

Static and Dynamic Convergence Behavior of Adaptive Blind Equalizers

Ye Li and K. J. Ray Liu

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 independence 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 stepsize, 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, which is contrary to what is conventionally conjectured. The theoretical analysis results are confirmed by computer simulations.

I. INTRODUCTION

SINCE THE pioneering work by Sato [18], many 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 intersymbol 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 understanding their performance. We may categorize the convergence analysis 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 proven 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 nonlinearity 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 equalizers by assuming that equalizer is in "open eye pattern." In [26], Weerackody and his co-authors 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. Since there are many initialization strategies [6], [7], [11] to make blind equalizer reach an open eye pattern, we will concentrate on the convergence analysis when the coefficient sets of equalizers are near the global minima of their cost functions. In the static analysis, we derive the close form solution for the coefficients of FIR blind equalizers from which we can evaluate the distortion caused by the finite length effect. In the dynamic analysis, we only use the *independence assumption* [9], which is widely used in the dynamic convergence analysis of adaptive algorithms [9], [14]–[16], [19], [26], [27]. Based on the independence assumption, together with the first-order approximation, we study the convergence and consistence of the equalizer coefficients. 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 a longer blind equalizer does not necessarily perform better than a 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 II briefly introduces the blind equalization in communication systems. Section III proves some properties of the prediction error function. Then, Section IV briefly analyzes the static convergence of blind equalizers. Next, Section V studies the dynamic convergence of blind equalizers. Finally, computer simulation results are presented in Section VI to demonstrate

Manuscript received June 29, 1995; revised May 27, 1996. This work was supported, in part, by the NSF grants MIP9309506 and MIP9457397. The associate editor coordinating the review of this paper and approving it for publication was Jose Carlos M. Bermudez.

Y. Li is with the Department of Electrical Engineering, University of Maryland, College Park, MD 20742 USA. He is now with the Wireless Systems Research Department, AT&T Laboratories—Research, Holmdel, NJ 07733-0400 USA.

K. J. R. Liu is with the Department of Electrical Engineering, University of Maryland, College Park, MD 20742 USA.

Publisher Item Identifier S 1053-587X(96)08240-2.

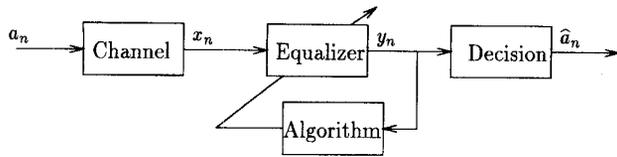


Fig. 1. PAM communication system with blind channel equalizer.

the consistency of our analysis results. Conclusion remarks are given in Section VII.

II. ADAPTIVE BLIND EQUALIZERS

Without loss of generality, we consider a baseband representation of the pulse-amplitude-modulation (PAM) communication system with blind channel equalizer as shown in Fig. 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 through a bounded-input bounded-output (BIBO) channel exhibiting linear distortion. The resulting output signal x_n can be expressed as

$$x_n = \sum_{k=-\infty}^{+\infty} a_k h_{n-k} + w_n \quad (2.1)$$

where h_n is the impulse response of the linear time-invariant (LTI) channel, and w_n is white Gaussian channel noise. In this paper, we will ignore the effects of the channel noise.

As shown in Fig. 1, a linear channel equalizer with parameters $\{c_n\}$ is used to remove the intersymbol interference caused by the channel distortion. The parameters $\{c_n\}$ are subject to adaptation via some algorithm to be determined. The equalizer output in Fig. 1 can then be written as

$$\begin{aligned} y_n &= \sum_{k=-\infty}^{\infty} c_k x_{n-k} \\ &= \sum_{k=-\infty}^{\infty} s_k a_{n-k} \end{aligned} \quad (2.2)$$

where $\{s_n\}$ is the impulse response of the equalized system related to h_n and c_n by

$$s_n \triangleq \sum_k h_k c_{n-k}. \quad (2.3)$$

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 $\{\dots, s_{-1}, s_0, s_1, \dots\}$ or $\{\dots, c_{-1}, c_0, c_1, \dots\}$. The cost function is usually of the form $E\{\Phi(y_n)\}$, where $\Phi(y_n)$, which is a function of y_n , is selected such that the cost function has the global minimum points at

$$\{s_n\} = \pm\{\delta[n - n_d]\} \quad \text{for all } n_d = 0, \pm 1, \pm 2, \dots \quad (2.4)$$

A stochastic gradient algorithm is used to minimize the cost function to obtain an on-line equalization algorithm, which

adjusts the k th parameter of the equalizer at time n by

$$\hat{c}_k^{(n+1)} = \hat{c}_k^{(n)} - \mu \phi(y_n) x_{n-k} \quad (2.5)$$

where μ is a small step size, and $\phi(\cdot)$ is the derivative of $\Phi(y_n)$, that is,

$$\phi(y_n) = \Phi'(y_n), \quad (2.6)$$

and it is sometimes called the *prediction error function*.

If an FIR filter is used as the equalizer, then (2.5) can be expressed as

$$\hat{\mathbf{c}}^{(n+1)} = \hat{\mathbf{c}}^{(n)} - \mu \mathbf{x}_n \phi(y_n) \quad (2.7)$$

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

$$\hat{\mathbf{c}}^{(n)} \triangleq [\hat{c}_{-N}^{(n)}, \dots, \hat{c}_0^{(n)}, \dots, \hat{c}_N^{(n)}]^T \quad (2.8)$$

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

$$\mathbf{x}_n \triangleq [x_{n+N}, \dots, x_n, \dots, x_{n-N}]^T. \quad (2.9)$$

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

$$\mathbf{x}_n = \mathcal{H}^T \mathbf{a}_n, \quad (2.10)$$

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

$$\mathcal{H} \triangleq \begin{pmatrix} h_{-M} & 0 & \cdots & 0 \\ \vdots & h_{-M} & \ddots & \vdots \\ h_0 & \vdots & \ddots & 0 \\ \vdots & h_0 & \ddots & h_{-M} \\ h_M & \vdots & \ddots & \vdots \\ 0 & h_M & \ddots & h_0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & h_M \end{pmatrix} \quad (2.11)$$

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

$$\mathbf{a}_n \triangleq [a_{n+(N+M)}, \dots, a_n, \dots, a_{n-(N+M)}]^T. \quad (2.12)$$

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

$$\begin{aligned} y_n &= \mathbf{a}_n^T \hat{\mathbf{s}}^{(n)} \\ &= \mathbf{a}_n^T \mathcal{H} \hat{\mathbf{c}}^{(n)} \end{aligned} \quad (2.13)$$

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

$$\hat{\mathbf{s}}^{(n)} \triangleq \mathcal{H} \hat{\mathbf{c}}^{(n)}. \quad (2.14)$$

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

$$\mathcal{H}\mathbf{c} = \mathbf{e}_{M+N} \quad (2.15)$$

where

$$\mathbf{e}_{M+N} = \underbrace{[0, \dots, 0]}_{M+N}, \underbrace{[1, 0, \dots, 0]}_{M+N}. \quad (2.16)$$

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

III. 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: The prediction error function $\phi(\cdot)$ 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

- a) $E\{\phi(\tilde{y}_n)\mathbf{x}_n\} = \mathbf{0}$, and
- b) $\mathcal{H}^T F \mathcal{H}$ is positive-definite

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

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

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

- 2) For all integers n and k

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

and

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

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

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

and

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

From (2.4) and (2.6),

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

Therefore, by (3.4), $E\{\phi(\tilde{y}_n) x_{n-i}\} = 0$, which is Lemma 1i). Using (2.4) and (2.6), direct calculation yields that

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

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

- 2) If a double-infinite length equalizer is used, then $\{\tilde{c}_i\} = \{\check{h}_i\}$, which is 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

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

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

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

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

IV. STATIC CONVERGENCE ANALYSIS

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

$$\{c_i\} = \{\pm \check{h}_{i-n_d}\} \quad (4.1)$$

for some integer n_d . However, only a 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 $\{\tilde{c}_n: n = -N, \dots, 0, \dots, N\}$ of the cost function near the channel inverse such that $\tilde{y}_n - a_n$ is very small. Using first-order approximation to $\phi(\cdot)$ at a_n , we can prove the following theorem.

Theorem 1: If an FIR equalizer is used to equalize an FIR channel, then at the minimum near the channel inverse, the equalizer coefficient vector $\tilde{\mathbf{c}} \triangleq [\tilde{c}_{-N}, \dots, \tilde{c}_0, \dots, \tilde{c}_N]^T$ can be expressed as

$$\tilde{\mathbf{c}} = f(0) R_f^{-1} \mathbf{h} \quad (4.2)$$

where

$$\mathbf{h} = [0, \dots, 0, h_M, \dots, h_0, \dots, h_{-M}, 0, \dots, 0]^T, \quad (4.3)$$

and

$$R_f \triangleq \mathcal{H}^T F \mathcal{H} \quad (4.4)$$

with

$$F = \text{diag}[\underbrace{f(1), \dots, f(1)}_{M+N}, f(0), \underbrace{f(1), \dots, f(1)}_{M+N}]. \quad (4.5)$$

and

$$f(0) \triangleq \frac{1}{\sigma^2} E\{\phi'(a_n) a_n^2\}, f(1) \triangleq E\{\phi'(a_n)\}. \quad (4.6)$$

Proof: Using the first-order approximation, we have

$$\phi(\tilde{y}_n) = \phi(a_n) + \phi'(a_n)(\tilde{y}_n - a_n). \quad (4.7)$$

According to the lemma,

$$\begin{aligned} E\{\mathbf{x}_n \phi(a_n)\} &= \mathcal{H}^T E\{\mathbf{a}_n \phi(a_n)\} \\ &= \mathbf{0} \end{aligned} \quad (4.8)$$

and

$$E\{\mathbf{x}_n \phi(\tilde{y}_n)\} = \mathbf{0}. \quad (4.9)$$

Consequently, from (4.7)

$$E\{\mathbf{x}_n \phi'(a_n)(\tilde{y}_n - a_n)\} = \mathbf{0}. \quad (4.10)$$

Substituting $\tilde{y}_n = \mathbf{x}_n^T \tilde{\mathbf{c}}$ and $\mathbf{x}_n = \mathcal{H}^T \mathbf{a}_n$ into (4.10), we can obtain that

$$\sigma^2 \mathcal{H}^T \frac{1}{\sigma^2} E\{\mathbf{a}_n \phi'(a_n) \mathbf{a}_n^T\} \mathcal{H} \tilde{\mathbf{c}} = \mathcal{H}^T E\{\mathbf{a}_n \phi'(a_n) \mathbf{a}_n\}. \quad (4.11)$$

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

$$E\{\phi'(a_n) a_m a_l\} = \begin{cases} f(0)\sigma^2 & \text{if } n = m = l \\ f(1)\sigma^2 & \text{if } n \neq m = l \\ 0 & \text{otherwise} \end{cases} \quad (4.12)$$

Therefore, $(1/\sigma^2)E\{\mathbf{a}_n \phi'(a_n) \mathbf{a}_n^T\}$ is a diagonal matrix that can be expressed as

$$\frac{1}{\sigma^2} E\{\mathbf{a}_n \phi'(a_n) \mathbf{a}_n^T\} = F \quad (4.13)$$

and (4.11) can be expressed as

$$\sigma^2 R_f \tilde{\mathbf{c}} = \mathcal{H}^T \mathbf{e}_{M+N} f(0). \quad (4.14)$$

Furthermore, from Lemma 2), $f(0) > 0$ and $f(1) > 0$; hence, F is positive definite; therefore, R_f is also positive definite since \mathcal{H} is of full (column) rank. From (4.14), the coefficient vector at the minimum of the cost function near the channel inverse is determined by (4.2). \square

From the above theorem, the equalizer coefficient vector at the minimum of the cost function near the channel inverse is determined by (4.2).

For the channel with impulse response vector \mathbf{h} , the optimum equalizer (Wiener-Hopf filter) coefficient vector to minimize $E\{(y_n - a_n)^2\}$ is given by [9]

$$\mathbf{c}_o = R^{-1} \mathbf{h} \quad (4.15)$$

where

$$R = \mathcal{H}^T \mathcal{H}. \quad (4.16)$$

Comparing (4.2) and (4.15), we have that the sufficient and necessary condition for $\tilde{\mathbf{c}} = \mathbf{c}_o$ for any FIR channel is

$$\mathcal{H}^T \mathcal{H} = \frac{1}{f(0)} \mathcal{H}^T F \mathcal{H}. \quad (4.17)$$

Since \mathcal{H} is of full column rank for all nonzero \mathbf{h} , (4.17) implies that

$$f(0) = f(1) \quad (4.18)$$

which means that

$$E\{\phi'(a_n) a_n^2\} = E\{\phi'(a_n)\} E\{a_n^2\}. \quad (4.19)$$

For the Sato algorithm [18] decision-directed equalizers [14], [16], $\phi'(a_n) = 1$, and therefore, $\tilde{\mathbf{c}} = \mathbf{c}_o$. For the Godard algorithm [8], $\phi(y) = y(y^2 - r)$ with $r = E\{a_n^4\}/E\{a_n^2\}$; therefore,

$$E\{\phi'(a_n) a_n^2\} = 3\sigma^4 - m_4 \quad (4.20)$$

and

$$E\{\phi'(a_n) a_n^2\} = 2m_4 \quad (4.21)$$

where

$$m_4 = E\{a_n^4\}. \quad (4.22)$$

Hence, if the channel input is binary, (4.19) 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 the equalizer is

$$\begin{aligned} D_f &\triangleq \|\tilde{\mathbf{s}} - \mathbf{e}_{M+N}\|^2 \\ &= \|\mathcal{H} \tilde{\mathbf{c}} - \mathbf{e}_{M+N}\|^2. \end{aligned} \quad (4.23)$$

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.

V. DYNAMIC CONVERGENCE ANALYSIS

When the blind equalization algorithms are implemented using stochastic gradient method, as are most blind equalizers, the blind equalizers will have an extra distortion $\epsilon_n \triangleq \hat{\mathbf{c}}^{(n)} - \tilde{\mathbf{c}}$ due to the gradient noise. Here, we study the stochastic dynamic convergence behavior of blind equalizers when the parameters of the blind equalizers are near the global minimum of the cost function. In our analysis, we will use the *independence assumption*, which assumes that a_n and ϵ_n are statistically independent. Similar assumptions have also been used in the convergence analysis of the LMS algorithm, the decision-directed equalizer, and the Sato algorithm. The literature [9], [16], [26], [27] has given some good justification on the validation of this assumption.

By means of the independence assumption, together with the first-order approximation, we are able to prove the following dynamic convergence theorem.

Theorem 2: Let

$$\tilde{R}_f \triangleq \mathcal{H}^T \tilde{F} \mathcal{H} \quad (5.1)$$

with the largest eigenvalue λ_{\max} , and

$$\tilde{R}_g \triangleq \mathcal{H}^T \tilde{G} \mathcal{H} \quad (5.2)$$

with

$$\tilde{G} = \frac{1}{\sigma^2} E\{\mathbf{a}_n \phi^2(\tilde{y}_n) \mathbf{a}_n^T\}. \quad (5.3)$$

- 1) For any FIR blind equalization algorithm, mean convergence behavior near the global minimum of the cost function satisfies

$$E\{\epsilon_n\} = (I - \mu\sigma^2\tilde{R}_f)^n E\{\epsilon_0\}. \quad (5.4)$$

If the step size μ in iteration formula (2.5) or (2.7) satisfies

$$0 < \mu < \frac{2}{\lambda_{\max}\sigma^2} \quad (5.5)$$

then

$$E\{\mathbf{c}^{(n)}\} \rightarrow \tilde{\mathbf{c}} \quad \text{and} \quad E\{\mathbf{s}^{(n)}\} \rightarrow \tilde{\mathbf{s}}. \quad (5.6)$$

- 2) The equalizer coefficient vector $\mathbf{c}^{(n)} \rightarrow \tilde{\mathbf{c}}$ 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

$$\tilde{R}_f R_\epsilon + R_\epsilon \tilde{R}_f = \mu \tilde{R}_g \quad (5.7)$$

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

Proof: From (2.15), the output of the equalizer can be expressed as

$$\begin{aligned} y_n &= \tilde{y}_n + (y_n - \tilde{y}_n) \\ &= \tilde{y}_n + \mathbf{a}_n^T \mathcal{H} \epsilon_n. \end{aligned} \quad (5.8)$$

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

$$\phi(y_n) = \phi(\tilde{y}_n) + \phi'(\tilde{y}_n) \mathbf{a}_n^T \mathcal{H} \epsilon_n. \quad (5.9)$$

Subtracting both sides of (2.7) by $\tilde{\mathbf{c}}$ and using (5.9), we have that

$$\epsilon_{n+1} = \epsilon_n - \mu(\mathbf{x}_n \phi(\tilde{y}_n) + \mathcal{H}^T \mathbf{a}_n \phi'(\tilde{y}_n) \mathbf{a}_n^T \mathcal{H} \epsilon_n). \quad (5.10)$$

Taking ensemble average on both sides of (5.10) and using the *independence assumption*, we have

$$\begin{aligned} E\{\epsilon_{n+1}\} &= E\{\epsilon_n\} - \mu(E\{\mathbf{x}_n \phi(\tilde{y}_n)\} \\ &\quad + \mathcal{H}^T E\{\mathbf{a}_n \phi'(\tilde{y}_n) \mathbf{a}_n^T\} \mathcal{H} E\{\epsilon_n\}). \end{aligned} \quad (5.11)$$

From Lemma 1i), $E\{\mathbf{x}_n \phi(\tilde{y}_n)\} = 0$. Thus, (5.11) can be simplified into

$$E\{\epsilon_{n+1}\} = (I - \mu\sigma^2 \mathcal{H}^T \tilde{F} \mathcal{H}) E\{\epsilon_n\}. \quad (5.12)$$

Hence,

$$E\{\epsilon_n\} = (I - \mu\sigma^2 \mathcal{H}^T \tilde{F} \mathcal{H})^n E\{\epsilon_0\}. \quad (5.13)$$

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 (5.13) that

$$E\{\Delta \mathbf{s}^{(n)}\} = \mathcal{H} (I - \mu\sigma^2 \mathcal{H}^T \tilde{F} \mathcal{H})^n (\mathcal{H}^T \mathcal{H})^{-1} \mathcal{H}^T E\{\Delta \mathbf{s}^{(0)}\}. \quad (5.14)$$

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

$$\lim_{n \rightarrow \infty} \|E\{\epsilon_n\}\| = 0 \quad (5.15)$$

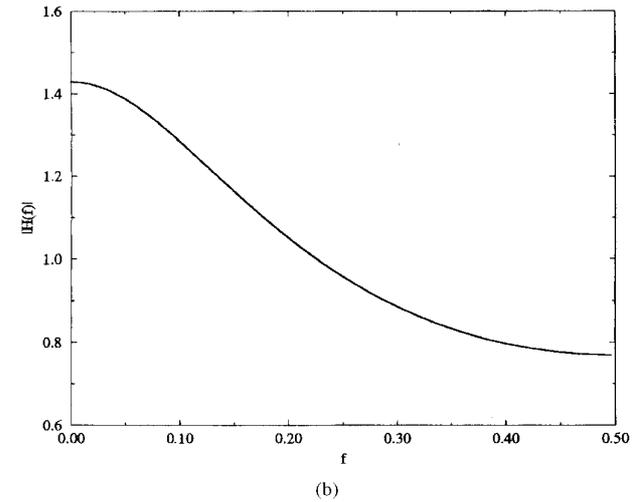
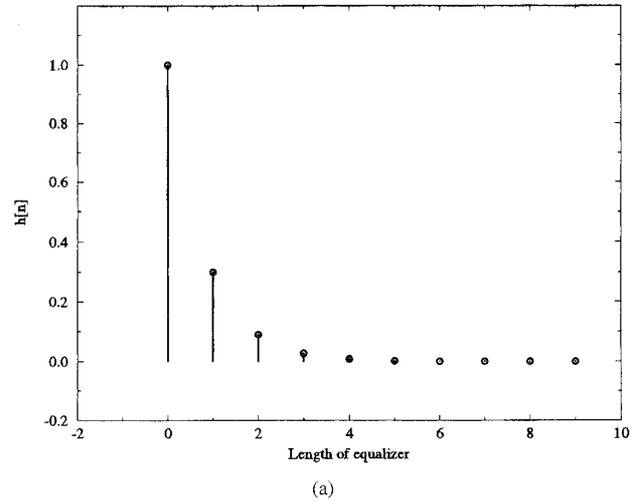


Fig. 2. (a) Impulse response and (b) the frequency response of channel I.

and

$$\lim_{n \rightarrow \infty} \|E\{\Delta \mathbf{s}^{(n)}\}\| = 0 \quad (5.16)$$

if μ satisfies $0 < \mu < 2/\lambda_{\max}\sigma^2$. Hence, Theorem 2.1 has been proven.

From (5.10), we can also obtain

$$E\{\epsilon_{n+1} \epsilon_{n+1}^T\} = E\{\epsilon_n \epsilon_n^T\} - \mu(I_1 + I_2 - \mu I_3) \quad (5.17)$$

where

$$I_1 = E\{(\mathcal{H}^T \mathbf{a}_n \phi(\tilde{y}_n) + \mathcal{H}^T \mathbf{a}_n \phi'(\tilde{y}_n) \mathbf{a}_n^T \mathcal{H} \epsilon_n)^T\}, \quad (5.18)$$

$$I_2 = E\{\epsilon_n (\mathcal{H}^T \mathbf{a}_n \phi(\tilde{y}_n) + \mathcal{H}^T \mathbf{a}_n \phi'(\tilde{y}_n) \mathbf{a}_n^T \mathcal{H} \epsilon_n)^T\} \quad (5.19)$$

and

$$\begin{aligned} I_3 &= E\{(\mathcal{H}^T \mathbf{a}_n \phi(\tilde{y}_n) + \mathcal{H}^T \mathbf{a}_n \phi'(\tilde{y}_n) \mathbf{a}_n^T \mathcal{H} \epsilon_n) \\ &\quad (\mathcal{H}^T \mathbf{a}_n \phi(\tilde{y}_n) \\ &\quad + \mathcal{H}^T \mathbf{a}_n \phi'(\tilde{y}_n) \mathbf{a}_n^T \mathcal{H} \epsilon_n)^T\}. \end{aligned} \quad (5.20)$$

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

$$I_1 = \sigma^2 \tilde{R}_f R_{\epsilon_n}, \quad I_2 = \sigma^2 R_{\epsilon_n} \tilde{R}_f \quad (5.21)$$

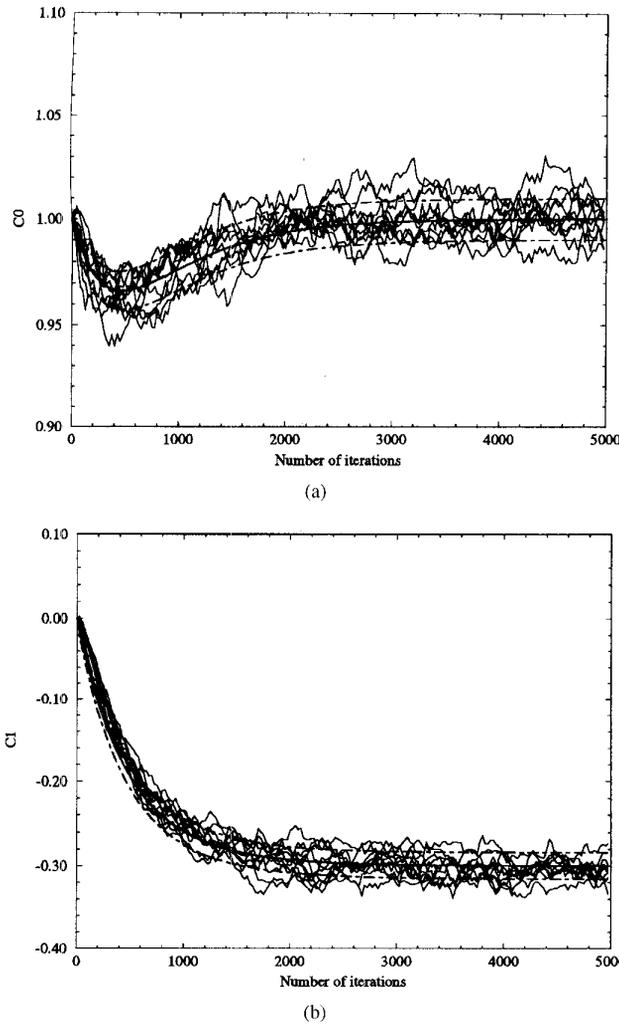


Fig. 3. Ten trials of learning curves of (a) c_0 and (b) c_1 for Sato algorithm using $\mu = 0.002$.

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

$$R_{\epsilon_n} = E\{\epsilon_n \epsilon_n^T\}. \quad (5.22)$$

The dominant term in I_3 is

$$I_3 = \sigma^2 \tilde{R}_g. \quad (5.23)$$

Substituting (5.21) and (5.23) into (5.17), we obtain that

$$R_{\epsilon_{n+1}} = R_{\epsilon_n} - \mu \sigma^2 (\tilde{R}_f R_{\epsilon_n} + R_{\epsilon_n} \tilde{R}_f - \mu \tilde{R}_g). \quad (5.24)$$

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

$$\tilde{R}_f R_\epsilon + R_\epsilon \tilde{R}_f = \mu \tilde{R}_g. \quad (5.25)$$

Then, (5.24) is equivalent to

$$R_{\epsilon_{n+1}} - R_\epsilon = \frac{1}{2} \Lambda (R_{\epsilon_n} - R_\epsilon) + \frac{1}{2} (R_{\epsilon_n} - R_\epsilon) \Lambda \quad (5.26)$$

where

$$\Lambda = I - 2\mu \sigma^2 \tilde{R}_f. \quad (5.27)$$

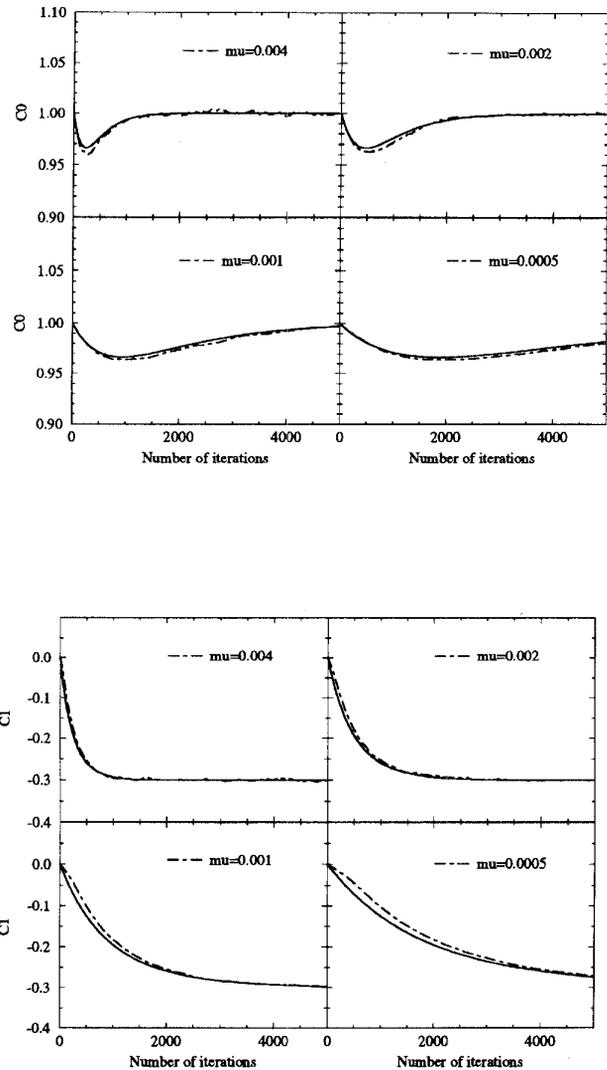


Fig. 4. Average of learning curves: Theoretical ones (solid lines) and simulated ones (dot-dash lines) based on 100 ensemble trials for Sato algorithm using different μ .

Hence,

$$\|R_{\epsilon_{n+1}} - R_\epsilon\| \leq \|\Lambda\| \|R_{\epsilon_n} - R_\epsilon\|. \quad (5.28)$$

Therefore,

$$\|R_{\epsilon_n} - R_\epsilon\| \leq \|\Lambda\|^n \|R_{\epsilon_0} - R_\epsilon\|. \quad (5.29)$$

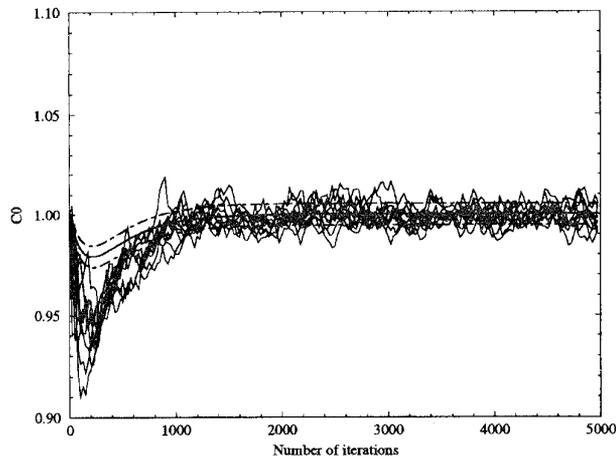
If $0 < \mu < 1/\lambda_{\max} \sigma^2$, then $\|\Lambda\| < 1$, and

$$\lim_{n \rightarrow \infty} \|R_{\epsilon_n} - R_\epsilon\| = 0. \quad (5.30)$$

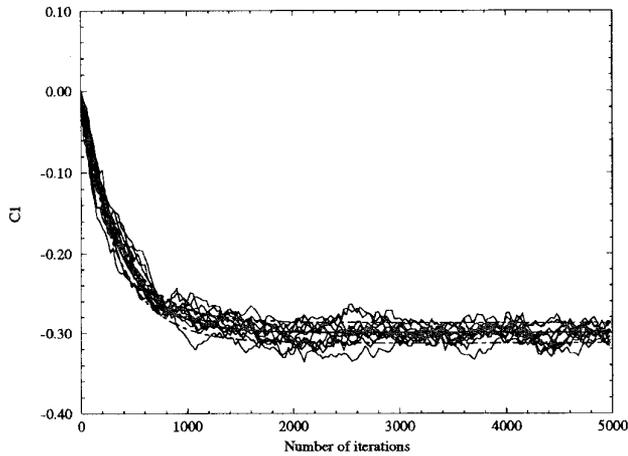
Therefore,

$$\lim_{n \rightarrow \infty} R_{\epsilon_n} = R_\epsilon. \quad (5.31)$$

Since $R_\epsilon \neq 0$, $c^{(n)} \rightarrow \tilde{c}$ is not consistent. \square



(a)



(b)

Fig. 5. Ten trials of learning curves of (a) c_0 and (b) c_1 for Godard algorithm using $\mu = 0.002$.

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

$$\begin{aligned} D_g &\triangleq E\{\|s - \tilde{s}\|^2\} \\ &= E\{\|\mathcal{H}\epsilon\|^2\} \\ &= \text{tr}[\mathcal{H}^T R_\epsilon \mathcal{H}] \\ &= \text{tr}[RR_\epsilon]. \end{aligned} \quad (5.32)$$

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

$$\tilde{R}_f \approx R_f, \quad \tilde{R}_g \approx R_g \quad (5.33)$$

where we have used the definitions

$$R_g \triangleq \mathcal{H}^T G \mathcal{H} \quad (5.34)$$

and

$$g(0) \triangleq \frac{1}{\sigma^2} E\{\phi^2(a_n) a_n^2\}, \quad g(1) = E\{\phi^2(a_n)\}, \quad (5.35)$$

$$G \triangleq \text{diag}[\underbrace{g(1), \dots, g(1)}_{M+N}, \underbrace{g(0), g(1), \dots, g(1)}_{M+N}]. \quad (5.36)$$

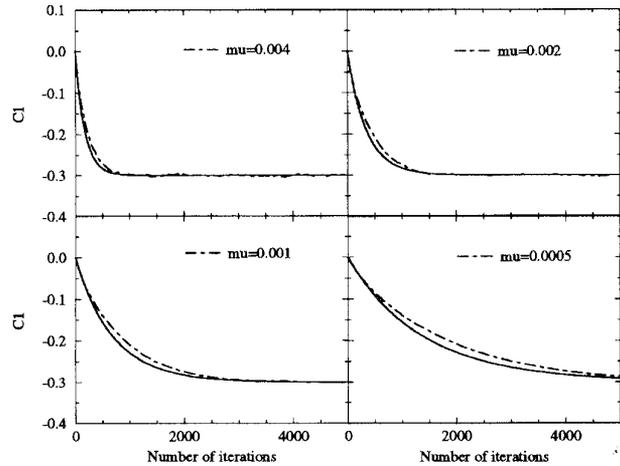
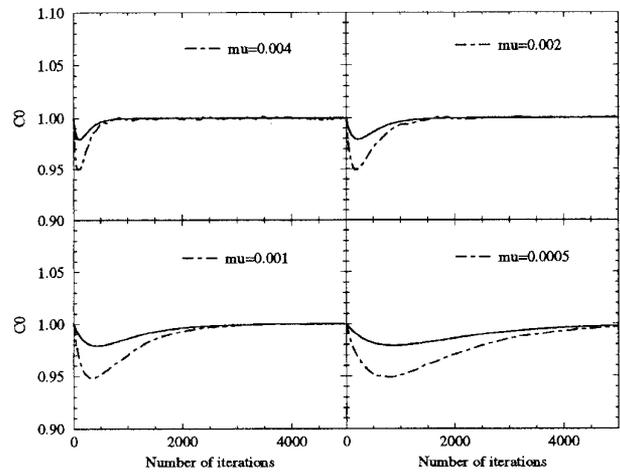


Fig. 6. Average of learning curves: Theoretical ones (solid lines) and simulated ones (dot-dash lines) based on 100 ensemble trials for Godard algorithm using different μ .

In this case, (5.7) becomes

$$R_f R_\epsilon + R_\epsilon R_f = \mu R_g. \quad (5.37)$$

For the blind equalization algorithms with $f(0) = f(1)$, $R_f = f(1)R$. Using (5.37), we have

$$D_g = \frac{1}{2f(1)} \mu \text{tr}[R_g] \approx \frac{1}{2} \mu (2N + 1) \frac{g(1)}{f(1)}. \quad (5.38)$$

For those blind equalizers with $f(0) \neq f(1)$, (5.38) can also be used to approximately estimate the average distortion introduced by gradient noise. According to (5.38), D_g is proportional to the step size μ and the length of equalizer N . On the other hand, however, step size affects the convergence speed of equalizers, i.e., the larger the μ , 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 tradeoff between these two factors.

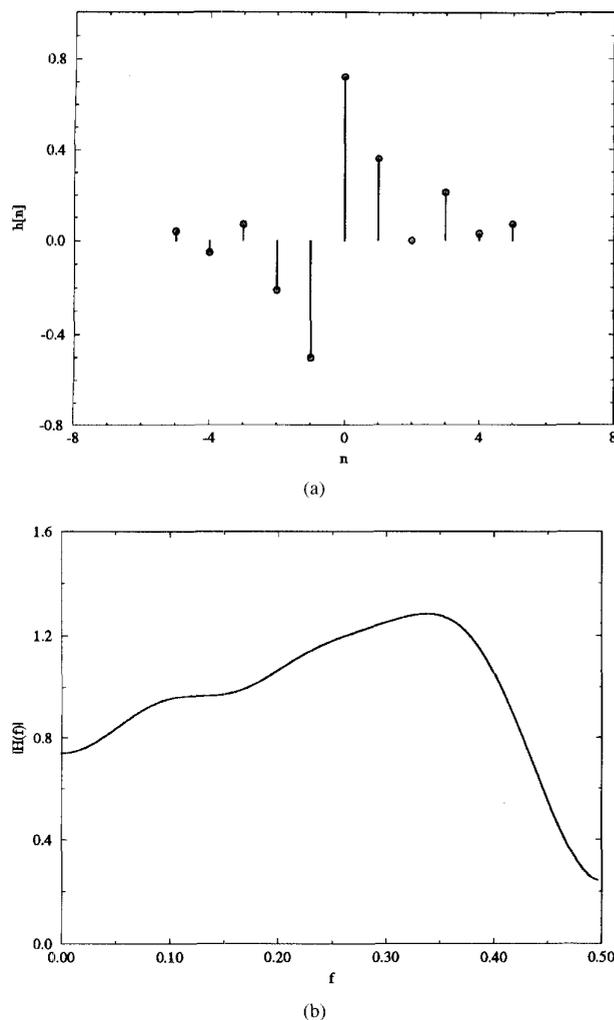


Fig. 7. (a) Impulse response and (b) the frequency response of channel II.

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, and another is D_g in (5.38) due to the gradient noise. Once the step size of a blind equalizer is set, there must be an optimum length that can be found 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.

VI. 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 Fig. 2. An FIR equalizer with coefficients c_0 and c_1 is used to compensate for the channel distortion. The initial value of

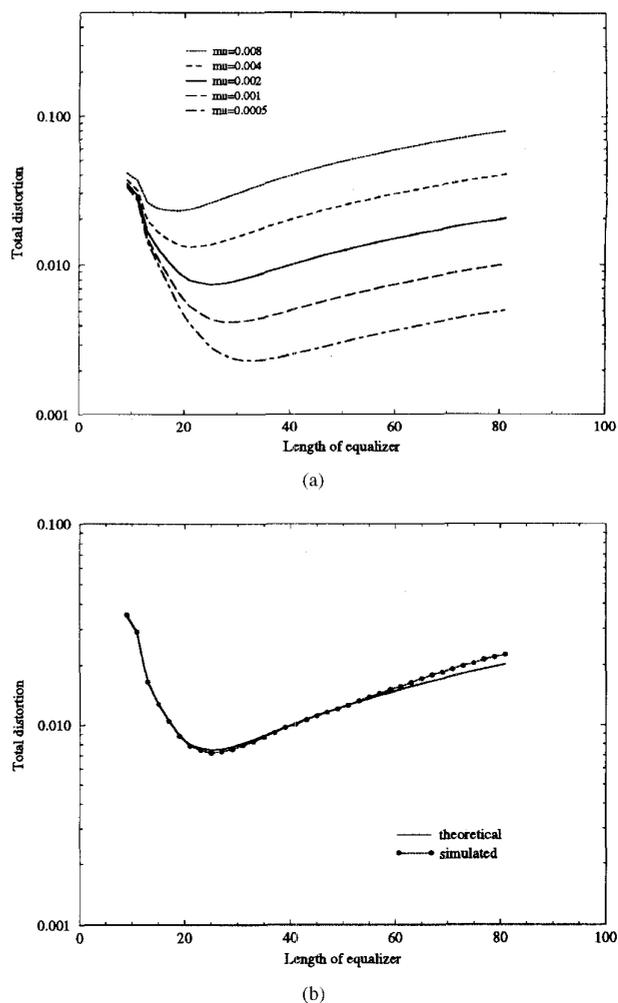


Fig. 8. Total distortion of equalized system (a) theoretical results for different step size μ . (b) Simulation results for $\mu = 0.002$, using Sato algorithm.

the equalizer coefficient vector is set to be

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

The Sato algorithm [18] is first used to adjust the coefficients of the equalizer. When the step size $\mu = 0.002$, there are 10 trials of learning curves of $c^{(n)}$, as shown in Fig. 3. In this figure, the thick solid line is the theoretical average learning curve calculated from (5.13), and the thick dot-dash lines are the theoretical one-standard-deviation lines determined by (5.26). According to this figure, 10 trials of learning curves are almost within one standard deviation of the theoretical average learning curves for Sato algorithm. Fig. 4 demonstrates the ensemble averages of learning curves for different step-sizes based on 100 trials. From Fig. 4, our theoretical results fit to the simulation results very well for the Sato algorithm.

Similar simulations have also been done for the Godard algorithm [8]. The simulation results are shown in Figs. 5 and 6, which also confirms 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 Fig. 7, which is

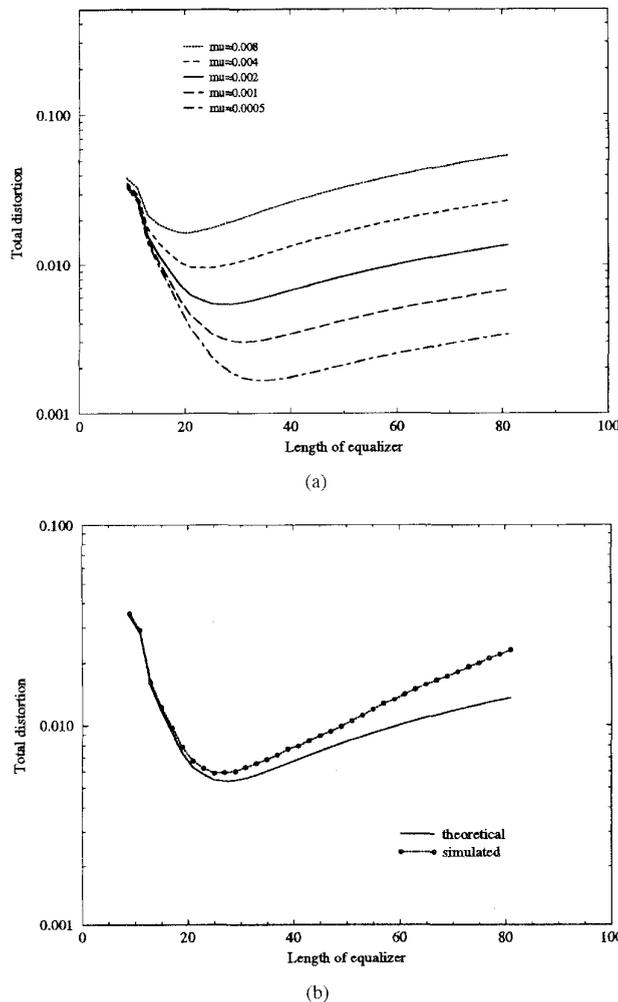


Fig. 9. Total distortion of equalized system (a) theoretical results for different step size μ . (b) Simulation results for $\mu = 0.002$, using Godard algorithm.

a typical telephone channel [20]. The center-tap initialization strategy [7] is used for the blind equalization algorithm.

When the Sato algorithm is used, the theoretical relationship between the total distortion and the length of the equalizer for different step sizes is illustrated in Fig. 8(a), which indicates that the optimum length of the Sato equalizer for this channel is between 15 and 25, depending on the step size. Fig. 8(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 Fig. 9 for the Godard algorithm. Because $g(1)/f(1)$ for the Godard algorithm (0.169) is less than that for the Sato algorithm (0.250) for four-level PAM input, the Godard algorithm should have less distortion than the Sato algorithm should according to (5.38), which is confirmed by comparing Figs. 8 and 9.

VII. CONCLUSION

We have studied the static and dynamic convergence behavior of adaptive blind equalizers in the PAM digital communication systems based on the first-order approximation to

the cost function of blind algorithms under the independence assumption. Most of the analysis results presented here can be extended to QAM digital communication systems. Our analysis result indicates that for a given channel and step size, there is an optimal length for an equalizer to minimize the intersymbol interference. The results imply that a longer length blind equalizer does not necessarily outperform a shorter one, which is contrary to what is conventionally conjectured. The theoretical results are confirmed by computer simulation examples.

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. Automat. Contr.*, vol. AC-25, pp. 385-399, June 1980.
- [2] A. Benveniste and M. Goursat, "Blind equalizers," *IEEE Trans. Commun.*, vol. COM-32, pp. 871-882, Aug. 1982.
- [3] R. Cusani and A. Laurenti, "Convergence analysis of the CMA blind equalizer," *IEEE Trans. Commun.*, vol. 43, pp. 1304-1307, Feb./Mar./Apr. 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. Commun.*, vol. 39, pp. 1313-1327, Sept. 1991.
- [5] ———, "On the (non)existence of undesirable equilibria of Godard blind equalizer," *IEEE Trans. Signal Processing*, vol. 40, pp. 2425-2432, Oct. 1992.
- [6] ———, "Local convergence of the Sato blind equalizer and generalization under practical constraints," *IEEE Trans. Inform. Theory*, vol. 39, pp. 129-144, Jan. 1993.
- [7] G. J. Foschini, "Equalization without altering or detecting data," *AT&T Tech. J.*, vol. 64, pp. 1885-1911, Oct. 1985.
- [8] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COM-28, pp. 1867-1875, Nov. 1980.
- [9] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1991.
- [10] R. A. Kennedy, B. O. Anderson, and R. R. Bitmead, "Stochastic dynamics of blind decision feedback equalizer adaptation," in *Proc. Adaptive Syst. Contr. Signal Processing 1989 IFAC Symp.*, 1990, pp. 579-584.
- [11] Y. Li and Z. Ding, "Convergence analysis of finite length blind adaptive equalizer," *IEEE Trans. Signal Processing*, vol. 43, pp. 2120-2129, Sept. 1995.
- [12] ———, "Global convergence of fractionally spaced Godard adaptive equalizers," *IEEE Trans. Signal Processing*, vol. 44, pp. 818-826, Apr. 1996.
- [13] Y. Li, K. J. R. Liu, and Z. Ding, "Length and cost dependent local minima of blind channel equalizers," *IEEE Trans. Signal Processing*, this issue, pp. 2726-2735.
- [14] G. Macchi and E. Eweda, "Convergence analysis of self-adaptive equalizers," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 161-176, Mar. 1984.
- [15] V. J. Mathews and H. Cho, "Improved convergence analysis of stochastic gradient adaptive filters using the sign algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 450-454, Apr. 1987.
- [16] J. E. Mazo, "Analysis of decision-directed equalizer convergence," *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1857-1876, Dec. 1980.
- [17] G. Picchi and G. Prati, "Blind equalization and carrier recovery using a 'stop-and-go' decision-directed algorithm," *IEEE Trans. Commun.*, vol. COM-35, pp. 877-887, Sept. 1987.
- [18] Y. Sato, "A method of self-recovering equalization for multi-level amplitude modulation," *IEEE Trans. Commun.*, vol. COM-23, pp. 679-682, June 1975.
- [19] C. K. Chan and J. J. Shynk, "Stationary points of the constant modulus algorithm for real gaussian signals," *IEEE Trans. Signal Processing*, vol. 38, pp. 2176-2180, 1990.
- [20] J. Proakis, *Digital Communications* (2nd ed.). New York: McGraw-Hill, 1989.
- [21] J. R. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 349-372, Apr. 1983.

- [22] J. R. Treichler, V. Wolff, and C. R. Johnson, Jr., "Observed misconvergence in the constant modulus adaptive algorithm," in *Proc. 25th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, 1991, pp. 663-667.
- [23] S. Verdu, B. D. O. Anderson, and R. A. Kennedy, "Blind equalization without gain identification," *IEEE Trans. Inform. Theory*, vol. 39, pp. 292-297, Jan. 1993.
- [24] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. Inform. Theory*, vol. 36, pp. 312-321, Mar. 1990.
- [25] J. K. Tugnait, O. Shalvi, and E. Weinstein, "Comments on 'New criteria for blind deconvolution of nonminimum phase systems (channels),' " *IEEE Trans. Inform. Theory*, vol. 38, pp. 210-213, Jan. 1992.
- [26] V. Weerackody, S. A. Kassam, and K. R. Laker, "Convergence analysis of an algorithm for blind equalization," *IEEE Trans. Commun.*, vol. 39, pp. 856-865, June 1991.
- [27] B. Windrow *et al.*, "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. IEEE*, vol. 64, pp. 1151-1162, Aug. 1976.

Ye Li, for photograph and biography, please see p. 2735 of this issue of this TRANSACTIONS.

K. J. Ray Liu, for photograph and biography, please see p. 2735 of this issue of this TRANSACTIONS.