Blind Signal Separation Using an Adaptive Generalized Continuous Distribution

Mohamed E. Waheed, Mohamed-H MOUSA, Mohamed-K HUSSEIN

Abstract. In this paper, we present an algorithm for the problem of independent component analysis (ICA) which can separate mixtures of sub- and super-Gaussian probability density distributions using a generalized continuous distribution source model. We use neural network representation to model the mixer and demixer respectively, and show how the parameters of the demixer can be adapted using a gradient descent algorithm incorporating the natural gradient extension. We also present a learning method for the unknown parameters of the generalized exponential source model. The nonlinear function in ICA algorithm is self-adaptive and is controlled by the shape parameters of generalized exponential density model. Computer simulation results confirm the validity and high performance of the proposed algorithm.

Keywords: Independent component analysis, Generalized exponential distribution, Maximum likelihood, sub- and super-Gaussian, Blind signal separation.

I. INTRODUCTION

The problem of independent component analysis (ICA) has received wide attention in various fields such as biomedical signal analysis and processing (EEG, MEG, ECG), geophysical data processing, data mining, speech recognition, image recognition and wireless communications (Babu and Narasimhan, 2012; Peng et al., 2009; Liu and Cheung, 2008; Amari and Cichocki, 1998; Amari et al., 1997; Bell and Sejnowski, 1995; Cichocki et al., 1997; Gardner, 1991; Slock, 1996). In many applications, the sensory signals (Observations obtained from multiple sensors) are generated by a linear generative model which is unknown to us. In other words, the observations are linear instantaneous mixtures of unknown source signals and the objective is to process the observations in such a way that the outputs correspond to the separate primary source signals. The operation starts with a random source vector \( U(n) \) defined by \( U(n) = [U_1, U_2, \ldots, U_m] \) where the \( m \) components are supplied by a set of independent sources. Temporal sequences are considered here; henceforth the argument \( n \) denotes discrete time. The vector \( U \) is applied to a linear system whose input-output characterization is defined by a nonsingular \( m \)-by-\( m \) matrix \( A \), called the mixing matrix. The result is an \( m \)-by-1 observation vector \( X(n) \) related to \( U(n) \) as follow \( X = AU \) where \( X = [X_1, X_2, \ldots, X_m]^T \). The source vector \( U \) and the mixing matrix \( A \) are both unknown. The only thing available to us is the observation vector \( X \). Given \( X \), the problem is to find a demixing matrix \( W \) such that the original source vector \( U \) can be recovered from the output vector \( Y \) defined by \( Y = WX \), where \( Y = [Y_1, Y_2, \ldots, Y_m]^T \). This is called the blind source separation. The solution to the blind source separation is feasible, except for an arbitrary scaling of each signal component and permutation of indices. In other words, it is possible to find a demixing matrix \( W \) whose individual rows are a rescaling and permutation of those of the matrix \( A \). That is, the solution may be expressed in the form \( Y = WX = WAU \rightarrow DPU \) where \( D \) is a nonsingular diagonal matrix and \( P \) is a permutation matrix.

Since Karhunen (1996) proposed a linear feedback network with a simple unsupervised learning algorithm, several methods have been developed. Cichocki et al. (1994) and Cichocki and Unbehauen (1996) proposed a robust, flexible algorithm with equivariant properties. Comon (1994) gave a good insight to ICA problem from the statistical point of view. Bell and Sejnowski (1995) adopted an information maximization principle to find a solution to ICA problem. Maximum likelihood estimation (Guidorzi et al., 2007; Diversi et al., 2005; Douglas et al., 1997; Pham et al., 1992) was proposed by Pham et al. (1992). The nonlinear extension of PCA was extensively studied in (Karhunen, 1996; Oja, 1995). Serial updating rule was introduced by Cardoso and Laheld (1996) and the resulting algorithm was shown to have equivariant performance. The natural gradient was proposed and applied to ICA by Amari et al. (1996). Conditions on cross cumulants for the separation of the source signals were investigated in (Peng et al., 2009; Liu and Cheung, 2008; F. Alberg et al. 2002; Babu and Narasimhan, 2012)

II. DERIVATION OF NATURAL GRADIENT BASIC LEARNING RULES

Let us consider a linear feedforward memoryless neural network which maps the observation \( X(n) \) to \( Y(n) \):

\[ Y(n) = WX(n) \]  \hspace{1cm} (1)

Where \((i,j)\)th element of the matrix \( W \), i.e., \( W_{ij} \) represents a synaptic weight between \( y(n) \) and \( x(n) \). In the limit of zero noise, for the square ICA problem (equal number of sources and sensors, the result can be extended to the case \( m > n \)) maximum likelihood or mutual information minimization suggest the following loss function A. Cichocki, et al. (1994).

\[ L(W) = -\log|\det W| - \sum_{i=1}^{m} \log p_i(y_i(n)), \]  \hspace{1cm} (2)
Where \( p_i(.) \) represent the probability density function. Let us define
\[
f_i(y_i(n)) = -\frac{d\log p_i(y_i(n))}{dy_i(n)} \tag{3}
\]
With this definition, the gradient of the loss function, Equation (2), is as follows:
\[
\nabla L(W(n)) = \frac{\partial L(W(n))}{\partial W(n)} = -W^{-T}(n) + f(y(n))X^T(n),
\]
Where \( f(y(n)) \) is the element-wise function whose \( i \)th component is \( f_i(y_i(n)) \). The natural Riemannian gradient (denoted by \( \nabla L(W(n)) \)) learning algorithm for \( W(n) \) is given by (S. Amari, 1998; J. F. Cardoso and B. H. Laheld, 1996; A. Cichocki and R. Unbehauen, 1996).
\[
W(n+1) = W(n) - \eta_1 \nabla L(W(n)) = W(n) - \eta_1 \frac{\partial L(W(n))}{\partial W(n)} W^T(n)W(n) \tag{5}
\]
\[
= W(n) + \eta_1 \left[ I - f(y(n))y^T(n) \right]
\]

III. GENERALIZED EXPONENTIAL MODEL FOR SOURCES

Optimal nonlinear activation function \( f_i(y_i(n)) \) is calculated by Equation (3). However, it required the knowledge of the probability distribution of source signals which are not available to us. A variety of hypothesized density model has been used. For example, for the sup-Gaussian source signals, unimodal or hyperbolic-Cauchy distribution model A. Bell and T. Sejnowski, (1995) leads to the nonlinear function given by:
\[
f_i(y_i(n)) = \tanh(\beta y_i(n)) \tag{6}
\]
Such sigmodal function was also used in A. Bell and T. Sejnowski, (1995). For sub-Gaussian source signals, cubic nonlinear function \( f_i(y_i(n)) = y_i^3(n) \) has been a favorite choice. For Mixtures of Sub- and super-Gaussian source signals, according to the estimated kurtosis of the expected signals, nonlinear function can be elected from two different choices (Guidorzi et al., 2007; R. Diversi, 2005; S. C. Douglas, et al. 1995). (for example, either \( f_i(y_i(n)) = y_i^2(n) \) or \( f_i(y_i(n)) = \tanh(\beta y_i(n)) \).)

This paper present a flexible nonlinear function derived using generalized exponential density model. It will be shown that the nonlinear function is self-adaptive and controlled by generalized exponential shape parameters. It is not a form of fixed nonlinear function.

A. The generalized exponential distribution

The generalized logistic probability distribution is a set of distributions parameterized by a positive real numbers \( p,q \) which is usually referred to as the shape parameters of the distribution. The shape parameters \( p,q \) control the peakiness of the distribution. The probability density function (PDF) for generalized exponential is described by
\[
p(y; p, q) = \frac{\exp(-qy)}{B(p,q)(1+\exp(-y))^{p+q}}, \tag{7}
\]
\(-\infty < y < \infty, p, q > 0 \)
The distribution of \( y \) now depend on the shape parameters \( p, q \). The plots of the density function in (7) at \( p=1, q=1...9 \) and \( p=1...9, q=1, 2 \) are presented in Figures 1 and 2 respectively.

Fig 1. The Plots Of The Generalized Exponential Density Function For P=1, Q=1,…,9

Fig 2. The Plots Of The Generalized Exponential Density Function For P=1...9, Q=1, 2
B. The Moments Of Generalized Exponential Distribution

In order to fully understand the generalized exponential distribution, it is useful to look at its moments (specially 2\textsuperscript{nd} and 4\textsuperscript{th} moments which give the kurtosis). The moment generating function of generalized exponential distribution is given by

$$E(e^{ty}) = \int_0^\infty e^{ty} p(y; \lambda, \alpha) dy = \frac{\Gamma(p+q)\Gamma(q-\theta)}{\Gamma(p)\Gamma(q)} e^{\theta t}, \quad -\lambda < \theta < q.$$  

(8)

Then

$$M_2 = E(y^2) = \frac{\Gamma(q)\Gamma(q-\theta)}{\Gamma(p)\Gamma(q)} - 2q\psi(p)\psi(q) \frac{\Gamma(q)}{\Gamma(p)} \quad (9)$$

$$M_3 = \frac{\Gamma(3)(p\Gamma(p)+\Gamma(p))}{\Gamma(p)\Gamma(q)} \bigg(-3\frac{\Gamma(q)}{\Gamma(p)}(p\Gamma(p)+\Gamma(p)) + 3\frac{\Gamma(1)(p\Gamma(p)+\Gamma(p))}{\Gamma(p)\Gamma(q)} \bigg) \quad (10)$$

$$M_4 = \frac{\Gamma(4)(p\Gamma(p)+\Gamma(p))}{\Gamma(p)\Gamma(q)}\bigg(-4\frac{\Gamma(q)}{\Gamma(p)}(p\Gamma(p)+\Gamma(p)) + 4\frac{\Gamma(1)(p\Gamma(p)+\Gamma(p))}{\Gamma(p)\Gamma(q)} + 6\frac{\Gamma(2)(p\Gamma(p)+\Gamma(p))}{\Gamma(p)\Gamma(q)} \bigg) \quad (11)$$

the moment ratios, coefficient of variation, and standard cumulants $k_r = \frac{k_r}{k_2^{\frac{1}{2}}}$ of the standard distribution in (8) are of course the same as those of the distribution in Equation (7).

C. Kurtosis and shape parameters

The kurtosis is a non-dimensional quantity. It measures the relative peakness or flatness of a distribution. A distribution with positive kurtosis is termed leptokurtic (super-Gaussian). A distribution with negative kurtosis is termed platykurtic (sub-Gaussian). The kurtosis of the distribution is defined in terms of the 2\textsuperscript{nd}- and 4\textsuperscript{th}-order moment as

$$k(y) = \frac{E((y-\mu)^4)}{E((y-\mu)^2)^2} = \frac{M_4 - 4M_3 + 6M_2 - 3M_1}{M_2 - M_1^2}.$$  

(12)

Where the constant term –3 makes the value zero for the standard normal distribution. For Weibull distribution, the kurtosis can be expressed in terms of the shape parameter, for generalized exponential distribution in Equation (3) as,

$$f_1(y_i) = -q_i + (p_i + q_i) \frac{\exp(-y_i)}{1 + \exp(-y_i)}.$$  

(13)

$$\frac{\partial \Delta p}{\partial q_i} = -q_i + (p_i + q_i) \Psi(p_i + q_i) - \Psi(p_i) - \ln(1 + \exp(-y_i))$$  

(14)

$$\frac{\partial \Delta q_i}{\partial q_i} = -y_i + \Psi(p_i + q_i) - \Psi(p_i) - \log(1 + \exp(-y_i))$$  

(15)

$$\Delta q_i = -(\Psi(q_i + q_i) - \Psi(q_i)) - \log(1 + \exp(-y_i))$$  

(16)

$$\Delta q_i = -\eta_1 \frac{\partial \Delta q_i}{\partial q_i} = -(\Psi(q_i + q_i) - \Psi(q_i)) - \log(1 + \exp(-y_i))$$  

(17)

Simulation Algorithm

1- Start with some initial point W (n) in the multi-dimensional parameter space

2- Obtain the gradient value $\nabla L(W(n))$

3- Compute the value W (n+1) by moving from W(n) along the gradient descent, i.e. along $\nabla L(W(n))$

$$W(n+1) = W(n) + \eta_1 [I - f(y(n))]y^2(n)$$

4- Test the stability of the parameters, i.e. if $|W(n+1) - W(n)| < \varepsilon$ (threshold)

IV. COMPUTER SIMULATION RESULTS

Example 1: Consider the system involving the following three independent sources

$$U_1(n) = \cos(0.8\pi n)\cos(0.1\pi n)$$

$$U_2(n) = 60^\circ \sin(20^\circ \pi^\circ n)\exp(-6n)$$

(18)

$$U_3(n) = \text{a periodic Gaussian pulse signal at 10 kHz, with 50\% bandwidth. The pulse repetition frequency is 1 kHz, sample rate is 50 kHz, and pulse train length is 10\text{ms}. The repetition amplitude should attenuate by 0.8 each time.}$$

The mixing matrix A is

$$A = \begin{bmatrix} 0.56 & 0.79 & -0.37 \\ -0.75 & 0.65 & 0.86 \\ 0.17 & 0.32 & -0.48 \end{bmatrix}.$$  

(19)

The algorithm was implemented using the following conditions:

- Initialization. The weights in the demixing matrix W were picked from a random number generator with a uniform distribution inside the range [0.0, 0.5].
- The learning rate parameter was fixed at $\eta = 0.1$
- Signal duration. The time series produced at the mixer output had a sampling period 10\text{ms} and contained N = 65,000 samples.

![Fig 3. The original and demixed signals](image-url)
Figure (3) displays the waveforms of the source signals and the signals produced at the output of the demixer. It can be observed that after 3000 iterations, source signals are well separated.

**Example 2:** We assume that the mixing system is an ARMA model of order 10, which is stable and minimum phase. The transfer function of the mixing system is plotted in Figure 4.

The demixing system is chosen to be a state space system of order 40 and the initial values for $W(n)$ and $\eta$ are in the form (5). From the simulation we see that if we do not update $W(n)$ and $\eta$, the outcome of the recovered signal cannot be perfect. After training $A$ by using algorithm (1), we can recover source signals quite well. Figure 5 plots the global transfer function $G(z) = W(z)H(z)$ up to order 60. From computer simulations we see that the overestimation of system order $N$ essentially do not the outcome of the learning algorithm, but it only increases the computing cost.

**Example 3:** In this experiment, we synthesize two source signals, from 6 sources: 4 signals $S'_i$, $i = 2,...,5$ are uniformly distributed, while the other two signals $S'_1$ and $S'_3$ are segments of speech signals are super-Gaussian while the uniformly distributed signals are super-Gaussian:

$$S'_1 \quad \text{if} \quad t < 1000$$

$$S'_2 \quad \text{if} \quad 1001 \leq t < 6000$$

$$S'_3 \quad \text{if} \quad 6001 \leq t < 7000$$

$$S'_4 \quad \text{if} \quad t < 2000$$

$$S'_5 \quad \text{if} \quad 2001 \leq t < 6000$$

$$S'_6 \quad \text{if} \quad 6001 \leq t < 8000$$

$$S'_7 \quad \text{if} \quad 8001 \leq t < 12000$$

Thus our source signals are non-stationary. We designed the source signals so that we can test the adaptability of the algorithms to the variation of signals. The observed signals $u(\square)$ are obtained by passing the source signals through a mixing linear dynamical system defined in (1), and (2), where the source number $\square = 2$, and system order $\square = 2$. The system matrix $W_{ij}$ is randomly selected, except that we guarantee $D^{-1}$ exists and the eigenvalues of $\square$ are within the unit circle.

**Fixed nonlinearity approach**

In this section, we will apply the fixed nonlinearity approach (J. F. Cardoso and B. H. Laheld, 1998) to recover the two non stationary signals. The results are shown in Figures 6 and 7. The Lee et al.’s Switching Nonlinearity method, we present results (Figures 8 to 10) in applying the method derived in Section 3, and the generalized exponential model approach, we will present
results using the generalized exponential model approach as proposed in this paper (Figures 11 to 13). Table 1 gives the mean squared errors among the three methods. It is noted that the generalized exponential model gives the smallest mean squared errors.

Table 1. A Table Showing the Mean Squared Errors As Obtained By Each of the Three Methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>Signal 1</th>
<th>Signal 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed nonlinearity model</td>
<td>0.0004</td>
<td>0.0001</td>
</tr>
<tr>
<td>Lee et al</td>
<td>0.0039</td>
<td>0.0021</td>
</tr>
<tr>
<td>Generalized exponential model</td>
<td>0.0002</td>
<td>0.00005</td>
</tr>
</tbody>
</table>

Fig 7. Signals of the fixed nonlinearity approach. The top two graphs show the sources; the second set of two graphs show the separated recovered signals; the next set of two graphs show the error between the recovered signals and the source signals; while the lowermost two graphs show the output of the mixer.

Fig 10. The switching function in Lee et al.’s method as estimated using the algorithm indicated in this paper. The upper graphs give the functions before the sign function is applied as indicated in Equation (12).

Fig 8. Parameter variations of D of Lee et al’s method.

Fig 11. Parameter Variations of D of Generalized Exponential Model.

Fig 9. Signals of Lee et al’s method. For explanation of these graphs please see caption of Figure 7.

Fig 12. Signals of the Generalized Exponential Model. For An Explanation Of The Graphs Please See The Caption Of Figure 7.
In this paper, we consider the possibility of “automatic” adaptation to the source model, whether it is super-Gaussian or sub-Gaussian. We use the fixed nonlinearity approach S. Choi and A. Cichocki, 1997; as the baseline for comparison. We extend the generalized exponential source model proposed in S. Amari, et al. 1997; to the MBD case. In addition, we have extended the switching nonlinearity method in A. Bell and T. Sejnowski, 1995; to the MBD case as well. The three methods are all applied to two synthesized signals made up of segments of speech signal and uniformly distributed signal. It is quite surprising to observe that the method proposed in Choi and A. Cichocki, 1997; works well even though it was not designed to work with non stationary signals. The method proposed in A. Bell and T. Sejnowski, 1995; works well, with the nonlinearities switching based on an estimation of the kurtosis of the recovered signals. The proposed generalized exponential model works best in that it gives the smallest mean squared error.

REFERENCES


