A Faster Maximum-Likelihood Modulation Classification in Flat Fading Non-Gaussian Channels

Wenhao Chen, Zhuochen Xie, Lu Ma, Jie Liu, and Xuwen Liang

***Abstract*—In this letter, we use squared iterative method with parameter checking to accelerate the convergence rate of expectation/conditional maximization (ECM) algorithm when estimating the channel parameters blindly in flat fading non-Gaussian channels, and further, we proposed automatic modulation classification (AMC) in flat fading non-Gaussian channels based on the proposed maximum likelihood estimator. The numerical results show that the proposed method can accelerate the convergence rate of ECM algorithm, and AMC based on the proposed method is faster than that based on ECM, while the accuracy of the former shows nearly no loss compared with that of the latter.**

***Index Terms*—Maximum likelihood estimation, expectation/conditional maximization algorithm, squared iterative method, Gaussian mixture model, automatic modulation classification.**

I. Introduction

AUTOMATIC modulation classification (AMC) can be defined as the process of identifying the modulation type of a noisy signal from a given set of possible modulation types [1]. It is an intermediate step between signal detection and demodulation, and plays a key role in various military and civilian applications, such as electronic warfare, software defined radio, interference identification, spectrum management and so on.

Recently, more and more literatures pay attentions to AMC under non-Gaussian noise because experimental researches have shown that most radio channels experience both natural and man-made noise, and that the combined noise is a highly non-Gaussian process. References [2] and [3] proposed AMC based on modulation dependent features. References [4] and [5] proposed fast and robust AMC based on distribution test. While feature-based and distribution test-based classifiers are generally easier to implement, they are sub-optimal. Likelihood-based classifiers are optimal in the Bayesian sense, as they minimize the probability of classification error [6]. In [7], the non-Gaussian noise was modeled by the N-term Gaussian mixture, then the hybrid likelihood ratio test (HLRT) was used for identifying the modulation formats. The unknown model parameters were estimated blindly by using the Expectation/Conditional Maximization (ECM) algorithm. Due to the slow convergence rate of ECM, AMC based on ECM is time-consuming.

In this letter, to accelerate AMC under non-Gaussian noise, we propose a novel method for estimating channel coefficient and parameters of Gaussian mixture model (GMM), which is based on squared iterative method (SQUAREM) to accelerate ECM [8]. In learning GMM, basic SQUAREM (bSQUAREM) will not preserve the constraints of some parameters, e.g. the mixture proportions are non-negative, covariance matrix of each Gaussian component is symmetric positive definite, which will lead to meaningless estimates. So we add parameters checking in each iteration of SQUAREM, we call it SQUAREM with Parameters Checking (SQUAREM-PC) algorithm, which can accelerate the convergence rate of ECM and converge to a stationary point of log-likelihood function.

The rest of this letter is organized as follows. Section II describes the channel model and modulation classifier. Section III presents the details of proposed algorithm. Numerical results are provided in Section IV. Finally, concluding remarks are given in Section V.

II. Model Description and Modulation Classifier

Let M be the set of linear modulation constellation points of size $m = |M|$. The transmitted symbols $\{s_t\}_{t=1}^{T}$ are uniformly and independently drawn from M. After preprocessing, the baseband complex envelope of the received signal sampled at one sample per symbol at the output of a matched filter can be written as

$$y_t = \alpha s_t + \varepsilon_t, \quad t = 1, \ldots, T, \tag{1}$$

where α is a complex factor used to represent both the flat fading experienced by the signal and the unknown power and carrier phase of the transmitted signal, T denotes the number of received symbols. As in [4], we assume that $\{\varepsilon_t\}_{t=1}^{T}$ is a set of independent identically distributed (i.i.d.) complex random variables, the real and imaginary part of ε_t are i.i.d.. ε_t can be seen as a 2-dimension vector including real and imaginary part of it. Here we choose N-term Gaussian mixture model to approximate the probability density function (pdf) of ε_t

$$p(\boldsymbol{x}) = \sum_{n=1}^{N} \frac{\lambda_n}{2\pi|\Sigma_n|^{\frac{1}{2}}} \exp(-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\Sigma_n^{-1}\boldsymbol{x}), \tag{2}$$

Manuscript received November 28, 2018; revised December 27, 2018; accepted January 16, 2019. Date of publication January 23, 2019; date of current version March 8, 2019. This work was supported by Natural Science Foundation of China (Grant No. 91738201) and Shanghai sailing program (Grant No. 17YF1418200). The associate editor coordinating the review of this letter and approving it for publication was F. Wang. *(Corresponding author: Xuwen Liang.)*

W. Chen is with the Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China, also with the Shanghai Engineering Center for Microsatellites, Chinese Academy of Sciences, Shanghai 201210, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: 18221505463@163.com).

Z. Xie, J. Liu, and X. Liang are with the Shanghai Engineering Center for Microsatellites, Chinese Academy of Sciences, Shanghai 201210, China (e-mail: xiezc.ac@hotmail.com; liujiezd@126.com; 18217631362@163.com).

L. Ma is with Shanghai Spaceok Technology Co., Ltd., Shanghai 201802, China (e-mail: e_wqs@hotmail.com).

Digital Object Identifier 10.1109/LCOMM.2019.2894400

where λ_n is the probability that ε_t is chosen from the n^{th} term in the pdf, with $0 \leq \lambda_n \leq 1$ and $\sum_{n=1}^{N} \lambda_n = 1$. $\Sigma_n = (\sigma_n^2/2)\boldsymbol{I}$, where $\boldsymbol{I}$ denotes identity matrix, and σ_n^2 is the energy of the n^{th} Gaussian component.

The main work in AMC is choosing the modulation format from $S = \{M_1, \ldots, M_C\}$ by analyzing received signals $\{y_t\}_{t=1}^{T}$. Maximum-likelihood modulation classification is a composite hypothesis testing problem in which the hypothesis that maximizes the log-likelihood of $\{y_t\}_{t=1}^{T}$ is chosen,

$$\hat{H} = \arg\max_{H_i} \ln p(y_1, \ldots, y_T|H_i), \tag{3}$$

where H_i be the hypothesis that the modulation scheme of received signal is $M_i, i = 1, \ldots, C$, where C is the total number of possible modulation types.

By analyzing the received signal model, the log-likelihood of $\{y_t\}_{t=1}^{T}$ can be calculated by

$$\ln p(y_1, \ldots, y_T|H_i) = \sum_{t=1}^{T} \ln\{\frac{1}{m_i} \sum_{j=1}^{m_i} \sum_{n=1}^{N} \frac{\lambda_n}{\pi\sigma_n^2} \exp(-\frac{|y_t - \alpha s_{tj}^i|^2}{\sigma_n^2})\}, \tag{4}$$

where s_{tj}^i is the j^{th} constellation point of modulation constellation M_i, $m_i = |M_i|$ denotes the size of M_i. Because $\alpha, \{\lambda_n\}_{n=1}^N, \{\sigma_n^2\}_{n=1}^N$ are unknown in advance, here we use the hybrid likelihood ratio test (HLRT) in which the unknown parameters are estimated by using maximum-likelihood estimation for each hypothesis. Let the estimates of $\alpha, \{\lambda_n\}_{n=1}^N, \{\sigma_n^2\}_{n=1}^N$ obtained for the hypothesis H_i be denoted by $\alpha^i, \{\lambda_n^i\}_{n=1}^N, \{(\sigma_n^i)^2\}_{n=1}^N$ respectively, then the modulation classifier can be represented as

$$\hat{H} = \arg\max_{H_i} \sum_{t=1}^{T} \ln\{\frac{1}{m_i} \sum_{j=1}^{m_i} \sum_{n=1}^{N} \frac{\lambda_n^i}{\pi(\sigma_n^i)^2} \times \exp(-\frac{|y_t - \alpha^i s_{tj}^i|^2}{(\sigma_n^i)^2})\}. \tag{5}$$

III. Proposed Method for Estimating Parameters

From the description of the HLRT classifier above, we can see that the most computation-intensive part is estimating the unknown parameters of the model for each modulation type hypothesis. In [4], the parameters are estimated via ECM. ECM is simplicity and stability, but its convergence rate is slow. To make the letter self-contained, we put the equations of parameters updating by ECM here.

Here we take the hypothesis H_i as example to show the updating equations by ECM. We define a new set of variables $\{z_t\}_{t=1}^{T}$, the value of z_t indicates the term of the Gaussian mixture which models the noise component of y_t, so $z_t \in \{1, 2, \ldots, N\}$. The vector $\boldsymbol{\theta} = [\mathcal{R}(\alpha), \mathcal{I}(\alpha), \lambda_1, \ldots, \lambda_N, \sigma_1^2, \ldots, \sigma_N^2]^{\mathrm{T}}$ denotes the parameters to be estimated, and $\boldsymbol{\theta^p} = [\mathcal{R}(\alpha^p), \mathcal{I}(\alpha^p), \lambda_1^p, \ldots, \lambda_N^p, (\sigma_1^p)^2, \ldots, (\sigma_N^p)^2]^{\mathrm{T}}$ stands for the current estimate of $\boldsymbol{\theta}$, where $\mathcal{R}(\cdot)$ and $\mathcal{I}(\cdot)$ denote the real and imaginary part of the complex number respectively, and we omit the superscript "i" of parameters for clarity here. The updating equations are as follows, the details of the derivation can be found in [7]. For the sake of brevity, we use ψ_{tnj}^p to represent the posterior conditional probability of hidden variables z_t, s_t given observed variables y_t, i.e. $P(z_t = n, s_t = s_{tj}^i|y_t, \boldsymbol{\theta^p}, H_i)$.

$$\alpha^{p+1} = \frac{\sum_{t=1}^{T} \sum_{n=1}^{N} \sum_{j=1}^{m_i} \frac{y_t (s_{tj}^i)^*}{(\sigma_n^p)^2} \psi_{tnj}^p}{\sum_{t=1}^{T} \sum_{n=1}^{N} \sum_{j=1}^{m_i} \frac{|s_{tj}^i|^2}{(\sigma_n^p)^2} \psi_{tnj}^p}, \tag{6}$$

$$\lambda_n^{p+1} = \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{m_i} \psi_{tnj}^p, \tag{7}$$

$$(\sigma_n^{p+1})^2 = \frac{\sum_{t=1}^{T} \sum_{j=1}^{m_i} |y_t - \alpha^{p+1} s_{tj}^i|^2 \psi_{tnj}^p}{\sum_{t=1}^{T} \sum_{j=1}^{m_i} \psi_{tnj}^p}. \tag{8}$$

By Bayes' theorem, ψ_{tnj}^p can be calculated as

$$\psi_{tnj}^p = \frac{\lambda_n^p}{m_i} \frac{p(y_t|z_t = n, s_t = s_{tj}^i, \boldsymbol{\theta^p})}{p(y_t|\boldsymbol{\theta^p}, H_i)}, \tag{9}$$

where

$$p(y_t|z_t = n, s_t = s_{tj}^i, \boldsymbol{\theta^p}) = \frac{1}{\pi(\sigma_n^p)^2} exp(-\frac{|y_t - \alpha^p s_{tj}^i|^2}{(\sigma_n^p)^2}),$$

and

$$p(y_t|\boldsymbol{\theta^p}, H_i) = \sum_{j=1}^{m_i} \sum_{n=1}^{N} \frac{\lambda_n^p}{m_i} p(y_t|z_t = n, s_t = s_{tj}^i, \boldsymbol{\theta^p}).$$

The ECM to update $\boldsymbol{\theta^p}$ implicitly defines a mapping F from the parameter space onto itself, i.e.: $\Omega \subset \mathbb{R}^{2+2N} \mapsto \Omega$, such that $\boldsymbol{\theta^{p+1}} = F(\boldsymbol{\theta^p}), p = 0, 1, \ldots$. So the goal of ECM iterations is to find the solution of equation $g(\boldsymbol{\theta}) \triangleq F(\boldsymbol{\theta}) - \boldsymbol{\theta} = \boldsymbol{0}$. Newton's method can provide quadratic convergence speed to locate the fixed point by finding the zero of a linear approximation of $g(\boldsymbol{\theta})$ around $\boldsymbol{\theta^p}$: $g(\boldsymbol{\theta}) = g(\boldsymbol{\theta^p}) + \boldsymbol{M^p}(\boldsymbol{\theta} - \boldsymbol{\theta^p})$, where $\boldsymbol{M^p} = J(\boldsymbol{\theta^p}) - \boldsymbol{I}$ and $J(\boldsymbol{\theta^p})$ is the Jacobian of $F(\boldsymbol{\theta})$ calculated at $\boldsymbol{\theta^p}$. Due to the complexity of computing $\boldsymbol{M^p}$, Steffensen-type method [8] proposed two-point, secant-like approximation of $\boldsymbol{M^p}$. By writing the linear approximation of $g(\boldsymbol{\theta})$ around $\boldsymbol{\theta^p}$ and $F(\boldsymbol{\theta^p})$ respectively as: $g_0(\boldsymbol{\theta}) = g(\boldsymbol{\theta^p}) + (1/\delta^p)(\boldsymbol{\theta} - \boldsymbol{\theta^p})$, $g_1(\boldsymbol{\theta}) = g(F(\boldsymbol{\theta^p})) + (1/\delta^p)(\boldsymbol{\theta} - F(\boldsymbol{\theta^p}))$, here $J(\boldsymbol{\theta^p})$ and $J(F(\boldsymbol{\theta^p}))$ are both approximated by $(1/\delta^p)\boldsymbol{I} + \boldsymbol{I}$. The two zeros of $g_0(\boldsymbol{\theta})$ and $g_1(\boldsymbol{\theta})$ are as follows, and they give two approximations of the fixed point.

$$\boldsymbol{\theta_0^{p+1}} = \boldsymbol{\theta^p} - \delta^p g(\boldsymbol{\theta^p}). \tag{10}$$

$$\boldsymbol{\theta_1^{p+1}} = F(\boldsymbol{\theta^p}) - \delta^p g(F(\boldsymbol{\theta^p})). \tag{11}$$

The steplength δ^p is chosen to minimize some measure of discrepancy between $\boldsymbol{\theta_0^{p+1}}$ and $\boldsymbol{\theta_1^{p+1}}$. In this letter, we choose the discrepancy measure as $-\|\boldsymbol{\theta_1^{p+1}} - \boldsymbol{\theta_0^{p+1}}\|^2/\delta^p$, and obtain $\delta^p = -\|\boldsymbol{r^p}\|/\|\boldsymbol{v^p}\|$, where $\boldsymbol{r^p} = F(\boldsymbol{\theta^p}) - \boldsymbol{\theta^p}$, $\boldsymbol{v^p} = F(F(\boldsymbol{\theta^p})) - 2F(\boldsymbol{\theta^p}) + \boldsymbol{\theta^p}$, and $\|\cdot\|$ denotes the Euclidean norm. Combining (10) with δ^p shows Steffensen-type methods to locate the fixed point.

$$\boldsymbol{\theta^{p+1}} = \boldsymbol{\theta^p} - \delta^p(F(\boldsymbol{\theta^p}) - \boldsymbol{\theta^p}). \tag{12}$$

Expand $F(\boldsymbol{\theta})$, $F^{(2)}(\boldsymbol{\theta}) \triangleq F(F(\boldsymbol{\theta}))$, etc., in Taylor series around the fixed point $\boldsymbol{\theta^*}$ and calculate them at $\boldsymbol{\theta^p}$ to obtain:

$$F^{(j)}(\boldsymbol{\theta^p}) = \boldsymbol{\theta^*} + \boldsymbol{J}^j \boldsymbol{e^p} + o(\boldsymbol{e^p}), \quad j = 1, 2, \ldots. \tag{13}$$

where $\boldsymbol{e^p} = \boldsymbol{\theta^p} - \boldsymbol{\theta^*}$ and $\boldsymbol{J}$ is the Jacobian of $F(\boldsymbol{\theta})$ calculated at $\boldsymbol{\theta^*}$. Based on (13), the relations between $\boldsymbol{r^p}$, $\boldsymbol{v^p}$ and $\boldsymbol{e^p}$ are as follows:

$$\boldsymbol{r^p} = (\boldsymbol{J} - \boldsymbol{I})\boldsymbol{e^p} + o(\boldsymbol{e^p}), \tag{14}$$

$$\boldsymbol{v^p} = (\boldsymbol{J} - \boldsymbol{I})^2\boldsymbol{e^p} + o(\boldsymbol{e^p}). \tag{15}$$

Meanwhile, based on (12) and (13), the recursive error equation can be obtained as follows:

$$\boldsymbol{e^{p+1}} = [\boldsymbol{I} - \delta^p(\boldsymbol{J} - \boldsymbol{I})]\boldsymbol{e^p} + o(\boldsymbol{e^p}). \tag{16}$$

By using the idea called 'squaring', the SQUAREM algorithm is obtained by demanding that the recursive error equation satisfies $\boldsymbol{e^{p+1}} = [\boldsymbol{I} - \delta^p(\boldsymbol{J} - \boldsymbol{I})]^2\boldsymbol{e^p} + o(\boldsymbol{e^p})$. Using (14) and (15), the iterative scheme of SQUAREM algorithm is derived as

$$\boldsymbol{\theta^{p+1}} = \boldsymbol{\theta^p} - 2\delta^p\boldsymbol{r^p} + (\delta^p)^2\boldsymbol{v^p}. \tag{17}$$

The bSQUAREM proposed in [8] have the limitation that it does not respect constraints of parameters explicitly which will lead to meanless estimates of parameters. For example, when $(\sigma_n^p)^2 < 0$, some $exp(-|y_t - \alpha^p s_{tj}^i|^2/(\sigma_n^p)^2)$ may be too large to exceed the precision range of computer and ψ_{tnj}^p calculated by (9) may be meanless. So we should check $\boldsymbol{\theta^{p+1}}$ achieved by (17) to see whether they obey the constraints, i.e. $0 \leq \lambda_n^{p+1} \leq 1$, $(\sigma_n^{p+1})^2 > 0$. If the updated parameters violate any one of these constraints, we will reject $\boldsymbol{\theta^{p+1}}$ and set $\boldsymbol{\theta^{p+1}} = F(F(\boldsymbol{\theta^p}))$. It means that whenever feasible, we will take the steplength δ^p to accelerate the rate of convergence, otherwise, we set $\delta^p = -1$ to update the parameters like ECM, so the scheme can accelerate the rate of convergence compared to ECM and converge to the stationary point of log-likelihood function. To complete one iteration of SQUAREM-PC, we calculate one more time F given $\boldsymbol{\theta^{p+1}}$ and define $\boldsymbol{\theta^{p+1}} = F(\boldsymbol{\theta^{p+1}})$ as the updated value of $\boldsymbol{\theta^p}$. The SQUAREM-PC algorithm is summarized as Algorithm 1. The complexity of SQUAREM-PC is $O(m_i NTf)$, where f denotes the number of calculating mapping function F when the algorithm achieves convergence.

Algorithm 1 SQUAREM-PC Algorithm

Input: $\boldsymbol{\theta^0}$, ρ.
for $p = 0, 1, 2, \ldots$ **do**
  $\boldsymbol{u_1} = F(\boldsymbol{\theta^p})$.
  $\boldsymbol{u_2} = F(\boldsymbol{u_1})$.
  $\boldsymbol{r^p} = \boldsymbol{u_1} - \boldsymbol{\theta^p}$.
  $\boldsymbol{v^p} = \boldsymbol{u_2} - 2\boldsymbol{u_1} + \boldsymbol{\theta^p}$.
  $\delta^p = -\|\boldsymbol{r^p}\|/\|\boldsymbol{v^p}\|$.
  $\boldsymbol{\theta'} = \boldsymbol{\theta^p} - 2\delta^p\boldsymbol{r^p} + (\delta^p)^2\boldsymbol{v^p}$.
  if $\boldsymbol{\theta'}$ violates any constraint of parameters **then**
    $\boldsymbol{\theta'} = \boldsymbol{u_2}$.
  end if
  $\boldsymbol{\theta^{p+1}} = \mathrm{F}(\boldsymbol{\theta'})$.
  if $\|\boldsymbol{\theta^{p+1}} - \boldsymbol{\theta^p}\| < \rho$ **then**
    exit the loop.
  end if
end for

TABLE I
16-QAM AS THE TRANSMITTED MODULATION TYPE, 5000 MONTE CARLO EXPERIMENTS, N_{cnt} OF bSQUAREM AND SQUAREM-PC

N_{cnt} \ SNR(dB) / Schemes	-3	1	5	9	13
bSQUAREM	271	428	780	1270	1661
SQUAREM-PC	0	0	0	0	0

TABLE II
THE AVERAGE PERCENTAGE OF VIOLATING TIMES IN ONE WHOLE ITERATIONS OF SQUAREM-PC

SNR(dB)	-3	1	5	9	13
Percentage	0.26%	0.66%	1.70%	5.11%	9.88%

IV. NUMERICAL RESULTS

In the following experiments, the real channel model parameters are defined as follows: The amplitude of α is assumed to be Rayleigh distributed, with $E[|\alpha|^2] = 2$, and the phase of α is uniformly distributed in $(0, 2\pi]$; The number of terms in the Gaussian mixture distribution is taken to be $N = 2, 3$ respectively. When $N = 2$, we set $\lambda_1 = 0.9$ and $\sigma_2^2/\sigma_1^2 = 100$, when $N = 3$, we set $\lambda_1 = 0.85, \lambda_2 = 0.1$ and $\sigma_2^2/\sigma_1^2 = 50$, $\sigma_3^2/\sigma_1^2 = 100$; The number of observed symbols $T = 500$. The energy of modulation constellation is normalized, and the received SNR is defined as SNR $= 10\log_{10}(E[|\alpha|^2]/\sigma_1^2)$.

Because the Expectation-Maximization (EM) and generalized EM algorithms (e.g. ECM) are sensitive to parameter initializations, here we perform initializations according to the guidelines given in [9]. The convergence label ρ is set to be 10^{-6}. The maximum number of iterations of ECM is set to be 6000, and because in one iteration of SQUAREM-PC, we need to calculate mapping function F three times, the maximum number of iterations of SQUAREM-PC is set to be 2000.

Firstly, the performances of bSQUAREM and SQUAREM-PC are investigated. Convergent results are analyzed by counting the number of meanless estimates (denoted by N_{cnt}). From Table I we can find that SQUAREM-PC is superior to bSQUAREM. Further, we research how many times on average parameters violate the constraints in one whole iterations of SQUAREM-PC. The average percentages of violating times in one whole iterations are shown in Table II. So we need the parameters checking step in each iteration to ensure that SQUAREM algorithm will converge to the stationary point of log-likelihood function.

Secondly, we compare the convergence rate of SQUAREM-PC with ECM. 16-QAM is chosen as the transmitted modulation type. The average number of calculating mapping function F (denoted by F_{cnt}) is taken into account as the performance index. It is seen in Fig.1 that convergence rate of SQUAREM-PC is faster than ECM. Complexity of the two algorithms are linear with F_{cnt}, so the complexity of SQUAREM-PC is lower than that of ECM.

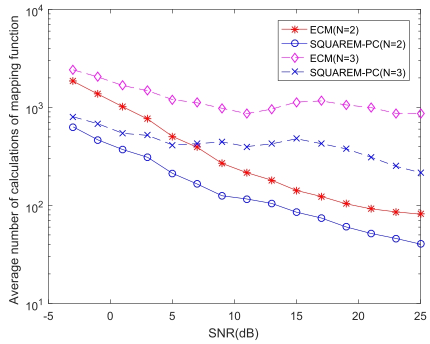

Fig. 1. F_{cnt} for SQUAREM-PC and ECM when $N = 2, 3$.

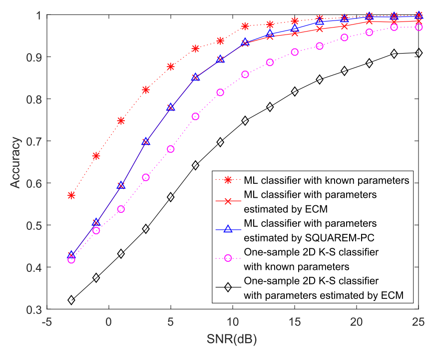

Fig. 2. The accuracy of classifier versus SNR when $N = 2$.

Thirdly, accuracy and speed of classifier based on SQUAREM-PC or ECM are researched. The modulation schemes considered are BPSK, QPSK, 8-PSK and 16-QAM, they are assumed to be equally likely. The accuracy of classifier is evaluated by using the average probability of correct classification P_{cc} defined by $P_{cc} = 1/C \sum_{i=1}^{C} P(\hat{H} = H_i|H_i)$, and H_i is the modulation type of transmitted signals. The accuracy of classifier based on SQUAREM-PC or ECM are presented in Fig.2 and Fig.3 when $N = 2, 3$ respectively. Meanwhile, the P_{cc} of maximum-likelihood (ML) classifier with known parameters, one-sample quadrature-based 2D K-S classifier [4] with known parameters and one-sample quadrature-based 2D K-S classifier with parameters estimated by ECM are researched. The speed of classifier is evaluated by using the average time of modulation classification t_{mc} defined by $t_{mc} = 1/C \sum_{i=1}^{C} t_{mc}^{i}$, where t_{mc}^{i} denotes the time of classifier when the modulation type of transmitted signals is H_i. t_{mc} of classifier based on SQUAREM-PC or ECM are listed in Table III. The simulation experiments are conducted by Matlab running on computer with Intel(R) Xeon(R) Platinum 8163 CPU with 2.50GHz CPU Clock Speed. So classifier based on SQUAREM-PC accelerates the speed of that based on ECM without decreasing the accuracy. From Fig 2 and Fig 3 we can find that accuracy of classifier based on SQUAREM-PC is even slightly higher than that based on ECM when SNR is high. The explanation for this phenomenon may be that SQUAREM-PC is more likely to converge to global maximum point of log-likelihood function than ECM using the same initializations when SNR is high. The theory for explaining it needs to be studied in the future.

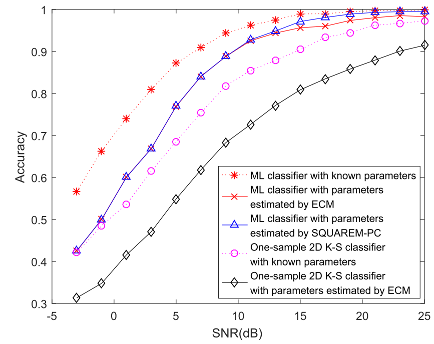

Fig. 3. The accuracy of classifier versus SNR when $N = 3$.

TABLE III
AVERAGE TIME OF CLASSIFIER VERSUS SNR

time(s) SNR(dB) / Schemes	-3	1	5	9	13	17
ECM(N=2)	10.72	6.87	3.93	2.00	1.15	0.67
SQUAREM-PC(N=2)	3.83	2.33	1.35	0.70	0.46	0.30
ECM(N=3)	15.92	11.52	8.35	6.15	5.02	5.42
SQUAREM-PC(N=3)	5.65	3.73	2.79	2.17	1.80	1.78

V. CONCLUSION

From experiments above, we can conclude that the convergence rate of SQUAREM-PC is faster than that of ECM when estimating the channel parameters blindly in flat fading non-Gaussian channels, and the AMC based on SQUAREM-PC is faster than AMC based on ECM without losing the accuracy of classification.

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 Commun.*, vol. 1, no. 2, pp. 137–156, Apr. 2007.

[2] Z. Zhu and A. K. Nandi, "Blind digital modulation classification using minimum distance centroid estimator and non-parametric likelihood function," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4483–4494, Aug. 2014.

[3] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Cyclostationarity-based modulation classification of linear digital modulations in flat fading channels," *Wireless Pers. Commun., Int. J.*, vol. 54, no. 4, pp. 699–717, Sep. 2010.

[4] F. Wang and X. Wang, "Fast and robust modulation classification via Kolmogorov–Smirnov test," *IEEE Trans. Commun.*, vol. 58, no. 8, pp. 2324–2332, Aug. 2010.

[5] F. Wang, O. A. Dobre, C. Chan, and J. Zhang, "Fold-based Kolmogorov–Smirnov modulation classifier," *IEEE Signal Process. Lett.*, vol. 23, no. 7, pp. 1003–1007, Jul. 2016.

[6] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, vol. 2. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[7] V. G. Chavali and C. R. C. M. D. Silva, "Maximum-likelihood classification of digital amplitude-phase modulated signals in flat fading non-Gaussian channels," *IEEE Trans. Commun.*, vol. 59, no. 8, pp. 2051–2056, Aug. 2011.

[8] R. Varadhan and C. Roland, "Simple and globally convergent methods for accelerating the convergence of any EM algorithm," *Scand. J. Statist.*, vol. 35, no. 2, pp. 335–353, 2008.

[9] R. J. Kozick and B. M. Sadler, "Maximum-likelihood array processing in non-Gaussian noise with Gaussian mixtures," *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3520–3535, Dec. 2000.