A Statistical Model for the Steady-State Behavior of the Multi-Split LMS Algorithm

L. S. RESENDE[1], R. D. SOUZA[1] and M. G. BELLANGER[2]

[1]Federal University of Santa Catarina, P. Box 476, 88040-900 Florianópolis-SC – BRAZIL, {leonardo,richard}@eel.ufsc.br

[2]Conservatoire National des Arts et Métiers, 292 rue Saint-Martin, 75141 Paris Cedex 03 – FRANCE, bellang@cnam.fr

Résumé – Ce travail presente un modèle de comportement en régime permanant de l'algorithme «multi-split LMS». Des expressions de récurrences sont obtenues pour le vecteur de coefficients moyens et pour l'erreur carré moyenne. Des résultats de simulation démontrent une conformité excellent avec les prédictions théoriques et nous permettent de valider le modèle proposé.

Abstract *– This paper presents a statistical model for the steady-state behavior of the multi-split LMS algorithm. Deterministic recursions are obtained for the mean weight vector and the mean square error. Simulation results display excellent agreement with the theoretical predictions and enable us to validate the proposed model.*

1. Introduction

The multi-split processing technique has been used in adaptive systems for improving the convergence behavior of the LMS algorithm [1-3]. It consists of a continued splitting process of the filter impulse response in symmetric and antisymmetric parts. The filter is then realized as a set of zero-order filters connected in parallel, and with each single coefficient independently updated. Such a technique can be viewed as a transform domain filter, in which multi-split preprocessing is applied to the input data vector.

An advantage of the multi-split transform is its ease of implementation. The computational burden is proportional to the number of filter coefficients N, and when N is equal to a power of two, the multi-split transform can be obtained by a butterfly computation scheme with no multiplication operation [1,4].

Recently, an analysis that justifies the improved performance of the multi-split LMS algorithm has been proposed in [4,5]. It is based on the fact that multi-split transform does not reduce the eigenvalue spread, but it does improve the diagonalization factor of the input signal correlation matrix, which is exploited by a power-normalized, time-varying step-size LMS algorithm for updating the filter coefficients in adaptive systems.

Our purpose in this paper is to start with a statistical analysis of the multi-split LMS algorithm and to present a model for the mean weight vector and the mean square error steady-state behavior. Deterministic recursions that predict such steady-state behaviors are derived, and their convergences towards the mean weight vector and the minimum mean-square error of the optimum filter are investigated.

The paper is organized as follows. In Section 2 we briefly describe the multi-split Wiener filtering and present the multi-split LMS algorithm for adaptive systems. Section 3 is dedicated to the statistical analysis of such an algorithm, developing a model for the mean weight vector and the mean square error steady-state behavior. In Section 4 we present simulation results that validate our analysis. Finally, in Section 5 we draw our conclusions.

2. Multi-split transversal filtering

2.1 Optimum multi-split Wiener filter

Consider initially the classical scheme of a nonadaptive transversal filter (Figure 1), where $\mathbf{w}$ denotes the N-by-1 tap-weight vector and

$$\mathbf{x}(n)=[x(n), x(n-1), \ldots, x(n-N+1)]^t \quad (1)$$

the tap-input vector. The input signal $x(n)$ and the desired response $d(n)$ are modeled as wide-sense stationary discrete-time stochastic processes of zero mean, Gaussian, with variance σ_x^2 and σ_d^2, respectively. The optimum weight vector $\mathbf{w}_{opt}$, called the Wiener vector, is given by [6-8]

$$\mathbf{w}_{opt}=\mathbf{R}^{-1}\mathbf{p}, \quad (2)$$

where $\mathbf{R}$ is the N-by-N correlation matrix of $\mathbf{x}(n)$, and $\mathbf{p}$ is the N-by-1 cross-correlation vector between $\mathbf{x}(n)$ and $d(n)$.

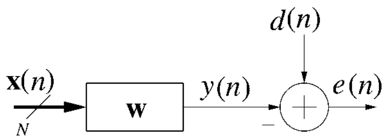


Figure 1: Transversal filtering.

For ease of presentation, let $N=2^L$, where L is an integer number greater than one. Without loss of generality, also consider that all the parameters are real-valued.

It has been shown in [3,4] that the multi-split filtering problem can be formulated and solved by using a linearly-constrained optimization, and can be implemented by means of a parallel GSC structure. The resulting multi-split filtering scheme can be represented by the block diagram in Figure 2, where

$$\mathbf{M}_N = \left[\begin{array}{c|c} \mathbf{M}_{N/2} & \mathbf{J}_{N/2}\mathbf{M}_{N/2} \\ \hline \mathbf{J}_{N/2}\mathbf{M}_{N/2} & -\mathbf{M}_{N/2} \end{array}\right], \quad (3)$$

$\mathbf{M}_1=[1]$ and $w_{\perp i}$, for $i=0, 1, \ldots, N-1$, are the single coefficients of the zero-order filters. It can be verified that $\mathbf{M}$ is a matrix of +1's and –1's, in which the inner product of any two distinct columns is zero. In fact, $\mathbf{M}$ is a nonsingular matrix and $\mathbf{M}^t\mathbf{M}=2^L\mathbf{I}$.

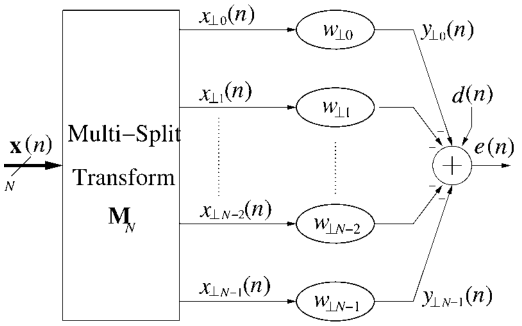


Figure 2: Multi-split transform of the input $\mathbf{x}(n)$.

The estimation error is then given by

$$e(n)=d(n)-\mathbf{w}_\perp^t\mathbf{x}_\perp(n), \quad (4)$$

where

$$\mathbf{w}_\perp=[w_{\perp 0}, w_{\perp 1}, \ldots, w_{\perp N-1}]^t \quad (5)$$

and

$$\mathbf{x}_\perp(n)=\mathbf{M}^t\mathbf{x}(n)=[x_{\perp 0}(n), x_{\perp 1}(n), \ldots, x_{\perp N-1}(n)]^t. \quad (6)$$

In the mean-squared-error sense, $\mathbf{w}_\perp$ is chosen to minimize the following cost function:

$$\xi(\mathbf{w}_\perp)=E\{e^2(n)\}=\sigma_d^2-2\mathbf{w}_\perp^t\mathbf{M}^t\mathbf{p}+\mathbf{w}_\perp^t\mathbf{M}^t\mathbf{RM}\mathbf{w}_\perp. \quad (7)$$

The optimum solution is given by

$$\mathbf{w}_{\perp opt}=[\mathbf{M}^t\mathbf{RM}]^{-1}\mathbf{M}^t\mathbf{p}=\mathbf{M}^{-1}\mathbf{R}^{-1}\mathbf{p}=(1/2^L)\mathbf{M}^t\mathbf{w}_{opt}, \quad (8)$$

and the scheme of Figure 2 corresponds to the optimum multi-split Wiener filter:

$$\mathbf{w}_{opt}=\mathbf{M}\mathbf{w}_{\perp opt}. \quad (9)$$

Substituting (8) in (7), the minimum mean-square error is found to be

$$\xi_{min}=\sigma_d^2-\mathbf{p}^t\mathbf{R}^{-1}\mathbf{p}=\mathbf{p}^t\mathbf{w}_{opt}=\mathbf{p}^t\mathbf{M}\mathbf{w}_{\perp opt}, \quad (10)$$

which is, therefore, equal to the minimum mean-square error of the optimum Wiener filter.

2.2 Adaptive multi-split filtering

It has been shown that the multi-split transform is not an input whitening transformation. Instead, it increases the diagonalization factor of the input signal correlation matrix without affecting its eigenvalue spread [4,5].

In the adaptive context, a power-normalized, time-varying step-size LMS algorithm, which exploits the nature of the transformed input correlation matrix, has been proposed for updating the single coefficients independently [3,4]. Table I presents a summary of the multi-split LMS algorithm.

3. Steady-state algorithm behavior

This section studies the limiting behavior of the converged multi-split LMS algorithm. To that end, let us assume that the variance estimates of $x_{\perp i}(n)$, for i=0, 1, …, N–1, have converged. Consequently, the step-sizes $\mu/r_i(n)$ are considered fixed. However, they can still be distinct to each other, which is one of the bases of the multi-split LMS algorithm.

3.1 Mean weight behavior

Taking into account the aforementioned assumption, the weight update equation for the multi-split LMS algorithm can then be rewritten as follows:

$$\mathbf{w}_\perp(n)=\mathbf{w}_\perp(n-1)+\mu\boldsymbol{\Sigma}^{-1}\mathbf{x}_\perp(n)e(n), \quad (11)$$

Table I: Multi-Split LMS (MS-LMS) algorithm

➢ <u>*Selection of parameters*</u>:

adaptation step-size: $0<\mu<\dfrac{1}{2^{2L}\sigma_x^2}$

forgetting factor: $0<<\gamma\leq 1$

➢ <u>*Initialization*</u>:

for i=0,1, …, N–1, set: $w_{\perp i}(0)=0$ and $r_i(0)=0$

set $\mathbf{M}_N$ as such (3)

➢ <u>*Updating*</u>:

for i=0,1, …, N–1 and n=1,2, …, compute:

1) *Linear transform*: $x_{\perp i}(n)=\mathbf{m}_i^t\mathbf{x}(n)$, where $\mathbf{m}_i$ is the $(i+1)^{\text{th}}$ column vector of $\mathbf{M}_N$

2) *Normalized LMS algorithm*:

$$y(n)=\sum_{i=0}^{N-1}x_{\perp i}(n)w_{\perp i}(n-1)$$

$$e(n)=d(n)-y(n)$$

$$r_i(n)=\gamma\, r_i(n-1)+\frac{1}{n}(x_{\perp i}^2(n)-\gamma\, r_i(n-1))$$

$$w_{\perp i}(n)=w_{\perp i}(n-1)+\frac{\mu}{r_i(n)}x_{\perp i}(n)e(n)$$

where

$$\Sigma=diag[\sigma_{x\perp 0}^2, \sigma_{x\perp 1}^2, \ldots, \sigma_{x\perp N-1}^2] \quad (12)$$

and

$$e(n)=d(n)-\mathbf{w}_\perp^t(n-1)\mathbf{x}_\perp(n), \quad (13)$$

The expected value of (11) leads to the recursion:

$$\begin{aligned}E\{\mathbf{w}_\perp(n)\}=&E\{\mathbf{w}_\perp(n-1)\}+\\&+\mu\boldsymbol{\Sigma}^{-1}E\{\mathbf{x}_\perp(n)d(n)\}+\\&-\mu\boldsymbol{\Sigma}^{-1}E\{\mathbf{x}_\perp(n)\mathbf{x}_\perp^t(n)\mathbf{w}_\perp(n-1)\}\\=&[\mathbf{I}-\mu\boldsymbol{\Sigma}^{-1}\mathbf{M}^t\mathbf{RM}]E\{\mathbf{w}_\perp(n-1)\}+\mu\boldsymbol{\Sigma}^{-1}\mathbf{M}^t\mathbf{p}, \quad (14)\end{aligned}$$

where, for sufficiently small μ and assuming that $\mathbf{x}(n)$ and $\mathbf{w}_\perp(n-1)$ are statistically independent, the following approximation has been used:

$$\begin{aligned}E\{\mathbf{x}_\perp(n)\mathbf{x}_\perp^t(n)\mathbf{w}_\perp(n-1)\}&\approx E\{\mathbf{x}_\perp(n)\mathbf{x}_\perp^t(n)\}E\{\mathbf{w}_\perp(n-1)\}\\&=\mathbf{M}^t\mathbf{RM}\,E\{\mathbf{w}_\perp(n-1)\}. \quad (15)\end{aligned}$$

Equation (14) is a deterministic recursion for the mean weight convergence of the multi-split LMS algorithm, in which its transient behavior with time-varying step-size has not been taken into account.

3.1.1 Steady-state mean weight

For the steady-state analysis, it is assumed that the algorithm converges as $n\to\infty$ and

$$\lim_{n\to\infty}\mathbf{w}(n)=\lim_{n\to\infty}E\{\mathbf{w}(n)\}=\mathbf{w}_\infty. \quad (16)$$

Replacing $\mathbf{w}(n)$ with $\mathbf{w}_\infty$ in (14) yields

$$\mathbf{w}_\infty=[\mathbf{M}^t\mathbf{RM}]^{-1}\mathbf{M}^t\mathbf{p}, \quad (17)$$

which corresponds to the optimum solution in (8).

3.2 Mean square error behavior

Squaring (13) and taking the expected value yields

$$\xi(n)=E\{e^2(n)\}=\sigma_d^2-2\mathbf{p}^t\mathbf{M}E\{\mathbf{w}_\perp(n-1)\}+ +tr[\mathbf{M}^t\mathbf{RM}E\{\mathbf{w}_\perp(n-1)\mathbf{w}_\perp^t(n-1)\}], \quad (18)$$

where the following approximation have been used:

$$E\{d(n)\mathbf{x}_\perp^t(n)\mathbf{w}_\perp(n-1)\}\approx E\{d(n)\mathbf{x}_\perp^t(n)\}E\{\mathbf{w}_\perp(n-1)\}, \quad (19)$$

$$E\{\mathbf{w}_\perp^t(n-1)\mathbf{x}_\perp(n)\mathbf{x}_\perp^t(n)\mathbf{w}_\perp(n-1)\}= tr[E\{\mathbf{x}_\perp(n)\mathbf{x}_\perp^t(n)\mathbf{w}_\perp(n-1)\mathbf{w}_\perp^t(n-1)\}] \approx tr[E\{\mathbf{x}_\perp(n)\mathbf{x}_\perp^t(n)\}E\{\mathbf{w}_\perp(n-1)\mathbf{w}_\perp^t(n-1)\}], \quad (20)$$

and $tr[\bullet]$ stands for the trace of the matrix.

3.2.1 Weight correlation matrix

Evaluation of (18) requires $E\{\mathbf{w}_\perp(n-1)\mathbf{w}_\perp^t(n-1)\}$. Post-multiplying (11) by its transpose and taking the expected value yields

$$\mathbf{K}(n)=E\{\mathbf{w}_\perp(n)\mathbf{w}_\perp^t(n)\}$$
$$=\mathbf{K}(n-1)+\mu\Sigma^{-1}\mathbf{M}^t\mathbf{p}E\{\mathbf{w}_\perp^t(n-1)\}-\mu\Sigma^{-1}\mathbf{M}^t\mathbf{RMK}(n-1)+$$
$$+\mu E\{\mathbf{w}_\perp(n-1)\}\mathbf{p}^t\mathbf{M}\Sigma^{-1}-\mu\mathbf{K}(n-1)\mathbf{M}^t\mathbf{RM}\Sigma^{-1}+$$
$$+\mu^2\Sigma^{-1}E\{\mathbf{x}_\perp(n)d^2(n)\mathbf{x}_\perp^t(n)\}\Sigma^{-1}+$$
$$-2\mu^2\Sigma^{-1}E\{\mathbf{x}_\perp(n)d(n)\mathbf{x}_\perp^t(n)\mathbf{w}_\perp(n-1)\mathbf{x}_\perp^t(n)\}\Sigma^{-1}+$$
$$+\mu^2\Sigma^{-1}E\{\mathbf{x}_\perp(n)\mathbf{w}_\perp^t(n-1)\mathbf{x}_\perp(n)\mathbf{x}_\perp^t(n)\mathbf{w}_\perp(n-1)\mathbf{x}_\perp^t(n)\}\Sigma^{-1}, \quad (21)$$

where the same approximation in (19) and (20) have been used. Now, since $x(n)$ and $d(n)$ are Gaussian, the last three expected values in (21) can be evaluated using the moment factoring theorem [8,9]. It can be shown that:

$$E\{\mathbf{x}_\perp(n)d^2(n)\mathbf{x}_\perp^t(n)\}=2\mathbf{M}^t\mathbf{pp}^t\mathbf{M}+\sigma_d^2\mathbf{M}^t\mathbf{RM}, \quad (22)$$

$$E\{\mathbf{x}_\perp(n)d(n)\mathbf{x}_\perp^t(n)\mathbf{w}_\perp(n-1)\mathbf{x}_\perp^t(n)\}\approx \mathbf{M}^t\mathbf{p}E\{\mathbf{w}_\perp^t(n-1)\}\mathbf{M}^t\mathbf{RM}+\mathbf{M}^t\mathbf{RM}E\{\mathbf{w}_\perp(n-1)\}\mathbf{p}^t\mathbf{M}+ +\mathbf{p}^t\mathbf{M}E\{\mathbf{w}_\perp(n-1)\}\mathbf{M}^t\mathbf{RM} \quad (23)$$

and

$$E\{\mathbf{x}_\perp(n)\mathbf{w}_\perp^t(n-1)\mathbf{x}_\perp(n)\mathbf{x}_\perp^t(n)\mathbf{w}_\perp(n-1)\mathbf{x}_\perp^t(n)\}\approx E\{\mathbf{w}_\perp^t(n-1)\}\mathbf{M}^t\mathbf{RM}E\{\mathbf{w}_\perp(n-1)\}\mathbf{M}^t\mathbf{RM}+ +2\mathbf{M}^t\mathbf{RMK}(n-1)\mathbf{M}^t\mathbf{RM}. \quad (24)$$

Thus, (21) is a recursion for the weight correlation matrix and (18) for the mean square error behavior. As in (14), it is also worth pointing out that these deterministic expressions do not take into account the time-varying step-size. Consequently, they cannot predict accurately the transient behavior of the multi-split algorithm.

3.2.2 Steady-state MSE

An expression for the steady-state MSE behavior is determined by replacing $\mathbf{w}_\perp(n-1)$ with the steady-state mean weight vector expression (17) in (18). It is given by

$$\lim_{n\to\infty}\xi(n)= \sigma_d^2-\mathbf{p}^t\mathbf{R}^{-1}\mathbf{p}=\xi_{min}, \quad (25)$$

which corresponds to the minimum mean-square error in (10), as expected.

4. Simulation results

In order to validate the proposed analysis, we consider the same equalization system in [8, chap.5] (Figure 3). The input channel is binary, with $b(n)=\pm1$, and the impulse response of the channel is described by the raised cosine:

$$c_j=\begin{cases}\frac{1}{2}(1+\cos(\frac{2\pi}{S}(j-2))), & j=1,2,3\\ 0, & \text{otherwise}\end{cases}, \quad (26)$$

where S controls the eigenvalue spread $\chi(\mathbf{R})$ of the correlation matrix of the tap inputs in the equalizer, with $\chi(\mathbf{R})$=6.0782 for S= 2.9 and $\chi(\mathbf{R})$=46.8216 for S= 3.5. The sequence $v(n)$ is an additive white noise that corrupts the channel output with variance σ_v^2=0.001, and the equalizer has eleven coefficients.

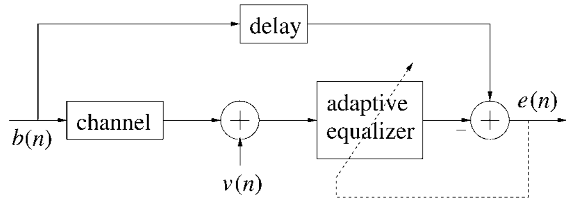


Figure 3: Adaptive equalizer for simulation.

Figure 4 compares the simulated (100 independent trials) mean weight behavior with the analytical model using (14). The comparison between the simulated mean-square error behavior and the analytical model using (18) and (21) is shown in Figure 5. It can be verified that the proposed statistical analysis predicts with good accuracy the steady-state behavior of the multi-split LMS algorithm, whereas the transient behavior has shown a lower convergence rate. It is due to the fact that the step-sizes have been considered fixed. The algorithm parameters were μ=0.0455 and γ=1.

In fact, a statistical model that enables us to predict the transient behavior of the multi-split LMS algorithm must take into account the time-varying step-size aspect, which necessarily makes the analysis more complex.

5. Conclusion

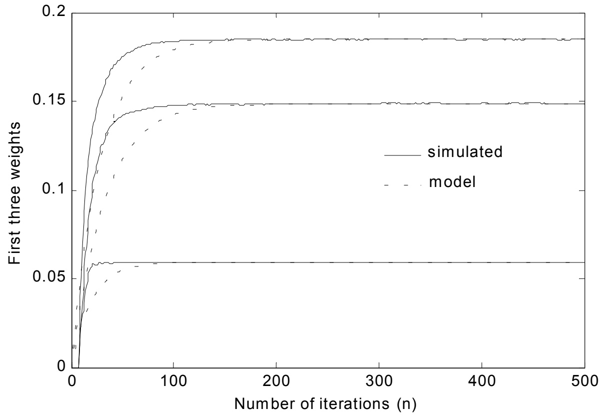
This paper has presented a statistical analysis for the steady-state behavior of the multi-split LMS algorithm. Deterministic recursions for the mean weight vector and the mean square error have been derived. The convergence of such recursions towards the mean weight vector and the minimum mean-square error of the optimum filter has been analytically demonstrated and confirmed by simulations. Even though the analysis carried out in this paper is based on some of the results in [4,5], its outcomes validate the assumptions considered therein and contribute to the understanding of multi-split filtering. This kind of analysis is also useful for adaptive algorithm design and evaluation.

Studies concerning to the development of a model for the transient behavior of the multi-split LMS algorithm are in progress.

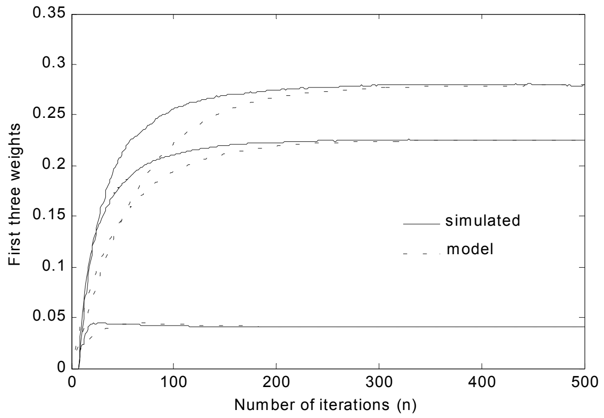
References

[1] K.F. Wan and P.C. Ching, *Multilevel split-path adaptive filtering and its unification with discrete Walsh transform adaptation*. IEEE Transactions on Circuits and Systems-Part II; vol. 44; no. 2; pp. 147-151; February 1992.

[2] R. D. Souza, L. S. Resende, C. A. F. da Rocha and M. G. Bellanger, *On split FIR filtering in blind equalization*. Proceedings of the IEEE ICC'02; New York; USA; April 2002.
[3] L. S. Resende, J. M. T. Romano and M. G. Bellanger, *Multi-split adaptive filtering*. Proceedings of the IEEE ICASSP'01; Salt Lake City; USA; May 2001.
[4] L. S. Resende, J. M. T. Romano and M. G. Bellanger, *Split Wiener filtering with application in adaptive systems*. Accepted for publication in IEEE Transactions on Signal Processing.
[5] R. D. Souza, L. S. Resende and M. G. Bellanger, *A Justification for the improved performance of the multi-split LMS algorithm*. Proceedings of the IEEE ICASSP'03; Hong Kong; May 2003.
[6] B. Widrow and S.D. Stearns, *Adaptive signal processing*. New Jersey: Prentice-Hall, 1985.
[7] M.G. Bellanger, *Adaptive digital dilters and signal analysis*, 2nd edition. New York and Basel: Marcel Dekker, Inc., 2001.
[8] S. Haykin, Adaptive filter theory, 4th edition. New Jersey: Prentice-Hall, 2002.
[9] A. Papoulis, *Probability, randow variables and stochastic processes*, 3rd edition. New York: McGraw-Hill, 1991.

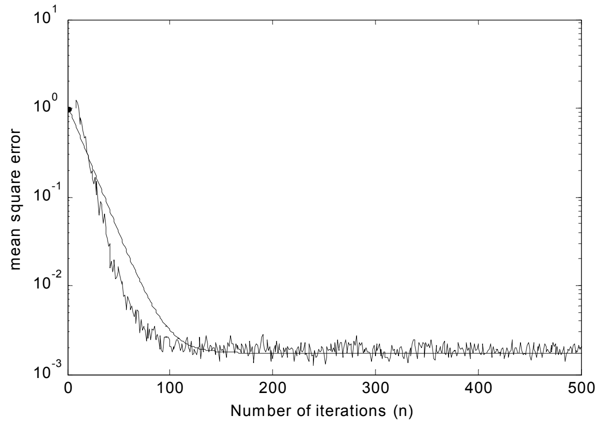


(a): $\chi(\mathbf{R})$=6.0782

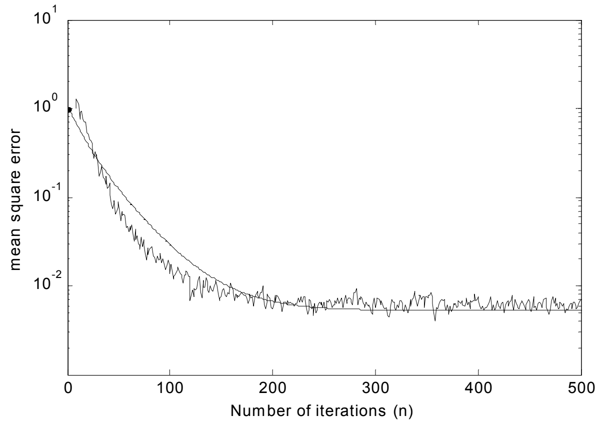


(b): $\chi(\mathbf{R})$=46.8216

Figure 4: Mean weight behavior.



(a): $\chi(\mathbf{R})$=6.0782



(b): $\chi(\mathbf{R})$=46.8216

Figure 5: Mean square error behavior.