Computationally Efficient Modulation Level Classification Based on Probability Distribution Distance Functions

Paulo Urriza, Eric Rebeiz, Przemysław Pawełczak, and Danijela Čabrić

Abstract—We present a novel modulation level classification (MLC) method based on probability distribution distance functions. The proposed method uses modified Kuiper and Kolmogorov- Smirnov (KS) distances to achieve low computational complexity and outperforms the state of the art methods based on cumulants and goodness-of-fit (GoF) tests. We derive the theoretical performance of the proposed MLC method and verify it via simulations. The best classification accuracy under AWGN with SNR mismatch and phase jitter is achieved with the proposed MLC method using Kuiper distances.

I. INTRODUCTION

Modulation level classification (MLC) is a process which detects the transmitter's digital modulation level from a received signal using a priori knowledge of the modulation class and signal characteristics needed for downconversion and sampling. Among many MLC methods [1], a cumulant (Cm) computation-based classification [2] is one of the most widespread. However, Cm-based classification possesses some limitations. First, MLC of some modulation classes (e.g. M-PSK) requires sixth and eighth order cumulants which are computationally expensive. Second, for multi-level modulations, the difference among cumulants is small and requires a large number of samples to achieve good accuracy. A recently proposed method [3] based on a goodness-of-fit (GoF) test using Kolmogorov-Smirnov (KS) statistic has been suggested as an alternative to the Cm-based classification. The KS method achieves better classification accuracy given a particular sample size, especially at high values of SNR. Both methods are equally computationally expensive [3, Sec. III-A3].

In this letter, we propose a novel MLC method based on distribution distance functions, mainly Kuiper (K) [4] and KS distances, which is a significant simplification of methods based on GoF. We show that using a classifier based only on K-distance achieves a better classification than the KS-based GoF classifier. At the same time our method requires only $2ML$ additions in contrast to $2M(\log 2M + 2K)$ additions for the KS-based GoF test, where K is the number of distinct modulation levels, M is the sample size and $L \ll M$ is the number of test points used by our method.

The authors are with the Department of Electrical Engineering, University of California, Los Angeles, 56-125B Engineering IV Building, Los Angeles, CA 90095-1594, USA (email: {pmurriza, rebeiz, przemek, danijela}@ee.ucla.edu).

II. PROPOSED MLC METHOD

A. System Model

Following [3], we assume a sequence of M discrete, complex, i.i.d. and sampled baseband symbols, $\mathbf{s}^{(k)} \triangleq [s_1^{(k)} \cdots s_M^{(k)}]$, drawn from a modulation order $\mathcal{M}_k \in \{\mathcal{M}_1, \ldots, \mathcal{M}_K\}$, transmitted over AWGN channel, perturbed by uniformly distributed phase jitter and attenuated by an unknown factor $A > 0$. Therefore, the received signal is given as $\mathbf{r} \triangleq [r_1 \cdots r_M]$, where $r_n = Ae^{j\Phi_n}s_n + g_n$, $\{g_n\}_{n=1}^M \sim \mathcal{CN}\left(0, \sigma^2\right)$ and $\{\Phi_n\}_{n=1}^M \sim \mathcal{U}\left(-\phi, +\phi\right)$. The task of the modulation classifier is to find $\mathcal{M}_k$, from which $\mathbf{s}^{(k)}$ was drawn, given $\mathbf{r}$. Without loss of generality, we consider unit power constellations and define SNR as $\gamma \triangleq A^2/\sigma^2$.

B. Classification based on Distribution Distance Function

The proposed method modifies MLC technique based on GoF testing using the KS statistic [3]. Since the KS statistic, which computes the minimum distance between theoretical and empirical cumulative distribution function (ECDF), requires *all* CDF points, we postulate that similarly accurate classification can be obtained by evaluating this distance using a smaller set of points in the CDF.

Let $\mathbf{z} \triangleq [z_1 \cdots z_N] = f(\mathbf{r})$ where $f(\cdot)$ is the chosen feature map and N is the number extracted features. Possible feature maps include $|\mathbf{r}|$ (magnitude, $N = M$) or the concatenation of $\Re\{\mathbf{r}\}$ and $\Im\{\mathbf{r}\}$ (quadrature, $N = 2M$). The theoretical CDF of z given $\mathcal{M}_k$ and γ, $F_0^k(z)$, is assumed to be known a priori (methods of obtaining these distributions, both empirically and theoretically, are presented in [3, Sec. III-A]). The K CDFs, one for each modulation level, define a set of test points

$$t_{ij}^{(\epsilon)} = \arg\max_z D_{ij}^{(\epsilon)}(z), \tag{1}$$

with the distribution distances given by

$$D_{ij}^{(\epsilon)}(z) = (-1)^\epsilon \left(F_0^i(z) - F_0^j(z) \right), \tag{2}$$

for $1 \le i, j \le K$, $i \ne j$, and $\epsilon \in \{0, 1\}$, corresponding to the maximum positive and negative deviations, respectively. Note the symmetry in the test points such that $t_{ji}^{(0)} = t_{ij}^{(1)}$. Thus, there are $L \triangleq 2\binom{K}{2}$ test points for a K order classification.

The ECDF, given as

$$F_N(t) = \frac{1}{N} \sum_{n=1}^N \mathbb{I}(z_n \le t), \tag{3}$$

is evaluated at the test points to form $\mathbf{F}_N \triangleq \{F_N(t_{ij}^{(\epsilon)})\}$, $1 \le i, j \le K, i \ne j$. Here, $\mathbb{I}(\cdot)$ equals to one if the input is

true, and zero otherwise. By evaluating $F_N(t)$ only at the test points in (1), we get

$$\hat{D}_{ij}^{(\epsilon)} = (-1)^\epsilon \left(F_N \left(t_{ij}^{(\epsilon)} \right) - F_0^j \left(t_{ij}^{(\epsilon)} \right) \right) \qquad (4)$$

which are then used to find an estimate of the maximum positive and negative deviations

$$\hat{D}_j^{(\epsilon)} = \max_{1 \le i \le K, i \ne j} \hat{D}_{ij}^{(\epsilon)}, \quad 1 \le j \le K, \qquad (5)$$

of the ECDF to the true CDFs. The operation of finding the ECDF at the given testpoints (4) can be implemented using a simple thresholding and counting operation and does not require samples to be sorted as in [3]. The metrics in (5) are used to find the final distribution distance metrics

$$\hat{D}_j = \max \left(\left| \hat{D}_j^{(0)} \right|, \left| \hat{D}_j^{(1)} \right| \right), \quad \hat{V}_j = \left| \hat{D}_j^{(0)} + \hat{D}_j^{(1)} \right|, \quad (6)$$

which are the reduced complexity versions of the KS distance (rcKS) and the K distance (rcK), respectively[1]. Finally, we use the metrics in (6) as substitutes to the true distance-based classifiers with the following rule: choose $\mathcal{M}_{\hat{k}}$ such that

$$\hat{k}_D = \arg \min_{1 \le j \le K} \hat{D}_j, \quad \hat{k}_V = \arg \min_{1 \le j \le K} \hat{V}_j. \qquad (7)$$

In the remainder of the letter, we define $h_{\hat{D}}(\mathbf{F}_N) = \hat{k}_D$ and $h_{\hat{V}}(\mathbf{F}_N) = \hat{k}_V$, where $\hat{k}_D, \hat{k}_V \in \{1, \dots, K\}$.

C. Analysis of Classification Accuracy

Let $\mathbf{t} \triangleq [t_1 \cdots t_L]$ denote the set of test points, $\{t_{ij}^{(\epsilon)}\}$, sorted in ascending order. For notational consistency, we also define the following points, $t_0 \triangleq -\infty$ and $t_{L+1} \triangleq +\infty$. Given that these points are distinct, they partition $\mathbf{z}$ into $L+1$ regions. An individual sample, z_n, can be in region l, such that $t_{l-1} < z_n \le t_l$, with a given probability, determined by $F_0^k(z)$.

Assuming z_n are independent of each other, we can conclude that given $\mathbf{z}$, the number of samples that fall into each of the $L+1$ regions, $\mathbf{n} \triangleq [n_1 \cdots n_{L+1}]$, is jointly distributed according to a multinomial PMF given as

$$f(\mathbf{n}|N, \mathbf{p}) = \begin{cases} \frac{N! p_1^{n_1} \cdots p_{L+1}^{n_{L+1}}}{n_1! \cdots n_{L+1}!}, & \text{if } \sum_{i=1}^{L+1} n_i = N, \\ 0, & \text{otherwise,} \end{cases} \qquad (8)$$

where $\mathbf{p} \triangleq [p_1 \cdots p_{L+1}]$, and p_l is the probability of an individual sample being in region l. Given that $\mathbf{z}$ is drawn from $\mathcal{M}_k$, $p_l = F_0^k(t_l) - F_0^k(t_{l-1})$, for $0 < l \le L+1$.

Now, with particular $\mathbf{n}$, the ECDF at all the test points is

$$\mathbf{F}_N(\mathbf{n}) \triangleq [F_N(t_1) \cdots F_N(t_L)], \quad F_N(t_l) = \frac{1}{N} \sum_{i=1}^{l} n_i. \quad (9)$$

Therefore, we can analytically find the probability of classification to each of the K classes as

$$\Pr(\hat{k} = \kappa | \mathcal{M}_k) = \sum_{\mathbf{n} \in \mathbb{N}^{L+1}} \mathbb{I}(h_{\hat{V}}(\mathbf{F}_N(\mathbf{n})) = \kappa) f(\mathbf{n}|N, \mathbf{p}), \quad (10)$$

for the rcK classifier. A similar expression can be applied to rcKS, replacing $h_{\hat{V}}(\cdot)$ with $h_{\hat{D}}(\cdot)$ in (10).

---

[1]Note, that other non-parametric distances used in hypothesis testing exist (see introduction in e.g. [4]), although for brevity they are not addressed here. We note, however, that our approach is easily applied to any assumed distance metric.

Method	Multiply	Add	Memory
Cm	$6M$	$6M$	K
rcKS/rcK	0	$2ML$	$WL(K+1)$
KS/K	0	$2M(\log 2M + 2K)$	$KW\bar{N}$
rcKS/rcK (mag)	$2M$	$M(L+1)$	$WL(K+1)$
KS/K (mag)	$2M$	$M(\log M + 2K + 1)$	$KW\bar{N}$

D. Complexity Analysis

Given that the theoretical CDFs change with SNR, we store W SNR profiles for each modulation level (impact of the selection of W on the accuracy is discussed further in Section III-B.) Further, we store KW theoretical CDFs of length $\bar{N}$ each. For the non-reduced complexity classifiers that require samples to be sorted, we also assume that we use a sorting algorithm whose complexity is $N \log N$. Based on the results shown in Table I, we see the rcK/rcKS tests use less addition operations than K/KS-based methods [3] and Cm-based classification [2]. Furthermore, their memory requirements are also smaller since $\bar{N}$ has to be large for a smooth CDF. It is worth mentioning that the authors in [3] used the theoretical CDF, but used $\bar{N}$ as the number of samples to generate the CDF in their complexity figures. The same observation favoring the proposed rcK/rcKS (mag) methods holds for the magnitude-based classifiers [3, Sec III-A].

III. RESULTS

As an example, we assume that the classification task is to distinguish between M-QAM, where M $\in \{4, 16, 64\}$. For comparison we also present classification result based on maximum likelihood estimation (ML).

A. Detection Performance versus SNR

In the first set of experiments we evaluate the performance of the proposed classification method for different values of SNR. The results are presented in Fig. 1. We assume fixed sample size of $M = 50$, in contrast to [3, Fig. 1] to evaluate classification accuracy for a smaller sample size. We confirm that even for small sample size, as shown in [3, Fig. 1], Cm has unsatisfying classification accuracy at high SNR. In (10,17) dB region rcK clearly outperforms all detection techniques, while as SNR exceeds ≈ 17 dB all classification methods (except Cm) converge to one. At low SNR region, 0–10 dB, KS, rcKS, rcK perform equally well, with Cm having comparable performance. The same observation holds for larger sample sizes, not shown here due to space constraints. Note that the analytical performance metric developed in Section II-C for rcK and rcKS matches perfectly with the simulations. For the remaining results, we set $\gamma = 12$ dB, unless otherwise stated.

B. Detection Performance versus Sample Size

In the second set of experiments, we evaluate the performance of the proposed classification method as a function of sample size M. The result is presented in Fig. 2. As observed in Fig. 1, also here Cm has the worst classification accuracy,
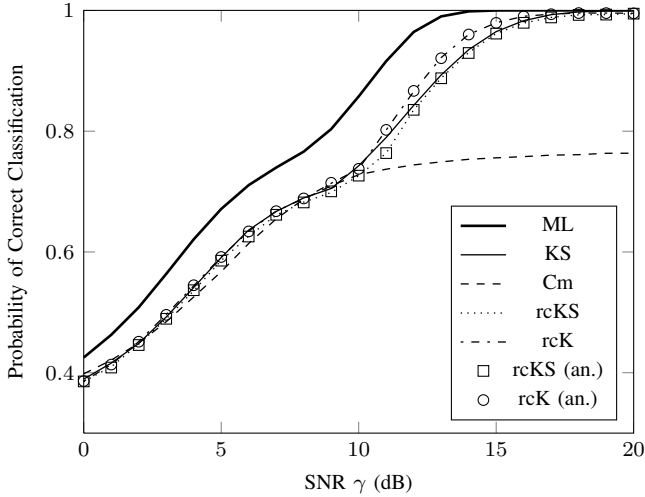
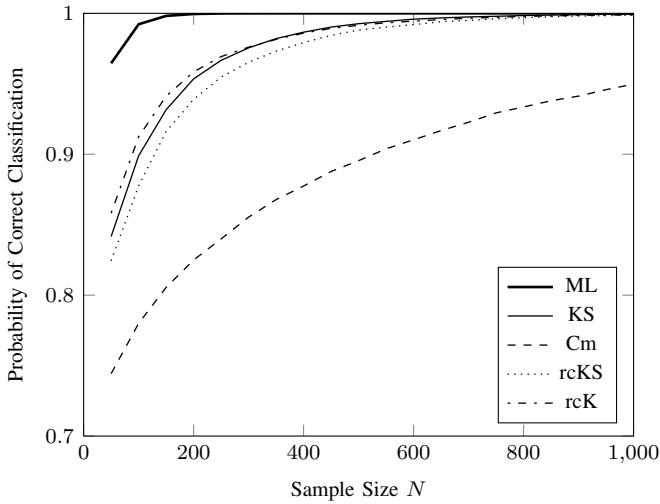Fig. 1. Effect of varying SNR on the probability of classification with M=50.

Fig. 2. Effect of varying sample size on the probability of classification with $\gamma = 12\,\mathrm{dB}$.

e.g. 5% below upper bound at $M = 1000$. The rcK method performs best at small sample sizes, $50 \leq M \leq 300$. With $M > 300$, the accuracy of rcK and KS is equal. Classification based on rcKS method consistently falls slightly below rcK and KS methods. In general, rcKS, rcK and KS converge to one at the same rate.

C. Detection Performance vs SNR Mismatch and Phase Jitter

In the third set of experiments we evaluate the performance of the proposed classification method as a function of SNR mismatch and phase jitter. The result is presented in Fig. 3. In case of SNR mismatch, Fig. 3(a), our results show the same trends as in [3, Fig. 4]; that is, all classification methods are relatively immune to SNR mismatch, i.e. the difference between actual and maximum SNR mismatch is less than 10% in the considered range of SNR values. This justifies the selection of limited set of SNR values W for complexity evaluation used in Section II-D. Interestingly, the ML method becomes
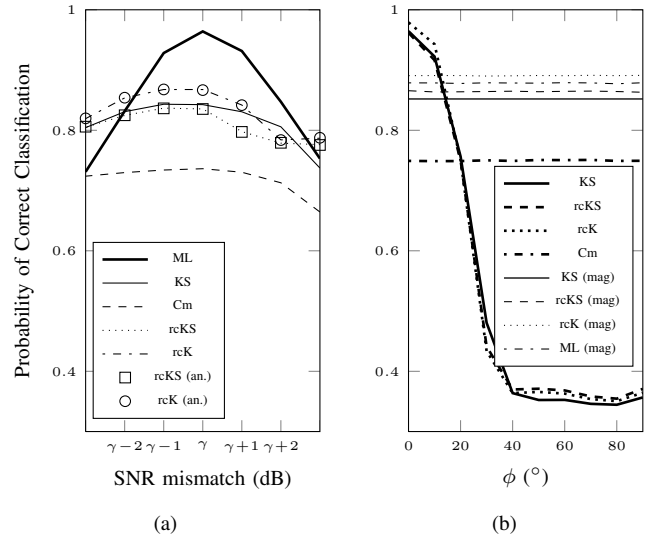

(a)   (b)

Fig. 3. Effect of (a) SNR mismatch and (b) phase jitter on classification accuracy with $M = 50$; (a) $\gamma = 12\,\mathrm{dB}$, (b) $\gamma = 15\,\mathrm{dB}$.

worse than methods based on distribution distance functions and GoF tests, at SNR mismatch larger than approximately $\pm 1\,\mathrm{dB}$. Note again the perfect match of analytical result of Section II-C with the simulations.

In the case of phase jitter caused by imperfect downconversion, we present results in Fig. 3(b) for $\gamma = 15\,\mathrm{dB}$ as in [2], in contrast to $\gamma = 12\,\mathrm{dB}$ used earlier, for comparison purposes. In contrast to [2, Fig. 5] our proposed classification method using signal magnitudes is invariant to phase jitter over the entire phase range. rcK, rcKS and ML perform almost equally well, while Cm is worse than the other three methods by $\approx 10\%$. Quadrature-based classifiers, as expected, are highly sensitive to phase jitter. Note that in the small phase jitter, $\phi < 10°$, quadrature-based classifiers perform better than others, since the sample size is twice as large as in the former case.

IV. CONCLUSION

In this letter we presented a novel, computationally efficient method for modulation level classification based on distribution distance functions. Specifically, we proposed to use a metric based on Kolmogorov-Smirnov and Kuiper distances which exploits the distance properties between CDFs corresponding to different modulation levels. The proposed method results in lower computational complexity compared to both cumulant-based and KS GoF-based methods.

REFERENCES

[1] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: Classical approaches and new trends," *IET Communications*, vol. 1, no. 2, pp. 137–156, Apr. 2007.
[2] A. Swami and B. M. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 416–429, Mar. 2000.
[3] F. Wang and X. Wang, "Fast and robust modulation classification via Kolmogorov-Smirnov test," *IEEE Trans. Wireless Commun.*, vol. 58, no. 8, pp. 2324–2332, Aug. 2010.
[4] G. A. P. Cirrone, S. Donadio, S. Guatelli, A. Mantero, B. Masciliano, S. Parlati, M. G. Pia, A. Pfeiffer, A. Ribon, and P. Viarengo, "A goodness of fit statistical toolkit," *IEEE Trans. Nucl. Sci.*, vol. 51, no. 5, pp. 2056–2063, Oct. 2004.