The Effect of the Forgetting Factor on the RI Adaptive Algorithm in System Identification

Mohammad Shukri Ahmad*†, Osman Kukrer* and Aykut Hocanin*

*Electrical and Electronic Engineering Department
Eastern Mediterranean University
Email: {mohammad.ahmad, osman.kukrer, aykut.hocanin}@emu.edu.tr

†Electrical and Electronic Engineering Department
European University of Lefke
Email: msalman@eul.edu.tr

Abstract— **The recently proposed Recursive Inverse (RI) algorithm was shown to have a similar mean-square-error (mse) performance as the Recursive-Least-Squares (RLS) algorithm with reduced complexity. The selection of the forgetting factor has a significant influence on the performance of the RLS algorithm. The value of the forgetting factor leads to a tradeoff between the stability and the tracking ability. In a system identification setting, both the filter length and a leakage phenomenon affect the selection of the forgetting factor. In this paper, we first analytically show that this leakage phenomenon and the filter length have much less influence on the performance of the RI algorithm. Simulation results, in a system identification setting, validate the theoretical results.**

I. Introduction

In adaptive filtering, the Recursive Least Squares (RLS) algorithm is known to be one of the most efficient algorithms, in its fast convergence rate and high performance, especially in highly correlated input data [1]-[3]. The RLS algorithm outperforms the Least-Mean-Square (LMS) algorithm [1], in both mean-square-error (mse) and speed of convergence. However, this high performance is at the expense of its high computational complexity.

In the RLS algorithm, the forgetting factor is chosen between zero and unity. However, there is a compromise between the performance criteria depending on the value of this forgetting factor [4]: When the forgetting factor is very close to unity, the RLS algorithm achieves good stability, whereas its tracking capabilities are reduced. A smaller value of the forgetting factor improves the tracking capabilities of the RLS algorithm but it decreases its probability of being stable [1].

Let us consider a system identification setting [1]. The output of the unknown system is assumed to be corrupted by an additive white Gaussian noise (AWGN). In this scenario, the aim of the adaptive filter is to recover the system noise from the error signal after it converges to the true solution. However the error signal must not go to zero, since this may introduce noise in the adaptive filter. Moreover, the RLS algorithm has a feature which makes the problem even more complicated. Under some conditions, it is able to cancel the error of the adaptive filter even in the presence of system noise. In other words, the system noise *leaks* into the output of the adaptive filter, which of course may lead to an incorrect solution to the problem. This *leakage* depends on the values of the forgetting factor and the length of the adaptive filter [6].

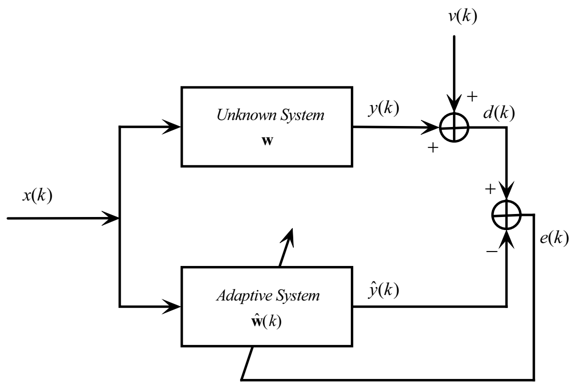

Fig. 1. Block diagram of adaptive system identification.

In this paper, we analyze this *leakage* phenomenon for the recently proposed RI algorithm [5], and give a theoretical estimation of it in terms of the forgetting factor β and the filter length N. These findings will be compared with those of the RLS algorithm introduced in [6]. The paper is organized as follows. In section II, the theoretical analysis of the *leakage* phenomenon, for the system identification setting of the RI adaptive algorithm, is introduced. In section III, simulation results that support the findings in section II are presented. Finally, in section IV, conclusions are drawn.

II. The Leakage Phenomenon

A. System Identification - Ideal Behavior

In the system identification setting shown in Fig. 1, our objective is to estimate the unknown system coefficients using an adaptive filter, both driven by the same input signal $x(k)$. The unknown and the adaptive systems are assumed to be, both, Finite-Impulse-Response (FIR) filters of length N, defined by the tap-weight vectors

978-1-61284-943-0

$$\mathbf{w} = [w_0 \; w_1 \; \ldots \; w_{N-1}]^T,$$

and

$$\hat{\mathbf{w}}(k) = [\hat{w}_0(k) \; \hat{w}_1(k) \; \ldots \; \hat{w}_{N-1}(k)]^T,$$

where T is the transposition operator.
As seen from Fig. 1, $y(k)$, the output of the unknown system is given by

$$y(k) = \mathbf{x}^T(k)\mathbf{w}, \tag{1}$$

where $\mathbf{x}(k)$ is the tap-input vector that contains the N most recent samples of the input signal given as,

$$\mathbf{x}(k) = [x(k) \; x(k-1) \; \ldots \; x(k-N+1)]^T.$$

The desired response $d(k)$ is defined as,

$$d(k) = y(k) + \nu(k) = \mathbf{x}^T(k)\mathbf{w} + \nu(k), \tag{2}$$

where $\nu(k)$ is the system (measurement) noise that corrupts the output of the unknown system, $y(k)$.

According to the block diagram in Fig. 1, the error signal $e(k)$ is

$$e(k) = [y(k) - \hat{y}(k)] + \nu(k). \tag{3}$$

From the Wiener-Hopf equation [1],

$$\mathbf{R}\hat{\mathbf{w}} = \mathbf{p}, \tag{4}$$

where $\mathbf{R}$ and $\mathbf{p}$ are the expected values of the autocorrelation matrix of the tap-input vector and cross-correlation vector between the desired output signal and the tap-input vector, respectively, and given by

$$\begin{aligned} \mathbf{R} &= E\left\{\mathbf{x}(k)\mathbf{x}^T(k)\right\}, &(5)\\ \mathbf{p} &= E\left\{\mathbf{x}(k)d(k)\right\}. &(6) \end{aligned}$$

Substituting (2) in (6) gives

$$\begin{aligned} \mathbf{p} &= E\left\{\mathbf{x}(k)d(k)\right\} \\ &= E\left\{\mathbf{x}(k)\left[y(k)+\nu(k)\right]\right\} \\ &= E\left\{\mathbf{x}(k)y(k)\right\} + E\left\{\mathbf{x}(k)\nu(k)\right\}, &(7) \end{aligned}$$

assuming that the input $\mathbf{x}(k)$ and the system noise $\nu(k)$ are statistically independent, then the second term of (7) will be eliminated and, therefore, substituting (1) in (7) gives

$$\begin{aligned} \mathbf{p} &= E\left\{\mathbf{x}(k)y(k)\right\} \\ &= E\left\{\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{w}\right\} \\ &= \mathbf{R}\mathbf{w}. &(8) \end{aligned}$$

According to (4)

$$\hat{\mathbf{w}} = \mathbf{w}, \tag{9}$$

and

$$\hat{y}(k) = \mathbf{x}^T(k)\hat{\mathbf{w}} = \mathbf{x}^T(k)\mathbf{w} = y(k). \tag{10}$$

Substituting the result of (10) in (3) yields,

$$e(k) = \nu(k). \tag{11}$$

Or in other words, $y(k)$ and $\nu(k)$ are correctly separated.

B. System Identification - RI Algorithm

In the case of the RI algorithm, (5) and (6) are replaced by the recursive estimates of the instantaneous correlations.

$$\begin{aligned} \mathbf{R}(k) &= \beta\mathbf{R}(k-1) + \mathbf{x}(k)\mathbf{x}^T(k), &(12)\\ \mathbf{p}(k) &= \beta\mathbf{p}(k-1) + \mathbf{x}(k)d(k), &(13) \end{aligned}$$

where β is the forgetting factor ($0 < \beta < 1$).
Reconsidering (3),

$$\begin{aligned} e(k) = d(k) - \hat{y}(k) &= \mathbf{x}^T(k)\mathbf{w} + \nu(k) - \mathbf{x}^T(k)\mathbf{w}(k) \\ &= \mathbf{x}^T(k)\left[\mathbf{w} - \mathbf{w}(k)\right] + \nu(k), &(14) \end{aligned}$$

where $\mathbf{w}(k)$ is the recursive estimate of $\mathbf{w}$ given by [5], [7]:

$$\mathbf{w}(k) = [\mathbf{I} - \mu(k)\mathbf{R}(k)]\mathbf{w}(k-1) + \mu(k)\mathbf{p}(k). \tag{15}$$

where $\mu(k)$ is the variable step-size [5], [7] defined as:

$$\mu(k) < \frac{2}{\lambda_{max}(\mathbf{R}(k))} = \left(\frac{1-\beta}{1-\beta^k}\right)\left(\frac{2}{\lambda_{max}(\mathbf{R}_{xx})}\right) = \frac{\mu_{max}}{1-\beta^k},$$

As $k \to \infty$,

$$\mathbf{w}(k) = \mathbf{w} + \delta\widetilde{\mathbf{w}}(k), \tag{16}$$

where $\mathbf{w}$ and $\delta\widetilde{\mathbf{w}}(k)$ are the optimum solution and the stochastic part of $\mathbf{w}(k)$, respectively.
Substituting (16) in (14) yields

$$e(k) = -\mathbf{x}^T(k)\delta\widetilde{\mathbf{w}}(k) + \nu(k), \tag{17}$$

For a good estimate of the unknown system, the difference between the output of the adaptive filter, $\hat{y}(k)$, and the output of the unknown system, $y(k)$, should approach zero in the steady state,

$$\begin{aligned} \hat{y}(k) - y(k) &= \hat{y}(k) - d(k) + \nu(k) \\ &= \hat{y}(k) - (e(k) + \hat{y}(k)) + \nu(k) \\ &= \nu(k) - e(k) = \mathbf{x}^T(k)\delta\widetilde{\mathbf{w}}(k). &(18) \end{aligned}$$

As a result of (18),

$$\hat{y}(k) = y(k) + \mathbf{x}^T(k)\delta\widetilde{\mathbf{w}}(k). \tag{19}$$

where $\delta\widetilde{\mathbf{w}}(k)$ [7] is given by

$$\delta\widetilde{\mathbf{w}}(k) = \left[\mathbf{I} - \mu_0\mathbf{R}_{xx}\right]\delta\widetilde{\mathbf{w}}(k-1) + \mu(k)\sum_{i=0}^{k}\beta^{k-i}\mathbf{x}(i)\nu(i). \tag{20}$$

where μ_0 is constant ($\mu_0 < \mu_{max}$) and $\mathbf{R}_{xx} = E\left[\mathbf{x}(k)\mathbf{x}^T(k)\right]$.

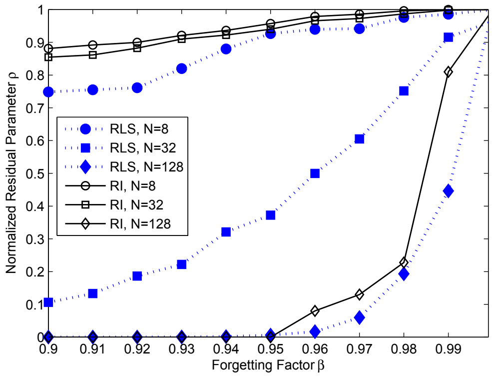

Fig. 2. Normalized residual parameter ρ for different values of the forgetting factor β and the filter length N.

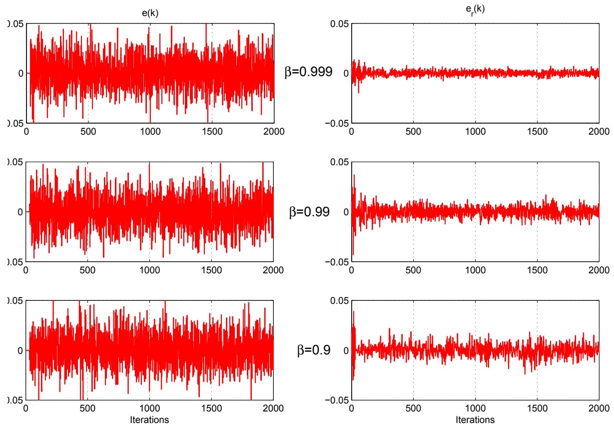

Fig. 3. The leakage phenomenon of The RI algorithm in time domain $N = 32$ and different values of β.

From (19) and (20), we note that, in a system identification setting, the output of the adaptive filter, $\hat{y}(k)$, always contains a component which is proportional to the system noise, $\nu(k)$. This phenomenon is a leakage of the system noise, $\nu(k)$, in the output of the adaptive filter, $\hat{y}(k)$.
By (19), this leakage value is given as,

$$r(k) = \mathbf{x}^T(k)\delta\widetilde{\mathbf{w}}(k). \tag{21}$$

A normalized residual parameter [6], $\rho(k)$, may be defined as

$$\rho(k) = \left|\frac{\nu(k) - r(k)}{\nu(k)}\right|. \tag{22}$$

When total leakage occurs, then $\nu(k) = r(k)$ and $\rho(k) = 0$. This usually occurs at low values of the forgetting factor and high values of the filter length, as shown in section III.

III. Simulation Results

In this section, we simulate the system identification problem shown in Fig. 1. The input signal was generated using [8]:

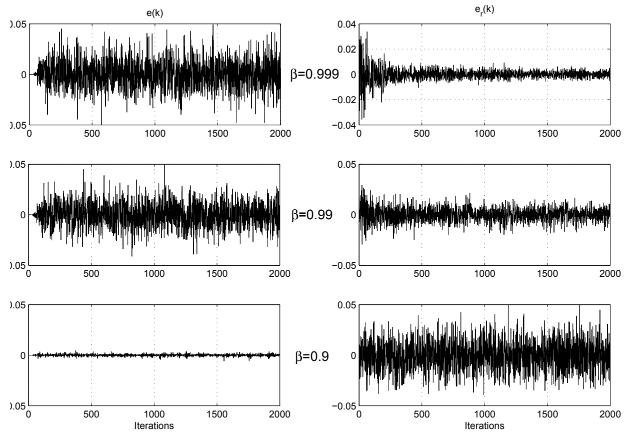

Fig. 4. The leakage phenomenon of The RLS algorithm in time domain $N = 32$ and different values of β.

$$\begin{aligned} x(k) &= 1.79x(k-1) - 1.85x(k-2) + 1.27x(k-3) \\ &- 0.41x(k-4) + v_0(k). \end{aligned} \tag{23}$$

where $v_0(k)$ is a Gaussian process with zero mean and variance $\sigma^2 = 0.3849$. The system noise $\nu(k)$ is assumed to be AWGN with zero mean and variance ($\sigma_\nu^2 = 0.015$).

Fig. 2 shows the normalized residual parameter estimate of the RI and RLS algorithms for different values of the filter length, N. This term was averaged over the last 500 samples of the steady-state of both algorithms. It is seen that, when the filter length is relatively small (8-32 taps) the RI algorithm is less sensitive to the changes in the forgetting factor, β, and the filter length, N, than the RLS algorithm. On the other hand, when the filter length is relatively high (i.e. $N = 128$ taps), the RI algorithm becomes sensitive to the value of β (the larger the N the larger the β should be).

In order to investigate the *leakage* phenomenon, two separate experiments were done. In order to study the effect of the forgetting factor β on the leakage phenomenon, the filter length is held constant ($N = 32$) and the forgetting factor, β, is assigned different values (i.e. $\beta = 0.9, 0.99, 0.999$). The recovered signal ($e(k)$) and the residual error ($e_r(k) = e(k) - \nu(k)$) are plotted for each value of β for both, the RI and the RLS, algorithms. It can be seen from Fig. 3 and Fig. 4 that the *leakage* in the RI algorithm case is much less than that of the RLS algorithm, especially, when the forgetting factor β is relatively small (i.e $\beta = 0.9, 0.99$). Also, to observe this *leakage* in the frequency domain, we repeat the previous experiments and plot the PSD's of both, $e(k)$ and $e_r(k)$ in Fig. 5 and Fig. 6. The conclusions are, almost, the same as those in the time domain; the recovered signal is strongly attenuated when the value of β is relatively small for the RLS algorithm. However, in the case of the RI algorithm, this attenuation become much less especially when the forgetting factor and/or the filter length are relatively small.

In order to study the effect of the filter length on the *leakage* phenomenon, now, the forgetting factor, β, is held constant

978-1-61284-943-0

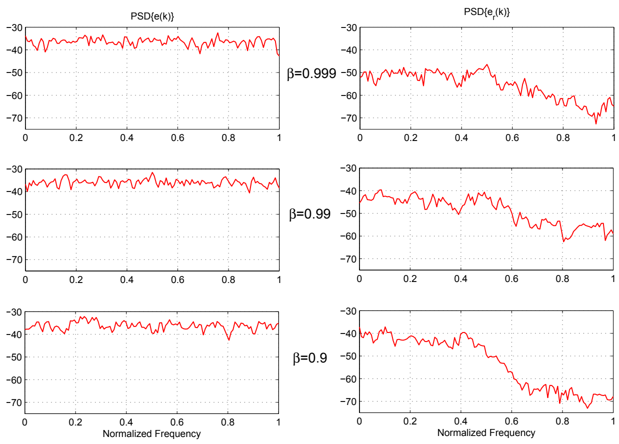

Fig. 5. Power Spectral Density of Fig. 3.

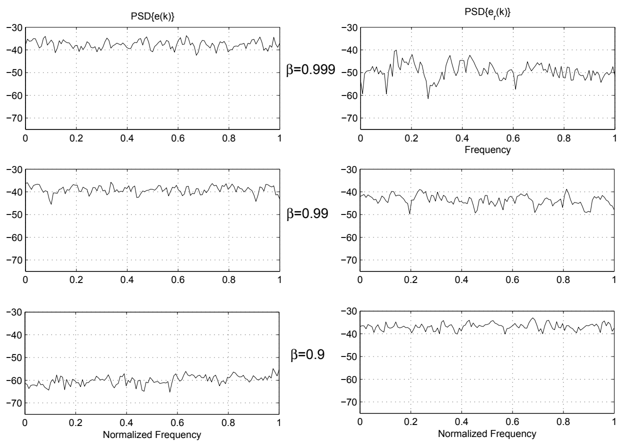

Fig. 6. Power Spectral Density of Fig. 4.

($\beta = 0.99$) and the filter length is assigned different values (i.e. $N = 8, 32, 64$). The recovered signal ($e(k)$) and the residual error ($e_r(k)$) are plotted for each value of N for the RI and the RLS algorithms. Fig. 7 and Fig. 8 show that the *leakage* in the RI algorithm case is much less than that of the RLS algorithm, especially, when the filter length N is relatively small (i.e $N = 8, 32$ taps).

IV. Conclusions

In this paper, we showed the existence of a *leakage* phenomenon that appears when the RI and RLS algorithms are used in a system identification setting. We theoretically proved that this *leakage* is proportional to the system noise, and highly dependent on the values of the forgetting factor and the filter length. Simulation results validate the theoretical results, and they show that for a relatively small forgetting factor and/or a relatively large filter length, this leakage is high or total in some cases. However, simulations show that the RI algorithm is much less sensitive to these parameters (β and N) than the RLS algorithm. Both the theoretical and experimental results lead to the conclusion that the *leakage* phenomenon can be avoided when the value of β is very close to unity, in both algorithms. Nevertheless, it should be noted that, when β is relatively small, the tracking capability of the algorithms is reduced.

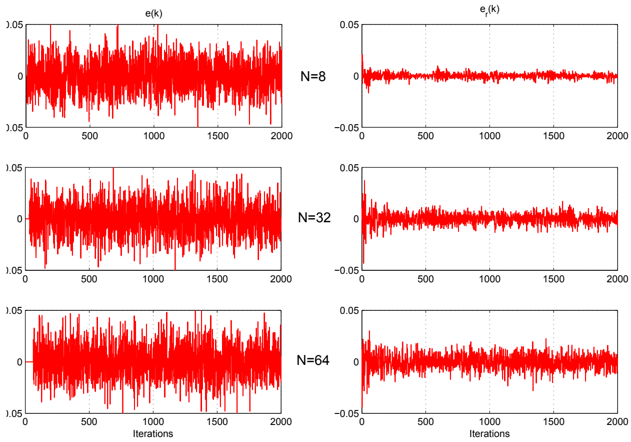

Fig. 7. The leakage phenomenon of the RI algorithm in time domain $\beta = 0.99$ and different values of N.

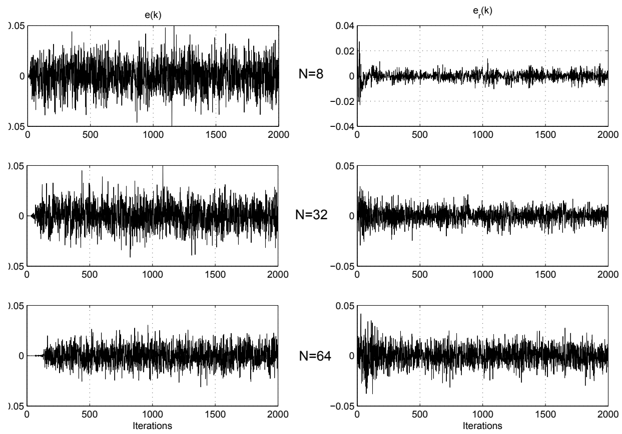

Fig. 8. The leakage phenomenon of the RLS algorithm in time domain $\beta = 0.99$ and different values of N.

References

[1] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Upper Saddle River, NJ, 2002.

[2] A. H. Sayed, *Adaptive Filters*, John Wily & Sons, NJ, 2008.

[3] P. S. R. Diniz, *Adaptive Filtering Algorithms and Practical Implementation*, 3^{rd} Ed., Springer, LLC, NY, 2008.

[4] C. Paleologu, J. Benesty, and S. Ciochina, “A robust variable forgetting factor recursive least-squares algorithm for system identification,” *IEEE Signal Processing Letters*, vol. 15, pp. 597-600, 2008.

[5] M. S. Ahmad, O. Kukrer and A. Hocanin, “An efficient Recursive Inverse adaptive filtering algorithm for channel equalization,” *European Wireless Conference* (EW 2010), pp. 88-92, April 2010.

[6] S. Ciochina, C. Paleologu, J. Benesty and A. A. Enescu, “On the influence of the forgetting factor of the RLS adaptive filter in system identification,” *International Symposium on Signals, Circuits and Systems (ISSCS 2009)*, pp.1-4, 2009.

[7] M. S. Ahmad, O. Kukrer and A. Hocanin, “Recursive Inverse adaptive filtering algorithm,” *Elsevier Journal of Digital Signal Processing*, DOI:10.1016/j.dsp.2011.03.001, March 2011.

[8] R. C. Bilcu, P. Kuosmanen and K. Egiazarian, “A Transform Domain LMS Adaptive Filter with Variable Step-Size”, *IEEE Signal Processing Letters*, vol. 9, no. 2, pp. 51-53, February. 2002.