CLASSIFICATION OF DIGITAL MODULATIONS BY MCMC SAMPLING

Stéphane Lesage*, Jean-Yves Tourneret* and Petar M. Djurić†

* ENSEEIHT/TéSA, 2 rue Charles Camichel, BP 7122, 31071 Toulouse Cedex 7, France
† Department of Electrical and Computer Engineering, SUNY at Stony Brook, NY 11794, USA

ABSTRACT

This paper addresses the problem of classification of digital modulations. The proposed solution uses the Bayes classifier, which is implemented by the Markov chain Monte Carlo scheme. In the proposed implementation, classifications in presence of phase and frequency offsets as well as residual filtering effects coming from imperfect channel equalization are considered. The proposed approach has been tested for many scenarios and its performance has been compared with the maximum likelihood classifier and the 4th order cumulant-based method. The obtained results show that our classifier outperforms the other methods considerably.

1. INTRODUCTION AND PROBLEM FORMULATION

The digital modulation classification problem consists of determining the underlying symbol constellation from observed noisy measurements. This problem has received much attention in the signal processing and communications literature (see for instance [1][2][3][4] and the references therein). The optimal Bayes classifier, which minimizes the average probability of error (based on a zero-one loss function), can be studied under appropriate conditions on the symbol sequence and the communication channel. Unfortunately, when these conditions are not satisfied, the Bayes classifier suffers from high computational complexity. This has motivated many authors to study suboptimal classifiers. Typical suboptimal classifiers consist of computing appropriate features from the observed data and applying standard classification rules (such as the nearest-mean rule of the k-nearest neighbor rule) on these features. The features, which have been used for classification of digital modulation, include moments of the extracted phase [5], estimates of the instantaneous amplitude, phase and frequency [6], wavelets coefficients [7] or more recently higher-order cumulants [3]. This paper studies an implementation of the Bayesian rule using Markov Chain Monte Carlo (MCMC) methods.

We assume that we can operate in a coherent and synchronous environment and that the carrier, timing, and waveform recovery have been accomplished. All problems concerning signal bandwidth, baud rate, pulse-shaping filter and noise variance estimations are not addressed in this paper. The baseband complex envelope of the modulated signal sampled at one sample per symbol can be written as:

$$x_k = m_k + n_k, \ k = 1, 2, \dots, N_S \qquad (1)$$

where n_k is an independent and identically distributed (i.i.d.) complex Gaussian noise sequence with variance σ_n^2 ($n_k \sim \mathcal{N}\left(0, \sigma_n^2\right)$),

with the real and imaginary components of n_k being independent and identically distributed, and

$$m_k = A e^{j(\pi \frac{k}{N_S} f_r + \phi)} \sum_{l=0}^{p} h_l s_{k-l}, \ k = 1, 2, \dots, N_S \qquad (2)$$

where s_k is an i.i.d. symbol sequence drawn from one of c constellations denoted $\{\omega_1, \omega_2, \dots, \omega_c\}$, and the constellation ω_j is a set of M_j complex numbers $\{S_1, S_2, \dots, S_{M_j}\}$. The remaining symbols have the following meaning: $\phi \in [-\pi/4, \pi/4]$ is a phase offset, A is the signal amplitude, $h = (h_0 = 1, h_1, ..., h_p)$ is the residual channel modeled as an FIR filter, and $f_r = 2N_S(f_c - \hat{f}_c) \in (-1/2, 1/2]$ is a normalized residual carrier frequency also called frequency offset (for $k = N_S$, f_r is the constellation rotation whose maximum value is $90°$).

Given N_S samples x_k, $k = 1, 2, \dots, N_S$ of a modulated signal, the problem of digital modulation classification consists of determining the underlying modulation ω_i represented by the samples x_k, where $\omega_i \in \{\omega_1, \omega_2, \dots, \omega_c\}$.

2. THE MAXIMUM LIKELIHOOD (ML) CLASSIFIER

Bayes theory provides a minimum error-rate classifier by finding the maximum among c *a posteriori* probabilities $\mathcal{P}(\omega_j|x)$, $j = 1, 2, \dots, c$. If all the modulations are equally-likely *a priori*, the optimal Bayes classifier reduces to the ML classifier. The ML classifier chooses the modulation of the samples $x = (x_1, x_2, \dots, x_{N_S})$ as the one that maximizes the probability density functions $p(x|\omega_j)$, using the I and Q samples as sufficient statistics. Such problem was studied in [8] in the ideal situation where A, h, f_r, θ and σ_n^2 are known. This unrealistic situation has provided an upper bound of the expected performance for a digital modulation classifier. The ML classifier can be summarized as follows:

Assign x to ω_i if $l(x|\omega_i) \geq l(x|\omega_j), \forall j$, where

$$l(x|\omega_j) = \sum_{k=1}^{N_s} \ln \left\{ \frac{1}{M_j} \sum_{i=1}^{M_j} \exp \left(-\frac{1}{\sigma_n^2} |x_k - S_i|^2 \right) \right\} \qquad (3)$$

is obtained after dropping constants in the log-likelihood of the observed signal.

3. THE BAYES MINIMUM ERROR-RATE CLASSIFIER

This paper generalizes the work carried out by Wei and Mendel to a more practical scenario, where the parameter vector denoted $\theta = (f_r, \phi, h)$ is unknown. In other words, the proposed algorithm mitigates the effects of phase and frequency offsets as well

as residual filtering effects coming from imperfect channel equalization. A common strategy for this kind of problem is to replace the unknown parameters in the density functions by their estimated values, resulting in the so-called ML plug-in classifier ([9], p. 267). Here, instead,we take a fully Bayesian route.

Under the assumption of uniform prior probabilities for the digital modulations, the posterior probabilities $\mathcal{P}(\omega_i|x)$ can be written as

$$\mathcal{P}(\omega_i|x) \propto p(x|\omega_i) \qquad (4)$$

where the proportionality constants of all the probabilities are the same and equal to $1/(cp(x))$, where c is the number of different classes and $p(x)$ is the probability density function of the observations x. The marginalized density of the data given the class ω_i can be obtained from

$$p(x|\omega_i) = \int p(x|\theta_i, \omega_i) p(\theta_i|\omega_i) d\theta_i \qquad (5)$$

where $p(\theta_i|\omega_i)$ is the prior density of θ_i, and $p(x|\theta_i, \omega_i)$ is the likelihood function. Thus, the classification of the modulations is carried out according to

$$\hat{\omega} = \arg\max_{\omega_i} \left(\int p(x|\theta_i, \omega_i) p(\theta_i|\omega_i) d\theta_i \right). \qquad (6)$$

Unfortunately, a closed-form expression of (6) can rarely be obtained. This paper proposes the use of MCMC methods to compute the multidimensional integrals in (5).

4. MCMC METHODS: THE METROPOLIS-HASTINGS (MH) ALGORITHM

MCMC methods have recently received much attention in signal processing applications [10] [11]. These numerical techniques consist of generating samples by running an ergodic Markov Chain whose target distribution is a desired distribution. A critical step in our classification procedure is the computation of integrals of the form given by (5). It can be shown that they can be evaluated by Monte Carlo integration using importance sampling according to

$$p(x|\omega_i) \simeq \frac{1}{N} \sum_{n=1}^{N} \frac{p(x|\theta_i^n, \omega_i) p(\theta_i^n|\omega_i)}{g(\theta_i^n|x, \omega_i)} \qquad (7)$$

where θ_i^n is the n-th sample drawn from $g(\cdot)$, i.e.,

$$\theta_i^n \sim g(\theta_i|x, \omega_i)$$

and $g(\cdot)$ is an importance density function.

The MCMC methods can play important role in the computation of integrals according to (7) in that they generate the samples θ_i^n from the importance density $g(\theta_i|x, \omega_i)$. In this paper we use the MH algorithm to draw samples distributed according to the *a posteriori* density of θ. The MH algorithm defines an irreducible, aperiodic Markov chain whose target distribution is the invariant distribution $p(\theta|x, \omega_i)$. It should be noted that not all the samples drawn by the chain are good samples. The set of samples in the beginning of the simulation, also known as "burn-in" samples, are discarded because the chain needs time to converge. The MH algorithm is summarized below. The Markov chain state space and current state are denoted by Ω and $\theta^n = (f_r^n, \phi^n, h_1^n, \ldots, h_p^n)$

$\in \Omega$, respectively. At each iteration, a candidate z is drawn according to an instrumental distribution $q(z|\theta^n)$. This candidate is accepted with the acceptance probability

$$\alpha(\theta^n, z) = \min\left\{1, \frac{p(z|x)q(\theta^n|z)}{p(\theta^n|x)q(z|\theta^n)}\right\}. \qquad (8)$$

A fundamental property of the MH algorithm is that any instrumental distribution $q(z|\theta^n)$ can be chosen, provided that the support of $p(\cdot|x)$ is contained in the support of $q(z|\theta^n)$ [12]. This paper proposes to draw z from a local perturbation of the previous sample, i.e., $z = \theta^n + \epsilon$, leading to the well-known random-walk MH algorithm. In this case, the instrumental distribution is of the form $q(z|\theta^n) = g(z - \theta^n)$. Interestingly, the choice of a symmetric distribution for g leads to an acceptance probability which is independent of q.

Instead of updating the whole of θ *en bloc*, it is often more convenient and computationally efficient to divide θ into k blocks and to update each block one-at-a-time. This procedure has been suggested by many authors (see [13] [14] for more details) and has been shown to improve the convergence of the sampler. This paper proposes to update each component of θ one-at-a-time. Such strategy exhibits good performance for the digital modulation classification, as shown in the next section.

5. SIMULATION RESULTS

Many simulation results have been performed to illustrate the performance of the MCMC based digital modulation classifier. For these experiments, the number of samples is $N_S = 250$, and the additive noise is complex white and Gaussian. The signal-to-noise ratio in decibels is defined as $SNR = \log_{10}(1/\sigma_n^2)$ (constellation symbols had unit energy). The MCMC sampler had the following properties:
- number of "burn-in" iterations: $N_{bi} = 500$,
- Markov chain length: $N = 3000$,
- instrumental distributions: $q(z|\theta_i^n) \sim \mathcal{N}(\theta_i^n, \sigma^2)$ where σ^2 was optimized to obtain an appropriate acceptance rate ($1/4$ to $1/2$, see ([12], p. 8) for more details).

The following priors have been used for the parameters:
- Uninformative independent uniform priors for the frequency and phase offsets : $p(f_r, \phi) = p(f_r)p(\phi)$ where $p(f_r) = I_{(-1/2, 1/2]}(f_r)$, $p(\phi) = \frac{M}{2\pi}I_{[-\pi/M, \pi/M]}(\phi)$ for an M-PSK modulation, $p(\phi) = \frac{2}{\pi}I_{[-\pi/4, \pi/4]}(\phi)$ for other modulations, and I is the indicator function.
- Independent Gaussian priors $\mathcal{N}(0, \sigma_h^2)$ for the residual channel FIR filter taps: a suitable choice of parameter σ_h^2 allows to incorporate vague prior information about the parameters h_l.

A. Convergence of the Sampler

Convergence assessments have to be investigated to determine whether the Markov chain has converged to the stationary distribution. Convergence diagnostics for MCMC methods have received increasing interest in the literature [12]. Gelman *et al.* [15] have suggested to study different key-parameters of interest with iteration number. Fig. 1 shows a typical Markov chain obtained for $h = (1, 0, 0)$, $\phi = 0$, and $f_r = -0.35$. Clearly, the Markov chain samples start to oscillate around a value which is quite close to the true value of f_r after the 200-th iteration. This result illustrates the Markov Chain convergence for the parameter f_r. Similar results have been obtained for h and ϕ. Therefore, for simplicity, all

simulations have been conducted with $N_{bi} = 500$ "burn-in" iterations, which provides some safety margin. (Of course, some of the convergence diagnosis methods such as those described in [12], could have been used instead.) Fig. 2 shows the mean square error (MSE) between the Markov Chain target distribution (computed from 100000 iterations) and the estimated frequency offset posterior distribution as a function of the iteration number. As can be seen, $N = 3000$ samples are sufficient to approximate this posterior distribution. Based on these results, all simulations presented in this paper have been carried out with Markov chains of length $N = 3000$ and $N_{bi} = 500$ "burn-in" iterations.

B. Comparisons

This section compares the MCMC based classifier with two other classifiers: the ML classifier [8] and a 4th order cumulant-based classifier [3]. Note that both classifiers do not take into account the presence of the residual effects (frequency offset, phase offset and residual channel), contrary to the proposed approach. The ML classifier has been presented in Section 2. The 4th order cumulant based classifier is summarized below for classification of BPSK, 4-PAM, 8-PSK and 16-QAM constellations:

$$|\hat{C}_{40}| < 0.34 \quad \Rightarrow \quad 8PSK$$
$$0.34 \leq |\hat{C}_{40}| < 1.02 \quad \Rightarrow \quad 16QAM$$
$$1.02 \leq |\hat{C}_{40}| < 1.68 \quad \Rightarrow \quad 4PAM$$
$$|\hat{C}_{40}| \geq 1.68 \quad \Rightarrow \quad BPSK$$

where $C_{40} = E[x^4(n)] - 3E[x^2(n)]^2$ (see [3] for motivations).

Table 1 shows representative confusion matrices for the classification of BPSK, 4-PAM, 8-PSK and 16-QAM modulations, obtained for a sample size $N_S = 250$, $SNR = 5dB$, $f_r = 0.4$, $\phi = 0$, and $h = (1, 0, 0)$. The MCMC classifier clearly outperforms the ML and HOS classifiers. The robustness of the classifier to frequency offset is depicted in Fig. 3. The MCMC classifier clearly outperforms the ML and HOS classifier for large values of the frequency offset. Figures 4 and 5 show better results for the MCMC classifier for all values of the frequency offset and $h = (1, 0.25, 0.15)$ for $SNR's = 5dB$ and $0dB$ respectively. Fig. 6 shows similar results for the robustness to residual channel effects.

Classifier Input	Classifier Output			
	BPSK	4-PAM	8-PSK	16-QAM
BPSK	0	0	99	1
4-PAM	0	0	0	100
8-PSK	0	0	99	1
16-QAM	0	0	18	82
BPSK	16	83	0	6
4-PAM	11	78	0	11
8-PSK	0	0	83	17
16-QAM	0	0	61	39
BPSK	100	0	0	0
4-PAM	24	76	0	0
8-PSK	0	0	100	0
16-QAM	0	0	37	63

Table 1. Confusion matrices for ML, HOS and MCMC classifiers (top to bottom), with $N_S = 250$, $SNR = 5dB$, 100 trials and $f_r = 0.4$, $\phi = 0$, $h = (1, 0, 0)$.

6. CONCLUSIONS

This paper addressed the problem of digital modulation classification in the presence of frequency and phase offsets and residual channel effects. The proposed approach consisted of estimating the class posterior probabilities by using samples generated with the Metropolis-Hastings algorithm. This strategy resulted in the so-called MCMC Bayes classifier. The MCMC Bayes classifier was shown to outperform two well known powerful digital modulation classifiers. Further generalization of the MCMC Bayes rule to the problem of digital modulation classification that includes analysis of signals propagated through non-linear channels is under consideration.

7. ACKNOWLEDGMENT

8. REFERENCES

[1] C. Le Martret and D. Boiteau, "A general maximum-likelihood classifier for modulation classification," in *Proc. EUSIPCO*, Rhodes, Greece, 1998, pp. 985–988.

[2] B. G. Mobasseri, "Digital modulation classification using constellation shape," *Signal Processing*, vol. 80, pp. 251–277, 2000.

[3] A. Swami and B. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Trans. Comm.*, vol. 48, no. 3, pp. 416–429, 2000.

[4] A. Swami and B. Sadler, "Modulation classification via hierarchical agglomerative cluster analysis," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications*, Paris, France, April. 1997, vol. 80, pp. 141–144.

[5] S. Soliman and S.-Z. Hsue, "Signal classification using statistical moments," *IEEE Trans. Signal Processing*, vol. 44, pp. 2793–2800, Nov. 1996.

[6] K. Assaleh, K. Farrell, and R.J. Mammone, "A new method of modulation classification for digitally modulated signals," in *Proc. MILCOM*, 1992, pp. 712–716.

[7] L. Hong and K. Ho, "Identification of digital modulation types using the wavelet transform," in *Proc. IEEE MILCOM 99*, 1999.

[8] W. Wei and J. M. Mendel, "Maximum likelihood classification for digital amplitude-phase modulations," *IEEE Trans. On Com*, Feb. 2000.

[9] L. Devroye, L. Gyorfi, and G. Lugosi, *A Probabilistic Theory of Pattern Recognition*, Berlin: Springer-Verlag, 1996.

[10] "Special issue on MCMC methods in signal processing," *Signal Processing*, to appear.

[11] "Special session on Bayesian methods," in *Proc. IEEE Workshop on Higher Order Statistics*, Caesarea, Israel, June 14-16 1999.

[12] C. P. Robert, *Discretization and MCMC Convergence Assessment*, Berlin: Springer-Verlag, 1998.

[13] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, "Introducing Markov chain Monte Carlo," in *Markov Chain Monte Carlo in Practice*, pp. 1–19. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Eds.. London, U.K.: Chapman & Hall, 1996.
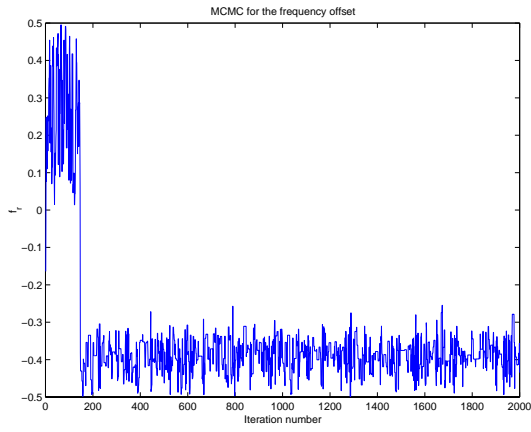
[14] S. Chib and E. Greenberg, "Understanding the Metropolis algorithm," *The American Statistician*, vol. 49, pp. 327–335, 1995.

[15] A. Gelman, "Inference and monitoring convergence," in *Markov Chain Monte Carlo in Practice*, pp. 132–143. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Eds.. London, U.K.: Chapman & Hall, 1996.
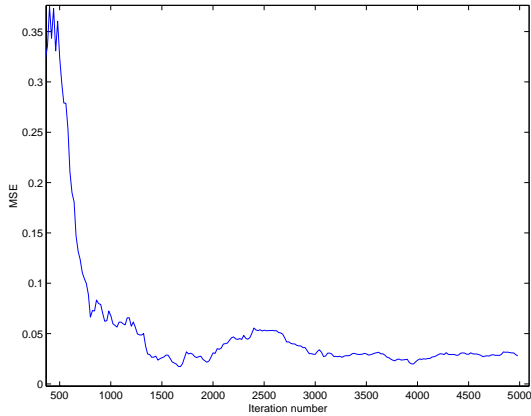
Fig. 1. MCMC samples for the frequency offset.

Fig. 2. MSE between target and estimated posterior distributions vs iteration number.
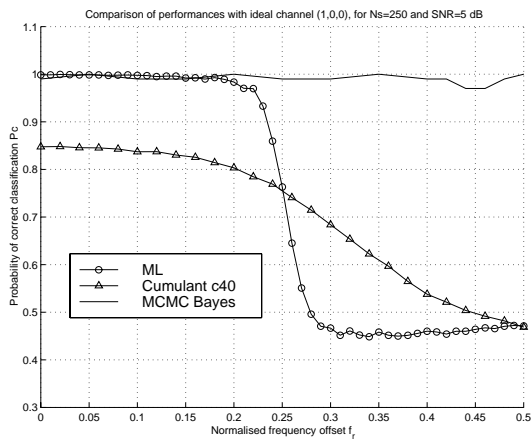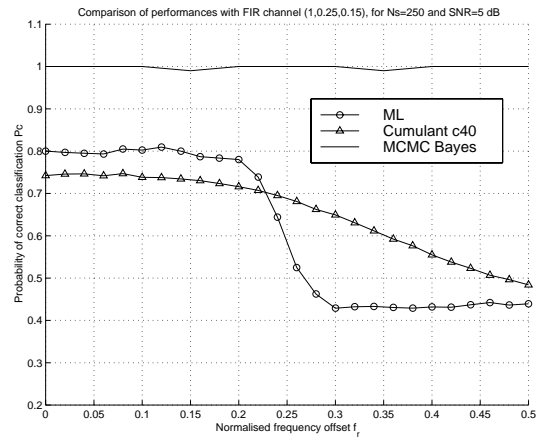

Fig. 3. Probability of correct classification vs frequency offset for the ML, HOS and MCMC classifiers (SNR=5dB, h and f_r are both estimated).

Fig. 4. Probability of correct classification vs frequency offset for the ML, HOS and MCMC classifiers (SNR=5dB, h and f_r are both estimated).
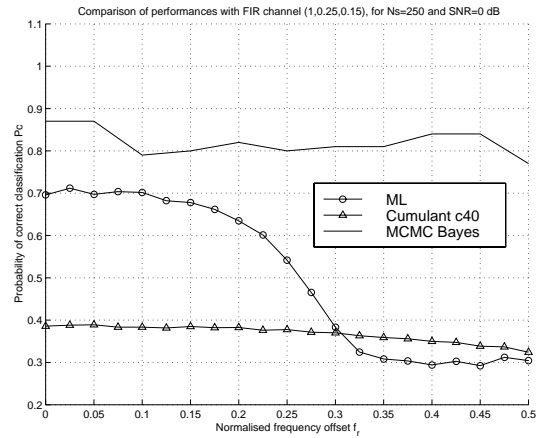

Fig. 5. Probability of correct classification vs frequency offset for the ML, HOS and MCMC classifiers (SNR=0dB, h and f_r are both estimated).
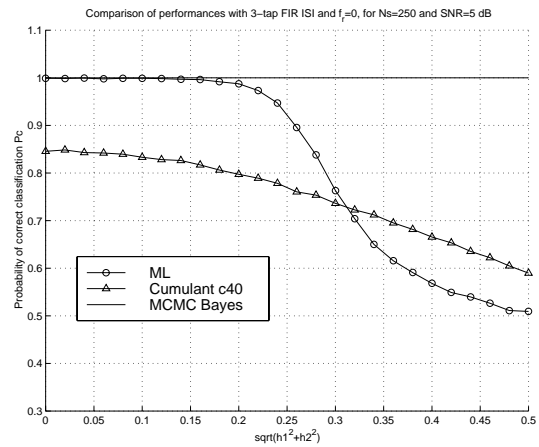

Fig. 6. Probability of correct classification vs residual channel modulus for the ML, HOS and MCMC classifiers (SNR=5dB, only h is estimated).