New algorithms for blind separation when sources have spatial variance dependencies

Motoaki Kawanabe

Fraunhofer FIRST, Kekuléstr. 7, 12489 Berlin, Germany
nabe@first.fhg.de

Abstract

Blind separation problem is discussed, when sources are not independent, but have spatial variance dependencies. Hyvärinen and Hurri (2003) proposed an algorithm which requires no assumption on distributions of sources and no parametric model of dependencies between components. In order to obtain semiparametric algorithms which give a consistent estimator regardless of the source densities and the dependency structure, we study estimating functions for this model by the statistical approach of Amari and Cardoso (1997). Unlike the ICA model, the maximum likelihood estimation is not a semiparametric method in this case. Therefore, we consider a class of estimating functions which contain the quasi maximum likelihood estimation of the ICA model and the nonstationary ICA algorithm by Pham and Cardoso (2000). By modifying the score function, we got an estimating function close to it and proposed semiparametric algorithms based on it. Our algorithms were compared to other BSS methods with several artificial examples and speech signals.

1. Introduction

Blind methods of source separation have been successfully applied to many areas of science. For example, factors like receptive fields of simple cells were obtained by an ICA algorithm in natural image analysis (Olshausen and Field, 1996). The basic model assumes that the observed signals are linear superpositions of underlying hidden source signals. Let us denote the n source signals by $\boldsymbol{s}(t) = (s_1(t), \ldots, s_n(t))^\top$ in a vector formula, and the observed signals by $\boldsymbol{x}(t) = (x_1(t), \ldots, x_n(t))^\top$. The mixing can be expressed as the equation

$$\boldsymbol{x}(t) = A\boldsymbol{s}(t), \tag{1}$$

where $A = (a_{ij})$ denote the mixing matrix. Both the sources $\boldsymbol{s}(t)$ and the mixing matrix A are unknown, and the goal is to estimate them based on the observation $\boldsymbol{x}(t)$ alone.

In most blind source separation (BSS) methods, the source signals are assumed statistically independent. Blind source separation based on such model is called independent component analysis (ICA). With nongaussianity of the sources, the mixing matrix can be estimated and the source signals can be extracted under appropriate conditions. Temporal correlation and nonstationarity have also been used for the basic BSS problem.

Among many extensions of the ICA models, several researchers have studied the case where the source signals are not independent (e.g. Cardoso, 1998, Hyvärinen et al., 2001a, Bach and Jordan, 2002, Valpola et al., 2003, see also references in Hyvärinen and Hurri, 2003). The dependencies either need to be exactly known beforehand, or they have to be simultaneously estimated by algorithms. Recently, a novel idea called double-blind approach was introduced by Hyvärinen and Hurri (2003). In contrast to previous researches, their method requires no assumption on distributions of the sources and no parametric model of dependencies between the components. They also proposed an algorithm based on lagged 4th-order cumulants and showed that it works well, provided that the sources are dependent only through their variances and that the sources have temporal correlation.

In this paper, we discuss semiparametric estimation in the double-blind case as Hyvärinen and Hurri (2003) and propose new algorithms for the BSS problem with variance-dependencies. The variance-dependent BSS model are defined in section 2. We follow the semiparametric statistical approach of Amari and Cardoso (1997). Estimating functions which play a fundamental role for discussing semiparametric estimation are explained in section 3. Then, in section 4, estimating functions for the variance-dependent BSS model are discussed. We show that the maximum likelihood method of this model does not give a semiparametric estimator, i.e. the maximum likelihood estimator is not consistent, if assumed densities and dependencies are wrong. Therefore, by modifying the maximum likelihood estimation, we propose new semiparametric algorithms for the model in section 5. In section 6, we show results of numerical experiments in order to compare the new algorithms to other semiparametric methods.

2. Variance-dependent BSS model

In this section, we explain the framework of Hyvärinen and Hurri (2003). Let us assume that each source signal $s_i(t)$ is a product of non-negative activity level $v_i(t)$ and underlying i.i.d. signal $z_i(t)$, i.e.

$$s_i(t) = v_i(t) z_i(t).$$

In practice, the activity levels $v_i(t)$ are often dependent among different signals. In their formulation, each observed signal is expressed as

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

where $v_i(t)$ and $z_i(t)$ satisfy:

(i) v_i's and z_j's are independent,

(ii) $z_i(t)$ is i.i.d. in time, z_i and z_j are mutually independent,

(iii) $z_i(t)$ have zero mean and unit variance.

No assumption on the distribution of z_i is made except (iii). Regarding the general activity levels v_i's, $v_i(t)$ and $v_j(t)$ are allowed to be statistically dependent, and furthermore, no particular assumption on their dependencies are made (double-blind situation). We refer this framework to the variance-dependent BSS model in this paper.

Hyvärinen and Hurri (2003) proposed an algorithm which can separate the sources under this situation. Let $\boldsymbol{u}(t)$ be preprocessed signal of $\boldsymbol{x}(t)$ by spatial whitening. Their method minimizes the objective function

$$J(W) = \sum_{i,j} [\widehat{\mathrm{cov}}([\boldsymbol{w}_i^\top \boldsymbol{u}(t)]^2, [\boldsymbol{w}_j^\top \boldsymbol{u}(t-\Delta t)]^2)]^2,$$

where $\widehat{\mathrm{cov}}$ denotes the sample covariance, $\Delta t > 1$ means the lag and $W = (\boldsymbol{w}_1, \dots, \boldsymbol{w}_n)^\top$ is a orthogonal matrix. It was proved that the objective function J is minimized when WA equals a signed permutation matrix, if $K_{ij} = \mathrm{cov}(s_i^2(t), s_j^2(t-\Delta t))$ is of full rank. This method works quite well, if there exist temporal variance dependencies and no outliers.

3. Semiparametric statistical models and estimating functions

Amari and Cardoso (1997) established a statistical basis of the ICA problem. They pointed out that the basic ICA model

$$p(X|B, \kappa_{\boldsymbol{s}}) = |\det B| \prod_{t=1}^{T} \prod_{i=1}^{n} \kappa_{s_i}\{\boldsymbol{b}_i^\top \boldsymbol{x}(t)\} \quad (3)$$

is an example of semiparametric statistical models (Bickel et al.,1993, Amari and Kawanabe, 1997a,b), where $B = (\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)^\top = A^{-1}$ is the demixing matrix to be estimated and $\kappa_{\boldsymbol{s}}(\boldsymbol{s}) = \prod_{i=1}^{n} \kappa_{s_i}(s_i)$ is the unknown density of the sources $\boldsymbol{s}$. As the function $\kappa_{\boldsymbol{s}}$ in this model, semiparametric statistical models contain infinite dimensional or functional nuisance parameters which are difficult to estimate. Moreover, they even disturb inference on parameters of interest.

In the variance-dependent BSS model, the sources $\boldsymbol{s}(t)$ are decomposed of two components, the normalized signals $\boldsymbol{z}(t) = (z_1(t), \dots, z_n(t))^\top$ and the general activity levels $\boldsymbol{v}(t) = (v_1(t), \dots, v_n(t))^\top$. Since the components of $\boldsymbol{z}(t)$ are mutually independent like the ICA model, the density of the data X is factorized as

$$p(X|V; B, \kappa) = |\det B| \prod_{t=1}^{T} \prod_{i=1}^{n} \frac{1}{v_i(t)} \kappa_i \left\{ \frac{\boldsymbol{b}_i^\top \boldsymbol{x}(t)}{v_i(t)} \right\}, \quad (4)$$

when $V = (\boldsymbol{v}(1), \dots, \boldsymbol{v}(T))^\top$ is fixed. Therefore, the marginal distribution can be expressed as

$$p(X|B, \kappa, \nu) = \int p(X|V; B, \kappa) \nu(V) \mathrm{d}V, \quad (5)$$

where the density ν of V becomes an extra nuisance function.

In order to construct valid estimators for such semiparametric models, estimating functions were introduced by Godambe. Let us consider a general semiparametric model $p(\boldsymbol{x}|\boldsymbol{\theta}, \kappa)$, where $\boldsymbol{\theta}$ is r-dimensional parameter of interest and κ is a nuisance parameter. An r-dimensional vector valued function $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta})$ is called an estimating function, when it satisfies the following conditions for any $\boldsymbol{\theta}$ and κ (Godambe, 1991),

$$\mathrm{E}[\, \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}) \,|\boldsymbol{\theta}, \kappa] = \boldsymbol{0}, \quad (6)$$

$$|\det Q| \neq 0, \quad \text{where } Q = \mathrm{E}\left[\left. \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}) \right| \boldsymbol{\theta}, \kappa \right], \quad (7)$$

$$\mathrm{E}\left[\, \|\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta})\|^2 \,\middle|\, \boldsymbol{\theta}, \kappa \right] < \infty, \quad (8)$$

where $\mathrm{E}[\cdot|\boldsymbol{\theta}, \kappa]$ means the expectation over $\boldsymbol{x}$ with the density $p(\boldsymbol{x}|\boldsymbol{\theta}, \kappa)$ and $\|\cdot\|$ denotes Euclidean norm. Suppose i.i.d. samples $\boldsymbol{x}(1), \dots, \boldsymbol{x}(T)$ are obtained from the model $p(\boldsymbol{x}|\boldsymbol{\theta}^*, \kappa^*)$. If such a function exists, an M-estimator $\hat{\boldsymbol{\theta}}$ is obtained by solving the estimating equation

$$\sum_{t=1}^{T} \boldsymbol{f}(\boldsymbol{x}(t), \hat{\boldsymbol{\theta}}) = \boldsymbol{0}. \quad (9)$$

The estimator $\hat{\boldsymbol{\theta}}$ is consistent regardless of the true nuisance parameter κ^*, when the sample size T goes to infinity. Moreover, it asymptotically distributes with Gaussian

$N(\boldsymbol{\theta}^*, \mathrm{Av})$, where

$$\begin{aligned}\mathrm{Av} &= \mathrm{Av}(\boldsymbol{\theta}^*, \kappa^*) \\ &= \frac{1}{T} Q^{-1} \mathrm{E}\left[\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}) \boldsymbol{f}^\top(\boldsymbol{x}, \boldsymbol{\theta}) \,\middle|\, \boldsymbol{\theta}^*, \kappa^* \right] (Q^{-1})^\top,\end{aligned}$$

and $Q = Q(\boldsymbol{\theta}^*, \kappa^*) = \mathrm{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}) \,\middle|\, \boldsymbol{\theta}^*, \kappa^* \right]$. However, it is not easy to find such estimating functions. Amari and Kawanabe (1997b) studied this problem from geometrical point of view and gave a guideline for discussing estimating functions.

4. Estimating functions for blind separation

Estimating functions for the ICA model (3) were discussed by Amari and Cardoso (1997), Cardoso (1997). In case of the BSS problem, the parameter of interest is the $n \times n$ matrix $B = A^{-1}$ and hence it is convenient to write estimating functions in $n \times n$ matrix form $F(\boldsymbol{x}, B)$. This function should satisfy the same conditions as (6)∼(8).

As an example, we explain the quasi maximum likelihood method which maximizes the likelihood function under a prefixed density κ_0. We remark that the assumed function κ_0 can be different from the true density. The estimator $\hat{B}$ is derived from the equation

$$\sum_{t=1}^{T} \left[I - \boldsymbol{\varphi}\{\hat{\boldsymbol{y}}(t)\} \hat{\boldsymbol{y}}^\top(t) \right] = 0, \tag{10}$$

where $\hat{\boldsymbol{y}}(t) = \hat{B}\boldsymbol{x}(t)$ is the estimator of the sources $\boldsymbol{s}(t)$, $\boldsymbol{\varphi}(\boldsymbol{y}) = (\varphi_1(y_1), \ldots, \varphi_n(y_n))^\top$ and

$$\varphi_i(y_i) = -\frac{\mathrm{d}}{\mathrm{d}y_i} \log \kappa_{0i}(y_i).$$

For the nonlinear function $\varphi_i(y_i)$,

$$\varphi_i(y_i) = \tanh(c y_i), \qquad c > 0, \tag{11}$$

$$\varphi_i(y_i) = y_i^3. \tag{12}$$

are often employed.

In general, the quasi maximum likelihood estimator is no longer consistent because of misspecified distribution. However, in the ICA model, Amari and Cardoso (1997) showed that the function

$$F(\boldsymbol{x}, B) = I - \boldsymbol{\varphi}(\boldsymbol{y})\boldsymbol{y}^\top$$

in equation (10) becomes an estimating function, where $\boldsymbol{y} = B\boldsymbol{x}$. This leads to the fact that the quasi maximum likelihood method and its online version (the natural gradient learning) give an asymptotically consistent estimator, even if the assumed distribution κ_0 is not equal to the true one. This research motivate us to investigate such semiparametric procedures for the variance-dependent BSS model (4) and (5). We have been studied estimating functions for the extended model and showed that many existing BSS algorithms including the quasi likelihood method (10) still give consistent estimators under the double-blind situation in our forthcoming paper.

In contrast to the ICA model, the data sequence $X = (\boldsymbol{x}(1), \ldots, \boldsymbol{x}(T))$ is not i.i.d. in time, but might have temporal dependencies in the variance-dependent BSS model. Therefore, we have to consider more general functions $\bar{F}(X, B)$ of the whole sequence X. General estimating functions $\bar{F}(X, B)$ must satisfy

$$\mathrm{E}[\, \bar{F}(X, B) \,| B, \kappa, \nu] = 0, \tag{13}$$

$$|\det Q| \neq 0, \quad \text{where } Q = \mathrm{E}\left[\frac{\partial \mathrm{vec}\{\bar{F}(X, B)\}}{\partial \mathrm{vec}(B)} \,\middle|\, B, \kappa, \nu \right], \tag{14}$$

$$\mathrm{E}\left[\, \|\bar{F}(X, B)\|_F^2 \,\middle|\, B, \kappa, \nu \right] < \infty, \tag{15}$$

for all (B, κ, ν). The operator

$$\mathrm{vec}(F) = (F_{11}, \ldots, F_{n1}, \ldots, F_{1n}, \ldots, F_{nn})^\top$$

means the vectorization of matrices and $\|\cdot\|_F$ denotes Frobenius norm. It should be noted that scales and orders of the sources cannot be determined, i.e., two matrices B and PDB indicate the same distribution, when P and D are a permutation and a diagonal matrix, respectively. Therefore, we can pick up any matrix in such equivalence class without loss of generality.

An M-estimator $\hat{B}$ derived from the estimating equation

$$\bar{F}(X, \hat{B}) = 0. \tag{16}$$

Suppose that the data X is subject to $p(X|B^*, \kappa^*, \nu^*)$ defined by (4) and (5). If appropriate regularity conditions hold, we can calculate the asymptotic distribution of the M-estimator $\hat{B}$.

Theorem 1. *If the function $\bar{F}(X, B)$ satisfies the conditions (13)∼(15) and appropriate regularity conditions also hold, the M-estimator $\hat{B}$ derived from the equation (16) is asymptotically distributed with Gaussian*

$$\mathrm{vec}(\hat{B}) \sim N(\mathrm{vec}(B^*), \mathrm{Av}),$$

where

$$\begin{aligned}\mathrm{Av} &= \mathrm{Av}(B^*, \kappa^*, \nu^*) = Q^{-1} \Sigma (Q^{-1})^\top, \\ \Sigma &= \Sigma(B^*, \kappa^*, \nu^*) \\ &= \mathrm{E}\left[\mathrm{vec}\{\bar{F}(X, B^*)\} \, \mathrm{vec}\{\bar{F}(X, B^*)\}^\top \right], \\ Q &= Q(B^*, \kappa^*, \nu^*) \\ &= \mathrm{E}\left[\frac{\partial \mathrm{vec}\{\bar{F}(X, B^*)\}}{\partial \mathrm{vec}(B)} \right].\end{aligned}$$

In this paper, we study functions of the form

$$\begin{aligned}&\bar{F}(X, B) \\ &= \sum_{t=1}^{T} \left[C - \boldsymbol{\varphi}_t\{\boldsymbol{y}(1), \ldots, \boldsymbol{y}(T)\} \, \boldsymbol{y}^\top(t) \right],\end{aligned} \tag{17}$$

where $\boldsymbol{y}(t) = B\boldsymbol{x}(t)$, C is an $n \times n$ diagonal matrix and

$$\boldsymbol{\varphi}_t\{\boldsymbol{y}(1), \dots, \boldsymbol{y}(T)\} = (\varphi_{t1}\{y_1(1), \dots, y_1(T)\}, \dots, \varphi_{tn}\{y_n(1), \dots, y_n(T)\})^\top.$$

The quasi maximum likelihood estimator of the ICA model is derived from the general estimating function

$$\sum_{t=1}^{T} \left[I - \boldsymbol{\varphi}\{\boldsymbol{y}(t)\}\, \boldsymbol{y}^\top(t) \right].$$

This is obviously an example of (17). Another example is the ICA algorithm for nonstationary signals by Pham and Cardoso (2000). Their method is based on a Gaussian source model with blockwise constant variance. Let us divide the whole sequence into L blocks, i.e. the l-th block T_l consists of time points $\{(l-1)\frac{T}{L}+1, \dots, l\frac{T}{L}\}$. The estimator $\hat{B}$ is derived from the equation

$$\sum_{t=1}^{T} \frac{\hat{y}_i(t)}{\frac{L}{T}\sum_{\tau\in T_{l(t)}} \hat{y}_i^2(\tau)}\, \hat{y}_j(t) = 0, \quad i \neq j, \tag{18}$$

where $T_{l(t)}$ means the block which includes time t. Taking $C = I$ and

$$\varphi_{ti}\{y_1(1), \dots, y_1(T)\} = \frac{\hat{y}_i(t)}{\frac{L}{T}\sum_{\tau\in T_{l(t)}} \hat{y}_i^2(\tau)},$$

we can see that the l.h.s. of (18) can be written in the form as (17).

It is not difficult to show that functions of the form (17) become candidates of estimating functions for the variance-dependent BSS model.

Theorem 2. *The function $\bar{F}(X, B)$ defined in (17) satisfies the conditions (13). If it also satisfies the other conditions (14), (15), it becomes an estimating function. Therefore, with appropriate regularity conditions, the M-estimator $\hat{B}$ derived from $\bar{F}(X, B) = 0$ is consistent regardless of the nuisance densities (κ, ν).*

5. New semiparametric algorithm

In the previous section, we derived a class of functions (17) which are candidates of estimating functions for the variance-dependent BSS model. Here we propose a semiparametric estimation procedure based on such an estimating function.

Let us begin with the maximum likelihood estimation of the model (4) and (5). The score function can be calculated as

$$\begin{aligned} &\frac{\partial}{\partial B}\log p(X|B,\kappa,\nu) \\ &= \mathrm{E}\left[\frac{\partial}{\partial B}\log p(X|V;B,\kappa) \middle| X \right] \\ &= \sum_{t=1}^{T}\left[I - \mathrm{E}\left[\boldsymbol{\psi}\{\boldsymbol{y}(t), \boldsymbol{v}(t)\} \middle| X \right] \boldsymbol{y}^\top(t) \right] (B^{-1})^\top, \end{aligned} \tag{19}$$

where $\boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{v}) = (\psi_1(y_1, v_1), \dots, \psi_n(y_n, v_n))$,

$$\psi_i(y_i, v_i) = -\frac{\partial}{\partial y_i}\log \kappa_i\left(\frac{y_i}{v_i}\right).$$

The conditional expectation

$$\begin{aligned} &\mathrm{E}\left[\psi_i\{y_i(t), v_i(t)\} \middle| X \right] \\ &= \mathrm{E}\left[\psi_i\{y_i(t), v_i(t)\} \middle| \boldsymbol{y}(1), \dots, \boldsymbol{y}(T) \right] \end{aligned} \tag{20}$$

depends not only on i-th component $y_i(1), \dots, y_i(T)$ but also on the other components $y_j(1), \dots, y_j(T)$ $(j \neq i)$, that is, the score function (19) does not belong to the class (17). This means that in general the maximum likelihood estimator based on (19) is not consistent, if the nuisance functions (κ, ν) are misspecified.

It is known that estimating functions closer to the score function show better performance (Godambe, 1991). Therefore, by modifying the score function (19), we consider the following estimating function of the form (17).

$$\begin{aligned} &\bar{F}_{ij}(X, B) \\ &= \sum_{t=1}^{T}\left[\delta_{ij} - \varphi_{ti}\{y_i(1), \dots, y_i(T)\}\, y_j(t) \right] \end{aligned} \tag{21}$$

$$\begin{aligned} &\varphi_{ti}\{y_i(1), \dots, y_i(T)\} \\ &= \mathrm{E}\left[\psi_i\{y_i(t), v_i(t)\} \middle| y_i(1), \dots, y_i(T) \right] \end{aligned} \tag{22}$$

We remark that the function (22) depends only on the i-th components $y_i(1), \dots, y_i(T)$, while the conditional expectation (20) is a function of all components.

Now we fix the nuisance densities (κ, ν) and compute the explicit form of the functions (22). Unfortunately, it is difficult to carry out the conditional expectations without coarse approximation. Hence we use the following simplification: (i) the sequences v_i and v_j are independent (like naive Bayes), (ii) $z_i(t) \sim N(0, 1/\tau_i(t))$ where $\tau_i(t) = 1/v_i^2(t)$, and (iii) $\tau_i(t)$ is independent in time with Gamma distribution $Ga(\lambda, \frac{1}{\alpha_i(t)})$. Then the conditional expectation (22) turns out to be

$$\begin{aligned} &\mathrm{E}\left[\psi_i\{y_i(t), v_i(t)\} \middle| y_i(1), \dots, y_i(T) \right] \\ &= \frac{y_i(t)}{\frac{2\lambda}{1+2\lambda}\,\alpha_i(t) + \frac{1}{1+2\lambda}\, y_i^2(t)}. \end{aligned}$$

Because the parameter $\alpha_i(t)$ controls magnitude of the activity level $v_i^2(t)$, we replace $\alpha_i(t)$ with an online estimator

$$\hat{v}_i^2(t) = (1-\varepsilon)\hat{v}_i^2(t-1) + \varepsilon y_i^2(t) \quad (23)$$

where $\varepsilon > 0$ denotes the learning rate. We remark that the estimator $\hat{v}_i^2(t)$ is a function of the i-th component $y_i(1), \ldots, y_i(T)$. To sum up, we got an estimating function of the form (21) with

$$\varphi_{ti}\{y_i(1), \ldots, y_i(T)\} = \frac{y_i(t)}{\frac{2\lambda}{1+2\lambda}\hat{v}_i^2(t) + \frac{1}{1+2\lambda}y_i^2(t)}. \quad (24)$$

You see that this estimating function is similar to the equation (18) of the algorithm by Pham and Cardoso (2000).

By regarding the conditional distribution $p(v_i(t)|y_i(1), \ldots, y_i(T)\,)$ as the delta function $\delta\{v_i(t) - \hat{v}_i(t)\}$, we can get other estimating functions of the type (21) with

$$\begin{aligned}\varphi_{ti}\{y_i(1), \ldots, y_i(T)\} &= \psi_i\{y_i(t), \hat{v}_i(t)\} \\ &= -\frac{\partial}{\partial y_i}\log\kappa_i\left(\frac{y_i(t)}{\hat{v}_i(t)}\right),\end{aligned}$$

where the estimator $\hat{v}_i(t)$ is learned by the equation (23). As an example, we took the nonlinear function

$$\varphi_{ti}\{y_i(1), \ldots, y_i(T)\} = \tanh\left(\frac{y_i(t)}{\hat{v}_i(t)}\right), \quad (25)$$

which is modified a little bit by practical reason. This method looks similar to the quasi maximum likelihood estimation (10) and (11), but the nonlinear function (25) takes the scale estimator $\hat{v}_i(t)$ into account.

6. Numerical experiments

We carried out experiments with several artificial data sets and then dealt with variance-dependent speech signals as more realistic examples. We compared the algorithms listed in Table 1. In our proposed methods, the estimating equations were solved by the fixed point algorithm. (Hyvärinen et al., 2001b)

Table 1. ICA algorithms used in the experiments

EFr	The proposed algorithm with (24), $\lambda = \frac{1}{2}$
EFt	The proposed algorithm with (25)
DB	the double blind algorithm by Hyvärinen and Hurri (2003)
FICt	FastICA with tanh nonlinearity
Gau	The nonstationary ICA by Pham and Cardoso (2000)

For evaluating the results, we employed the index used in Amari et al. (1996)

$$\mathrm{AmariIndex}(B, A^*) = \sum_{i=1}^{n}\left\{\frac{\sum_{j=1}^{n} C_{ij}}{\max_k Ckj} - 1\right\} + \sum_{j=1}^{n}\left\{\frac{\sum_{i=1}^{n} C_{ij}}{\max_k Cik} - 1\right\},$$

where A^* is the true mixing matrix and $C = BA^*$. If $B = PD(A^*)^{-1}$ with a permutation matrix P and a diagonal matrix D, then $\mathrm{AmariIndex}(B, A^*) = 0$.

6.1. Artificial data sets

In all artificial data sets, five source signals of various types were generated and data after multiplying a random 5×5 mixing matrix were observed. We prepared seven datasets **ar_subG**, **ar_uni**, **sin_supG**, **sin_subG**, **com_supG**, **exp_supG** and **uni_subG**. In the first two datasets, the activity levels $\boldsymbol{v}(t)$ were generated from a multivariate AR(1) model, where outliers were excluded in an appropriate manner (see Hyvärinen and Hurri, 2003). The normalized signals z_i's were i.i.d. subgaussian random variables which are signed fourth roots of zero-mean uniform variables in **ar_subG**, while i.i.d. uniform random variables were used in **ar_uni**. In the third and the fourth datasets, the activity levels $v_i(t)$ were sinusoidal functions with different frequencies

$$v_i(t) = 1 + 0.9\sin\left(\frac{(13+i)\pi t}{8000}\right), \quad (26)$$

for $i = 1, \ldots, 5$. For the normalized signals z_i's, Laplace (resp. the subgaussian) i.i.d. random variables were used in **sin_supG** (resp. **sin_subG**). We also tested the case where all activity levels $v_i(t)$ ($i = 1, \ldots, 5$) were the same sinusoidal function

$$v_i(t) = 1 + 0.9\sin\left(\frac{\pi t}{500}\right). \quad (27)$$

As in the third example, Laplace i.i.d. random variables were used for the normalized signals z_i's. Because all algorithms did not work for the activity levels (27) combined with the subgaussian random variables, we omit the results here. In the last two datasets, the activity levels $\boldsymbol{v}(t)$ were i.i.d. in time t. In **exp_supG**, we transformed 5 independent exponential random variables linearly such that v_i and v_j have correlation 0.9, and z_i's were i.i.d Laplace random variables. On the other hand, in **uni_subG**, $\boldsymbol{v}(t)$ was generated from 5 uniform random variables by the same linear transformation and z_i's were the i.i.d subgaussian random variables.

We made 100 replications for each dataset and estimated the demixing matrix B for each replication. The results are summarized in Table 2. The proposed methods were comparable to or slightly better than **FICt** in all cases. As reported in Hyvärinen and Hurri (2003), all algorithms except **DB** did not work for **ar_subG**, but in other examples our algorithms were at least not worse than **DB**. Although for the data with the activity levels (26) **Gau** performed quite well, our algorithms were better in four cases.

Table 2. AmariIndex of the estimators. The values are the medians of 100 replications with the measure of deviation, $(\text{3rd-quantile} - \text{1st-quantile})/2$

	ar_subG	**ar_uni**	**sin_supG**
EFr	5.45 (1.03)	0.31 (0.04)	0.16 (0.02)
EFt	6.36 (1.76)	0.36 (0.05)	0.16 (0.02)
DB	0.52 (0.10)	0.70 (0.16)	0.79 (0.13)
FICt	9.25 (1.98)	0.38 (0.05)	0.23 (0.03)
Gau	1.19 (0.48)	0.85 (0.22)	0.08 (0.01)
	sin_subG	**com_supG**	**exp_supG**
EFr	0.38 (0.04)	0.37 (0.05)	0.39 (0.05)
EFt	0.22 (0.02)	0.42 (0.07)	0.43 (0.06)
DB	0.27 (0.03)	6.45 (1.56)	7.63 (1.88)
FICt	0.68 (0.14)	0.48 (0.07)	0.44 (0.06)
Gau	0.08 (0.01)	1.28 (0.19)	1.28 (0.20)
	uni_subG	**sss**	**v12**
EFr	0.17 (0.03)	0.01	0.04
EFt	0.19 (0.05)	0.01	0.04
DB	18.56 (1.66)	0.02	0.21
FICt	0.18 (0.03)	0.19 (0.04)	0.17 (0.02)
Gau	27.08 (0.33)	0.01	0.01

6.2. Variance-dependent speeches

We also dealt with more realistic data sets with speech signals which have strong variance dependencies, even though the mixing process is instantaneous. In the first experiment, we took two speech signals, where one speaker says digits from 1 to 10 in English, and the other counts at the same time in Spanish [1]. Figure 1 shows the sources and the estimators of their activity levels with an appropriate smoother. We inserted one short pause at different positions of both sequences to make the variances of the modified signals more correlated (0.65). In the second experiments, two speech signals from Japanese text were used [2]. Figure 2 shows the sources and the estimators of the activity levels. We extended and shortened each syllable of the second sequence and tuned its amplitude such that the two sources have high variance dependency. Correlation of the variances becomes 0.74 in the arranged signals.

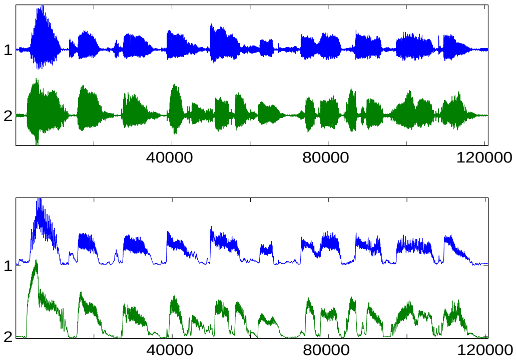

Figure 1. The sources and the estimators of their activity levels (dataset **sss**). The upper panel is the signals which are counting from 1 to 10 in English and Spanish. The lower panel shows their activity levels with an appropriate smoother.

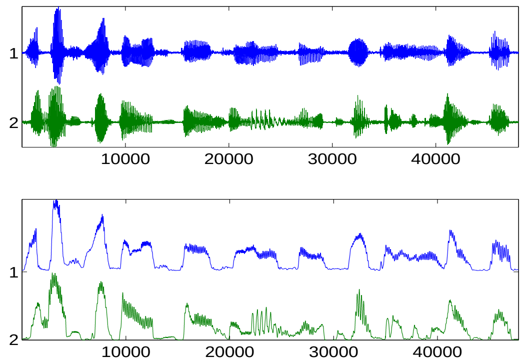

Figure 2. The sources and the estimators of their activity levels (dataset **v12**). The upper panel is the signals which are Japanese sentences. The lower panel shows their activity levels with an appropriate smoother.

A 2×2 mixing matrix A was randomly generated 100 times and 100 different mixtures of the source signals were made. Because most algorithms are equivariant, we can get the same results for any mixing matrix A in principle. In our result, except **FICt** this property held. The proposed algorithms were comparable to **Gau**, which showed the best performance.

7. Conclusions

In this paper, we discussed semiparametric estimation for blind separation, when sources have spatial variance dependencies. Hyvärinen and Hurri (2003) introduced the

[1] The signals were downloaded from `http://inc2.ucsd.edu/ tewon/`. We used the separated signals of their second demo as the sources, because they are good enough

[2] The signals can be downloaded by `http://www.islab.brain.riken.go.jp/ mura/ica/v1.wav` and `v2.wav`

double blind setting, where, in addition to source distributions, dependencies between components are not restricted by any parametric model. By the semiparametric approach of Amari and Cardoso (1997), we discussed estimating functions for this variance-dependent BSS model. We presented a class of estimating functions which contain the quasi maximum likelihood estimation of the ICA model and the nonstationary ICA algorithm by Pham and Cardoso (2000). Then, we proved that the maximum likelihood estimation of the variance-dependent BSS model is not a semiparametric methods, because the score function does not satisfy the unbiasedness property of estimating functions. Therefore, by modifying the score function, we got an estimating function close to it and proposed semiparametric algorithms based on it. Our algorithms were tested with several artificial examples and speech signals with strong variance dependencies. They gave at least comparable results to other BSS algorithms.

Our algorithm and **Gau** contain parameters which affect their performance. In practice, it is important to discuss how to select the best method and such tuning parameters for specific data. We think that suitable algorithms might be chosen with resampling method by Meinecke et al. (2002). Further works to this direction are necessary. Finally, Interesting applications should also be found.

Acknowledgements

The author appreciates K.-R. Müller, S. Harmeling, A. Ziehe and F. Meinecke for valuable discussion.

References

S. Amari and J.-F. Cardoso. Blind source separation—semiparametric statistical approach. *IEEE Trans. on Signal Processing*, 45(11):2692–2700, 1997.

S. Amari, A. Cichocki, and H.H. Yang. A new learning algorithm for blind source separation. In *Advances in Neural Information Processing Systems 8*, pages 757–763. MIT Press, 1996.

S. Amari and M. Kawanabe. Estimating functions in semiparametric statistical models. In I.V. Basawa et al., editor, *Selected Proceedings of the Symposium on Estimating Functions*, volume 32 of *IMS Lecture Notes–Monograph Series*, pages 65–81, 1997a.

S. Amari and M. Kawanabe. Information geometry of estimating functions in semiparametric statistical models. *Bernoulli*, 3:29–54, 1997b.

F. R. Bach and M. I. Jordan. Tree-dependent component analysis. In *Uncertainty in Artificial Intelligence: Proceedings of the Eighteenth Conference (UAI-2002)*, 2002.

J.-F. Cardoso. Estimating equations for source separation. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP9́7)*, volume 5, pages 3449–3452, Munich, Germany, 1997.

J.-F. Cardoso. Multidimensional independent component analysis. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP9́8)*, Seattle, WA, 1998.

V.P. Godambe, editor. *Estimating Functions*. Oxford Univ. Press, New York, 1991.

A. Hyvärinen, P. O. Hoyer, and M. Inki. Topographic independent component analysis. *Neural Computation*, 13(7), 2001a.

A. Hyvärinen and J. Hurri. Blind separation of sources that have spatiotemporal variance dependencies. *Signal Processing*, 2003. to appear.

A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley, 2001b.

F. Meinecke, A. Ziehe, M. Kawanabe, and K.-R. Müller. A resampling approach to estimate the stability of one- or multidimensional independent components. *IEEE Transactions on Biomedical Engineering*, 49 (12):1514–1525, 2002.

B.A. Olshausen and D.J. Field. Emergence of simple-cell receptive field by learning a sparse code for natural images. *Nature*, pages 607–609, 1996.

D.-T. Pham and J.-F. Cardoso. Blind separation of instantaneous mixtures of non-stationary sources. In *Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)*, pages 187–193, Helsinki, Finland, 2000.

H. Valpola, M. Harva, and J. Karhunen. Hierachical models of variance sources. In *Proc.Int. Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, Nara, Japan, 2003.