❏ 548

# Blind Signal Processing Algorithms based on Recursive Gradient Estimation

**Namyong Kim\*, Mingoo Kang\*\***

\* Division of Electronic, Information and Commnication Engineering, Kangwon National University, Korea
\*\* Division of Information and Telecommnication Engineering, Hanshin University, Korea

| Article Info | ABSTRACT |
|---|---|
| | Blind algorithms based on the Euclidean distance (ED) between the output distribution function and a set of Dirac delta functions have a heavy computational burden of $O(NM)$ due to some double summation operations for the sample size $N$ and $M$ symbol points. In this paper, a recursive approach to the estimation of the ED and its gradient is proposed to reduce the computational complexity for efficient implementation of the algorithm. The ED of the algorithm is comprised of information potentials (IPs), and the IPs at the next iteration can be calculated recursively based on the currently obtained IPs. Utilizing the recursively estimated IPs, the next step gradient for the weight update of the algorithm can be estimated recursively with the present gradient. With this recursive approach, the computational complexity of gradient calculation has only $O(N)$. The simulation results show that the proposed gradient estimation method holds significantly reduced computational complexity keeping the same performance as the block processing method.<br><br> |

*Corresponding Author:*

Namyong Kim,
Division of Electronic, Information and Communication Engineering,
Kangwon National University,
346 Jungang Road, Samcheok city, Kangwon, 245-711, Korea.
Email: namyong@kangwon.ac.kr

## 1. INTRODUCTION

The information-theoretic learning (ITL) method has been developed based on a combination of probability density functions (PDFs) and a procedure to compute information potential (IP) that is defined based on the concept that data samples interact with each other as a pair of physical particles in a potential field [1]. The construction of PDFs is done by the kernel density estimation method employing Gaussian kernel [2].

The ITL method has yielded many performance criteria such as Kullback-Leibler (KL) divergence to estimate mutual information and Euclidian distance (ED) between two PDFs to measure their similarity [3], [4]. Through minimizing the KL divergence or the ED, algorithms for supervised training have been developed for classification in biomedical problems [4], [5].

In blind signal processing where training data sets are not available the PDF of desired signal needs be generated without knowing the exact desired signal. For communication applications where the symbol points to be transmitted are identically and independently distributed, the desired PDF can be replaced with a set of Dirac-delta functions [6]. The blind algorithm developed based on the ED criterion between the PDF of output samples and Dirac-delta functions in place of the desired PDF has shown superior learning performance in various environments.

However, in the ED estimation of the blind algorithm, there is a computational problem that hinders the efficient implementation of the algorithm. The method requires heavy computational complexity due to

double summation operations at each iteration time. In our initial work [7], a recursive approach to the ED estimation has been introduced as an efficient method reducing the computational complexity.

The ED estimation can be used to check how well the output PDF matches the desired PDF. That is, the estimated ED can be used as an indicator of the convergence state of the employed algorithm. Therefore the recursive estimation of ED in [7] does not play the role of weight adjustment in the blind algorithm. For that purpose, gradient estimation essential for the weight update process of the blind algorithm is needed to be dealt with.

After derivatives of the ED are taken, the relationship between the current and the next time components produced from the derivation is investigated to figure out whether gradient estimation can be carried out recursively with reduced computational complexity. And then it is experimented to prove that the proposed method produces the same equalization performance with the significantly reduced computations compared to the block processing method introduced in [6].

This paper is organized as follows. Section 2 presents the definition of Euclidean distance between the PDF of the equalizer output samples and a set of Dirac delta functions. The estimation of the ED is shown to be able to be carried out recursively in Section 3, and the recursive calculation of the gradient of the ED is proposed for the weight update in Section 4. Section 5 reports simulation results and discussions. Finally, concluding remarks are presented in Section6.

## 2. ED BETWEEN THE EQUALIZER OUTPUT AND A SET OF DIRAC DELTA FUNCTIONS

The similarity measure between two distribution functions of the desired symbols and equalizer output samples can be estimated through the ED calculation. The ED between $f_D$ the PDF of the desired symbol and $f_Y$ the PDF of the equalizer output is defined as

$$ED(f_D, f_Y) = \int [f_D(\alpha) - f_Y(\alpha)]^2 d\alpha = \int f_D^2(\alpha) d\alpha + \int f_Y^2(\alpha) d\alpha - 2\int f_D(\alpha) f_Y(\alpha) d\alpha \quad (1)$$

The output PDF can be constructed by the kernel density estimation method for output samples at time k $\{y_k, y_{k-1}, ..., y_{k-N+1}\}$ with the sample size $N$ [2].

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

The PDF of the desired symbols can be expressed as Dirac delta functions when the $M$ symbol points $\{s_1, s_2, ..., s_M\}$ are equally likely [6].

$$f_D(d) = \frac{1}{M}[\delta(d - s_1) + \delta(d - s_2) + ... + \delta(d - s_m) + ... + \delta(d - s_M)] \quad (3)$$

Substituting (2) and (3) into (1), each term of (1) becomes

$$\int f_D^2(\alpha) d\alpha = \frac{1}{M} \quad (4)$$

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

$$\int f_D(\alpha) f_Y(\alpha) d\alpha = \frac{1}{M} \sum_{m=1}^{M} f_Y(s_m) = \frac{1}{M}\frac{1}{N} \sum_{m=1}^{M} \sum_{i=k-N+1}^{k} \frac{1}{\sigma\sqrt{2\pi}} \exp[\frac{-(s_m - y_i)^2}{2\sigma^2}] \quad (6)$$

For convenience, the equation (5) and (6) are defined as $A_k$ and $B_k$, respectively, then $ED(f_D, f_Y)_k$ at time k becomes

$$ED(f_D, f_Y)_k = \frac{1}{M} + A_k - B_k \tag{7}$$

It is noticeable that $A_k$ in (5) and $B_k$ in (6) are calculated through double summation operations so that this block processing method demands heavy computational burden.

One of the basic ideas of the ITL method is that data samples can be considered as physical particles in physics. That is, the sample values $y_i$ and $y_j$ in (5) are considered as information particles placed at the locations of $y_i$ an $y_j$. The Gaussian kernel $\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(y_j - y_i)^2}{4\sigma^2}]$ produces an exponential decay with the distance between the particles. This leads us to consider the Gaussian kernel as a potential field inducing interaction among the information particles. Then $\sum_{j=k-N+1}^{k} \frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(y_j - y_i)^2}{4\sigma^2}]$ is corresponding to the sum of interactions on the i th particle, and $\frac{1}{N^2}\sum_{i=k-N+1}^{k}\sum_{j=k-N+1}^{k}\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(y_j - y_i)^2}{4\sigma^2}]$ in (5) is the sum of all pairs of interactions. This overall potential energy is referred to as information potential [1].

According to the concept of information potential, the ED (7) can be regarded as a combination of information potentials; $\frac{1}{M}$ caused by the pairs of transmitted symbol points, $A_k$ by the pairs of output samples, and $B_k$ by the pairs of symbol points and output samples. When a new output sample $y_{k+1}$ comes into the combined potential field at time k+1, $A_{k+1}$ and $B_{k+1}$ of $ED(f_D, f_Y)_{k+1}$ are renewed by computing interactions among all corresponding sample pairs. We can observe that the old output sample $y_{k-N+1}$ leaves the field while the new sample $y_{k+1}$ comes into it. This indicates that the interactions between $y_{k-N+1}$ and the other samples are discarded while the new interactions between $y_{k+1}$ and the others are added to the combined potential field. It is noticeable that each new information potential at time k+1, $A_{k+1}$ and $B_{k+1}$ might be calculated just by adding new interactions related with $y_{k+1}$ to the current information potentials $A_k$ and $B_k$, and subtracting old interactions related with $y_{k-N+1}$ from the current information potentials. This points out that a simpler method for $ED(f_D, f_Y)$ calculation can be possible than the block processing method of carrying out double summation operations for (5) and (6) at each iteration time.

## 3. RECURSIVE ED ESTIMATION

During the time interval $1 \leq k \leq N$, only k output samples are available at time k. Therefore two cases will be dealt with, that is, the $A_k^I$ and $B_k^I$ are for the initial state $1 \leq k \leq N$, and $A_k^S$ and $B_k^S$ are for the steady state $k > N$, as follows.

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

$$B_k^I = \frac{2}{kN}\sum_{i=i}^{k}\sum_{j=1}^{N}\frac{1}{\sigma\sqrt{2\pi}}\exp[\frac{-(s_m - y_i)^2}{2\sigma^2}] \tag{9}$$

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

$$B_k^S = \frac{1}{M} \frac{1}{N} \sum_{m=1}^{M} \sum_{i=k-N+1}^{k} \frac{1}{\sigma\sqrt{2\pi}} \exp[\frac{-(s_m - y_i)^2}{2\sigma^2}] \tag{11}$$

In the initial state $1 \le k \le N$, the terms of $ED(f_D, f_Y)_{k+1}$, $A_{k+1}^I$ and $B_{k+1}^I$ can be estimated recursively as

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

$$= \frac{k^2}{(k+1)^2 k^2} \left[ \sum_{i=1}^{k} \sum_{j=1}^{k+1} \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_j - y_i)^2}{4\sigma^2}] + \sum_{j=1}^{k+1} \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_j - y_{k+1})^2}{4\sigma^2}] \right]$$

$$= \frac{k^2}{(k+1)^2 k^2} \left[ \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_j - y_i)^2}{4\sigma^2}] + \sum_{i=1}^{k} \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_{k+1} - y_i)^2}{4\sigma^2}] \right]$$

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

$$= \frac{k^2}{(k+1)^2} A_k^I + \frac{1}{(k+1)^2} \sum_{i=1}^{k} \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_{k+1} - y_i)^2}{4\sigma^2}] + \frac{1}{(k+1)^2} \left[ \sum_{j=1}^{k} \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_j - y_{k+1})^2}{4\sigma^2}] \right.$$

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

$$A_{k+1}^I = \frac{k^2}{(k+1)^2} A_k^I + \frac{2}{(k+1)^2} \sum_{i=1}^{k} \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_{k+1} - y_i)^2}{4\sigma^2}] + \frac{1}{(k+1)^2} \frac{1}{2\sigma\sqrt{\pi}} \tag{13}$$

where the equality $\exp[\frac{-(y_{k+1} - y_i)^2}{4\sigma^2}] = \exp[\frac{-(y_i - y_{k+1})^2}{4\sigma^2}]$ is utilized in (12) for (13). Similarly,

$$B_{k+1}^I = \frac{2}{(k+1)M} \sum_{i=i}^{k+1} \sum_{m=1}^{M} \frac{1}{\sigma\sqrt{2\pi}} \exp[\frac{-(s_m - y_i)^2}{2\sigma^2}]$$

$$= \frac{2k}{(k+1)kM} \sum_{i=i}^{k} \sum_{m=1}^{M} \frac{1}{\sigma\sqrt{2\pi}} \exp[\frac{-(s_m - y_i)^2}{2\sigma^2}]$$

$$+ \frac{2}{(k+1)M} \sum_{m=1}^{M} \frac{1}{\sigma\sqrt{2\pi}} \exp[\frac{-(s_m - y_{k+1})^2}{2\sigma^2}]$$

$$= \frac{k}{(k+1)} B_k^I + \frac{2}{(k+1)M} \sum_{m=1}^{M} \frac{1}{\sigma\sqrt{2\pi}} \exp[\frac{-(s_m - y_{k+1})^2}{2\sigma^2}] \tag{14}$$

The equations (13) and (14) show that $ED(f_D, f_Y)_{k+1}$ in the initial state can be calculated with $y_{k+1}$ and the current ED terms $A_k^I$ and $B_k^I$.

In the steady state, each component $A_{k+1}^S$ and $B_{k+1}^S$ of $ED(f_D, f_Y)_{k+1}$ can be divided into the $y_{k+1}$ related terms, the $y_{k-N+1}$ related terms and the remaining.

$$
\begin{aligned}
A_{k+1}^S &= \frac{1}{N^2} \sum_{i=k-N+2}^{k+1} \sum_{j=k-N+2}^{k+1} \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_j - y_i)^2}{4\sigma^2}] \\
&= \frac{1}{N^2} \sum_{i=k-N+1}^{k} \left[ \sum_{j=k-N+1}^{k} \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_j - y_i)^2}{4\sigma^2}] \right. \\
&\quad \left. + \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_i - y_{k+1})^2}{4\sigma^2}] - \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_i - y_{k-N+1})^2}{4\sigma^2}] \right] \\
&\quad + \frac{1}{N^2} \left[ \sum_{j=k-N+1}^{k} \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_{k+1} - y_j)^2}{4\sigma^2}] + \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_{k+1} - y_{k+1})^2}{4\sigma^2}] \right. \\
&\quad \left. - \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_{k+1} - y_{k-N+1})^2}{4\sigma^2}] \right] - \frac{1}{N^2} \left[ \sum_{j=k-N+1}^{k} \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_{k-N+1} - y_j)^2}{4\sigma^2}] \right. \\
&\quad \left. + \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_{k-N+1} - y_{k+1})^2}{4\sigma^2}] - \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_{k-N+1} - y_{k-N+1})^2}{4\sigma^2}] \right] \quad (15)
\end{aligned}
$$

Utilizing the equity $\exp[\frac{-(y_{k+1} - y_{k-N+1})^2}{4\sigma^2}] = \exp[\frac{-(y_{k-N+1} - y_{k+1})^2}{4\sigma^2}]$ in (15), $A_{k+1}^S$ can be rewritten as

$$
\begin{aligned}
A_{k+1}^S &= A_k^S + \frac{2}{N^2} \sum_{j=k-N+1}^{k} G_{\sigma\sqrt{2}}(y_i - y_{k+1}) - \frac{2}{N^2} \sum_{j=k-N+1}^{k} \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_i - y_{k-N+1})^2}{4\sigma^2}] \\
&\quad - \frac{2}{N^2} \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_{k+1} - y_{k-N+1})^2}{4\sigma^2}] + \frac{2}{N^2} \frac{1}{2\sigma\sqrt{\pi}} \quad (16)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
B_{k+1}^S &= \frac{2}{NM} \sum_{i=k-N+2}^{k+1} \sum_{m=1}^{M} \frac{1}{\sigma\sqrt{2\pi}} \exp[\frac{-(s_m - y_i)^2}{2\sigma^2}] \\
&= \frac{2}{NM} \sum_{i=k-N+1}^{k} \sum_{m=1}^{M} \frac{1}{\sigma\sqrt{2\pi}} \exp[\frac{-(s_m - y_i)^2}{2\sigma^2}] \\
&\quad + \frac{2}{NM} \sum_{m=1}^{M} \frac{1}{\sigma\sqrt{2\pi}} \exp[\frac{-(s_m - y_{k+1})^2}{2\sigma^2}] - \frac{2}{NM} \sum_{m=1}^{M} \frac{1}{\sigma\sqrt{2\pi}} \exp[\frac{-(s_m - y_{k-N+1})^2}{2\sigma^2}] \\
&= B_k^S + \frac{2}{NM} \sum_{m=1}^{M} \left[ \frac{1}{\sigma\sqrt{2\pi}} \exp[\frac{-(s_m - y_{k+1})^2}{2\sigma^2}] - \frac{1}{\sigma\sqrt{2\pi}} \exp[\frac{-(s_m - y_{k-N+1})^2}{2\sigma^2}] \right] \quad (17)
\end{aligned}
$$

The equations (16) and (17) proves that $ED(f_D, f_Y)_{k+1}$ in the steady state can be calculated with interactions with $y_{k+1}$ and $y_{k-N+1}$, based on the current information potentials $A_k^S$ and $B_k^S$ [7].

In summary, the equations (13), (14), (16) and (17) are a recursive version of the block-processed calculation of $ED(f_D, f_Y)_{k+1}$. That is, information potentials in the Euclidean distance can be estimated

recursively. Furthermore, the computational complexity of its recursive version has only $O(N)$ which is compared apparently with $O(NM)$ of the block-processed ED.

## 4.    BLIND ALGORITHMS BASED ON RECURSIVE GRADIENT OF ED

The output $y_k$ at time $k$ of linear equalizer structures such as a tapped delay line (TDL) with L weights, the input vector $\mathbf{X}_k = [x_k, x_{k-1}, ..., x_{k-L+1}]^T$ and $\mathbf{W}_k = [w_{0,k}, w_{1,k}, ..., w_{L-1,k}]^T$ can be expressed as $y_k = \mathbf{W}_k^T \mathbf{X}_k$. Blind algorithms searching for its optimum weight vector are derived from the minimization process of performance criteria, for which the Euclidean distance $ED(f_D, f_Y)$ in (7) is employed in this paper. That is, the minimization of $\dfrac{1}{M} + A_k - B_k$ with respect to the equalizer weights can be carried out recursively by using the gradient of the criteria and the steepest descent method as

$$\mathbf{W}_{k+1} = \mathbf{W}_k - \mu \cdot \frac{\partial(A_k - B_k)}{\partial \mathbf{W}} \tag{18}$$

where $\dfrac{1}{M}$ is omitted because it's not a function of weight and $\mu$ is the step-size for convergence control.

When the gradient at time k $\dfrac{\partial(A_k - B_k)}{\partial \mathbf{W}} = \dfrac{\partial A_k}{\partial \mathbf{W}} - \dfrac{B_k}{\partial \mathbf{W}}$ is calculated by the block-processing method directly on (5) and (6), it is obtained as

$$\frac{\partial A_k}{\partial \mathbf{W}} = \frac{1}{2N^2\sigma^2} \sum_{i=k-N+1,}^{k} \sum_{j=k-N+1}^{k} (y_j - y_i) \cdot \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_j - y_i)^2}{4\sigma^2}] \cdot (\mathbf{X}_i - \mathbf{X}_j) \tag{19}$$

$$\frac{\partial B_k}{\partial \mathbf{W}} = \frac{2}{NM\sigma^2} \sum_{i=k-N+1,}^{k} \sum_{m=1}^{M} (s_m - y_i) \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp[\frac{-(s_m - y_i)^2}{2\sigma^2}] \cdot \mathbf{X}_i \tag{20}$$

The weight update equation (18) with gradients (19) and (20) has been introduced in the work [6]. In this section, a recursive version of the algorithm is proposed by using a similar approach that was used in the recursive ED estimation introduced in the previous section.

It can be noticed that the gradients (19) and (20) are described for the steady state so that they are corresponding to $\dfrac{\partial A_k^S}{\partial \mathbf{W}}$ and $\dfrac{\partial B_k^S}{\partial \mathbf{W}}$, respectively. The gradients in the initial state $\dfrac{\partial A_k^I}{\partial \mathbf{W}}$ and $\dfrac{\partial B_k^I}{\partial \mathbf{W}}$ are written as

$$\frac{\partial A_k^I}{\partial \mathbf{W}} = \frac{1}{2k^2\sigma^2} \sum_{i=1}^{k} \sum_{j=1}^{k} (y_j - y_i) \cdot \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_j - y_i)^2}{4\sigma^2}] \cdot (\mathbf{X}_i - \mathbf{X}_j) \tag{21}$$

$$\frac{\partial B_k^I}{\partial \mathbf{W}} = \frac{2}{kM\sigma^2} \sum_{i=1}^{k} \sum_{m=1}^{M} (s_m - y_i) \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp[\frac{-(s_m - y_i)^2}{2\sigma^2}] \cdot \mathbf{X}_i \tag{22}$$

Then

$$\frac{\partial A_{k+1}^I}{\partial \mathbf{W}} = \frac{1}{2(k+1)^2\sigma^2} \sum_{i=1}^{k+1} \sum_{j=1}^{k+1} (y_j - y_i) \cdot \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_j - y_i)^2}{4\sigma^2}] \cdot (\mathbf{X}_i - \mathbf{X}_j) \tag{23}$$

$$\frac{\partial B^I_{k+1}}{\partial \mathbf{W}} = \frac{2}{(k+1)M\sigma^2} \sum_{i=1}^{k+1}\sum_{m=1}^{M}(s_m - y_i) \cdot \frac{1}{\sigma\sqrt{2\pi}}\exp[\frac{-(s_m - y_i)^2}{2\sigma^2}]\cdot \mathbf{X}_i \tag{24}$$

Using the similar approach as in (12), the gradients $\dfrac{\partial A^I_{k+1}}{\partial \mathbf{W}}$ and $\dfrac{\partial B^I_{k+1}}{\partial \mathbf{W}}$ can be divided into the terms related with $y_{k+1}$ and the remaining.

$$\begin{aligned}
\frac{\partial A^I_{k+1}}{\partial \mathbf{W}} =& \frac{k^2}{(k+1)^2 2k^2\sigma^2}\sum_{i=1}^{k}\left[\sum_{j=1}^{k+1}(y_j - y_i)\cdot\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(y_j - y_i)^2}{4\sigma^2}]\cdot(\mathbf{X}_i - \mathbf{X}_j)\right] \\
&+ \frac{k^2}{(k+1)^2 2k^2\sigma^2}\sum_{j=1}^{k+1}(y_j - y_{k+1})\cdot\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(y_j - y_{k+1})^2}{4\sigma^2}]\cdot(\mathbf{X}_{k+1} - \mathbf{X}_j) \\
=& \frac{k^2}{(k+1)^2 2k^2\sigma^2}\sum_{i=1}^{k}\left[\sum_{j=1}^{k}(y_j - y_i)\cdot\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(y_j - y_i)^2}{4\sigma^2}]\cdot(\mathbf{X}_i - \mathbf{X}_j)\right] \\
&+ \frac{k^2}{(k+1)^2 2k^2\sigma^2}\sum_{i=1}^{k}(y_{k+1} - y_i)\cdot\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(y_{k+1} - y_i)^2}{4\sigma^2}]\cdot(\mathbf{X}_i - \mathbf{X}_{k+1}) \\
&+ \frac{k^2}{(k+1)^2 2k^2\sigma^2}\sum_{j=1}^{k}(y_j - y_{k+1})\cdot\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(y_j - y_{k+1})^2}{4\sigma^2}]\cdot(\mathbf{X}_{k+1} - \mathbf{X}_j) \\
&+ \frac{k^2}{(k+1)^2 2k^2\sigma^2}(y_{k+1} - y_{k+1})\cdot\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(y_{k+1} - y_{k+1})^2}{4\sigma^2}]\cdot(\mathbf{X}_{k+1} - \mathbf{X}_{k+1}) \\
=& \frac{k^2}{(k+1)^2}\frac{\partial A^I_k}{\partial \mathbf{W}} + \frac{1}{(k+1)^2\sigma^2}\sum_{i=1}^{k}(y_{k+1} - y_i)\cdot\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(y_{k+1} - y_i)^2}{4\sigma^2}]\cdot(\mathbf{X}_i - \mathbf{X}_{k+1})
\end{aligned} \tag{25}$$

The equation (25) indicates that the next gradient can divided into the term of the current gradient and the other term of $y_{k+1}$.

On the other hand,

$$\begin{aligned}
\frac{\partial B^I_{k+1}}{\partial \mathbf{W}} =& \frac{2}{(k+1)M\sigma^2}\left[\sum_{i=1}^{k}\sum_{m=1}^{M}(s_m - y_i)\cdot\frac{1}{\sigma\sqrt{2\pi}}\exp[\frac{-(s_m - y_i)^2}{2\sigma^2}]\cdot\mathbf{X}_i\right. \\
&\left.+ \sum_{m=1}^{M}(s_m - y_{k+1})\cdot\frac{1}{\sigma\sqrt{2\pi}}\exp[\frac{-(s_m - y_{k+1})^2}{2\sigma^2}]\cdot\mathbf{X}_{k+1}\right] \\
=& \frac{2}{(k+1)M\sigma^2}\left[\frac{Mk\sigma^2}{2}\cdot\frac{\partial B^I_k}{\partial \mathbf{W}} + \sum_{m=1}^{M}(s_m - y_{k+1})\cdot\frac{1}{\sigma\sqrt{2\pi}}\exp[\frac{-(s_m - y_{k+1})^2}{2\sigma^2}]\cdot\mathbf{X}_{k+1}\right] \\
=& \frac{k}{k+1}\cdot\frac{\partial B^I_k}{\partial \mathbf{W}} + \frac{2}{(k+1)M\sigma^2}\sum_{m=1}^{M}(s_m - y_{k+1})\cdot\frac{1}{\sigma\sqrt{2\pi}}\exp[\frac{-(s_m - y_{k+1})^2}{2\sigma^2}]\cdot\mathbf{X}_{k+1}
\end{aligned} \tag{26}$$

Through (25) and (26), the initial state gradients $\dfrac{\partial A^I_{k+1}}{\partial \mathbf{W}}$ and $\dfrac{\partial B^I_{k+1}}{\partial \mathbf{W}}$ can be obtained recursively.

Similarly, the steady state gradients $\dfrac{\partial A^S_{k+1}}{\partial \mathbf{W}}$ and $\dfrac{\partial B^S_{k+1}}{\partial \mathbf{W}}$ can be investigated whether they can be separated into the terms related with $y_{k+1}$ and the ones related with $y_{k-N+1}$.

From (19),

$$\frac{\partial A_{k+1}^S}{\partial \mathbf{W}} = \frac{1}{2N^2\sigma^2} \sum_{i=k-N+2,}^{k+1} \sum_{j=k-N+2}^{k+1} (y_j - y_i) \cdot \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_j-y_i)^2}{4\sigma^2}] \cdot (\mathbf{X}_i - \mathbf{X}_j)$$

$$= \frac{1}{2N^2\sigma^2} \sum_{i=k-N+1,}^{k} \sum_{j=k-N+2}^{k+1} (y_j - y_i) \cdot \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_j-y_i)^2}{4\sigma^2}] \cdot (\mathbf{X}_i - \mathbf{X}_j)$$

$$+ \frac{1}{2N^2\sigma^2} \sum_{j=k-N+2}^{k+1} (y_j - y_{k+1}) \cdot \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_j-y_{k+1})^2}{4\sigma^2}] \cdot (\mathbf{X}_{k+1} - \mathbf{X}_j)$$

$$- \frac{1}{2N^2\sigma^2} \sum_{j=k-N+2}^{k+1} (y_j - y_{k-N+1}) \cdot \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_j-y_{k-N+1})^2}{4\sigma^2}] \cdot (\mathbf{X}_{k-N+1} - \mathbf{X}_j)$$

$$= \frac{1}{2N^2\sigma^2} \sum_{i=k-N+1,}^{k} \left[ \sum_{j=k-N+1}^{k} (y_j - y_i) \cdot \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_j-y_i)^2}{4\sigma^2}] \cdot (\mathbf{X}_i - \mathbf{X}_j) \right.$$

$$+ (y_{k+1} - y_i) \cdot \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_{k+1}-y_i)^2}{4\sigma^2}] \cdot (\mathbf{X}_i - \mathbf{X}_{k+1})$$

$$\left. - (y_{k-N+1} - y_i) \cdot \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_{k-N+1}-y_i)^2}{4\sigma^2}] \cdot (\mathbf{X}_i - \mathbf{X}_{k-N+1}) \right] + \frac{1}{2N^2\sigma^2} \sum_{j=k-N+1}^{k} (y_j - y_{k+1})$$

$$\cdot \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_j-y_{k+1})^2}{4\sigma^2}] \cdot (\mathbf{X}_{k+1} - \mathbf{X}_j) + \frac{1}{2N^2\sigma^2} (y_{k+1} - y_{k+1})$$

$$\cdot \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_{k+1}-y_{k+1})^2}{4\sigma^2}] \cdot (\mathbf{X}_{k+1} - \mathbf{X}_{k+1})$$

$$- \frac{1}{2N^2\sigma^2} (y_{k-N+1} - y_{k+1}) \cdot \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_{k-N+1}-y_{k+1})^2}{4\sigma^2}] \cdot (\mathbf{X}_{k+1} - \mathbf{X}_{k-N+1})$$

$$- \frac{1}{2N^2\sigma^2} \sum_{j=k-N+1}^{k} (y_j - y_{k-N+1}) \cdot \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_j-y_{k-N+1})^2}{4\sigma^2}] \cdot (\mathbf{X}_{k-N+1} - \mathbf{X}_j)$$

$$- \frac{1}{2N^2\sigma^2} (y_{k+1} - y_{k-N+1}) \cdot \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_{k+1}-y_{k-N+1})^2}{4\sigma^2}] \cdot (\mathbf{X}_{k-N+1} - \mathbf{X}_{k+1})$$

$$+ \frac{1}{2N^2\sigma^2} (y_{k-N+1} - y_{k-N+1}) \cdot \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_{k-N+1}-y_{k-N+1})^2}{4\sigma^2}] \cdot (\mathbf{X}_{k-N+1} - \mathbf{X}_{k-N+1})$$

$$= \frac{\partial A_k^S}{\partial \mathbf{W}} + \frac{1}{2N^2\sigma^2} \sum_{j=k-N+1}^{k} (y_{k+1} - y_i) \cdot \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_{k+1}-y_i)^2}{4\sigma^2}] \cdot (\mathbf{X}_i - \mathbf{X}_{k+1})$$

$$- \frac{1}{2N^2\sigma^2} \sum_{j=k-N+1}^{k} (y_{k-N+1} - y_i) \cdot \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_{k-N+1}-y_i)^2}{4\sigma^2}] \cdot (\mathbf{X}_i - \mathbf{X}_{k-N+1})$$

$$+ \frac{1}{2N^2\sigma^2} \sum_{j=k-N+1}^{k} (y_j - y_{k+1}) \cdot \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_j-y_{k+1})^2}{4\sigma^2}] \cdot (\mathbf{X}_{k+1} - \mathbf{X}_j)$$

$$- \frac{1}{2N^2\sigma^2} (y_{k-N+1} - y_{k+1}) \cdot \frac{1}{2\sigma\sqrt{\pi}} \exp[\frac{-(y_{k-N+1}-y_{k+1})^2}{4\sigma^2}] \cdot (\mathbf{X}_{k+1} - \mathbf{X}_{k-N+1})$$

$$-\frac{1}{2N^2\sigma^2}\sum_{j=k-N+1}^{k}(y_j - y_{k-N+1})\cdot\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(y_j - y_{k-N+1})^2}{4\sigma^2}]\cdot(\mathbf{X}_{k-N+1}-\mathbf{X}_j)$$

$$-\frac{1}{2N^2\sigma^2}(y_{k+1} - y_{k-N+1})\cdot\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(y_{k+1} - y_{k-N+1})^2}{4\sigma^2}]\cdot(\mathbf{X}_{k-N+1}-\mathbf{X}_{k+1})$$

$$=\frac{\partial A_k^S}{\partial\mathbf{W}}+\frac{1}{N^2\sigma^2}\sum_{j=k-N+1}^{k}(y_{k+1} - y_i)\cdot\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(y_{k+1} - y_i)^2}{4\sigma^2}]\cdot(\mathbf{X}_i - \mathbf{X}_{k+1})$$

$$-\frac{1}{N^2\sigma^2}\sum_{j=k-N+1}^{k}(y_{k-N+1} - y_i)\cdot\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(y_{k-N+1} - y_i)^2}{4\sigma^2}]\cdot(\mathbf{X}_i - \mathbf{X}_{k-N+1})$$

$$-\frac{1}{N^2\sigma^2}(y_{k-N+1} - y_{k+1})\cdot\frac{1}{2\sigma\sqrt{\pi}}\exp[\frac{-(y_{k-N+1} - y_{k+1})^2}{4\sigma^2}]\cdot(\mathbf{X}_{k+1} - \mathbf{X}_{k-N+1}) \qquad (27)$$

From (27), we can observe that the next gradient of information potential $\dfrac{\partial A_{k+1}^S}{\partial\mathbf{W}}$ can be acquired recursively from the current gradient $\dfrac{\partial A_k^S}{\partial\mathbf{W}}$ and other terms related with $y_{k+1}$ and $y_{k-N+1}$. Similarly, $\dfrac{\partial B_{k+1}^S}{\partial\mathbf{W}}$ is analyzed to see if it can be obtained recursively as described below.

$$\frac{\partial B_{k+1}^S}{\partial\mathbf{W}}=\frac{2}{NM\sigma^2}\sum_{i=k-N+2}^{k+1}\sum_{m=1}^{M}(s_m - y_i)\cdot\frac{1}{\sigma\sqrt{2\pi}}\exp[\frac{-(s_m - y_i)^2}{2\sigma^2}]\cdot\mathbf{X}_i$$

$$=\frac{2}{NM\sigma^2}\left[\sum_{i=k-N+1}^{k}\sum_{m=1}^{M}(s_m - y_i)\cdot\frac{1}{\sigma\sqrt{2\pi}}\exp[\frac{-(s_m - y_i)^2}{2\sigma^2}]\cdot\mathbf{X}_i\right.$$

$$+\sum_{m=1}^{M}(s_m - y_{k+1})\cdot\frac{1}{\sigma\sqrt{2\pi}}\exp[\frac{-(s_m - y_{k+1})^2}{2\sigma^2}]\cdot\mathbf{X}_{k+1}$$

$$\left.-\sum_{m=1}^{M}(s_m - y_{k-N+1})\cdot\frac{1}{\sigma\sqrt{2\pi}}\exp[\frac{-(s_m - y_{k-N+1})^2}{2\sigma^2}]\cdot\mathbf{X}_{k-N+1}\right]$$

$$=\frac{\partial B_k^S}{\partial\mathbf{W}}+\frac{2}{NM\sigma^2}\sum_{m=1}^{M}\left[(s_m - y_{k+1})\cdot\frac{1}{\sigma\sqrt{2\pi}}\exp[\frac{-(s_m - y_{k+1})^2}{2\sigma^2}]\cdot\mathbf{X}_{k+1}\right.$$

$$\left.-(s_m - y_{k-N+1})\cdot\frac{1}{\sigma\sqrt{2\pi}}\exp[\frac{-(s_m - y_{k-N+1})^2}{2\sigma^2}]\cdot\mathbf{X}_{k-N+1}\right] \qquad (28)$$

In summary, the weight update in (18) can be carried out with the gradient $\dfrac{\partial ED(f_D, f_Y)}{\partial\mathbf{W}}$ that is recursively obtained through (25), (26), (27) and (28), as opposed to the block-processed gradients (19) and (20). Also the recursive equations show that the recursive version of the gradient has significantly reduced computational complexity of $O(N)$ while the block-processed ED has $O(NM)$.

In the following section, it will be investigated if the proposed ED estimation method in which the information potentials as components of the ED are calculated recursively yields the same estimation results as the block-processed ED estimation method does. And then it will be shown that the gradient of the ED for the weight update has equal results for the two cases of the block-processing method and the recursive calculation method.

## 5. RESULTS AND ANALYSIS

In this section, simulation results in the same environment of blind equalization using the MED2 algorithm as in [6] are illustrated for the two approaches of the block-processing and recursive methods to estimate the ED being minimized and the gradient of ED coming to zero.

The transmitted symbol points are $\{\pm3,\pm1\}$ so that the probability density function (3) becomes $f_D(d) = \frac{1}{4}[\delta(d+3)+\delta(d+1)+\delta(d-1)+\delta(d-3)]$. The channel impulse response $h_i$ is chosen to be $h_i = 1/2\{1+\cos[2\pi(i-2)/3.3]\}$, $i=1,2,3$ [8]. The data-block size $N=20$, the convergence parameter $\mu = 0.005$, and the kernel size is set to be $\sigma = 0.6$.

In Figure 1, the trace of Euclidean distance for the two methods is depicted where the block-processing method is by (4), (5) and (6), and the recursive calculation is by (13), (14), (16) and (17). As the MED2 algorithm designed to minimize the ED converges, the distance decreases rapidly with iterations. The two methods yield the same distance estimation performance showing no difference between the two methods, but we can observe a slight difference in Figure 2 that is focused in the initial part of iteration $(1 \le k < N)$. That is, the two methods produce different estimation results but getting closer to each other until the iteration number 20, and in the steady state $(k \ge N)$ they yield exactly the same estimation performance. We see that the difference during the time interval $1 \le k \le N$ is due to the number of output samples being available at time k.

For fair comparison, we need to present convergence curves for all $(L=11)$ weight gradients, but only the gradient of the center tap weight is presented just for the limited room of this paper. Similar results are observed in Figure 2 that shows the trace of center tap gradient calculated by the two methods; the block-processing method by (19) and (20), and the recursive calculation by (25), (26), (27) and (28). As the gradient converges by the MED2 algorithm, fluctuations of the gradient decrease and the gradient value concentrates at zero after the iteration number 4000, from which the ED converges also as in Figure 1. Though a slight difference is observed in the initial part of iteration $(1 \le k < 20)$ as shown in Figure 4, the two methods give the same gradient estimation results in the steady state $(k \ge 20)$ as observed in ED estimation.
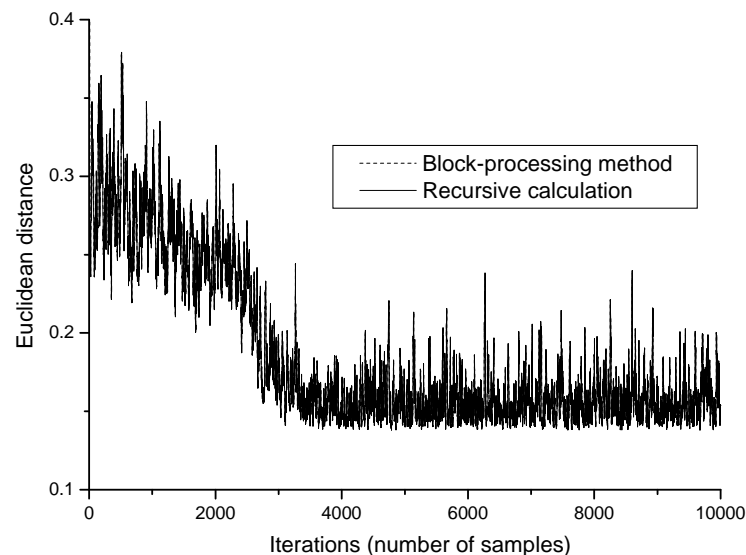


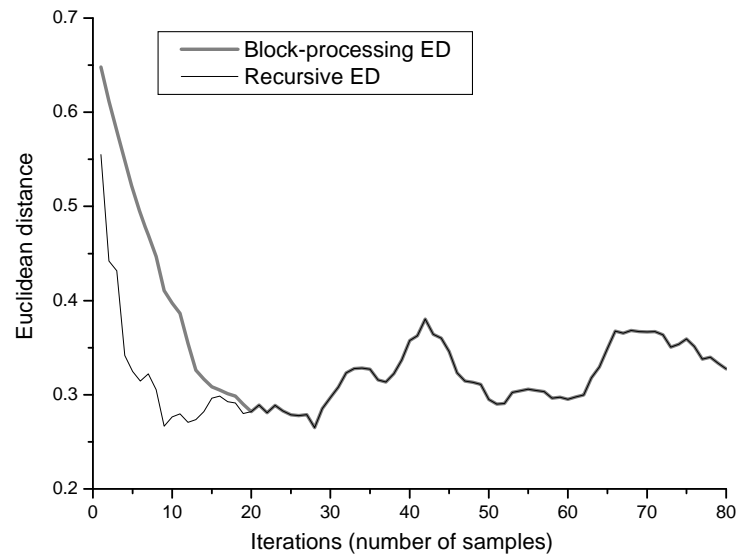Figure 1. Euclidean distance for the block-processing method and the recursive one

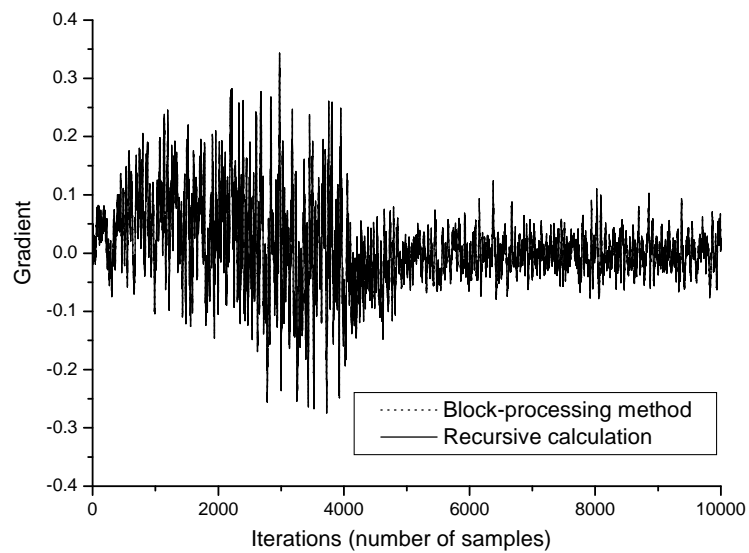Figure 2. The comparison of Euclidean distance in the initial part of iteration



Figure 3. Gradient for the center weight update for the two calculation methods; the block-processing method and the recursive one
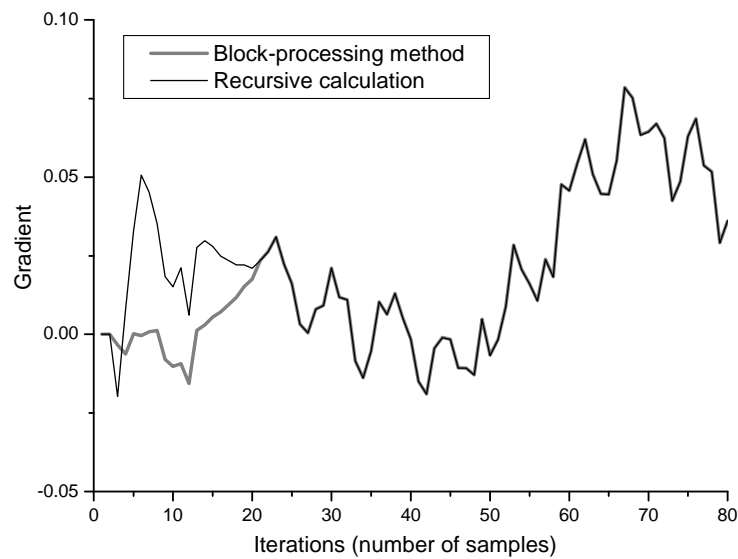
Figure 4. Gradient comparison in the initial part of iteration

The computational complexity analysis could be carried out in the aspects of time complexity and number of operations (usually multiplications), but in this paper the number of operations is analyzed according to conventional complexity analysis in the research field of equalization [9], [10].

For the sake of convenience and comparison, we consider the terms $\dfrac{1}{2N^2\sigma^3\sqrt{\pi}}\exp[.]$ in (19)

and $\dfrac{2}{NM\sigma^3}\dfrac{1}{\sqrt{2\pi}}\exp[.]$ in (20) as a constant. Then the number of multiplications for the conventional

gradient estimation of (19) and (20) is $2(N^2+NM)$. Similarly, the number of multiplications for the proposed estimation of (27) and (28) becomes $2(N+M)+6$. The comparison of multiplication operations with respect to the sample size $N$ is described in Figure 5. It is noticeable that the number of multiplications of the conventional method increases like a quadratic function with respect to $N$ while that of the proposed estimation remains linear. For $N=20$, $M=4$ which are used in the simulation, the complexity ratio of the conventional method to the proposed estimation is $960:54$.
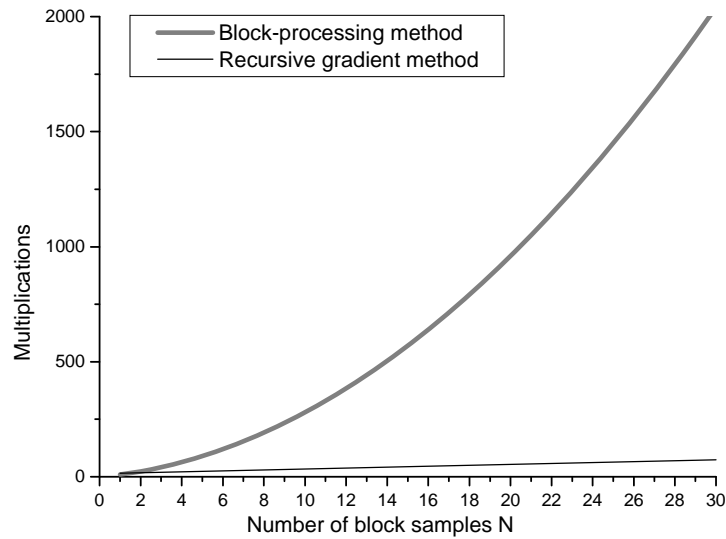
Figure 5. Comparison of the number of multiplications for sample size N

## 6.    CONCLUSION

The blind algorithm based on the criterion of ED minimization between the output distribution function and a set of Dirac delta functions located at the symbol points yields superior performance. However, it has a great deal of computational burden due to some double summation operations. For efficient implementation of the algorithm, a recursive approach to gradient calculation for reduced computational complexity is proposed in this paper. The Euclidean distance to be minimized is comprised of information potentials, and it can be calculated recursively based on the currently obtained information potentials. Also, the weight update of the algorithm can be carried out recursively based on the current gradient. The computational complexity of each recursive approach to gradient calculation has only $O(N)$ which is significantly reduced when compared with $O(NM)$ of the block-processed method. From the simulation results showing that the proposed gradient estimation method in the steady state yields the same estimation results as the block-processing method, we can conclude that the proposed recursive gradient estimation method is significantly efficient with regard to the computational complexity that is crucial to efficient implementation.

**REFERENCES**
[1] J. Principe, D. Xu and J. Fisher, Information Theoretic Learning in: S. Haykin, *Unsupervised Adaptive Filtering*, Wiley, (New York, USA), pp. 265-319, 2000.
[2] E. Parzen, "On the estimation of a probability density function and the mode", *Ann. Math. Stat.* vol. 33, p.1065, 1962.
[3] S. Kullback, *Information Theory and Statistics*, Dover Publications. (New York, USA), 1968.
[4] 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.
[5] D. Erdogmus, Y. Rao and J. Principe, "Supervised training of adaptive systems with partially labeled data", *Proceedings of the International Conference on ASSP*, pp. v321-v324, Apr. 2005.
[6] N. Kim, "Euclidian distance minimization of probability density functions for blind equalization", *Journal of Communications and Networks*, vol. 12. pp. 399-405, 2010.
[7] A. Andonova1, N. Kim, M. Kang, S. Han, and K. Lee, "Euclidean distance estimation of distribution functions for blind signal processing", *Proceeding of ICONI 2014 Symposium*, Taipei, Taiwan, pp. 157-160, Dec. 2014.
[8] S. Haykin, *Adaptive Filter Theory,* Prentice Hall. (Upper Saddle River), 4th edition, 2001.
[9] S. Yassin, H. Tawfik, "Reduced complexity decision feedback channel equalizer using series expansion division", *AICT 2013: The Ninth Advanced International Conference on Telecommunications,* Rome, Italy, pp. 219-223, June 2013.

[10]  J. Park, Y. Whang, and K. Kim, "Low complexity MMSE-SIC equalizer employing time-domain recursion for OFDM  systems", *IEEE Signal Processing Letters*, vol. 15, pp. 633-636, Oct. 2008.

## BIOGRAPHIES OF AUTHORS
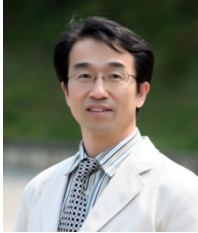
**Namyong Kim** received the B. S., M. S. and Ph.D degrees from Yonsei University, all in Electronic Engineering in 1986, 1988 and 1991, respectively. From 1992 to 1997, he was with Kwandong University, Korea. Currently, he serves as a Professorat Division of Electronics, Information and Communication Engineering, Kangwon National University, Korea. His research interests are in adaptive signal processingin mobile communications and odor sensing - identification algorithms.

**Mingoo Kang** received the B.S., M.S., and Ph.D. degrees in electronic engineering from Yonsei University, Seoul, Korea, in 1986, 1989 and 1994, respectively. He was a research engineer at Samsung Electronics from 1985 to 1997. He is in Hanshin University, Osan, Korea, from 2000 as a Professor in the Division of Inform. & Telecomm. His research interests include wireless communication devices, and Smart UX for Mobile TV & DTV.