Statistical properties of a blind source separation estimator for stationary time series

Jari Miettinen[a,*], Klaus Nordhausen[b], Hannu Oja[b], Sara Taskinen[a]

[a] *Department of Mathematics and Statistics, 40014 University of Jyväskylä, Finland*
[b] *School of Health Sciences, 33014 University of Tampere, Finland*

Abstract

In this paper, we assume that the observed p time series are linear combinations of p latent uncorrelated weakly stationary time series. The problem is then, using the observed p-variate time series, to find an estimate for a mixing or unmixing matrix for the combinations. The estimated uncorrelated time series may then have nice interpretations and can be used in a further analysis. The popular AMUSE algorithm finds an estimate of an unmixing matrix using covariances and autocovariances of the observed time series. In this paper, we derive the limiting distribution of the AMUSE estimator under general conditions, and show how the results can be used for the comparison of estimates. The exact formula for the limiting covariance matrix of the AMUSE estimate is given for general MA(∞) processes.

Keywords: AMUSE, asymptotic normality, autocovariance matrix, MA(∞) processes, minimum distance index

1. Introduction

Blind source separation (BSS) is a very general method which in its most basic form solves the following problem: Assume that the components of an observed p-variate vector $\boldsymbol{x}$ are linear combinations of the components of a latent unobserved p-variate source vector $\boldsymbol{z}$. The BSS model can then

*Corresponding author
Email addresses: `jari.p.miettinen@jyu.fi` (Jari Miettinen), `klaus.nordhausen@uta.fi` (Klaus Nordhausen), `hannu.oja@uta.fi` (Hannu Oja), `sara.l.taskinen@jyu.fi` (Sara Taskinen)

be written as $\boldsymbol{x} = \Omega \boldsymbol{z}$ with an unknown full rank $p \times p$ mixing matrix Ω, and the aim is, based on the observations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, to find an estimate of the mixing matrix Ω (or its inverse). In the independent component analysis (ICA), which is perhaps the most popular BSS approach, it is further assumed that the components of $\boldsymbol{z}$ are mutually independent and at most one of them is gaussian. The BSS approach has been widely used in biomedical signal analysis, brain imaging, and economic time series applications, for example.

For BSS applications, it is often assumed that the observation vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are independent and identically distributed (iid) observations from the distribution F_x. In reality, however, there is often temporal (e.g. time series) or spatial (e.g. image analysis) dependence between the observations, and one should naturally utilize this dependence in the analysis of data. In the time series context, the model may be written as

$$\boldsymbol{x}_t = \Omega \boldsymbol{z}_t, \quad t = 0, \pm 1, \pm 2, \ldots$$

where $\boldsymbol{z} = (\boldsymbol{z}_t)_{t=0,\pm 1,\pm 2,\ldots}$ is a p-variate time series satisfying some general assumptions, and the aim is again, based on observed p-variate time series $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T$, to estimate a full-rank $p \times p$ matrix Ω (or its inverse). In the signal processing community, an analysis tool called a second order source separation was developed for this problem. In this paper, we consider the second order source separation model and find the limiting properties of the estimate obtained from the popular AMUSE (Algorithm for Multiple Unknown Signals Extraction) (Tong et al., 1990, 1991) algorithm, one of the first solutions of the problem. The algorithm uses covariances and autocovariances with different lags $\tau = 1, 2, \ldots$ of the observed multivariate time series.

The structure of the paper is as follows. In Section 2, the BSS model and autocovariance matrix functionals with different lags τ are discussed and then used for the definition of the AMUSE unmixing matrix functional and estimator. The asymptotic distribution of the AMUSE estimator based on the covariance matrix and autocovariance matrix with lag τ is derived in Section 3. In Section 4, the asymptotic as well as finite sample behavior of the the AMUSE estimators is considered in some selected AR, MA, and ARMA models. The paper ends with a short discussion in Section 5.

2. Notation and definitions

2.1. Blind source separation model

We start with the definition of the blind source separation model.

Definition 1. *A p-variate time series* $\boldsymbol{x} = (\boldsymbol{x}_t)_{t=0,\pm1,\pm2,\ldots}$ *follows a blind source separation (BSS) model if*

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

where Ω *is a full-rank* $p \times p$ *mixing matrix and, for all* $t = 0, \pm1, \pm2, \ldots$*, the p-variate time series* $\boldsymbol{z}$ *satisfies*

(A1) $E(\boldsymbol{z}_t) = 0$,

(A2) $E(\boldsymbol{z}_t \boldsymbol{z}_t') = I_p$, *and*

(A3) $E(\boldsymbol{z}_t \boldsymbol{z}_{t+\tau}') = D_\tau$ *is diagonal for all* $\tau = 1, 2, \ldots$.

Note that the assumptions in Definition 1 imply the (second-order) weak stationarity and uncorrelatedness of the p time series in $\boldsymbol{z}$. Note also that Ω and $\boldsymbol{z}$ in the definition are confounded in the sense the signs and order of the components of $\boldsymbol{z}$ (and the signs and order of the columns of Ω, respectively) are not uniquely defined. Given an observed time series $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ from the BSS model (1), the aim is to find estimate $\hat{\Gamma}$ of an unmixing matrix Γ such that $\Gamma \boldsymbol{x}$ has uncorrelated components.

Remark 1. *The semiparametric BSS model is very flexible, and contains, for example, a multivariate MA(*∞*) process* $\Omega \boldsymbol{z}$ *where*

$$\boldsymbol{z}_t = \sum_{j=-\infty}^{\infty} \Psi_j \boldsymbol{\epsilon}_{t-j}, \tag{2}$$

and Ψ_j, $j = 0, \pm1, \pm2, \ldots$, *are diagonal matrices with diagonal elements* $\psi_{j1}, \ldots, \psi_{jp}$ *such that* $\sum_{j=-\infty}^{\infty} \Psi_j^2 = I_p$, *and* $\boldsymbol{\epsilon}_t$ *are p-variate iid random vectors with* $E(\boldsymbol{\epsilon}_t) = \boldsymbol{0}$ *and* $Cov(\boldsymbol{\epsilon}_t) = I_p$. *Hence*

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

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

Tong et al. (1990) proposed a simple two-step algorithm to obtain an unmixing matrix estimate $\hat{\Gamma}$. In their AMUSE algorithm, the p-variate observed time series is first standardized using the sample covariance matrix. Second, a sample autocovariance matrix with some delay τ is used to find the rotation that transforms standardized time series into uncorrelated time series. In the next section, we write the AMUSE algorithm in the form of a well-defined statistical functional.

2.2. Unmixing matrix functionals based on autocovariance matrices

Let $\boldsymbol{x}$ be a time series obeying the semiparametric BSS model (1) such that, for some $\tau > 0$, the diagonal elements of the autocovariance matrix $E(\boldsymbol{z}_t \boldsymbol{z}'_{t+\tau}) = D_\tau$ are distinct. Write

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

for the autocovariance matrices.

Definition 2. *Assume that the diagonal elements of D_τ, $\tau > 0$, are distinct. The unmixing matrix Γ_τ is then a $p \times p$ matrix that satisfies*

$$\Gamma_\tau S_0 \Gamma'_\tau = I_p \quad \textit{and} \quad \Gamma_\tau S_\tau \Gamma'_\tau = \Lambda_\tau,$$

where Λ_τ is a diagonal matrix with the diagonal elements in a decreasing order.

Note that S_0 is the regular covariance matrix. It is easy to see that the diagonal elements of Λ_τ are the diagonal elements of D_τ, but possibly reordered. The unmixing matrix Γ_τ based on S_0 and S_τ may be seen as a statistical functional satisfying that $\Gamma_\tau \boldsymbol{x}$ recovers $\boldsymbol{z}$ up to signs and order of the time series components.

The functional Γ_τ is affine equivariant in the sense that, if Γ_τ and Γ^*_τ are the values of the functionals at $\boldsymbol{x}$ and $\boldsymbol{x}^* = A\boldsymbol{x}$ in the BSS model (1) then $\Gamma_\tau \boldsymbol{x} = \Gamma^*_\tau \boldsymbol{x}^*$ (up to sign changes of the components). Note also that, for different τ's, the components of $\Gamma_\tau \boldsymbol{x}$ may be given in a different order.

The (population) autocovariance matrices S_τ can be estimated by the sample autocovariance matrices

$$\hat{S}_\tau = \frac{1}{T-\tau}\sum_{t=1}^{T-\tau} \boldsymbol{x}_t \boldsymbol{x}'_{t+\tau}, \quad \tau = 0, 1, 2, \ldots$$

Write also

$$\hat{S}^S_\tau = \frac{1}{2}(\hat{S}_\tau + \hat{S}'_\tau).$$

for a symmetrized version of the autocovariance matrix. (This is natural as as we are estimating the population quantity S_τ which is symmetrical.)

The unmixing matrix estimate corresponding to the functional Γ_τ is then given in the following.

Definition 3. *The unmixing matrix estimate $\hat{\Gamma}_\tau$ is then a $p \times p$ matrix that satisfies*

$$\hat{\Gamma}_\tau \hat{S}_0 \hat{\Gamma}'_\tau = I_p \quad \textit{and} \quad \hat{\Gamma}_\tau \hat{S}^s_\tau \hat{\Gamma}'_\tau = \hat{\Lambda}_\tau,$$

where $\hat{\Lambda}_\tau$ is a diagonal matrix with diagonal elements a in a decreasing order.

3. Limiting distributions

Similarly as in Ilmonen et al. (2010a), it is possible to prove the following theorem and corollary. We first derive the limiting distributions using the BSS model with $\Omega = I_p$. The limiting distributions in general case then follow from the affine equivariance of $\hat{\Gamma}$ and the affine invariance of $\hat{\Lambda}$.

Theorem 1. *Assume that $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ is an observed p-variate time series obeying the BSS model (1) with $\Omega = I_p$. Assume also that $\sqrt{T}(\hat{S}_0 - I_p) = O_p(1)$ and $\sqrt{T}(\hat{S}^S_\tau - \Lambda_\tau) = O_p(1)$, where Λ_τ is a diagonal matrix with diagonal elements $\lambda_1 > \cdots > \lambda_p$. Then, for a sequence of solutions $\hat{\Gamma}$ and $\hat{\Lambda}$,*

$$\begin{aligned}
\sqrt{T}\,\mathrm{diag}(\hat{\Gamma}_\tau - I_p) &= -\frac{1}{2}\sqrt{T}\,\mathrm{diag}(\hat{S}_0 - I_p) + o_p(1) \\
\sqrt{T}\,\mathrm{off}(\hat{\Gamma}_\tau) &= \sqrt{T}\,H \odot ((\hat{S}^S_\tau - \Lambda_\tau) - (\hat{S}_0 - I_p)\Lambda_\tau) + o_p(1) \\
\sqrt{T}\,(\hat{\Lambda}_\tau - \Lambda_\tau) &= \sqrt{T}\,\mathrm{diag}((\hat{S}^S_\tau - \Lambda_\tau) - (\hat{S}_0 - I_p)\Lambda_\tau) + o_p(1),
\end{aligned}$$

as $T \to \infty$, where $\mathrm{diag}(\Gamma)$ is a diagonal matrix with the same diagonal elements as Γ, $\mathrm{off}(\Gamma) = \Gamma - \mathrm{diag}(\Gamma)$, $\odot$ is the Hadamard (entrywise) product and H is a $p \times p$-matrix with elements

$$H_{ii} = 0 \quad \textit{and} \quad H_{ij} = (\lambda_i - \lambda_j)^{-1}, \ \textit{if } i \neq j.$$

The limiting distributions of $\hat{\Gamma}_\tau$ and $\hat{\Lambda}_\tau$ are thus given by the limiting distributions of $\hat{S}_0$ and $\hat{S}_\tau^S$ and, based on Theorem 1, one can write

$$\sqrt{T}\begin{pmatrix}\text{vec}(\hat{\Lambda}_\tau - \Lambda_\tau)\\ \text{vec}(\hat{\Gamma}_\tau - I_p)\end{pmatrix} = B\sqrt{T}\begin{pmatrix}\text{vec}(\hat{S}_0 - I_p)\\ \text{vec}(\hat{S}_\tau^S - \Lambda_\tau)\end{pmatrix} + o_p(1),$$

where

$$B = \begin{pmatrix} B_{11} & B_{12}\\ B_{21} & B_{22}\end{pmatrix} \tag{3}$$

with

$$\begin{aligned}
&B_{11} = -\text{diag}(\text{vec}(I_p))(\Lambda_\tau \otimes I_p), \ B_{12} = \text{diag}(\text{vec}(I_p)),\\
&B_{21} = -\frac{1}{2}\text{diag}(\text{vec}(I_p)) - \text{diag}(\text{vec}(H))(\Lambda_\tau \otimes I_p) \ \ \text{and}\\
&B_{22} = \text{diag}(\text{vec}(H)).
\end{aligned}$$

Note that, using the affine equivariance of $\hat{\Gamma}_\tau$ and the affine invariance of $\hat{\Lambda}_\tau$, the limiting distribution of these sample statistic are easily found for any Ω, see the next corollary.

Corollary 1. *Assume that $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ is an observed time series from the BSS model (1). Assume also that, for $\Omega = I_p$, the joint limiting distribution of*

$$\sqrt{T}\begin{pmatrix}\text{vec}(\hat{S}_0 - I_p)\\ \text{vec}(\hat{S}_\tau^s - \Lambda_\tau)\end{pmatrix}$$

is $N_{2p^2}(\mathbf{0}, V)$ where the diagonal elements of Λ_τ are in a decreasing order. Then, for any full-rank Ω, the limiting distribution of

$$\sqrt{T}\begin{pmatrix}\text{vec}\big(\hat{\Lambda}_\tau - \Lambda_\tau\big)\\ \text{vec}\big(\hat{\Gamma}_\tau - \Omega^{-1}\big)\end{pmatrix}$$

is a $2p^2$-variate normal distribution with zero mean vector and covariance matrix $ABVB'A'$, where B is as in (3), and

$$A = \begin{pmatrix} I_{p^2} & 0\\ 0 & (\Omega^{-1})' \otimes I_p\end{pmatrix}.$$

Assume next that $\boldsymbol{z}_t$ are uncorrelated multivariate MA(∞) processes defined in (2) that satisfy assumptions (A1)-(A3). First note that the autocovariance matrices of $\boldsymbol{z}$ are

$$S_\tau = E(\boldsymbol{z}_t \boldsymbol{z}_{t+\tau}') = \sum_{t=-\infty}^{\infty} \Psi_t \Psi_{t+\tau}, \quad \tau = 0, 1, 2, ...$$

The series is standardized as clearly $S_0 = I_p$. The limiting distributions of the sample autocovariance matrices $\hat{S}_\tau$ depend on the Ψ_t, $t = 0, \pm 1, \pm 2, ...$, through autocovariances of the parameter vector series,

$$F_\tau = \sum_{t=-\infty}^{\infty} \boldsymbol{\psi}_t \boldsymbol{\psi}_{t+\tau}', \quad \tau = 0, 1, 2, ...$$

where $\boldsymbol{\psi}_t = (\psi_{t1}, ..., \psi_{tp})'$ and $\psi_{t1}, ..., \psi_{tp}$ are the diagonal elements of Ψ_t, $t = 0, \pm 1, \pm 2, ...$

To derive the joint limiting distribution of $\hat{S}_0$ and $\hat{S}_\tau$ we also assume that

(A4) The components of $\boldsymbol{\epsilon}_t$ have finite fourth moments.

(A5) The components of $\boldsymbol{\epsilon}_t$ are exchangeable and marginally symmetric, that is,

$$JP\boldsymbol{\epsilon}_t \sim \boldsymbol{\epsilon}_t$$

for all sign-change matrices J and for all permutation matrices P.

Now the joint limiting normality of $\hat{\Gamma}_\tau$ and $\hat{\Lambda}_\tau$ follows from the next results.

Theorem 2. *Let $\boldsymbol{x}_t = \Omega \boldsymbol{z}_t$, where $\boldsymbol{z}$ is a multivariate MA(∞) process satisfying (A1)-(A5). Assume (w.l.o.g.) that $\Omega = I_p$ and $S_\tau^S = \Lambda_\tau$ with the distinct diagonal elements in a decreasing order. Also, write $E[\epsilon_{ti}^4] = \beta_{ii}$ and $E[\epsilon_{ti}^2 \epsilon_{tj}^2] = \beta_{ij}$. Then the joint limiting distribution of*

$$\sqrt{T} \begin{pmatrix} \mathrm{vec}(\hat{S}_0 - I_p) \\ \mathrm{vec}(\hat{S}_\tau^s - \Lambda_\tau) \end{pmatrix}$$

is $2p^2$-variate normal with mean value zero and covariance matrix

$$V = \begin{pmatrix} V_{00} & V_{0\tau} \\ V_{\tau 0} & V_{\tau\tau} \end{pmatrix}. \tag{4}$$

The submatrices of V are given by

$$V_{lm} = \mathrm{diag}(\mathrm{vec}(D_{lm}))(K_{p,p} - D_{p,p} + I_{p^2}),$$

where $K_{p,p} = \sum_i \sum_j (\boldsymbol{e}_i \boldsymbol{e}_j^T) \otimes (\boldsymbol{e}_j \boldsymbol{e}_i^T)$, $D_{p,p} = \sum_i (\boldsymbol{e}_i \boldsymbol{e}_i^T) \otimes (\boldsymbol{e}_i \boldsymbol{e}_i^T)$ *and* D_{lm} *is a* $p \times p$ *matrix with elements with*

$$(D_{lm})_{ii} = (\beta_{ii} - 3)(F_l)_{ii}(F_m)_{ii} + \sum_{k=-\infty}^{\infty} \left((F_{k+l})_{ii}(F_{k+m})_{ii} + (F_{k+l})_{ii}(F_{k-m})_{ii}\right),$$

$$\begin{aligned}(D_{lm})_{ij} &= \frac{1}{2} \sum_{k=-\infty}^{\infty} \left((F_{k+l-m})_{ii}(F_k)_{jj} + (F_k)_{ii}(F_{k+l-m})_{jj}\right) \\ &\quad + \frac{1}{4}(\beta_{ij} - 1)(F_l + F_l')_{ij}(F_m + F_m')_{ij}, \quad i \neq j.\end{aligned}$$

Remark 2. *First note that, if ϵ_t are iid from $N_p(\mathbf{0}, I_p)$ then $\beta_{ii} = 3$ and $\beta_{ij} = 1$ for all $i \neq j$, and the variances and covariance in Theorem 2 become much simpler. If we assume (A4) but replace (A5) by*

(A6) The components of $\boldsymbol{\epsilon}_t$ are mutually independent.

then, in this independent component model case, the joint limiting distribution of $\hat{S}_0$ and $\hat{S}_\tau^S$ is still as in Theorem 2 with $\beta_{ij} = 1$ for $i \neq j$.

Remark 3. *The joint limiting distribution of two autocovariance matrices in non-Gaussian case is also derived in Su and Lund (2012). The limiting distribution of symmetrized autocovariance matrices follows by noticing that*

$$\mathrm{vec}(\hat{S}_\tau^s) = \frac{1}{2}(I_p^2 + K_{p,p})\,\mathrm{vec}(\hat{S}_\tau).$$

4. The behavior of the estimates $\hat{\boldsymbol{\Gamma}}_{\boldsymbol{\tau}}$ in ARMA models

4.1. Minimum distance index

In simulation studies the mixing matrix Ω is naturally known. For any reasonable unmixing matrix estimate $\hat{\Gamma}$, the so called gain matrix $\hat{G} = \hat{\Gamma}\Omega$ converges to some C in

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

The minimum distance index (Ilmonen et al., 2010b) for the comparison of the estimates is then defined as

$$\hat{D} = D(\hat{\Gamma}\Omega) = \frac{1}{\sqrt{p-1}} \inf_{C \in \mathcal{C}} \|C\hat{\Gamma}\Omega - I_p\|,$$

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

The minimum distance index is affine invariant and has the following nice properties, see (Ilmonen et al., 2010b):

(i) $0 \leq \hat{D} \leq 1$,

(ii) $\hat{D} = 0$ if and only if $C\hat{\Gamma}\Omega = I_p$ for some $C \in \mathcal{C}$,

(iii) $\hat{D} = 1$ if and only if $C\hat{\Gamma}\Omega = \mathbf{1}_p \boldsymbol{a}'$ for some p-vector $\boldsymbol{a}$ and some $C \in \mathcal{C}$, and

(iv) if $\Omega = I_p$ and $\sqrt{T}\,\mathrm{vec}(\hat{\Gamma} - I_p) \to N_{p^2}(0, \Sigma)$, then

$$T\hat{D}^2 = \frac{T}{p-1}||\mathrm{off}(\hat{\Gamma})||^2 + o_p(1),$$

and the limiting distribution of $T\hat{D}^2$ is that of weighted sum of independent chi squared variables. The limiting distribution has the expected value

$$\frac{1}{p-1}\mathrm{tr}\left((I_{p^2} - D_{p,p})\Sigma(I_{p^2} - D_{p,p})\right). \quad (5)$$

Notice that $\mathrm{tr}\left((I_{p^2} - D_{p,p})\Sigma(I_{p^2} - D_{p,p})\right)$ is the sum of the limiting variances of the off-diagonal elements of $\sqrt{T}\,\mathrm{vec}(\hat{\Gamma} - I_p)$ and therefore provides a global measure of the variation of the estimate $\hat{\Gamma}$. Note also that, for different estimates$\hat{\Gamma}_\tau$, the limiting distribution of the diagonal elements is the same. See Theorem 1.

4.2. The efficiency results for some ARMA models

To compare unmixing matrix estimates $\hat{\Gamma}_\tau$ with different lags τ, we calculated the limiting values of $T(p-1)E(\hat{D}^2)$ for different lags τ and for different multivariate AR, MA and ARMA models. We also performed a small simulation study in the same cases to see how fast is the convergence to these limiting values.

We considered the case where the uncorrelated time series were given by

(i) three independent AR(1) processes with parameters 0.55, 0.35 and 0.15, respectively, and normally distributed innovations (AR(1)-N),

(ii) three AR(1) processes as in (i) but with spherical three-variate t_5-distributed innovations (AR(1)-t_5),

(iii) three independent AR(2) processes with parameters $(0.25, 0.6)$, $(0.15, 0.4)$ and $(0.1, 0.2)$, respectively, and normally distributed innovations (AR(2)-N),

(iv) three AR(2) processes as in (iii) but with spherical three-variate t_5-distributed innovations (AR(2)-t_5),

(v) three independent causal MA(25) processes with normal innovations,

(vi) three independent ARMA(3,3) processes with normal innovations,

(vii) four independent AR(10) processes, where the innovations were χ^2_4, exponentially, normally, and normally distributed, respectively.

Notice that in (ii) and (iv) the time series are not independent, only uncorrelated. The innovations were standardised to have mean equal to zero and variance equal to one. The number of parameters in (v)-(vii) were 75, 18, and 30, respectively, and the values of the parameters are therefore not reported here. They can be obtained from the first author upon request. The mixing matrix in all simulation settings was $\Omega = I_p$. Notice that due to affine invariance, the performance of the methods do not depend on the choice of Ω.

The limiting values of $T(p-1)E(\hat{D}^2)$, that is, global measures of the efficiencies of the estimates, for different setups were computed using (5). The results for lags $\tau = 1, \ldots, 10$ are listed in Table 1. As an example, consider simulation setting (i) and case $\tau = 1$. To obtain (5), only the limiting variances of the elements of $\sqrt{T}\,\mathrm{vec}(\hat{\Gamma}_1 - I_p)$ need to be computed. They were computed numerically using the results given in Corollary 1 and Theorem 2. The limiting variances can be collected into a matrix

$$\Sigma_1 = \begin{pmatrix} 0.93 & 12.46 & 3.79 \\ 7.96 & 0.64 & 13.12 \\ 2.04 & 10.62 & 0.52 \end{pmatrix},$$

where $(\Sigma_1)_{ij}$ is the limiting variance of $\sqrt{T}(\hat{\Gamma}_1 - I_p)_{ij}$, $i, j = 1, 2, 3$. The sum of the limiting variances of the off-diagonal elements now equals

$$\mathrm{tr}\left(ASCOV(\sqrt{T}\mathrm{vec}(\mathrm{off}(\hat{\Gamma}_1)))\right) = 49.99.$$

Other values can be computed in a similar way.

The values for $\hat{\Gamma}_\tau$ naturally depend strongly on the model as well as on the chosen parameters in the model. The results in Table 1 clearly indicate that there are major differences between the expected performances for different lags. One can see, for example, that using $\tau = 1$ and $\tau = 2$ in considered AR(1) and AR(2) models, respectively, gives the smallest limiting variances for the estimate. Also, in ARMA$(3,3)$ model $\tau = 3$ seems to be the best choice. For models with large number of parameters, it seems difficult to give a general rule for the choice of τ.

Table 1: The limiting values of $T(p-1)E(\hat{D}^2)$ for for $\hat{\Gamma}_\tau$ for $\tau = 1, \ldots, 10$, and for the models (i)-(vii).

τ	AR(1)-N	AR(1)-t_5	AR(2)-N	AR(2)-t_5	MA(25)	ARMA(3,3)	AR(10)
1	50.0	52.7	136.3	138.7	1 723	48.5	252.2
2	163.4	166.1	32.8	35.0	123.3	32.7	675.5
3	848.1	850.8	117.4	119.7	7 635	63.6	55.3
4	5 674	5 677	70.3	72.5	53.1	48.9	432.0
5	42 649	42 652	215.8	218.1	334.5	543.6	102.7
6	$3.4 \cdot 10^5$	$3.4 \cdot 10^5$	206.7	208.9	32.5	4250	4179
7	$2.7 \cdot 10^6$	$2.7 \cdot 10^6$	572.2	574.4	57.6	1139	562.8
8	$2.2 \cdot 10^7$	$2.2 \cdot 10^7$	725.2	727.5	567.1	19 325	71.0
9	$1.8 \cdot 10^8$	$1.8 \cdot 10^8$	1 824	1 826	2 520	51 742	180.5
10	$1.5 \cdot 10^9$	$1.5 \cdot 10^9$	2 748	2 750	316.0	$1.0 \cdot 10^6$	109.4

We next illustrate the finite sample behavior of unmixing matrix estimates $\hat{\Gamma}_\tau$ using simulations. The programs for computing AMUSE estimates as well as minimum distance index are available in the R-package JADE (Nordhausen et al., 2012). The simulations were carried out by generating 10 000 repetitions of observed time series with selected lengths T. The comparisons in the considered cases were made through average values of $T(p-1)\hat{D}^2$ over 10 000 repetitions. In Figure 1, the finite sample behavior of $\hat{\Gamma}_1$, $\hat{\Gamma}_2$ and $\hat{\Gamma}_3$ is illustrated using models (i) and (ii). Figure 1 shows that $\hat{\Gamma}_1$ and $\hat{\Gamma}_2$ are much more efficient than $\hat{\Gamma}_3$ when the observations come from the AR(1)-N model, and the convergence of the average to $T(p-1)E(\hat{D}^2)$ seem slowest for $\hat{\Gamma}_3$. One also sees that the estimate $\hat{\Gamma}_1$ is better in model (i) than in model (ii). Notice that the second moments are the same in these two models; the difference is that innovations in model (ii) have heavier tails and are dependent.

In Figure 2, the performance of different unmixing matrix estimates $\hat{\Gamma}_\tau$ is illustrated using models (iv)-(vii). It is again seen that there are major differences between the performances for different lags, and that the convergence is slower the higher the value $T(p-1)E(\hat{D}^2)$ is.

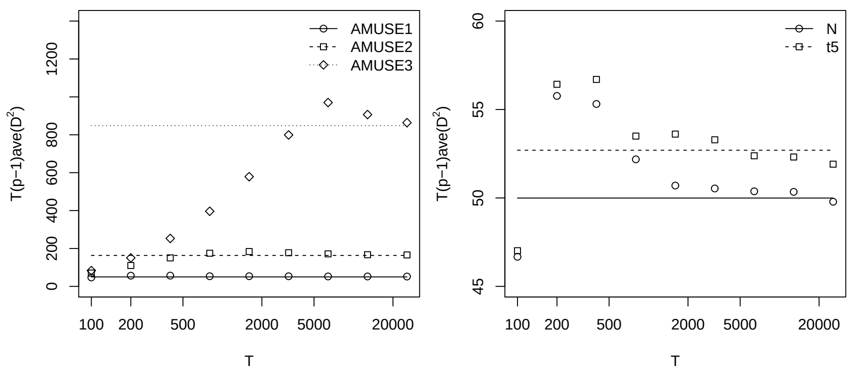

Figure 1: The averages of $T(p-1)\hat{D}^2$ for $\hat{\Gamma}_1$, $\hat{\Gamma}_2$, and $\hat{\Gamma}_3$ from 10 000 repetitions of observed time series with length T from models (i) (AR(1)-N) and (ii) (AR(1)-t_5. In the left panel $\hat{\Gamma}_1$, $\hat{\Gamma}_2$, and $\hat{\Gamma}_3$ are compared in model (i). In the right panel, the behavior of $\hat{\Gamma}_1$ is illustrated in models (i) and (ii). The horizontal lines give the expected values of the limiting distributions of $T(p-1)\hat{D}^2$.

5. Final remarks

The AMUSE procedure was originally presented as a method which jointly diagonalizes the covariance matrix and an autocovariance matrix with a chosen lag τ. The solution can, however, be found as a solution in a principal components analysis (PCA) problem (Cichocki & Amari, 2002) or as a solution in a canonical correlation analysis (CCA) problem (Liu et al., 2007) as follows. The PCA solution is found if one first finds standardized p-variate times series $\boldsymbol{y}_t = Cov(\boldsymbol{x}_t)^{-1/2}\boldsymbol{x}_t$, which is then rotated using the matrix of

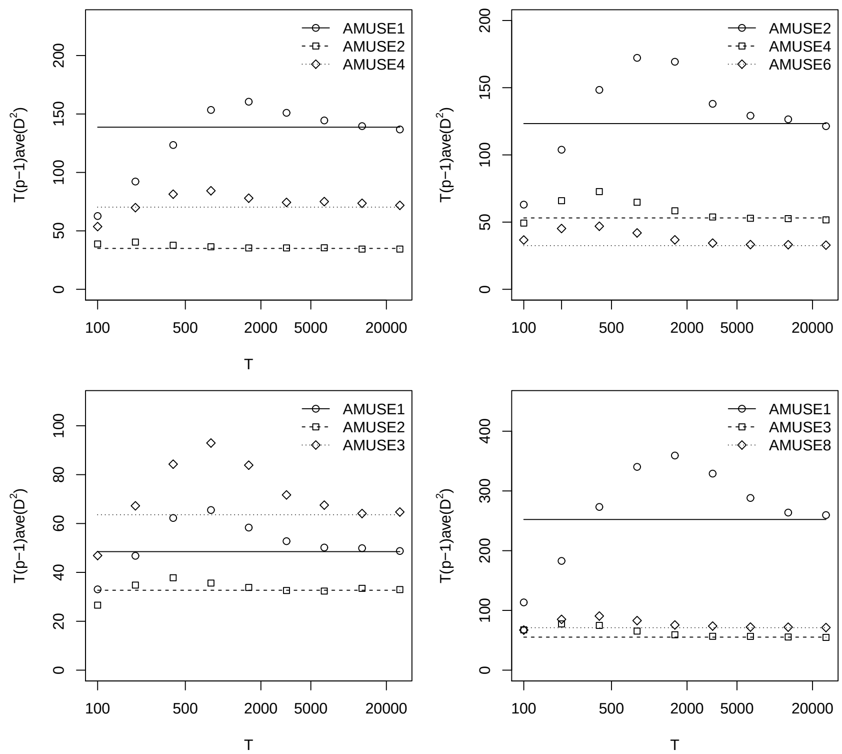

Figure 2: The averages of $T(p-1)\hat{D}^2$ for $\hat{\Gamma}_\tau$ with different values of τ from 10 000 repetitions of observed time series with length T. In the top left panel $\hat{\Gamma}_1$, $\hat{\Gamma}_2$, and $\hat{\Gamma}_4$ are compared in model (iv). In the top right panel $\hat{\Gamma}_2$, $\hat{\Gamma}_4$, and $\hat{\Gamma}_6$ are compared in model (v). In the bottom left panel $\hat{\Gamma}_1$, $\hat{\Gamma}_2$, and $\hat{\Gamma}_3$ are compared in model (vi). In the bottom right panel $\hat{\Gamma}_1$, $\hat{\Gamma}_3$, and $\hat{\Gamma}_8$ are compared in model (vii). The horizontal lines give the expected values of the limiting distributions of $T(p-1)\hat{D}^2$.

eigenvectors of $Cov(\boldsymbol{y}_t + \boldsymbol{y}_{t+\tau})$. The CCA solution finds the canonical vectors for

$$Cov \begin{pmatrix} \boldsymbol{x}_t \\ \boldsymbol{x}_{t+\tau} \end{pmatrix} = \begin{pmatrix} S_0 & S_\tau \\ S'_\tau & S_0 \end{pmatrix}$$

In the paper, we showed that the efficiency of the estimate $\hat{\Gamma}_\tau$ depends strongly on the model as well as on the unknown parameters in the model. It seems that in practice there is no good rule for the choice of τ without using any preliminary or additional information on the underlying process. Cichocki & Amari (2002), for example, simply recommend to start with the choice $\tau = 1$, and then try another value for τ if $\hat{\Lambda}_\tau$ has diagonal elements that are too close together.

Without any knowledge about good choices of τ, another approach is to jointly diagonalize several autocovariance matrices $S_0, S_{\tau_1}, ..., S_{\tau_k}$ as is done by the popular SOBI (Second Order Blind Identification) algorithm (Belouchrani et al., 1997). In simulations and practical applications, SOBI seems to perform better than AMUSE but no (asymptotic) efficiency results are available so far. It would be interesting also to robustify the AMUSE and SOBI procedures by replacing the regular covariance and autocovariance matrices by some robust functionals. Robust AMUSE and SOBI procedures has been proposed by Theis et al. (2010) but, unfortunately, these procedures lack the affine equivariance property.

Acknowledgements

The authors wish to thank the referee for comments on this paper. This research was supported by the Academy of Finland.

References

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.

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

Cichocki, A., Amari, S., 2002. Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications. Wiley , Cichester.

Ilmonen, P., Nevalainen, J., 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., Ollila, E., 2010b. A new performance index for ICA: properties computation and asymptotic analysis. In Vigneron, V., Zarzoso, V., Moreau, E., Gribonval, R. and Vincent, E. (Eds.), Latent Variable Analysis and Signal Separation. 229–236, Springer, Heidelberg.

Liu, W, Mandic, D.P., Cichocki, A., 2007. Analysis and online realization of the CCA approach for blind source separation. IEEE Transactions on Neural Networks. 18, 1505–1510.

Nordhausen, K., Cardoso, J-F., Miettinen, J., Oja, H., Ollila, E., Taskinen, S., 2012. JADE: JADE and other BSS methods as well as some BSS performance criteria. R package version 1.1.

Su, N. and Lund, R., 2012. Multivariate versions of Bartlett's formula. Journal of Multivariate Analysis. 105, 18–31.

Theis, F.J., Müller, N.S., Plant, C., Böhm, C., 2010. Robust second-order source separation identifies experimental responses in biomedical imaging. In Vigneron, V., Zarzoso, V., Moreau, E., Gribonval, R. and Vincent, E. (Eds.), Latent Variable Analysis and Signal Separation. 466–473, Springer, Heidelberg.

Tong, L., Liu, R., Soon, V.C., Huang, Y.F., 1991. Indeterminacy and identifiability of blind identification. IEEE Transactions on Circuits and Systems. 38, 499–509.

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

Appendix: Proofs of the results

Proof of Theorem 1. The proof is similar to the proof of Theorem 3.1 in Ilmonen et al. (2010).

To prove Theorem 2, we need the following lemma.

Lemma 1. *Assume that $\boldsymbol{x}_t$ is a multivariate linear process satisfying assumptions (A1)-(A5), $S_\tau^S = \Lambda_\tau$, where Λ_τ is a diagonal matrix with diagonal elements $(\Lambda_\tau)_{ii} = E[x_{ti}x_{t+\tau,i}] = (F_\tau)_{ii}$, where $F_\tau = \sum_{t=-\infty}^{\infty} \boldsymbol{\psi}_t \boldsymbol{\psi}'_{t+\tau}$. Denote*

$$\hat{S}_\tau = ave\{\boldsymbol{x}_t \boldsymbol{x}_{t+\tau}\} =: \begin{pmatrix} (\hat{S}_\tau)_{11} & \dots & (\hat{S}_\tau)_{1p} \\ \vdots & \ddots & \vdots \\ (\hat{S}_\tau)_{p1} & \dots & (\hat{S}_\tau)_{pp} \end{pmatrix}.$$

Then the elements of symmetrised matrix $\hat{S}_\tau^S = (\hat{S}_\tau + \hat{S}'_\tau)/2$ are given by

$$(\hat{S}_\tau^S)_{ij} = \begin{cases} T^{-1}\sum_t x_{ti}x_{t+\tau,i}, & if \;\; i = j \\ T^{-1}\sum_t (x_{t+\tau,i}x_{tj} + x_{ti}x_{t+\tau,j})/2, & if \;\; i \neq j. \end{cases} \tag{6}$$

Write now ASCOV for the asymptotic covariance. Then for any $i \neq j$,

$$\begin{aligned} &ASCOV(\sqrt{T}(\hat{S}_l^S)_{ii}, \sqrt{T}(\hat{S}_m^S)_{ii}) \\ &= (\beta_{ii} - 3)(F_l)_{ii}(F_m)_{ii} + \sum_{k=-\infty}^{\infty} \big[(F_{k+l})_{ii}(F_{k+m})_{ii} + (F_{k+l})_{ii}(F_{k-m})_{ii}\big]. \end{aligned}$$

and

$$\begin{aligned} &ASCOV(\sqrt{T}(\hat{S}_l^S)_{ij}, \sqrt{T}(\hat{S}_m^S)_{ij}) \\ &= 2^{-1} \sum_{k=-\infty}^{\infty} \big[(F_{k+l})_{ii}(F_{k+m})_{jj} + (F_{k+l})_{ii}(F_{k-m})_{jj}\big] \\ &\quad + (\beta_{ij} - 1)(F_l + F'_l)_{ij}(F_m + F'_m)_{ij}. \end{aligned}$$

Proof of Lemma 1. The first expression is proved in Propositions 7.3.1 and 7.3.4 of Brockwell & Davis (1991), and the second one can be proved using similar technique.

Proof of Theorem 2. Assume that the components of $\boldsymbol{x}_t$ are sign-symmetric linear zero mean processes such that $E[\epsilon_{ti}^2\epsilon_{tj}^2] = \beta_{ij}$, $S_0 = I_p$ and $S_\tau^S = \Lambda_\tau$, where Λ is a diagonal matrix with diagonal elements $(\Lambda_\tau)_{ii} = (F_\tau)_{ii} = E[x_{t+\tau,i}x_{ti}]$. One can easily show that the asymptotic covariance matrix of $\sqrt{T}(\text{vec}(\hat{S}_0 - I_p)', \text{vec}(\hat{S}_\tau^S - \Lambda_\tau)')$ is given by (4), where the non-zero elements can be expressed using covariances computed in Lemma 1.

To prove the asymptotic normality, we proceed as in section 7.3 of Brockwell & Davis (1991), that is, we first prove the normality using m-dependent moving average processes $\boldsymbol{x}_t^m = \sum_{j=-m}^{m} \Psi_j \boldsymbol{\epsilon}_{t-j}$ and then extend the result to case $m \to \infty$.

Write now $\boldsymbol{y}_t^m = (vec(\boldsymbol{x}_t^m(\boldsymbol{x}_t^m)'), vec(\boldsymbol{x}_t^m(\boldsymbol{x}_{t+\tau}^m)' + \boldsymbol{x}_{t+\tau}^m(\boldsymbol{x}_t^m)'))$. Then $\boldsymbol{y}_t^m$ is a strictly stationary $(2m+\tau)$ -dependent sequence such that

$$T^{-1} \sum_{t=1}^{T} \boldsymbol{y}_t^m = ((\hat{S}_0^m)_{11}, (\hat{S}_0^m)_{12}, \ldots, (\hat{S}_0^m)_{pp}, (\hat{S}_\tau^{S,m})_{11}, \ldots, (\hat{S}_\tau^{S,m})_{pp})',$$

where $(\hat{S}_\tau^{S,m})_{ij}$ are obtained from (6) replacing x_{ti} with x_{ti}^m. Then proceeding as in the proof of Proposition 7.3.2 in Brockwell & Davis (1991), one can show that $T^{-1}\sum_{t=1}^{T} \boldsymbol{y}_t^m$ has a limiting normal distribution with mean $(\text{vec}(I_p)', \text{vec}(\Lambda_\tau^m)')$ and covariance V^m/T, where Λ_τ^m and V^m are the corresponding matrices to Λ_τ and V, where each $(F_\tau)_{ii}$ is replaced with $(F_\tau^m)_{ii}$.

The result can be extended to MA(∞) processes by following the steps in the proof of Proposition 7.3.3 in Brockwell & Davis (1991).