Malaysian Journal of Mathematical Sciences 10(S) March : 67-77 (2016)
Special Issue: The 10th IMT-GT International Conference on Mathematics, Statistics and its Applications 2014 (ICMSA 2014)

MALAYSIAN JOURNAL OF MATHEMATICAL SCIENCES

Journal homepage: http://einspem.upm.edu.my/journal

Kernel Based Learning for Nonlinear System Identification

Shazia Javed [1*] and Noor Atinah Ahmad [2]

[1]Department of Mathematics,
Lahore College for Women University, 54000 Lahore, Pakistan

[2] School of Mathematical Sciences,
Universiti Sains Malaysia, 10800 Penang, Malaysia

E-mail: shaziafateh@hotmail.com

*Corresponding author

ABSTRACT

In this paper, an efficient Kernel based algorithm is developed with application in nonlinear system identification. Kernel adaptive filters are famous for their universal approximation property with Gaussian kernel, and online learning capabilities. The proposed adaptive step-size KLMS (ASS-KLMS) algorithm can exhibit universal approximation capability, irrespective of the choice of reproducing kernel. Performance evaluation of proposed nonlinear adaptive filter is carried out for an unknown plant with an additive white Gaussian noise in the expected output. In comparative study, learning parameters are computed using KLMS, NKLMS and ASS-KLMS algorithms for polynomial and Gaussian kernels. Simulation results are presented for the performance analysis of the algorithm in terms of signal to noise ratio (SNR) and MSE [dB], showing preference of ASS-KLMS algorithm over the rest.

Keywords: On-line learning, Kernel filter, adaptive step-size, system modeling.

1. Introduction

Adaptive modeling of unknown systems is a rich and challenging task in various fields of signal processing and control. Adaptive filters can be identified as on-line learning systems that rely on a recursive algorithm to become able to perform adequately in an environment where knowledge of the relevant statistics is not available. They have ability to detect time varying potentials and to track the dynamic variations of various kinds of signals. Linear adaptive filters have vast application in signal processing and control, including noise cancellation (Javed and Ahmad, 2014a, 2014b), system identification (Javed and Ahmad, 2014c) and channel equalization etc. (Diniz, 2008). However, for situations where nonlinear phenomenon appear, the performance of linear filters becomes poor, and the development of nonlinear filters is thus necessary (Liu, Principe, & Haykin, 2011; Mathews, 1991). Adaptive kernel filter is one of the most popular types of nonlinear filters that transforms the nonlinear data from input space into a higher dimensional feature space to treat with a reproducing kernel function. These filters preserve the conceptual simplicity of standard linear adaptive filters, while inheriting the rich articulateness from the kernel methods. Liu (Liu et al., 2011) and his companions proposed an iterative method based kernel LMS algorithm (KLMS) that is a nonlinear modification of conventional LMS algorithm into a higher dimensional reproducing kernel Hilbert space (RKHS). However, convergence of KLMS is highly sensitive to the value of step-size, and amplitude of data signal. Normalized KLMS (NKLMS) algorithm is designed predominately to reduce such problems (Modaghegh, Khosravi, Manesh and Yazdi, 2009), however, it is observed that if kernel is shift invariant, like Gaussian kernel, then KLMS and NKLMS algorithms are the same (Chen, Zhao, Zhu, and Príncipe, 2012; Liu et al., 2011). It is therefore needed to develop a new KLMS based algorithm with better convergence properties to identify characteristics of unknown system models.

Most of the work in kernel filtering is done by taking a constant step-size for KLMS algorithm that is chosen so as to have a trade-off between the speed of adaptation and mean square error (MSE). In common practice, step-size changes according to the change in MSE, permitting the adaptive filter to modify parameters in an attempt to identify system with minimum MSE. A variable step-size has been helpful in having minimum misadjustment of LMS algorithm (Kwong and Johnston, 1992), and having steady state convergence. Since KLMS algorithm is a non-linear modification of LMS algorithm in a higher dimensional RKHS, variable step-size may help in improving convergence behaviour of KLMS algorithms as well. To date authors have not seen any work regarding variable step-size of KLMS algorithms.

This paper presents a new and efficient adaptive step-size based KLMS (ASS-KLMS) algorithm that is suitable for applications requiring large signal to noise ratios and with less computational complexity. A detailed comparative study of the system identification problem for nonlinear plants, employing the new ASS-KLMS algorithm, is presented with different choices of reproducing kernels. The adaptive step-size of the

ASS-KLMS algorithm is realized by optimized power of input signal vector and current iteration number. In initial iterations, the MSE is normally high, which decreases gradually to zero with successive iterations to yield a steady state convergence. To make choice of step-size coherent with MSE behavior, adaptive step-size of proposed algorithm is set to decrease with an increase in the iteration number, hence ensuring a reduced MSE in successive iterations. This property has made proposed ASS-KLMS algorithm capable to be applied in any signal processing application, including signal prediction, noise cancelation and system identification. For experimental results, an unknown system identification application is simulated here, followed by a comparative study with KLMS and NKLMS algorithms. Performance of all kernel based adaptive algorithms is analyzed by using polynomial kernel and Gaussian kernel. Observations are made in terms for their learning behavior for minimum mean squares error (MSE), and processing time and improvement is signal to noise ratio (SNR). Better performance of ASS-KLMS algorithm is observed is all the experiments.

2. System Identification Using Kernel Adaptive Filter

The block diagram of the used nonlinear system identification model is shown in Figure 1. Input $u(n)$ is passed through an unknown plant to get an output signal $s(n)$. The desired output $\hat{s}(n)$ of the system is obtained by adding a white Gaussian noise $\sigma(n)$ in $s(n)$, such that the added noise is uncorrelated with $s(n)$. The error signal $e(n)$ is the difference between $\hat{s}(n)$ and the estimated output $y(n)$ determined by a nonlinear adaptive filter.

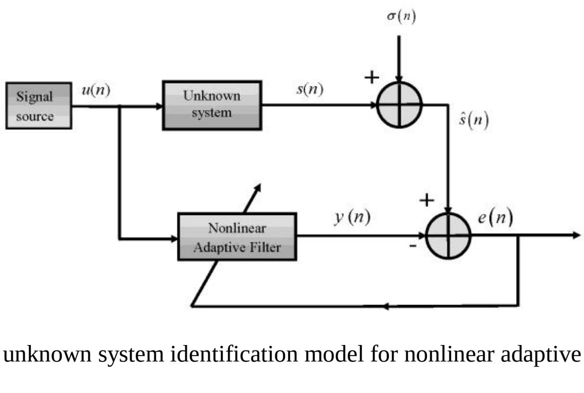

Figure 1: An unknown system identification model for nonlinear adaptive filtering

The purpose of study it to derive a new efficient ASS-KLMS algorithm, and check its ability for nonlinear system identification. ASS-KLMS is adaptive step size KLMS algorithm that is a modified version of normalized KLMS algorithm.

2.1 Kernel LMS (KLMS) Algorithms

Consider the input-output data sequence $\{\mathbf{a}_n, s(n)\}$, with an N dimensional input vector

$$\mathbf{a}_n = \left[u(n), u(n-1), \cdots, u(n-N+1)\right] \in \mathbf{R}^N,$$

and one dimensional output signal $s(n)$ at time n. If $\varphi(n) = \varphi(\mathbf{a}_n)$ is the transformed input into a high dimensional feature space $\mathbf{F}$, then a kernel is a symmetric, positive definite, continuous function $\kappa : \mathbf{R}^N \times \mathbf{R}^N \rightarrow \mathbf{R}$, defined by $\kappa(\mathbf{a}_n, \mathbf{a}_m) = \langle \varphi(n), \varphi(m) \rangle$. In kernel learning the feature space $\mathbf{F}$ is a reproducing kernel Hilbert space (RKHS) induced by the kernel κ, if $\varphi : \mathbf{R}^N \rightarrow \mathbf{F}$ is identified as $\varphi(\mathbf{a}_n) = \kappa(\mathbf{a}_n, .)$. Kernel is an essential component of any kernel method in the sense that it defines the similarity between the data points (Liu et al., 2011). Among various reproducing kernels, the Gaussian kernel is usually a default choice, because of its universal approximating capability and numerical stability.
The kernel LMS algorithm can be viewed as an LMS algorithm in a high dimensional feature space $\mathbf{F}$. With transformed input-output data sequence $\{\varphi(n), s(n)\}$, the stochastic gradient algorithm yields:

(a)
$$\left.\begin{aligned} \mathbf{w}_{init} &= \mathbf{0} \\ e(n) &= s(n) - \mathbf{w}_{n-1}^T \varphi(n) \\ \mathbf{w}_n &= \mathbf{w}_{n-1} + \mu e(n) \varphi(n) \end{aligned}\right\}.$$

If $f_n = \mathbf{w}_n^T \varphi(.)$ denotes an estimate of input-output nonlinear learning rule at time n, then sequential learning rule for KLMS algorithm can be written as

(b)
$$\left.\begin{aligned} f_{init} &= \mathbf{0} \\ e(n) &= s(n) - f_{n-1}(\mathbf{a}_n) \\ f_n &= f_{n-1} + \mu e(n) \kappa(\mathbf{a}_n, .) \end{aligned}\right\}.$$

The step-size parameter μ controls the convergence speed of the algorithm and for a stable algorithm it must satisfy the following condition:

$$\mu < \frac{1}{\varsigma_{\max}},$$

where $\varsigma_{\max}$ is the largest eigenvalue of the transformed data autocorrelation matrix.

2.2 Normalized KLMS (NKLMS) Algorithms

Although KLMS is a proficient algorithm, it is found sensitive to step-size and signal strength. NKLMS algorithm has controlled this problem to a great extent (Modaghegh et al., 2009). For $\varphi(n) = \kappa(\mathbf{a}_n, .)$, update equation for NKLMS algorithm is given as,

$$f_n = f_{n-1} + \frac{\mu}{\psi + \left\|\varphi(n)\right\|^2} e(n)\varphi(n) \ ,$$

where ψ is a small positive constant, used to avoid division by zero.

2.3 The Adaptive Step-size Kernel LMS (ASS-KLMS) Algorithm

ASS-KLMS algorithm is a modification of NKLMS algorithm. It is observed that the step-size of NKLMS algorithm, given by

$$\mu_N(n) = \frac{\mu}{\psi + \left\|\varphi(n)\right\|^2},$$

changes adaptively with incoming signals, and algorithm have different step-size in each update. However, if kernel is shift invariant, like Gaussian kernel, then KLMS and NKLMS algorithm are same (Chen et al., 2012).

Since there is always a need to normalize data to improve an adaptive algorithm, therefore, a new time dependent step-size based ASS-KLMS algorithms is presented in this section.

The adaptive step-size of ASS-KLMS is motivated from data dependent step-size of NKLMS algorithm. Spectral power of an input signal vector is measured by $\left\|\varphi(n)\right\|^2$, and excellent convergence behaviour of a learning algorithm is guaranteed for spectral power close to 1. To maintain this power close to 1, adaptive step-size is designed as

$$\text{(c)} \quad \mu_n = \frac{\mu_{n-1}}{\varepsilon n + c},$$

where ε is a small constant, c is a real number close to 1 and n time instant. For a given value of initial step-size μ_{init}, the ASS-KLMS algorithm is summarized as

$$\text{(d)} \qquad \left.\begin{aligned} & f_{init} = \mathbf{0}; \quad 0 < \mu_{init} < 1; c = 1; \\ \mu_n = \frac{\mu_{n-1}}{\varepsilon n + c}, & \, e(n) = s(n) - f_{n-1}(\mathbf{a}_n) \\ & f_n = f_{n-1} + \mu_n e(n) \varphi(n) \end{aligned}\right\},$$

2.4 Analysis of ASS-KLMS Algorithm

Convergence: For $0 < \mu_{init} < 1$, and $c = 1$ equation (c) yields $0 < \mu_n < 1$ for all n. Therefore, convergence is guaranteed for all the algorithm.

Universal Approximation Property: From (Liu et al., 2011), it appears that KLMS has excellent performance with Gaussian Kernel, while employing its universal approximation property. However, in practice, its behaviour is found to be sensitive to the choice of kernel. With Gaussian Kernel, NKLMS and KLMS algorithms have similar convergence behaviour. It is because of the fact that Gaussian kernel normalizes the power of signal to unity, giving $\|\varphi(n)\|^2 = 1$, and KLMS algorithm becomes NKLMS. But, no such thing happens with other kernels, like polynomial kernel etc.

On the other hand, if $c = 1$, the adaptive step-size of ASS-KLMS algorithm depends upon n only and $\|\varphi(n)\|^2$ is normalized to 1 for all types of kernel. This fact establishes universal approximation property in ASS-KLMS algorithm irrespective of the choice of reproducing kernel, provided that μ_{init} is suitably chosen.

3. Simulation Results and Discussions

Assume that the output signal $s(n)$ for the nonlinear system identification problem is given by

$$s(n)=x_0\sin(u(n))+x_1u(n-1)+x_2u^2(n-2)+...+x_{N-1}u^2(n-N+1)$$

where $\mathbf{x}=[x_0,x_1,...,x_{N-1}]^T\in\mathbf{R}^N$ is the parameter vector. The desired output signal $\hat{s}(n)$ is the sum of $s(n)$ and a zero mean, white Gaussian noise $\sigma(n)$ of $SNR\approx 30\,dB$. In our simulations, we choose

$$\mathbf{x}=\left[1.0,0.5,-0.1,(-0.1)^2,...,(-0.1)^{N-2}\right]^T$$

and obtain the reference signal $u(n)$ by passing the random input signal through an unknown filter, with frequency response given by (Farhang-Boroujeny, 1998)

$$H(z)=\frac{\sqrt{1-0.01^2}}{1-0.01z^{-1}}.$$

To avoid the increased computational complexity of nonlinear adaptive filters, filter length is set at $N=3$. Experiments are performed with polynomial kernel first and then with Gaussian kernel. For all simulations, averaged experimental results are presented for 200 independent runs/experiments of 2500 iterations for each run. Furthermore, adaptive step-size of ASS-KLMS algorithm is set with $\varepsilon=8\times10^{-8}$ and $c=1$, while for NKLMS algorithm $\psi=8\times10^{-8}$.

3.1 Polynomial Kernel

The first simulation results correspond to the polynomial kernel

$$\varphi(n,m)=\left(1+\mathbf{a}_n^T\mathbf{a}_m\right)^5.$$

The initial value of step-size parameter for ASS-KLMS algorithm is $\mu_{init}=0.002$, and same value is taken as step-size for KLMS and NKLMS algorithms. Figure 3 shows that how adaptive step-size of ASS-KLMS algorithm decreases with successive iterations, ensuring decrease in corresponding value of MSE. This is quite visible in the learning curves of MSE (in dB) of Figure 2. Furthermore, in Table 1, ASS-KLMS shows highest improvement in SNR in shortest computational time (see Table 2) for polynomial kernel. Hence, with polynomial kernel, performance of

ASS-KLMS algorithm is far better than other two algorithms and this betterment increases with an increase in the iteration number, just because step-size decreases adaptively with a decrease in MSE.

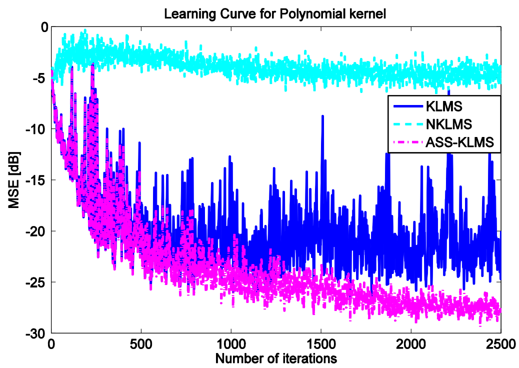

Figure 2: Learning curves of MSE [dB] for KLMS, NKLMS, & ASS-KLMS algorithms with Polynomial kernel.

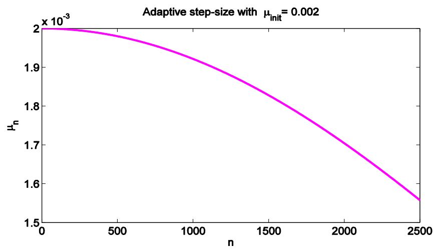

Figure 3: Adaptive step-size for Polynomial kernel.

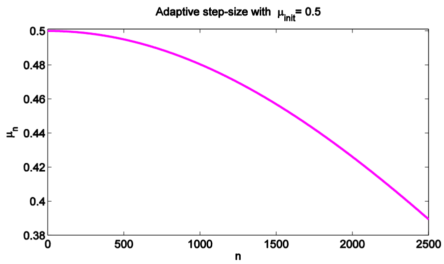

Figure 4: Adaptive step-size for Gaussian kernel.

TABLE 1: Improvement in SNR using Kernel based learning algorithms.

Learning algorithm	Polynomial Kernel	Gaussian Kernel
KLMS	-0.11265	16.60819
NKLMS	-27.47110	16.60819
ASS- KLMS	3.40980	16.71565

TABLE 2: Average computational time (in sec.) for one independent run of 2500 iterations.

Learning algorithm	Polynomial Kernel	Gaussian Kernel
KLMS	5.04023	3.78459
NKLMS	5.08002	3.74103
ASS- KLMS	5.02637	3.72772

3.2 Gaussian Kernel

The second simulation results correspond to the Gaussian kernel

$$\varphi(n,m) = \exp\left(-\alpha \|\mathbf{a}_n - \mathbf{a}_m\|^2\right),$$

for $\alpha = 0.5$. For this simulation μ_{init} of ASS-KLMS algorithm, and step-size of KLMS and NKLMS algorithms are taken as 0.5. Adaptive decrease in step-size of ASS-KLMS algorithm is shown in Figure 4, with $\mu_{init} = 0.5$. Learning curves of MSE [dB] in Figure 5 show similar convergence behavior for the three kernel filters, however, Table 1 shows slight better improvement in SNR with ASS-KLMS algorithm and Table 2 shows least computational time. To make these results more evident, learning curves of MSE [dB] for the last 100 iterations are shown in Figure 6. These curves show a better convergence of ASS-KLMS algorithm, and same behavior of NKLMS and KLMS algorithm verifies the statement about auto-normalized behavior of KLMS algorithm with Gaussian kernel(Chen et al., 2012).

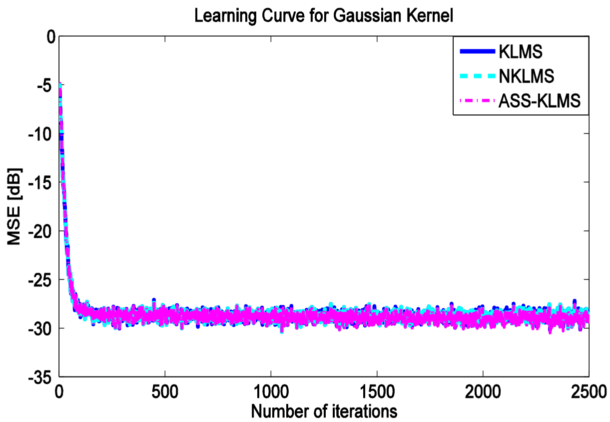

Figure 5: Learning curves of MSE [dB] for KLMS, NKLMS, & ASS-KLMS algorithms with Gaussian kernel

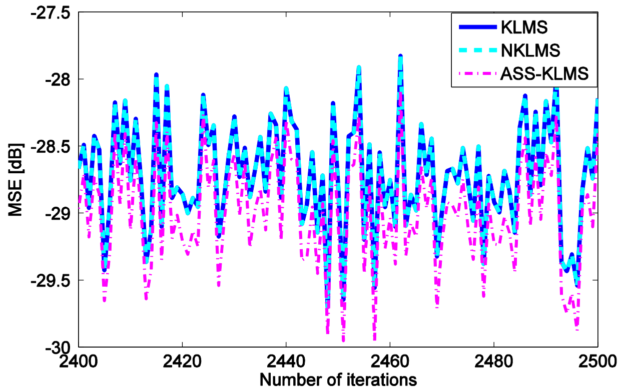

Figure 6: Learning curves of MSE [dB] for last 100 iterations with Gaussian kernel.

4. Conclusion

The new ASS-KLMS algorithm is an efficient on-line learning algorithm that is shown to exhibit universal approximation property, irrespective of the choice of reproducing kernel. It is an improved kernel based adaptive filtering algorithm that is able to perform well in nonlinear system identification problem, and exhibit better convergence as compared with conventional KLMS and NKLMS algorithms.

A comparative study is carried out in terms of steady state mean square error (MSE) performances, and improvement in SNR. This study shows that ASS-KLMS adaptive filter is capable of producing better and accurate system identification results for different reproducing kernels, compared to KLMS and NKLMS adaptive filters. In future, ASS-KLMS algorithm can be applied to various nonlinear adaptive filtering applications, where conventional KLMS algorithm fails to perform well because of the reproducing kernel used.

Acknowledgments

The authors would like to acknowledge the financial support of the School of Mathematical Sciences, Universiti Sains Malaysia (USM).

References

Chen, B., Zhao, S., Zhu, P., and Príncipe, J. C. (2012). Mean square convergence analysis for kernel least mean square algorithm. *Signal Processing,* **92**(11): 2624-2632.

Diniz, P. S. R. (2008). *Adaptive filtering: algorithms and practical implementation* (Vol. 694), Springer Verlag.

Farhang-Boroujeny, B. (1998). *Adaptive filters: theory and applications*, John Wiley & Sons, Inc.

Javed, S. and Ahmad, N. A. (2014a). *An Adaptive Noise Cancelation Model for Removal of Noise from Modeled ECG Signals.* Paper presented at the 2014 IEEE Region 10 Symposium.

Javed, S. and Ahmad, N. A. (2014b). Noise Cancelation From ECG Signals Using Householder-RLS Adaptive Filter. *The 8th International Conference on Robotic, Vision, Signal Processing & Power Applications* (pp. 73-79), Springer Singapore.

Javed, S., and Ahmad, N. A. (2014c). *Performance study of LMS based adaptive algorithms for unknown system identification.* Paper presented at the Proceedings of the 21st National Symposium on Mathematical Sciences (SKSM21): Germination of Mathematical Sciences Education and Research towards Global Sustainability.

Kwong, R. H. and Johnston, E. W. (1992). A variable step size LMS algorithm. *Signal Processing, IEEE Transactions on.* **40**(7): 1633-1642.

Liu, W., Principe, J. C. and Haykin, S. (2011). *Kernel Adaptive Filtering: A Comprehensive Introduction* (Vol. 57): John Wiley & Sons.

Mathews, V. J. (1991). Adaptive polynomial filters. *Signal Processing Magazine, IEEE.* **8**(3): 10-26.

Modaghegh, H., Khosravi, R., Manesh, S. A. and Yazdi, H. S. (2009). *A new modeling algorithm-Normalized Kernel Least Mean Square.* Paper presented at the International Conference on Innovations in Information Technology, 2009 (IIT'09).