Signal Processing 113 (2015) 95–103

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Model selection using limiting distributions of second-order blind source separation algorithms

Katrin Illner [a,b], Jari Miettinen [c], Christiane Fuchs [a,b], Sara Taskinen [c], Klaus Nordhausen [d], Hannu Oja [d], Fabian J. Theis [a,b,*]

[a] Institute of Computational Biology, German Research Center for Environmental Health, 85764 Helmholtz Zentrum München, Germany
[b] Chair of Mathematical Modeling of Biological Systems, Center for Mathematics, 85748 Garching, Technische Universität München, Germany
[c] Department of Mathematics and Statistics, 40014 University of Jyväskylä, Finland
[d] Department of Mathematics and Statistics, 20014 University of Turku, Finland

ARTICLE INFO

Article history:
Received 3 October 2014
Accepted 21 January 2015
Available online 30 January 2015

Keywords:
Joint diagonalization
SOBI
Asymptotic normality
Pattern identification

ABSTRACT

Signals, recorded over time, are often observed as mixtures of multiple source signals. To extract relevant information from such measurements one needs to determine the mixing coefficients. In case of weakly stationary time series with uncorrelated source signals, this separation can be achieved by jointly diagonalizing sample autocovariances at different lags, and several algorithms address this task. Often the mixing estimates contain close-to-zero entries and one wants to decide whether the corresponding source signals have a relevant impact on the observations or not. To address this question of model selection we consider the recently published second-order blind identification procedures SOBIdef and SOBIsym which provide limiting distributions of the mixing estimates. For the first time, such distributions enable informed decisions about the presence of second-order stationary source signals in the data. We consider a family of linear hypothesis tests and information criteria to perform model selection as second step after parameter estimation. In simulations we consider different time series models. We validate the model selection performance and demonstrate a good recovery of the true zero pattern of the mixing matrix.

1. Introduction

Time resolved signals appear in a large variety of contexts, and often one observes a multivariate mixture of different signals rather than separated ones. In blind source separation (BSS) we assume a linear and instantaneous mixing model and aim to estimate the underlying source signals together with the mixing weights. In case of weakly stationary time series with uncorrelated source signals, a mixing matrix can be estimated based on the second-order statistics of the observations. The problem then reduces to jointly diagonalizing sample autocovariances at different lags. Many existing BSS algorithms are based on this idea [1–5]. A review on joint diagonalization algorithms is given in [6]. Applications range from audio recordings to biomedical signal or image data. For the latter the assumption of uncorrelated source components can be extended to the spatial dimension of the data [7,8]. In an application to high dimensional functional magnetic resonance imaging (fMRI), for example, patients alternately passed through periods of rest and photic stimulus.

* Corresponding author at: Institute of Computational Biology, German Research Center for Environmental Health, 85764 Neuherberg, Helmholtz Zentrum München, Germany.
E-mail address: fabian.theis@helmholtz-muenchen.de (F.J. Theis).

http://dx.doi.org/10.1016/j.sigpro.2015.01.017
0165-1684/

In comparison to other BSS methods, joint diagonalization could identify a signal with high coherence to the stimulus [8]. Another widely used measuring technique is electroencephalography (EEG). Here, the brain's electrical activity is recorded and joint diagonalization could successfully separate artifacts like eye movement or blinking from the data [9]. If the EEG signals arise from correlated stimulation of the left and right somatosensory cortices a large number and wide range of time delays is preferable [10].

To draw further conclusions from source separation, one often wants to know whether single source signals are present in a specific observation. More precisely, one wants to decide whether close-to-zero entries of the mixing estimate are actually zero or not. This is commonly done by thresholding which lacks statistical motivation. To provide informative decisions we develop suitable model selection criteria. To that end, we consider the recently published second-order blind identification versions SOBIdef [11] and SOBIsym [12]. For both algorithms the authors showed that the (un-)mixing estimates are asymptotically normally distributed under mild conditions, and they derived limiting variances of the estimates when the time series length goes to infinity. Based on these distributions we create a framework to perform model selection on the mixing estimates. Here, we use a family of linear hypothesis tests and different information criterions including the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). To speed up the selection process we also consider an alternative information criterion that does not require the maximum likelihood parameter estimates.

In the first part, we state the second-order source separation problem (Section 2) and shortly review the algorithms SOBIdef and SOBIsym (Section 3). Their practical estimation performance has not been evaluated yet. To figure it out, we compare both algorithms to the established methods SOBI [2] and the non-orthogonal ACDC [4]. We find that SOBIsym achieves the same estimation results as SOBI but with the gain of knowing the limiting distribution of the (un-)mixing estimates (Section 4). In the main part, we then demonstrate how the additional information about the distribution can be used to choose between different candidates for the mixing matrix (Section 5). In simulations we consider a BSS model where the mixing matrix contains zero and close-to-zero entries (Section 6). For both algorithms SOBIdef and SOBIsym the testing performance could be validated and we show the percentages of correctly reconstructed zero-patterns among different time series models and for the different selection approaches.

Throughout the paper we use bold symbols to denote random variables and solid symbols to denote parameters and realizations of random variables.

2. A second-order blind source separation model

Let $\{\boldsymbol{x}(t)\}_{t\in\mathbb{Z}}$ be a p-variate observable time series that is weakly stationary. This means that the mean and the autocovariance at any lag $\tau\in\mathbb{N}$ do not change with respect to time. After mean-removal we assume a zero-centered process that is generated by the following linear mixing model:

$$\boldsymbol{x}(t)=\Omega\boldsymbol{z}(t),\quad t\in\mathbb{Z}. \tag{1}$$

Here, Ω denotes a deterministic full rank $p\times p$ mixing matrix and $\{\boldsymbol{z}(t)\}_{t\in\mathbb{Z}}$ is a p-variate unobservable time series that is weakly stationary as well and has uncorrelated components. More precisely, we assume

(A1) $E(\boldsymbol{z}(t))=0$,
(A2) $\mathrm{Cov}(\boldsymbol{z}(t),\boldsymbol{z}(t))=I_p$,
(A3) $\mathrm{Cov}(\boldsymbol{z}(t),\boldsymbol{z}(t+\tau))=\mathrm{Cov}(\boldsymbol{z}(t+\tau),\boldsymbol{z}(t))=\Lambda_\tau$ is diagonal for all lags $\tau\in\mathbb{N}$, and
(A4) for all $i\neq j\in\{1,\dots,p\}$ there exists a lag $\tau\in\mathbb{N}$ such that $\lambda_{\tau i}\neq\lambda_{\tau j}$ with $\lambda_{\tau i}$ and $\lambda_{\tau j}$ being the ith and jth diagonal entries of Λ_τ, respectively.

With the scaling to unit variance in $(A2)$ and the assumption $(A4)$ the mixing becomes unique up to a sign-changing permutation: if $\boldsymbol{x}(t)=\Omega_1\boldsymbol{z}_1(t)=\Omega_2\boldsymbol{z}_2(t)$, then $\Omega_2=\Omega_1 B$ and $\boldsymbol{z}_2(t)=B^{-1}\boldsymbol{z}_1(t)$, where B contains exactly one non-zero entry per row and column and these entries equal ± 1. This restriction on B follows from the spectral theorem [13].

In second-order source separation we consider the second-order statistics of the observable process, and with these we estimate the mixing matrix Ω as well as the unobservable process $\{\boldsymbol{z}(t)\}_{t\in\mathbb{Z}}$. The autocovariance of $\{\boldsymbol{x}(t)\}_{t\in\mathbb{Z}}$ at lag $\tau\in\mathbb{N}$ is of the form:

$$\mathrm{Cov}(\boldsymbol{x}(t),\boldsymbol{x}(t+\tau))=\Omega\Lambda_\tau\Omega',$$

where $\Lambda_0=I_p$ at lag zero. Let now $x(1),\dots,x(T)$ be observations at subsequent time points. The sample autocovariance at lag τ is then given as

$$S_\tau=\frac{1}{T-\tau}\sum_{t=1}^{T-\tau}x(t)x(t+\tau)'.$$

To determine an unmixing estimate we jointly diagonalize sample autocovariances at distinct lags $\tau_1,\dots,\tau_K$. We assume that $\{\tau_1,\dots,\tau_K\}\subseteq\mathbb{N}$ is such that $(A4)$ also holds for $\{\tau_1,\dots,\tau_K\}$ instead of $\mathbb{N}$. For better readability, we denote the corresponding autocovariances as $S_1,\dots,S_K$ even if the lags are different from $1,\dots,K$. An unmixing estimate is then a $p\times p$ matrix $\Gamma=(\gamma_1,\dots,\gamma_p)'$ that minimizes the off-diagonal elements of $\Gamma S_k\Gamma'$ for all k=1,…,K in the sense that

$$f^*(\Gamma)=\sum_{k=1}^{K}\|\mathrm{off}(\Gamma S_k\Gamma')\|_F^2$$

is minimized under the constraint $\Gamma S_0\Gamma'=I_p$. Here, $\mathrm{off}(M)=M-\mathrm{diag}(M)$ with $\mathrm{diag}(M)$ being a diagonal matrix consisting of the diagonal entries of M, and $\|\,.\,\|_F$ denotes the Frobenius norm of a matrix. The above minimization is equivalent to the maximization of

$$f(\Gamma)=\sum_{k=1}^{K}\|\mathrm{diag}(\Gamma S_k\Gamma')\|_F^2=\sum_{j=1}^{p}\sum_{k=1}^{K}(\gamma_j{}'S_k\gamma_j)^2 \tag{2}$$

under the same constraint. From the spectral theorem it follows that an optimal solution Γ is indeed an estimate of the unmixing matrix.