# An Improved Algorithm for Dynamic Independent Component Analysis

Authors:
**Evan Hamzh – University of Pitesti**
**Doru Constantin – University of Pitesti**

**Abstract**

The inner time-structure information is considered for the dynamic independent component analysis (DICA) model, and the new concept of improved and corresponding algorithm are presented. When the improved is applied in DICA, a new linear mixing model of improved DICA (IDICA) can be obtained in which the assumptions of independence and non-Gaussianity for the basic ICA model are also satisfied, and the non-Gaussianity of latent components increase. What's more, the mixing matrix (or demixing matrix) estimated from IDICA model is also available for original DICA model. Then the new algorithm of (IDICA) is presented, and the effects of IDICA are discussed. It is validated that it can accelerate the convergence rate with the increasing of the latent component's non-Gaussianity, but has little influence on the convergence accuracy. In performing improved, the adaptive filter based on Minimum Square Error (MSE) is proposed. The experimental results show the new method's efficient and available.

## I INTRODUCTION

Independent Component Analysis (ICA) is an important high-order sta- tistical (HOS) method in signal processing and widely applied in blind source separation, feature analysis [l-3]. It follows the linear system as

$$x = As \tag{1}$$

where x the random vector with elements $x_1, x_2, ..., x_n$ which are $n$ different mixtures, and s the random vector with elements $s_1, s_2, ..., s_n$ the independent and non-Gaussian sources (almost one Gaussian). Frequently ICA is also defined as the dynamic model [2]

$$x(t) = As(t) \tag{2}$$

For this model, most conventional discovering methods are based on second-order statistics (SOS), in which different time serials (sources) usually have different time and second-order spectral structures. In the case that more than two sources have the same power spectra shape, or practically noise and other interferences are introduced, the identification abilities degenerates swiftly [2]. This problem turns to another method based on HOS, for the reason that different sources are often independent and the HOS parameters are different. This is one of the origins of ICA theory, where only HOS information of different sources is considered, and the inner time structure information of each source is ignored [2, 3]. And compared with the SOS methods, ICA algorithms cost much more time [4]. Even though these two methods can both be applied in the identification of DICA model, as far as the research in [4], the time structure information is first considered as a method to promote the estimate accuracy of DICA.

In the paper, a further search is given in considering the time structure information in DICA. The concept of imroved model is described and corresponding algorithm is presented as well. First, the improved version of DICA (IDICA) model is put forward, and it shows that the assumptions of ICA are also suitable to IDICA model. It is demonstrated that the convergence accuracy does not increase since the independency of corresponding improved equals to that of dynamic independent component, which is on the contrary to the results in [4]; while the non-Gaussianity of latent independent component does increase swiftly, which can directly accelerate the convergence rate. Experimental results show that improved process is an efficient preprocessing step in DICA but has little influence on the estimate accuracy of DICA model.

## II THE IMPROVED DICA MODEL

Define $\tilde{s}(t)$ and $\tilde{x}(t)$ the improved for $s(t)$ and $x(t)$, respectively

$$\tilde{s}(t) = s(t) - E(s(t)|t, x(t-1), x(t-2), \ldots) \tag{3}$$

$$\tilde{x}(t) = x(t) - E(x(t)|t, x(t-1), x(t-2), \ldots) \tag{4}$$

That's $\tilde{s}(t)$ and $\tilde{x}(t)$ are respectively the errors of the best prediction of $s(t)$ and $x(t)$ [5]. Then the following formula can be obtained from

(2) and (3)

$$\begin{aligned} \mathrm{A}\tilde{\mathrm{s}}(t) &= \mathrm{As}(t) - \mathrm{AE}(\mathrm{s}(t)|t, \mathrm{s}(t-1), \mathrm{s}(t-2), \ldots) \\ &= \mathrm{As}(t) - \mathrm{E}(\mathrm{As}(t)|t, \mathrm{s}(t-1), \mathrm{s}(t-2), \ldots) \qquad (5) \\ &= \mathrm{x}(t) - \mathrm{E}(\mathrm{x}(t)|t, \mathrm{x}(t-1), \mathrm{x}(t-2), \ldots) \end{aligned}$$

Since A is an invertible matrix, according to information theory, we can get the following formula

$$\begin{aligned} &\mathrm{E}(\mathrm{x}(t)|t, \mathrm{x}(t-1), \mathrm{x}(t-2), \ldots) \qquad (6) \\ &= \mathrm{E}(\mathrm{s}(t)|t, \mathrm{s}(t-1), \mathrm{s}(t-2), \ldots) \end{aligned}$$

Then from (3)-(6), we can get the linear mixing model for improved form
as

$$\tilde{\mathrm{x}}(t) = \mathrm{A}\tilde{\mathrm{s}}(t) \qquad (7)$$

Here formula (7) is defined improved DICA (IDICA) model, in which the mixing matrix in (7) is the same as in (2), and mixtures and sources in (7) are the improveds of those in (2), respectively.

## III BASIC ASSUMPTIONS ANALYSIS AND IDICA ALGORITHM

Denote by MI the symbol of mutual information. From information theory [5], we can see that

$$\mathrm{MI}(s_i(t), s_j(t)) = \mathrm{MI}(\tilde{s}_i(t), \tilde{s}_j(t)), \text{pentru } 0 < i, j < m \qquad (8)$$

That means when the sources in (2) are independent, the corresponding improved forms are also mutually independent.

Denote by $\mathrm{Ng}(s_i(t))$ the non-Gaussianity measure of $s_i(t)$, and likewise $\mathrm{Ng}(\tilde{s}_i(t))$ that of $\tilde{s}_i(t)$. According to the Central Limit Theorem (CLT), the sums of non-Gaussian variables tends to be more Gaussian than the original variables [2, 5], we can get the relation between $\mathrm{Ng}(s_i(t))$ and $\mathrm{Ng}(\tilde{s}_i(t))$:

$$\mathrm{Ng}(\tilde{s}_i(t)) = \mathrm{Ng}(s_i(t)) \qquad (9)$$

It means that the non-Gaussianity increases when a certain approximate of improved process (such as the AR model) is selected, and the final effect mainly depends on the practical implementation.

Now from (8) and (9), we can see that the IDICA model also satisfies the basic assumptions of basic ICA, that is the latent improved form of sources are mutually independent and of non-Gaussian distribution (at most one Gaussian).

Therefore, to recover sources $s_1, s_2, \dots, s_n$ from mixtures $x_1, x_2, \dots, x_n$ in DICA, we can first estimate the mixing matrix A from the IDICA model, and put A into DICA model. Thus the following algorithm of IDICA can be obtained in the following steps.

1) Get the O-mean, I-variance form of mixtures $\mathrm{x}(t)$;

2) Calculate the improved process of $\mathrm{x}(t)$;

3) Estimate the mixing matrix A (or the separation matrix W) from the IDICA model by traditional ICA algorithm;

4) Put A or W into DICA model, then obtain the estimation $\tilde{\mathrm{s}}(t)$ of original sources $\mathrm{s}(t)$.

Here the improved form can be given in several forms, such as the first-order or second-order differential filter [5]. The aim is to keep the independence of time serial's sampling. Some nonlinear methods can also be considered as MPL and REF, while it also makes the calculation much more complex [4]. Here, an adaptive filter based on the Minimum Square Error (MSE) [5] is proposed.

The iteration of $\tilde{\mathrm{x}}(t)$ is given as

$$\begin{aligned} \tilde{\mathrm{x}}(1) &= \mathrm{x}(1) \\ \tilde{\mathrm{x}}(t) &\leftarrow \mathrm{x}(t) - b(t)\mathrm{x}(t-1) \ \text{ pentru } 2 \leq t \leq n \end{aligned} \tag{10}$$

in which parameter $b(t)5$ is as

$$\begin{aligned} b(2) &= 1 \\ b(t+1) &\leftarrow b(t) + \eta \frac{\partial J(b(t))}{\partial b(t)} \\ &= b(t) - \eta \tilde{\mathrm{x}}(t)^{\mathrm{T}} \mathrm{x}(t-1) \end{aligned} \tag{11}$$

where $J(b(t)) = \tilde{\mathrm{x}}(t)^{\mathrm{T}}\tilde{\mathrm{x}}(t) = \tilde{\mathrm{x}}(t)^{\mathrm{T}}(\mathrm{x}(t) - b(t)(x)(t-1)), 2 \leq t \leq n$, and $\eta$ the factor of step size.

# IV CONVERGENCE PERFORMANCE ANALYSIS

For simplicity, the object function of kurtosis form is selected, which is a special case of negentropy when the nonlinear activation function is selected as $g(x) = x^3$ [2]. Denote by $\mathrm{kurt}(s_i)$ as the kurtosis of $s_i$ likewise $\mathrm{kurt}(y)$ the kurtosis of $y$. And each of the independent components is given in the estimation [2,6] by

$$y = \mathrm{w}^{\mathrm{T}}\mathrm{A}\mathrm{s} = \mathrm{q}^{\mathrm{T}}\mathrm{s}\,,\quad 1 \le \boldsymbol{i} \le m \qquad (12)$$

The object function is as follows:

$$\begin{cases} J = |\,kurt(y)\,| \\ \sum_{i=1}^{m} q_i^2 \end{cases} \qquad (13)$$

Optimal algorithms are considered, including the stochastic gradient algorithm and the popular fast fixed-point algorithm [1, 2].
For the stochastic gradient algorithm

$$\Delta \mathrm{q} \;\leftarrow\; \eta\,\mathrm{sign}(\mathrm{kurt}(\mathrm{q}^{\mathrm{T}}\mathrm{s}))\;\frac{\partial \mathrm{kurt}(\mathrm{q}^{\mathrm{T}}\mathrm{s})}{\partial \mathrm{q}(t)} \qquad (14)$$

When $q_i \approx \pm 1$, $q_j \to 0$, $j = 0(0 < i, j \leq m)$ and $q(t)$ is fixed, the iteration is

$$\begin{aligned} \Delta q_i &\leftarrow \eta\, 4\mathrm{sign}(\mathrm{kurt}(q_i^T s))q^3\mathrm{kurt}(s_i) \\ \Delta q_i &\infty\ \mathrm{kurt}(s_i) \end{aligned} \tag{15}$$

For the fast fixed-point algorithm

$$\begin{aligned} \Delta q_i &\leftarrow 4q_i^3\mathrm{kurt}(s_i) \\ \Delta q_i &\infty\ \mathrm{kurt}(s_i) \end{aligned} \tag{16}$$

Formulas (15) and (16) show that the increasing of non-Gaussianity does lead to a faster convergence.

As for the convergence accuracy of $q$, we know that it mostly relies on the convergence condition, the convergence threshold, which also decides the estimate accuracy of set). As the improved process does not change the convergence condition, IDICA also has no other special influence on the convergence accuracy [6]. This opinion is on the contrary to that in [5] (the accuracy increases with the increasing of non-Gaussianity).

## V EXPERIMENTAL RESULTS

The original sources are representations of 2 human faces, shown in Fig. 1. In Fig. 2 are illustrated the mixtures with linear mixing matrix A.

Fig. 1. The original sources

Fig. 2. The mixtures.

First we try to recover the sources from mixtures by applying FastICA. The results are presented in Fig. 3.

$$A = \begin{pmatrix} 0.3 & 0.1 \\ 0.3 & 0.4 \end{pmatrix}$$

Fig. 3. The components estimated by FastICA.

Then it was implemented the improved process and the results are as depicted in Fig. 4.

Fig. 4. Improved process of mixtures.

Fig. 5 reveals the independent components obtained after applying IDICA.

Fig. 5. Independent components by IDICA.

Changing the mixing matrix, and repeating the experiments many times, the results are similar.

# VI DISCUSSIONS

The new concept of improved process is introduced in dynamic ICA model, and the corresponding algorithm IDICA is put forward. It shows that the IDICA is a promising method for DICA model for its outstand- ing performance of better convergence rate, but has little influence on the estimate accuracy.

However, it should be pointed out that in information theory, improved process is a concept, and not a fixed and practical form that can be used directly. Different realization of improved process may bring different results. And the final estimate performance is also related with the sources, especial the independent assumption [6].

# References


[1] P. Comon, ”Independent component analysis - a new concept”, Signal
Processing, Vo1. 36, No. 3, pp. 287-314, 1994.

[2] A. Cichocki, S. Amari, Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications, Wiley, 2003

[3] A. Hyvarinen, ”Independent Component Analysis for Time - dependent Stochastic Processes”, http://www.bsp.brain.rikenjp/ICApub/TNN99.pdf

[4] A. Hyvarinen, ”Fast and Robust Fixed-Point Algorithms for Independent
Component Analysis”, IEEE Transactions on Neural Networks Vol. 10 No. 3, pp. 626-634, 1999.

[5] S. Haykin, Adaptive filter theory, Prentice-Hall International, 3rd edition, 1996.

[6] G. Wang, X. Xu, D. Hu, ”Global convergence of FastICA: theoretical analysis and practical considerations, ”ICNC2005: 700-705, 2005.