Classification of linear and non-linear modulations using the Baum–Welch algorithm and MCMC methods ☆

Anchalee Puengnim [a], Nathalie Thomas [a,*], Jean-Yves Tourneret [a,♣], Josep Vidal [b]

[a] IRIT/ENSEEIHT/TéSA, 2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France
[b] Universitat Politècnica de Catalunya, Campus Nord-UPC, Jordi Girona 1-3, 08034 Barcelona, Spain

Keywords:
Digital modulation classification
Linear and non-linear modulations
Bayesian classifiers
Markov chain Monte Carlo methods
Baum–Welch algorithm

A B S T R A C T

Satellite transmissions classically use constant amplitude linear modulation schemes, such as M-state phase shift keying (M-PSK), because of their high robustness to amplifier non-linearities. However, other modulation formats are interesting in a satellite transmission context. For instance, non-linear modulations such as Gaussian minimum shift keying (GMSK) present a higher spectral efficiency and appear in new standards for telemetry/telecommand satellite links. Another example is offset-QPSK (OQPSK) modulation that allows one to decrease the out-of-band interference due to band limiting and the non-linearity of the amplifier. To get a compromise between the robustness to amplifier non-linearities provided by MPSK modulation and the spectral efficiency given by QAM modulation, the recent broadcasting satellite standard (DVB-S2) proposes new modulation schemes called APSK. Obviously, all satellite systems that use various modulation schemes will have to co-exist. In this context, modulation recognition using the received communication signal is essential. In that context, this paper studies two Bayesian classifiers to recognize linear and non-linear modulations. These classifiers estimate the posterior probabilities of the received signal, given each possible modulation, and plug them into the optimal Bayes decision rule. Two algorithms are used for that purpose. The first one generates samples distributed according to the posterior distributions of the possible modulations using Markov chain Monte Carlo (MCMC) methods. The second algorithm estimates the posterior distribution of the possible modulations using the Baum–Welch (BW) algorithm. The performance of the resulting classifiers is assessed through several simulation results.

1. Introduction

Satellite transmissions classically use constant amplitude modulation schemes because of their high robustness to amplifier non-linearities. Linear phase modulation, or M-state phase shift keying (M-PSK), are the most widely used. However, some non-linear modulation formats appear in new standards for satellite communications. For example, Gaussian minimum shift keying (GMSK) is a new modulation standard for telemetry/telecommand satellite links. This class of modulation is very interesting for satellite transmissions. It is actually very robust to amplifier non-linearities. With proper choice of parameters, it also allows one to obtain higher spectral efficiency than that obtained with the traditional M-PSK schemes. Particularly, the choice of a pre-modulation Gaussian filter, associated to a modulation index $h=\frac{1}{2}$,

☆ This work was partially supported by the French Space Agency (CNES, Toulouse, France).
* Corresponding author.
E-mail addresses: Anchalee.Puengnim@enseeiht.fr (A. Puengnim), nathalie.thomas@enseeiht.fr (N. Thomas), jean-yves.tourneret@enseeiht.fr (J.-Y. Tourneret), josep.vidal@upc.edu (J. Vidal).
♣ A member of EURASIP.

leads to a modulation with a very interesting compact power spectral density. Two different GMSK schemes, characterized by two different normalized bandwidths BT, have been adopted by the consultative committee for space data system (CCSDS) for future space missions [1]. More precisely, for the packet telemetry for space-to-earth links, the CCSDS recommends the GMSK modulation with BT=0.25 (denoted GMSK25) for spacecrafts orbiting at the altitude below 2×10^6 km and the GMSK modulation with BT=0.5 (denoted GMSK50) at the altitude above 2×10^6 km. Another interesting modulation for satellite transmissions is the offset-QPSK or OQPSK. This modulation is a linear phase modulation similar to QPSK, except that some phase changes are not allowed between two consecutive symbols. This property leads to OQPSK modulated signals that are less sensitive to spectral sidelobe spreading than QPSK signals.

Thus the out-of-band interfering due to band limiting and the non-linearity of the amplifier is decreased. Quadrature amplitude modulations (QAM) are very interesting compared to PSKs because of their higher spectral efficiency. However, they are not use in a satellite context because of their sensitivity to amplifier non-linearities. The new satellite standard for broadcasting referred to as DVB-S2 defines a kind of compromise between PSK and QAM modulations called APSK. APSK are amplitude and phase modulations defined by a reduced set of possible amplitude values compared to QAM. As all satellite systems using various modulation formats will have to co-exist, it is important to be able to recognize the modulation associated to the received communication signal. In particular, identifying the modulation is important for spectrum monitoring to check whether the user is authorized to send this modulation. The modulation classification problem is also interesting in non-cooperative scenarios where there is an emerging need for intelligent modem capable of quickly discriminated signal types [2]. The application considered in this paper is spectrum monitoring. Our objective is to identify the constellation of a received communication signal assuming this constellation belongs to a known dictionary. Various strategies have been proposed in the literature for the classification of linear modulations. The most popular modulation classifier (often referred to as *optimal classifier*) is probably the Bayes classifier that minimizes the average probability of error (or an appropriate average cost function). However, the Bayes classifier is difficult to implement due to its high computational complexity. Moreover, it is not robust to model mismatch due to transmission impairments, such as synchronization errors or residual channel. To overcome the difficulties inherent to the Bayesian classifier, several suboptimal likelihood-based classifiers have been proposed in the signal processing and communication literature (see for example [3–6]). An alternative to likelihood based classifiers is to extract interesting features from the observations and to use these features for classification. In this case, the key point is to find the "appropriate" set of features depending on the considered communication system. Many features have been proposed in the literature, including statistical moments [7] or higher-order statistics [2].

This paper studies new strategies to classify linear and non-linear modulations. The first strategy is based on a practical suboptimal Bayes classifier using a "plug-in" rule initially proposed in [8]. It can be applied to recognize M-PSK classical schemes, as well as M-QAM (M states QAM) or M-APSK (M states APSK) modulations. The main idea is to estimate the unknown model parameters by Bayesian estimation combined with Markov chain Monte Carlo (MCMC) methods. The estimated parameters are then plugged into the posterior probabilities of the received modulated signal (conditionally to each class). The classical maximum *a posteriori* (MAP) classification rule is finally implemented with these estimated probabilities. Unfortunately, the complexity of this MCMC classifier may be prohibitive for some practical applications. To overcome this difficulty, we consider a new digital modulation classifier based on hidden Markov models (HMMs) to classify linear modulations transmitted through an unknown finite memory channel and corrupted by additive white Gaussian noise (AWGN). This classifier is based on a state trellis representation, allowing one to use a modified version of the Baum–Welch (BW) algorithm (proposed in [9] for speech recognition) to estimate the posterior probabilities of the possible modulations. These posterior probabilities are then plugged into the optimal Bayes decision rule. This BW classifier, initially introduced in [10], is interesting since it can be used to recognize OQPSK modulation from other linear phase modulations. Indeed, since some transitions are not allowed in case of OQPSK, a distinct state trellis representation from QPSK can be defined. The BW classifier then exploits this state trellis representation for modulation classification.

Classifying non-linear modulations has received less attention in the literature. Several methods for classifying full response binary CPMs with rectangular pulse shape and different modulation indexes have been studied in [11,12]. A classifier based on an approximate likelihood function for a multiple M-ary frequency shift keying (MFSK) signal propagating through a Rayleigh fading channel has been developed in [13]. However, classification problems involving GMSK modulations have not been considered in the literature (to the best of our knowledge). Exploiting the fact that GMSK modulation is a modulation with memory, the BW classifier can be used for classifying the two non-linear GMSK modulations recommended by CCSDS. Finally, we show that linear modulations used in satellite systems (BPSK, QPSK, 8PSK, OQPSK), as well as the non-linear standardized GMSK modulation schemes, can be identified using the same recognition process (the problem was initially introduced in [14]). By associating a first order HMM to the received baseband communication signal, the BW classifier can be used to estimate the posterior distribution of the received GMSK communication signals. The BW method for HMM can also provide as a by-product the sequence of estimated transmitted symbols using the MAP criterion.

This paper is organized as follows: Section 2 gives some useful information regarding the linear and non-linear modulations considered in this study. Section 3 presents a model of the received baseband communication

signal, including some practical channel impairments. Section 4 recalls the classical MAP classification rule. Sections 5 and 6 present the MCMC classifier and the BW classifier. Simulation results and conclusions are reported in Sections 7 and 8.

2. Linear and non-linear modulations

The emitted signal $s(t)$ can be written as

$$s(t) = \mathrm{Re}[\tilde{s}(t)e^{j\omega_c t}],$$

where $\tilde{s}(t) = I_s(t,\mathbf{a}) + jQ_s(t,\mathbf{a})$ is the complex envelope (or equivalent low-pass signal) associated to $s(t)$, $\mathbf{a}=\{a_k, k=1,\ldots\}$ is the independent identically distributed (i.i.d.) complex symbol sequence to be transmitted and $\omega_c = 2\pi f_c$, where f_c is the carrier frequency. Note also that $\mathrm{Re}(z)$ denotes the real part of the complex number z. The modulation is called *linear* when $\tilde{s}(t)$ linearly depends on $\mathbf{a}$, and *non-linear* in the other cases.

2.1. Linear modulations

The baseband complex envelope of a linearly modulated signal can be written as

$$\tilde{s}(t) = \sum_k a_k h(t-kT),$$

where $h(t)$ is the pulse shape filter impulse response and T is the symbol duration. The i.i.d. complex symbol sequence $\mathbf{a}$ to be transmitted takes its values into a set of M_j complex numbers $\{S_0, S_1, \ldots, S_{M_j}\}$ called *constellation* representing a particular modulation.

2.1.1. Linear M-PSK modulations

classical M-PSK modulations are defined for $M \geq 4$ by

$$S_m = \exp\left[j\left(2\pi\frac{m}{M} + \frac{\pi}{M}\right)\right], \quad m = 0,\ldots,M-1,$$

whereas BPSK modulation (M=2) is defined by $S_m = \exp(jm\pi), m = 0,\ldots,M-1$. For instance, BPSK ($M$=2), QPSK ($M$=4) and 8-PSK ($M$=8) constellations that will be considered in this paper are displayed in Fig. 1.

2.1.2. OQPSK modulation

Another important linear phase modulation is the offset-QPSK or OQPSK. OQPSK is similar to QPSK except that the I- and Q-channel pulses are offset in time by $T/2$ s avoiding a simultaneous change at the symbol boundaries. Unlike QPSK signal whose phase changes at the symbol boundaries can be 0°, $\pm 90^\circ$, and 180°, the phase changes at the symbol boundaries of OQPSK signal can only be 0° and $\pm 90^\circ$. The consequence is to obtain an OQPSK signal that is less sensitive to spectral sidelobe spreading than a QPSK signal. Thus the out-of-band interfering due to band limiting and the non-linearity of the amplifier is decreased. As an example, Fig. 2 compares the possible transitions from the state $S_2 = \exp(j\pi/2)$ for QPSK and OQPSK modulations. OQPSK is a particular case of phase modulation in the sense that it is a modulation with memory.

2.1.3. Linear M-QAM modulation

Quadrature amplitude modulations with M possible symbols (M-QAM) are defined by $S_m = I_m + jQ_m$, where I_m and Q_m are independent and take their values in $\{\pm 1, \pm 3, \ldots, \pm(\sqrt{M}-1)\}$. This paper will consider 16QAM (M=16) constellation that was considered in [2] in a

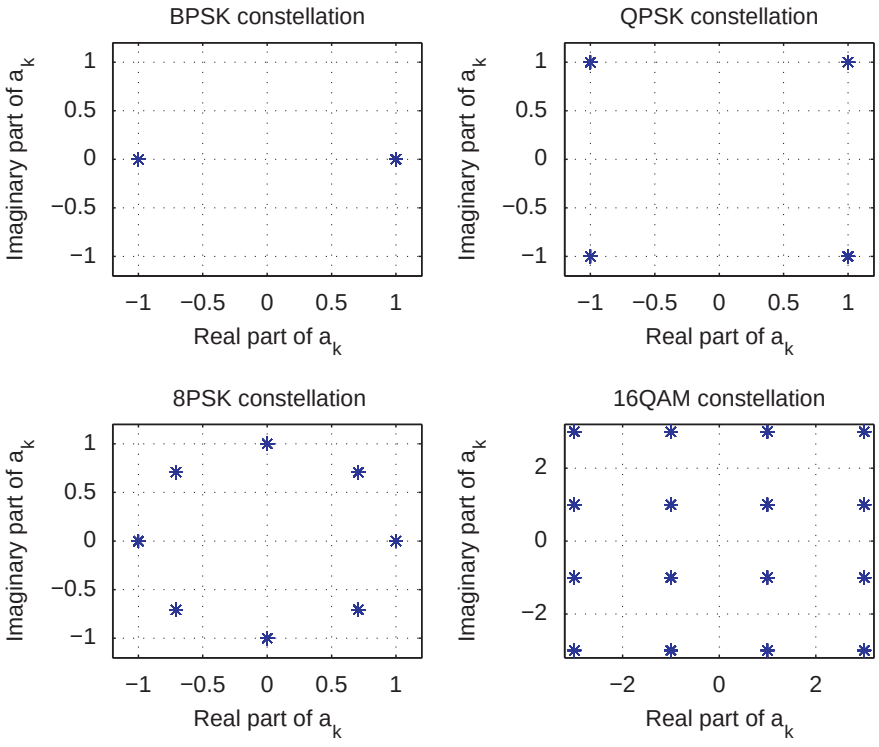

Fig. 1. Classical linear modulation constellations.

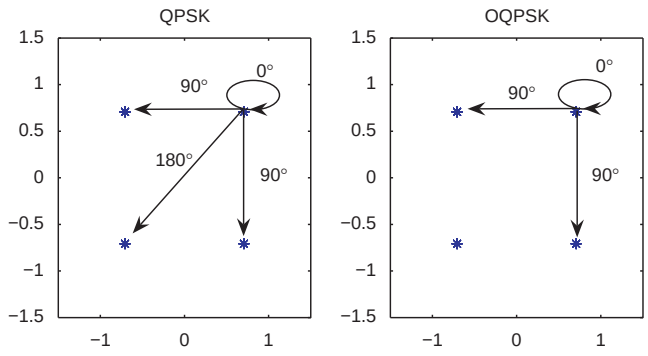

Fig. 2. Constellations and phase changes for QPSK and OQPSK.

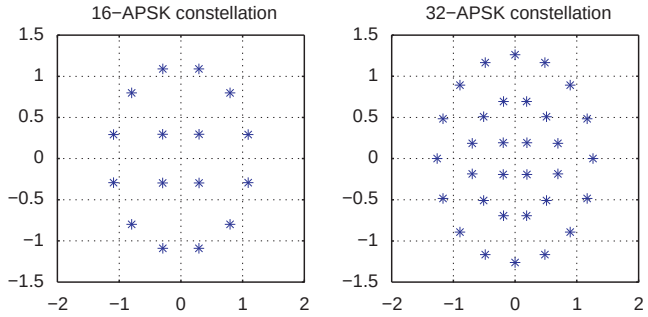

Fig. 3. APSK modulation constellations.

context of digital modulation classification. This constellation is illustrated in Fig. 1.

2.1.4. APSK modulation

Amplitude and phase shift keying modulations with M possible symbols (M-APSK) are defined by

$$S_m = R_i \exp\left[j\left(2\pi\frac{m}{n_i}+\theta_i\right)\right], \quad m=0,\ldots,n_i-1,\ i=0,\ldots,R-1,$$

where n_i is the number of symbols having the same amplitude value R_i (they are differentiated thanks to the phase $2\pi m/n_i+\theta_i$) and R is the number of possible amplitude values for the given APSK constellation. The standard DVB-S2 uses 16-APSK (or 4-12-APSK, with R=2, n_0=4 and n_1=12) and 32-APSK (or 4-12-16-APSK, with R=3, n_0=4, n_1=12 and n_2=16) constellations that are illustrated in Fig. 3.

2.2. Non-linear GMSK modulations

This section recalls the principles of GMSK modulations (the reader is invited to consult [15,16] for more details). The GMSK signal is a partial response CPM signal with modulation index $\frac{1}{2}$ and a smooth shape frequency pulse $g(t)$ of length LT, where $L \in \mathbb{N}$. The function $g(t)$ is the global impulse response of two consecutive filters. The first filter is rectangular of length T whereas the second one is Gaussian with a normalized 3 dB bandwidth BT. The global impulse response of the two filters

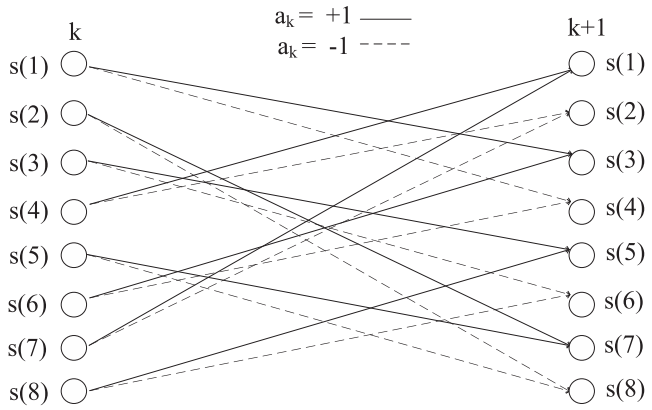

Fig. 4. State trellis diagram of GMSK signal, BT=0.5, L=2, M=2.

can be written:

$$g(t)=\frac{1}{2T}\left\{Q\left(2\pi BT\frac{t-\frac{T}{2}}{T\sqrt{\ln 2}}\right)-Q\left(2\pi BT\frac{t+\frac{T}{2}}{T\sqrt{\ln 2}}\right)\right\},$$

where $Q(t)=\int_t^{\infty}(1/\sqrt{2\pi})\exp(-(\tau^2/2))\,d\tau$.

The complex envelope of the GMSK signal can be written as $\tilde{s}(t)=e^{j\Phi(t,\mathbf{a})}$, where the transmitted i.i.d. symbol sequence $\mathbf{a}$ taken from $\{\pm 1,\ \pm 3,\ldots,\pm(M-1)\}$ is embedded in the time-varying phase

$$\Phi(t,\mathbf{a})=\pi\sum_k a_k q(t-kT),$$

and where $q(t)=\int_{-\infty}^{t} g(\tau)\,d\tau$. For $t\in[kT,(k+1)T]$, the time-varying phase can be written

$$\Phi(t,\mathbf{a})=\theta_k(t,\mathbf{a})+\phi_k,$$

where

- $\theta_k(t,\mathbf{a})=\pi\sum_{i=k-L+1}^{k} a_i q(t-iT)$ represents the changing part of the time-varying phase in $[kT,(k+1)T]$ and
- $\phi_k=[(\pi/2)\sum_{i=-\infty}^{k-L} a_i] \bmod (2\pi)$ is the cumulative phase (where $[x]\bmod(2\pi)$ denotes the angle of x modulo 2π). The phase ϕ_k represents the constant part of the time-varying phase in $[kT,(k+1)T]$ that can be recursively computed as $\phi_{k+1}=\phi_k+\pi a_{k-L+1}$.

A *state* of the GMSK signal is classically defined at $t=kT$ as $s_k=\{\phi_k,a_{k-1},a_{k-2},\ldots,a_{k-L+1}\}$. The state s_k corresponds to a specific value of the time-varying phase $\Phi(kT,\mathbf{a})$. The number of states of a GMSK signal is $N=4M^{L-1}$ and the different states will be denoted as $s(1),s(2),\ldots,s(N)$. Fig. 4 displays an example of the state trellis representing a GMSK modulation with parameters BT=0.5, L=2 and M=2, i.e., the transitions from time instant kT to time instant $(k+1)T$ (the number of states is N=8 for this example). The set of possible states provides a set of possible values for the complex envelop of the GMSK modulated signal taken at $t=kT$. This set can be assimilated to a kind of *constellation*, as shown in Fig. 5 for the two standardized GMSK modulations corresponding to BT=0.5 and 0.25.

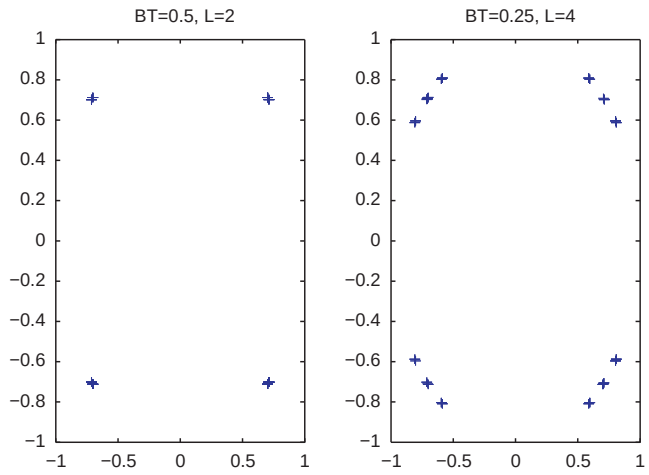

Fig. 5. GMSK *constellations* displayed for the two standardized cases: BT=0.5, L=2, M=2 and BT=0.25, L=4, M=2.

3. Model of the received signal

3.1. AWGN transmission channel

The AWGN channel assumes that the emitted signal $s(t)$ is corrupted by a white Gaussian noise $w(t)$ with power spectral density $N_0/2$. The associated complex baseband Gaussian noise process will be denoted as $\tilde{w}(t)$. The received signal $r(t)$ is first down-converted by the receiver to recover its complex envelope $\tilde{r}(t)$. Fig. 6 recalls the structure of the standard down-convertor used in this study [16]. After down-conversion, the received baseband signal $\tilde{r}(t)$ can be written

$$\tilde{r}(t) = I_r(t,\mathbf{a}) + jQ_r(t,\mathbf{a}) = \tilde{s}(t) * f(t) + z(t), \quad t \in \mathbb{R},$$

where $f(t)$ is the impulse response of the two low-pass (LP) filters, $z(t) = \tilde{w}(t) * f(t)$ is a normalized complex-valued additive Gaussian noise with variance σ_z^2 and $*$ denotes convolution.

Assuming a perfect synchronization between the emitter and the receiver, the complex envelope of the received modulated signal, sampled at one sample per symbol (t=kT), can be written

$$\tilde{r}(k) = \tilde{s}(k) * f(k) + z(k), \quad k = 1, \ldots, N_s, \tag{1}$$

where N_s is the number of symbols in the observation interval. In absence of noise, the received constellations for linear modulations are exactly the same as the emitted ones when the Nyquist criterion is satisfied. The situation is different when the emitted signals are GMSK modulated. Fig. 7 shows emitted and received *constellations* associated to the two standardized GMSK modulations, when square root raised cosine LP filters are used in the down conversion process. Note that these constellations have been obtained in absence of noise, with a roll-off factor $\alpha = 0.35$ and a cutoff frequency adapted to symbol duration. Fig. 7 indicates that the received signals corresponding to GMSK modulations and QPSK modulations are very similar in the presence of additive noise. However, the classification rule proposed in this paper will allow us to distinguish these modulations.

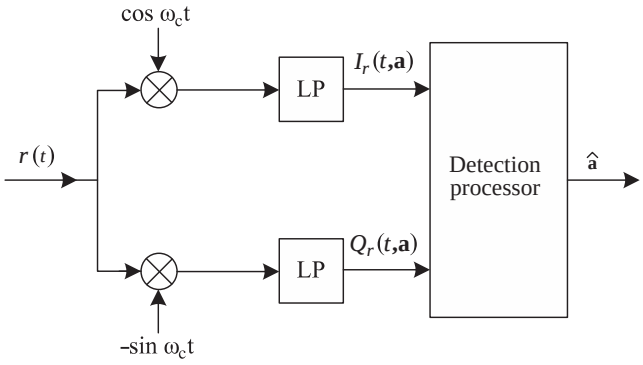

Fig. 6. Basic quadrature receiver.

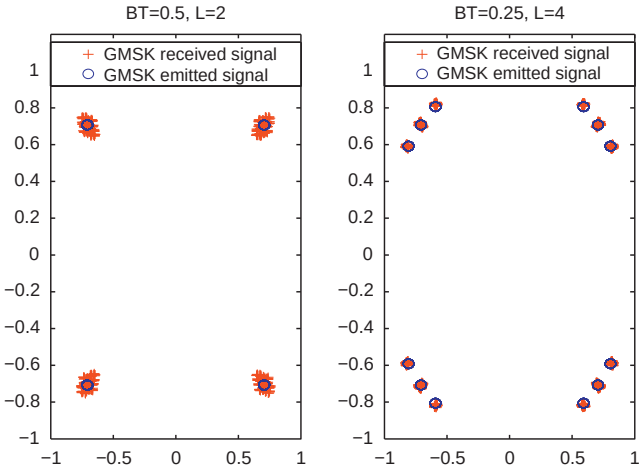

Fig. 7. Emitted (circles) and received (crosses) GMSK *constellations* displayed for the two standardized cases BT=0.5, L=2, M=2 and BT=0.25, L=4, M=2.

The baseband complex envelope of the received modulated signal in (1) can be rewritten as

$$\tilde{r}(k) = d(k) + z(k), \quad k = 1, \ldots, N_s \tag{2}$$

where $d(k) = \tilde{s}(k) * f(k)$. In the case of a linearly modulated signal, when the transmitter and receiver filters are matched, $d(k)$ is an i.i.d. symbol sequence taking its values in the set of the jth emitted constellation points $\{S_1, S_2, \ldots, S_{M_j}\}$. For a GMSK modulated signal, $d(k)$ takes its values in a different set (with $4M^{L-1}$ points) depicted in Fig. 7.

3.2. Channel impairments

In some scenarios, the received signal may suffer from various impairments leading to more complex classification problems. As an example, the noisy received communication signal may be affected by carrier frequency and phase errors and a residual channel due to imperfect equalization. In that case, the baseband complex envelope of the received signal sampled at one sample per symbol at the output of a matched filter (linear modulation case) can be written as in [2]

$$\tilde{r}(k) = e^{j(\pi(k/N_s)f_r + \phi)} \sum_{l=0}^{q} h_l d(k-l) + z(k), \quad k = 1, \ldots, N_s, \tag{3}$$

where

- $f_r = 2N_s(f_c - \hat{f}_c) \in [-\frac{1}{2}, \frac{1}{2}]$ is a carrier frequency offset (f_c is the carrier frequency of the received signal and $\hat{f}_c$ is the frequency of the local oscillator in the receiver). Note that these notations imply that f_r is the constellation rotation whose maximum value is $\pi/2$ for $k=N_s$,
- $\mathbf{h}=[1,h_1,\ldots,h_q]$ is the residual channel coefficient vector, and
- ϕ is the carrier phase offset.

4. MAP classifier

For linear modulations and an AWGN channel, the maximum likelihood (ML) classifier developed by Wei and Mendel [17] minimizes the probability of classification error. Thus this classifier achieves the optimum performance. However, the performance of the ML classifier is significantly reduced in presence of channel impairments. This section briefly recalls the principle of the Bayesian classifier that will be considered in this paper.

Given the samples $\tilde{\mathbf{r}} = [\tilde{r}(1), \ldots, \tilde{r}(N_s)]$, Bayes theory provides a minimum error-rate classifier from the maximum a posteriori probabilities $P(\lambda_j|\tilde{\mathbf{r}})$, $j = 1, \ldots, c$, where $\lambda_1, \ldots, \lambda_c$ denote the possible constellations (c is the number of constellations to be classified). More precisely, the Bayesian classifier is defined by the following rule:

$$\text{assign } \tilde{\mathbf{r}} \text{ to } \lambda_i \text{ if } R(\lambda_i|\tilde{\mathbf{r}}) \leq R(\lambda_j|\tilde{\mathbf{r}}) \quad \forall j = 1, \ldots, c, \tag{4}$$

where $R(\lambda_i|\tilde{\mathbf{r}}) = \sum_{j=1}^{c} c_{i,j} P(\lambda_j|\tilde{\mathbf{r}})$ is the cost function of constellation λ_i and $c_{i,j}$ is the cost of deciding λ_i given that $\tilde{\mathbf{r}} \in \lambda_j$. The MAP classifier is obtained in the special case of $0-1$ cost functions

$$c_{i,j} = \begin{cases} 0 & \text{if } i = j, \\ 1 & \text{if } i \neq j. \end{cases}$$

In that case, the Bayes decision rule (4) can be expressed as

$$\text{assign } \tilde{\mathbf{r}} \text{ to } \lambda_i \text{ if } P(\lambda_i|\tilde{\mathbf{r}}) \geq P(\lambda_j|\tilde{\mathbf{r}}) \quad \forall j. \tag{5}$$

If all modulations are equally likely

$$P(\lambda_j) = \frac{1}{c} \quad \forall j,$$

the MAP classifier reduces to the ML classifier defined as

$$\text{assign } \tilde{\mathbf{r}} \text{ to } \lambda_i \text{ if } p(\tilde{\mathbf{r}}|\lambda_i) \geq p(\tilde{\mathbf{r}}|\lambda_j) \quad \forall j. \tag{6}$$

The ML classifier selects the modulation of the samples $\tilde{\mathbf{r}}$ as the one that maximizes the probability density function (pdf) $p(\tilde{\mathbf{r}}|\lambda_j)$ using the I and Q samples as sufficient statistics (where $\tilde{r}(k) = I(k) + jQ(k)$).

5. Markov chain Monte Carlo (MCMC) classifier

In absence of channel impairment, the ML classifier (6) can be rewritten as follows:

$$\text{assign } \tilde{\mathbf{r}} \text{ to } \lambda_i \text{ if } l(\tilde{\mathbf{r}}|\lambda_i) \geq l(\tilde{\mathbf{r}}|\lambda_j) \quad \forall j, \tag{7}$$

where $l(\tilde{\mathbf{r}}|\lambda_j)$ is obtained by computing the logarithm of the likelihood associated to class λ_j (whose constellation consists of M_j symbols $S_1, \ldots, S_{M_j}$) and by removing additive and multiplicative constants. Straightforward computations allow us to express $l(\tilde{\mathbf{r}}|\lambda_j)$ as follows

$$l(\tilde{\mathbf{r}}|\lambda_j) = \sum_{k=1}^{N_s} \ln\left\{ \frac{1}{M_j} \sum_{i=1}^{M_j} \exp\left(-\frac{1}{\sigma_z^2} \|\tilde{r}(k) - S_i\|^2 \right) \right\}. \tag{8}$$

The ML classifier (7) achieves the optimal solution for the AWGN channel since it minimizes the average classification error. It provides an upper bound of the expected performance for a digital linear modulation classifier. Note that criterion (8) can be viewed as a kind of distance measure between the received symbols contained in $\tilde{\mathbf{r}}$ and the symbol constellation $S_1, \ldots, S_{M_j}$.

In the presence of transmission impairments, a practical solution consists of estimating the likelihoods of the different classes and plugging the resulting estimation in (7). This paper concentrates on the model (3) defined by the unknown parameter vector $\boldsymbol{\theta} = (f_r, \phi, \mathbf{h})$. Eq. (3) can be rewritten as follows:

$$\mathcal{F}^{-1}[e^{-j(\pi(k/N_s)f_r + \phi)} \tilde{r}(k)] = d(k) + \mathcal{F}^{-1}[e^{-j(\pi(k/N_s)f_r + \phi)} z(k)], \tag{9}$$

where $\mathcal{F}^{-1}$ represents the inverse filter associated to the residual channel.

5.1. Approximate likelihood

The first term of the right hand side of (9) is the kth transmitted symbol, whereas the second term is a colored noise whose distribution depends on the parameter vector $\boldsymbol{\theta}$. Neglecting the noise correlations, we propose to approximated the log-likelihood $l(\tilde{\mathbf{r}}|\lambda_j)$ as follows:

$$l(\tilde{\mathbf{r}}|\lambda_j) \simeq \sum_{k=1}^{N_s} \ln\left\{ \frac{1}{M_j} \sum_{i=1}^{M_j} \exp\left(-\frac{1}{\sigma_z^2} \|\tilde{r}^*(k) - S_i\|^2 \right) \right\}, \tag{10}$$

where

$$\tilde{r}^*(k) = \mathcal{F}^{-1}[e^{-j(\pi(k/N_s)f_r + \phi)} \tilde{r}(k)]. \tag{11}$$

This approximation considerably reduces the computational complexity of the log-likelihood. Indeed, an exact computation of $l(\tilde{\mathbf{r}}|\lambda_j)$ would require to evaluate all possible values of $\sum_{l=0}^{q} h_l d(k-l)$ which is clearly too expensive for large values of q (there are M^{q+1} possible values of $\sum_{l=0}^{q} h_l d(k-l)$). Of course the determination of the approximate likelihood in (10) requires to know $\boldsymbol{\theta} = (f_r, \phi, \mathbf{h})$. The estimation of $\boldsymbol{\theta}$ is addressed in the next section.

5.2. Parameter estimation

Estimating the parameter vector $\boldsymbol{\theta}$ can be made by using the method of moments as in [18]. However, ML or Bayesian estimators are often preferred because of their nice asymptotic properties. We propose to estimate the unknown parameter vector $\boldsymbol{\theta} = (f_r, \phi, \mathbf{h})$, associated to impairments described in Section 3, using the minimum mean square error (MMSE) principle. The MMSE estimator of $\boldsymbol{\theta}$ minimizing the quadratic cost function $E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^2]$ is defined as the mean of the posterior distribution $p(\boldsymbol{\theta}|\tilde{\mathbf{r}})$

$$\hat{\boldsymbol{\theta}}_{\text{MMSE}} = E[\boldsymbol{\theta}|\tilde{\mathbf{r}}]. \tag{12}$$

The determination of $p(\boldsymbol{\theta}|\tilde{\mathbf{r}})$ relies on the Bayes rule

$$p(\boldsymbol{\theta}|\tilde{\mathbf{r}}) = \frac{p(\tilde{\mathbf{r}}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\tilde{\mathbf{r}})}, \tag{13}$$

where $p(\boldsymbol{\theta})$ is the prior distribution of $\boldsymbol{\theta}$. The priors considered in this paper for the unknown parameters f_r, ϕ and $\mathbf{h}$ are summarized below.

- Uninformative independent uniform priors are chosen for the frequency and phase offsets, i.e., $p(f_r,\phi) = p(f_r)p(\phi)$ with $p(f_r)=I_{[-1/2,1/2]}(f_r)$ and

$$p(\phi) = \begin{cases} \frac{M}{2\pi} I_{[-\pi/M,\pi/M]}(\phi) & \text{for an M} - \text{PSK modulation,} \\ \frac{2}{\pi} I_{[-\pi/4,\pi/4]}(\phi) & \text{for other modulations,} \end{cases} \tag{14}$$

  where I is the indicator function.
- Independent normal prior distribution $\mathcal{N}(0,\sigma_h^2)$ are selected for the residual channel FIR filter taps. A suitable choice of parameter σ_h^2 allows us to incorporate vague prior information about the parameter h_l ($\sigma_h^2 = 0.01$ will be chosen for the simulations conducted in this paper).

A closed form expression for $\hat{\boldsymbol{\theta}}_{\text{MMSE}}$ using the approximate likelihood defined by (10) and the priors defined above cannot be obtained easily. Instead, we propose to generate N_θ samples $\boldsymbol{\theta}^i, i = 1,\ldots,N_\theta$ distributed according to the posterior $p(\boldsymbol{\theta}|\tilde{\mathbf{r}})$ and to estimate $\boldsymbol{\theta}$ as follows

$$\hat{\boldsymbol{\theta}}_{\text{MMSE}} = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\tilde{\mathbf{r}}) d\boldsymbol{\theta} \simeq \frac{1}{N_\theta} \sum_{i=1}^{N_\theta} \boldsymbol{\theta}^i. \tag{15}$$

The generation of $(\theta^1,\ldots,\theta^{N_\theta})$ distributed according to $p(\boldsymbol{\theta}|\tilde{\mathbf{r}})$ can be achieved by many different simulation methods. This paper proposes to use the Metropolis–Hastings (MH) algorithm whose main principles are recalled in the next section (the reader is invited to consult [19] for more details).

5.3. The Metropolis–Hastings algorithm

The MH algorithm is one of the most popular MCMC methods. It allows one to draw samples distributed according to $p(\boldsymbol{\theta}|\tilde{\mathbf{r}})$ by running an ergodic Markov chain whose stationary distribution is the target distribution $p(\boldsymbol{\theta}|\tilde{\mathbf{r}})$. The Markov chain state space and current state are denoted by Ω and $\theta^n = (f_r^n,\phi^n,\mathbf{h}^n) \in \Omega$, respectively. At each iteration, a candidate z is drawn according to an instrumental distribution $q(z|\theta^n)$. This candidate is accepted with the following acceptance probability:

$$\alpha(\theta^n,z) = \min\left\{1, \frac{p(z|\tilde{\mathbf{r}})q(\theta^n|z)}{p(\theta^n|\tilde{\mathbf{r}})q(z|\theta^n)}\right\}. \tag{16}$$

A fundamental property of the MH algorithm is that any instrumental distribution $q(z|\theta^n)$ can be chosen, provided that the support of $p(z|\tilde{\mathbf{r}})$ is contained in the support of $q(z|\theta^n)$ [19]. This paper proposes to draw z from a local perturbation of the previous sample, i.e., $z = \theta^n + \varepsilon$, 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 on q. The first generated samples $\boldsymbol{\theta}^i$ are usually not considered for the estimation of $\hat{\boldsymbol{\theta}}_{\text{MMSE}}$. These first samples belong to the so called burn in phase.

Instead of updating the whole of $\boldsymbol{\theta}$ *en bloc*, it is often more convenient and computationally efficient to divide $\boldsymbol{\theta}$ into k blocks and to update each block one-at-a-time. This procedure has been suggested by many authors (see [20] for more details) and has been shown to improve the mixing property of the sampler. Thus, we propose here to update $\boldsymbol{\theta}$ one component at-a-time. This strategy has shown good performance for the classification of modulations.

5.4. Plug-in classification rule

The classification rule resulting from the previous generation of samples can be summarized as follows:

$$\text{assign } \tilde{\mathbf{r}} \text{ to } \lambda_i \text{ if } \hat{l}(\tilde{\mathbf{r}}|\lambda_i) \geq \hat{l}(\tilde{\mathbf{r}}|\lambda_j) \quad \forall j, \tag{17}$$

where $\hat{l}(\tilde{\mathbf{r}}|\lambda_j)$ is given by

$$\hat{l}(\tilde{\mathbf{r}}|\lambda_j) = \sum_{k=1}^{N_s} \ln\left\{\frac{1}{M_j}\sum_{i=1}^{M_j} \exp\left(-\frac{1}{\sigma_z^2}\|\tilde{r}^*(k) - S_i\|^2\right)\right\}, \tag{18}$$

and where $\tilde{r}^*(k) = \hat{\mathcal{F}}^{-1}[\tilde{r}(k)e^{-j(\pi(k/N_s)\hat{f}_r + \hat{\phi})}]$. Note that the classification rule explicitly depends on the parameter estimates $\hat{f}_r$, $\hat{\phi}$ and $\hat{\mathbf{h}}$ ($\hat{f}_r$ is the estimate of the carrier frequency, $\hat{\phi}$ is the estimate of the carrier phase and $\hat{\mathcal{F}}^{-1}$ represents the inversion of the residual channel using $\hat{\mathbf{h}}$).

6. The Baum–Welch (BW) classifier

The BW classifier is based on a state trellis representation allowing one to estimate the posterior probabilities of the received modulated signal conditionally to each class. These posterior probabilities are then plugged into the optimal Bayes decision rule. An interesting property of the BW algorithm is that unknown parameter estimates, such as the noise variance, can also be obtained as side results. For modulation recognition purpose, the BW algorithm requires to associate a first order HMM to the received baseband communication signal. Section 6.1 describes this model, while Subsection 6.2 recalls the main steps of the BW algorithm.

6.1. Hidden Markov model

The received baseband signal $\tilde{r}(k)$ in (2) can be modeled as a probabilistic function of an hidden state at time k which is represented by a first order HMM whose characteristics are summarized below.

- The state of the HMM at time instant k is $s_k = a_k$ for MPSK and OQPSK modulated signals, whereas $s_k = (\phi_k,a_{k-1},a_{k-2},\ldots,a_{k-L+1})$ for GMSK modulated signals. The state vector s_k takes its values in a finite alphabet denoted as $\{s_k(1),s_k(2),\ldots,s_k(N)\}$ ($s_k(j)$ is the jth possible value of s_k). The size of this alphabet is $N = M$

for linear MPSK and OQPSK modulations and $N=4M^{L-1}$ for GMSK signals.

- The state transition probability is defined by

  $$d_{ij} = P[s_{k+1} = s_{k+1}(j) | s_k = s_k(i)],$$

  and equals $1/M$ when all symbols are equally likely for MPSK and GMSK modulated signals, whereas it equals $1/3$ for OQPSK modulated signals.
- The initial state distribution vector $\pi = (\pi_1, \ldots, \pi_N)^T$ is defined by

  $$\pi_i = P[s_1 = s_1(i)] = 1/N \quad \text{for } i = 1, \ldots, N.$$
- Based on (2), the pdf of the observation $\tilde{r}(k)$ conditioned on state i, denoted as $p_i[\tilde{r}(k)] \triangleq p[\tilde{r}(k)|s_k(i)]$ can be written

  $$p_i[\tilde{r}(k)] = \frac{1}{\pi\sigma_z^2} \exp\left(-\frac{|\tilde{r}(k) - m_i|^2}{\sigma_z^2} \right),$$

  where m_i denotes the ith *constellation* point. In the case of an AWGN channel, $m_i = S_i$ for MPSKs and OQPSK modulations and m_i is approximated by the ith value of $e^{j\Phi(kT,\mathbf{a})}$ for GMSK signals, for $i = 1, \ldots, N$. In the presence of a residual channel, m_i can be written as $m_i = \mathbf{h}^T \mathbf{S}_i$, where $\mathbf{S}_i$ is a $q \times 1$ vector containing q possible values of the constellation points.

6.2. The standard BW algorithm

The BW algorithm proposed in [9] for speech recognition can be used to determine the posterior probability of the observation sequence $P(\lambda|\tilde{\mathbf{r}}, \boldsymbol{m}, \sigma_z^2)$, given a model $\lambda \in \{\lambda_1, \ldots, \lambda_c\}$ (representing a modulation among the set of all c possible modulations), where $\boldsymbol{m}$ denotes the vector containing all possible *constellation* points ($\boldsymbol{m} = [m_1, \ldots, m_N]^T$ in the case of an AWGN channel, while $\boldsymbol{m} = [m_1, \ldots, m_{Mq}]^T$ in the presence of a residual channel). The probability of the observation sequence $p(\tilde{\mathbf{r}}|\boldsymbol{m}, \sigma_z^2, \lambda)$ for a given modulation is classically defined as a summation covering all possible state sequences. However, the direct computation of this summation requires high computational cost. The main idea of the BW algorithm is to use a forward-backward procedure which ensures a very efficient computation. The forward-backward procedure repeats the following three steps until convergence.

- *Step* 0: Initialization.
  The noise variance σ_z^2 and the constellation point vector $\boldsymbol{m}$ are randomly generated according to their prior distributions. In the presence of a residual channel, the channel coefficients are initialized using a fourth order cumulant based estimation [21]

  $$\hat{h}(k) = \frac{\hat{c}_{4,\tilde{r}}(q,0,k)}{\hat{c}_{4,\tilde{r}}(q,0,0)}, \quad k = 0, \ldots, q, \tag{19}$$

  where $\hat{c}_{4,\tilde{r}}(t_1,t_2,t_3)$ is an estimate of

  $$c_{4,\tilde{r}}(t_1,t_2,t_3) = \mathrm{cum}(\tilde{r}^*(t), \tilde{r}(t+t_1), \tilde{r}(t+t_2), \tilde{r}^*(t+t_3))$$

  with

  $$\mathrm{cum}(w,x,y,z) = E(wxyz) - E(wx)E(yz) - E(wy)E(xz) - E(wz)E(xy).$$
- *Step* 1: Compute the normalized forward variable $\alpha_i(k)$.
  Initialization

  $$\alpha_i(1) = \pi_i p_i(\tilde{r}(1)), \quad 1 \le i \le N,$$

  $$c(1) = \left(\sum_{i=1}^{N} \alpha_i(1) \right)^{-1}.$$

  Induction for $k = 1, \ldots, N_s - 1$, $j = 1, \ldots, N$

  $$\alpha_j(k+1) = c(k) p_j[\tilde{r}(k+1)] \sum_{i=1}^{N} \alpha_i(k) d_{ij},$$

  $$c(k+1) = \left(\sum_{i=1}^{N} \alpha_i(k+1) \right)^{-1}.$$
- *Step* 2: Compute the normalized backward variable $\beta_i(k)$.
  Initialization $\beta_i(N_s) = c(N_s)$, $1 \le i \le N$,
  Induction for $k = N_s - 1, \ldots, 1$, $i = 1, \ldots, N$,

  $$\beta_i(k) = c(k) \sum_{j=1}^{N} d_{ij} p_j[\tilde{r}(k+1)] \beta_j(k+1).$$
- *Step* 3: Estimate the model parameters.

  $$\hat{m}_i = \frac{\sum_{k=1}^{N_s} \gamma_i(k) \tilde{r}(k)}{\sum_{k=1}^{N_s} \gamma_i(k)},$$

  $$\hat{\sigma}_z^2 = \frac{1}{N_s} \sum_{k=1}^{N_s} \sum_{i=1}^{N} \gamma_i(k) |m_i - \tilde{r}(k)|^2,$$

  where $\gamma_i(k) = \alpha_i(k)\beta_i(k)$.

In a batch mode implementation, steps 1–3 are carried out iteratively with updated values of $p_j[\tilde{r}(k)]$ until convergence. The posterior probability of the observation sequence given the model is then estimated as follows:

$$\hat{P}(\lambda|\tilde{\mathbf{r}}, \boldsymbol{m}, \sigma_z^2) = \frac{\sum_{i=1}^{N} \alpha_i(N_s)}{\sum_{i=1}^{N_s} c(i)}, \tag{20}$$

and will be used in the plug-in classification rule (see Section 6.4).

Different modifications have been applied to the standard BW algorithm to improve its performance or reduce its computation complexity. One of these modifications presented in Section 6.3 is the adaptive BW algorithm.

6.3. The adaptive BW algorithm

An adaptive version of the BW algorithm was proposed in [22] to improve performance in terms of memory and computation speed. This LMS-type update algorithm is based on the following recursions:

$$\hat{m}_i(k) = \hat{m}_i(k-1) + \mu_m \gamma_i(k) e_i(k),$$

$$\hat{\sigma}_z^2(k) = (1 - \mu_s)\hat{\sigma}_z^2(k-1) + \mu_s \left(\sum_{i=1}^{N} \gamma_i(k) |e_i(k)|^2 \right),$$

where $e_i(k)=\tilde{r}(k)-\hat{m}_i(k-1)$ for $i=1,\ldots,N$. The initialization and time-induction steps for the forward variable can be computed as in the standard BW algorithm. The calculation of the backward variable can be obtained by using the fixed-lag or sawtooth-lag schemes [23]. In this paper, we have used the fixed-lag scheme.

6.4. Plug-in classification rule

Once the posterior probabilities of the observation sequence given each possible model have been estimated from (20), they are plugged into the optimal Bayes decision rule yielding

$$\text{assign } \tilde{\mathbf{r}} \text{ to } \lambda_i \text{ if } \hat{P}(\lambda_i|\tilde{\mathbf{r}}) \geq \hat{P}(\lambda_j|\tilde{\mathbf{r}}) \quad \forall j=1,\ldots,c, \tag{21}$$

where $\hat{P}(\lambda_i|\tilde{\mathbf{r}}) \triangleq \hat{P}(\lambda_i|\tilde{\mathbf{r}},\boldsymbol{m},\sigma_z^2)$ is obtained from (20). This strategy, consisting of replacing the class posterior probabilities in the optimal Bayesian classifier by their estimates, is sometimes referred to as plug-in MAP rule [24].

Note that the whole sequence of length N_s is required to estimate the posterior probabilities of the different modulations even if the on-line LMS-type update algorithm has been used for the computation of $\hat{m}_i(k)$ and $\hat{\sigma}_z^2(k)$. Note also that the observation length N_s required to properly identify the different modulations should be greater than the maximum number of HMM states in the class dictionary to ensure that any possible state can be reached by the algorithm. Note also that this paper has assumed that the different modulation formats are equally likely, i.e., $P(\lambda_i)=1/c$ for $i=1,\ldots,c$ (where c is the number of possible constellation, i.e., the number of classes).

7. Simulation results

Many simulations have been carried out to evaluate the performance of the proposed plug-in MAP classifiers. All constellations have been normalized to unit energy. The signal to noise ratio per bit is defined as E_b/N_0 and the signal to noise ratio per symbol as E_s/N_0, where E_b is the energy per bit and E_s is the energy per symbol at the input of the receiver. The classification performance is the average probability of correct classification defined as

$$P_{cc}=\frac{1}{c}\sum_{i=1}^{c} P[\text{assigning } \tilde{\mathbf{r}} \text{ to } \lambda_i|\tilde{\mathbf{r}}\in\lambda_i].$$

The receiver is supposed to be dedicated to a specific transmission system with a known particular set of possible constellations. We have considered different scenarios for our simulations. The first set of modulations $\lambda=\{\text{BPSK,QPSK,8PSK,16QAM}\}$ has been considered for comparison purposes with other methods available in the literature. The second set of modulations $\lambda=\{\text{QPSK, 8PSK,16APSK,32APSK}\}$ is appropriate to DVB-S2 (the proposed paper considers satellite transmission systems). OQPSK modulation format has been added in the third set of modulations because it is an interesting modulation for satellite transmissions. The other sets of constellations have been considered to show that the proposed classifier can be applied to various communication systems. Note finally that all described modulation formats (BPSK, QPSK, OQPSK, 8PSK, 16QAM, 16APSK, 32APSK, GMSK25, GMSK50) could be considered in the same set of constellations. However, we have not considered this case since all these constellations cannot currently appear in the same transmission system.

7.1. Classification of linear modulations

This section studies the performance of the two proposed classifiers to recognize linear modulations. The MCMC classifier is able to cope with channel impairments including noise, carrier phase, carrier frequency offset and residual channel due to imperfect equalization. However, it is not able to distinguish between OQPSK and QPSK modulations. Indeed, the MCMC classifier is based on the log-likelihood (8) which assumes that $\tilde{r}(i)$ and $\tilde{r}(j)$ are independent for $i\neq j$. Thus it cannot be applied to OQPSK modulations where successive symbols are correlated. Conversely, the BW classifier allows one to estimate the probabilities of moving from the different states of the trellis. Thus, it can be used for constellations with correlated symbols. However, the BW classifier is only able to cope with some channel impairments (noise, carrier phase offset and residual channel due to imperfect equalization) and does not take into account carrier frequency offset. Thus, it requires to use a preprocessing step that compensates the carrier frequency offset.

7.1.1. MCMC classifier

This section first considers a classical problem introduced in [2] for which

$$\lambda=\{\text{BPSK,QPSK,8PSK,16QAM}\}.$$

The transmission impairments described in Section 5 include a carrier frequency offset f_r, a three-tap residual channel whose channel coefficient vector is $\mathbf{h}=[1,h_1,h_2]$ and a carrier phase offset ϕ. All simulations have been obtained with 1000 trials belonging to each class λ_i (i.e., a total of 4000 trials). The number of symbols in each observation interval is $N_s=250$. The MCMC sampler has the following characteristics:

- number of burn-in iterations: $N_{bi}=500$,
- number of iterations: $N_\theta=1500$, and
- Instrumental distributions: $q(z|\theta_i^n)\sim\mathcal{N}(\theta_i^n,\sigma^2)$ where $\sigma=0.03$.

The simulations compare the performance of the following classifiers:

- the ML classifier (labeled ML) has been derived assuming $f_r=\phi=0$ and $\mathbf{h}=[1,0,0]$,
- the MCMC classifier (labeled MCMC), *Step*
- the classifier derived in [2] (labeled HOS since it is based on higher-order statistics).

The robustness of the MCMC plug in classifier to the carrier frequency offset is illustrated in Fig. 8 without any

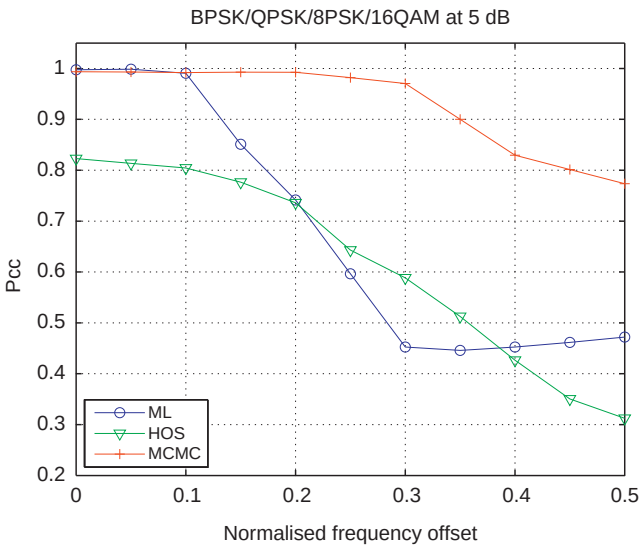

Fig. 8. Average probability of correct classification versus f_r ($\mathbf{h}$=[1, 0, 0], E_s/N_0=5 dB).

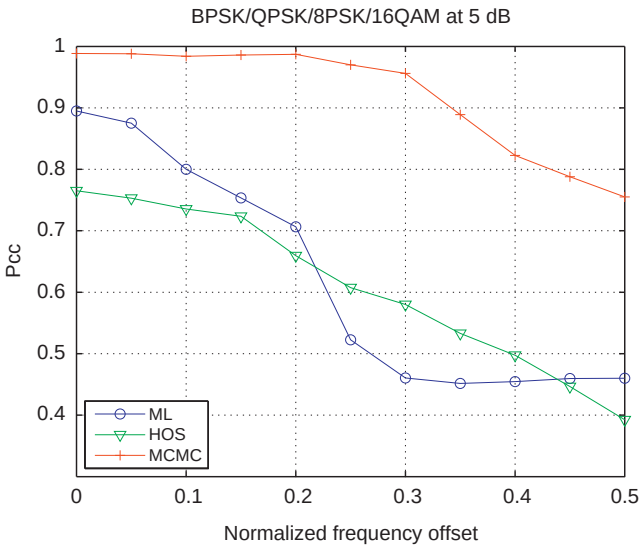

Fig. 9. Average probability of correct classification versus f_r ($\mathbf{h}$=[1, 0.25, 0.15], E_s/N_0=5 dB).

residual channel, i.e., with h=[1,0,0] (for E_s/N_0=5 dB and $\phi = 0$) and in Fig. 9 in the presence of a three-tap FIR channel h=[1,0.25,0.15] (for E_s/N_0=5 dB and $\phi = 0$). Fig. 10 studies the performance of the classifiers as a function of the norm of the residual channels $\sqrt{h_1^2+h_2^2}$ (for E_s/N_0=5 dB, f_r=0 and $\phi = 0$). The effect of the phase offset, keeping other parameters constant, is not presented here because the HOS and MCMC classifiers are insensitive to the phase offset. Moreover, the estimation of frequency offset is in general more difficult than that of phase offset. The MCMC classifier is clearly more robust to frequency offset and residual channel than the ML and HOS classifiers, at the price of an higher computational cost.

The second set of simulation considers modulations related to the DVB-S2 standard, i.e.,

$\lambda = \{\text{QPSK,8PSK,16APSK,32APSK}\}$.

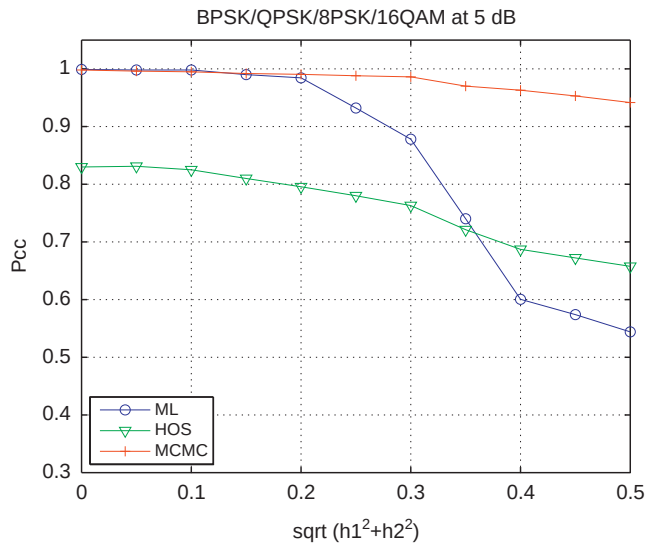

Fig. 10. Average probability of correct classification versus residual channel modulus (only $\mathbf{h}$ is estimated).

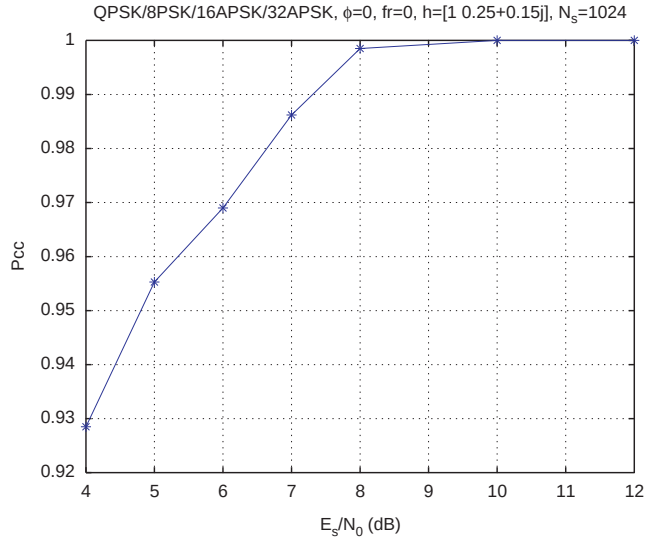

Fig. 11. Average probability of correct classification for the MCMC classifier versus E_s/N_0 in the presence of a residual channel (DVB-S2 modulations).

Fig. 11 studies the performance of the MCMC plug in classifier as a function of E_s/N_0, in the presence of a residual channel defined by h=[1, 0.25+0.15j]. The number of observed symbols is N_s=1024, $\phi = 0$ and f_r=0. The classification performance is satisfactory for this example, i. e., for the operating points of DVB-S2 systems.

7.1.2. BW classifier

This section studies a three-class problem

$\lambda = \{\text{BPSK,QPSK,OQPSK}\}$.

Since some particular transitions are not allowed for OQPSK (see Section 2) the BW classifier can discriminate OQPSK and QPSK signals even though they have the same constellations. This simulation considers that BPSK, QPSK, and OQPSK signals have a common baud-time (defined as the minimum time between data transition). The same

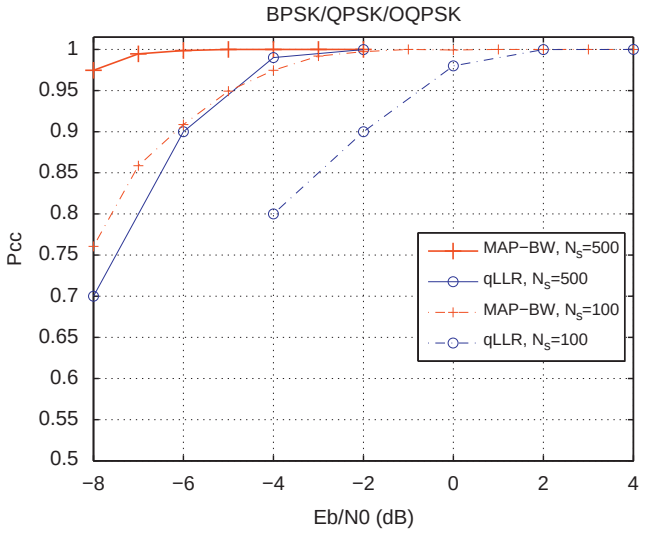

Fig. 12. Average probability of correct classification versus E_b/N_0.

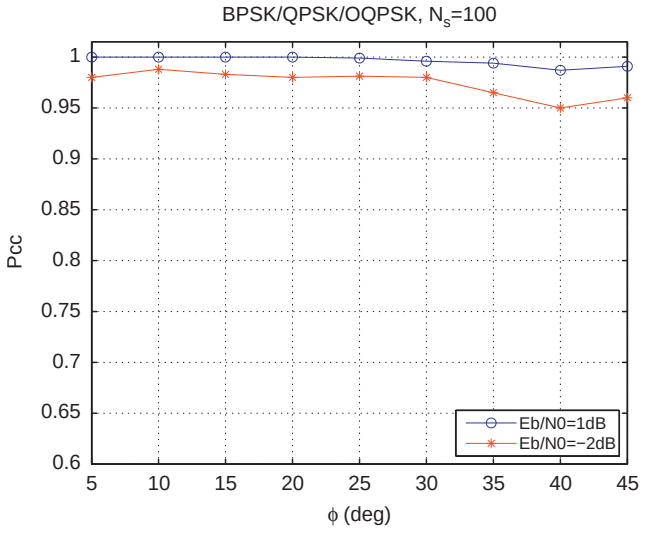

Fig. 13. Average probability of correct classification versus ϕ for the BW classifier.

observation interval is simulated as it is more realistic. The number of symbols of BPSK and QPSK signals is twice the number of symbols of OQPSK signals. It is required to adjust the value of the LMS step-size parameter μ_m for each constellation. The values of μ_m used in this work have been obtained by minimizing the average MSE of the estimated parameters. The following results have been obtained: $\mu_m = 0.3$ for BPSK, $\mu_m = 0.6$ for QPSK and OQPSK. Fig. 12 shows a performance comparison between the proposed BW strategy and the method of Chugg and Polydoros [5]. The authors proposed a qLLR classifier to identify BPSK/QPSK/OQPSK modulation types. Due to difficulty in setting thresholds resulting from the approximation of ALRT, the classifier works in two stages. First, it distinguishes between {OQPSK} and {BPSK, QPSK}. If the received signal is not OQPSK type, then it classifies between BPSK and QPSK formats. Note that there is no phase offset in this simulation. The threshold of the qLLR classifier in Fig. 12 is an *ideal threshold* obtained by

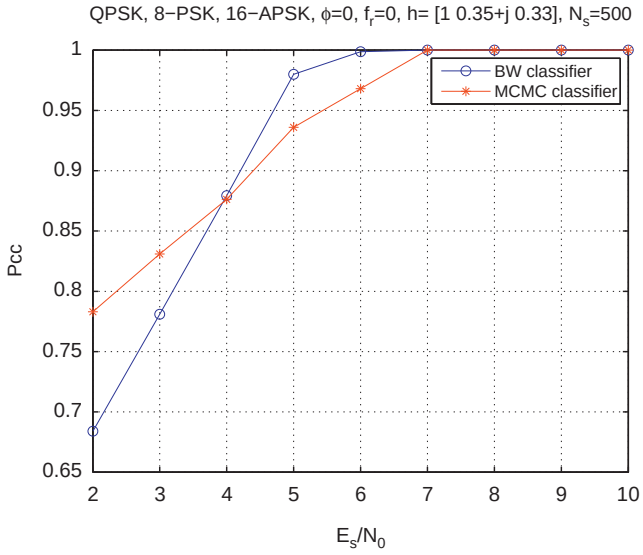

Fig. 14. Average probability of correct classification for MCMC and BW classifiers versus E_s/N_0 in the presence of a residual channel (DVB-S2 standard modulations).

maximizing P_{cc} over a large number of data and noise realizations. Although this threshold setting is not practical, it gives the best performance for the qLLR classifier. It is obviously seen that the BW classifier outperforms the qLLR classifier (even if the latter has been used with the best threshold setting). The effect of a phase offset to the BW classifier is also investigated. The phase offset represents a synchronization error of the local oscillator at the receiver and is obtained by rotating the constellation with an angle ϕ. The classification performance for different E_b/N_0 versus ϕ is illustrated in Fig. 13. Two values of the number of observed symbols have been tested: N_s=100 and 500. Note that the same methodology can be directly extended to discriminate between $\pi/4$−QPSK and 8PSK signals and, of course, to classify QPSK/OQPSK/$\pi/4$−QPSK/8PSK signals.

7.1.3. MCMC/BW classifier comparison

Fig. 14 compares the performance of MCMC and BW classifiers in the presence of a residual channel defined by $\mathbf{h}$=[1,0.35+0.33j]. The set of considered modulations is related to the DVB-S2 standard

$\lambda = \{\text{QPSK,8PSK,16APSK}\}$.

The 32-APSK was not considered here since the corresponding computational cost was prohibitive for the BW algorithm. The number of observed symbols is N_s=500, $\phi = 0$, f_r=0. The classification performance is similar for both classifiers.

Table 1 compares MCMC and BW classifiers in terms of computation time using Matlab 7.4.0.287 (R 2007a). The set of considered modulations is related to the DVB-S2 standard

$\lambda = \{\text{QPSK,8PSK,16APSK,32APSK}\}$.

A QPSK constellation has been emitted and E_s/N_0 has been set to 10 dB. The needed time to test each possible constellation of the dictionary has been measured and is

Table 1
Computation time comparison between BW and MCMC classifiers—E_s/N_0=10 dB.

Tested constellation	QPSK (s)	8-PSK (s)	16-APSK (s)	32-APSK (s)
BW classifier	1.47	4.9	20.88	111.9
MCMC classifier	10.6	14.4	23.46	50.18

Table 2
BW classifier—confusion matrix for E_b/N_0=0 dB.

In/Out	GMSK25	GMSK50	BPSK	QPSK	8PSK
GMSK25	449	51	0	0	0
GMSK50	13	487	0	0	0
BPSK	0	0	500	0	0
QPSK	0	0	0	498	2
8PSK	0	0	0	0	500

given in seconds. The number of observed symbols is N_s=500. It can be noted that the computational cost of the BW classifier increases rapidly with the modulation order. This is due to the increasing number of states in the state trellis representation. When compared to the MCMC classifier, BW classifier allows to reduce the computational cost for small constellations but is more expensive when higher order modulations are considered in the dictionary.

7.2. Classification of linear and non-linear modulation

This section adresses the problem of classifying GMSK25, GMSK50, BPSK, QPSK and 8PSK modulations with the BW algorithm. Actually as it can be observed in Fig. 5, the *constellations* of GMSK25 and GMSK50 are very similar (even without noise) and also very close to a QPSK constellation. As a consequence, the MCMC classifier would not be able to distinguish these two modulations. Tables 2–4 present the confusion matrices of the classifiers for different values of E_b/N_0 (note that f_r = 0, $\phi = 0$ and that no residual channel has been considered). The number of symbols in each observation interval is N_s=500. It can be observed that the two GMSK signals as well as the MPSK signals can be distinguished even at very low values of E_b/N_0. However, to distinguish among linear modulations, the required operating E_b/N_0 has to be much higher especially when 8PSK modulations are present in the dictionary.

Fig. 15 displays the classification performance as a function of E_b/N_0, for different values of the number of symbols in the observation interval N_s (with f_r = 0, $\phi = 0$ and no residual channel). A good classification performance can be observed especially for small values of E_b/N_0 that are typical for satellite space communications.

Classical linear receivers implement low-pass filters that are square root raised cosine filters. These filters depend on the same parameter called roll-off factor (matched filtering for linear modulations). Obviously, in the case of modulation recognition the emitted roll-off factor is unknown. In this context, it is interesting to study

Table 3
BW classifier—confusion matrix for E_b/N_0= −2 dB.

In/Out	GMSK25	GMSK50	BPSK	QPSK	8PSK
GMSK25	406	94	0	0	0
GMSK50	46	454	0	0	0
BPSK	0	0	500	0	0
QPSK	0	0	0	457	43
8PSK	0	0	0	5	495

Table 4
BW classifier—confusion matrix for E_b/N_0= −6 dB.

In/Out	GMSK25	GMSK50	BPSK	QPSK	8PSK
GMSK25	334	164	1	0	1
GMSK50	123	375	0	1	1
BPSK	0	0	488	4	8
QPSK	0	0	0	313	187
8PSK	0	0	0	81	419

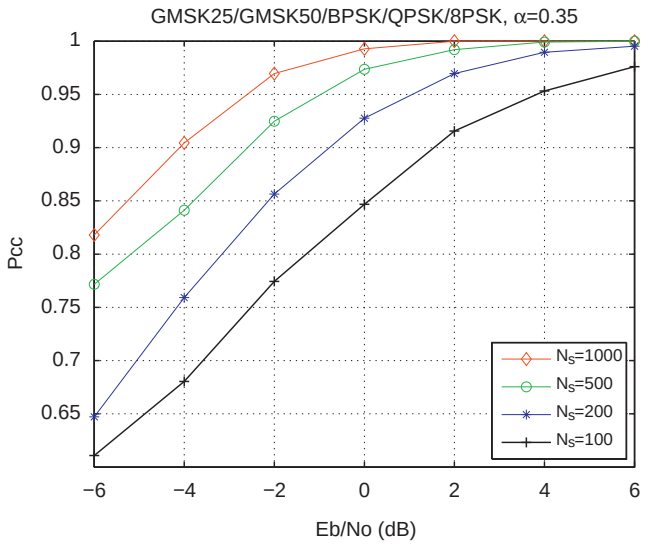

Fig. 15. Average probability of correct classification versus E_b/N_0 for the BW classifier.

the effect of a roll-off mismatch on classification performance. Fig. 16 displays the classification performance for several values of the roll-off factors α of the square root raised cosine filters used in the transmitter and the receiver (α_T and α_R denote the transmitter and receiver roll-off factors, respectively). The number of symbols in each observation interval is N_s=500 and f_r=0, $\phi = 0$, **h**=[1,0,0]. The proposed classifier seems to perform similarly for the different roll-off factor combinations.

The last simulations study the effect of a phase offset obtained by rotating the constellation with an angle ϕ (this phase offset is due to synchronization errors at the receiver) and f_r = 0, h=[1,0,0]. The number of symbols in each observation interval is N_s=500. Fig. 17 shows that the classification performance seems to be robust to moderate values of phase offset.

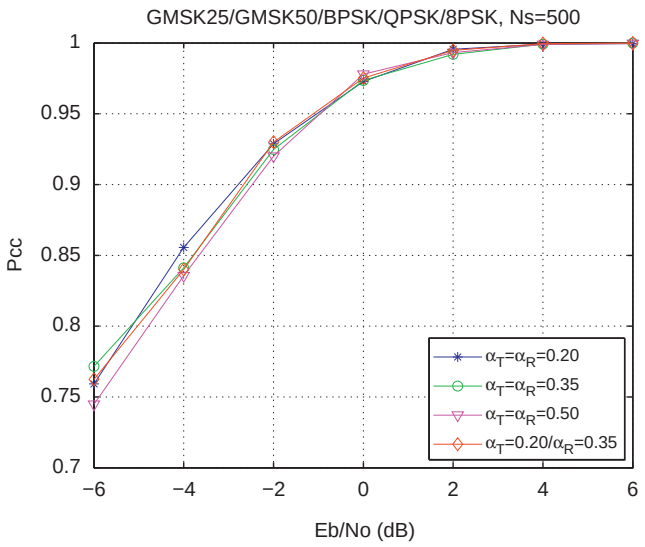

Fig. 16. Average probability of correct classification versus E_b/N_0 for different values of the roll-off factor α and for the BW classifier.

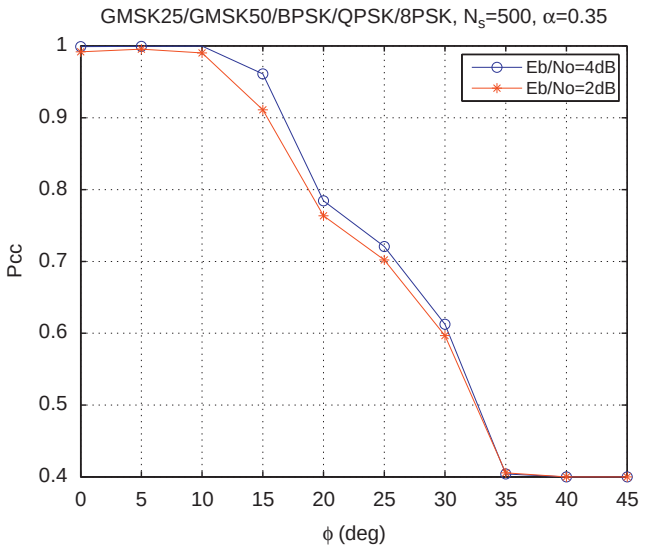

Fig. 17. Average probability of correct classification versus phase offset for the BW classifier.

8. Conclusions

This paper addressed the problem of classifying linear and non-linear modulations in the presence of noise and of different channel impairments including carrier frequency errors, carrier phase errors and residual channel due to imperfect equalization. Two Bayesian classifiers, referred to as MCMC classifier and BW classifier, were studied. These classifiers estimated the posterior probabilities of the possible modulations, and plugged them into the optimal Bayes decision rule. The MCMC classifier generated samples distributed according to the posterior using Markov chain Monte Carlo (MCMC) methods. These generated samples were used to estimate the unknown model parameters that were plugged into the optimal Bayesian classification rule. The second classifier estimated the posterior distribution of the received communication signal using the Baum–Welch (BW) algorithm. Several simulations showed the good performance of the proposed classifiers. The MCMC classifier appeared to be robust to channel impairments like noise, carrier phase and frequency offset or a residual channel. However, it is not able to recognize between QPSK and OQPSK modulations and cannot handle non-linear GMSK modulations. The BW algorithm was able to recognize classical linear modulations but also OQPSKs and GMSKs. However, carrier frequency errors have to be corrected in a preprocessing step. Future works include the classification of modulations appearing in new satellite communication standards. These modulations include orthogonal frequency division multiplexing (OFDM) modulation for mobile digital video broadcasting on handled receivers (DVB-SH standard).

References

[1] Consultative Committee for Space Data Systems (CCSDS), Radio Frequency and Modulation Systems, CCSDS, no. 401, 2001.
[2] A. Swami, B. Sadler, Hierarchical digital modulation classification using cumulants, IEEE Trans. Commun. 48 (3) (March 2000) 416–429.
[3] O.A. Dobre, A. Abdi, Y. Bar-Ness, W. Su, Survey of automatic modulation classification techniques: classical approaches and new trends, IET Commun. 1 (2) (April 2007) 137–156.
[4] N. Lay, A. Polydoros, Modulation classification of signals in unknown ISI environments, in: Proceedings of the IEEE Milcom vol. 1, San Diego, CA, USA, November 1995, pp. 170–174.
[5] K.M. Chugg, C.S. Long, A. Polydoros, Combined likelihood power estimation and multiple hypothesis modulation classification, in: Asilomar Conference on Signals, Systems, Computers, vol. 2, November 1995, pp. 1137–1141.
[6] L. Hong, K.C. Ho, An antenna array likelihood modulation classifier for BPSK and QPSK signals, in: Proceedings of the IEEE Milcom, vol. 1, Anaheim, CA, USA, October 2002, pp. 647–651.
[7] S. Soliman, S.Z. Hsue, Signal classification using statistical moments, IEEE Trans. Commun. 40 (May 1992) 908–916.
[8] S. Lesage, J.-Y. Tourneret, P.M. Djuric, Classification of digital modulation by MCMC sampling, in: Proceedings of the IEEE International Confence on Acoustics, Speech, and Signal Processing (ICASSP), vol. 4, May 2001, pp. 2553–2555.
[9] L. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proc. IEEE 77 (2) (1989) 257–286.
[10] A. Puengnim, T. Robert, N. Thomas, J. Vidal, Hidden Markov models for digital modulation classification in unknown ISI channels, in: Proceedings of the European Signal Processing Conference (EUSIPCO), Poznan, Poland, September 2007, pp. 1882–1885.
[11] C.Y. Huang, A. Polydoros, Two small SNR classification rules for CPM, Proc. IEEE Milcom 3 (October 1992) 1236–1240.
[12] C.D. Chung, A. Polydoros, Envelope based classification schemes for continuous phase binary frequency shift keyed modulation, in: Proceedings of the IEEE Milcom Fort Monmouth, NJ, October 1994, pp. 796–800.
[13] A.E. El-Mahdy, N.M. Namazi, Classification of multiple M-ary frequency-shift keying over a rayleigh fading channel, IEEE Trans. Commun. 50 (6) (June 2002) 967–974.
[14] A. Puengnim, N. Thomas, J. Tourneret, J. Vidal, Classification of linear and nonlinear modulations using the Baum–Welch algorithm, in: Proceedings of the European Signal Processing Conference (EUSIPCO), Lauzanne, Switzerland, August 2008.
[15] J.G. Proakis, Digital Communications, McGraw-Hill, New York, 2001.
[16] F. Xiong, Digital Modulation Techniques, Artech House, Inc, Norwood, 2000.
[17] W. Wei, J.M. Mendel, Maximum likelihood classification for digital amplitude-phase modulation, IEEE Trans. Commun. 48 (February 2000) 189–193.

[18] A. Abdi, O.A. Dobre, R. Cauchy, Y. Bar-Ness, W. Su, Modulation classification in fading channels using antenna arrays, Proc. IEEE Milcom 1 (October 2004) 211–217.
[19] C.P. Robert, G. Casella, Monte Carlo Statistical Methods, second ed., Springer-Verlag, New York, 2004.
[20] W.R. Gilks, S. Richardson, D.J. Spiegelhalter, Introducing Markov chain Monte Carlo, in: W.R. Gilks, S. Richardson, D.J. Spiegelhalter (Eds.), Markov chain Monte Carlo in Practice, Chapman & Hall, London, UK, 1996, pp. 1–19.
[21] J.M. Mendel, Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications, Proc. IEEE 79 (3) (March 1991) 278–305.
[22] J.A.R. Fonollosa, J. Vidal, Application of hidden Markov models to blind channel characterization and data detection, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), vol. 4, Adelaide, Australia, April 1994, pp. 185–188.
[23] V. Krishnamurthy, J. Moore, On-line estimation of hidden Markov model parameters based on the Kullback–Leibler information measure, IEEE Trans. Signal Process. 41 (11) (1993) 2557–2573.
[24] A.K. Jain, R.P.W. Duin, J. Mao, Statistical pattern recognition: a review, IEEE Trans. Pattern Anal. Mach. Intell. 22 (1) (2000) 4–37.