# 블라인드 등화를 위한 최소 에러 엔트로피 성능기준들에 관한 연구

김남용, 권기현

# A Study on the Minimum Error Entropy - related Criteria for Blind Equalization

Namyong Kim*, Kihyun Kwon**


요 약

정보이론적 학습 기법에 해당하는 에러 엔트로피 최소화 (MEE) 성능기준과 상호 상관 엔트로피 최대화 (MCC) 성능기준은 그 동안 깊이 있게 많은 연구가 이루어져 왔다. 에러 엔트로피 최소화 성능기준은 정보 포텐셜을 최대화하는 것으로 귀결되고 상호 상관 엔트로피 최대화 성능기준은 시스템의 출력과 원신호의 상호 상관도를 최대화하는 것으로 정의된다. 이 두 성능기준을 적정 가중치를 두고 합성한 것이 기준점을 내포한 에러 엔트로피 최소화 기법 (MEEF) 인데 이 또한 많은 연구가 이루어지고 있다. 이 논문에서는 블라인드 채널 등화를 위해 CMA에 쓰이는 상수 모듈러스 에러 (CME)를 도입하여 이 정보이론적 학습기법에 적용하고자 그 가능성과 문제점을 찾고자 연구하였다. 또한 MEEF 성능기준에도 이 CME 적용가능성을 연구하였다. 연구결과로부터 CME를 적용한 MEE (MEE-CME)는 상수 모듈러스 정보를 잃게 되는 결과를 낳았다. 이 결과 MEE-CME나 MEE를 사용하는 MEEF-CME 모두에게서 수렴하지 못하거나 CME를 사용하는 다른 방식과 비교할 때 수렴이 늦게 되는 문제점을 발견하게 되었다.

ABSTRACT

As information theoretic learning techniques, error entropy minimization criterion (MEE) and maximum cross correntropy criterion (MCC) have been studied in depth for supervised learning. MEE criterion leads to maximization of information potential and MCC criterion leads to maximization of cross correlation between output and input random processes. The weighted combination scheme of these two criteria, namely, minimization of Error Entropy with Fiducial points (MEEF) has been introduced and developed by many researchers. As an approach to unsupervised, blind channel equalization, we investigate the possibility of applying constant modulus error (CME) to MEE criterion and some problems of the method. Also we study on the application of CME to MEEF for blind equalization and find out that MEE-CME loses the information of the constant modulus. This leads MEE-CME and MEEF-CME not to converge or to converge slower than other algorithms dependent on the constant modulus.



Key Word

MEE, MEEF, Blind equalization, ITL, CMA, Constant modulus.


---

* 강원대학교 전자정보통신공학부 (namyong@kangwon.ac.k, kweon@kangwon.ac.kr)
#논문번호 : KIIECT2009-03-13 #접수일자 : 2009.08.26 #최종논문접수일자 : 2009.09.15

## I. Introduction

In broadcast networks, multipoint communication networks and mobile networks, blind equalizers are very useful to counteract multipath effects since they do not require a training sequence to start up or to restart after a communications breakdown [1][2]. Recently new blind equalization techniques have been developed through the use of information theoretic optimization criteria. This technique called information - theoretic learning (ITL) has been introduced by Princepe [3]. This approach is to choose the parameters W of the mapping g(.) such that a figure of merit based on information theory is optimized at the output space of the mapper. 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). The difficulty in approximating Shannon's entropy is overcome by utilizing Renyi's generalized entropy. Estimating the data PDF nonparametrically is based on the Parzen window method using a Gaussian kernel. The combination of Renyi's quadratic entropy with the Parzen window 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.

Error entropy minimization (MEE) criterion introduced by Erdogmus and his coworkers leads to maximization of information potential. Instead of using entropy minimization in blind equalization, a new method in which Euclidian distance between two PDFs is minimized has also been introduced [4]. The authors investigated the interactions among not only output samples but also ramdomly generated desired samples at the receiver by utilizing Euclidian distance (ED) in their previous works [5]. As another approach in supervised learning, the Euclidian distance between the PDF of error and a delta function can be minimized with respect to system weights, and we can get two information potentials, one for MEE and another for MCC (maximum cross correntropy) [6]. In that approach, the information potentials for MEE and MCC are in discord, that is, the information potential for MEE is to be maximized, and the information potential for MCC is to be minimized. On the other hand, the authors in [6] tried to unify MEE and MCC where both information potentials are to be minimized under some weighted combination schemes. This cost function is named as Minimization of Error Entropy with Fiducial points (MEEF). The MEEF has shown enhanced performance in a robust regression example and nonlinear short term prediction of the Mackey-Glass time series.

As an approach to unsupervised, blind channel equalization, we can adopt the strategy that the constant modulus error (CME) becomes minimum or zero. In this paper, we investigate the possibility of applying CME to MEE criterion and some problems of the method. Also we study on the application of CME to MEEF for blind equalization and find out any obstacles or problems for that approach.

## II. Euclidian Distance of PDFs

Recently, Erdogmus introduced an information theoretic framework based on Kullback-Leibler (KL) divergence [7] minimization for training adaptive systems in supervised learning settings using both labeled and unlabeled data [4]. The KL divergence is a way to estimate mutual information which is capable of quantifying the entropy between pairs of random variables. The KL divergence between two PDFs, $f_x$ and $f_y$ is

$$KL[f_x, f_y] = \int f_x(\xi) \log[f_x(\xi)/f_y(\xi)]d\xi \quad (1)$$

Since it is not quadratic in the PDFs, it can not be easily integrated with the information potential [3]. Based on the quadratic entropy theory, a new difference measure between the desired and output samples has been introduced as follows.

## III. Supervised MEE Criterion

Entropy is a scalar quantity that provides a measure for the average information contained in a given PDF. When 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

$$H(e) = -\log(\int f_E(\xi)^2 d\xi) \quad (2)$$

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

$$H(e) = -\log(IP_e), \quad (3)$$

where

$$IP_e = \frac{1}{N^2} \sum_{i=k-N+1}^{k} \sum_{j=k-N+1}^{k} G_{\sigma\sqrt{2}}(e_j - e_i) \quad . \quad (4)$$

Obviously, minimizing the error entropy $H(e)$ is equivalent to maximizing the information potential $IP_e$. This criterion maximizing $IP_e$ is referred to as MEE [8].
By applying gradient ascent method to maximization of $IP_e$, supervised MEE algorithm in [8][9] can be obtained as

$$W_{k+1} = W_k + \frac{\mu_{MEE}}{2\sigma^2 N^2} \sum_{i=k-N+1}^{k} \sum_{j=k-N+1}^{k} (e_j - e_i) \cdot G_{\sigma\sqrt{2}}(e_j - e_i)[X_j - X_i] \quad . \quad (5)$$

## IV. MEE Criterion based on CME

Supervised MEE criterion in (8) deals with $e_j - e_i$. By replacing $e_i$ with CME $|y_i|^2 - R_2$, information potential using constant modulus error $IP_{CME}$ becomes independent of the constant modulus $R_2$ as

$$IP_{CME} = \frac{1}{N^2} \sum_{i=k-N+1}^{k} \sum_{j=k-N+1}^{k} G_{\sigma\sqrt{2}}(|y_i|^2 - |y_j|^2) \quad (6)$$

To maximize the cost function (6) we adopt the gradient ascent method. The gradient is evaluated from

$$\frac{\partial IP_{CME}}{\partial W} = \frac{1}{N^2\sigma^2} \sum_{i=k-N+1}^{k} \sum_{j=k-N+1}^{k} G_{\sigma\sqrt{2}}(|y_i|^2 - |y_j|^2) \cdot (|y_i|^2 - |y_j|^2) \cdot (y_j X_j^* - y_i X_i^*) \quad . \quad (7)$$

MEE-CME can be written using the gradient as following.

$$W_{k+1} = W_k + \mu_{MEECME} \cdot \frac{\partial IP_{CME}}{\partial W} \quad , \quad (8)$$

where $\mu_{MEECME}$ is the step-size for MEE-CME.

$IP_{CME}$ is maximized when equalizer output powers are the same $|y_i|^2 = |y_j|^2$. 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 the equalizer tries to cluster the outputs to 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 obtain the same output power $|y_i|^2 = |y_j|^2$ in spite of different desired powers. Consequently, MEE-CME loses the information of the constant modulus $R_2$. This may lead MEE-CME not to converge or to converge slower than other algorithms dependent on the constant modulus $R_2$.

## V. MEEF Criterion based on CME

In the unified version of MEE and MCC, both information potentials are to be minimized under some weighted combination schemes. This supervised cost function MEEF has shown enhanced performance in a robust regression example and nonlinear short term prediction of the Mackey-Glass time series.

The supervised cost function MEEF is

$$MEEF_e = \lambda \cdot \sum G_{\sigma\sqrt{2}}(e_i) + (1-\lambda) \cdot IP_e. \quad (9)$$

where is a weighting constant between 0 and 1. Unifying two cost functions actually retains all the merits of being robust with outlier resistance and kernel size resilience [6].

Now introducing constant modulus error signals $e_{CME} = |y_k|^2 - R_2$ to the MEEF cost function, we can obtain the unsupervised MEEF cost function as follows

$$MEEF_{CME} = \lambda \cdot \sum G_{\sigma\sqrt{2}}(e_{CME}) + (1-\lambda) \cdot IP_{CME} \quad (10)$$

Minimization of the cost function leads to the following algorithm (we will call this MEEF-CME in this paper).

$$\boldsymbol{W}_{k+1} = \boldsymbol{W}_k - \mu_{MEEFCME} \frac{1}{N^2 \sigma^2}$$
$$\cdot [\lambda \cdot N \sum_{i=k-N+1}^{k} G_{\sigma}(|y_i|^2 - R_2)$$
$$\cdot (R_2 - |y_i|^2) \cdot y_i \cdot \boldsymbol{X}_i^*$$
$$+ (1-\lambda) \sum_{i=k-N+1}^{k} \sum_{j=k-N+1}^{k} G_{\sigma\sqrt{2}}(|y_i|^2 - |y_j|^2)$$
$$\cdot (|y_i|^2 - |y_j|^2) \cdot (y_j \boldsymbol{X}_j^* - y_i \boldsymbol{X}_i^*)]. \quad (11)$$

## VI. Results and Discussion

In this section we present and discuss simulation resultsthat illustrate the comparative performance of the MEE-CME and MEEF-CME for blind equalization. They

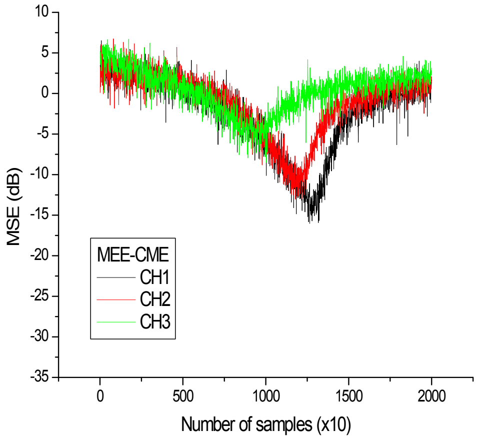



Fig. 1. MSE convergence of MEE-CME.

are studied for the three channel models in [10].The transfer functions of each channel models are

CH1: $H_1(z) = 0.26 + 0.93z^{-1} + 0.26z^{-2}$ .(12)

CH2:

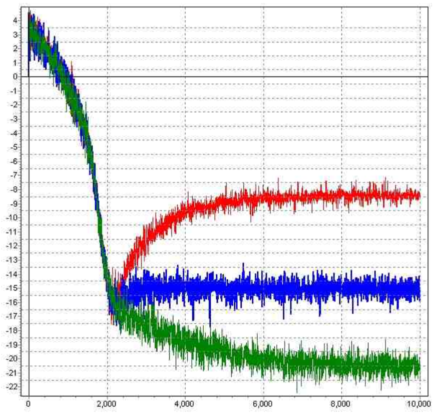
$H_2(z) = 0.304 + 0.903z^{-1} + 0.304z^{-2}$ .(13)

CH3: $H_3(z) = 0.407 + 0.815z^{-1} + 407z^{-2}$ . (14)

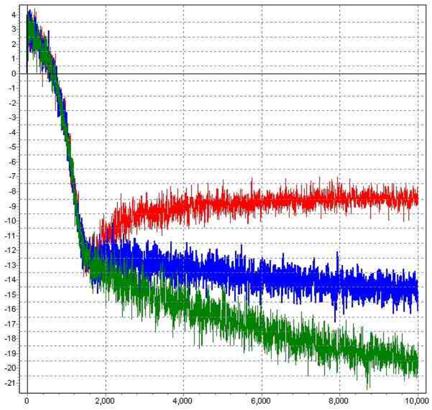
These channel models are typical multipath channel models and result in severe inter-symbol interference. Especially the channel model 3, CH3 poses worst spectral nulls in spectral characteristics.

The number of weights in the linear TDL equalizer structure is set to 11. The channel noise for MSE convergence performance is zero mean white Gaussian with the 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 4 level ( $M = 4$ ) random signal {-3, -1, 1, 3} is transmitted to the channel.

We use a common data-block size $N = 20$ for ITL-type blind algorithms. For MEE-CME, we use $\sigma = 3.0$ and $\mu_{MEECME} = 0.03$. The parameters for ITL algorithms are



(a)



(b)

Fig. 2. MSE convergence of MEEF-CME:
(a) CH1, (b) CH2. Red, blue and green lines are for lamda=0.3, 0.5, and 0.7, respectively.

commonly used in three channel models CH1, CH2, and CH3.

As discussed previously, MEE-CME loses the information of the constant modulus $R_2$. In those channel models, MEE-CME show ill-convergence as depicted in Fig. 1. This result indicates that MEE-CME can not be used in blind equalization due to the absence of the information on constant modulus.

MEEF (maximum error entropy with fiducial points) is a method of weighted combination of MEE and MCC. As a part of the MEEF, MEE works well in supervised equalization but in blind equalization applications based on constant modulus error, MEE-CME loses the information of the constant modulus $R_2$. In simulation MEE-CME shows ill-convergence as depicted in Fig. 1. MEE-CME is considered not appropriate in blind equalization due to the absence of the information on constant modulus $R_2$.As a result, MEE-CME can play a negative role in the combination of MEE-CME MCC-CME as shown in the following figure of learning performance with the variation of the balancing weight. (a) is for channel model 1 and (b) is for channel model 2. Red lines are for lamda=0.3, blue lines are for lamda=0.5, and green lines are for lamda=0.7. The case of lamda=0.3 means the portion of MEE-CME is bigger than MCC-CME and the case of lamda=0.7 means the portion of MEE-CME is smaller than MCC-CME. According these results, we could include MEEF based on constant modulus error can not be applied to blind equalization which requires rigorous performance.

## Ⅵ. Conclusion

MEE criterion has been a robust ITL criterion for many machine learning applications. MEE leads to maximization of information potential and MCC criterion leads to maximization of cross correlation between output and input random processes. As an approach to blind channel equalization, we investigate the possibility of applying constant modulus error (CME) to MEE criterion and some problems of the method and also the application of CME to MEEF for blind equalization. From the results, we find out that MEE-CME loses the information of the constant modulus. This leads MEE-CME and MEEF-CME not to converge or to converge slower than other algorithms dependent on the constant modulus. This implicates that other compensation techniques are needed to be developed for blind equalization.

## References

[1] L. M. Garth, " A dynamic convergence analysis of blind equalization algorithms," IEEE Trans. on Comm., Vol. 49, April. 2001, pp. 624-634.

[2] F. Mazzenga, " Channel estimation and equalization for M-QAM transmission with a hidden pilot sequence," IEEE Trans. on Broadcasting, Vol. 46, June. 2000, pp. 170-176.

[3] J.C. Principe, D. Xu and J. Fisher, Information Theoretic Learning, in: S. Haykin, Unsupervised Adaptive Filtering, Wiley, New York, Vol. I, 2000, pp 265-319.

[4] D. Erdogmus, Y. Rao and J. C. Principe, " Supervised Training of Adaptive Systems with Partially Labeled Data" Proceedings of the International Conference on ASSP, Apr. 2005. pp. v321-v324.

[5] N. Kim, and K. Kwon, "A Study on the Weighting Effect on Information Potentials in Blind Equalizers for Multipoint Communication," International Conference on Information Science and Technology, APIC-IST 2008, (Indang, Philippines), Dec. 18-19, 2008, pp. 103-108.

[6] Weifeng Liu, P. P. Pokharel, and J. C. Principe, "Error Entoropy, Correntropy, Metric, and M-Estimation," IEEE Int. Workshop on Machine Learning for Signal Processing, Maynooth, Ireland, 2006, pp. 179-184.

[7] S. Kullback, Information Theory and Statistics, Dover Publications, New York, 1968.

[8] D. Erdogmus, and J.C. Principe, "An Entropy Minimization algorithm for Supervised Training of Nonlinear Systems," IEEE

Trans. Signal Processing, vol. 50, July, 2002, pp. 1780-1786.
[9] I. Santamaria, D. Erdogmus, and J. C. Principe, "Entropy Minimization for Supervised Digital Communica tions Channel Equalization, IEEE Trans. Signal Processing, vol. 50, May, 2002, pp. 1184-1192.
[10] J. G. Proakis, Digital Communications, McGraw-Hill, 2nd edition, 1989.

## 저자약력

김남용(Namyong Kim)

1986년 세대학교
전자공학과 학사
1988년 연세대학교
전자공학과 석사
1991년 연세대학교
전자공학과 박사
1992년~1998년
관동대학교 전자통신공학과 부교수
1998년~현재 강원대학교
공학대학
전자정보통신공학부
교수
<관심분야> Adaptive equalization, RBFN algorithms, Odor sensing systems.

권기현 (Kwon, Kihyun)

1993년 강원대학교 컴퓨터과학과학사
1995년 강원대학교 컴퓨터과학과석사
2000년 강원대학교 컴퓨터과학과박사
1998년~2001년
동원대학교 조교수
2002년~현재 강원대학교 공학대학 전자정보통신공학부 조교수
<관심분야> 임베디드 소프트웨어, Odor sensing embedded systems.