Robust Blind Learning Algorithm for Nonlinear Equalization Using Input Decision Information

Published in:
IEEE Transactions on Neural Networks and Learning Systems

DOI:
10.1109/TNNLS.2015.2399499

Document Version
Peer reviewed version

Link to publication in the UWA Research Repository

General rights
Copyright owners retain the copyright for their material stored in the UWA Research Repository. The University grants no end-user rights beyond those which are provided by the Australian Copyright Act 1968. Users may make use of the material in the Repository providing due attribution is given and the use is in accordance with the Copyright Act 1968.

Take down policy
If you believe this document infringes copyright, raise a complaint by contacting repository-lib@uwa.edu.au. The document will be immediately withdrawn from public access while the complaint is being investigated.

Download date: 29. Nov. 2017
Robust Blind Learning Algorithm for Nonlinear Equalization Using Input Decision Information

Lu Xu, Member, IEEE, Defeng (David) Huang, Senior Member, IEEE, Yingjie Jay Guo, Fellow, IEEE

Abstract—In this paper, we propose a new blind learning algorithm, namely, the Benveniste-Goursat input-output-decision (BG-IOD), to enhance the convergence performance of neural network based equalizers for nonlinear channel equalization. In contrast to conventional blind learning algorithms, where only the output of the equalizer is employed for updating system parameters, the BG-IOD exploits a new type of extra information, the input decision information obtained from the input of the equalizer, to mitigate the influence of the nonlinear equalizer structure on parameters learning, thereby leading to improved convergence performance. We prove that, with the input decision information, a desirable convergence capability that the output symbol error rate (SER) is always less than the input SER if the input SER is below a threshold, can be achieved. Then, the BG soft-switching technique is employed to combine the merits of both input and output decision information, where the former is used to guarantee SER convergence and the latter is to improve SER performance. Simulation results show that the proposed algorithm outperforms conventional blind learning algorithms such as stochastic quadratic distance (SQD) and dual mode constant modulus algorithm (DM-CMA), in terms of both convergence performance and SER performance, for nonlinear equalization.

Index Terms—SER convergence, Benveniste-Goursat input-output-decision (BG-IOD), input decision information, blind learning, nonlinear equalization.

I. INTRODUCTION

CHANNEL equalization plays an important role in modern digital communication systems. Typically, the physical channel introduces many impairments such as inter-symbol interference (ISI), nonlinear distortion, noise, etc. An equalizer is then necessary to be employed to mitigate those effects upon the transmitted signals. An equalizer can be characterized as linear or nonlinear by its structure. The most popular linear equalizer is the linear transversal equalizer (LTE) [1]. While for nonlinear equalization, neural networks have received a great amount of attention for their superior performance in a variety of applications [2]. Due to their simple structures and good performance, the functional link artificial neural networks (FLANNs) [3]–[7], which are suitable for real-time signal processing, are selected to accomplish nonlinear equalization in this paper.

This work was supported in part by Australian Research Council’s Discovery Projects DP1093000 and DP140100522. Part of this work has been presented in WCSP 2013.

L. Xu and D. Huang are with School of Electrical, Electronic and Computer Engineering, University of Western Australia, WA 6009, Australia (e-mail: lu.xu@uwa.edu.au; david.huang@uwa.edu.au).

Y. J. Guo is with University of Technology, Sydney (UTS), Australia (email: jay.guo@uts.edu.au).

When a reference (training) sequence is not available for the equalizer, blind learning algorithms must be employed. In practical applications, one of the main challenges of blind learning algorithms is to maintain convergence of symbol error rate (SER), demonstrating that the output SER of the equalizer is lower than the input SER. Actually, if the algorithm converges falsely, the SER of the equalizer output may be worse than that of the input (as will be shown in this paper). In other words, the SER performance can even be degraded due to the employment of the equalizer.

For blind learning of the coefficients of linear equalizers, the well-known least mean square (LMS) algorithm was commonly employed for its computational efficiency. However, to guarantee SER convergence, LMS requires a low level of output decision errors at the initial acquisition state, which is a hard task for most applications [8]. Two classes of schemes have been proposed to cope with this problem: methods based on extra information and methods based on dual mode switching techniques. By exploiting extra information other than the output decision used in LMS, some excellent algorithms have been developed. For instance, the stochastic quadratic distance (SQD) algorithm, in which the probability density function (pdf) of the equalizer output was employed to approximately match the estimated constellation pdf, was investigated in [9]. [10] extracted information about the time structure as well as the statistical distribution for coded transmitted signals and presented an algorithm based on the correntropy function. Also, the well-known Sato algorithm [11] and Godard algorithms [12] can be regarded as using the high order moments of the output. Indeed, [13] proved that Godard and constant modulus algorithm (CMA) for fractionally spaced equalizers (FSEs) can be globally convergent under a finite-length channel satisfying a length-and-zero condition. Using the dual mode switching methods, an acquisition algorithm (like CMA) is employed to ensure SER convergence during the initial acquisition state and then switched to a tracking algorithm (like LMS) when the output decision error rate is sufficiently low to achieve an accurate final solution. Some examples are the classical Benveniste-Goursat (BG) algorithm [14], the dual mode CMA (DM-CMA) [15] and its Stop-and-Go (SAG) extension SAG-DM-CMA [16].

Compared to linear equalization, the convergence problem of nonlinear equalization is more challenging, as there exists more unpredictable false minima on the performance surfaces, owing to the inherent nonlinear structures employed [17]. Actually, even for supervised learning with a reference sequence
on nonlinear equalizers, substantial efforts have been made towards achieving good convergence results [17]–[25]. In [17], [18], an error-entropy minimization algorithm considering the entropy of the output error was proposed and its generalized pattern was presented in [19]. For equalizers with feedforward neural networks, the extreme learning machine (ELM) [20], [21] and its variants [22]–[25], which tend to reach the smallest training error, have been investigated. While for blind learning of nonlinear equalization, the convergence problem gets even worse and as a result it becomes more difficult to provide a solution. With the nonlinear structure and without a prior supervised training, the learning process can be easily trapped into a false minimum, which then results in a totally false convergence. However, to the best of our knowledge, no blind learning algorithm that guarantees SER convergence has been proposed for nonlinear equalization so far. This paper aims at filling in the niche, and presents an alternative blind learning algorithm with SER convergence for nonlinear cases.

To guarantee SER convergence, a new type of extra information, the input decision information, is proposed. In contrast to the output decision information (as used in LMS) coming from the equalizer output signals, this input decision information comes directly from the input signals and is obviously not influenced by the equalizer configurations. Then, to demonstrate the convergence capability of employing the new information, an input-directed (ID) algorithm by replacing the output decision in LMS with the input decision information only is theoretically analyzed, and an SER convergence result is proved.

The rest of the paper is organized as follows. Section II presents a blind learning model for nonlinear equalization using FLANN and formulates ID using the input decision information and several other blind algorithms. In Section III, some assumptions are listed and then a theoretical analysis for the SER convergence property of ID is provided. A new blind learning algorithm, BG-IOD, is proposed in Section IV. Simulation results are presented in Section V. Finally, Section VI concludes the paper.

II. BLIND LEARNING MODEL FOR NONLINEAR EQUALIZATION

We consider a blind learning model for nonlinear equalization using FLANN as shown in Fig. 1. At the i-th step, let $X_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,N})^T$ be the input signal of dimension $N$, $\Phi_i = (\phi_{i,1}, \phi_{i,2}, \ldots, \phi_{i,M})^T$ of dimension $M$ the expanded signal after input transformation, $H_i = (h_{i,1}, h_{i,2}, \ldots, h_{i,M})^T$ the parameter vector, $y_i$ the output signal, and $d_i = \text{dec}(y_i)$ the output decision with the detection function $\text{dec}(\cdot)$. In Fig. 1, $F(\cdot)$ is the input transformation function with $N$ inputs and $M$ outputs, and $g(\cdot)$ is the activation function. $F(\cdot)$ could be Chebyshev polynomials [3], Legendre polynomials [4], or some other expansion functions. And the well-known hyperbolic tangent activation function $g(x) = \tanh(x)$ is usually adopted in [30], [31]. Actually, the nonlinear nature of FLANN is embodied in the nonlinear $F(\cdot)$ and $g(\cdot)$ employed.

Let $a_i$ be the desired signal of $X_i$, the goal of the equalizer is to generate $y_i$ using $X_i$ and recursively updated $H_i$ to approximate $a_i$ with the lowest or even vanishing error, as the index $i$ increases. Clearly, there are two procedures to achieve this goal, the feedforward procedure to generate $y_i$ and the contributions of this paper are listed as follows.

1) The convergence property of employing the input decision information only is theoretically analyzed, and an SER convergence result is proved.
2) The BG-IOD learning algorithm in [29] is further improved to be more practical. Now there are no parameters that need to be manually tuned in the proposed algorithm.
3) For simulations, more linear and nonlinear distortion scenarios are considered to evaluate the convergence performance and SER performance of the proposed algorithm.
feedback procedure to update $H_i$. The feedforward part can be illustrated by the following equation

$$y_i = g(\Phi_i^TH_i).$$

(1)

The feedback part, which is a realization of some blind learning algorithm, can be written as

$$H_{i+1} = H_i + ue_i g(\Phi_i^TH_i)\Phi_i, \quad i = 1, 2, \ldots$$

(2)

where $u$ is the learning rate and $e_i$ is determined by the blind algorithm employed.

For FLANN with fourth order Chebyshev expansion and $g(x) = \tanh(x)$, we have

$$y_i = \tanh(\Phi_i^TH_i)$$

(3)

and

$$H_{i+1} = H_i + ue_i(1 - y_i^2)\Phi_i,$$

(4)

where

$$\Phi_i = (1, T_1(x_{i1}), T_2(x_{i1}), T_3(x_{i1}), T_4(x_{i1}), \ldots, T_1(x_{iN}), T_2(x_{iN}), T_3(x_{iN}), T_4(x_{iN}))^T$$

with $T_0(x) = 1$, $T_1(x) = x$ and $T_{n+1}(x) = 2x T_n(x) - T_{n-1}(x)$.

Note that the constant element 1.0 in $\Phi_i$ is used for the compensation of a DC bias if it exists, and the factor $(1 - y_i^2)$ in (4) comes from the derivative of $\tanh()$.

It is shown in Fig. 1 that for existing blind algorithms, such as LMS and DM-CMA, the only output decision $d_i$ and the output signal $y_i$ are employed. For the LMS algorithm, $e_i$ in (2) is

$$e_i = d_i - y_i,$$

(5)

and for DM-CMA

$$e_i = \begin{cases} 
(d_i^2 - y_i^2)y_i, & \text{if } |y_i - d_i| < \delta_e \\
(R - y_i^2)y_i, & \text{otherwise}
\end{cases}$$

(6)

where $\delta_e$ is empirically determined and

$$R = \frac{E[a_i^4]}{E[a_i^2]}.$$}

Since the output signals are affected by the parameter vectors and equalizer structures, it is reasonable to conclude that blind algorithms based on information coming from output signals may lead to poor convergence performance for complicated equalization structures. To solve this problem, we propose to use a new type of information, that is, the input decision information. In Fig. 1, $b_i = \text{dec}(x_{i,l})$ with $l \in \{1, \ldots, N\}$ refers to the new information. Note that $x_{i,l}$ is the element of the input signal $X_i$ that is used for detecting $a_i$ for $i = 1, 2, \ldots$, if no equalizer exists. In this paper, we choose $l = (N + 1)/2$, as it is common to employ the middle element of each $X_i$ for detection in most practical applications without equalization. It’s clear that compared to existing blind algorithms, the algorithm proposed in this paper employ not only the output decision $d_i$ but also this input decision $b_i$. As the input signal $X_i$, is obviously not affected by the equalizer configurations, the blind learning algorithms using this input decision $b_i$ can maintain good convergence property. To verify this claim, we first propose an input-directed (ID) algorithm, in which only the input decision $b_i$ is employed

$$e_i = b_i - y_i.$$  

(7)

And we show in the next Section that an SER convergence property exists for the ID algorithm, which can well demonstrate the convergence capability of employing the input decision information.

### III. CONVERGENCE PROPERTY OF ID

The case of perfect modeling [8], [32] is considered in the analysis, that is, for the expanded sequence $P \overset{\Delta}{=} \{\Phi_i\}_{i=1}^{\infty}$, there exists at least one optimal solution $H_\ast$, such that

$$a_i = g(\Phi_i^TH_\ast), \quad i = 1, 2, \ldots.$$  

(8)

This special case can be satisfied. For example, [33] proved that there exists a converged parameter vector $H_\ast$, such that an FLANN with sufficient input dimension can approximate the mapping of any input-output set with arbitrarily small error. Furthermore, the universal approximation analysis conducted in [34] and [35] shows that, using a proper parameter vector, an FLANN can approximate any continuous target function with vanishing output deviation.

#### A. Assumptions

A list of assumptions for the analysis are established as follows.

1) If the dimension of the expanded sequence $P$ is $N_P$ (i.e., the maximum number of linear independent elements in $P$ is $N_P$), then there exists a finite constant $K < \infty$, such that the dimension of any subsequence $P_m \overset{\Delta}{=} \{\Phi_i\}_{i=m+1}^{m+K}$ for $m = 0, 1, \ldots, N_P$.

2) The parameter sequence $\{H_i\}_{i=1}^{\infty}$ and the optimal solution $H_\ast$ are uniformly upper bounded, i.e. there exists a $B_H \in \mathbb{R}$, such that $\|H_i\| \leq B_H < \infty$ for $i = 1, 2, \ldots$ and $\|H_\ast\| \leq B_H < \infty$.

3) The expanded sequence $\{\Phi_i\}_{i=1}^{\infty}$ is uniformly upper bounded, i.e. there exists a $B_\Phi \in \mathbb{R}$, such that $\|\Phi_i\| \leq B_\Phi < \infty$ for $i = 1, 2, \ldots$.

4) The activation function $g(\cdot)$ is continuously differentiable, and $g'(\cdot)$ is positive and uniformly upper bounded, i.e. there exists a $B_{g'} \in \mathbb{R}$, such that $0 < g'(x) \leq B_{g'} < \infty$.

5) For all bounded $x$ of $|x| \leq B_x < \infty$, $g(x)$ is lower bounded by a small positive number, i.e. $g(x) \geq D_g(B_x) > 0$, where $D_g(B_x) \overset{\Delta}{=} \min(g(x) : |x| \leq B_x)$.

The notation $\|\|_\|_\|$ here is the Euclidean norm for vectors and the matrix norm induced by the Euclidean norm for matrices as

$$\|A\| = \max \|AX\|, \quad \forall A \in \mathbb{R}^{m \times n}, X \in \mathbb{R}^n.$$
1. Assumption 2 appears slightly restrictive at first sight. However, using the boundedness analysis of [28] and Assumption 1, some improvement methods can be used for achieving the upper boundedness of \( \{ \mathbf{H}_i \}_{i=1}^{\infty} \) [28].

For Assumptions 3, with the input signals \( \{ X_i \}_{i=1}^{\infty} \) and a specific input transformation function \( F(\cdot) \), the upper boundedness of the expanded sequence can be readily justified in practice. And it is not difficult to see that the widely used hyperbolic tangent activation function \( g(x) = \tanh(x) \), is consistent with Assumptions 4 and 5.

**B. Convergence property of ID**

We consider the convergence of SER in this paper, which guarantees that the output SER \( p_o \) is lower than the input SER \( p_i \). Let \( c_i \) denote the input decision error

\[
c_i = b_i - a_i.
\]

As \( b_i \) is obtained from the input signal \( X_i \), it follows from (9) that \( c_i \) includes all of the effects due to ISI, nonlinear distortion, noise, and other impairments introduced by the transmission channel. As these effects together are commonly characterized as an additive white noise \([10],[18],[36]–[38]\), we can model the input decision error sequence \( \{ c_i \}_{i=1}^{\infty} \) as an ergodic random process. And the input SER \( p_i \) can be defined as follows

\[
p_i = \lim_{L \to \infty} \frac{1}{L} \sum_{j=1}^{L} \eta_j
\]

with

\[
\eta_j = \begin{cases} 
0, & c_j = 0 \\
1, & \text{otherwise.}
\end{cases}
\]

The output SER \( p_o \) can be calculated when the final solution \( \mathbf{H}_f = \lim_{i \to \infty} \mathbf{H}_i \) updated by an algorithm of (2) is available. We denote the final output error as follows

\[
\omega_i = g(\Phi_i^T \mathbf{H}_f) - a_i, \quad i = 1, 2, \ldots
\]

Using the Differential Mean Value Theorem [39] and (8), \( \omega_i \) can be rewritten as follows

\[
\omega_i = g(\Phi_i^T \mathbf{H}_f) - g(\Phi_i^T \mathbf{H}_s)
\]

\[
= g'(\Phi_i^T \mathbf{H}_f) \Phi_i^T (\mathbf{H}_f - \mathbf{H}_s)
\]

\[
= g'(\Phi_i^T \mathbf{H}_f) \Phi_i^T \mathbf{V}_f
\]

where \( \Phi_i^T \mathbf{H}_f \) satisfies

\[
\min(\Phi_i^T \mathbf{H}_f, \Phi_i^T \mathbf{H}_s) \leq \Phi_i^T \mathbf{H}_f \leq \max(\Phi_i^T \mathbf{H}_f, \Phi_i^T \mathbf{H}_s),
\]

and

\[
\mathbf{V}_f = \mathbf{H}_f - \mathbf{H}_s
\]

is the final parameter estimation error.

It is clear that after equalization, the final output error sequence \( \Delta = \{ \omega_i \}_{i=1}^{\infty} \) contains all the impairments including the residual ISI, nonlinear distortion, and noise. Assuming that none of the impairments dominates, according to the Central Limit Theorem [40], we can model \( \omega \) as a zero-mean white Gaussian process with variance \( \sigma^2_\omega \). Actually, this Gaussian approximation has been widely used in the literature \([10],[18],[36]–[38]\). The pdf of this Gaussian process \( \xi(\omega) \) can then be written as

\[
\xi(\omega) = \frac{1}{\sqrt{2\pi\sigma_\omega}} \exp\left\{ -\frac{\omega^2}{2\sigma^2_\omega} \right\},
\]

and the output SER \( p_o \) can be calculated as follows

\[
p_o = \int_{-\infty}^{\frac{1}{2}} \xi(\omega) d\omega + \int_{\frac{1}{2}}^{\infty} \xi(\omega) d\omega
\]

where \( \delta \) is the minimum distance for detection of the symbol set \( \Gamma = \{ \tau_1, \ldots, \tau_S \} \) (\( S \) is the set size) from which the desired sequence \( \{ a_i \}_{i=1}^{\infty} \) take values, and is defined as follows

\[
\delta = \min_{1 \leq i, j \leq S, i \neq j} |\tau_i - \tau_j|.
\]

In contrast to the minimum distance \( \delta \), the average distance between elements in the symbol set \( \Gamma \) is also used in the analysis (see Appendix A), and can be calculated as

\[
\delta_0 = \frac{1}{S(S-1)} \sum_{1 \leq i, j \leq S, i \neq j} |\tau_i - \tau_j|.
\]

For example, with a symbol set \( \Gamma = \{-0.3, -0.1, 0.1, 0.3\} \) of 4-PAM signals, we have \( \delta = 0.1 \) and \( \delta_0 \approx 0.333 \). Note that (15) is a generic representation of \( p_o \) for analysis, some changes to the range of integration may be necessary for a specific symbol set.

The variance \( \sigma^2_\omega \) of \( \omega \) is now an important number to calculate the output SER \( p_o \). From (12), it is clear that \( \omega \) can be investigated through the analysis of the final parameter estimation error \( \mathbf{V}_f \). We define the parameter estimation error at step \( i \) as

\[
\mathbf{V}_i = \mathbf{H}_i - \mathbf{H}_s
\]

and we have

\[
\lim_{i \to \infty} \mathbf{V}_i = \mathbf{V}_f.
\]

Then the recently proposed extended difference equation [28] of \( \mathbf{V}_i \), which is applicable to the case with nonlinear activation functions \( g(\cdot) \), can be employed here.

Using (2) and (7), the parameter sequence \( \{ \mathbf{H}_i \}_{i=1}^{\infty} \) can be recursively updated by ID algorithm as follows

\[
\mathbf{H}_{i+1} = \mathbf{H}_i + u(b_i - g(\Phi_i^T \mathbf{H}_i)) g'(\Phi_i^T \mathbf{H}_i) \Phi_i, \quad i = 1, 2, \ldots
\]

With \( \mathbf{V}_i \), we can rewrite the recursion (20) as follows

\[
\mathbf{V}_{i+1} = \mathbf{V}_i + u(b_i - g(\Phi_i^T \mathbf{H}_i)) g'(\Phi_i^T \mathbf{H}_i) \Phi_i, \quad i = 1, 2, \ldots
\]

Using the Differential Mean Value Theorem and (8), we have

\[
g(\Phi_i^T \mathbf{H}_i) = g(\Phi_i^T \mathbf{H}_s) + g'(\Phi_i^T \mathbf{H}_i)(\Phi_i^T \mathbf{H}_i - \Phi_i^T \mathbf{H}_s)
\]

\[
= a_i + g(\Phi_i^T \mathbf{H}_i) \Phi_i \mathbf{V}_i
\]

where \( \Phi_i^T \mathbf{H}_i \) satisfies

\[
\min(\Phi_i^T \mathbf{H}_i, \Phi_i^T \mathbf{H}_s) \leq \Phi_i^T \mathbf{H}_i \leq \max(\Phi_i^T \mathbf{H}_i, \Phi_i^T \mathbf{H}_s),
\]

and

\[
\mathbf{V}_f = \mathbf{H}_f - \mathbf{H}_s
\]
Inserting (22) into (21), we now have an extended difference equation as follows
\[ V_{i+1} = (I - ut_i A_i) V_i + uc_i g \left( \Phi_i^T H_i \right) \Phi_i \] (23)
where
\[ t_i = g \left( \Phi_i^T H_i \right) g \left( \Phi_i^T H_i \right) \]
and
\[ A_i = \Phi_i \Phi_i^T. \] (25)

Using the extended difference equation (23), after \( i \) recursions, we can obtain the following expression of \( V_{i+1} \)
\[ V_{i+1} = F_i + S_i, \quad i = 1, 2, \ldots \] (26)
where
\[ F_i = U_{0,i} V_1 \]
and \( S_i \) is the transition matrix defined as
\[ S_i = \sum_{j=1}^{i} c_j g \left( \Phi_j^T H_j \right) U_{j,i} \Phi_j, \] (28)
and \( U_{j,i} \) is the transition matrix defined as
\[ U_{j,i} = \begin{cases} (I - ut_i A_i) \ldots (I - ut_{j+1} A_{j+1}), & j < i \\ I, & j = i. \end{cases} \] (29)

We use Lemma 1 and Lemma 2 of [28] to support our analysis in this paper. For convenience, both Lemmas are copied here.

**Lemma 1:** Under Assumptions 3 and 4, let \( Z \) be an \( N \)-dimensional vector, and \( \rho_i \) the correlation coefficient between \( \Phi_i \) and \( Z \), i.e. \( \rho_i = \langle \Phi_i, Z \rangle / (\| \Phi_i \| \| Z \|) \) (\( \langle \cdot, \cdot \rangle \) denotes inner product) for \( \Phi_i \neq 0 \) and \( Z \neq 0 \), and \( \rho_i = 1 \) for \( \Phi_i = 0 \) or \( Z = 0 \). There exists a positive constant
\[ u_0 = \frac{1}{B g^2 B_k^2}, \] (30)
such that for all \( 0 < u \leq u_0 \), we have
\[ \| (I - ut_i A_i) Z \| \leq \sqrt{1 - u \gamma^2} t_i \| Z \| \]
where \( t_i \) is defined in (24), \( A_i \) is defined in (25), and \( \gamma = \lambda_i = \| \Phi_i \| \).

**Proof:** See [28].

**Lemma 2:** Under Assumptions 2-5, let \( Z \) be an upper bounded \( N \)-dimensional vector that can be linearly combined by some elements of the sequence \( \{ \Phi_i \}_{i=0}^{M} \) with a finite size of \( i - j \). Then for all \( 0 < u \leq u_0 \), we have
\[ \| U_{j,i} Z \| \leq \sqrt{1 - u \gamma^2} \| Z \| \]
where
\[ \gamma = \alpha \Omega g^2 D \Phi^2 > 0 \] (31)
with \( \Omega \) being the lower bound of \( g(x) \) when \( |x| \leq B \Phi \Phi H_i \) and the minimal value of the norms of non-zero elements of the expanded sequence \( P \), and \( \alpha \in (0, 1] \) a constant.

**Proof:** See [28].

For \( 0 < u \leq u_0 \), from (30) and (31), it can be seen that \( 0 < u \gamma \leq 1 \). Using Lemmas 1 and 2 and (26), the upper bound of the final parameter estimation error \( V_j \) can be obtained as follows.

**Lemma 3:** Under Assumptions 1-5, for all \( 0 < u \leq u_0 \), the final parameter estimation error \( V_j \) obtained by ID algorithm is upper bounded as follows
\[ \| V_j \| \leq KB g p_i \delta_0 C(u) \] (32)
where
\[ C(u) = \frac{u}{\sqrt{1 - u \gamma}} + u. \] (33)

**Proof:** See Appendix A.

Using (12) to (15) and Lemma 3, we can now derive an important Theorem for the convergence property of ID.

**Theorem 1:** Under Assumptions 1-5, if the input SER \( p_i \) satisfies the following inequality
\[ f(p_i) = p_i^2 \ln(p_i) > D \] (34)
where
\[ D = -\frac{2g^2 \delta^2}{8K^2 B g^4 B_k^2 (1 - \rho_i^2)} \] (35)
then for all \( 0 < u \leq u_0 \), the output SER \( p_o \) obtained by the ID algorithm can always be less than the input SER \( p_i \).

**Proof:** Using (12), the variance \( \sigma^2 \) of the final output error sequence \( \omega = \{ \omega_i \}_{i=1}^{M} \) can be obtained as
\[ \sigma^2 \omega = \lim_{M \to \infty} \frac{1}{M} \sum_{i=1}^{M} \| \omega_i \|^2 \]
\[ = \lim_{M \to \infty} \frac{1}{M} \sum_{i=1}^{M} \| g \left( \Phi_i^T H_i \right) V_f \|^2 \]
\[ \leq B g^2 B_k^2 \| V_f \|^2. \] (36)
Given the upper bound of the final parameter estimation error \( V_f \) obtained in Lemma 3, and plugging (32) into (36), we have
\[ \sigma^2 \omega \leq \sigma^2 m = \Delta \frac{K^2 B g^4 B_k^2 (1 - \rho_i^2) C(u)}{2 \sigma^2}. \] (37)

Given (14), the output SER \( p_o \) can be calculated as
\[ p_o \approx \int_{-\omega}^{\omega} \xi(\omega) d\omega + \int_{\omega}^{\infty} \xi(\omega) d\omega \]
\[ = \int_{-\omega}^{\omega} \frac{1}{\sqrt{2\pi\sigma}} \exp \left( -\frac{\omega^2}{2\sigma^2} \right) d\omega \]
\[ = 2Q \left( \frac{\delta}{2\sigma^2} \right) \] (38)
where \( Q(\cdot) \) is the Q-function. Then, using the following improved exponential-type bound of the Q-function [41, 42]
\[ Q(x) \leq \frac{1}{2} \exp \left( -\frac{x^2}{2} \right), \quad x > 0, \]
we have
\[ p_o \approx 2Q\left(\frac{\delta}{2\sigma_o}\right) \leq \exp\left\{ -\frac{\delta^2}{8\sigma_m^2} \right\} \]
\[ \leq \exp\left\{ -\frac{\delta^2}{8\sigma_m^2} \right\} \]  
(39)
where the last inequality follows from (37).

It can be seen from (39) that \( p_o \) is upper bounded by
\[ \exp\{-\delta^2/(8\sigma_m^2)\} \]. Now, if we have
\[ \exp\{-\frac{\delta^2}{8\sigma_m^2}\} < p_i, \]  
(40)
then the output \( p_o \) obtained by ID algorithm will be less than the input \( p_i \). Using (37), and after some routine rearrangement, (40) changes into
\[ p_i^2 \ln(p_i) > -\frac{\delta^2}{8K^2B^4_gB^2_{\Phi}\delta_0^2C^2(u)}. \]  
(41)

Now we further investigate \( C(u) \) of (33). We have
\[ C'(u) = \frac{(6 - 2u\gamma)\sqrt{1 - u\gamma} - 6 + 5u\gamma}{(4 - 2u\gamma)\sqrt{1 - u\gamma} - 4 + 4u\gamma}. \]  
(42)
Then from (42), it is not difficult to verify that \( C(u) \) has a stationary point at \( u = 3/(4\gamma) \), and \( C'(u) > 0 \) for \( u < 3/(4\gamma) \), \( C'(u) < 0 \) for \( u > 3/(4\gamma) \). Then, as
\[ 0 < u \leq u_0 = \frac{1}{B^2_gB^2_{\Phi}} \leq \frac{1}{\alpha\Omega^2_gD^2_{\Phi}} = \frac{1}{\gamma}, \]
we can prove that
\[ \frac{2}{\gamma} \leq C(u) \leq \frac{9}{4\gamma} \]  
(43)
where the upper bound is obtained when \( u = 3/(4\gamma) \), and the lower bound is obtained when \( u = 1/\gamma \).

Inserting (43) into (41), we obtain the following inequality
\[ p_i^2 \ln(p_i) > D \]  
(44)
where
\[ D = \frac{-2\gamma^2\delta^2}{81K^2B^4_gB^2_{\Phi}\delta_0^2}. \]  
(45)
This completes the proof.

For a specific input sequence \( \{X_i\}_{i=1}^\infty \) and equalization structure, \( D \) in (35) is a negative constant. Now we investigate the existence of a valid input \( p_i \in [0, 1] \) satisfying the inequality (34). It is not difficult to verify that
\[ f(p_i) \mid p_i = 1 = f(p_i) \mid p_i = 0 = 0. \]
\[ f(p_i) \] for \( p_i \in [0, 1] \) is shown in Fig. 2. It can be seen that \( f(p_i) \) has a stationary point at \( p_i = 0.6065 \), and \( f'(p_i) < 0 \) for \( p_i \in (0, p_h) \) and \( f'(p_i) > 0 \) for \( p_i \in (p_h, 1) \). Also, as \( f(p_i) \) is a continuous function, it is clear that given a negative constant \( D \), there exists a threshold \( p_h \in (0, 1) \), such that all \( p_i < p_h \) satisfy the inequality (34). Then, using Theorem 1, it can be concluded that \( p_o < p_i \) as long as \( p_i < p_h \).

To further illustrate the convergence property, the value of the threshold \( p_h \) for a specific example is provided as follows.

Considering a nonlinear \( g(x) = tanh(x) \) and normalized 4-PAM signals, we have \( B_{\Phi} = D_{\Phi}, \delta_0 \approx 3.335 \) and we assume \( 10\Omega^2_g = B_g \cdot K = 50, \alpha = 0.5 \). Then by numerical calculation, we can obtain \( p_h = 1.25E - 6 \).

From Fig. 2, it is clear that, for \( p_i \in (0, p_h) \), the larger the constant \( D \), the smaller the threshold \( p_h \). The constant \( D \), which is obtained by using several bounds (as shown in the proof), is actually an upper bound. Accordingly, the value of \( p_h \) derived by Theorem 1 is a lower bound for the threshold of guaranteeing the SER convergence. As a result, \( p_h \) derived by Theorem 1 is a lower bound to the worst scenario, thereby it is conservative compared to the threshold values obtained on an average basis in our simulation results to be presented in Section V. Indeed, our simulation results show that the threshold values could be large (larger than 1E-1 for our simulation scenarios), which then makes the SER convergence property suitable for many practical applications.

IV. PROPOSED BLIND ALGORITHM

It follows from the SER convergence analysis that, the blind algorithm ID using the input decision information only can guarantee SER convergence for nonlinear equalization. However, simulation results in Section V will also show that the ID has a poor SER performance. On the other side, it is commonly approved that algorithms employing the output decision information can obtain good SER performance as long as no false convergence exists [14]–[16]. In this paper, we propose to combine the merits of input and output decision information for blind learning, where the former is used to ensure SER convergence and the latter is to improve SER performance. And a well-known soft-switching technique, the BG, is employed to accomplish this combination. As a result, we propose a new blind algorithm, the BG-IOD, for nonlinear equalization, where \( e_i \) in (2) is given as follows
\[ e_i = (d_i - y_i) + |d_i - y_i| (b_i - y_i). \]  
(46)

It’s clear that for BG-IOD, in the initial acquisition state, when the output error \( |d_i - y_i| \) is large, the second term of (46)
of amplifiers used in the transmitting system, and NL also can occur in a channel due to saturation. Note that the tap-centered initial conditions are employed here.

Numerous experiments have been carried out to give the best results of the learning rate and some other parameters of these blind algorithms. The learning rate for all algorithms is set to 1E-4. For SQD, we set \( a = 0.5, b = 0.0, \alpha = 0.999995 \) and \( E_1 = 1.0 \) for simulations. For DM-CMA, \( \delta_c \) in (6) is set to 0.0333.

The input SER and output SER of the ID algorithm are first investigated to verify the SER convergence property of ID. Then, convergence behaviors and SER performance of the proposed BG-IOD are studied and compared with existing algorithms like LMS, SQD and DM-CMA, for nonlinear equalization. In each case, 100 independent runs each consisting of \( 10^7 \) iterations for each SNR value, are carried out. To obtain the steady state value of SER, SER value is calculated using the output decision information, there exists a convergence region of parameter vectors and once the recursively updated parameter vector \( \mathbf{H} \) falls into this region, good SER performance can be obtained. However, in practical applications, due to the effects of noise and wrong detection, \( \mathbf{H} \) updated by LMS often wanders far away from the convergence region, which then results in a false convergence. In this sense, the ID part of the proposed algorithm, which has a desirable SER convergence property, may constrain the wandering range of \( \mathbf{H} \) around the convergence region, and consequently let the LMS part take effect. In Section V, we will use extensive simulation results, which show that the proposed algorithm can always obtain good SER performance, to empirically verify this claim.

V. SIMULATION RESULTS

Monte Carlo simulations have been carried out for nonlinear FLANN equalization using ID, BG-IOD, LMS, SQD and DM-CMA. In our study, the transmission channel consists of linear and nonlinear components. The linear component is characterized by the following impulse response [43]

\[
g(i) = \begin{cases} 
\frac{1}{2} \left[ 1 + \cos \left( \frac{2\pi i}{\Lambda} \right) \right], & i = 1, 2, 3 \\
0, & \text{otherwise}. 
\end{cases} 
\]  

\tag{47}

The parameter \( \Lambda \) in (47) is set to 2.9 or 3.3, which produces eigenvalue ratio (EVR) value of 6.08 or 21.71. Then we can obtain the following normalized transfer functions in the z-transform form

\[
EV R = 2.9 : 0.209 + 0.995z^{-1} + 0.209z^{-2} \\
EV R = 3.3 : 0.304 + 0.903z^{-1} + 0.304z^{-2}.
\]

Considering the nonlinear component which is added to the output of the linear component, the following models are employed [4], [44]

\[
NL = 1 : Non_k = \tanh(Lin_k) \\
NL = 2 : Non_k = Lin_k + 0.2Lin_k^2 - 0.1Lin_k^3.
\]

Note that \( NL = 1 \) may occur in a channel due to saturation of amplifiers used in the transmitting system, and \( NL = 2 \) is arbitrarily selected.

The transmitted message is with a 4-PAM signal and in the form of \( \{-0.3, -0.1, +0.1, +0.3\} \). A zero mean white Gaussian noise is added to the channel output.

A 7-dimensional FLANN with 4th order Chebyshev expansion and hyperbolic tangent activation function, is employed. In this FLANN, the input pattern is expanded into a 29-dimensional pattern (including the constant bias 1.0) by Chebyshev polynomials.

ID algorithm is given by (7), BG-IOD algorithm is by (46), and DM-CMA algorithm is by (6). SQD algorithm used here is consistent with that in [9] except some changes of parameters. Note that the tap-centered initial conditions are employed here.

Further investigation of the proposed algorithm is provided as follows. The algorithm contains two parts, the part \( (d_i - y_i) \) from the LMS algorithm and the part \( (b_i - y_i) \) from the ID algorithm. It was proved in [8] that, for the LMS algorithm using the output decision information, there exists a convergence region of parameter vectors and once the recursively updated parameter vector \( \mathbf{H} \) falls into this region, good SER performance can be obtained. However, in practical applications, due to the effects of noise and wrong detection, \( \mathbf{H} \) updated by LMS often wanders far away from the convergence region, which then results in a false convergence. In this sense, the ID part of the proposed algorithm, which has a desirable SER convergence property, may constrain the wandering range of \( \mathbf{H} \) around the convergence region, and consequently let the LMS part take effect. In Section V, we will use extensive simulation results, which show that the proposed algorithm can always obtain good SER performance, to empirically verify this claim.

Fig. 3 and Fig. 4 show the input SER \( p_i \) and the output SER \( p_o \) obtained by the ID algorithm with \( NL = 1, 2 \) for
SNR from 12 to 28 dB, using linear distortion of $EVR = 2.9$ and $EVR = 3.3$, respectively. It is clear that for these various nonlinear equalization scenarios, the output SER $p_o$ obtained by the ID algorithm can always be less than the input SER $p_i$, which verifies the SER convergence property of ID.

And it is further shown that the threshold $p_h$ obtained is larger than 1E-1 in these cases, which is larger than that calculated by Theorem 1. However, it can be also seen that the ID algorithm which only employs the input decision information, can not achieve a desired SER performance.

The input SER $p_i$ and the output SER $p_o$ obtained by three existing blind algorithms, LMS, SQD, and DM-CMA, with $EVR = 3.3$, $NL = 2$, for SNR from 12 to 28 dB, are shown in Fig. 5. It is clear that due to false convergence, the output SER $p_o$ obtained can be larger than the input SER $p_i$, and as a result the SER performance is degraded after the employment of the equalizer.

**B. Convergence Behaviors**

The convergence behaviors of the four blind learning algorithms, LMS, SQD, DM-CMA, and BG-IOD, are depicted in Fig. 6 and Fig. 7 at $EVR = 3.3$ for SNR of 20 dB, with nonlinear distortion of $NL = 1$ and $NL = 2$, respectively. Among 100 runs, the times of the final SER falling into specified intervals are shown, which illustrate the performance of convergence of SER using blind algorithms.

It follows from Fig. 6 and Fig. 7 that, the proposed BG-IOD can always guarantee SER convergence for both cases. While for the case of $NL = 1$ in Fig. 6, DM-CMA can ensure SER convergence but maintains a poor SER performance, LMS leads to false convergence occasionally. For the case of $NL = 2$ in Fig. 7, it is clear that the learning processes by all three existing algorithms, LMS, SQD, and DM-CMA, are all trapped into false convergence.

**C. SER Performance**

SER performance of the five blind learning algorithms, ID, BG-IOD, LMS, SQD, and DM-CMA, and the supervised learning algorithm (48), is plotted in Fig. 8 and Fig. 9 at $EVR = 3.3$ for SNR from 12 to 28 dB, with nonlinear distortion of $NL = 1$ and $NL = 2$, respectively. For both cases, it can be seen that the SER performance of the proposed BG-IOD can reach the same level of that of the supervised learning algorithm with a reference sequence. And it is clear that BG-IOD is superior to all the three existing algorithms, LMS, SQD, and DM-CMA, in terms of SER performance for nonlinear equalization. Note that LMS, SQD and DM-CMA always suffer from false convergence due to nonlinearity contained in the channel mode $NL = 2$, and it is verified.
again that the ID algorithm is not sufficient to achieve good SER performance.

VI. Conclusion

In order to carry out robust and high-performance blind learning for nonlinear equalization, we have proposed a new blind algorithm, the BG-IOD. In the new algorithm, a new type of extra information, the input decision information, which is not affected by the nonlinear equalizer structure, has been presented and theoretically proved having SER convergence capability. Then the BG soft-switching technique has been employed to accomplish the transition from the acquisition state using the input decision information to ensure good convergence, to the tracking state using the output decision information to improve SER performance. Simulation results have demonstrated that, for nonlinear equalization with FLANN, the proposed BG-IOD algorithm can always guarantee SER convergence and maintain excellent SER performance, which outperforms existing blind algorithms.

Appendix A

Proof of Lemma 3: From (19), we have
\[ \|V_f\| = \lim_{i \to \infty} \|V_i\|. \] (49)

Note that \( V_{i+1} = F_i + S_i \) [see (26)], then the upper bound of \( V_f \) is obtained by
\[ \|V_f\| \leq \lim_{i \to \infty} \|F_i\| + \lim_{i \to \infty} \|S_i\|. \] (50)

Let us investigate the limits of \( \|F_i\| \) and \( \|S_i\| \) as \( i \to \infty \) separately.

A. The limit of \( \|F_i\| \).

Let \( i = nK + k \) for \( n = 0, 1, \ldots \) and \( k = 1, \ldots, K \), we have
\[ F_i = U_{0,i}V_1 = U_{nK,nK+k}U_{(n-1)K,nK} \cdots U_{0,K}V_1. \] (51)

We first investigate \( V_1 = H_1 - H_s \). The initial condition \( H_1 \) can be separated into two components as \( H_1 = H_1^\perp + H_1^\parallel \), where \( H_1^\perp \) is the component of \( H_1 \) which can be linearly combined by some elements of the expanded sequence \( P \), and \( H_1^\parallel \) is the component that is orthogonal to \( P \). For \( H_s \), it can be seen that the optimal solutions in (8) may not be unique, as we have
\[ g(\Phi_i^T H_s) = g(\Phi_i^T (H_s + H_1^\parallel)), \quad i = 1, 2, \ldots \]
where \( H_1^\parallel \) is a vector orthogonal to the expanded sequence \( P (H_1^\parallel = 0 \text{ when } N_p = M) \). It is not difficult to verify that there exist optimal solutions which can be linearly combined by some elements of \( P \). For the sake of clarification, we let \( H_s^\parallel \) denote one of those optimal solutions.

Without loss of generality, we choose the optimal solution as
\[ H_s = H_s^\parallel + H_1^\perp. \]

Then we can obtain \( V_1 \) as
\[ V_1 = H_1' - H_s^\parallel. \] (52)

From (52), it is clear that \( V_1 \) can also be linearly combined by some elements of \( P \). From Assumption 2, the upper bound of \( V_1 \) can be obtained as
\[ \|V_1\| \leq \|H_1\| + \|H_s\| \leq 2BH < \infty. \] (53)

For (51), in each transition \( U_{mK,(m+1)K} \) of \( m = 0, \ldots, n-1 \), the vectors of the expanded subsequence \( P_m = \{\Phi_{i,m+1}K\}_{i=0}^{nK+1} \) have been used. Assumption 1 states that the dimension of each subsequence \( P_m \) is equal to that of \( P \). Then, it can be concluded that \( V_1 \) can be linearly combined by some elements of \( P_m \) of \( m = 0, \ldots, n-1 \) as well. Using Lemma 2, we have
\[ \|U_{0,K}V_1\| \leq \sqrt{1 - u\gamma} \|V_1\| \]
where \( 0 < u\gamma \leq 1 \).
From the proof of Lemma 1 in [28], we can see that \((I - ut_jA_j)\) is obtained by reducing the projection of \(Z\) on \(\Phi_i\). Then \(U_{0,K}V_1\) is obtained by reducing the projection on each vector of \(P_0 = \{\Phi_j\}_{j=1}^{K}\) from \(V_1\). Then it is not difficult to verify that \(U_{0,K}V_1\) can also be linearly combined by some elements of \(P_1 = \{\Phi_{j}1_{j=K+1}^{2K}\}.\) And we have
\[
\|U_{K,2K}U_{0,K}V_1\| \leq \sqrt{1 - u\gamma}\|U_{0,K}V_1\| \leq \left(\sqrt{1 - u\gamma}\right)^2\|V_1\|.
\]

In a similar vein, and note that \(\|U_{n,K,n,K+k}\| \leq 1\) from Lemma 1, we have
\[
\|F_i\| \leq \left(\sqrt{1 - u\gamma}\right)^n\|V_i\|.
\]

Since \(0 \leq \sqrt{1 - u\gamma} < 1\) and \(\|V_i\|\) is upper bounded by \(2B_H\) in (53), the limit of \(\|F_i\|\) as \(i \to \infty\) can be obtained as
\[
0 \leq \lim_{i \to \infty} \|F_i\| \leq \lim_{n \to \infty} \left(\sqrt{1 - u\gamma}\right)^n\|V_i\| = 0,
\]
and consequently
\[
\lim_{i \to \infty} \|F_i\| = 0. \tag{54}
\]

B. The limit of \(\|S_i\|\).

Recall that \(i = nK + k\) for \(n = 1, 2, \ldots\) and \(k = 1, \ldots, K\), \(S_i\) of (28) can be rewritten as
\[
u K \sum_{j=1}^{K} \sum_{m=0}^{n-1} c_{mK+j} t_j (\Phi^T_{mK+j} H_{mK+j}) M_{n,m}^{(j)} M_{mK+j} + u K \sum_{j=1}^{K} c_{mK+j} t_j (\Phi^T_{mK+j} H_{mK+j}) U_{nK+j, nK+K} \Phi_{nK+j} \tag{55}
\]

where
\[
M_{n,m}^{(j)} = \begin{cases} U_{(n-1)K,nK} \cdots U_{(m+1)K,(m+2)K} U_{mK+j,(m+1)K}, & m = 0, 1, \ldots, n - 2 \smallskip \ U_{mK+j,(m+1)K}, & m = n - 1. \end{cases} \tag{56}
\]

We first investigate the upper bound of the first term of (55). It follows from Lemma 1 and Assumption 4 that this term is upper bounded by
\[
u K \sum_{j=1}^{K} \sum_{m=0}^{n-1} c_{mK+j} M_{n,m}^{(j)} M_{mK+j} \tag{57}
\]

In (57), when \(m = n - 1\), from Lemma 1 and Assumption 3, we have
\[
\|M_{n,m}^{(j)}\| = \|U_{(n-1)K,nK} \Phi_{(n-1)K+j}\| \leq \|\Phi_{(n-1)K+j}\| \leq \|B_{\Phi}||
\]

Then, for \(m = 0, \ldots, n - 2\), it is clear that \(U_{mK+j,(m+1)K} \Phi_{mK+j}\) is obtained by reducing the projections of \(\Phi_{mK+j}\) on \(\{\Phi_{mK+j+1}, \ldots, \Phi_{(m+1)K}\}\). As \(\Phi_{mK+j} \in P\) can be linearly combined by some elements of \(P\), and
\[
\{\Phi_{mK+j+1}, \ldots, \Phi_{(m+1)K}\} \subseteq P,
\]

it is not difficult to verify that \(U_{mK+j,(m+1)K} \Phi_{mK+j}\) can be linearly combined by some elements of \(P\), and consequently by some elements of \(P_m\) of \(m = 0, 1, \ldots\) as well. Then, using Lemma 2 and Assumption 3, we have
\[
\|U_{(m+1)K,(m+2)K} U_{mK+j,(m+1)K} \Phi_{mK+j}\| \leq \sqrt{1 - u\gamma} \|U_{mK+j,(m+1)K} \Phi_{mK+j}\| \leq \sqrt{1 - u\gamma} \|\Phi_{mK+j}\| \leq \sqrt{1 - u\gamma} B_{\Phi}, \quad m = 0, 1, \ldots, n - 2
\]
\noindent where \(0 < \sqrt{1 - u\gamma} \leq 1\).

In the same vein as above, using Lemma 2, the upper bound of \(M_{n,m}^{(j)} \Phi_{mK+j}\) can be derived as
\[
\|M_{n,m}^{(j)} \Phi_{mK+j}\| \leq \left(\sqrt{1 - u\gamma}\right)^{n-m-1} B_{\Phi}, \quad m = 0, \ldots, n - 1. \tag{58}
\]

Then, (57) can be upper bounded by
\[
u K \sum_{j=1}^{K} \sum_{m=0}^{n-1} c_{mK+j} M_{n,m}^{(j)} M_{mK+j} \leq \nu K \sum_{j=1}^{K} B_{\Phi} \sum_{m=0}^{n-1} \left(\sqrt{1 - u\gamma}\right)^{n-m-1} \sum_{j=1}^{K} c_{mK+j}. \tag{59}
\]

As the sequence \(\{c_i\}_{i=1}^{\infty}\) is modeled as an ergodic process in this analysis. For \(c_i = \beta_i - \alpha_i\) and \(\beta_i, \alpha_i \in \mathbb{G}\), using the average distance \(\delta_0\) of elements in the symbol set \(\mathbb{G}\) (defined by (17)), if \(K\) is large enough, we have
\[
\sum_{j=1}^{K} c_{mK+j} \approx Kp_{j}\delta_0. \tag{60}
\]

Inserting (60) into (59), we have
\[
u K \sum_{j=1}^{K} \sum_{m=0}^{n-1} c_{mK+j} M_{n,m}^{(j)} M_{mK+j} \leq \nu K \sum_{j=1}^{K} B_{\Phi} p_{j}\delta_0 \sum_{m=0}^{n-1} \left(\sqrt{1 - u\gamma}\right)^{n-m-1} \sum_{j=1}^{K} c_{mK+j} \approx \nu K \sum_{j=1}^{K} B_{\Phi} p_{j}\delta_0 \left(\frac{1}{1 - \sqrt{1 - u\gamma}}\right)^n. \tag{61}
\]

From Lemma 1 and using the same analysis for the first term, the second term of (55) is upper bounded by
\[
u K \sum_{j=1}^{K} \sum_{m=0}^{n-1} c_{mK+j} M_{n,m}^{(j)} M_{mK+j} \leq \nu K \sum_{j=1}^{K} B_{\Phi} p_{j}\delta_0 \left(\frac{1}{1 - \sqrt{1 - u\gamma}}\right)^n. \tag{62}
\]

Combining (61) and (62) together, as \(0 < \sqrt{1 - u\gamma} \leq 1\), we can
obtain the limit of $\|S_i\|$ as follows
\[
\lim_{i \to \infty} \|S_i\| \\
\leq \lim_{n \to \infty} \left( uKB_g B_p \rho_0 + \frac{1}{1 - \sqrt[1-u]} \right)^n + uKB_g B_p \rho_0 \delta_0 C(u)
\]
(63)

where
\[
C(u) = \left( u + \frac{u}{1 - \sqrt[1-u]} \right).
\]
(64)

Given the limit of $\|F_i\|$ in (54) and the limit of $\|S_i\|$ in (63), from (50), we can get an upper bound of $\|F_j\|$ as follows
\[
\|F_j\| \leq KB_g B_p \rho_0 \delta_0 C(u).
\]
(65)

This completes the proof.

**REFERENCES**


