Blind Modulation Classification: A Concept Whose Time Has Come

Octavia A. Dobre[1], Ali Abdi[1], Yeheskel Bar-Ness[1], and Wei Su[2]

[1]CCSPR, Dept. of ECE, New Jersey Institute of Technology, Newark, NJ 07102, USA
[2]RDECOM, Fort Monmouth, NJ 07703, USA

octavia.a.dobre@njit.edu, ali.abdi@njit.edu, barness@yegal.njit.edu, wei.su@mail1.monmouth.army.mil

Abstract – In this paper we address the problem of identifying the modulation format of an incoming signal. We review many existing techniques for digital modulation recognition in a systematic way, which helps the reader to see the main features of each technique. The goal is to provide useful guidelines for choosing appropriate classification algorithms for different modulations, from the large pool of available techniques. Furthermore, the performance of a benchmark classifier is presented, as well as its sensitivity to several model mismatches. At the end, open problems and possible directions for further research are briefly discussed.

I. Introduction

Blind modulation classification (MC) is an intermediate step between signal detection and demodulation, with application in both commercial and military communication systems. In civilian systems, MC is specially applied in software defined radio (SDR), to cope with the variety of communication systems. In military communication systems, advanced techniques are required for real-time signal interception and processing, which are vital for decisions involving electronic warfare operations and other tactical actions. This has emerged the need for smart receivers, which employ blind signal processing techniques. A major task of such systems is the blind recognition of the modulation of an incoming signal. MC is a challenging problem, particularly in a non-cooperative environment, since in addition to complex channels, there are many unknown parameters, such as carrier frequency, symbol timing, etc., that have to be extracted from the received signal. The design of a classifier essentially involves two steps: preprocessing of the incoming signal and proper selection of the classification algorithm. Preprocessing tasks can include noise removal, estimation of carrier frequency, symbol period, signal and noise powers, equalization, etc. The required preprocessing and its accuracy depend on the classification algorithm chosen in the second step. For the classification part there are two general approaches: likelihood-based (LB) [1]-[8] and feature-based (FB) methods [9]-[20]. A good classifier has to provide a high correct classification rate for a large range of signal-to-noise ratio (SNR), in a short observation interval. In addition, it needs to fulfill these requirements: robustness with respect to (w.r.t.) preprocessing inaccuracy, capability to recognize a large number of modulations in environments with different propagation characteristics, real-time functionality, and low computational complexity. A classifier should either decide what modulation has been received, out of N_{mod} equally likely candidates, or that the modulation cannot be recognized as one of the catalog. The former case is considered here, as being the only treated in the literature. Due to the lack of space, we focus on classification algorithms for phase shift keying (PSK), quadrature amplitude modulations (QAM), amplitude shift keying (ASK), and frequency shift keying (FSK).

The rest of the paper is organized as follows. A general signal model is introduced in Section II. Algorithms derived within the LB and FB frameworks are presented in Section III and Section IV, respectively. Numerical results are provided in Section V, and conclusions are finally drawn in Section VI.

II. Signal Model

The MC algorithms proposed in the literature employ information extracted from the baseband [1]-[9], [13]-[20] or intermediate frequency band [11]-[12]. A general expression for the baseband received waveform is given by

$$r(t) = s(t;\mathbf{u}_i) + n(t)\,,\ 0 \le t \le KT, \qquad (1)$$

where

$$s(t;\mathbf{u}_i) = a_i e^{j\theta} e^{j2\pi\Delta f t} \sum_{k=1}^{K} s_k^{(i)} e^{j\phi_k} g(t-(k-1)T-\varepsilon T) \qquad (2)$$

is the noise-free baseband complex envelope of the received signal, resulting from the ith modulation format, $i = 1,\ldots,N_{\mathrm{mod}}$, $a_i = \sqrt{E_s/\sigma^2_{s^{(i)}} E_p}$, $\sigma^2_{s^{(i)}} = M_i^{-1}\sum_{m=1}^{M_i} |s_m^{(i)}|^2$ is the variance for the ideal noise-free constellation corresponding to the i th modulation format, M_i is the number of equi-probable points in the ith signal constellation, $E_p = \int_{-\infty}^{\infty} |p_{TX}(t)|^2\, dt$ represents the pulse energy, with $p_{TX}(t)$ the transmitter pulse shape, E_s is the baseband signal energy, Δf is the frequency offset, θ is the time-invariant carrier phase, $\{\phi_k\}_{k=1}^{K}$ represents the phase jitter, $g(t) = p_{TX}(t) \otimes h(t)$ [1], with $h(t)$ as the channel impulse response and $\otimes$ denoting convolution, T is the symbol period, $[0, KT]$ is the K symbol observation interval, ε denotes the timing offset w.r.t. the receiver reference clock ($0 \le \varepsilon < 1$), $s_k^{(i)} = s_{k,I}^{(i)} + j s_{k,Q}^{(i)}$ [2] is the symbol transmitted within the kth period and $n(t)$ is aggregate noise: receiver noise, as well as cochannel interference and jamming. We adopt the notation $s(t;\mathbf{u}_i)$ to stress the signal dependence on the unknown quantities, $\mathbf{u}_i = [a_i\ \Delta f\ \theta\ T\ \varepsilon\ g(t)\ \{\phi_k\}_{k=1}^{K}\ \{s_k^{(i)}\}_{k=1}^{K}]^{\dagger}$, where $\dagger$ is the transpose operator. Note that the term

[1] The pulse shape $g(t)$ contains information of both transmitter pulse shape and channel. For example, the additive white Gaussian noise (AWGN) channel is characterized by $h(t) = \delta(t)$, with $\delta(.)$ as the Dirac delta function, whereas for a flat block fading $h(t) = \alpha e^{j\varphi}\delta(t)$, with α and φ the channel amplitude and phase, respectively, constant over the observation interval.

[2] For FSK the data symbols depend on t. For others, such as ASK, PSK and QAM, $s_k^{(i)}$'s do not depend upon t. To simplify the notation, we most often drop the t-dependence, unless otherwise specified.

"unknown quantity" refers to both unknown signal parameters, such as carrier frequency and timing offsets, as well as unknown data symbols. The data symbols $\{s_k^{(i)}\}_{k=1}^K$ are taken from a finite-alphabet specific to the i th modulation format, $i = 1,...,N_{\text{mod}}$. For alphabets associated to different digital modulations see, for example, [21] Ch. 4. Without loss of generality, we consider here unit variance constellations, i.e, $\{s_k^{(i)}\}_{k=1}^K / \sigma_{s^{(i)}}$. To simplify the notation, only $\{s_k^{(i)}\}_{k=1}^K$ will be used in the sequel.

III. Likelihood Based Approach To MC

Within the LB framework, MC is formulated as a multiple composite hypothesis-testing problem. Under the hypothesis H_i, the i th modulation is assigned to the incoming signal, $i = 1,...,N_{\text{mod}}$. This approach is based on the likelihood-ratio test (LRT), which uses the likelihood function (LF) of $r(t)$ over the interval $[0, KT]$. Depending on the model chosen for the unknown quantities, three LB-MC techniques were proposed in the literature: average LRT (ALRT) [1], [8], generalized LRT (GLRT) [6] and hybrid LRT (HLRT) [6]-[7].

With ALRT, the unknown quantities are treated as random variables (r.v.'s) and the LF is computed by averaging over them,

$$\Lambda_A^{(i)}[r(t)] = \int \Lambda[r(t) \mid \mathbf{v}_i, H_i] p(\mathbf{v}_i \mid H_i) d\mathbf{v}_i, \tag{3}$$

where $\Lambda[r(t) \mid \mathbf{v}_i, H_i]$ is the conditional LF of $r(t)$, conditioned on the unknown vector $\mathbf{v}_i$, and $p(\mathbf{v}_i \mid H_i)$ is the a priori probability density function (PDF) of $\mathbf{v}_i$ under H_i. If $p(\mathbf{v}_i \mid H_i)$ coincides with the true PDF, ALRT results in an optimal classifier in the Bayesian sense, viz., it maximizes the average probability of correct classification.

In a two-hypothesis classification problem, the decision is made according to

$$\Lambda_l^{(1)}[r(t)] / \Lambda_l^{(2)}[r(t)] \underset{H_2}{\overset{H_1}{\gtrless}} \eta_l, \tag{4}$$

where η_l is a threshold, $l = A(ALRT), G(\text{GLRT}), H(\text{HLRT})$. Extension of (4) to multiple hypotheses is straightforward (see, for example, [23] Ch. 2).

For AWGN, using the complex Gaussian distribution of $n(t)$ and for the i th hypothesis H_i, one can show that the conditional LF is given by (see, for example, [21] Ch.6)[3]

$$\Lambda[r(t) \mid \mathbf{u}_i, N_0, H_i] = \exp\left\{ 2N_0^{-1} \operatorname{Re}\left[\int_0^{KT} r(t) s^*(t; \mathbf{u}_i) dt \right] - N_0^{-1} \int_0^{KT} \left| s(t; \mathbf{u}_i) \right|^2 dt \right\}, \tag{5}$$

where $\operatorname{Re}[.]$ is the real part and N_0 (W/Hz) is the two-sided power spectral density (PSD) of the complex AWGN $n(t)$. The autocorrelation function of $n(t)$ is given by $\mathrm{E}[n(t) n^*(t+\tau)] = N_0 \delta(\tau)$, where $\mathrm{E}[.]$ is the mathematical expectation, and * denotes complex conjugation.

In the available literature, the mathematical expression of the LF, $\Lambda_A^{(i)}[r(t)]$, was derived for certain simplified scenarios, i.e., AWGN channel, with $\mathbf{v}_i = [\{s_k^{(i)}\}_{k=1}^K]^{\dagger}$ [4] [1], $\mathbf{v}_i = [\theta \ \{s_k^{(i)}\}_{k=1}^K]^{\dagger}$ [5] [8], and $\mathbf{v}_i = [\{\phi_k\}_{k=1}^K \ \{s_k^{(i)}\}_{k=1}^K]^{\dagger}$ [6] [4], respectively, as well as flat block Rayleigh fading channel, with $\mathbf{v}_i = [\alpha \ \varphi \ \{s_k^{(i)}\}_{k=1}^K]^{\dagger}$ [8]. Note that the pulse shape was assumed rectangular[7], i.e., $p_{TX}(t) = 1$ for $0 \le t < T$ and zero otherwise, and the data symbols $\{s_k^{(i)}\}_{k=1}^K$ were treated as independent and identical distributed r.v.'s. The LFs derived for $\mathbf{v}_i = [\{s_k^{(i)}\}_{k=1}^K]^{\dagger}$ [1] and $\mathbf{v}_i = [\theta \ \{s_k^{(i)}\}_{k=1}^K]^{\dagger}$ [8], are given respectively by

$$\Lambda_A^{(i)}[r(t)] = \prod_{k=1}^K \mathrm{E}_{s_k^{(i)}} \left[\exp\left[2\sqrt{S} N_0^{-1} \operatorname{Re}[R_k^{(i)}] - STN_0^{-1} \mid s_k^{(i)} \mid^2 \right] \right], \tag{6}$$

and

$$\Lambda_A^{(i)}[r(t)] = \mathrm{E}_{\{s_k^{(i)}\}_{k=1}^K} \left[e^{-STN_0^{-1} \eta_K^{(i)}} I_0 (2\sqrt{S} N_0^{-1} \mid \xi_K^{(i)} \mid) \right], \tag{7}$$

where $S = E_s / T$ is the signal power, $I_0(.)$ is the zero order modified Bessel function of the first kind, $\mathrm{E}_{s_k^{(i)}}$ is nothing but a finite summation over all the M_i possible constellation points of the i th modulation, divided by M_i, for the k th interval and $\mathrm{E}_{\{s_k^{(i)}\}_{k=1}^K}[.]$ is the averaging performed over K data symbols. Also, $\xi_K^{(i)} = \sum_{k=1}^K R_k^{(i)}$ and $\eta_K^{(i)} = \sum_{k=1}^K | s_k^{(i)} |^2$, where

$$R_k^{(i)} = \int_{(k-1)T}^{kT} r(t) s_k^{(i)*}(t) dt, \ k = 1,...,K. \tag{8}$$

For linear modulations $s_k^{(i)}(t)$ is constant over the period $(k-1)T$ to kT, and thus $R_k^{(i)} = s_k^{(i)*} r_k$, where $r_k = \int_{(k-1)T}^{kT} r(t) dt$ is the output of the matched filter at kT. The decision was made based on (4), with the threshold set to one. By comparing (6) and (7), one can easily notice that the complexity in computing the LF increases with the unknown carrier phase θ.

For many cases of interest, the computational complexity and even mathematical intractability of the ALRT-based classifier, as well as the need for prior knowledge, can render the ALRT impractical. Hence, approximations of the LF were investigated, leading to the so-called quasi-ALRT classifiers [3]-[5]. Such algorithms were derived for linear modulation classification in AWGN channel, with $\mathbf{v}_i = [\theta \ \{s_k^{(i)}\}_{k=1}^K]^{\dagger}$ [5] [2]-[3] and $\mathbf{v}_i = [\theta \ \varepsilon \ \{s_k^{(i)}\}_{k=1}^K]^{\dagger}$ [8] [2], and for FSK signal identification, with $\mathbf{v}_i = [\{\phi_k\}_{k=1}^K \ \{s_k^{(i)}\}_{k=1}^K]^{\dagger}$ [6] [4] and $\mathbf{v}_i = [\varepsilon \ \{\phi_k\}_{k=1}^K \ \{s_k^{(i)}\}_{k=1}^K]^{\dagger}$ [8] [5]. For example, the approximation of the LF used in [3][9] for PSK and QAM signal, with $\mathbf{v}_i = [\theta \ \{s_k^{(i)}\}_{k=1}^K]^{\dagger}$ and AWGN, is

$$\Lambda_A^{(i)}[r(t)] \approx \exp\left\{ \sum_{n=1}^{\infty} (\sqrt{S} N_0^{-1})^n K \sum_{q=0}^{\lfloor n/2 \rfloor} \left[\upsilon_{n-2q} \left(q!(n-q)! \right)^{-1} \mid m_{s^{(i)},n,q} \| \hat{m}_{r,n,q}(\mathbf{0}_{n-1}) \mid \right] \right\}, \tag{9}$$

where $m_{s^{(i)},n,q} = \mathrm{E}[(s^{(i)})^{n-q} (s^{(i)*})^q]$ is the n th-order/ q conjugate moment of the i th constellation, $\hat{m}_{r,n,q}(\mathbf{0}_{n-1}) = K^{-1} \sum_{k=1}^K r_k^{n-q} (r_k^*)^q$ is the sample estimate of the n th order/ q -conjugate moment at the zero-delay vector $\mathbf{0}_{n-1}$ [10], $\lfloor . \rfloor$ denotes rounding to the nearest integer, υ_{n-2q} is 1 if $q = n/2$ and 2 for $q < n/2$, and ! denotes factorial. Eq. (9) can be further simplified for symmetric constellations, as the n th order moments are equal to zero for n odd. Furthermore, one can easily show that the lowest order

---

[3] Note that $\mathbf{v}_i = [\mathbf{u}_i^{\dagger} \ N_0]^{\dagger}$, as N_0 is also unknown.

[4] In AWGN, with all parameters perfectly known, we set $\theta = \Delta f = \varepsilon = \{\phi_k\}_{k=1}^K = 0$. The unknowns are specified in $\mathbf{v}_i$, and here $\mathbf{v}_i = \mathbf{u}_i$.

[5] The carrier phase θ was modeled as a r.v. $U[-\pi, \pi)$.

[6] With ϕ_k shown in (2) as a phase jitter and modeled as a r.v. $U[-\pi, \pi)$.

[7] A rectangular pulse shape was assumed in the scanned literature, unless otherwise mentioned.

[8] The timing offset ε is modeled as a r.v. $U[0,1)$.

[9] Eq. (9) was obtained by writing (5) given in [3] in terms of signal moments.

[10] For the definition of the n th-order/ q conjugate moment, $m_{r,n,q}(\tau_{n-1})$, and the cumulant, $c_{r,n,q}(\tau_{n-1})$, as well as the relations between the moments and cumulants, see [22] Ch.2. A $\tau_{n-1} = \mathbf{0}_{n-1}$ delay vector is an $(n-1) \times 1$ vector, with all the elements equal to zero.

statistic to distinguish between M-PSK and M'-PSK ($M'>M$) is the M th-order/ zero-conjugate moment, $m_{s^{(i)},M,0}$. This property does not hold for QAM, for which $m_{s^{(i)},n,q}=0$, where n is a multiple of four and q odd, or n is not a multiple of four and q even. By resorting to only the lowest order statistic, suboptimal, but implementable and manageable classifiers were proposed to discriminate PSK and QAM signals in [2] and [3], respectively. For example, the metric $\nu_M = |\sum_{k=1}^{K} r_k^M|$ was used to distinguish between M-PSK and M'-PSK ($M'>M$). The decision was made by comparing the chosen statistic against a threshold η, which is set either theoretically[11] or empirically[12]. By comparing the metric ν_M with (7), one can notice that the complexity of the quasi-ALRT classifier is much lower than that of the ALRT classifier.

The following approximation of the LF was used with the classifiers designed for unknown timing offset ε [8] [5]

$$\Lambda_A^{(i)}[r(t)] \approx D^{-1}\sum_{d=0}^{D-1}\Lambda[r(t)\mid \varepsilon_d, H_i], \tag{10}$$

where D is the number of levels to which the timing offset uncertainty interval is quantized and $\varepsilon_d = d/D$, $d=0,...,D-1$. This approximation improves as $D\to\infty$, since then the summation converges to an integral. The value of D directly determines the classifier complexity, as it introduces more terms in (10). Interestingly, the quasi-ALRT classifiers are actually FB classifiers derived within the LB framework.

When a PDF cannot be assigned to the unknown parameters, a logical procedure is to estimate the unknown parameters assuming H_i is true and use these estimates in the LRT as if they were correct. If maximum likelihood (ML) estimates are used, the result is called GLRT. Then, the LF is given by

$$\Lambda_G^{(i)}[r(t)] = \max_{\mathbf{v}_i}\Lambda[r(t)\mid \mathbf{v}_i, H_i]. \tag{11}$$

HLRT is a mixture of ALRT and GLRT, with the LF given by

$$\Lambda_H^{(i)}[r(t)] = \int \max_{\mathbf{v}_{i_1}}\Lambda[r(t)\mid \mathbf{v}_i, H_i]p(\mathbf{v}_{i_2}\mid H_i)d\mathbf{v}_{i_2}. \tag{12}$$

where $\mathbf{v}_i = [\mathbf{v}_{i_1}^\dagger \mathbf{v}_{i_2}^\dagger]^\dagger$.

GLRT and HLRT were examined for linear modulation classification in AWGN ($\mathbf{v}_i = [\theta\ \{s_k^{(i)}\}_{k=1}^{K}]^\dagger$ [5]) [6], with the LFs given respectively by

$$\Lambda_G^{(i)}[r(t)] = \max_{\theta}\left\{\sum_{k=1}^{K}\max_{s_k^{(i)}}\left\{\mathrm{Re}[s_k^{(i)*}r_k e^{-j\theta}] - 2^{-1}\sqrt{S}T\,|s_k^{(i)}|^2\right\}\right\}, \tag{13}$$

and

$$\Lambda_H^{(i)}[r(t)] = \max_{\theta}\left\{\prod_{k=1}^{K}\mathrm{E}_{s_k^{(i)}}\left[\exp\left[2\sqrt{S}N_0^{-1}\,\mathrm{Re}[s_k^{(i)*}r_k e^{-j\theta}] - STN_0^{-1}\,|s_k^{(i)}|^2\right]\right]\right\}. \tag{14}$$

[11] In order to maximize the probability of correct classification, the threshold η is chosen to satisfy $p^{(i)}(\eta) = p^{(i')}(\eta)$, where $p^{(i)}(\eta)$ and $p^{(i')}(\eta)$ are the PDFs of the chosen metric under the hypotheses H_i and $H_{i'}$, respectively.

[12] This threshold is set to maximize the average probability of correct classification over a large number of data and noise realizations. It is assumed that such simulations should be run off-line and the threshold can be stored as a function of the noise and signal parameters. In a practical implementation, the threshold is therefore obtained from a look-up table.

Eq. (4) was used for decision, with the thresholds set based on the empirical histogram method[12].

An advantage of GLRT and HLRT over ALRT is that as by-products, the estimates of the unknown quantities are of interest for data demodulation. In addition, GLRT and HLRT can be applicable to different environments, e.g., Rician and Rayleigh fading [7]. Also, as one can easily notice that GLRT displays some implementation advantages over ALRT and HLRT, as it avoids the calculation of exponential functions and does not require the knowledge of noise power to compute the LF. However, with GLRT maximization over data symbols can lead to equal LFs for nested signal constellations, e.g., 16-QAM and 64-QAM, which in turn leads to incorrect classification [6]. Averaging over data symbols in HLRT removes the nested constellations problem of GLRT [6], though, with several unknown parameters, HLRT does not seem to be a good solution either, as finding the ML estimates of several parameters can be very time consuming. Low-complexity estimators can be used instead, leading to the so-called quasi-HLRT classifiers. For example, methods of moments were employed in [8] to estimate the channel parameters, when discriminating QAM signals in flat block fading channels. Then, (4) was used for decision, with the threshold set to one. A quasi-HLRT multi-antenna classifier, which exploits receive diversity for performance enhancement, was also reported in [8].

We summarize the afore-mentioned LB classifiers in Table I, emphasizing the type of modulations, unknown parameters and channel used.

IV. Feature Based Approach To MC

The design of a FB-MC algorithm essentially needs feature selection and decision-making. Features common to different modulations are selected in an ad-hoc way and the decision is made based on their differences. Some examples of features are statistics of the instantaneous amplitude, phase and frequency [9], [14], statistics of the signal itself [13], [16]-[20], the number of extrema in the PDF of the magnitude and peak magnitude of the signal wavelet transform [11], etc. Different methods can be used for decision-making, such as PDF-based [13]-[14], Euclidian distance [17]-[20], binary decision tree [9]-[10], neural networks [9], etc. The FB classifiers are summarized in Table II, including the features used, modulation types, unknown parameters and channel. Subsequently, the FB algorithms are presented from the perspective of a hierarchical approach, i.e., the modulation class of the incoming signal is first identified (e.g., QAM, PSK, ASK, FSK), and then the modulation order of each (M).

A. FB Algorithms for Modulation Class Identification

• ***Algorithms based on the instantaneous amplitude, phase and frequency***

The most intuitive way to identify the modulation class of an incoming signal is by using information extracted from the instantaneous amplitude, phase and frequency. In [9], the maximum of the discrete Fourier transform (DFT) of the normalized centered[13] amplitude was used to discriminate FSK and (ASK and PSK) classes, the variance of the absolute normalized centered phase was used to distinguish between M-PSK ($M > 2$) and real-valued constellation signals (BPSK and ASK), and the variance of direct (not absolute) normalized centered phase to distinguish between BPSK and ASK classes. A binary decision tree structure was employed to discriminate between classes. Discrimination within each class will be briefly discussed in Subsection B. At each tree node, the decision was made by comparing the selected statistic against a threshold[12]. A decision-tree classifier was also proposed in [10], and a practical system was developed based on this algorithm, at the Communication Research Center, Canada [24].

• ***Algorithms based on the wavelet transform***

The utility of the wavelet transform (WT) to localize transients in the instantaneous frequency, amplitude and phase of the received signal was also exploited for MC. The distinct behavior of the Haar WT (HWT) magnitude for PSK, QAM and FSK signals was explored for class identification in [11]-[12]. For a PSK signal the HWT magnitude is a constant, with peaks occurring at phase changes. On the other hand, because of the frequency and amplitude variation in an FSK and QAM signal, respectively, the HWT magnitude is a staircase function with peaks at phase changes, although these peaks do not provide useful information for non-continuous phase FSK signals. If only the phase is retained for a QAM signal, it behaves like a PSK signal and thus, the HWT magnitude is constant. On the other hand, as PSK and FSK signals are of constant amplitude, amplitude normalization has no effect on their HWT magnitude. After peaks removal, the variance of the HWT magnitude with amplitude normalization was used to distinguish between FSK and (QAM and PSK) classes, while the variance of HWT magnitude without amplitude normalization to distinguishing between QAM and PSK classes. The decision was made by comparing the selected feature against a threshold [11]-[12].

• ***Algorithm based on cumulants***

To discriminate amongst BPSK, ASK, M-PSK ($M > 2$) and QAM classes, the normalized fourth-order/two-conjugate cumulant of the received signal, $c_{r,4,2}(\mathbf{0}_3)/c_{r,2,1}^2(0)$ [10], was investigated in [13]. The PDF of the sample estimate was used for decision-making.

B. FB Algorithms for Modulation Order Identification

• ***Algorithms based on the instantaneous amplitude, phase and frequency***

Information extracted from the instantaneous amplitude, phase and frequency was also exploited to identify the modulation order (M) [9]. The variance of the absolute value of the normalized centered instantaneous amplitude (frequency) was used to distinguish between 2-ASK and 4-ASK (2-FSK and 4-FSK) [9]. The feature was compared against a threshold[12] for decision making, at a node, as part of the binary decision tree classifier mentioned in Section IV.A.

Statistical moments of the phase were investigated to identify PSK signals in [14], with the PDF of the sample estimates used for decision-making.

• ***Algorithms based on the wavelet transform***

Different PSK signals give rise to different sets of peak values in the magnitude of the Haar wavelet transform. The modulation order (M) of the PSK signal was selected by matching the histogram of the peaks with theoretical PDFs corresponding to different orders [11]. The number of modes in the PDF of the HWT magnitude was investigated for FSK signal recognition in [11]; the input was identified as M-FSK if there were $M/2+1$ to M modes in the histogram.

• ***Algorithms based on cumulants***

To identify the modulation order of ASK, PSK and QAM signals, features based on the fourth-order cumulants of the received signal were investigated in [13]. For example, the normalized cumulant of fourth-order/ zero-conjugate, $c_{r,4,0}(\mathbf{0}_3)/c_{r,2,1}^2(0)$ [10], was employed to identify the order of QAM signals. The PDF of the sample estimate was used for decision-making.

• ***Algorithms based on cyclostationarity***

Signal cyclostationarity was exploited for classification, with two manifestations, i.e., spectral line generation when passing the signal through different nonlinearities [15] and periodical fluctuation of cumulants up to the nth-order ($n = 8,6,4,2$) with time [16]-[20]. A generic classifier based on a feature vector whose components are the magnitudes of the cyclic cumulants (CCs)[14] up to the nth-order (q-conjugate, $q = 0,...,n$), raised to the power of $2/n$ [15], as n goes to infinity, and computed at all possible cycle frequencies (CFs) and delay vectors, $\boldsymbol{\tau}_{n\text{-}1}$, was proposed in [17]. The difficulty of implementing such a classifier led to simplified algorithms [17]-[18]. For example, the features investigated in [18] were based on the magnitudes of the CCs up to the sixth-order ($n = 6,4,2$, $q = 0,...,n$), at the CF $(n-2q)\Delta f + 1/T$ and a delay vector $\boldsymbol{\tau}_{n\text{-}1} = \mathbf{0}_{n\text{-}1}$, which ensures the maximum of the features. Such features are robust to the carrier phase and timing offset [17]-[19]. The decision was made based on the distance between the vector of sample estimates of the features and the vector of prescribed features, with the Euclidian norm employed as metric. The CC-based features were estimated from $K\rho$ [16] samples, taken over the observed K symbol interval [25]. Classification of single or multiple incoming signals present at the receiver was investigated in [18]. Multiple signals which overlap in time and frequency but have distinct CFs, can be distinguished using CCs (selectivity property of CCs) [17]. Eight-order CC-based features were investigated in [19] for classifying real- and complex-valued constellations, respectively. Features based on the nth-order CCs ($n = 4,6,8$, $q = n/2$), robust to carrier phase and timing offset, as well as frequency offset and phase jitter, were proposed for QAM recognition in [20]. The minimum Euclidian distance was also employed for decision in [19]-[20]. A raised-cosine pulse shape was considered in [17]-[20].

[13] The term "centered" specifies that the average is removed from the data set.

[14] For the definition of the nth-order CCs of a cyclostationary process, as well as the theoretical expression for the nth-order CCs of linear modulations see, for example, [17], [19] and [25].

[15] Raising the nth-order CC magnitudes to the power of $2/n$ forces the feature to take values within the same order of magnitude. Therefore, the classic Euclidian distance can be use for decision.

[16] The received signal is oversampled (multiple samples over a symbol period) in order to exploit signal cyclostationarity. The sampling frequency is ρ/T, with ρ a positive integer, called the oversampling factor.

V. NUMERICAL RESULTS

Subsequently, we present numerical results for QAM signal identification, with 16-QAM, 32-QAM and 64-QAM as candidate modulations ($N_{\mathrm{mod}} = 3$)[17]. In classifying N_{mod} modulations, we use the average probability of correct classification as a performance measure. This is defined as $P_{cc} = N_{\mathrm{mod}}^{-1}\sum_{i=1}^{N_{\mathrm{mod}}} P_c^{(i|i)}$, where $P_c^{(i|i)}$ is the probability to declare that the i th modulation is received when indeed it has been originally transmitted. The $P_c^{(i|i)}$ is estimated based on 1000 Monte Carlo simulations. The number of processed symbols is $K = 100$, the pulse shape is rectangular, and the SNR per symbols is defined as $\gamma_s = ST/N_0$. We set $T = 1$ and $S = 1$, and change the SNR by varying N_0.

The ALRT-based classifier, as defined by (4) and (6), with $\eta_A = 1$, serves as a benchmark, against which performances of other classifiers are compared. We present in Table III the performance of this classifier, as well as a sensitivity effect to several model mismatches. Such analysis provides bounds of performance, imposed by different preprocessing operations.

In an ideal scenario (AWGN and all parameters assumed perfectly known), a P_{cc} of one is attained at 14dB SNR (III-1), and a P_{cc} of 0.9 at 11dB (III-2). Henceforth, the SNR is set to 14dB, unless otherwise mentioned. An acceptable performance (P_{cc} above 0.9) is reached for a normalized carrier frequency offset ΔfT lower than 3.3x10^{-4} (III-3). When investigated for model mismatch, the carrier phase is fixed over a realization, but varies randomly, $U[-\theta_{\lim},\theta_{\lim})$ from realization to realization. An acceptable performance is achieved for $\theta_{\lim} \leq 9^\circ$ (III-4). A similar result ($\phi_{\lim} \leq 12^\circ$) is obtained when studying performance degradation due to a phase jitter, with the phase ϕ_k modeled as a r.v. uniformly distributed over $[-\phi_{\lim},\phi_{\lim})$, which varies from symbol to symbol (III-5). For a rectangular pulse shape, one can easily show that a synchronization error of ε translates, after matched filtering, to an equivalent two-path channel $[1-\varepsilon\ \varepsilon]$. An acceptable P_{cc} is still achieved for $\varepsilon = 0.074$ (III-6). With an error ΔS in estimating the signal power S, an acceptable performance is achieved for $|\Delta S / S| \leq 22\%$ (III-7). The effect of the impulsive noise was investigated using a contaminated Gaussian noise PDF, $(1-p)\mathcal{N}(0,\sigma_n^2) + p\mathcal{N}(0,100\sigma_n^2)$ [26]. Here the addition refers to "mixture", i.e., the process is realized from $\mathcal{N}(0,\sigma_n^2)$ with probability $1-p$ and from $\mathcal{N}(0,100\sigma_n^2)$ with probability p. The impulsive noise causes occasional impulsive events of a magnitude that is considerably greater than the background noise. We set the SNR based on the total noise variance, i.e., $N_0 = (1-p)\sigma_n^2 + p(100\sigma_n^2)$, and $p = 10^{-2}$. Another type of non-Gaussian noise used in the sensitivity analysis was generated by a gun fired repeatedly (Machine gun noise) [27]. With non-Gaussian noise, 13.5dB SNR is required to achieve a P_{cc} of 0.9 (III-8). With a PSK signal as interference, i.e., transmitted on the same carrier frequency and at the same symbol rate as the QAM signal, a signal to interference ratio (SIR) greater than 18.4 dB is required to identify the QAM signals with a P_{cc} higher than 0.9 (III-9). The SIR is defined as the signal power over the interference power. Interestingly, we have noticed that classification performance degradation does not depend on the modulation order of the interference.

[17] Of course, when higher order modulations are included in the modulation pool, higher SNRs and/or a larger number of symbols are needed to achieve the same performance. We have simulated these modulations to draw some basic, yet insightful conclusions.

In summary, results of the sensitivity analysis, i.e., the need for high SNR (14dB and 13.5dB compared to 11dB) show that the ALRT-based classifier is not robust to model mismatches.

VI. CONCLUSION

A useful summary of many existing LB- and FB-MC algorithms is provided in this paper, which gives the reader an overview of the approaches used so far. The LB approach provides an optimal solution to the MC problem (ALRT), in the sense that it maximizes the average probability of correct classification. However, the complexity of the optimal solution, in many cases of interest naturally, gives rise to proposing suboptimal algorithms, e.g., the quasi-ALRT classifiers. Using ML estimates of the unknown quantities, GLRT and HLRT were investigated as two alternatives. Although GLRT has some advantages, it fails in identifying nested constellations. On the other hand, HLRT can be implemented with less complexity than ALRT, and still achieving a reasonable performance. The complexity is further reduced in quasi-HLRT classifiers, which rely on low-complexity yet accurate parameter estimators. Obviously, there is a trade-off between the complexity and performance, which depends on the estimation method. Although a FB method may not be optimal, it is usually simple to implement, with near-optimal performance, when designed properly.

Accurate preprocessing is required for the effective implementation of most of the known MC algorithms. Devising low-complexity blind algorithms for joint parameter estimation is a topic of interest in MC. In addition, development of classification methods which rely less on preprocessing is another topic for further investigation. New classification problems have raised as a result of emerging wireless technologies, such as single carrier versus multicarrier modulation recognition, classification of signals transmitted using single and multiple antennas, identification of space-time modulation formats, etc. These issues mean that MC in real-world environments continues to be a dynamic research field.

REFERENCES

[1] W. Wei and J. M. Mendel, "Maximum-likelihood classification for digital amplitude-phase modulations," *IEEE Trans. Commun*., vol. 48, pp. 189-193, 2000.

[2] C. Y. Huang and A. Polydoros, "Likelihood methods for MPSK modulation classification," *IEEE Trans. Commun*., vol. 43, pp. 1493-1504, 1995.

[3] C. Long, K. Chugg, and A. Polydoros, "Further results in likelihood classification of QAM signals," *in Proc. MILCOM*, 1994, pp. 57-61.

[4] B. F. Beidas and C. L. Weber, "Higher-order correlation-based approach to modulation classification of digitally frequency-modulated signals," *IEEE Journal on Sel. Areas in Commun*, vol. 13, pp. 89-101, 1995.

[5] B. F. Beidas and C. L. Weber, "Asynchronous classification of MFSK signals using the higher order correlation domain," *IEEE Journal on Sel. Areas in Commun*, vol. 46, pp. 480-493, 1998.

[6] P. Panagiotou, A. Anastasoupoulos, and A. Polydoros, "Likelihood ratio tests for modulation classification," in *Proc. MILCOM*, 2000, pp. 670-674.

[7] O. A. Dobre, J. Zarzoso, Y. Bar-Ness, and W. Su, "On the classification of linearly modulated signals in fading channel", in *Proc. CISS*, 2004.

[8] A. Abdi, O. A. Dobre, R. Chauchy, Y. Bar-Ness, and W. Su, "Modulation classification in fading channels using antenna arrays," in *Proc. MILCOM*, 2004.

[9] E. E. Azzouz and A. K. Nandi, *Automatic Modulation Recognition of*

Communication Signals. Kluwer Academic, 1996.
[10] D. Boudreau et al., “A fast automatic modulation recognition algorithm and its implementation in a spectrum monitoring application,” *in Proc. MILCOM*, 2000, pp. 732-736.
[11] K. C. Ho, W. Prokopiw, and Y. T. Chan, “Modulation identification of digital signals by the wavelet transform,” *IEE Proc. Radar, Sonar and Navig.*, vol. 47, pp. 169-176, 2000.
[12] L. Hong and K. C. Ho, “Identification of digital modulation types using the wavelet transform,” in *Proc. MILCOM*, 1999, pp. 427-431.
[13] A. Swami and B. M. Sadler, “Hierarchical digital modulation classification using cumulants,” *IEEE Trans. Commun.*, vol. 48, pp. 416-429, 2000.
[14] S. S. Soliman and S. Z. Hsue, “Signal classification using statistical moments,” *IEEE Trans. Commun.*, vol. 40, pp. 908-916, 1992.
[15] J. Reichert, “Automatic classification of communication signals using higher order statistics,” in *Proc. ICASSP*, 1992, pp. 221-224.
[16] P. Marchand, J. L. Lacoume, and C. Le Martret, “Multiple hypothesis classification based on cyclic cumulants of different orders,” in *Proc. ICASSP*, 1998, pp. 2157-2160.
[17] C. M. Spooner, “Classification of co-channel communication signals using cyclic cumulants,” in *Proc. ASILOMAR*, 1996, pp.531-536.
[18] C. M. Spooner, “On the utility of sixth-order cyclic cumulants for RF signal classification,” in *Proc. ASILOMAR,* 2001, pp. 890-897.
[19] O. A. Dobre, Y. Bar-Ness, and W. Su, “Higher-order cyclic cumulants for high order modulation classification,” in *Proc. MILCOM*, 2003, pp. 112-117.
[20] O. A. Dobre, Y. Bar-Ness, and W. Su, “Robust QAM modulation classification algorithm based on cyclic cumulants,” in *Proc. WCNC*, 2004, pp. 745-748.
[21] J. G. Proakis, *Digital Communications*, 4th ed., New York: McGraw-Hill, 2001.
[22] C. L Nikias and A. P. Petropulu, *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[23] H. L. Van Trees, *Detection, Estimation and Modulation Theory- Part I*. New York: Wiley, 1968.
[24] Communication signal analyzer, Communication Research Center, Canada, May 2000, www-ext.crc.ca.
[25] V. Dandawade and G. B. Giannakis, ”Asymptotic theory of mixed time averages and kth- order cyclic-moment and cumulant statistics”, *IEEE Trans. Inform. Theory*, vol. 41, no.1, pp. 216-232, 1995.
[26] M. E. Johnson, *Multivariate Statistical Simulation*. New York: Wiley, 1987.
[27] Signal Processing Information Base, Rice University, Houston, TX, URL http://spie.ece.rice.edu/spib/data/signals/.

Table I. A SUMMARY OF LIKELIHOOD-BASED CLASSIFIERS.

Proposed by	Classifier	Modulations	Unknown parameters	Channel
Wei and Mendel [1]	ALRT	Linear modulations	-	AWGN
Huang and Polydoros [2]	Quasi-ALRT	PSK signals	Carrier phase and timing offset	AWGN
Long et al. [3]	Quasi-ALRT	QAM signals	Carrier phase	AWGN
Beidas and Weber [4]-[5]	ALRT, Quasi-ALRT	FSK signals	Carrier phase and timing offset	AWGN
Panagiotu et al. [6]	GLRT, HLRT	Linear modulations	Carrier phase	AWGN
Dobre et al. [7]	HLRT	Linear modulations	Channel amplitude and phase	Flat fading
Abdi et al. [8]	ALRT	Linear modulations, FSK signals	Channel amplitude and phase	Flat fading
	Multi-antenna Quasi-HLRT	QAM signals		

Table II. A SUMMARY OF FEATURE-BASED CLASSIFIERS.

Proposed by	Features	Modulations	Unknown parameters	Channel
Azzouz and Nandi [9]	Maximum DFT of normalized centered amplitude, variance of normalized centered (absolute) amplitude, phase and frequency	2ASK,4ASK,BPSK, QPSK,2FSK,4FSK	-	AWGN
Ho et al. [11]	Variance of the HWT magnitude, the HWT magnitude and peak magnitude histograms	FSK and PSK classes, as well as within each classes	-	AWGN
Hong and Ho [12]	Variance of the HWT magnitude	FSK, PSK and QAM classes	-	AWGN
Swami and Sadler [13]	Normalized fourth-order cumulants of the received signal	Linear modulations	Carrier phase, frequency and timing offsets	AWGN, impulsive noise, cochannel interferences
Spooner [17]-[18]	The $2/n$ th powers of the magnitudes of the n th-order CCs of the received signal, $n = 6,4,2$ and $q = 0,...,n$	Linear modulations and minimum shift keying (MSK)	Carrier phase, frequency and timing offsets, symbol period, signal amplitude and pulse shape	AWGN, cochannel interferences
Dobre et al. [20]	The $2/n$ th powers of the magnitudes of the n th-order CCs of the received signal, $n = 8,6,4$ and $q = n/2$	QAM signals	Carrier phase, frequency and timing offsets, and phase jitter	AWGN, impulsive noise

Table III. RESULTS OF THE SENSITIVITY ANALYSIS OF THE ALRT-BASED CLASSIFIER TO MODEL MISMATCHES.

	Classifier	Model mismatch	SNR(dB)	P_{cc}
1	ALRT, (4) and (6), with $\eta_A = 0$	Ideal case	14	1
2	ALRT (as in first row)	Ideal case	11	0.9
3	ALRT (as in first row)	Carrier frequency offset ($\Delta fT = 3.3\times10^{-4}$)	14	0.9
4	ALRT (as in first row)	Carrier phase ($\theta_{\lim} = 9^\circ$, with θ r.v. $U[-\theta_{\lim}, \theta_{\lim})$)	14	0.9
5	ALRT (as in first row)	Phase jitter ($\phi_{\lim} = 12^\circ$, with ϕ_k r.v. $U[-\phi_{\lim}, \phi_{\lim})$)	14	0.9
6	ALRT (as in first row)	Timing offset ($\varepsilon = 0.074$)	14	0.9
7	ALRT (as in first row))	Error in estimating the signal power ($\Delta S / S = 22\%$)	14	0.9
8	ALRT (as in first row))	Impulsive noise (mixture model, as well a machine gun noise)	13.5	0.9
9	ALRT (as in first row)	Cochannel interferences (PSK signal as interference, SIR=18.4dB)	14	0.9