the observation period $T = 150$. We assume that the transmitters sequentially become active with random initial instant and burst duration, ensuring that the burst consists in the transmission of at least 30 symbols since shorter bursts are unusual in a real communication system. As we described in Section 2, the channel is assumed to be Rayleigh, i.e., the channel coefficients are Gaussian distributed with zero mean and variance equal to one, and the observations are corrupted by Gaussian additive noise with zero mean and variance σ_n^2.

We evaluate the performance of the model in terms of detection error probability (DEP), defined as the error probability of detecting both the true number of transmitters and the true channel length. Additionally, for those cases where the true values for the number of transmitters and channel length are recovered, we also evaluate the symbol error rate (SER), the activity detection error rate (ADER), and the mean square error (MSE) of the channel coefficient estimates. When computing the SER, an error is computed at time t whenever the estimated symbol for a transmitter differs from the actual transmitted symbol, given that the transmitter is active. Regarding the ADER, it is the probability of detecting activity (inactivity) in a transmitter while that transmitter is actually inactive (active). Additionally, if we denote the $N_t \times N_r$ matrix that contains the estimated channel coefficients corresponding to tap k ($k \in \{0, \ldots, K-1\}$) by $\hat{\mathbf{H}}^{(k)}$ and its elements by $\hat{h}_{ji}^{(k)}$, we can compute the MSE as

$$\text{MSE} = \frac{1}{KN_tN_r} \sum_{j,i,k} \left(h_{ji}^{(k)} - \hat{h}_{ji}^{(k)} \right)^2, \quad (12)$$

where the estimated coefficients $\hat{h}_{ji}^{(k)}$ are obtained using the MAP (*maximum a posteriori*) solution of the matrices $\mathbf{\Phi}_q$.

For each scenario, we run 500 simulations for each combination of the SNR and the N_r values. In each experiment we run 50 iterations of the inference algorithm presented in Section 4 to infer the latent matrix $\mathbf{S}$. The hyperparameters are set to $\alpha = 1$, $\gamma = 1$, $\beta_0 = 0.1$, $\beta = 10$, $\lambda = 1$,

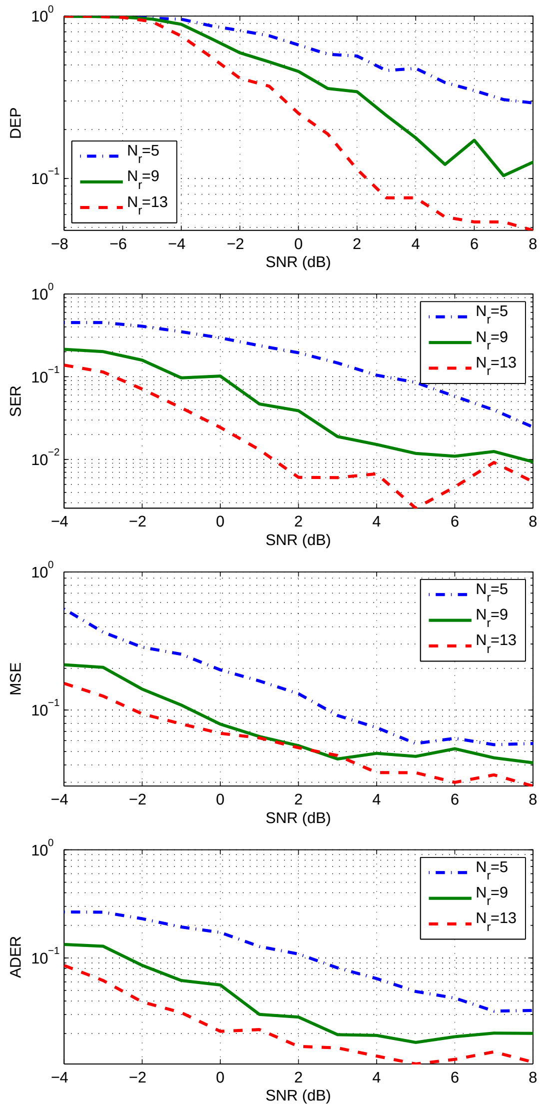

Fig. 2. Results for the Scenario A.

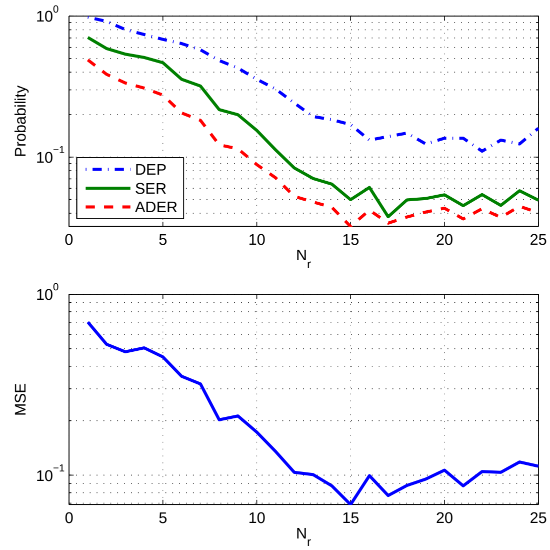

Fig. 3. Results for the Scenario B.

$\tau = 1$, $\nu = 0.1$ and $\xi = 2\sigma_n^2$ (the SNR is known at the receiver because it usually has a SNR estimator device). Note that, although we have adapted the observation model to properly fit MIMO systems, the proposed model still suffers from several limitations because, although the number of possibles states in the channel is a power of two (i.e., 2^K), our model allows any integer value above 1. Then, we resort to an additional post-processing of the inference results to account for the prior knowledge of the communication system. Specifically, we rearrange the elements of the inferred matrix $\mathbf{S}$, so that the inferred matrices $\mathbf{\Phi}_q$ properly recover the channel coefficients. We repeat the previous procedure, consisting on the 50 iterations of the inference algorithm and post-processing, initialized with the results of the first post-processing.

For the Scenario A, we show in Fig. 2 the DEP, the SER, the MSE and the ADER as functions of the SNR for several values of the number of receivers N_r. In all these plots, we observe that the performance of the proposed algorithm (initialized with $Q = 2$ states) improves as the SNR or the number of receivers increase.

For the Scenario B, we show in Fig. 3 the DEP, the SER, the MSE and the ADER as functions of the number of receivers N_r for SNR $= 0$dB and initializing the algorithm with $Q = 6$ states. Note that the behavior of the inference improves as the number of receivers increases until $N_r = 15$, but for higher values of N_r the DEP is around 10% and the SER, the MSE and the ADER also remain approximately constant.

Finally, let us analyze the performance of the inference algorithm in the Scenario C. To this end, in Fig. 4 we plot the DEP, the SER, the MSE and the ADER as functions of the number of receivers N_r for SNR $= 0$dB and several initializations of the number of states Q (denoted by Q_{ini} in the plots). The top plot shows that the DEP is much higher when the algorithm is initialized with $Q_{ini} = 2$ states. This behavior is due to the fact that the algorithm falls in a local optimum different from the ground truth. However, once the algorithm finds the true values of both the number of transmitters and the number of states, the performance of the model is similar regardless of the initialization. In the cases in which Q is initialized to 4 or 6, we find similar results, being $Q_{ini} = 6$ slightly better in terms of DEP.

Under the three scenarios, we observe in the SER and the ADER plots the presence of an error floor (of the order of 10^{-2}) corresponding to the errors caused by the active-to-inactive (and inactive-to-active) transitions, which are not taken into account in our model. These error floors can be decreased by transmitting larger bursts of symbols, since the number of transitions becomes negligible compared to the total number of transmitted bits.

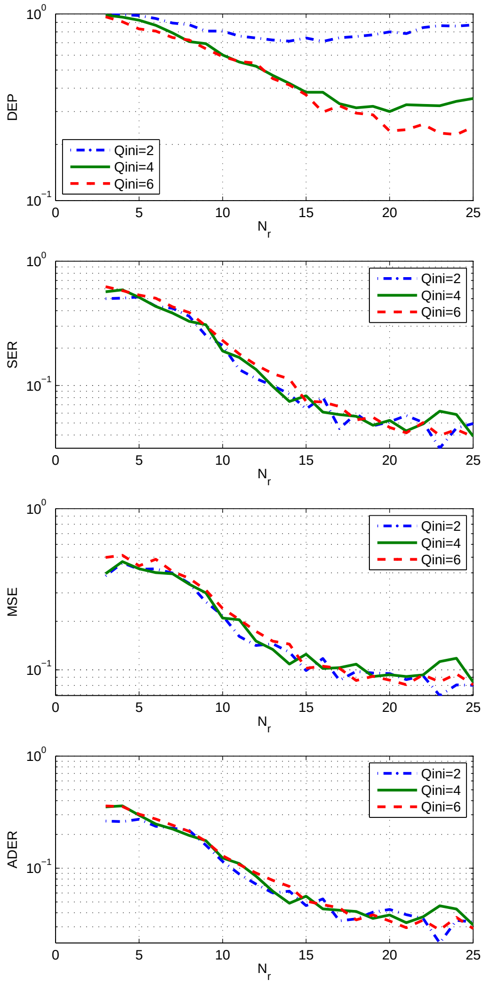

Fig. 4. Results for the Scenario C.

6. CONCLUSIONS

We have extended the existent binary IFHMM [6] to allow for any number of states in the Markov chains and developed an MCMC algorithm that learns both the number of parallel chains and the number of hidden states in the factorial HMM. Our algorithm effectively deals with the trade-off problem between the number of chains and the number of states, avoiding the model selection. We have applied the model to the joint activity detection and channel parameter estimation problem in multiuser MIMO systems. Simulation results show that our algorithm properly recovers the number of active transmitters as well as the transmitted symbols while estimating both the channel length and the channel coefficients.

7. REFERENCES

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, nov 1999.

[2] A. Scherb and K.-D. Kammeyer, "Bayesian channel estimation for doubly correlated MIMO systems," in *International ITG/IEEE Workshop on Smart Antennas (WSA 07)*, Vienna, Austria, Feb 2007.

[3] L.R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, feb 1989.

[4] Y. W. Teh and M. I. Jordan, "Hierarchical Bayesian nonparametric models with applications," in *Bayesian Nonparametrics: Principles and Practice*, N. Hjort, C. Holmes, P. Müller, and S. Walker, Eds. Cambridge University Press, 2010.

[5] Z. Ghahramani and M. I. Jordan, "Factorial hidden Markov models," *Machine Learning*, vol. 29, no. 2-3, pp. 245–273, 1997.

[6] J. Van Gael, Y. W. Teh, and Z. Ghahramani, "The infinite factorial hidden Markov model," in *Advances in Neural Information Processing Systems (NIPS)*, 2009, vol. 21.

[7] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: information-theoretic and communications aspects," *Information Theory, IEEE Transactions on*, vol. 44, no. 6, pp. 2619–2692, oct 1998.

[8] Hao Zhu and G. B. Giannakis, "Exploiting sparse user activity in multiuser detection," *Communications, IEEE Transactions on*, vol. 59, no. 2, pp. 454–465, february 2011.

[9] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the k -user interference channel," *Information Theory, IEEE Transactions on*, vol. 54, no. 8, pp. 3425–3441, aug. 2008.

[10] A. Angelosante, E. Biglieri, and M. Lops, "Multiuser detection in a dynamic environment - Part II: Joint user identification and parameter estimation," *Information Theory, IEEE Transactions on*, vol. 55, pp. 2365–2374, May 2009.

[11] D. Angelosante, E. Biglieri, and M. Lops, "Low-complexity receivers for multiuser detection with an unknown number of active users," *Signal Processing*, vol. 90, pp. 1486–1495, May 2010.

[12] M. A. Vázquez and J. Míguez, "A per-survivor processing receiver for MIMO transmission systems with one unknown channel order per output," *Vehicular Technology, IEEE Transactions on*, vol. 60, no. 9, pp. 4415–4426, nov 2011.