A Comparative Analysis for WLMS Algorithm in System Identification

[1]Sidhartha Dash and [2]Mihir Narayan Mohanty
Electronics & Communication Engg. Department
ITER, Siksha 'O' Anusandhan University Bhubaneswar, Odisha, India
[1]sidharthadashiter@gmail.com, [2]mihir.n.mohanty@gmail.com

***Abstract*— System Identification is an important area of research in signal processing to design an unknown system. The identification task covers almost all the areas of engineering application such as problem of building models of systems. When insignificant prior information is available and system's properties are known up to a few parameters, identification is most useful. This paper approaches the system identification problem using WLMS Algorithm (Wilcoxon based LMS) in presence of outliers. Also the result is compared with the conventional LMS. In addition to it, the error is analyzed for the deviation factor at the time of analysis. The result shows an excellent performance with minimum training samples.**

***Index Terms*— Wilcoxon norm; Least Mean Square (LMS); WLMS algorithm; system identification.**

I. INTRODUCTION

In machine learning, incomplete data is a big problem. There are many possibilities that can cause the training data to be incomplete, such as mislabeling, biases, non-sufficiency, imbalance, noise, outliers, etc. The Least Mean Square (LMS) algorithm is generally used as a learning tool for optimization technique. The LMS model is to minimize the Euclidean norm by the help of a conventional least square fit analysis. LMS uses a gradient-based method of steepest decent and it uses the estimates of the gradient vector from the available data [1]. It incorporates an iterative procedure that makes successive corrections to the weight vector in the direction of the negative of the gradient vector which eventually leads to the minimum mean square error [2]. But the resulting model obtained by this approach is not effective against outliers.

The convergence of the LMS algorithm is inversely proportional to the eigenvalue spread of the correlation matrix. When the eigenvalues are widespread, convergence may be very slow [3].The eigenvalue spread of the correlation matrix is estimated by computing the ratio of the largest eigenvalue to the smallest eigenvalue of the matrix. The step-size parameter or the convergence factor μ is the basis for the convergence speed of the LMS algorithm. When μ is small the convergence rate is slow, the error is still quite large. A moderately large value of μ leads to faster convergence. However, when the value of μ is too high it leads to the instability of the algorithm and leads to an erroneous result [2].

In most of the real time application, the objective is to model the structure and estimate the parameters effectively in presence of outliers [4]. An outlier is a pattern that was either mislabeled in the training data, or inherently ambiguous and hard to recognize, therefore, it usually brings extra trouble for a learning task, either in debasing the performance or leading the learning process to be more complicated [5]. These outliers are observations that are separated in some fashion from the rest of the data. Hence, outliers are data points that are not typical of the rest of the data. Outliers can occur by chance in any distribution, but they are often indicative either of measurement error or that the population has a heavy-tailed distribution. Depending on their location, outliers may have moderate to severe effects on the regression model. It is always required to discard them or use statistics that are robust to outliers. A regressor or a learning machine is said to be robust if it is insensitive to outliers in the set of input data [6].

To achieve improved and dynamic performance, regression based on the Wilcoxon norm minimization is also suggested in literature [7]. Most of the application areas such as in Communication, machine learning, data mining, adaptive control etc., it has used successfully. In this paper, Wilcoxon norm is used as cost function instead of MSE function for LMS algorithm. The application is considered as the system identification and has analyzed using proposed adaptive algorithm. The error curve as well as the performance is compared between LMS and WLMS in presence of outliers.

II. LMS ALGORITHM

Least mean squares (LMS) algorithms are used in adaptive systems to find the coefficients that relate to producing the least mean squares of the error signal (difference between the desired and the actual signal). It is a stochastic gradient descent method in which the filter is only adapted, based on the error at the current time [2]. In LMS Algorithm the simplified cost function, $\xi_{LMS}(n)$ is given by

$$\xi_{LMS}(n) = 1/2e^2(n) \tag{1}$$

where *e(n)* is defined as the error function which is the difference between the original output of the unknown system and estimated output of the adaptive filter.

The cost function in (1) can be thought of as an instantaneous estimate of the MSE (mean square error) cost function which is extremely useful for practical applications and the weight update LMS equation is given by

Manuscript received January 25, 2013; revised March 29, 2013.

doi: 10.12720/lnit.1.3.95-98

$$w(n+1) = w(n) + \mu e(n)x(n) \quad (2)$$

where *w(n)* and *x(n)* represents the weight of the FIR adaptive filter and input samples respectively.

Equation (2) requires only multipliers and adders to implement. In fact, the number and type of operations needed for the LMS algorithm is nearly the same as that of the FIR filter structure with fixed coefficient values, which is one of the reasons for the algorithm's popularity. The behavior of the LMS algorithm had been widely studied before, and numerous results concerning its adaptation characteristics under different situations have already been developed.

However, the average behavior of the LMS algorithm is quite similar to that of the steepest descent algorithm in which it depends explicitly on the statistics of the input and desired response signals. In effect, the iterative nature of the LMS coefficient updates is a form of time-averaging that smoothes the errors in the instantaneous gradient calculations to obtain a more reasonable estimate of the true gradient. The problem is that gradient descent is a local optimization technique, which is limited because it is unable to converge to the global optimum on a multimodal error surface if the algorithm is not initialized in the basin of attraction of the global optimum. As well as the conventional LMS algorithm tends to converge very slowly, for which we need to apply more number of input training samples [8]. Also the LMS algorithm is very sensitive to outliers i.e. the performance of LMS algorithm is affected severely in presence of outliers [9].

In direct modeling process using LMS Algorithm we are updating the weights of adaptive filter so that the result is optimized and simultaneously the error is minimized. Here we are taking the input samples within a predefined range. But in any practical system outliers are always expected. In presence of outliers the direct modeling using LMS is not a robust one i.e. the error increases rapidly with the increase in percentage of outliers. Wilcoxon Algorithm can overcome this problem.

III. WILCOXON LMS ALGORITHM

Wilcoxon Algorithm is one of the effective methods of robust identification [7]. The cost function taken in the proposed model is a robust norm called Wilcoxon norm. The weights of the models are updated using conventional LMS, which progressively reduces the norm [9].

The Wilcoxon Norm of a vector is analyzed in terms of a score function which is defined as

$$\Phi(u) : [0,1] \rightarrow R \quad (3)$$

which is non-decreasing such that

$$\int_0^1 \Phi^2(u)du < \infty \quad (4)$$

Usually the score function is standardized such that

$$\int_0^1 \Phi^2(u)du = 1 \quad (5)$$

and $$\int_0^1 \Phi(u)du = 0 \quad (6)$$

Let the error vector of pth particle at kth generation due to application of N input samples to the model be represented as $[e_{1,p}(k), e_{2,p}(k), \ldots\ldots e_{N,p}(k)]^T$. The errors are then arranged in an increasing manner from which the rank $R[e_{n,p}(k)]$ each l^{th} error term is obtained. The score associated with each rank of the error term is evaluated as

$$\begin{aligned} a(i) &= \Phi(u) = \sqrt{12}(u) \\ &= \sqrt{12}\left(\frac{i}{N+1} - 0.5\right) \end{aligned} \quad (7)$$

where N is a fixed positive number. and ($1 \le i \le N$) denotes the rank associated with each error term. At k^{th} generation of each p^{th} particle, the Wilcoxon norm is calculated as

$$C_p(k) = \sum_{i=1}^{N} a(i)e_{i,p}(k) \quad (8)$$

The learning strategy using LMS Algorithm continues until the cost function $C_p(k)$ in (8) decreases to the possible minimum value [10].

IV. SYSTEM IDENTIFICATION

The process of going from observed data to a mathematical model is fundamental in science and engineering. In the control area, this process has been termed system identification. The identification task is to determine a suitable estimate of finite dimensional parameters, which completely characterize the plant. The selection of the estimate is based on comparison between the actual output sample and a predicted value on the basis of input data up to that instant [11].

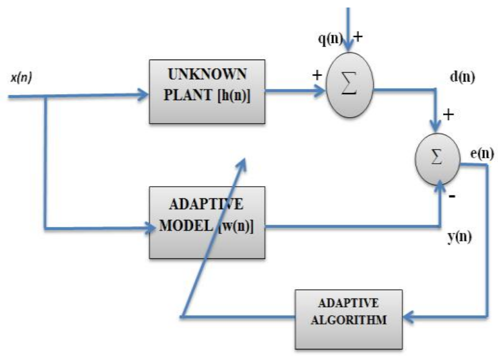

Figure 1. Block diagram for system identification

The block diagram for system identification is shown in Fig. 1. The model is placed parallel to the linear plant and same input is given to the plant as well as the model. The impulse response of the linear segment of the plant is represented by *h(n)*. A white Gaussian noise *q(n)* is added with the linear output accounts for measurement noise. The desired output *d(n)* is compared with the estimated output *y(n)* of the identifier to generate the error *e(n)*, which is used by some adaptive algorithm for updating the weights of the model.

Here the model chosen for the adaptive filter is a typical linear filter. So the practical goal of the adaptive filter is to determine the best linear model that describes the input-output relationship of the unknown system. Such a procedure makes the most sense when the unknown system is also a linear model of the same structure as the adaptive filter.

The training of the filter weights is continued until the error becomes minimum and does not decrease further. At this stage the correlation between input signal and error signal is minimum. Here the training is stopped and the weights are stored for testing. For testing purpose new samples are passed through both the plant and the model and their responses are compared. The optimal model or solution is attained when this function of the error is minimized [12]. The model of system identification can be expressed in terms of mathematical equations.
The error signal is expressed mathematically as

$$e(n) = d(n) - y(n) \qquad (9)$$

The estimated output of the adaptive filter is found to be

$$y(n) = w^T(n).x(n) \qquad (10)$$

where $w(n)$ are the parameters correspond to the weights the adaptive filter at time n.

When $e(n) = 0$, then $d(n) = y(n)$ (11)

The above said condition occurs at a particular set of weights that is known as optimized value i.e.

$$w(n) = w_{opt}(n) \; for \;\; n \to \infty \qquad (12)$$

where $w_{opt}(n)$ is an optimum set of filter coefficients for the unknown system at time n. When the adaptive filter weights are optimized, at that time the model provides the best performance.

V. Result & Discussion

It has been simulated by taking learning rate parameter μ as 0.2 and SNR as 30db. The weight update equation is used for system identification. The analysis is carried without outliers as well as in presence of outliers and the simulation results are compared. The original weight of the system is taken as a set of standard weights [0.2600 0.9300 0.2600].

A. Comaparison of Performance

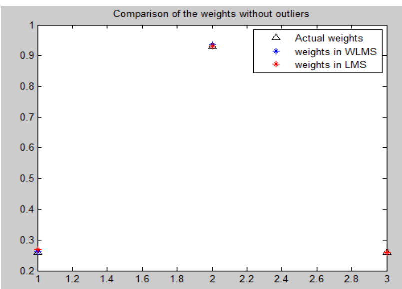

Figure 2. Comparison of the actual weight with the estimated weight using LMS and WLMS without outliers

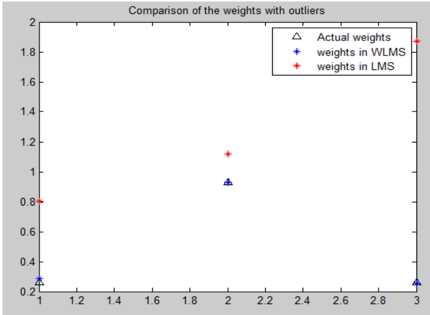

Figure 3. Comparison of the actual weight with the estimated weight using LMS and WLMS in presence of outliers

From the simulation results (Fig. 2 & Fig. 3), it is seen that both LMS and WLMS converges effectively without outliers but in presence of outliers WLMS converges whereas LMS produces more deviation.

B. Effect on Performance Due to the Variation of the no. of Input training Samples in Presence of Outliers:

Number of training samples - 200

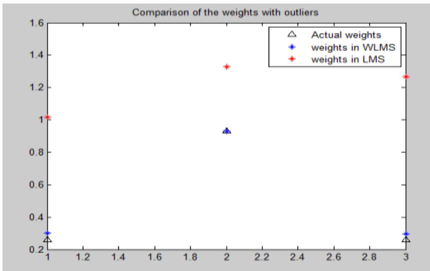

Figure 4. Comparison of the actual weight with the estimated weight using LMS and WLMS in presence of outliers

Number of training samples - 300

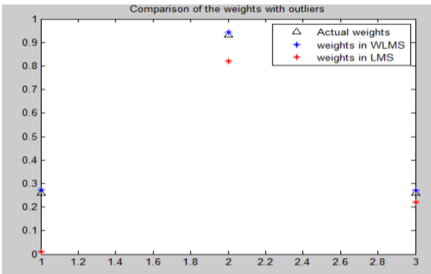

Figure 5. Comparison of the actual weight with the estimated weight using LMS and WLMS in presence of outliers

Number of training samples - 400

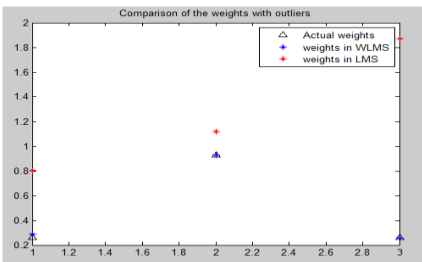

Figure 6. Comparison of the actual weight with the estimated weight using LMS and WLMS in presence of outliers

For 200 number of training input samples (Fig. 4, Fig. 5, Fig. 6), WLMS Algorithm provides optimized result with very less deviation whereas the result is diverged in case of LMS Algorithm with high deviation. As the number of input samples increases, the performance of WLMS is unaffected but LMS output remains diverged with variable deviation. It is concluded that WLMS convergence faster compared to LMS. The deviation values have been mentioned in the Table I.

C. Error Analysis & Comparison

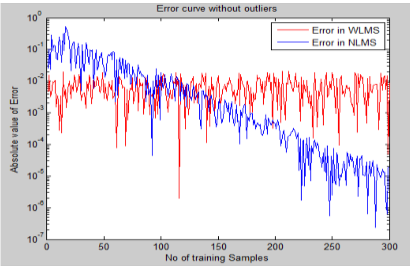

Figure 7. Comparison of the error curve using LMS and WLMS without outliers.

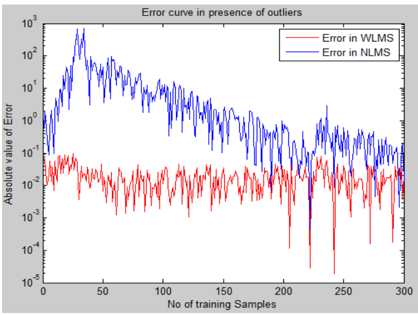

Figure 8. Comparison of the error curve using LMS and WLMS in presence of outliers

From Fig. 7 and Fig. 8, without outliers LMS gives better error response where as in presence of outliers WLMS produces better result as compared to LMS. Also, it converges quickly whereas LMS produces more deviation.

TABLE I. [DEV.=(W1-H1)^2+(W2-H2)^2+(W3-H3)^2]

NO. OF TRAINING INPUT SAMPLES	DEVIATION WITHOUT OUTLIERS		DEVIATION IN PRESENCE OF OUTLIERS	
	LMS	WLMS	LMS	WLMS
200	$3.1*10^{-5}$	$7.9*10^{-6}$	1.74	$3.2*10^{-3}$
300	$8.1*10^{-5}$	$8.9*10^{-6}$	$7.6*10^{-2}$	$4.3*10^{-4}$
400	$8.1*10^{-5}$	$8.9*10^{-6}$	2.93	$8.1*10^{-4}$

VI. CONCLUSION

The paper has proposed for WLMS algorithm. The application is considered and tested for system identification. Also, the error rate has been analyzed by considering the deviation parameter. Simulation study clearly demonstrates that the WLMS based method achieves better results than its LMS counterparts, both in terms of speed and minimum MSE. The further work can be applied with the variance factor as well as the step size parameter.

REFERENCES

[1] B. Chen, J. Hu, H. Li, and Z. Sun, "A joint stochastic gradient algorithm and its application to system identification with RBF networks," in *Proc. 6th World Congress On Intelligent Control and Automation*, Dalian, China, June 21 - 23, 2006, vol. 1, pp. 1754-1758.

[2] S. Haykin, *Adaptive Filter Theory*, 4th ed., PHI Publication, 2001, Chapter 20, pp. 855-874.

[3] S. Haykin, *Neural Networks, A Comprehensive Foundation*, 2nd ed., Pearson Education, 1998, pp. 112-125.

[4] J. W. Mckean, "Robust analysis of linear models," in *Statistical Science*, vol. 19, no. 4, 2004, pp. 562-570.

[5] D. Luo, X. Wang, X. Wu,and H. Chi, "Outliers learning and its applications," in *Proc. International Conference on Neural Networks & Brain*, Oct 13-15, 2005, vol. 2, pp. 661-666.

[6] B. Majhi, G. Panda, and B. Mulgrew, "Robust identification using new Wilcoxon least mean square algorithm," *Electronics Letters*, 12th March 2009, vol. 45, no. 6, pp. 334-335.

[7] J. G. Hseih, Y. L. Lin, and J. H. Jeng, "Preliminary study on wilcoxon learning machines," *IEEE Transactions on Neural Networks*, vol. 19, no. 2, pp. 201-211, Feb 2008.

[8] S. Kumar, *Neural Networks, A Classroom Approach*, TMH Publication, 2004.

[9] U. K. Sahoo, G. Panda, and B. Mulgrew, "Sign-regressor wilcoxon and sign-sign wilcoxon," in *Proc. International Conference on Advances in Recent Technologies in Communication and Computing*, Kerala, Oct 16-17, 2010, pp. 35-39.

[10] B. Majhi, G. Panda, and B. Mulgrew, "Robust identification and prediction using wilcoxon norm and particle swarm optimization," in *Proc. 17th European Signal Processing Conference*, Glasgow, Scotland, August 24-28, 2009, pp.1695-1699.

[11] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Transactions on Neural Networks*, vol. 1, no. 1, pp. 04-27, March, 1990.

[12] X. Deng, "System identification based on particle swarm optimization algorithm," in *Proc. International Conference on Computational Intelligence and Security*, Dec 11-14, 2009, vol. 1, pp. 259-263.

Mihir N. Mohanty is presently an Associate Professor with the Department of Electronics and Communication Engineering, Institute of Technical Education and Research, Bhubaneswar, Odisha, India. He has over 30 papers in International/National Journals and conferences. His area of research interests includes Digital signal/image processing, Biomedical signal processing and Bioinformatics.
E-mail id: mihir.n.mohanty@gmail.com

Sidhartha Dash is presently working as an Assistant Professor in the Department of Electronics and Communication Engineering, Institute of Technical Education and Research, Bhubaneswar, Odisha, India. He received his B.E degree in Electronics and Tele-Communication Engineering from B.I.E.T, Odisha, India in the year 2001. He obtained his M-Tech degree from KIIT, BBSR in 2010. He has 5 papers in International / National Journals and conferences. His research interests are in the areas of Digital signal Processing, soft computing application and Image Processing.
E-mail id: sidharthadashiter@gmail.com.