A New Normalized Minimum Error Entropy Algorithm with Reduced Computational Complexity

Md. Zulfiquar A. Bhotto and Andreas Antoniou

Dept. of Elec. and Comp. Eng.
University of Victoria
Victoria, BC, Canada
Email: zbhotto@ece.uvic.ca
aantoniou@ieee.org

IEEE International Symposium on Circuits and Systems

May 24–27, 2009

Outline

- Mean-Square Error Criterion

Outline

- Mean-Square Error Criterion
- Entropy-of-Error Criterion
  - Minimum Entropy-of-Error (MEE) Algorithm (Erdogmus and Principe)
  - Minimum Entropy-of-Error with Self-Adjusting Step Size (MEE-SAS) Algorithm (Han, Erdogmus, and Principe)
  - Normalized Minimum Entropy-of-Error (NMEE) Algorithm (Han, Rao, Jeong, and Principe)
  - Comparison

University of Victoria

Outline

- Mean-Square Error Criterion
- Entropy-of-Error Criterion
  - Minimum Entropy-of-Error (MEE) Algorithm (Erdogmus and Principe)
  - Minimum Entropy-of-Error with Self-Adjusting Step Size (MEE-SAS) Algorithm (Han, Erdogmus, and Principe)
  - Normalized Minimum Entropy-of-Error (NMEE) Algorithm (Han, Rao, Jeong, and Principe)
  - Comparison
- Proposed NMEE Algorithm
- Simulation Results
- Conclusions

Mean-Square Error Criterion *Cont'd*

- Algorithms for conventional adaptive filters that use the MSE criterion yield an *optimal solution* only if the input signal is *Gaussian* and the desired signal is *a linear function* of the input signal.

University of Victoria

Mean-Square Error Criterion *Cont'd*

- Algorithms for conventional adaptive filters that use the MSE criterion yield an *optimal solution* only if the input signal is *Gaussian* and the desired signal is *a linear function* of the input signal.

  Such algorithms include the least mean-square error (LMS) algorithms and normalized LMS (NLMS) algorithms.

Mean-Square Error Criterion *Cont'd*

- Algorithms for conventional adaptive filters that use the MSE criterion yield an *optimal solution* only if the input signal is *Gaussian* and the desired signal is *a linear function* of the input signal.

  Such algorithms include the least mean-square error (LMS) algorithms and normalized LMS (NLMS) algorithms.

- Often in practice, the input signal *is not Guassian* and/or the desired signal is *not a linear function* of the input signal.

Mean-Square Error Criterion *Cont'd*

- Algorithms for conventional adaptive filters that use the MSE criterion yield an *optimal solution* only if the input signal is *Gaussian* and the desired signal is *a linear function* of the input signal.

  Such algorithms include the least mean-square error (LMS) algorithms and normalized LMS (NLMS) algorithms.

- Often in practice, the input signal *is not Guassian* and/or the desired signal is *not a linear function* of the input signal.
- Under these circumstances, better performance can be achieved by using algorithms based on *information theoretic* criteria, which can extract information based on higher-order statistics.

Information-Potential Function Criterion

- In MEE adaptive filters, the so-called *information potential function* of the error signal is maximized in order to achieve *maximum mutual information content* between the input and output signals.

University of Victoria

Information-Potential Function Criterion

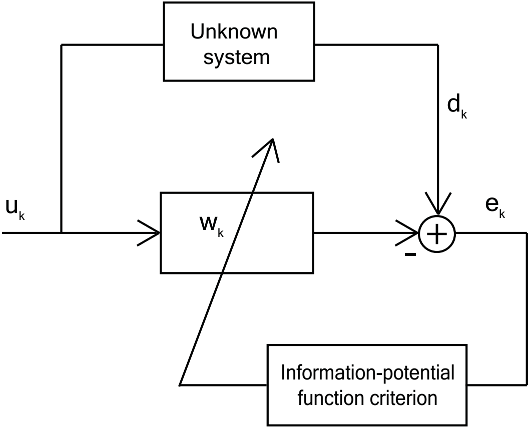
- In MEE adaptive filters, the so-called *information potential function* of the error signal is maximized in order to achieve *maximum mutual information content* between the input and output signals.
- A discretized estimate of the information potential function over k samples can be constructed as

$$V_k(x) = \frac{1}{(k)^2} \sum_{j=1}^{k} \sum_{i=1}^{k} \kappa_\sigma(x_j - x_i)$$

where $\kappa_\sigma(x)$ is the *Gaussian kernel*.

University of Victoria

Adaptive Filter Based on Information-Potential Function Criterion



Basic MEE Algorithm

- A recursive update equation for the information potential of the error can be constructed as

$$V_k(e) = \frac{1}{k^2}\left\{(k-1)^2 V_{k-1}(e) + 2\left[\sum_{i=1}^{k-1} \kappa_\sigma(e_k - e_i) + \kappa_\sigma(0)\right]\right\}$$

University of Victoria

Basic MEE Algorithm

- A recursive update equation for the information potential of the error can be constructed as

$$V_k(e) = \frac{1}{k^2}\left\{(k-1)^2 V_{k-1}(e) + 2\left[\sum_{i=1}^{k-1} \kappa_\sigma(e_k - e_i) + \kappa_\sigma(0)\right]\right\}$$

- In order to control the increasing memory requirement in the above estimator, a *forgetting factor* is introduced as

$$V_k(e) = (1-\lambda)V_{k-1}(e) + \frac{\lambda}{L}\sum_{i=k-L}^{k-1} \kappa_\sigma(e_k - e_i)$$

where L is the so-called *Parzen window length*.

University of Victoria

Basic MEE Algorithm *Cont'd*

- If we now use a forgetting factor $\lambda = 1$, the information potential simplifies to

$$V_k(e) = \frac{1}{L} \sum_{i=k-L}^{k-1} \kappa_\sigma(e_k - e_i)$$

Basic MEE Algorithm *Cont'd*

- If we now use a forgetting factor $\lambda = 1$, the information potential simplifies to

$$V_k(e) = \frac{1}{L} \sum_{i=k-L}^{k-1} \kappa_\sigma(e_k - e_i)$$

- The weight update equation assumes the form

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu \boldsymbol{\nabla} V_k(e)$$

where μ is the *step size* and

$$\boldsymbol{\nabla} V_k(e) = \frac{1}{L} \sum_{i=k-L}^{k-1} \kappa_\sigma(e_k - e_i)(\mathbf{u}_k - \mathbf{u}_i)$$

is the gradient of $V_k(e)$. ($\mathbf{u}_k$ and $\mathbf{u}_i$ are input vectors.)

University of Victoria

MEE Algorithm with Self-Adjusting Step Size

- An MEE adaptive filter with self-adjusting step size solves the optimization problem

$$\underset{\mathbf{w}}{\text{minimize}}\ [V(0) - V(e)]^2$$

where $V(0)$ is the value of $V_k(e)$ at $e = 0$, i.e., this is the *maximum value* of $V_k(e)$.

MEE Algorithm with Self-Adjusting Step Size

- An MEE adaptive filter with self-adjusting step size solves the optimization problem

$$\underset{\mathbf{w}}{\text{minimize}}\ [V(0) - V(e)]^2$$

where $V(0)$ is the value of $V_k(e)$ at $e = 0$, i.e., this is the *maximum value* of $V_k(e)$.

- The update equation for the coefficient vector is

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu[V(0) - V_k(e)]\boldsymbol{\nabla} V_k(e)$$

and the gradient of $V_k(e)$, $\boldsymbol{\nabla} V_k(e)$, is the same as for the basic MEE algorithm.

University of Victoria

Normalized MEE Algorithm

- The NMEE algorithm solves the optimization problem

$$\underset{\mathbf{w}}{\text{minimize}} \; ||\mathbf{w}_{k+1} - \mathbf{w}_k||^2$$

$$\text{subject to: } V(e_{p,k}) - V(0) = 0$$

where $V(e_{p,k})$ is the *a posteriori* information potential function.

University of Victoria

Normalized MEE Algorithm

- The NMEE algorithm solves the optimization problem

$$\underset{\mathbf{w}}{\text{minimize}} \ ||\mathbf{w}_{k+1} - \mathbf{w}_k||^2$$

$$\text{subject to:} \ V(e_{p,k}) - V(0) = 0$$

where $V(e_{p,k})$ is the *a posteriori* information potential function.

- The update equation for the weight vector assumes the form

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu \frac{e_{a,k} \boldsymbol{\nabla} V(e_{p,k})}{\boldsymbol{\nabla} V(e_{p,k})^T \mathbf{x}_k}$$

where $e_{a,k} = d_k - \mathbf{w}_k^T \mathbf{u}_k$ is the *a priori* error,

$$\boldsymbol{\nabla} V(e_{p,k}) = \frac{1}{2L\sigma^2} \sum_{i=k-L}^{k-1} (e_{p,k} - e_{p,i}) \, k_\sigma(e_{p,k} - e_{p,i}) \, (\mathbf{u}_k - \mathbf{u}_i)$$

and $e_{p,i} = d_i - \mathbf{w}_{k+1}^T \mathbf{u}_i$ is the *a posteriori* error.

Comparison

- The performance of the basic MEE algorithm depends on the power of the input signal and the kernel size σ.

Comparison

- The performance of the basic MEE algorithm depends on the power of the input signal and the kernel size σ.

  If the input power changes, the kernel size σ and step size μ *need to be adjusted* in order to achieve optimal performance.

Comparison

- The performance of the basic MEE algorithm depends on the power of the input signal and the kernel size σ.

  If the input power changes, the kernel size σ and step size μ *need to be adjusted* in order to achieve optimal performance.

- The MEE-SAS and NMEE algorithms *are faster and yield reduced residual error* relative to that of the basic MEE algorithm.

Comparison

- The performance of the basic MEE algorithm depends on the power of the input signal and the kernel size σ.

  If the input power changes, the kernel size σ and step size μ *need to be adjusted* in order to achieve optimal performance.

- The MEE-SAS and NMEE algorithms *are faster and yield reduced residual error* relative to that of the basic MEE algorithm.

- *Increased computational complexities* are associated with the MEE-SAS and NMEE algorithms.

University of Victoria

Comparison

- The performance of the basic MEE algorithm depends on the power of the input signal and the kernel size σ.

  If the input power changes, the kernel size σ and step size μ *need to be adjusted* in order to achieve optimal performance.

- The MEE-SAS and NMEE algorithms *are faster and yield reduced residual error* relative to that of the basic MEE algorithm.

- *Increased computational complexities* are associated with the MEE-SAS and NMEE algorithms.

- The performance of the NMEE algorithm is not affected by changes in the input signal power and is not sensitive to changes in the kernel size.

University of Victoria

Proposed NMEE Algorithm

- The proposed NMEE algorithm is derived by minimizing the *Euclidean distance* between the weight vectors at iterations k and $k+1$ subject to the constraint that the *a posteriori* information potential function is equal to V(0).

Proposed NMEE Algorithm

- The proposed NMEE algorithm is derived by minimizing the *Euclidean distance* between the weight vectors at iterations k and $k+1$ subject to the constraint that the *a posteriori* information potential function is equal to V(0).
- The associated optimization problem can be stated as

$$\underset{\mathbf{z}}{\text{minimize}} \; ||\mathbf{w}_{k+1} - \mathbf{w}_k||^2$$

$$\text{subject to: } \; V(0) - V(e_{p,k}) = 0$$

where

$$V(e_{p,k}) = \frac{1}{L} \sum_{i=k-L}^{k-1} k_\sigma(e_{p,k} - e_{p,i})$$

is the *a posteriori* information potential function.

University of Victoria

Proposed NMEE Algorithm *Cont'd*

- Solving the associated optimization problem, the update equation can be obtained as

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \frac{\sum_{i=k-L}^{k-1} \nu_{k,i}\mathbf{u}_{k,i} \sum_{i=k-L}^{k-1}(e_{a,k} - e_{a,i})}{\left(\sum_{i=k-L}^{k-1} \nu_{k,i}\mathbf{u}_{k,i}\right)^T \left(\sum_{i=k-L}^{k-1} \mathbf{u}_{k,i}\right)}$$

where $\mathbf{u}_{k,i} = \mathbf{u}_k - \mathbf{u}_i$ and $\nu_{k,i} = (e_{p,k} - e_{p,i})k_\sigma(e_{p,k} - e_{p,i})$.

Proposed NMEE Algorithm *Cont'd*

- Solving the associated optimization problem, the update equation can be obtained as

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \frac{\sum_{i=k-L}^{k-1} \nu_{k,i} \mathbf{u}_{k,i} \sum_{i=k-L}^{k-1} (e_{a,k} - e_{a,i})}{\left(\sum_{i=k-L}^{k-1} \nu_{k,i} \mathbf{u}_{k,i}\right)^T \left(\sum_{i=k-L}^{k-1} \mathbf{u}_{k,i}\right)}$$

where $\mathbf{u}_{k,i} = \mathbf{u}_k - \mathbf{u}_i$ and $\nu_{k,i} = (e_{p,k} - e_{p,i}) k_\sigma (e_{p,k} - e_{p,i})$.

- For a sufficiently small kernel size and window length, the scaling factor $\nu_{k,i}$ can be assumed to be constant over the Parzen window and, therefore, it can be factored out in the numerator and denominator to obtain the *simplified update equation*

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \frac{\bar{\mathbf{u}}_{k,i}}{\bar{\mathbf{u}}_{k,i}^T \bar{\mathbf{u}}_{k,i}} \bar{e}_{a,k,i}$$

where $\bar{\mathbf{u}}_{k,i} = \sum_{i=k-L}^{k-1} \mathbf{u}_{k,i}$, and $\bar{e}_{a,k,i} = \sum_{i=k-L}^{k-1} (e_{a,k} - e_{a,i})$.

University of Victoria

Proposed NMEE Algorithm *Cont'd*

- A step size μ can be introduced in the update equation as

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu \frac{\bar{\mathbf{u}}_{k,i}}{\bar{\mathbf{u}}_{k,i}^T \bar{\mathbf{u}}_{k,i}} \bar{e}_{a,k,i}$$

to control the learning rate and the residual error.

University of Victoria

- A step size μ can be introduced in the update equation as

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu \frac{\bar{\mathbf{u}}_{k,i}}{\bar{\mathbf{u}}_{k,i}^T \bar{\mathbf{u}}_{k,i}} \bar{e}_{a,k,i}$$

to control the learning rate and the residual error.

- The proposed NMEE algorithm will converge for values of μ in the range $0 < \mu < 2$ as can be easily verified.

Proposed NMEE Algorithm *Cont'd*

- A step size μ can be introduced in the update equation as

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu \frac{\bar{\mathbf{u}}_{k,i}}{\bar{\mathbf{u}}_{k,i}^T \bar{\mathbf{u}}_{k,i}} \bar{e}_{a,k,i}$$

to control the learning rate and the residual error.

- The proposed NMEE algorithm will converge for values of μ in the range $0 < \mu < 2$ as can be easily verified.

- It will converge to the solution only when

$$\bar{e}_{a,k,i} = \sum_{i=k-L}^{k-1} (e_{a,k} - e_{a,i}) = 0$$

which will ensure that *the maximum value* of the information potential function is achieved.

University of Victoria

Proposed NMEE Algorithm *Cont'd*

- A step size μ can be introduced in the update equation as

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu \frac{\bar{\mathbf{u}}_{k,i}}{\bar{\mathbf{u}}_{k,i}^T \bar{\mathbf{u}}_{k,i}} \bar{e}_{a,k,i}$$

to control the learning rate and the residual error.

- The proposed NMEE algorithm will converge for values of μ in the range $0 < \mu < 2$ as can be easily verified.

- It will converge to the solution only when

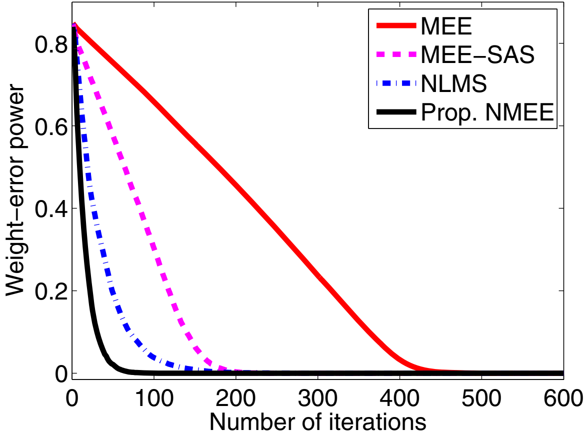
$$\bar{e}_{a,k,i} = \sum_{i=k-L}^{k-1} (e_{a,k} - e_{a,i}) = 0$$

which will ensure that *the maximum value* of the information potential function is achieved.

- The simplified update equation *reduces the computational complexity* significantly.

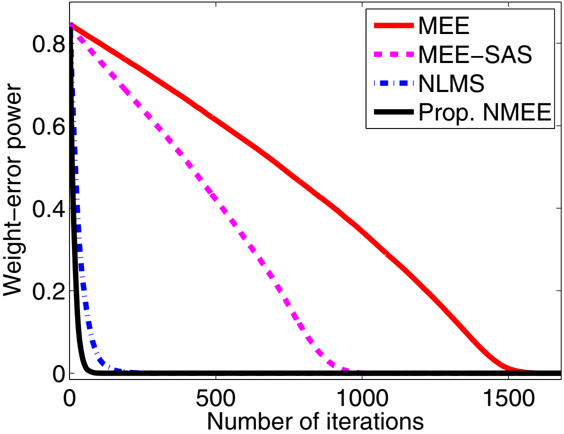
Simulation Results

- Convergence plots for a plant identification problem:



University of Victoria

- Convergence plots with the input power increased by a factor of 10:



Simulation Results *Cont'd*

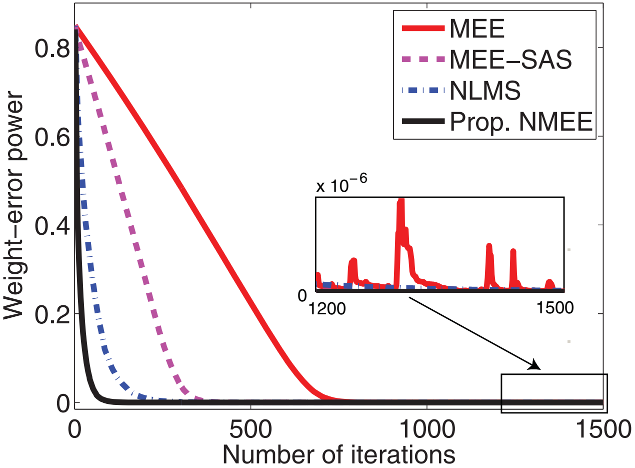
- Residual error and required number of iterations for a white input signal (1p=unit power, 10p=10 times unit power)

	MEE	MEE-SAS	NLMS	Prop. NMEE
Res. Err.	1.115×10^{-25}	1.888×10^{-29}	2.221×10^{-14}	3.741×10^{-32}
Iters., 1p	450	200	200	90
Res. Err.	1.944×10^{-7}	2.675×10^{-31}	2.211×10^{-28}	3.833×10^{-32}
Iters., 10p	1550	950	280	100

University of Victoria

Simulation Results *Cont'd*

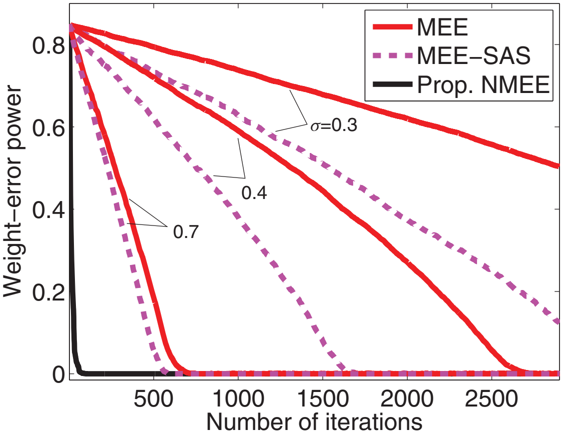
- Convergence plots for a color input signal of unit power:



- Residual error for colored input signal of unit power:

	MEE	MEE-SAS	NLMS	Prop. NMEE
Res. Err.	2.343×10^{-8}	1.389×10^{-11}	1.033×10^{-7}	3.474×10^{-32}

- Convergence plots for different kernel sizes for a color input signal with the power level increased by a factor of 10:

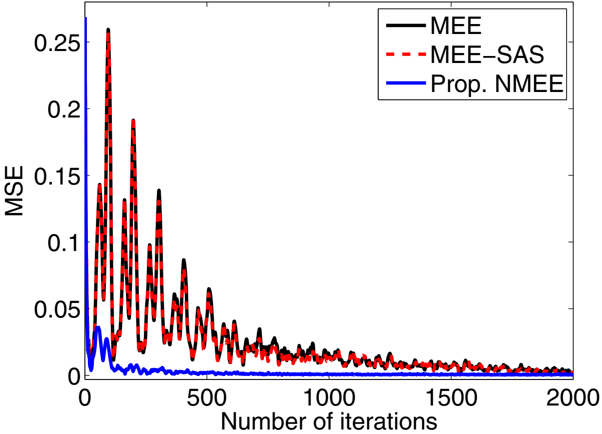


- Residual error for a colored input signal with the power increased by a factor of 10:

kernel size	MEE	MEE-SAS	Prop. NMEE
0.4	1.061×10^{-3}	1.129×10^{-3}	2.960×10^{-32}
0.7	4.937×10^{-4}	1.720×10^{-5}	3.192×10^{-32}

University of Victoria

- Convergence plots for a signal prediction application: We used an FIR filter of length 6 to predict a Mackey-Glass time series with delay parameter $\tau = 30$.



University of Victoria

Conclusions

- The proposed NMEE algorithm is *faster* than the known MEE algorithms and *its performance is not sensitive* to the power of the input signal.

Conclusions

- The proposed NMEE algorithm is *faster* than the known MEE algorithms and *its performance is not sensitive* to the power of the input signal.
- It offers *reduced computational complexity* as well as *better tracking* performance.

Conclusions

- The proposed NMEE algorithm is *faster* than the known MEE algorithms and *its performance is not sensitive* to the power of the input signal.
- It offers *reduced computational complexity* as well as *better tracking* performance.
- It yields *reduced residual error* relative to the known MEE algorithms.

Conclusions

- The proposed NMEE algorithm is *faster* than the known MEE algorithms and *its performance is not sensitive* to the power of the input signal.
- It offers *reduced computational complexity* as well as *better tracking* performance.
- It yields *reduced residual error* relative to the known MEE algorithms.
- It offers improved performance for the case of a color input signal as it reuses the past L input signal vectors and the corresponding error signals in every iteration.

Thank you for your attention.