MULTICHANNEL BLIND DECONVOLUTION USING A GENERALIZED GAUSSIAN SOURCE MODEL

A. S. Abu-Taleb¹, E. M. E. Zayed², W. M. El-Sayed², A. M. Badawy¹ and O. A. Mohammed²

¹Dept. of System and Bioengineering. Faculty of Engineering, Cairo University, Egypt. ²Dept. of Mathematics, Faculty of Science, Zagazig University, Egypt. <u>mewahed@yahoo.com</u>, <u>usama2005</u> <u>1969@yahoo.com</u>

Abstract– In this paper, we present an algorithm for the problem of multi-channel blind deconvolution which can adapt to un-known sources with both sub-Gaussian and super-Gaussian probability density distributions using a generalized gaussian source model. We use a state space 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 Gaussian source model. The performance of the proposed generalized Gaussian source model on a typical example is compared with those of other algorithm, viz the switching nonlinearity algorithm proposed by Lee *et al.* [8].

Keywords– blind deconvolution, blind source separation, generalized Gaussian source model, multichannel blind deconvolution.

1. INTRODUCTION

Recently, the problems of Blind Source Separation (BSS) or Independent Component Analysis [2, 3, 9], and multichannel blind deconvolution (MBD) [6, 10] have been extensively studied. For BSS problems, source signals are instantaneously mixed by an unknown matrix, while for MBD problems, the source signals are mixed through a series of unknown dynamical systems. Since the mixing involves dynamical systems in MBD problems, delay is inevitably present, thus rendering it more complex to analyze than BSS. In BSS and MBD problems, both the sources and the mixing system are unknown, it is intuitively clear that the sources cannot be recovered without some necessary restriction on the nature of the sources.

Assumptions on the nature of the sources. The usual assumptions [2] place on the problems of BSS and MBD are as follows:

- 1. The source signals are statistically independent.
- 2. At the most one of the source signals is Gaussian distributed.
- 3. It is only possible to recover the source signals modulo scale, and polarity.

In addition, in this paper, we assume that the dynamical mixing systems in MBD problems are linear time invariant (LTI) systems which are modeled by constant parameters. There are many algorithms for tackling BSS and MBD problems. A popular assumption on the source signals is that they have sharply symmetrical distributions, i.e., the source signals are super-Gaussian with positive kurtosis. The classic Bell-Sejnowski infomax algorithm [2] works well when it is applied to mixtures of super-Gaussian signals, but it becomes inefficient when the mixtures include sub-Gaussian

signals. The reason of the inefficiency in source separation in this case is that the assumed source model is invalid for the sub-Gaussian sources. This gives impetus to find more appropriate source models to BSS and MBD problems. Most researchers focus on modelling the sources with symmetrical unimodal probability density distributions because they are simple to analyze and they can model the source signals well in practice. Douglas et al. [6] presented a simple and efficient extension of a family of algorithms in BSS and MBD problems with mixtures of arbitrary non-Gaussian sources. Their algorithms [6] monitor the statistics of each output signal of the demixer, then selects the appropriate nonlinearity for each recovered source. Their algorithm is based on some necessary BSS stability conditions [3]. Choi et al. [4] derived a learning algorithm, called flexible ICA, with a "flexible" nonlinearity. This nonlinear function is controlled by the Gaussian exponent according to the estimated kurtosis of the recovered signals. Hence the algorithm can successfully separate the mixture of both super-Gaussian and sub-Gaussian sources simultaneously. In Lee et al. [8], an extension of the infomax algorithm is derived which is able to separate both super- and sub-Gaussian independent components. A symmetrical strictly sub-Gaussian density is modeled using a symmetrical form of Pearson model or hyperbolic Cauchy density model. A switching criterion based on BSS stability analysis [3] is obtained. The above mentioned methods work well as they assume that the characteristics of the signals do not alter rapidly. In these methods, there is a certain detection latency as most of them require a finite time to elapse before they can react to the characteristics of the recovered source signals. For example, in [4] the kurtosis of the recovered signals need to be estimated while in [6], statistics of the recovered signals need to be monitored. Hence for signals whose characteristics vary rapidly, the above mentioned methods might not be sufficiently sensitive. In this paper, we will consider the generalized Gaussian source model originally proposed in [7] to "automatically" estimate the source signals in the MBD problems. Here "automatic" is taken to mean that our algorithm will estimate the parameters of the generalized gaussian source model from the recovered signals directly. The nonlinearity used in the separation of signals is a function of these estimated parameters. The parameters of the generalized gaussian source model can be easily interpreted, thus adding transparency to the estimated parameters. Hence the proposed algorithm works more "flexibly" in its ability to adapt to the changing nature of the source signals. The paper is organized as follows: State space Approach to Multichannel Blind Deconvolution is described in Section 2. For the sake of completeness, in Section 3 we include a brief description of the switching nonlinearity algorithm [10] adapted to the MBD case. This section also contains a special form of the switching nonlinearity algorithm, viz., the fixed nonlinearity algorithm. Generalized gaussian Source Model is introduced in Section 4. Some experimental results are given in Section 5. Finally we draw some conclusions in Section 6.

2. STATE SPACE APPROACH TO MULTICHANNEL BLIND DECONVOLUTION

Given a vector of observed signals u(k); $k \in [0,N]$, we wish to recover the source signals s(k) based on the assumption that the sources are statistically Independent. If we assume that the observations are convolutive version of the sources, the problem can be tackled using state space approach [10]. We consider the state space

approach instead of the transfer function approach, as the state space approach can be easily extended to nonlinear mixing systems. Moreover, the state space approach not only gives an efficient internal description of the dynamic systems, but also there exist different possible equivalent state space realizations, for instance, canonical controller form [10] which allows us to find "efficient" representations of the demixer . We model the mixing environment of the MBD problem as follows:

$$x(k+1) = Ax(k) + Bs(k) + L\xi_{p}(k)$$
(1)

$$u(k) = \overline{C}x(k) + \overline{D}u(k) + \theta(k)$$
⁽²⁾

where $s \in \mathbb{R}^n$, $u \in \mathbb{R}^n$, and $x \in \mathbb{R}^{\overline{N}}$ are the source signals, the observations, and the state of the LTI dynamical system, respectively. The system matrices $\overline{A} \in \mathbb{R}^{\overline{NxN}}$, $\overline{B} \in \mathbb{R}^{\overline{Nxn}}$, $\overline{C} \in \mathbb{R}^{nx\overline{N}}$, and $\overline{D} \in \mathbb{R}^{nxn}$, which are assumed to be constant, are state mixing matrix, input mixing matrix, output mixing matrix, and input-output mixing matrix respectively. The system matrices in the demixer are similarly defined. For simplicity, we assume the number of source n equals to the number of sensors. Normally, the system order N is unknown; we need to estimate its value from the observed data. Here we assume that the system order is "known". Correspondingly, we model the demixer using a similar discrete time dynamical system:

$$\mathbf{x}(\mathbf{k}+1) = \mathbf{A}\mathbf{x}(\mathbf{k}) + \mathbf{B}\mathbf{u}(\mathbf{k}) + \mathbf{L}\boldsymbol{\xi}_{R}(\mathbf{k})$$
(3)

$$\mathbf{y}(\mathbf{k}) = \mathbf{C}\mathbf{x}(\mathbf{k}) + \mathbf{D}\mathbf{u}(\mathbf{k}) \tag{4}$$

where the input vector $u \in \mathbb{R}^n$ is the output of the mixer, $y \in \mathbb{R}^n$ is the recovered signal vector, and $x \in \mathbb{R}^N$ is the system states of the demixer. For successful separation and deconvolution, we need $N \ge \overline{N}$. Here we assume both the mixer and the demixer exist, in particular, D^{-1} exists. The condition of the existence of solution in multichannel blind deconvolution is studied in [10].



Fig.1. General linear state space model for blind deconvolution

We measure the dependence among the recovered sources y using mutual information. Given P(y), the probability density function (PDF) of the recovered signal vector y, the mutual information between the recovered signals can be defined as follows:

$$I(y) = \int p(y) \frac{p(y)}{\prod_{k=1}^{n} p(y_k)} dy$$
(5)

A. S. Abu-Taleb, E. M. E. Zayed, W. M. El-Sayed, A. M. Badawy and O. A. Mohammed

$$= -H(y) + \sum_{k=1}^{n} H(y_{q}),$$
(6)

where H(y) = -E[log(P(y)]] is the entropy of y, $H(y_q) = -E[log(p(y_q))]$ is the marginal entropy of y_q . For simplicity, for the remaining part of this paper, the time index k is dropped if there is no risk of confusion. Observe that I(y) = 0, and I(y) = 0 if and only if the components of vector y are statistically independent. Therefore I(y) is an appropriate measurement of the dependence among the recovered signals. Unfortunately, mutual information is difficult to compute explicitly, hence we use a cost function similar to [10]:

$$\max_{\theta} I(y,\theta) = \max_{\theta} \left[\log |\det(D)| - \sum_{q=1}^{n} \log p(y_q) \right]$$
(7)

where θ is the set of system parameters and source model parameters, which we will study in Section 3 and 4, det is the determinant. There exist various ways to tackle the optimization problem [10]. Here we follow the derivation of information back-propagation approach given in [8].

2.1. Gradient-based learning rules

Based on the cost function (7), we can easily obtain the following updating rules. For matrices D and

$$D(k+1) = D(t) + \eta(k)(I - \varphi(y)u^T D^T)D$$

$$C(k+1) = C(k) - \eta(k)\varphi(y)x^T$$
(8)

where $\varphi(y)$ is a vector nonlinearity related to the source model. This will be discussed further in Section 3. Note, natural gradient [1] is used in (8) to improve the performance of the learning process. Similarly, for matrices A and B, we have:

$$A_{ij}(k+1) = A_{ij}(k) - \eta(k)\varphi^T(y)\sum_{l=1}^N C_{\bullet l} \frac{\partial x_l}{\partial A_{ij}}$$
(9)

$$B_{iq}(k+1) = B_{iq}(k) - \eta(k)\varphi^{T}(y)\sum_{l=1}^{N} C_{\bullet l} \frac{\partial x_{l}}{\partial B_{iq}}$$
(10)

where $l, i, j = 1, 2, ..., N, q = 1, 2, ..., n, C_{\bullet l}$ denotes the l-th column vector of the matrix C $\frac{\partial x_l}{\partial A}, \frac{\partial x_l}{\partial B}$ can be obtained from following:

$$\frac{\partial x_l(k+1)}{\partial A_{ij}} = \sum_{m=1}^N A_{lm} \frac{\partial x_m(k)}{\partial A_{ij}} + \delta_{li} x_j$$
(11)

$$\frac{\partial x_l(k+1)}{\partial B_{iq}} = \sum_{m=1}^N A_{lm} \frac{\partial x_m(k)}{\partial B_{iq}} + \delta_{li} u_q$$
(12)

where δ_{li} is the Kronecker delta function.

3. LEE ET AL. SWITCHING NONLINEARITY ALGORITHM

For the sake of completeness, we will briefly describe the algorithm due to Lee et al. [8] on detecting the kurtosis of the recovered signals, and then switch the nonlinearity based on this information. We will follow closely the development in [8], but adapting their notation to the notations in this paper for its applications to MBD. Consider the learning equation for D, as shown in (8). Consider the reconstructed signal y. In general, a symmetrical form for modelling a strictly sub-Gaussian density can be obtained using Pearson's mixture model:

$$p(y_q) = \frac{1}{2} \left[N(\mu, \sigma^2) + N(-\mu, \sigma^2) \right]$$
(13)

where $N(\mu, \sigma^2)$ denotes the Gaussian distribution with mean μ and variance σ^2 . For various values of μ this can be either a unimodal distribution, or a bimodal distribution. The nonlinearity in (8) can be expressed as:

$$\varphi(y) = -\frac{\partial \log p(y_q)}{\partial y_q} = \frac{y_q}{\sigma^2} - \frac{\mu}{\sigma^2} \tanh\left(\frac{\mu}{\sigma^2}y_q\right)$$
(14)

if $\mu = 1$, and $\sigma^2 = 1$, and substituting this into (8), we have:

$$D(k+1) = D(k) + \eta(k)(I + \tanh(y)u^{T}D^{T} - yu^{T}D^{T})D$$
(15)

This is the learning algorithm derived by assuming that the source model is a sub-Gaussian model. And in a similar fashion, we can derive a learning rule for super-Gaussian distributions. The two can be combined together to give:

$$D(k+1) = D(k) + \eta(k)(I - K \tanh(y)u^{T}D^{T} - yu^{T}D^{T})D$$
(16)

where K is a diagonal matrix with elements kq = 1 if the recovered signal is super-Gaussian, and kq = -1 if the recovered signal is sub-Gaussian. If we assume that K is an identity matrix, then this will be referred to as the fixed nonlinearity case. This is the algorithm derived in [10]. Using stability arguments [3], it is possible to estimate the coefficient kq as follows:

$$k_{q} = sign\{E\{\sec h^{2}(y_{q})\}E\{y_{q}^{2}\} - E\{\tanh(y_{q})y_{q}\}\}$$
(17)

Thus, k_q can be estimated for each recovered signal.

4. GENERALIZED GAUSSIAN SOURCE MODEL

A generalized Gaussian source model is introduced in [5]. This model encompasses both super- and sub-Gaussian sources. The generalized gaussian density is expressed as follows:

$$p(y_q) = \frac{r_q}{2\sigma_q \Gamma(1/r_q)} e^{\left(\frac{1}{r_q} \left| \frac{y_q}{\sigma_q} \right|^{r_q}\right)}$$
(18)

where $r_q > 0$ is a variable parameter, $\Gamma(r) = \int_{0}^{\infty} y^{r-1} \exp(-y) dy$ is the gamma function and

 $\sigma_i^r = E\{y|^r\}$ is a generalized measure of variance known as the dispertion of the distribution. The parameter r_q can change from zero to, through 1 (the Laplace distribution) and $r_q = 2$ (standard Gaussian distribution), to r_q going to infinity (for uniform distribution). Once we model the source with generalized gaussian density, we can choose the nonlinearity as follows:

$$\varphi(y_q) = -\frac{\partial \log p(y_q)}{\partial y_q} = \left(\frac{\left|y_q\right|^{r_q-1} sign(y_q)}{\left|\sigma_q\right|^{r_q}}\right), \ r_q \ge 1$$
(19)

where the sign operator can be defined as follows:

$$sign(y_q) = \begin{cases} 1 & y_q > 0 \\ 0 & y_q = 0 \\ -1 & y_q < 0 \end{cases}$$
(20)

The parameters of the model can be estimated online as follows: Consider the cost function (7), we need to learn the parameters of the generalized gaussian source model to maximize the cost function. We can derive the gradient of ℓ with respect to with respect to r_q , σ_q as follows [7]:

$$\frac{\partial \ell}{\partial r_{q}} = \left[\frac{-\ln(2) + 1 + \ln(r_{q}) - \ln(\sigma_{q}) + \frac{\psi(1/r_{q})}{r_{q}} - \ln(\Gamma(1/r_{q})) - \left|\frac{y_{q}}{\sigma_{q}}\right|^{r_{q}} \ln\left(\left|\frac{y_{q}}{\sigma_{q}}\right|\right)\right]}{r_{q}} - \left[\frac{\ln(2) \cdot r_{q} + \ln(r_{q}) \cdot r_{q} - \ln(\sigma_{q}) \cdot r_{q} - \ln\left(\Gamma(1/r_{q}) \cdot r_{q} - \left|\frac{y_{q}}{\sigma_{q}}\right|^{r_{q}}\right)}{r_{q}^{2}}\right]$$
(21)

$$\frac{\partial \ell}{\partial \sigma_q} = \frac{1}{\sigma_q} - \left| y_q \right|^{r_q} \left| \sigma_q \right|^{-r_q - 1} sign(\sigma_q)$$
(22)

where $\psi(\bullet) = \frac{\Gamma'(\bullet)}{\Gamma(\bullet)}$ is the digamma function. Based on the above gradient, we have the following learning rules for r_q , σ_q

$$\sigma(k+1) = \sigma(k) + \eta(k) \left[\frac{\left[\frac{-\ln(2) + 1 + \ln(r_{q}(k)) - \ln(\sigma_{q}(k)) + \frac{\psi(1/r_{q}(k))}{r_{q}(k)} - \ln(\Gamma(1/r_{q}(k)))) - \left| \frac{y_{q}(k)}{\sigma_{q}(k)} \right|^{r_{q}(k)} \right]}{r_{q}(k)} \right] - \frac{1}{\sigma_{q}(k) + \ln(r_{q}(k))r_{q}(k) - \ln(\sigma_{q}(k))r_{q}(k) - \ln(\Gamma(1/r_{q}(k))r_{q}(k) - \left| \frac{y_{q}(k)}{\sigma_{q}(k)} \right|^{r_{q}(k)} \right]}{r_{q}^{2}(k)} \right]$$

$$\sigma(k+1) = \sigma(k) + \eta(k) \left(\frac{1}{\sigma_{q}(k)} - \left| y_{q}(k) \right|^{r_{q}(k)} \left| \sigma_{q}(k) \right|^{-r_{q}(k)-1} sign(\sigma_{q}(k)) \right)$$

$$(24)$$

5. EXPERIMENTAL RESULTS

In this experiment, consider the system described by the following two independent soources:

$$u_1(n) = 0.1\sin(400n)\cos(30n)$$

$$u_2(n) = 0.01\sin(500n + 9\cos(40n)))$$
(25)

The observed signals u(k) are obtained by passing the source signals through a mixing linear dynamical system defined in (1), and (2), where the source number n = 2, and system order N = 2. The system matrices $\overline{A}, \overline{B}, \overline{C}$ and \overline{D} are randomly selected, except that we guarantee \overline{D}^{-1} exists and the eigenvalues of \overline{A} are within the unit circle.

5.1. Lee et al.'s Switching Nonlinearity method

In this section, we present results (shown in Figure 1)in applying the method derived in Section 3.

5.2. Generalized Gaussian model approach

In this section, we will present results using the generalized gaussian model approach as proposed in this paper (see Figures 2). Table 1 gives the mean squared errors among the two methods. It is noted that the generalized Gaussian model gives the smallest mean squared errors.



Fig. 2 Signals of Lee *et a*l's method. For explanation of these graphs please see caption for Figure 2



Fig. 3. Signals of the generalized Gaussian model. (for an explanation of the graphs please see the caption for Figure 3)

Table 1: A table showing the mean squared errors as obtained by each of the two methods.

Method	Signal 1	Signal 2
Lee et al	0.00029	0.00021
	1	6
generalized gaussian	0.00012	0.00007
model	67	82

6. CONCLUSION

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 [10] as the baseline for comparison. We extend the generalized gaussian source model proposed in [5] to the MBD case. In addition, we have extended the switching nonlinearity method in [8] to the MBD case as well. The two methods are all applied to

two synthesized signals. It is quite surprising to observe that the method proposed in [10] works well even though it was not designed to work with non stationary signals. The method proposed in [8] works well, with the nonlinearities switching based on an estimation of the kurtosis of the recovered signals. The proposed generalized Gaussian model works best in that it gives the smallest mean squared error.

REFERENCES

- 1. Amari, S., Natural Gradient Working Efficiently in Learning, *Neural Computation*, 10, 251–276, 1998.
- 2. Bell, A.J. and Sejnowski, T.J., An Information Maximization Approach to Blind Separation and Blind Deconvolution, *Neural Computation*, 7, 1129–1159, 1995.
- 3. Cardoso, J-F. and Laheld, B., Equivalent Adaptive Source Separation, *IEEE Trans.* on Signal Processing, 44, 3017–3030, 1996.
- 4. Choi, S., Cichocki, A. and Amari, S., Flexible Independent Component Analysis, *Proc. IEEE, Workshop on Neural Networks for Signal Processing (NNSP'98)*, Cambridge, England, 83–92, 1998.
- 5. Choi, S., Cichocki, A. and Amary, S., Flexible Independent Component Analysis, *Journal of VLSI Signal Processing*, 26(1/2), 25-38, 2000.
- 6. Douglas, S.C., Cichocki, A., and Amari, S., Multichannel Blind Separation and Deconvolution of Sources with Arbitrary Distributions, *IEEE Workshop on Neural Networks for Signal Processing*, 436–445, 1997.
- 7. Everson, R.M. and Roberts, S.J., Independent Component Analysis: A Flexible Nonlinearity and Decorrelating Manifold Approach, *Neural Computation*, 11(8), 1957– 1983, 1999.
- 8. Lee, T-W., Girolami, M., and Sejnowski, T.J., Independent Component Analysis using an Extended Infomax Algorithm for Mixed Sub-Gaussian and Super-Gaussian Sources, *Neural Computation*, 11(2), 417–441, 1999.
- 9. Woo, W.L. and Dlay, S.S., Neural Network Approach to Blind Signal Separation of Mono-nonlinearly Mixed Signals, *IEEE Transactions on Circuits and System*, 52(1/6), 1232-1247, 2005.
- 10. Zhang, L-Q., and Cichocki, A., Blind Deconvolution of Dynamical Systems: A State Space Approach, *Journal of Signal Processing*, 4(2), 111–130, 2000.