

A GENERALIZED LEAST SQUARES APPROACH TO BLIND SEPARATION OF SOURCES WHICH HAVE VARIANCE DEPENDENCIES

Shohei Shimizu *, Aapo Hyvärinen

Helsinki Institute for Information Technology
Basic Research Unit
PO Box 68, FIN-00014
University of Helsinki, Finland

Yutaka Kano

Division of Mathematical Science
Osaka University
Toyonaka, Osaka, 560-8531, Japan

ABSTRACT

We discuss the blind source separation problem where the sources are not independent but are dependent only through their variances. Some estimation methods have been proposed on this line. However, most of them require some additional assumptions, a parametric model for their dependencies or a temporal structure of the sources, for example. In this article, we propose a generalized least squares approach to the blind source separation problem in the general case where those additional assumptions do not hold.

1. INTRODUCTION

In blind source separation methods, the observed signals $x_i(t)$ ($i = 1 \dots m$) are typically assumed to be linear mixtures of sources $s_j(t)$ ($j = 1 \dots n$). Let a_{ij} denote the coefficients in the linear mixing between the sources $s_j(t)$ and the observed signals $x_i(t)$. Then the mixing can be expressed as

$$x_i(t) = \sum_{j=1}^n a_{ij} s_j(t). \quad (1)$$

The number of sources n is usually assumed to be equal or less than the number of observed signals m , which is assumed here as well. The problem of blind source separation is now to estimate both the source signals $s_i(t)$ and the mixing coefficient a_{ij} based on observations of the $x_i(t)$ alone [1].

The model (1) is called independent component analysis (ICA) model if $s_j(t)$ are assumed to be non-gaussian and independent [2]. The ICA model has been extensively studied for last two decades, and many estimation techniques for the model are available [3].

Recently, many extensions of the ICA model have started to be considered [4, 5, 6]. A quite interesting extension

among them is the case where the source signals are not independent but dependent only through their variances [6]. To model such dependencies, [7] assumed that each source signal $s_i(t)$ can be represented as a product of two random signals $v_i(t)$ and $y_i(t)$:

$$x_i(t) = \sum_{j=1}^n a_{ij} v_j(t) y_j(t), \quad (2)$$

where $v_i(t)$ and $y_i(t)$ are independent, $y_i(t)$ are independent over time and are mutually independent of each other. No assumption on the distribution of $y_i(t)$ is made other than $y_i(t)$ have zero means. The variance signals $v_i(t)$ are nonnegative signals giving general activity levels and are allowed to be statistically dependent. Thus, the $v_i(t)$ could produce dependencies between sources $s_i(t) = v_i(t)y_i(t)$. No particular assumptions on the dependencies between $v_i(t)$ are made. This setting was called double-blind source separation problem because one neither observe the source signals $s_i(t)$ nor postulate a parametric model of their dependencies.

In [7], it was further assumed that the source signals have some time dependencies (autocorrelations) and proposed a method that utilizes the time structure of the observed signals for separating the source signals. The time dependency assumption is the key to their method. Therefore, the method is not applicable to the case where the sources are not temporally structured and has a more limited domain of applications, since many kinds of data do not have temporal structure in practical cases,

In [8], estimating functions for the model (2) was studied, and the quasi maximum likelihood estimation that does not require the time dependencies was proposed. However, one has to appropriately choose the nonlinearity depending on whether the underlying independent signals $y_i(t)$ are supergaussian or subgaussian as in maximum likelihood methods for the ordinary ICA model. Moreover, they have to make certain extra assumptions on the signs of certain complicated nonlinear cross-moments of the sources, and it is

*This research was supported by the Grant-in-Aid for Scientific Research from the Ministry of Education, Culture and Sports #16-7933.

not clear when these are fulfilled.

In this article, we propose a generalized least squares approach using second- and fourth-order moment structures of observed signals in the general case where no temporal structure is available and it is unknown whether the underlying signals are supergaussian or subgaussian.

2. MODEL

We shall define the following model, which we will refer to as variance dependent component analysis (VDCA) here. Let us collect the source signals in a vector $\mathbf{s} = [s_1, \dots, s_n]^T$, and also construct the observed signal vector \mathbf{x} in the same manner. (We omit the time indices in the subsequent part since we do not consider time structures.) Also let us collect the mixing coefficients in a matrix $A = [a_{ij}]$. The VDCA model for the m -dimensional observed vector \mathbf{x} is written as

$$\mathbf{x} = A\mathbf{s}, \quad (3)$$

where non-gaussian components s_i can be expressed as products of two signals v_i and y_i , $s_i = v_i y_i$, as in (2), where the y_i are zero-mean and mutually independent, and that the set of the y_i is independent from the set of the v_j . An important point in the VDCA model is that no temporal structure is assumed, which is different from [7].

It is rather difficult to develop identification conditions for the VDCA model. We leave it open here. We assume the model is identifiable using second- and fourth-order moments of observed signals, which is empirically confirmed by the simulation experiments in Section 4.

3. ESTIMATION

In [9, 10, 11], the generalized least squares approach (GLS) in estimation that utilizes higher-order moments were studied. We also employ the GLS approach to estimate A in (3).

First, let us introduce some notations. Let us denote by $\sigma_2(\boldsymbol{\tau})$ the vector that consists of elements of the covariance matrix based on the model where any duplicates due to symmetry have been removed, by $\sigma_4(\boldsymbol{\tau})$ the vector that consists of the tensor of fourth-order (cross-) moments where duplicate entries have been removed and by $\boldsymbol{\tau}$ the vector of source statistics and mixing coefficients that uniquely determines the second- and fourth-order moment structures of the model $\sigma_2(\boldsymbol{\tau})$ and $\sigma_4(\boldsymbol{\tau})$. Then the $\sigma_2(\boldsymbol{\tau})$, $\sigma_4(\boldsymbol{\tau})$ and $\boldsymbol{\tau}$ can be written as

$$\begin{aligned} \sigma_2(\boldsymbol{\tau}) &= H_2 E[\mathbf{x} \otimes \mathbf{x}], \\ \sigma_4(\boldsymbol{\tau}) &= H_4 E[\mathbf{x} \otimes \mathbf{x} \otimes \mathbf{x} \otimes \mathbf{x}], \end{aligned}$$

where the symbol \otimes denotes the Kronecker product and H_i is a selection matrix of order $\binom{p+k-1}{k} \times p^k$ ($k = 2, 4$)

that selects non-duplicated elements. The parameter vector $\boldsymbol{\tau}$ consists of A , $E(s_i^4)$ and $E(s_i^2 s_j^2)$ ($i \neq j$).

We shall provide an example of the second- and fourth-order moment structures of the VDCA model. For saving space, let us consider the two variables case. Then the second- and fourth-order moment structures $\sigma_2(\boldsymbol{\tau})$ and $\sigma_4(\boldsymbol{\tau})$ are written as

$$\begin{aligned} \sigma_2(\boldsymbol{\tau}) &= \begin{bmatrix} E(x_1^2) \\ E(x_1 x_2) \\ E(x_2^2) \end{bmatrix} \\ &= \begin{bmatrix} a_{11}^2 & a_{12}^2 \\ a_{11} a_{21} & a_{12} a_{22} \\ a_{21}^2 & a_{22}^2 \end{bmatrix} \begin{bmatrix} E(s_1^2) (= 1) \\ E(s_2^2) (= 1) \end{bmatrix}, \\ \sigma_4(\boldsymbol{\tau}) &= \begin{bmatrix} E(x_1^4) \\ E(x_1^3 x_2) \\ E(x_1^2 x_2^2) \\ E(x_1 x_2^3) \\ E(x_2^4) \end{bmatrix} \\ &= \begin{bmatrix} a_{11}^4 & 6a_{11}^2 a_{12}^2 & a_{12}^4 \\ a_{11}^3 a_{21} & 3a_{11} a_{12} (a_{11} a_{22} + a_{21} a_{12}) & a_{12}^3 a_{22} \\ a_{11}^2 a_{21}^2 & a_{11}^2 a_{22}^2 + a_{21}^2 a_{12}^2 + 4a_{11} a_{21} a_{12} a_{22} & a_{12}^2 a_{22}^2 \\ a_{11} a_{21}^3 & 3a_{21} a_{22} (a_{11} a_{22} + a_{21} a_{12}) & a_{12} a_{22}^3 \\ a_{21}^4 & 6a_{21}^2 a_{22}^2 & a_{22}^4 \end{bmatrix} \\ &\quad \times \begin{bmatrix} E(s_1^4) \\ E(s_1^2 s_2^2) \\ E(s_2^4) \end{bmatrix}, \end{aligned}$$

where the parameter vector $\boldsymbol{\tau}$ is

$$\boldsymbol{\tau} = [a_{11} \ a_{12} \ a_{21} \ a_{22} \ E(s_1^4) \ E(s_1^2 s_2^2) \ E(s_2^4)]^T.$$

Due to our assumptions of the dependencies (i.e., dependence only through variances), $E(s_1 s_2)$, $E(s_1^3 s_2)$ and $E(s_1 s_2^3)$ are zeroes, but $\text{cov}(s_1^2, s_2^2) = E(s_1^2 s_2^2) - 1$ is not zero.

We now propose that the model is estimated using the principle of generalized least-square estimation. This is a method of matching the moments of the observed data and those based on the model in a weighted least-squares sense. In other words, we minimize the distance between the $\sigma_2(\boldsymbol{\tau})$ and $\sigma_4(\boldsymbol{\tau})$ and their sample counterparts.

Let $\mathbf{x}_1, \dots, \mathbf{x}_N$ be a random sample from the VDCA model (3), and define the sample counterparts to the moments in (4) as

$$\begin{aligned} \mathbf{m}_2 &= \frac{1}{N} H_2 \sum_{j=1}^N \mathbf{x}_j \otimes \mathbf{x}_j, \\ \mathbf{m}_4 &= \frac{1}{N} H_4 \sum_{j=1}^N \mathbf{x}_j \otimes \mathbf{x}_j \otimes \mathbf{x}_j \otimes \mathbf{x}_j. \end{aligned}$$

Let us denote by $\boldsymbol{\tau}_0$ the true parameter vector. The $\sigma_i(\boldsymbol{\tau}_0)$ can be estimated by the \mathbf{m}_i when N is enough large:

$$\sigma_i(\boldsymbol{\tau}_0) \approx \mathbf{m}_i \quad (i = 2, 4).$$

The GLS estimator of $\boldsymbol{\tau}$ is obtained as

$$\hat{\boldsymbol{\tau}} = \arg \min_{\boldsymbol{\tau}} \left\| \begin{bmatrix} \mathbf{m}_2 \\ \mathbf{m}_4 \end{bmatrix} - \begin{bmatrix} \boldsymbol{\sigma}_2(\boldsymbol{\tau}) \\ \boldsymbol{\sigma}_4(\boldsymbol{\tau}) \end{bmatrix} \right\|_{\hat{U}^{-1}}^2. \quad (4)$$

(For simplicity, the norm $\mathbf{y}^T W \mathbf{y}$ of a vector \mathbf{y} associated with a nonnegative definite matrix W is here expressed as $\|\mathbf{y}\|_W^2$.) Here \hat{U} is a weight matrix in GLS estimation and converges in probability to a certain positive definite matrix U (see below on how to choose U). The resultant GLS estimator $\hat{\boldsymbol{\tau}}$ determined by (4) is then consistent and asymptotic normal [12].

Assume that the eighth-order moments of \mathbf{x}_i are finite, and let

$$V = \lim_{N \rightarrow \infty} N \cdot \text{cov}[\mathbf{m}_2^T, \mathbf{m}_4^T]^T.$$

Let V be a sample covariance matrix for $[\mathbf{m}_2^T, \mathbf{m}_4^T]^T$. One often takes a consistent estimator \hat{V} as \hat{U} in estimation to achieve the best asymptotic efficiency among the class of all estimators that are functions of $[\mathbf{m}_2^T, \mathbf{m}_4^T]^T$ (see, e.g., [12]). In the case, however, one has to compute the eighth-order moments to implement this estimation, and it is known that one needs to have large samples to stably estimate these higher-order moments and that inverting such a huge matrix may cause difficulties in solving the optimization problem [13]. Also, use of eighth-order moments could cause serious outlier problems. Thus, it would be more effective in practice to simply choose the identity matrix (or some other diagonal matrix, e.g., a diagonal matrix with the diagonal elements of \hat{V}) as \hat{U} particularly for moderate sample sizes. In this paper we report a simulation result where the identity matrix as \hat{U} is used to compute the estimates.

4. SIMULATIONS

We conducted small simulations to study the empirical performance of the GLS method above. The simulation consisted of 100 source separation trials with three different methods: 1) the GLS method proposed in this paper; 2) FastICA using kurtosis [14] and 3) FastICA using hyperbolic tangent function [15]. For the two FastICA, the symmetric orthogonalization was made. (The FastICA with the symmetric orthogonalization using hyperbolic tangent function as the nonlinearity is basically the same as the quasi-maximum likelihood estimation [16].)

In each trial, we generated four sources that were dependent through their variances and created observed signals following the VDCA model (3).

First, we created a random signal v_0 with the sample size $N = 20,000$ where their components were independently distributed according to the gaussian distribution with zero mean and unit variance. The variance signals v_i ($i = 1, \dots, 4$) were then defined as the absolute values of the

signal, that is, $v_i = |v_0|$ ($i = 1, \dots, 4$). These variance signals were completely dependent on each other since they were identical, but were independent over time. (Therefore, the double-blind method [7] that utilized temporal correlations was not applicable to this case.)

Next the source signals s_i were created by multiplying the variance signals v_i by four-dimensional random signals y_i , that is, $s_i = v_i y_i$, as in (2). Here, the four underlying signals y_i ($i = 1, \dots, 4$) were i.i.d. (white) zero-mean subgaussian random processes to create enough variance dependencies [7]. (The subgaussian signals were signed fourth root of zero mean-uniform variables.) The source signals were normalized to have zero means and unit variances. Finally, a random mixing matrix A was created, and the signals were mixed to provide the observed signals $x_i, i = 1, \dots, 4$.

The three methods were then applied on the data after it was prewhitened. The performance of each method was assessed as follows. Denoting by W the obtained estimate of the inverse of the mixing matrix (with permutation and sign indeterminacies), we looked at the matrix WA . In the ideal case where WA is a signed permutation matrix, there would be a single element whose absolute value is one and the other elements that are zeroes in each row. Therefore, we computed how many rows of this matrix had one element whose absolute value was larger than 0.95 and the other elements whose absolute values were smaller than 0.05. This gave a measure of how many source signals had been separated.

The results are shown in Figure 1. Our method separated 95.0% of the components, whereas both FastICAs could not separate the components at all (0%) since FastICA is based on independence of sources. Thus, our method was quite good, while not being perfect.

Our simulation results thus show that ordinary FastICA is not able to separate the sources. This shows that the results in [8] have limited applicability: it is necessary to develop new separation algorithms for variance-dependent sources, since ordinary ICA algorithms do not necessarily work.

5. DISCUSSION

We have proposed a GLS approach utilizing second- and fourth-order moment structures of observed signals to the blind source separation of source signals that are dependent only through their variances.

In the GLS approach, we do not have to assume that the sources have some temporal structures nor postulate any parametric models for their dependencies. This could be a big advantage of our approach over the conventional methods.

There would be some potential problems in our method.

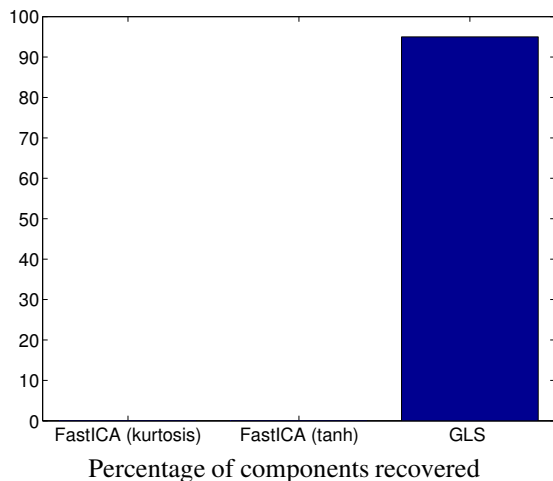


Fig. 1. Percentages of components recovered by the FastICA using kurtosis [14] and hyperbolic tangent function [15] and the GLS method

For example, the method is very sensitive to outliers because fourth-order moments are utilized. The problem of outliers is quite serious when applying estimation methods based on higher-order moments (cumulants) to many applications. We need more robust methods. They will be discussed in a forthcoming paper.

6. REFERENCES

- [1] C. Jutten and J. Héroult, "Blind separation of sources, part i: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, no. 1, pp. 1–10, 1991.
- [2] P. Comon, "Independent component analysis. a new concept?," *Signal Processing*, vol. 36, pp. 62–83, 1994.
- [3] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*, Wiley, New York, 2001.
- [4] F. R. Bach and M. I. Jordan, "Tree-dependent component analysis," in *Uncertainty in Artificial Intelligence: Proceedings of the Eighteenth Conference (UAI-2002)*, 2002.
- [5] A. Hyvärinen, "A unifying model for blind separation of independent sources," *Signal Processing*, Conditionally accepted.
- [6] A. Hyvärinen, P. O. Hoyer, and M. Inki, "Topographic independent component analysis," *Neural Computation*, vol. 13, pp. 1525–1558, 2001.
- [7] A. Hyvärinen and J. Hurri, "Blind separation of sources that have spatiotemporal dependencies," *Signal Processing*, vol. 84, pp. 247–254, 2004.
- [8] M. Kawanabe and K.-R. Müller, "Estimating functions for blind separation when sources have variance-dependencies," in *Proceedings of 5th International Conference on ICA and Blind Source Separation, Granada, Spain, 2004*, pp. 136–143.
- [9] P. M. Bentler, "Some contributions to efficient statistics in structural models: specification and estimation of moment structures," *Psychometrika*, vol. 48, pp. 493–517, 1983.
- [10] A. Mooijaart, "Factor analysis for non-normal variables," *Psychometrika*, vol. 50, pp. 323–342, 1985.
- [11] S. Shimizu and Y. Kano, "Non-normal structural equation modeling," 2003, submitted manuscript.
- [12] T. S. Ferguson, "A method of generating best asymptotically normal estimates with application to estimation of bacterial densities," *Annals of Mathematical Statistics*, vol. 29, pp. 1046–1062, 1958.
- [13] L. Hu, P. M. Bentler, and Y. Kano, "Can test statistics in covariance structure analysis be trusted?," *Psychological Bulletin*, vol. 112, pp. 351–362, 1992.
- [14] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, pp. 1483–1492, 1997.
- [15] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Networks*, vol. 10, pp. 626–634, 1999.
- [16] A. Hyvärinen, "The fixed-point algorithm and maximum likelihood estimation for independent component analysis," *Neural Processing Letters*, vol. 10, pp. 1–5, 1999.