A New Learning Algorithm For Blind 2-channel Identification

Jie Luo, Qing Zhao and Xieting Ling
Fudan University, Shang Hai, 200433, P.R.China
Guojie Dong, Ruey-wen Liu
University of Notre Dame, Notre Dame, In 46556, U.S.A.
(email: xtling@fudan.ihep.ac.cn)

ABSTRACT: *A new algorithm is proposed for blind identification of non-minimum phase linear time-variant channels. It is proved to be globally convergent in both continuous and discrete domains. The fast convergence speed makes this algorithm very useful for tracing a time-variant channel in mobile communication. Some simulations are given to show its tracking ability. Also a constructive approach is presented to determine the channel order.*

1. INTRODUCTION

In order to solve the non-minimum phase problem, conventional approaches for blind channel identification are tend to use the high order cumulant (HOC) of the signal. Such HOC based methods, as expected, suffer from computational intensity, unreliability of high order statistics, and slow convergence rate. Recently, Dong *et al*[1] proposed a new method using two receivers and a neural network with orthogonal learning rule for blind channel identification. The algorithm is proved to be globally convergent in [1]. However, due to the difference between continuous domain and discrete domain, computer simulations show that divergence occurs when signal amplitude is relatively large or the learning step is not very small. Meanwhile, the algorithm needs a precise estimation of the channel order to design the neural network. This problem is vital but not discussed in [1].

In this paper, a new learning rule based on the same network of [1] is proposed. Proofs show that the algorithm is globally convergent in both continuous and discrete domains. The convergence speed of the network with the new rule is much faster than that of the previous one. Some simulation results are given to show its tracking ability and its anti-noise capacity. The problem of channel order determination is also discussed in this paper. A method of precise order estimation for multi-channel is presented after the discussion.

2. BRIEF DESCRIPTION OF THE NETWORK

Suppose that the communication channel can be described by MA model, the structure of the network proposed in [1] is shown in figure 1. Where $s(k)$ is the common input signal sequence, H_j $(j=1, 2)$ describe the two communication channels. $x_j(k)$ $(j=1, 2)$ is the output sequence of the j th channel received by the j th sensor. p_j $(j=1,\dots,2L+2)$ present the weights of the neural network whose output is y. Since both channels can be described as FIR filters, the channel's transfer function in the domain of Z can be written as

$$H_j(z)=\sum_{n=0}^{L}h_j(n)z^{-n} \qquad (j=1, 2) \tag{1}$$

We assume that the higher order of the two channels has already been known to be L. The order determination problem will be discussed later in section 5. Define two arrays in space R^{2L+2} here

$$\vec{X}_k=[x_2(k),\dots x_2(k-L), -x_1(k),\dots -x_1(k-L)]^T$$
$$\vec{H}=[h_1(0),\dots h_1(L), h_2(0),\dots h_2(L)]^T$$

and consider the input signal $s(k)$ as random sequence. The following theorem will then be obtained.

Theorem 1: The correlation matrix of $\vec{X}_k$ will have a distinctive zero eigenvalue whose corresponding eigenvector will just be the normalized vector of $\vec{H}$, if the following conditions are satisfied

1. $H_1(z)$ and $H_2(z)$ have no common zeros.
2. At least the higher order of the channels is L.

Proof: First, in the aspect of existence,

The relationship between $x_j(k)$ and $s(k)$ can be described as

$$\begin{cases} x_1(k)=\sum_{j=0}^{L}s(k-j)h_1(j) \\ x_2(k)=\sum_{j=0}^{L}s(k-j)h_2(j) \end{cases} \tag{2}$$

Then, we have

$$\vec{X}_k{}^T\vec{H}=\sum_{j=0}^{L}x_2(k-j)h_1(j)-\sum_{j=0}^{L}x_1(k-j)h_2(j)\equiv 0 \tag{3}$$

If the correlation matrix of $\vec{X}_k$ is denoted as R, a direct conclusion of (3) will be

$$R\vec{H}=E[\vec{X}_k\vec{X}_k{}^T]\vec{H}=E[\vec{X}_k\vec{X}_k{}^T\vec{H}]\equiv\vec{0} \tag{4}$$

That is to say, R do have a zero eigenvalue while the normalized vector of $\vec{H}$ is an eigenvector corresponding to it.

Second, in the aspect of uniqueness,

The relation between $\vec{X}_k$ and $s(k)$ can be described as

0-7803-3679-8/96/$5.00 © 1996 IEEE

$$\vec{X}_k = \begin{bmatrix} h_2(0) & h_2(1) & \cdots & 0 & \cdots & 0 \\ 0 & h_2(0) & \cdots & h_2(L) & \cdots & 0 \\ 0 & \cdots & h_2(0) & \cdots & \cdots & h_2(L) \\ -h_1(0) & -h_1(1) & \cdots & 0 & \cdots & 0 \\ 0 & -h_1(0) & \cdots & -h_1(L) & \cdots & 0 \\ 0 & \cdots & -h_1(0) & \cdots & \cdots & -h_1(L) \end{bmatrix} \begin{bmatrix} S(k) \\ S(k-1) \\ \vdots \\ \vdots \\ \vdots \\ S(k-2L) \end{bmatrix} \quad (5)$$

It has been proved that when condition 1,2 is satisfied, the transfer matrix will have full column rank[2]. Since $S(k)$ is random, the space spanned by $\vec{X}_k$ will have a dimension of $2L+1$. Thus the rank of R will exactly be $2L+1$, indicating that the zero eigenvalue is distinctive. Proof completed.

Considering that R is quasi-positive definite, so the zero eigenvalue is the minimum eigenvalue of R. If we call the eigenvector corresponding to the minimum eigenvalue as "minimum eigenvector", the purpose of the neural network is to extract the unique minimum eigenvector.

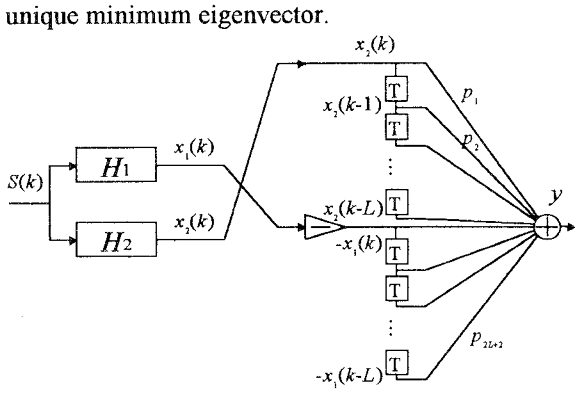

Figure 1: Structure of the network proposed in [1]

3. LEARNING ALGORITHM AND ANALYSIS OF CONVERGENCE IN CONTINUOUS DOMAIN

We propose the new learning algorithm using anti-hebbian rule in continuous domain.

$$\dot{\vec{P}} = b\left\{-y\vec{X} + \frac{y^2}{|\vec{P}|^2}\vec{P} - a\left(|\vec{P}|^2 - 1\right)\vec{P}\right\} \quad (6)$$

where $b, a \in R$ are both positive. The convergence of the network will then be discussed in both aspects of amplitude and direction.

Considering $y = \vec{P}^T\vec{X} = \vec{X}^T\vec{P}$, we come to the following theorem.

Theorem 2: If $\vec{P}_0$ is an eigenvector of R and $|\vec{P}_0| = 1$, then $\vec{P}_0$ is an equilibrium point of the network.

Proof: Denote the corresponding eigenvalue of $\vec{P}_0$ as λ_0, when $\vec{P} = \vec{P}_0$ we have

$$\dot{\vec{P}} = b\left\{-R\vec{P}_0 + \left[\vec{P}_0^T R\vec{P}_0\right]\vec{P}_0\right\} = b\left\{-\lambda_0\vec{P}_0 + \lambda_0\vec{P}_0\right\} = 0 \quad (7)$$

Proof completed.

Next, we will give another theorem that ensures the convergence in the aspect of direction.

Theorem 3: If $\dot{\vec{P}}$ can be expressed as $b\left\{-y\vec{X} + k\vec{P}\right\}$ (k is an arbitrary scalar), the weight vector $\vec{P}$ will converge to the minimum eigenvector of R in the aspect of direction.

Proof: Assume $\vec{C}_1 \cdots \vec{C}_n$ to be the n normalized eigenvectors of R, while their corresponding eigenvalues are denoted as $\lambda_1 \cdots \lambda_n$ respectively. Without losing generality, we assume λ_n to be the minimum one. If $\vec{C}_n$ is distinctive, we can give the expansion of $\vec{P}$ on $\vec{C}_1 \cdots \vec{C}_n$ as

$$\vec{P} = \sum_{i=0}^{n} a_i\vec{C}_i \quad (8)$$

Meanwhile, if $a_n \neq 0$, define $\xi_i = a_i / a_n$, thus, from (9) we obtain

$$\frac{d\xi_i}{dt} = \frac{1}{a_n}\frac{da_i}{dt} - \frac{\zeta_i}{a_n}\frac{da_n}{dt} = -b(\lambda_i - \lambda_n)\xi_i \quad (9)$$

Since $\vec{C}_n$ being distinctive and $a_n \neq 0$ can always be satisfied in practice, and from the above assumption, $\lambda_i - \lambda_n > 0$, we can get $\xi_i \to 0$ when $t \to \infty$. That is to say, vector $\vec{P}$ have the tendency to converge to the minimum eigenvector in the aspect of direction.

Finally, in the aspect of amplitude, we have following theorem to affirm convergence.

Theorem 4: With the learning rule of (6), if $b > 0, a > 0$, then $|\vec{P}|^2 = 1$ will be the only stable solution of the neural network.

Proof: Because

$$\frac{d|\vec{P}|^2}{dt} = -2ba\left(|\vec{P}|^2 - 1\right)|\vec{P}|^2 \quad (10)$$

from $\frac{d|\vec{P}|^2}{dt} = 0$ we can get the two solutions of the network to be $|\vec{P}|^2 = 1$ and $|\vec{P}|^2 = 0$. It can be easily proved $|\vec{P}|^2 = 0$ is not a stable solution. When come to $|\vec{P}|^2 = 1$, we define $Q = |\vec{P}|^2 - 1$ and construct a Lyapunov function $E = \frac{1}{2}Q^2$, then we have

$$\frac{dE}{dt} = Q\frac{dQ}{dt} = -2ba\left(|\vec{P}|^2 - 1\right)^2|\vec{P}|^2 \leq 0 \quad (11)$$

So $|\vec{P}|^2 = 1$ is the only stable solution.

With the above theorems, we have proved in continuous domain that with the learning algorithm (6) the weight vector $\vec{P}$ will globally converge to the normalized minimum eigenvector of R.

4. DISCRETE LEARNING ALGORITHM AND CONVERGENCE DISCUSSION

In this section, we will do our discussion in the discrete domain. As given in [3], how to select a learning step in practice is a very difficult problem. In the following passage, we will first propose our discrete learning algorithm and then give a thorough analysis of the network's convergence property. It will be seen from the proof that if we apply the learning rule of continuous domain to discrete domain directly, convergence will no longer be affirmed.

Suppose $0<b\le 1$, we define $\xi=\frac{|\bar{P}(t)|^2-1}{|\bar{P}(t)|^2+1}$, $q=|\bar{X}(t)|^2-\frac{y^2}{|\bar{P}(t)|^2}$. Since $\xi<1$ and $q\ge 0$, $\frac{1-\xi}{q}=\frac{2}{\left(|\bar{P}(t)|^2+1\right)q}>0$ can always be satisfied. Then, our discrete learning rule can be written as

$$\bar{P}(t+1)=\bar{P}(t)+b\left(\frac{1-\xi}{q}\right)\left(-y\bar{X}(t)+\frac{y^2}{|\bar{P}(t)|^2}\bar{P}(t)\right)-\xi\bar{P}(t) \tag{12}$$

Here $0<b\le 1$. Similar to the discussion in section III, the convergence property will be analyzed in both aspects of amplitude and direction.

First, in the aspect of direction, we denote the angle between $\bar{P}(t)$ and $\bar{X}(t)$ in R^{2L+2} as $\theta(t)$, and the angle between $\bar{P}(t+1)$ and $\bar{X}(t)$ as $\theta(t+1)$. To extract the minimum eigenvector, $\theta(t)\to\frac{\pi}{2}$ is what we expect. If we define

$$\rho(b)=\frac{\cos^2\theta(t+1)}{\cos^2\theta(t)}=\left[\frac{|\bar{P}^T(t+1)\bar{X}(t)|^2}{|\bar{P}(t+1)|^2}\right]\Bigg/\left[\frac{|\bar{P}^T(t)\bar{X}(t)|^2}{|\bar{P}(t)|^2}\right] \tag{13}$$

Then $\rho(b)<1$ should be ensured for every learning step to make the network converge directly to the final result. Define $\Delta=\bar{X}(t)-\frac{y}{|\bar{P}(t)|^2}\bar{P}(t)$ and considering $|\Delta|^2=q=|\bar{X}|^2\sin^2\theta(t)$, we can get from (12)

$$\rho(b)=|\bar{P}(t)|^2(1-\xi)^2(1-b)^2\Bigg/(1-\xi)^2\left[|\bar{P}(t)|^2+\frac{b^2y^2}{q}\right] \tag{14}$$
$$\le |\bar{P}(t)|^2(1-b)^2\Big/|\bar{P}(t)|^2=(1-b)^2<1$$

Especially when $b=1$, we get $\rho(b)=0$, indicating the network has reached its highest convergence speed in the aspect of direction.

In the aspect of amplitude, the analysis becomes more complicated to some extend. Here we just give the relation between $|\bar{P}(t+1)|^2$ and $|\bar{P}(t)|^2$ by

$$|\bar{P}(t+1)|^2=\frac{4|\bar{P}(t)|^2}{\left(|\bar{P}(t)|^2+1\right)^2}\left(1+b^2ctg^2\theta(t)\right) \tag{15}$$

Obviously, what we have already gotten is

$$|\bar{P}(t+1)|^2\le\frac{4|\bar{P}(t)|^2}{\left(2|\bar{P}(t)|\right)^2}\left(1+b^2ctg^2\theta(t)\right)=1+b^2ctg^2\theta(t) \tag{16}$$

Thus, an upper bound of the amplitude has already been given. Then let us consider the lower bound. If $|\bar{P}(t+1)|^2<1$, from (15), we have

$$|\bar{P}(t+1)|^2>|\bar{P}(t)|^2\left(1+b^2ctg^2\theta(t)\right) \tag{17}$$

indicating that the amplitude has a tendency of increasing. Meanwhile, if $|\bar{P}(t+1)|^2\ge 1$, due to (15),(16)

$$|\bar{P}(t+1)|^2\ge\left(1+b^2ctg^2\theta(t)\right)\Big/|\bar{P}(t)|^2\ge 1 \tag{18}$$

can be obtained. (17),(18) assured us that the lower bound of the amplitude should finally be 1. Since $\theta(t)\to\frac{\pi}{2}$, so $\left(1+b^2ctg^2\theta(t)\right)\to 1$ surely holds, accordingly affirming the convergent point of the amplitude to be 1.

Now we have proved the globally convergent property of the discrete learning rule(12).

5. CHANNEL ORDER DETERMINATION

From Theorem 1 we can see that accurately estimating the channel order is vital to our proposed algorithm. Some HOC based methods for order determination have already been proposed before, but computer simulations show that they often fail to ensure a precise order estimation. In this section, we will present an approach not appealing in appearance but efficient to determine the higher order of the channels.

Suppose we have two neural networks, each is the same as the one shown in Fig. 1. The two networks are both designed according to a supposed order L, and their inputs are the same ($\bar{X}(t)$). The only difference between them is the initial condition. On one hand, if the supposed order L is higher, from (5) we can see the minimum engivector of R will no longer be distinctive. Since the two neural networks are different in initial condition, their learning result will also be different. So the crosscorrelation value of the two weight vectors will have a large probability to be small, at least not too high to be above 0.95 for example. (Though the possibility of misjudgment is very small, we can simply calculate twice to reduce the probability.) On the other hand, if the supposed order L is lower, $\bar{X}_k$ will then span the entire space of dimension $2L+2$. Though the minimum eigenvalue may no longer be zero, most probably, the minimum eigenvector will remain distinctive. As we have discussed before, the two networks will still converge to the normalized minimum eigenvector. Their crosscorrelation of weight vector will keep a very high value close to 1. From this phenomenon, whether the supposed order is higher or not can be easily judged. That will be enough for us to find the proper order.

We design the two neural networks in an arbitrarily high order first, then let them have a downward search to find the proper channel order. Meanwhile, every time when the current order L of the two NN (neural network) is not higher than that of the real channel, we increase the design order to $L+1$. Thus, if the higher order of the physical channels is L_0, the design order of the two NN will finally oscillate between L_0 and L_0+1. Of course, the proper order can be picked out easily. Because of the avoiding of high order statistics calculation and the high convergence speed of the algorithm, our order determination method is very accurate and efficient to be put into practice. Furthermore, it can even trace the channel order varying in a relatively low speed.

6. SIMULATION RESULTS

Without losing the high convergence speed, we add a moving average filter behind each neural network to reduce the effect of noise on the output. The final construction of the network with order determination part is shown in Fig. 2. Where NN1 is used to identify the coefficients while NN2 and NN3 are used to estimate the channel order.

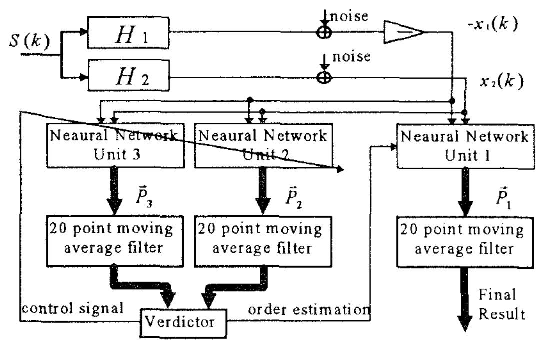

Figure 2: structure of the network in practice ($\vec{P}_1$, $\vec{P}_2$, $\vec{P}_3$ respectively present the weight vector of the neural networks)

In this section, two simulations are given. The inputs $\vec{X}(t)$ of the neural networks are the outputs of two channels with PAM scheme driven by a randomly distributed 4-ary transmitting symbols $S(k)$. According to the analysis in section IV, we let $b=1$ to set the convergence rate to its fastest point. In each simulation, we initiate the weights of the neural networks by a normalized random vector. For convenience, we omit the order determination course in the following simulations.

Simulation 1: The actual 2-channel impulse responses are set as

$$\vec{H}_1 = [1.0 \quad 0.366 \quad -0.183 \quad -0.125]^T,$$
$$\vec{H}_2 = [1.0 \quad -0.2 \quad -0.3 \quad 0.4]^T$$

We reduce SNR of the receivers to 14db. Line I and II in Fig. 3 show the correlation between the expected result and $\vec{P}(t)$ gotten by the new learning rule and that of [1] respectively.

Simulation 2: The tracking ability of the network is simulated by setting the symbol rate at 300kHz with

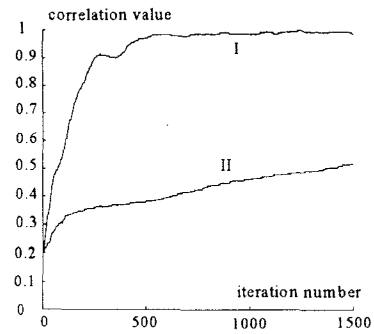

Figure 3: Network performance of different learning algorithm

channel Doppler spread 100Hz, corresponding to a speed of a mobile receiver up to about 70mph. For PAM signals, the time-varying channel coefficients are simulated by low frequency Gaussian noise with the bandwidth equal to the Doppler spread. The trajectories of the first three coefficients of the first channel are given by the dotted lines in Fig. 4, while their estimations are given by the solid lines.

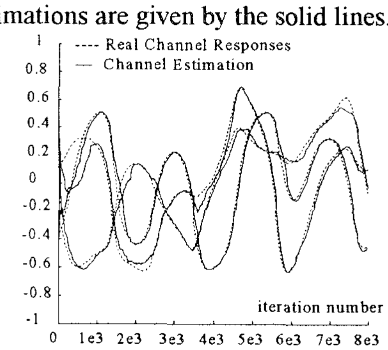

Figure 4: The trajectories of channel coefficients and their estimation

7. CONCLUSIONS

A new algorithm for blind 2-channel identification is proposed. Proof and computer simulations show its high convergence speed and anti-noise capacity. It is very efficient in tracking time-variant MA channels in mobile communication.

8. REFERENCES

[1] Guojie Dong, Ruey-wen Liu, "An Orthogonal Learning Rule for Null-Space Tracking with Implementation to Blind 2-Channel Identification," Technical Report, University of Notre Dame, Dec. 1995.

[2] E. Moulines, P. Duhamel, J-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," IEEE Trans. Signal processing, Vol. 43, 1995, pp. 516-525.

[3] Lei Xu, Krkki Oja, and Ching Y. Suen. "Modified Hebbian Learning for Curve and Surface Fitting," Neural Networks, Vol. 5, 1992, pp. 441-457.