Euclidian Distance Minimization of Probability Density Functions for Blind Equalization

Namyong Kim

Abstract: **Blind equalization techniques have been used in broadcast and multipoint communications. In this paper, two criteria of minimizing Euclidian distance between two probability density functions (PDFs) for adaptive blind equalizers are presented. For PDF calculation, Parzen window estimator is used. One criterion is to use a set of randomly generated desired symbols at the receiver so that PDF of the generated symbols matches that of the transmitted symbols. The second method is to use a set of Dirac delta functions in place of the PDF of the transmitted symbols. From the simulation results, the proposed methods significantly outperform the constant modulus algorithm in multipath channel environments.**

Index Terms: **Blind equalizer, Dirac delta, Euclidian distance function, information theoretic learning (ITL), Parzen window, probability density function (PDF).**

I. INTRODUCTION

Multipoint communication has been an increasingly focused topic in broadcasting systems, computer communication networks, and the wireless/mobile networks [1]. In applications such as broadcast and multipoint networks, blind equalizers to counteract multipath effects are very useful since they do not require a training sequence to start up or to restart after a communications breakdown [2], [3].

Problems involving the training of adaptive equalizers have been solved usually through the use of mean squared error (MSE) criterion. As another way for solving these problems, information-theoretic learning (ITL) has been introduced by Princepe [4]. Unlike the MSE criterion that utilizes error energy, ITL algorithms are based on a combination of a nonparametric probability density function (PDF) estimator and a procedure to compute entropy or information potential (IP).

As an application of ITL, Erdogmus *et al.* introduced an information theoretic approach based on Kullback-Leibler (KL) divergence [5] minimization for training adaptive systems in supervised learning settings using both labeled and unlabeled data [6]. The KL divergence is a way to estimate mutual information which is capable of quantifying the entropy between pairs of random variables. But, it is not quadratic in the PDFs, so it can not be easily integrated with the information potential.

Another measure of merit which can be expressed as information potential is the minimum error entropy (MEE) criterion that has been compared in terms of the error distributions [4]. The combination of Renyi's quadratic entropy with the Parzen window using a Gaussian kernel [7] leads to an estimation of entropy or information potential by computing interactions among pairs of output samples which is a practical cost function for ITL. In their work, it has also been shown that Renyi's quadratic entropy expression with Parzen PDF estimator is negatively proportional to the logarithmic value of the information potential of error samples. Since logarithm is a monotonic function, information potential is maximized in MEE instead of minimizing Renyi's entropy. Therefore, MEE criterion can be considered as maximization of information potential.

The study demonstrated that the error samples of the entropy-trained systems exhibit a more concentrated density function and the distribution of the produced outputs is also closer to that of the desired signals compared to MSE. So, MEE criterion can be a promising alternative to MSE in supervised channel equalization applications [8].

However, MEE criterion has been born as a measure for only supervised learning. In its final cost function, MEE criterion is a nonlinear function of the difference between two error samples. For blind signal processing, the error values should be replaced with constant modulus errors as in constant modulus algorithm (CMA) [9]. Then, the MEE cost function using constant modulus error becomes independent of the constant modulus $R_2 = E[|A_m|^4]/E[|A_m|^2]$, where A_m is the transmitted symbol level and $E[\cdot]$ is the expectation operator. As a result, the MEE cost function forces equalizer output powers to have the same value. In M-ary modulation schemes, the power of each desired signal has different values. The force induced from maximizing the MEE cost function will lose its target direction because the cost function forces the equalizer outputs to obtain the same output power in spite of different desired powers. Consequently, MEE cost function loses the information of the constant modulus R_2. This may lead the MEE cost function to ill-convergence.

As another measure for unsupervised adaptive signal processing based on ITL, the Euclidian distance (ED) between two PDFs contains the desired information within the desired PDF not like in the MEE measure. Also the ED criterion contains only quadratic terms to be utilized very easily. Furthermore, whenever we have the shape information of desired PDF, we can construct the desired PDF in various ways so that we can apply the ED criterion to any adaptive signal processing applications aimed at obtaining desired output. These characteristics can be the most important advantage of using ED criterion. Recently, the ED-based PDF matching algorithm was introduced by Jeong *et al.* and applied successfully to the classification problem with a real biomedical data set [10]. In that method, the authors proposed to reuse the previously acquired training-phase output samples in the test phase so that the test-phase output PDF follows the training-phase output PDF.

In this paper, based on the criterion of ED minimization of

Manuscript received March 04, 2009; approved for publication by Huaiyu Dai, Division I Editor, June 02, 2010.

N. Kim is with the Dept. of Information & Communication Eng., Kangwon National University, email: namyong@kangwon.ac.kr.

1229-2370/10/$10.00 © 2010 KICS

two PDFs, we investigate not only the interactions between output samples and randomly generated desired samples but also the interactions between desired level values and output samples for application to blind equalization. In our previous work [11] for blind equalization, we created desired symbols for the desired PDF construction by utilizing the PDF information of the transmitted symbols. The idea was to generate, at the receiver, random symbol values that have the same PDF shape as the transmitted symbol PDF shape. This algorithm uses the Parzen window method for the desired PDF construction. On the other hand, in this work, the Parzen window method is no longer used in the desired PDF construction. That is, instead of using self-generated samples, the desired PDF of the proposed cost function is created only with a set of Dirac-delta functions that complies well with the modulation schemes.

This paper is organized as follows. Section II presents the baseband pulse amplitude modulation (PAM) data transmission system model and blind equalizer structure considered. In Section III, we briefly describe the constant modulus blind equalizer algorithm which is based on MSE criterion. In Section IV, the first ED minimization method using randomly generated symbols at the receiver is introduced. The second method for blind equalization using a set of Dirac delta functions is proposed in Section V and the robustness of the proposed method is analyzed in Section VI. Section VII reports simulation results and discussions. Finally, concluding remarks are presented in Section VIII.

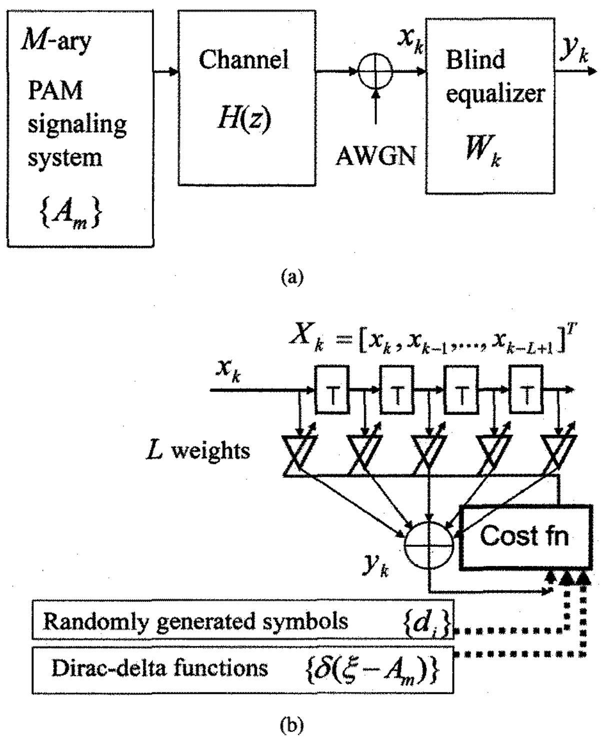

Fig. 1. (a) Simplified baseband PAM communication system and (b) blind equalizer structure.

II. SYSTEM MODEL

For simplicity, we consider the baseband-equivalent data transmission system of Fig. 1 with transmitted symbol set $\{A_m\}$, multi-path channel $H(z)$ with the impulse response h_i, received signal x_k at time k and the equalizer output y_k. The linear tapped delay line (TDL) equalizer structure is employed with L weights. For the weight adaptation without the aid of a training sequence d_k, blind algorithms are used according to proposed cost functions. CMA uses output y_k and R_2 based on MSE criterion. Fig. 1 depicts briefly what the two different PDF construction methods utilize for ED minimization. Our initial algorithm uses N random symbol values of d_i generated at the receiver, and the algorithm proposed in this paper uses a set of Dirac-delta functions for the desired PDF construction.

III. CMA BASED ON MSE CRITERION

Channel equalization without the aid of a training sequence is referred to as blind channel equalization. Unlike traditional trained equalization algorithms, many of the widely employed blind equalization algorithms employ nonlinearity at the equalizer output to generate the error signal for weight updates based on MSE criterion. One of the well known blind equalization algorithms based on MSE criterion is CMA which minimizes the following cost function [9]

$$P_{\text{CMA}} = E\left[\left(|y_k|^2 - R_2\right)^2\right]. \tag{1}$$

The minimization of P_{CMA} with respect to the equalizer coefficients can be performed recursively according to the steepest descent method

$$\mathbf{W}_{\text{new}} = \mathbf{W}_{\text{old}} - \mu_{\text{CMA}} \frac{\partial P_{\text{CMA}}}{\partial \mathbf{W}} \tag{2}$$

where μ_{CMA} is the step-size parameter for CMA.

In case of on-line linear equalization, a TDL can be used for input $\mathbf{X}_k = [x_k, x_{k-1}, \cdots, x_{k-L+1}]^T$ and output $y_k = \mathbf{W}_k^T \mathbf{X}_k$ at time k. By differentiating P_{CMA} and dropping the expectation operation, we obtain the following least mean square (LMS)-type algorithm for adjusting the blind equalizer coefficients

$$\mathbf{W}_{k+1} = \mathbf{W}_k - 2\mu_{\text{CMA}} \mathbf{X}_k^* y_k (|y_k|^2 - R_2). \tag{3}$$

We assume that M-ary PAM signaling systems are employed and all M levels are equally likely to be transmitted a priori with a probability $1/M$, and the transmitted levels A_m takes the following discrete values:

$$A_m = 2m - 1 - M, \ m = 1, 2, \cdots, M. \tag{4}$$

Then, the constant modulus R_2 becomes

$$R_2 = \frac{E\left[|A_m|^4\right]}{E\left[|A_m|^2\right]}. \tag{5}$$

IV. BLIND EQUALIZATION BASED ON ED MINIMIZATION USING RANDOMLY GENERATED SYMBOLS AT THE RECEIVER

The ED between the desired symbol PDF f_d and the equalizer output PDF f_y can be minimized with respect to weight $\mathbf{W}$ as

$$\min_{\mathbf{W}}(ED[f_d, f_y]) = \min_{\mathbf{W}}\left(\int f_d^2(\xi)d\xi + \int f_y^2(\xi)d\xi - 2\int f_d(\xi)f_y(\xi)d\xi\right). \quad (6)$$

If the two distributions are close to each other, the ED cost function (6) minimizes the divergence between the desired symbols and equalizer output samples. In other words, we create desired symbols for the equalizer input signal by utilizing the PDF information of the transmitted symbols. Our initial idea was to generate, at the receiver, random symbols that have the same PDF of the transmitted symbols [10], which is introduced as follows.

Given a set of randomly generated N symbols $D_N = \{d_1, d_2, \cdots, d_N\}$, the PDF based on Parzen window method can be approximated by

$$\begin{aligned} f_d(\xi) &= \frac{1}{N}\sum_{i=1}^{N} G_\sigma(\xi - d_i) \\ &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma\sqrt{2\pi}}\exp\left[\frac{-(\xi - d_i)^2}{2\sigma^2}\right]. \end{aligned} \quad (7)$$

Under the assumption that all M levels are equally likely, the number of random symbols corresponding to each level A_m is N/M.

The point noticeable here is that the random symbols used for Parzen PDF calculation have the same PDF pattern as the transmitted symbols, but the symbols are randomly generated ones at the receiver, not the exact training symbols. This approach makes blind equalization possible. Then, the integrals of the multiplication of two PDFs in (6) become

$$\int f_d^2(\xi)d\xi = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} G_{\sigma\sqrt{2}}(d_j - d_i), \quad (8)$$

$$\int f_y^2(\xi)d\xi = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} G_{\sigma\sqrt{2}}(y_j - y_i), \quad (9)$$

$$\int f_d(\xi)f_y(\xi)d\xi = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} G_{\sigma\sqrt{2}}(d_j - y_i). \quad (10)$$

Equation (8) can be the IP of randomly generated symbols, which is denoted as $IP_1(d,d)$ in this paper, where the subscript 1 indicates the method 1. We note that (8) is not a function of weight. By summing the interactions among pairs of output samples we can obtain the $IP_1(y,y)$ as in (9). Equation (10) defined as $IP_1(d,y)$ indicates the interactions between the two different variables d and y. So, the cost function (6) can be reduced as P_1 in (11).

$$P_1 = IP_1(y,y) - 2IP_1(d,y). \quad (11)$$

Now a gradient descent method can be applied for the minimization of the cost function (11) with respect to equalizer weight as follows

$$\mathbf{W}_{\text{new}} = \mathbf{W}_{\text{old}} - \mu_1\frac{\partial P_1}{\partial \mathbf{W}}. \quad (12)$$

The gradient is evaluated from

$$\begin{aligned} \frac{\partial P_1}{\partial \mathbf{W}_k} = &\frac{1}{2N^2\sigma^2}\sum_{i=k-N+1,}^{k}\sum_{j=k-N+1}^{k}(y_j - y_i) \\ &G_{\sigma\sqrt{2}}(y_j - y_i)(\mathbf{X}_i - \mathbf{X}_j) \\ &-\frac{1}{N^2\sigma^2}\sum_{i=k-N+1,}^{k}\sum_{j=1}^{N}(d_j - y_i) \\ &G_{\sigma\sqrt{2}}(d_j - y_i)\mathbf{X}_i. \end{aligned} \quad (13)$$

For convenience sake, this method 1 shall be referred to here as minimum ED 1 (MED1) algorithm.

Normally, modulation schemes are known to receivers. Furthermore, most transmitters use independent and identically distributed symbols. Under these considerations, we propose a new method of using a set of Dirac delta functions described in the following section.

V. BLIND EQUALIZATION USING A SET OF DIRAC DELTA FUNCTIONS

According to our assumption that all the M levels of M-ary PAM signaling systems are equally likely with a probability $1/M$, we can construct PDF of the desired symbols without knowing the exact training symbols as

$$\begin{aligned} f_d(\xi) = &\frac{1}{M}[\delta(\xi - A_1) + \delta(\xi - A_2) + \cdots \\ &+ \delta(\xi - A_m) + \cdots + \delta(\xi - A_M)]. \end{aligned} \quad (14)$$

Substituting (14) into (6), the information potential $IP_2(d,d)$ can be expressed as

$$\begin{aligned} IP_2(d,d) &= \int f_d^2(\xi)d\xi \\ &= \frac{1}{M^2}\sum_{m=1}^{M}\sum_{l=1}^{M}\int \delta(\xi - A_l)\delta(\xi - A_m)d\xi \\ &= \frac{1}{M}. \end{aligned} \quad (15)$$

Accordingly, the information potential $IP_2(d,y)$ becomes

$$IP_2(d,y) = \int f_d(\xi)f_y(\xi)d\xi = \frac{1}{M}\sum_{m=1}^{M} f_y(A_m). \quad (16)$$

By using Parzen window method (7) in (16), we have

$$IP_2(d,y) = \frac{1}{M}\frac{1}{N}\sum_{m=1}^{M}\sum_{i=1}^{N} G_\sigma(A_m - y_i). \quad (17)$$

Since (9) and (17) which contain system output are a function of weight but (15) is not a function of weight, ED minimization (6) can reduce to the following criterion

$$\min_{\mathbf{W}}[IP_1(y,y) - 2IP_2(d,y)]. \tag{18}$$

With the cost function P_2 defined as

$$P_2 = IP_1(y,y) - 2IP_2(d,y),$$

the gradient is calculated from

$$\frac{\partial P_2}{\partial \mathbf{W}_k} = \frac{1}{2N^2\sigma^2} \sum_{i=k-N+1,}^{k} \sum_{j=k-N+1}^{k} (y_j - y_i) G_{\sigma\sqrt{2}}(y_j - y_i)(\mathbf{X}_i - \mathbf{X}_j) - \frac{2}{MN\sigma^2} \sum_{i=k-N+1,}^{k} \sum_{m=1}^{M} (A_m - y_i) G_\sigma(A_m - y_i)\mathbf{X}_i. \tag{19}$$

Finally, the weight update of the method 2 (MED2) can be obtained as

$$\mathbf{W}_{\text{new}} = \mathbf{W}_{\text{old}} - \mu_2 \frac{\partial P_2}{\partial \mathbf{W}}. \tag{20}$$

Comparing (14) and (7), we can notice the essential relationship between MED1 and MED2. When we let the size σ of the Gaussian kernel in (7) go to zero, the Gaussian function $G_\sigma(\xi - d_i)$ becomes a Dirac-delta function $\delta(\xi - d_i)$. This observation leads us to finding out the relationship between MED1 and MED2 that MSE2 can be considered as an asymptotic case of MED1.

VI. ANALYSIS OF ROBUSTNESS OF THE PROPOSED METHOD OVER CMA

MSE criterion contains only second-order statistics defined by its mean and variance. When ITL learning criteria are considered, more information is utilized [12]. Using a Taylor series expansion for the Gaussian kernel, $IP_2(d,y)$ of MED2 criterion can be rewritten as

$$IP_2(d,y) = \frac{1}{MN} \sum_{m=1}^{M} \sum_{i=1}^{N} G_\sigma(A_m - y_i) = \frac{1}{MN\sigma\sqrt{2\pi}} \sum_{m=1}^{M} \sum_{i=1}^{N} \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n \sigma^{2n} n!} (A_m - y_i)^{2n}. \tag{21}$$

We can notice from (21) that all even moments of the error, not only the second moments, are constrained. On the other hand, the constant modulus algorithm is based on the MSE criterion which contains only second-order statistics. MSE criterion, therefore, would be able to extract all possible information from a signal whose statistics are only defined by its mean and variance. This suggests that the proposed method using the Gaussian kernel exploits more information than MSE criterion, on which CMA is based.

To provide some more insight on the soundness of the proposed method, we rewrite the term $IP_2(d,y)$ as a set of partitioned functions. Considering 4PAM signaling, as used in our simulation in Section VII, the set of outputs $y_i$can be partitioned in accordance with the transmitted symbol set $A_m = \{\pm 3, \pm 1\}$ into four subsets as

$$R^{(+3)} = \{y_i, A_m = 3\}, \tag{22}$$

$$R^{(+1)} = \{y_i, A_m = 1\}, \tag{23}$$

$$R^{(-1)} = \{y_i, A_m = -1\}, \tag{24}$$

$$R^{(-3)} = \{y_i, A_m = -3\}. \tag{25}$$

Then, the information potential $IP_2(d,y)$ can be expressed as

$$IP_2(A,y) = \sum_{i\in R^{(+3)}} G_\sigma(3 - y_i) + \sum_{i\in R^{(+1)}} G_\sigma(1 - y_i) + \sum_{i\in R^{(-1)}} G_\sigma(-1 - y_i) + \sum_{i\in R^{(-3)}} G_\sigma(-3 - y_i). \tag{26}$$

Noticing that $IP_2(d,y)$ is maximized in (18), each term in (26) is maximized when $y_i = 3$ for $i \in R^{(+3)}$, $y_i = 1$ for $i \in R^{(+1)}$, $y_i = -1$ for $i \in R^{(-1)}$, and $y_i = -3$ for $i \in R^{(-3)}$, respectively. This process can be viewed that the criterion forces the output signal to have correct symbol values.

On the other hand, the CMA cost function (1) can be partitioned using a sample mean estimator as

$$P_{\text{CMA}} = \sum_{i\in R^{(+3)}} (R_2 - y_i^2)^2 + \sum_{i\in R^{(+1)}} (R_2 - y_i^2)^2 + \sum_{i\in R^{(-1)}} (R_2 - y_i^2)^2 + \sum_{i\in R^{(-3)}} (R_2 - y_i^2)^2 \tag{27}$$

where $R_2 = E[|A_m|^4]/E[|A_m|^2] = 8.2$. Similar to the process in (26), each term in (27) is minimized when $y_i^2 = 8.2$ for all symbol regions: $i \in R^{(+3)}$, $i \in R^{(+1)}$, $i \in R^{(-1)}$, and $i \in R^{(-3)}$. This implies that the cost function of CMA pushes output samples to have a constant power 8.2 regardless of symbol classes. This analysis suggests that the proposed blind approach uses more rigorous constraints than CMA in order to produce output samples closer to their correct symbols.

Performance difference between MED2 and MED1 in PAM systems is considered due largely to the difference of the kernel size σ between the two algorithms. The kernel size in $IP_1(d,y)$ of MED1 is $\sigma\sqrt{2}$, whereas the kernel size in $IP_2(d,y)$ of MED2 is σ. The kernel size determines the accuracy of the solution [13]. A small kernel size means a small amount of overlapping in Parzen PDF estimation and the desired solution is very near the minimum of P_2.

Though MED2 has enhanced performance, it is not readily applicable to other adaptive systems in which desired PDF is not expressed as a set of Dirac delta functions. On the other hand, MED1 can be used in any adaptive systems where random symbols can be generated so as to conform to the desired PDF.

Thus, we are lead to believe that our approaches in this blind equalizer application using the information theoretic learning produce superior performance to CMA based on MSE criterion.

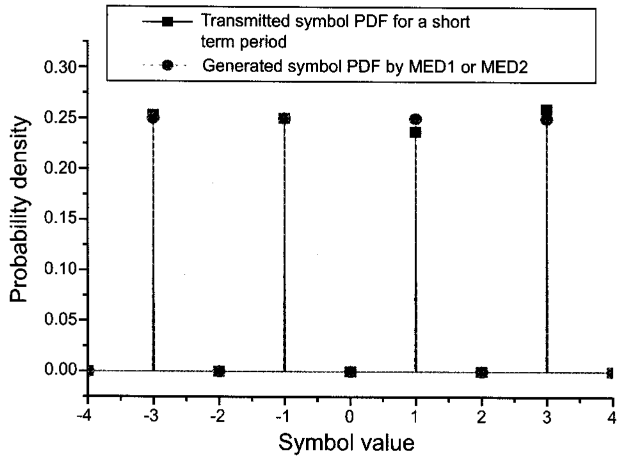

Fig. 2. Probability density distributions of generated symbols and transmitted symbols for a short term period.

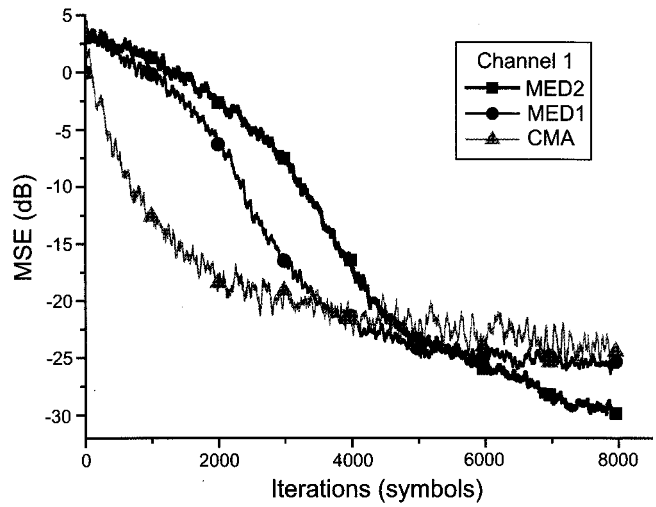

Fig. 3. MSE convergence comparison for channel 1.

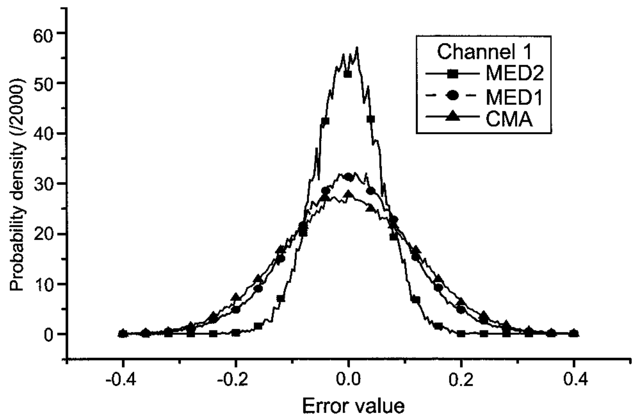

Fig. 4. Probability density for errors for channel 1.

VII. RESULTS AND DISCUSSION

In this section, we present and discuss simulation results that illustrate the comparative performance of the proposed MED algorithms versus CMA in blind equalization for two linear channels. The 4 level random symbols $\{\pm 3, \pm 1\}$ are transmitted through a channel and the impulse response, h_i of the channel model $H(z)$ [14] is dependent on the channel parameter BW as $h_i = 1/2\{1 + \cos[2\pi(i-2)/BW]\}$, $i = 1, 2, 3$. The parameter BW determines the channel bandwidth and controls the eigenvalue spread ratio (ESR) of the correlation matrix of the inputs in the equalizer. The number of weights in the linear TDL equalizer structure is set to $L = 11$. The channel noise is additive zero mean white Gaussian (AWGN) with a variance of 0.001. As measures of equalizer performance, we use probability densities for errors at that noise variance of CMA, MED1, and MED2, and then bit error rate (BER) versus Eb/No curves of CMA, MED2, and a training-aided algorithm (LMS) with $L = 11$. The convergence parameters for CMA are 0.00001 and 0.0000005 for channel 1 ($BW = 3.1$, ESR $= 11.12$) and channel 2 ($BW = 3.3$, ESR $= 21.71$), respectively. For MED1 and MED2, we set the data-block size $N = 20$, and the convergence parameter $\mu_1 = \mu_2 = 0.005$. The kernel size is chosen as $\sigma = 0.6$ for MED1 and MED2. All the convergence parameters are obtained when the algorithms show the lowest steady-state error performance, and proper kernel sizes are chosen based on Silverman's rule [15].

The transmitted 4PAM symbols takes the values of $\{-3, -1, +1, +3\}$ which are generated by computer as $2 \times \text{random}(4) - 3$. To investigate that the generated PDFs conform to the transmitted PDF, the randomly generated desired symbols $D_N = \{d_1, d_2, \cdots, d_j, \cdots, d_N\}$ for MED1 are compared with the transmitted 4PAM symbols. The random symbol d_j for MED1 is generated according to the following rule.

$$d_j = \begin{cases} +3: & j = 1, 2, 3, \cdots, N/4, \\ +1: & j = N/4+1, N/4+2, \cdots, N/2, \\ -1: & j = N/2+1, N/2+2, \cdots, 3N/4, \\ -3: & j = 3N/4+1, 3N/4+2, \cdots, N. \end{cases} \quad (28)$$

Clearly, the mathematical PDF expression for the set D_N in (28) is $f_d(\xi) = \frac{1}{4}[\delta(\xi-3) + \delta(\xi-1) + \delta(\xi+1) + \delta(\xi+3)]$. Referring to (14), we notice that MED2 has exactly the same probability distribution $f_d(\xi)$ as MED1 does.

In Fig. 2, the PDFs of transmitted symbols, MED1, and MED2 are depicted. For a short term period of transmitted symbols (about 10000 symbols), PDF values have shown a very slight difference among them but the same in the long run.

The main drawback of CMA-type equalization algorithms is that they require a long sequence of data to converge. So, we give a comparison on the convergence rate in Figs. 3 and 6 for channel 1 and channel 2, respectively. We see that increasing the ESR has the effect of increasing the steady-state error of CMA in comparison of convergence rate and PDF distribution. In both channel environments, MED2 has shown significantly fast convergence and lower steady state MSE. In case of channel 1, the error performance of MED1 shows a slightly increased performance in comparison with CMA. In Fig. 5, BER performance for channel 1 shows that MED2 has 1 dB performance enhancement compared to CMA. For channel 2, CMA shows severe performance degradation in Fig. 7, but the error performance of MED1 and MED2 is superior.

CMA was proposed for constant magnitude signals such as frequency modulation (FM) and quadratic phase shift keying

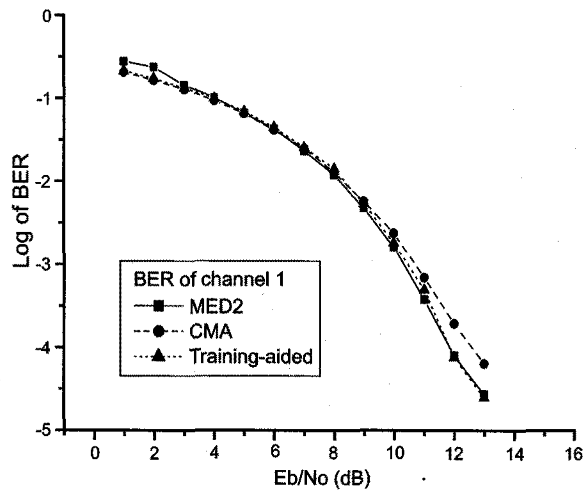

Fig. 5. BER performance for channel 1.

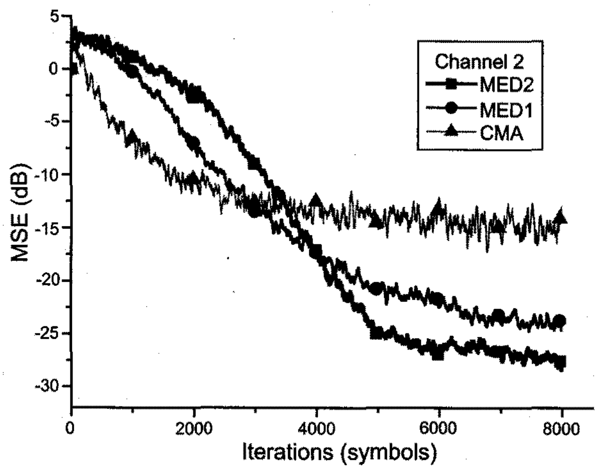

Fig. 6. MSE convergence comparison for channel 2.

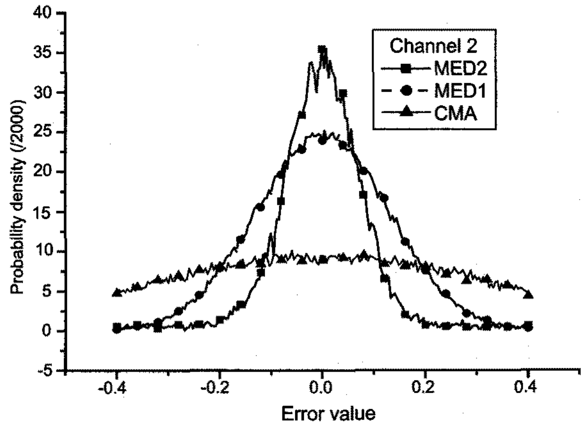

Fig. 7. Probability density for errors for channel 2.

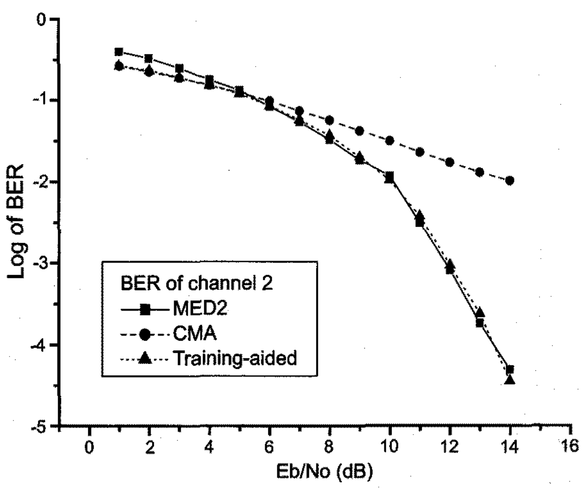

Fig. 8. BER performance for channel 2.

(QPSK). It is worthwhile to compare the proposed methods, other algorithms and CMA in QPSK modulation scheme. As a figure of merit, MSE convergence performance is compared in Fig. 9. The quadratic distance cost function (QD) is a blind algorithm based on the minimization of ED between the two PDFs of output power and transmitted symbol power [16]. In the aspect of using output symbol power, QD can be considered to be closely related with CMA.

The kernel size for QD and MED2 is set to 0.8 and 0.5, respectively. The convergence parameter for both MED2 and QD is 0.005, and for CMA, 0.0005 is used. In QPSK environment, QD has shown small performance enhancement in convergence rate compared to CMA, but it has shown less variance in steady state MSE than CMA. The proposed MED2 converges in about 2000 samples and its steady state MSE is about 2 dB lower than QD. The training-aided algorithm converges in about 1000 samples. Though MED2 is a little slower than the training-aided algorithm, it has the same steady state MSE performance as the training-aided algorithm, and furthermore it is noticeable that the MSE curves of training-aided algorithm and CMA fluctuate more than MED2 in the steady state. This indicates that the proposed MSE2 is considered to have less excess MSE in QPSK modulation scheme. In Figs. 8 and 9, the proposed MED2 method illustrates almost the same BER or minimum MSE performance as the training-aided equalizer algorithm does. The training-aided algorithm is based on the MSE criterion, whereas the proposed algorithms are based on information potential induced from Euclidian PDF distance.

VIII. CONCLUSION

For blind equalization, we introduced the criterion of ED minimization between the output PDF and the PDF of the transmitted symbols. In creating the PDF of the transmitted symbols at the receiver, we proposed two methods. One method is to use a set of randomly generated desired symbols that the PDF of the generated symbols matches that of the transmitted symbols, and the second one is to use a set of Dirac delta functions as the PDF of the transmitted symbols. In both channel models, the proposed methods, MED1 and MED2, show enhanced performance without significant performance degradation in comparison with CMA. This implies that the proposed methods can be considered relatively insensitive to ESR variations compared to CMA. The BER performance comparison reveals that MED2 has very close performance to the training-aided LMS equalizer, so the proposed method MED2 can be successfully employed in blind equalizer applications. In future work, research

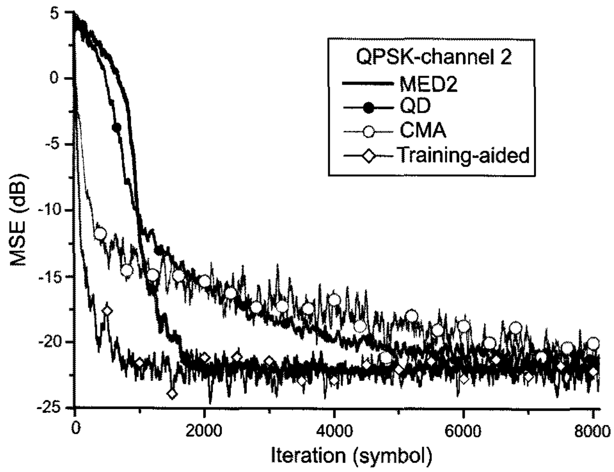

Fig. 9. MSE convergence comparison in QPSK scheme.

for reduced computational complexity is considered for efficient implementation.

REFERENCES

[1] W. Moh and Y. Chen, "Multicasting flow control for hybrid wired/wireless ATM networks," *Performance Evaluation*, vol. 40, pp. 161–194, Mar. 2000.

[2] L. M. Garth, "A dynamic convergence analysis of blind equalization algorithms," *IEEE Trans. on Comm.*, vol. 49, pp. 624–634, Apr. 2001.

[3] F. Mazzenga, "Channel estimation and equalization for M-QAM transmission with a hidden pilot sequence," *IEEE Trans. Broadcast.*, vol. 46, pp. 170–176, June 2000.

[4] J. C. Principe, D. Xu and J. Fisher, "Information theoretic learning," *Unsupervised Adaptive Filtering*, vol. I, S. Haykin(ed.) NY: USA, Wiley, 2000, pp. 265-319.

[5] S. Kullback, *Information Theory and Statistics*, NY: USA, Dover Publications, 1968.

[6] D. Erdogmus, Y. Rao, and J. C. Principe, "Supervised training of adaptive systems with partially labeled data," in *Proc. ASSP*, Apr. 2005, pp.v321–v324.

[7] 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.

[8] 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.

[9] 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, Nov. 1983, pp. 349–372.

[10] K. H. Jeong, J. W. Xu, D. Erdogmus, and J. C. Principe, "A new classifier based on information theoretic learning with unlabeled data," *Neural Networks*, vol. 18, pp. 719–726, 2005.

[11] N. Kim, K. H. Jeong, and K. Kwon, "A study on the weighting effect on information potentials in blind equalizers for multipoint communication," in *Proc. APIC-IST*, 2008, pp. 103–108.

[12] S. Haykin and and J. C. Principe, "Dynamic modeling with neural networks," *IEEE Trans. Signal Process. Mag.*, vol. 15, p. 66, Mar. 1998.

[13] L. Wasserman, *All of Statistics: A Concise Course in Statistical Inference.* Springer Texts in Statistics, 2005.

[14] S. Haykin, *Adaptive Filter Theory.* , 4th ed., Prentice Hall. 2001.

[15] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, London: UK, Chapman and Hall, 1986.

[16] 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.

Namyong Kim received the B. S., M. S., and Ph. D. degree 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.