A STATISTICAL ANALYSIS OF THE MULTI-SPLIT LMS ALGORITHM

L. S. Resende[1], C. A. F. Rocha[1], J. C. M. Bermudez[1] and M. G. Bellanger[2]

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

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

ABSTRACT

This paper presents a statistical analysis 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 models for both transient and steady-state behaviors.

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. However, an analytical model for such an algorithm has not yet been provided in the literature.

Our purpose in this paper is to introduce a statistical analysis of the multi-split LMS algorithm. Deterministic recursions that predict the transient and steady-state behaviors of the mean weight vector and mean square error are derived. Furthermore, their convergences towards the mean weight vector and the minimum mean square error of the optimum filter are also investigated. Finally, simulation results validate our analysis.

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

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

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]

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

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 $d(n)$.

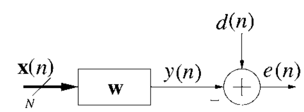


Figure 1: Transversal filtering.

For ease of presentation of the multi-split filtering scheme, 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

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

$\mathbf{J}_{N/2}$ is the $N/2$-by-$N/2$ exchange matrix, which has unit elements along the cross diagonal and zeros elsewhere, $\mathbf{M}_1=[1]$ and $w_{\perp i}$, for i=0, 1, …, $N-1$, 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 $\mathbf{M}^t\mathbf{M}=2^L\mathbf{I}$.

The estimation error is then given by

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

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:

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

The optimum solution is given by

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

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

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

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

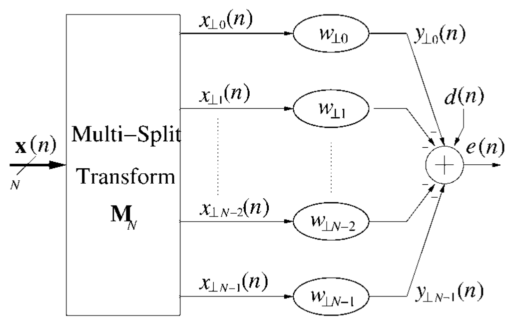


Figure 2: Multi-split transform of the input $\mathbf{x}(n)$.

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]:

$$w_{\perp i}(n)=w_{\perp i}(n-1)+\frac{\mu}{\sigma_i^2(n)}x_{\perp i}(n)e(n), \quad (11)$$

where

$$\sigma_i^2(n)=\gamma\,\sigma_i^2(n-1)+\frac{1}{n}(x_{\perp i}^2(n)-\gamma\,\sigma_i^2(n-1)), \quad (12)$$

μ is the adaptation step-size and γ is the forgetting factor ($0<<\gamma\leq 1$).

3. STATISTICAL ANALYSIS

3.1 TRANSIENT BEHAVIOR

A. Mean Weight Vector

The expected value of the weight adaptation equation leads to the recursion:

$$E\{w_{\perp i}(n)\}=E\{w_{\perp i}(n-1)\}+\mu E\{\frac{x_{\perp i}(n)e(n)}{\sigma_i^2(n)}\}. \quad (13)$$

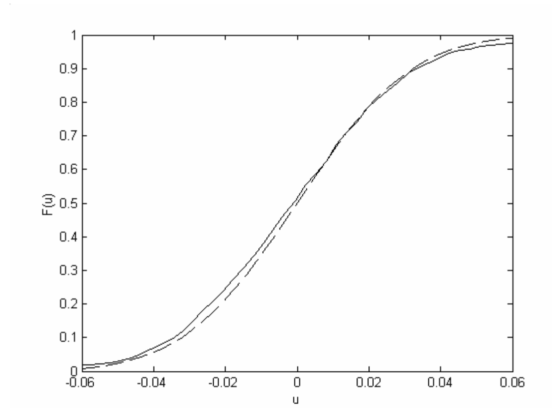
In order to evaluate the last expectation in (13), let us treat the argument as a ratio of two random variables:

$$\frac{x_{\perp i}(n)e(n)}{\sigma_i^2(n)}=\frac{u}{v}=z. \quad (14)$$

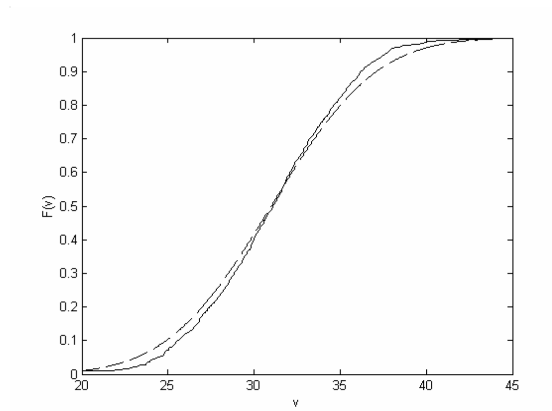
Thus, if we assume that u and v are jointly Gaussian with zero-mean, the expected value of z is given by [9]:

$$E\{z\}=\frac{E\{uv\}}{E\{v^2\}}=\frac{E\{uv\}}{\sigma_v^2}. \quad (15)$$

The assumption that $x_{\perp i}(n)e(n)$ and $\sigma_i^2(n)$ are jointly Gaussian can be validated through comparisons of their distribution functions with those of a Gaussian (hypothesis testing). For the equalization system considered in Section 4, such comparisons are shown in Figure 3, which have also been tested successfully by the Kolmogoroff-Smirnov test [9].



(a): random variable *u*



(b): random variable *v*

Figure 3: Comparison of the distribution functions of *u* and *v* in (14) with the Gaussian distribution function. The dashed lines correspond to Gaussian distributions with same mean values as *u* and *v*.

Although the mean value of v is not equal to zero, we proceed under the assumption that (15) holds.

Based on aforementioned assumptions, the following approximation is used for the last expectation in (13):

$$E\{\frac{x_{\perp i}(n)e(n)}{\sigma_i^2(n)}\} \cong \frac{E\{x_{\perp i}(n)e(n)\sigma_i^2(n)\}}{E\{[\sigma_i^2(n)]^2\}-[E\{\sigma_i^2(n)\}]^2}\,. \quad (16)$$

Proceeding:

$$E\{x_{\perp i}(n)e(n)\sigma_i^2(n)\} = \frac{1}{n}\sum_{l=1}^{n}\gamma^{n-l}E\{x_{\perp i}(n)e(n)x_{\perp i}^2(l)\}$$

$$=\frac{1-\gamma^n}{n(1-\gamma)}[E\{x_{\perp i}(n)d(n)x_{\perp i}^2(l)\}+$$

$$-\sum_{j=0}^{N-1}E\{x_{\perp i}(n)x_{\perp j}(n)x_{\perp i}^2(l)\}E\{w_{\perp j}(n-1)\}]\,, \quad (17)$$

where the effects of the statistical dependence between $x_{\perp i}(n)$ and $w_{\perp i}(n\text{-}1)$ have been neglected. Now, since $x_{\perp i}(n)$ and $d(n)$ are Gaussian, it can be shown, using the moment factoring theorem, that:

$$E\{x_{\perp i}(n)d(n)x_{\perp i}^2(l)\} = 3p_i\sigma_{x\perp i}^2 \quad (18)$$

and

$$E\{x_{\perp i}(n)x_{\perp j}(n)x_{\perp i}^2(l)\} = 3r_{ij}\sigma_{x\perp i}^2\,, \quad (19)$$

where p_i=$E\{x_{\perp i}(n)d(n)\}$ and r_{ij}= $E\{x_{\perp i}(n)\ x_{\perp j}(n)\}$. For the denominator in (16), we obtain:

$$E\{[\sigma_i^2(n)]^2\}-[E\{\sigma_i^2(n)\}]^2 = \frac{2}{n^2}(\frac{1-\gamma^n}{1-\gamma})^2(\sigma_{x\perp i}^2)^2\,. \quad (20)$$

Thus, we find the following deterministic recursion for the mean weight convergence:

$$E\{w_{\perp i}(n)\} = E\{w_{\perp i}(n-1)\}+$$

$$+\frac{3n(1-\gamma)\mu}{2(1-\gamma^n)\sigma_{x\perp i}^2}[p_i - \sum_{j=0}^{N-1} r_{ij}E\{w_{\perp j}(n-1)\}] \quad (21)$$

or

$$E\{\mathbf{w}_\perp(n)\}=[\mathbf{I}-\beta\mathbf{\Sigma}^{-2}\mathbf{M}^t\mathbf{RM}]E\{\mathbf{w}_\perp(n-1)\}+\beta\mathbf{\Sigma}^{-2}\mathbf{M}^t\mathbf{p}, \quad (22)$$

where

$$\beta = \frac{3n(1-\gamma)\mu}{2(1-\gamma^n)} = \frac{3\mu}{2}\,, \quad (23)$$

for γ=1, and

$$\mathbf{\Sigma}^2 = diag[\sigma_{x\perp 0}^2,\ \sigma_{x\perp 1}^2,\ \ldots,\ \sigma_{x\perp N-1}^2]. \quad (24)$$

B. Mean Square Error

Squaring the estimation error and taking the expected value yields:

$$\xi(n) = E\{e^2(n)\} = \sigma_d^2 - 2\sum_{i=0}^{N-1} p_i E\{w_{\perp i}(n-1)\} +$$

$$+\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} r_{ij}E\{w_{\perp i}(n-1)w_{\perp j}(n-1)\}$$

$$=\sigma_d^2 - 2\mathbf{p}^t\mathbf{M}E\{\mathbf{w}_\perp(n-1)\}+$$

$$+tr[\mathbf{M}^t\mathbf{RMK}(n-1)], \quad (25)$$

where $\mathbf{K}(n-1)=E\{\mathbf{w}_\perp(n-1)\mathbf{w}_\perp^t(n-1)\}$ is the tap-weight correlation matrix, and the effects of the statistical dependence between $d(n)$ and $w_{\perp i}(n\text{-}1)$ have also been neglected. Proceeding:

$$k_{ij}(n-1) = E\{w_{\perp i}(n-1)w_{\perp j}(n-1)\}$$

$$= k_{ij}(n-2) + \mu E\{w_{\perp i}(n-2)\}E\{\frac{x_{\perp j}(n-1)d(n-1)}{\sigma_j^2(n-1)}\} +$$

$$-\mu\sum_{m=0}^{N-1}E\{\frac{x_{\perp j}(n-1)x_{\perp m}(n-1)}{\sigma_j^2(n-1)}\}E\{w_{\perp i}(n-2)w_{\perp m}(n-2)\}+$$

$$+\mu E\{\frac{x_{\perp i}(n-1)d(n-1)}{\sigma_i^2(n-1)}\}E\{w_{\perp j}(n-2)\}+$$

$$-\mu\sum_{m=0}^{N-1}E\{\frac{x_{\perp i}(n-1)x_{\perp m}(n-1)}{\sigma_i^2(n-1)}\}E\{w_{\perp j}(n-2)w_{\perp m}(n-2)\}+$$

$$+\mu^2 E\{\frac{x_{\perp i}(n-1)x_{\perp j}(n-1)e^2(n-1)}{\sigma_i^2(n-1)\sigma_j^2(n-1)}\}\,. \quad (26)$$

The expected values in the terms in μ have been evaluated. For the term in μ^2, the following approximation is used:

$$E\{\frac{x_{\perp i}(n-1)x_{\perp j}(n-1)e^2(n-1)}{\sigma_i^2(n-1)\sigma_j^2(n-1)}\} \cong$$

$$\frac{E\{x_{\perp i}(n-1)x_{\perp j}(n-1)e^2(n-1)\}}{E\{\sigma_i^2(n-1)\sigma_j^2(n-1)\}}\,, \quad (27)$$

seeing that μ is sufficiently small and n is large. The final recursive equation for $\mathbf{K}(n-1)$ is given by:

$$\mathbf{K}(n-1)=\mathbf{K}(n-2)+$$
$$+2\beta\mathbf{\Sigma}^{-2}[\mathbf{M}^t\mathbf{p}E\{\mathbf{w}_\perp^t(n-2)\}-\mathbf{M}^t\mathbf{RMK}(n-2)]+$$
$$+(4/9)\beta^2\mathbf{\Sigma}^{-2}[2\mathbf{M}^t\mathbf{pp}^t\mathbf{M}+\sigma_d^2\mathbf{M}^t\mathbf{RM}]\mathbf{\Sigma}^{-2}+$$
$$-(8/9)\beta^2\mathbf{\Sigma}^{-2}[2\mathbf{M}^t\mathbf{p}E\{\mathbf{w}_\perp^t(n-2)\}\mathbf{M}^t\mathbf{RM}+$$
$$\mathbf{p}^t\mathbf{M}E\{\mathbf{w}_\perp(n-2)\}\mathbf{M}^t\mathbf{RM}]\mathbf{\Sigma}^{-2}+$$
$$+(4/9)\beta^2\mathbf{\Sigma}^{-2}[2\mathbf{M}^t\mathbf{RMK}(n-2)\mathbf{M}^t\mathbf{RM}+$$
$$E\{\mathbf{w}_\perp^t(n-2)\}\mathbf{M}^t\mathbf{RM}E\{\mathbf{w}_\perp(n-2)\}\mathbf{M}^t\mathbf{RM}]\mathbf{\Sigma}^{-2}. \quad (28)$$

3.2 Steady-State Behavior

A. Mean Weight Vector

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

$$\lim_{n\to\infty}\mathbf{w}(n)=\lim_{n\to\infty}E\{\mathbf{w}(n)\}=\mathbf{w}_\infty. \quad (29)$$

Replacing $\mathbf{w}(n)$ with $\mathbf{w}_\infty$ in (22) yields

$$\mathbf{w}_\infty=[\mathbf{M}^t\mathbf{RM}]^{-1}\mathbf{M}^t\mathbf{p}, \quad (30)$$

which corresponds to the optimum solution in (8).

B. Mean Square Error

An expression for the steady-state MSE behavior is determined by replacing $\mathbf{w}_\perp(n-1)$ with $\mathbf{w}_\infty$ in (25). It is given by

$$\lim_{n\to\infty}\xi(n)= \sigma_d^2-\mathbf{p}^t\mathbf{R}^{-1}\mathbf{p}=\xi_{min}, \tag{31}$$

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 4). The input channel is binary, with $b(n)=\pm1$, and the impulse response of the channel is described by the raised cosine:

$$c_j=\begin{cases}\frac{1}{2}\left(1+\cos\left(\frac{2\pi}{S}(j-2)\right)\right), & j=1,2,3\\ 0, & \text{otherwise}\end{cases}, \tag{32}$$

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.

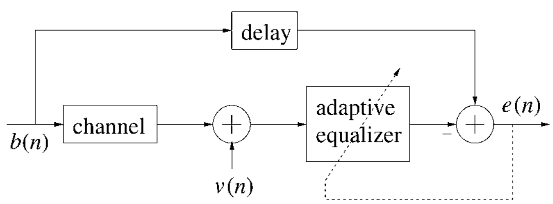


Figure 4: Adaptive equalizer for simulation.

Figure 5 compares the simulated (100 independent trials) mean square error behavior with the analytical model. It can be verified that the proposed statistical analysis predicts with good accuracy the transient and steady-state behaviors of the multi-split LMS algorithm. The algorithm parameters were μ=0.0455 and γ=1.

5. SUMMARY

This paper has presented a statistical analysis for the transient and steady-state behaviors of the multi-split LMS algorithm. Deterministic recursions have been derived for the mean weight vector and the mean square error. 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. This kind of analysis is useful for adaptive algorithm design and evaluation.

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*. To appear in IEEE Transactions on Signal Processing; vol. 52; no. 3; March 2004.

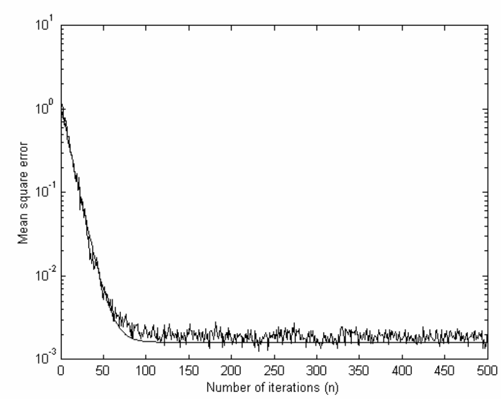
[5] R. D. Souza, L. S. Resende and M. G. Bellanger, *A Justification for the improved performance of the multi-split LMS algorithm*. Proceedings of the IEEE ICASSP'03; Hong Kong; May 2003.

[6] B. Widrow and S.D. Stearns, *Adaptive signal processing*. New Jersey: Prentice-Hall, 1985.

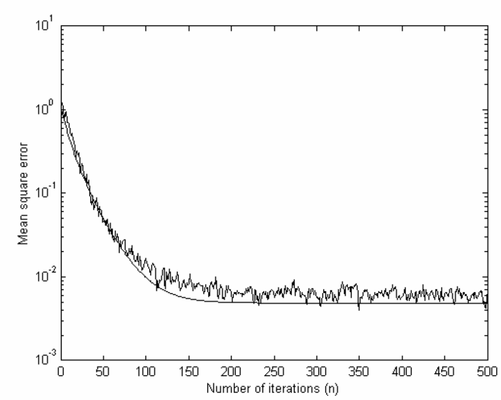
[7] M.G. Bellanger, *Adaptive digital dilters and signal analysis*, 2nd edition. New York and Basel: Marcel Dekker, Inc., 2001.

[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 5: Mean square error behavior.