Separation of uncorrelated stationary time series using autocovariance matrices

Jari Miettinen*[1], Katrin Illner[2], Klaus Nordhausen[3], Hannu Oja[3], Sara Taskinen[1] and Fabian J. Theis[4]

[1]University of Jyväskylä

[2]Helmholtz Center Munich, Germany

[3]University of Turku, Finland.

[4]Technical University of Munich, Germany

May 15, 2014

Abstract

Blind source separation (BSS) is a signal processing tool, which is widely used in various fields. Examples include biomedical signal separation, brain imaging and economic time series applications. In BSS, one assumes that the observed p time series are linear combinations of p latent uncorrelated weakly stationary time series. The aim is then to find an estimate for an unmixing matrix, which transforms the observed time series back to uncorrelated latent time series. In SOBI (Second Order Blind Identification) joint diagonalization of the covariance matrix and autocovariance matrices with several lags is used to estimate the unmixing matrix. The rows of an unmixing matrix can be

*Address for correspondence: Jari Miettinen, Department of Mathematics and Statistics, 40014 University of Jyväskylä, Finland. E-mail: jari.p.miettinen@jyu.fi

derived either one by one (deflation-based approach) or simultaneously (symmetric approach). The latter of these approaches is well-known especially in signal processing literature, however, the rigorous analysis of its statistical properties has been missing so far. In this paper, we fill this gap and investigate the statistical properties of the symmetric SOBI estimate in detail and find its limiting distribution under general conditions. The asymptotical efficiencies of symmetric SOBI estimate are compared to those of recently introduced deflation-based SOBI estimate under general multivariate MA(∞) processes. The theory is illustrated by some finite-sample simulation studies as well as a real EEG data example.

Keywords: Asymptotic normality; Blind source separation; Joint diagonalization; MA(∞); SOBI

1 Introduction

In blind signal separation or blind source separation (BSS) one assumes that a p-variate observable random vector $\boldsymbol{x}$ is a linear mixture of p-variate latent source vector $\boldsymbol{z}$. The model can thus be written as $\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{\Omega}\boldsymbol{z}$, where the full rank $p \times p$ matrix $\boldsymbol{\Omega}$ is so called *mixing matrix* and $\boldsymbol{z}$ is a random p-vector with certain preassigned properties. The p-vector $\boldsymbol{\mu}$ is a location parameter and usually considered as a nuisance parameter, since the main goal in BSS is to find an estimate for an *unmixing matrix* $\boldsymbol{\Gamma}$ such that $\boldsymbol{\Gamma}\boldsymbol{x} \sim \boldsymbol{z}$, based on a $p \times n$ data matrix $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ from the distribution of $\boldsymbol{x}$.

The BSS model was formulated for signal processing and computer science applications in the early 1980's and since then, many approaches have been suggested to solve the problem under various assumptions on $\boldsymbol{z}$. For an account of the early history of BSS, see Jutten and Taleb (2000). The most popular BSS approach is *independent component analysis (ICA)*, which assumes that $E(\boldsymbol{z}) = 0$ and $E(\boldsymbol{z}\boldsymbol{z}') = \boldsymbol{I}_p$, and that the components of $\boldsymbol{z}$ are mutually independent. The ICA model is a semiparametric model as the marginal distributions of the components of $\boldsymbol{z}$ are not specified at all. For identifiability of the parameters, one has to assume, however, that at most one of the components is normally distributed. Typical algorithms for ICA use centering and whitening as preprocessing steps: Write $\boldsymbol{\Omega} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}'$ for the singular value decomposition (SVD) of the mixing matrix $\boldsymbol{\Omega}$. Then, under the above mentioned assumptions on $\boldsymbol{z}$, $E(\boldsymbol{x}) = \boldsymbol{\mu}$ and $Cov(\boldsymbol{x}) = \boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{\Lambda}^2\boldsymbol{U}'$, and therefore $\boldsymbol{V}\boldsymbol{U}'\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x}-\boldsymbol{\mu}) = \boldsymbol{z}$. An iterative algorithm can then be applied to $\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x}-\boldsymbol{\mu})$ to find the orthogonal matrix $\boldsymbol{V}\boldsymbol{U}'$. For an overview of ICA from a signal processing perspective, see for example Hyvärinen et al. (2002) and Comon and Jutten (2010).

Since the late 1990's, there has been an increasing interest in ICA methods among statisticians. Oja et al. (2006), for example, used two scatter matrices, $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$, with the independence property to solve the ICA problem. We say that a $p \times p$ matrix valued functional $\boldsymbol{S}(F)$ is a *scatter matrix* if it is symmetric, positive definite and affine equivariant in the sense that $\boldsymbol{S}(F_{\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}}) = \boldsymbol{A}\boldsymbol{S}(F_{\boldsymbol{x}})\boldsymbol{A}'$ for all full-rank $p \times p$ matrices $\boldsymbol{A}$ and for all p-vectors $\boldsymbol{b}$. Moreover, a scatter matrix $\boldsymbol{S}(F)$ has the *independence property* if $\boldsymbol{S}(F_{\boldsymbol{z}})$ is a diagonal matrix for all $\boldsymbol{z}$ with independent components. An unmixing matrix $\boldsymbol{\Gamma}$ and a diagonal matrix $\boldsymbol{\Lambda}$ (diagonal elements in a decreasing order) then solve the estimating equations

$$\boldsymbol{\Gamma}\boldsymbol{S}_1\boldsymbol{\Gamma}' = \boldsymbol{I}_p \quad \text{and} \quad \boldsymbol{\Gamma}\boldsymbol{S}_2\boldsymbol{\Gamma}' = \boldsymbol{\Lambda}.$$

The independent components in $\boldsymbol{\Gamma}\boldsymbol{x}$ are thus standardized with respect to $\boldsymbol{S}_1$ and uncorrelated with respect to $\boldsymbol{S}_2$, and $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}$ can be found as eigenvector-eigenvalue solutions for $\boldsymbol{S}_1^{-1}\boldsymbol{S}_2$. An example of such ICA methods is the classical FOBI (fourth order blind identification) method which uses $\boldsymbol{S}_1(F_{\boldsymbol{x}}) = Cov(\boldsymbol{x})$ and

$$\boldsymbol{S}_2(F_{\boldsymbol{x}}) = E\left[(\boldsymbol{x}-E(\boldsymbol{x}))(\boldsymbol{x}-E(\boldsymbol{x}))'Cov(\boldsymbol{x})^{-1}(\boldsymbol{x}-E(\boldsymbol{x}))(\boldsymbol{x}-E(\boldsymbol{x}))'\right].$$

See Oja et al. (2006), for example. The unmixing matrix estimate $\hat{\mathbf{\Gamma}}$ is naturally obtained by replacing the population values by their sample counterparts.

The limiting statistical properties of several ICA unmixing matrix estimates have been developed quite recently and mainly for iid data: For the limiting behavior of the unmixing matrix estimate based on two scatter matrices, see Ilmonen et al. (2010a). For other recent work and different estimation procedures for ICA, see for example Hastie and Tibshirani (2003), Chen and Bickel (2006), Bonhomme and Robin (2009), Ilmonen and Paindaveine (2011), Allassonniere and Younes (2012), Samworth and Yuan (2013) and Hallin and Mehta (2014).

In applications such as the analysis of medical images or signals (EEG, MEG or fMRI) and financial or geostatistical times series, the assumption of independent observations does not usually hold. Nevertheless, ICA has been considered in this context in Chen et al. (2007), Garcia-Ferrer et.al (2011), Garcia-Ferrer et.al (2012), Lee et al. (2011), Poncela (2012) and Schachtner et al. (2008) among others. See also Matteson and Tsay (2011) for a slightly different model. Apart from these results, other BSS models have been developed for time series data in signal processing literature. The so called AMUSE (Algorithm for Multiple Unknown Signals Extraction) and SOBI (Second Order Blind Identification) procedures for the stationary time series BSS models were suggested by Tong et al. (1990) and Belouchrani et al. (1997), respectively. Miettinen et al. (2012, 2014) provided careful analysis of the statistical properties of the AMUSE and so called deflation-based SOBI estimates. Nordhausen (2014) considered methods that assume only local stationarity.

In this paper we will continue the work of Miettinen et al. (2012, 2014) and derive the statistical properties of so called symmetric SOBI estimate. We will use a real EEG data example to illustrate how the theoretical results derived in this paper may be used to measure the accuracy of the unmixing matrix estimates. The structure of this paper is as follows. In Section 2 we introduce the blind source separation model which assumes second order stationary components. In Section 3 we recall the definition for the deflation-based SOBI functional and define symmetric SOBI functional using the Lagrange multiplier technique. The theoretical properties of deflation-based and symmetric SOBI estimators are given in general case in Section 4. Further, in Section 5 the limiting distributions of the two SOBI estimators will be more concretely compared under the assumption of $MA(\infty)$ processes. The theoretical results are illustrated using simulation studies in Section 6.1 and a real EEG data example in Section 7 before the paper is concluded in Section 8. Asymptotical results for the symmetric SOBI estimates are proven in the appendix.

2 Second order source separation model

We assume that the observable p-variate time series $\boldsymbol{x} = (\boldsymbol{x}_t)_{t=0,\pm1,\pm2,\ldots}$ are distributed according to

$$\boldsymbol{x}_t = \boldsymbol{\mu} + \boldsymbol{\Omega} \boldsymbol{z}_t, \quad t = 0, \pm1, \pm2, \ldots, \tag{1}$$

where $\boldsymbol{\mu}$ is a p-vector, $\boldsymbol{\Omega}$ is a full-rank $p \times p$ mixing matrix and $\boldsymbol{z} = (\boldsymbol{z}_t)_{t=0,\pm1,\pm2,\ldots}$ is a p-variate latent time series that satisfies

(A1) $E(\boldsymbol{z}_t) = 0$ and $E(\boldsymbol{z}_t \boldsymbol{z}_t') = \boldsymbol{I}_p$.

(A2) $E(\boldsymbol{z}_t \boldsymbol{z}_{t+\tau}') = E(\boldsymbol{z}_{t+\tau} \boldsymbol{z}_t') = \boldsymbol{\Lambda}_\tau$ is diagonal for all $\tau = 1, 2, \ldots$.

This is again a semiparametric model as only the moment assumptions (A1)-(A2) of the time series in $\boldsymbol{z}$ are made. The assumptions state that the p time series in $\boldsymbol{z}$ are weakly stationary and uncorrelated. This model is called the *second order source separation (SOS) model*. A model with stronger assumptions, i.e. the *independent component time series model*, is obtained if the condition (A2) is replaced by the condition

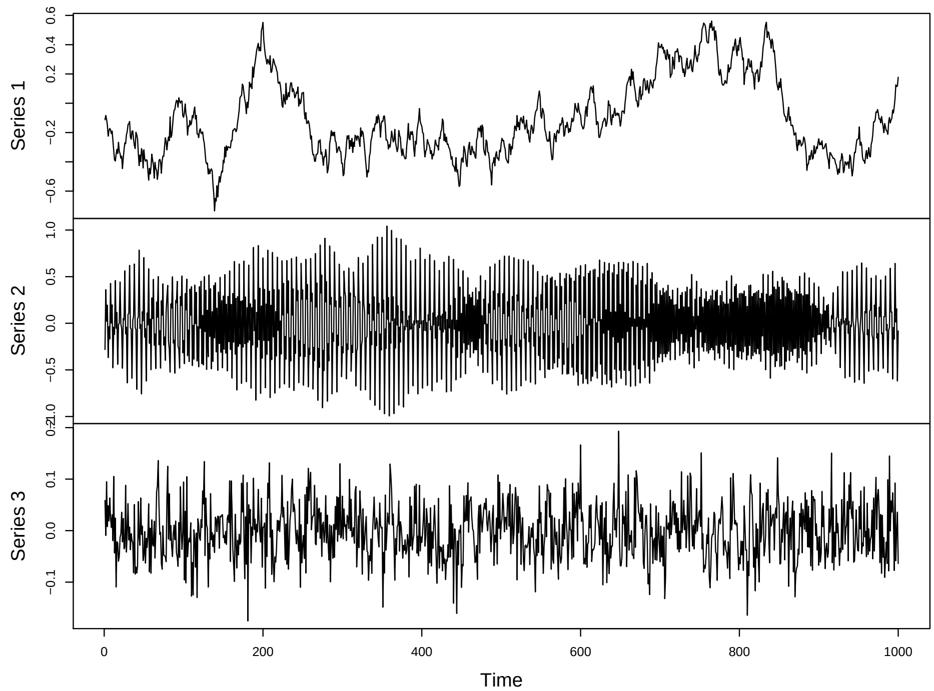

Figure 1: Example time series $\boldsymbol{z}$: Three independent stationary AR series.

(A2*) the p times series in $\boldsymbol{z}$ are mutually independent and $E(\boldsymbol{z}_t\boldsymbol{z}'_{t+\tau}) = E(\boldsymbol{z}_{t+\tau}\boldsymbol{z}'_t) = \boldsymbol{\Lambda}_\tau$ is diagonal for all $\tau = 1, 2, \ldots$.

Figure 1 serves as an example of a 3-variate time series $\boldsymbol{z}$ with three independent components, namely AR processes with coefficient vectors $(0.9, 0.09)$, $(0, 0, -0.99)$ and $(0, 0.3)$. The observable 3-variate time series $\boldsymbol{x}$ consisting of three different mixtures of the latent time series in $\boldsymbol{z}$ are shown in Figure 2. Given the observed time series $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$, the aim is to find an estimate $\hat{\boldsymbol{\Gamma}}$ of an unmixing matrix $\boldsymbol{\Gamma}$ such that $\boldsymbol{\Gamma x}$ has uncorrelated components. Clearly, $\boldsymbol{\Gamma} = \boldsymbol{C}\boldsymbol{\Omega}^{-1}$ is an unmixing matrix for any $p \times p$ matrix $\boldsymbol{C}$ with exactly one nonzero element in each row and in each column. Notice that the signs and order of the components of $\boldsymbol{z}$ and the signs and order of the columns of $\boldsymbol{\Omega}$ are confounded also in the BSS model. Additional assumptions are therefore needed in order to study the consistency and asymptotical properties of $\hat{\boldsymbol{\Gamma}}$. Contrary to ICA in the iid case, the mixing matrix may now be identifiable for any number of gaussian components. However, as we will see later in this paper, weak assumptions on the autocovariance matrices $\mathbf{\Lambda}_\tau$, $\tau = 1, 2, \ldots$, have to be made for the identifiability of our functionals and for the study of the asymptotic properties of corresponding estimates.

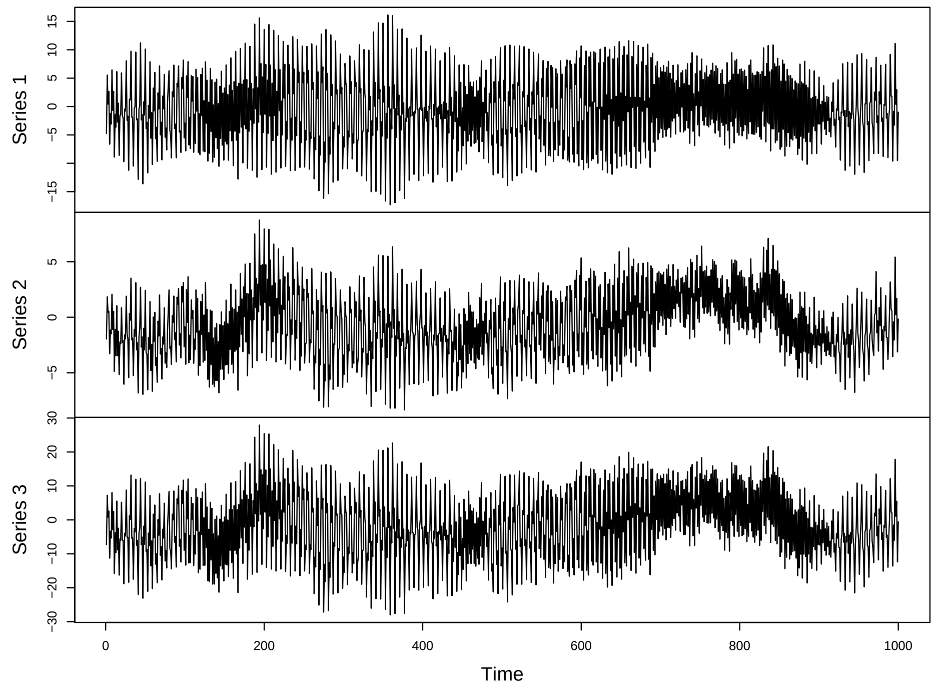

Figure 2: Time series $\boldsymbol{x}$: Mixtures of the three independent stationary AR series in Figure 1.

3 BSS functionals based on autocovariance matrices

3.1 Joint diagonalization of autocovariance matrices

The separation of uncorrelated stationary time series can be solely based on autocovariances and cross-autocovariances of the p time series. Assume that $\boldsymbol{x}$ follows a centered SOS model with $\boldsymbol{\mu} = \mathbf{0}$. This is not a restriction in our case as the asymptotical properties of the estimated autocovariances are the same for known and unknown

$\boldsymbol{\mu}$. It then follows that

$$E(\boldsymbol{x}_t\boldsymbol{x}'_{t+\tau}) = \boldsymbol{\Omega}\boldsymbol{\Lambda}_\tau\boldsymbol{\Omega}', \quad \tau = 0, 1, 2, \ldots.$$

Assume also that, for some lag $\tau > 0$, the diagonal elements of the autocovariance matrix $E(\boldsymbol{z}_t\boldsymbol{z}'_{t+\tau}) = \boldsymbol{\Lambda}_\tau$ are distinct. An unmixing matrix functional $\boldsymbol{\Gamma}_\tau$ is then defined as a $p \times p$ matrix that satisfies

$$\boldsymbol{\Gamma}_\tau E(\boldsymbol{x}_t\boldsymbol{x}'_t)\boldsymbol{\Gamma}'_\tau = \boldsymbol{I}_p \quad \text{and} \quad \boldsymbol{\Gamma}_\tau E(\boldsymbol{x}_t\boldsymbol{x}'_{t+\tau})\boldsymbol{\Gamma}'_\tau = \boldsymbol{P}_\tau\boldsymbol{\Lambda}_\tau\boldsymbol{P}'_\tau,$$

where $\boldsymbol{P}_\tau\boldsymbol{\Lambda}_\tau\boldsymbol{P}'_\tau$ is a diagonal matrix with the same diagonal elements as in $\boldsymbol{\Lambda}_\tau$ but in a decreasing order. (As in PCA and ICA, the signs of the rows of $\boldsymbol{\Gamma}_\tau$ are not fixed in this definition.) The components of $\boldsymbol{\Gamma}_\tau\boldsymbol{x}$ are thus the components of $\boldsymbol{z}$ in a permuted order. The permutation matrix $\boldsymbol{P}_\tau$ remains unidentifiable. Notice that $\boldsymbol{\Gamma}_\tau$ is affine equivariant, that is, the transformation $\boldsymbol{x} \to \boldsymbol{A}\boldsymbol{x}$ with a full-rank $p \times p$ matrix $\boldsymbol{A}$ induces the transformation $\boldsymbol{\Gamma}_\tau \to \boldsymbol{\Gamma}_\tau\boldsymbol{A}^{-1}$. This implies that $\boldsymbol{\Gamma}_\tau\boldsymbol{x}$ does not depend on the mixing matrix $\boldsymbol{\Omega}$ at all. The corresponding sample statistic, the so called AMUSE (Algorithm for Multiple Unknown Signals Extraction) estimator, was proposed by Tong et al. (1990). Figure 3 shows the estimated latent sources obtained with AMUSE, $\tau = 1$, from the data in Figure 2. See Miettinen et al. (2012) for a recent study of the statistical properties of the AMUSE estimate.

The drawback of the AMUSE procedure is the assumption that, for the chosen lag τ, the eigenvalues in $\boldsymbol{\Lambda}_\tau$ must be distinct. This is of course never known in practice. Therefore, the choice of τ may have a huge impact on the performance of the method, as only information coming from $S_0 = E(\boldsymbol{x}_t\boldsymbol{x}'_t)$ and $S_\tau = E(\boldsymbol{x}_t\boldsymbol{x}'_{t+\tau})$ is used. To overcome this drawback, Belouchrani et al. (1997) proposed the SOBI (Second Order Blind Indentification) algorithm that aims to jointly diagonalize several autocovariance matrices as follows. Let $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_K$ be K autocovariance matrices

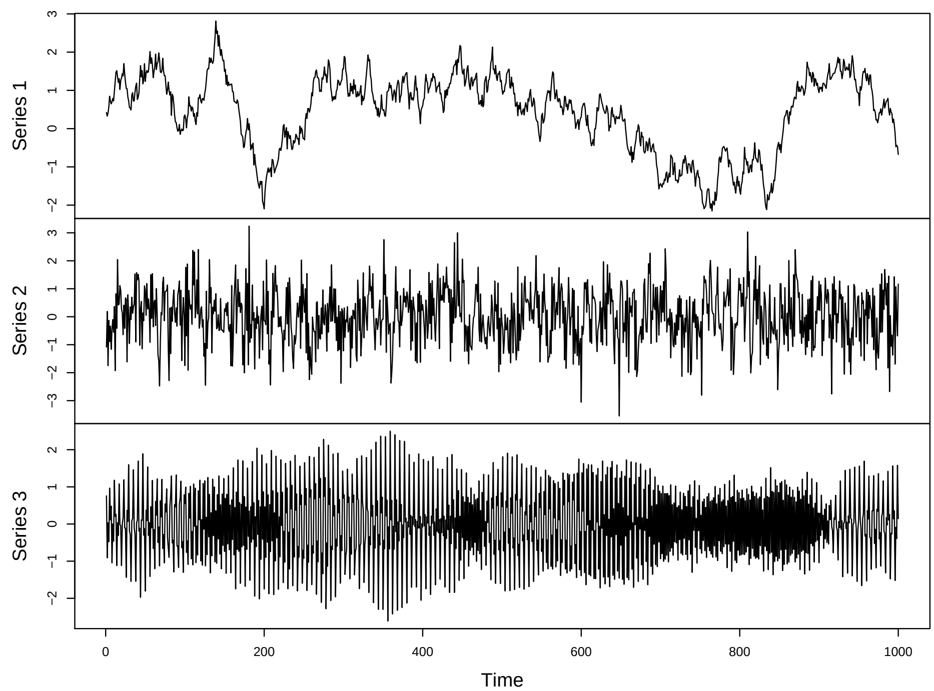

Figure 3: Time series $\hat{\boldsymbol{\Gamma}}_1\boldsymbol{x}$ obtained with AMUSE from $\boldsymbol{x}$ in Figure 2.

with distinct lags $\tau_1, \ldots, \tau_K$. The $p \times p$ unmixing matrix functional $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_p)'$ is the matrix that minimizes

$$\sum_{k=1}^{K} ||\text{off}(\boldsymbol{\Gamma}\boldsymbol{S}_k\boldsymbol{\Gamma}')||^2$$

under the constraint $\boldsymbol{\Gamma}\boldsymbol{S}_0\boldsymbol{\Gamma}' = \boldsymbol{I}_p$, or, equivalently, maximizes

$$\sum_{k=1}^{K} ||\text{diag}(\boldsymbol{\Gamma}\boldsymbol{S}_k\boldsymbol{\Gamma}')||^2 = \sum_{j=1}^{p}\sum_{k=1}^{K}(\boldsymbol{\gamma}_j'\boldsymbol{S}_k\boldsymbol{\gamma}_j)^2$$

under the same constraint. Here we write $\text{diag}(\boldsymbol{S})$ for a $p \times p$ diagonal matrix with the diagonal elements as in $\boldsymbol{S}$ and $\text{off}(\boldsymbol{S}) = \boldsymbol{S} - \text{diag}(\boldsymbol{S})$.

Next notice that, as $\boldsymbol{\Gamma}\boldsymbol{S}_0\boldsymbol{\Gamma}' = \boldsymbol{I}_p$, then $\boldsymbol{\Gamma} = \boldsymbol{U}\boldsymbol{S}_0^{-1/2}$ for some orthogonal $p \times p$ matrix $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p)'$. If then $\boldsymbol{R}_k = \boldsymbol{S}_0^{-1/2}\boldsymbol{S}_k\boldsymbol{S}_0^{-1/2}$, $k = 1, \ldots, K$, are the autocorrelation matrices, the solution for $\boldsymbol{U}$ can be found by maximizing

$$\sum_{k=1}^{K} ||\text{diag}(\boldsymbol{U}\boldsymbol{R}_k\boldsymbol{U}')||^2 = \sum_{j=1}^{p}\sum_{k=1}^{K}(\boldsymbol{u}_j'\boldsymbol{R}_k\boldsymbol{u}_j)^2 \tag{2}$$

under the orthogonality constraints $\boldsymbol{U}\boldsymbol{U}' = \boldsymbol{I}_p$.

In the literature, several algorithms to solve the maximization problem in (2) are proposed: In deflation-based approach, the rows of $\boldsymbol{U}$ are found one by one using some pre-assigned rule. In the symmetric approach, the rows are found simultaneously. The solution $\boldsymbol{\Gamma} = \boldsymbol{U}\boldsymbol{S}_0^{-1/2}$ naturally depends on the approach as well as on the concrete algorithm used in the optimization. In the following we consider deflation-based and symmetric approaches in more detail.

3.2 Deflation-based approach

In the deflation-based approach, the rows of an unmixing matrix functional $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_p)'$ are found one by one so that $\boldsymbol{\gamma}_j$, $j = 1, \ldots, p-1$, maximizes

$$\sum_{k=1}^{K}(\boldsymbol{\gamma}_j'\boldsymbol{S}_k\boldsymbol{\gamma}_j)^2, \tag{3}$$

under the constraints $\boldsymbol{\gamma}_i'\boldsymbol{S}_0\boldsymbol{\gamma}_j = \delta_{ij}$, $i = 1, \ldots, j$. Recall that the Kronecker delta $\delta_{ij} = 1$ (0) as $i = j$ ($i \neq j$).

The solution $\boldsymbol{\gamma}_j$ then optimizes the Lagrangian function

$$L(\boldsymbol{\gamma}_j, \boldsymbol{\theta}_j) = \sum_{k=1}^{K}(\boldsymbol{\gamma}_j'\boldsymbol{S}_k\boldsymbol{\gamma}_j)^2 - \theta_{jj}(\boldsymbol{\gamma}_j'\boldsymbol{S}_0\boldsymbol{\gamma}_j - 1) - \sum_{i=1}^{j-1}\theta_{ji}\boldsymbol{\gamma}_i'\boldsymbol{S}_0\boldsymbol{\gamma}_j,$$

where $\boldsymbol{\theta}_j = (\theta_{j1}, \ldots, \theta_{jj})'$ are the Lagrangian multipliers. Write

$$\boldsymbol{T}(\boldsymbol{\gamma}) = \sum_{k=1}^{K}(\boldsymbol{\gamma}'\boldsymbol{S}_k\boldsymbol{\gamma})\boldsymbol{S}_k\boldsymbol{\gamma}. \tag{4}$$

The unmixing matrix functional $\boldsymbol{\Gamma}$ found in this way then satisfies the following estimating equations (Miettinen et al., 2014).

Definition 1. *The deflation-based unmixing matrix functional* $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_p)'$ *solves the* $p-1$ *estimating equations*

$$\boldsymbol{T}(\boldsymbol{\gamma}_j) = \boldsymbol{S}_0 \left(\sum_{r=1}^{j} \boldsymbol{\gamma}_r \boldsymbol{\gamma}_r' \right) \boldsymbol{T}(\boldsymbol{\gamma}_j), \quad j = 1, \ldots, p-1.$$

Recall that $\boldsymbol{\Gamma} = \boldsymbol{U}\boldsymbol{S}_0^{-1/2}$ with some orthogonal matrix $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p)'$ and, in the deflation-based approach, the rows of $\boldsymbol{U}$ are found one by one as well. The estimating equations then suggest the following fixed point algorithm for the deflation-based solution. After finding $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{j-1}$, the following two steps are repeated until convergence to get $\boldsymbol{u}_j$.

$$\text{step 1:} \quad \boldsymbol{u}_j \leftarrow \left(\boldsymbol{I}_p - \sum_{i=1}^{j-1} \boldsymbol{u}_i \boldsymbol{u}_i' \right) \boldsymbol{T}(\boldsymbol{u}_j).$$

$$\text{step 2:} \quad \boldsymbol{u}_j \leftarrow ||\boldsymbol{u}_j||^{-1} \boldsymbol{u}_j.$$

Here $\boldsymbol{T}(\boldsymbol{u}) = \sum_{k=1}^{K} (\boldsymbol{u}'\boldsymbol{R}_k\boldsymbol{u})\boldsymbol{R}_k\boldsymbol{u}$. Notice that the algorithm naturally needs initial values for each $\boldsymbol{u}_j$, $j = 1, \ldots, p-1$, and different initial values may change the rows of the estimate and produce them in a permuted order. Therefore, for $\boldsymbol{u}_j$, one should use several randomly selected initial values to guarantee that the true maximum in (3) is attained at each stage. For a more detailed study of this algorithm, see Appendix A in Miettinen et al. (2014).

3.3 Symmetric approach

In the symmetric approach, the rows of an unmixing matrix functional $\mathbf{\Gamma} = (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_p)'$ are found simultaneously. We then consider the maximization of

$$\sum_{j=1}^{p} \sum_{k=1}^{K} (\boldsymbol{\gamma}_j' \boldsymbol{S}_k \boldsymbol{\gamma}_j)^2,$$

under the constraint $\mathbf{\Gamma} \boldsymbol{S}_0 \mathbf{\Gamma}^T = \boldsymbol{I}_p$. The matrix $\mathbf{\Gamma}$ now optimizes the Lagrangian function

$$L(\mathbf{\Gamma}, \Theta) = \sum_{j=1}^{p} \sum_{k=1}^{K} (\boldsymbol{\gamma}_j' \boldsymbol{S}_k \boldsymbol{\gamma}_j)^2 - \sum_{j=1}^{p} \theta_{jj} (\boldsymbol{\gamma}_j' \boldsymbol{S}_0 \boldsymbol{\gamma}_j - 1) - \sum_{j=1}^{p} \sum_{i=1}^{j-1} \theta_{ij} \boldsymbol{\gamma}_i' \boldsymbol{S}_0 \boldsymbol{\gamma}_j,$$

where the symmetric matrix $\mathbf{\Theta} = (\theta_{ij})$ contains the $p(p+1)/2$ Lagrangian multipliers of the optimization problem. At the solution $\mathbf{\Gamma}$ we then have

$$2\boldsymbol{T}(\boldsymbol{\gamma}_j) = \boldsymbol{S}_0 \left(2\theta_{jj} \boldsymbol{\gamma}_j + \sum_{i=1}^{j-1} \theta_{ij} \boldsymbol{\gamma}_i + \sum_{i=j+1}^{p} \theta_{ji} \boldsymbol{\gamma}_i \right),$$

where $\boldsymbol{T}(\boldsymbol{\gamma})$ is as in (4). Multiplying both sides from the left by $\boldsymbol{\gamma}_i'$ gives $2\boldsymbol{\gamma}_i' \boldsymbol{T}(\boldsymbol{\gamma}_j) = \theta_{ij}$, for $i < j$, and $2\boldsymbol{\gamma}_i' \boldsymbol{T}(\boldsymbol{\gamma}_j) = \theta_{ji}$, for $i > j$. Hence the solution $\mathbf{\Gamma}$ must satisfy the following estimating equations.

Definition 2. *The symmetric unmixing matrix functional* $\mathbf{\Gamma} = (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_p)'$ *solves the estimating equations*

$$\boldsymbol{\gamma}_i' \boldsymbol{T}(\boldsymbol{\gamma}_j) = \boldsymbol{\gamma}_j' \boldsymbol{T}(\boldsymbol{\gamma}_i) \quad \textit{and} \quad \boldsymbol{\gamma}_i' \boldsymbol{S}_0 \boldsymbol{\gamma}_j = \delta_{ij}, \quad i, j = 1, \dots, p.$$

Notice that the exact joint diagonalization is possible only if the matrices $\boldsymbol{R}_1, \dots, \boldsymbol{R}_K$ have the same sets of eigenvectors. This is naturally true for the population matrices in the SOS model. For estimated autocorrelation matrices from a continuous SOS

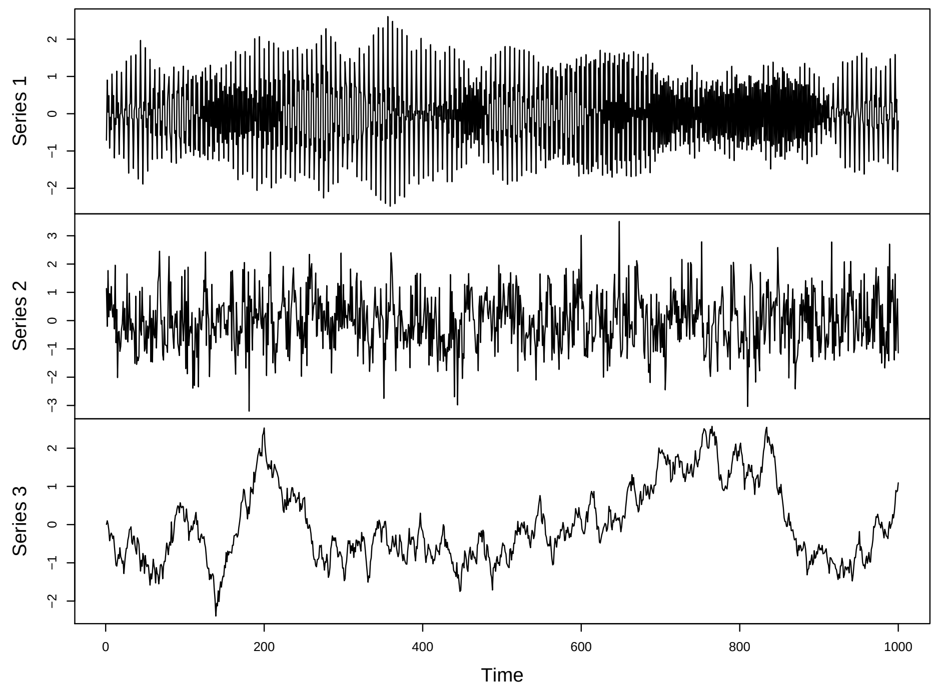

Figure 4: Time series $\hat{\mathbf{\Gamma}}\boldsymbol{x}$ obtained with symmetric SOBI from $\boldsymbol{x}$ in Figure 2.

model, the eigenvectors are however almost surely different. As $\mathbf{\Gamma} = \boldsymbol{U}\boldsymbol{S}_0^{-1/2}$, the estimating equations for $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p)'$ are

$$\boldsymbol{u}_i'\boldsymbol{T}(\boldsymbol{u}_j) = \boldsymbol{u}_j'\boldsymbol{T}(\boldsymbol{u}_i) \;\text{ and }\; \boldsymbol{u}_i'\boldsymbol{u}_j = \delta_{ij}, \quad i, j = 1, \ldots, p,$$

where again $\boldsymbol{T}(\boldsymbol{u}) = \sum_{k=1}^{K}(\boldsymbol{u}'\boldsymbol{R}_k\boldsymbol{u})\boldsymbol{R}_k\boldsymbol{u}$. The equations then suggest a new fixed point algorithm with the two steps

$$\text{step 1:} \quad \boldsymbol{T} \leftarrow (\boldsymbol{T}(\boldsymbol{u}_1), \ldots, \boldsymbol{T}(\boldsymbol{u}_p))'$$

$$\text{step 2:} \quad \boldsymbol{U} \leftarrow (\boldsymbol{T}\boldsymbol{T}')^{-1/2}\boldsymbol{T}.$$

Figure 4 shows the estimated latent sources obtained from the data in Figure 2. In the signal processing literature, there are several other algorithms available for approximate simultaneous diagonalization of K matrices. The most popular one for SOBI is based on Jacobi rotations (Clarkson, 1988). Surprisingly, our new algorithm and the algorithm based on Jacobi rotations seem to yield exactly the same solutions for practical data sets. Notice also that the symmetric procedure does not fix the order of the rows of $\boldsymbol{U}$. To guarantee that the deflation-based and symmetric procedures estimate the same $\boldsymbol{U}$, we can reorder the rows of $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p)'$ so that

$$\sum_{k=1}^{K} (\boldsymbol{u}_1' \boldsymbol{R}_k \boldsymbol{u}_1)^2 \geq \cdots \geq \sum_{k=1}^{K} (\boldsymbol{u}_p' \boldsymbol{R}_k \boldsymbol{u}_p)^2.$$

4 Asymptotical properties of the SOBI estimators

In this section we derive the asymptotical properties of the two competing SOBI estimates under SOS model (1). We may assume without loss of generality that $\boldsymbol{\mu} = \boldsymbol{0}$ and consider the limiting properties of the deflation-based and symmetric SOBI estimates based on the autocovariance matrices $\boldsymbol{S}_0, \boldsymbol{S}_1, \ldots, \boldsymbol{S}_K$ with lags $0, 1, \ldots, K$. We then need some additional assumptions that are specific for these choices. First, we assume

(A3) the diagonal elements of $\sum_{k=1}^{K} \boldsymbol{\Lambda}_k^2$ are strictly decreasing.

Assumption (A3) guarantees the identifiability of the mixing matrix (with the specified autocovariance matrices) and fixes the order of the component time series in our model. Our unmixing matrix estimate is based on the sample autocovariance matrices $\hat{\boldsymbol{S}}_0, \hat{\boldsymbol{S}}_1, \ldots, \hat{\boldsymbol{S}}_K$. We then further assume that the estimates of the autocovariance matrices are root-T consistent, that is,

(A4) $\boldsymbol{\Omega} = \boldsymbol{I}_p$ and $\sqrt{T}(\hat{\boldsymbol{S}}_k - \boldsymbol{\Lambda}_k) = O_p(1)$, $k = 0, 1, \ldots, K$ as $T \to \infty$.

Whether (A4) is true or not depends on the distribution of the latent p-variate time series $\boldsymbol{z}$.

For the estimated autocovariance matrices, write then

$$\hat{\boldsymbol{T}}(\gamma) = \sum_{k=1}^{K} (\gamma' \hat{\boldsymbol{S}}_k \gamma) \hat{\boldsymbol{S}}_k \gamma.$$

The deflation-based and symmetric unmixing matrix estimates are obtained when the functionals are applied to estimated autocovariance matrices and, consequently, they solve the following estimating equations.

Definition 3. *The unmixing matrix estimate* $\hat{\boldsymbol{\Gamma}} = (\hat{\boldsymbol{\gamma}}_1, \ldots, \hat{\boldsymbol{\gamma}}_p)'$ *based on* $\hat{\boldsymbol{S}}_0$ *and* $\hat{\boldsymbol{S}}_1, \ldots, \hat{\boldsymbol{S}}_K$ *solves the estimating equations*

$$\hat{\boldsymbol{T}}(\hat{\boldsymbol{\gamma}}_j) = \hat{\boldsymbol{S}}_0 (\sum_{r=1}^{j} \hat{\boldsymbol{\gamma}}_r \hat{\boldsymbol{\gamma}}_r') \hat{\boldsymbol{T}}(\hat{\boldsymbol{\gamma}}_j), \quad j = 1, \ldots, p-1, \quad \textit{(deflation-based)}$$

or

$$\hat{\boldsymbol{\gamma}}_i' \hat{\boldsymbol{T}}(\hat{\boldsymbol{\gamma}}_j) = \hat{\boldsymbol{\gamma}}_j' \hat{\boldsymbol{T}}(\hat{\boldsymbol{\gamma}}_i) \quad \textit{and} \quad \hat{\boldsymbol{\gamma}}_i' \hat{\boldsymbol{S}}_0 \hat{\boldsymbol{\gamma}}_j = \delta_{ij}, \quad i, j = 1, \ldots, p \quad \textit{(symmetric).}$$

Using the estimating equations and assumptions (A1)-(A4), one easily derives the following results. The first part was already proven in Miettinen et al. (2014). For the proof of the second part, see the Appendix.

Theorem 1. *Under the assumptions (A1)-(A4) we have*

(i) the deflation-based $\hat{\boldsymbol{\Gamma}} = (\hat{\boldsymbol{\gamma}}_1, \ldots, \hat{\boldsymbol{\gamma}}_p)' \to_p \boldsymbol{I}_p$*, and for* $j = 1, \ldots, p$*,*

$$\begin{aligned}
\sqrt{T}\hat{\gamma}_{ji} &= -\sqrt{T}\hat{\gamma}_{ij} - (\sqrt{T}\hat{\boldsymbol{S}}_0)_{ij} + o_p(1), && i < j, \\
\sqrt{T}(\hat{\gamma}_{jj} - 1) &= -\frac{1}{2}\sqrt{T}((\hat{\boldsymbol{S}}_0)_{jj} - 1) + o_p(1), && i = j, \\
\sqrt{T}\hat{\gamma}_{ji} &= \frac{\sum_k \lambda_{kj} \left[(\sqrt{T}\hat{\boldsymbol{S}}_k)_{ji} - \lambda_{kj} (\sqrt{T}\hat{\boldsymbol{S}}_0)_{ji} \right]}{\sum_k \lambda_{kj} (\lambda_{kj} - \lambda_{ki})} + o_p(1), && i > j,
\end{aligned}$$

(ii) the symmetric $\hat{\mathbf{\Gamma}} = (\hat{\boldsymbol{\gamma}}_1, \ldots, \hat{\boldsymbol{\gamma}}_p)' \to_p \boldsymbol{I}_p$*, and for* $j = 1, \ldots, p$*,*

$$\sqrt{T}\hat{\gamma}_{jj} = -\frac{1}{2}\sqrt{T}((\hat{\boldsymbol{S}}_0)_{jj} - 1) + o_p(1), \qquad i = j$$

$$\sqrt{T}\hat{\gamma}_{ji} = \frac{\sum_k (\lambda_{kj} - \lambda_{ki}) \left[(\sqrt{T}\hat{\boldsymbol{S}}_k)_{ji} - \lambda_{kj}(\sqrt{T}\hat{\boldsymbol{S}}_0)_{ji} \right]}{\sum_k (\lambda_{kj} - \lambda_{ki})^2} + o_p(1), \qquad i \neq j.$$

First note that, for $\mathbf{\Omega} = \boldsymbol{I}_p$, the limiting distribution of the diagonal element $\sqrt{T}(\hat{\gamma}_{jj} - 1)$ only depends on the limiting distribution of $\sqrt{T}((\hat{\boldsymbol{S}}_0)_{jj} - 1)$, $j = 1, \ldots, p$. Hence, the comparison of the estimates should be made only using the off-diagonal elements. Also,

$$\sqrt{T}(\hat{\mathbf{\Gamma}} + \hat{\mathbf{\Gamma}}' - 2\boldsymbol{I}_p) = -\sqrt{T}(\hat{\boldsymbol{S}}_0 - \boldsymbol{I}_p) + o_p(1),$$

and the limiting behavior of $\sqrt{T}(\hat{\mathbf{\Gamma}} + \hat{\mathbf{\Gamma}}' - 2\boldsymbol{I}_p)$ is therefore similar for both approaches. If the joint limiting distribution of the (vectorized) autocovariance matrices is multivariate normal, Slutsky's theorem implies that the same is true also for the unmixing matrix estimates.

Corollary 1. *Under the assumptions (A1)-(A4), if the joint limiting distribution of*

$$\sqrt{T}\left[\mathrm{vec}(\hat{\boldsymbol{S}}_0, \hat{\boldsymbol{S}}_1, \ldots, \hat{\boldsymbol{S}}_K) - \mathrm{vec}(\boldsymbol{I}_p, \mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_K)\right]$$

is a (singular) $(K+1)p^2$*-variate normal distribution with mean value zero, then the joint limiting distribution of* $\sqrt{T}\mathrm{vec}(\hat{\mathbf{\Gamma}} - \mathbf{\Gamma})$ *is a singular* p^2*-variate normal distribution.*

In Section 5 we consider multivariate $MA(\infty)$ processes since their autocovariance matrices have limiting joint multivariate normal distribution. So far, we have assumed that the true value of $\mathbf{\Omega}$ is $\boldsymbol{I}_p$. Due to the affine equivariance of $\hat{\mathbf{\Gamma}}$, the limiting distribution of $\sqrt{T}\text{vec}(\hat{\mathbf{\Gamma}}\mathbf{\Omega} - \boldsymbol{I}_p)$ does not depend on $\mathbf{\Omega}$. If, for $\mathbf{\Omega} = \boldsymbol{I}_p$,

$$\sqrt{T}\text{vec}(\hat{\mathbf{\Gamma}} - \boldsymbol{I}_p) \to_d N_{p^2}(\mathbf{0}, \mathbf{\Sigma}),$$

then, for any full-rank true $\mathbf{\Omega}$, $\hat{\mathbf{\Gamma}} - \mathbf{\Gamma} = (\hat{\mathbf{\Gamma}}\mathbf{\Omega} - \boldsymbol{I}_p)\mathbf{\Gamma}$ and

$$\sqrt{T}\text{vec}(\hat{\mathbf{\Gamma}} - \mathbf{\Gamma}) \to_d N_{p^2}\left(0, (\mathbf{\Gamma}' \otimes \boldsymbol{I}_p)\mathbf{\Sigma}(\mathbf{\Gamma} \otimes \boldsymbol{I}_p)\right).$$

Moreover, for any true $\mathbf{\Omega}$ and $\hat{\mathbf{\Omega}} = \hat{\mathbf{\Gamma}}^{-1}$,

$$\mathbf{0} = \sqrt{T}\left(\hat{\mathbf{\Gamma}}\mathbf{\Omega}\mathbf{\Gamma}\hat{\mathbf{\Omega}} - \boldsymbol{I}_p\right) = \sqrt{T}\left(\hat{\mathbf{\Gamma}}\mathbf{\Omega} - \boldsymbol{I}_p\right) + \sqrt{T}(\mathbf{\Gamma}\hat{\mathbf{\Omega}} - \boldsymbol{I}_p) + o_P(1),$$

which implies that

$$\sqrt{T}\text{vec}(\hat{\mathbf{\Omega}} - \mathbf{\Omega}) \to_d N_{p^2}\left(0, (\boldsymbol{I}_p \otimes \mathbf{\Omega})\mathbf{\Sigma}(\boldsymbol{I}_p \otimes \mathbf{\Omega}')\right).$$

5 An example: MA(∞) processes

5.1 $MA(\infty)$ model

An example of multivariate time series having a limiting multivariate normal distribution is a MA(∞) process. From now on we assume that $\boldsymbol{z}_t$ are uncorrelated multivariate MA(∞) processes, that is,

$$\boldsymbol{z}_t = \sum_{j=-\infty}^{\infty} \Psi_j \boldsymbol{\epsilon}_{t-j}, \tag{5}$$

where $\boldsymbol{\epsilon}_t$ are standardized iid p-vectors and Ψ_j, $j = 0, \pm 1, \pm 2, \ldots$, are diagonal matrices satisfying $\sum_{j=-\infty}^{\infty} \Psi_j^2 = \boldsymbol{I}_p$. Hence

$$\boldsymbol{x}_t = \boldsymbol{\Omega} \boldsymbol{z}_t = \sum_{j=-\infty}^{\infty} (\boldsymbol{\Omega} \Psi_j) \boldsymbol{\epsilon}_{t-j} \tag{6}$$

is also a multivariate MA(∞) process. Notice that every second-order stationary process is either a linear process (MA(∞)) or can be transformed to a such one using Wold's decomposition. Notice also that causal ARMA(p, q) processes are MA(∞) processes (see for example Chapter 3 in (Brockwell and Davis, 1991)).

For our assumptions, we need the following notation and definitions. We say that a $p \times p$ matrix $\boldsymbol{J}$ is a sign-change matrix if it is a diagonal matrix with diagonal elements ± 1, and $\boldsymbol{P}$ is a $p \times p$ permutation matrix if it is obtained from an identity matrix by permuting its rows and/or columns. For the iid $\boldsymbol{\epsilon}_t$, we then assume that

(B1) $\boldsymbol{\epsilon}_t$ are iid with $E(\boldsymbol{\epsilon}_t) = 0$ and $Cov(\boldsymbol{\epsilon}_t) = \boldsymbol{I}_p$ and with finite fourth order moments, and

(B2) the components of $\boldsymbol{\epsilon}_t$ are exchangeable and marginally symmetric, that is, $\boldsymbol{J} \boldsymbol{P} \boldsymbol{\epsilon}_t \sim \boldsymbol{\epsilon}_t$ for all sign-change matrices $\boldsymbol{J}$ and for all permutation matrices $\boldsymbol{P}$.

Assumption (B1) implies that $E(\epsilon_{ti}^3 \epsilon_{tj}) = 0$ and that $E(\epsilon_{ti}^4) = \beta_{ii}$ and $E(\epsilon_{ti}^2 \epsilon_{tj}^2) = \beta_{ij}$ are bounded for all $i, j = 1, \ldots, p$. The above assumptions also imply that the model (6) satisfies assumptions (A1)-(A2).

5.2 Limiting distributions of the SOBI estimates

To obtain the limiting distributions of the (symmetrized) sample autocovariance matrices

$$\hat{\boldsymbol{S}}_k = \frac{1}{2(T-k)} \sum_{t=1}^{T-k} \left[\boldsymbol{x}_t \boldsymbol{x}_{t+k}' + \boldsymbol{x}_{t+k} \boldsymbol{x}_t' \right],$$

we define

$$\boldsymbol{F}_k = \sum_{t=-\infty}^{\infty} \boldsymbol{\psi}_t \boldsymbol{\psi}'_{t+k}, \quad k = 0, \pm 1, \pm 2, \dots$$

where $\boldsymbol{\psi}_t = (\psi_{t1}, \dots, \psi_{tp})'$ is the vector of the diagonal elements of $\boldsymbol{\Psi}_t$, $t = 0, \pm 1, \pm 2, \dots$. The diagonal elements of F_k are the autocovariances of the components of $\boldsymbol{z}$ at lag k. We also define the $p \times p$ matrices $\boldsymbol{D}_{lm}$, $l, m = 0, \dots, K$, with elements

$$\begin{aligned}
(\boldsymbol{D}_{lm})_{ii} &= (\beta_{ii} - 3)(\boldsymbol{F}_l)_{ii}(\boldsymbol{F}_m)_{ii} + \sum_{k=-\infty}^{\infty} ((\boldsymbol{F}_{k+l})_{ii}(\boldsymbol{F}_{k+m})_{ii} + (\boldsymbol{F}_{k+l})_{ii}(\boldsymbol{F}_{k-m})_{ii}), \\
(\boldsymbol{D}_{lm})_{ij} &= \frac{1}{2} \sum_{k=-\infty}^{\infty} ((\boldsymbol{F}_{k+l-m})_{ii}(\boldsymbol{F}_k)_{jj} + (\boldsymbol{F}_k)_{ii}(\boldsymbol{F}_{k+l-m})_{jj}) \\
&\qquad + (\beta_{ij} - 1)(\boldsymbol{F}_l + \boldsymbol{F}'_l)_{ij}(\boldsymbol{F}_m + \boldsymbol{F}'_m)_{ij}, \quad i \neq j.
\end{aligned}$$

The ijth element of $\boldsymbol{D}_{lm}$ is the limiting covariance of $(\hat{\boldsymbol{S}}_l)_{ij}$ and $(\hat{\boldsymbol{S}}_m)_{ij}$. The following lemma is proved in Miettinen et al. (2012).

Lemma 1. *Assume that $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_T)$ is a multivariate $MA(\infty)$ process defined in (5) that satisfies (B1) and (B2). Then the joint limiting distribution of*

$$\sqrt{T}(\mathrm{vec}(\hat{\boldsymbol{S}}_0, \hat{\boldsymbol{S}}_1, \dots, \hat{\boldsymbol{S}}_K) - \mathrm{vec}(\boldsymbol{I}_p, \boldsymbol{\Lambda}_1, \dots, \boldsymbol{\Lambda}_K))$$

is a singular $(K+1)p^2$-variate normal distribution with mean value zero and covariance matrix

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_{00} & \dots & \boldsymbol{V}_{0K} \\ \vdots & \ddots & \vdots \\ \boldsymbol{V}_{K0} & \dots & \boldsymbol{V}_{KK} \end{pmatrix},$$

with submatrices of the form

$$\boldsymbol{V}_{lm} = \mathrm{diag}(\mathrm{vec}(\boldsymbol{D}_{lm}))(\boldsymbol{K}_{p,p} - \boldsymbol{D}_{p,p} + \boldsymbol{I}_{p^2})$$

where

$$\boldsymbol{K}_{p,p} = \sum_i \sum_j (\boldsymbol{e}_i \boldsymbol{e}_j^T) \otimes (\boldsymbol{e}_j \boldsymbol{e}_i^T) \quad and \quad \boldsymbol{D}_{p,p} = \sum_i (\boldsymbol{e}_i \boldsymbol{e}_i^T) \otimes (\boldsymbol{e}_i \boldsymbol{e}_i^T).$$

Remark 1. *If we assume (B1) but replace (B2) by*

(B2) the components of $\boldsymbol{\epsilon}_t$ are mutually independent,*

then, in this independent component model case, the joint limiting distribution of $(\hat{\boldsymbol{S}}_0, \hat{\boldsymbol{S}}_1, \ldots, \hat{\boldsymbol{S}}_K)$ is again as given in Lemma 1 but with $\beta_{ij} = 1$ for $i \neq j$. If we further assume that innovations $\boldsymbol{\epsilon}_t$ are iid from $N_p(\boldsymbol{0}, \boldsymbol{I}_p)$, then $\beta_{ii} = 3$ and $\beta_{ij} = 1$ for all $i \neq j$, and the variances and covariances in Lemma 1 become even more simplified.

The first part of the next theorem was presented in Miettinen et al. (2014), the second part is new.

Theorem 2. *Assume that $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ is an observed time series from the $MA(\infty)$ process (6) that satisfies (B1), (B2) and (A3). Assume (wlog) that $\boldsymbol{\Omega} = \boldsymbol{I}_p$. If $\hat{\boldsymbol{\Gamma}} = (\hat{\boldsymbol{\gamma}}_1, \ldots, \hat{\boldsymbol{\gamma}}_p)'$ is the SOBI estimate, then the limiting distribution of $\sqrt{T}(\hat{\boldsymbol{\gamma}}_j - \boldsymbol{e}_j)$ is a p-variate normal distribution with mean zero and covariance matrix*

(i) (deflation-based case)

$$ASV(\hat{\boldsymbol{\gamma}}_j) = \sum_{r=1}^{j-1} ASV(\hat{\gamma}_{jr}) \boldsymbol{e}_r \boldsymbol{e}_r' + ASV(\hat{\gamma}_{jj}) \boldsymbol{e}_j \boldsymbol{e}_j' + \sum_{t=j+1}^{p} ASV(\hat{\gamma}_{jt}) \boldsymbol{e}_t \boldsymbol{e}_t',$$

where

$$
\begin{aligned}
ASV(\hat{\gamma}_{jj}) &= \frac{1}{4}(\boldsymbol{D}_{00})_{jj}, \\
ASV(\hat{\gamma}_{ji}) &= \frac{\sum_{l,m} \lambda_{li}\lambda_{mi}(\boldsymbol{D}_{lm})_{ji} - 2\mu_{ij}\sum_k \lambda_{ki}(\boldsymbol{D}_{k0})_{ji} + \mu_{ij}^2(\boldsymbol{D}_{00})_{ji}}{(\mu_{ij}-\mu_{ii})^2}, \\
&\qquad \text{for } i<j \\
ASV(\hat{\gamma}_{ji}) &= \frac{\sum_{l,m} \lambda_{lj}\lambda_{mj}(\boldsymbol{D}_{lm})_{ji} - 2\mu_{jj}\sum_k \lambda_{kj}(\boldsymbol{D}_{k0})_{ji} + \mu_{jj}^2(\boldsymbol{D}_{00})_{ji}}{(\mu_{jj}-\mu_{ji})^2}, \\
&\qquad \text{for } i>j
\end{aligned}
$$

with $\mu_{ij} = \sum_k \lambda_{ki}\lambda_{kj}$, *or*

(ii) (symmetric case)

$$ASV(\hat{\boldsymbol{\gamma}}_j) = \sum_{r=1}^{p} ASV(\hat{\gamma}_{jr})\boldsymbol{e}_r\boldsymbol{e}_r'$$

where, for $i \neq j$,

$$
\begin{aligned}
ASV(\hat{\gamma}_{jj}) =& \frac{1}{4}(\boldsymbol{D}_{00})_{jj}, \\
ASV(\hat{\gamma}_{ji}) =& \frac{\sum_{l,m}(\lambda_{lj}-\lambda_{li})(\lambda_{mj}-\lambda_{mi})(\boldsymbol{D}_{lm})_{ji}}{(\sum_k(\lambda_{kj}-\lambda_{ki})^2)^2} \\
&+ \frac{-2\nu_{ji}\sum_k(\lambda_{kj}-\lambda_{ki})(\boldsymbol{D}_{k0})_{ji} + \nu_{ji}^2(\boldsymbol{D}_{00})_{ji}}{(\sum_k(\lambda_{kj}-\lambda_{ki})^2)^2},
\end{aligned}
$$

with $\nu_{ji} = \sum_k(\lambda_{kj}^2 - \lambda_{kj}\lambda_{ki})$.

6 Efficiency comparisons

6.1 Performance indices

In this section we compare asymptotic and finite-sample efficiencies of the two SOBI estimates. The performance of the estimates in simulation studies can be measured using for example the minimum distance index (MDI) (Ilmonen et al., 2010)

$$\hat{D} = D(\hat{\mathbf{\Gamma}}\mathbf{\Omega}) = \frac{1}{\sqrt{p-1}} \inf_{\boldsymbol{C}\in\mathcal{C}} \|\boldsymbol{C}\hat{\mathbf{\Gamma}}\mathbf{\Omega} - \boldsymbol{I}_p\|$$

where $\|\cdot\|$ is the matrix (Frobenius) norm and

$$\mathcal{C} = \{\boldsymbol{C} \ : \ \text{each row and column of } \boldsymbol{C} \text{ has exactly one non-zero element.}\}$$

The minimum distance index is invariant with respect to the change of the mixing matrix, and it is scaled so that $0 \le \hat{D} \le 1$. It is also surprisingly easy to compute. The smaller the MDI-value, the better is the performance.

From the asymptotic point of view, the most attractive property of the minimum distance index is that for an estimate $\hat{\mathbf{\Gamma}}$ with $\sqrt{T}\,\mathrm{vec}(\hat{\mathbf{\Gamma}}\mathbf{\Omega} - \boldsymbol{I}_p) \to N_{p^2}(\mathbf{0}, \mathbf{\Sigma})$, the limiting distribution of $T(p-1)\hat{D}^2$ is that of a weighted sum of independent chi squared variables with the expected value

$$\mathrm{tr}\left((\boldsymbol{I}_{p^2} - \boldsymbol{D}_{p,p})\mathbf{\Sigma}(\boldsymbol{I}_{p^2} - \boldsymbol{D}_{p,p})\right).$$

Notice that $\mathrm{tr}((\boldsymbol{I}_{p^2} - \boldsymbol{D}_{p,p})\mathbf{\Sigma}(\boldsymbol{I}_{p^2} - \boldsymbol{D}_{p,p}))$ equals the sum of the limiting variances of the off-diagonal elements of $\sqrt{T}\,\mathrm{vec}(\hat{\mathbf{\Gamma}} - \boldsymbol{I}_p)$ and therefore provides a global measure of the variation of the estimate $\hat{\mathbf{\Gamma}}$.

If for example PCA, FOBI, AMUSE, deflation-based SOBI and symmetric SOBI are used to find the latent times series based on $\boldsymbol{x}$ given in Figure 2, the minimum distance index gets the values

$$0.914, \quad 0.628, \quad 0.246, \quad 0.062 \quad \text{and} \quad 0.058,$$

respectively. As PCA and FOBI are solely based on the 3-variable marginal distribution of the observations, they ignore time order and temporal dependence present in data. Of course, there is no reason why PCA should perform well here. Similarly, FOBI can be used for independent time series only if the latent series have distinct kurtosis values. The failure of these two methods in this example is clearly demonstrated by their high MDI values. AMUSE performs better here than PCA and FOBI, but is still much worse than symmetric SOBI. Clearly the first lag is not a good choice for the separation in this example.

Finally recall that, in the signal processing literature, several other indices have been proposed for the finite sample comparisons of the performance of the unmixing matrix estimates (for an overview see for example Nordhausen et al. (2011a)). One of the most popular performance indices, the Amari index (Amari et al., 1996), is defined as

$$\frac{1}{p}\left[\sum_{i=1}^{p}\frac{\sum_{j=1}^{p}|\hat{\boldsymbol{G}}_{ij}|}{\max_j|\hat{\boldsymbol{G}}_{ij}|}+\sum_{j=1}^{p}\frac{\sum_{i=1}^{p}|\hat{\boldsymbol{G}}_{ij}|}{\max_i|\hat{\boldsymbol{G}}_{ij}|}\right]-2,$$

where $\hat{\boldsymbol{G}}=\hat{\boldsymbol{\Gamma}}\boldsymbol{\Omega}$. The index is invariant under permutations and sign changes of the rows and columns of $\hat{\boldsymbol{G}}$. However, heterogeneous rescaling of the rows (or columns) on $\hat{\boldsymbol{G}}$ changes its value. Therefore, for the comparisons, the rows of $\hat{\boldsymbol{\Gamma}}$ should be rescaled in a similar way. We prefer MDI, since the Amari index is based on the L_1 norm and cannot be easily related to the limiting distribution of the unmixing matrix estimate.

6.2 Four models for the comparison

The following four models were chosen for the comparison of the deflation-based and symmetric SOBI estimates. The components of the source vectors are

(a) three MA(10)-series with coefficient vectors

$(0.8, 3.8, 1.2, 1.4, 1.1, 0.5, 0.7, 0.3, 0.5, 1.8)$,

$(-0.6, 1.3, -0.1, 1.3, 1.6, 0.4, 0.5, -0.4, 0.1, 2.8)$ and

$(-0.4, -1.5, 0, -1.1, -1.9, 0, -0.7, -0.4, -0.2, 0.4)$,

respectively, and normal innovations,

(b) three AR-series with coefficient vectors (0.6), $(0, 0.6)$ and $(0, 0, 0.6)$, respectively, and normal innovations,

(c) three ARMA-series with AR-coefficient vectors $(0.3, 0.3, -0.4)$, $(0.2, 0.1, -0.4)$, $(0.2, 0.2, 0.4)$, and MA-coefficient vectors $(-0.6, 0.3, 1.1, 1.0, -1.1, -0.3)$, $(1.2, 2.8, -1.0, -1.0, 0.1, 0.1)$, $(-1.4, -1.9, -0.5, -0.3, -0.4, 0.4)$, respectively, and normal innovations,

(d) three AR(1)-series with coefficients 0.6, 0.4 and 0.2, respectively, and normal innovations.

Each component is scaled to unit variance. Due to the affine equivariance of the estimates, it is not a restriction to use $\boldsymbol{\Omega} = \boldsymbol{I}_3$ in the comparisons.

6.3 Asymptotic efficiency

The asymptotic efficiency of the estimates can be compared using the sum of the limiting variances of the off-diagonal elements of $\sqrt{T}\,\mathrm{vec}(\hat{\boldsymbol{\Gamma}} - \boldsymbol{I}_p)$. In Table 1, these values are listed for the symmetric and deflation-based SOBI estimates, when both methods are using lags $k = 1, \ldots, 10$ in all four models.

Table 1: The sums of the estimated variances of selected rows of $\hat{\boldsymbol{\Gamma}}$ for the symmetric SOBI estimates utilizing four candidate sets of lags.

	model			
	(a)	(b)	(c)	(d)
deflation-based	46.5	31.8	11.0	61.6
symmetric	24.1	10.6	9.4	75.1

First notice that model (d) is the only one, where the deflation-based estimate outperforms the symmetric estimate. Also in model (c) the deflation-based method has quite competitive variances whereas in models (a) and (b) the symmetric estimate is much more accurate than the deflation-based estimate.

6.4 Finite sample behavior

For finite sample efficiencies, we use simulated time series from models (a)-(d). For each model and for each value of T, we have 10 000 repetitions of the simulated time series. The averages of $T(p-1)\hat{D}^2$ were then computed for the symmetric SOBI and deflation-based SOBI estimates. Again, we use lags $k = 1, \ldots, 10$ for both estimates. The results are plotted in Figure 5.

As explained in Section 6.1, the limiting distribution of $T(p-1)\hat{D}^2$ has expected values as given in Table 1. We then expect that the averages of $T(p-1)\hat{D}^2$ converge to these expected values in all four models. As seen in Figure 5, this seems to happen but the convergence is quite slow in all cases. Notice also that the finite sample efficiencies are higher than their approximations given by the limiting results.

Functions to compute deflation-based and symmetric SOBI estimates and their theoretic and finite sample efficiencies are provided in the R packages JADE (Nordhausen et al., 2013) and BSSasymp (Miettinen et al., 2013).

7 EEG example

Let us now illustrate how the theoretical results derived in Section 5.2 can be used to select a good set of autocovariance matrices to be used in SOBI method for a problem at hand. We consider EEG (electroencephalography) data recorded at the Department of Psychology, University of Jyväskylä, from an adult using 129 electrodes placed on the scalp and face. The EEG data was online bandpass filtered at 0.1 - 100 Hz and sampled at 500 Hz. The length of the data set equals $T = 100000$. An extract of five randomly selected EEG components is plotted in Figure 6.

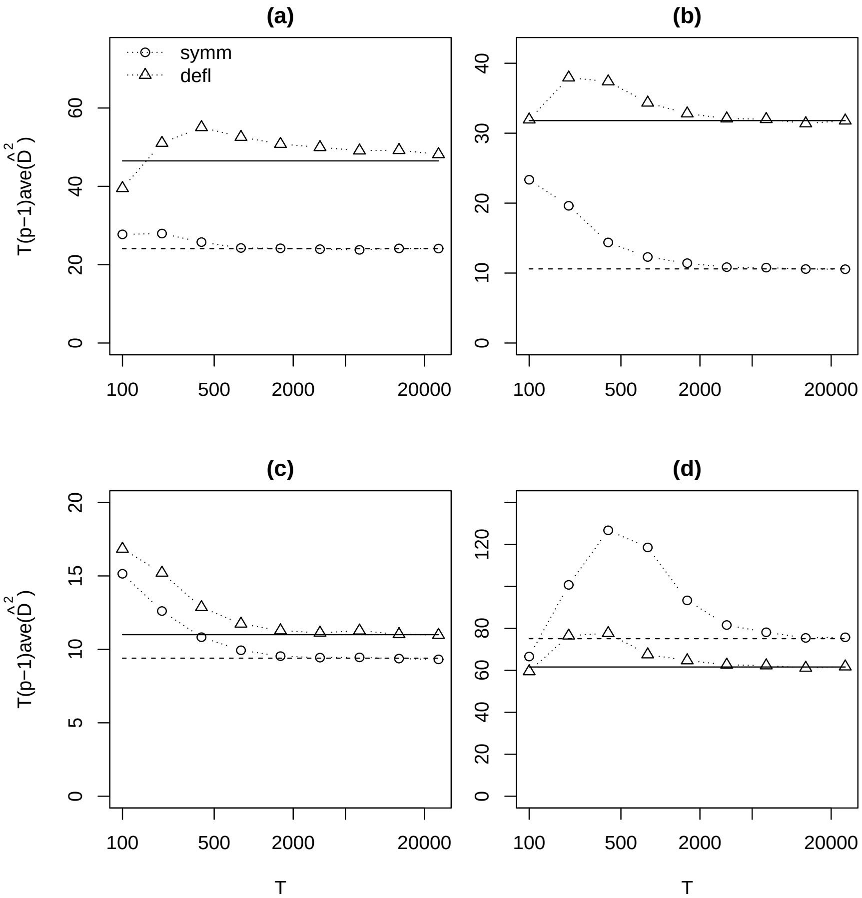

Figure 5: The averages of $T(p-1)\hat{D}^2$ for the symmetric and deflation-based SOBI estimates from 10 000 repetitions of observed time series with length T from models (a)-(d). The horizontal lines give the expected values of the limiting distributions of $T(p-1)\hat{D}^2$.

The goal in EEG data analysis is to measure the brain's electrical activity. As the measurements are made along the scalp and face, a mixture of unknown source signals is observed. Moreover, EEG recordings are often contaminated by non-cerebral artifacts, such as eye movements and muscular activity. Our aim is to use SOBI methods to separate few clear artifact components from EEG data. When choosing the set of autocovariance matrices to be diagonalized in SOBI, we follow the recommendations of Tang et al. (2005) and compare the results given by the following four sets of lags.

(1) $1, 2, \ldots, 10, 12, \ldots, 20, 25, \ldots, 100, 120, \ldots, 300$.

(2) $1, 2, \ldots, 10, 12, \ldots, 20$.

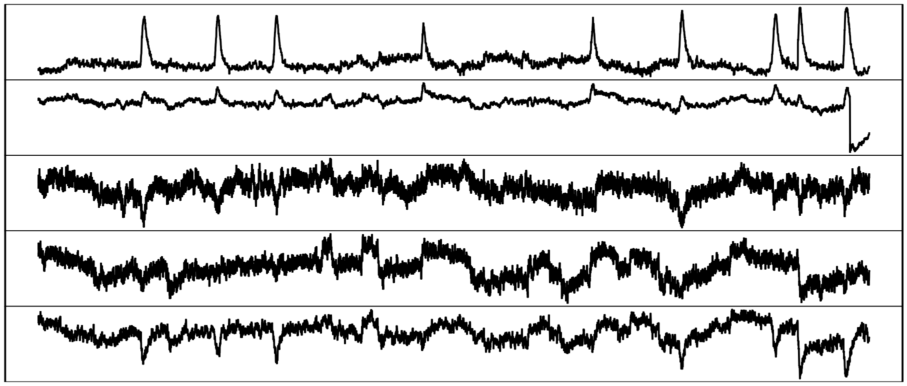

Figure 6: Five randomly selected EEG components of length $T = 10000$, from 20001 to 30000.

(3) $1, 2, \ldots, 10, 12, \ldots, 20, 25, \ldots, 100$.

(4) $25, 30, \ldots, 100, 120, \ldots, 300$.

We applied both, deflation-based and symmetric, SOBI methods to the EEG data using the four different autocovariance sets. However, as the deflation-based method gave significantly larger variance estimates for the unmixing matrix estimate, we only report the results based on symmetric SOBI. We chose three recognizable components, i.e. eye blink, horizontal eye movement and muscle activity, and estimated the sum of variances of the corresponding rows of the unmixing matrix under the assumption that the source signals are generated by $MA(\infty)$ processes with normal innovations. The resulting variance estimates are reported in Table 2.

Table 2: The sums of the estimated variances of selected rows of $\hat{\mathbf{\Gamma}}$ for the symmetric SOBI estimates utilizing four candidate sets of lags.

	set of lags			
	(1)	(2)	(3)	(4)
eye blink	0.052	0.021	0.060	0.091
horizontal eye movement	0.234	0.192	0.337	0.517
muscle activity	0.058	0.051	0.057	0.123

As the results of Table 2 indicate, the symmetric SOBI method that used autocovariance matrices with lags in set (2) gave best separation results for the three components of interest. These findings are confirmed by the time plots of separated components in Figure 7. The good performance of symmetric SOBI based on lags in set (2) is especially visible when looking at the corresponding eye blink and horizontal eye movement components.

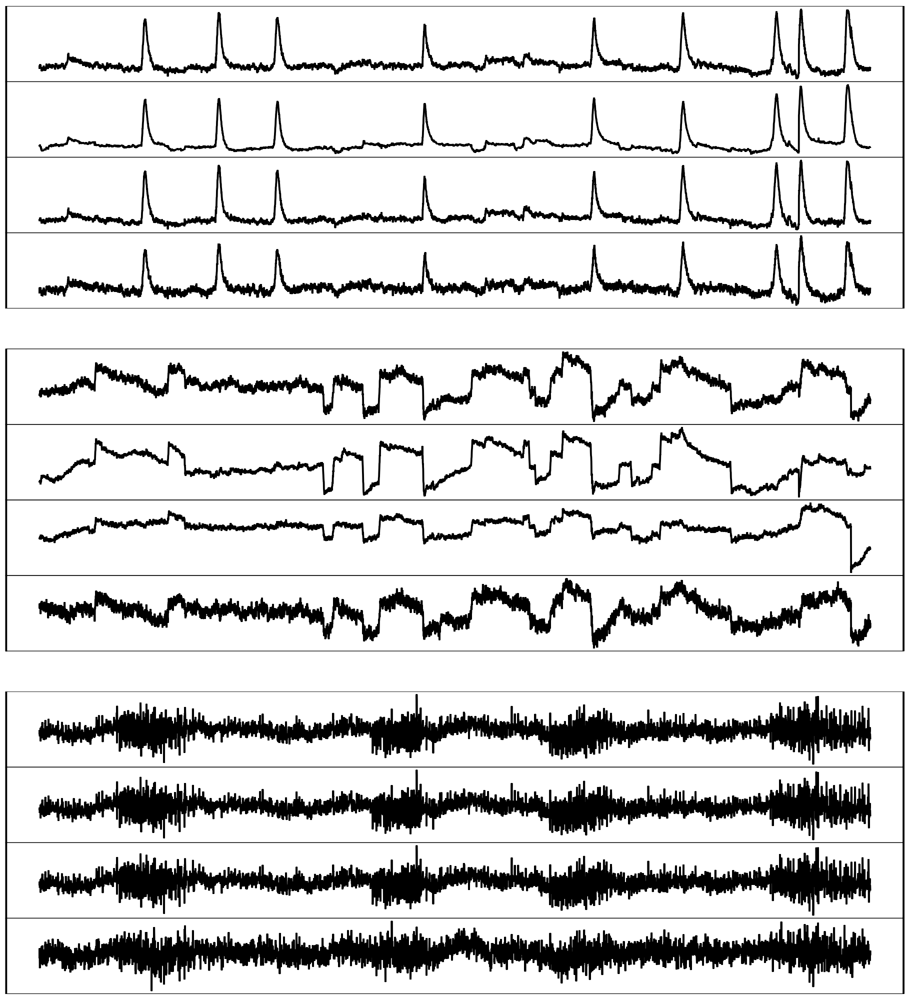

Figure 7: Estimated eye blink (top figure), horizontal eye movement (middle figure) and muscle activity (bottom figure) components of length $T = 10000$, from 20001 to 30000. The components are estimated using symmetric SOBI method with lags given in sets (1)-(4), respectively.

8 Concluding remarks

Symmetric SOBI is a popular blind source separation method but a careful analysis of its statistical properties has been missing so far. The theoretical results for deflation-based SOBI were presented only recently in Miettinen et al. (2014). There is a lot of empirical evidence that symmetric BSS methods perform better than their deflation-based competitors. In this paper we used the Lagrange multiplier technique to derive estimating equations for symmetric SOBI that allowed a thorough theoretic analysis of the properties of the estimate. In most cases we studied, the limiting efficiency of the symmetric SOBI estimate was better than that of the deflation-based estimate. The estimating equations also suggested a new algorithm for symmetric SOBI. Such an algorithm gave, in all our simulations, exactly the same results as the most popular algorithm based on Jacobi rotations. In a separate paper, these and other algorithms with associated estimates are compared in various settings with different values of p and K.

The problem corresponding to the selection of lags is still open; only few ad-hoc guidelines are available in the literature, see for example Tang et al. (2005). We followed these guidelines in our EEG data example. Notice however that the results presented in this paper can be used to build a two-stage estimation procedure where, at stage 2, the final SOBI estimate is selected among all SOBI estimates using their estimated efficiencies in a model determined by a preliminary SOBI estimate applied at stage 1. The results derived here can also be applied to different inference procedures, including hypothesis testing and model selection.

Acknowledgements

This research was supported by the Academy of Finland (grants 256291 and 268703). We thank Dr. Jarmo Hämäläinen for providing us with the EEG data.

References

Allassonniere, S. and Younes, L. (2012). A stochastic algorithm for probabilistic independent component analysis. *Annals of Applied Statistics* **6**, 125–160.

Amari, S. I., Cichocki, A., Yang H. H.(1996). A new learning algorithm for blind source separation. *Advances in Neural Information Processing Systems* **8**, 757–763

Belouchrani, A., Abed-Meraim, K., Cardoso, J.-F. & Moulines, E. (1997). A blind source separation technique using second-order statistics. IEEE Transactions on *Signal Processing* **45**, 434–444.

Bonhomme, S. and Robin, J.M. (2009). Consistent noisy independent component analysis. *Journal of Econometrics* **149**, 12–25.

Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods.* Second edition, Springer-Verlag, New York.

Chen, A. and Bickel, P. J. (2006). Efficient independent component analysis. *Ann. Statist.* **34**, 2825–2855.

Chen, Y., Härdle, W. and Spokoiny, V. (2007). Portfolio value at risk based on independent component analysis. *Journal of Computational and Applied Mathematics* **205**, 594–607.

Clarkson, D.B. (1988). A least squares version of algorithm AS 211: The F-G diagonalization algorithm. *Appl. Statist.* **37**, 317–321.

Comon P. and Jutten C. (2010). *Handbook of Blind Source Separation. Independent Component Analysis and Applications.* Academic Press, Amsterdam

GarciaFerrer A., Gonzalez-Prieto, E. and Pena, D. (2011). *Exploring ICA for time series decomposition.* Working paper 11-16, Statistics and Econometrics Series 11, Universidad Carlos III de Madrid.

GarciaFerrer A., Gonzalez-Prieto, E. and Pena, D. (2012). A conditionally heteroscedastic independent factor model with an application to finacial stock returns. *International Journal of Forecasting* **28**, 70–93.

Hallin, M. and Mehta, C. (2014). R-estimation for asymmetric independent component analysis. *Journal of the American Statistical Association*, DOI: 10.1080/01621459.2014.909316.

Hastie, T. and Tibshirani, R. (2003). Independent Component Analysis through Product Density Estimation. In Becker, S. and Obermayer, K. (editors) *Advances in Neural Information Processing Systems 15.* 649–656, MIT Press, Cambridge, MA.

Hyvärinen, A. Karhunen, J. and Oja, E. (2001). *Independent Component Analysis.* John Wiley & Sons, New York.

Ilmonen, P., Nevalainen, J. and Oja, H. (2010a). Characteristics of multivariate distributions and the invariant coordinate system. *Statistics and Probability Letters* **80**, 1844–1853.

Ilmonen, P., Nordhausen, K., Oja, H. and Ollila, E. (2010). A new performance index for ICA: properties computation and asymptotic analysis. In Vigneron, V., Zarzoso, V., Moreau, E., Gribonval, R. and Vincent, E. (editors) *Latent Variable Analysis and Signal Separation.* 229–236, Springer, Heidelberg.

Ilmonen, P. and Paindaveine, D. (2011). Semiparametrically efficient inference based on signed ranks in symmetric independent component models. *Ann. Statist.* **39**, 2448–2476.

Jutten, C. and Taleb, A. (2000). Source separation: From dusk till dawn. *Proceedings of the International Symposium on Independent Component Analysis and Blind Signal Separation*, 15–26.

Lee, S., Shen, H., Truong, Y., Lewis, M. and Huang, X. (2010). Independent component analysis involving autocorrelated sources with an application to functional magnetic resonance imaging. *J. Am. Statist. Ass.* **106**, 1009–1024.

Matteson, D. S. and Tsay, R. S. (2011). Dynamic orthogonal components for multivariate time series. *J. Am. Statist. Ass.* **106**, 1450–1463.

Miettinen, J., Nordhausen, K., Oja, H. and Taskinen, S. (2012). Statistical properties of a blind source separation estimator for stationary time series. *Statistics and Probability Letters* **82**, 1865–1873.

Miettinen, J., Nordhausen, K., Oja, H. and Taskinen, S. (2013). BSSasymp: Covariance matrices of some BSS mixing and unmixing matrix estimates. R package version 1.0-0. http://cran.r-project.org/web/packages/BSSasymp.

Miettinen, J., Nordhausen, K., Oja, H. and Taskinen, S. (2014). Deflation-based separation of uncorrelated stationary time series. *Journal of Multivariate Analysis* **123**, 214–227.

Nordhausen, K. (2014). On robustifying some second order blind source separation methods for nonstationary time series. *Statistical Papers* **55**, 141-156.

Nordhausen K., Cardoso J.F., Miettinen, J., Oja H., Ollila, E. and Taskinen, S. (2013). JADE: JADE and other BSS methods as well as some BSS performance criteria. R package version 1.9-91. http://cran.r-project.org/web/packages/JADE.

Nordhausen, K., Ilmonen, P., Mandal, A., Oja, H. and Ollila, E. (2011). Deflation-based FastICA Reloaded. In the Proceedings of *19th European Signal Processing Conference 2011 (EUSIPCO 2011)*, 1854–1858.

Nordhausen K., Ollila E., Oja H. (2011). On the performance indices of ICA and blind source separation. In the Proceedings of *2011 IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2011)*, 486–490.

Oja, H., Sirkiä, S. and Eriksson, J. (2006). Scatter matrices and independent component analysis. *Austrian Journal of Statistics* **35**, 175–189.

Poncela, P. (2012). Further research on independent component analysis. *International Journal of Forecasting* **28**, 94–96.

Samworth, R. J. and Yuan M. (2013). Independent component analysis via nonparametric maximum likelihood estimation. *Ann. Statist.* **40**, 2973–3002.

Schachtner, R., Lutter, D., Knollmueller, P., Tome, A., Theis, F., Schmitz, G., Stetter, M., Vilda, P. G. and Lang, E. (2008). Knowledge-based gene expression classification via matrix factorization. *Bioinformatics* **24**(15):16881697.

Tang, A.C., Liu, J.-Y. and Sutherland M.T. (2005). Recovery of correlated neuronal sources from EEG: The good and bad ways of using SOBI. *NeuroImage* **7**, 507–519.

Tong, L., Soon, V.C., Huang, Y.F. and Liu, R. (1990). AMUSE: a new blind identification algorithm. In Proceedings of IEEE International Symposium on *Circuits and Systems 1990*, 1784–1787.

Appendix

Proof of Theorem 1

(i) The proof for the consistency and limiting behavior of the deflation-based SOBI estimate can be found in Miettinen et al. (2014).

(ii) We first prove the consistency of the estimate. Let $\boldsymbol{U}$ be the compact set of all $p \times p$ orthogonal matrices. For $\boldsymbol{U} = (\boldsymbol{u}_1, \dots, \boldsymbol{u}_p)' \in \boldsymbol{U}$, write

$$\Delta(\boldsymbol{U}) = \sum_{i=1}^{p} \sum_{k=1}^{K} (\boldsymbol{u}_i' \boldsymbol{R}_k \boldsymbol{u}_i)^2 \quad \text{and} \quad \hat{\Delta}(\boldsymbol{U}) = \sum_{i=1}^{p} \sum_{k=1}^{K} (\boldsymbol{u}_i' \hat{\boldsymbol{R}}_k \boldsymbol{u}_i)^2$$

As $(\boldsymbol{u}' \hat{\boldsymbol{R}}_k \boldsymbol{u})^2 - (\boldsymbol{u}' \hat{R} \boldsymbol{u})^2 = (\boldsymbol{u}'(\hat{\boldsymbol{R}}_k - \boldsymbol{R}_k)\boldsymbol{u})(\boldsymbol{u}'(\hat{\boldsymbol{R}}_k + \boldsymbol{R}_k)\boldsymbol{u})$ and $\hat{\boldsymbol{R}}_k \to_P \boldsymbol{R}_k$, for $k = 1, \dots, K$,

$$M = \sup_{\boldsymbol{U} \in \boldsymbol{U}} |\Delta(\boldsymbol{U}) - \hat{\Delta}(\boldsymbol{U})| \leq 2 \sum_{k=1}^{K} ||\hat{\boldsymbol{R}}_k - \boldsymbol{R}_k|| \to_P 0.$$

Under our assumptions, $\boldsymbol{U} = \boldsymbol{I}_p$ is the unique maximizer of $\Delta(\boldsymbol{U})$ in the subspace

$$\boldsymbol{U}_0 = \{\boldsymbol{U} \in \boldsymbol{U} \ : \ \boldsymbol{u}_i \boldsymbol{R}_i \boldsymbol{u}_i \text{ descending and } \boldsymbol{u}_i' \boldsymbol{1}_p \geq 0, \, i = 1, \dots, p\}.$$

Notice that, in this subspace $\boldsymbol{U}$, the order and signs of the rows of $\boldsymbol{U}$ are fixed. For all $\epsilon > 0$, write next

$$\boldsymbol{U}_\epsilon = \{\boldsymbol{U} \in \boldsymbol{U}_0 \ : \ ||\boldsymbol{U} - \boldsymbol{I}_p|| \geq \epsilon\}$$

and

$$\delta_\epsilon = \Delta(\boldsymbol{I}_p) - \sup_{\boldsymbol{U} \in \boldsymbol{U}_\epsilon} \Delta(\boldsymbol{U}).$$

Clearly, $\delta_\epsilon > 0$ and $\delta_\epsilon \to 0$ as $\epsilon \to 0$. Let $\hat{\boldsymbol{U}}$ be the unique maximizer of $\hat{\Delta}(\boldsymbol{U})$ in $\boldsymbol{U}_0$.

Then

$$P\left(||\hat{\boldsymbol{U}} - \boldsymbol{I}_p|| < \epsilon\right) \geq P\left(\hat{\Delta}(\boldsymbol{I}_p) > \sup_{\boldsymbol{U}\in\boldsymbol{U}_\epsilon} \hat{\Delta}(\boldsymbol{U})\right) \geq P\left(M \leq \delta_\epsilon/3\right) \to 1, \quad \forall\epsilon$$

and the convergence $\hat{\boldsymbol{U}} \to_P \boldsymbol{I}_p$ follows. Thus $\hat{\boldsymbol{\Gamma}} = \hat{\boldsymbol{U}}\hat{\boldsymbol{S}}_0^{-1/2} \to_P \boldsymbol{I}_p$ also holds true.

To prove the second part of the result (ii), notice that the estimating equations give

$$0 = \sqrt{T}\hat{\boldsymbol{\gamma}}_i'\hat{\boldsymbol{S}}_0\hat{\boldsymbol{\gamma}}_j = \sqrt{T}\hat{\gamma}_{ij} + \sqrt{T}\hat{\gamma}_{ji} + \sqrt{T}(\hat{\boldsymbol{S}}_0)_{ij} + o_p(1), \tag{7}$$

for $i \neq j$, and

$$0 = \sqrt{T}(\hat{\boldsymbol{\gamma}}_i'\hat{\boldsymbol{S}}_0\hat{\boldsymbol{\gamma}}_i - 1) = 2\sqrt{T}(\hat{\gamma}_{ii} - 1) + \sqrt{T}((\hat{\boldsymbol{S}}_0)_{ii} - 1) + o_p(1). \tag{8}$$

Next note that

$$\hat{\boldsymbol{\gamma}}_i'\hat{\boldsymbol{T}}(\hat{\boldsymbol{\gamma}}_j) - \boldsymbol{e}_i'\boldsymbol{T}(\boldsymbol{e}_j) = \hat{\boldsymbol{\gamma}}_j'\hat{\boldsymbol{T}}(\hat{\boldsymbol{\gamma}}_i) - \boldsymbol{e}_j'\boldsymbol{T}(\boldsymbol{e}_i) \tag{9}$$

and

$$\begin{aligned}\sqrt{n}\left(\hat{\boldsymbol{\gamma}}_i'\hat{\boldsymbol{T}}(\hat{\boldsymbol{\gamma}}_j) - \boldsymbol{e}_i'\boldsymbol{T}(\boldsymbol{e}_j)\right) &= \sum_{k=1}^{K}\lambda_{kj}^2\sqrt{n}\hat{\gamma}_{ij} + \sum_{k=1}^{K}\lambda_{kj}\sqrt{n}(\hat{\boldsymbol{S}}_k)_{ij} \\ &\quad + \sum_{k=1}^{K}\lambda_{ki}\lambda_{kj}\sqrt{n}\hat{\gamma}_{ji} + o_p(1).\end{aligned} \tag{10}$$

for all $i \neq j$. The result then follows from equations (7)-(10).