On the INFOMAX algorithm for blind signal separation

J. Xi* J. F. Chicharo†
A. C. Tsoi‡ W. C. Siu**

*University of Wollongong, jiangtao@uow.edu.au
†University of Wollongong, chicharo@uow.edu.au
‡University of Wollongong, act@uow.edu.au
**Hong Kong Polytechnic University

This paper originally appeared as: Xi, J, chicharo, JF, Tsoi, AC & Siu, WC, On the INFOMAX algorithm for blind signal separation, WCCC-ICSP 2000. 5th International Conference on Signal Processing Proceedings, 21-25 August 2000, vol 1, 425-428. Copyright IEEE 2000.

ON THE INFOMAX ALGORITHM FOR BLIND SIGNAL SEPARATION

Jiangtao Xi, J. F. Chicharo and Ah Chung Tsoi
The University of Wollongong
Australia
jiangtao@uow.edu.au

Wan-Chi Siu
Hong Kong Polytechnic University
Hong Kong

Abstract
This paper provides an analytical examination of the INFOMAX algorithm [3] and establishes its effectiveness for blind signal separation using extensive simulation results. Results obtained show that the INFOMAX is not able to separate signal sources unless signal preprocessing is carried-out whereby the data to train the separating matrix is decorrelated. Further, results also show that if one uses the decorrelation preprocess alone it is able to effectively separate signal sources in many instances.. Hence we conclude that the INFOMAX algorithm may not be a useful approach for signal separation.

1. INTRODUCTION

Blind Signal Separation (BSS) has been proposed to separate or estimate the waveforms of unknown signal sources based on measurements from an array of sensors. A BSS problem can be modeled by the system depicted in Figure 1 where$N$ statistically independent signal sources are mixed through NxN unknown channels to produce N observations at the sensors. BSS is in effect a multiple-input multiple-output (MIMO) system referred to as a separating network which can be used to extract each of the original signal sources. BSS is a challenging problem due to its *blind* nature. That is, neither the original signal sources nor the transmission channels from the sources to the sensors are known *apriori*.

There exists two main approaches for BSS, these include: Statistically based algorithms [1] and neural network based BSS algorithms [2]. Compared to the statistical approach, neural network methods are considered to be more computational efficient, although their performance may not be guaranteed in some instances.

One of the most popular neural network-based approaches is called INFOMAX algorithm proposed by Bell and Sejnowski [3]. The INFOMAX method uses a gradient-based algorithm which leads to low complexity in terms of implementation.

The theoretical basis for the claimed success of the INFOMAX algorithm is not entirely clear from existing literature. For instance the role and effectiveness of the preprocessing stage prior to using the INFOMAX algorithm needs to be established. The pre-processing usually performs DC component removal and de-correlation. In other words, we still do not know the answer to the following questions: what is the effect of the pre-processing andcan the INFOMAX algorithm work without the pre-processing stage? This paper addresses these two important issues.

For the purposes of the our study we only consider the instantaneous mixing cases, that is, where all the mixing channels are instantaneous scalars. In such cases the separating network also contains instantaneous channels only. Consider the system depicted in Figure 1 where a column vector S is used to denote N signal sources, X denotes the measurements, while the column vector U denotes the separated signals, and Y denotes the auxiliary outputs as follows:

$$X = AS \tag{1}$$
$$U = WX \tag{2}$$
$$Y = G(U) \tag{3}$$

where A and W are N by N matrices, representing the mapping from S to X and from X to U

0-7803-5747-7/00/$10.00©2000 IEEE.

respectively, and $G(.)$ is the non-linear mapping from U to Y.

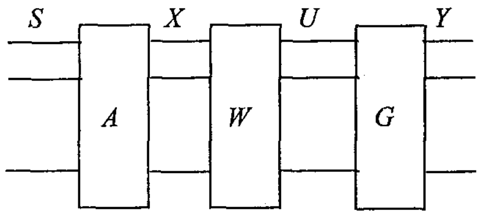

Figure 1. A blind separation system

2. REVISITING THE INFORMAX ALGORITHM [3]

The INFORMAX method is based on concept of maximizing the entropy (ME) of the above auxiliary output Y. The rational is based on the assumption that the expectation of the joint entropy of signal vector Y is maximized when the components of Y are independent. In order to study the INFOMAX method more closely it is necessary to briefly reexamine its derivation.

The joint entropy of Y is defined by:

$$H(Y) = -\int f_y(Y) \log f_y(Y) dY \qquad (4)$$

Where $f_y(Y)$ is the joint probability density function (pdf) of Y. It can be shown that,

$$H(Y) = H(X) + E(\log J_{Unmix}) + E(\log J_{Nonlin}) \qquad (5)$$

where $f_u(U)$ is the joint pdf of U, J_{Nonlin} is the Jacobian determinant of the transformation from U to Y, and J_{Unmix} is the Jacobian determinant of the transformation from X to U.

A gradient-based algorithm is used to maximize the entropy in Equation (5). However, it is a difficult task, because the first term, the entropy of the measurements, is unknown and therefore nothing can be done with it in the maximization. In this situation a sub-optimization approach was proposed, in which the gradient was evaluated based only on the remaining two terms [3],

$$\nabla W(k) \propto \frac{\partial[\log J_{Unmix} + \log J_{Nonlin}]}{\partial W(k)} = W^{-1} + \frac{\partial}{\partial W(k)} \sum_{i=1}^{N-1} \log \frac{\partial y_i}{\partial u_i} \qquad (6)$$

Assuming a sigmoid nonlinear function for y_n

$$y_n = g(u_n) = \frac{1}{1+e^{-u_n}} \qquad (7)$$

The instantaneous gradient is given by

$$\nabla W(k) \propto \left[W^T(k)\right]^{-1} + [I - 2Y(k)]X^T(k) \quad (8)$$

It has been claimed that the above algorithm is able to separate the sources mixed by instantaneous network [3-4]. Note that the INFOMAX principle has also been extended to convolutive mixing cases [5-7].

3. PERFORMANCE EVALUATION OF THE INFOMAX ALGORITHM

In this Section we examine the theoretical validity of the ME principle together with the gradient algorithm as presented in [3-7]. We first question the issue where it is assumed, without theoretical proof, the claim that maximization of the entropy will lead to an independent output [4]. Secondly, we examine the gradient given in Equation (8) and look for the justification that it will indeed result in Maximization of the joint entropy. It is evident that only the last two terms are considered in determining the gradient, and t is also clear that the entropy will depend on the source pdf as well as the mixing matrix A. In other words, the optimal solution for W also depends on the mixing matrix A. Hence it is evident that Equation (8) is not a perfect solution to the underlying ME problem. The above observations and discussion leads to the natural question: how the gradient algorithm achieves the blind separation? In other words, what is the major factor that steers the separating network toward a good solution? We examine this issue further by using extensive simulations.

We evaluate the separation performance of [3-7] by considering speech signals as an example. Consider the following matrix:

$$E = WA = \begin{bmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \end{bmatrix} \qquad (9)$$

where e_{ij} represents the of source s_j to output u_i. The diagonal elements of the above matrix can be used to establish the performance of signal separation.

A. Simulations on INFOMAX algorithm Without Pre-Processing (De-correlation)

For the sake of simplicity, we only study cases where two signal sources are mixed through a scalar (instantaneous) mixing network. The approach is to use two given signal sources, pass them through a artificial mixing network to get the two measurements, train the separation network using the measurement data, and pass the measurement data through the trained separation network to obtain the separated outputs.

The first example used in the simulation is as follows. Two speech signal sources are mixed through the mixing matrix given by:

$$A=\begin{bmatrix}1 & 0.7\\ 0.5 & 1\end{bmatrix} \tag{10}$$

The waveforms of the two original speech signals and the two mixed measurements are depicted in Figure 2. Obviously A is of full rank which guarantees the existence of a separation matrix. In our simulation, the measurement data samples are scaled into the range [-1,1], the step size is 0.01, the separation matrix is updated for every 30 data samples, and a total 16000 data samples are used for training the separation network. The resulting output waveforms are depicted in Figure 2. Note that the results obtained show poor performance in the sense that the signal are still mixed and signal sources remain largely unseperated as indicated by:

$$E=\begin{bmatrix}3.3399 & 2.3098\\ 1.6410 & 3.3379\end{bmatrix}$$

It may be argued that the above example may be a special case and may not mean much. In order to address this we tried a variaty of mixing systems with different, initial values of separation matrix as well as various signal sources. The results were found to always be consistent with the above example. That is, the INFOMAX algorithm does not result in a solution for signal source separation, irrespective of signal sources being clean speech, noisy speech or noise and speech (?).

B. Simulation Results in Literature

As mentioned above signal separation could not be achieved. However, a lot of very good separation results have been reported in literature, which are seems to be contrary to the results depicted above. The question is why the published results are so good. In trying to answer this, we down loaded the original code used by [4]. Careful comparison the code used in [4] and our own code revealed that the data is pre-processed before feeding it into the adaptive learning algorithm. The pre-processing contains DC component removal and de-correlation (making the cross-correlation matrix to be 4I). Simulations were carried out for the same signal sources and mixing matricies as discussed in Section B except that now we inserted the pre-processing stage. The results are shown in Figure 3, and the resulting E matrix is

$$E=\begin{bmatrix}23.3079 & -0.0314\\ -0.5952 & 23.4237\end{bmatrix}$$

Clearly good separation has been achieved.

C. The Effect of Pre-processing

It is evident from the above that the pre-processing stage seems to be crucial in order for the INFOMAX algorithm [?] to provide a good solution. The next question is "what kind of role does the pre-processing (de-correlation) play in the process of separation? What is the effect of INFOMAX algorithm on the signal separation?" In trying to answer these two questions, let us see what happens to the measurements by the pre-processing stage. We consider the case where the measurements are zero mean:

$$\hat{X}=X-E\{X\} \tag{15}$$

and un-correlated:

$$\hat{X}=C\hat{X} \tag{16}$$

where C is the de-correlation matrix given by:

$$CC=E(XX^T) \tag{17}$$

Clearly the above pre-processing will lead to identy cross-covariance matrices:

$$E\{\hat{X}\hat{X}\}=E\{C\hat{X}\hat{X}C\}=I \tag{18}$$

Computer simulations were also performed to investigate the effect of the above pre-processing. We used the same signal sources and mixing network as those in Section B. The results are shown in Figure 4. It is clear that the signal sources are separated by the pre-processing stage. The E matrix and L are given as follows.

$$E=\begin{bmatrix}10.3024 & 0.6671\\ -0.5523 & 11.9845\end{bmatrix}$$

The above results are not surprising because de-correlation itself has already been shown to capable of achieve signal separation [5].

4 CONCLUSIONS AND CONSIDERATIONS

In conclusion the following observations have been established:

- The de-correlation in pre-processing stage plays a very important role in the separation performance. The INFOMAX algorithm itself is not able to separate signal sources.

In most cases, the de-correlation is able to separate the signal sources itself.

REFERENCES

[1]. S. Amari and A. Cichocki, "Adaptive blind signal separation --- Neural network approaches," Proc. Of IEEE, Vol.86, No.10, October 1998, pp.2026-2048.

[2]. J.-F. Cardoso, "Bind signal separation: Statistical principles," Proc. Of IEEE, Vol.86, No.10, October 1998, pp.2009-2047.

[3]. A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," Neural Computation, Vol.7, 1995, pp.1129-1159.

[4]. "Paris' Independent Component Analysis & Blind Source Separation page" http://sound.media.mit.edu/~paris/ica.html

[5]. E. Weinstein, M. feder, and A. V. Oppenheim, "Mult-channel signal separation by decorrelation," IEEE Trans. Speech and Audio Processing, Vol.1, no.4, Oct.1993, pp.405-413.

[6]. D.C.B Chan, P.J.W. Rayner and S.J.Godsill, "Multi-channel blind signal separation by decorrelation," Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Mohonk Mountain, House, New Paltz, New York, pp.15-18, Oct. 1995.

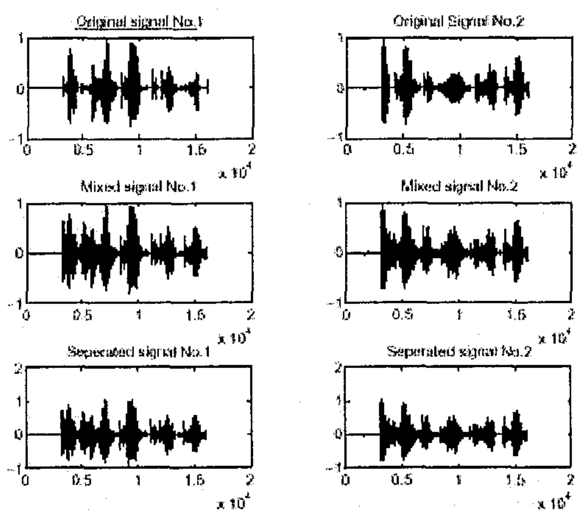

Figure 2. Simulations on INFOMAX without Decorrelation

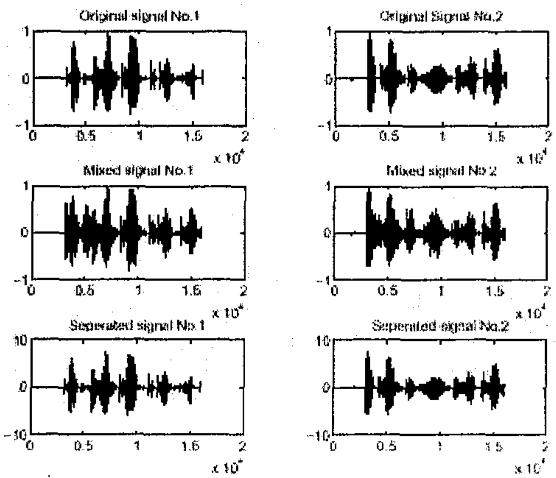

Figure 3. Simulations on INFOMAX with Decorrelation

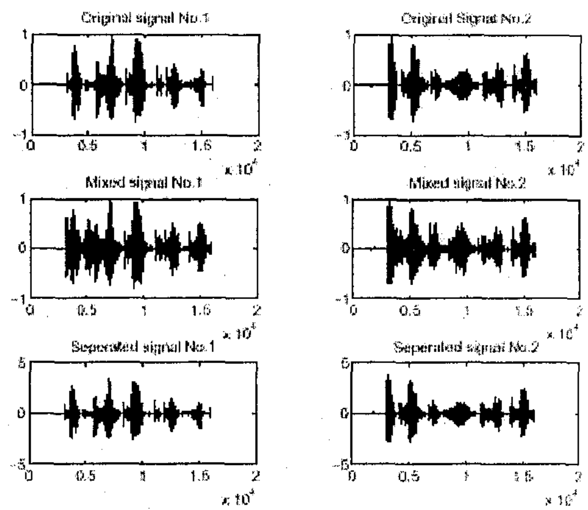

Figure 4. Separation results using decorrelation only