

# Variable Step Size Maximum Correntropy Criteria Based Adaptive Filtering Algorithm

S. Radhika

Research Scholar,

Faculty of Electrical and Electronics Engineering  
Sathyabama University, Chennai, India  
radhikachandru79@gmail.com

A. Sivabalan

Manager, Research

NEC Mobile Networks Excellence Centre  
Chennai, India  
sivabalan.a@gmail.com

**Abstract**—Maximum correntropy criterion (MCC) based adaptive filters are found to be robust against impulsive interference. This paper proposes a novel MCC based adaptive filter with variable step size in order to obtain improved performance in terms of both convergence rate and steady state error with robustness against impulsive interference. The optimal variable step size is obtained by minimizing the Mean Square Deviation (MSD) error from one iteration to the other. Simulation results in the context of a highly impulsive system identification scenario show that the proposed algorithm has faster convergence and lesser steady state error than the conventional MCC based adaptive filters.

**Keywords**—Maximum correntropy criteria; adaptive filter; variable step size; mean square deviation.

## I. INTRODUCTION

The minimum mean square error (MSE) criterion is a widely used criterion for adaptive filters due to its simplicity and easy implementation [1]. Recently, information theory criterion based adaptive filters have received much attention due to their robustness against impulsive noises, in which case MSE criterion performs poorly. Correntropy is one such information theory criterion which is a non linear measure of similarity between two random variables [2, 3]. The correntropy is defined as:

$$V(X, Y) = E[\kappa(X, Y)] = \int \kappa(x, y) dF_{XY}(x, y) \quad (1)$$

Where  $\kappa$  is a shift-invariant Mercer kernel and  $F_{XY}(x, y)$  denotes the joint distribution function of  $(X, Y)$ . If the kernel is Gaussian, then it is defined as:

$$\kappa(x, y) = G_e(e) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-e^2}{2\sigma^2}\right) \quad (2)$$

where  $e = x - y$  and  $\sigma > 0$  is the kernel width. The cost function under MCC criterion is defined as

$$J_{MCC} = E\left[\exp\left(\frac{-e^2}{2\sigma^2}\right)\right] \quad (3)$$

The adaptive filter using  $J_{MCC}$  as the cost function is defined [4] as:

$$W(n+1) = W(n) + \mu \exp\left(\frac{-e^2(n)}{2\sigma^2}\right) e(n)X(n) \quad (4)$$

The steady state MSE of MCC algorithm is derived in [5] using energy conservation argument. From the analysis it is evident that MCC based algorithm is an excellent candidate for impulsive interference and its performance is dependent on the step size and kernel width. Also from [5] it is found that a large value of step size increases the convergence at the cost of more steady state error and vice versa. This conflicting tradeoff between convergence speed and steady state error is resolved in [6] using a combinational approach. But the major disadvantage with this approach is that the complexity is higher as two adaptive filters are used in parallel. Adaptive kernel width is proposed in [7] where lesser steady state error is obtained but convergence speed is not improved especially when the kernel width is larger.

The variable step size approach has long been used for better performance as it gives a more appreciable solution than the fixed step size approach [8-10]. Hence, in this paper a variable step size for MCC based adaptive filter is proposed so as to have better performance in terms of both faster convergence and lesser steady state error. The criterion for variation is based on the minimization of MSD as it indicates how close the filter is to optimal performance. Thus the step size is increased or decreased based on the square of the instantaneous error obtained. Finally the performance of the proposed algorithm is verified through simulation results for the identification of highly impulsive unknown system.

## II. PROPOSED ALGORITHM

Consider an unknown system with input signal vector given by  $X(n) = [x(n), x(n-1), \dots, x(n-N+1)]^T$  where  $x(n)$  is the input signal. Let  $W_o = [w_{o,1}, w_{o,2}, w_{o,3}, \dots, w_{o,N}]^T$  be the optimal weights of length  $N$ . The desired response  $d(n)$  is given by  $d(n) = W_o^T X(n) + v(n)$  where  $v(n)$  is a noise source with variance  $\sigma_v^2$  which is made of both background noise  $v_b(n)$  and impulsive noise  $v_i(n)$ . The error signal  $e(n)$  is given as:

$$e(n) = d(n) - W^T(n)X(n)$$

where  $W(n) = [w_1(n), w_2(n), w_3(n) \dots w_N(n)]^T$  is the estimate of  $W_0$  at the  $n^{\text{th}}$  iteration. If the weight error vector be defined as  $\tilde{W}(n) = W_0 - W(n)$  then (4) can be written in terms of weight error vector as:

$$\tilde{W}(n+1) = \tilde{W}(n) - \mu \exp\left(\frac{-e^2(n)}{2\sigma^2}\right) e(n)X(n) \quad (5)$$

Let  $f(e(n)) = \exp\left(\frac{-e^2(n)}{2\sigma^2}\right)$ . Thus (5) becomes:

$$\tilde{W}(n+1) = \tilde{W}(n) - \mu f(e(n))e(n)X(n) \quad (6)$$

Squaring on both sides of (6) we get:

$$\begin{aligned} E[\|\tilde{W}(n+1)\|^2] &= \\ &= E[\|\tilde{W}(n)\|^2] - 2\mu \text{Re}[E[\tilde{W}^T(n)f(e(n))e(n)X(n)]] \quad (7) \\ &+ \mu^2 E[X^T(n)e^T(n)f^T(e(n))f(e(n))e(n)X(n)] \end{aligned}$$

We can write (7) further as:

$$E[\|\tilde{W}(n+1)\|^2] = E[\|\tilde{W}(n)\|^2] - \Delta(\mu) \quad (8)$$

where

$$\begin{aligned} \Delta(\mu) &= 2\mu \text{Re}[E[\tilde{W}^T(n)f(e(n))e(n)X(n)]] \\ &- \mu^2 E[X^T(n)e^T(n)f^T(e(n))f(e(n))e(n)X(n)] \end{aligned}$$

Thus in order to reduce the MSD from one iteration to the other,  $\Delta(\mu)$  should be maximized. Thus differentiating  $\Delta(\mu)$  respect to  $\mu$  and equating it to zero gives optimal value of step size as

$$\mu^o(n) = \frac{\text{Re} E\left(\tilde{W}^T(n)f(e(n))e(n)X(n)\right)}{E\left(X^T(n)e^T(n)f^T(e(n))f(e(n))e(n)X(n)\right)} \quad (9)$$

In order to further simplify (9) we use of the following assumptions:

1. The input  $X(n)$  is independent and identically distributed (i.i.d) with zero mean and  $E(X(n)X^T(n)) = R$ .
2. The noise  $v(n)$  is i.i.d with zero mean and variance  $\sigma_v^2$  and is independent of the input  $X(n)$
3. The error  $e(n)$  is independent of  $X(n)$
4. The error non linearity  $f(e(n))$  is independent of  $X(n)$

Assumption 1 and 2 are commonly used to analyze adaptive filters. Assumption 3 is used in [11] and Assumption 4 is used in [12] and they are reasonable when the actual weights converge to the optimal value. Using assumptions 1 to 4 we can write (9) as:

$$\mu(n) = \frac{E(f(e(n)))E(|e_a(n)|^2)}{E(f^2(e(n)))\text{Tr}(R)E(e(n)^2)} \quad (10)$$

If  $e(n) = \tilde{W}^T(n)X(n) + v(n) = e_a(n) + v(n)$  where  $e_a(n)$  is the a priori error, then squaring on both sides of  $e(n)$  and it expectation is taken we get  $E|e(n)|^2 = E|e_a(n)|^2 + \sigma_v^2$ . Thus (10) is written in terms of  $e(n)$  as:

$$\mu(n) = \frac{E(f(e(n)))}{E(f^2(e(n)))\text{Tr}(R)} \left\{ \frac{E(|e(n)|^2) - \sigma_v^2}{E(|e(n)|^2)} \right\} \quad (11)$$

If we assume that the weights  $W(n)$  are independent of the input  $X(n)$ , then:

$$\mu(n) = \frac{1}{E(f(e(n)))\text{Tr}(R)} \left\{ 1 - \frac{\sigma_v^2}{E(|e(n)|^2)} \right\} \quad (12)$$

#### A. Practical step size

From (12) it is found that variable step size may prove difficult because of the presence of expectation. Therefore we replace the expectation by an instantaneous approximate value. Finally using time averaging as adopted in [9], the update equation becomes:

$$\mu(n) = \alpha\mu(n-1) + (1-\alpha)\beta(n) \quad (13)$$

where  $\beta(n) = \frac{1}{f(e(n))\text{Tr}(R)} \left\{ 1 - \frac{\sigma_v^2}{|e(n)|^2} \right\}$  and  $\alpha$  is the smoothing factor that ranges between  $0 < \alpha < 1$ . Thus we can say that whenever the instantaneous error is more or during the

initial stage of the algorithm, the quantity  $\frac{\sigma_v^2}{|e(n)|^2}$  becomes smaller as  $|e(n)|^2 = \sigma_v^2 + |e_a(n)|^2$  and  $\left\{ 1 - \frac{\sigma_v^2}{|e(n)|^2} \right\}$  becomes larger (nearer to but less than unity). During the same time

$f(e(n)) = \exp\left(\frac{-e^2(n)}{2\sigma^2}\right)$  becomes smaller ( $\exp\left(\frac{-e^2(n)}{2\sigma^2}\right) < 1$ ) so that  $1/f(e(n))$  is more. Therefore the overall step size becomes larger so as to have faster convergence. On the other hand during the steady state condition or whenever the instantaneous error becomes smaller, and

$\lim_{n \rightarrow \infty} |e(n)|^2 = \lim_{n \rightarrow \infty} |e_a(n)|^2 + \sigma_v^2 = \varepsilon + \sigma_v^2$  where  $\varepsilon$  is the value

of steady state excess mean square error. Therefore  $\frac{\sigma_v^2}{|e(n)|^2}$  becomes larger such that  $\frac{\sigma_v^2}{|e(n)|^2} < 1$ . Similarly, under the same condition  $f(e(n))$  becomes larger than the value obtained during the initial stage with  $f(e(n)) < 1$ . Thus  $\left\{1 - \frac{\sigma_v^2}{|e(n)|^2}\right\}$  becomes very small and  $\frac{1}{f(e(n))}$  also becomes small. As a result, the step size becomes smaller so as to have a smaller steady state error. During the intermediate value of error, the step size is increased or decreased based on the value of the instantaneous error. Thus both improvement in convergence and steady state error can be obtained simultaneously.

### III. SIMULATION RESULTS

In this section, we present the simulation results to prove the performance improvement of the proposed algorithm against the fixed step size MCC based algorithm. For this purpose an unknown system with 16 coefficients is taken. It is assumed that both the filter and the system have the same number of coefficients. The input is a zero mean white Gaussian noise with unity variance. The noise is made of both background noise and impulsive noise. The background noise is Gaussian with variance  $\sigma_{v_b}^2$ . The impulsive noise is generated  $k(n)A(n)$  where  $k(n)$  is a Bernoulli process with probability of success  $P[k(n)=1]=p_r$ . The performance is evaluated by normalized MSD (NMSD) which is given as  $NMSD = 20\log_{10}\left(\frac{\|\tilde{W}(n)\|_2}{\|W_0\|_2}\right)$ . The kernel width is fixed as  $\sigma = 2$ . The performance curves are obtained by ensemble averages over 50 independent runs. It is assumed that the variance of the noise source is known [ 9].

Figure 1 examines the performance of the proposed algorithm with signal to noise ratio (SNR) taken as 30 dB, signal to interference ratio (SIR) as -10 dB and  $p_r = 0.001, 0.01, 0.1$  respectively. For comparison, fixed step size MCC algorithm with  $\mu = 0.05, 0.01$  and  $\mu = 0.001$  with the same parameters were also plotted. The plot confirms that the proposed algorithm can achieve both faster convergence and smaller steady state error than the fixed step size counterparts. Also it is evident from Figure 1 that the variable step size approach is robust against changing values of  $p_r$ . Similar conclusions can be drawn from Figure 2 and Figure 3 which were simulated for different values of SNR, SIR with  $p_r$  as 0.001.

The convergence behavior of our proposed algorithm can be analyzed if the time evolution of  $E[\mu(n)]$  is made. Fig 4 shows the plot of  $E[\mu(n)]$  Vs  $n$ . As per (13) when the instantaneous error is larger, the value of  $\mu$  is expected to be larger in order to obtain faster convergence. On the other hand, when the instantaneous error is smaller, then  $\mu$  is expected to become smaller in order to have lesser steady state error.

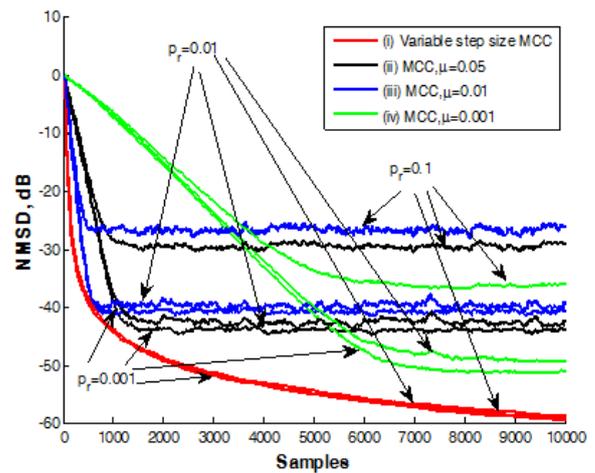


Fig. 1. Performance of proposed variable step size MCC and fixed step size MCC for different values of  $p_r$ . The input is white with SNR=30 dB, SIR=-10 dB.

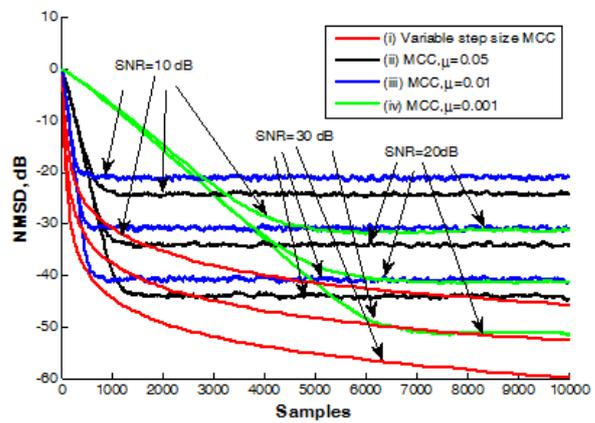


Fig. 2. Performance of proposed variable step size MCC and fixed step size MCC for different values of SNR. The input is white with SIR=-10 dB and  $p_r=0.001$

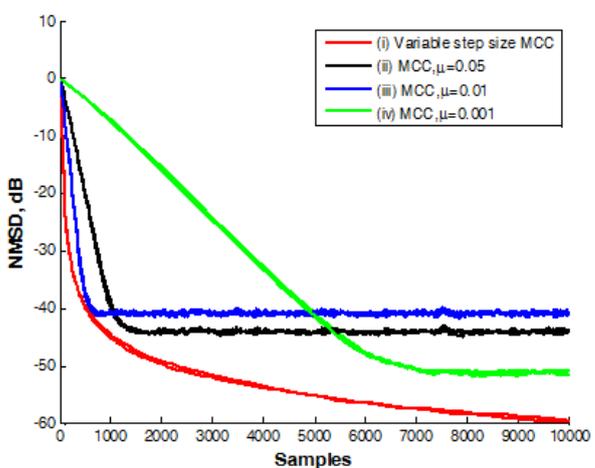


Fig. 3. Performance of proposed variable step size MCC and fixed step size MCC for different values of SIR. The input is white with SNR=30 dB and  $p_r=0.001$

Thus, Figure 4 confirms the predicted results where  $E[\mu(n)]$  is initially larger and starts to decrease and reaches its minimum during steady state which is required for improved performance. In Figure 5, the performance of the proposed algorithm is tested for colored input for the same system as above with SNR=30 dB, SIR=-10 dB and  $p_r=0.001$ . The input is obtained by passing white Gaussian noise through a first order system with pole at 0.8. Thus, it is found from Figure 5 that the proposed algorithm works well even for colored input. The tracking performance of the proposed algorithm is analyzed in Figure 6 where the system is made to change abruptly after 10,000 samples with  $p_r=0.001$ , SNR=30dB, SIR=-10 dB.

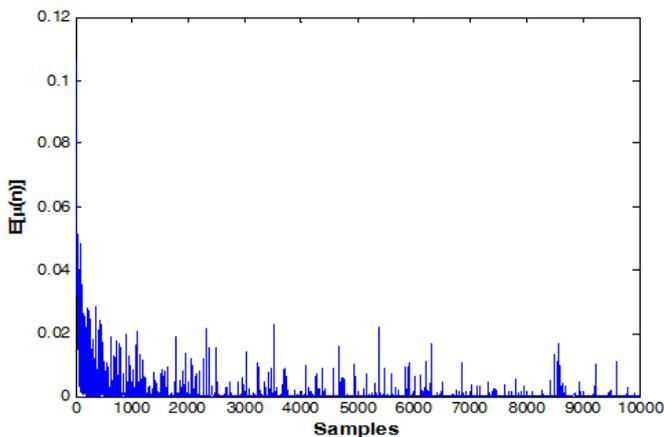


Fig. 4. Time evolution of  $E[\mu(n)]$

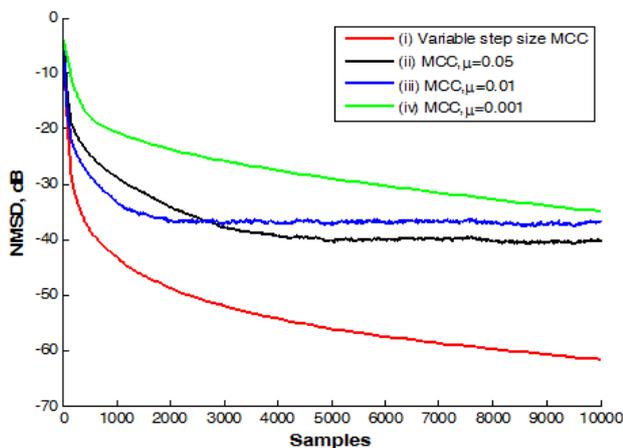


Fig. 5. Performance of proposed variable step size MCC and fixed step size MCC for colored input with pole at 0.8 with SNR=30 dB, SIR=-10 dB and  $p_r=0.001$ .

#### IV. CONCLUSION

Step size is an important parameter which affects the performance of MCC based adaptive filters. A fixed step size gives rise to the tradeoff between convergence speed and steady state error. This paper presented a variable step size approach for MCC based adaptive filters based on largest decrease in MSD. Here the step size is increased or decreased based on the instantaneous error square obtained. Therefore

improved filter performance in terms of convergence speed and steady state error is obtained. The simulations done in the context of a system identification scenario that includes impulsive noises confirms that the proposed algorithm has better performance than the fixed step size MCC based adaptive filters.

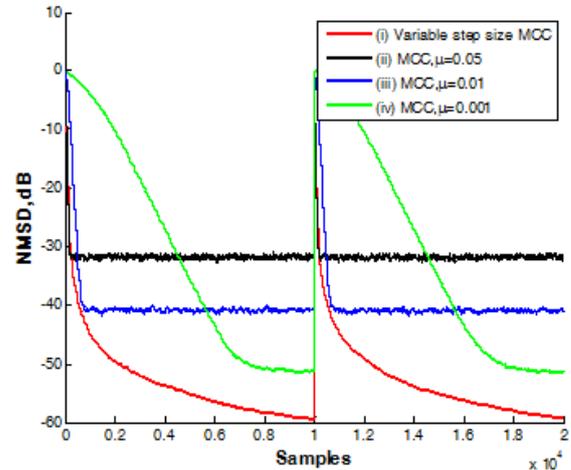


Fig. 6. Tracking performance of proposed variable step size MCC and fixed step size MCC for white input with SNR=30 dB, SIR=-10 dB and  $p_r=0.001$

#### REFERENCES

- [1] S. Haykin, Adaptive filtering theory, Prentice-Hall, New York, NY, USA, 1996.
- [2] J. C. Principe, Information theoretic learning: Renyi's entropy and kernel perspectives, Springer Verlag, Berlin, Germany, 2010
- [3] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: Properties, and applications in non-Gaussian signal processing", IEEE Trans. Signal Process., Vol. 55, No. 11, pp. 5286-5298, 2007
- [4] A. Singh, J. C. Principe, "Using correntropy as a cost function in adaptive filters", International Joint Conference on Neural Networks, Atlanta, USA, pp. 2950-2955, June 14-19, 2009
- [5] B. Chen, L. Xing, J. Liang, N. Zheng, J. C. Principe, "Steady-state mean-square error analysis for adaptive filtering under the maximum correntropy criterion", IEEE Signal Processing Letters, Vol. 21, No. 7, pp. 880-884, 2014
- [6] L. Shi, Y. Lin, "Convex combination of adaptive filters under the maximum correntropy criterion in impulsive interference," IEEE Signal Processing Letters, Vol. 21, No. 11, pp. 1385-1388, 2014
- [7] S. Zhao, B. Chen, J. C. Principe, "An adaptive kernel width update for correntropy", The 2012 International Joint Conference on Neural networks, Brisbane, Australia, pp. 1-5, June 10-15, 2012
- [8] R. H. Kwong, E. W. Johnston, "A variable step size LMS algorithm", IEEE Transactions on Signal Processing, Vol. 40, No. 7, pp. 1633-1642, 1992
- [9] H. C. Shin, A. H. Sayed, W. J. Song, "Variable step-size NLMS and affine projection algorithms", IEEE Signal Processing Letters, Vol. 11, No. 2, pp. 132-135, 2004
- [10] X. D. Luo, Z. H. Jia, Q. Wang, "A new variable step size LMS adaptive filtering algorithm", Acta Electronica Sinica, Vol. 34, No. 6, pp. 1123-1126, 2006
- [11] A. H. Sayed, Adaptive filters, John Wiley & Sons, 2008
- [12] W. Ma, H. Qua, G. Gui, L. Xu, J. Zhaoa, B. Chen, "Maximum correntropy criterion based sparse adaptive filtering algorithms for robust channel estimation under non-Gaussian environments", Journal of the Franklin Institute, Vol. 352, No. 7, pp. 2708-2727, 2015