Maximization of Zero-Error Probability for Adaptive Channel Equalization

Namyong Kim, Kyu-Hwa Jeong, and Liuqing Yang

Abstract: **A new blind equalization algorithm that is based on maximizing the probability that the constant modulus errors concentrate near zero is proposed. The cost function of the proposed algorithm is to maximize the probability that the equalizer output power is equal to the constant modulus of the transmitted symbols. Two blind information-theoretic learning (ITL) algorithms based on constant modulus error signals are also introduced: One for minimizing the Euclidean probability density function distance and the other for minimizing the constant modulus error entropy. The relations between the algorithms and their characteristics are investigated, and their performance is compared and analyzed through simulations in multi-path channel environments. The proposed algorithm has a lower computational complexity and a faster convergence speed than the other ITL algorithms that are based on a constant modulus error. The error samples of the proposed blind algorithm exhibit more concentrated density functions and superior error rate performance in severe multi-path channel environments when compared with the other algorithms.**

Index Terms: **Blind, constant modulus error, equalization, information-theoretic learning (ITL), Parzen window, zero-error probability.**

I. INTRODUCTION

Unlike the mean square error (MSE) criterion that is based on error energy, the information-theoretic learning (ITL) method is based on the combined use of a nonparametric probability density function (PDF) estimator and a procedure to compute entropy [1]. As a robust ITL algorithm, a minimization of error entropy (MEE) algorithm has been developed by Principe and Erdogmus [2]. In their work it has also been shown that the combination of Renyi's quadratic entropy expression with the Parzen PDF estimator [1] is negatively proportional to the logarithmic value of the information potential [1] of error samples. Since logarithm is a monotonic function, it maximizes the information potential in the MEE, instead of minimizing Renyi's entropy. Therefore, the MEE criterion can be considered as the maximization of information potential. The MEE criterion has shown superior performance when compared to the MSE criterion in supervised channel equalization applications [3]. Another cost function based on the quadratic distance between two distributions has been considered by Principe *et al.* [4]. The objective of the algorithm that is based on the minimization of Euclidean distance (MED) is to adjust the system such that the error PDF is as close as possible to the delta distribution. This approach can be considered as a method that matches the error PDF with the delta distribution. Some of the demerits of the MED and MEE algorithms include the computational complexity that arises from the estimation of the entropy or the quadratic distance and their incompatibility with unsupervised signal processing.

In order to realize unsupervised, blind channel equalization, we can adopt the strategy that maximizes the probability that the constant modulus error (CME) becomes zero. Firstly, we present a new blind method to use the CME in the minimization of the Euclidean distance (MED-CME) that forces the PDF of the CME to match with the delta distribution. In the process of MED-CME, two conflicting terms are observed. The term for the maximization of the CME entropy forces the CME samples to have a dispersed distribution. This is in discord with the goal that the error samples should be near zero. In order to prevent this conflict, we propose a method to maximize only the other term, the term for zero-CME probability. Thirdly, the possibility of applying the CME to the MEE criterion is investigated and certain problems of the method are discussed.

The extension of ITL to blind methods by substituting the trained error with the CME has been considered in references [5] and [6], but the algorithms operate under heavy computational burdens. In this paper, we show that the implementation of the new method of maximization of zero-error probability based on the CME requires a significant reduction in computational complexity and is superior in performance in the blind equalization of multi-path channel models.

This paper is organized as follows. Section II presents the supervised ITL criteria that are related with Euclidean PDF distance and their algorithms. New blind equalizer algorithms based on CME, MED, and maximum zero-error probability criterion are proposed in Section III. Section IV reports and discusses the simulation results. Finally, concluding remarks are presented in Section V.

II. SUPERVISED ITL CRITERIA RELATED TO THE EUCLIDEAN PDF DISTANCE AND THEIR ALGORITHMS

A. Supervised MED Criterion

In this sub-section we introduce supervised finite impulse response (FIR) filter algorithms in order to create a concentration of error samples near zero by using the ITL criteria that are related to the Euclidean PDF distance. With this as the objective, we first minimize the Euclidean distance $ED[f_E(e), \delta(e)]$ between the two PDFs, the error signal PDF $f_E(e)$, and the Dirac-

Manuscript received February 09, 2009; approved for publication by Chungyong Lee, Division II Editor, March 09, 2010.

N. Kim is with the Department of Information and Communication Engineering, Kangwon National University, Korea, email: namyong@kangwon.ac.kr.

K.-H. Jeong is with the Intel Mask Operation, Santa Clara, CA, USA, email: kyu-hwa.jeong@intel.com.

L. Yang is with the Department of Electrical and Computer Engineering, University of Florida, FL, USA, email: lqyang@ece.ufl.edu.

1229-2370/10/$10.00 © 2010 KICS

delta function $\delta(e)$ so that the error PDF shows an impulse peak at the origin.

$$ED[f_E(e),\delta(e)] = \int \left(f_E^2(\xi) + \delta^2(\xi) - 2f_E(\xi)\delta(\xi)\right) d\xi. \tag{1}$$

Substituting IP_e for $\int f_E^2(\xi)d\xi$ in (1), where IP_e is defined as the information potential in reference [1], we obtain

$$ED[f_E(e),\delta(e)] = IP_e + c - 2f_E(0). \tag{2}$$

The square of the Dirac distribution, $\int \delta^2(\xi)d\xi$, is mathematically undefined, but the term can be treated as a constant since it does not depend on the weights of the adaptive system.

This supervised Euclidean distance criterion between the error PDF and delta distribution has been proposed for infinite impulse response (IIR) adaptive filtering in reference [4].

We now derive a FIR adaptive equalizer algorithm that is based on the supervised Euclidean distance criterion between the error PDF and the delta distribution. In the case of FIR linear equalization, a tapped delay line (TDL) with L taps can be used for the input vector $\mathbf{X}_k = [x_k\ x_{k-1}\ x_{k-2} \cdots x_{k-L+1}]^T$ and the output sample $y_k = \mathbf{W}_k^T\mathbf{X}_k$, where $\mathbf{W}_k$ is the weight vector at time k. Let us define the error $e_k = d_k - y_k$, where d_k is the desired value at time k; we can then adopt a gradient descent method for the minimization of the cost function (2).

$$\mathbf{W}_{\text{new}} = \mathbf{W}_{\text{old}} - \mu_{\text{MED}} \frac{\partial ED[f_E(e),\delta(e)]}{\partial \mathbf{W}}. \tag{3}$$

In order to calculate the error PDF $f_E(e)$ non-parametrically, we need the Parzen estimator [1] that contains a Gaussian kernel and a block of N error samples as follows

$$\begin{aligned} f_E(e) &= \frac{1}{N}\sum_{i=1}^{N} G_\sigma(e-e_i) \\ &= \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(e-e_i)^2}{2\sigma^2}\right]. \end{aligned} \tag{4}$$

In online systems that operate on a sample-by-sample basis, we can use a small sliding window and then evaluate the gradient from

$$\begin{aligned} \frac{\partial ED[f_E(e),\delta(e)]}{\partial \mathbf{W}} = &-\frac{1}{2\sigma^2N^2}\sum_{i=k-N+1}^{k}\sum_{j=k-N+1}^{k}(e_j-e_i) \\ &\cdot G_{\sigma\sqrt{2}}(e_j-e_i)[\mathbf{X}_j-\mathbf{X}_i] \\ &+ \frac{2}{\sigma^2 N}\sum_{i=k-N+1}^{k} e_i G_\sigma(-e_i)\frac{\partial y_i}{\partial \mathbf{W}}. \end{aligned} \tag{5}$$

Then, the MED algorithm for the supervised FIR adaptive equalizer can be expressed as

$$\begin{aligned} \mathbf{W}_{k+1} = \mathbf{W}_k - \frac{\mu_{\text{MED}}}{\sigma^2 N}\Bigg[\frac{1}{2N}\sum_{i=k-N+1}^{k}\sum_{j=k-N+1}^{k}(e_j-e_i) \\ \cdot G_{\sigma\sqrt{2}}(e_j-e_i)[\mathbf{X}_j-\mathbf{X}_i] - 2\sum_{i=k-N+1}^{k} e_i G_\sigma(-e_i)\mathbf{X}_i\Bigg] \end{aligned} \tag{6}$$

where μ_{MED} is the step-size for the adaptation control of the supervised MED algorithm.

B. Supervised MEE Criterion

Entropy is a scalar quantity that provides a measure for the average information contained in a given PDF. When the error entropy is minimized, the error distribution of adaptive systems is concentrated. Renyi's quadratic error entropy which is effectively used in ITL methods is defined as

$$H(e) = -\log\left(\int f_E(\xi)^2 d\xi\right). \tag{7}$$

Substituting the information potential IP_e, for $\int f_E^2(\xi)d\xi$ in (1), we obtain

$$H(e) = -\log(IP_e) \tag{8}$$

where $IP_e = N^{-2}\sum_{i=k-N+1}^{k}\sum_{j=k-N+1}^{k} G_{\sigma\sqrt{2}}(e_j-e_i)$. Obviously, minimizing the error entropy $H(e)$ is equivalent to maximizing the information potential IP_e. This criterion of maximizing IP_e is referred to as MEE [2].

By applying a gradient ascent method to the maximization of IP_e, the supervised MEE algorithm in references [2] and [3] can be obtained as

$$\begin{aligned} \mathbf{W}_{k+1} = \mathbf{W}_k + \frac{\mu_{\text{MEE}}}{2\sigma^2N^2}\sum_{i=k-N+1}^{k}\sum_{j=k-N+1}^{k}(e_j-e_i) \\ \cdot G_{\sigma\sqrt{2}}(e_j-e_i)[\mathbf{X}_j-\mathbf{X}_i]. \end{aligned} \tag{9}$$

C. Supervised MZEP Criterion

Minimizing $ED[f_E(e),\delta(e)]$ in (2) leads to the simultaneous minimization and maximization of IP_e and $f_E(0)$, respectively, since they have opposite signs. It is noticeable that the minimization of IP_e, which the indicates the maximization of the error entropy, forces the error samples to have a dispersed distribution. This is in discord with the MEE criterion that maximizes IP_e in (8) to force the error samples to be near zero [2]. From the aspect of physical meaning, minimizing IP_e can be considered as applying a force opposite to the direction of the overall force of the MED by pushing the error samples apart and causing them to be diffused. To avoid this conflict, we propose to maximize only the third term, $f_E(0)$, while omitting the error information potential IP_e from (2). We can also remove the constant term from (2), because it does not depend on the equalizer weight $\mathbf{W}$. By adopting this procedure, a criterion for errors, maximization of zero-error probability (MZEP), can be obtained as follows

$$\max_{\mathbf{W}} f_E(0). \tag{10}$$

Although (9) has its origin in the Euclidean distance criterion, this new cost function no longer uses the Euclidean distance, because it only contains the term $f_E(0)$. With $e = 0$ in (4), the zero-error probability $f_E(0)$ reduces to $f_E(0) = N^{-1}\sum_{i=k-N+1}^{k} G_\sigma(-e_i)$. We now derive a gradient ascent method for the maximization of the cost function (10) as follows $\mathbf{W}_{\text{new}} = \mathbf{W}_{\text{old}} + \mu_{\text{MZEP}}\partial f_E(0)/\partial \mathbf{W}$. The gradient is evaluated from

$$\frac{\partial f_E(0)}{\partial \mathbf{W}} = \frac{1}{\sigma^2 N}\sum_{i=k-N+1}^{k} e_i G_\sigma(-e_i)\frac{\partial y_i}{\partial \mathbf{W}}. \tag{11}$$

The MZEP algorithm can be expressed as

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \frac{\mu_{\text{MZEP}}}{\sigma^2 N} \sum_{i=k-N+1}^{k} e_i G_\sigma(-e_i)\mathbf{X}_i. \quad (12)$$

This expression for the resultant supervised method of maximizing the zero-error probability is exactly the same as that for the maximization of the correntropy criterion (MCC), which has been derived by using a different approach, described as follows [7], [8].

D. Supervised MCC Criterion

Correntropy is a similarity measure that is analogous to the autocorrelation of two random processes. Let a nonlinear mapping Φ transform the data to an infinite dimensional reproducing kernel Hilbert space F. The auto-correntropy function $V_X(t, s)$ for a random process $X(t)$ is then defined as

$$V_X(t,s) = E[< \Phi(X(t), \Phi(X(s)) >_F] \quad (13)$$

where $E[\cdot]$ and $< \cdot, \cdot >_F$ denote statistical expectation and inner product in F, respectively.

Cross-correntropy is a generalized version of the similarity measure between two scalar random variables X and Y defined by

$$V(X,Y) = E[< \Phi(X), \Phi(Y) >_F]. \quad (14)$$

When the Gaussian kernel $G_\sigma(x - x_i)$, which satisfies Mercer's Theorem [9], is used, (14) can be rewritten as $V_\sigma(X,Y) = E[G_\sigma(X - Y)]$. In practice, the sample estimator can be used in place of statistical expectation, and we obtain the following cross-correntropy function $V_\sigma(X,Y) = N^{-1}\sum_{i=1}^{N} G_\sigma(x_i - y_i)$. Replacing x_i with d_i, we can obtain the cross-correntropy for the error signal [7], [8].

$$V_\sigma(D,Y) = \frac{1}{N}\sum_{i=1}^{N} G_\sigma(d_i - y_i) = \frac{1}{N}\sum_{i=1}^{N} G_\sigma(e_i). \quad (15)$$

This is of exactly the same form as that of (10). Consequently, we can determine that the maximizing cross-correntropy between the desired signal and the system output leads to the same criterion as the MZEP, which has been induced in the process of mediating the conflict between the two opposite potentials in the MED criterion. Since the two criteria have the same form, from this point on in this paper, the MZEP approach will also be referred to as the MCC.

To prove that (10) has a global maximum when the error equals zero, $e = 0$, we need to show that non-zero error sample values result in a smaller value of (10), i.e.,

$$f_E(0) = \frac{1}{N} \sum_{i=k-N+1}^{k} G_\sigma(-e_i) \le G_\sigma(0) \quad (16)$$

or equivalently $\sum_{i=k-N+1}^{k} G_\sigma(-e_i) \le N G_\sigma(0)$.

Since the maximum value of the Gaussian kernel with zero mean is achieved when the error is zero, this inequality is readily satisfied. This proves that the supervised MZEP (MCC) criterion with the Parzen window preserves the global maximum.

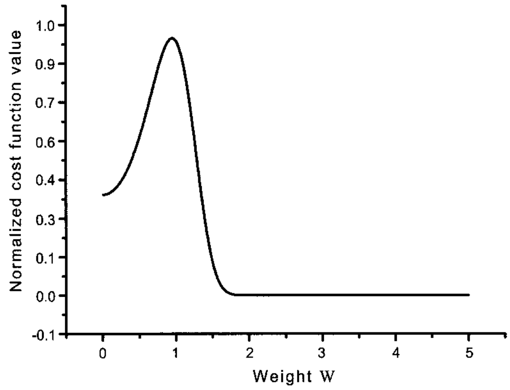

Fig. 1. Normalized cost function output as a function of equalizer weight.

In order to verify that (10) plays the role of a cost function in adaptive equalization and that it has a single maximum (global optimum) with respect to the equalizer weights, the value of the cost function is tested for a simple case with an all pass channel having a delta function $\delta(k)$ as its impulse response and an equalizer with a single weight. In Fig. 1, the normalized cost surface is plotted against various weight values. It shows that the supervised MZEP (MCC) cost function has a single maximum point. It is noticeable that the MEE criterion in (9) deals with $e_j - e_i$, whereas the MZEP (MCC) criterion in (12) deals only with e_i. Assuming e_i to be the zero mean and uncorrelated, the variance of $e_j - e_i$ becomes

$$\text{Var}_E(e_i - e_j) = E[(e_i - e_j)^2] = 2E[e_i^2] = 2\text{Var}_E(e_i). \quad (17)$$

This relationship indicates that the MEE or MED algorithms may suffer performance degradation when compared to the MZEP (MCC) algorithms in certain environments.

III. CMA AND NEW BLIND EQUALIZER ALGORITHMS BASED ON CME AND ITL CRITERIA

A. CMA

Many blind equalization algorithms employ nonlinearity at the equalizer output y_k in order to generate the error signal for weight updates. One of the well-known blind equalization algorithms is the constant modulus algorithm (CMA) that minimizes the CME $e_{\text{CME}} = |y_k|^2 - R_2$ on the basis of the MSE criterion [10]. The cost function P_{CMA} is $P_{\text{CMA}} = E[(|y_k|^2 - R_2)^2]$, where $|y_k|^2$ is the equalizer output power and $R_2 = E[|d_k|^4]/E[|d_k|^2]$. We assume that M-ary pulse amplitude modulation (PAM) signaling systems [11] are employed and all the M levels are equally likely to be transmitted a priori with a probability of $1/M$, and the transmitted levels A_m take the following discrete values

$$A_m = 2m - 1 - M, m = 1, 2, \cdots, M. \quad (18)$$

Then, the constant modulus R_2 becomes

$$R_2 = E[|A_m|^4]/E[|A_m|^2]. \quad (19)$$

The minimization of P_{CMA} with respect to the equalizer weights can be performed recursively by using the steepest descent method as follows $\mathbf{W}_{\text{new}} = \mathbf{W}_{\text{old}} - \mu_{\text{CMA}} \partial P_{\text{CMA}} / \partial \mathbf{W}$, where μ_{CMA} is the step-size parameter of the CMA. By differentiating P_{CMA} and dropping the expectation operation, we obtain the CMA for adjusting the blind equalizer weights: $\mathbf{W}_{k+1} = \mathbf{W}_k - 2\mu_{\text{CMA}} \mathbf{X}_k^* y_k (|y_k|^2 - R_2)$.

B. MCC-CME

Minimizing P_{CMA} is equivalent to minimizing only the variance (second order statistics) of the CME. As an alternative approach, we can adopt a new strategy that maximizes the probability that the CME becomes zero for blind channel equalization. For this purpose, by inserting the CME $e_{\text{CME}} = |y_k|^2 - R_2$ into (4) and using a block of past output samples $\{y_k, y_{k-1}, \cdots, y_{k-N+1}\}$, we have

$$f_E(e_{\text{CME}}) = \frac{1}{N} \sum_{i=k-N+1}^{k} G_\sigma(e_{\text{CME}} - [|y_i|^2 - R_2]) = \frac{1}{N} \sum_{i=k-N+1}^{k} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(e_{\text{CME}} - [|y_i|^2 - R_2])^2}{2\sigma^2}\right]. \tag{20}$$

Considering e_{CME} to be zero, the probability $f_E(e_{\text{CME}} = 0)$ reduces to

$$f_E(e_{\text{CME}} = 0) = \frac{1}{N} \sum_{i=k-N+1}^{k} G_\sigma(-[|y_i|^2 - R_2]). \tag{21}$$

We now derive a gradient ascent method for the maximization of the cost function (20).

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mu_{\text{MCC-CME}} \frac{2}{\sigma^2 N} \sum_{i=k-N+1}^{k} G_\sigma(|y_i|^2 - R_2) \cdot (R_2 - |y_i|^2) y_i \mathbf{X}_i^* \tag{22}$$

where $\mu_{\text{MCC-CME}}$ is the step-size for convergence control of the proposed blind algorithm. For convenience, this proposed blind algorithm will be referred to as MCC-CME.

The proposed MCC-CME is based on ITL, whereas the CMA is based on the MSE criterion, which contains only second-order statistics. The MSE criterion, therefore, would be able to extract all possible information from a signal whose statistics are only defined by its mean and variance. On the other hand, in the case of the proposed MCC-CME that uses the Gaussian kernel, all the moments of the PDF (not only the second moments) are constrained. Expanding the Gaussian kernel by using a Taylor series expansion, the criterion of MCC-CME can be rewritten as

$$f_E(e_{\text{CME}} = 0) = \frac{1}{N} \sum_{i=0}^{N-1} G_\sigma(-[|y_{k-i}|^2 - R_2]) = \frac{1}{N\sigma\sqrt{2\pi}} \sum_{i=0}^{N-1} \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n \sigma^{2n} n!} (|y_{k-i}|^2 - R_2)^{2n}. \tag{23}$$

This involves all the even-order moments of the error. Specifically, the term corresponding to $n = 1$ in (23) is $\frac{-1}{2\sigma^3\sqrt{2\pi}} \frac{1}{N} \sum_{i=0}^{N-1} (|y_{k-i}|^2 - R_2)^2 = \frac{-1}{2\sigma^3\sqrt{2\pi}} \text{Var}_{\text{CME}}$ where Var_{CME} is the variance of the CME, with the sample mean estimate used in place of the statistical expectation; this shows that the information provided by the second-order statistics is included within the MCC-CME criterion. Noting that this uses the higher-order information contained in the data, instead of only using the second-order information, as in the case of the MSE criterion, one can expect this criterion to provide more meaningful representations of the CME data, which may result in improved performance.

C. MEE-CME

The supervised MEE criterion in (8) deals with $e_j - e_i$. By replacing e_i with CME $|y_i|^2 - R_2$, the information potential that is based on the CME, IP_{CME}, becomes independent of the constant modulus R_2, as shown below. $IP_{\text{CME}} = N^{-2} \sum_{i=k-N+1}^{k} \sum_{j=k-N+1}^{k} G_{\sigma\sqrt{2}}(|y_i|^2 - |y_j|^2)$. To maximize the cost function (23), we adopt the gradient ascent method. The gradient is evaluated from

$$\frac{\partial IP_{\text{CME}}}{\partial \mathbf{W}} = \frac{1}{N^2\sigma^2} \sum_{i=k-N+1}^{k} \sum_{j=k-N+1}^{k} G_{\sigma\sqrt{2}}(|y_i|^2 - |y_j|^2) \cdot (|y_i|^2 - |y_j|^2)(y_j \mathbf{X}_j^* - y_i \mathbf{X}_i^*). \tag{24}$$

MEE-CME can be written using the gradient as follows. $\mathbf{W}_{k+1} = \mathbf{W}_k + \mu_{\text{MEE-CME}} \partial IP_{\text{CME}} / \partial \mathbf{W}$, where $\mu_{\text{MEE-CME}}$ is the step-size for MEE-CME.

IP_{CME} is maximized when the equalizer output powers are the same, i.e., $|y_i|^2 = |y_j|^2$. In binary modulation, each desired signal $d_i = \pm 1$ has the same absolute value. That is, the power of each desired signal has a common value $|d_i|^2 = 1$. This can be viewed as an attempt by the equalizer to cluster the outputs such that they have their desired power values. However, in M-ary modulation schemes, the power of each desired signal has different values. The force induced from maximizing IP_{CME} will lose its target direction, because the cost function forces the equalizer outputs to have the same output power $|y_i|^2 = |y_j|^2$, in opposition to their desire to have different powers. Consequently, MEE-CME loses the information of the constant modulus R_2. This may cause MEE-CME to not converge or to converge slower than other algorithms, depending son the constant modulus R_2.

D. MED-CME

On the other hand, minimization of the Euclidean distance $ED[f_E(e_{\text{CME}}), \delta(e_{\text{CME}})]$ between the two PDFs around the CME signals $e_{\text{CME}} = |y_k|^2 - R_2$ results in a match between $f_E(e_{\text{CME}})$ and the Dirac-delta function. This process tries to create a concentration of CME samples near zero. The process of minimizing $ED[f_E(e_{\text{CME}}), \delta(e_{\text{CME}})]$, also involves two different simultaneous forces: minimization of IP_{CME} and maximization of $2f_E(0)$. The cost function ED_{CME} for MED-CME can be expressed as

$$ED_{\text{CME}} = IP_{\text{CME}} - 2f_E(e_{\text{CME}} = 0). \tag{25}$$

Minimization of the cost function ED_{CME} leads to the following algorithm (this is referred to as MED-CMA in this pa-

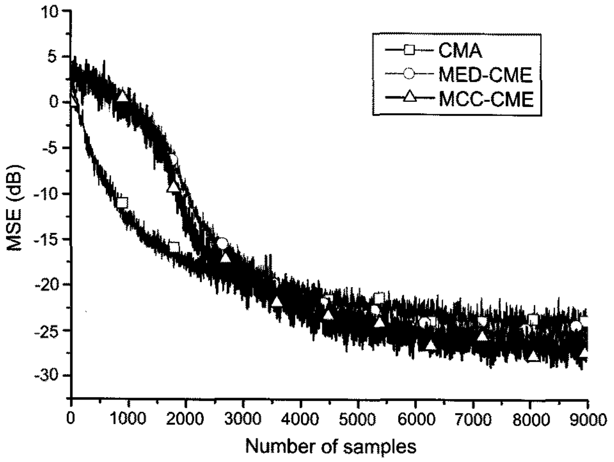

Fig. 2. MSE convergence performance in CH1.

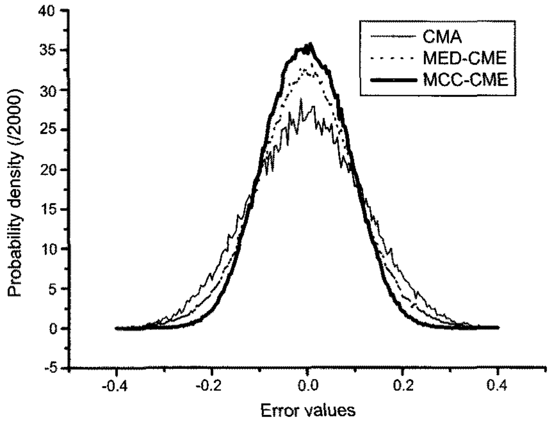

Fig. 3. Probability density for errors in CH1.

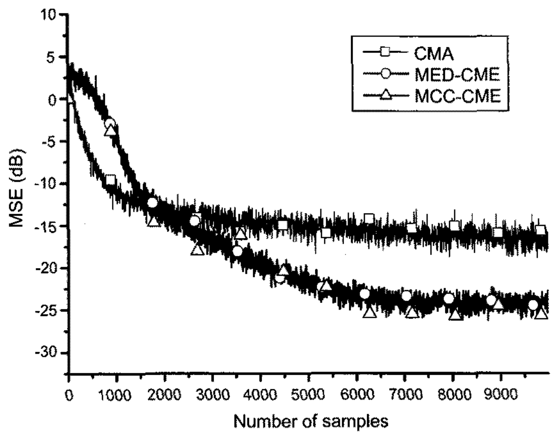

Fig. 4. MSE convergence performance in CH2.

per).

$$\mathbf{W}_{k+1} = \mathbf{W}_k - \mu_{\text{MED-CME}} \frac{1}{N^2\sigma^2} \cdot \left[\sum_{i=k-N+1}^{k} \sum_{j=k-N+1}^{k} G_{\sigma\sqrt{2}}(|y_i|^2 - |y_j|^2) \cdot (|y_i|^2 - |y_j|^2)(y_j \mathbf{X}_j^* - y_i \mathbf{X}_i^*) - 2N \sum_{i=k-N+1}^{k} G_\sigma(|y_i|^2 - R_2)(R_2 - |y_i|^2) y_i \mathbf{X}_i^* \right]. \tag{26}$$

From the aspect of implementation, MEE-CME in (24) and MED-CME in (26) are computationally cumbersome because of the complexity of $O(N^2)$, whereas the implementation of MCC-CME in (22) requires only $O(N)$. From the aspect of accuracy, there is a difference in the kernel size σ between MCC-CME and MED-CME. The kernel size in the information potential IP_{CME} of MED-CME is $\sigma\sqrt{2}$, whereas the kernel size in $f_E(e_{\text{CME}} = 0)$ of MCC-CME is σ. The kernel size usually determines the accuracy of the solution [12]. A small kernel size implies a small extent of overlapping in Parzen PDF estimation, which, in turn, implies that the desired solution is very near the optimum point. From this point of view, MCC-CME can be expected to show a higher performance when compared to MED-CME or MEE-CME.

IV. RESULTS AND DISCUSSION

In this section, we present and discuss the simulation results that illustrate the comparative performance of the proposed MCC-CME and MED-CME versus the CMA for blind equalization. They are studied for the three channel models in [11]. The transfer functions of each of the channel models are

$$\begin{aligned} &\text{CH1: } H_1(z) = 0.26 + 0.93z^{-1} + 0.26z^{-2}, \\ &\text{CH2: } H_2(z) = 0.304 + 0.903z^{-1} + 0.304z^{-2}, \\ &\text{CH3: } H_3(z) = 0.389 + 0.835z^{-1} + 0.389z^{-2}. \end{aligned} \tag{27}$$

These channel models are typical multipath channel models, and they result in severe inter-symbol interference. Especially in terms of spectral characteristics, the channel model 3, CH3, shows the worst spectral nulls.

The number of weights in the linear TDL equalizer structure is set to 11. For the worst channel, CH3, the number of weights is 31. The channel noise for MSE convergence performance is zero mean white Gaussian with a variance of 0.001. As measures of equalizer performance, we use MSE convergence, probability densities for errors and error rate versus signal to noise ratio (SNR).

The four-level ($M = 4$) random signal $\{-3, -1, 1, 3\}$ is transmitted to the channel. The convergence parameters of the CMA that show the least steady-state MSE are 0.00001, 0.000005, and 0.0000007 for CH1, CH2, and CH3, respectively. We use a common data-block size of $N = 20$ for the ITL-type blind algorithms. For MED-CME, a kernel size of $\sigma = 6.0$ and the convergence parameter $\mu_{\text{MED-CME}} = 0.03$ are used. For MCC-CME, a kernel size of $\sigma = 6.0$ and the convergence parameter $\mu_{\text{MCC-CME}} = 0.03$ are used. The parameters for both the ITL algorithms are common for all the three channel models (CH1, CH2, and CH3).

As discussed previously, MEE-CME loses the information of the constant modulus R_2. In all the three channel models, MEE-CME showed ill-convergence. The MSE convergence perfor-

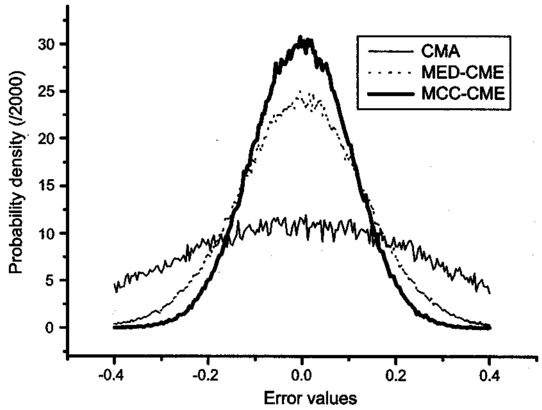

Fig. 5. Probability density for errors in CH2.

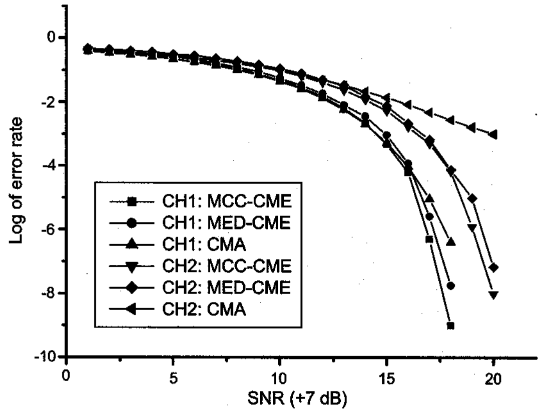

Fig. 6. Error-rate performance comparison for CH1 and CH2.

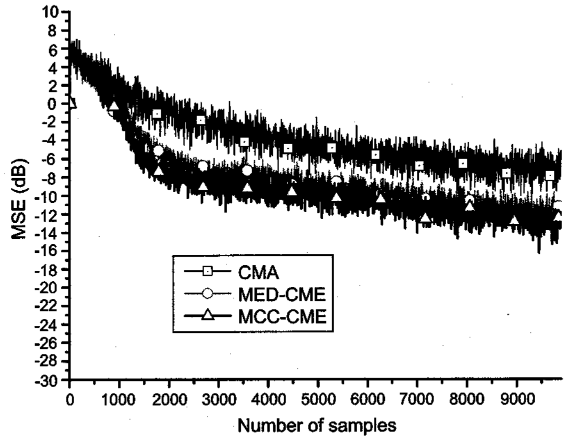

Fig. 7. MSE convergence performance in CH3.

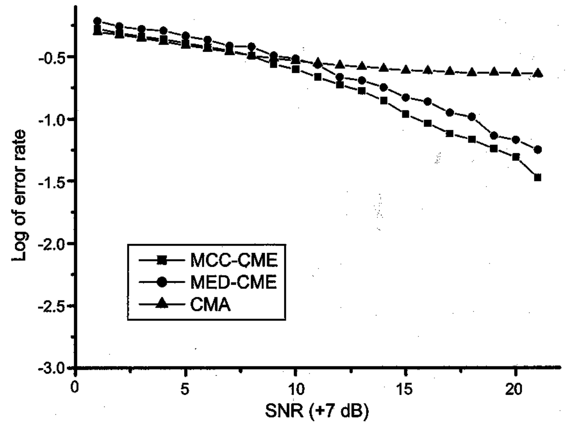

Fig. 8. Error-rate performance comparison for CH3.

mance, error-PDF, and BER performance for CH1 are shown in Figs. 2–4, respectively. The MSE performance in Fig. 2 shows that the proposed MCC-CME and MED-CME have a slightly enhanced performance when compared to the CMA. From the error PDF estimates in Fig. 3, the error distribution of MCC-CME is shown to be more concentrated around zero. In the channel model, CH2, the CMA shows a significant performance degradation in Fig. 4. On the other hand, the steady-state error-performance of MCC-CME and MED-CME is similar to that in CH1. From these results, the ITL-type algorithms can be considered to be relatively insensitive to channel variations when compared to the CMA based on the MSE criterion. Fig. 5 depicts the estimated probability densities of the algorithms in CH2.

The differences in the performance of the algorithms in CH2 are more prominent. The error values of the CMA do not appear to be concentrated around zero, whereas the distribution from MCC-CME and MED-CME is still concentrated around zero. In particular, MCC-CME shows a superior error-PDF performance. In order to show the merits of the blind algorithms that are based on the CME and ITL, we compare and present the error rate performance for CH1 and CH2 in Fig. 6.

In terms of the error rate performance for CH1, MCC-CME shows a performance enhancement of 1 dB when compared to MED-CME and a performance enhancement of approximately 2 dB when compared to the CMA for an error rate of 10^{-6}. In the case of CH2, the CMA shows a significant degradation in error rate performance. However, MCC-CME and MED-CME still show a good error rate performance and the difference of 1 dB in performance is still maintained. For the worst channel model, CH3, which exhibits spectral nulls, the CMA shows unsatisfying performance and the minimum MSE of the algorithm does not fall below -6 dB in Fig. 7. On the other hand, the MSE of MED-CME converges to -12 dB and the MSE of the proposed MCC-CME reaches to less than -14 dB. In Fig. 8, the error rate performance for CH3 is compared, and the error rate of the CMA is found to be above 0.2. However, MED-CME and MCC-CME show enhanced error performance, their error rates being less than 0.05 and 0.03, respectively.

In order to investigate the effect of the kernel size on the equalizer performance, the minimum MSE is plotted for different kernel sizes ranging from 0.5 to 15 for CH2 in Fig. 9. The figure shows that the kernel size chosen for our simulation (6.0) corresponds to the lowest steady-state MSE performance.

V. CONCLUSION

In this paper, a new approach to blind equalization that is based on maximizing the concentration of CMEs near zero is proposed. The derivation of the proposed method is different from that of the MCC criterion that maximizes the correntropy between desired symbols and output symbols. The cost func-

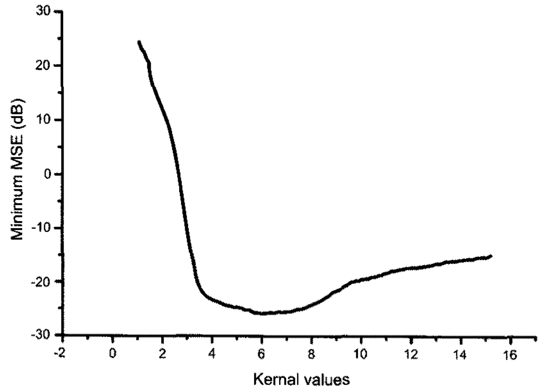

Fig. 9. Minimum MSE versus kernel sizes for the proposed method.

tion of the proposed algorithm is to maximize the probability that the equalizer output power is equal to the constant modulus of the transmitted symbols. Two blind ITL algorithms based on CME signals are also introduced: one for minimizing the Euclidean PDF distance (MED-CME) and the other for minimizing the CME entropy. It is found that the method of minimizing the error entropy based on the CME (MEE-CME) loses the constant modulus information and shows ill-convergence in the simulation environments. This result indicates that MEE-CME cannot be used in blind equalization because of the absence of the constant modulus information.

For a data block size of N in the Parzen window method, MEE-CME and MED-CME are computationally cumbersome because of the $O(N^2)$ complexity; however, MCC-CME requires only $O(N)$. In the case of the kernel size, it is found that the kernel size in MED-CME is $\sigma\sqrt{2}$, whereas the kernel size in MCC-CME is σ. Due partly to the difference in the kernel sizes, MCC-CME shows enhanced performance in the simulation. Further, MCC-CME and MED-CME, which are both based on ITL, utilize higher order information contained in the data, instead of only using second-order information, as in the case of the MSE criterion. This results in improved simulation performance in terms of convergence speed, error distribution, and error rate versus SNR curves when compared to the CMA, which is based on MSE. These results indicate that the proposed MCC-CME is a successful candidate for use in blind equalization systems.

REFERENCES

[1] J. C. Principe, D. Xu and J. Fisher, *Information Theoretic Learning in: S. Haykin (Ed.), Unsupervised Adaptive Filtering*. Wiley, New York, vol. I, 2000, pp. 265–319.

[2] D. Erdogmus and J. C. Principe, "An entropy minimization algorithm for supervised training of nonlinear systems," *IEEE Trans. Signal Process.*, vol. 50, pp. 1780–1786, July 2002.

[3] I. Santamaria, D. Erdogmus, and J. C. Principe, "Entropy minimization for supervised digital communications channel equalization," *IEEE Trans. Signal Process.*, vol. 50, pp. 1184–1192, May 2002.

[4] Ching-an Lai, *Global Optimization Algorithms for Adaptive Infinite Impulse Response Filters*, Ph. D. thesis, Dec. 2002, University of Florida.

[5] M. Lazaro, I. Santamaria, D. Erdogmus, K. Hild, C. Pantaleon, and J. C. Principe, "Stochastic blind equalization based on PDF fitting using parzen estimator," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 696–704, Feb. 2005.

[6] I. Santamaria, C. Pantaleon, L. Vielva, and J. C. Principe, "A fast algorithm for adaptive blind equalization using order-alpha Renyi entropy," in *Proc. IEEE ICASSP*, Orlando, USA, 2002, pp. 2657–2660.

[7] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: Properties and applications in non-Gaussian signal processing," *IEEE Trans. Signal Process.*, vol. 55, no. 11, Nov. 2007, pp. 5286–5298.

[8] W. Liu, P. P. Pokharel, and J. C. Principe, "Error entropy, correntropy, metric, and M-estimation," in *Proc. IEEE MLSP*, Maynooth, Ireland, 2006, pp. 179–184.

[9] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer Verlag, New York: NY, 1995.

[10] J. R. Treichler and B. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-31, pp. 349–372, Nov. 1983.

[11] J. G. Proakis, *Digital Communications*. 2nd ed., McGraw-Hill, 1989.

[12] L. Wasserman, *All of Statistics: A Concise Course in Statistical Inference*. Springer, 2004.

Namyong Kim received the B.S., M.S., and Ph.D degrees from Yonsei University, all in electronic engineering in 1986, 1988 and 1991, respectively. From 1992 to 1997, he was with Kwandong University, Korea. In 1998–2002, he was an Associate Professor in the Department of Information and Communications Engineering at Kangwon National University and currently he serves as a Professor. His current research interests are in adaptive signal processing in mobile communications and odor sensing-identification systems.

Kyu-Hwa Jeong received the Ph.D. degree in Electrical and Computer Engineering from the University of Florida, Gainesville, USA in 2007 and the M.S. degree in Electronics Engineering from Yonsei University, Seoul, Korea, in 1997. In 1997–2003, he was a Senior Research Engineer with Digital Media Research Lab. in LG Electronics, Seoul, Korea. In 2004–2007, he was a Research Assistant with the Computational Neuro-Engineering Lab (CNEL) in the University of Florida and pursued the Ph.D. degree. In 2007, he joined Intel Mask Operation, Santa Clara, CA, where he is currently a Process and Automation Engineer. His research interests are in the field of signal processing, machine learning, and its applications to image analysis and pattern recognition.

Liuqing Yang received her B.S. degree in Electrical Engineering from Huazhong University of Science and Technology, Wuhan, China, in 1994, and her M.Sc. and Ph.D. degrees in the Electrical and Computer Engineering from University of Minnesota, Minneapolis, in 2002 and 2004, respectively. Since August 2004, she has been with the Department of Electrical and Computer Engineering at University of Florida, Gainesville and currently as an Associate Professor. Her current research interests include signal processing, communications theory, networking, and UWB systems. She serves as an Associate Editor of the IEEE Transactions on Communications, IEEE Transactions on Wireless Communications, and IEEE Transactions on Intelligent Transportation Systems.