Computationally Efficient Methods for Decision Feedback Algorithms based on Minimum Error Entropy

Namyong Kim and Anna Vladova Andonova

Abstract - **The algorithm of minimum error entropy with decision feedback (MEE-DF) has robustness to impulsive noise and severe multipath fading. However it has heavy computational burden being an obstacle for practical implementation. In this paper, a recursive gradient estimation method is proposed. The recursive gradient calculation method reduces the computations of to for the data block size. This indicates that the proposed recursive method is more appropriate for practical implementations of the MEE-DF algorithm**

Keywords – **Entropy, MEE, Decision-feedback, Recursive gradient, Computation, Complexity**

I. INTRODUCTION

Multipath fading and impulsive noise degrade performance in many communication systems such as satellite-mobile radio link, power line and underwater communication channels [1][2][3].

Typical adaptive equalizer algorithms to counteract multipath fading are MSE-based ones being highly sensitive to impulsive noise. But ITL based methods utilizing nonparametric probability density function (PDF) estimation and error entropy are robust to impulsive noise. Minimization of error entropy (MEE) has shown superior performance compared to MSE-based methods in supervised channel equalization applications [4]. While the MSE criterion uses only second order statistics being adequate under the assumptions of linearity and Gaussianity, error entropy considers all the higher order statistics of the error signal and is a scalar quantity that provides a measure for the average information contained in a given error distribution. When error entropy is minimized, all higher order moments are minimized and the error samples of adaptive systems are concentrated on zero as well. As a useful alternative definition of entropy, Renyi's quadratic error entropy is effectively used in ITL methods [4].

In the work [5], the performance of supervised linear MEE algorithm under impulsive noise environments has been studied, and for enhanced performance the MEE algorithm with decision feedback (MEE-DF) was proposed. It has shown significantly improved convergence in the situation of severely distorted channel and impulsive noise. One of the problems can be the computational burden of double summations induced from estimating the gradient for weight update at each iteration time. This computational burden must be an obstacle for practical implementation. In order to reduce the computational complexity of the MEE-DF algorithm for practical implementation a new method of computing each gradient recursively is proposed in this paper.

N. Kim is with Division of Electronic, Information and Communication Engineering, Kangwon National University, Samcheok, Gangwon-do, Korea, email: namyong@kangwon.ac.kr

A. Andonova is with the Department of Microelectronics, Faculty of Electronic Engineering and Technologies, Technical University - Sofia, 8 Kliment Ohridski blvd., 1000 Sofia, Bulgaria, e-mail: ava@ecad.tu-sofia.bg

II. MEE ALGORITHM WITH DECISION FEEDBACK

The Reny's quadratic entropy is associated with a probability density function as

$$H(e) = -\log(\int f_E(e)^2 de) \tag{1}$$

For the form of the PDF, the Parzen estimator [6] with Gaussian kernel $G_\sigma(\cdot)$ and a block of N past error samples can be employed as

$$f_E(e) = \frac{1}{N}\sum_{i=k-N+1}^{k} G_\sigma(e-e_i) = = \frac{1}{N}\sum_{i=k-N+1}^{k} \frac{1}{\sigma\sqrt{2\pi}}\exp[\frac{-(e-e_i)^2}{2\sigma^2}] \tag{2}$$

Then the combination of Reny's quadratic entropy with the Parzen window leads to an estimation of entropy by computing interactions among pairs of error samples. The sum of all error pairs of interactions is called information potential, IP_e [4].

$$H(e) = -\log\left[\frac{1}{N^2}\sum_{i=k-N+1}^{k}\sum_{j=k-N+1}^{k} G_{\sigma\sqrt{2}}(e_j-e_i)\right] = = -\log(IP_e) \tag{3}$$

Obviously, MEE is equivalent to maximizing the information potential IP_e.

In order for the MEE criterion to be applied to tapped delay line structures with DF that consist of a feed-forward filter with weight vector $\mathbf{W}_k^F$ and a feedback filter with weight vector $\mathbf{W}_k^B$, the algorithm has to have the DF part using produced decisions $\hat{d}_k$. While the feed-forward filter receives input x_k to produce output y_k, the feedback filter receives the sequence of decisions.

When the number of weights in feed-forward is A and the one of feedback filter section is B, respectively, then output y_k becomes

$$y_k = \sum_{a=0}^{A-1} w_{k,a}^F x_{k-a} + \sum_{b=0}^{B-1} w_{k,b}^B \hat{d}_{k-b-1} \quad (4)$$

where $\{w_{k,0}^F, w_{k,1}^F, w_{k,2}^F, ..., w_{k,A-1}^F\}$ are elements of feed-forward weight vector $\mathbf{W}_k^F$, $\{w_{k,0}^B, w_{k,1}^B, w_{k,2}^B, ..., w_{k,B-1}^B\}$ are elements of feedback weight vector $\mathbf{W}_k^B$. The input vector is $\mathbf{X}_k = [x_k, x_{k-1}, x_{k-2}, ..., x_{k-A+1}]^T$ and the elements of vector $\hat{\mathbf{D}}_{k-1}$, $\{\hat{d}_{k-1}, \hat{d}_{k-2}, ..., \hat{d}_{k-B-2}\}$ are previously decided symbols.

The filter weights are adjusted recursively to minimize the cost function (6) (i,e, to maximize IP_e) using the calculated error $e_k = d_k - y_k$ in training mode and $e_k = \hat{d}_k - y_k$ in decision directed mode. Then the feed-forward weight vector $\mathbf{W}_k^F$ and the feedback weight vector $\mathbf{W}_k^B$ are updated based on steepest ascent method with the step size μ_{MEE-DF} and the gradients

For the use of the gradient ascent method in maximization of IP_e in [4], we have the gradient of the cost function IP_e with respect to $\mathbf{W}^F$ and $\mathbf{W}^B$ as

$$\frac{\partial IP_e}{\partial \mathbf{W}^F} = \frac{1}{2\sigma^2 N^2} \sum_{i=k-N+1}^{k} \sum_{j=k-N+1}^{k} (e_j - e_i). \cdot G_{\sigma\sqrt{2}}(e_j - e_i)[\mathbf{X}_j - \mathbf{X}_i] \quad (5)$$

$$\frac{\partial IP_e}{\partial \mathbf{W}^B} = \frac{1}{2\sigma^2 N^2} \sum_{i=k-N+1}^{k} \sum_{j=k-N+1}^{k} (e_j - e_i). \cdot G_{\sigma\sqrt{2}}(e_j - e_i)[\hat{\mathbf{D}}_{j-1} - \hat{\mathbf{D}}_{i-1}] \quad (6)$$

Using these gradients as introduced in [5] the MEE-DF algorithm with step size μ_{MEE-DF} becomes

$$\mathbf{W}_{k+1}^F = \mathbf{W}_k^F + \mu_{MEE-DF} \frac{\partial IP_e}{\partial \mathbf{W}^F} \quad (7)$$

$$\mathbf{W}_{k+1}^B = \mathbf{W}_k^B + \mu_{MEE-DF} \frac{\partial IP_e}{\partial \mathbf{W}^B} \quad (8)$$

As shown in (5) and (6), it is observed that the gradient vector $\frac{\partial IP_e}{\partial \mathbf{W}^F} = \nabla_k^F$ and $\frac{\partial IP_e}{\partial \mathbf{W}^B} = \nabla_k^B$ at each iteration are estimated through the computation of $O(N^2)$ (due to double summations) for each filter section. This computational burden must be an obstacle for practical implementation. In order to reduce the computational complexity of the MEE-DF algorithm for practical implementation a new method of computing each gradient recursively through utilizing the previously calculated gradient and current data is proposed.

III. Recursive Gradient Calculation

Each gradient vector ∇_k^F and ∇_k^B can be rewritten in two cases; the initial and steady state. In the initial state for the time $1 \le k \le N$ each gradient vector can be expressed as

$$\nabla_k^F = \frac{1}{2\sigma^2 k^2} \sum_{i=1}^{k} \sum_{j=1}^{k} (e_j - e_i). \cdot G_{\sigma\sqrt{2}}(e_j - e_i)[\mathbf{X}_j - \mathbf{X}_i] \quad (9)$$

and

$$\nabla_k^B = \frac{1}{2\sigma^2 k^2} \sum_{i=1}^{k} \sum_{j=1}^{k} (e_j - e_i). \cdot G_{\sigma\sqrt{2}}(e_j - e_i)[\hat{\mathbf{D}}_{j-1} - \hat{\mathbf{D}}_{i-1}] \quad (10)$$

Separating the data related with time k from each summation leads to

$$\nabla_k^F = \frac{(k-1)^2}{2\sigma^2 k^2 (k-1)^2} \sum_{i=1}^{k} \Big(\sum_{j=1}^{k-1} (e_j - e_i). \cdot G_{\sigma\sqrt{2}}(e_j - e_i)[\mathbf{X}_j - \mathbf{X}_i] + + (e_k - e_i) \cdot G_{\sigma\sqrt{2}}(e_k - e_i)[\mathbf{X}_k - \mathbf{X}_i] \Big) = = \frac{(k-1)^2}{k^2} \nabla_{k-1}^F + \frac{1}{\sigma^2 k^2} \sum_{i=1}^{k-1} (e_k - e_i). \cdot G_{\sigma\sqrt{2}}(e_k - e_i)[\mathbf{X}_k - \mathbf{X}_i] \quad (11)$$

Similarly, the backward gradient becomes

$$\nabla_k^B = \frac{(k-1)^2}{2\sigma^2 k^2 (k-1)^2} \sum_{i=1}^{k} \Big(\sum_{j=1}^{k-1} (e_j - e_i). \cdot G_{\sigma\sqrt{2}}(e_j - e_i)[\hat{\mathbf{D}}_{j-1} - \hat{\mathbf{D}}_{i-1}] + + (e_k - e_i) \cdot G_{\sigma\sqrt{2}}(e_k - e_i)[\hat{\mathbf{D}}_{k-1} - \hat{\mathbf{D}}_{i-1}] \Big) = = \frac{(k-1)^2}{k^2} \nabla_{k-1}^B + \frac{1}{\sigma^2 k^2} \sum_{i=1}^{k-1} (e_k - e_i). \cdot G_{\sigma\sqrt{2}}(e_k - e_i)[\hat{\mathbf{D}}_{k-1} - \hat{\mathbf{D}}_{i-1}] \quad (12)$$

In the steady state for $k \ge N+1$, the gradient for the feed forward section at time $k-1$ is

$$\nabla_k^F = \frac{1}{2\sigma^2 N^2} \sum_{i=k-N+1}^{k} \sum_{j=k-N+1}^{k} (e_j - e_i). \cdot G_{\sigma\sqrt{2}}(e_j - e_i)[\mathbf{X}_j - \mathbf{X}_i] = = \nabla_{k-1}^F + \frac{1}{\sigma^2 N^2} \sum_{i=k-N}^{k-1} (e_k - e_i). \cdot G_{\sigma\sqrt{2}}(e_k - e_i)[\mathbf{X}_k - \mathbf{X}_i] - - \frac{1}{\sigma^2 N^2} \sum_{i=k-N}^{k-1} (e_{k-N} - e_i). \cdot G_{\sigma\sqrt{2}}(e_{k-N} - e_i)[\mathbf{X}_{k-N} - \mathbf{X}_i] \quad (13)$$

Similarly, the gradient at time $k-1$ for the feedback section is

$$\nabla_k^B = \frac{1}{2\sigma^2 N^2} \sum_{i=k-N+1}^{k} \sum_{j=k-N+1}^{k} (e_j - e_i).$$
$$\cdot G_{\sigma\sqrt{2}}(e_j - e_i)[\hat{\mathbf{D}}_{j-1} - \hat{\mathbf{D}}_{i-1}] =$$
$$= \nabla_{k-1}^B + \frac{1}{\sigma^2 N^2} \sum_{i=k-N}^{k-1} (e_k - e_i).$$
$$\cdot G_{\sigma\sqrt{2}}(e_k - e_i)[\hat{\mathbf{D}}_{k-1} - \hat{\mathbf{D}}_{i-1}] -$$
$$- \frac{1}{\sigma^2 N^2} \sum_{i=k-N}^{k-1} (e_{k-N} - e_i).$$
$$\cdot G_{\sigma\sqrt{2}}(e_{k-N} - e_i)[\hat{\mathbf{D}}_{k-N-1} - \hat{\mathbf{D}}_{i-1}] \qquad (14)$$

IV. RESULTS AND DISCUSSIONS

It is noticeable when we consider the computational complexity $O(N^2)$ of the original gradient (5) and (6) that the resulting recursive estimation of initial state gradients (11) and (12) for $1 \le k \le N$ and the steady state gradients (13) and (14) for $k \ge N+1$ obviously has reduced computations $O(N)$ which is more appropriate to practical implementations.

In this section we investigate how much the proposed recursive estimation of (13) and (14) reduces computational burden in the aspect of multiplications compared to the original gradient estimation of (5) and (6). For convenience, the Gaussian kernel $G_{\sigma\sqrt{2}}(e_j - e_i)$ is treated as a function value for $(e_j - e_i)$ and the steady state estimation is dealt with. Also $\frac{1}{2\sigma^2 N^2}$ or $\frac{1}{\sigma^2 N^2}$ are treated as constants.

Then the equation (5) and (6) require $2N^2 + 1$ multiplications for each forward and backward weight.

However, the equation (13) and (14) demands $4N + 2$ multiplications for each forward and backward weight. This can be illustrated in the Fig. 1 for clearer comparison.

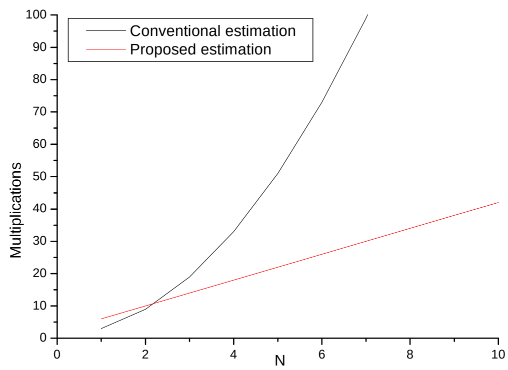

Fig. 1. Comparison of the number of multiplications for various data block size N

From the Fig. 1, we can mention that when the data block size N is smaller than 2 the computational complexity is lower than the proposed method, but in the case of the data block size N being greater than 2 the computational complexity of the proposed one is significantly low showing a linear increase while the conventional method increases by the square of the block size N. Because large data block sizes ensure reliable density estimates as depicted in [7], they must be selected under the consideration of a computational cost and reliable performance. So the proposed method provides a significantly wide range of choice for the data block size N.

V. CONCLUSION

It has been revealed that decision feedback MEE algorithms have robustness against impulsive noise and severe multipath fading. However the gradient of MEE-DF algorithm is estimated through the computation of double summations at each iteration time for each filter section. This computational burden must be an obstacle for practical implementation. In this paper, for the advantages in implementation, a recursive gradient estimation method is proposed. The recursive gradient calculation method reduces the computational burden of $O(N^2)$ to $O(N)$. This indicates that the proposed recursive method is efficient in channel equalization and adaptive signal processing applications in which the MEE-DF can be adopted. And the gradient estimation of the decision feedback MEE algorithm is more appropriate for practical implementations.

REFERENCES

[1] M. Chitre, S. Shahabudeen, L. Freitag, and M. Stojanovic, *Recent advances in underwater acoustic communications & networking*, Proc. MTS/IEEE OCEANS 2008, Quebec City, QC, Canada: IEEE, Sep. 2008. pp. 1-10.

[2] M. Richharia, *Satellite communication systems: design principles*, Technology & Engineering, 1999.

[3] T. Faber, T. Scholand, H. Bruck, P.Ju ng, , *Application of joint source-channel decoding to impulsive noise environments*, Vehicular Technology Conference, 2004. VTC2004-Fall. 2004 IEEE 60th , vol.2, no., pp.1238-1242 Vol. 2, 26-29 Sept. 2004.

[4] 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.

[5] N. Kim, *Decision feedback equalizer algorithms based on error entropy criterion*, Journal of Korean Society for Internet Information, vol. 12, 2011, pp. 27-33.

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

[7] M. Girolami and C. He, *Probability density estimation from optimally condensed data samples*, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 25, 2003, pp. 1253-1264.