INDEPENDENT COMPONENT ANALYSIS IN THE PRESENCE OF GAUSSIAN NOISE BASED ON ESTIMATING FUNCTIONS

Motoaki Kawanabe

Faculty of Engineering, University of Tokyo
7-3-1 Hongo, Bunkyo-ku,
Tokyo 113-8656, Japan

Noboru Murata

RIKEN Brain Science Institute
2-1 Hirosawa, Wako-shi,
Saitama 351-0198, Japan

ABSTRACT

The problem of estimating the statistical model of independent component analysis in the presence of Gaussian noise is considered. Because of the additive noise, a combination of factor analysis and a noise-free ICA algorithm doesn't give a consistent estimator of the mixing matrix. In this paper, following the semiparametric statistical approach to the noise-free ICA model by Amari and Cardoso [1], we propose a method of estimating the mixing matrix consistently even if the additive noise exists. The proposed algorithm consists of two stages: First find the factor subspace by means of factor analysis, and then determine the directions of independent components based on an estimating function in this semiparametric model.

1. INTRODUCTION

Independent component analysis (ICA) uses a statistical model where observed data are expressed as a linear combination of statistically independent random variables. Since Jutten & Hérault published the first algorithm for the blind source separation, a lot of new ideas and algorithms have been proposed by researchers on signal processing and neural networks. These algorithms were rationalized theoretically by Amari and Cardoso [1] in the framework of semiparametric statistical models [4].

Many papers on ICA treat simple cases where no measurement noises are taken into account. However in realistic situations such as in MEG data analysis, it is not rare that certain measurement noises are added after mixing source signals. Ordinal ICA algorithms perform worse as the noise level increases and it is very difficult to derive meaningful outcomes. Therefore investigation of the ICA model with additive noise becomes one of the most important topics now. There exist several papers which handled ICA models with measurement noise. Both the maximum joint likelihood method [8] and the maximum marginal likelihood method [3] work on condition that the distributions of the source signals are known. The bias removal learning algorithm proposed by Chichocki et al. [6] assumes that the amplitude of noise is small. The JADE algorithm [5] and the Fast ICA algorithm with Gaussian moments [9] are semiparametric methods which give an estimate of the mixing matrix without knowledge of the unobserved source distributions.

The purpose of this paper is to explain semiparametric methods for the noisy ICA model in terms of estimating functions and propose alternative noisy ICA algorithms. The algorithm presented here consists of two stages: First find the factor subspace by means of factor analysis, and then determine the directions of independent components based on an estimating function in this semiparametric model.

2. ICA MODEL WITH MEASUREMENT NOISES

Let us consider a sequence of n-dimensional random vectors generated from an ICA model with additive noise

$$\boldsymbol{x}(t) = A\boldsymbol{s}(t) + \boldsymbol{\xi}(t), \qquad t = 1, \ldots, T. \tag{1}$$

where A is an unknown $n \times m$ matrix, and the vector $\boldsymbol{\xi}$ is measurement noise. And $\boldsymbol{s}(t) = (\, s_1(t), \ldots, s_m(t)\,)^{\mathrm{T}}$ is a sequence of m unobserved source signals which are mutually independent. Although there are a lot of papers considering time-dependent sources in the noise-free ICA model, we assume for simplicity that the source signal vectors $\boldsymbol{s}(t)$ are independent and identically distributed in time. So sometimes the index of time is omitted in the following. The joint probability density function $\kappa(\boldsymbol{s})$ of $\boldsymbol{s}$ is factorized as

$$\kappa(\boldsymbol{s}) = \prod_{i=1}^{n} \kappa_i(s_i) \tag{2}$$

where $\kappa_i(s_i)$ is the density function of the ith signal s_i. We consider the semiparametric situation that the function forms of $\kappa_1, \ldots, \kappa_m$ is unknown except for

$$\mathrm{E}_{\kappa_i}[s_i] = 0, \qquad i = 1, \ldots, m. \tag{3}$$

With respect to the additive noise, we assume that the random vector $\boldsymbol{\xi}$ is independent from the sources $\boldsymbol{s}$ and subjects to a multivariate normal distribution $N(\mathbf{0}, \Sigma)$ where $\Sigma = \mathrm{diag}(\sigma_i^2)$.

The density function of observed data $\boldsymbol{x}$ can be expressed as

$$p(\boldsymbol{x}; A, \Sigma, \kappa) = \int p(\boldsymbol{x}|\boldsymbol{s}; A, \Sigma)\, \kappa(\boldsymbol{s}) \mathrm{d}\boldsymbol{s}, \tag{4}$$

$$p(\boldsymbol{x}|\boldsymbol{s}; A, \Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \times \exp\left[-\frac{1}{2}(\boldsymbol{x} - A\boldsymbol{s})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x} - A\boldsymbol{s})\right]. \tag{5}$$

This is a semiparametric model, where the mixing matrix A and the noise variance Σ are parameters of interest and the density κ is a nuisance parameter in a function space. We note that this parameterization is redundant. Therefore, it is necessary to impose appropriate restrictions on the mixing matrix A or the scales of the source signals $\boldsymbol{s}$. For instance, we can add restrictions

$$\mathrm{E}_{\kappa_i}\left[s_i^2\right] = 1, \qquad i = 1, \ldots, m. \tag{6}$$

But it does not matter that the scales of estimated signals do not satisfy these restrictions.

It should be remarked that factor analysis model is also expressed as (1). However, distributional assumption on the latent variables $\boldsymbol{s}$ are different. We assume that at least $m-1$ signals subject to non-normal distributions in the noisy ICA model, while the factors $\boldsymbol{s}$ are often supposed to have a normal distribution $N(\mathbf{0}, I_m)$ in case of factor analysis.

3. ESTIMATING FUNCTIONS AND MATHEMATICAL PRELIMINARIES

Estimating functions introduced by Godambe [7] provide a general framework for discussing semiparametric estimators. Let us consider general semiparametric models in the form of $\{\, p(x; \boldsymbol{\theta}, \kappa) \,\}$ where $\boldsymbol{\theta}$ is the r-dimensional parameter to be estimated and κ is a nuisance parameter which belongs to an infinite dimensional or a function space. A r-dimensional vector function $\boldsymbol{f}(x, \boldsymbol{\theta})$ that does not depend on κ is called an estimating function when the following conditions are satisfied for all $\boldsymbol{\theta}$ and all κ.

$$\mathrm{E}_{\theta,\kappa}[\boldsymbol{f}(x, \boldsymbol{\theta})] = \mathbf{0} \tag{7}$$

$$\det|K| \neq 0, \quad \text{where } K = \mathrm{E}_{\theta,\kappa}\left[\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{f}(x, \boldsymbol{\theta})\right] \tag{8}$$

$$\mathrm{E}_{\theta,\kappa}\left[\boldsymbol{f}(x, \boldsymbol{\theta})\, \boldsymbol{f}^{\mathrm{T}}(x, \boldsymbol{\theta})\right] < \infty \tag{9}$$

If such an estimating function $\boldsymbol{f}(x, \boldsymbol{\theta})$ exists, we can obtain an M-estimator from given i.i.d. data $x_1, \ldots, x_n$ by solving the estimating equation

$$\sum_{i=1}^{n} \boldsymbol{f}(x_i, \widehat{\boldsymbol{\theta}}) = \mathbf{0}. \tag{10}$$

It can be shown under some additional regularity conditions that the M-estimator is consistent whatever κ is. Its covariance matrix is given asymptotically by

$$V\left[\widehat{\boldsymbol{\theta}}\right] = \frac{1}{n}K^{-1}\mathrm{E}_{\theta,\kappa}\left[\boldsymbol{f}\,\boldsymbol{f}^{\mathrm{T}}\right]\left(K^{-1}\right)^{\mathrm{T}}. \tag{11}$$

The model (1) can be regarded as a linear regression model with heteroschedastic noises after the parameters A and Σ are fixed. Following the theory of regression analysis, the data $\boldsymbol{x}$ is orthogonally decomposed as

$$\boldsymbol{x} = A\boldsymbol{y} + \boldsymbol{z}, \tag{12}$$

$$\boldsymbol{y}(\boldsymbol{x}; A, \Sigma) \equiv (A^{\mathrm{T}}\Sigma^{-1}A)^{-1}A^{\mathrm{T}}\Sigma^{-1}\boldsymbol{x}, \tag{13}$$

$$\boldsymbol{z}(\boldsymbol{x}; A, \Sigma) \equiv \left\{I - A(A^{\mathrm{T}}\Sigma^{-1}A)^{-1}A^{\mathrm{T}}\Sigma^{-1}\right\}\boldsymbol{x}, \tag{14}$$

where $\boldsymbol{y}$ is the generalized least square estimator (GLS) of $\boldsymbol{s}$, $\boldsymbol{z}$ is the residual, and orthogonality means $(A\boldsymbol{y})^{\mathrm{T}}\Sigma^{-1}\boldsymbol{z} = 0$. Furthermore, $\boldsymbol{y}$ and $\boldsymbol{z}$ are independent of each other, because the density function of $\boldsymbol{x}$ can be decomposed as follows.

$$p(\boldsymbol{x}; A, \Sigma, \kappa) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}\boldsymbol{z}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{z}\right] \times \int \exp\left[-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{s})^{\mathrm{T}}V^{-1}(\boldsymbol{y} - \boldsymbol{s})\right]\kappa(\boldsymbol{s})\mathrm{d}\boldsymbol{s} \tag{15}$$

For given $\boldsymbol{s}$, $\boldsymbol{y}$ subject to $N(\boldsymbol{s}, V)$ where $V = (v_{ij}) \equiv (A^{\mathrm{T}}\Sigma^{-1}A)^{-1}$. On the other hand, $\boldsymbol{z}$ does not depend κ and distributes with an $(n-m)$-dimensional degenerated normal distribution $(A^{\mathrm{T}}\Sigma^{-1}\boldsymbol{z} = 0)$.

$$\boldsymbol{y} = \boldsymbol{s} + \boldsymbol{\zeta}, \quad \boldsymbol{\zeta} \equiv (A^{\mathrm{T}}\Sigma^{-1}A)^{-1}A^{\mathrm{T}}\Sigma^{-1}\boldsymbol{\xi} \sim N(\mathbf{0}, V) \tag{16}$$

$$\boldsymbol{z} \sim N(\mathbf{0}, \Gamma), \quad \Gamma = (\gamma_{ij}) = \Sigma - AVA^{\mathrm{T}} \tag{17}$$

Let us define Multivariate Hermite polynomials which are extensions of well known Hermite polynomials to multivariate normal distributions. We express the density function of the m-variate normal distribution with mean $\mathbf{0}$ and covariance matrix $V = (v_{ab})$ as

$$\phi(\boldsymbol{y}; V) = \frac{1}{(2\pi)^{m/2}|V|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{y}^{\mathrm{T}} V^{-1} \boldsymbol{y}\right). \quad (18)$$

Let us put two differential operators, $D_a \equiv \partial/\partial y_a$ and $\widetilde{D}_a \equiv \partial/\partial \widetilde{y}_a$, where $\widetilde{\boldsymbol{y}} = V^{-1}\boldsymbol{y}$. Then two types of multivariate Hermite polynomials are defined as follows.

Definition 1 Covariant Hermite Polynomials H and Contravariant Hermite Polynomials $\widetilde{H}$

$$\begin{aligned} & H_{i_1 \dots i_m}(\boldsymbol{y}; V) \\ & \quad = (-D_1)^{i_1} \cdots (-D_m)^{i_m} \phi(\boldsymbol{y}; V)/\phi(\boldsymbol{y}; V) \quad (19) \\ & \widetilde{H}_{i_1 \dots i_m}(\boldsymbol{y}; V) \\ & \quad = (-\widetilde{D}_1)^{i_1} \cdots (-\widetilde{D}_m)^{i_m} \phi(\boldsymbol{y}; V)/\phi(\boldsymbol{y}; V) \quad (20) \end{aligned}$$

4. ESTIMATING FUNCTIONS IN THE NOISY ICA MODEL

In this section we will explain estimating functions in the noisy ICA model briefly. As the most important property is unbiasedness (7) for any nuisance parameter κ, we devote this part to characterize scalar functions which satisfy the same unbiasedness as (7). These unbiased functions are candidates for components of estimating functions.

We impose following there constraints on the density function κ_i of the source s_i $(i = 1, \dots, m)$.

$$\int \kappa_i(s_i) \mathrm{d}s_i = 1, \quad (21)$$

$$\int s_i \kappa_i(s_i) \mathrm{d}s_i = 0, \quad (22)$$

$$\int s_i^2 \kappa_i(s_i) \mathrm{d}s_i = 1, \quad (23)$$

The first two constraints (21) and (22) are necessary. On the other hand, the last one (23) is added in order to restrict scale of the source signal.

Let us define $F^{\perp}_{A,\Sigma}$ as the set of functions whose conditional expectations for given $\boldsymbol{y} = \boldsymbol{y}(\boldsymbol{x}; A, \Sigma)$ vanish.

$$F^{\perp}_{A,\Sigma} \equiv \{f(\boldsymbol{x});\ \mathrm{E}_{A,\Sigma}[f(\boldsymbol{x})|\,\boldsymbol{y}] = 0\} \quad (24)$$

We remark that the conditional expectation of $\boldsymbol{x}$ for given $\boldsymbol{y}$ does not depend on κ. Then we can characterise the set of the unbiased scalar functions as follows.

Theorem 1 In the case that only the necessary constraints (21) and (22) are imposed, the set of the unbiased functions are expressed as

$$F^{\perp}_{A,\Sigma} \oplus \{f(\boldsymbol{y}; A, \Sigma)\,;\ \text{satisfy (26)}\} \quad (25)$$

$$\mathrm{E}_{A,\Sigma}\left[f(\boldsymbol{y}; A, \Sigma)|\boldsymbol{s}\right] = \sum_{i=1}^{m} s_i\, \nu_i(\boldsymbol{s}_{-i}) \quad (26)$$

where $\oplus$ means the direct sum and ν_i is an arbitrary function of $\boldsymbol{s}_{-i} \equiv (s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_m)$.

It is difficult to determine general form of functions of $\boldsymbol{y}$ whose conditional expectation for given $\boldsymbol{s}$ can be expressed as (26). However we can describe polynomials concretely which have this property. For simplicity, we assume that any moment of the source signals s_i exist in this section.

Theorem 2 Under the same condition as Theorem1, the set of the unbiased polynomials of $\boldsymbol{y}$ are expressed as

$$\begin{aligned} I^{Y}_{A,\Sigma} = \operatorname{span}\{&\widetilde{H}_{\boldsymbol{i}}(\boldsymbol{y}; V);\ \boldsymbol{i} = (i_1, \dots, i_m), \\ &\text{at least one of the indices } \boldsymbol{i} \text{ is equal to } 1\} \quad (27) \end{aligned}$$

Remark In the case that the constraints (23) are also added, for instance $y_j^2 - v_{jj} - 1$ satisfies the unbiasedness.

If we can collect $n \times m + n$ unbiased functions (the same number as the parameter to be estimated) which satisfy the other conditions (8) and (9) in all, then the set of the functions becomes an estimating function. A necessary condition of (8) is that any function included in the set is not orthogonal to all score functions of the parameter to be estimated

$$\begin{aligned} U_A &= \frac{\partial}{\partial A} \log p \\ &= \Sigma^{-1}\boldsymbol{x}\, \mathrm{E}[\boldsymbol{s}^{\mathrm{T}}|\boldsymbol{y}] - \Sigma^{-1} A\, \mathrm{E}[\boldsymbol{s}\boldsymbol{s}^{\mathrm{T}}|\boldsymbol{y}], \quad (28) \\ u_{\sigma_i^2} &= \frac{\partial}{\partial \sigma_i^2} \log p = -\frac{1}{2\sigma_i^2} + \frac{x_i^2}{2\sigma_i^4} - \frac{x_i}{\sigma_i^4} \sum_j a_{ij} \mathrm{E}[s_j|\boldsymbol{y}] \\ &\quad + \frac{1}{2\sigma_i^4} \sum_{j,k} a_{ij} a_{ik} \mathrm{E}[s_j s_k|\boldsymbol{y}]. \quad (29) \end{aligned}$$

Therefore, we must pick up available Hermite polynomials which is not orthogonal to all score functions. It is possible to calculate the inner product of Hermite polynomials and score functions. If $s_1, \dots, s_m$ have symmetric distributions around the origin, we can show that available Hermite polynomials are specied as follows.

Theorem 3 Suppose $s_1, \dots, s_m$ are symmetrically distributed around the origin. Score functions and $\widetilde{H}_{\boldsymbol{r}}$ are orthogonal except for

$$\boldsymbol{r} = (1, r_2, r_3, \dots, r_m), \quad r_2 : \text{odd},\ r_3, \dots, r_m : \text{even}. \quad (30)$$

or its permutation. If $\boldsymbol{r}$ is (30), the inner products of score functions and $\widetilde{H}_{\boldsymbol{r}}$ are expressed as

$$\begin{aligned} \mathrm{E}\left[\widetilde{H}_{\boldsymbol{r}} u_{a_{i1}}\right] &= r_2 w_{2i}\, \mu_1^{(2)} \mu_2^{(r_2-1)} \mu_3^{(r_3)} \cdots \mu_m^{(r_m)}, & (31) \\ \mathrm{E}\left[\widetilde{H}_{\boldsymbol{r}} u_{a_{i2}}\right] &= w_{1i}\, \mu_2^{(r_2+1)} \mu_3^{(r_3)} \cdots \mu_m^{(r_m)}, & (32) \\ \mathrm{E}\left[\widetilde{H}_{\boldsymbol{r}} u_{a_{ij}}\right] &= 0, \qquad (j \geq 3), & (33) \\ \mathrm{E}\left[\widetilde{H}_{\boldsymbol{r}} u_{\sigma_i^2}\right] &= r_2 w_{1i} w_{2i}\, \mu_2^{(r_2-1)} \mu_3^{(r_3)} \cdots \mu_m^{(r_m)}, & (34) \end{aligned}$$

where $W = (w_{ij}) = VA^{\mathrm{T}}\Sigma^{-1}$ and $\mu_i^{(r_i)} = \mathrm{E}\,[s_i^{r_i}]$.

Examples of polynomials which satisfy the condition (30) are

$$\begin{aligned} & y_j y_k - v_{jk}, \quad j < k, & (35) \\ & y_j^3 y_k - 3v_{jj} y_j y_k - 3v_{jk} y_j^2 + 3v_{jj} v_{jk}, \quad j \neq k, & (36) \\ & y_j^2 y_k y_l - v_{jj} y_k y_l - v_{kl} y_j^2 - 2v_{jk} y_j y_l \\ & \qquad -2v_{jl} y_j y_k + v_{jj} v_{kl} + 2v_{jk} v_{jl}. & (37) \end{aligned}$$

We call them $(1,1)$-type, $(3,1)$-type and $(2,1,1)$-type respectively.

We can construct unbiased functions other than polynomials. Some functions which are product of a polynomial and a Gaussian density become unbiased. These functions appeared in the Fast ICA with Gaussian moments [9]

Theorem 4 For any $d > 0$ and any integer i,

$$\{y_j H_i(y_k; d) + v_{jk} H_{i+1}(y_k; d)\}\, \phi(y_k; d), \quad j \neq k \qquad (38)$$

is unbiased for all κ.

5. A NOISY ICA ALGORITHM BASED ON ESTIMATING FUNCTIONS

Now we propose an algorithm which is a combination of factor analysis and an estimating function method for the noisy ICA model.

1. Find the factor subspace by using factor analysis such as the unweighted least squares method (ULS) or the maximum likelihood method (ML). Let the solution be $\left(A^{(0)}, \Sigma^{(0)}\right)$.

2. Calculate initial estimates of source signals and their conditional covariances.

$$\begin{aligned} \boldsymbol{y}^{(0)}(t) &= V^{(0)} (A^{(0)})^{\mathrm{T}} (\Sigma^{(0)})^{-1} \boldsymbol{x}(t) & (39) \\ V^{(0)} &= \left\{ (A^{(0)})^{\mathrm{T}} (\Sigma^{(0)})^{-1} A^{(0)} \right\}^{-1} & (40) \end{aligned}$$

3. Let Q be $m \times m$ transformation matrix to the direction of the independent components, that is, the mixing matrix is expressed as $A = A^{(0)} Q^{-1}$ and

$$\begin{aligned} \boldsymbol{y}(t) &= VA^{\mathrm{T}} (\Sigma^{(0)})^{-1} \boldsymbol{x}(t) \\ &= Q \boldsymbol{y}^{(0)}(t) & (41) \\ V &= \left\{ A^{\mathrm{T}} (\Sigma^{(0)})^{-1} A \right\}^{-1} = Q V^{(0)} Q^{\mathrm{T}} & (42) \end{aligned}$$

The matrix Q can be determined by the following estimating equations $(i \neq j)$

$$\begin{aligned} & \sum_{t=0}^{T} \{ y_i^3(t) y_j(t) - 3v_{ij} y_i^2(t) \\ & \qquad -3v_{ii} y_i(t) y_j(t) + 3v_{ii} v_{ij} \} = 0 \qquad (43) \end{aligned}$$

and appropriate additional constraints such as

$$\sum_{j=1}^{m} q_{ij}^2 = 1, \qquad i = 1, \ldots, m. \qquad (44)$$

The m-dimensional vector $\boldsymbol{y}^{(0)}$ constructed by factor analysis can be regarded as quasi-whitened data of $\boldsymbol{x}$. The Quasi-whitening is also used in noisy ICA algorithms proposed so far, though the noise variance Σ is assumed to be known or sphere. As described here, we can carry out the quasi-whitening by factor analysis even if Σ is unknown.

Finally, we will explain the asymptotic behavior of this algorithm briefly. We express the true parameter with superscript $*$ which indicates the generating model of the observed samples. Let us express the sample covariance matrix as $S = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}(t)\boldsymbol{x}(t)^{\mathrm{T}}$ and the model covariance matrix as $\Psi = AA^{\mathrm{T}} + \Sigma$. In the ULS method for factor analysis, estimators are obtained by minimizing the quadratic loss criterion

$$F_u(\Psi) = \mathrm{tr}(S - \Psi)^2, \qquad (45)$$

with respect to (A, Σ) under the constraint offdiag$(A^{\mathrm{T}}A)$ $= 0$. On the other hand, the ML method gives estimators which minimize the following loss criterion

$$F_M(\Psi) = \mathrm{tr}(S\Psi^{-1}) - \log|S\Psi^{-1}| - n, \qquad (46)$$

under the the constraint offdiag$(A^{\mathrm{T}}\Sigma^{-1}A) = 0$. Let us define a transformation $A^{\dagger} = A^* Q^*$ where for each procedure Q^* is an orthogonal matrix defined in the following eigenvalue decomposition. If the ULS method is used, Q^* consists of the eigen vectors of $(A^*)^{\mathrm{T}} A^*$, i.e.

$$(A^*)^{\mathrm{T}} A^* = Q^* \Delta (Q^*)^{\mathrm{T}}. \qquad (47)$$

where Δ is a diagonal matrix. When the ML method is used, we define Q^* as the eigen vectors of $(A^*)^{\mathrm{T}}(\Sigma^*)^{-1}A^*$, i.e.

$$(A^*)^{\mathrm{T}}(\Sigma^*)^{-1}A^* = Q^*\Delta(Q^*)^{\mathrm{T}}. \qquad (48)$$

where Δ is a diagonal matrix. The sample covariance matrix S converges to the covariance matrix $\Psi^* = A^*(A^*)^{\mathrm{T}} + \Sigma^* = A^\dagger(A^\dagger)^{\mathrm{T}} + \Sigma^*$ of the true model as the sample size T goes to infinity. Therefore it can be shown that the parameter $(A^\dagger, \Sigma^*)$ asymptotically minimizes the criterion F_u (or F_M) and $(A^\dagger)^{\mathrm{T}}A^\dagger = \Delta$ (or $(A^\dagger)^{\mathrm{T}}(\Sigma^*)^{-1}A^\dagger = \Delta$) becomes a diagonal matrix. Suppose that $\Psi^* = A^*(A^*)^{\mathrm{T}} + \Sigma^*$ is identifiable, that is, parameter (A, Σ) that satisfies $AA^{\mathrm{T}} + \Sigma = \Psi^*$ can be determined uniquely except for a rotation matrix.

Theorem 5 The estimator $(A^{(0)}, \Sigma^{(0)})$ derived by the ULS method or the ML method converges to $(A^\dagger, \Sigma^*)$ as T goes to infinity.

After estimating the factor subspace and the noise variance, we must determine the directions of the independent components. Assuming that T is very large, we consider for simplicity that $(A^{(0)}, \Sigma^{(0)}) = (A^\dagger, \Sigma^*)$ holds. Since

$$A^* = A^\dagger(Q^*)^{\mathrm{T}} = A^{(0)}(Q^*)^{\mathrm{T}}, \qquad (49)$$

the correct transformation is expressed as $Q = Q^*$ or $Q = PDQ^*$ where D is any diagonal matrix and P is any permutation matrix (remember $A = A^{(0)}Q^{-1}$). Due to the additional conditions (44) we can derive Q^* except for indefiniteness of sign and order of independent components. We ignore this indefiniteness here. It is also possible to construct estimation procedures on the restricted set of orthogonal matrices.

It can be shown that the solution of the estimating equation is guaranteed to converge to the correct transformation matrix Q^* because of property of estimating functions.

Theorem 6 The transformations $Q = PDQ^*$ for any diagonal matrices D and any permutation matrix P satisfy unbiasedness of the non-diagonal terms of the estimating function. From the additional constraints (44), the solution Q^* is selected in these transformations.

6. NUMERICAL EXPERIMENTS

In order to test the proposed algorithm, we conducted a numerical experiment. As source signals we used three speech signals whose sizes are 48000. The signals which were normalised so that they had unit variances were plotted in Figure.1. The dimension of the observed data was 10 and at each trial a 10×3 mixing matrix was randomly generated so that each component was subject to the standard normal distribution independently. The noise covariance was identity matrix I_{10}. In this experiment, we compared three algorihtms. FastICA, JADE, and estimating function method (EF) were used after quasi-whitening by factor analysis. We generated 500 sets of such data and estimated the mixing matrices and the noise covariances by the three algorihtms. We used the crosstalk ratio to compare the performance of the estimators obtained by these algorithms. The results are summarized in Table.1. Because ordinal FastICA does not take measurement noises into account, the crosstalk ratios were rather big. Both JADE and the proposed algorithm are semiparametric methods in the noisy ICA model, the crosstalk ratios were about 1/3 compared to the noise-free ICA algorithm. In this experiment, the performance of the proposed algorithm was almost same as that of JADE.

Table.1. the results of the numerical experiment (crosstalk ratio)

method	mean	s.d.	min	max
FastICA	0.0594	0.0631	0.0051	0.6634
JADE	0.0184	0.0044	0.0067	0.0539
EF	0.0186	0.0041	0.0072	0.0389

7. CONCLUSIONS

In this paper, we discussed a general form of estimating functions in the noisy ICA model, following the semiparametric statistical approach by Amari and Cardoso [1]. From this results we can see that the combination of factor analysis and a noise-free ICA algorithm is not guaranteed to give a consistent estimator. Then a noisy ICA algorithm which gives a consistent estimator of the mixing matrix and the noise variance was proposed. This algorithm consists of two steps: prewhitening by factor analysis and pursuit of the independent component directions with the estimating equations. It is also possible to construct noisy ICA algorithms by combining factor analysis and other estimating functions. Although we assumed that the additive noise $\boldsymbol{\xi}$ is subject to the normal distribution $N(\mathbf{0}, \Sigma)$, the algorithm proposed here still has consistency in the semiparametric sense under the weaker assumption that $\boldsymbol{\xi}$ has the same 4th order moments structure as the normal distribution.

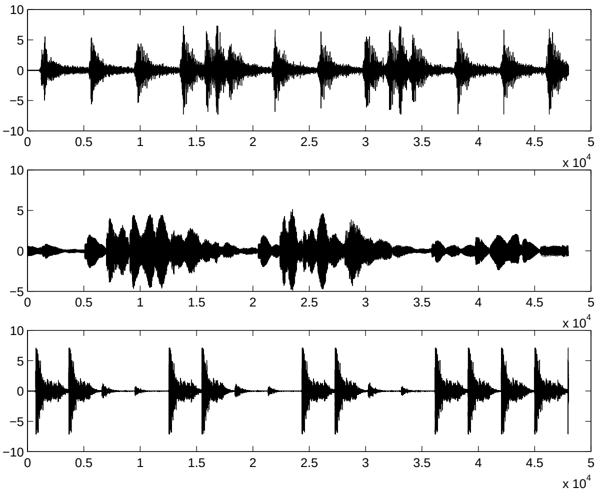

Figure 1: source signals

8. REFERENCES

[1] S.Amari, and J.-F.Cardoso, "Blind source separation — semiparametric statistical approach," *IEEE Trans. on Signal Processing*, vol.45, pp.2692–2700, 1997.

[2] S.Amari and M.Kawanabe, "Information geometry of estimating functions in semiparametric statistical models," *Bernoulli*, vol.3, pp.29–54, 1997.

[3] H.Attias, "Independent Factor Analysis," *Neural Computation*, vol.11, pp.803-851, 1999.

[4] P.J.Bickel, C.A.J.Klaassen, V.Ritov and J.A.Wellner, *Efficient and Adaptive Estimation in Semiparametric Models*. Baltimore: Johns Hopkins University Press, 1993.

[5] J.-F.Cardoso and A.Souloumiac, "Blind beamforming for non Gaussian signals," *IEE-Proceedings-F*, vol.140, pp.362-370, 1993.

[6] A.Cichocki, S.C.Douglas, and S.Amari, "Robust techniques for independent component analysis with noisy data", *Neurocomputing*, vol.22, pp.113-129, 1998.

[7] V.P.Godambe, Ed., *Estimating Functions*. New York: Oxford Univ. Press, 1991.

[8] A.Hyvärinen, "Independent Component Analysis in the Presence of Gaussian Noise by Maximizing Joint Likelihood," *Neurocomputing*, vol.22, pp.49-67, 1998.

[9] A.Hyvärinen, "Gaussian Moments for Noisy Independent Component Analysis," *IEEE Signal Processing Letters*, Vol.6, pp.145–147, 1999.

[10] A.Hyvärinen, "Survey on independent component analysis," *Neural Computing Surveys*, vol.2, pp.94–128, 1999.