Robust Blind Equalization for NB-IoT Driven by QAM Signals

Jin Li, Wei Xing Zheng, *Fellow, IEEE*, Mingqian Liu, *Member, IEEE*, Yunfei Chen, *Senior Member, IEEE*, Nan Zhao, *Senior Member, IEEE*,

***Abstract*—The expansion of data coverage and the accuracy of decoding of the narrowband-internet of things (NB-IOT) mainly depend on the quality of channel equalizers. Without using training sequences, blind equalization is an effective method to overcome adverse effects in the internet of things (IoT). The constant modulus algorithm (CMA) has become a favorite blind equalization algorithm due to its least mean square (LMS)-like complexity and desirable robustness property. However, the transmission of high-order quadrature amplitude modulation (QAM) signals in the IoT can degrade its performance and the convergence speed. This paper investigates a family of modified constant modulus algorithms for blind equalization of IoT using high-order QAM. Our theoretical analysis for the first time illustrates that the classical CMA has the problem of artificial error using high-order QAM signals. In order to effectively deal with these issues, a modified constant modulus algorithm (MCMA) is proposed to decrease the modulus matched error, which can efficiently suppress the artificial error and misadjustment at the expense of reduced sample usage rate. Moreover, a generalized form of the MCMA (GMCMA) is developed to improve the sample usage rate and guarantee the desirable equalization performance. Two modified Newton methods (MNMs) for the proposed MCMA and GMCMA are constructed to obtain the optimal equalizer. Theoretical proofs are presented to show the fast convergence speed of the two MNMs. Numerical results show that our methods outperform other methods in terms of equalization performance and convergence speed.**

***Index Terms*—Artificial error, steady-state misadjustment, modified constant modulus algorithm, generalized modified constant modulus algorithm, modulus matched error.**

I. Introduction

NARROWBAND internet of things (NB-IoT) is a sustainable technology for connecting billions of devices from a great range of utilities, logistics to the industrial applications, playing an important role in future IoT business [1]. Since NB-IoT devices often transmit very short message blocks without strong error correction coding or pilots for channel equalization. The transmitted signals encounter a severe multipath channel, the receiver is often unable to equalize the resulting inter-symbol interference (ISI) via traditional methods, leading to a high retransmission rate [2]. The accuracy of channel equalization will significantly affect the performance of the NB-IoT system [3], [4]. Therefore, it is necessary to study efficient channel equalization methods for the NB-IoT.

The trained equalization approach uses the repeated transmission of a pseudo-random pattern of bits (training sequence) known to both the transmitting and receiving ends [5].Various algorithms exist to perform trained equalization. The least mean squares (LMS) [6] and minimum mean squared error (MMSE) [7] equalizations are the most representative ones among them. In the LMS equalization, the receiver computes the error between the output of the equalizer and the training data, namely the LMS error, and then the equalizer updates its taps by moving them in the direction that, on average, reduces the LMS error. The MMSE equalization adjusts the taps of an equalizer to minimize the average error between the output of the equalizer and the training data. Although a training sequence provides the equalizer with a helpful reference, it consumes valuable bandwidth. Even worse, a training sequence is often inadequate and sometimes even infeasible in certain communication systems [8].

A blind equalizer (BE) performs channel equalization without a training sequence. Without using training symbols, no bandwidth is consumed by its transmission. More importantly, in the uncooperative or point-to-multipoint communication scenarios, a blind algorithm is the only feasible solution to achieve system equalization [9], [10]. In the past, many blind equalization algorithms [11]–[14], [16]–[20] have been proposed since [21] in 1975.

Up to now, the most popular blind equalization approaches for two-dimensional (2D) modulation schemes, such as quadrature amplitude modulation (QAM) and carrier-less amplitude and phase (CAP) modulation, are the constant modulus algorithm (CMA) [11] and its variants [22]–[24]. On the one hand, the CMA, whose cost function attempts to minimize the difference between the outputs' squared magnitude and the Godard dispersion constant, has less local minima and reliable convergence [25], [26]. On the other hand, the CMA changes the value of its taps in time and has an LMS-like complexity for ease of implementation [21]. Moreover, with the above-mentioned characterization, the CMA can explicitly or implicitly provide a good initial state for two-stage blind algorithms [18]–[20] or dual mode blind algorithms [16], [17],

This work was supported by the National Natural Science Foundation of China under Grant 62231027 and 62071364, in part by the Natural Science Basic Research Program of Shaanxi under Grant 2024JC-JCQN-63, in part by the Key Research and Development Program of Shaanxi under Grant 2023-YBGY-249, and in part by the Guangxi Key Research and Development Program under Grant 2022AB46002. *(Corresponding author: Mingqian Liu.)*

J. Li and M. Liu are with the State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China (e-mail: lijin342@163.com; mqliu@mail.xidian.edu.cn).

Wei Xing Zheng is with the School of Computer, Data and Mathematical Sciences, Western Sydney University, Sydney, NSW 2751, Australia (e-mail: w.zheng@westernsydney.edu.au).

Y. Chen is with Department of Engineering, University of Durham, Durham, UK DH1 3LE (e-mail: Yunfei.Chen@durham.ac.uk).

N. Zhao is with the School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, China (e-mail: zhaonan@dlut.edu.cn).

allowing them to obtain better performances. Although the CMA and its various variants [27]–[29] have so many preferable properties, some shortcomings limit their applications.

Firstly, the primary concern of blind equalization is its performance. The CMA cost function only exploits a part of amplitude information of high-order QAM signals. It means that some knowledge about the signal constellations is discarded, leading to the relatively poor performance [30], [31], especially for the high-order QAM signals with non-constant modulus. Moreover, the CMA requires an additional procedure to restore the revolved phase. To overcome this shortcoming, many researchers developed the modified CMA (MCMA) [14], [32]–[35], the dual-mode schemes (DMSs) [16], [17], [36], [37], and the two-stage schemes [18]–[20]. The MCMA mainly uses three ways to improve the equalization performance. The first one, such as the multi-modulus algorithm (MMA) [14], [15], uses the information of the imaginary part and the real part simultaneously to overcome the problem of phase rotation. The second one, such as the constant norm algorithms (CNAs) [32], creates new norms by combining several existing norms in order to benefit from the advantages of each original norm, and then this kind of methods achieves better mean square error (MSE) performance than the CMA. The third kind of MCMA, such as [34] and [35], directly or indirectly revises the equalizer output error with nonlinear transformation, and then the maladjustment can be relieved to a certain extent. However, the existing MCMAs cannot completely remove the maladjustment because of the inherent properties of its cost function. The DMSs add a term of constellation matched error (CME) to the constant modulus loss function (CMLF). The CME term can improve the equalization performance and the CMLF guarantees the reliable convergence. For example, the additive CME term was designed in [16] to have the sinusoidal form. Although such kind of schemes can achieve the desirable error level, the computational complexity increases significantly. The two-stage schemes perform prefiltering of the received signal based on stable convergent algorithms, and then implements constellation matching algorithms. In the first stage, the ISI is mitigated and a good initial value for the constellation matching algorithm is provided. In the second stage, the constellation matching algorithm further relieves the ISI and a desirable equalization performance is achieved. Taking the work [19] for instance, the joint generalized multilevel modulus algorithm and modified soft decision-directed (SDD) equalization were applied at the first convergence stage. When the convergence process reached the steady state, the equalizer changed the first equalization stage to the second stage. At the second stage, the modified SDD scheme reduced the MSE further. However, this kind of schemes cannot provide the attainable switching threshold. To guarantee the stable convergence, the scheme may switch to the second stage later, but then it converges slowly and consumes unnecessary computational cost.

Secondly, the convergence of the equalizer using a blind tap updating algorithm is concerning. Although the CMA is noted for its LMS-like complexity, it exhibits slow convergence speed. As in the classical LMS theory, the selection of step size becomes a tradeoff between the convergence rate and MSE [30]. Even worse, in order to reduce the well-known steady-state misadjustment and avoid the initial instability, the step size of the CMA, which is usually set at 10^{-5} orders of magnitude or less, is much smaller than that of its LMS counterpart, which is usually set at 10^{-2} orders of magnitude [38]. Hence, the CMA has a much slower convergence speed than other LMS-type algorithms. In contrast, it is widely known that the Newton method possesses a fast convergence speed. However, the Newton method requires the computation of the Hessian matrix of the cost function for its implementation. It should be noted that we are working with complex signals and, as a consequence, the Hessian matrix of the constant modulus loss function is always singular [39], which means that the fast converging Newton method can hardly be used in practice without modification. Taking all these into accounts, it is essential to improve the equation accuracy and convergence speed of the CMA.

This work derives a new family of modified constant modulus algorithms (MCMAs) for blind equalization of high-order QAM systems. The main contributions are listed as follows. (i) By selecting the samples corresponding to the symbols with the same modulus, a new MCMA is proposed, which can completely remove the misadjustment and the artificial error at the expense of reduced sample usage rate. (ii) Based on the MCMA, a new GMCMA is developed to improve the sample usage rate and meantime preserve the desirable equalization performance of the MCMA. (iii) According to the optimization algorithm proposed in [40], the MNMs associated with the new MCMA and GMCMA are constructed to fast find the optimal equalizer. (iv) We prove that the classical CMA causes the artificial error (excess error) and steady-state misadjustment due to the use of a single modulus for high-order QAM signals. We prove that the proposed MNMs have fast convergence speed and can converge to the optimal equalizer $\hat{\mathbf{w}}$ at the step size of $\frac{1}{2}\|\mathbf{w}_k - \hat{\mathbf{w}}\|_2$. (v) We illustrate that the MCMA can effectively suppress the artificial error and misadjustment by removing the modulus matched error. Second, a theoretical analysis is made to show that the proposed MNMs have much less computational load than other Newton methods due to its fixed Hessian matrix.

The remainder of this article is outlined as follows. Section II introduces the system model for blind equalization. In Section III, we discuss the MCMA and its fast optimization for blind equalization of high-order QAM systems, and we also describe the GMCMA and its fast solution for blind equalization. Following that, we analyze the computational complexity of the benchmark schemes and the proposed algorithms in Section IV. Next, we present the performance evaluation results of the MCMA and GMCMA through extensive experiments in Section V, followed by conclusions given in the final section.

Throughout the paper, boldface capitals and lower-case letters stand for matrices and vectors, respectively. Given a matrix $\mathbf{A}$, symbols $\mathbf{A}^*$, $\mathbf{A}^T$, $\mathbf{A}^H$ and $\mathbf{A}^{-1}$ denote the complex conjugation, the transpose, the Hermitian transpose and the inverse, respectively. Moreover, $|\cdot|$ represents the magnitude, $E[\cdot]$ the expectation operator, $\lfloor\cdot\rfloor$ the round down operator, $\otimes$ the discrete convolution, and $\mathbb{C}$ the complex number set. Furthermore, $\mathbf{e}(\tau)$ indicates a column vector whose $(\tau+1)$th

element is 1 and all other elements are equal to 0. Finally, $j = \sqrt{-1}$.

II. System Description

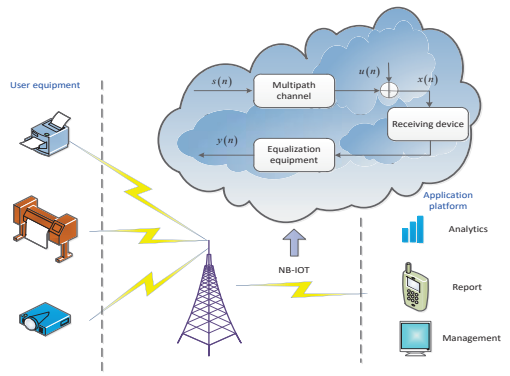

Fig. 1. The model of NB-IoT and its communication system.

The model of NB-IOT system is shown in Fig. 1. In the system, the user equipment collects the concerned data, then transmits the information to the base station. Finally, the information is transmitted to the application platform to manage equipment. As seen in the Fig. 1, the transmitted signals pass through wireless channel and experience multipath fading. To abstain accuracy information, the channel equalization must be implemented. Assuming the transmitted signal $s(n)$ is independently and identically distributed (i.i.d.) and it takes values from high-order QAM symbols with equal probability. The channel impulse response vector with length $\bar{L}$ is $\mathbf{h} = \left[h\left(0\right), h\left(1\right), \cdots, h\left(\bar{L}-1\right)\right]^T$ is, and $h(n)$ are the channel coefficients. Let $x(n)$ be the receive data as

$$\begin{aligned} x(n) &= h(n) \otimes s(n) + u(n) \\ &= \sum_{l=0}^{\bar{L}-1} h(l)s(n-l) + u(n) \\ &= \mathbf{h}^T \mathbf{s}(n) + u(n) \end{aligned} \tag{1}$$

where $\mathbf{s}(n) = \left[s(n), s(n-1), \cdots, s(n-\bar{L}+1\right]^T$ is the signal sequence, and $u(n)$ is the complex-valued Gaussian white noise with variance $E[|u(n)|^2] = \sigma_u^2$ and mean zero. Then the output of the equalizer is described by

$$\begin{aligned} y(n) &= \sum_{l=0}^{L-1} w^*(l)x(n-l) = \mathbf{w}^H \mathbf{x}(n) \\ &= \mathbf{w}^H \left(\mathbf{H}\bar{\mathbf{s}}(n) + \mathbf{u}(n)\right) \\ &= \bar{\mathbf{w}}^T \bar{\mathbf{s}}(n) + \mathbf{w}^H \mathbf{u}(n) \end{aligned} \tag{2}$$

where $\mathbf{w} = [w(0), w(1), \cdots, w(L-1)]^T$ is the equalizer, L is the number of taps in the equalizer, and $\mathbf{x}(n) = [x(n), x(n-1), \cdots, x(n-L+1)]^T$ is the regression vector of channel observations. Moreover, we have $\bar{\mathbf{s}}(n) = \left[s(n), s(n-1), \cdots, s(n-\bar{L}-L+2)\right]^T$, $\mathbf{u}(n) = [u(n), u(n-1), \cdots, u(n-L+1)]^T$, and $\bar{\mathbf{w}} = \mathbf{H}^T\mathbf{w}^* \in \mathbb{C}^{(L+\bar{L}-1)\times 1}$, where $\bar{\mathbf{w}}$ is the combined impulse response of the channel and the equalizer, and $\mathbf{H}$ is the channel matrix defined as

$$\mathbf{H} = \begin{bmatrix} h(0) & h(1) & \cdots & h(\bar{L}-1) & 0 & \cdots & 0 \\ 0 & h(0) & h(1) & \cdots & h(\bar{L}-1) & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & h(0) & h(1) & \cdots & h(\bar{L}-1) \end{bmatrix}. \tag{3}$$

A blind equalization algorithm is employed to adjust the tap weights of the equalizer $\mathbf{w}$ depending on the equalizer output $y(n)$. Therefore, $y(n)$ provides an estimate of the source signal $s(n)$ with some inherent indeterminacies, i.e., $Cy(n) \approx s(n-\tau)$, where C is a constant and τ the time delay resulted by the inherent ambiguity from blind signal processing.

III. Modified Constant Modulus Algorithm

A. Constant Modulus Algorithm

The CMA is perhaps the most popular scheme among all the blind channel equalization algorithms [11]. The CMA tries to solve the following optimization problem

$$\min_{\mathbf{w}} J(\mathbf{w}) = E\left[\left(|y(n)|^p - R\right)^2\right] \tag{4}$$

where $R = E\left[|s(n)|^{2p}\right] / E\left[|s(n)|^p\right]$ and p is a positive integer. If the implementation method is realized by gradient descent based on an adaptive scheme, then the equalizer taps are updated according to the following equation

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu E\left[\left(|y(k)|^p - R\right)|y(k)|^{p-2}y^*(k)\mathbf{x}(k)\right] \tag{5}$$

where μ is the step size governing the speed of convergence and the level of steady-state equalizer performance. For simplicity, the expectation of the gradient is replaced by its instantaneous value [41]. Hence, the adaptation equation (5) is recast as

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu\left(|y(k)|^p - R\right)|y(k)|^{p-2}y^*(k)\mathbf{x}(k). \tag{6}$$

The cost function $J(\mathbf{w})$ is an expression for implicitly embedded higher-order statistics of the equalizer output $y(n)$. Ideally, the minimization of $J(\mathbf{w})$ aligns the statistics of $y(n)$ with the transmitted signals. The equalization is accomplished when the equalized sequence $y(n)$ acquires an identical distribution of the channel input $s(n)$ [5]. However, in practice, the statistics of $y(n)$ is estimated by using sample data, while the statistics of $s(n)$ is provided by the theoretical value. This inconsistency results in an error called the artificial error. Evidently, the artificial error can cause a slight performance loss. In addition, the stochastic gradient-based adaptive approach contains large misadjustments under high-order QAM signals environment. We have the following proposition.

Proposition 1: In the high-order QAM systems, if the BE converges to the optimal solution $\hat{\mathbf{w}}$ and completely compensates the channel distortion, i.e.,

$$\hat{\bar{\mathbf{w}}} = \mathbf{H}^T\hat{\mathbf{w}}^* = \mathbf{e}(\tau)$$

then the instantaneous gradient $\nabla J(\mathbf{w})|_{\mathbf{w}=\hat{\mathbf{w}}}$, which is calculated by sample data, is unequal to $\mathbf{0}$ under the noise-free

environment. Moreover, the instantaneous gradient satisfies the inequality

$$\left(|y(k)|^p - R\right)|y(k)|^{p-2}y^*(k)\mathbf{x}(k)|_{\mathbf{w}=\hat{\mathbf{w}}} \neq 0.$$

Proof: See Appendix A. ■

It is seen that $\nabla J(\mathbf{w})|_{\mathbf{w}=\hat{\mathbf{w}}} \neq \mathbf{0}$ results from the artificial error as shown in the proof of Proposition 1. Because the gradient is unequal to $\mathbf{0}$, the BE $\mathbf{w}$ will be continuously adjusted even when it achieves the optimal solution $\hat{\mathbf{w}}$, which then produces an extra error (artificial error). Moreover, the instantaneous gradient $\left(|y(k)|^p - R\right)|y(k)|^{p-2}y^*(k)\mathbf{x}(k)|_{\mathbf{w}=\hat{\mathbf{w}}} \neq 0$ implies that the CMA will continue to adjust the BE when the BE converges to the optimal solution in its implementation. This is called the misadjustment, which causes fluctuation of the CMA in the steady state. To overcome these issues, we propose a new MCMA in the next subsection.

B. Modified Constant Modulus Algorithm

1) Cost function of the MCMA: From the Proposition 1, the artificial error and misadjustment of the CMA are caused by the fact that the amplitudes of the high-order QAM signals have several different constants and we use the statistical value to replace the real value. If the transmitted signals have the same amplitudes, such as the low-order 4-QAM signals, then these issues can be avoided. Thus, we try to improve the equalization performance of the CMA by transforming the high-order signal input into a constant modulus signal input. Fortunately, it is generally believed that an equalizer is dependent on the channel but independent of the input signal. Therefore, if we can identify the received data corresponding to a specific modulus from all the received data and discard the other data, then the input signal can be regarded as a constant modulus signal and the equalization performance can be improved. In this subsection, a new modified constant modulus algorithm is designed for high-order QAM signals.

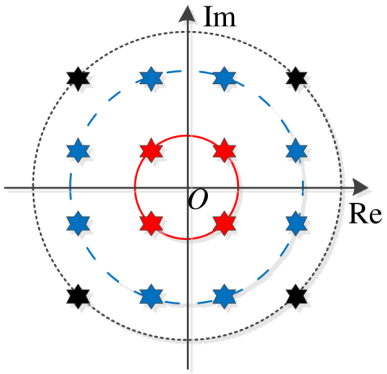

Fig. 2. Schematic diagram of subset division for 16-QAM constellation.

High-order QAM signals can be divided into different subsets according to their moduli. As shown in Fig. 2, 16-QAM constellation points can be divided into three subsets which are marked as red, blue and black colors, respectively. The constellation point sets $\{\pm 1 \pm 1j\}$, $\{\pm 1 \pm 3j, \pm 3 \pm 1j\}$ and $\{\pm 3 \pm 3j\}$ correspond to the moduli of $\sqrt{2}$, $\sqrt{10}$ and $3\sqrt{2}$, respectively. The set composed of the possible moduli of the signals is defined as $\Omega = \{R_i\}\,(i = 1, 2, \cdots, I)$. Because the steady-state outputs $(\hat{y}(n))$ of the equalizer are the estimates of the transmitted signals, these outputs can as well be divided into different subsets according to their moduli, i.e.,

$$\bar{\Omega}_i = \{\hat{y}(n) \mid |\hat{y}(n)| = R_i\}. \tag{7}$$

Following the classification of the outputs $\hat{y}(n)$, the corresponding regression vector of the channel observations $\mathbf{x}(n)$ can also divided into different subsets

$$\Omega_i = \left\{\mathbf{x}(n) \mid \left|\hat{\mathbf{w}}^H\mathbf{x}(n)\right| = R_i\right\} \tag{8}$$

where $\hat{\mathbf{w}}$ is the ideal equalizer. It is easy to validate that the outputs belonging to the set $\bar{\Omega}_i$ satisfy the constant modulus property in the strict sense. Therefore, if we only use the outputs $\hat{y}(n)$ that correspond to the channel observations $\mathbf{x}(n)$ belonging to the set Ω_i to adjust the equalizer, then the input signals can be considered as constant modulus signals and the artificial error and misadjustment can be completely avoided.

Thus, the CMA cost function can be modified as

$$\begin{aligned} &\min_{\mathbf{w}} J_{\mathrm{MCMA}}(\mathbf{w}) = E\left[\left(\left|\mathbf{w}^H\mathbf{x}(n)\right| - R_i\right)^2\right]. \\ &s.t. \quad \mathbf{x}(n) \in \Omega_i \end{aligned} \tag{9}$$

It is noteworthy that the set $\Omega_i = \{\mathbf{x}_i(1), \mathbf{x}_i(2), \cdots, \mathbf{x}_i(N_i)\}$ for any $i = 1, 2, \cdots, I$ is unknown beforehand, where N_i denotes the cardinality of the set Ω_i. Therefore, we need to find it out from the sample matrix $\mathbf{X} = [\mathbf{x}(1), \mathbf{x}(2), \cdots, \mathbf{x}(n)]$ at first. In order not to affect the readability of the paper, the determination of the set Ω_i is presented in Subsection III-B3.

In the following, to express in a more convenient way, we symbolize $\mathbf{x}(n) \in \Omega_i$ as $\mathbf{x}_i(n)$ and let $y_i(n) = \mathbf{w}^H\mathbf{x}_i(n)$. Then, according to the statistical gradient algorithm, the corresponding updating formula of the proposed MCMA is given by

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu\left(|y_i(k)| - R_i\right)|y_i(k)|^{-1}y_i^*(k)\mathbf{x}_i(k). \tag{10}$$

Remark 1: Comparing the updating formula (10) with (6), the differences between the MCMA and the CMA include: (i) The MCMA just uses the samples $\mathbf{x}_i(n)\,(n = 1, 2, \cdots, N_i)$, where $\mathbf{x}_i(n)$ is the sample belonging to the set Ω_i and N_i is its size; (ii) The parameter R in (6) is replaced by R_i in (10); (iii) The parameter p is usually set to be 2 for the CMA, whilst the MCMA selects the parameter p as 1. These differences bring the following three merits. Firstly, the steady-state misadjustment is avoided by adopting the MCMA. It is due to the fact that $\left|\hat{\mathbf{w}}^H\mathbf{x}_i(n)\right| - R_i = R_i - R_i = 0$ (where $R_i = E\left[|s_i(k)|^2\right]/E\left[|s_i(k)|\right] = \sum_k |s_i(k-\tau)|^2 / \sum_k |s_i(k-\tau)|$) when the BE converges to the optimal solution. Secondly, the MCMA can eliminate the artificial error due to $\nabla J_{\mathrm{MCMA}}(\mathbf{w})|_{\mathbf{w}=\hat{\mathbf{w}}} = \mathbf{0}$. Thirdly, the MCMA has a typical quadratic structure, which facilitates designing a fast convergence algorithm to find the optimal BE.

2) Modified Newton method for MCMA: Although the MCMA (9) can be optimized by a gradient descent approach like other LMS methods, it converges much slowly. Fortunately, since the constructed cost function $\left(\left|\mathbf{w}^H\mathbf{x}(n)\right| - R_i\right)^2$ has a

typical quadratic structure, we can easily design a Newton-type method which has the quadratic order or asymptotically quadratic order convergence speed to search for the optimal BE.

In order to construct the MNM associated with the MCMA conveniently, the statistical average in (9) is replaced with the time average and $\mathbf{x}(n) \in \Omega_i$ is replaced by $\mathbf{x}_i(n)$. Then the MCMA cost function can be rewritten as

$$J_{\mathrm{MCMA}}(\mathbf{w}) = \sum_{n=1}^{N_i} \left(\left|\mathbf{w}^H \mathbf{x}_i(n)\right| - R_i\right)^2 \tag{11}$$

where N is the length of available samples which can be utilized to search the optimal BE. Moreover, it is reasonable to set N_i to be $\left\lfloor \frac{P_i}{Q} \times N \right\rfloor$, where P_i is the number of the constellation points with modulus R_i, Q is the order of the QAM signal, and P_i/Q is the prior probability of the transmitted signals with modulus R_i. Taking 16-QAM signals for instance, if $R_i = \sqrt{10}$, then $P_i/Q = 8/16 = 1/2$.

Now, differentiating $J_{\mathrm{MCMA}}(\mathbf{w})$ with respect to $\mathbf{w}$, we obtain the following gradient expression

$$\nabla J_{\mathrm{MCMA}}(\mathbf{w}) = \sum_{n=1}^{N_i} \mathbf{x}_i(n)\mathbf{x}_i^H(n)\mathbf{w} - R_i \frac{y_i^*(n)}{|y_i(n)|}\mathbf{x}_i(n). \tag{12}$$

Let the sample matrix be $\mathbf{X}_i = [\mathbf{x}_i(1), \mathbf{x}_i(2), \cdots, \mathbf{x}_i(N_i)]$ and the normalized output vector be $\mathbf{y}_i = \left[\frac{y_i^*(1)}{|y_i(1)|}, \frac{y_i^*(2)}{|y_i(2)|}, \cdots, \frac{y_i^*(N_i)}{|y_i(N_i)|}\right]^T$. Then the gradient $\nabla J_{\mathrm{MCMA}}(\mathbf{w})$ is simplified as

$$\nabla J_{\mathrm{MCMA}}(\mathbf{w}) = \mathbf{X}_i\mathbf{X}_i^H\mathbf{w} - R_i\mathbf{X}_i\mathbf{y}_i. \tag{13}$$

It can be seen that the gradient $\nabla J_{\mathrm{MCMA}}(\mathbf{w})$ can be further decomposed into $\mathbf{A}(\mathbf{w})\mathbf{w} - \mathbf{b}(\mathbf{w})$, which is required by the MNM proposed in [40]. Here, the matrix $\mathbf{A}(\mathbf{w})$, which is a positive definite matrix, can be considered as $\mathbf{X}_i\mathbf{X}_i^H$, and the vector $\mathbf{b}(\mathbf{w})$ can be referred to as $R_i\mathbf{X}_i\mathbf{y}_i$. According to the MNM [40], $\mathbf{w}_{k+1} = \mathbf{A}^{-1}(\mathbf{w}_k)\,\mathbf{b}(\mathbf{w}_k)$. Hence, the updating formula of the MCMA, which is named the MNM based MCMA (MCMA-MNM), is expressed as

$$\mathbf{w}_{k+1} = R_i\left(\mathbf{X}_i\mathbf{X}_i^H\right)^{-1}\mathbf{X}_i\mathbf{y}_{i,k} = R_i\mathbf{R}_i^{-1}\mathbf{X}_i\mathbf{y}_{i,k} \tag{14}$$

where $y_{i,k}(n) = \mathbf{w}^H\mathbf{x}_i(n)|_{\mathbf{w}=\mathbf{w}_k}$, $\mathbf{y}_{i,k} = \left[\frac{y_{i,k}^*(1)}{|y_{i,k}(1)|}, \frac{y_{i,k}^*(2)}{|y_{i,k}(2)|}, \cdots, \frac{y_{i,k}^*(N_i)}{|y_{i,k}(N_i)|}\right]^T$, and $\mathbf{R}_i = \mathbf{X}_i\mathbf{X}_i^H$. Moreover, a equalizer is usually initialized with central single-spike.

Remark 2: It is widely known that Newton methods may not stable due to their indefinite or approximately singular Hessian matrix. In addition, Newton methods usually have high computational complexity to calculate the inverse of Hessian matrix at each iteration. In contrast, the MCMA-MNM adopts the positive definite modified Hessian matrix $\mathbf{X}_i\mathbf{X}_i^H$, making it stable. The matrix $\mathbf{X}_i$ should be unchanged with different $\mathbf{w}_k$ in theory, since it is predetermined that which signal constellation can be recovered by a distinct sample. Thus, the term $\mathbf{R}_i^{-1} = \left(\mathbf{X}_i\mathbf{X}_i^H\right)^{-1}$ can be calculated in advance. The MCMA-MNM mainly requires computing $\mathbf{y}_{k,i}$ and implementing an operation $R_i\mathbf{R}_i\mathbf{X}_i\mathbf{y}_{k,i}$ at each iteration step as long as $\mathbf{R}_i$ is obtained. This greatly reduces the computational cost of the proposed algorithm.

Also, the proposed MCMA-MNM is fast convergent, as shown in the following proposition.

Proposition 2: If $\hat{\mathbf{w}}$ is the optimal solution of (11) and $\mathbf{w}_k \approx \hat{\mathbf{w}}$, i.e.,

$$\mathbf{w}_k \in \zeta(\hat{\mathbf{w}}, \delta) = \{\mathbf{w} \mid \|\mathbf{w} - \hat{\mathbf{w}}\|_2 \le \delta\}$$

where δ is a small positive constant, then the sequence $\mathbf{w_k}$ in the iteration formula (14) converges to the optimal solution $\hat{\mathbf{w}}$ at the step size of $\frac{1}{2}\|\mathbf{w}_k - \hat{\mathbf{w}}\|_2$.

Proof: See Appendix B. ■

3) Determination of the set Ω_i: It is noteworthy that the set $\Omega_i = \{x_i(1), x_i(2), \cdots, x_i(N_i)\}$ is unknown beforehand. Therefore, we need to find it out from the sample matrix $\mathbf{X} = [\mathbf{x}(1), \mathbf{x}(2), \cdots, \mathbf{x}(n)]$ first. Now, we develop a sample selecting method based on the following theorem.

Theorem 1: If the BE converges to the optimal solution, i.e., $\mathbf{w} = \hat{\mathbf{w}}$, such that there is the mathematical relationship $|y_i(n)| = \left|\hat{\mathbf{w}}^H\mathbf{x}_i(n)\right| \approx R_i$, then the following inequality holds:

$$\left|\left|\hat{\mathbf{w}}^H\mathbf{x}_i(n)\right| - R_i\right| < \left|\left|\hat{\mathbf{w}}^H\mathbf{x}_j(n)\right| - R_i\right| \tag{15}$$

for all $\mathbf{x}_i(n) \in \Omega_i$ and $\mathbf{x}_j(n) \notin \Omega_i$.

One has $\left|\hat{\mathbf{w}}^H\mathbf{x}_i(n)\right| - R_i = 0$ in the ideal case according to the definition of the set Ω_i. Moreover, $\left|\hat{\mathbf{w}}^H\mathbf{x}_j(n)\right| \neq R_i$ and $\left|\left|\hat{\mathbf{w}}^H\mathbf{x}_j(n)\right| - R_i\right|$ is much larger than 0. Hence, the inequality (15) holds.

According to the inequality (15), we sort out the equalizer output errors $\left|\left|\hat{\mathbf{w}}^H\mathbf{x}(n)\right| - R_i\right| = \hat{e}(n)\,(n = 1, 2, \cdots, N)$ in the ascending order. Then it seems safe to assume that the samples corresponding to the first N_i small errors $\hat{e}(n)\,(n = 1, 2, \cdots, N_i)$ are the selected samples which are thought to make up the set Ω_i. However, the optimal equalizer $\hat{\mathbf{w}}$ is unknown before achieving the channel equalization. To overcome this issue, the kth iteration value $\mathbf{w_k}$ is used to take the place of $\hat{\mathbf{w}}$, and then the samples corresponding to the first N_i small errors $e_k(n) = \left|\left|\mathbf{w}_k^H\mathbf{x}(n)\right| - R_i\right|\,(n = 1, 2, \cdots, N)$ are considered as the selected samples required.

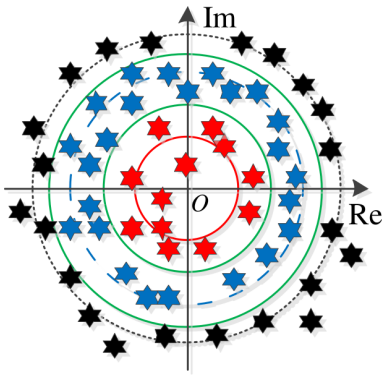

Fig. 3. Constellation of the output of the kth iteration for 16-QAM constellation.

Taking the 16-QAM signal into consideration, Fig. 3 shows the outputs of the kth iteration $y_k(n) = \mathbf{w}_k^H\mathbf{x}(n)$. If R_i is set to be $\sqrt{10}$, then $e_k(n) = \left||y_k(n)| - \sqrt{10}\right|$. According to the above principle, the errors are sorted out in the ascending order, then the outputs corresponding to the first $N/2$ (for the

16-QAM signal and $R_i = \sqrt{10}$, $P_i = 8$ and $Q = 16$, so $P_i/Q = 1/2$) small errors are considered to be the selected outputs $y_{i,k}(n) = \mathbf{w}_k^H \mathbf{x}_i(n)\,(n = 1, 2, \cdots, N_i = N/2)$. As shown in Fig. 3, the outputs marked in blue color (located in the region between the two green circles) are considered as the selected outputs corresponding to $R_i = \sqrt{10}$. The corresponding samples and the set of these samples are symbolized as $\mathbf{x}_{i,k}(n)\,(n = 1, 2, \cdots, N_i)$ and $\Omega_{i,k}$, respectively, and the latter one can be considered as the substitute of the set Ω_i.

According to the above determination method of the set Ω_i, the selected modulus R_i is an important parameter. The following two rules are used to choose R_i. Although the receiver don't know which symbol was sent, the set of all symbols is known, namely the R_i (for all i) is known to the receiver. Firstly, as shown in Fig. 3, the circle with the selected modulus (radius) should pass through as many points as possible. Secondly, the distance between the selected circle and its adjacent circle should be as large as possible.

In the light of the above analysis, the realizable MCMA-MNM is as

$$\mathbf{w}_{k+1} = R_i\big(\mathbf{X}_{i,k}\mathbf{X}_{i,k}^H\big)^{-1}\mathbf{X}_{i,k}\mathbf{y}_{k,i} = R_i\mathbf{R}_{i,k}^{-1}\mathbf{X}_{i,k}\mathbf{y}_{i,k}, \quad (16)$$

where $\mathbf{X}_{i,k} = [\mathbf{x}_{i,k}(1), \mathbf{x}_{i,k}(2), \cdots, \mathbf{x}_{i,k}(N_i)]$ and $\mathbf{R}_{i,k} = \mathbf{X}_{i,k}\mathbf{X}_{i,k}^H$.

Note that the sample matrices $\mathbf{X}_{i,k}$ and $\mathbf{R}_{i,k}$ will change more or less with iterations. Thus, we need updating $\mathbf{R}_{i,k}$ in each iteration, which results in a significant increase in the computational load compared with the unchanged $\mathbf{R}_i$ in the theoretical analysis. Fortunately, according to our theoretical analysis in Subsection III-B1, $\mathbf{X}_i$ is predetermined and independent of the BE. This suggests that the change between $\mathbf{X}_{i,k}$ and $\mathbf{X}_{i,k+1}$ is trivial, especially when the BE is close to convergence. Therefore, the Hessian matrix (correlation matrix) $\mathbf{R}_{i,k+1}$ can be fast calculated by the following formula

$$\mathbf{R}_{i,k} = \mathbf{R}_{i,k} - \sum_{n\in\{n|\mathbf{x}(n)\in\bar{\Omega}_{i,k+1}\}} \mathbf{x}(n)\mathbf{x}^H(n) + \sum_{n\in\{n|\mathbf{x}(n)\in\bar{\Omega}_{i,k}\}} \mathbf{x}(n)\mathbf{x}^H(n) \quad (17)$$

where $\bar{\Omega}_{i,k+1} = \{\mathbf{x}(n) \mid \mathbf{x}(n) \notin \Omega_{i,k+1}, \mathbf{x}(n) \in \Omega_{i,k}\}$ and $\bar{\Omega}_{i,k} = \{\mathbf{x}(n) \mid \mathbf{x}(n) \notin \Omega_{i,k}, \mathbf{x}(n) \in \Omega_{i,k+1}\}$. Because there are only a few elements in the sets $\bar{\Omega}_{i,k}$ and $\bar{\Omega}_{i,k+1}$, it requires a very small amount of calculation to update $\mathbf{R}_{i,k}$ based upon the formula (17), and then the computational volume of the proposed MCMA-MNM is greatly reduced.

To sum up, minimizing CF (11) by using the proposed MNM (14), a optimal BE can be obtained when the solution $\mathbf{w}_{k+1}$ converges. The detailed rationale of the proposed MCMA-MNM is summarized in Algorithm 1, where accuracy parameter η is a sufficiently small positive value, K is the maximum iteration times.

Algorithm 1: The Proposed MCMA-MNM

1: **Initialization**: We find the optimal solution by using the proposed MNM. First, the $\mathbf{w}_0$ is initialized by unit center criteria, i.e. $\mathbf{w}_0 = [0, \cdots, 1, \cdots, 0]^{\mathrm{H}}$, where the element 1 is in the center of the vector. The iteration index k is set to be 0 initial error e is set larger than then accuracy parameter ε.
2: **While** $e > \eta$ and $k < K$ **do**
3: Calculate $y(n)n = 1, \cdots, N$ by using $\mathbf{w} = \mathbf{w}_k$. Then determine $\mathbf{X}_{i,k}$ according to (15) and $\mathbf{R}_{i,k}$ according to (17). On this basis, we update $\mathbf{w}_{k+1}$ by using iteration formula (24).
4: Update iteration error $e = \|\mathbf{w}_{k+1} - \mathbf{w}_k\|_2$.
5: Update iteration index $k = k + 1$ and let $\mathbf{w}_k = \mathbf{w}_{k+1}$.
6: **End While**
7: **Return** $\mathbf{w}_k$

C. Generalized Modified Constant Modulus Algorithm

Based on the analysis made in Remark 1, we know that the proposed MCMA can efficiently overcome the misadjustment and artificial error. However, these advantages are achieved at the cost of reduced sample usage rate. In 16-QAM signal, the sample usage rate is $1/2$ when $R_i = \sqrt{10}$. In this subsection, the generalized modified constant modulus algorithm (GMCMA) is proposed to improve the sample usage rate as well as the convergence speed.

The reason for the reduced sample usage rate is that only the samples corresponding to a special R_i (signal modulus) are used to adjust the BE, whilst the other samples are discarded. To improve the sample usage rate, the samples corresponding to several $R_i \in \Omega\,(i = 1, 2, \cdots, I'; I' \le I)$ are used in the GMCMA. Then, the following optimization problem is constructed:

$$\min_{\mathbf{w}} J_{\mathrm{GMCMA}}(\mathbf{w}) = \sum_{i=1}^{I'} E\left[\left(\left|\mathbf{w}^H\mathbf{x}_i(n)\right| - R_i\right)^2\right]. \quad (18)$$

Remark 3: It can be seen from (18) that the GMCMA is different from the MCMA. The former utilizes the samples belonging to the sets $\Omega_1, \Omega_2, \cdots, \Omega_{I'}$ and penalizes the deviations of the equalized signals' magnitude from the associated modulus which is determined by the sample $\mathbf{x}_i(n)$. In contrast, the latter one only uses the samples belonging to one of the sets $\Omega_1, \Omega_2, \cdots, \Omega_I$, named Ω_i, and penalizes the deviations of the equalized signals' magnitude from the modulus corresponding to Ω_i. Therefore, the GMCMA can preserve all the merits of the MCMA mentioned in Remark 1. Furthermore, it greatly improves the sample usage rate. Particularly, in the extreme case of $I' = I$, the sample usage rate will approach 100%.

Similar to the MCMA, we use the time average to replace the ensemble average. Then the cost function $J_{\mathrm{GMCMA}}(\mathbf{w})$ is expressed as

$$J_{\mathrm{GMCMA}}(\mathbf{w}) = \sum_{i=1}^{I'}\sum_{n=1}^{N_i}\left(\left|\mathbf{w}^H\mathbf{x}_i(n)\right| - R_i\right)^2 \quad (19)$$

Now, differentiating $J_{\mathrm{GMCMA}}(\mathbf{w})$ with respect to $\mathbf{w}$ gives the gradient

$$\begin{aligned}\nabla J_{\mathrm{GMCMA}}(\mathbf{w}) &= \sum_{i=1}^{I'} \mathbf{X}_i\mathbf{X}_i^H\mathbf{w} - R_i\mathbf{X}_i\mathbf{y}_i \\ &= \mathbf{X}'\mathbf{X}'^H\mathbf{w} - \mathbf{X}'\mathbf{y}' \end{aligned} \quad (20)$$

where $\mathbf{X}' = [\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_I]$ and $\mathbf{y}' = \left[R_1\mathbf{y}_1^T, R_2\mathbf{y}_2^T, \cdots, R_{I'}\mathbf{y}_{I'}^T\right]^T$. Finally, the updating formula of the GMCMA

based upon the MNM [40], which is named the MNM based GMCMA (GMCMA-MNM), is derived as

$$\mathbf{w}_{k+1} = \left(\mathbf{X}'\mathbf{X}'^{H}\right)^{-1}\mathbf{X}'\mathbf{y}'_k = \mathbf{R}'^{-1}\mathbf{X}'\mathbf{y}'_k \tag{21}$$

where $y_{i,k}(n) = \mathbf{w}^H\mathbf{x}_i(n)|_{\mathbf{w}=\mathbf{w}_k}$, $\mathbf{y}'_k = \Big[R_1\frac{y^*_{1,k}(1)}{|y_{1,k}(1)|},\cdots, R_1\frac{y^*_{1,k}(N_1)}{|y_{1,k}(N_1)|},\cdots,R_{I'}\frac{y^*_{I',k}(N_{I'})}{|y_{I',k}(N_{I'})|},\cdots,R_{I'}\frac{y^*_{I',k}(N_{I'})}{|y_{I',k}(N_{I'})|}\Big]^T$, and $\mathbf{R}' = \mathbf{X}'\mathbf{X}'^H$.

Remark 4: From the updating formula (21), we can deduce that the proposed GMCMA-MNM persists all the advantages of the MCMA-MNM. Firstly, The GMCMA-MNM also adopts the positive definite Hessian matrix $\mathbf{X}'\mathbf{X}'^H$ and converges stably. Secondly, the matrix $\mathbf{X}'$ is predetermined and then $\mathbf{R}'$ and its corresponding inverse matrix $\mathbf{R}'^{-1}$ can be calculated beforehand. Thus, the computational volume of the GMCMA-MNM is much less than its Newton-type counterpart. Thirdly, the GMCMA-MNM has a fast convergence speed. In fact, we have the following proposition.

Proposition 3: If $\hat{\mathbf{w}}$ is the optimal solution of (19) and $\mathbf{w}_k \approx \hat{\mathbf{w}}$, i.e.,

$$\mathbf{w}_k \in \zeta(\hat{\mathbf{w}},\delta) = \{\mathbf{w} \mid \|\mathbf{w}-\hat{\mathbf{w}}\|_2 \le \delta\}$$

where δ is a small positive constant, then $\mathbf{w}_{k+1}$ in the iteration formula (21) converges to the optimal solution $\hat{\mathbf{w}}$ at the step size of $\frac{1}{2}\|\mathbf{w}_k-\hat{\mathbf{w}}\|_2$.

Proof: This proposition can be similarly proved by the method provided in Appendix B. ∎

In practice, the matrix $\mathbf{X}'$ is replaced by the available matrix $\mathbf{X}'_k$ and the iteration formula of the GMCMA-MNM is modified to

$$\mathbf{w}_{k+1} = \left(\mathbf{X}'_k\mathbf{X}'^{H}_k\right)^{-1}\mathbf{X}'_k\mathbf{y}'_k = \mathbf{R}'^{-1}_k\mathbf{X}'_k\mathbf{y}'_k \tag{22}$$

where $\mathbf{X}'_k = [\mathbf{X}_{1,k},\mathbf{X}_{2,k},\cdots,\mathbf{X}_{I',k}] = \mathbf{X}'|_{\mathbf{w}=\mathbf{w}_k}$ and $\mathbf{y}'_k = \mathbf{y}'|_{\mathbf{w}=\mathbf{w}_k}$.

It is worth mentioning that all the matrixes $\mathbf{X}_{k,i}\big(i=1,2,\cdots,I'\big)$ are obtained recursively. When the required samples $\mathbf{x}_{1,k}(n)(n=1,\cdots,N_1),\cdots,\mathbf{x}_{i-1,k}(n)(n=1,\cdots,N_{i-1})$ have been selected from the sample matrix $\mathbf{X}=[\mathbf{x}(1),\mathbf{x}(2),\cdots,\mathbf{x}(n)]$, $\mathbf{x}_{i,k}(n)\,(n=1,\cdots,N_i)$ are chosen from the remaining samples with the length of $\bar{N}_i = N-N_1-N_2\cdots-N_{i-1}$ via the method suggested in Subsection III-B3. In addition, the Hessian matrix of the GMCMA-MNM can be efficiently updated according to the generalized form of (17):

$$\mathbf{R}'_{k+1} = \mathbf{R}'_k - \sum_{n\in\{n|\mathbf{x}(n)\in\bar{\Omega}'_{k+1}\}}\mathbf{x}(n)\mathbf{x}^H(n) + \sum_{n\in\{n|\mathbf{x}(n)\in\bar{\Omega}'_k\}}\mathbf{x}(n)\mathbf{x}^H(n) \tag{23}$$

where

$$\begin{aligned}\bar{\Omega}'_{k+1} = \big\{\mathbf{x}(n) \mid \mathbf{x}(n)\notin\Omega_{i,k+1}\,(i=1,2,\cdots,I'),\\ \mathbf{x}(n)\in\Omega_{i,k}\,(i=1,2,\cdots,I')\big\}\\ \bar{\Omega}'_k = \big\{\mathbf{x}(n) \mid \mathbf{x}(n)\notin\Omega_{i,k}\,(i=1,2,\cdots,I'),\\ \mathbf{x}(n)\in\Omega_{i,k+1}\,(i=1,2,\cdots,I')\big\}.\end{aligned}$$

Similarly, there are only a few elements contained in the sets $\bar{\Omega}'_{k+1}$ and $\bar{\Omega}'_k$ because of the predetermination of $\Omega_i\,(i=1,2,\cdots,I)$ and the independence between the BE and Ω_i. The computational burden of calculating $\mathbf{R}'_{k+1}$ can be noticeably decreased through the utilization of (23), and furthermore the computational burden of the proposed GMCMA-MNM is significantly reduced. Especially, in the extreme case of $I'=I$, both the sets $\bar{\Omega}'_{k+1}$ and $\bar{\Omega}'_k$ are empty and $\mathbf{R}=\mathbf{X}\mathbf{X}^H=\mathbf{R}'_k$ for all $k\,(k=1,2,\cdots)$. The matrix $\mathbf{R}'_k$ and its inverse matrix $\mathbf{R}'^{-1}_k$ will be unchanged and can be pre-computed.

To sum up, the detailed rationale of the proposed GMCMA-MNM is summarized in Algorithm 2.

Algorithm 2: The Proposed GMCMA-MNM

1: **Initialization**: We find the optimal solution by using the proposed MNM. First, the $\mathbf{w}_0$ is initialized by unit center criteria, i.e. $\mathbf{w}_0=[0,\cdots,1,\cdots,0]^{\mathrm{H}}$, where the element 1 is in the center of the vector. The iteration index k is set to be 0 initial error e is set larger than then accuracy parameter ε.
2: **While** $e>\eta$ and $k<K$ **do**
3: Calculate $y(n)n=1,\cdots,N$ by using $\mathbf{w}=\mathbf{w}_k$. Then determine $\mathbf{X}_{i,k}$ for $i=1,2,\cdots,I'$ according to (15) and $\mathbf{R}_{i,k}$ for $i=1,2,\cdots,I'$ according to (23). Finally, we update $\mathbf{w}_{k+1}$ by using iteration formula (24).
4: Update iteration error $e=\|\mathbf{w}_{k+1}-\mathbf{w}_k\|_2$.
5: Update iteration index $k=k+1$ and let $\mathbf{w}_k=\mathbf{w}_{k+1}$.
6: **End While**
7: **Return** $\mathbf{w}_k$

Interestingly, the proposed method can be easily converted into an adaptive algorithm and the can be applied to a time-vary channel via the following updating formula

$$\mathbf{w}_{k+1} = \mathbf{R}^{-1}_{k+1}\left(\lambda\mathbf{b}_k + \frac{y^*(k+1)}{|y(k+1)|}R_i\mathbf{x}(k+1)\right),$$

where $\mathbf{x}\,(k+1)\in\Omega_i$,

$$\mathbf{R}^{-1}_{k+1} = \lambda^{-1}\mathbf{R}^{-1}_k - \frac{\lambda^{-2}\mathbf{R}^{-1}_k\mathbf{x}(k+1)\mathbf{x}^H(k+1)\mathbf{R}^{-1}_k}{1+\lambda^{-1}\mathbf{x}^H(k+1)\mathbf{R}^{-1}_k\mathbf{x}(k+1)},$$

and

$$\mathbf{b}_{k+1} = \lambda\mathbf{b}_k + \frac{y^*(k+1)}{|y(k+1)|}R_i\mathbf{x}(k+1).$$

The initial values are taken as $\mathbf{R}^{-1}_0=\mathbf{I}$ and $\mathbf{b}_0=\mathbf{0}$.

IV. Computational Complexity

In this section, we discuss the computational complexity of CMA [11], MMA [14], CNA [32], DMS [17] and the proposed MCMA and GMCMA. To simplify the analysis, the computational cost lower than L is omitted in the following discussion.

For CMA/MMA, the precalculation of the output requires L complex multiplications. Once the output is provided, updating the blind equalizer at each iteration requires L multiplications approximately. Thus, the CMA/MMA requires a computational complexity of approximately $2L$ to update the BE once.

For CNA, the computational complexity is similar to the CMA except with 7 additional exponents at each iteration. Accordingly, if the computational cost lower than L is disregarded, the CNA requires $2L$ complex multiplications and 7 exponents at each iteration.

TABLE I
COMPUTATIONAL COMPLEXITY OF THE RELATED METHODS

Methods	CMA	MMA	CNA	DMS	MCMA	GMCMA
Complexity	$14000L$	$14000L$	$50000L$	$21000L$	$NL^2+210N_iL+210NL$	$NL^2+210NL+210\sum_{i=1}^{I'}N_iL$

The DMS decomposes the BE into two components which are related to the CMA and the soft decision-directed algorithm (SDDA), respectively. If the output of the BE is calculated in advance, then the DMS approximately requires L multiplications associated with the CMA, and L multiplications and 4 exponential operations associated with the SDDA. Hence, the computational complexity of the DMS at each iteration is approximately $3L$ multiplications and 4 exponential operations.

Finally, for the proposed MCMA (GMCMA), all the $\mathbf{x}(n)\mathbf{x}^H(n)$ $(n=1,2,\cdots,N)$ can be calculated in advance, which requires NL^2 multiplications. Then, $o(L^3)$ multiplications is needed to calculate $\mathbf{R}_{i,k}^{-1}$ $\left(\mathbf{R}_k'^{-1}\right)$. Once $\mathbf{R}_{i,k}^{-1}$ $\left(\mathbf{R}_k'^{-1}\right)$ is obtained, the MCMA (GMCMA) only requires to update $\mathbf{y}_{k,i}$ $(\mathbf{y}_k')$ and implement $R_i\mathbf{R}_{i,k}^{-1}\mathbf{X}_{i,k}\mathbf{y}_{i,k}$ $\left(\mathbf{R}_k'^{-1}\mathbf{X}_k'\mathbf{y}_k'\right)$ at each iteration, which takes approximately NL multiplications and an approximate computational complexity of N_iL+L^2 $\left(\sum_{i=1}^{I'}N_iL+L^2\right)$, respectively. It is worth noting that $L\ll N_i\,(i=1,2,\cdots,I')$, thus L^2 is considerably less than N_iL and can be omitted. Accordingly, the total computational cost of the MCMA (GMCMA) at each iteration is $NL+N_iL\left(NL+\sum_{i=1}^{I'}N_iL\right)$ when $\mathbf{R}_{i,k}^{-1}$ $\left(\mathbf{R}_k'^{-1}\right)$ is calculated in advance.

For the simulations described in the next section, the iteration times for the CMA, MMA, CNA, DMS, MCMA and GMCMA for the 36-QAM system are 7000, 7000, 25000, 7000, 210 and 140, respectively. Therefore, their total computational volumes of multiplications are $14000L$, $14000L$, $50000L$, $21000L$, $NL^2+210N_iL+210NL$ and $NL^2+210\sum_{i=1}^{I'}N_iL+210NL$, respectively. Moreover, the CNA requires 175000 additional exponentiations and the DMS requires 28000 exponential operations. Without considering the exponentiation and exponent operations, the comparison of the computational complexity of the methods is listed in Table I.

V. SIMULATION RESULTS AND DISCUSSION

To verify the effectiveness of the proposed algorithms, we compare the proposed MCMA and GMCMA with the conventional CMA ($p=2$) [11], MMA, CNA DMS and improved DMS (IDMS) [37] using the symbol error rate (SER), the MSE and the ISI in this section. The MSE is defined as

$$\mathrm{MSE}=E\left[\left|Cy(k)-s(k-\tau)\right|^2\right]$$

where $[C,\tau]=\arg\min_{C,\tau}E\left[\left|Cy(n)-s(n-\tau)\right|^2\right]$. Moreover, the ISI is defined as

$$\mathrm{ISI}=\sum_{n=0,n\neq n_{\max}}^{\bar{L}+L-2}\left|\bar{w}(n)\right|^2\Big/\left|\bar{w}\left(n_{\max}\right)\right|^2$$

in which $\bar{\mathbf{w}}(n)=\left[\bar{w}\left(0\right),\bar{w}(1),\cdots,\bar{w}\left(L+\bar{L}-2\right)\right]^T$ is the combined impulse response of the channel and the equalizer defined in Section II, and $n_{\max}=\arg\max_n|\bar{w}(n)|$.

In the simulation, we consider the QAM signaling over a complex-valued frequency-selective channel with Gaussian noise. The channel impulse response with order $\bar{L}=5$ is assumed to be $\mathbf{h}_1=[0.250+j0.201,\ 0.153+j0.171,\ 0.100+j0.097,\ 0.073+j0.062,\ 0.041+j0.063]^T$, corresponding to the channel gain of $[-9.8758,\ -12.7860,\ -17.1200,\ -20.3749,\ -22.4795]$ in dB. To demonstrate the universality of these methods, the typical data-quality telephone channel $\mathbf{h}_2=[0.04,-0.05,0.07,-0.21,-0.5,0.72,0.36,0,0.21,0.03,0.07]\times exp(-j\pi/5)$ [42] and the non-minimum phase channels $\mathbf{h}_3=[0.0545+0.05j,0.2832-0.11971j,-0.7676+0.2788j,-0.0641-0.0576j,0.0466-0.02275j]$ [43] are used to future illustrate the good performance of the proposed algorithms.

Moreover, a six-tap equalizer is used and initialized with central single-spike. Two modulation schemes are analyzed: first the 16-QAM for the special case of high-order modulation scheme and then the 36-QAM for the general case. The set Ω is defined as $\left\{R_1=\sqrt{10},R_2=\sqrt{2},R_3=3\sqrt{2}\right\}$ for the 16-QAM and $\left\{R_1=\sqrt{26},R_2=\sqrt{34},R_3=\sqrt{10},R_4=\sqrt{18},R_5=\sqrt{2},R_6=5\sqrt{2}\right\}$ for the 36-QAM.

Firstly, we compare the CMA, MMA, DMS, IDMS, MCMA and GMCMA under h_1. The step size of the CMA, MMA and IDMS is chosen as 5×10^{-5}, 8×10^{-4} and 5×10^{-4}, respectively. For the DMS, the step size associated with CMLF is set to be 5×10^{-5}, and the other one associated with CME is taken as 5×10^{-3}. Additionally, the parameters R_i of the MCMA and I' of the GMCMA is taken as $R_i|_{i=1}=\sqrt{10}$ and $I'=2$, respectively. Lastly, the sample number $N=1500$ is adopted in the following except for the MSE versus the number of samples.

Firstly, the simulation results under channel $\mathbf{h}_1$ is presented.

Fig. 4 and Fig. 5 show the MSE and SER versus the SNR, respectively. It can be seen from these two figures that the proposed method nearly approaches the optimal (trained) MMSE equalizer and has better performance than the other methods. This is due to the following two reasons. 1) The MCMA and GMCMA can efficiently suppress the artificial error and steady-state misadjustment as analyzed in Remark 1. 2) Since the proposed algorithms use a large number of sam-

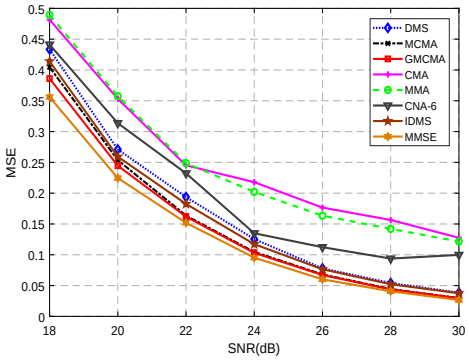

Fig. 4. MSE versus SNR for CMA, MMA, DMS, IDMS, MCMA and GMCMA used for 16-QAM system.

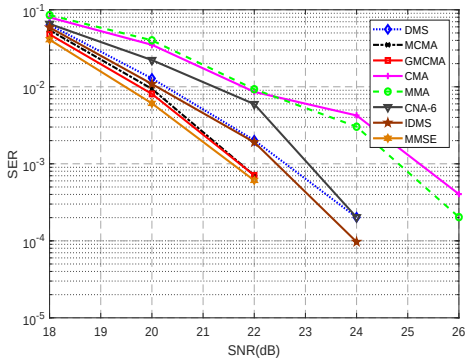

Fig. 5. SER versus SNR for CMA, MMA, DMS, IDMS, MCMA and GMCMA used for 16-QAM system.

ples simultaneously, they avoid the well-known excess error caused by the adaptive methods that only adopt one sample per iteration. Furthermore, the performance of GMCMA is slightly better than that of MCMA, because GMCMA utilizes more samples than MCMA.

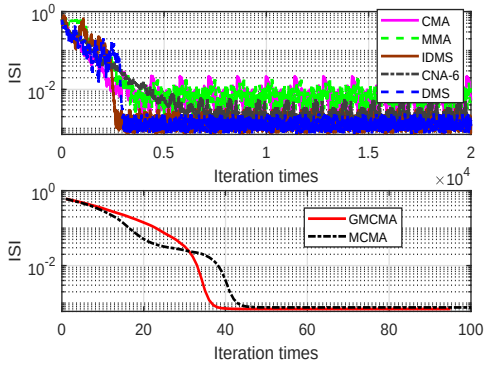

Fig. 6. ISI versus iteration times for CMA, MMA, DMS, IDMS, MCMA and GMCMA used for 16-QAM system.

Given SNR = 28dB and $N = 1500$, the convergence performance of the CMA, MMA, DMS, IDMS, MCMA and GMCMA in terms of the ISI is depicted in Fig. 6 for the 16-QAM system. From this figure, it is observed that the proposed MCMA and GMCMA converge much faster than the other three methods. The reasons are listed as follows: 1) The proposed algorithms efficiently suppress the artificial error and steady-state misadjustment and can steadily converge without fluctuation accordingly. 2) A comparison of (14) and (21) with (6) reveals that an iteration of the proposed algorithms is approximately equivalent to calculating a large number of samples by the adaptive algorithm. 3) The proposed algorithms adopt the constructed MNM, so their convergence speed is faster. More specifically, the GMCMA has a slightly faster convergence speed than the MCMA because of its higher sample usage rate. On the other hand, the proposed methods can also converge to a much lower steady-state ISI than the CMA and MMA, and has a similar steady-state ISI with the DMS and IDMS. The good ISI performance is also due to the facts mentioned in the fourth paragraph of Section V.

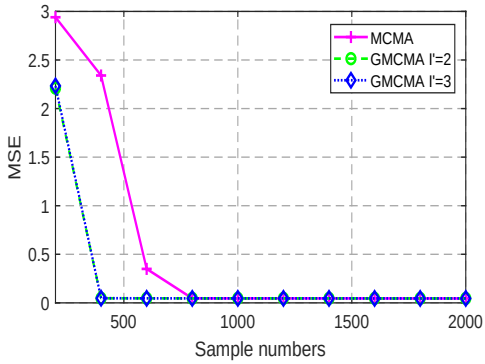

Fig. 7. MSE versus the number of samples for MCMA and GMCMA when SNR = 28dB for 16-QAM system.

For SNR = 28dB, Fig. 7 depicts the variation of the MSE versus the number of samples for the MCMA and GMCMA with different I'. It is clear that the GMCMA needs fewer samples to achieve a good MSE than the MCMA due to its improvement in the sample usage rate. Furthermore, if the sample sequence is long enough, then both the MCMA and GMCMA have similar low MSE. In this case, the samples are sufficient even for the MCMA which has a low sample usage rate to impose a strict constraint to the BE. Therefore, the MCMA and GMCMA are slightly affected by the length of samples when it is longer than 800 as seen in Fig. 7.

Secondly, we compare the CMA, MMA, DMS, IDMS, MCMA and GMCMA for 36-QAM. The step size of the CMA and MMA is set to be 6×10^{-6} and 2×10^{-4}, respectively. In the DMS, the step size of the CMLF and CME is taken as 6×10^{-6} and 2×10^{-3}, respectively. Moreover, the sample number $N = 6000$ is used in the simulations related to the MSE/SER versus the SNR and the ISI versus iteration times. Additionally, the other parameters are chosen as the same as the 16-QAM system.

Fig. 8 and Fig. 9 show the MSE and SER versus the SNR for 36-QAM signals, respectively. It can be seen from Fig. 8 and Fig. 9 that the proposed MCMA and GMCMA have 1dB gain at SER = 0.01 compared with the CMA and MMA. Moreover, their MSE and SER are even lower than the DMS/IDMS and is closest to the optimal MMSE. The reasons for the good performance of the proposed algorithms

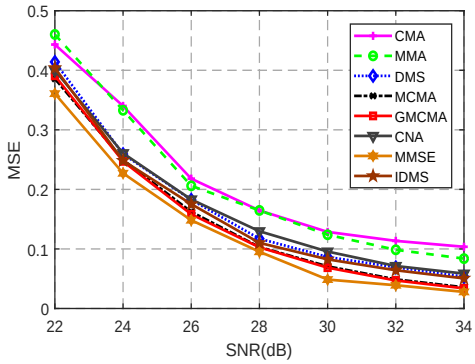

Fig. 8. MSE versus SNR for CMA, MMA, DMS, IDMS, MCMA and GMCMA used for 36-QAM system.

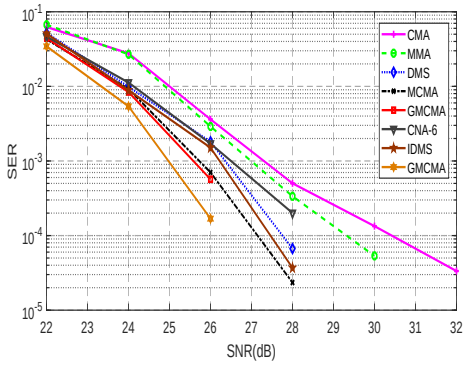

Fig. 9. SER versus SNR for CMA, MMA, DMS, IDMS, MCMA and GMCMA used for 36-QAM system.

used for the 36-QAM symbol are partly identical as those for the 16-QAM symbol.The misadjustment caused by the CMA or MMA increases rapidly with the increase of the signal modulation order, whereas the proposed algorithms can effectively eliminate the misadjustment no matter how high the signal order is.

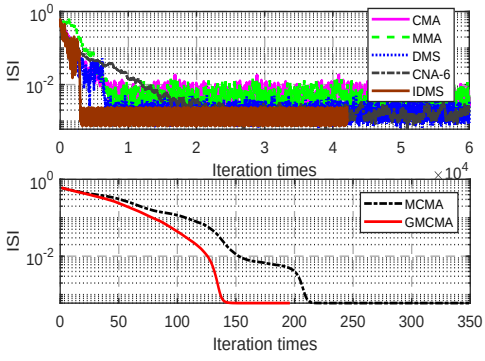

Fig. 10. ISI versus iteration times for CMA, MMA, DMS, IDMS, MCMA and GMCMA used for 36-QAM system.

When SNR = 30dB, the convergence performances of all the five methods are presented in Fig. 10 for the 36-QAM system. It can be seen that both the MCMA and GMCMA still converge faster than the other methods obviously due to the same reasons that we have concluded for the fast convergence speed of the 16-QAM system. Comparing Fig. 10 with Fig. 6, we can observe that all the methods used for the 36-QAM system converge at a slower speed than the case when they are used for the 16-QAM system. This is because the discrepancy between the initial output of the BE and the desirable symbol of the 36-QAM system is larger than that of the 16-QAM system under the condition of equal initial ISI. Moreover, it is worth noting that the gap of the convergence speeds between the MCMA and GMCMA are further enlarged. This is because the sample usage rate of the MCMA of the 36-QAM system is rather low, namely 2/9, while the GMCMA has a relatively much higher sample usage rate, namely 4/9.

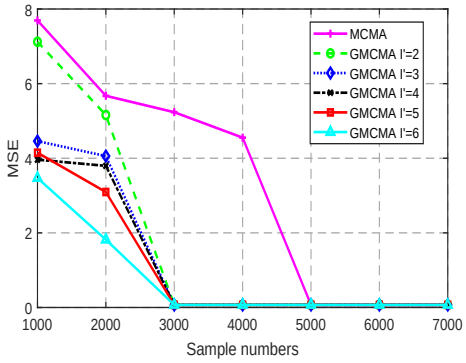

Fig. 11. MSE versus the number of samples for MCMA and GMCMA under the condition of SNR = 30dB for 36-QAM system.

Fig. 11 depicts the MSE versus the sample number when SNR = 30dB for the MCMA and GMCMA. The result indicates that the GMCMA requires fewer samples to achieve desirable equalization performance than the MCMA and both of them can approach the satisfactory performance when the sample sequence is long enough. This result is similar to that of the 16-QAM system. However, it is generally agreed that the proportion of the signals with the same amplitude falls quickly with the rise of the signal modulation order. Therefore, it can be deduced that the MCMA for the 36-QAM symbol needs much more samples to obtain a small MSE than the 16-QAM symbol. Fortunately, comparing with the MCMA, the increased amount of samples needed by the GMCMA is relatively small. More important, there are always enough samples for a blind method, since all received data can be used as samples in practice. Therefore, although the MCMA may need a large number of samples to complete channel equalization, its requirement can always be satisfied. Moreover, we can employ the GMCMA to improve the sample usage rate and then a desired equalization performance can be achieved in the case of a small number of samples.

In the following, the SER of all the method under channel $\mathbf{h}_2$ and $\mathbf{h}_3$ is presented.

As seen in the Fig. and Fig. 12 and Fig. 13, the proposed MCMA and GMCMA still have lower SER than the other methods under channel $\mathbf{h}_2$ and $\mathbf{h}_3$.

Finally, the equalization performance in term of SER is given under time-varying channel. The time variation of channel is approximated by autoregressive model [44]. The initial channel is set to be $\mathbf{h}_1$.

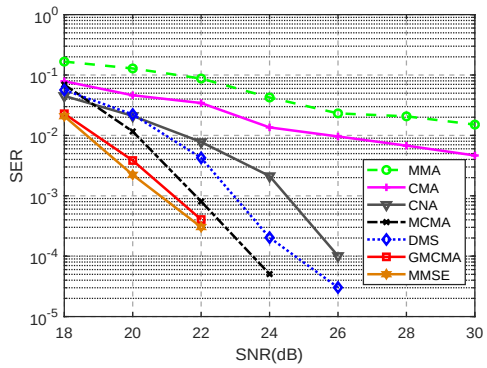

Fig. 12. SER versus SNR for CMA, MMA, DMS, IDMS, MCMA and GMCMA used for 16-QAM system under $\mathbf{h}_2$.

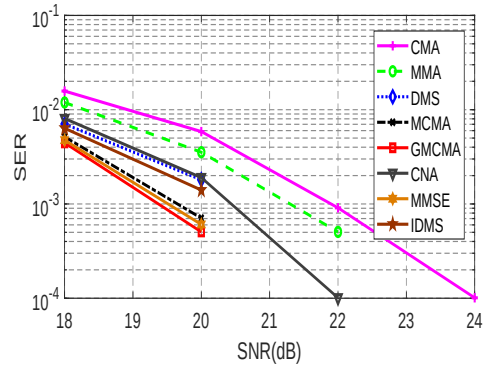

Fig. 13. SER versus SNR for CMA, MMA, DMS, IDMS, MCMA and GMCMA used for 16-QAM system under $\mathbf{h}_3$.

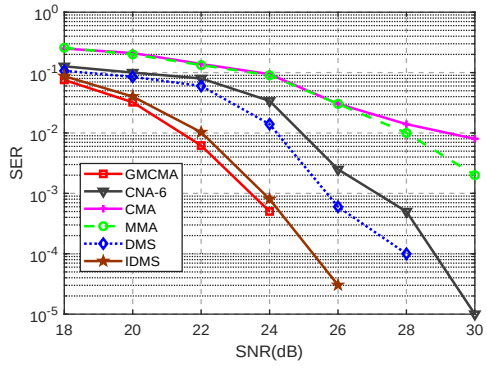

Fig. 14. SER versus SNR for CMA, MMA, DMS, IDMS, MCMA and GMCMA used for 16-QAM system under time-vary channel.

It can be seen from Fig. 14 that the proposed GMCMA has the lowest SER under time-vary channel. This is due to its fast convergence and good original equalization performance.

VI. Conclusion

In this paper, two modified constant modulus algorithms have been proposed for channel equalization of BN-IOT. The proposed MCMA can efficiently deal with these issues at the cost of reduction in the sample usage rate. Moreover, the proposed GMCMA can guarantee the desirable error performance of the MCMA whilst preserving the sample usage rate. Simulation results have demonstrated that the proposed MCMA and GMCMA have more preferable equalization performances than the other existing methods and then the reliability of the IoT can be guaranteed.

Appendix A
Proof of Proposition 1

Proposition 1: If the BE converges to the optimal solution $\hat{\mathbf{w}}$, then the gradient calculated by K samples can be obtained as

$$\begin{aligned}\nabla J(\mathbf{w})|_{\mathbf{w}=\hat{\mathbf{w}}} &= \frac{1}{K}\sum_k \left(|y(k)|^p - R\right)|y(k)|^{p-2}y^*(k)\mathbf{x}(k)|_{\mathbf{w}=\hat{\mathbf{w}}} \\ &= \frac{1}{K}\sum_k |y(k)|^{2p-2}y^*(k)\mathbf{x}(k) \\ &\quad - R|y(k)|^{p-2}y^*(k)\mathbf{x}(k)|_{\mathbf{w}=\hat{\mathbf{w}}}. \end{aligned} \tag{24}$$

Substituting $y(n) = \mathbf{w}^H\mathbf{H}\bar{\mathbf{s}}(n)$ and $\hat{\hat{\mathbf{w}}} = \mathbf{H}^T\hat{\mathbf{w}}^* = \mathbf{e}(\tau)$ into (24) generates

$$\begin{aligned}\nabla J(\mathbf{w})|_{\mathbf{w}=\hat{\mathbf{w}}} &= \frac{1}{K}\sum_k \left|\hat{\mathbf{w}}^H\mathbf{H}\bar{\mathbf{s}}(k)\right|^{2p-2}\left(\hat{\mathbf{w}}^H\mathbf{H}\bar{\mathbf{s}}(k)\right)^*\mathbf{x}(k) \\ &\quad - R\left|\hat{\mathbf{w}}^H\mathbf{H}\bar{\mathbf{s}}(k)\right|^{p-2}\left(\hat{\mathbf{w}}^H\mathbf{H}\bar{\mathbf{s}}(k)\right)^*\mathbf{x}(k) \\ &= \frac{1}{K}\sum_k \left|\mathbf{e}^T(\tau)\bar{\mathbf{s}}(k)\right|^{2p-2}\left(\mathbf{e}^T(\tau)\bar{\mathbf{s}}(k)\right)^*\mathbf{x}(k) \\ &\quad - R\left|\mathbf{e}^T(\tau)\bar{\mathbf{s}}(k)\right|^{p-2}\left(\mathbf{e}^T(\tau)\bar{\mathbf{s}}(k)\right)^*\mathbf{x}(k). \end{aligned} \tag{25}$$

It is a clear derivation that $\mathbf{e}^T(\tau)\bar{\mathbf{s}}(k) = s(k-\tau)$. Taking it into (25), the gradient $\nabla J(\mathbf{w})|_{\mathbf{w}=\hat{\mathbf{w}}}$ can be simplified as

$$\begin{aligned}\nabla J(\mathbf{w})|_{\mathbf{w}=\hat{\mathbf{w}}} &= \mathbf{H}\times\frac{1}{K}\sum_k |s(k-\tau)|^{2p-2}s^*(k-\tau)\bar{\mathbf{s}}(k) \\ &\quad - R|s(k-\tau)|^{p-2}s^*(k-\tau)\bar{\mathbf{s}}(k). \end{aligned} \tag{26}$$

Obviously, the formula $\frac{1}{K}\sum_k |s(k)|^{2p-2}s^*(k)s(j) \approx 0$ holds for $k \neq j$. Thus, the gradient $\nabla J(\mathbf{w})|_{\mathbf{w}=\hat{\mathbf{w}}}$ can be reduced to

$$\nabla J(\mathbf{w})|_{\mathbf{w}=\hat{\mathbf{w}}} = \mathbf{H}\mathbf{e}(\tau)\times\frac{1}{K}\sum_k |s(k-\tau)|^{2p} - R|s(k-\tau)|^p. \tag{27}$$

Making formula (26) be equal to $\mathbf{0}$, then $\sum_k |s(k-\tau)|^{2p} - R|s(k-\tau)|^p$ must be equal to 0, i.e.,

$$R = \frac{\sum_k |s(k-\tau)|^{2p}}{\sum_k |s(k-\tau)|^p}. \tag{28}$$

However,

$$R = \frac{E\left[|s(k)|^{2p}\right]}{E\left[|s(k)|^p\right]} \approx \frac{\sum_k |s(k-\tau)|^{2p}}{\sum_k |s(k-\tau)|^p}$$

for a large sample number. Hence, $\nabla J(\mathbf{w})|_{\mathbf{w}=\hat{\mathbf{w}}}$ is very close to $\mathbf{0}$ but not equal to $\mathbf{0}$.

Moreover, according to (27), the instantaneous gradient is equal to

$$\begin{aligned}&\left(|y(k)|^p - R\right)|y(k)|^{p-2}y^*(k)\mathbf{x}(k)|_{\mathbf{w}=\hat{\mathbf{w}}}\\&= \mathbf{H}\bar{\mathbf{s}}(k)s^*(k-\tau)\left(|s(k-\tau)|^{2p-2} - R|s(k-\tau)|^{p-2}\right).\end{aligned} \tag{29}$$

There is no doubt that $|s(k-\tau)|^{2p-2} - R|s(k-\tau)|^{p-2} \neq 0$. Thus, $\left(|y(k)|^p - R\right)|y(k)|^{p-2}y^*(k)\mathbf{x}(k)|_{\mathbf{w}=\hat{\mathbf{w}}} \neq 0$ always holds. This completes the proof of Proposition 1. ■

APPENDIX B
PROOF OF PROPOSITION 2

Proposition 2: It is clear that $\lim\limits_{k\to\infty}\frac{\|\mathbf{w}_{k+1}-\hat{\mathbf{w}}\|_2}{\|\mathbf{w}_k-\hat{\mathbf{w}}\|_2} = \frac{1}{2}$ indicates that the sequence $\mathbf{w}_k$ converges to $\hat{\mathbf{w}}$ at the step size of $\frac{1}{2}\|\mathbf{w}_k - \hat{\mathbf{w}}\|_2$ per iteration. The limit can be completely proved by considering

$$\lim_{k\to\infty}\left(\mathbf{w}_{k+1} - \hat{\mathbf{w}}\right) = \frac{1}{2}\lim_{k\to\infty}\left(\mathbf{w}_k - \hat{\mathbf{w}}\right)$$

or

$$\mathbf{w}_{k+1} - \hat{\mathbf{w}} = \frac{1}{2}\left(\mathbf{w}_k - \hat{\mathbf{w}}\right). \tag{30}$$

If follows from (30) that

$$\frac{1}{2}\mathbf{w}_k - \mathbf{w}_{k+1} = -\frac{1}{2}\hat{\mathbf{w}}. \tag{31}$$

Now, adding $\frac{1}{2}\mathbf{w}_k$ to both sides of (31) and combining the like terms together, (31) can be rewritten as

$$\mathbf{w}_k - \mathbf{w}_{k+1} = \frac{1}{2}\left(\mathbf{w}_k - \hat{\mathbf{w}}\right). \tag{32}$$

To prove that the above equation holds, we start derivation from the left-hand side of (32). Replacing $\mathbf{w}_{k+1}$ with $R_i\mathbf{R}_i^{-1}\mathbf{X}_i\mathbf{y}_{i,k}$ in (32) yields

$$\begin{aligned}\mathbf{w}_k - \mathbf{w}_{k+1} &= \mathbf{w}_k - R_i\mathbf{R}_i^{-1}\mathbf{X}_i\mathbf{y}_{i,k}\\&= \mathbf{R}_i^{-1}\left(\mathbf{R}_i\mathbf{w}_k - R_i\mathbf{X}_i\mathbf{y}_{i,k}\right).\end{aligned} \tag{33}$$

It is easily deduced from (13) that

$$\mathbf{R}_i\mathbf{w}_k - R_i\mathbf{X}_i\mathbf{y}_{i,k} = \nabla J_{\mathrm{MCMA}}(\mathbf{w}_k). \tag{34}$$

Then (33) can be rewritten as

$$\mathbf{w}_k - \mathbf{w}_{k+1} = \mathbf{R}_i^{-1}\nabla J_{\mathrm{MCMA}}(\mathbf{w}_k). \tag{35}$$

Let us now find the Taylor's series expansion of $\nabla J_{\mathrm{MCMA}}(\mathbf{w}_k)$ up to the first-order terms at the point $\hat{\mathbf{w}}$. Noticing that $\nabla J_{\mathrm{MCMA}}(\hat{\mathbf{w}}) = \mathbf{0}$, then we obtain

$$\begin{aligned}&\nabla J_{\mathrm{MCMA}}(\mathbf{w}_k)\\&= \nabla J_{\mathrm{MCMA}}(\hat{\mathbf{w}}) + \left.\frac{\partial\nabla J_{\mathrm{MCMA}}(\mathbf{w})}{\partial\mathbf{w}^T}\right|_{\mathbf{w}=\hat{\mathbf{w}}}\left(\mathbf{w}_k - \hat{\mathbf{w}}\right)\\&\quad + o\left(\mathbf{w}_k - \hat{\mathbf{w}}\right)\\&= \mathbf{R}_i\left(\mathbf{w}_k - \hat{\mathbf{w}}\right) - R_i\mathbf{X}_i\left.\frac{\partial\mathbf{y}_i}{\partial\mathbf{w}^T}\right|_{\mathbf{w}=\hat{\mathbf{w}}}\left(\mathbf{w}_k - \hat{\mathbf{w}}\right)\\&\quad + o\left(\mathbf{w}_k - \hat{\mathbf{w}}\right)\end{aligned} \tag{36}$$

where $o\left(\mathbf{w}_k - \hat{\mathbf{w}}\right)$ is a quadratic function with respect to $\mathbf{w}_k - \hat{\mathbf{w}}$. To calculate $\frac{\partial\mathbf{y}_i}{\partial\mathbf{w}^T}$, we first differentiate $R_i\frac{y_i^*(n)}{|y_i(n)|}$ with respect to $\mathbf{w}^T$ as follows:

$$\begin{aligned}&\frac{\partial R_i\frac{y_i^*(n)}{|y_i(n)|}}{\partial\mathbf{w}^T}\\&= R_i\left(\frac{1}{|y_i(n)|^2}\left(|y_i(n)|\frac{\partial y_i^*(n)}{\partial\mathbf{w}^T} - y_i^*(n)\frac{\partial|y_i(n)|}{\partial\mathbf{w}^T}\right)\right)\\&= R_i\left(\frac{1}{|y_i(n)|^2}\left(|y_i(n)|\,\mathbf{x}_i^H(n) - \frac{1}{2}y_i^*(n)\frac{y_i(n)\mathbf{x}_i^H(n)}{|y_i(n)|}\right)\right)\\&= R_i\left(\frac{1}{|y_i(n)|^2}\left(|y_i(n)|\,\mathbf{x}_i^H(n) - \frac{1}{2}|y_i(n)|^2\frac{\mathbf{x}_i^H(n)}{|y_i(n)|}\right)\right)\\&= \frac{1}{2}R_i\frac{\mathbf{x}_i^H(n)}{|y_i(n)|}.\end{aligned} \tag{37}$$

It is undoubted that $|y_i(n)|$ is close enough to R_i when $\mathbf{w}$ approaches $\hat{\mathbf{w}}$, i.e.,

$$|y_i(n)|\big|_{\mathbf{w}=\hat{\mathbf{w}}} \approx R_i. \tag{38}$$

Hence, differentiating $R_i\frac{y_i^*(n)}{|y_i(n)|}$ with respect to $\mathbf{w}^T$ at the point $\hat{\mathbf{w}}$ gives the below expression

$$\left.\frac{\partial R_i\frac{y_i^*(n)}{|y_i(n)|}}{\partial\mathbf{w}^T}\right|_{\mathbf{w}=\hat{\mathbf{w}}} = \frac{1}{2}\mathbf{x}_i^H(n). \tag{39}$$

Moreover, because $\mathbf{y}_i = \left[\frac{y_i^*(1)}{|y_i(1)|}, \frac{y_i^*(2)}{|y_i(2)|}, \cdots, \frac{y_i^*(N_i)}{|y_i(N_i)|}\right]^T$ and $\mathbf{X}_i^H = \left[\mathbf{x}_i(1), \mathbf{x}_i(2), \cdots, \mathbf{x}_i\left(N_i\right)\right]^H$, the gradient expression $R_i\mathbf{X}_i\frac{\partial\mathbf{y}_i}{\partial\mathbf{w}^T}\big|_{\mathbf{w}=\hat{\mathbf{w}}}$ is given as

$$R_i\mathbf{X}_i\left.\frac{\partial\mathbf{y}_i}{\partial\mathbf{w}^T}\right|_{\mathbf{w}=\hat{\mathbf{w}}} = \frac{1}{2}\mathbf{X}_i\mathbf{X}_i^H = \frac{1}{2}\mathbf{R}_i. \tag{40}$$

Furthermore, substituting the above derivation result into (36), the gradient of $\nabla J_{\mathrm{MCMA}}(\mathbf{w}_k)$ can be eventually expressed by Taylor's series expansion as

$$\nabla J_{\mathrm{MCMA}}(\mathbf{w}_k) = \frac{1}{2}\mathbf{R}_i\left(\mathbf{w}_k - \hat{\mathbf{w}}\right) + o\left(\mathbf{w}_k - \hat{\mathbf{w}}\right). \tag{41}$$

Finally, based on the above derivation, (35) can be reexpressed s

$$\begin{aligned}\mathbf{w}_k - \mathbf{w}_{k+1} &= \mathbf{R}_i^{-1}\left(\frac{1}{2}\mathbf{R}_i\left(\mathbf{w}_k - \hat{\mathbf{w}}\right) + o\left(\mathbf{w}_k - \hat{\mathbf{w}}\right)\right)\\&= \frac{1}{2}\left(\mathbf{w}_k - \hat{\mathbf{w}}\right) + \mathbf{R}_i^{-1}\times o\left(\mathbf{w}_k - \hat{\mathbf{w}}\right).\end{aligned} \tag{42}$$

Since $o\left(\mathbf{w}_k - \hat{\mathbf{w}}\right)$ is a quadratic function with respect to $\mathbf{w}_k - \hat{\mathbf{w}}$, the term $\mathbf{R}_i^{-1}\times o\left(\mathbf{w}_k - \hat{\mathbf{w}}\right)$ can often be ignoblue when $\mathbf{w}_k$ is close enough to the optimal point $\hat{\mathbf{w}}$. Then the above mathematical relationship can be directly simplified as

$$\mathbf{w}_k - \mathbf{w}_{k+1} = \frac{1}{2}\left(\mathbf{w}_k - \hat{\mathbf{w}}\right). \tag{43}$$

This means that $\mathbf{w}_{k+1} - \hat{\mathbf{w}} = \frac{1}{2}\left(\mathbf{w}_k - \hat{\mathbf{w}}\right)$ when $\mathbf{w}_k$ is close enough to the optimal point $\hat{\mathbf{w}}$. Additionally, it is defaulted that $\mathbf{w}_k$ belongs to the set $\zeta\left(\hat{\mathbf{w}}, \delta\right) = \{\mathbf{w} \mid \|\mathbf{w} - \hat{\mathbf{w}}\|_2 \leq \delta\}$,

i.e., $\mathbf{w}_k$ is close enough to the optimal point $\hat{\mathbf{w}}$, when k tends to infinity. Therefore, we have the following conclusion

$$\lim_{k\to\infty} \frac{\|\mathbf{w}_{k+1} - \hat{\mathbf{w}}\|_2}{\|\mathbf{w}_k - \hat{\mathbf{w}}\|_2} = \frac{1}{2}. \tag{44}$$

The above limit shows that the sequence $\mathbf{w_k}$ converges to the optimal solution $\hat{\mathbf{w}}$ at the step size of $\frac{1}{2}\|\mathbf{w}_k - \hat{\mathbf{w}}\|_2$ per iteration. This completes the proof of Proposition 2. ■

REFERENCES

[1] GSMA, "3GPP low power wide area technologies," *GSMA white paper*, London, U.K., Tech. Rep., Oct. 2016.

[2] Z. Na, Z. Pan, M. Xiong, J. Xia and W. Lu, "Soft Decision Control Iterative Channel Estimation for the Internet of Things in 5G Networks," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 5990–5998, 2019.

[3] L. Zhang, A. Ijaz, P. Xiao and R. Tafazolli, "Channel Equalization and Interference Analysis for Uplink Narrowband Internet of Things (NB-IoT)," *IEEE Communications Letters*, vol. 21, no. 10, pp. 2206-2209, 2017.

[4] S. Jing, J. Hall, Y. R. Zheng and C. Xiao, "Signal Detection for Underwater IoT Devices With Long and Sparse Channels," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 6664-6675, 2020.

[5] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ: Prentice Hall, 2002.

[6] T. Ikuma and A. A. Beex, "Improved mean-square error estimate for the LMS transversal equalizer with narrowband interference," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5273–5277, 2008.

[7] C. Krall, K. Witrisal, G. Leus, and H. Koeppl, "Minimum mean-square error equalization for second-order Volterra systems," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4729–4737, 2008.

[8] A. Ahmed, "A convex approach to blind MIMO communications," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 812–815, 2018.

[9] C. Yu and L. Xie, "On recursive blind equalization in sensor networks," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 662–672, 2015.

[10] M. Komatsu, N. Tanabe, and T. Furukawa, "Direct blind equalization corresponding to noisy environment using Rayleigh quotient," in *Proc. 2019 IEEE 15th Int. Colloquium Signal Process. Its Appl.* (CSPA), Penang, Malaysia, pp. 35–38, 2019.

[11] N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communications systems," *IEEE Trans. Commun.*, vol. COM-28, no. 11, pp. 1867–1875, 1980.

[12] J. Li, D. Feng, and W. X. Zheng, "A robust decision directed algorithm for blind equalization under α-stable noise," *IEEE Trans. Signal Process.*, vol. 69, pp. 4949–4960, 2021.

[13] Q. Han, L. Yang, J. Du, and L. Cheng, "Blind equalization for chaotic signals based on echo state network and Kalman filter under nonlinear channels," *IEEE Commun. Lett.*, vol. 25, no. 2, pp. 589–592, 2021.

[14] J. Yuan and K. Tsai, "Analysis of the multimodulus blind equalization algorithm in QAM communication systems," *IEEE Trans. Commun.*, vol. 53, no. 9, pp. 1427–1431, 2005.

[15] K. Nam and Y. O. Chin, "Modified constant modulus algorithm: blind equalization and carrier phase recovery algorithm," *IEEE International Conference on Communications ICC '95*, Seattle, WA, USA, vol. 1, pp. 498-502, 1995.

[16] L. He, M. Amin, C. Reed and R. Malkemes, "A hybrid adaptive blind equalization algorithm for QAM signals in wireless communications," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 2058–2069, 2004.

[17] N. Xie, H. Hu, and H. Wang, "A new hybrid blind equalization algorithm with steady state performance analysis," *Digital Signal Process*, vol. 22, no. 2, pp. 233–237, 2012.

[18] J. Sun, X. Li, K. Chen, W. Cui and M. Chu, "A Novel CMA+DD_LMS Blind Equalization Algorithm for Underwater Acoustic Communication," *Computer Journal*, vol. 63, no. 1, pp. 974-981, Jan. 2020.

[19] C. Fan, C. Fang, H. Hu, and W. Hsu, "Design and analyses of a fast feed-forward blind equalizer with two-stage generalized multilevel modulus and soft decision-directed scheme for high-order QAM cable downstream receivers," *IEEE Trans. Consum. Electron.*, vol. 56, no. 4, pp. 2132–2140, 2010.

[20] J. Li, D. Feng, and B. Li, "Space-time blind equalization of dispersive MIMO systems driven by QAM signals," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4136–4148, 2018.

[21] Y. Sato, "A method of self-recovering equalization for multilevel amplitude modulation systems," *IEEE Trans. Commun.*, vol. COM-23, no. 6, pp. 679–682, 1975.

[22] Y. Chen, Y. Wang, K. Wang, "FPGA-Based Demonstration of WDM-PAM4-PON System With Low-Complexity CMA," *IEEE Photonics Technology Letters*, vol. 35, no. 11, pp. 621-624, 2023.

[23] Y. Xiao and J. Sun, "RLS CMA blind equalization with adaptive forgetting factor controlled by energy steady state," in *Proc. 2016 9th Int. Congress Image Signal Process., Biomed. Eng. Informat.* (CISP-BMEI), Datong, China, pp. 935–939, 2016.

[24] K. Maruta and C. Ahn, "Uplink interference suppression by semi-blind adaptive array with decision feedback channel estimation on multicell massive MIMO systems," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6123–6134, 2018.

[25] O. Dabeer and E. Masry, "Convergence analysis of the constant modulus algorithm," *IEEE Trans. Inf. Thoery*, vol. 49, no. 6, pp. 1447–1464, 2003.

[26] S. Alsaid, E. Abdel-Raheem, and K. Mayyas, "New constant modulus blind equalizer with optimum tap-length for QAM signals," in *Proc. 2019 IEEE Int. Sym. Signal Process. Inf. Technol.* (ISSPIT), Ajman, UAE, pp. 1–4, 2019.

[27] X. Gu, Z. Wang, R. Cao, Y. Hu, and L. Hao, "Research on blind equalization algorithm of multipath interference PCM-FM signal based on CMA," in *Proc. 2019 IEEE 2nd Int. Conf. Inf. Commun. Signal Process.* (ICICSP), Weihai, Shandong, China, pp. 67–71, 2019.

[28] V. Savaux, F. Bader, and J. Palicot, "OFDM/OQAM blind equalization using CNA approach," *IEEE Trans. Signal Process.*, vol. 64, no. 9, pp. 2324–2333, 2016.

[29] A. M. Ragheb, M. Shoaib, S. Alshebeili, and H. Fathallah, "Enhanced blind equalization for optical DP-QAM in finite precision hardware," *IEEE Photon. Technol. Lett.*, vol. 27, no. 2, pp. 181–184, 2015.

[30] S. Lambotharan, J. Chambers, and C. R. Johnson, "Attractions of saddles and slow convergence in CMA adaptation," *Signal Process.*, vol. 59, no. 3, pp. 335–340, 1997.

[31] J. Li, D. Feng, and W. X. Zheng. "An efficient soft decision-directed algorithm for blind equalization of 4-QAM systems," in *Proc. 2016 49th IEEE Int. Symp. Circuits Syst.* (ISCAS), Montréal, Canada, pp. 1726–1729, 2016.

[32] A. Goupil and J. Palicot, "New algorithms for blind equalization: The constant norm algorithm family," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1436–1444, 2007.

[33] K. K. Delgado and Y. Isukapalli, "Use of the Newton Method for Blind Adaptive Equalization Based on the Constant Modulus Algorithm," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3983–3995, 2008.

[34] P. Priyadarshi and C. S. Rai, "Modified constant modulus type (MCMT) algorithm for blind channel equalization," in *Proc. 2017 Int. Conf. Wireless Commun., Signal Process. Netw.* (WiSPNET), Chennai, India, pp. 2517–2520, 2017.

[35] J. Ma, T. Qiu, and Q. Tian, "Fast blind equalization using bounded nonlinear function with non-Gaussian noise," *IEEE Commun. Lett.*, vol. 24, no. 8, pp. 1812–1815, Aug. 2020.

[36] J. Sun, X. Li, K. Chen, W. Cui, and M. Chu, "A novel CMA+DD_LMS blind equalization algorithm for underwater acoustic communication," *J. Comput.*, vol. 63, no. 1, pp. 974–981, Jan. 2020.

[37] Z. Lv, B. Feng and L. Tan, "An Improved Dual-mode Blind Equalization Algorithm for QAM Signals," *2022 6th International Conference on Robotics and Automation Sciences (ICRAS)*, Wuhan, China, pp. 283-287, 2022.

[38] M. Xiang, Y. Xia, and D. P. Mandic, "Performance analysis of deficient length quaternion least mean square adaptive filters," *IEEE Trans. Signal Process.*, vol. 68, pp. 65–80, 2020.

[39] K. K. Delgado and Y. Isukapalli, "Use of the Newton method for blind adaptive equalization based on the constant modulus algorithm," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3983–3995, 2008.

[40] J. Li, D. Feng, and W. X. Zheng, "Space-time semi-blind equalizer for dispersive QAM MIMO system based on modified Newton method," *IEEE Trans. Wireless Commun.*, vol. 13, no. 6, pp. 3244–3256, 2014.

[41] A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of a nonminimum phase system: blind adjustment of a linear equalizer in data communication," *IEEE Trans. Autom. Control*, vol. 25, no. 3, pp. 385–399, 1980.

[42] G. Yan and H. Fan, "A newton-like algorithm for complex variables with applications in blind equalization," IEEE Trans. Signal Process., vol. 48, no. 2, pp. 553C556, 2000.

[43] A. Caciularu and D. Burshtein, "Unsupervised Linear and Nonlinear Channel Equalization and Decoding Using Variational Autoencoders," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 3, pp. 1003-1018, 2020.

[44] Y. Yu, "An Improved Particle Smoother for Blind Equalization in Time-Varying MIMO Channels," *IEEE Communications Letters*, vol. 19, no. 6, pp. 929-932, 2015.