TECHNICAL RESEARCH REPORT

Static and Dynamic Convergence Behavior of Adaptive Blind Equalizers

by Y. Li and K.J.R. Liu

T.R. 95-62

INSTITUTE FOR SYSTEMS RESEARCH

Sponsored by
the National Science Foundation
Engineering Research Center Program,
the University of Maryland,
Harvard University,
and Industry

Static and Dynamic Convergence Behavior of Adaptive Blind Equalizers*

Ye Li and K. J. Ray Liu
Department of Electrical Engineering and Institute for System Research
University of Maryland at College Park
College Park, MD 20742

June 26, 1995

Abstract: *This paper presents a theoretical analysis of the static and dynamic convergence behavior for a general class of adaptive blind equalizers. We first study the properties of prediction error functions of blind equalization algorithms, and then we use these properties to analyze the static and dynamic convergence behavior based on the independent assumption. We prove in this paper that with a small step-size, the ensemble average of equalizer coefficients will converge to the minimum of the cost function near the channel inverse. However, the convergence is not consistent. The correlation matrix of equalizer coefficients at equilibrium is determined by a Lyapunov equation. According to our analysis results, for a given channel and step-size, there is an optimal length for an equalizer to minimize the intersymbol interference. This result implies that a longer-length blind equalizer does not necessarily outperform a shorter one, as contrary to what conventionally conjectured. The theoretical analysis results are confirmed by computer simulations.*

SP EDICS: SP 2.6.4 and SP 2.6.2.

*The work was supported in part by the NSF grants MIP9309506 and MIP9457397.

1 Introduction

Since the pineering work by Sato[18], lots of blind channel equalization algorithms have been proposed[1, 2, 17, 21, 23, 24]. They have been effectively used in digital communication systems to cancel the inter-symbol interference (ISI). Blind equalization algorithms are usually designed to minimize some cost functions consisting of higher-order statistics of the channel output, without using the channel input. They are implemented mostly by stochastic gradient algorithms. The convergence analysis of blind equalization algorithms is very important to understand their performance. We may categorize the convergence analyses into two different kinds, *static convergence analysis* and *dynamic convergence analysis.*

The static convergence analysis studies the positions of the minimum points of the cost functions under various conditions. It has been proved that undesirable local minima may exist for Godard algorithms [8] implemented with FIR equalizers [4, 5, 11], and for BGR algorithms [1] and decision-directed algorithms [17] even if implemented with IIR equalizers[6, 10, 14, 16]. Recently, we have found that almost all cost functions of blind equalization algorithms may have undesirable local minima[13] due to the finite length of equalizers. Although undesirable local minima exist for blind equalization algorithms, they may be effectively avoided by smart initialization strategies[7, 11].

On the other hand, the dynamic convergence analysis addresses the stochastic dynamics of equalization algorithms. Because of the non-linearity in adaptive blind equalization algorithms, the exact dynamic convergence analysis is often very difficult. Almost all dynamic convergence analyses are conducted under some assumptions. Several papers [9, 10, 14, 16, 27] have studied the dynamic convergence of the decision-directed equalizer by assuming that equalizer is in "open eye pattern". In [26], Weerackody and his coauthers have presented dynamic convergence analysis of Sato equalizer. Chan and Shynk [19] have studied the dynamic convergence of the constant modulus algorithm by assuming that the channel output is Gaussian. Recently, Cusani and Laurenti [3] have given some new results on the dynamic convergence analysis of the constant modulus algorithm.

Unlike most of the previous convergence analysis works which specifically focused on some blind equalization algorithm, we will present the static and dynamic convergence analysis for almost all adaptive blind equalization algorithms when the coefficient sets of equalizers are near the global minima of their cost functions. In our analysis, we only use the *independent assumption*, which is widely used in the dynamic convergence analysis of adaptive algorithms [9, 14, 15, 16, 19, 26, 27]. Our analysis indicates that for a given channel and step-size, there is an optimum length of equalizer minimizing the intersymbol interference, which implies that the longer blind equalizer does not necessarily perform better than of the shorter one. This result can be applied to the design of blind equalizers used in digital communication systems.

This paper is organized as following. Section 2 briefly introduces the blind equalization in communication systems. Section 3 proves some properties of the prediction error function. Then, Section 4 briefly analyzes the static convergence of blind equalizers. Next, Section 5 studies the dynamic convergence of blind equalizers. Finally, computer simulation results are presented in Section 6 to demonstrate the consistency of our analysis results. Conclusion remarks are given in Section 7.

2 Adaptive Blind Equalizers

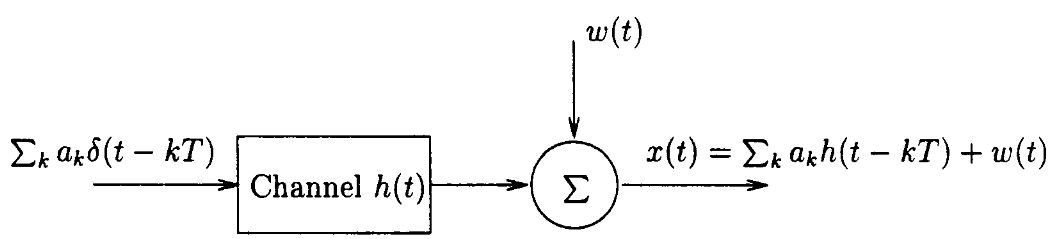


Figure 1: Baseband representation of a PAM communication system.

Without loss of generality, we consider a baseband representation of the pulse-amplitude-modulation (PAM) communication system as shown in Figure 1. A sequence of independent, identically distributed (i.i.d.) digital signal $\{a_n \in \mathcal{R}\}$ with zero-mean and variance σ^2 is sent by the transmitter at the symbol rate of $1/T$ through a channel exhibiting linear distortion. The resulting output signal $x(t)$ can be expressed as

$$x(t) = \sum_{k=-\infty}^{+\infty} a_k h(t-kT) + w(t), \tag{2.1}$$

where $w(t)$ is white Gaussian channel noise and $h(t)$ is the impulse response of the linear time-invariant (LTI) channel. If the channel output is sampled at the baud rate $1/T$, a stationary sequence is obtained, which can be expressed as

$$x_n \triangleq x(nT) = \sum_{k=-\infty}^{+\infty} a_{n-k}h_k + w_n, \tag{2.2}$$

in which we have used the definitions

$$h_n \triangleq h(nT) \qquad \text{and} \qquad w_n \triangleq w(nT). \tag{2.3}$$

The equalizer parameters $\{c_n\}$ are subject to adaptation via some algorithm to be determined. The equalizer output as shown in Figure 2 can then be written as

$$\begin{aligned} y_n &= \sum_{k=-\infty}^{\infty} c_k x_{n-k} \\ &= \sum_{k=-\infty}^{\infty} s_k a_{n-k}, \end{aligned} \tag{2.4}$$

where the channel noise is ignored and $\{s_n\}$ is the impulse response of the equalized system which can be expressed as

$$s_n \triangleq \sum_k h_k c_{n-k}. \tag{2.5}$$

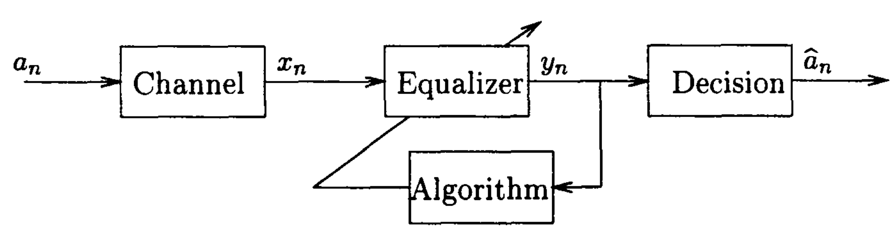


Figure 2: Diagram of typical channel equalization system.

In blind equalization, the original sequence is unknown to the receiver except for its probabilistic or statistical properties. A blind equalization algorithm is usually devised by minimizing a cost function consisting of the statistics of the output of the equalizer y_n, which is a function of $\{\cdots, s_{-1}, s_0, s_1, \cdots\}$ or $\{\cdots, c_{-1}, c_0, c_1, \cdots\}$, The cost function is usually of the form

$E\{\Phi(y_n)\}$, where $\Phi(y_n)$ is a function of y_n, which is selected such that the cost function has the global minimum points at

$$\{s_n\} = \pm\{\delta[n-n_d]\} \text{ for all } n_d = 0, \pm 1, \pm 2, \cdots \tag{2.6}$$

A stochastic gradient algorithm is used to minimize the cost function to obtain on-line equalization algorithm, which adjusts the k-th parameter of the equalizer at time n by

$$\widehat{c}_k^{(n+1)} = \widehat{c}_k^{(n)} - \mu\phi(y_n)x_{n-k}, \tag{2.7}$$

where μ is a small step size, $\phi(.)$ relates to the cost function by

$$\Phi(y_n) = \int_0^{y_n} \phi(x)dx, \tag{2.8}$$

and it is sometimes called *prediction error function.*

If an FIR filter is used as the equalizer, then (2.7) can be expressed as

$$\widehat{\mathbf{c}}^{(n+1)} = \widehat{\mathbf{c}}^{(n)} - \mu\mathbf{x}_n\phi(y_n), \tag{2.9}$$

where $\widehat{\mathbf{c}}^{(n)}$ is the *coefficient vector* of a blind equalizer after n-th iteration defined as

$$\widehat{\mathbf{c}}^{(n)} \triangleq [\widehat{c}_{-N}^{(n)}, \cdots, \widehat{c}_0^{(n)}, \cdots, \widehat{c}_N^{(n)}]^T \tag{2.10}$$

and $\mathbf{x}_n$ is the *channel output vector* at time n defined as

$$\mathbf{x}_n \triangleq [x_{n+N}, \cdots, x_n, \cdots, x_{n-N}]^T. \tag{2.11}$$

Since all channels can be approximated as a moving-average model with appropriate impulse response $\{h_{-M}, \cdots, h_0, \cdots, h_M\}$, the channel output vector can be expressed as

$$\mathbf{x}_n = \mathcal{H}^T\mathbf{a}_n, \tag{2.12}$$

where $\mathcal{H}$ is a $(2N+2M+1) \times (2N+1)$ *channel matrix* defined as

$$\mathcal{H} \triangleq \begin{pmatrix} h_{-M} & 0 & \cdots & 0 \\ \vdots & h_{-M} & \ddots & \vdots \\ h_0 & \vdots & \ddots & 0 \\ \vdots & h_0 & \ddots & h_{-M} \\ h_M & \vdots & \ddots & \vdots \\ 0 & h_M & \ddots & h_0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & h_M \end{pmatrix}, \tag{2.13}$$

and $\mathbf{a}_n$ is the *input symbol vector* at time n defined as

$$\mathbf{a}_n \triangleq [a_{n+(N+M)}, \cdots, a_n, \cdots, a_{n-(N+M)}]^T. \tag{2.14}$$

With the above definitions, the channel output can be expressed in a compact form:

$$\begin{aligned} y_n &= \mathbf{a}_n^T \widehat{\mathbf{s}}^{(n)} \\ &= \mathbf{a}_n^T \mathcal{H} \widehat{\mathbf{c}}^{(n)} \end{aligned} \tag{2.15}$$

where $\widehat{\mathbf{s}}^{(n)}$ is the equalized system vector at time n defined as

$$\widehat{\mathbf{s}}^{(n)} \triangleq [s_{-(N+M)}^{(n)}, \cdots, s_0^{(n)}, \cdots, s_{N+M}^{(n)}]^T. \tag{2.16}$$

It is obvious that an FIR channel can not be perfectly equalized by an FIR equalizer, that is, there is no equalization vector $\mathbf{c}$ such that

$$\mathcal{H}\mathbf{c} = \mathbf{e}_{M+N}, \tag{2.17}$$

where

$$\mathbf{e}_{M+N} = [\underbrace{0, \cdots, 0}_{M+N}, \ 1, \ \underbrace{0, \cdots, 0}_{M+N}]^T. \tag{2.18}$$

But when the length of the equalizer is very large, there exists a $\tilde{\mathbf{c}}$ such that $\| \mathcal{H}\tilde{\mathbf{c}} - \mathbf{e}_{M+N} \|$ is very small.

3 Properties of Prediction Error Function

Before analyzing the convergence behavior of blind equalizers, we first introduce some properties of the prediction error function here. The following lemma considers two important properties to be used in subsequent discussions.

Lemma 3.1 *The prediction error function $\phi(.)$ has the following two properties:*

1) When the parameters of a finite-length equalizer make its cost function attain one of its minima, the output of the equalized system, $\tilde{y}_n$, satisfies

(i) $E\{\phi(\tilde{y}_n)\mathbf{x}_n\} = \mathbf{0}$, and

(ii) $\mathcal{H}^T \tilde{F} \mathcal{H}$ is positive-definite,

where the $(2M+2N+1) \times (2M+2N+1)$ *matrix* $\tilde{F}$ *is defined as*

$$\tilde{F} = \frac{1}{\sigma^2} E\{\mathbf{a}_n \phi'(\tilde{y}_n) \mathbf{a}_n^T\}, \tag{3.1}$$

with $\phi'(.)$ *being the derivative of* $\phi(.)$, $\tilde{y}_n = \sum_k \tilde{c}_k x_{n-k}$ *and* $\tilde{c}_k$ *being the equalizer coefficients making the cost function attain a minimum.*

2) For all integers n and k

$$E\{\phi(a_n) a_k\} = 0, \tag{3.2}$$

and

$$E\{\phi'(a_n) a_k^2\} > 0. \tag{3.3}$$

Proof:

1) Let $\{\tilde{c}_n\}$ be the coefficients of an FIR blind equalizer which make the cost function $E\{\Phi(y_n)\}$ attain one of its minima, then

$$\frac{\partial}{\partial \tilde{c}_i} E\{\Phi(\tilde{y}_n)\} = 0 \tag{3.4}$$

and

$$\frac{\partial^2}{\partial \tilde{c}_i \partial \tilde{c}_j} E\{\Phi(\tilde{y}_n)\} \text{ is positive-definite.} \tag{3.5}$$

From (2.4) and (2.8),

$$\frac{\partial}{\partial \tilde{c}_i} E\{\Phi(\tilde{y}_n)\} = E\{\phi(\tilde{y}_n) x_{n-i}\}, \tag{3.6}$$

therefore, by (3.4), $E\{\phi(\tilde{y}_n) x_{n-i}\} = 0$, which is Lemma 3.1 1) (i). Using (2.4) and (2.8), direct calculation yields that

$$\begin{aligned} \frac{1}{\sigma^2}\left[\frac{\partial^2}{\partial \tilde{c}_i \partial \tilde{c}_j} E\{\Phi(\tilde{y}_n)\}\right] &= \frac{1}{\sigma^2} E\{\mathbf{x}_n \phi(\tilde{y}_n) \mathbf{x}_n^T\} \\ &= \frac{1}{\sigma^2} E\{\mathcal{H}^T \mathbf{a}_n \phi'(\tilde{y}_n) \mathbf{a}_n^T \mathcal{H}\} \\ &= \mathcal{H}^T \tilde{F} \mathcal{H}. \end{aligned} \tag{3.7}$$

According to (3.5), $\mathcal{H}^T \tilde{F} \mathcal{H}$ is positive-definite.

2) If a double-infinite length equalizer is used, then $\{\tilde{c}_i\} = \{\check{h}_i\}$, the channel inverse, is a global minimum of the cost function, and $\tilde{y}_n = a_n$. From the proof of the first part, for any integer k,

$$\begin{aligned} E\{\phi(a_n)a_k\} &= E\{\phi(\tilde{y}_n)\sum_i x_{k-i}\check{h}_i\} \\ &= \sum_i \check{h}_i E\{\phi(\tilde{y}_n)x_{k-i}\} \\ &= 0. \end{aligned} \tag{3.8}$$

Since matrix $[E\{\phi'(\tilde{y}_n)x_{n-i}x_{n-j}\}]$ is positive-definite, $\sum_i \sum_j \check{h}_i\check{h}_j E\{\phi'(\tilde{y}_n)x_{k-i}x_{k-j}\} > 0$, therefore

$$\begin{aligned} E\{\phi'(a_n)a_k^2\} &= E\{\phi'(\tilde{y}_n)\sum_i x_{k-i}\check{h}_i \sum_j x_{k-j}\check{h}_j\} \\ &= \sum_i \sum_j \check{h}_i\check{h}_j E\{\phi'(\tilde{y}_n)x_{k-i}x_{k-j}\} \\ &> 0 \end{aligned} \tag{3.9}$$

□

With the above lemma, we are now able to analyze the static and dynamic convergence of adaptive blind equalizers.

4 Static Convergence Analysis

If the equalizer is double-infinite, then at the global minimum of the cost function, the parameters of the equalizer

$$\{c_i\} = \{\pm\check{h}_{i-n_d}\}, \tag{4.1}$$

for some integer n_d. However, only FIR blind equalizer is used in practical systems. In this case, smart initialization strategies [6, 7, 11] will make the equalizer coefficients converge to a minimum $\{\tilde{c}_n : n = -N, \cdots, 0, \cdots, N\}$ of the cost function near the channel inverse such that $\tilde{y}_n - a_n$ is very small. Therefore, using the Taylor expansion, we have

$$\phi(\tilde{y}_n) = \phi(a_n) + \phi'(a_n)(\tilde{y}_n - a_n). \tag{4.2}$$

According to Lemma 3.1,

$$\begin{aligned} E\{\mathbf{x}_n\phi(a_n)\} &= \mathcal{H}^T E\{\mathbf{a}_n\phi(a_n)\} \\ &= \mathbf{0}, \end{aligned} \tag{4.3}$$

and

$$E\{\mathbf{x}_n\phi(\tilde{y}_n)\} = \mathbf{0}. \tag{4.4}$$

Consequently, from (4.2)

$$E\{\mathbf{x}_n\phi'(a_n)(\tilde{y}_n - a_n)\} = 0. \tag{4.5}$$

If we denote

$$\tilde{\mathbf{c}} \triangleq [\tilde{c}_{-N}, \cdots, \tilde{c}_0, \cdots, \tilde{c}_N]^T, \tag{4.6}$$

then

$$\tilde{y}_n = \mathbf{x}_n^T\tilde{\mathbf{c}}. \tag{4.7}$$

Substituting $\mathbf{x}_n = \mathcal{H}^T\mathbf{a}_n$ into (4.5), we can obtain that

$$\sigma^2 R_f\tilde{\mathbf{c}} = \mathcal{H}^T E\{\mathbf{a}_n\phi'(a_n)a_n\}, \tag{4.8}$$

where we have used the definitions

$$R_f \triangleq \mathcal{H}^T F\mathcal{H} \tag{4.9}$$

and

$$F \triangleq \frac{1}{\sigma^2}E\{\mathbf{a}_n\phi'(a_n)\mathbf{a}_n^T\}. \tag{4.10}$$

Since we have assumed that $\{a_n\}$ is an i.i.d. sequence with zero-mean and variance σ^2,

$$E\{\phi'(a_n)a_m a_l\} = \begin{cases} f(0)\sigma^2 & \text{if } n = m = l \\ f(1)\sigma^2 & \text{if } n \neq m = l \\ 0 & \text{otherwise} \end{cases}, \tag{4.11}$$

with

$$f(0) \triangleq \frac{1}{\sigma^2}E\{\phi'(a_n)a_n^2\} \ \text{ and } \ f(1) \triangleq E\{\phi'(a_n)\}. \tag{4.12}$$

Hence, F is a diagonal matrix and

$$F = \mathrm{diag}[\underbrace{f(1), \cdots, f(1)}_{M+N}, f(0), \underbrace{f(1), \cdots, f(1)}_{M+N}]. \tag{4.13}$$

Furthermore, from Lemma 3.1 2), $f(0) > 0$ and $f(1) > 0$, hence, F is positive-definite, therefore, R_f is also positive-definite for $\mathcal{H}$ is full rank. From (4.8), the coefficient vector at the minimum of the cost function near the channel inverse is

$$\tilde{\mathbf{c}} = f(0)R_f^{-1}\mathbf{h}, \tag{4.14}$$

where $\mathbf{h}$ is a $(2N+1) \times 1$ vector given by

$$\mathbf{h} = [0, \cdots, 0,\ h_M, \cdots, h_0, \cdots, h_{-M},\ 0, \cdots,\ 0]^T. \tag{4.15}$$

The above discussion can be summerized into the following theorem.

Theorem 4.1 *If an FIR equalizer is used to equalize an FIR channel, then at the minima near the channel inverse, the equalizer coefficient vector can be expressed as*

$$\tilde{\mathbf{c}} = f(0)R_f^{-1}\mathbf{h}. \tag{4.16}$$

The optimum equalizer (Wiener-Hopf filter) coefficient vector that minimizes $E\{(y_n - a_n)^2\}$ is given by [9]

$$\mathbf{c}_o = R^{-1}\mathbf{h}, \tag{4.17}$$

where

$$R = \mathcal{H}^T\mathcal{H}. \tag{4.18}$$

Hence, the sufficient and necessary condition for $\tilde{\mathbf{c}} = \mathbf{c}_o$ for any FIR channel is

$$f(0) = f(1), \tag{4.19}$$

which means

$$E\{\phi'(a_n)a_n^2\} = E\{\phi'(a_n)\}E\{a_n^2\}. \tag{4.20}$$

For Sato algorithm[18], decision-directed equalizers[14, 16], $\phi'(a_n) = 1$, and therefore, $\tilde{\mathbf{c}} = \mathbf{c}_o$. For Godard algorithm, $\phi(y) = y(y^2 - r)4$ with $r = \frac{E\{a_n^4\}}{E\{a_n^2\}}$, therefore,

$$E\{\phi'(a_n)a_k^2\} = 3\sigma^4 - m_4, \tag{4.21}$$

and

$$E\{\phi'(a_n)a_n^2\} = 2m_4, \tag{4.22}$$

where

$$m_4 = E\{a_n^4\}. \tag{4.23}$$

Hence, if the channel input is binary, (4.20) is true and $\tilde{\mathbf{c}} = \mathbf{c}_o$. Otherwise, $\tilde{\mathbf{c}} \neq \mathbf{c}_o$.

The distortion due to the finite-length of equalizer is

$$\begin{aligned} D_f &\triangleq \| \tilde{\mathbf{s}} - \mathbf{e}_{M+N} \|^2 \\ &= \| \mathcal{H}\tilde{\mathbf{c}} - \mathbf{e}_{M+N} \|^2 . \end{aligned} \tag{4.24}$$

With the increase of the length of the blind equalizer, the global minimum of the cost function adopted by the equalization algorithm will be closer to the channel inverse. Hence, the distortion D_f will decrease.

5 Dynamic Convergence Analysis

In this section, we study the dynamic convergence behavior of blind equalizers when the parameters of blind equalizers near the global minimum of the cost function.

From (2.15) and (4.7), the output of the equalizer can be expressed as

$$\begin{aligned} y_n &= \tilde{y}_n + (y_n - \tilde{y}_n) \\ &= \tilde{y}_n + \mathbf{a}_n^T \mathcal{H} \epsilon_n, \end{aligned} \tag{5.1}$$

where we have used the notation that

$$\epsilon_n \triangleq \widehat{\mathbf{c}}^{(n)} - \tilde{\mathbf{c}}. \tag{5.2}$$

Around the minima of the cost function of the equalizer $\| \epsilon_n \|$ is small, therefore $y_n - \tilde{y}_n$ is also small. Applying the Taylor expansion to $\phi(y_n)$ at $\tilde{y}_n$, we have

$$\phi(y_n) = \phi(\tilde{y}_n) + \phi'(\tilde{y}_n)\mathbf{a}_n^T \mathcal{H} \epsilon_n. \tag{5.3}$$

Subtracting both sides of (2.9) by $\tilde{\mathbf{c}}$ and using (5.3), we have that

$$\epsilon_{n+1} = \epsilon_n - \mu(\mathbf{x}_n \phi(\tilde{y}_n) + \mathcal{H}^T \mathbf{a}_n \phi'(\tilde{y}_n)\mathbf{a}_n^T \mathcal{H} \epsilon_n). \tag{5.4}$$

This is the key identity that we are going to apply in the dynamic convergence analysis in this section.

In our analysis, we will use the *independent assumption* [9, 16, 27, 26] which assumes that a_n and ϵ_n are statistically independent. Similar assumptions have also been used in the convergence analysis of LMS algorithm, decision-directed equalizer, and Sato algorithm. The references [9, 16, 27, 26] have given some good justification on the validation of this assumption.

5.1 Convergence in Mean

To study the mean convergence of adaptive blind equalizers, we first take ensemble average on both sides of (5.4),

$$E\{\epsilon_{n+1}\} = E\{\epsilon_n\} - \mu(E\{\mathbf{x}_n\phi(\tilde{y}_n)\} + \mathcal{H}^T E\{\mathbf{a}_n\phi'(\tilde{y}_n)\mathbf{a}_n^T\}\mathcal{H}E\{\epsilon_n\}). \tag{5.5}$$

in which we have used the *independent assumption.* From Lemma 3.1 1) (i), $E\{\mathbf{x}_n\phi(\tilde{y}_n)\} = 0$. Thus, (5.5) can be simplified into

$$E\{\epsilon_{n+1}\} = (I - \mu\sigma^2\mathcal{H}^T\tilde{F}\mathcal{H})E\{\epsilon_n\}, \tag{5.6}$$

with

$$\tilde{F} \triangleq \frac{1}{\sigma^2}E\{\mathbf{a}_n\phi'(\tilde{y}_n)\mathbf{a}_n^T\}. \tag{5.7}$$

Hence,

$$E\{\epsilon_n\} = (I - \mu\sigma^2\mathcal{H}^T\tilde{F}\mathcal{H})^n E\{\epsilon_0\}. \tag{5.8}$$

Let $\Delta\mathbf{s}^{(n)} \triangleq \mathbf{s}^{(n)} - \tilde{\mathbf{s}}$, then $\Delta\mathbf{s}^{(n)} = \mathcal{H}\epsilon_n$ and $\epsilon_n = (\mathcal{H}^T\mathcal{H})^{-1}\mathcal{H}^T\Delta\mathbf{s}^{(n)}$. It follows from (5.8) that

$$E\{\Delta\mathbf{s}^{(n)}\} = \mathcal{H}(I - \mu\sigma^2\mathcal{H}^T\tilde{F}\mathcal{H})^n(\mathcal{H}^T\mathcal{H})^{-1}\mathcal{H}^T E\{\Delta\mathbf{s}^{(0)}\}. \tag{5.9}$$

Since $\mathcal{H}^T\tilde{F}\mathcal{H}$ is positive-definite according to Lemma 3.1 1) (i),

$$\lim_{n\to\infty} \| E\{\epsilon_n\} \| = 0 \tag{5.10}$$

and

$$\lim_{n\to\infty} \| E\{\Delta\mathbf{s}^{(n)}\} \| = 0 \tag{5.11}$$

if μ satisfies $0 < \mu < \frac{2}{\lambda_{max}\sigma^2}$, where λ_{max} is the largest eigenvalue of $\mathcal{H}^T\tilde{F}\mathcal{H}$. The above discussion can be summerized into the following theorem.

Theorem 5.1 *For any FIR blind equalization algorithm, mean convergence behavior near the global minimum of the cost function satisfies*

$$E\{\epsilon_n\} = (I - \mu\sigma^2 \tilde{R}_f)^n E\{\epsilon_0\}, \tag{5.12}$$

where

$$\tilde{R}_f \triangleq \mathcal{H}^T \tilde{F} \mathcal{H}. \tag{5.13}$$

If the step-size μ in iteration formula (2.7) or (2.9) satisfies

$$0 < \mu < \frac{2}{\lambda_{max}\sigma^2}, \tag{5.14}$$

then

$$E\{\mathbf{c}^{(n)}\} \to \tilde{\mathbf{c}}^{(n)} \quad and \quad E\{\mathbf{s}^{(n)}\} \to \tilde{\mathbf{s}}^{(n)}. \tag{5.15}$$

5.2 Consistency

Having proved $E\{\mathbf{c}^{(n)}\} \to \tilde{\mathbf{c}}^{(n)}$ and $E\{\mathbf{s}^{(n)}\} \to \tilde{\mathbf{s}}^{(n)}$, we are going to study the consistency of the convergence here.

From (5.4), we have

$$E\{\epsilon_{n+1}\epsilon_{n+1}^T\} = E\{\epsilon_n \epsilon_n^T\} - \mu(I_1 + I_2 - \mu I_3), \tag{5.16}$$

where

$$I_1 = E\{(\mathcal{H}^T \mathbf{a}_n \phi(\tilde{y}_n) + \mathcal{H}^T \mathbf{a}_n \phi'(\tilde{y}_n) \mathbf{a}_n^T \mathcal{H} \epsilon_n)\epsilon_n^T\}, \tag{5.17}$$

$$I_2 = E\{\epsilon_n (\mathcal{H}^T \mathbf{a}_n \phi(\tilde{y}_n) + \mathcal{H}^T \mathbf{a}_n \phi'(\tilde{y}_n) \mathbf{a}_n^T \mathcal{H} \epsilon_n)^T\}, \tag{5.18}$$

and

$$I_3 = E\{(\mathcal{H}^T \mathbf{a}_n \phi(\tilde{y}_n) + \mathcal{H}^T \mathbf{a}_n \phi'(\tilde{y}_n) \mathbf{a}_n^T \mathcal{H} \epsilon_n)(\mathcal{H}^T \mathbf{a}_n \phi(\tilde{y}_n) + \mathcal{H}^T \mathbf{a}_n \phi'(\tilde{y}_n) \mathbf{a}_n^T \mathcal{H} \epsilon_n)^T\}. \tag{5.19}$$

Using Lemma 3.1 1) (i), I_1 and I_2 can be simplified as

$$I_1 = \sigma^2 \tilde{R}_f R_{\epsilon_n}, \quad I_2 = \sigma^2 R_{\epsilon_n} \tilde{R}_f \tag{5.20}$$

where R_{ϵ_n} denotes the correlation matrix of ϵ_n, i.e.

$$R_{\epsilon_n} = E\{\epsilon_n \epsilon_n^T\}. \tag{5.21}$$

The dominant term in I_3 is

$$I_3 = \sigma^2 \tilde{R}_g, \tag{5.22}$$

in which

$$\tilde{R}_g = \{\mathcal{H}^T \tilde{G} \mathcal{H}\}, \tag{5.23}$$

and

$$\tilde{G} = \frac{1}{\sigma^2} E\{\mathbf{a}_n \phi^2(\tilde{y}_n) \mathbf{a}_n^T\}. \tag{5.24}$$

It is obvious that both $\tilde{G}$ and $\tilde{R}_g$ are positive-definite matrix. Substituting (5.20) and (5.22) into (5.16), we obtain that

$$R_{\epsilon_{n+1}} = R_{\epsilon_n} - \mu\sigma^2(\tilde{R}_f R_{\epsilon_n} + R_{\epsilon_n}\tilde{R}_f - \mu\tilde{R}_g). \tag{5.25}$$

Let R_ϵ be the unique positive-definite solution of the Lyapunov equation

$$\tilde{R}_f R_\epsilon + R_\epsilon \tilde{R}_f = \mu \tilde{R}_g, \tag{5.26}$$

then (5.25) is equivalent to

$$R_{\epsilon_{n+1}} - R_\epsilon = \frac{1}{2}\Lambda(R_{\epsilon_n} - R_\epsilon) + \frac{1}{2}(R_{\epsilon_n} - R_\epsilon)\Lambda, \tag{5.27}$$

where

$$\Lambda = I - 2\mu\sigma^2 \tilde{R}_f. \tag{5.28}$$

Hence,

$$\| R_{\epsilon_{n+1}} - R_\epsilon \| \leq \| \Lambda \| \| R_{\epsilon_n} - R_\epsilon \| . \tag{5.29}$$

Therefore,

$$\| R_{\epsilon_n} - R_\epsilon \| \leq \| \Lambda \|^n \| R_{\epsilon_0} - R_\epsilon \| . \tag{5.30}$$

If $0 < \mu < \frac{1}{\lambda_{max}\sigma^2}$, then $\| \Lambda \| < 1$, and

$$\lim_{n\to\infty} \| R_{\epsilon_n} - R_\epsilon \| = 0. \tag{5.31}$$

Therefore,

$$\lim_{n\to\infty} R_{\epsilon_n} = R_\epsilon. \tag{5.32}$$

Since $R_\epsilon \neq \mathbf{0}$, $E\{\mathbf{c}^{(n)}\} \to \tilde{\mathbf{c}}$ is not consistent. Summerizing the above analysis, the following theorem is obtained.

Theorem 5.2 *The equalizer coefficient vector $\mathbf{c}^{(n)} \to \tilde{\mathbf{c}}$ is not consistent and at the equilibrium near the minimum of the cost function, the correlation matrix R_ϵ of ϵ is uniquely determined by the following Lyapunov equation*

$$\tilde{R}_f R_\epsilon + R_\epsilon \tilde{R}_f = \mu \tilde{R}_g, \tag{5.33}$$

if $0 < \mu < \frac{1}{\lambda_{max}\sigma^2}$.

From the above discussion, the distortion of the equalized system due to gradient noise is

$$\begin{aligned} D_g &\triangleq E\{\| \mathbf{s} - \tilde{\mathbf{s}} \|^2\} \qquad (5.34) \\ &= E\{\| \mathcal{H}\epsilon \|^2\} \\ &= \mathrm{tr}[\mathcal{H}^T R_\epsilon \mathcal{H}] \\ &= \mathrm{tr}[R R_\epsilon]. \end{aligned}$$

When an FIR equalizer is so long that $\{\tilde{c}_n \approx \check{h}_n\}$, $\{\tilde{y}_n \approx a_n\}$, then

$$\tilde{R}_f \approx R_f, \quad \tilde{R}_g \approx R_g, \tag{5.35}$$

where we used the definitions

$$R_g \triangleq \mathcal{H}^T G \mathcal{H}, \tag{5.36}$$

and

$$g(0) \triangleq \frac{1}{\sigma^2} E\{\phi^2(a_n) a_n^2\}, \quad g(1) = E\{\phi^2(a_n)\}, \tag{5.37}$$

$$G \triangleq \mathrm{diag}[\underbrace{g(1), \cdots, g(1)}_{M+N}, g(0), \underbrace{g(1), \cdots, g(1)}_{M+N}]. \tag{5.38}$$

In this case, (5.33) becomes

$$R_f R_\epsilon + R_\epsilon R_f = \mu R_g. \tag{5.39}$$

For the blind equalization algorithms with $f(0) = f(1)$, $R_f = f(1)R$. Using (5.39), we have

$$D_g = \frac{1}{2f(1)} \mu \mathrm{tr}[R_g] \approx \frac{1}{2}\mu(2N+1)\frac{g(1)}{f(1)}. \tag{5.40}$$

For those blind equalizers with $f(0) \neq f(1)$, (5.40) can also be used to approximately estimate the average distortion introduced by gradient noise. According to (5.40), D_g is proportional to the step-size μ and the length of equalizer N. But, on the other hand step-size affects the convergence speed of equalizer, i.e. the larger the faster it converges if μ is in the allowable range. Hence, when we select the step-size of an equalizer, we have to consider the trade-off between these two factors.

As we have seen, there are two sources of distortion. One is D_f in (4.24) due to the finite length of an equalizer, another is D_g in (5.40) due to the gradient noise. Once the step-size of a blind equalizer is set, there must be an optimum length for an FIR equalizer to minimize the total distortion $D = D_f + D_g$ since with the increase of the equalizer length, D_f decreases while D_g increases.

6 Computer Simulations

Since approximation has been used in our theoretical analysis, we shall check the validity of our theory by computer simulations. Two computer simulation examples are presented in this section.

Example 1:

The channel input sequence $\{a_n\}$ is independent, uniformly distributed over $\{\pm a, \pm 3a\}$ ($a = 1/\sqrt{5}$ to make $E\{a_n^2\} = 1$). The impulse response of the channel is $h_n = 0.3^n u[n]$ with $u[n]$ being unit step function. The channel impulse response and frequency response are shown in Figure 3. An FIR equalizer with coefficients c_0 and c_1 is used to compensate the channel distortion. The initial value of the equalizer coefficient vector is set to be

$$\mathbf{c}^{(0)} = [c_0^{(0)},\ c_1^{(0)}]^T = [1,\ 0]^T. \tag{6.1}$$

The Sato algorithm[18] is first used to adjust the coefficients of the equalizer. When the step-size $\mu = 0.002$, 10 trails of learning curves of $\mathbf{c}^{(n)}$ are shown in Figure 4. In this figure, the thick solid line is the theoretical average learning curve calculated from (5.8), the thick dot-dash lines are the theoretical one-standard-deviation lines determined by (5.27). According to this figure, 10 trails of learning curves are almost within one standard deviation of the theoretical average learning curves for Sato algorithm. Figure 5 demonstrates the ensemble averages of learning curves for different step-size based on 100 trails. From Figure 5, our theoretical results fit to the simulation results very well for Sato algorithm.

Similar simulations have also been done for Godard algorithm[8]. The simulation results are shown in Figure 6 and 7, which also confirm our theoretical analysis.

Example 2:

The channel input sequence in this example has the same statistical property as in Example 1. The channel impulse and frequency response are shown in Figure 8, which is a typical telephone channel [20]. The center-tap initialization strategy[7] is used for blind equalization algorithm.

When the Sato algorithm is used, the theoretical relationship between the total distortion and the length of equalizer for different step sizes are illustrated in Figure 9 (a), which indicates that the optimum length of Sato equalizer for this channel is between 15 and 25 dependent upon the step-size. Figure 9 (b) demonstrates the comparison between the theoretical results of $D_f + D_g$ and simulated results for step size $\mu = 0.002$.

The calculation and simulation results are given in Figure 10 for Godard algorithm. Because $g(1)/f(1)$ for Godard algorithm (0.169) is less than that for Sato algorithm (0.250) for 4-level PAM input, Godard algorithm should have less distortion than Sato should according to (5.40), which is confirmed by comparing Figure 9 and 10.

7 Conclusion

We have studied the static and dynamic convergence behavior of adaptive blind equalizers based on the first-order approximation to the cost function of blind algorithms under the independent assumption. Our analysis indicates that for a given channel and step-size, there is an optimal length for an equalizer to minimize the intersymbol interference. This result implies that a longer-length blind equalizer does not necessarily outperform a shorter one, as contrary to what conventionally conjectured. The theoretical results are confirmed by computer simulation examples. The analysis method used in this paper can also be applied in the analysis of tracking performance of the blind equalizers used in time-varying channels, such as mobile radio channel and HF channel. The analysis results presented in this paper can be directly employed in the design of blind equalizer in practical communication systems.

References

[1] A. Benveniste, M. Goursat, and G. Ruget. "Robust identification of a nonminimum phase system: blind adjustment of a linear equalizer in data communications," *IEEE Trans. on Automatic Control*, AC-25:385–399, June 1980.

[2] A. Benveniste and M. Goursat, "Blind equalizers," *IEEE Transactions on Communications*, COM-32: 871-882, August, 1982.

[3] R. Cusani and A. Laurenti, "Convergence analysis of the CMA blind equalizer," *IEEE Transactions on Communications*, COM-43:1304-1307, Feb./March/April 1995.

[4] Z. Ding, R. A. Kennedy, B. D. O. Anderson, and C. R. Johnson, Jr.. "Ill-convergence of Godard blind equalizers in data communication systems," *IEEE Trans. on Communications*, COM-39: 1313-1327, Sept. 1991.

[5] Z. Ding, R. A. Kennedy, B. D. O. Anderson and C. R. Johnson, Jr.. "On the (non)existence of undesirable equilibria of Godard blind equalizer," *IEEE Trans on Signal Processing*, ASSP-40: 2425-2432, Oct. 1992.

[6] Z. Ding, R. A. Kennedy, B. D. O. Anderson and C. R. Johnson, Jr.. "Local convergence of the Sato blind equalizer and generalization under practical constrains," *IEEE Trans on Information Theory*, IT-39: 129–144, Jan. 1993.

[7] G. J. Foschini, "Equalization without altering or detecting data," *AT&T Technical Journal*, 64: 1885-1911, October, 1985.

[8] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. on Communications*, COM-28, pp. 1867-1875, Nov. 1980.

[9] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, 1991

[10] R. A. Kennedy, B. O. Anderson and R. R. Bitmead, "Stochastic Dynamics of Blind Decision Feedback Equalizer Adaptation," *Adaptive Systems in Control and Signal Processing 1989* pp.579-584, IFAC Symposia Series, 1990.

[11] Y. Li and Z. Ding, "Convergence analysis of finite length blind adaptive equalizer," to appear on *IEEE Trans on Signal Processing*.

[12] Y. Li and Z. Ding, 'Global convergence of fractionally spaced Godard equalizer', *The 26th Asiloma Conference on Signal, Systems & Computers*, California, October 1994

[13] Y. Li, K. J. R. Liu, and Z. Ding, "On the convergence of blind channel equalization," Technical Report of SRC of University of Maryland, T.R. 95-47.

[14] O. Macchi and E. Eweda, "Convergence analysis of self-adaptive equalizers," *IEEE Transactions on Information Theory*, IT-30:161-176, March 1984.

[15] V. J. Mathews and H. Cho, "Improved convergence analysis off stochastic gradient adaptive filters using the sign algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-35: 450-454, April, 1987.

[16] J. E. Mazo, "Analysis of Decision-Directed Equalizer Convergence," *The Bell System Technical Journal*, Vol.59, No.10, pp1857-1876, December 1980.

[17] G. Picchi and G. Prati, "Blind equalization and carrier recovery using a "Stop-and-Go" decision-directed algorithm," *IEEE Trans. on Comm.* COM-35: 877-887, Sept. 1987

[18] Y. Sato, "A method of self-recovering equalization for multi-level amplitude modulation," *IEEE Trans. on Comm.* COM-23: 679-682, June 1975

[19] C. K. Chan and J. J. Shynk, "Stationary Points of the Constant Modulus Algorithm for Real Gaussian Signals," *IEEE Transactions on Signal Processing*, ASSP-38:2176-2180, 1990.

[20] J. Proakis, *Digital Communications* (2nd Edition), McGraw-Hill 1989.

[21] J.R. Treichler and B.G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-31: 349-372, April, 1983.

[22] J. R. Treichler, V. Wolff and C. R. Johnson, Jr., "Observed misconvergence in the constant modulus adaptive algorithm," *Proc. 25th Asilomar Conference on Signals, Systems and Computers*, pp. 663-667, Pacific Grove, CA, 1991.

[23] S. Verdu, B. D. O. Anderson, R. A. Kennedy, "Blind equalization without gain identification," *IEEE Transactions on Information Theory*, IT-39:292-297, Jan. 1993.

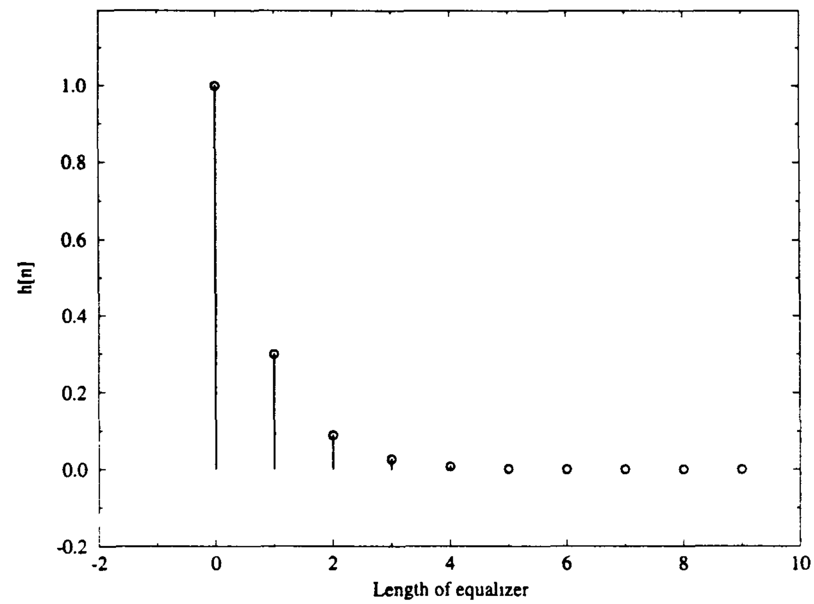
[24] O. Shalvi and E. Weinstein, "New criteria for blinα deconvolution of non-minimum phase systems (channels)," *IEEE Transactions on Information Theory*, IT-36:312-321, March 1990.

[25] J. K. Tugnait, O. Shalvi, and E. Weinstein, "Comments on 'New Criteria for Blind Deconvolution of Nonminimum Phase Systems (Channels)," *IEEE Transactions on Information Theory*, IT-38:210-213, Jan. 1992.

[26] V. Weerackody, S. A. Kassam, and K. R. Laker, "Convergence analysis of an algorithm for blind equalization," *IEEE Transactions on Communications*, Vol.39:856-865, June, 1991.

[27] B. Windrow *et al*, "Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter," *Proceedings of the IEEE*, Vol.64:1151-1162, August 1976.

(a)



(b)

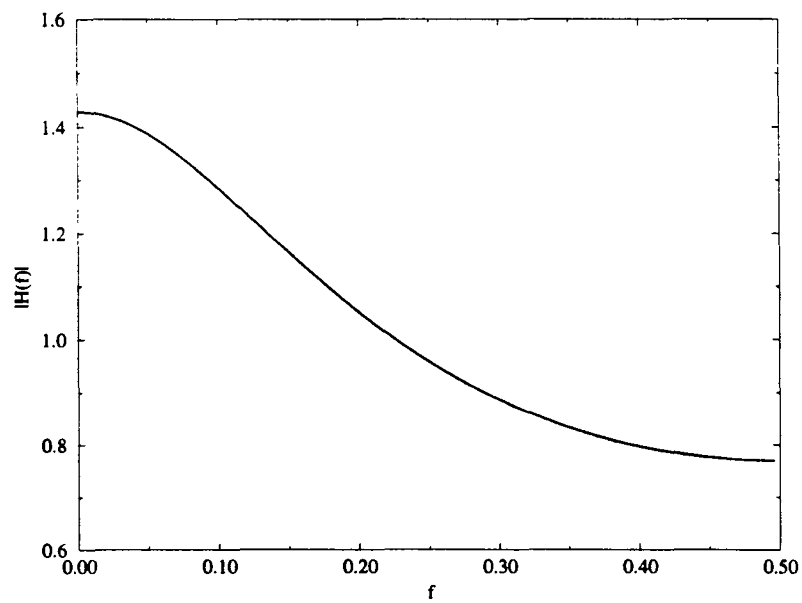


Figure 3: (a) The impulse response, and (b) the frequency response of channel I.

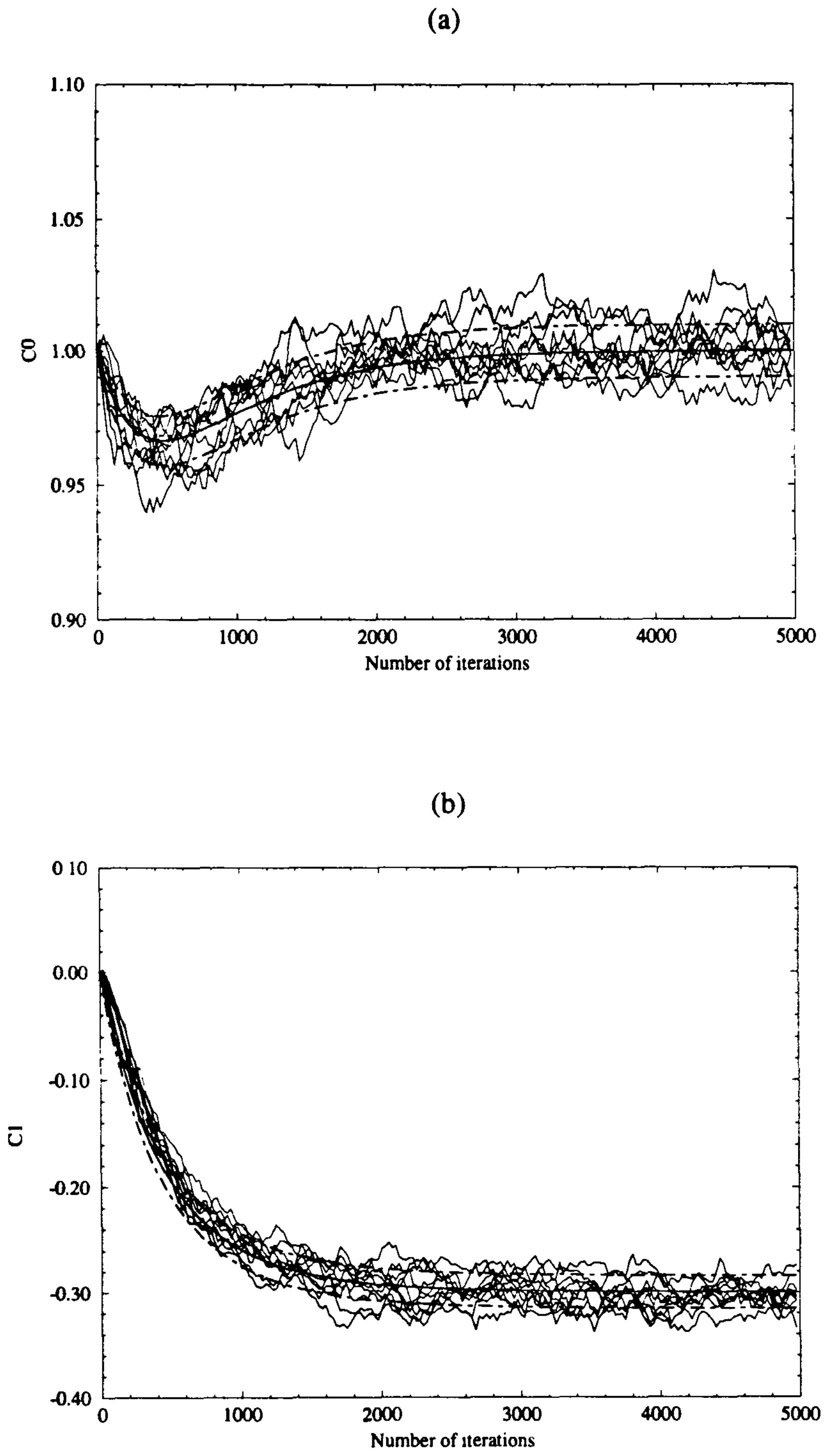


Figure 4: 10 trails of learning curves of (a) c_0, and (b) c_1 for Sato algorithm using $\mu = 0.002$.

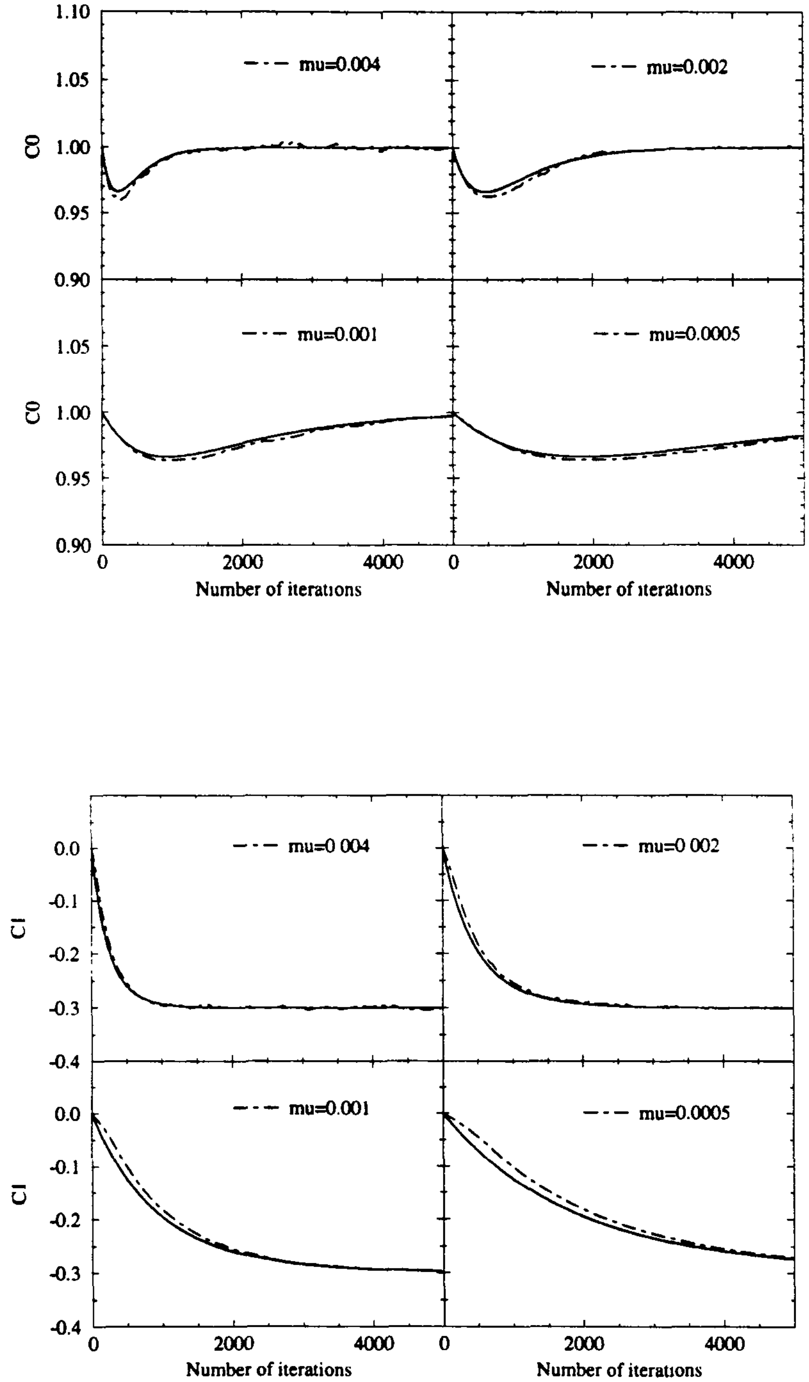


Figure 5: Average of learning curves: theoretical ones (solid lines) and simulated ones (dot-dash lines) based on 100 ensemble trails for Sato algorithm using different μ.

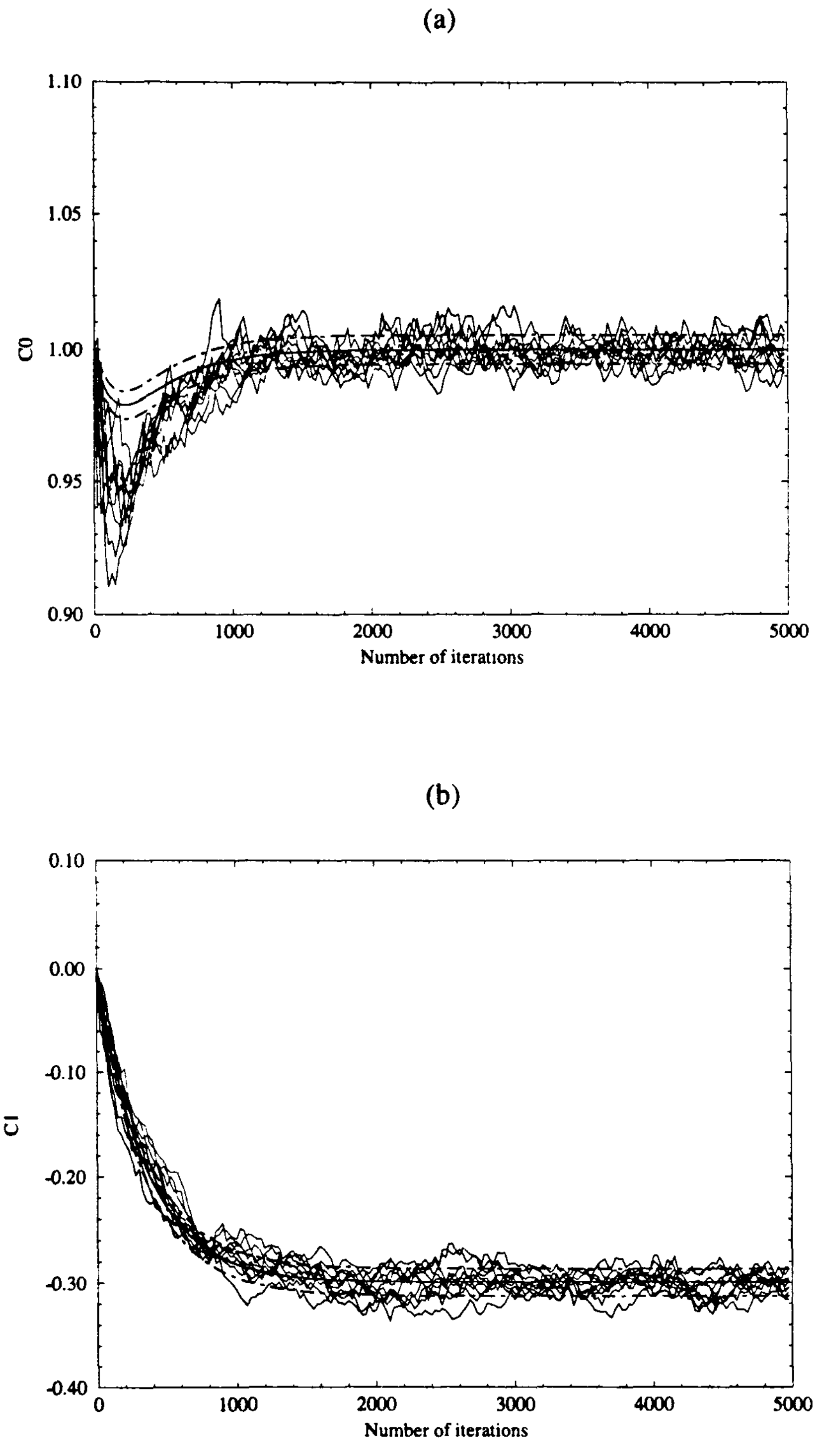


Figure 6: 10 trails of learning curves of (a) c_0, and (b) c_1 for Godard algorithm using $\mu = 0.002$.

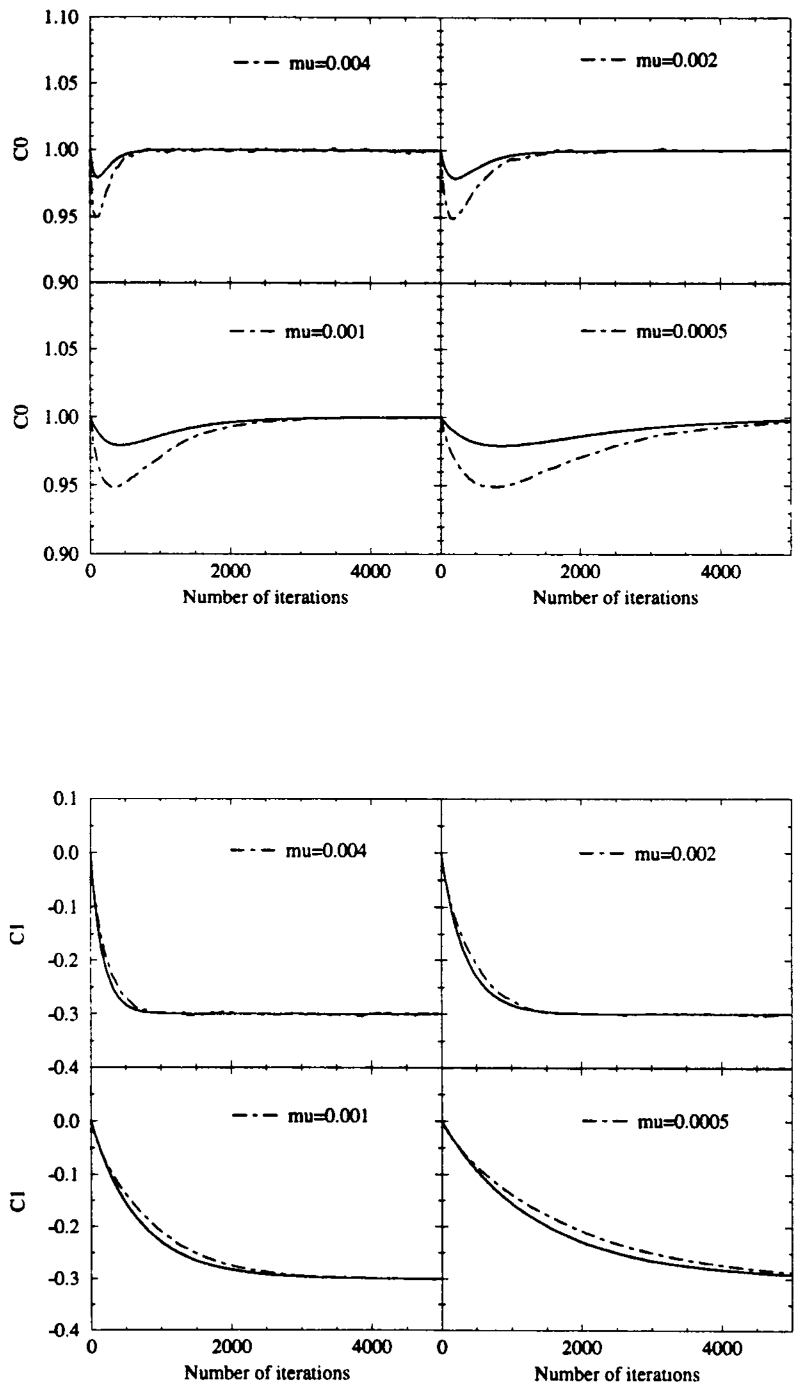
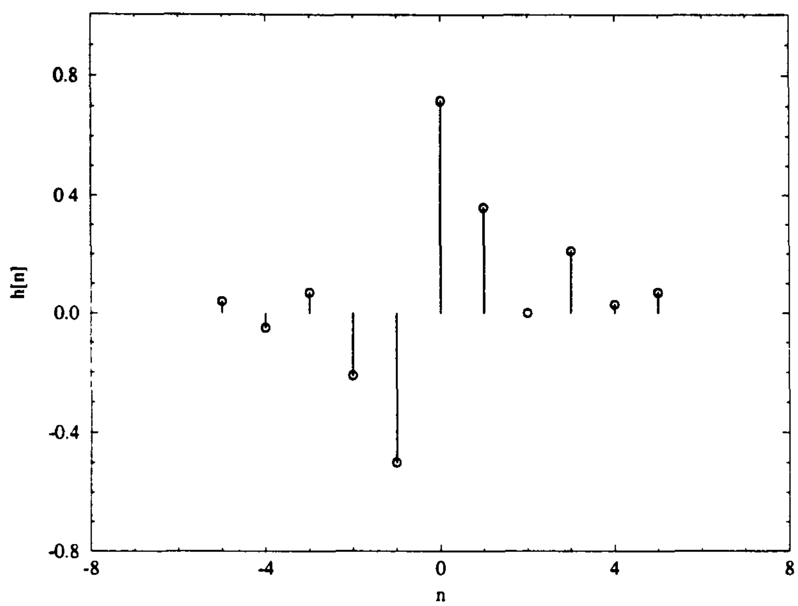


Figure 7: Average of learning curves: theoretical ones (solid lines) and simulated ones (dot-dash lines) based on 100 ensemble trails for Godard algorithm using different μ.

(a)



(b)

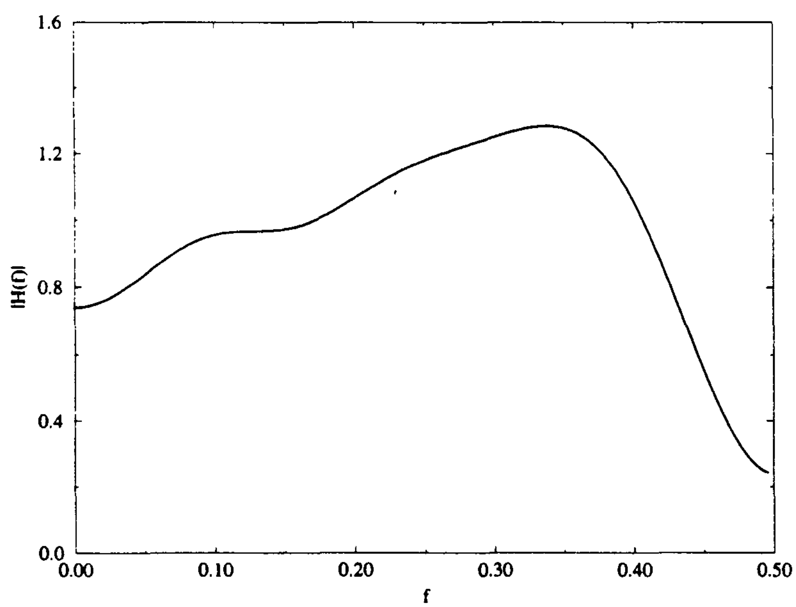
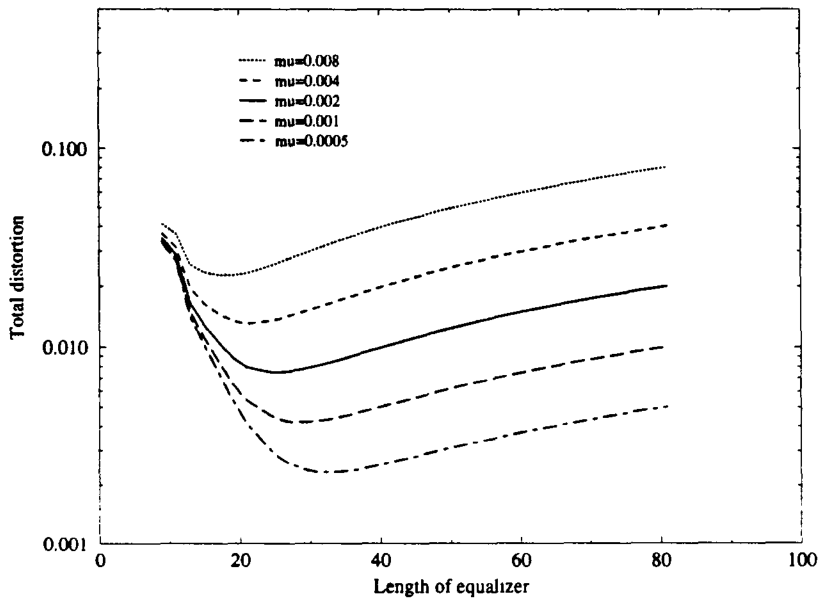


Figure 8: (a) The impulse response, and (b) the frequency response of channel II.

(a)



(b)

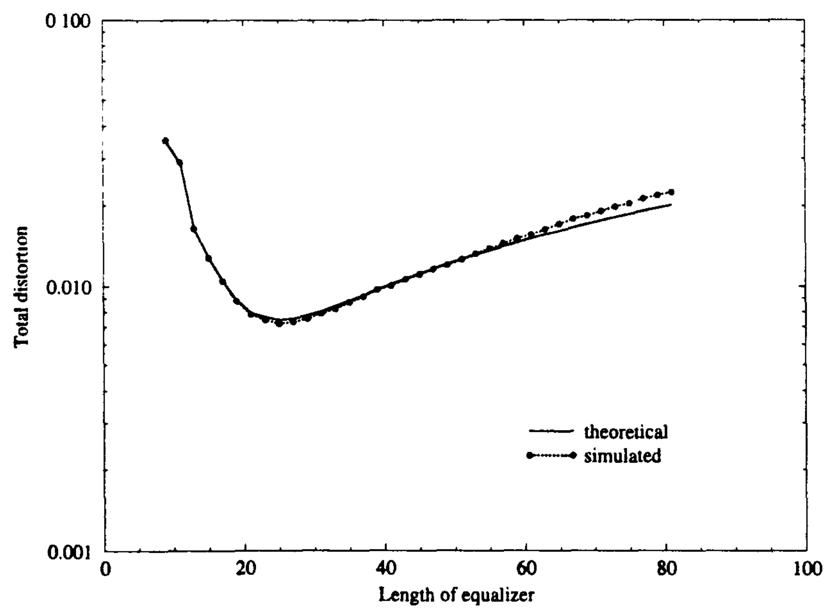
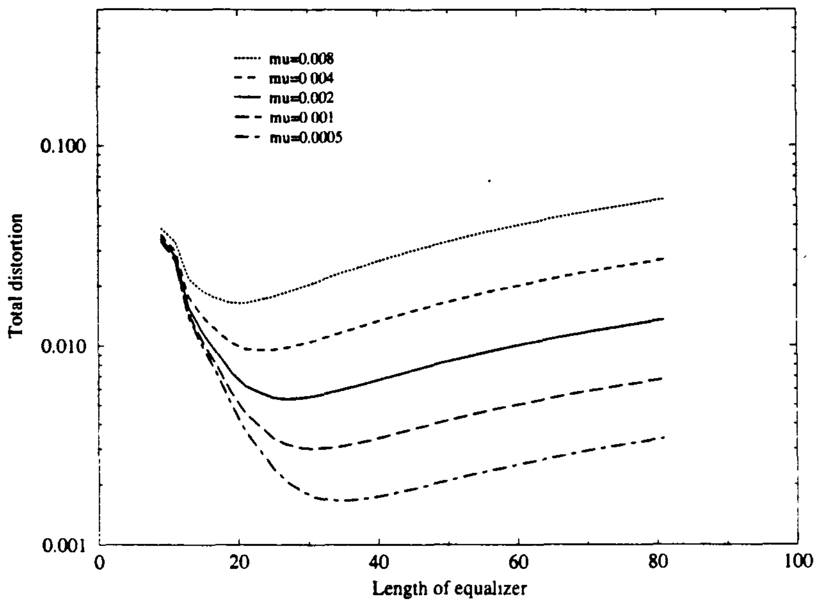


Figure 9: Total distortion of equalized system (a) theoretical results for different step size μ, (b) simulation results for $\mu = 0.002$, using Sato algorithm.

(a)



(b)

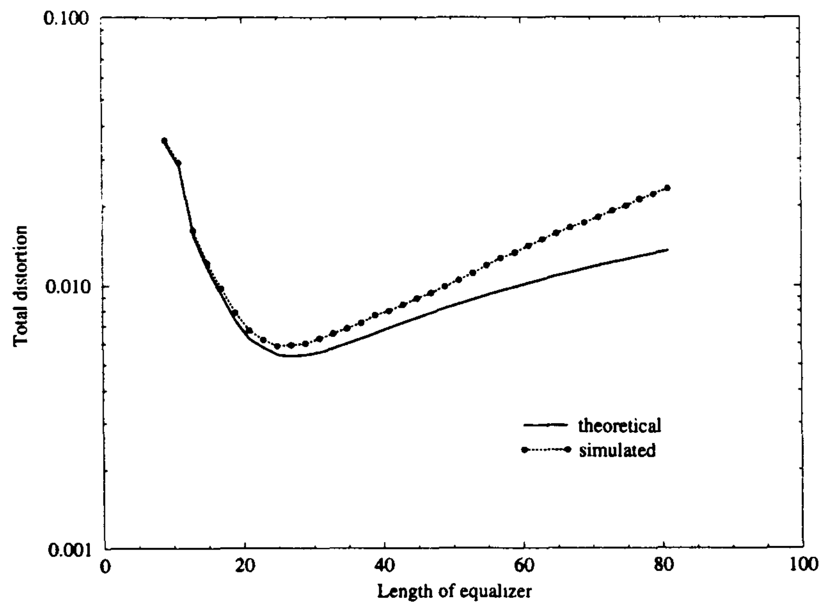


Figure 10: Total distortion of equalized system (a) theoretical results for different step size μ, (b) simulation results for $\mu = 0.002$, using Godard algorithm.