ON PERFORMANCE BOUNDS FOR AN AFFINE COMBINATION OF TWO LMS ADAPTIVE FILTERS

N. J. Bershad, J. C. M. Bermudez† and J.-Y. Tourneret⋆*

* University of California, Irvine, 1621 Santiago Drive, Newport Beach, CA, 92660, USA
† Federal University of Santa Catarina, Florianopolis, SC Brazil
⋆ IRIT/ENSEEIHT/TéSA, 2 rue C. Camichel, BP 7122, 31071 Toulouse cedex 7, France
bershad@ece.uci.edu, j.bermudez@ieee.org, jean-yves.tourneret@enseeiht.fr

ABSTRACT

This paper studies the statistical behavior of an affine combination of the outputs of two LMS adaptive filters that simultaneously adapt using the same white Gaussian input. The purpose of the combination is to obtain an LMS adaptive filter with fast convergence and small steady-state mean-square error (MSE). The linear combination studied is a generalization of the convex combination, in which the combination factor is restricted to the interval $(0, 1)$. The viewpoint is taken that each of the two filters produces dependent estimates of the unknown channel. Thus, there exists a sequence of optimal affine combining coefficients which minimizes the MSE. The optimal unrealizable affine combiner is studied and provides the best possible performance for this class. Then, a new scheme is proposed for practical applications. It is shown that the practical scheme yields close-to-optimal performance when properly designed (as suggested by the theoretical optimal).

***Index Terms*—** Adaptive filters, affine combination, convex combination, LMS algorithm, stochastic algorithms.

1. INTRODUCTION

Adaptive filter design usually requires a fundamental trade-off between convergence speed and steady-state MSE. A faster (slower) convergence speed yields a larger (smaller) steady-state mean-square deviation (MSD) and MSE. This trade-off is usually controlled by design parameters such as the step-size in the LMS algorithm. Variable step-size modifications of the basic adaptive algorithms offer a possible solution to this design problem [1, 2]. Recently, a novel scheme has been proposed that uses a convex combination of two fixed step-size adaptive filters as shown in Fig. 1 [3]. The adaptive filter $W_1(n)$ uses a larger step-size than the adaptive filter $W_2(n)$. The key to this scheme is the selection of the scalar mixing parameter $\lambda(n)$ for combining the two filter outputs. For instance, the mixing parameter is defined in [3] as a sigmoid function. The free parameter is then adaptively optimized using a stochastic gradient search which minimizes the quadratic error of the overall filter. The steady-state performance of this adaptive scheme has been studied in [3]. The convex combination performed as well as the best of its components in the mean-square sense. The results indicate that a combination of adaptive filters can lead to fast convergence rates and good steady-state performance, an attribute usually obtained only in variable step-size algorithms. Thus, there is great interest in learning more about the properties of such adaptive structures.

This paper provides new results on the performance of the combined structure which supplement the work presently available in the literature. The achievable performance is studied for an affine combination of two LMS adaptive filters with white Gaussian inputs and the structure shown in Fig. 1. The parameter $\lambda(n)$ is not restricted to the range $(0, 1)$ as it was in [3]. Thus, the output $y(n)$ is an affine combination of the individual outputs $y_1(n)$ and $y_2(n)$. Since each adaptive filter is estimating the unknown optimal impulse response using the same input data, $W_1(n)$ and $W_2(n)$ are statistically dependent estimates of the unknown response. There exists a single combining parameter sequence which minimizes the MSE.

The adaptive scheme is first studied from the viewpoint of an optimal affine combiner. The value of $\lambda(n)$ that minimizes the MSE for each time instant n (conditioned on the filter parameters at iteration n) is determined as a function of the unknown system response. This leads to an optimal affine sequence. The statistical properties of the optimal combiner are then studied. It is shown that the optimal $\lambda(n)$ can be outside the interval $(0, 1)$ for several iterations. More importantly, the optimal $\lambda(n)$ is usually negative in steady-state. Although the optimal combiner is unrealizable, its performance provides an upper bound on the performance of any realizable linear combiner. If a suboptimal (but realizable) algorithm has a performance close to that of the optimal combiner, sufficient motivation exists for a more detailed study of the algorithm with respect to analysis and implementation issues. The performance of the adaptive filter using a suboptimal but realizable adjustment algorithm for $\lambda(n)$ is compared to that of the optimal combiner. The realizable scheme is based upon the relative values of time-averaged estimates of the individual adaptive filter error powers. It is shown that this realizable combiner rule can be designed so that the adaptive filter performance is comparable to the optimal combiner.

2. THE OPTIMAL AFFINE COMBINER

2.1. The Affine Combiner

The system under investigation is shown in Fig. 1. Each filter uses the LMS adaptation rule but with different step sizes μ_i, $i = 1, 2$. For an input vector $\boldsymbol{U}(n) = [u(n), ..., u(n-N+1)]^T$,

$$\boldsymbol{W}_i(n+1) = \boldsymbol{W}_i(n) + \mu_i e_i(n)\boldsymbol{U}(n), \tag{1}$$

for $i = 1, 2$, where

$$e_i(n) = d(n) - \boldsymbol{W}_i^T(n)\boldsymbol{U}(n), \tag{2}$$

$$d(n) = e_0(n) + \boldsymbol{W}_0^T\boldsymbol{U}(n), \tag{3}$$

and $e_0(n)$ is white and uncorrelated with $u(n)$. Thus, the $\boldsymbol{W}_i(n)$ are coupled both deterministically and statistically through $\boldsymbol{U}(n)$ and

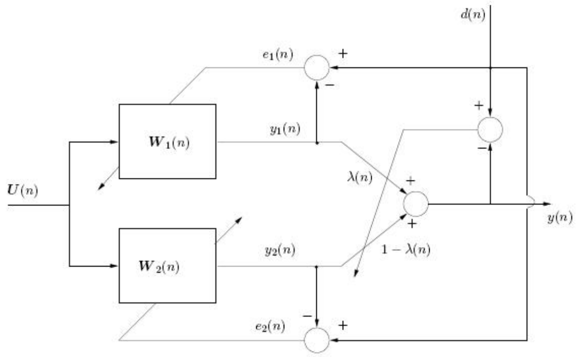

Fig. 1. Adaptive combining of two transversal adaptive filters.

$e_0(n)$. The stochastic behavior of (1) is well-known [4–6]. The outputs of the two filters are combined as

$$y(n) = \lambda(n)y_1(n) + [1-\lambda(n)]y_2(n), \tag{4}$$

where $y_i(n) = \boldsymbol{W}_i^T \boldsymbol{U}(n)$ for $i = 1, 2$. Here, $\lambda(n)$ is not constrained to $(0, 1)$. An overall system error is defined as

$$e(n) = d(n) - y(n). \tag{5}$$

2.2. Optimal Combining Rule

Letting $\boldsymbol{W}_{12}(n) = \boldsymbol{W}_1(n) - \boldsymbol{W}_2(n)$, Eq. (4) can be written

$$\begin{aligned} y(n) &= \lambda(n)\boldsymbol{W}_1^T(n)\boldsymbol{U}(n) + [1-\lambda(n)]\boldsymbol{W}_2^T(n)\boldsymbol{U}(n) \\ &= \big\{\lambda(n)\big[\boldsymbol{W}_1(n) - \boldsymbol{W}_2(n)\big] + \boldsymbol{W}_2(n)\big\}^T\boldsymbol{U}(n) \\ &= \big\{\lambda(n)\boldsymbol{W}_{12}(n) + \boldsymbol{W}_2(n)\big\}^T\boldsymbol{U}(n). \end{aligned} \tag{6}$$

Subtracting the versions of (1) for $i = 2$ and $i = 1$ yields:

$$\boldsymbol{W}_{12}(n+1) = \big[\boldsymbol{I} - \mu_1\boldsymbol{U}(n)\boldsymbol{U}^T(n)\big]\boldsymbol{W}_{12}(n) + (\mu_1 - \mu_2)e_2(n)\boldsymbol{U}(n). \tag{7}$$

Next, consider an instantaneous rule for choosing $\lambda(n)$ that minimizes the conditional MSE $E[e^2(n)|\boldsymbol{W}_2(n), \boldsymbol{W}_{12}(n)]$ at time n. Defining $\boldsymbol{W}_{02}(n) = \boldsymbol{W}_0(n) - \boldsymbol{W}_2(n)$, $e(n)$ in (5) can be written

$$\begin{aligned} e(n) &= e_0(n) + [\boldsymbol{W}_0 - \lambda(n)\boldsymbol{W}_{12}(n) - \boldsymbol{W}_2(n)]^T\boldsymbol{U}(n), \\ &= e_0(n) + [\boldsymbol{W}_{02}(n) - \lambda(n)\boldsymbol{W}_{12}(n)]^T\boldsymbol{U}(n), \end{aligned} \tag{8}$$

and equating the gradient of the conditional MSE with respect to $\lambda(n)$ to zero yields

$$\frac{\partial E[e^2(n)|\boldsymbol{W}_2(n), \boldsymbol{W}_{12}(n)]}{\partial \lambda(n)} = 0, \tag{9}$$

or equivalently

$$-2E\left[e(n)\boldsymbol{W}_{12}^T(n)\boldsymbol{U}(n)|\boldsymbol{W}_2(n), \boldsymbol{W}_{12}(n)\right] = 0. \tag{10}$$

Using (8), taking the expectation over $\boldsymbol{U}(n)$ and defining the input conditional autocorrelation matrix

$$\boldsymbol{R}_u = E\left[\boldsymbol{U}(n)\boldsymbol{U}^T(n)|\boldsymbol{W}_2(n), \boldsymbol{W}_{12}(n)\right] \tag{11}$$

results in

$$[\boldsymbol{W}_{02}(n) - \lambda(n)\boldsymbol{W}_{12}(n)]^T\boldsymbol{R}_u\boldsymbol{W}_{12}(n) = 0. \tag{12}$$

Solving (12) for $\lambda(n) = \lambda_0(n)$ yields

$$\lambda_0(n) = \frac{\boldsymbol{W}_{02}^T(n)\boldsymbol{R}_u\boldsymbol{W}_{12}(n)}{\boldsymbol{W}_{12}^T(n)\boldsymbol{R}_u\boldsymbol{W}_{12}(n)}. \tag{13}$$

In the subsequent analysis, $\boldsymbol{R}_u = E\left[\boldsymbol{U}(n)\boldsymbol{U}^T(n)\right] = \sigma_u^2\boldsymbol{I}$ since the input at time n is assumed statistically independent of the weights at time n (Independence Theory [6]). Note that Eq. (13) requires knowledge of $\boldsymbol{W}_0$ and imposes no constraint on $\lambda_0(n)$. This optimal combining rule can be used as a benchmark for evaluation of realizable schemes for online adaptation of $\lambda(n)$.

3. PERFORMANCE OF THE OPTIMAL COMBINER

3.1. Steady-state behavior of $\lambda_0(n)$

Approximating the expected value of (13) by the ratio of the expected values[1], knowing that

$$\lim_{n\to\infty} E[\boldsymbol{W}_i(n)] = \boldsymbol{W}_0, \tag{14}$$

for $i = 1, 2$, and neglecting the correlations between input signal and weight vectors in both filters, it can be shown that the steady-state value of $E[\lambda_0(n)]$ is approximately given by [8]

$$\lim_{n\to\infty} E[\lambda_0(n)] \approx \frac{\mu_2/\mu_1}{2\,(\mu_2/\mu_1 - 1)}. \tag{15}$$

3.2. Mean Square Deviation

Equation (13) yields the minimum MSE at each time instant and also the MSD of the optimal combiner $\text{MSD}_c(n)$ where

$$\begin{aligned} \text{MSD}_c(n) & \\ &= E\Big\{\big[\boldsymbol{W}_{02}(n) - \lambda_0(n)\boldsymbol{W}_{12}(n)\big]^T\big[\boldsymbol{W}_{02}(n) - \lambda_0(n)\boldsymbol{W}_{12}(n)\big]\Big\} \\ &= E\big[\boldsymbol{W}_{02}^T(n)\boldsymbol{W}_{02}(n)\big] + E\Big[\lambda_0^2(n)\boldsymbol{W}_{12}^T(n)\boldsymbol{W}_{12}(n)\Big] \\ &\qquad - 2E\big[\lambda_0(n)\boldsymbol{W}_{02}^T(n)\boldsymbol{W}_{12}(n)\big]. \end{aligned} \tag{16}$$

Inserting (13) in (16) yields

$$\text{MSD}_c(n) = E\big[\boldsymbol{W}_{02}^T(n)\boldsymbol{W}_{02}(n)\big] - E\left\{\frac{\big[\boldsymbol{W}_{02}^T(n)\boldsymbol{W}_{12}(n)\big]^2}{\boldsymbol{W}_{12}^T(n)\boldsymbol{W}_{12}(n)}\right\}. \tag{17}$$

The first term of (17) is $\text{MSD}_2(n)$, the MSD of the second adaptive filter. Since the second term of (17) is a positive quantity, $\text{MSD}_c(n)$ is always less than $\text{MSD}_2(n)$. Similarly, it can be shown that the optimum linear combiner leads to a $\text{MSD}_c(n)$ that is smaller than $\text{MSD}_1(n)$ [8].

[1] There are two main justifications for this approximation: 1) Evaluating expectations of quotients of correlated random variables is usually a very difficult undertaking. Approximations are often made in order to make progress in the analysis. We have chosen the approximation that the expectation of the ratio is approximately the ratio of the expectations (see for instance [7] and reference [13] within), 2) The latter simulation results support this approximation.

Figures 2 and 3 display simulation results obtained with 50 Monte Carlo (MC) runs for $10\log_{10}(\mathrm{MSD}_i(n))$ and $\widetilde{\lambda}_0(n)$ (MC average of $\lambda_0(n)$) for $\sigma_0^2 = 10^{-4}$, $\mu_1 = 1/(N+2)$, $\mu_2 = 0.1\mu_1$ (assuming $\boldsymbol{W}_0$ is known). Figure 2 shows that $\mathrm{MSD}_c(n)$ is always less than either $\mathrm{MSD}_1(n)$ or $\mathrm{MSD}_2(n)$ as expected from an optimal affine combiner. The solid curve is the best performance that could be obtained using two LMS adaptive filters in the system in Fig. 1 (because $\boldsymbol{W}_0$ is assumed known). Figure 3 displays a smooth transition for $\widetilde{\lambda}_0(n)$ from slightly larger than one to a small negative number as filter two takes over from filter one. The steady-state value of $E[\lambda_0(n)]$ is equal to -0.056, which matches the theoretical prediction of Eq. (15). The latter implies that the estimate of $\boldsymbol{W}_0$ obtained using filter one should be subtracted from the estimate of $\boldsymbol{W}_0$ obtained using filter two. The reason for this is that the estimates from the two filters are correlated [8].

The theoretical behavior of the MSD for the optimal affine combiner is determined next. Using Eq. (17), $\mathrm{MSD}_c(n)$ can be written

$$\mathrm{MSD}_c(n) = \mathrm{MSD}_2(n) - \frac{E\left[\boldsymbol{W}_{02}^T(n)\boldsymbol{W}_{12}(n)\boldsymbol{W}_{12}^T(n)\boldsymbol{W}_{02}(n)\right]}{E\left[\boldsymbol{W}_{12}^T(n)\boldsymbol{W}_{12}(n)\right]}. \tag{18}$$

It is shown in [8] that

$$\mathrm{MSD}_c(n) = \frac{\mathrm{MSD}_2(n)\dfrac{\mu_1\sigma_0^2(N-1)}{2-\mu_1(N+2)}}{\mathrm{MSD}_2(n) + \dfrac{\mu_1\sigma_0^2 N}{2-\mu_1(N+2)}}. \tag{19}$$

Equation (19) agrees with physical intuition. The right hand side of (19) is always less than either $\mathrm{MSD}_1(n)$ or $\mathrm{MSD}_2(n)$. Figure 4 shows the theory as compared to the MC simulations shown in Fig. 2. Excellent agreement can be observed. It can be shown [8] that the reduction in steady-state MSD as compared to that of the convex combiner [3] is a monotonic increasing function of μ_2/μ_1 and can be as large as 2dB as μ_2/μ_1 approaches unity.

4. AN ITERATIVE ALGORITHM

The previous derivation of the optimal linear combiner was based upon prior knowledge of the unknown optimal system response $\boldsymbol{W}_0$. Clearly, this is not the case in reality. However, the theoretical model and its derived properties can be used to upper bound the performance of practical algorithms for adjusting $\lambda(n)$ without such knowledge. Algorithms that yield close-to-optimal performance for typical unknown responses can be considered as good candidates for practical applications. This section studies an algorithm for the adjustment of $\lambda(n)$ which is based on the ratio of the average error powers from each individual adaptive filter. Its performance is then compared to the optimal performance. The performance of other algorithms applicable to the system in Fig. 1, such as the algorithm studied in [3], can also be compared with the optimum performance.

A function of the time averaged error powers could be a good candidate for an estimator of the optimum $\lambda(n)$ for each n. The individual adaptive error powers are good indicators of the contribution of each adaptive output to the quality of the present estimation of $d(n)$. These errors are readily available. Consider a uniform sliding time average of the instantaneous errors

$$\hat{e}_1^2(n) = \frac{1}{K}\sum_{m=n-K+1}^{n} e_1^2(m), \tag{20}$$

$$\hat{e}_2^2(n) = \frac{1}{K}\sum_{m=n-K+1}^{n} e_2^2(m), \tag{21}$$

where K is the averaging window. Then, consider the instantaneous values of $\lambda(n)$ given by

$$\widehat{\lambda}(n) = 1 - \kappa\ \mathrm{erf}\left(\frac{\hat{e}_1^2(n)}{\hat{e}_2^2(n)}\right), \tag{22}$$

where

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2/2}\,dt. \tag{23}$$

Equation (22) allows $\widehat{\lambda}(n)$ to vary smoothly over $(1-\kappa, 1)$. A proper choice of κ can lead to close-to-optimal performance [8]. Figure 5 compares $10\log_{10}(\mathrm{MSD}_c(n))$ obtained from 50 MC simulations using both (13) and (22) for updating n and the theory for the optimal combiner using (19). For this example, $\kappa = 1.125$ has been used so that $E[\widehat{\lambda}(n)] \approx E[\lambda_0(n)]$ in steady-state [8]. Excellent agreement can be seen for this suboptimal combining scheme. Figure 6 compares the average values of $\widehat{\lambda}(n)$ and $\lambda_0(n)$ over the 50 MC simulations. There is a reasonably good match between the two curves.

5. CONCLUSIONS

This paper studied the statistical behavior of an affine combination of the outputs of two LMS adaptive filters that simultaneously adapt using the same input. The purpose was to obtain a fast convergence and a small steady-state MSE. The scheme studied generalizes the convex combination so that the combination factor $\lambda(n)$ is not restricted to the interval $(0, 1)$. A sequence of unconstrained optimal combining coefficients (minimizing the MSE) was determined. The optimal unrealizable combiner was studied and provided the best possible performance. Then, a new scheme was proposed for practical applications. The scheme depended upon the time-averaged instantaneous squared error of each adaptive filter. This new scheme was designed using design information from the optimal combiner. Its performance was very close to the optimum. Monte Carlo simulations were in close agreement with the theoretical predictions.

6. REFERENCES

[1] T. Aboulnasr and K. Mayyas, "A robust variable step-size LMS-type algorithm: Analysis and simulations," *IEEE Trans. Signal Process.*, vol. 45, pp. 631–639, March 1997.

[2] H. C. Shin, A. H. Sayed, and W. J. Song, "Variable step-size NLMS and affine projection algorithms," *IEEE Signal Process. Lett.*, vol. 11, pp. 132–135, Feb. 2004.

[3] J. Arenas-Garcia, A. R. Figueiras-Vidal, and A. H. Sayed, "Mean-Square performance of a convex combination of two adaptive filters," *IEEE Trans. Signal Process.*, vol. 54, pp. 1078–1090, March 2006.

[4] S. Haykin, *Adaptive Filter Theory*. New Jersey: Prentice-Hall, 4th ed., 2002.

[5] N. J. Bershad and A. Bist, "Fast coupled adaptation for sparse impulse responses," *IEEE Trans. Signal Process.*, vol. 53, pp. 966–976, March 2005.

[6] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New York: Wiley-Interscience, 2003.

[7] S. J. M. Almeida, J. C. M. Bermudez, and N. J. Bershad, “A statistical analysis of the affine projection algorithm for unity step size and autoregressive inputs,” *IEEE Transactions on Circuits and Systems I - Fundamental Theory and Applications*, vol. 52, pp. 1394–1405, July 2005.

[8] N. J. Bershad, J. C. M. Bermudez, and J.-Y. Tourneret, “An affine combination of two LMS adaptive filters-Transient mean-square analysis,” *to appear in the IEEE Trans. Signal Process.*, 2008.

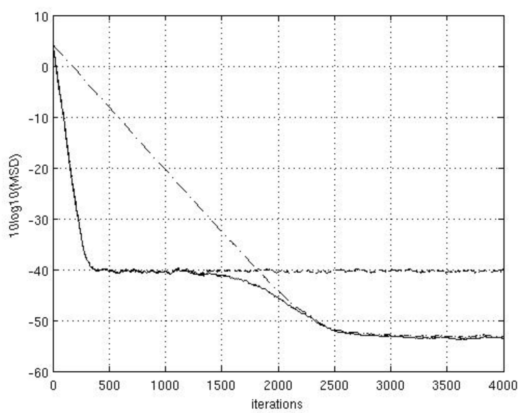

Fig. 2. $10\log_{10}(\mathrm{MSD}(n))$ versus n for $\boldsymbol{W}_1(n)$ (long dashes), $\boldsymbol{W}_2(n)$ (short dashes) and combined filter (solid line) (50 MC's and $\sigma_0^2 = 10^{-4}$).

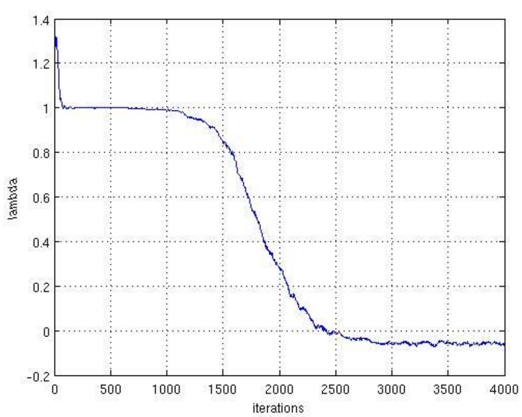

Fig. 3. $\widehat{\lambda}(n)$ versus n (50 MC's and $\sigma_0^2 = 10^{-4}$). Note that $\widehat{\lambda}(4000) = -.0573 < 0$.

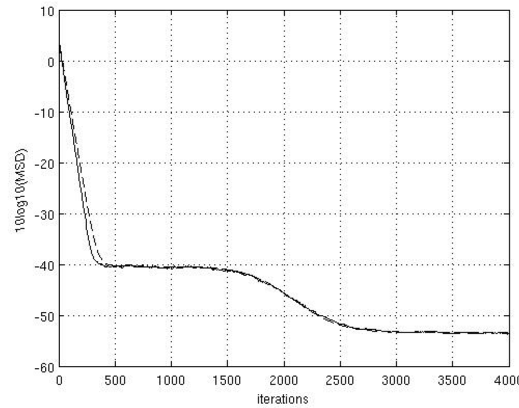

Fig. 4. $10\log_{10}(\mathrm{MSD}_c(n))$ for $\sigma_0^2 = 10^{-4}$, Theory (dashed) and Simulations (100 MC's, solid).

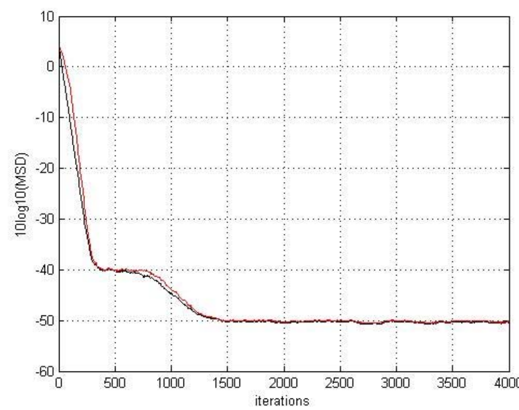

Fig. 5. $10\log_{10}(\mathrm{MSD}_c(n))$ for $\sigma_0^2 = 10^{-4}$, Theory (black) and Simulations (50 MC's, top curve, red).

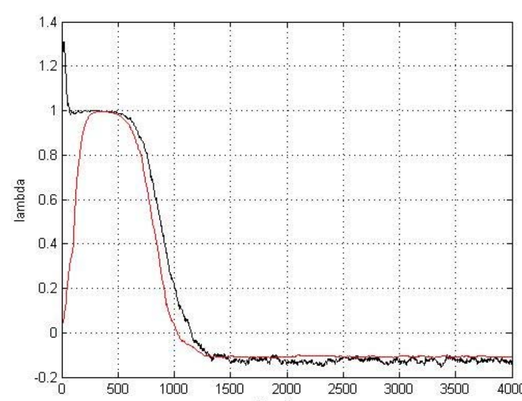

Fig. 6. $\lambda(n)$ (Theory, black) and $\widehat{\lambda}(n)$ (50 MC's, red) for $\sigma_0^2 = 10^{-4}$.