NONLINEAR RLS ALGORITHM USING VARIABLE FORGETTING FACTOR IN MIXTURE NOISE

S. H. Leung and C. F. So

Department of Electronic Engineering, City University of Hong Kong
83 Tat Chee Avenue, Hong Kong
e-mail: EEEUGSHL@CITYU.EDU.HK

ABSTRACT

In impulsive noise environment, most learning algorithms are encountered difficulty in distinguishing the nature of large error signal, whether caused by the impulse noise or model error. Consequently, they suffer from large misadjustment or otherwise slow convergence. A new nonlinear RLS (VFF-NRLS) adaptive algorithm with variable forgetting factor for FIR filter is introduced. In this algorithm, the autocorrelations of non-zero lags, which is insensitive to white noise, is used to control forgetting factor of the nonlinear RLS. This scheme makes the algorithm have fast tracking capability and small misadjustment. By experimental results, it is shown that the new algorithm can outperform other RLS algorithms.

1. INTRODUCTION

Recursive least squares (RLS) algorithms have been used extensively in adaptive filtering, self-tuning control systems and system identification [1]. The standard RLS is well known for its good convergence property and small mean square error when the system is time-invariant. However RLS is shown not effective for tracking time-varying parameters because it is difficult to find a suitable forgetting factor to provide good tracking in dealing with large model variations.

Many efforts have been directed to the development of modified RLS algorithm. To maintain the tracking capability of RLS algorithm, modification on the inverse of the covariance matrix are proposed [2-3]. In this scheme, an additional term is added to the inverse that results in improving the tracking and giving good noise immunity. Others try to control the forgetting factor or the effective data window length [4-5]. This approach can maintain the form of the RLS algorithm derived from the least square minimization. However, the control of the forgetting factor in most of these algorithms is sensitive to disturbance and noise.

Most of the noise sources in many practical environments are found to be non-Gaussian in nature [6-7]. Due to some natural and man-made sources, they may exhibit impulsive characteristics. Identification of time varying system in impulsive noise could impose difficulty to most of adaptive systems as their performances may seriously be deteriorated. The reason is due to the fact that the adaptive filters are easily confused by the errors caused by the impulse and model variations. It is shown in the literature that the performance of RLS will be degraded in the presence of impulse noise [8]. The performance of the standard RLS algorithm can be improved by using a nonlinear function in the weight update to limit the estimation error.

In this paper, we introduce a new nonlinear RLS algorithm with variable forgetting factor (VFF) of which the control of the forgetting factor is much less sensitive to impulse noise but can response well to model variations. Unlike other algorithms, the control is based on the autocorrelation values of the error of nonzero lags and constrained by a sigmoidal function. This approach can reduce the affect of the impulse noise and make the change of the forgetting factor directly response to the model variations. Based on the mean square analysis, a control scheme is devised. In this paper, we apply the new VFF scheme to the nonlinear RLS in [8]. Experimental results are presented to illustrate the performance of the new adaptive filter and other VFF RLS are compared.

2. NONLINEAR RLS ALGORITHM

In the linear RLS, the update of the weight vector, $\mathbf{W}(n)$, is described as

$$\mathbf{W}(n+1) = \mathbf{W}(n) + k(n)e(n) \tag{1}$$

where the error signal, $e(n)$, and the Kalman gain vector, $\boldsymbol{k}(n)$, are given by

$$e(n) = d(n) - \mathbf{W}^T(n)\mathbf{X}(n) \tag{2a}$$

$$k(n) = \frac{\mathbf{P}(n-1)\mathbf{X}(n)}{\lambda + \mathbf{X}^T(n)\mathbf{P}(n-1)\mathbf{X}(n)} \tag{2b}$$

$$\mathbf{P}(n) = \lambda^{-1}\left[\mathbf{P}(n-1) - k(n)\mathbf{X}^T(n)\mathbf{P}(n-1)\right] \tag{2c}$$

where $\mathbf{X}(n) = [x(n),..., x(n-N+1)]^T$ is the data vector of length N, $\lambda \in (0,1]$ is the forgetting factor, and $\mathbf{P}(n)$ is the inverse of the correlation matrix given by

$$\mathbf{P}(n) = \left(\sum_{i=1}^{n} \lambda^{n-i}\mathbf{X}(i)\mathbf{X}^T(i) + \lambda^n \delta^{-1} I\right)^{-1} \tag{3}$$

where δ is the initial value. In (2a), the desired signal $d(n)$ for system identification can be written as

$$d(n) = \mathbf{W}_0{}^T\mathbf{X}(n) + \eta(n) \tag{4}$$

where $\mathbf{W}_0$ is the desired weight vector. In (4), the noise component $\eta(n)$ is assumed to be independent and identically distributed (i.i.d.) with zero mean and variance

σ_η^2. The noise model is a mixture density with the *pdf* defined as

$$f_\eta(x) = (1-A) f_0(x) + A f_1(x) \quad (5)$$
$$= \sum_{m=0}^{1} p_m f_m(x), \quad p_0 = 1-A, \quad p_1 = A$$

where A is the impulse index, $f_m(x)$ is taken to be a Gaussian *pdf* with zero mean and variance given by

$$\sigma_m^2 = \sigma_g^2 + \sigma_i^2, \quad m = 0,1 \quad (6)$$

where σ_g^2 and σ_i^2 are the variances of the nominal Gaussian component and the impulsive component, respectively. The ratio of the power in the nominal Gaussian component to that in the impulse component is defined as $\Gamma = \sigma_g^2 / A\sigma_i^2$. The variance σ_η^2 is given by

$$\sigma_\eta^2 = (1-A)\sigma_0^2 + A\sigma_1^2 . \quad (7)$$

The nonlinear RLS of interest has the weight vector update described by

$$\mathbf{W}(n+1) = \mathbf{W}(n) + k(n) g\{e(n)\} \quad (8)$$

where $g\{x\}$ is an odd-symmetric error-saturation nonlinear function. The nonlinear function is the generalized clipping function defined by [8]

$$g\{x\} = \begin{cases} rT_0 & x > T_0 \\ x & |x| \le T_0 \\ -rT_0 & x < -T_0 \end{cases} \quad (9)$$

where r is the clipping parameter between zero and one. The attraction of using this nonlinearity is its simplicity.

In next section, we will discuss about a strategy to control the forgetting factor in (2c) to make the algorithm functioning well in time varying environments.

3. VARIABLE FORGETTING FACTOR

The general strategy for the control of variable forgetting factor (VFF) can be described as follows. Large forgetting factor (effectively large memory of data) is used when the learning is in the steady state and also there is no obvious model variation, while small one (to fade away the very old data) is applied when the model error is large. In time varying environment, the control should be able to sense the change of the model and reduce the disturbance from the noise.

In the environment with impulsive noise, at the incident of large error signal, there could be two possibilities. The error is due to either large model variation or impulse noise. In case the former one occurred, the forgetting factor should be adjusted to make the filter response to the change; otherwise, the forgetting factor should remain large to neglect the effect of the impulse noise.

Before we discuss about the control scheme, let us observe how the value of the forgetting factor affects the mean square error.

3.1 Mean square error and forgetting factor

For the sake of brevity, we state only the mean square error of the standard RLS algorithm without the detail of the derivation. The mean square error is defined as $\sigma_e^2(n) = E\{e^2(n)\}$ For standard RLS and sufficiently large n, the mean square error is recursively given by

$$\sigma_e^2(n+1) = A_c \sigma_e^2(n) + (1-A_c)\sigma_\eta^2 + (a+bN)^{-1} N \sigma_\eta^2 \quad (10)$$

where

$$A_c = 1 - 2\rho^{-1}(n) + (a+bN)^{-1}(N+2)\sigma_x^4$$
$$a = (\tilde{\rho}(n) + \rho^2(n))\sigma_x^4$$
$$b = \tilde{\rho}(n)\sigma_x^4, \qquad \rho(n) = (1-\lambda^n)/(1-\lambda) \approx 1/(1-\lambda)$$
$$\tilde{\rho}(n) = (1-\lambda^{2n})/(1-\lambda^2) \approx 1/(1-\lambda^2)$$

The mean square error is the most relevant objective to select the forgetting factor. One way is to find the forgetting factor to minimize $\sigma_e^2(n+1)$ in (10). Let $s = \sigma_e^2(n)/\sigma_\eta^2$ be the ratio between the mean square error at the n-th iteration and the noise variance. For a given ratio s, we can find an optimum λ to minimize $\sigma_e^2(n+1)$. To show the relation between the mean square error and forgetting factor for a given ratio s, we plot the value of the expression on the right hand side of (10) versus against λ for s=1.1, 2, 4, 6 with $\sigma_\eta^2 = 1$ and N=9 in Fig.1. It is observed that for large ratios, the smaller the forgetting factor, the smaller is the mean square error. On the other hand, for small ratio (s<4), the larger the forgetting factor, the smaller is the mean square error. Hence when there is large model error, we should set the forgetting factor to the permissible minimum value $\lambda_{\min}$ (when $\lambda_{\min}$ is too small, there will be nearly zero memory). Based on the concept of exponential time constant, we can express $\lambda = \exp(-1/\tau_0)$, where τ_0 is the effective time constant roughly related to data memory length. It is recommended to set $\lambda_{\min}$ =0.6 at which τ_0 =2.

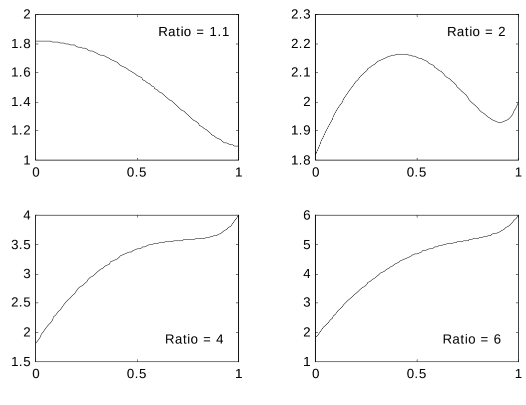

Fig. 1. Plot of $\sigma_e^2(n+1)$ versus λ for different ratios s

In order to avoid the disturbance by the impulse noise, we propose to use the autocorrelations of nonzero lags to measure the model error and control the forgetting factor for those model errors not very large.

3.2 Control scheme for forgetting factor

Based on the previous discussion about the relation between the mean square error and the forgetting factor, we introduce a control scheme for the forgetting factor. The scheme is basically composed of two parts. When the error signal is very large, we set the forgetting factor to $\lambda_{\min}$; otherwise the forgetting factor is governed by a sigmoidal function as given in (10)

$$\lambda(n) = \begin{cases} \lambda_{\min} & \textit{large error} \\ \dfrac{1}{1+e^{-1/\bar{R}_{ee}(n)}} & \textit{otherwise} \end{cases}, \qquad (11)$$

In (10) the modified autocorrelation $\bar{R}_{ee}(n)$ is defined as

$$\bar{R}_{ee}(n) = \frac{1}{T(\tau_2 - \tau_1)} \sum_{m=\tau_1}^{\tau_2} \left| \sum_{i=n-T+1}^{n} e(i)e(i-m) \right|, \qquad (12)$$

where T is the length of the short-time window and the lags in the autocorrelations are nonzero.

To define the large error in the scheme, we will use the modified autocorrelation $\bar{R}_{ee}(n)$ to measure against its average $R_{av}(n)$. Whenever $\bar{R}_{ee}(n)$ is larger than four times of the average $R_{av}(n-1)$, the forgetting factor is set equal to $\lambda_{\min}$. The average is calculated recursively by

$$R_{av}(n) = \beta R_{av}(n-1) + (1-\beta)\bar{R}_{ee}(n) \qquad (13)$$

Considering independent noise with short correlation lag, the autocorrelation $\bar{R}_{ee}(n)$ will not be so affected by the noise if the lags in (11) are larger than the correlation lag of the noise. In other words, $\bar{R}_{ee}(n)$ is a term suitable for the tracking of the change in the time varying model. In our experiments, we consider only white noise. Hence, we set $\tau_1 = 1$ and $\tau_2 = 2$. The window length T in the correlation is not necessary too large and T=5 is sufficient to provide satisfactory performance.

4. EXPERIMENTAL RESULTS

Experiments on system identification are carried out to evaluate the performance of the proposed nonlinear RLS algorithm. The new algorithm is denoted by VFF-NRLS. Two other algorithms as described in [2] and [4] are compared and they are respectively denoted by SPRLS and FKY. The standard RLS with VFF will be denoted by VFF-RLS.

In the experiments, the system is time varying which is switched between

$$\mathbf{W}_0 = \{0.2,-0.4,0.6,-0.8,1,-0.8,0.6,-0.4,0.2\}$$

and $\mathbf{W}_1 = \{1,-0.8,0.6,-0.4,0.2,-0.4,0.6,-0.8,1\}$ at every 200 iterations. The initial value δ is set equal to 1.

Two mixture noise models, { $A = 0.1$, $\Gamma = 0.1$ } and {A=0.01, Γ =1}, are considered. The learning curves and simulation results are averaged over 200 runs. In the control of forgetting factor, we set τ_1=1, τ_2=2, and the window length T=5. The clipper function parameter r is set equal to 1. The threshold T_0 in (9) is given by

$$T_0 = T_0' \tilde{\sigma}_e^2(n) \qquad (14)$$

where the variance of the error is computed by

$$\tilde{\sigma}_e^2(n+1) = \alpha\tilde{\sigma}_e^2(n) + (1-\alpha)e^2(n) \qquad (15)$$

The smoothing parameter α is set equal to 0.9 and the parameter T_0' is fixed to 1.

In Fig.2, we plotted the learning curves on the mean square model error of VFF-NRLS for SNR=19.63dB. The smoothing parameter β in (13) is set equal to 0.995. The mean square model error is defined as

$$\sigma_w^2 = E\left\{ \left\| \mathbf{W}(n) - \mathbf{W}_0 \right\|_2^2 \right\} \sigma_x^2 \qquad (16)$$

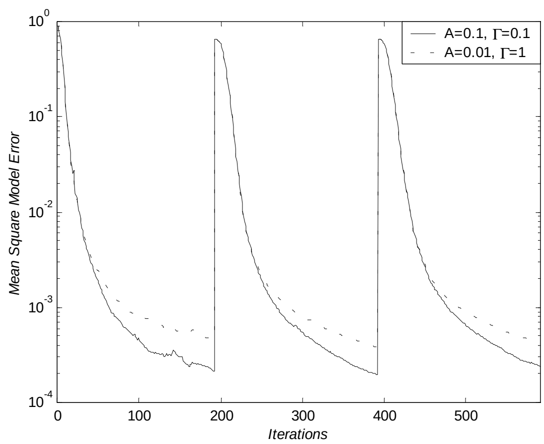

Fig. 2.Mean square convergence of VFF-NRLS in mixture noise for two different model parameters

The results show that the VFF-NRLS algorithm can provide a robust tracking performance in different mixture noise and give smaller misadjustment for large A.

In Fig.3, we plotted $\bar{R}_{ee}(n)$ to illustrate the tracking of the model error of using the modified autocorrelations in mixture noise. The results verify that $\bar{R}_{ee}(n)$ is able to identify the model errors and is insensitive to the disturbance of the impulsive noise. Especially the $\bar{R}_{ee}(n)$ can follow along with the large model errors quite well.

In Fig.4, we compare the mean square model error of VFF-NRLS with SPRLS and FKY for { $A = 0.1$, $\Gamma = 0.1$ }. The results show that the new algorithm has better tracking capability and much smaller misadjustment than other variable forgetting factor RLS. In this experiment, all the

VFF algorithms are applied to the nonlinear RLS algorithm. On the other hand, in Fig.5, we compare the new variable forgetting factor algorithm and other two VFF algorithms on the standard RLS in the same mixture noise. Comparing the results in Fig.4 to that of Fig.5, it is observed that the nonlinear RLS can yield much smaller model error than the standard RLS in the steady state, while the standard RLS can converge slightly faster than the nonlinear RLS algorithm.

5. CONCLUSIONS

A new variable forgetting factor scheme for RLS algorithm in impulsive noise is presented. The scheme is basically derived from the minimization of the mean square error. Using autocorrelations of nonzero lags is shown effective to track model. Simulation results show that the new algorithm yields faster convergence and much smaller steady state mean square error than the existing variable forgetting factor RLS algorithms.

6. REFERENCES

[1] S. Haykin, "Adaptive Filter Theory", 2nd Ed., Prentice Hall, New York, 1991.

[2] D. J. Park and B. E. Jun, "Self-perturbing RLS algorithm with fast tracking capability", Electronic Letters, vol.28, pp.558-559, March 1992.

[3] J. Jiang and R. Cook, "Fast Parameter Tracking RLS Algorithm with High Noise Immunity," Electronics Letters, vol.28, pp.2043-2045, Oct. 1992.

[4] T. R. Fortescue, L. S. Kershenbaum and B. E. Ydstie, "Implementation of self-tuning regulators with variable forgetting factors", Automatica, vol.17, pp.831-835, 1981.

[5] S. Song, et al, "Gauss Newton Variable Forgetting Factor Recursive Least Squares for Time Varying Parameter Tracking," Electronics Letters, vol.36, pp.988-990, May 2000.

[6] D. Middleton, "Non-Gaussian Noise Models in Signal Processing for Telecommunications: New Methods and Results for Class A and Class B Noise Models," IEEE Trans. Inform. Theory, vol.45, pp.1129-1149, May 1999.

[7] C. L. Nikias and Min Shao, Signal Processing with Alpha-Stable Distributions and Applications, John Wiley, 1995.

[8] S. H. Leung and J. F. Weng, "Nonlinear RLS algorithm for amplitude estimation in class A noise", Communications, IEE Proceedings-Commun., vol.147, pp.81-86 April 2000.

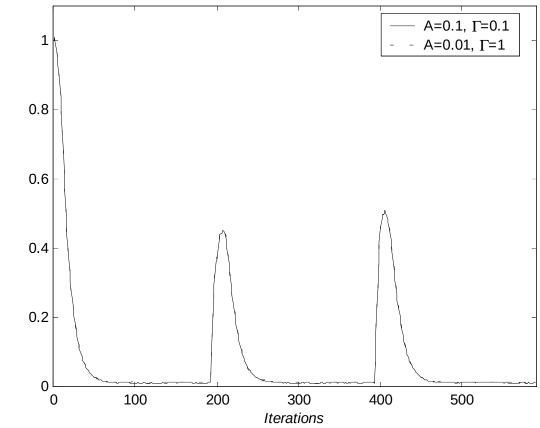

Fig. 3. Plot of $\overline{R}_{ee}(n)$ for two mixture noise models

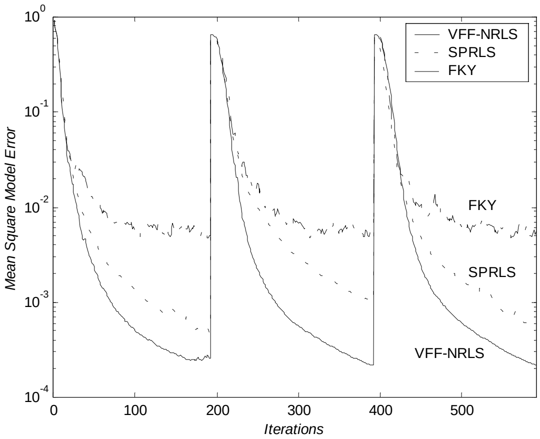

Fig. 4 Performance comparison of VFF-NRLS, SPRLS and FKY algorithms

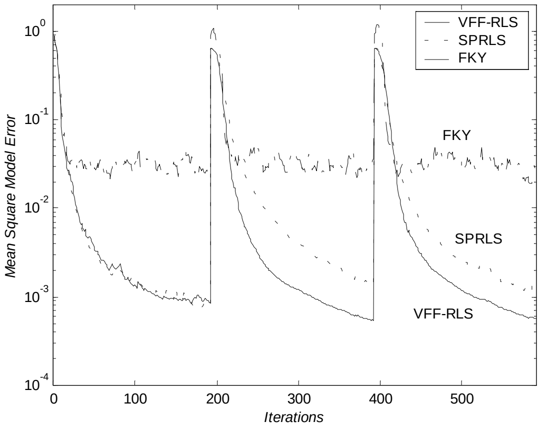

Fig. 5. Performance comparison of VFF-RLS, SPRLS and FKY algorithms