논문 2014-51-8-14

오차확률분포 사이 유클리드 거리의 새로운 기울기 추정법

(A New Gradient Estimation of Euclidean Distance between Error Distributions)

김 남 용*

(Namyong Kim©)

요 약

오차 신호의 확률분포 사이의 유클리드 거리 (Euclidean distance between error probability density functions, EDEP)는 충격성 잡음 환경의 적응 신호 처리를 위한 성능 지수로 사용되었다. 이 EDEP 알고리듬의 단점 중의 하나로 각 반복 시간마다 수행하는 이중적분에 의해 과다한 계산상의 복잡성이 있다. 이 논문에서는 EDEP 와 그 기울기 계산에서 계산상의 부담을 줄일 수 있는 반복적 추정 방법을 제안하였다. 데이터 블록 크기 N에 대하여, 기존의 추정 방식에 의한 EDEP와 그 기울기 계산량은 $O(N^2)$ 인 반면, 제안한 방식의 계산량은 $O(N)$ 이다. 성능 시험에서 제안한 방식의 EDEP와 그 기울기는 정상상태에서 기존의 블록 처리 방식과 동일한 추정결과를 나타냈다. 이러한 시뮬레이션 결과로부터, 제안한 방식이 실제 적응신호처리 분야에서 효과적인 방식임을 알 수 있다.

Abstract

The Euclidean distance between error probability density functions (EDEP) has been used as a performance criterion for supervised adaptive signal processing in impulsive noise environments. One of the drawbacks of the EDEP algorithm is a heavy computational complexity due to the double summation operations at each iteration time. In this paper, a recursive method to reduce its computational burden in the estimation of the EDEP and its gradient is proposed. For the data block size N, the computational complexity for the estimation of the EDEP and its gradient can be reduced to $O(N)$ by the proposed method, while the conventional estimation method has $O(N^2)$. In the performance test, the proposed EDEP and its gradient estimation yield the same estimation results in the steady state as the conventional block-processing method. The simulation results indicates that the proposed method can be effective in practical adaptive signal processing.

Keywords : Computational complexity, Error distribution, Euclidian distance, Gradient, Impulsive noise

Ⅰ. INTRODUCTION

Adaptive signal processing is a computation process that attempts to produce desired output signal through adjusting system weights to process the input signal according to the reference signal (the desired signal to be produced) in an iterative manner. The input signal to be processed can be signals from receptors like sensors and receivers[1]. Commonly the input signal suffers time-varying distortions so that adaptive filters are used to compensate the distortion

* 정회원, 강원대학교 공학대학 전자정보통신공학부 (Division of Electronic, Information and Communications Eng. Kangwon National University)
© Corresponding Author(E-mail: namyong@kangwon.ac.kr)
접수일자: 2014년03월15일, 수정일자: 2014년07월11일
수정완료: 2014년07월28일

and transform it into the desired output signal. The adaptive algorithms employed to the adaptive filter determine how the weights are adjusted from one time instant to the next. The adaptive algorithm is derived as a form of optimization procedure that minimizes or maximizes a chosen performance criterion[2].

Among many performance criteria, the mean squared error (MSE) measures the average of the squares of the error signal. The error signal is the difference between the reference signal and the output produced by the adaptive filter[3]. For Gaussian noise cases, averaging squared error samples taken from different time instants discards the effects of the Gaussian noise, but a single large, impulsive noise sample can dominate these sums and defeat the averaging.

Unlike the MSE-type learning methods, the information-theoretic learning (ITL) method is based on the information potential concept defined based on the assumption that data samples in a space interact with each other like physical particles in a potential field[4]. The information potential consists of probability density functions (PDFs) constructed by the kernel density estimation method with the Gaussian kernel[5].

The ITL method has given birth to many different performance criteria such as Kullback-Leibler (KL) divergence to estimate mutual information, and Euclidian distance (ED) between two PDFs to measure their similarity[6~7]. By minimizing the ED between PDFs obtained from real biomedical data sets, supervised training methods for medical diagnosis have been developed[7].

In the work [8] the researchers proposed to minimize the ED of error PDFs (EDEP) for FIR adaptive filter structures and derived adaptive algorithms adjusting the weights so that the error PDF matches a delta function. This algorithm based on the minimization of the EDEP has shown superior learning performance even in impulsive noise environments[9].

Some of the difficulties of the EDEP algorithm include a heavy computational complexity due to the double summation operations at each iteration time for the estimation of the EDEP or weight adjustment, being considered inappropriate to practical signal processing. To the best of our knowledge any efforts to reduce the computational complexity of the EDEP-type algorithms[8~9] have not been reported in the scientific literature.

In this paper, a recursive approach to the EDEP and its gradient estimation for the weight update of the EDEP algorithm is proposed in order to reduce the computational complexity. For that purpose, the relationship between the current components of the EDEP and the next time components is investigated to see if their estimation can be done recursively. Through simulations in equalization under an impulsive noise environment, it will be shown that the proposed method of recursive estimation of the EDEP and its gradient yields the same results with significantly reduced computations compared to the conventional EDEP method presented in [8].

This paper is organized as follows. Section II presents the definition of MSE criterion. The EDEP cost function is explained in Section III, and the recursive estimation of the EDEP is introduced in Section IV. In Section V a recursive gradient calculation method based on the recursive EDEP estimation is proposed. Section VI presents simulation results and discussions. Finally, concluding remarks are given in Section VII.

Ⅱ. SUPERVISED MSE CRITERION FOR ADAPTIVE SIGNAL PROCESSING

If we define d_k and y_k as a desired signal and the output of the adaptive system at time k, respectively, the error signal is calculated as $e_k = d_k - y_k$. The most widely used criterion, MSE is statistical average of error power expressed $MSE = E[e_k^2]$. Instead of estimating the expected value of error power, we can use the instant squared error $ISE = e_k^2$ as a cost function for practical implementation. Assuming the structure of a linear combiner with the input vector $\mathbf{X}_k = [x_k, x_{k-1}, ..., x_{k-L+1}]^T$ and system weight $\mathbf{W}_k = [w_{0,k}, w_{1,k}, ..., w_{L-1,k}]^T$ is employed, the output is $y_k = \mathbf{W}_k^T \mathbf{X}_k$. For minimization of the ISE, we adopt the steepest descent method using the gradient $\frac{\partial e_k^2}{\partial \mathbf{W}}$. This method leads to the well known least mean square (LMS) algorithm with convergence parameter μ_{LMS} [2].

$$\mathbf{W}_{k+1} = \mathbf{W}_k - \mu_{LMS} \frac{\partial e_k^2}{\partial \mathbf{W}}$$
$$= \mathbf{W}_k + 2\mu_{LMS} e_k \mathbf{X}_k \quad (1)$$

It is noticeable in (1) that large error samples due to impulsive noise can make the weight adjustment of the LMS algorithm unstable.

Ⅲ. EUCLIDEAN DISTANCE OF ERROR PDFS AND ITS ESTIMATION

Error PDFs can be constructed non-parametrically by way of the kernel density estimation method with Gaussian kernel and a block of N error samples $\{e_k, e_{k-1}, ..., e_{k-N+1}\}$ [5].

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

The Euclidean distance of error PDFs (EDEP) proposed in [8] is defined as $EDEP[f_E(e), \delta(e)]$, the distance between the PDF of error signal $f_E(e)$ and a Dirac-delta function $\delta(e)$. By minimizing $EDEP[f_E(e), \delta(e)]$, the PDF of system error $f_E(e)$ forms an impulse shape located at the origin.

$$EDEP[f_E(e), \delta(e)] = \int [f_E(\alpha)d\alpha - \int \delta(\alpha)]^2 d\alpha$$
$$= \int f_E^2(\alpha)d\alpha + \int \delta^2(\alpha)d\alpha - 2\int f_E(\alpha)\delta(\alpha)d\alpha \quad (3)$$

The term $\int \delta^2(\alpha)d\alpha$ in (3) is not related with the weights of the adaptive system so that it can be treated as a constant c. Then

$$EDEP[f_E(e), \delta(e)] = \int f_E^2(\alpha)d\alpha + c - 2\int f_E(\alpha)\delta(\alpha)d\alpha \quad (4)$$

Substituting (2) into (4), each term of (4) can be expressed as

$$\int f_E^2(\alpha)d\alpha$$
$$= \frac{1}{N^2} \sum_{i=k-N+1}^{k} \sum_{j=k-N+1}^{k} \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(e_j - e_i)^2}{4\sigma^2}] \quad (5)$$

$$\int f_E(\alpha)\delta(\alpha)d\alpha = \frac{1}{N} \sum_{i=k-N+1}^{k} \frac{1}{\sigma\sqrt{2\pi}} \exp[\frac{-(0-e_i)^2}{2\sigma^2}] \quad (6)$$

Defining A_k and B_k as $\int f_E^2(\alpha)d\alpha$ and $\int f_E(\alpha)\delta(\alpha)d\alpha$, respectively, $EDEP[f_E(e), \delta(e)]_k$ at time k becomes

$$EDEP[f_E(e), \delta(e)]_k = A_k + c - 2B_k \quad (7)$$

One of the basic ideas of the ITL method is that each data sample is considered as a physical particle. That is, the error samples e_i and e_j in (5) are treated as particles containing information of their locations. Since the Gaussian kernel $\frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(e_j - e_i)^2}{4\sigma^2}]$ functions an exponential decay of the distance between the two particles, it can treated as a potential field with interaction among the information particles.

Then $\sum_{j=k-N+1}^{k}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_j-e_i)^2}{4\sigma^2}]$ is corresponding to the sum of interactions on the i-th particle, and the Eq. (5) is the sum of interactions between all error sample pairs. This concept of overall potential energy is referred to as information potential (IP) and the EDEP (4) can be regarded as a combination of IPs[4]. Then $A_k=\int f_E^2(\alpha)d\alpha$ is the IP of the pairs of error samples and $B_k=\int f_E(\alpha)\delta(\alpha)d\alpha$ is the IP of the pairs of the origin and error samples. When a new error sample e_{k+1} comes into the combination of IPs (7) at time k+1, the old error sample e_{k-N+1} leaves the combined field, and A_{k+1} and B_{k+1} are estimated for $EDEP[f_E(e),\delta(e)]_{k+1}$. That e_{k-N+1} leaves the combined field indicates that the interactions between e_{k-N+1} and the other samples are discarded. The event that e_{k+1} comes into the combination of IPs implies that the new interactions between e_{k+1} and the others are counted in. From observing this process we can come up with the idea that A_{k+1} and B_{k+1} might be calculated just by adding new interactions related with e_{k+1} and subtracting old interactions related with e_{k-N+1} from the current information potentials A_k and B_k. That is, a more estimation-efficient method for $EDEP[f_E(e),\delta(e)]$ can be possible than the conventional block processing method for (5) and (6) at each iteration time.

Ⅳ. PROPOSED ESTIMATION OF EDEP

The summation operations can be divided into two cases according to whether the buffer for the summation is filled full $(k>N)$ or not $(1\le k\le N)$. Letting A_k^I and B_k^I be defined for the initial state $1\le k\le N$, and A_k^S and B_k^S be for the steady state $k>N$, we have

$$A_k^I=\frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_j-e_i)^2}{4\sigma^2}] \quad (8)$$

$$B_k^I=\frac{1}{k}\sum_{i=1}^{k}\frac{1}{\sigma\sqrt{2\pi}}\exp[\frac{-e_i^2}{2\sigma^2}] \quad (9)$$

$$A_k^S=\frac{1}{N^2}\sum_{i=k-N+1}^{k}\sum_{j=k-N+1}^{k}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_j-e_i)^2}{4\sigma^2}] \quad (10)$$

$$B_k^S=\frac{1}{N}\sum_{i=k-N+1}^{k}\frac{1}{\sigma\sqrt{2\pi}}\exp[\frac{-e_i^2}{2\sigma^2}] \quad (11)$$

Firstly in the initial state $1\le k\le N$, A_{k+1}^I and B_{k+1}^I of $EDEP[f_E(e),\delta(e)]_{k+1}$ are divided into some terms related with e_{k+1} and the remaining.

$$A_{k+1}^I=\frac{1}{(k+1)^2}\sum_{i=1}^{k+1}\sum_{j=1}^{k+1}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_j-e_i)^2}{4\sigma^2}]$$
$$=\frac{k^2}{(k+1)^2k^2}\left[\sum_{i=1}^{k}\sum_{j=1}^{k+1}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_j-e_i)^2}{4\sigma^2}]\right.$$
$$\left.+\sum_{j=1}^{k+1}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_j-e_{k+1})^2}{4\sigma^2}]\right]$$
$$=\frac{k^2}{(k+1)^2k^2}\left[\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_j-e_i)^2}{4\sigma^2}]\right.$$

$$+\sum_{i=1}^{k}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_{k+1}-e_i)^2}{4\sigma^2}]\Bigg]$$

$$+\frac{k^2}{(k+1)^2k^2}\sum_{j=1}^{k+1}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_j-e_{k+1})^2}{4\sigma^2}]$$

$$=\frac{k^2}{(k+1)^2}\cdot\frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_j-e_i)^2}{4\sigma^2}]$$

$$+\frac{1}{(k+1)^2}\sum_{i=1}^{k}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_{k+1}-e_i)^2}{4\sigma^2}]$$

$$+\frac{1}{(k+1)^2}\sum_{j=1}^{k+1}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_j-e_{k+1})^2}{4\sigma^2}]$$

$$=\frac{k^2}{(k+1)^2}A_k^I+\frac{1}{(k+1)^2}\sum_{i=1}^{k}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_{k+1}-e_i)^2}{4\sigma^2}]$$

$$+\frac{1}{(k+1)^2}\sum_{j=1}^{k+1}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_j-e_{k+1})^2}{4\sigma^2}]$$

$$=\frac{k^2}{(k+1)^2}A_k^I+\frac{1}{(k+1)^2}\sum_{i=1}^{k}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_{k+1}-e_i)^2}{4\sigma^2}]$$

$$+\frac{1}{(k+1)^2}\left[\sum_{j=1}^{k}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_j-e_{k+1})^2}{4\sigma^2}]\right.$$

$$\left.+\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_{k+1}-e_{k+1})^2}{4\sigma^2}]\right] \quad (12)$$

Since $\exp[\frac{-(e_{k+1}-e_i)^2}{4\sigma^2}]=\exp[\frac{-(e_i-e_{k+1})^2}{4\sigma^2}]$, A_{k+1}^I can be rewritten in the following recursive form.

$$A_{k+1}^I=\frac{k^2}{(k+1)^2}A_k^I+\frac{2}{(k+1)^2}\sum_{i=1}^{k}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_{k+1}-e_i)^2}{4\sigma^2}]$$

$$+\frac{1}{(k+1)^2}\frac{1}{2\sigma\sqrt{\pi}} \quad (13)$$

Similarly, the terms related with e_{k+1} can be divided from B_{k+1}^I as

$$B_{k+1}^I=\frac{1}{k+1}\sum_{i=1}^{k+1}\frac{1}{\sigma\sqrt{2\pi}}\exp[\frac{-e_i^2}{2\sigma^2}]$$

$$=\frac{k}{k+1}\frac{1}{k}\sum_{i=1}^{k+1}\frac{1}{\sigma\sqrt{2\pi}}\exp[\frac{-e_i^2}{2\sigma^2}]$$

$$=\frac{k}{k+1}[\frac{1}{k}\sum_{i=1}^{k}\frac{1}{\sigma\sqrt{2\pi}}\exp(\frac{-e_i^2}{2\sigma^2})$$

$$+\frac{1}{k}\frac{1}{\sigma\sqrt{2\pi}}\exp(\frac{-e_{k+1}^2}{2\sigma^2})]$$

$$=\frac{k}{k+1}[B_k^I+\frac{1}{k}\frac{1}{\sigma\sqrt{2\pi}}\exp(\frac{-e_{k+1}^2}{2\sigma^2})]$$

$$=\frac{k}{k+1}\cdot B_k^I+\frac{1}{k+1}\cdot\frac{1}{\sigma\sqrt{2\pi}}\exp(\frac{-e_{k+1}^2}{2\sigma^2}) \quad (14)$$

The Eq. (13) and Eq. (14) show that $EDEP[f_E(e),\delta(e)]_{k+1}$ in the initial state can be estimated with the current information potentials(A_k^I and B_k^I) and the next error sample e_{k+1}.

Secondly in the steady state, A_{k+1}^S and B_{k+1}^S can be divided into terms related with e_{k+1}, the ones related with e_{k-N+1}, and the remaining.

$$A_{k+1}^S=\frac{1}{N^2}\sum_{i=k-N+2}^{k+1}\sum_{j=k-N+2}^{k+1}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_j-e_i)^2}{4\sigma^2}]$$

$$=\frac{1}{N^2}\left[\sum_{i=k-N+1}^{k}\sum_{j=k-N+2}^{k+1}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_j-e_i)^2}{4\sigma^2}]\right.$$

$$+\sum_{j=k-N+2}^{k+1}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_{k+1}-e_j)^2}{4\sigma^2}]$$

$$\left.-\sum_{j=k-N+2}^{k+1}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_{k+N-1}-e_j)^2}{4\sigma^2}]\right]$$

$$=\frac{1}{N^2}\sum_{i=k-N+1}^{k}\left[\sum_{j=k-N+1}^{k}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_j-e_i)^2}{4\sigma^2}]\right.$$

$$+\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_i-e_{k+1})^2}{4\sigma^2}]$$

$$\left.-\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_i-e_{k-N+1})^2}{4\sigma^2}]\right]$$

$$+\frac{1}{N^2}\left[\sum_{j=k-N+1}^{k}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_{k+1}-e_j)^2}{4\sigma^2}]\right.$$

$$+\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_{k+1}-e_{k+1})^2}{4\sigma^2}]$$

$$\left.-\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_{k+1}-e_{k-N+1})^2}{4\sigma^2}]\right]$$

$$-\frac{1}{N^2}\left[\sum_{j=k-N+1}^{k}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_{k-N+1}-e_j)^2}{4\sigma^2}]\right.$$

$$+\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_{k-N+1}-e_{k+1})^2}{4\sigma^2}]$$

$$\left.-\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_{k-N+1}-e_{k-N+1})^2}{4\sigma^2}]\right] \quad (15)$$

Since $\exp[\frac{-(e_{k+1}-e_{k-N+1})^2}{4\sigma^2}]=\exp[\frac{-(e_{k-N+1}-e_{k+1})^2}{4\sigma^2}]$ in (15), A_{k+1}^S becomes

$$A_{k+1}^S=A_k^S+\frac{2}{N^2}\sum_{j=k-N+1}^{k}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_i-e_{k+1})^2}{4\sigma^2}]$$

$$-\frac{2}{N^2}\sum_{j=k-N+1}^{k}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_i-e_{k-N+1})^2}{4\sigma^2}]$$

$$-\frac{2}{N^2}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_{k+1}-e_{k-N+1})^2}{4\sigma^2}]+\frac{2}{N^2}\frac{1}{2\sigma\sqrt{\pi}} \quad (16)$$

The Eq. (16) shows that a component A_{k+1}^S of $EDEP\ [f_E(e),\delta(e)]_{k+1}$ can be estimated based on the current value A_k^S and the interactions with e_{k+1} and e_{k-N+1}.

Similarly, B_{k+1}^S can also be divided into the terms related with e_{k+1} and e_{k-N+1}, separately.

$$B_{k+1}^S=\frac{1}{N}\sum_{i=k-N+2}^{k+1}\frac{1}{\sigma\sqrt{2\pi}}\exp[\frac{-e_i^2}{2\sigma^2}]$$

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

$$=\frac{1}{N}[\sum_{i=k-N+1}^{k}\frac{1}{\sigma\sqrt{2\pi}}\exp(\frac{-e_i^2}{2\sigma^2})+\frac{1}{\sigma\sqrt{2\pi}}\exp(\frac{-e_{k+1}^2}{2\sigma^2})$$

$$-\frac{1}{\sigma\sqrt{2\pi}}\exp(\frac{-e_{k-N+1}^2}{2\sigma^2})]=B_k^S$$

$$+\frac{1}{N}[\frac{1}{\sigma\sqrt{2\pi}}\exp(\frac{-e_{k+1}^2}{2\sigma^2})$$

$$-\frac{1}{\sigma\sqrt{2\pi}}\exp(\frac{-e_{k-N+1}^2}{2\sigma^2})] \quad (17)$$

The Eq. (17) indicates that the other component B_{k+1}^S can be estimated based on the current B_k^S and the interactions with e_{k+1} and e_{k-N+1}.

In summary, the Eq. (13), (14), (16) and (17) are a recursive estimation of $EDEP[f_E(e),\delta(e)]$, contrary to the conventional block-processed calculation of it. One of the merits of the proposed estimation method is that its computational complexity is only $O(N)$, which is compared apparently with $O(N^2)$ of the block-processed method in (5), (6) and (7).

V. THE GRADIENT OF EDEP FOR ADAPTIVE ALGORITHMS

Adaptive algorithms developed to force error samples to be concentrated at zero through minimization of the criterion EDEP can be derived utilizing the derivative of EDEP with respect to the system weight. For that purpose, the gradient of the criterion and the steepest descent method are used as

$$\mathbf{W}_{k+1}=\mathbf{W}_k-\mu\cdot\frac{\partial EDEP\ [f_E(e),\delta(e)]_k}{\partial\mathbf{W}} \quad (18)$$

$$=\mathbf{W}_k-\mu\cdot\frac{\partial(A_k-2B_k)}{\partial\mathbf{W}} \quad (19)$$

where μ is the step-size for convergence control.

When the gradient $\frac{\partial(A_k-2B_k)}{\partial\mathbf{W}}=\frac{\partial A_k}{\partial\mathbf{W}}-2\frac{B_k}{\partial\mathbf{W}}$ is calculated by the block-processing method in (5) and (6), we have

$$\frac{\partial A_k}{\partial\mathbf{W}}=\frac{1}{2N^2\sigma^2}\sum_{i=k-N+1,}^{k}\sum_{j=k-N+1}^{k}(e_j-e_i)\cdot\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_j-e_i)^2}{4\sigma^2}]\cdot(\mathbf{X}_i-\mathbf{X}_j) \quad (20)$$

$$\frac{\partial B_k}{\partial\mathbf{W}}=\frac{1}{\sigma^2N}\sum_{i=k-N+1}^{k}\frac{1}{\sigma\sqrt{2\pi}}\exp(\frac{-e_i^2}{2\sigma^2})\cdot e_i\mathbf{X}_i \quad (21)$$

The weight update Eq. (19) with gradients (20) and (21) has been proposed in the work[8] and will be referred to as EDEP algorithm in this paper.

In this section, by using the similar approach to the EDEP estimation introduced in the previous section, a recursive estimation of the gradient (19) is proposed.

The gradients in the initial state $\frac{\partial A_{k+1}^I}{\partial\mathbf{W}}$ and $\frac{\partial B_{k+1}^I}{\partial\mathbf{W}}$ can be obtained by directly differentiate (13) and (14) with respect to the system weight.

$$\begin{aligned}\frac{\partial A_{k+1}^I}{\partial\mathbf{W}}&=\frac{k^2}{(k+1)^2}\frac{\partial A_k^I}{\partial\mathbf{W}}+\frac{2}{(k+1)^2}\sum_{i=1}^{k}\frac{1}{2\sigma\sqrt{\pi}}\frac{\partial}{\partial\mathbf{W}}\exp[\frac{-(e_{k+1}-e_i)^2}{4\sigma^2}]\\&=\frac{k^2}{(k+1)^2}\frac{\partial A_k^I}{\partial\mathbf{W}}+\frac{1}{(k+1)^2}\sum_{i=1}^{k}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_{k+1}-e_i)^2}{4\sigma^2}]\cdot\frac{(e_{k+1}-e_i)(\mathbf{X}_{k+1}-\mathbf{X}_i)}{\sigma^2}\end{aligned} \quad (22)$$

$$\begin{aligned}\frac{\partial B_{k+1}^I}{\partial\mathbf{W}}&=\frac{k}{k+1}\cdot\frac{\partial B_k^I}{\partial\mathbf{W}}+\frac{1}{k+1}\cdot\frac{1}{\sigma\sqrt{2\pi}}\frac{\partial}{\partial\mathbf{W}}\exp(\frac{-e_{k+1}^2}{2\sigma^2})\\&=\frac{k}{k+1}\cdot\frac{\partial B_k^I}{\partial\mathbf{W}}+\frac{1}{k+1}\cdot\frac{1}{\sigma\sqrt{2\pi}}\exp(\frac{-e_{k+1}^2}{2\sigma^2})\cdot\frac{e_{k+1}\mathbf{X}_{k+1}}{\sigma^2}\end{aligned} \quad (23)$$

Similarly, in the steady state,

$$\begin{aligned}\frac{\partial A_{k+1}^S}{\partial\mathbf{W}}=\frac{\partial A_k^S}{\partial\mathbf{W}}&+\frac{2}{N^2}\sum_{j=k-N+1}^{k}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_i-e_{k+1})^2}{4\sigma^2}]\cdot\frac{2(e_i-e_{k+1})}{4\sigma^2}(\mathbf{X}_i-\mathbf{X}_{k+1})\\&-\frac{2}{N^2}\sum_{j=k-N+1}^{k}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_i-e_{k-N+1})^2}{4\sigma^2}]\cdot\frac{2(e_i-e_{k-N+1})}{4\sigma^2}(\mathbf{X}_i-\mathbf{X}_{k-N+1})\\&-\frac{2}{N^2}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_{k+1}-e_{k-N+1})^2}{4\sigma^2}]\cdot\frac{2(e_{k+1}-e_{k-N+1})}{4\sigma^2}(\mathbf{X}_{k+1}-\mathbf{X}_{k-N+1})\end{aligned} \quad (24)$$

$$\begin{aligned}\frac{\partial B_{k+1}^S}{\partial\mathbf{W}}&=\frac{\partial B_k^S}{\partial\mathbf{W}}+\frac{1}{N}[\frac{1}{\sigma\sqrt{2\pi}}\frac{\partial}{\partial\mathbf{W}}\exp(\frac{-e_{k+1}^2}{2\sigma^2})-\frac{1}{\sigma\sqrt{2\pi}}\frac{\partial}{\partial\mathbf{W}}\exp(\frac{-e_{k-N+1}^2}{2\sigma^2})]\\&=\frac{\partial B_k}{\partial\mathbf{W}}+\frac{1}{N}[\frac{1}{\sigma\sqrt{2\pi}}\exp(\frac{-e_{k+1}^2}{2\sigma^2})\frac{-1}{2\sigma^2}\frac{\partial e_{k+1}^2}{\partial\mathbf{W}}-\frac{1}{\sigma\sqrt{2\pi}}\exp(\frac{-e_{k-N+1}^2}{2\sigma^2})\frac{-1}{2\sigma^2}\frac{\partial e_{k-N+1}^2}{\partial\mathbf{W}}]\\&=\frac{\partial B_k}{\partial\mathbf{W}}+\frac{1}{N}[\frac{1}{\sigma\sqrt{2\pi}}\exp(\frac{-e_{k+1}^2}{2\sigma^2})\cdot\frac{e_{k+1}\mathbf{X}_{k+1}}{\sigma^2}-\frac{1}{\sigma\sqrt{2\pi}}\exp(\frac{-e_{k-N+1}^2}{2\sigma^2})\cdot\frac{e_{k-N+1}\mathbf{X}_{k-N+1}}{\sigma^2}]\end{aligned} \quad (25)$$

Adopting the gradient descent method of (19) for the minimization of $EDEP[f_E(e),\delta(e)]$ and utilizing the recursive gradient (24) and (25), we can update the system weights for supervised adaptive signal processing.

For convenience, the Gaussian kernel $\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(e_j-e_i)^2}{4\sigma^2}]$ is treated as a function value for (e_j-e_i) and $\frac{1}{2N^2\sigma^3\sqrt{\pi}}$ is treated as a constant, then the single summation contains elements with 3 multiplications each in (24). The remaining term also has 3 multiplications. So the gradient calculation in (24) requires equation (5) and (6) require $3N+3$ multiplications. Similarly, the Eq. (25) demands 6 multiplications. Therefore the proposed gradient estimation requires only $3N+9$ multiplications. However, the Eq. (20) and (21) demand $3N^2$ and $3N$, respectively, so that the conventional method requires $3N^2+3N$ multiplications.

As a result, the proposed algorithm requires only $O(N)$ for recursive update of $\frac{\partial A_k^S}{\partial \mathbf{W}}$ in (24) while the block-processed algorithm needs $O(N^2)$ for $\frac{\partial A_k}{\partial \mathbf{W}}$ in (20).

VI. RESULTS AND DISCUSSION

In this section simulation results in the same environment of channel equalization as in [9] are compared for the two methods of the conventional block-processing and the proposed recursive estimation of the EDEP and its gradient. The symbol points to be transmitted are {-3, -1, 1, 3}. The multipath channel model with impulsive noise is $H_2(z)=0.304+0.903z^{-1}+0.304z^{-2}$. The power of back ground white noise is $\sigma_1^2=0.001$ and the variance of impulse noise plus back ground noise is $\sigma_2^2=50.001$. The impulse occurrence rate is $\varepsilon=0.03$. The number of weights is 11 and data-block size N is 20. The parameters of the step-size μ, the kernel size σ are 0.04 and 0.7, respectively. These parameters are selected when the EDEP algorithm yields the lowest steady-state MSE.

The trace of EDEP for the two methods is shown in Fig. 1. The EDEP of the conventional block-processing method is obtained by (7), and that of the recursive estimation is by (13), (14), (16) and (17). As the EDEP algorithm (19) converges, the distance EDEP decreases rapidly with iterations. The two methods yield the same EDEP estimation results

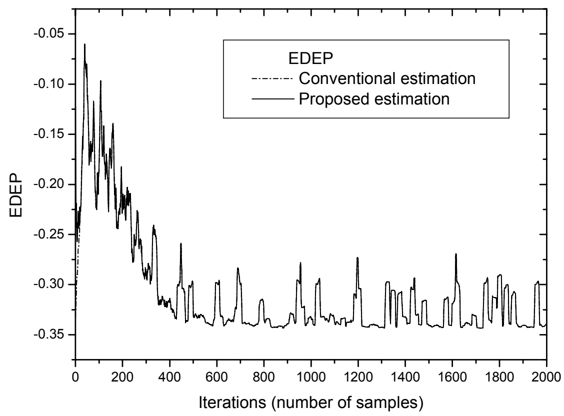

그림 1. 기존 방식과 제안한 방식의 EDEP 추정값
Fig. 1. EDEP estimation for the conventional block-processing method and the proposed one.

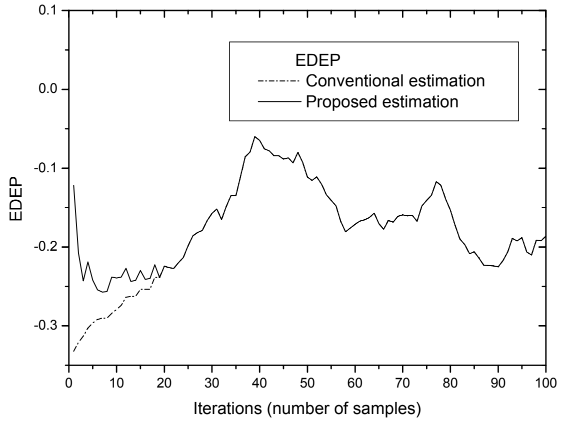

그림 2. 초기 단계에서 EDEP 추정값 비교
Fig. 2. The comparison of EDEP estimation in the initial part of iteration.

except a slight difference in the initial state. For better observation, Fig. 2 is depicted focused in the initial part of iteration $(1 \le k < N)$. Consequently the two methods produce different estimation results initially but getting closer to each other until the iteration number 20, and then in the steady state $(k \ge N)$ they yield exactly the same estimation results. It can be considered that the difference comes according to the occupation state of the buffer for the data block for the summation operation.

The trace of gradient is compared in Fig. 3 estimated by the two methods of the block-processing method by (20) and (21), and the recursive estimation by (22), (23), (24) and (25). For fair comparison, convergence curves for all $(L=11)$ weight gradients are needed to be presented, but only that of the center tap weight is compared just for the limited space of this paper. Similar results are observed in Fig. 3. As the EDEP algorithm converges, fluctuations of the gradient decrease and after about the iteration number 4000 the gradient concentrates at about zero. When we compare Fig. 1 and 3, it can be observed that the EDEP and its gradient converge accordingly except a slight difference in the initial state $(1 \le k < 20)$.

Figure 4 shows the comparison of the two gradient estimation methods for the early stage of iteration for better observation. As observed in EDEP estimation in Fig. 2, the two methods produce the same gradient estimation results in the steady state $(k \ge 20)$ though slightly different traces are observed in the initial stage $(k < 20)$.

On the other hand, letting the two constants $\frac{1}{2N^2\sigma^2}\frac{1}{2\sigma\sqrt{\pi}}$ in (20) and $\frac{1}{\sigma^2 N}\frac{1}{\sigma\sqrt{2\pi}}$ in (21) be C_1 and C_2, respectively, the number of multiplications for the conventional estimation of (20)

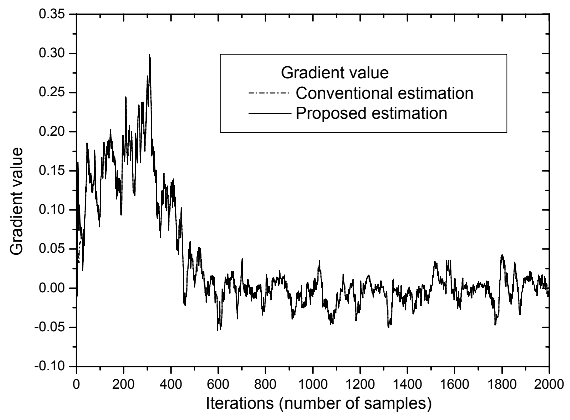

그림 3. 중앙 가중치 갱신을 위한 기울기 추정값
Fig. 3. Gradient estimation for the center weight update.

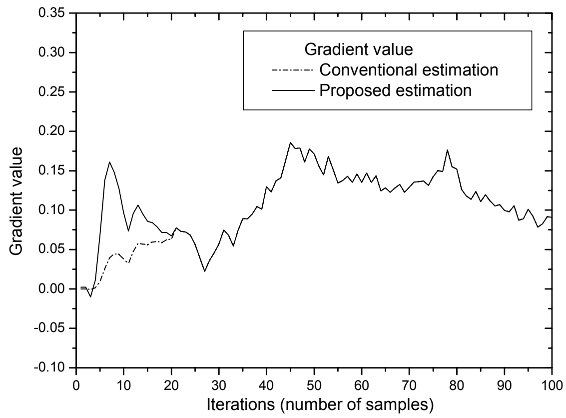

그림 4. 초기 단계에서 기울기 비교
Fig. 4. Gradient comparison in the initial part of iteration.

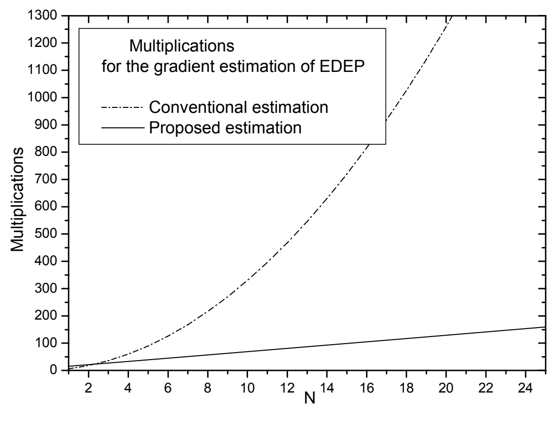

그림 5. 데이터 블록 크기 N에 대한 곱셈 연산량 비교
Fig. 5. Comparison of the number of multiplications for the data block size N.

and (21) is $3N^2 + 3N$. Similarly, defining $\frac{2}{N^2}\frac{1}{2\sigma\sqrt{\pi}}\frac{1}{2\sigma^2}$ in (24) as a constant C_3 and $\frac{1}{N}\frac{1}{\sigma\sqrt{2\pi}}\frac{1}{\sigma^2}$ in (25) as a constant C_4, the number of multiplications for the proposed estimation of (24) and (25) becomes $6N + 9$. The difference of multiplication numbers with respect to the data block size N is depicted in Fig. 5. It is noticed that the number of multiplications of the conventional method increases cubically. It is a very rapid way of growing. But that of the proposed method increases simply linearly with a very low slope. When $N = 20$, the number of multiplications of the proposed method is 129, while that of the conventional method is 1260.

VII. CONCLUSIONS

The EDEP algorithm known to be robust to impulsive noise has a heavy computational complexity of $O(N^2)$ due to the double summation operations at each iteration time. In this paper, the components of the EDEP are shown to be estimated recursively. Through this approach, the computational complexity for the estimation of the EDEP and gradient can be reduced to $O(N)$. In the performance test, the proposed EDEP and its gradient estimation that have significantly reduced complexity yield the same estimation results in the steady state as the conventional block-processing method. From the results, we can conclude that the proposed adaptive method can be a great candidate for practical adaptive signal processing under impulsive noise environments.

REFERENCES

[1] S. Haykin, Adaptive Filter Theory. Prentice Hall. 4th edition, 2001.
[2] B. Widrow, Adaptive Signal processing, Prentice – Hall, New Jersey, 1985.
[3] J. Proakis, Digital Communications, McGraw – Hill, 2nd edition, 1989.
[4] J. Principe, D. Xu and J. Fisher, Information Theoretic Learning in: S. Haykin, Unsupervised Adaptive Filtering, Wiley, (New York, USA), pp. 265-319, 2000.
[5] E. Parzen, "On the estimation of a probability density function and the mode," Ann. Math. Stat. Vol. 33, p.1065, 1962.
[6] S. Kullback, Information Theory and Statistics, Dover Publications. (New York, USA), 1968.
[7] K. Jeong, J. Xu, D. Erdogmus, and J. Principe, "A new classifier based on information theoretic learning with unlabeled data," Neural Networks, Vol. 18, pp. 719-726, 2005.
[8] N. Kim, K. Jung, and L. Yang, "Maximization of zero-error probability for adaptive channel equalization," JCN, Vol. 12. pp. 459-465, Oct. 2010.
[9] N. Kim and Y. Hwang, "Communication equalizer algorithms with decision feedback based on error probability," Journal of the Korea Academia -Industrial cooperation society, Vol. 12, pp. 2390-2395, May, 2011.

저 자 소 개

김 남 용(정회원)
1986년 연세대학교 전자공학과 학사 졸업.
1988년 연세대학교 전자공학과 석사 졸업.
1991년 연세대학교 전자공학과 박사 졸업.
<주관심분야 : 통신신호처리, 후각센싱알고리듬>