A NEW ADAPTIVE FILTER ALGORITHM FOR SYSTEM IDENTIFICATION USING INDEPENDENT COMPONENT ANALYSIS

Jun-Mei Yang and Hideaki Sakai

Graduate School of Informatics, Kyoto University,
Kyoto, 606-8501, Japan
Email: {jmyang, hsakai}@sys.i.kyoto-u.ac.jp

ABSTRACT

This paper proposes a new adaptive filter algorithm for system identification using independent component analysis (ICA), which separates the signal from noisy observation under the assumption that the signal and noise are independent. We first introduce an augmented state-space expression of the observed signal, representing the problem in terms of ICA, and then use an adaptive gradient descent algorithm to separate the noise from the signal. A local convergence condition is also shown. The proposed algorithm can be applied to the acoustic echo cancellation problem directly and some simulations have been carried out to illustrate its effectiveness.

Index Terms— Adaptive filter, System identification, Information theory, Nonlinear estimation

1. INTRODUCTION

The adaptive filter technique has been applied to many system identification problems in communications and noise control [1][2]. The LMS (least mean square) algorithm is extensively used due to its low computational complexity and the robustness of its performance [1]. The RLS (recursive least squares) algorithm is also often used due to its fast convergence property. These two popular algorithms for system identification are based on the idea that the effect of additive observation noise is to be suppressed in the least square sense. But if the noise is non-Gaussian, the performances of the above algorithms degrade significantly. The other class of non-linear algorithms has been derived based on the robust estimation theory [3], but these algorithms are a little bit heuristic.

On the other hand, in recent years, independent component analysis (ICA) has been attracting much attention in many fields such as signal processing and machine learning [4]. However, in the adaptive filter area, there have been only a few papers which try to derive adaptive algorithms from the view point of ICA. The authors in [5] tried to formulate the conventional system identification problem in the ICA context, but the proposed algorithm is nothing but the QR type RLS algorithm. In [6] a truly ICA type algorithm has been derived for identification of multivariate autoregressive models.

In this paper, by combining the approaches in [5] and [6], we derive a new adaptive algorithm for system identification using the technique of ICA. We try not to suppress the noise in the least mean square sense but to maximize the independence (i.e. minimize the dependence) between the signal part and the noise.

The organization of this paper is as follows: In Section 2 we introduce an augmented linear model representing the problem in the frame work of ICA and then propose a new adaptive algorithm in Section 3 by using the ICA technique. Section 4 is devoted to the convergence analysis of the new algorithm. In Section 5, the new ICA-based method is applied to acoustic echo cancellation. Finally, some numerical simulations are demonstrated to show the superiority of our ICA-based method to the conventional NLMS algorithm.

2. PROBLEM FORMULATION

In the basic ICA model, the observation variables are denoted as $x_i(t), (i = 1, \cdots, n)$, where t is the time or sample index. We assume that they are generated as a linear mixture of the independent components $s_i(t), (i = 1, \cdots, n)$,

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$$

where

$$\mathbf{x}(t) \triangleq (x_1(t), \cdots, x_n(t))^T,\ \mathbf{s}(t) \triangleq (s_1(t), \cdots, s_n(t))^T$$

and $\mathbf{A}$ is some unknown matrix. The problem is to recover original source signals $\mathbf{s}(t)$ from the observations $\mathbf{x}(t)$. Independent component analysis consists of finding a linear transform $\hat{\mathbf{s}} = \mathbf{W}\mathbf{x}$ so that the components $\hat{s}_i$ are as independent as possible, in the sense of maximizing some function that measures independence.

We consider the problem of identifying a linear system described by

$$y(n) = \mathbf{w}_0^T \mathbf{x}(n), \tag{1}$$

where

$$\begin{aligned} &\mathbf{w}_0 \triangleq [w_0\ w_1 \cdots w_{m-1}]^T, \\ &\mathbf{x}(n) \triangleq [x(n)\ x(n-1) \cdots x(n-m+1)]^T. \end{aligned} \tag{2}$$

$x(n)$ is the zero mean input signal. The measurement of the system output $y(n)$ is corrupted by the noise $e(n)$, that is,

$$d(n) = y(n) + e(n). \tag{3}$$

We assume the noise $e(n)$ is zero mean and statistically independent with the system input $\mathbf{x}(n)$. Statistical independence is a much stronger condition than uncorrelatedness. As a result, statistics of order higher than the second has to be considered for non-Gaussian signals.

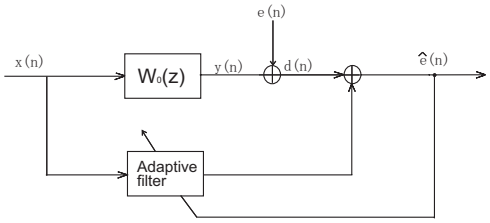

Fig.1. General configuration of system identification

We now introduce the following augmented linear model to formulate the problem of system identification in the frame work of ICA:

$$\begin{bmatrix} \mathbf{x}(n) \\ d(n) \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{w}_0^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{x}(n) \\ e(n) \end{bmatrix}, \tag{4}$$

where $\mathbf{I}$ denotes the identity matrix.

The noise signal $e(n)$, which is assumed to be independent of the input signal $\mathbf{x}(n)$, is expected to be separated from the observation signal. So we may consider the system identification problem as an ICA problem, although the input signals $x(n-i), i = 0, 1, \cdots, m-1$ are heavily autocorrelated.

On the basis of model (4), we introduce the following system

$$\begin{bmatrix} \mathbf{x}(n) \\ \hat{e}(n) \end{bmatrix} = \hat{W} \begin{bmatrix} \mathbf{x}(n) \\ d(n) \end{bmatrix} \tag{5}$$

where

$$\hat{W} \triangleq \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \hat{\mathbf{w}}^T & a \end{bmatrix} \tag{6}$$

$$\hat{\mathbf{w}} = [\hat{w}_0\ \hat{w}_1 \cdots \hat{w}_{m-1}]^T \tag{7}$$

$$\hat{e}(n) = \hat{\mathbf{w}}^T \mathbf{x}(n) + ad(n) \tag{8}$$

and a is a nonzero scalar quantity. In the usual adaptive filtering problem a is set to 1.

In the following, we will find a good estimate of the matrix $\hat{W}$ so that $\mathbf{x}(n)$ and $\hat{e}(n)$ are as independent as possible.

3. DERIVATION OF THE PROPOSED ALGORITHM

Mutual information is a natural measure of the statistical dependency between random variables. It is a special case of Kullback-Leibler divergence (KLD), when one measures the distance between the joint probability distribution and the product of the marginal distributions:

$$I(\mathbf{x}; \mathbf{y}) = \int p(\mathbf{x}, \mathbf{y}) \log \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{x})p(\mathbf{y})} \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}$$

It takes into account the whole dependence structure of the variables, not just the covariance. It is always nonnegative, and zero if and only if the variables are statistically independent. Therefore it is a very natural way to estimate the independent components by finding a transform that minimizes the mutual information of their estimates. The mutual information between $\mathbf{x}(n)$ and $\hat{e}(n)$ can be expressed as:

$$I(\mathbf{x}, \hat{e}) = H(\mathbf{x}) + H(\hat{e}) - H(\mathbf{x}, \hat{e})$$

where $H(\cdot)$ denotes the differential entropy,

$$H(\boldsymbol{y}) \triangleq -\int p(\boldsymbol{y}) \log p(\boldsymbol{y}) d\boldsymbol{y} = E_p[-\log p(\boldsymbol{y})].$$

By the properties of differential entropy, for the invertible linear transformation (5) the following equation is verified:

$$H(\mathbf{x}, \hat{e}) = H(\mathbf{x}, d) + \log|a|.$$

Hence, we have

$$I(\mathbf{x}, \hat{e}) = H(\mathbf{x}) - H(\mathbf{x}, d) + H(\hat{e}) - \log|a|. \tag{9}$$

To minimize the mutual information, first we need to find its gradient. Since the elements of $\hat{W}$ in (6), except those of the last row, are fixed quantities, we define the gradient of $I(\mathbf{x}, \hat{e})$ with respect to $\hat{W}$ as

$$\frac{\partial I(\mathbf{x}, \hat{e})}{\partial \hat{W}} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \frac{\partial I}{\partial \hat{\mathbf{w}}^T} & \frac{\partial I}{\partial a} \end{bmatrix}. \tag{10}$$

Using the same argument in [7], we have

$$\frac{\partial I}{\partial \hat{\mathbf{w}}^T} = \mathbf{E}[\mathbf{x}^T \phi(\hat{e})], \quad \frac{\partial I}{\partial a} = \mathbf{E}[d\phi(\hat{e})] - \frac{1}{a}. \tag{11}$$

where the function $\phi(\cdot)$ is defined as

$$\phi(\hat{e}) \triangleq -\frac{\mathrm{d} \log p_{\hat{e}}(\hat{e})}{\mathrm{d}\hat{e}}. \tag{12}$$

Instead of using this usual gradient, we use the so-called relative (natural) gradient ([8]) defined by

$$\tilde{\nabla} I(\mathbf{x}; \hat{e}) = \frac{\partial I(\mathbf{x}, \hat{e})}{\partial \hat{W}} \hat{W}^T \hat{W}. \tag{13}$$

Hence from (6), (10), (11), (13) and (8), we have

$$-\tilde{\nabla} I(\mathbf{x}; \hat{e}) = -\begin{bmatrix} 0 & 0 \\ \Psi(\mathbf{x}, \hat{e}) & \mathbf{E}[\hat{e}\phi(\hat{e})]a - a \end{bmatrix}, \tag{14}$$

where $\Psi(\cdot)$ is defined as

$$\Psi(\mathbf{x}, \hat{e}) \triangleq \mathbf{E}[\mathbf{x}^T \phi(\hat{e})] + \mathbf{E}[\hat{e}\phi(\hat{e})]\hat{\mathbf{w}}^T - \hat{\mathbf{w}}^T.$$

By using the instantaneous values of $\mathbf{x}^T\phi(\hat{e})$ and $\hat{e}\phi(\hat{e})$ instead of their expectations, we propose the following stochastic gradient adaptive algorithm :

$$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \mu_1[(1-\phi(\hat{e}(n))\hat{e}(n))\hat{\mathbf{w}}(n) - \phi(\hat{e}(n))\mathbf{x}(n)], \quad (15)$$

$$a(n+1) = a(n) + \mu_2[1-\phi(\hat{e}(n))\hat{e}(n)]a(n), \quad (16)$$

where μ_1 and μ_2 are the update step sizes and

$$\hat{e}(n) = a(n)d(n) + \mathbf{x}^T(n)\hat{\mathbf{w}}(n). \quad (17)$$

By using the averaging method in [9], we consider the following averaged system corresponding to (15), (16)

$$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \mu_1\mathbf{E}[(1-\phi(\hat{e}(n))\hat{e}(n))\hat{\mathbf{w}}(n) - \phi(\hat{e}(n))\mathbf{x}(n)], \quad (18)$$

$$a(n+1) = a(n) + \mu_2\mathbf{E}[1-\phi(\hat{e}(n))\hat{e}(n)]a(n). \quad (19)$$

The desired equilibrium point of (18) and (19) is

$$a(n) = a_0, \quad \hat{\mathbf{w}}(n) = -a_0\mathbf{w}_0, \quad (20)$$

where a_0 is determined by

$$\mathbf{E}[\phi(a_0e(n))a_0e(n)] = 1. \quad (21)$$

This is because at this point, from (1), (3) and (17), we have

$$\hat{e}(n) = a_0e(n)$$

and the second terms in the right hand side of (18) and (19) are zero. From (20), we can identify $\mathbf{w}_0$.

If the noise signal is Gaussian with mean 0 and variance σ^2, from (12) we have

$$\phi(e) = -\frac{\mathrm{d}\log(\exp(-e^2/2\sigma^2)/\sqrt{2\pi}\sigma)}{\mathrm{d}e} = e/\sigma^2.$$

Substituting this into (21), then we get $a_0 = \pm 1$ which is quite natural.

The configuration of the new adaptive noise canceller is shown in Fig.2.

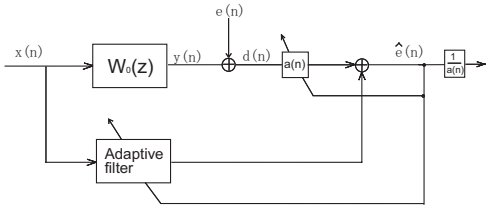

Fig.2 Configuration of the new adaptive filter.

Remark 1: If we do not introduce the quantity a, but use a constant "1" in (8) as is usual in the adaptive filter area, then the second partial derivative in (11) with respect to a will be zero. Hence, the measurement $d(n)$ can not be explicitly used in the algorithm.

Remark 2: It is quite difficult to calculate function $\phi(\hat{e})$ since it involves the probability density function. Fortunately, in practice the selection of such a function is not too stringent. We can use the tanh function and sometimes the sigmoid function.

4. CONVERGENCE ANALYSIS

The algorithm is highly nonlinear and is difficult to analyze its convergence property rigorously. So the local stability property of this nonlinear algorithm near the equilibrium point is considered here.

Let $\hat{\mathbf{w}}(n) = -a_0\mathbf{w}_0 + \Delta\hat{\mathbf{w}}(n)$, $a(n) = a_0 + \Delta a(n)$, from (17) we have

$$\hat{e}(n) = a_0e(n) + \mathbf{x}^T(n)\Delta\hat{\mathbf{w}}(n) + \Delta a(n)d(n).$$

Discarding the higher order terms, and noting the noise $e(n)$ is zero mean and statistically independent with $\mathbf{x}(n)$, the averaged system (18), (19) can be linearized around the equilibrium point $(-a_0\mathbf{w}_0, a_0)$ as follows:

$$\begin{bmatrix} \Delta\hat{\mathbf{w}}(n+1) \\ \Delta a(n+1) \end{bmatrix} = \begin{bmatrix} G_{11} & G_{12} \\ 0 & G_{22} \end{bmatrix} \begin{bmatrix} \Delta\hat{\mathbf{w}}(n) \\ \Delta a(n) \end{bmatrix}, \quad (22)$$

where

$$G_{11} = I - \mu_1\boldsymbol{E}\big[\phi'(a_0e(n))\big]R_{xx},$$

$$G_{12} = \mu_1\Big[a_0\Big(\boldsymbol{E}\big[\phi(a_0e(n))e(n)\big] + \boldsymbol{E}\big[\phi'(a_0e(n))a_0e^2(n)\big]\Big)\mathbf{w}_0 - \boldsymbol{E}\big[\phi'(a_0e(n))\big]R_{xx}\mathbf{w}_0\Big],$$

$$G_{22} = 1 - \mu_2a_0\Big(\boldsymbol{E}\big[\phi(a_0e(n))e(n)\big] + \boldsymbol{E}\big[\phi'(a_0e(n))a_0e^2(n)\big]\Big),$$

with the covariance matrix $R_{xx} = \boldsymbol{E}\big[\mathbf{x}(n)\mathbf{x}^T(n)\big]$.
The algorithm converges to its equilibrium point $(-a_0\mathbf{w}_0, a_0)$ iff all the eigenvalues of G_{11} and G_{22} are strictly inside the unit circle. Hence, the following step size conditions are needed:

$$0 < \mu_1 < \frac{2}{\mathbf{E}[\phi'(a_0e(n))]\delta_{max}},$$

$$0 < \mu_2 < \frac{2}{a_0\big(\mathbf{E}[\phi(a_0e(n))e(n)] + \mathbf{E}[\phi'(a_0e(n))a_0e^2(n)]\big)}.$$

where δ_{max} represents the maximum eigenvalue of R_{xx}. Of course, these upper limits are approximate ones, since the averaging method assumes the slow adaptation limit ($\mu_1 \to 0, \mu_2 \to 0$). According to the above two conditions, the function $\phi(x)$ should be chosen so that the denominators of the above expressions of the upper bound are positive, that is,

$$\boldsymbol{E}\big[\phi'(a_0e(n))\big] > 0,$$

$$\boldsymbol{E}\big[\phi(a_0e(n))a_0e(n) + \phi'\big(a_0e(n)\big)(a_0e(n))^2\big] > 0. \quad (23)$$

5. APPLICATION TO ACOUSTIC ECHO CANCELLATION

The proposed algorithm can be directly implemented to acoustic echo cancellation (AEC).

Fig.3 illustrates the general configuration of an acoustic echo canceller. Let $x(n)$ be the far-end signal going to the loudspeaker and $y(n)$ be the echo signal due to leakage from the loudspeaker to the microphone, which is heavily correlated with $x(n)$; $d(n)$ is the signal picked up by the microphone, which will be called the desired signal. Here the echo signal $y(n)$ and the near-end signal $e(n)$ are the independent sources to be separated from the microphone signal $d(n)$. We assume a linear echo

$$y(n) = \sum_{i=0}^{m-1} w_i x(n-i),$$

where the echo is a weighted sum of echo components $x(n-i), i = 0, 1, \cdots, m-1$. The echo components are delayed versions of the far-end signal and the microphone signal $d(n)$ is a sum of the near-end signal $e(n)$ and the echo $y(n)$.

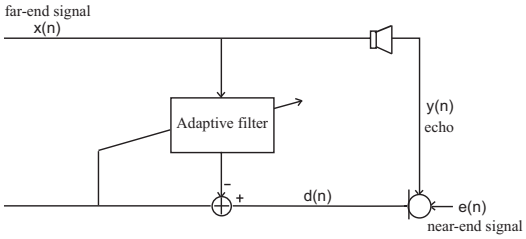

Fig.3. General configuration of an acoustic echo canceller

In a common telephone call, double-talk is found to occur 20 percent of the time. Double-talk is any period during a call when both the near-end signal and the far-end signal contain speech. During a double talk period, standard adaptive filtering schemes tend to get confused and can not track the echo path properly.

The near-end signal is usually independent of the far-end signal and the echo signal. Also it may be of much larger amplitude than the echo. So we apply the proposed ICA-based algorithm (15) (16) to the acoustic echo canceller without double talk detector.

6. SIMULATION RESULTS

We simulate the performance of the acoustic echo canceller by using the NLMS and ICA methods, respectively.

The echo path impulse response used in the simulation has a length of 128ms, consisting of 1024 coefficients at 8 kHz sampling rate. We use a male speech sound as the near-end signal and a female speech as the far-end signal. Both of them are sampled by 8kHz. The near-end signal is added at the midpoint. The quantity "error" is defined as the difference between the near-end signal $e(n)$ and its estimate $\hat{e}(n)/a(n)$. The last two parts of the figure show the MSE values for the NLMS and ICA algorithm respectively. We take $\phi(x) = \tanh(x)$ in the ICA method. For this choice we note $\phi'(x) > 0, \phi(x)x + \phi'(x)x^2 > 0$ $(x \neq 0)$, so that (23) is satisfied. The simulation result shows the NLMS algorithm gets confused when the double talk occurs, however our ICA algorithm works considerably well.

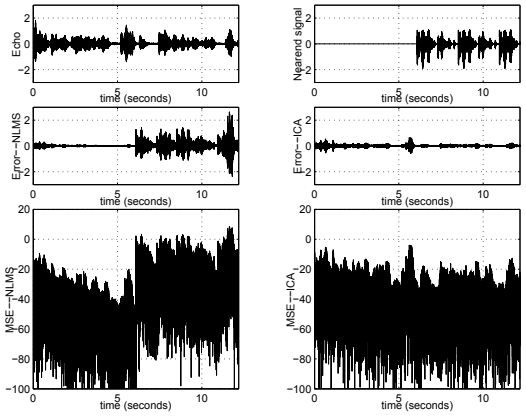

Fig.4 Performance comparison of the NLMS and ICA algorithm

7. CONCLUSION

In this paper we have derived a new adaptive filter algorithm for system identification. The new algorithm is based on the idea of ICA. The local stability of the algorithm has also been analyzed. The proposed algorithm is applied to the acoustic echo canceller. Simulation results show that the approach based on ICA performs much better than the conventional NLMS algorithm.

It is a future work to investigate the effect of the choice of $\phi(\cdot)$ on the performance.

8. REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[2] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, 1985.

[3] P. J. Huber, *Robust Statistics*, Wiley, New York, 1981.

[4] A. Hyvarinen, J. Karhunen and E. Oja, *Independent Component Analysis*, Wiley, New York, 2001.

[5] M. Magadán, R. Niemistö, U. Ruotsalainen and J. Möttönen, "ICA for acoustic echo control", *Proc. EUSIPCO*, 2002, Vol. I, pp.503-506.

[6] M. Nitta, K. Sugimoto and A. Satoh, "Blind system identification of autoregressive model using independent component analysis", *Transactions of SICE*, Vol.41, No.5, pp.444-451(2005). (in Japanese)

[7] H. Yang and S. Amari, "Adaptive On-line Learning Algorithms for Blind Separation: Maximum Entropy and Minimum Mutual Information", *Neural Computation*, Vol.9, pp.1457-1482, 1997.

[8] S. Amari, "Natural Gradient Works Efficiently in Learning", *Neural Computation*, 10, 251-276 (1998).

[9] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms: Stability and Performance*, Prentice-Hall,. 1995.